The causal effects of an industrial policy

Chiara Criscuolo

Centre for Economic Performance, London School of Economics

Ralf Martin

Centre for Economic Performance, London School of Economics

Henry Overman

Centre for Economic Performance, London School of Economics, CEPR

John Van Reenen

Centre for Economic Performance, London School of Economics, NBER

This draft: February 9th 2009, Preliminary and Incomplete

Abstract

Industrial or business support policies designed to raise productivity and employment are a common feature of the policy landscape, especially in the wake of the financial crisis. Rigorous micro-econometric evaluation of their causal effects is rare primarily because of the difficulty of achieving credible identification. We exploit multiple changes in the area-specific eligibility criteria for a major UK program ("Regional Selective Assistance"). Eligibility changes across areas arose from changes to the pan-European state aid rules. We match twenty years of administrative panel data on the population of plants to the population of program participants to investigate the causal impact of the policy on jobs, investment, productivity and entry/exit. Using an instrumental variable approach we find that the program has had a positive effect on both employment and investment, which naïve estimators underestimate. There is no statistically significant effect on total factor productivity. Single plant firms have a much larger treatment effect than plants belonging to multi-plant firms who may be effectively "gaming" the system. There is also some evidence that the program, by supporting less efficient enterprises, may slow down reallocation from less efficient plants, negatively affecting aggregate productivity growth.

JEL classification: H25, L52, L53, O47
Keywords: industry policy, treatment effects, employment, investment, productivity

Acknowledgements: Helpful comments have come from seminar participants in Essex, LSE, NBER, HECER and Stanford, Financial support is from the ESRC through the CEP and the British Academy. We would like to thank the DBERR for the SAMIS database and Marjorie Roome, Beatrice Parrish, Alex Wilson, David Southworth and Fernando Galindo-Rueda for useful insights. The ONS Virtual Microdata Lab ensured access to ONS Data and Alberta Criscuolo helped with the EU legislation. Mehtap Beyza Polat provided excellent research assistance. Errors in use of these data are our own. 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.
Corresponding author: c.criscuolo@lse.ac.uk; Centre for Economic Performance, LSE, Houghton Street, London, WC1E 2AE, UK.

I. INTRODUCTION

The global downturn has brought industrial policy back into fashion. Part of the stimulus packages of governments around the world has been subsidies for firms, most dramatically in the US car industry[1]. But business support policies are not new – most governments have business subsidies that claim to foster productivity and employment, particularly in disadvantaged areas. Traditionally they have been a mainstay of developing economies, but are also common in developed nations. Despite the ubiquity of such schemes, rigorous micro-econometric evaluation of their causal effect is rare (e.g. Klette, Moen and Griliches, 2000). This is somewhat surprising as the sums allocated are large[2] and now growing.

The methods to analyse program treatment effects have advanced significantly in the last ten years or so. Labor economists have used these techniques most intensively, for evaluating a variety of government programs (e.g. Heckman, LaLonde and Smith, 1999). The basic concern these techniques try to address it that government programmes might simply finance activities the recipients – individuals in the labor literature and businesses in our case – would have undertaken anyways in absence of the programme. If this is the case, large amounts of taxpayer dollars could simply be wasted, even before we take into account the deadweight costs of taxation and other distortions induced by the program's design. The consensual view among economists is that industrial policy is a failure, but the econometric basis for this is weak. As Rodrik (2007) emphasises much of these policies are targeted on firms and industries that would be in difficulties in the absence of the program so an OLS regression of (say) jobs growth on subsidies is likely to be seriously biased downwards[3].

To address this concern requires the construction of a counter-factual: what would have happened in the absence of the support programme? Comparison to non-treated firms is one possibility but, of course, the problem is that in the absence of experimental data those who participate in such programs are not random firms but are heavily selected, and thus participation is endogenous. Matching is another strategy used to deal with this problem, but it relies on the strong assumption of conditional independence which is unlikely to hold unless we have access

---

[1] *The Economist* 29th January 2009 reports estimated that current commitments in 2009 were an extra \$21bn in loans for autos and \$135bn of tax cuts for firms. In the UK the numbers were \$3bn for autos and \$14bn in working capital for smaller firms. In Canada it was also \$3bn for autos and another \$2bn in business tax cuts.

[2] For example, in 2005 the 25 European Union countries were conservatively estimated to spend 0.6 of GDP on state aid to industries as a conservative estimate. See http://ec.europa.eu/comm/competition/state_aid/studies_reports/key_indicators.xls

[3] For examples see Krueger and Tuncer (1982), Harrison (1994), Beason and Weinstein (1996) and Lawrence and Weinstein (2001).

to an unusually rich set of covariates. Another solution, and the one we adopt in this paper, is to look for instrumental variables correlated with the likelihood of treatment, but not with individual firm performance.

One factor holding back the industrial policy evaluation literature has been the absence of obvious instruments. In this paper we tackle this problem by exploiting a quasi-experiment that induced exogenous changes in the eligibility criteria governing the receipt of investment subsidies under the Regional Selective Assistance (RSA) program in the UK. RSA is administered by the Department for Business (DBERR) that gives grants to firms for investment in selected (economically disadvantaged) areas of Britain. Grants totalling about $220m were disbursed in the UK in 1998-1999 alone (DBERR, 2001). Crucially for our identification strategy, new European Union guidelines caused a change in the areas eligible to receive these grants in 1993 and in 2000. We exploit the change in these "maps of assistance" to generate instrumental variables for the receipt of investment grants. This enables us (under certain assumptions) to estimate the causal effect of the program on employment, investment, productivity, exit and entry.

Our data set is constructed by linking observations from three very rich administrative data sources. First, from the DBERR, we have data on the population of all firms who applied to the RSA program since 1972 and the amount of assistance they received if they were successful. Second, from the Annual Business Inquiry (ABI) we have panel data on a large stratified sample of UK manufacturing plants, covering 90% of manufacturing employment in the UK. Finally, from the Interdepartmental Business Register (IDBR) we have location, entry and exit information for the population of all manufacturing plants in the UK. Further details are provided below.

Our substantive conclusion is first, that there is a large and significant "average effect of treatment on the treated" for employment and investment. These effects are seriously underestimated if endogeneity is ignored, as the participants in the scheme are firms who would otherwise perform badly given their observable and unobservable characteristics. Second, we find that significant increases in employment are confined to single plant firms. Subsidies to plants in multiplant firms are largely ineffective suggesting that these firms are "gaming" the system. Thirdly, there appear to be no additional effects on productivity after controlling for the investment effects. Since the less productive plants receive subsidies this implies aggregate productivity is lower due to the program as it slows reallocation.

The paper is structured as follows: the next section describes the policy in more detail and outlines how eligibility changes over time. Section III describes the econometric modelling strategy. In Section IV we describe the data we use and the characteristics of treated and non-treated firms in our sample. In Section V we report our results and a number of robustness checks. Finally we provide some conclusions and a discussion on how we intend to forward the research. In Appendices we report more details on the data matching procedure and issues involved; the description of the variables used and we discuss some of the existing literature on evaluations of RSA and similar policies.

II. INSTITUTIONAL FRAMEWORK: DESCRIPTION OF THE REGIONAL SELECTIVE ASSISTANCE POLICY

IIA. Overview

During the period of our study, (1985-2004) Regional Selective Assistance (RSA) was the main regional business support scheme in the UK.[4] From the early 1970s it provided discretionary grants to companies in Assisted Areas. These are disadvantaged regions typically characterised by relatively high levels of unemployment and deprivation. It was designed to "create and safeguard employment". Emphasis was given to internationally mobile investments, new products and processes and the manufacturing sector. Assistance could be provided to establish a new business; to expand, modernise or rationalise an existing business; to set up research and development facilities or enable businesses to take the next step from development to production.

Because RSA has the potential to distort competition and trade between European countries it must comply with European Union legislation concerning state aid. In general, this type of assistance is prohibited by European law except in certain cases. In particular, Article 87 of the Treaty of Amsterdam allows for some state aid in support of the European Union's regional development policies. The guidelines designate very deprived "Tier 1 Areas" (previously called "Development Areas") in which higher rates of grant can be offered and slightly less deprived "Tier 2 Areas" (previously called "Intermediate Areas").[5] There is an upper threshold of support

[4] We discuss our choice of study period below. According to Harris and Robinson (2004), in 1998-9 RSA represented 19% of the UK's industrial policy spending. In April 2004, the RSA scheme was replaced by the Selective Finance for Investment (SFI) scheme. Productivity became an official objective in April 2004, when RSA was replaced by SFI which explicitly requires that projects yield productivity improvements. We discuss the difference between the two schemes below.

[5] Article 87(3) of the Treaty of Amsterdam defines conditions where State aid may be compatible with EU laws. Article 87(3) (a) allows for "aid to promote the economic development of areas where the standard of living is

that is allowed, referred to as Net Grant Equivalent (NGE),[6] which essentially sets a maximum proportion of the firm's investment that can be subsidised by the member state government.

Since the main formulae which determine eligibility are decided at the European level at fixed seven year intervals and not at the UK level, this mitigates concern of endogeneity of policy decisions. And although the UK finance ministry has latitude to decide the overall amount of the annual budget for RSA they are not able to change the rules over which areas are eligible to receive some RSA. Thus, area-level eligibility is the key form of identification in our paper.

IIB. Changes in eligibility over time

The map of the areas eligible for RSA changed twice during our study period: first in 1993 and then again in 2000. There were also changes in 1986 before our sample period begins and in 2006, after our sample period ends. These changes happen every seven years in conjunction with the periodic revision of the Structural Funds, the European Union's main policy for supporting economic development in less prosperous regions.

The map of the eligible areas is proposed by the UK but needs to be approved by the EU in accordance with the EU regional guidelines and in respect of Article 87 of the Amsterdam Treaty. The main criteria are (i) that only areas with underemployment and a low standard of living are eligible (ii) any map or change in the map must satisfy the threshold imposed by the EU defining the proportion of UK population that may be covered by Assisted Areas Status.

The eligibility criteria are outlined in the regional guidelines which are published two years before the implementation of the map (in our case 1991 and 1998). The UK government will then gather quantitative information on indicators of employment level and deprivation at the relevant regional level based on the previous three years data where possible and will propose a new map.

Figures 1 through 4 show that eligibility changes in the maps for all areas in the UK.

Below we discuss each of the changes in turn.

---

abnormally low or where there is serious underemployment" [Tier1/Development Areas] and Article 87(3) (c) allows for: "aid to facilitate the development of economic activities or of certain economic areas, where such aid does not adversely affect trading conditions to an extent contrary to the common interest." [Tier 2 or intermediate Areas] Additional restrictions apply to sectors with over-capacity: motor vehicles, synthetic fibres and yarns, iron and steel, coal, fishery and agricultural products.

[6] The Net Grant Equivalent (NGE) of aid is the benefit accruing to the recipient from the grant after payment of taxes on company profits. RSA grants must be entered in the accounts as income and are made subject to tax. Details for calculations of NGEs are available in OJ C74/19 10.03.1998.

(a) The 1993 change

The assisted area map for RSA was redrawn in 1993 on the basis of the new 1991 guidelines using "Travel to Work Areas" as the underlying spatial units.[7] The selection of Assisted Areas was based on several factors using a quantitative formula. The first set of factors used indicators of bad labor market conditions, such as persistently high unemployment, the proportion of long-term unemployed, participation rates and the likely future demand for jobs (based on growth/decline in local industries, demographic changes and expected major firm closures). The second set related to geographic features such as distance from major markets, low population density and urban problems.

The Assisted Areas fell into two categories: (a) Development Areas where aid could be granted up to a maximum of 30% NGE (Net Grant Equivalent - see above) and (b) Intermediate Areas where aid was limited to 20% NGE. The new 1993 maps implied a net reduction in the number of assisted areas with Development Areas covering 17%, and Intermediate Areas covering 19%, of the total UK population.

(b) The change in 2000

The EU Commission introduced new guidelines for State Aid in 1998 and the UK responded to that with the introduction of a new Assisted Area map in 2000. The maximum investment subsidy allowed for in these areas is 35% NGE for the most deprived (Tier 1) Areas.[8] These areas are the four eligible for funding under Objective 1 of the EU Structural Funds: Cornwall & the Isles of Scilly, Merseyside, South Yorkshire and West Wales & the Valleys.

The Tier 2 areas are more scattered. These 65 zones are constructed on the basis of groups of electoral wards.[9] Each grouping must have a population of at least 100,000 and the wards were selected according to four statistical indicators. Although the main criteria to decide eligibility were still labor market performance and the share of manufacturing, the precise

---

[7] Travels to Work Areas are defined by the UK Census Bureau (Office for National Statistics). The fundamental criterion is that, of the resident economically active population, at least 75% actually work in the area, and also, that of everyone working in the area; at least 75% actually live in the area. Thus, in terms of definition, they are similar to the US Metropolitan Statistical Areas.

[8] Special status and a higher allowance are awarded to Northern Ireland, which is not included in our analysis.

[9] The data used for the zone boundaries come from the 1991 Census of Population. A detailed list of the assisted wards by local authority within regions and the NGEs to which they are eligible is available from the authors upon request.

indicators differed from 1993. We discuss below how this change in decision rules affects the areas eligibility independently of areas economic conditions.[10]

Within Tier 2 Areas the map identified four sub-tier areas eligible for different level of maximum NGE. The level of aid intensities proposed for these areas vary according to the seriousness and intensity of the problems in each region relative to the Community context, in particular as regards other EU countries.

For the most disadvantaged sub-tier areas, that were geographically distant and sparsely populated, a maximum subsidy rate of 30% NGE was allowed[11]. The maximum NGE level for relatively less deprived areas was 10%.[12] However, if those (less deprived) areas are adjoining to Tier 1 areas they have a 20% ceiling. The rest of the eligible areas aid ceilings are either an NGE of 20% or 15% (with the decision as to which applies made by referring to current conditions as well as the NGE in the 1993 map).

