INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2010-47

Assessing the Effects of Local Taxation Using Microgeographic Data

G. DURANTON[1] − L. GOBILLON[2]
H. G. OVERMAN[3]

1 University of Toronto, Canada
2 Institut National d'Etudes Démographiques (INED).
3 London School of Economics.

Assessing the effects of local taxation using microgeographic data [§]

Gilles Duranton[‡]
University of Toronto
Laurent Gobillon[§]
Institut National d'Etudes Démographiques
Henry G. Overman[¶]
London School of Economics

Revised version: 30 June 2010

ABSTRACT: We study the impact of local taxation on the location and growth of firms. Our empirical methodology pairs establishments across jurisdictional boundaries to estimate the impact of taxation. Our approach improves on existing work as it corrects for unobserved establishment heterogeneity, for unobserved time-varying site-specific effects, and for the endogeneity of local taxation. Applied to data for English manufacturing establishments, we find that local taxation has a negative impact on employment growth, but no effect on entry.

Key words: Local taxation, spatial differencing, borders, regression discontinuity.

JEL classification: H22, H71, R38.

[§]Thanks to the editor, Antonio Ciccone, two anonymous referees, Tom Davidoff, Fritz Foley, Thomas Klier, Thierry Mayer, John Quigley, Helen Simpson and to seminars and conferences participants in Berkeley, Boston, Ghent, Glasgow, Hamburg, Kansas City, Kiel, Kyoto, London, Oslo, Paris, Seattle, Vancouver, and Venice. We also owe special gratitude to Tim Besley and Tom Holmes. Financial support from the Economic and Social Research Council (ESRC Grant R000239878) and the Leverhulme Trust is gratefully acknowledged. This work contains statistical data from ONS which is Crown copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

[‡]Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: `gilles.duranton@utoronto.ca`; website: `http://individual.utoronto.ca/gilles/default.html`). Also affiliated with the Center for Economic Policy Research, and the Center for Economic Performance at the London School of Economics.

[§]Institut National d'Etudes Démographiques, 133 bd Davout, 75980 Paris Cédex 20, France (e-mail: `laurent.gobillon@ined.fr`; website: `http://laurent.gobillon.free.fr/`). Also affiliated with the Center for Economic Policy Research, PSE, and CREST.

[¶]Department of Geography and Environment, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom (e-mail: `h.g.overman@lse.ac.uk`; website: `http://personal.lse.ac.uk/OVERMAN/`). Also affiliated with the Centre for Economic Policy Research, and the Centre for Economic Performance at the London School of Economics.

1. Introduction

This paper develops new empirical methodologies to identify the effects of local taxation on the location and growth of firms. This issue has been the focus of an extensive theoretical literature and our paper is not the first paper to consider these issues empirically.[1] Bartik (1991) summarises the earlier literature. Evidence from the 1960s and 1970s suggested there was no effect of taxes on firm location decisions. Bartik's own work focusing on the 1980s suggested a negative relationship and a number of subsequent papers have confirmed that finding. Much of this work used fairly large spatial units (mostly US states). More recent work (e.g., Guimaraes, Figueiredo, and Woodward, 2004) has moved towards smaller spatial units such as US counties with similar results.

The existing literature, however, has failed to resolve three main problems when assessing this impact. First, firms are faced with the choice of a large number of heterogenous locations. Many of these site's characteristics are unobserved and likely to be correlated with other explanatory variables such as plant characteristics and local taxation, thus biasing the results. Second, firms themselves are heterogenous. Again much of this heterogeneity is unobservable so that the sorting of firms according to these characteristics provides another source of bias. Third, aspects of the tax system may be endogenous to firm decisions, which may lead to a reverse causality bias.

Our approach is to use spatial differencing, time differencing and instrumenting to solve for these three identification problems. Our use of panel techniques to condition out firm heterogeneity is standard. To solve for unobserved time-varying site characteristics and the endogeneity of local taxation, we show that neither spatial differencing nor instrumenting alone suffices. Instead, a combination of the two is needed. Spatial differencing conditions out local characteristics. In turn this makes the exclusion restriction associated

[1] One can identify at least three strands of theoretical literature. The issue of tax competition has received considerable attention. See Wilson (1999) for a review. A similarly lengthy debate has centred around local public good provision and the Tiebout (1956) hypothesis that inter-jurisdictional competition helps achieve the efficient provision of local public goods. See Epple and Nechyba (2004) for a recent review. Finally, capitalisation of local taxes and the efficient taxation of land are key concerns for urban economists. This has been a subject of discussion since George (1884). See Fujita (1989) for a presentation of the arguments and Arnott (2004) for a recent discussion of its applicability.

with our instrumentation strategy plausible. That is, without spatial differencing the absence of correlation between our political instruments and unobserved time-varying local conditions would be unlikely. When applying our methodology to data for English manufacturing establishments we find that local taxation of non-residential property has a sizeable negative impact on employment growth, but no effect on entry. We show that methodologies that do not address these three problems give substantively different results.

That a tax on non-residential property (which applies to both land and buildings) should affect *employment* might seem surprising. Our model shows how well documented rigidities in rents imply that tax increases may not fully capitalise and thus negatively affect employment. In addition, our model shows how revaluations of properties in case of expansions can magnify those effects and explain why local employment may be severely affected by local taxation.

The rest of the paper is structured as follows. In section 2 we outline our methodology and relate it to the existing literature. Section 3 outlines our data. Section 4 presents our findings on the impact of local taxation on employment while section 5 presents results for entry. Section 6 concludes.

2. Methodology

A. *Model*

To start, it is useful to recall some standard results about the taxation of fixed factors.[2] Taxes on land and *existing* buildings are expected to be fully capitalised into property values. As a result, when buildings remain in fixed supply, rental prices for non-residential properties should be unaffected by local taxation. In turn, this suggests that establishments' employment choices should also be unaffected. The only channel through which property taxation (which taxes both land and buildings) matters is that it acts as a tax on

[2]See, among others, Wellisch (2000) for a detailed discussion.

the construction of new buildings and the extension of existing ones. If the supply of new buildings is perfectly elastic everywhere and firms can move at no cost, all construction (and thus employment growth) will take place where taxes are lowest. Hence, small differences in property taxes can have large employment effects in the simplest of settings.

In reality, we expect two complications that should greatly attenuate this result. First, firms face significant moving costs. Second, the excess burden of property tax increases should be small since property taxation only affects investment in buildings, arguably a small fraction of firms' expenditures. In such a setting, it is hard to imagine that local taxation has a sizeable effect on employment.

For the sake of even greater realism and to explain how local taxation may have first-order effects on employment, we consider a richer model that embeds several aspects of our institutional setting.[3] Specifically, we allow for rigidities in rent setting which imply that property taxation affects the *overall* rental costs of properties (and not only the costs of expansion). Property revaluations (which determine the tax base) occur in case of building expansion. They magnify the effects of an increase in property tax so that property taxation can indeed act as a major break on local employment growth by limiting the expansion of establishments if they stay or by forcing them to move away if they want to grow.

We now describe the details of our model. Establishments use labour and building space to produce a final good.[4] To keep things simple assume that building space and labour are perfect complements.[5] By choice of units, one unit of building and one unit of labour are needed to produce one unit of the final good. The unit rental cost of buildings in jurisdiction a comprises two elements: the building rent, b_a, and the local tax, referred to in the UK as business rates, r_a. Building rents are endogenously determined in a way described below. Consistent with our empirical setting we assume that the tax is paid by

[3] The UK rates system is further described below. See also Hale and Travers (1994).

[4] We use the term establishments to refer to our basic units of observation. In the data that we use, these are sometimes part of larger multi-establishment firms. We ignore the complications this introduces.

[5] As long as labour and buildings remain gross complements, qualitatively similar results are obtained. Gross complementarity between building and labour is a natural assumption for most sectors.

occupiers, not landlords. It is the product of a rate, which varies across jurisdictions, times the historical value of the building.

We take labour as numeraire. The price of the final good is exogenous, common across all jurisdictions, and equal to $1 + p$. Establishments live for up to two periods and new establishments can enter every period. During an establishment's first period, demand for their good is normalised to unity. Period 1 profit of an establishment i initially located in jurisdiction a at time t is given by

$$\pi^1_{at}(i) = (p - b_{at} - r_{at}) \tag{1}$$

In their second period, firms face a shock to the demand for their final good and a tax shock.[6] The second period tax, r_{at+1}, is the realisation of a random draw $\tilde{r}_{at+1}$ from a distribution $f(.)$ over $[\underline{r}, \overline{r}]$ at the beginning of period $t+1$. This is consistent with our empirical setting where the tax is fully decentralised and set freely by each jurisdiction every year.[7] We assume that $\underline{r} \geq 0$ and that $\overline{r}$ is not too high in a sense made clear below. Demand in the second period is given by $1 + \rho_{it+1}$ where ρ_{it+1} is a realisation of a random draw $\tilde{\rho}_{it+1}$ from a continuous distribution $g(.)$ over the support $[-1, \overline{\rho}]$ with $\overline{\rho} > 0$ taking place again at the beginning of period $t+1$. Firms respond differently to negative and positive shocks and we consider each in turn.

We rule out the possibility that establishments can renegotiate rents downwards in response to either a tax or a demand shock. This helps limit the number of possible firm responses that we need to consider. It is also consistent with the institutional setting during our period of analysis when most industrial establishments sign long term rental contracts of 20 years or more with a review of the rent only every five years or so. Even at the time of rent reviews, scope for adjustment is limited because virtually all commercial contracts in the UK include an upward only clause (Crosby, Lizieri, Murdoch, and Ward, 1998). In light of this, we assume that new establishments sign a lease in t which sets building rent for both t and $t+1$. For simplicity and without affecting the qualitative

[6] To keep our model transparent we use shocks affecting demand for final goods. We could instead use productivity shocks with a fully specified demand system to derive similar results.

[7] While we consider an exogenous tax shock, our empirical analysis worries about endogeneity.

nature of our results, we assume $b_{at} = b_{at+1}$. We will return to this issue below in discussing when rents respond to changes in local taxes.

Given that rent cannot be reduced, after a negative shock, $\rho_i < 0$, an establishment wants to either exit or downsize.[8] Exit leads to second period profit $\pi^2_{t+1}(i) = \pi_E(i) = 0$ where the subscript E stands for exit. Downsizing occurs through subletting part of the original building unit. In the UK, commercial contracts give tenants "the right to sell or sublet the unexpired term of the lease, with landlords being unable to withhold their consent unreasonably" (Crosby *et al.*, 1998, p. 3). We assume that subletting part of the initial unit of building yields a unit rent $\underline{b}$ where $\underline{b} < b_{at} + r_{at+1}$ in equilibrium. That is, firms cannot make profits by subletting. Downsizing to satisfy the new level of demand and renting unused building space implies second period profit $\pi^2_{t+1}(i) = \pi_D(i) = (1 + \rho_i)p - \rho_i\underline{b} - b_{at} - r_{at+1}$.

Following a positive demand shock, $\rho_i \geq 0$, an establishment has three options. First, it can grow by expanding employment and building space on its current site.[9] The unit rent for extra building space is b_{at+1}, the same as for the original unit of building. The main issue with increasing the amount of building space on a given site is that, in our setting, it typically led to a revaluation of the building (Hale and Travers, 1994). This change in the tax base implies that the unit rate increases for the expansion as well as the original unit. We assume that revaluation leads the value of the tax base to be multiplied by $\delta > 1$. Profit in this case is $\pi^2_{t+1}(i) = \pi_G(i) = (1 + \rho_i)(p - b_{at} - \delta r_{at+1})$.

