Longitudinal Micro Data Study of Regional Selective Assistance in Wales

Final Report Stage 2

Chiara Criscuolo, Ralf Martin

Centre for Economic Performance, London School of Economics

Acknowledgements

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.

Executive Summary

This report describes the results of our analysis of the Regional Selective Assistance (RSA) Scheme in Wales. Our analysis is based on a combination of Department Business, Enterprise and Regulatory Reform (BERR, formerly DTI) data on participation in the RSA scheme with independent performance data for both participants and non-participants from the Annual Respondents Database (ARD), a dataset maintained by the Office of National Statistics (ONS).

We have looked at the impact of RSA on three aspects of firm performance:

- Employment
- Investment
- Productivity

Our focus is on whether (other things equal) firms that receive RSA grants performed better than firms that did not receive RSA grants. This is consequently not an evaluation of the overall welfare effects of the RSA scheme – for that we would have to assess among other things the tax distortions implied by raising the money used in the scheme – or the overall employment impact of RSA, which would require looking at possible displacement effects as well. However, to establishing that there are impacts at the level of participating firms is the first step and a necessary condition for any wider impacts.

The basic findings are as follows:

- RSA participation leads to higher employment and investment but not productivity in participating firms.
- Looking at differences in the effectiveness of RSA across industries we find that RSA has an impact in low tech sectors, and a less robust effect in the medium tech sectors but not in high tech sectors. However, the estimates are not statistically different across the different sectors considered.
- Looking at differences in the effectiveness of RSA across firms of different size we find larger effects for smaller firms.
- Quantitatively, when we find consistent evidence that RSA affects firm outcomes the effects are economically important.
- The size of impacts is very similar to values found in an earlier study for the UK as a whole.

The key challenge in an analysis of this kind is to account for selection effects. Since participation in the scheme is voluntary the sample of firms participating might be a subsample with specific characteristics which we cannot observe in our data. Consequently, observed performance differences between participants and non-participants could be driven by these unobserved characteristics rather than being a causal effect of RSA participation. We employ two econometric techniques to deal with this problem. Firstly, we use the panel structure of our data to make our results robust to time unvarying unobserved firm characteristics. Secondly, we try to exploit a "quasi-natural experiment": only firms located in certain disadvantaged areas are eligible for participation in RSA. Which area qualifies as disadvantaged depends on rules dictated by the EU which changed twice during our sample period. Given that unobserved firm level characteristics driving participation are unlikely to influence changes in EU rules we can use these eligibility changes as an instrumental variable for participation to compute RSA impact effects that are robust also to time varying unobserved characteristics. While this strategy worked very well when working with the whole UK sample, the eligibility changes become too weak as instrument for RSA (i.e. did not have enough explanatory power to explain variation in RSA participation) when we restrict the analysis to Wales. The reason could simply be that the sample is smaller and there are fewer changes than for the UK as a whole. Reassuringly, however, the instrumental variable estimates for the UK sample were larger in magnitude than the non-IV estimates (obtained using fixed effects). Since the non-IV estimates for the UK as a whole and for Wales alone are very similar we are confident that if we had sufficiently strong instruments we would find stronger results also in the analysis for Wales.

Table of Contents

List of Tables

List of Figures

1. Introduction

This report contains the results from the econometric evaluation of Regional Selective Assistance (RSA) in Wales. The RSA involves large amounts of direct financial support for individual firms with the aim of safeguarding or creating jobs: over the 1992-2002 decade, RSA cumulative offers amounted to £3.2 Billions for the whole of Great Britain and to £ 785.7 millions in Wales (see Secretary of State for Trade and Industry et al., 2002). In Wales RSA is administered by the Welsh Assembly (WA); before the creation of WA the National Assembly for Wales and previous to that the Welsh Office were running the scheme. More details on the scheme can be found in the Stage 1 report of this project.

The basis for the analysis in this report is a combination of data on participation in the RSA (the SAMIS database), provided by BERR and company performance data from the Annual Respondents Database (ARD) of the Office of National Statistics (ONS).