Finally, note that in assisted areas, a higher rate may be paid to small and medium-sized enterprises.[13] We plan to exploit this size difference in future work.

IIC. Formal criteria for receipt of RSA

During our study period (1988-2003), RSA traditionally targeted manufacturing sectors, although support was also available for services sector firms that served national or foreign markets (i.e. not just the local market). The grants were discretionary and firms could only apply if the supported project satisfied the following criteria. (a) **Location**: The project had to be undertaken within one of the Assisted Areas. (b) **Investment:** The project had to involve capital expenditure on property, plant or machinery; (c) **Jobs:** The project should normally have been

---

[10] The indicators used are employment and unemployment rates and manufacturing share of employment. In the Appendix we report the results of analyzing why areas switched eligibility status. Although the weights given to the indicators are not published we can estimate these empirically using lagged values of the indicators (i.e. using 1989-1991 values of the indicators for the 1993 rule change and 1996-1998 to predict the 2000 rule change. These do a good job at predicting changes in eligibility. Estimating the probability of becoming eligible for all areas both in 1993 using lagged values (e.g. 1989-1991) of the indicators used in the 2000 eligibility rules and vice versa and in 2000 using lagged values (e.g. 1996-1998) of the indicators used in the 1993 eligibility rules we show that about 40% of areas that were eligible in 1993 (2000) would have not been eligible according to the 2000 (1993) rules. This suggests that becoming eligible/ineligible is affected by factors beyond the economic conditions of the area.

[11] These areas have a population density of less than 12.5 inhabitants per square kilometre and are mainly the Highlands in Scotland (1.2% of assisted areas population were in these areas).

[12] These are areas with a higher GDP per capita and lower unemployment rate than the Community average (covering 4.2% of assisted areas population).

[13] The additional support for smaller enterprises consists of fifteen percentage points gross in Tier 1 areas, and ten percentage points gross in Tier 2 areas.

expected to lead to the creation of new employment or directly protect some or all of the jobs of existing workers which, without the project, would otherwise have been lost; (d) **Viability:** The project should have had good prospects of viability and should have been expected to help the business become more competitive; (e) **Need:** The applicant had to demonstrate that assistance was necessary to enable the project to proceed as envisaged in terms of nature, scale, timing or location;[14] (f) **Prior Commitments:** As RSA could only be offered where it would make the difference between the project going ahead and not proceeding, there should have been no prior commitment to the project, i.e. the DBERR must have completed its appraisal of the project and issued a formal offer of assistance before the applicant entered into a commitment to proceed with the project; (g) **Other Funding:** The greater part of the funding for the project should have been expected to be met by the applicant or come from other sources in the private sector.[15]

Location, which also forms the basis for our instrumental variables, is objective, clearly defined and enforceable.

The process for application was as follows. Firms needed to fill in an application form, in which they needed to prove additionality, to provide business plans, accounts and reasons for wanting the grant. They then submitted this to the local DBERR regional agency. The lag between the day the application was submitted and the decision depended on the amount that the firm was applying for. During the period analysed, the lag was normally between 35 and 60 days, and 100 days or more for grants above £2 million (about $4 million). The lag also depended on the time needed to appraise the application to ensure that all of the criteria were met and on negotiations between the government agency and the firm on the terms of assistance. If the application was accepted, the firm was paid the minimum necessary to get the project going. Additional payments started only after the jobs were created/safeguarded and the capital expenditure defrayed and were based on agreed fixed capital expenditure and job targets. The payments were given in installments – between two and seven and in the vast majority of cases these were spread across more than one financial year. The government agency monitored the project with visits (normally one per year, but more frequently for risky projects).

[14] This may be to meet a funding gap, to reduce the risks associated with the project, or to influence the choice of location of a mobile project. It might also be to obtain parent company approval by meeting established investment criteria; or for some other acceptable reason – each case is considered on its own merits.

[15] These may include bank borrowings, hire purchase or lease finance, equity and loan finance from existing or new shareholders and loans from other organisations or institutions. Additional public sector assistance may however, be available towards the project. Any additional assistance must be cumulated with the RSA support and must not breach the European Union State Aid limits.

III. ECONOMETRIC MODELLING STRATEGY

We next consider the econometric modeling strategy, starting with our basic approach and then considering extensions.

IIIA. Basic Approach

Consider the outcome equation:

$$y_{it} = \alpha D_{it} + \beta X_{it} + u_{it} \tag{1}$$

where y_{it} is the outcome of interest for plant ("local unit") *i* at time *t*. Note that a plant ("local unit") is uniquely located in an area, *r*, and in a firm ("reporting unit"), *j*. We suppress the sub-scripts on area and firm for notational simplicity unless needed. D_{it} is the participation indicator which we will initially assume to be binary (see below for an extension to the continuous treatment intensity setting). Thus D_{it} = 1 if a plant is in the program (RSA) in year *t* and zero otherwise. Initially we assume a homogeneous treatment effect, α, but we relax this assumption below. X_{it} are other covariates used as controls such as age, industry, area, whether the plant belongs to a larger firm (and if so whether this firm is domestic or foreign). Outcomes include employment, investment and productivity; the precise set of X_{it} will depend on the outcome of interest. The u_{it} is an error term which we decompose into a correlated fixed effect, η_i, a set of time dummies, τ_t, and v_{it}, a plant-specific time varying error term:

$$y_{it} = \alpha D_{it} + \beta X_{it} + \eta_i + \tau_t + v_{it} \tag{2}$$

Estimation of equation (2) through OLS will purge our estimate of α of biases due to common macro-economic shocks (through τ_t) and permanent correlated unobserved heterogeneity (through η_i), it will still be inconsistent if there are unobserved transitory shocks v_{it} correlated with D_{it}. This is likely to be the case as areas and firms who are facing difficulties are targeted by the policy to receive subsidies. In this case E(v_{it} D_{it}) < 0 and OLS will underestimate α.[16]

[16] Working in the opposite direction is the fact that a second objective of RSA is to create new jobs, which may increase the likelihood of receiving a grant for firms who have experienced some positive shock v_{it}.

Consequently we consider instrumental variables, Z_{it}, for program participation, D_{it}. The instrument we will construct exploits the fact that only plants located in certain areas of the UK are eligible for RSA (see Figures 1 to 4). Although these areas are fixed at a given point in time, the UK government changed the map of eligible areas twice (in 1993 and 2000) to comply with European Commission State Aid legislation. As a result, some areas ceased to be eligible and other areas that were ineligible became eligible. For example, Darlington a small city in the North-East of England was eligible pre-1993, but became ineligible in 1993. The changes were driven by a new European-wide formula for calculating which regions were eligible to receive subsidies under state aid rules (see Section III for more details). In addition to discrete indicators constructed on the basis of eligibility, the maximum investment subsidy also differs across eligible areas in a way that has changed over time. This allows us to construct measures for the intensity of treatment. Finally, there are size and industry-specific components of the subsidy which can be used to improve the efficiency of the estimates.

With the instruments we can estimate equation (2) by instrumental variables. As reported below, we look carefully at the first stage to check for weak instruments issues. We also consider the reduced form:

$$y_{it} = \pi_1 Z_{it} + \pi_2 X_{it} + \tilde{\eta}_i + \tilde{\tau}_t + \tilde{v}_{it} \tag{3}$$

Under the covariance assumption $E(Z_{it}\ \tilde{v}_{it}) = 0$ the estimate of π_1 by OLS is the "intent to treat" effect, which is of interest in its own right.

When moving from theory to implementation, one complication arises because of the unit of observation in the available data. We have written the analysis at the plant level, however the main data used for the analysis (the UK Census Bureau's ABI dataset) is collected at the firm-level rather than at the plant -level. Although for most firms in the ABI the two levels of aggregation coincide (on average 80% of reporting units sampled are single plant firms), measures of investment, output and materials are only available at the firm level.[17] Employment and location are always available at the local unit level, even for multi-plant firms.

To deal with this issue we simply aggregate the relevant equation across all plants in the same firm. For firms, equation (2) becomes:

$$y_{jt} = \alpha D_{jt} + \beta X_{jt} + \eta_j + \tau_t + v_{jt} \tag{4}$$

[17] We call this the firm level, *j*, but there could be many reporting units in one large firm.

For example, when y_{it} is total employment in the plant, y_{jt} is simply employment in the firm, summing across all plants *i* in firm *j*, i.e. $y_{jt} = \sum_{i,i\in j} y_{it}$. All other variables are defined similarly, although there are some issues around the definition of D_{jt} and the instruments, Z_{jt}. For the participation dummy we mainly continue to use a simple binary indicator if any plant in the firm received any treatment. But we also present checks on alternatives such as the amount of money received expressed as the fraction of project total costs covered by the grant. For the firm-level instruments we have $Z_{jt} = \sum_{i,i\in j} w_{it}^{j} Z_{it}$. We consider several different weighting factors, w_{it}^{j} , but one important issue is that choice of the weights could induce an endogeneity bias. For example, the current distribution of firm employment across plants across areas could be affected by the eligibility to RSA. Consequently we only ever use lagged data to construct weights (see sub-section IIIC below for a more detailed discussion).

Note that the interpretation of α subtly changes in the aggregated regression. Consider employment outcomes and assume that the number of plants is fixed. If a firm has two plants in two areas and then one area becomes ineligible for RSA, the firm could substitute employees from the plant in the ineligible area to the plant in the eligible area without changing total employment. Analysis at the plant level in equation (2) would find a positive program effect. Analysis at the firm level in equation (4) would find zero effect. In theory, program rules are meant to stop firms engaging in such switching, but in practice this is hard to enforce as the firm has more private information on the true counterfactual than the government agency. Given our data, and the fact that equation (4) is arguably of more direct policy interest, we focus on firm level results in what follows.[18]

[18] Comparison between the two estimates would be informative as regards such intra-firm switching behavior. In the paper we report our analysis for all of the firms in the sample and then separately for single plant firms. A comparison of the results for these two samples could give some indication of how this behaviour affects our results. However, we might be introducing additional selection bias as the single plant firms are a subset of the sample. Ideally, we would want to use the information on employment reported in the business register, (IDBR). We intend to exploit this information as a robustness check but we are worried about measurement error issues when using the IDBR employment information (see Data section for a description of the IDBR and related issues).

IIIB. Extensions

(a) Heterogeneous Treatment Effects

If we relax the assumption that the response to participating is the same across firms we can re-write the plant-level equation of interest as:

$$y_{it} = \alpha_i D_{it} + \beta X_{it} + \eta_i + \tau_t + v_{it} \qquad (5)$$

where α_i is now the plant specific effect of treatment.

There is much discussion in the evaluation literature over the interpretation of IV estimation when the true model is equation (5) rather than equation (2).[19] The essential problem is that using observations from the whole population may give a selection of non-treated plants that provides a poor comparison group for those who participate. To address this we apply matching techniques to link our set of participants with a set of control observations using the propensity score, trimming the sample of participants and controls so we have a common support. The combination of using program eligibility as an instrument combined with matching[20] is proposed by Blundell, Costa-Dias, Meghir and Van Reenen (2004).

One interpretation of the resulting IV estimates (even with matching) is a Local Average Treatment Effect (LATE). To understand this consider the IV estimate of equation (5) dropping all covariates. Let D_{0i} and D_{1i} denote potential treatment assignments defined relative to the binary instrument Z_i. D_{0i} indicates what treatment *i* would receive if $Z_i = 0$ and D_{1i} indicates what treatment *i* would receive if $Z_i = 1$. The LATE is defined as $E(\mathrm{Y}_{1i} - \mathrm{Y}_{0i} \mid \mathrm{D}_{1i} > \mathrm{D}_{0i})$ the effect on the outcome of interest for the individuals in the population who are induced into treatment as a result of the instrument. If the instruments satisfy the usual rank and order conditions, plus a monotonicity condition (i.e. that the instrument never decreases treatment probability for any individual) then Imbens and Angrist (1994) show that the IV estimator identifies the LATE, i.e.

$$\frac{\mathrm{E}(\mathrm{Y}_i \mid Z_i = 1) - \mathrm{E}(\mathrm{Y}_i \mid Z_i = 0)}{\mathrm{E}(\mathrm{D}_i \mid Z_i = 1) - \mathrm{E}(\mathrm{D}_i \mid Z_i = 0)} = E(\mathrm{Y}_{1i} - \mathrm{Y}_{0i} \mid \mathrm{D}_{1i} > \mathrm{D}_{0i})$$

---

[19] For some examples see Angrist (2004), Imbens and Angrist (1994) or Heckman et al (1997, 1999).

[20] Using the propensity score function we restrict the sample to those firms that have a predicted propensity score larger than the 10th percentile of the propensity score distribution of treated firms and lower than the 90th percentile of non treated firms. We check the robustness of these results to more conservative thresholds. There are other ways to match including matching by area (looking at ineligible areas that are closer in observed characteristics to eligible areas) and matching within area by plant and firm observables.

Where the left hand side is the population analog to the Wald estimator and the right hand side is LATE. If we consider the OLS estimate of equation (5) as picking up the average marginal effect of the subsidy program, what does the LATE estimate reflect? Note that we are estimating in differences so identification is essentially between areas who become eligible and those that were always eligible. In both areas there are plants in different states of program participation. The group of plants who become participants in the recently eligible areas (relative to the plants who became participants in always eligible areas) are likely to be more financially constrained and therefore have larger returns to the program than the average treated plant. This is a reason why we might expect a larger IV coefficient than the OLS coefficient (see Card, 2001, for an analogous argument in the education literature).

(b) Aggregation to the area level

We also examine the impact of treatment at the small regional level of a ward rather than firms or plants and at more aggregate regional level such as districts; travel to work areas (TTWAs) and counties.

$$y_{rt} = \alpha D_{rt} + \beta X_{rt} + \eta_r + \tau_t + v_{rt} \tag{6}$$

where $D_{rt} = \sum_{i, i \in r} w_{it}^R \; D_{rt}$ is the weighted total number of treatments in an area r at time t. We look at several alternatives for w_{it}^R.