Alternatively, an establishment can decide to forego this expansion opportunity and continue to produce only one unit of final good. Profit when staying the same is $\pi^2_{t+1}(i) = \pi_S(i) = p - b_{at} - r_{at+1}$. Finally an establishment can incur cost c and relocate to another jurisdiction where the unit rental cost of buildings is equal to $\overline{b}$ such that in equilibrium $\underline{b} < \overline{b} < b_{at} + \delta r_{at+1}$. Profit in case of relocation is $\pi^2_{t+1}(i) = \pi_R(i) = (1 + \rho_i)(p - \overline{b}) - c$.

[8] Another alternative would be to relocate and leave the existing lease. To keep the number of options manageable, we assume that moving cost, c, would make this alternative prohibitively costly. More precisely, we assume $c > b_{at} + \overline{r} - \underline{b}$ where $\underline{b}$ is the unit rental costs of buildings when subletting (as defined below).

[9] It could also occur by renting adjacent sites. We ignore this possibility here.

Hence, depending on its demand shock, ρ_{it+1}, and the level of tax in period $t+1$, r_{at+1}, an establishment faces five possible choices: exit (leading to second period profit π_E), downsize (leading to π_D), stay put (π_S), grow (π_G), or relocate (π_R).

We now define four thresholds, $\rho_{ED}(r_{at+1})$, $\rho_{DS}(r_{at+1})$, $\rho_{SG}(r_{at+1})$, and $\rho_{GR}(r_{at+1})$. They correspond to particular realisations of ρ such that an establishment is indifferent between exit and downsize, between downsize and stay at its original size, between stay at its original size and grow locally, or between grow locally and relocate, respectively. Provided the tax, r_{at+1}, is not too high and relocation costs, c, are large enough, the above expressions for π_E, π_D, π_S, π_G, and π_R imply $-1 \leq \rho_{ED}(r_{at+1}) < \rho_{DS}(r_{at+1}) < \rho_{SG}(r_{at+1}) < \rho_{GR}(r_{at+1})$. Establishments that face very negative demand shocks prefer to exit. For a less negative shock, they remain in business but downsize. For a small positive shock, establishments retain their original size. For an intermediate positive shock, they grow. For a large positive shock, they choose to relocate. We briefly explain the ranking of each of these thresholds in turn.

Because establishments that remain in business are stuck with their lease and because they can sublet below the rental cost they face, establishments which experience a large negative demand shock exit rather than downsize to a very small size. Interestingly, $\partial\rho_{ED}(r_{at+1})/\partial r_{at+1} > 0$. That is, a higher tax in $t+1$ induces more exits. More exits also imply that surviving establishments are those that have experienced a less negative demand shock.

Next, it is easy to show that $\rho_{DS}(r_{at+1}) = 0$. Establishments that face a small negative shock are left with unnecessary building space. They prefer to sublet it rather than leave it empty. Establishments that face a small positive shock would like to expand. However, adding building space implies a revaluation of the tax base and a higher tax. Whenever the demand shock is positive, but not large enough to offset this increase in the tax, establishments prefer to keep their original size instead of expanding. Hence, $\rho_{SG}(r_{at+1}) > 0$. It is also easy to see that $\partial\rho_{SG}(r_{at+1})/\partial r_{at+1} > 0$. A higher tax in $t+1$ makes the cost of a revaluation higher and thus can only be justified for larger demand shocks. That is, for establishments that remain in jurisdiction a, a higher rate of taxation leads to lower employment growth on average.

Finally, establishments that face a large enough positive demand shock prefer to relocate. The fixed cost of relocation can only be justified for those establishments that need to expand a lot. Because a higher tax implies a greater gain from relocation, we have $\partial\rho_{GR}(r_{at+1})/\partial r_{at+1} < 0$.

To close the model, we assume a competitive land market for new establishments. Free entry means that establishments enter until they make zero expected profit $E(\Pi_{at}) = \pi^1_{at} + E(\pi^2_{t+1}) = 0$ and b_{at} adjust to ensure this holds across all jurisdictions.[10] Put differently, when new entrants sign contracts current and expected future taxes are fully capitalised in b_{at}. However, the realisation of r_{at+1} is not capitalised for continuing establishments. Loosely speaking, taxes are capitalised in the long run but not in the short run. Furthermore, clearing on the land market implies that more exits and more relocations in a jurisdiction are matched by more entries when the supply of sites is fixed (a reasonable assumption for the UK).

The main result of our model is thus that the tax rate affects the use of building space by establishments. In turn, the complementarity between factors implies that taxation affects the employment decision of establishments. There are two separate channels. Higher taxes lead to both a *growth slow-down* and a *selection* effect. More precisely, higher taxes reduce employment growth by inducing more establishments to keep their size constant instead of growing. The growth slow-down effect is driven entirely by building revaluations that lead to higher taxes and make small expansions unprofitable. In terms of selection, higher taxes in $t+1$ imply more exits for establishments with a negative shock and more relocations for establishments with a positive shock. The selection effect is thus ambiguous. The selection effect is driven by imperfect capitalisation of higher taxes into rents for existing establishments. This affects the local profitability of establishments and,

[10]Developing this expression is not very enlightening. When r_{at+1} and ρ_{t+1} are independent, expected profit is well defined and equal to $E(\Pi_{at}) = \pi^1_{at} + \int_{\underline{r}}^{\overline{r}} \int_{-1}^{\overline{\rho}} \pi^{2*}_{t+1}(r_{at+1}, \rho_{t+1}) f(r_{at+1}) g(\rho_{t+1}) \mathrm{d}r_{at+1} \mathrm{d}\rho_{t+1}$ where $\pi^{2*}_{t+1}(r_{at+1}, \rho_{t+1})$ is optimal second-period profit.

in turn, their location choice.[11]

Overall, the effect of higher taxes on employment in the establishments that stay is thus ambiguous. Higher taxes lead to less employment when selection through exits is dominated by selection through relocation and the growth slow-down effect.[12] In our empirical analysis, we can estimate an overall effect but cannot identify the selection and growth slow down effects separately.

Our model also offers some predictions about entry. Through the selection effect, higher taxes imply more exits and more relocations. In turn, this should lead to more entries. In the second part of the paper, we assess the effect of taxes on entries.

We could extend our framework to incorporate other factors of production. The tax could cause establishments to change their use of these other inputs. In turn, this could impact on employment. Because of these omitted inputs, we need to be cautious about the interpretation of our results. Specifically, we are only able to estimate the overall effect of taxation on employment even though other cross-factor effects may be at work.[13]

Because higher taxes affect the employment of establishments, a naive reading of our model would lead us to estimate

$$e_{it} = \alpha r_{at} + \epsilon_{it} \tag{2}$$

where e_{it} is the log employment of establishment i at time t and ϵ_{it} an error term that captures the demand shock ρ_{it} in reduced form. The main parameter of interest is α. It captures the (net) effect on employment of the (log) local tax, r_{at}.

[11] As discussed at the start of this subsection, there is imperfect capitalisation because of the combination of rigid rents, uncertainty about future taxes, and costly relocation. Without rent rigidity, renegotiation in a competitive land market implies full capitalisation of r_{at+1} into b_{at+1}. Without uncertainty, r_{at+1} would be capitalised into the initial level of rent b_{at} just like r_{at} since new establishments face a competitive land market. Finally, without costly adjustment, establishments could leave their lease and negotiate a new one competitively.

[12] We note that these effects can be quantitatively large. If in a jurisdiction there are lots of establishments just below ρ_{GR}, a further unexpected increase in taxation can lead to many relocations and a much lower growth rate in employment for remaining establishments.

[13] The effects of taxation may also vary across places. See, for instance, Baldwin, Forslid, Martin, Ottaviano, and Robert-Nicoud (2004). In this paper we only estimate an average effect.

B. ***Heterogenous establishments and heterogenous locations***

Unlike in the model, we expect production establishments to be very heterogenous ex-ante and much of this heterogeneity to be unobservable. This is a first likely source of bias if establishments with different unobserved characteristics sort across jurisdictions with different tax rates. We can enrich the specification (2) and add establishment characteristics:

$$e_{it} = \alpha r_{at} + X_{it}\beta + \mu_i + \epsilon_{it} \tag{3}$$

where β captures the effect of time-varying establishment-specific observable variables, X_{it}. An establishment fixed effect, μ_i, captures the impact of unobservable time invariant establishment characteristics.

The second issue to consider is that not all sites are the same. There are a large number of heterogenous sites and these sites come at very different rental prices. For instance, Thompson and Tsolacos (2001) document a sixfold difference in the rental price between industrial sites close to Heathrow airport and those in the suburbs of Leeds. It is unlikely that these differences in rents only reflect differences in local taxation.

Traditional empirical approaches have worried about the fact that the costs of factors of production differ across geographical regions. We are also concerned with heterogeneity at a much finer geographical scale. A wide variety of factors affect both the attractiveness of sites and the success of establishments once they choose their site. For example, the attractiveness of a site may depend on access to the road network while improvements to that network may affect the performance of establishments at that site. Similarly, changes in congestion can have different implications for establishments located close, but not very close to one another as can the entry or exit of big buyers or suppliers. Overall, there are many reasons to expect considerable site heterogeneity, possibly at a very fine spatial scale, and varying over time.

If this is the case, assessing the impact of taxation requires us to control for both fixed and time-varying site characteristics since they are likely to be correlated with taxes or the demand shock for establishments. Some of these site characteristics may be observable and thus can be controlled for directly. However many are likely to be unobservable. This implies that we are interested in a specification like

$$e_{it} = \alpha r_{at} + X_{it}\beta + \mu_i + \gamma_a + \theta_{zt} + \epsilon_{it} \tag{4}$$

where γ_a is a time-invariant effect for jurisdiction a and θ_{zt} is a time-varying effect for location z, possibly at a finer spatial scale than a. Note that the establishment fixed effects also control for unobserved time-invariant site-specific effects (if establishments do not move) leaving θ_{zt} to control for unobserved time-*varying* site-specific effects.[14]

Estimating (4) by OLS ignoring the unobservable effects gives a consistent estimate of α and β only if $\text{Cov}\left([r_{at}, X_{it}], \mu_i + \gamma_a + \theta_{zt} + \epsilon_{it}\right) = 0$. This condition is unlikely to hold, if only because observable establishment characteristics X_{it} are likely to be correlated with unobservable establishment characteristics μ_i. Hence, estimating (4) by OLS is unlikely to yield a consistent estimate of α.

C. *Time differencing*

As a first step, and to control for establishment and jurisdictional fixed effects we can use the panel dimension of our data to calculate the *within* estimator (alternatively, we could calculate the first-difference estimator). The *within* transformation is obtained, as usual, by centring all observations around their mean. For any variable y, for observation i, let $\overline{y}_i$ denote the time average and define $\tilde{y}_{it} \equiv y_{it} - \overline{y}_i$. We can then rewrite equation (4) as:

$$\tilde{e}_{it} = \alpha \tilde{r}_{at} + \tilde{X}_{it}\beta + \tilde{\theta}_{zt} + \tilde{\epsilon}_{it} \tag{5}$$

So far our approach for dealing with heterogenous establishments and time-invariant spatial heterogeneity is standard. Because we have a panel of establishment data, we are able to control for observed time-varying characteristics of establishments, as well as condition out unobserved time-invariant characteristics of both establishments and jurisdictions through the inclusion of establishment and jurisdiction fixed effects.