The remainder of this document is organised as follows: Section 2 discusses methodology. Section 3 describes the data used, focusing on the important variables derived from the ARD. The Stage 1 report included a detailed discussion of the DTI administrative dataset on business support (SAMIS) and how this was matched to the ARD. This Stage 2 report only provides a brief summary and the reader is referred to the Stage 1 report for details. Section 4 contains the results of the analysis. Section 5 concludes.

2. Methodology

2.1. Terminology

The objective of this study is to determine the causal impact of Regional Selective Assistance on various aspects of firm performance. Following the econometric evaluation literature we refer to receipt of an RSA grant as “treatment” or support programme participation, the set of businesses receiving RSA grants as “the treated” or “participants” and certain subgroups of non-treated businesses as “the control group”. By comparing the performance of the treated firms with those in the control group we derive a measure of the impact of RSA. The precise composition of the control group and the type of comparison varies across a number of different methods we employ.

2.2. Methodological approach

Our method for evaluating RSA is based on data derived by combining BERR administrative data on recipients of RSA (the SAMIS database) with independent performance data for these businesses from other more representative data sources. For the methodology to be useful we need the matched data to satisfy several criteria

- We must be able to match a sufficiently large number of firms that are recorded as receiving treatment in the SAMIS data, to the performance data set.
- For matched firms, performance data needs to be available both before and after treatment.

As discussed in the Stage 1 report, there are two main advantages of such an evaluation study. First, the availability of independent performance data means we can compare the performance of the firm before and after its exposure to the scheme. In contrast, administrative data on scheme participants is frequently only available after they have joined the scheme. Secondly and more importantly, we can compare the change in the participating firms’ performance to a “control group” of firms who did not participate in the program. Comparing the change in the treatment group’s outcomes to the change in the control group’s outcomes is the standard “difference in difference” approach common in the evaluation literature. The first “difference” (before and after comparison for the same firm) allows us to control for unobserved characteristics of participants. For example participants in the scheme may have systematically poorer performance before treatment. The second “difference” (comparison to control group) allows us to control for other factors, such as business cycles, that may affect all firms regardless of their participation in the scheme.

In contrast, more conventional evaluation exercises such as performance questions on a scheme application form or a survey among scheme participants after completion are unlikely to provide sufficient data to control for these two problems. In addition independent performance data is less likely to be affected by strategic reporting by firms.

2.3. Matching the data sets

The Stage 1 report explains the process of matching the SAMIS and ARD datasets in some detail. The fundamental unit of observation in the Annual Respondents Database (ARD) is the reporting unit. We want to consider the impact of support scheme participation on reporting unit performance.

To match an incidence of participation to an ARD reporting unit, SAMIS is matched to the Inter-Departmental Business Register (IDBR) and then matched from the IDBR to the ARD. Not all cases of scheme participation could be matched to reporting units in the ARD. Even when they could be matched, the ARD does not collect performance data on all reporting units in all years. Of particular concern is the fact that the ARD stratified sampling leads to small reporting units being under-represented in the performance data. In addition, as discussed in detail in the Stage 1 report, there is some ambiguity regarding which ARD units have been affected by the business support recorded in SAMIS. In some cases, several ARD reporting units could potentially be affected by a particular treatment. Further, because reporting units may be comprised of several local units, it might also be the case that only a part of a particular reporting unit is affected. This problem can arise even with unique matches. In practice, when multiple reporting units are matched to a particular piece of business support we assume that all reporting units have been treated.

To address these concerns we report robustness checks restricting attention to the smallest firms (less than 100 employees) only. To address the issue of multiple matching we report robustness checks based on a sample of firms having only one local unit. As shown later in the report, these two robustness checks do not weaken our key results.

2.4. Definition and timing of treatment

Aside from matching problems, there are also questions regarding the timing of any effects. Firms may receive multiple treatments or schemes may be designed to last over extended time periods. In these cases, when should we expect any impact to take place?