Analogously with equation (4) we are able to see what the area-wide effect is of the policy taking into account any possible substitution effects between participating and non-participating plants. For example, although the UK Department of Business performs a market and competition assessment prior to awarding subsidies, it is still possible that, ex-post, plants that receive subsidies grow at the expense of other non-participants in the same area.

More generally, the area-level analysis enables us to look at several aspects of the effect of program that we cannot identify at the plant or firm level. For example, at the firm level, we can only capture entry of new plants for incumbent multi-plant firms while at the area level we can examine plant entry (single or multi) by non-incumbents. In addition, the aggregate growth in employment in the area can be decomposed into the components coming from incumbent growth, exit and entry.

(c) Indirect Effects

Estimation of equation (2) by IV makes the usual Stable Unit Treatment Value Assumption (SUTVA), i.e. the treatment of unit *i* only affects the outcome of unit *i*. A violation of SUTVA would be when there are "indirect" effects of the program on non-treated plants. In the previous sections we have already considered some violations of this assumption: within a firm, there may be substitution away from some plants in non-eligible areas towards other eligible plants; within an area there may be substitution away from non-participants to participants. As discussed above, these indirect effects can be tested for by comparing estimation results for the same specification, but at different levels of aggregation (see Griliches, 1979 for a similar discussion in the context of R&D spillovers).

A more general indirect effect, that we have not yet discussed, may arise from substitution of activity between *different* areas for *different* firms (contrast equation (4) which picks up within-firm substitution and equation (6) which picks up the within area). For example, when area A becomes eligible and area B remains ineligible, does employment fall in area B? A way to test this is to examine the impact of switches in eligibility in A on outcomes in neighbouring areas to A. Effectively we intend to use "border effects" to see if there are strong indirect effects.

(d) Dynamics

Changes in eligibility are unlikely to have an immediate impact on outcomes. Most important is the fact that the need to collect information and properly assess applications means that there can be considerable delay between the application for a grant, its approval and eventual receipt of the subsidy. In addition to this administrative delay there may be the normal lags between investing and observing changes in production and employment (delivery lags, time to build, etc.). Consequently, our main instruments use eligibility dates at t-2, but we also experiment with lagging the eligibility by one or three years. This should also reduce further any concern that the selection of areas is endogenous to shocks to individual firm outcomes.

(e) Expectations

We are assuming that, from the firm's perspective, the change in eligibility is unexpected. Although there was some uncertainty over the exact placing of re-drawn boundaries and revised levels of subsidy it is possible that many firms knew about changes in advance. Although grants are only paid in arrears, the increased probability of subsidy could alter firm behaviour in advance of the change. For example, some firms may have delayed investment as they would wait to become eligible to the subsidy. We can check for the magnitude of these effects by including *future* Z in the reduced form of equation (3) that captures the intent to treat. The significance of these (in the presence of our preferred current or lagged Z's) would signal the possibility of future looking behaviour by firms.

(f) Continuous Treatment Intensity

So far, we have focused on the discrete treatment case, but we can also exploit more information by using a continuous measure of treatment intensity. Our main continuous measure is simply to calculate the proportion of investment that is paid for by the program. If we denote the amount of grant received as R then in this case the participation variable is: $\left(\frac{R}{I}\right)_{it}$ where I is the total investment cost of the project. This investment subsidy can be directly calculated from available data.[21]

IIIC. More detail on the Instrumental Variables

In the dataset we have information at the postcode level on whether plants located there are eligible and if so the Net Grant Equivalent.[22] As already discussed the map of eligibility and RSA rates changed in 1993 and in 2000 so the data contains variation in both the cross-section and the time-series dimensions. Consider first the discrete variation in eligibility: when we use the cross-sectional variation, identification comes from firms located in eligible areas who did not get treated. When we use time series variation we use information from firms whose eligibility changes as a result of changes to the RSA map. Second, we exploit variation in the RSA rate since the higher the RSA rate the higher the returns to applying for RSA. Identification is similar to that for eligibility except we now use differences in RSA rates or changes in RSA rates. Table 3 reports the number of such changes in eligibility and RSA rates both at the plant and the more aggregated firm and ward level in our sample over both the 1993 and the 2000 eligibility map changes.

[21] An alternative is to include two variables, one a dummy for participation as before (D_{it}) and the second $D_{it} * \ln(R_{it} / \overline{R})$ where $\overline{R}$ is the sample mean of R conditional on receiving any grant (to enable the average marginal effect to be read off more easily for the participants. Since the logarithmic transformation is undefined at zero we set this to an arbitrary value (0) which will be picked up by the participation dummy.

[22] In the UK, postcodes typically refer to one property or a very small group of dwellings.

(a) Endogenous Eligibility?

One concern is that areas that lose eligibility are also those who have improving economic conditions, thus generating a bias on our instrument. Consider the first differenced equivalent of the reduced form, equation (2), and ignoring time dummies for simplicity:

$$\Delta y_{it} = \pi_1 \Delta Z_{it} + \pi_2 \Delta X_{it} + \Delta w_{it} + \Delta \nu_{it} \tag{7}$$

We have decomposed the error term into two components, Δw_{it} which is correlated with the eligibility changes and a truly idiosyncratic error, $\Delta \nu_{it}$ which is not. The first thing to note is that since areas who are doing better are more likely to be made ineligible for RSA, i.e. E($\Delta Z_{it} \Delta \nu_{it}$) <0, this will lead to a *downwards* bias on the coefficient of interest, π_1 , and make it harder to identify a policy effect.

Recall from the discussion in Section II, however, that the determination of area eligibility status depends on the European Commission's Regional Guidelines which are published two years prior to the map changes. The implementation of the guidelines, in turn, depend on data that available, at most, three to five years before the map changes (for example, in the 2000 change most of the indicators were actually based on the 1991 Census – nine years previous). So the magnitude of this possible bias will depend upon the correlation between variables like unemployment rates three years ago and current unobserved area-specific shocks. Note that variation in Z_{it} is also driven by changes in the EU wide average GDP per capita and unemployment which change dramatically as new countries have entered the EU and the changes in the criteria/indicators used to determine eligibility.

Although we think the size and direction of such biases are likely to be second order, we consider some checks of this. First, we include area-specific trends to proxy $\Delta \nu_{it}$ which are likely to pick up any longer run decline in an area that are not reflected in the covariates. Second, since we know the rules that are used to designate areas we can condition directly on the area-specific indicators used to determine eligibility, such as lagged area GDP per capita and lagged area unemployment rates. Since eligibility is now only driven by the exogenous decision of the EU commission, Z_{it} is identified.

(b) *Weighting used in constructing the instrument*

As discussed above, information on the ARD is recorded at the reporting unit level rather than at the local unit level (i.e. for multi-plant firms information is aggregated at the firm level), so we are also faced with the additional issue of how to use the information on eligibility and rates for reporting units that have several local units some of which are eligible and others of which are not (and similarly for firms whose local units face different RSA rates).

As noted above an obvious concern is that using the current distribution of employment within the firm could create an endogeneity bias as this distribution could be affected by RSA eligibility. Also the location of plants within the firm could be affected by RSA eligibility. Consequently we used only lagged information on the location of plants. To further reduce the risk that forward-looking firms take into account future changes in eligibility in deciding where to locate their plants our main results use the location of the oldest plant in the firm (i.e. the local units owned by the firm for the longest amount of time) to calculate eligibility for RSA. The past geographical location of this plant is least likely to be affected by current changes in the eligibility map.

We also investigate the robustness of our results to less conservative alternatives constructions of the weight for the instrument such as using all plants of whatever age and employment distributions across regions. Generally the results are even stronger, which we can't exclude results from potential endogeneity of the instruments.

(c) Functional form of the instrument

Our instrument is the level of the maximum investment subsidy, the Net Grant Equivalent, available in the area. This variable takes on a number of discrete values ranging from zero in ineligible areas to 35% in the most deprived areas after 2000. Our baseline results for the analysis use mutually exclusive dummies for two aggregations of the different rates (with zero being the baseline), in order not to impose too much structure on the non-linear schedule. We also show the robustness to more parametric definitions of the instrument using a continuous measure.

The definitions of the instrument are therefore $NGE <= x$ where x ranges from greater than zero to 20%. For example, $0<NGE <= 0.2$ indicates that the firm's oldest plant is in an area where the maximum investment subsidy (Net Grant Equivalent) is 20% or less; and a second dummy $0.2<NGE<=0.35$ indicates that the firm's oldest plant is in an area where the maximum investment subsidy (Net Grant Equivalent) is greater than 20%,.

IV. DATA

We combine administrative data on support scheme participants[23] with independent business performance data. This involves matching Selective Assistance Management Information System (SAMIS) database of participants, the Interdepartmental Business Register (IDBR) and the Annual Respondents Database (ARD) which we describe in more detail below. We also describe the characteristics of the firms in our sample while we address in more detail the issues related to the matching procedure to the Appendix.

There are two main advantages of such an approach relative to evaluations based on industrial surveys. First, we can compare the firm before and after its exposure to the program – frequently data on program participants is only available after they have joined the program. Secondly and more importantly, we can compare the change in the participating firms' performance to a "control group" of firms who did not participate or were not eligible to participate in the program. Finally, independent performance data is less likely to be affected by strategic reporting by surveyed firms.

IVA. Administrative Data on program participants

The Selective Assistance Management Information System (SAMIS) was used to monitor RSA projects. It contains information on more than 50,000 applications from 1972 to 2003. It includes for all applications information on the name, date and address of the applicant, a project description; the amount applied for, aims and date of application. For successfully completed applications it provides the date in which and the amount of the grant offered and paid (since 1988 additional payment information is available containing date and amount of first and last instalments). For those that were not completed it contains information on why; i.e. whether the project was withdrawn; was accepted but then the firm did not proceed; was not accepted by the firm; or was rejected by the DBERR and if so for which reason[24].

Since the payment information from the SAMIS database is not always accurate we prefer to use additional information with more detailed payment information from the Payment RSA database available from 1988.

[23] As described in more detail below we also have information on applicants to the scheme that were rejected for various reasons or had withdrawn their application.

[24] Note that only about 10% of all applications were rejected.

For reasons due to the quality of the match with the other data sources used, as described in more detail in the Appendix, and in order to have accurate information on payments and some information on post-treatment performance we use the applications made between 1st January 1988 and 31st December 2003.

IVB. Interdepartmental Business Register (IDBR)

In order to be able to match the administrative information with production data the records from the SAMIS database needed to be matched with the Interdepartmental Business Register (IDBR), which contains both the names of the businesses and the identification numbers used by the Office for National Statistics to conduct the Annual Business Inquiry. The Interdepartmental Business Register (IDBR)[25] is essentially a list of all businesses in the UK, their addresses, type of activity and ownership/control structure compiled using a combination of tax records on VAT and PAYE, information lodged at Companies House, Dun and Bradstreet data, and data from other surveys using three aggregation categories: "local units" (plants), "enterprises" and "enterprise groups".[26] A plant or "local unit" is defined as "an enterprise or part thereof (e.g. a workshop, factory, warehouse, office, mine or depot) situated in a geographically identified place" and is identified by a unique identifier. A major advantage of the IDBR is that information is available at many disaggregated levels. For our analysis this is particularly useful since we also look at the effect of the policy at the regional level on employment and entry/exit. We therefore need employment and entry/exit information at the local unit level rather than at the enterprise level since enterprises can consist of local units in different regions.[27]

[25] The IDBR was introduced between 1994 and 1995. Previously, that sampling was on the basis of a Business Register maintained by the Office of National Statistics (the UK Census).

[26] Criscuolo et al (2003) report that in the 1998 IDBR the vast majority of enterprise groups and Reporting Units consist of just one local unit (92%, 149,326 out of 162,477 and 93%, 158,727 out of 171,271 respectively).

[27] Employment information on the IDBR comes from PAYE data if that is the source of the original inclusion and the enterprises operate a PAYE scheme, which in turn if operated at the local unit level, provides independent local unit employment data. Also the IDBR gathers and updates information on employment from the Annual Register Inquiry (see Criscuolo et al., 2003 for details) and the Annual Business Inquiry (ABI). However, employment data is required to construct sampling frames and hence it will be interpolated from turnover data. The IDBR turnover information comes from VAT records if the original source of business information was VAT data; however this information is quite limited as it is only available for single-local unit enterprises that are large enough to pay VAT (the threshold was £52,000 in 2000–01) at both the enterprise and local unit level. For multi-local unit enterprises, no turnover information will be available for local units, since most multi-local unit enterprises do not pay VAT at the local unit level.

A stratified[28] random sample of enterprises is drawn every year from the IDBR to form the sampling frame for the Annual Business Inquiry (ABI), which provides information on employment, investment, materials, etc. and is described next.

IVC. Annual Respondents Database (ARD)

The Annual Respondents Database (ARD)[29] is the UK equivalent of the US Longitudinal Respondents Database and is made available by the Office for National Statistics (ONS) based on information from the Annual Business Inquiry (ABI),[30] the mandatory annual survey of UK businesses. The ARD unit of observation is defined by the ONS as an autonomous business unit (also referred to as "reporting units").

Some of these business units are spread across several sites but in about eighty percent of all cases a business unit is located entirely at a single mailing address. We call this unit a "firm".

It is important to note that the ARD does not consist of the complete population of all UK businesses, since the sample is stratified with smaller businesses sampled randomly. It contains the population of larger businesses however (those over than 100 or 250 employees depending on the exact year). Each year the sampled firms account for around 90% of total UK manufacturing employment. For our analysis we use the unbalanced panel between 1985 and 2004.