This specification will give consistent estimates of α and β if $\text{Cov}\left(\left[\tilde{r}_{at}, \tilde{X}_{it}\right], \tilde{\theta}_{zt} + \tilde{\epsilon}_{it}\right) = 0$. This condition, although weaker than that necessary for consistency of OLS, is still

[14] Empirically, we cannot separately identify the time-invariant establishment-specific and jurisdiction-specific effects because we condition both out using establishment fixed effects.

unlikely to hold because, in a spatial context, the site-specific effect θ_{zt} is likely to be correlated across neighbouring sites. This raises the possibility that, within a jurisdiction, there could be omitted variables driving both the average site-specific effect and the tax rate. That is, r_{at} is likely to be correlated with θ_{zt}. Hence, although it conditions out time-invariant characteristics of both establishments and jurisdictions, time differencing is not enough to obtain consistent estimates of α.

D. *Instrumenting*

The standard way to deal with the correlation between r_{at} and θ_{zt} would be to find a suitable instrument for the tax rate. In our context, one possibility is to use local political variables (denoted s_{at}) to instrument for tax rates. We expect political parties to set local taxes differently. For instance, local authorities controlled by the (left-wing) Labour party are likely to set higher taxes than (right-wing) Conservative ones. More subtly, we also expect a local authority with a strong Labour majority to set higher taxes than a local authority with small Labour majority. Changes in these political variables are highly likely to cause changes in local tax rates (i.e., they satisfy the relevance condition for a suitable instrument). The crucial issue is whether they satisfy the exogeneity condition $\text{Cov}(\tilde{s}_{at}, \tilde{\theta}_{zt} + \tilde{\epsilon}_{it}) = 0$. In this respect note that changes in θ_{zt} are likely to be correlated across sites within jurisdictions while changes to the 'average' θ_{zt} in a jurisdiction may be correlated with changes in voting behaviour. That votes at local elections should be determined by local conditions is a very real possibility and would suggest a correlation between political variables, $\tilde{s}_{at}$, and unobserved local effects, $\tilde{\theta}_{zt}$. In turn, this would violate our exclusion restriction and yield inconsistent estimates for α. Since it is unclear to us what would be a variable that determines local taxation but is uncorrelated with local conditions, we conclude that instrumenting for local taxation while estimating (5) is unlikely to solve our inference problem.

E. *Spatial differencing*

An alternative to instrumenting is spatial differencing in the spirit of Holmes (1998) or Black (1999).[15] Define Δ_d as the spatial difference operator which takes the difference between each establishment and any other establishment located at distance less than d from that establishment. Applying this spatial difference operator to (5) gives:

$$\Delta_d \tilde{e}_{it} = \alpha \Delta_d \tilde{r}_{at} + \Delta_d \tilde{X}_{it} \beta + \Delta_d \tilde{\theta}_{zt} + \Delta_d \tilde{\epsilon}_{it} \tag{6}$$

Now, we impose the crucial identifying assumption that site-specific effects change smoothly across space. That is, for d sufficiently small $\Delta_d \tilde{\theta}_{zt} \approx 0$. Noting also that taxes will be the same for establishments within the same jurisdiction, this gives us:

$$\Delta_d \tilde{e}_{it} = \Delta_d \tilde{X}_{it} \beta + \Delta_d \tilde{\epsilon}_{it} \tag{7}$$

for establishments in the same jurisdiction and:

$$\Delta_d \tilde{e}_{it} = \alpha \Delta_d \tilde{r}_{at} + \beta \Delta_d \tilde{X}_{it} + \Delta_d \tilde{\epsilon}_{it} \tag{8}$$

for establishments across jurisdictional boundaries. This shows that we can use neighbouring establishments located across jurisdictional boundaries to identify the effects of taxation. We can also use neighbouring establishments within the same jurisdiction to improve our estimates of the effect of establishment-specific variables. Estimating equation (6) will give consistent estimates of α and β if $\mathrm{Cov}\left(\left[\Delta_d \tilde{r}_{at}, \Delta_d \tilde{X}_{it}\right], \Delta_d \tilde{\epsilon}_{it}\right) = 0$.

Theoretically, assuming $\tilde{\theta}_{zt}$ varies continuously across space, then $\Delta_d \tilde{\theta}_{zt} = 0$ will hold for arbitrarily small distances and spatial differencing alone will provide consistent estimates. In practice, however, because we need enough observations to conduct our estimation we may have to spatial difference across establishments that are too far apart to ensure $\Delta_d \tilde{\theta}_{zt} = 0$. Continuity of $\tilde{\theta}_{zt}$ ensures that shocks to neighbouring establishments are correlated *within* jurisdictions so if tax rates are endogenous to local conditions, then in

[15] The methodology proposed by Holmes (1998) and Black (1999) has been repeated elsewhere, but most applications only use cross-sectional data and do not address endogeneity. The three exceptions are Gibbons and Machin (2003) who consider endogeneity and Kahn (2004) and Chirinko and Wilson (2008) who use some longitudinal information. Our analysis improves on these existing methodologies by incorporating instrumented panel data techniques into the spatial discontinuity approach.

practice spatial differencing alone will reduce, but not eliminate the correlation between changes to tax rates and local shocks.

F. *Instrumenting and spatial differencing*

Instrumenting and spatial differencing, when used 'alone' are not fully satisfactory. Instrumenting is imperfect because instruments that determine local tax rates are likely to be correlated with unobserved time varying local effects. Spatial differencing is imperfect because, in practice, it may not remove all the endogeneity of local taxes. This suggests that combining both approaches can allow the proper identification of α. To do so we need to use appropriately transformed political variables: $\Delta_d \tilde{s}_{at}$.

Using these instruments together with spatial differencing gives consistent estimates provided $\text{Cov}(\Delta_d \tilde{s}_{at}, \Delta_d \tilde{\theta}_{zt} + \Delta_d \tilde{\epsilon}_{it}) = 0$. This condition is weaker than the corresponding condition when using IV without spatial differencing because Δ_d removes any component that varies smoothly across space and drives both political shares and site-specific effects. It is also weaker than the corresponding condition when using spatial differencing without IV because we use only the variation in changes in tax rates that is explained by changes in local elections.

Before turning to the implementation, note that spatial differencing and the *within* transformation have implications for the error structure. As usual, implementing (8) yields consistent estimates for the coefficients but does not give the correct standard errors. Appendix A shows how to correct the standard errors.[16]

3. Data

To implement our methodology, we need data satisfying a number of requirements. First, we need to have a panel of individual establishment level data. Cross-sectional data do not allow us to use the *within* transformation to remove establishment and jurisdiction-

[16] As we discussed above we ignore the complication introduced by multi-plant firms. We also do not correct for the fact that taxes vary by jurisdiction rather than establishment (Moulton, 1990).

specific effects. We then need to be able to precisely locate these establishments so that we can identify which pairs of establishments are neighbours. Finally, we need to identify a local tax which is time-varying. We would prefer this tax to be economically significant to increase the chances of detecting any impact on location and employment decisions. Data satisfying all of these requirements is available for England for the six year period from 1984 to 1989. We first describe the establishment level data set we use before turning to details of the particular tax that we consider.

A. *Establishment data*

Establishment level data for years 1984 to 1989 come from the Annual Respondent Database (ARD) which underlies the Annual Census of Production in the UK. Collected by the Office for National Statistics (ONS), the ARD is a rich data set providing information on all UK establishments from 1973 onwards. We face two restrictions on time period. Changes to the tax system restrict our focus to years before 1990, while changes to the ARD restrict us to years after 1983. We also restrict ourselves to English manufacturing establishments.[17] For every establishment, we know postcode, five-digit industrial classification, and number of employees. In our analysis of employment we use only the information coming from a sub-sample of 'selected' establishments required to make a detailed return in any year. These establishments are generally larger and their employment information is of better quality than for non selected establishments.[18] The precise sampling frame for selected establishments and further description of this data can be found in Griffith (1999). In our analysis of entry we use the exhaustive data (that includes 'selected' and 'non-selected' establishments).

The postcode reported in the ARD is very useful for locating establishments. In the UK, postcodes typically refer to one property or a very small group of dwellings. The Ord-

[17]We ignore Scotland because it operated a different local tax system, Wales because it is not covered by the data set that provides our instruments, and Northern Ireland because special permission is required to access ARD data for establishments located there.

[18]The better quality of data is not the only reason to focus on selected establishments. Implementation of the algorithm to identify neighbours and calculate data for pairs of neighbouring establishments is infeasible with the large samples that include non-selected establishments.

nance Survey (os) Code-Point data set gives spatial coordinates for all uk postcodes. By merging this data with the ard we generate very detailed information about the location of all English manufacturing establishments. For all but a tiny percentage of matched establishments the os acknowledges a potential location error below 100 metres. For the remaining observations, the maximum error is a few kilometres. Overall, we expect a very high level of precision for our location data (see Duranton and Overman, 2005, for further discussion).

As the data requirements for our spatial differencing methodology are already somewhat restrictive, we only consider establishment specific variables that can be calculated for all establishments. For the ard, this means only controlling for establishment age. For establishments already in the panel in 1976 (the earliest year for which we have information), we are unable to assess exact age. With regressions running over 1984-1989, this truncation censors age for establishments older than 8 years in 1984. For consistency across years, we thus construct a dummy for establishments that are 8 years or older in 1984. We also use age and age squared interacted with a dummy for being an establishment younger than 8 years old.

There are 25,579 establishments in the ard that are located in England and that report employment at least once within our study period. We make several sample restrictions. We deleted 122 establishments where the Local Authority code was missing since we need this to append the tax rate. We then dropped 374 establishments where we could not identify coordinates.[19] A further 723 establishments change sector during the period and 2,547 move site (i.e., change coordinates). In both cases these changes frequently reflect coding errors rather than genuine changes in activity or location, so we drop these establishments. This leaves us with a sample of 21,813 establishments and a total of 61,785

[19]We directly identified coordinates for around 90% of establishments. The main problem for the remaining 10% results from the creation of new postcodes. To increase matching rates, we checked our data against a data set of postcode updates. For observations with postcodes still unmatched, we imputed co-ordinates as follows. In the uk, a postcode is either six or seven digits. We first drop the last digit and assign establishments the mean coordinates of all postcodes sharing the same truncated postcode. If this failed to produce a match, we dropped two, then three digits from the postcode and again matched on mean coordinates where possible. This left us with 1.4% of establishments that could not be given a grid reference.

observations.

B. *Local taxation and political variables*

As already outlined above, the tax that we consider is a property tax on non-residential property known as the UK business rate, set and collected by jurisdictions called Local Authorities (hereafter LAs). Before 1990, tax rates varied over time and jurisdictions. We are fortunate because this tax consisted only of a flat rate that applied to all occupied non-residential properties. This simplicity implies that the only source of variation is the tax rate itself.