We address these issues by making a set of simple assumptions:

- For reporting units with multiple treatments we only consider the first treatment assuming that this has the potential to make the largest impact.
- We assume that effects occur from the first ARD reporting period that occurs after the start date of the treatment.

2.5. Basic micro level evaluation analysis (Impact on the firm)

The objective of econometric evaluation techniques is to estimate the causal impact of a particular scheme on an outcome variable such as employment; i.e. we try to answer the question: what would employment have been if the scheme had not taken place? In this section we explain the methodology used to examine whether there is an impact at the individual reporting unit level (micro impact). Note that we are assuming that the reader has read the Stage 1 report of the evaluation study and in particular that the reader is familiar with Appendix A of the report. For convenience, appendix A is repeated as an appendix to this Stage 2 report.

2.5.1. Regression Analysis

Our basic approach is based on linear regression analysis. For a given outcome variable y_{it} for reporting unit *i* at time *t*, we begin by running regressions of the form:

$$y_{it} = \beta_0 + \alpha RSA_{it} + \sum_k X_{kit}\beta_k + \varepsilon_{it} \qquad (1)$$

where the *k* explanatory variables are X_{it}, a vector of control variables (e.g. a polynomial in age; a dummy for whether the firm is part of a domestic group; a dummy for whether the firm is foreign owned; a set of industry dummies and a set of years dummies), and a dummy variable RSA_{it} which is set equal to 1 in all years in and after the firm receives treatments which we use to estimate the long run causal effect of treatment. In such a specification the impact of the RSA is measured by the coefficient estimate for α. The specification is run on all reporting units whether or not they receive treatment. The implied control group is made up of all businesses that are similar along the dimensions defined by the variables in X_{it} but that do not receive treatment.

In such a specification the coefficient estimates for α can be used to trace the impact of RSA on the outcome variable y_{it} while netting out the impact of control variables X_{it}.

While this approach has the attraction of simplicity, such a specification can produce biased estimates of scheme impact because there are other unobserved characteristics of firms that affect both outcomes and scheme participation.

To deal with this problem we start by estimating a fixed effects specification. Such a specification will control for all time invariant unobserved firm characteristics. Formally this implies that in equation 1 we assume that the error term can be written as:

$$\varepsilon_{it} = \eta_i + v_{it} \qquad (2)$$

where η_i is the fixed effect that captures all time invariant unobserved characteristics. This modification will adequately control for the fact that treated businesses might have systematically poorer performance than other businesses. It will not control for unobserved characteristics that vary over time nor will it provide correct estimates if programme participants businesses have different outcome trends.

This still leaves us with the problem that unobserved time varying firm characteristics may affect both firm performance and the decision to participate in the scheme. To deal with the role of unobserved time varying characteristics we can use information on exogenous variation in scheme rules that affects the probability for firms of being treated, but not the outcome variables. In this report, we exploit such exogenous variation in the context of RSA. We discuss this in turn.

2.5.2. Instrumental Variables approach

2.5.2.1. Using regional information to identify the effect of RSA on firm performance

The reason why we need to look for instrumental variables is that although fixed effects estimates are purged of bias due to common macro-economic shocks (through time dummies) and permanent correlated unobserved heterogeneity (through η_i), it will still be inconsistent if there are unobserved transitory shocks v_{it} correlated with RSA_{it} .Consequently we consider instrumental variables, Z_{it}, for program participation, RSA_{it} . In this subsection we describe how we use the fact that RSA eligibility depends on location to help address this problem of unobserved heterogeneity.

As discussed in the Stage 1 report, a firm's location in different areas will determine both whether the firm is entitled to apply to RSA and, if successful, the percentage of the investment the firm can claim from RSA. Thus, these criteria partially determine the likelihood and amount of treatment any given firm receives, but are determined independently of the performance of individual firms. As a result, they satisfy the requirements for instrumental variables that can help us deal with unobserved heterogeneity; i.e. they are exogenous and - conditional on treatment - are not correlated with individual firm's outcomes. Particularly useful is the fact that the criteria change over time thus allowing us to deal with unobserved time varying heterogeneity using these instruments while using fixed effects to deal with time invariant heterogeneity.