The ARD contains a wealth of information, but most importantly for our study it contains information on employment, investment, intermediate inputs and gross output. We are particularly interested in the effect of RSA for employment, investment and productivity.

IVD. The samples

The details of the linking procedure are described in more detail in the Appendix. Here we describe briefly the difficulties encountered and the characteristics of the sample focusing on the differences between firms that receive RSA and those who do not receive RSA in our sample.

28 Stratification is broadly based on industry affiliation; regional location and size. For details see Criscuolo et al. (2003).

[29] More extensive description of the ARD can be found in Criscuolo, Haskel and Martin (2003), Griffith (1999) and Oulton (1997).

[30] Called the Annual Census of Production until 1998.

(a) Plant Sample

For the whole IDBR we have essentially all businesses and are able to match all the plants to firms who participated in RSA. The exception to this is where there are issues such as variations in spelling of names, changes in postcodes or typos in either of the databases. From manual examination of the merge results this does not seem to be a major issue.

A second problem is that for a large firm that consist of several plants it is not always clear which plant has been participating in the programme and which has not. This is not a problem for the firm-level analysis of course, but means that there is measurement error in classifying plants as "participating" or not. To solve this problem, we use postcode information at the plant level to identify participating plants in the firms. Secondly we rerun our regressions excluding ambiguous matches. Finally, our instrumentation strategy should deal with this measurement error problem.

We call this the plant sample and we can use this sample to conduct regional level analysis on exit rates and employment since when we conduct regressions at the area level as with equation (6), we aggregate the plant level data within an area to form an area panel.

(b) Firm Sample

For analysis of investment and productivity we need to use the match between the SAMIS and ARD sample. This is a (large) subset of the total population of all businesses in Britain so we miss some businesses that participate in a programme but might not be sampled in the ARD in a particular year. Moreover since for our analysis we need to have observations both pre- and post-treatment this issue might be more severe for smaller plants. However, since sampling of the ARD is random this is not a problem for consistency of the estimates, although it does mean we miss out on some of the smaller firms.

IVE. Some Data Description: Characteristics of RSA recipients

Table 1 reports descriptive statistics distinguishing between plants and firms which never participate in RSA and firms that participate to the RSA programs. For the participants we report characteristics before they joined the program.

Rows 1 and 2 show that both plants and firms that participate into the program are on average significantly larger than non participants (column 1). The average participating plant is more than 3 times larger than the average non participant; the average participating firm is 1 and half time larger than the average non participating plant. These differences remain also when comparing the medians (column 3). At the firm level we can also compare other measures of size: gross output (row 3) and investment (row 4). According to both these measures participants are larger than non participants both at the mean and the median. Finally in row 5 and 6 we compare firms in terms of labor productivity, measured as value added per employee (VA/L), and Total Factor Productivity (TFP). We show that participating firms are significantly less productive when they enter the program.

V. RESULTS

We report plant and firm level results for employment. For firms we additionally report results on investment and productivity. We then present some robustness tests before moving on to the area level analysis.

VA. Plant Level Analysis –Results

We start by reporting the results from estimating equation (5) using simple OLS regressions for employment at the plant level in Table 4. Our programme variable, RSA is a dummy which is equal to unity for all the periods in and after a firm has participated in the RSA program and zero otherwise.[31]

The first four columns do not include fixed effects whereas the last four columns do include fixed effects. The first column simply reports the basic OLS regression results. The RSA program participation dummy is positive and highly significant RSA coefficient. This very high value reflects the fact that program participants are larger as shown in Table 1. Column (2) reports the reduced form where we regress plant employment on the policy instruments – two dummies for the level of investment subsidy available to the plant; the first dummy switches on if the plant is located in an area where the level of investment subsidy is above zero and not larger than 20%; while the second dummy is unity in areas where the subsidy is larger than 20%. The omitted category is for plants that are in areas ineligible for an investment subsidy. The policy dummies are both positive and jointly and individually significant. They show a sensible pattern rising monotonically with the generosity of the subsidy. Column (3) reports the first stage of the 2SLS estimates where we regress the RSA dummy on all the exogenous covariates. The policy variables are again jointly significant with the largest effect from the area with the most

[31] Given the large number of observations (almost 3 millions) and given the limited calculating power we can only include year dummies and drop any other variable in the vector of explanatory variables. In all regressions, we control for year dummies and allow for clustering at the area level.

generous subsidy.[32] In column (4) we present the instrumental variable results. The coefficient on the RSA dummy is much larger than in column (1) suggesting substantial downwards bias in OLS.

To control for permanent unobserved heterogeneity we include a full set of firm dummies and repeat the same specifications. Column (5) reports OLS-FE estimates where the coefficient on the RSA dummy is a tenth of the size that of column (1), reflecting the fact that most characteristics that we could not control for in the OLS regressions are captured by the fixed effects but remains a positive 15.2% and significant at the 1% level. Columns (6) and (7) report the reduced form and first stage respectively. The policy instruments are positive and strongly significant, and again present a monotonic increasing pattern both in the reduced form and the first stage estimates. Finally in our preferred specification of column (8) the IV results show a strongly significant positive RSA coefficient estimate, larger than the OLS-FE column (by a factor of two). This confirms the evidence of a downward bias which is consistent with the view that RSA is being awarded to those firms who face negative shocks, exactly what one would expect from the policy aimed at under-performing firms in deprived areas.[33] The magnitude of this coefficient suggests that plants that enter into the program observe an increase in employment of 30.6%. Given our mean and median sized participating plants this would imply an increase of 24 employees (from 79) at the mean and of 2 (from 6) at the median plant.

VB. Firm Level Analysis – Basic Results

At the firm level we can estimate equation (5) for each of the following outcome variables: employment; investment and productivity. The treatment variable RSA is defined as before as a dummy which is equal to unity for all the periods in and after a firm has participated in the RSA program and zero otherwise. This time we can include in the vector of explanatory variables a dummy for whether the firm is part of a domestic group or of a foreign group, a quadratic polynomial in firm age, a dummy for firms that was born before 1985 to control for left

[32] The F-statistics of the excluded instruments confirms that the two eligibility dummies are strong instruments.

[33] Another possible explanation of why the OLS-FE estimates are lower than the OLS-IV estimates is random measurement error in the endogenous variable RSA. This could be true for at least two reasons in our case. Firstly there could be some degree of error when doing name-postcode matching the treatment information from the SAMIS database to the plant information from the business register. Secondly there might be some mismeasurement of the starting date of the treatment.

censoring of the age variable, a full set of four-digit industry dummies, regional dummies and time dummies. Note that in all the regressions we allow for clustering at the area level.

We first turn to analyzing employment in Table 5. All columns include firm fixed effects. The first columns report regression estimates using the whole sample of firms in the ARD. The first column simply reports the OLS-FE regression results. The RSA program participation dummy is positive and significant with a coefficient that indicates that RSA participation is associated with a 14.1% increase in employment; this coefficient is similar in magnitude and significance to the coefficient estimated at the plant level (15.2%). Column (2) reports the reduced form where we regress firm employment on the policy instruments – the two dummies for the level of investment subsidy are one if the oldest plant(s) in the firm is in an areas that is eligible to no more than (or more than) 20% investment subsidy.[34] The omitted category is if the firm's oldest plant(s) is not located in an area that is eligible for an investment subsidy. The policy dummies are all positive and jointly and individually significant. They show a sensible pattern rising monotonically with the generosity of the subsidy. Column (3) reports the first stage of the 2SLS estimates where we regress the RSA dummy on all the exogenous covariates. The policy variables are again jointly significant with the largest effect from the area with the most generous subsidy. In column (4) we present the instrumental variable results. The coefficient on the RSA dummy is much larger than in column (1) suggesting substantial downwards bias in OLS. This is again consistent with RSA being awarded to those firms who face negative shocks.

With heterogeneous treatment effects, it is important to examine whether our estimates of the average effect of treatment on the treated (ATT) could be biased due to a non-overlapping support of the distributions of participants and non-participants. Columns 5 to 8 repeat the same order of specifications as columns 1 to 4; but now the regression sample is limited to firms that are on the same common support through using propensity score matching based on pre-treatment characteristics. In columns 5 to 8 we restrict the sample to those firms that have a predicted propensity score larger than the 25 percentile of the propensity score distribution of participating firms and lower than the 75 percentile of non-participating firms. We are left with about a quarter of the initial sample and 16% of the initial firms by imposing this constraint.

The qualitative results are similar to those presented in columns 1 to 4. There is an increase in the coefficient between OLS and IV, the reduced form and the first stage confirm a monotonic

[34] Note that if the firm has several plants that have been with the firm the longest then we take an average of the eligibility rates of each of these plants' locations.

pattern in the policy eligibility coefficients; the instruments have power and we obtain a significant and positive treatment effect. For employment, we obtain a lower estimate of the treatment effect using the common support restriction than in column 4 (0.33 rather than 0.444). This does suggest some evidence of heterogeneous treatment effects and we regard the lower estimate of 33% as our current preferred estimates.

Table 6 repeats the same order of specifications as Table 5 but uses investment as the dependent variable. The broad pattern of results is similar to that for employment. First, the policy instruments are informative in both the reduced firm and first stage. Secondly, and most importantly, the IV results are positive and significant being larger in magnitude than the OLS results (0.445 in the final column). When in the last column we report the estimates obtained from the common support sample defined as above we find that the magnitude of the RSA coefficient is reduced to 0.445 but not significantly so relative to the estimate obtained from the whole sample – 0.468 - and reported in column 4. This seems to suggest that heterogeneous treatment effects might be less important for investment effects.

Table 7 reports some estimates for production functions. Note that for brevity sake we only report results from estimating the regressions on the common support sample.[35] The dependent variable in all columns is real gross output. In the first 4 columns we control for whether the firm is part of a domestic group or of a foreign group, a quadratic polynomial in firm age, a dummy for firms that entered before 1985 to control for left censoring of the age variable, a full set of four-digit industry dummies, regional dummies and time dummies but we do not control for the inputs of production labor, capital and materials. The coefficient on the RSA dummy is significant in the both the OLS-FE regression of column (1) and the IV-FE regression in column 4 and it confirm the downward bias of the OLS-FE with the IV-FE estimates being more than three times the OLS estimates. This likely reflects the positive effect of the policy on the size of the business now measured in terms of gross output. The first stage and the reduced form are very much in line with the results reported in the previous tables. The next four columns control for capital; employment and materials. Once we do this, the RSA variable is insignificant even after using instrumental variables. The main reason for this is conditioning on capital. We saw from Table 6 that RSA increases investment, once we control for this effect there is no impact of RSA on productivity.

[35] Estimates on the whole sample are available from the authors upon request.

We confirmed the absence of an RSA effect on total factor productivity (TFP) in a number of ways such as calculating TFP as a residual and regressing this on (instrumented) RSA. We also estimated a value added per worker regression instead of a gross output regression.

In summary, our basic results suggest that RSA has a causal effect in increasing the employment of participating plants and firms and the investment of participating firms. After controlling for the increase in these factor inputs, however, the policy has no impact on productivity.

VB. Firm Level Analysis – Robustness Tests

(a) Continuous Treatment Intensity

[TO be completed]

(b) Heterogeneous Treatment Effects

Tables 5 to 7 examine whether our estimates of the average effect of treatment on the treated (ATT) could be biased due to a non-overlapping support of the distributions of participants and non-participants due to heterogeneous treatment effects. We do so by restricting ourselves to having a common support through using propensity score matching based on pre-treatment characteristics keeping those firms that have a predicted propensity score larger than the 25th percentile of the propensity score distribution of participating firms and lower than the 75th percentile of non-participating firms. In tables 5 to 7 we have shown that this constraint leaves us with a quarter of the initial sample, but that across all the dependent variables – employment; investment and gross output - the qualitative results are similar to those obtained using the whole sample. There is an increase in the coefficient between OLS and IV, the instruments have power and we obtain a significant and positive treatment effect. Generally, we obtain a lower estimate of the treatment effect using the common support restriction thus suggesting some heterogeneous treatment effects; especially for employment effects.

(c) Plants in Single vs. multi plant firms

To investigate further the importance of heterogeneous effects we analyze the difference in the treatment effect across two groups of plants: plants that are single establishments and plants that are part of a larger group. About twenty percent of plants in our data are part of multiplant firms. Table 8 reports the estimates of equation 5 with log employment as dependent variable for the single plants (left panel columns 1 to 4) and for plants of multiplant firms (right panel columns 5 to 8). We start by comparing column 1 and 5 and note that there is no difference in the OLS_FE estimate of the treatment effect. Similarly the first stage estimates are also very similar between the two groups of firms (column 3 vs. 7). However, both the reduced form coefficient and the IV estimates differ starkly across the two groups of firms: for the subsample of single plants both the reduced form and the IV coefficients of columns 3 and 4 are positive and strongly significant with the IV being significantly larger than the OLS-FE estimates as in the previous tables. In columns 6 and 8 the estimated reduced form and IV coefficients are not significantly greater than zero and the IV point estimates is smaller than the OLS-FE estimate. This suggests that amongst multiplant firms it is not the plants hit by a negative shock that get the grant and/or that there is very little evidence for additionality for these plants.

VC. Area-level analysis

The firm level regressions reported earlier suggest that the causal effect of RSA is to increase employment both at the plant and firm level and investment but not total factor productivity. A problem with this identification strategy is that it relies entirely on within incumbent firm variation and therefore cannot capture the effect of RSA on firm entry.[36] Another issue is that we only capture the direct effect of RSA on participating plants in the eligible areas but not the indirect impact of RSA on the non-participating firms in the eligible areas. RSA support might give supported firms a competitive edge so that their market share in the eligible areas increases at the expense of other plants.[37] Finally, we aggregated from plants in multiplant firms where we don't observe any employment effect to the level of (multiplant) firm where we see a positive employment effect because of the increase in the number of plants belonging to the firm rather than in the employment of incumbent plants belonging to the firm. However, this increase in employment in multiplant firms might not translate in an increase of employment/number of plants at the ward level if the increase in the number of plants derives from an increase in the takeover activity of the multiplant firms. We will investigate these issues below.