The UK business rates were subject to a major reform in 1990 which essentially eliminated jurisdictional variation. This reform provides a large amount of exogenous variation in the tax rate. Unfortunately we cannot exploit it because all properties were also revalued in 1990. Since we do not know by how much each property was revalued, we cannot compute the change in taxation faced by each establishment. Hence, as mentioned above, while the production data restricts our study period to start in 1984, the reforms to the tax system mean it must end in 1989.[20]

UK business rates represented a considerable tax on business. In 1990 the average annual rates bill per square foot was over £4, compared with an average rent bill per square foot of just over £13 (Bond, Denny, Hall, and McCluskey, 1996). Hence rates bills make up roughly 25 per cent of the total occupancy costs of rented commercial property. In 1992 business paid £13bn in local rates, almost equal to the £15bn that they paid in corporation tax for the same year.[21]

As a first approximation, none of this money was used to finance local services for businesses. During the 1980s, the responsibilities of LAs include social and health services for their residents, community safety, some aspects of education, housing, arts, culture,

[20]For a comprehensive discussion of the reforms see Hale and Travers (1994).

[21]Additional evidence that the tax is significant is provided by Bennett (1986) using a cost of capital approach.

and environment. Very little is directly provided to businesses.[22] Even basic services such as refuse collection need to be separately organised by businesses. The only LA responsibility that can affect businesses directly is planning (and some trading regulations that affect mostly retail). Despite a national set of planning regulations, LAs differ in their speed of processing applications. It is hard to believe, however, that there were substantial within jurisdictional changes in the efficiency of local planning offices over the time period we consider. Instead, differences with respect to planning efficiency and restrictiveness, if they matter at all, are expected to be part of the fixed effect of these jurisdictions. Of course, if changes to taxes are used to fund services that directly or indirectly benefit businesses, then it is still of interest to understand the net effect.

During our period of study, tax rates were set locally by the 366 LAs in England, which entirely covered its land area. Large cities comprise several LAs (e.g., 33 for London). The tax paid depended on the value of the buildings occupied. The tax rates were known as 'rate poundages' and the value of the property including buildings as the 'rateable value'. We use the rate poundages as explanatory variable in our regressions. They were changed yearly. Unless some construction took place, rateable values of buildings were fixed in 1973 and did not change until 1990. The average rate for 1987 was 229.5 pence per pound of 1970 rateable value with a standard deviation of 32.2 pence. The ratio between the highest business rate (Oldham in Manchester) and the lowest (Kensington and Chelsea in London) was about three. These two LAs are far from each other and this difference will not be taken into account when we spatial difference. The largest ratio between two *neighbouring* LAs is nearly two. There was also significant time-series variation. There was for instance an average increase in the business rates of nearly 10 per cent between 1988 and 1989. Over the entire period ($1984 - 89$) the average increase was 45 per cent with a standard deviation of 0.15. Again looking at neighbouring LAs, we observe a 25 per cent decrease in Kensington and Chelsea against a 31 percent increase in neighbouring

[22]Clearly, there may be some indirect effects through, for instance, crime or education. However we expect these effects to spill over across jurisdictions so that establishments on both sides of any border face similar conditions.

Hammersmith over the period. Overall, there is more than sufficient variation in the data to perform our estimations.

Finally, the data on shares of political parties that we use for instrumenting, come from the British Local Election Database which is available through the UK Data Archive. See Rallings and Thrasher (2004).

4. Results

To show the impact of spatial differencing, we also estimate employment effects of local taxation using standard techniques. Results for the non-spatial specifications (i.e., using standard techniques and ignoring the micro-spatial nature of the data) are reported in table 1. We use log employment (e_{it}) as the dependent variable. As explanatory variables we include log tax rate (r_{at}), the three variables described above to capture the impact of age (X_{it}) and a full set of industry-year dummies (two-digit industries) for which we do not report coefficients.[23] Table 2 reports results for comparable specifications that use spatial differencing for establishments matched by industry and years. For the sake of comparison, we restrict the sample to be the same for all specifications by only including those establishments that we use to estimate our preferred specification: the instrumented, fixed effect spatial-differencing specification discussed below (the final specification in table 2). This restriction requires establishments to be in a pair in which both establishments simultaneously report employment in at least two years.[24]

[23]Industry-year dummies allow for industries to have different μ, γ and θ. In the spatial difference specifications we allow for this by only pairing establishments if they are part of the same industry. This restriction in defining pairs limits the degrees of freedom sufficiently that we cannot allow α and β to differ across industries. Thus, we impose the same assumption here to facilitate comparison. We could drop this assumption if we paired establishments across industries. But this would impose an identical distribution of θ across industries. As we are interested in the econometric issues arising because of the presence of θ we prefer to impose the restriction on α and β.

[24]Consider an establishment that reports employment for (say) only 1987 and 1988. It is in the sample if it has at least one neighbour and that neighbour also reports employment for these two years. We impose this restriction because our implementation of spatial-differencing uses fixed effects for pairs. An establishment can be part of more than one pair satisfying this condition. However, we only use each establishment once in estimating the non-spatial specifications.

Table 1: Non-spatial regression results

	OLS	WITHIN	WITHIN IV
(log) tax rate	0.210[a]	0.071[b]	0.363[a]
	(0.060)	(0.030)	(0.104)
age censored dummy	0.569[a]	0.214[a]	0.223[a]
	(0.091)	(0.033)	(0.034)
age	0.023	0.036[a]	0.036[a]
	(0.047)	(0.011)	(0.011)
age squared	0.005	-0.001	-0.001
	(0.005)	(0.001)	(0.001)
Adjusted R-squared	0.12		
Number of observations	13490	13490	13490
Number of establishments	4414	4414	4414

Notes: Regression of (log) employment on (log) local tax rates and age variables. First column reports results from OLS, second column (WITHIN) allows for establishment specific fixed effects, third column (WITHIN IV) further instruments local tax using local political variables. Standard errors under coefficients. [a], [b] and [c] denote significance at the 1%, 5%, and 10% level respectively.

Starting with 21,813 establishments, 7,938 of them are dropped because they only report employment once. Of the remaining 13,875 establishments, 8,792 have a neighbour within 1 kilometre (the distance threshold that we use for our preferred reported results). However, only 4,414 of these establishments are in at least one pair where both establishments simultaneously report employment in at least two years. To show that this sample is representative, Appendix B reports results for the same specifications as table 1 but different sampling rules; namely pairing establishments within 2 kilometres (instead of 1 kilometre), using all establishments that can be paired when spatially differencing, and using all available establishments in the data.

The first regression in table 1 reports results from estimating equation (4) using pooled OLS for 1984-1989. The results for the age variables show that, as expected, older establishments have higher employment. Our main focus, however, is on the role of taxation. In the cross-section, higher tax rates are associated with higher employment.

As we noted above, one possible explanation of this positive correlation is that some establishments are larger than others for unobserved reasons and larger establishments happen to be located in higher tax jurisdictions. The second specification in table 1 allows for this possibility by introducing an establishment-specific fixed effect and calculating the *within* estimator. The coefficient on tax rate is divided by three suggesting that much of the positive correlation between employment and tax rate is indeed due to unobserved characteristics of establishments. Note, however, that the effect remains positive and significant.

The remaining problem that we need to tackle is that the tax rate may be correlated with the error in equation (5). There are two possible sources for this correlation. First, there may be a feedback from employment to tax rate. This feedback will be positive when local politicians tax local business more when it is doing well. Alternatively, and working in the opposite direction, it could be that LAs can afford to keep taxes low during 'good times'. In the UK context, this alternative may arise because good times imply a lesser need for social expenditure. This being said, we expect the first effect to dominate and taxation to go up when local business is doing well. Second, there may be other *time-varying* characteristics that are positively correlated with both employment and tax rates and that we do not control for through the use of establishment-specific fixed effects.

To solve these problems we use local political variables to instrument for tax rates. The full set of instruments includes the share of local politicians affiliated with the three main political parties (Conservative, Labour and Liberal Democrat), a set of dummies indicating whether the LA is controlled by one of the three main parties and a set of interactions giving the share of the three main parties if they control the LA. There are many smaller parties that play a role in local politics and we aggregate these in to 'other' and treat them as the omitted category. As argued above, we expect both the identity of the party which controls the LA and its margin of control to matter. The R-squared of the first stage regression is 69%. For the interested reader Appendix C provides further details.[25]

[25] One might be tempted to test for the validity of instruments given that we have more instruments than endogenous regressors. We would argue, however, that such a test is invalid because all our instruments are based on the same underlying assumption (that local politics is independent of changes to establishments' employment).

The third column in table 1 shows what happens when we use these instruments for the level of taxation in the fixed effects specification that we reported in column 2. Surprisingly, perhaps, the coefficient on taxation increases after instrumenting. It would be tempting to conclude that, contrary to most priors, LAS taxed businesses according to jurisdictions (social) needs rather than local business' capacity to pay. That is, there is a negative correlation between tax rates and the omitted variables so that instrumenting leads to a higher coefficient on taxation. Nonetheless, a positive effect with an elasticity close to 0.4 strikes us as implausible. As we will see below it appears that this increase in the coefficient on taxation following instrumentation occurs because the unobserved time-varying site-specific effect is correlated with both employment and local political variables.

Table 2 presents two sets of results (with and without corrected standard errors) for three different spatial specifications that parallel those presented in table 1. In the first column, we spatial difference equation (4) and estimate using OLS. In the second, we spatial difference (5) and estimate using a fixed effect for each pair of establishments. Finally in the third column, we instrument the tax rate in the spatially differenced *within* specification using spatially differenced political variables as in the non-spatial specification.

The results use a distance threshold of 1 kilometre to identify neighbours. In our choice of threshold, we face a tradeoff between sample size and the extent to which the spatially varying site-specific effects are equal across neighbouring sites. We chose the minimum threshold which gives sufficient observations to identify the effect of local taxes (remembering that identification comes from cross jurisdiction border pairs). We use neighbouring establishments within the same jurisdiction to improve our estimates of the effect of establishment-specific variables.[26]

Note that, although the overall sample of establishments is restricted to be identical for

[26]We get the same results if we restrict attention only to the 164 establishments that are part of cross-jurisdiction border pairs. The spatially differenced within IV specification gives a coefficient of -1.072 as opposed to the -1.024 reported in the text. Both coefficients are significant at the 1% level.

Table 2: Spatial differencing regression results, one kilometre threshold

Uncorrected standard errors			
spatial difference of	OLS	WITHIN	WITHIN IV
(log) tax rate	0.846[a]	0.111	-1.024[a]
	(0.225)	(0.119)	(0.314)
age censored dummy	0.738[a]	0.134[a]	0.132[a]
	(0.076)	(0.028)	(0.028)
age	0.068[c]	0.042[a]	0.041[a]
	(0.039)	(0.009)	(0.009)
age squared	-0.003	-0.003[a]	-0.003[a]
	(0.004)	(0.001)	(0.001)
Corrected standard errors			
spatial difference of	OLS	WITHIN	WITHIN IV
(log) tax rate	0.846[b]	0.111	-1.024[b]
	(0.379)	(0.167)	(0.420)
age censored dummy	0.738[a]	0.134[a]	0.132[b]
	(0.138)	(0.051)	(0.049)
age	0.068	0.042[a]	0.041[a]
	(0.071)	(0.016)	(0.015)
age squared	-0.003	-0.003[c]	-0.003[a]
	(0.008)	(0.002)	(0.002)
Adjusted R-squared	0.04		
Number of observations	18370	18370	18370
Number of establishments	4414	4414	4414
Number of pairs	6087	6087	6087

Notes: Regression of spatial difference of (log) employment on spatial difference of (log) local tax rates and age variables. First column reports results from OLS, second column (WITHIN) allows for establishment specific fixed effects, third column (WITHIN IV) further instruments local tax using local political variables. Standard errors under coefficients. First block of results report uncorrected standard errors. Second block of results report standard errors corrected according to Appendix A. [a], [b] and [c] denote significance at the 1%, 5% and 10% level respectively.

both the spatial and non-spatial specifications, the number of observations is higher for the spatially differenced specifications (18,370 compared to 13,490). This is because each establishment can be involved in more than one pair. Specifically, we have 6,087 unique pairs as compared to 4,414 unique establishments suggesting that each establishment has on average three neighbours.[27]

Before turning to the individual coefficients, comparing the two blocks of results (with and without corrected standard errors), we see the corrections outlined in Appendix A generally increase standard errors by around 50%. In our context, this results in minor changes in significance, but does not change overall findings. In other contexts it could, suggesting that the correction should usually be implemented. Turning to the coefficients, we see that, apart from changes in significance, the results on the age variables are essentially unchanged. As before, we focus on the tax rate and note that after spatial differencing we get a higher correlation between (spatially differenced) tax rate and (spatially differenced) employment than previously, with a coefficient of 0.846 compared to 0.210 with OLS. These coefficients have the same probability limit when there are no unobserved establishment or site-specific effects, or when these effects are uncorrelated with tax rates or other included explanatory variables. A possible reason for the higher correlation between employment and taxation after controlling for site-specific effects is that areas with poor sites had higher tax rates thus biasing downward the coefficient on taxes in the non-spatial OLS estimation. A correlation between having 'poor sites' and higher taxes is certainly believable given that de-industrialising LAs tended to vote for very left-wing councils that then greatly increased taxation.[28]

However, this interpretation assumes that establishment fixed effects are either absent or uncorrelated with local tax rates. To account for possible correlation we use, as before, the panel dimension of our data to control for unobserved establishment heterogeneity.

[27]Working against this, is the fact that both establishments in the pair must simultaneously report employment data in at least two years. For this distance threshold the first effect dominates. That need not be the case for other distance thresholds.

[28]Sheffield under the leadership of the (then) leftwing firebrand David Blunkett had the highest business rate in the country in 1990 and Liverpool led by the notorious Derek Hatton ranked 12th.

Column 2 of table 2 reports results when we both spatially difference and allow for fixed effects for pairs of establishments. These pair fixed effects not only control for time-invariant unobserved establishment heterogeneity but also for other time-invariant local effects such as the propensity of some jurisdictions to provide better services and thus have consistently better performing establishments.

Comparing results across the non-spatial and spatially differenced specifications we see that after controlling for establishment fixed effects, the coefficient on the tax rate is again higher after spatial differencing. However because of higher standard errors, it is hard to provide a definitive interpretation for this comparison. More significantly, comparing across the spatially differenced specifications, we see that allowing for pair fixed effects reduces our estimate of the positive correlation between taxation and employment relative to the spatially differenced OLS results. The coefficient even becomes insignificant. This confirms our finding from the non-spatial specifications that establishment fixed effects appear to be positively correlated with tax rates: LAs with 'good' establishments charge higher taxes.

As with the non-spatial fixed effects specifications, we still want to control for the fact that tax rates may be endogenous. To do this, we instrument using the spatial difference of the same set of political variables that we use for the non-spatial specification. Results are shown in column 3 of table 2. This is our preferred specification. Taxation now has a negative effect on employment. As argued above, the difference between the spatial and non-spatial results suggests that unobserved time-varying site-specific effects are correlated with both employment and local political variables. Spatial differencing removes these site-specific effects, ensuring that our instruments are valid and thus allowing us to identify the negative effect of taxes on establishment employment.

To assess the robustness of our results, table 3 reports results for alternative sets of instruments. In our preferred specification, reported in column 3 of table 2 we take "no overall control" as the omitted control dummy. An alternative would be to exclude one of the main political parties allowing the coefficients on the other parties to be directly compared to those for the omitted party. Unfortunately, this introduces an additional complication because there is then no "share" variable that corresponds to no overall control. It turns out that our results are not very sensitive to this as shown in table 3. In column 1 of table 3 we repeat our preferred specification taking Labour as the omitted category for both control and share. In column 2 we omit Conservatives. Both columns leave our main result virtually unchanged. In column 3, we restrict our set of instruments to the control dummies. This yields a coefficient on the tax rate that is larger in magnitude. Because of larger standard errors, it is statistically undistinguishable from that in our preferred specification. Finally, in column 4 we use only the shares as instruments. This yields results that are similar to those of our preferred specification.

Table 3: Alternative instrumentation strategies, WITHIN IV

	Labour omitted	Conservative omitted	only party control	only shares
(log) tax rate	-0.953[a]	-0.865[a]	-3.240[c]	-0.822[a]
	(0.319)	(0.317)	(1.765)	(0.338)
Number of observations	18370	18370	18370	18370
Number of pairs	6087	6087	6087	6087

Notes: Regression of (log) employment on (log) local tax rates and age variables. Coefficients for the age censored dummy, age, and age squared not reported. All four columns report WITHIN IV results. In column 1, the instruments are as in column 3 of table 2 except that Labour is the omitted category. Column 2 duplicates regression using Conservative as the omitted variable. Column 3 only uses party control as instruments. Column 4 only uses party shares as instruments. Standard errors (uncorrected) under coefficients. [a], [b] and [c] denote significance at the 1%, 5%, and 10% level respectively.

Table 4 duplicates the three regressions of table 2 for different distance thresholds. In the first block of the table, the estimated coefficient on the tax rate for a threshold of 0.5 km (instead of 1 km) is statistically insignificant in all three regressions. This is unsurprising because this coefficient is estimated from only 83 cross border pairs instead of 298 when using a 1 km threshold. In the second block, the results for a 1.5 km threshold are close to those of table 2 except that the coefficient on the tax rate estimated with spatial within IV is still negative but lower and marginally insignificant. In the third block, at the 2 km threshold, the point estimate for this coefficient is close to zero. In the fourth block, at the 3 km threshold, it becomes positive. This gradual increase suggests that as we increase the distance threshold there is a smooth transition from our coefficient of -1.02 obtained with a 1 km threshold to the coefficient of +0.36 obtained without spatially differentiating. At the same time, the within coefficient in the second column remains stable and does not appear to depend on our sample selection rule. The differences between these two series of regressions are thus driven by the instrumentation itself and appear consistent with our conjecture above that the exclusion restriction is more likely to hold for short distances.

Pulling the results together, spatial differencing offers two improvements over existing methodologies. First, comparing the non-instrumented regressions with and without spatial differencing allows us to identify the nature of the relationship between site-specific effects and local taxation. Second, and more importantly, because spatial differencing removes unobserved time-varying site-specific effects it makes it far easier to find valid instruments that allows us to identify the negative relationship between local taxation and employment. Quantitatively, spatial differencing greatly affects the results. In what is arguably the best estimation using standard techniques (that does not control for time-varying site-specific effects), we find a positive elasticity around 0.4. Our preferred specification with spatial differencing reverses the sign of the coefficient and leads to an elasticity around -1 (although not very precisely estimated). In a nutshell, instead of a positive effect of taxation on employment spatial differencing shows a negative effect of taxation on employment. From our model we know that this occurs because of some combination of the growth-slowing and selection effects of higher taxes. Focusing on entries allows us to focus specifically on the second of these effects since we expect firms that leave a jurisdiction to be replaced by new 'entrants' in the data. We now turn to this issue.

Table 4: Spatial differencing regression results for alternative distance thresholds

spatial difference of	OLS	WITHIN	WITHIN IV
0.5 kilometre threshold			
(log) tax rate	-0.755	0.172	0.224
	(0.542)	(0.265)	(0.541)
Adjusted R-squared	0.03		
Number of observations	9530	9530	9530
Number of establishments	3226	3226	3226
Number of pairs	3112	3112	3112
1.5 kilometre threshold			
(log) tax rate	0.582[b]	0.106	-0.454
	(0.273)	(0.131)	(0.288)
Adjusted R-squared	0.04		
Number of observations	35762	35762	35762
Number of establishments	6386	6386	6386
Number of pairs	11753	11753	11753
2 kilometre threshold			
(log) tax rate	0.426[c]	0.156	-0.040
	(0.231)	(0.108)	(0.249)
Adjusted R-squared	0.04		
Number of observations	49578	49578	49578
Number of establishments	7273	7273	7273
Number of pairs	16654	16654	16654
3 kilometre threshold			
(log) tax rate	0.448[b]	0.131	0.117
	(0.188)	(0.088)	(0.198)
Adjusted R-squared	0.04		
Number of observations	81211	81211	81211
Number of establishments	8518	8518	8518
Number of pairs	26661	26661	26661

Notes: Regression of spatial difference of (log) employment on spatial difference of (log) local tax rates and age variables. Coefficients for the age censored dummy, age, and age squared not reported. First column reports results from OLS, second column (WITHIN) allows for establishment specific fixed effects, third column (WITHIN IV) further instruments local tax using local political variables. Corrected standard errors (according to Appendix A) under coefficients. First block reports results for 0.5 km threshold, second block for 1.5 km, third block for 2 km, and fourth block for 3 km. [a], [b] and [c] denote significance at the 1%, 5% and 10% level respectively.

5. Entries

A. *Methodology and data*

We now turn to the effects of local taxation on entry.[29] This is an important issue for three reasons. First, the rate at which new establishments enter LAs is an important outcome that deserves attention. Second, our employment estimates confound selection and growth slow-down effects as made clear by our model. Assuming the supply of sites is fixed and that the land market clears so that exits are matched with entries, an analysis of entries offers an opportunity to focus specifically on selection effects. Third spatial differencing provides solutions to the same three problems (establishment heterogeneity, site heterogeneity, and endogeneity) that we addressed when looking at employment growth. However, there are subtle differences which are worth highlighting.

Consider establishment i that wishes to enter in year t. It can choose between all available sites, indexed by z. As before, the jurisdiction that sets the tax for the establishment depends on the site occupied, and is indexed by a. Profit maximisation can be performed in two stages. First, establishment i computes the highest profit it can achieve, Π_{iz}, at each site z. It then selects the site offering the highest profit.[30] We assume that the highest profit for establishment i entering in year t at site z can be written as:

$$\Pi_{izt} = \lambda r_{at} + Z_{it}\zeta + \nu_i + \kappa_a + \varphi_{zt} + \epsilon_{izt} \tag{9}$$

where ν_i is an establishment fixed effect, κ_a is a jurisdiction fixed effect, Z_{it} are explanatory variables at the establishment level, φ_{zt} is a site-specific effect, and ϵ_{izt} is an establishment site-specific shock.[31] Establishment i will choose the site z that gives the highest expected

[29] Rathelot and Sillard (2008) develop an analysis of the effect of a local tax on capital on entries using a similar approach. Details of the implementation differ.

[30] We ignore any possible interaction between the location decisions of entrants. This seems reasonable in established manufacturing industries where existing establishments drive local wages, determine product market competition etc. As discussed in the text, the fact that the effect on profits of these factors will be highly correlated across neighbouring sites then justifies our approach.