We make a first attempt to estimate the net impact of both of these effects at the level of the ward (henceforth "area"). We therefore estimate equation (6) where the outcome of interest that

[36] We do however capture plant entry within existing reporting units. In fact initial results seem to suggest that there is a difference between the effect on employment through increase in employment at incumbent plants within the firm and the opening up of new plants in existing firms.

[37] A further type of indirect effect is the impact of RSA on non-eligible areas. Part of this is captured when we look at the firm level, but part may be missed due to (e.g.) lower entry in non-eligible areas due to RSA. We are examining these effects by looking at the effect of RSA changes to plants in *neighbouring* areas.

we focus on is aggregate employment and the number of plants at the area level (investment is not available at the local unit level). Our treatment variables become the number of treatments an area receives at a given point in time and the amount of support money that flows in the area. Table 2 reports descriptive statistics at the area level in terms of number of treatments and amount of financial support received expressed also in average Net Grant Equivalent terms for all eligible areas in our sample over the period analysed 1986 to 2004.

Table 9 reports reduced form regressions for this type of analysis; i.e. the explanatory variable of interest becomes the support intensity (Net Grant Equivalent rate, NGE) of an area. In column (1) we find a positive and significant effect on area employment. The coefficient implies that increasing the maximum support level (NGE) by ten percentage points is associated with a 2.3 percent increase in employment. The bottom panel of Table 9 uses the total number of plants as the outcome – RSA appears to also increase the number of plants in an area. When we distinguish between single plants firms and plants belonging to multiplant firms we find that the significant positive effects only remains when looking at the single plants while we cannot find any significant effect on employment/active number of plants for multiplant firms.

In Table 10 we continue the analysis at the area level and we analyse how the increase in employment is driven by an increase in the share of net entry employment (defined as the difference between employment in entry plants at time t and exiting plants at the end of t-1) and employment in incumbent plants. We find that overall there is an increase in the share of net entry employment at the area level. When we separately analyse war level employment in single vs. multiplants we find that the share between net entry and incumbents do not change for single plants suggesting that both increase proportionally. In the last panel we report the estimates for plants of multiplant firms: the results suggest that although the overall contribution to war level employment is insignificant there is an increase in the share of employment in net entry relative to incumbents.

The results as a whole tell a very straightforward story. RSA increases employment and investment[38] for the plants/firms who receive it. This is disguised in the standard OLS results because the firms tend to be subject to negative unobservable (to the econometrician) shocks causing a downwards bias on the treatment effect. We find some evidence that RSA increased the employment and number of plants in an area, our results also suggest that the employment

[38] In unreported analysis we also look at the impact of RSA on survival of firms. However, we did not report the results of the analysis here as we could not control for unobserved heterogeneity and therefore the results might be biased. However the results obtained without controlling for unobserved heterogeneity suggest that RSA treatment is positively correlated with survival of treated firms.

effects on incumbents are weaker than any positive effect on net entry. This area level growth is more from new entry than incumbents. Consequently, since there is no productivity effect from receiving RSA grants on treated firms, the fact that they are relatively large with low productivity prior to treatment implies that RSA dampens reallocation effects from more productive to less productive plants. This is likely to dampen aggregate productivity growth both in the affected areas and in the economy as a whole. The results at the area level also confirm that the reduced form effect of the policy is much larger for single firms than for plants that belong to multi plant firms whose contribution to area level employment is not significantly different from zero.

VI. CONCLUSIONS

There are surprisingly few micro-econometric analyses of the causal effects of industrial policy, despite their ubiquity in policy making. In this paper we have examined one business support policy – Regional Selective Assistance. We use exogenous changes in the eligibility of businesses to receive support driven by policy changes at the European level. These changes are based on areas and are exogenous to firm characteristics and the party in power in the UK.

We find that OLS and matching techniques that fail to account for the endogenous selection of firms contain a large downward bias. When we correct for this we find evidence for a positive effect of treatment on the treated in terms of employment and investment. We find no effects on (total factor) productivity, however. When we investigate possible heterogeneity in the effects of the policy we find that the employment effect of the policy is much stronger for single plant firms than for plants belonging to multiplant firms. This might indicate that the latter might be more successful at gaming the system. Finally, since participants tend to have below average productivity and RSA helps them to expand in size this might yield a negative effect on aggregate productivity. When we analyze the effects of the policy at the area level we find a positive effect of the policy on employment; we also find that this effect is mainly driven by increase in employment in single plant firms and that this increase in employment is associated with a n increase in the share of net entry employment. In terms of further work we aim to investigate more carefully the indirect effects on other plants outside the eligible areas. Secondly, we intend to use the policy to obtain credible identification of structural parameters in the production function. The coefficient on capital in firm or plant-level production functions is difficult to estimate due to selection and endogeneity problems[39]. Investment subsidies can be an external instrument that shifts the capital stock exogenously under the assumption that RSA does not have a direct effect on TFP (consistent with what we are finding here). Finally, we have not attempted a full cost-benefit analysis, due to concerns that we have not incorporated general equilibrium effects. Nevertheless we plan some simple policy simulations in future work.

[39] See inter alia Marschak and Andrews (1944), Griliches and Mairesse (1998); Olley and Pakes (1996) and Ackerberg et al (2007).

REFERENCES

Ackerberg, Daniel, A.; Kevin Caves and Garth Frazer (2007) "Structural Identification of Production Functions", available from http://www.econ.ucla.edu/ackerber/

Angrist, Joshua D. (2004) "Treatment effect heterogeneity in theory and practice" *Economic Journal*, vol. 114(494), pp. C52-C83.

Armstrong, Harvey W. (2001) "Regional Selective Assistance: Is the Spend Enough and Is It Targeting the Right Places?" *Regional Studies*, vol. 35(3) May, pp. 247-57.

Beason Richard and David E. Weinstein (1996) "Growth, Economies of Scale and Targeting in Japan (1955-1990)", *Review of Economics and Statistics*, 78(2), pp. 286-295.

Blundell, Richard, Monica Costa Dias, Costas Meghir and John Van Reenen (2004) "Evaluating the Employment Impact of a Mandatory Job Search Program" *Journal of the European Economic Association*, vol. 2(4), pp. 569-606.

Card, David (2002), Handbook of Labour Economics

Criscuolo Chiara; Jonathan Haskel and Ralf Martin, (2003) "Building the evidence base for productivity policy using business data linking", *Economic Trends* vol. 600, November, pp. 39-51, www.statistics.gov.uk/articles/economic_trends/ETNov03Haskel.pdf.

Department of Trade and Industry (2001) "DEPARTMENT OF TRADE AND INDUSTRY (2001). The government's expenditure plans, 1999–2000, the regions, p.41

Devereux, Griffith and Simpson (2007)

Foster, Lucia; John Haltiwanger and C.J. Krizan (1998), "Aggregate productivity growth: lessons from microeconomic evidence", NBER Working Paper (6803).

Griffith Rachel S. (1999) "Using the ARD establishment level data to look at foreign ownership and productivity in the UK." *Economic Journal* vol. 109, pp F416-F442.

Griliches, Zvi (1979) "Issues in Assessing the Contribution of Research and Development to Productivity Growth" *Bell Journal of Economics*, The RAND Corporation, vol. 10(1), pp. 92-116, Spring.

Griliches, Zvi, and Jacques Mairesse (1998) "Production Functions: The Search for Identification," in Z. Griliches, *Practicing Econometrics: Essays in Method and Application*, Cheltenham, UK: Elgar.

Harrison, Ann E. (1994) "An empirical Test of the Infant Industry Argument: Comment", The *American Economic Review*, 84(4), pp.1090-1095.

Heckman, James J.; Hidehiko Ichimura, and Petra Todd (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program", *Review of Economic Studies*, vol. 64, pp. 605-654.

Heckman, James J., Hidehiko Ichimura, Jeffrey A. Smith, and Petra Todd (1998) "Characterizing selection bias using experimental data" *Econometrica* vol. 66, pp.1017–1098.

Heckman, James J.; Robert, J. LaLonde and Jeffrey A. Smith (1999) "The Economics and Econometrics of Active Labor Market Programs" in Handbook of Labor Economics, Volume 3 O. Ashenfelter and D. Card eds., pp. 1865-2097.

Heckman, James J., Jeffrey A. Smith, Nancy Clements (1997). "Making the most out of social experiments: accounting for heterogeneity in programme impacts" *Review of Economic Studies* vol. 64, pp. 487–535.

Holmes, Thomas (1998) “The Effects of State Policies on the Location of Industry: Evidence from State Borders,” *Journal of Political Economy* 106, (4), August, pp. 667-705.

Imbens, Guido W and Angrist, Joshua D. (1994) “Identification and Estimation of Local Average Treatment Effects” *Econometrica*, Econometric Society, vol. 62(2), pp. 467-75, March.

Jones, Jonathan and Colin Wren (2004) “Do Inward Investors Achieve their Job Targets?” *Oxford Bulletin of Economics and Statistics*, vol. 66(4), pp. 483-514.

Jones, Jonathan and Colin Wren (2004) “Inward Foreign Direct Investment and Employment: A Project-Based Analysis in North-East England” *Journal of Economic Geography*, vol. 5(4), pp. 517-44.

Irwin, Douglas A. and Pete Klenow (1996) "High-Tech R&D Subsidies: Estimating the Effects of Sematech" *Journal of International Economics* 40, May, pp.323-344.

Klette Tor Jakob; Jarle Møen and Zvi Griliches (2000) “Do subsidies to commercial R&D reduce market failures?” *Research Policy* vol. 29, 4-5, pp. 471-496.

Krueger Ann O. and Baran Tuncer (1982) “An empirical test of the Infant Industry Argument”, *American Economic Review* 72(5), pp. 1142-1152.

Lawrence, Robert Z. and David E. Weinstein (2001) “Trade and Growth: Import Led or Export Led? Evidence from Japan and Korea” in Joseph E. Stiglitz and Shahid Yusuf eds., *Rethinking the East Asia Miracle*, Oxford: Oxford University Press.

Marschak, Jacob and Andrews, William H. (1944) “Random Simultaneous Equations and the Theory of Production” *Econometrica*, July-October, vol. 12(3-4): pp. 143-205

Olley, Steve and Ariel Pakes (1996) “The dynamics of Productivity in the Telecommunications equipment industry”, *Econometrica* vol. 64 (6), pp. 1263-1297

Oulton Nicholas (1997) The ABI Respondents Database: a new resource for industrial economics research. *Economic Trends* vol. 528, pp. 46–57.

Rodrik, Dani (2007) “Normalizing Industrial Policy”, mimeo, Harvard University, September.

Roome, Marjorie (2005),'Linking DBERR Business Support Schemes data to the IDBR and ARD with implications for the future use of the administrative data in scheme assessment', DBERR report.

Stock, James H. and Motohiro Yogo (2002) “Testing for weak instruments in linear regressions”, NBER Technical Working Paper 284 http://www.nber.org/papers/T0284

Wren, Colin (2005) “Regional Grants: Are They Worth It?” *Fiscal Studies*, 26(2), pp. 245-75.

Wren, Colin (2003) “Investment Scale as a Signal in Industrial Assistance Schemes with Employment Objectives” *Economica*, vol. 70, pp. 331-52.

Wren, Colin and David Storey (2002). “Evaluating the Effect of ‘Soft’ Business Support upon Small Firm Performance” *Oxford Economic Papers*, vol. 54, pp. 334-65.

Wren, Colin and J. Taylor. (1999) “Industrial Restructuring and Regional Policy” *Oxford Economic Papers*, vol. 51, pp. 487-516.

Table 1: Summary statistics for RSA participants and non-participants

		(1)		(2)	(3)	(4)
Variable		mean		Sd	median	Obs.
Plant Employment	non treated	22.25	***	118.92	2	3,193,504
	Treated before	79.39		241.45	6	136,488
Firm Employment	non treated	253	***	737	111	145389
	Treated before	417		957	171	8209
Gross output	non treated	26774	***	136448	6622	136524
	Treated before	39401		151614	10256	7247
Investment	non treated	1082.76	***	8471.20	147.70	145382
	Treated before	1624.35		7204.89	310.03	8209
Real Value added per worker	non treated	31.05	**	162.51	24.27	136524
	Treated before	26.32		23.51	22.38	7247
Total Factor Productivity	non treated	0.02	***	0.33	0.01	134755
	Treated before	-0.03		0.29	-0.03	7925

Notes: Column 1, mean, reports the mean of the variables of interest separately for non treated plants/firms and for the group of treated for the period before treatment. We also report the significance of a t-test of equality between the values for treated plants/firms before treatment relative to the group of non treated plants/firms. Column 2, "sd" reports standard deviations, while column (3) reports the median of the variables for non-treated and treated plants/firms in the pre-treatment period. Finally, column 4 reports the number of observations for each cell.

Source: Authors' calculation using the IDBR ARD SAMIS matched data.

Table 2: Descriptive statistics across areas (WARDS)

	Share Eligible wards (%)	Plants in RSA		share of employment in RSA		Eligible wards	
		All wards	Eligible wards	All wards	Eligible wards	Average RSA Payments	NGE rate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1985	29.57	0.055	0.168	0.007	0.017	40615	0.248
1986	29.57	0.069	0.225	0.009	0.027	59242	0.248
1987	29.57	0.096	0.316	0.012	0.039	64001	0.248
1988	29.57	0.140	0.468	0.014	0.046	75164	0.248
1989	29.57	0.119	0.391	0.014	0.047	74665	0.248
1990	29.57	0.088	0.296	0.014	0.046	55627	0.248
1991	29.57	0.106	0.348	0.016	0.053	52690	0.248
1992	29.57	0.094	0.314	0.017	0.056	49308	0.248
1993	29.57	0.103	0.338	0.018	0.059	63077	0.248
1994	32.00	0.122	0.345	0.017	0.049	57037	0.241
1995	32.00	0.131	0.391	0.017	0.051	46537	0.241
1996	32.00	0.138	0.432	0.019	0.057	62401	0.241
1997	32.00	0.137	0.432	0.019	0.060	71329	0.241
1998	32.00	0.114	0.361	0.018	0.055	64976	0.241
1999	32.00	0.104	0.327	0.015	0.047	47968	0.241
2000	32.00	0.093	0.304	0.013	0.045	44259	0.237
2001	26.27	0.069	0.235	0.010	0.035	43902	0.237
2002	26.27	0.045	0.161	0.009	0.030	45901	0.237
2003	26.27	0.026	0.094	0.006	0.020	24063	0.237
2004	26.27	0.018	0.066	0.004	0.014	14788	0.237
2005	26.27	0.012	0.048	0.002	0.008	7559	0.237
Total	29.32	0.089	0.296	0.013	0.042	51584	0.243
Obs	226044	216901	61782	216873	61768	61782	66285

Notes: Column (1) reports the share of wards that are eligible to RSA in each year. Column 2 reports the average number of firms that are receiving RSA grants in each year; Column 3 reports the average number of firms that are receiving RSA grants in each year and restricts the sample only to eligible areas. Columns (3) and (4) report the average share of ward employment in plants that receive RSA grants both for all wards (column 3) and for the subsample of eligible wards (column 4). Column 5 and 6 report the average annualised deflated RSA grants values given within eligible areas and the average NGE rate. Note that there are 10,764 areas (wards) in Great Britain.

Source: Authors' calculation using the IDBR ARD SAMIS matched data.

Table 3: Source of identification

		(1)	(2)
	year	**Change in eligibility**	**Increase in eligibility**
Plants	1993	23225	14369
	2000	50920	14967
Firms	1993	19866	12505
	2000	45692	13520
Firms ARD Sample	1993	864	502
	2000	2057	672
Areas (wards)	1993	1893	1034
	2000	4048	1424

Notes: The first two rows in the table report the number of plants ("local units") in the business register (IDBR) for which a change in eligibility (column 1) is observed in 1993 (row 1) and in 2000 (row 2); column 2 reports the number of plants that have experienced an increase in their eligibility rate. For example, in 1993 there were 23,225 changes in support status at the plant level. 62% of these changes (14,369/23,225) were increases in eligibility rates. The second panel reports the same figures at the firm level and the third panel restricts the sample of firms to the ARD sample. Finally, the last two rows in the bottom panel report these statistics at the areas (ward) level.
Source: Authors' calculation using the IDBR ARD SAMIS matched data.

Table 4: ln(Employment) Regressions: plant level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**ln(EMP)**	**ln(EMP)**	**RSA**	**ln(EMP)**	**ln(EMP)**	**ln(EMP)**	**RSA**	**ln(EMP)**
	OLS	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
RSA	1.435***			3.296***	0.152***			0.306***
	(0.027)			(0.142)	(0.007)			(0.057)
0<NGE<=20%		0.265***	0.074***			0.019***	0.049***	
		(0.018)	(0.002)			(0.004)	(0.002)	
NGE>20%		0.354***	0.113***			0.021***	0.082***	
		(0.021)	(0.004)			(0.005)	(0.003)	
Observations	2903819	2903819	2903819	2903819	2903819	2903819	2903819	2903819
Number of clusters	10703	10703	10703	10703	10703	10703	10703	10703
Number of plants	494832	494832	494832	494832	494832	494832	494832	494832
P-value for F test of excluded instruments			p<0.001				p<0.001	
Fixed effects	NO	NO	NO	NO	YES	YES	YES	YES

Notes: Dependent variable is Log employment in all columns except columns 3 and 7 where the dependent variable is RSA. *RSA* equals unity for all the periods in and after a plant has participated in the program and zero otherwise. *NGE* is Net Grant Equivalent (maximum investment subsidy) at the area-level. Eligibility for investment subsidies used as an instrumental variable in columns (4) and (8). *NGE = x%* indicates that the plant is in an area that is eligible for up to x% in investment subsidy. All columns include a full set of time dummies. Standard errors below coefficients are robust to heteroscedacity and arbitrary serial correlation (they are clustered by area in all columns). The last four columns also include a full set of plant dummies. Time period is 1985-2004. In columns (2) (3) (6) and (7) the reference category is NGE=0.

Source: Authors' calculation using the IDBR ARD SAMIS matched data.

Table 5: ln(Employment) Regressions; firm level ARD sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All ARD sample				Common Support 25-75			
Dependent Variable	**ln(EMP)**	**Ln(EMP)**	**RSA**	**ln(EMP)**	**ln(EMP)**	**ln(EMP)**	**RSA**	**ln(EMP)**
	FE	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
RSA	0.141***			0.444***	0.120***			0.330**
	(0.010)			(0.067)	(0.020)			(0.132)
0<NGE<=20%		0.052***	0.120***			0.045**	0.138***	
		(0.010)	(0.006)			(0.019)	(0.012)	
NGE>20%		0.080***	0.178***			0.061**	0.185***	
		(0.013)	(0.008)			(0.028)	(0.016)	
Observations	154715	154715	154715	154485	37943	37943	37943	37925
Number of clusters	6034	6034	6034	6014	2327	2327	2327	2324
Number of firms	28313	28313	28313	28313	4534	4534	4534	4534
F-stats for excluded instruments			354.023				96.006	
Fixed effects	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Dependent variable is Log employment in all columns except columns 3 and 7 where the dependent variable is RSA. *RSA* equals unity for all the periods in and after a firm has participated in the program and zero otherwise. *NGE* is Net Grant Equivalent (maximum investment subsidy) at the area-level. Eligibility for investment subsidies used as an instrumental variable in columns (4) and (8). *NGE = x%* indicates that the firm has the oldest plant(s) in an area that is eligible for up to x% in investment subsidy. All columns include controls for whether a firm is foreign owned, whether it is part of a domestic multi-firm group, a quadratic in age; an age censoring dummy for firms born before 1985 and a full set of four digit industry dummies, regional and time dummies. All columns also include a full set of firm dummies. Standard errors below coefficients are robust to heteroscedacity and arbitrary serial correlation (they are clustered by area in all columns). Columns 1 to 4 include all firms in the ARD sample. Columns 5 to 8 restrict the sample to those firms that have a predicted propensity score larger than the 25th percentile of the propensity score distribution of participating firms and lower than the 75th percentile of non-participating firms. Time period is 1985-2004. In columns (2) (3) (6) and (7) the reference category is NGE=0.

Source: Authors' calculation using the IDBR ARD SAMIS matched data.

Table 6: ln(Real Investment) Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Whole sample				Common Support 25-75		
Dependent Variable	**ln(INV)**	**ln(INV)**	**RSA**	**ln(INV)**	**ln(INV)**	**ln(INV)**	**RSA**	**ln(INV)**
	OLS	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
RSA	0.190***			0.468***	0.176***			0.445**
	(0.021)			(0.120)	(0.038)			(0.225)
0<NGE<=20%		0.036*	0.123***			0.048	0.139***	
		(0.019)	(0.006)			(0.036)	(0.012)	
NGE>20%		0.100***	0.180***			0.095**	0.183***	
		(0.024)	(0.008)			(0.046)	(0.016)	
Observations	140667	140667	140667	140667	35707	35707	35707	35601
Number of clusters	5955	5955	5955	5955	2326	2326	2326	2304
Number of firms	27691	27691	27691	27691	4533	4533	4533	4533
F-stats for excluded instruments			334.757				91.747	
Fixed effects	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Dependent variable is Log real investment in all columns except columns 3 and 7 where the dependent variable is RSA. *RSA* equals unity for all the periods in and after a firm has participated in the program and zero otherwise. *NGE* is Net Grant Equivalent (maximum investment subsidy) at the area-level. Eligibility for investment subsidies used as an instrumental variable in columns (4) and (8). *NGE = x%* indicates that the firm has the oldest plant(s) in an area that is eligible for up to x% in investment subsidy. All columns include controls for whether a firm is foreign owned, whether it is part of a domestic multi-firm group, a quadratic in age; an age censoring dummy for firms born before 1985 and a full set of four digit industry dummies, regional and time dummies. All columns also include a full set of firm dummies. Standard errors below coefficients are robust to heteroscedacity and arbitrary serial correlation (they are clustered by area in all columns). Columns 1 to 4 include all firms in the ARD sample. Columns 5 to 8 restrict the sample to those firms that have a predicted propensity score larger than the 25th percentile of the propensity score distribution of participating firms and lower than the 75th percentile of non-participating firms. Time period is 1985-2004. In columns (2) (3) (6) and (7) the reference category is NGE=0.

Source: Authors' calculation using the IDBR ARD SAMIS matched data.

Table 7: Productivity Regressions
ln(Real Gross Output)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Common Support 25-75				**Common Support 25-75**			
Dependent Variable	**ln(GO)**	**ln(GO)**	**RSA**	**ln(GO)**	**ln(GO)**	**ln(GO)**	**RSA**	**ln(GO)**
	FE	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
RSA	0.079***			0.331***	-0.003			0.025
	(0.020)			(0.123)	(0.009)			(0.078)
0<NGE<=20%		0.046**	0.134***			0.003	0.105***	
		(0.018)	(0.013)			(0.010)	(0.019)	
NGE>20%		0.059**	0.181***			0.004	0.145***	
		(0.025)	(0.017)			(0.012)	(0.025)	
ln(Capital)					0.012***	0.012***	0.001	0.012***
					(0.001)	(0.001)	(0.002)	(0.001)
ln(Materials)					0.499***	0.499***	0.020***	0.498***
					(0.012)	(0.012)	(0.008)	(0.012)
ln(Employment)					0.336***	0.336***	0.025***	0.335***
					(0.014)	(0.014)	(0.009)	(0.014)
Observations	36351	36351	36351	36329	24403	24403	24403	24141
Number of clusters	2326	2326	2326	2322	2295	2295	2295	2210
Number of firms	4532	4532	4532	4532	4445	4445	4445	4445
F-stats for excluded instruments			83.849				19.26	
Fixed effects	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Notes: Dependent variable is Log real gross output in all columns except columns 3 and 7 where the dependent variable is RSA. *RSA* equals unity for all the periods in and after a firm has participated in the program and zero otherwise. *NGE* is Net Grant Equivalent (maximum investment subsidy) at the area-level. Eligibility for investment subsidies used as an instrumental variable in columns (4) and (8). *NGE = x%* indicates that the firm has the oldest plant(s) in an area that is eligible for up to x% in investment subsidy. All columns include controls for whether a firm is foreign owned, whether it is part of a domestic multi-firm group, a quadratic in age; an age censoring dummy for firms born before 1985 and a full set of four digit industry dummies, regional and time dummies. All columns also include a full set of firm dummies. Standard errors below coefficients are robust to heteroscedacity and arbitrary serial correlation (they are clustered by area in all columns). All columns report estimates on a sample of firms that have a predicted propensity score larger than the 25th percentile of the propensity score distribution of participating firms and lower than the 75th percentile of non-participating firms. Time period is 1985-2004.

Source: Authors' calculation using the IDBR ARD SAMIS matched data.

Table 8: ln(Employment) – single-plants vs. multi-plants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Single plants				Plants of multiplant firms			
Dependent Variable	**ln(EMP)**	**ln(EMP)**	**RSA**	**ln(EMP)**	**ln(EMP)**	**ln(EMP)**	**RSA**	**ln(EMP)**
	FE	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
RSA	0.156***			0.474***	0.169***			0.097
	(0.008)			(0.070)	(0.014)			(0.089)
0<NGE<=20%		0.024***	0.043***			0.012	0.069***	
		(0.004)	(0.002)			(0.008)	(0.004)	
NGE>20%		0.029***	0.068***			0.009	0.122***	
		(0.006)	(0.003)			(0.012)	(0.005)	
Observations	2329112	2329112	2329112	2329112	574707	574707	574707	574707
Clusters	10681	10681	10681	10681	8533	8533	8533	8533
Firms	417419	417419	417419	417419	81074	81074	81074	81074
F-stats for excluded instruments			1084.513				522.444	
Fixed effects	YES	YES	YES	YES	YES	YES	YES	YES

Notes: *RSA* equals unity for all the periods in and after a firm has participated in the program and zero otherwise. *NGE* is Net Grant Equivalent (maximum investment subsidy) at the area-level. Eligibility for investment subsidies used as an instrumental variable in columns (4) and (8). *NGE = x%* indicates that the plant is in an area that is eligible for up to x% in investment subsidy. All columns include a full set of time dummies and plant fixed effects. Standard errors below coefficients are robust to heteroskedacity and arbitrary serial correlation (they are clustered by area). Time period is 1985-2004. Columns 1 to 4 restrict the sample to all plant that have always been single plants; columns 5 to 8 restrict the sample to firms that have been at least once owned by a multiplant firm.

Source: Authors' calculation using the IDBR ARD SAMIS matched data.

Table 9: Regressions at the ward level: employment; number of plants

	(1)	(2)	(3)
	Log employment		
	All	Single Plants	Multi plants
NGE	0.233**	0.493***	-0.011
	(0.115)	(0.122)	(0.214)
Observations	186267	184872	104638
Number of areas	10737	10725	8818
	Log number of plants		
	All	Single Plant	Multi plants
NGE	0.214***	0.275***	-0.037
	(0.054)	(0.056)	(0.083)
Observations	186294	184902	104791
Number of areas	10737	10725	8828

Notes: *NGE* is Net Grant Equivalent (maximum investment subsidy) at the area-level (ward). *NGE = x%* indicates that plants in that area are eligible for up to x% in investment subsidy. All columns include a full set of time dummies and ward fixed effects. Standard errors below coefficients are robust to heteroscedacity and arbitrary serial correlation (they are clustered by area). Time period is 1986-2004. Single plants employment is defined as employment aggregated from plants that were single plant firm in the previous period; Multi plants employment/number of plants is aggregated from plants that in the previous period belonged to multiplant firms.