[31] The establishment fixed effect, ν_i, mirrors μ_i in (4). Similarly the jurisdiction effect, κ_a, and the site-specific effect, φ_{zt}, are the counterparts of γ_a and θ_{zt}. Finally both (9) and (4) contain coefficients for the effect of local taxation (λ and α) and establishment-level variables (ζ and β). Note that our approach for entries is also consistent with a more general setting where the specification includes variables that are individual-site-specific, ϕ_{izt}, in addition to individual effects, ν_i, and site effects, φ_{zt}.

profit. When the shocks ϵ_{izt} follow an appropriate *iid* extreme value distribution, the probability of choosing site z, P_{izt}, is logistic and is given by

$$P_{izt} = \frac{\exp E(\Pi_{izt})}{\sum_{z=1}^{Z} \exp E(\Pi_{izt})} \tag{10}$$

where $E(\cdot)$ is the expectation operator and the summation is across all possible sites Z.

The standard approach to estimating the coefficients λ and ζ is to ignore the site-specific effect, φ_{zt}, and estimate a conditional logit model. To do this, one creates a set of establishment-jurisdiction observations and defines $c_{ia} = 1$ if establishment i locates in jurisdiction a and $c_{ia} = 0$ otherwise. The coefficients can then be estimated by maximising the log likelihood of the conditional logit model:

$$\log L_{cl} = \sum_{i=1}^{I} \sum_{a=1}^{A} c_{ia} \log P_{ia} \tag{11}$$

where for simplicity we drop the time subscripts as establishments only enter once.

As is well recognised, application of the conditional logit model can be problematic when the set of possible jurisdictions is large. One solution is to take a random sub-sample of jurisdictions, although this has implications for the efficiency of the estimator and the small sample properties are unknown. Another possibility, recently proposed by Guimaraes, Figueiredo, and Woodward (2003) is to use the fact that, under certain conditions, the log likelihood of the Poisson model is identical to that of the conditional logit. Estimating a Poisson regression is computationally much easier, though the equivalence between the likelihoods only holds in the absence of establishment specific variables (i.e., Z_{it}). In any case both solutions ignore the site-specific effects. As we now show, spatial differencing provides an alternative which controls for site-specific effects and which, in other contexts, would allow for the inclusion of establishment specific variables.

Our approach is as follows. Consider two neighbouring sites, z_1 and z_2, close to the border between two jurisdictions a_1 and a_2. z_1 is located in jurisdiction a_1 and z_2 in a_2. Since the two locations are close, we assume $\varphi_{z_1 t} \approx \varphi_{z_2 t}$. This is the same identification assumption made in section 2 to derive our employment specification (8). To repeat, this assumption is justified by the fact that site-specific effects (labour market conditions, access to markets and major facilities, etc) vary smoothly across space. The probability of choosing z_1 conditional on locating in one of these two neighbouring sites is:

$$P\left(i \in z_1 \,|i \in \{z_1,z_2\}\right) = \frac{P_{iz_1}}{P_{iz_1} + P_{iz_2}} \tag{12}$$

When the shocks ϵ_{izt} follow an appropriate *iid* extreme value distribution, the probability of choosing one of the sites is logistic and is given by:

$$P\left(i \in z_1 \,|i \in \{z_1,z_2\}\right) = \frac{1}{1 + \exp(\lambda\left(r_{a_2t} - r_{a_1t}\right) + \kappa_{a_2} - \kappa_{a_1})} \tag{13}$$

Note that unlike equation (10) above, this specification conditions out both establishment- and site-specific effects because these effects are the same at locations z_1 and z_2.[32] Recall that in standard conditional logit models observed site-specific factors are computationally hard to deal with. By contrast our approach directly conditions out both observed and unobserved site-specific factors in a way that is easy to implement. In addition, we do not need to rely on the assumption of the Independence of Irrelevant Alternatives which underlies the conditional logit model. This is a distinct advantage as this assumption is unlikely to hold in spatial settings (see, for example, Head and Mayer, 2004).

Equation (13) only involves jurisdictional level variables so we can estimate λ directly from an aggregate logit model where the observation units are the (border-side, time) pairs.[33] In the estimation, the observations are weighted by the number of entrants for consistency with equation (13).

In a nutshell, we select entrants located close to jurisdictional boundaries and examine their decision to choose to locate on one particular side of the border. The main conceptual difference with our employment regression (8) is that we consider the location decision of a new establishment choosing between neighbouring *jurisdictions* rather than comparing the employment outcomes for (existing) neighbouring establishments. For entry, this is

[32] We assumed in (9) that the effect of establishment characteristics did not depend on location. However our spatial differentiation approach is more powerful than this since any interaction between establishment and site characteristics is conditioned out by spatial differencing provided the characteristics of the local environment vary smoothly over space.

[33] We need to estimate 366 jurisdiction dummies. These dummies are identified from the time variation in entry rates at the borders.

the appropriate way to control for both time invariant establishment specific effects and any unobserved site-specific effects common to both sides of a border.[34] As local tax rates do not vary smoothly across space at jurisdictional boundaries, we can use entrants on either side of these boundaries to identify the effect of taxes after conditioning out time-varying site-specific factors and time-invariant establishment-specific effects. Interestingly one step is enough to eliminate both establishment and site effects for entry, whereas two steps are necessary when studying employment. This reflects the fact that instead of comparing each establishment with a matched establishment on the other side of the border we compare each establishment with itself on the two sides of the border making the *within* transformation redundant. In passing, we note that the same idea cannot be used for exits.[35]

As with the employment specifications, endogeneity must also be addressed since the local tax rate may be simultaneously determined with the rate of entry. To control for this, we estimate a two-stage IV *logit* model instrumenting the difference in tax rate, $r_{a_2 t} - r_{a_1 t}$, in the logit specification, with the predicted difference in tax rate from a first stage regression using spatially-differenced local political variables. The simplicity of this method comes at a price because correcting the errors to allow for possible correlation between the residual of the instrumentation equation and that of the entry equation is not straightforward.[36]

[34]To see this, note that when choosing between two neighbouring sites, the entrant compares profits between them. Time-varying site-specific factors, which vary smoothly across space, will affect profits at both sites in the same way. Hence, these factors do not enter into the location decision. A similar argument applies to unobserved time-invariant establishment characteristics.

[35]Indeed, we can observe the exit of an establishment only on the side of the border where it entered. Hence, unlike with entry, we cannot compare an establishment with itself on both sides of a border. Alternatively, one might want to apply the same methodology as for employment by matching establishments with their closest neighbour(s). However, since establishments exit only once, we would not be able to control for establishment unobserved heterogeneity.

[36]As an alternative, we used an IV *probit* model that estimates the entry and instrumenting equation simultaneously. This approach corrects standard errors but at the cost that it is not fully consistent with the theoretical specification. The results are the same as with the two-stage *logit* model.

B. ***Results***

To construct the entry data, we need to detect all entrants in the ARD. Because of a change in 1984 in the way the registry of establishments was constructed, there is a large amount of artificial entry in 1984 and 1985 (i.e., establishments enter the data set for the first time during those two years even though they already existed prior to 1984). See Griffith (1999) for further discussion. As a result, we ignore entries for these two years and focus instead on 81,042 newly reporting establishments between 1986 and 1989.

For consistency with the employment regressions we would like to identify all entrants within 1 kilometre of jurisdiction boundaries. The easiest way to do this would be to draw 1 kilometre buffers around boundaries. Lacking a set of digital boundaries for this time period, we instead proceed indirectly and identify the set of border entrants from the ARD itself. To do this, for each entrant we searched for the closest establishment located in each of the neighbouring LAs and retained only those entrants that had such a neighbour within one kilometre. Since this detection procedure is only meant to compute distances to the LA border for each entrant (rather than find a match for a pair), we considered all possible establishments in all sectors and all years as potential neighbours. We expect this procedure to catch nearly all entrants located within one kilometre of a border.

As for employment, we implement both the standard methodology (i.e., conditional logit) and our spatial differencing approach. Results for two non-spatial specifications, are given in columns 1 and 2 of table 5 while columns 3 and 4 report results for two comparable spatial specifications. Again, for the sake of comparison, we have restricted the sample to be the same for all specifications. This restriction requires establishments to locate within 1 kilometre of a boundary between two English LAs between 1986 and 1989. Imposing this requirement leaves us with a sample of 19,337 establishments. To show that this sample is representative, Appendix B once again reports results for the same specifications but without imposing this restriction.

Starting with the conditional logit we see, from the results reported in column 1, that there appears to be a positive effect of tax rates on entry. Column 2 shows what happens

Table 5: Results for entries

	CL	CL IV	LOGIT	LOGIT IV
(log) tax rate	0.397[a]	0.521	0.108	0.809
	(0.079)	(0.883)	(0.177)	(0.921)
Number of entrants	19,337	19,337	19,337	19,337

Notes: Number of entrants as a function of local tax rates. First column (CL) reports results for conditional logit, second column (CL IV) instruments local taxes using political variables, third column (LOGIT) reports results from a logit model for spatially differenced variables; fourth column (LOGIT IV) instruments local taxes. Standard errors under coefficients. [a], denotes significance at the 1% level. Estimates in the first two columns are from a Poisson regression using the equivalence result from Guimaraes *et al.* (2003).

when we correct for endogeneity by replacing the actual tax rates with the predicted tax rates from a first stage regression of tax rates on political variables. Once instrumenting, we find that the coefficient increases but it also becomes insignificant because of a much higher standard error. Correcting the standard errors to allow for the fact that we are instrumenting would only reinforce this finding.[37] Columns 3 and 4 show that we reach the same conclusions for entry using our spatially differenced approach.

A positive effect of local taxation on entries (although insignificant) might seem surprising in light of its negative effect on employment. Nonetheless, having a positive effect of taxation on entries is consistent with our model. Recall that higher taxes lead to more exits and relocations. This creates vacant sites. In turn, these vacant sites are occupied by new entrants. Overall, our findings about entries are suggestive that the selection effect highlighted by the model plays a role. This selection effect might be more complicated than in our model. In particular, it could be that exiting and relocating establishments in high tax jurisdictions are large and capital intensive. They might be replaced by smaller and less capital intensive firms which are less sensitive to high local taxes. We also note that our model assumes that the supply of sites is fixed. The development of new sites might be negatively affected by high local taxes (a force countering our selection effect).[38]

[37]Hence our decision only to implement the theoretically consistent two stage conditional logit procedure rather than an instrumented probit specification.

[38]We could also imagine that jurisdictions where more new sites can be developed raise taxes to maximise tax revenue. This could lead to a positive correlation between new developments and taxes. This should normally be corrected by our instrumentation strategy.

6. Conclusion

We propose a new approach to assess the effects of local taxation. Our results show the importance of controlling for both unobserved establishment-specific and unobserved site-specific characteristics and possible simultaneity. Simple OLS results suggest a positive relationship between employment and taxes. Allowing for unobserved establishment-specific effects and instrumenting for local taxation, we still find a positive relationship between employment and taxes. Allowing for unobserved location-specific effects and instrumenting for local taxation, we find a negative significant relationship between employment and taxes. By contrast we find that local taxation has no effect on the entry of new establishments.

Beyond our methodological contribution, this analysis also suggests that the study of local taxation and, more broadly, that of decentralised public intervention faces serious endogeneity problems whereby local public decisions depend strongly on very local conditions, which are extremely difficult to control for. As shown here, properly controlling for such local conditions is a necessary condition to obtain reliable estimates.