Table 10: Regressions at the ward level: entry and exit shares

	(1)	(2)	(3)
	Employment	Net entry Share	Incumbent Share
NGE	0.233**	0.289*	-0.217
	(0.115)	(0.157)	(0.156)
Observations	186267	184386	184386
Areas	10737	10713	10713
		Single Plants	
NGE	0.493***	0.037	-0.018
	(0.122)	(0.067)	(0.082)
Observations	184872	184386	184386
Areas	10725	10713	10713
		Multiplants	
NGE	-0.011	0.252*	-0.199
	(0.214)	(0.142)	(0.139)
Observations	104638	184386	184386
Areas	8818	10713	10713

Notes: *NGE* is Net Grant Equivalent (maximum investment subsidy) at the area-level (ward). *NGE = x%* indicates that plants in that area are eligible for up to x% in investment subsidy. All columns include a full set of time dummies and ward fixed effects. Standard errors below coefficients are robust to heteroscedacity and arbitrary serial correlation (they are clustered by area). Time period is 1986-2004. Single plants employment is defined as employment aggregated from plants that were single plant firm in the previous period; Multi plants employment/number of plants is aggregated from plants that in the previous period belonged to multiplant firms. Net entry share is the share of ward employment in new plants in year t (entrants in year t) – employment in firms that exit between t-1 and t (exitors at time t-1). Incumbent share of employment in the area is the share of employment in firms that existed areald in t-1.

Table 11: Unemployment regressions Area level

[to be completed]

Table 12: ln(Employment) - Continuous Treatment Intensity

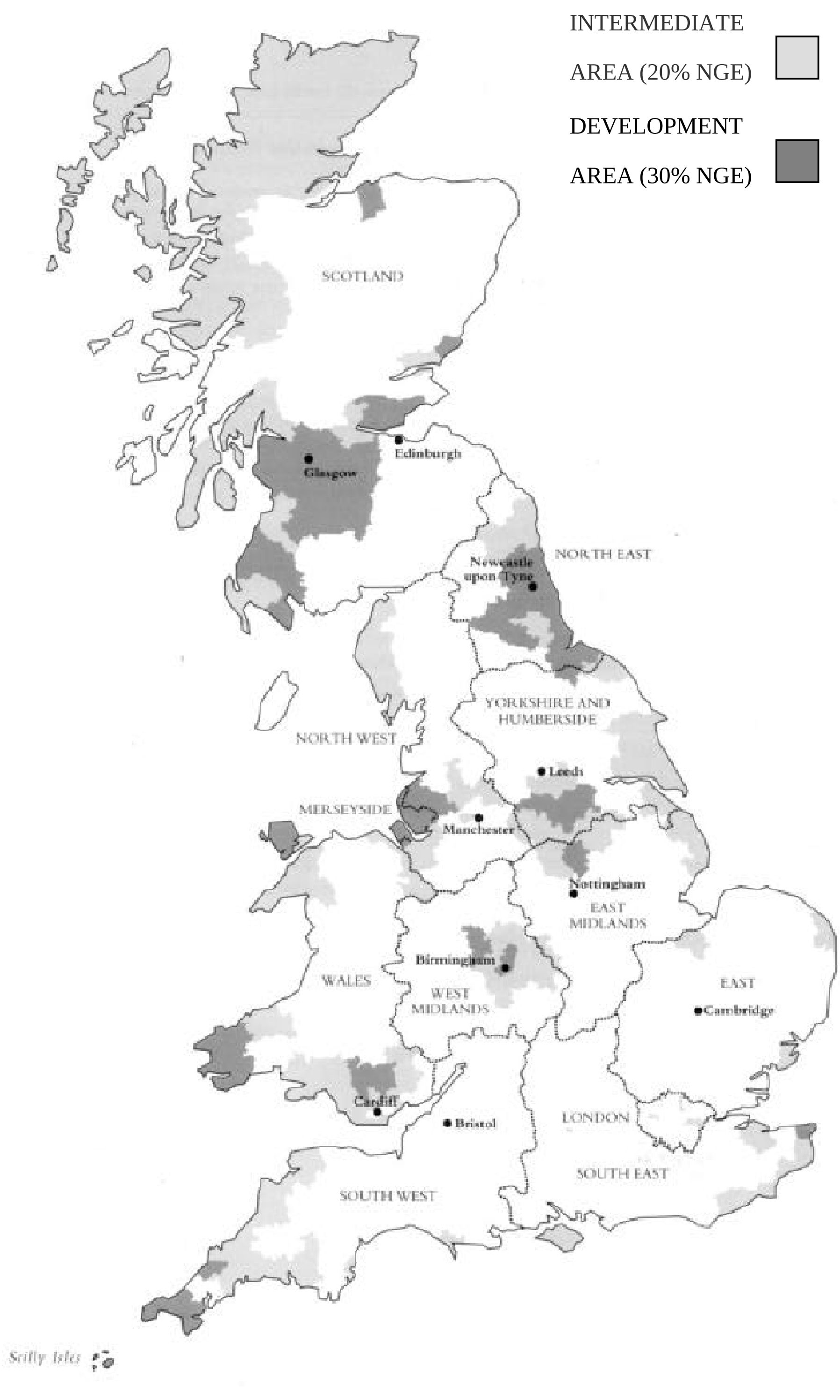
[to be completed]

Notes: *NGE_RSA* is the investment subsidy received by the firm (total payment divided by total project costs). *NGE* is Net Grant Equivalent (maximum investment subsidy) at the area-level. Eligibility for investment subsidies used as an instrumental variable in columns (4) and (8). *NGE = x%* indicates that the firm's has a reference plant in an area that is eligible for up to x% in investment subsidy. All columns include controls for whether a firm is foreign owned, whether it is part of a domestic multi-firm group, a quadratic in age; an age censoring dummy for firms born before 1980 and a full set of four digit industry dummies, regional and time dummies. Standard errors below coefficients are robust to heteroskedacity and arbitrary serial correlation (they are clustered by firm in columns (1) and (5) and by area in columns (2), (3) (4), (6), (7) and (8)). The last four columns include a full set of firm dummies. Time period is 1985-2004.

Figure 1: Assisted Areas Map prior to August 1st 1993

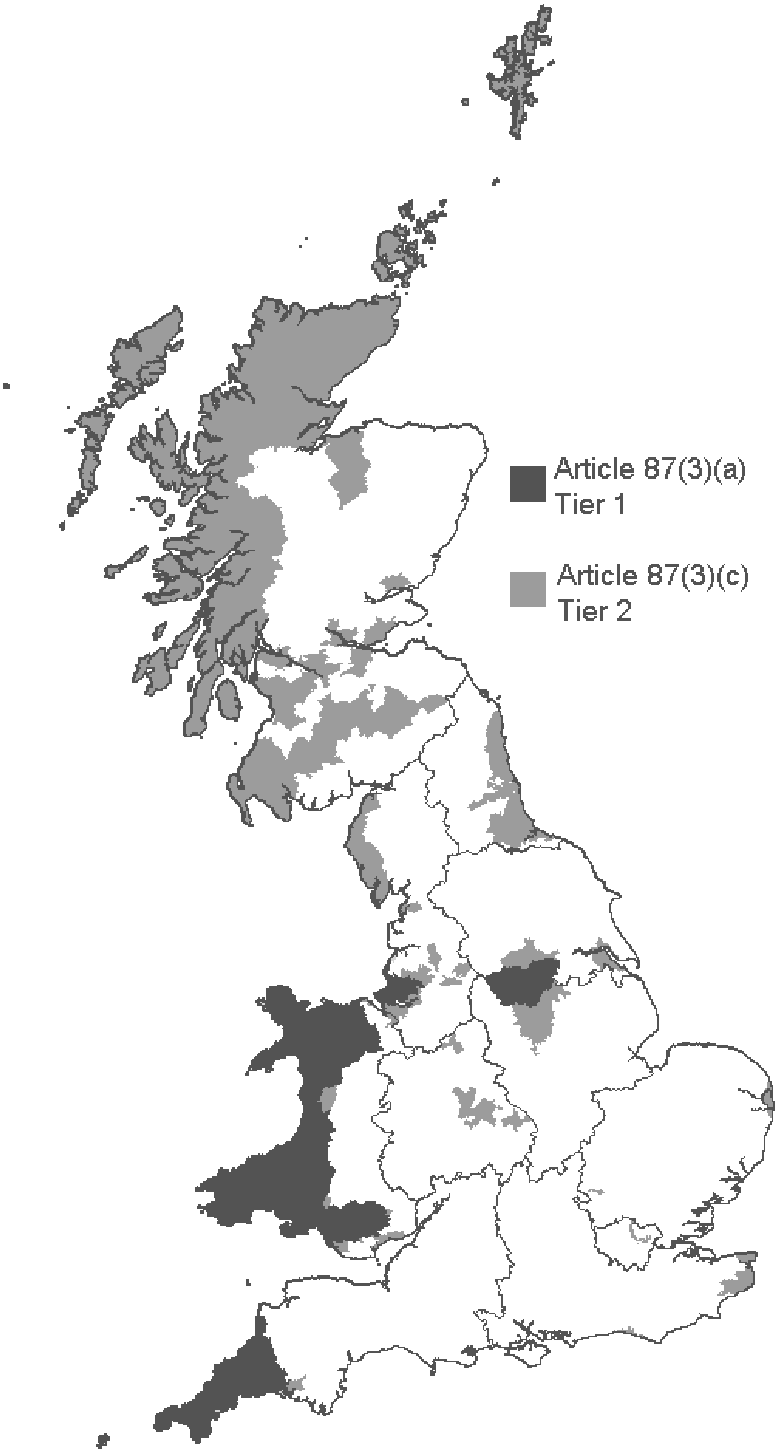
Notes: The shaded areas are those which are eligible for some Regional Selective Assistance. The dark shaded areas are the very deprived areas eligible for an investment subsidy of up to 30% NGE (Net Grant Equivalence). The light shaded areas are eligible for up to 20% NGE.
Source: Department of Trade and Industry

Figure 2: Assisted Areas Map after August 1st 1993 and prior to January 1st 2000



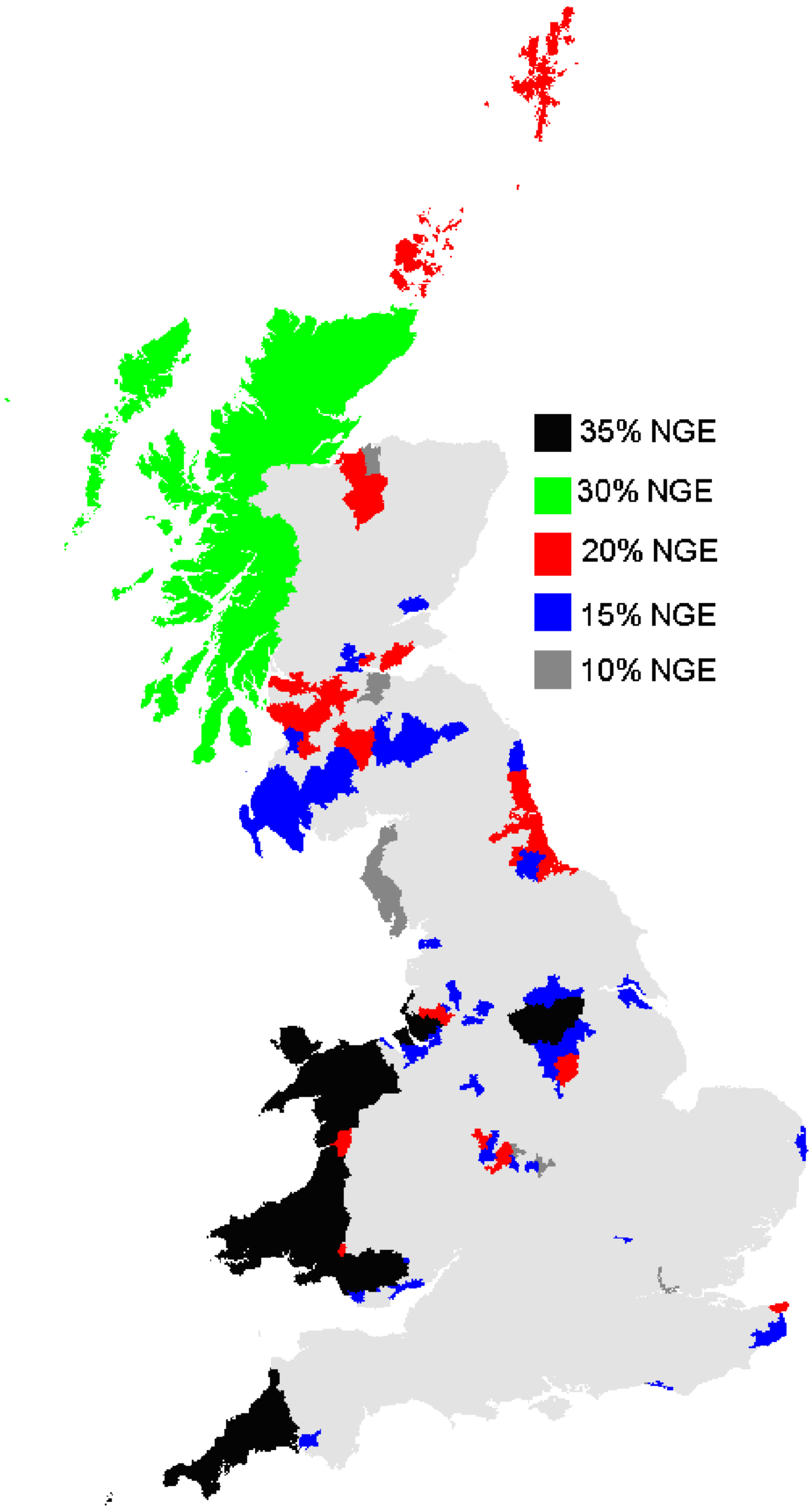
Notes: The shaded areas are those which are eligible for some Regional Selective Assistance. The dark shaded areas are the very deprived areas eligible for an investment subsidy of up to 30% NGE (Net Grant Equivalence). The light shaded areas are eligible for up to 20% NGE.
Source: Department of Trade and Industry

Figure 3: Assisted Areas Map after January 1st 2000



Notes: The shaded areas are those which are eligible for some Regional Selective Assistance.
Source: Department of Trade and Industry

Figure 4: Assisted Areas Map with detailed NGE rates after January 1st 2000



Source: Department of Trade and Industry
Notes: This shows all the different levels of NGE by area

Appendices

A. Details of data matching

Since the performance data comes from data sources unrelated to programme participation several problems arise in matching. Firstly, our dataset with productivity information, the ARD, is a survey with stratified random sampling (i.e. not a census) so programme participants might not be sampled in every year. Secondly, it may be that the company identifier in the program applicants' database does not give a unique match to the ARD.