The second broad lesson is that even taxes that are seemingly close to an 'ideal' tax that would be free of distortion can in practice generate significant distortions. Despite the fact that the UK business rates were close to George's (1884) 'pure' land tax, revaluations in case of expansions and frictions in the rental market implied that increases in local taxation had an adverse effect on employment.

References

Arnott, Richard. 2004. Does the Henry George Theorem provide a practical guide to optimal city size? *American Journal of Economics and Sociology* 63(5):1057–1090.

Baldwin, Richard E., Rikard Forslid, Philippe Martin, Gianmarco I. P. Ottaviano, and Frederic Robert-Nicoud. 2004. *Economic Geography and Public Policy*. New Jersey: Princeton University Press.

Bartik, Timothy. 1991. *Who benefits from state and local economic development policies?* Kalamazoo, Michigan: W. E. Upjohn Institute for Employment Research.

Bennett, Robert. 1986. The impact of non-domestic rates on profitability and investment. *Fiscal Studies* 7(1):34–50.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics* 119(1):249–275.

Black, Sandra E. 1999. Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics* 114(2):587–599.

Bond, Stephen, Kevin Denny, John Hall, and William McCluskey. 1996. Who pays business rates? *Fiscal Studies* 17(1):19–35.

Chirinko, Robert S. and Daniel J. Wilson. 2008. State investment tax incentives: A zero-sum game? *Journal of Public Economics* 92(12):2362–2384.

Crosby, Neil, Colin Lizieri, Sandi Murdoch, and Charles Ward. 1998. Implications of changing lease structures on the pricing of lease contracts. The cutting edge, Royal Institution of Chartered Surveyors.

Duranton, Gilles and Henry G. Overman. 2005. Testing for localization using micro-geographic data. *Review of Economic Studies* 72(4):1077–1106.

Epple, Dennis and Thomas Nechyba. 2004. Fiscal decentralization. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland.

Fujita, Masahisa. 1989. *Urban Economic Theory: Land Use and City Size*. Cambridge: Cambridge University Press.

George, Henry. 1884. *Progress and Poverty: An Inquiry into Causes of Industrial Depressions, and of Increase of Want with Increase of Wealth. The Remedy*. London: W. Reeves.

Gibbons, Steve and Stephen Machin. 2003. Valuing english primary schools. *Journal of Urban Economics* 53(2):197–219.

Griffith, Rachel. 1999. Using the ARD establishment level data: an application to estimating production functions. *Economic Journal* 109(456):F416–F442.

Guimaraes, Paulo, Octavio Figueiredo, and Douglas Woodward. 2003. A tractable approach to the firm location decision problem. *Review of Economics and Statistics* 85(1):201–204.

Guimaraes, Paulo, Octavio Figueiredo, and Douglas Woodward. 2004. Industrial location modeling: Extending the random utility framework. *Journal of Regional Science* 44(1):1–20.

Hale, Rita and Tony Travers. 1994. *A National or Local Tax: A Study of the Non Domestic Rate*. London: The Chartered Institute of Public Finance and Accountancy.

Head, Keith and Thierry Mayer. 2004. Market potential and the location of Japanese investment in Europe. *Review of Economics and Statistics* 86(4):959–972.

Holmes, Thomas J. 1998. The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy* 106(4):667–705.

Kahn, Matthew E. 2004. Domestic pollution havens: Evidence from cancer deaths in border counties. *Journal of Urban Economics* 56(1):51–69.

Moulton, Brent R. 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics* 72(2):334–338.

Rallings, C. and M. Thrasher (eds.) . 2004. *Local Elections in Britain: a statistical digest*. Plymouth: Local Government Chronicle Elections Centre.

Rathelot, Roland and Patrick Sillard. 2008. The importance of local corporate taxes in business location decisions: Evidence from French micro data. *Economic Journal* 118(527):499–514.

Thompson, Robert and Sotiris Tsolacos. 2001. Industrial land values – a guide to future markets? *Journal of Real Estate Research* 21(1/2):55–76.

Tiebout, Charles M. 1956. A pure theory or local expenditures. *Journal of Political Economy* 64(5):416–424.

Wellisch, Dietmar. 2000. *Theory of Public Finance in a Federal State*. Cambridge UK: Cambridge University Press.

Wilson, John D. 1999. Theories of tax competition. *National Tax Journal* 52(2):269–304.

Appendix A. Correction of the standard errors

When spatially differencing, an establishment i that has n neighbours will be in n pairs. This induces correlation in the error for all n of these pairs. The correlation arises because ϵ_{it} (the error of establishment i) enters the error of each pair. This imposes a particular structure to the covariance matrix which we use to correct the standard error. This appendix gives the details of that correction. Note that Bertrand, Duflo, and Mullainathan (2004) consider similar issues when proposing their correction for the standard errors of difference-in-difference estimators to measure treatment effects. There are several key differences between our correction and the one proposed there. First, their correction is difficult to apply to specifications with a large number of establishments and locations because they estimate the treatment effect directly without rewriting the model in difference-in-difference. Second, their correction requires the covariance matrix to be block diagonal. This means that it is not applicable to situations, like ours, where there is no obvious way to construct closed sets of neighbours (because establishment A may be a neighbour to establishment B, who may be neighbour to C, etc). In short we deal with an error structure which is considerably more complex. This comes at a cost: we ignore issues arising from serial correlation of the errors that are the key concern of that paper.

After spatial-differentiation and within-pair projection, the model can be re-written as follows:

$$\begin{aligned} W\Delta e &= \alpha W\Delta r + W\Delta X\beta + W\Delta\epsilon & \text{(A1)} \\ &= Z\gamma + W\Delta\epsilon & \text{(A2)} \end{aligned}$$

where for any variable v_{it}, observations have been stacked in *pair* and time order, $\gamma' = (\alpha,\beta')'$ and $Z = (W\Delta r, W\Delta X)$. The OLS estimator is then:

$$\widehat{\gamma} = (Z'Z)^{-1} Z'\Delta e = \gamma + (Z'Z)^{-1} Z'\Delta\epsilon \qquad \text{(A3)}$$

We suppose that the residuals ϵ_{it} are *iid* with variance σ^2. The variance of the OLS estimator is:

$$V(\widehat{\gamma}) = \sigma^2 ABA \qquad \text{(A4)}$$

where $A = (Z'Z)^{-1}$ and $B = Z'\Delta\Delta'Z$. Matrix A is easy to compute since Z can be obtained after spatial-differenciation and a projection *within*-pair of the explanatory variables. Matrix B can also be computed after using an algorithm to obtain $\Delta'Z$. This algorithm relies on the fact that Δ has a simple structure. Indeed, denote $p \in \{1,...,P\}$ where P is the number of pairs and N_p the number of years that pair p appears in the data. We can decompose Δ in blocks such that $\Delta = (\Delta'_1,..., \Delta'_P)'$ where, for instance, block Δ_p writes:

$$\begin{pmatrix} ... & 0 & 1 & 0 & ... & 0 & -1 & 0 & & ... & ... \\ ... & 0 & 0 & 1 & 0 & ... & 0 & & -1 & 0 & ... \\ ... & 0 & 0 & 0 & 1 & 0 & ... & 0 & & -1 & ... \end{pmatrix} \tag{A5}$$

(supposing $N_p = 3$). The first line corresponds to the first year that the pair is in the data, the second line to the second year, etc... Each column of Δ can contain the values 1 and -1 several times depending on the number of times an establishment has been matched with neighbours in the corresponding year. For a column i of Δ, denote j_i the identifier of the corresponding establishment and t_i the corresponding date. If *establishment* is the vector containing all the establishment identifiers and *year* is the vector containing the years, j_i and t_i can be retrieved from the first element in column i that takes a value different from 0. For example, in the GAUSS language, column i of Δ is then of the form $(establishment\ .eq\ establishment\,[j_i]).*(year\ .eq\ year[t_i]) - (establishment_n\ .eq\ establishment\,[j_i]).*(year_n\ .eq\ year[t_i])$ where $establishment_n$ contains the identifier of the neighbouring establishments and $year_n$ their years. Element (n,k) of $\Delta'Z$ can be computed using column n of Δ and column k of Z. The whole matrix $\Delta'Z$ is obtained from a loop over n and k.

We now propose an estimator of σ^2. Denote $\widehat{W\Delta\epsilon}$ the vector of residuals from the OLS estimation. We have:

$$\widehat{W\Delta\epsilon} = W\Delta e - Z\widehat{\gamma} = M_Z W\Delta\epsilon \tag{A6}$$

where M_Z is the projector in the dimension orthogonal to Z. We then get:

$$\widehat{W\Delta\epsilon}'\widehat{W\Delta\epsilon} = \epsilon'\Delta' W M_Z W\Delta\epsilon \tag{A7}$$

From this formula, we obtain:

$$\begin{aligned}
E\left(\widehat{W\Delta\epsilon}'\widehat{W\Delta\epsilon}\right) &= Etr\left(\widehat{W\Delta\epsilon}'\widehat{W\Delta\epsilon}\right) && \text{(A8)}\\
&= \sigma^2 tr\left(\Delta' W\Delta\right) - \sigma^2 tr\left[Z\left(Z'Z\right)^{-1} Z'\Delta\Delta' W\right] && \text{(A9)}\\
&= \sigma^2 tr\left(W\Delta\Delta'\right) - \sigma^2 tr\left(AB\right) && \text{(A10)}
\end{aligned}$$

We can recover $tr\left(AB\right)$ very easily from A and B. It is also possible to simplify the expression: $tr\left(W\Delta\Delta'\right)$. We can write $\Delta\Delta'$ in blocks corresponding to pairs. Indeed, the (p,q)-block writes: $\Delta_p\Delta'_q$. W is block diagonal. Thus, the (p,q)-block of $W\Delta\Delta'$ writes $W_p\Delta_p\Delta'_q$ where W_p is the (p,p)-block of W. Hence, we get: $tr\left(W\Delta\Delta'\right) = \sum_p tr\left(W_p\Delta_p\Delta'_p\right)$. As we have $\Delta_p\Delta'_p = 2I_{T_p}$ and $tr\left(W_p\right) = T_p - 1$ (where T_p is the number of years that pair p appears in the data), we finally get: $tr\left(W\Delta\Delta'\right) = 2\left(N-P\right)$. An unbiased *(and consistent)* estimator of σ^2 is then:

$$\widehat{\sigma}^2 = \frac{1}{2\left(N-P\right) - tr\left(AB\right)}\widehat{W\Delta\epsilon}'\widehat{W\Delta\epsilon} \tag{A11}$$