For any dataset it is useful to keep in mind the unit of observation (i.e. the unit that defines a row in the dataset). The ARD unit of observation is referred to as a "reporting unit" (RU) and provides data on these units on an annual basis. Reporting units are composed of many local units (LUs) and we have population information (on employment and location) on these local units through the Inter-Departmental Business Register (IDBR) which is linked to the ARD. We refer to the RUs as firms and the LUs as plants. Some firms are part of larger "enterprise groups" and this is also recorded (and controlled for in the regressions). The DBERR administrative data on the other hand reports on applicants and participants' in the RSA programme. The key difficulty is to identify which ARD reporting unit or IDBR local unit has potentially been affected by a particular incidence of programme participation.

DBERR uses name and postcode from its administrative data to match a list of participants (and possibly applicants) to the Inter-Departmental Business Register. This matching may occur at the local unit, reporting unit, enterprise unit, and sometimes PAYE unit and Company's House Reference Numbers levels. We exploit the match at each level to get a pointer from the administrative record to the reporting units that potentially could have been affected by the RSA subsidies. There are three possible outcomes: (i) we cannot map the support to any reporting unit in the ARD; (ii) we map the support uniquely to one reporting unit in the ARD or (iii) we map the support to multiple ARD RUs. Figure A1 illustrates this graphically. For each ARD reporting unit that is matched to a record of DBERR support, we then examine if we have sufficient pre- and post- programme participation performance data to perform econometric evaluation analysis[40].

This raises a number of issues. Consider a binary treatment (i.e. simply a one for getting an incident or a zero for not getting it)[41]. Firstly, as Figure A1 illustrates, even if we have a unique match between an observation in the DBERR's dataset and an ARD reporting unit that unit might report for several local units and only a subset of them may actually be affected by the programme. The fact that it only affects a subset of local units may lead us to understate the programme impact if we do not control for this in the evaluation. Similarly, for records of DBERR support matched with multiple ARD RUs we might have the problem that not all ARD RUs are affected by the programme.

The fundamental problem with these issues is that we generally do not know which case applies, which makes it hard to control for it in an econometric model. RSA should by definition only apply to local units. However, if the head office applies for the RSA on behalf of one of its local units then the administrative dataset would hold the

[40] This is not an issue for the area-level analysis where there are many observations per area (see Figure 2).

[41] These problems are less severe if the treatment is continuous and measured in cash terms. For example, if a firm received an RSA grant of £100,000 then this is equivalent to two grants of £50,000. It is not so obvious how to aggregate a binary treatment (employment weights are the practical solution in many cases).

postcode of the head office. As a consequence we would associate the administrative record with the head office local unit instead of the programme participating local unit.

Our main response to this is to estimate at the reporting unit level where we effectively aggregate over all local units. Being aware of the problem in each case we consider many robustness tests. This can be done for example by looking at differences in evaluation results when using only a sample of RUs with unique match to a treatment unit or by aggregating over several RUs in the case of multiple matches.

In the SAMIS database, there is information on 54,322 program applications, whether or not the application has been successful from 1972 to 2004. Using name, postcode and CRN numbers, the information in the DBERR files was matched to the IDBR for 68% of the cases. The improvement of matching rate over time, shown in Table A1, reflects the fact that the IDBR was introduced in 1994 and does not contain information for "units" that have closed down before 1993. Given the lower match rate in earlier years and fewer observations in the last few years we focus only on data between 1988 through 2003 in the econometric analysis.

Concerning the matching to the IDBR, there are three main issues. First, Name and address identifiers from the DBERR SAMIS database for one investment project are sometimes attributed to several IDBR units. Secondly, there may be only one unique match but the company name and postcode does not necessarily refer to the local unit that is affected by the support scheme. Sometimes, especially if the grant is given to a group to open up a new plant in a deprived area, where the application is likely handled by the Headquarters the recorded address and postcode will not correspond to the plant which got the money. Thirdly, one "IDBR unit" might have applied for and received several grants.

We start by considering all matches as valid. In a few instances "IDBR units" might have received more than one grant. We proceed as follows in those cases: when the IDBR unit has applied several times to RSA we keep this information separately in the data; however if there are more than one application in the same year we consider the sum of these applications (i.e. the total sum awarded) for that year, since our production and employment data is yearly.

Finally, we conducted a detailed comparison of the characteristics of projects and project participants of firms that DBERR matched with IDBR relative to all the projects in the database. We do not report these here (available from the authors) but the analysis shows that the set of "IDBR matches" do not significantly differ from the rest of the projects in the database and this is the case for both unsuccessful and successful applications. The variables we considered are: application amounts, headquarter location, a dichotomous variable which is one if the application was handled by the London office of the DBERR, foreign owned, and a DBERR code that seeks to identify "internationally mobile" jobs.

A.II Matching with ARD Sample

When matching to the ARD sample, the aim is to investigate whether there is sufficient information in the matched RSA-ARD sample to undertake meaningful econometric evaluation. Sample sizes for number of successful firms matched are larger than those for unsuccessful despite the fact that both outcomes are roughly equally likely in the DBERR data.

We find that the numbers appear sufficiently large to examine programme effects with at least a five year window and possibly beyond. From our analysis we conclude that a large fraction of units in our sample are unique matches and while there is quite a significant number of administrative records that are linked to several RUs in most cases only one of these RUs has sampled information in the ARD.

B. Variable Definitions

Regression based TFP

There are numerous ways to obtain a TFP measure, a subject of ongoing debate in the economic literature. For the purpose of this study we experiment with a number of different TFP measures to see if our results are sensitive to the method chosen. The simplest measure we use is regression based TFP. This involves running a regression of gross output per employee deflated using 2-digit sectoral producer price indices (source EUKLEMS database) on capital stock per employee, deflated material inputs using 2-digit sectoral material price indices (source EUKLEMS database) per employee and employment:

$$\ln \frac{GO_{it}}{L_{it}} = \beta_K \ln \frac{K_{it}}{L_{it}} + \beta_M \ln \frac{M_{it}}{L_{it}} + \beta_L \ln L_{it} + TFP_{it} \qquad (1)$$

where GO is gross output, K is capital, M is inputs and L is employment.

The TFP measure is then obtained by calculating the residual from that regression:

$$\widehat{TFP}_{it} = \ln \frac{GO_{it}}{L_{it}} - \beta_K \ln \frac{K_{it}}{L_{it}} - \beta_M \ln \frac{M_{it}}{L_{it}} - \beta_L \ln L_{it}$$

To analyse if treatment has any effects on TFP we could then run a second regression of estimated TFP on the treatment indicator; i.e.

$$\widehat{TFP}_{it} = \delta D_{it} + \varepsilon_{it}$$

It is more convenient, however, to run both regressions in one step as

$$\ln \frac{GO_{it}}{L_{it}} = \beta_K \ln \frac{K_{it}}{L_{it}} + \beta_M \ln \frac{M_{it}}{L_{it}} + \beta_L \ln L_{it} + \delta D_{it} + \varepsilon_{it} \qquad (2)$$

These two procedures are equivalent if D_{it} is not correlated with any of the production factor variables. It is very likely that this condition is not met, however, because treatment may both shift TFP and lead to adjustments of the factors mix. In this case using two steps is not only less convenient but may also give biased results. For this reason, we focus only on the one-step regressions below.

A.III Previous evaluations of the policy

Most of the previous evaluation studies of RSA are based on “industrial survey” techniques where senior personnel of a randomly drawn sample of assisted firms are asked to give their subjective assessment of what the counterfactual situation would have been had they not received the grant (see AEP NERA 2003, Cambridge Economics). Few other studies have used firm-level econometric techniques to evaluate the direct impact of RSA (Wren, 1994).

Devereux, Griffith and Simpson (2007) look at the role of RSA in affecting location decision of greenfield investments by foreign-owned multinationals and UK-owned multi-plant groups across different counties using information from matching the Annual Respondents Database (which we describe below) with publicly available information published by DBERR on the subset of the largest RSA grants' offers of above £75,000 over the period 1986-1992.[42] Using econometric methods[43] to solve the problem of endogeneity of firm and industry characteristics in the location choice equation their analysis suggests that grants are a very poor predictors of firms' location choices relative to agglomeration effects, such as locating near other foreign-owned plants in the same industry and natural advantages. In fact they find that an increase in the expected grant of £100,000 raises the probability of a Greenfield locating in an assisted area from 1% to 1.01%; this rises to 1.03% when taking into account that the location incentive of grant offers increases as the economic activity in the entrant's industry increases in the assisted area.

Harris and Robinson use the ARD matched the SAMIS database (which we describe in more detail below) over the period 1990-1998 to look at differences in survival rates for RSA recipients vs. non-recipients using a hazard model and sources of productivity growth in assisted vs. non-assisted areas using the productivity decomposition techniques taken by Haltiwanger (1997). Their results show that RSA recipients have significantly higher survival rates and that while in terms of labor productivity growth RSA recipients make a significant contribution to aggregate growth; in terms of TFP growth RSA treated plants experienced negative growth, mainly because plants with low initial TFP increased their market share. The authors therefore conclude that "plants in receipt of RSA generally experience market share growth despite having relatively lower productivity" (*ibidem* p.763).

Harris and Robinson look at how RSA is related to aggregate productivity growth using a decomposition technique

Finally, Jones and Wren (2003) look at differences in survival between "treated plants", i.e. recipients of RSA grants and non-treated firms and find that treated firms have shorter survival durations.

Therefore the econometric evidence seems to suggest some positive effect of RSA on employment but very small effect on location and on productivity (growth) and survival.

A.IV The role of changes in the criteria in determining eligibility

[to be completed]

[42] Not that their matching is coarser than our as they can only match the two datasets using postcode and industry affiliation of the plant, while we also use information on the name of the applicant.

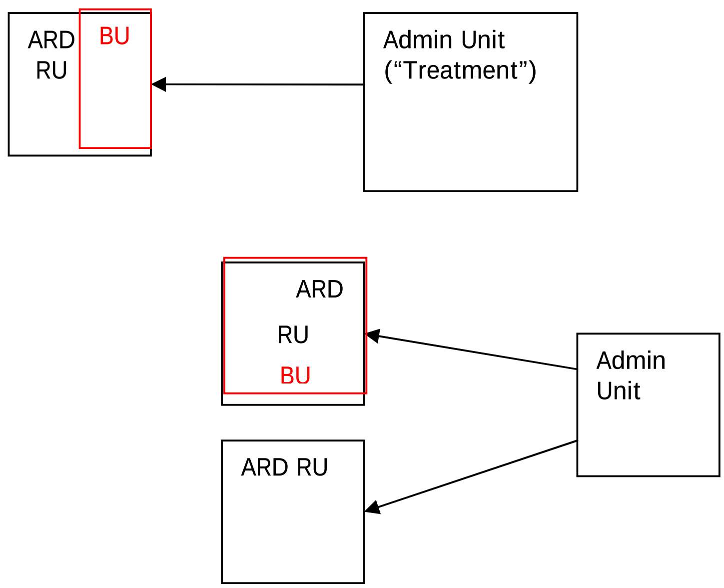
[43] Thy estimate a conditional logit equation of location choices –where any variable that do not vary across region drop out - on the predicted value of grant. The latter is estimated in a first stage where non-plant specific explanatory variables are included and corrected for self-selection using a Heckman selection equation method. The selection equation is estimated using data on a set of potential applicants and including firm-specific variables that affect the probability of application but not the amount of grant offered.

Table A1: the RSA database and the success rate of IDBR matching

Year	Applications	Linkage %
1972	357	34.73
1973	1,458	39.64
1974	1,119	42.45
1975	872	41.40
1976	1,024	47.07
1977	1,214	46.05
1978	1,199	51.79
1979	1,418	52.75
1980	584	47.60
1981	743	43.20
1982	1,305	47.51
1983	1,144	46.50
1984	1,363	54.22
1985	1,437	63.12
1986	2,034	62.73
1987	3,357	61.45
1988	3,119	64.67
1989	2,553	61.22
1990	2,782	63.66
1991	2,589	70.14
1992	2,336	75.30
1993	2,732	80.78
1994	2,710	80.89
1995	2,900	81.24
1996	2,516	84.70
1997	2,219	87.07
1998	1,775	86.99
1999	1,913	86.83
2000	1,003	89.13
2001	874	90.50
2002	675	92.00
2003	705	89.79
2004	222	91.89
2005	71	97.18
Total	54,322	67.83

Notes: This Table shows the absolute number of RSA applications and the proportion of those applications we were able to match into other datasets. In the regressions we only use data from 1988 to 2003.

Figure A1: Matching unique and multiple matches



Notes: This shows heuristically the matching of the different datasets. See text for details