We can finally deduce an estimator of the variance of $\widehat{\gamma}$:

$$\widehat{V}\left(\widehat{\gamma}\right) = \widehat{\sigma}^2 ABA \tag{A12}$$

We now compute the standard errors when instrumenting. The model is:

$$\begin{aligned}
W\Delta e &= \alpha W\Delta r + W\Delta X\beta + W\Delta\epsilon && \text{(A13)}\\
W\Delta r &= Y\delta + \xi && \text{(A14)}
\end{aligned}$$

with $Y = W\Delta P$ where P are some political variables, and $cov\left(Y,\xi\right) = cov\left(Y,W\Delta\epsilon\right) = 0$ by assumption. Denote $\widehat{\delta}$ the OLS estimator of δ obtained from equation (A14) and $V = \widehat{V}\left(\widehat{\delta}\,|X\right)$ an estimator of its covariance. This covariance estimator may simply be the usual OLS estimator. It may also take into account clusters at the jurisdiction level. Equation (A13) rewrites:

$$\begin{aligned}
W\Delta e &= \alpha Y\widehat{\delta} + W\Delta X\beta + W\Delta\epsilon + \phi && \text{(A15)}\\
&= \widetilde{Z}\gamma + W\Delta\epsilon + \phi && \text{(A16)}
\end{aligned}$$

with $\widetilde{Z} = \left(Y\widehat{\delta}, W\Delta X\right)$ and $\phi = \alpha\left(Y\delta - Y\widehat{\delta}\right)$. ϕ is such that $E\left(\phi\,|X,Y\right) = 0$ and $V\left(\phi\,|X,Y\right) = \alpha^2 YVY'$ with $V = V\left(\widehat{\delta}\,|Y\right)$. The IV estimator is:

$$\widehat{\gamma}_{IV} = \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\widetilde{Z}'W\Delta e = \gamma + \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\widetilde{Z}'\left(W\Delta\epsilon + \phi\right) \tag{A17}$$

Assuming that $\widehat{\delta}$ is known, the variance of the IV estimator can be approximated by:

$$\begin{aligned} V\left(\widehat{\gamma}_{IV}\right) &\approx \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\widetilde{Z}'\left(\sigma^2 W\Delta\Delta'W + \alpha^2 YVY'\right)\widetilde{Z}\left(\widetilde{Z}'\widetilde{Z}\right)^{-1} && \text{(A18)} \\ &\approx \widetilde{A}\left(\sigma^2\widetilde{B} + \alpha^2\widetilde{C}\right)\widetilde{A} && \text{(A19)} \end{aligned}$$

with $\widetilde{A} = \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}$, $\widetilde{B} = \widetilde{Z}'W\Delta\Delta'W\widetilde{Z}$ and $\widetilde{C} = \widetilde{Z}'YVY'\widetilde{Z}$. $\widetilde{A}$ and $\widetilde{B}$ are easy to compute (see the OLS case). $\widetilde{C}$ is also easy to compute since we can first compute $Y'\widetilde{Z}$, which has a small dimension. We now propose an estimator of σ^2. Denote $\widehat{W\Delta\epsilon + \phi}$ the vector of residuals from the IV second-stage estimation. We have:

$$\widehat{W\Delta\epsilon + \phi}'\,\widehat{W\Delta\epsilon + \phi} = \left(W\Delta\epsilon + \phi\right)' M_{\widetilde{Z}}\left(W\Delta\epsilon + \phi\right) \tag{A20}$$

Consider for a while that Z is non random (i.e., $\widehat{\delta}$ is non random). We have:

$$\begin{aligned} E(\widehat{W\Delta\epsilon + \phi}'\,\widehat{W\Delta\epsilon + \phi}) &= trE\left(\epsilon'\Delta'WM_{\widetilde{Z}}W\Delta\epsilon\right) + trE\left(\phi'M_{\widetilde{Z}}\phi\right) && \text{(A21)} \\ &= \sigma^2\left[2(N-P) - tr(\widetilde{A}\widetilde{B})\right] + \alpha^2\left[tr(VY'Y) - tr(\widetilde{A}\widetilde{C})\right] && \text{(A22)} \end{aligned}$$

A *(consistent)* estimator of σ^2 is then:

$$\widehat{\sigma}^2_{IV} = \frac{\widehat{W\Delta\epsilon + \phi}'\,\widehat{W\Delta\epsilon + \phi} - \widehat{\alpha}^2_{IV}\left[tr\left(\widehat{V}Y'Y\right) - tr\left(\widetilde{A}\widehat{\widetilde{C}}\right)\right]}{2\left(N-P\right) - tr\left(\widetilde{A}\widetilde{B}\right)} \tag{A23}$$

with $\widehat{V}$ an estimator of V obtained from the first-stage equation, $\widehat{\widetilde{C}} = \widetilde{Z}'Y\widehat{V}Y'\widetilde{Z}$. Note that when the residuals ξ_{it} are *iid* with variance θ^2, we have $tr\left(\widehat{V}Y'Y\right) = N\widehat{\theta}^2$ and $tr\left(\widetilde{A}\widehat{\widetilde{C}}\right) = \widetilde{\theta}^2 tr\left(P_{\widetilde{Z}}P_Y\right)$. The estimator of σ^2 becomes:

$$\widehat{\sigma}^2_{IV} = \frac{\widehat{W\Delta\epsilon + \phi}'\,\widehat{W\Delta\epsilon + \phi} - \widehat{\alpha}^2_{IV}\widehat{\theta}^2\left[N - tr\left(P_{\widetilde{Z}}P_Y\right)\right]}{2\left(N-P\right) - tr\left(\widetilde{A}\widetilde{B}\right)} \tag{A24}$$

Finally, the variance of the IV estimator can be approximated by:

$$\widehat{V}\left(\widehat{\gamma}_{IV}\right) = \widetilde{A}\left(\widehat{\sigma}^2_{IV}\widetilde{B} + \widehat{\alpha}^2_{IV}\widehat{\widetilde{C}}\right)\widetilde{A} \tag{A25}$$

Appendix B. Non-spatial results for different samples of establishments

Tables 6, 7, and 8 present results for the non-spatial employment specifications using less restrictive samples than those used in the text. These results should be compared to those in table 1, where establishments must be (i) in a pair with a matched establishment within one kilometre and this pair must be such that (ii) both establishments simultaneously report employment in at least two years (the sample used for our preferred specification). Table 6 uses a 2 kilometre cutoff (instead of 1 km) to pair establishments. That is, we relax restriction (i) and retain restriction (ii). Table 7 uses the largest possible sample of establishments that are part of a pair. That is, we impose restriction (i) but not restriction (ii). Table 8 uses the largest possible sample for each specification imposing neither restriction (i) or (ii). As can be seen from the comparison of tables 1, 6, 7, and 8 all point estimates are of the same sign and do not significantly differ from each other.

Table 6: Non-spatial regression results for all establishments with a 2 km pairing cutoff

	OLS	WITHIN	WITHIN IV
(log) tax rate	0.139^a	0.117^a	0.337^a
	(0.047)	(0.023)	(0.078)
age censored dummy	0.584^a	0.295^a	0.300^a
	(0.070)	(0.025)	(0.025)
age	0.038	0.057^a	0.059^a
	(0.037)	(0.008)	(0.008)
age squared	0.003	-0.003^a	-0.003^a
	(0.004)	(0.001)	(0.001)
Adjusted R-squared	0.13		
Number of observations	22387	22387	22387
Number of establishments	7273	7273	7273

Notes: Standard errors under coefficients. ^{a}, ^{b} and ^{c} denote significance at the 1%, 5% and 10% level respectively. Sample restricted as in table 1 except for the use of a two kilometre cutoff (instead of 1 km).

Table 7: Non-spatial regression results for largest possible samples of establishments in pairs

	OLS	WITHIN	WITHIN IV
(log) tax rate	0.156^a	0.105^a	0.472^a
	(0.042)	(0.023)	(0.080)
age censored dummy	0.662^a	0.208^a	0.211^a
	(0.054)	(0.025)	(0.025)
age	0.004	0.046^a	0.047^a
	(0.030)	(0.008)	(0.008)
age squared	0.008^b	-0.002^a	-0.003^a
	(0.003)	(0.001)	(0.001)
Adjusted R-squared	0.13		
Number of observations	25579	22803	22803
Number of establishments	5564	5852	5852

Notes: Standard errors under coefficients. ^{a}, ^{b} and ^{c} denote significance at the 1%, 5% and 10% level respectively. Sample restricted to establishments that are part of a pair. First column (OLS) only restricted by data availability. Second and third column require at least two observations per establishment. Compare to table 1 where establishments must be (i) in a pair in which (ii) both establishments simultaneously report employment in at least two years.

Table 8: Non-spatial regression results for largest possible samples

	OLS	WITHIN	WITHIN IV
(log) tax rate	0.222^a	0.131^a	0.481^a
	(0.029)	(0.016)	(0.062)
age censored dummy	0.652^a	0.290^a	0.294^a
	(0.033)	(0.016)	(0.015)
age	-0.015	0.053^a	0.055^a
	(0.018)	(0.005)	(0.005)
age squared	0.011^a	-0.002^a	-0.003^a
	(0.002)	(0.001)	(0.001)
Adjusted R-squared	0.13		
Number of observations	61785	53684	53684
Number of establishments	21813	13875	13875

Notes: Standard errors under coefficients. ^{a}, ^{b} and ^{c} denote significance at the 1%, 5% and 10% level respectively. Largest possible sample. First column (OLS) only restricted by data availability. Second and third column require at least two observations per establishment. Compare to table 1 where establishments must be (i) in a pair in which (ii) both establishments simultaneously report employment in at least two years.

Table 9 present results for the non-spatial entry specifications using all entrants. These results should be compared to those in table 5, where establishments must enter within 1 kilometre of an LA boundary to be part of the sample. All point estimates are of the same sign and do not significantly differ from each other.

Table 9: Non-spatial regression results for largest possible sample

	CL	CL IV
(log) tax rate	0.633[a]	0.267
	(0.070)	(0.244)
Number of establishments	81,042	81,042

Notes: Number of entrants as a function of local tax rates. First column (CL) reports results for conditional logit, second column (CL IV) instruments local taxes using political variables. Largest possible sample. Standard errors under coefficients. [a], denotes significance at the 1% level. Estimates are from a Poisson regression using the equivalence result from Guimaraes *et al.* (2003).

Appendix C. First stage regression

Table 10 presents the results from the first stage regression of spatially difference (log) tax rates on the exogenous variables and spatially differenced instruments. We present the *within* version that is used for instrumenting the spatially differenced specification in the text and report both corrected and uncorrected standard errors.

Table 10: First stage regression results

spatial difference of	(1)	(2)
age censored dummy	-0.0029	-0.0029
	(0.0036)	(0.0036)
age	-0.0006	-0.0006
	(0.0011)	(0.0011)
age squared	0.0001	0.0001
	(0.0001)	(0.0001)
share Conservative	0.1279[a]	0.1279[a]
	(0.0436)	(0.0482)
share Labour	0.2466[a]	0.2466[a]
	(0.0429)	(0.0470)
share Liberals	0.1236[a]	0.1236[a]
	(0.0394)	(0.0426)
Conservative controlled	0.0813[a]	0.0813[a]
	(0.0178)	(0.0226)
Labour controlled	-0.1715[a]	-0.1715[a]
	(0.0154)	(0.0173)
Liberal controlled	-0.1376[a]	-0.1376[a]
	(0.0244)	(0.0281)
share Conservative (if control)	-0.2278[a]	-0.2278[a]
	(0.0325)	(0.0431)
share Labour (if control)	0.2645[a]	0.2645[a]
	(0.0317)	(0.0355)
share Liberal (if control)	0.2659[a]	0.2659[a]
	(0.0428)	(0.0497)
Number of observations	18370	18370
Number of establishments	6087	6087

Notes: Standard errors under coefficients. [a], [b] and [c] denote significance at the 1%, 5% and 10% level respectively. Results from first stage regression of spatial difference of tax rates on establishment fixed effects, exogenous variables and political instruments. Column (1) presents results with uncorrected standard errors. Column (2) presents results with standard errors corrected according to Appendix A.