BERR provided us with a dataset with information at the postcode level on whether firms located at that particular postcode are eligible and if so what is the percentage of investment for which they can claim (we call this "RSA rates").

As discussed in more detail below, the map of eligibility and RSA rates has changed overtime 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. The changes were driven by a new European-wide formula for calculating which regions were eligible to receive subsidies under state aid rules, so that we can identify three sub-periods: the period per-1993, 1994-1999 and after 2000. Thus, the data contains variation in both the cross-section and the time-series dimension.

Since information on the ARD is recorded at the reporting unit level rather at the local unit level, we were 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). We take this into account when constructing our instruments.

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, in detail, the reduced form:

$$y_{it} = \pi_1 Z_{it} + \pi_2 X_{it} + \tilde{\eta}_i + \tilde{v}_{it} \qquad (3)$$

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 (ABI) is collected at the firm (reporting unit) level rather than at the plant (local unit) 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), for the other firms measures of investment, output and materials are only available at the "reporting unit" level which combines several plants.[1] 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. Equation (2), then becomes:

$$y_{jt} = \alpha RSA_{jt} + \beta X_{jt} + \eta_j + v_{jt} \qquad (4)$$

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.

For the participation dummy we mainly continue to use a simple binary indicator if the firm received any treatment. But we also present checks on alternatives such as the amount of money received expressed as the fraction of total project 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 weighting factors, w_{it}^{j}, to minimise the risk that 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. Even the location of plants within the firm could be affected by RSA eligibility (although this is less likely to be a problem due to sunk costs of plant entry and exit). 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 unit 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.

Note that the interpretation of the RSA coefficient 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 might be 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.[2]

---

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

[2] Comparison between the two estimates would be informative as regards such intra-firm switching behaviour. 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

Below we discuss in more detail the changes in eligibility and the rationale for using them in our identification strategy and the details of the instrumental variables approach.

2.5.2.2. Changes in eligibility over time

Since 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"). There is an upper threshold of support that is allowed, referred to as Net Grant Equivalent (NGE), which essentially sets a maximum proportion of the firm's investment that can be subsidised by the member state government.

In fact 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 criterion is that only areas with underemployment and a low standard of living are eligible.
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 then gathers 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.
Since the main formulae which determine eligibility are decided at the European level at fixed seven yearly 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.

(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. The selection of Assisted Areas was based on several factors using a quantitative formula. The first set of factors used indicators of bad labour 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.

---

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).

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) 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. 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. Each grouping must have a population of at least 100,000 and the wards were selected according to four statistical indicators. Although the precise indicators differed from 1993, the main criteria were still labour market performance and the share of manufacturing.
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 neighbouring 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. The maximum NGE level for relatively less deprived areas was 10%. 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).
In the dataset we have information at the postcode level on whether plants located there are eligible and if so the Net Grant Equivalent. As already discussed above, 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 dimension. 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 participating in 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.
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 (3), and ignoring time dummies for simplicity:

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

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.

However, 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.
Finally, 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. Concerning the functional form of the instrument, our baseline results for the analysis use mutually exclusive dummies for three different rates level (with zero being the baseline), less other equal 20% and more than 20%. The definitions of the instrument are therefore NGE <= 20% and NGE>20%. So that our instrument mainly identifies if the firm's oldest plant is in an area where the maximum investment subsidy (Net Grant Equivalent) is 20%; i.e. in a tier 2 or intermediate area or is in an area where the maximum investment subsidy (Net Grant Equivalent) is more than 20%; i.e. either 30 or 35%; i.e. a Tier 1 area or development area.

2.5.3. Robustness checks and extensions

2.5.3.1. 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.

2.5.3.2. Heterogenous 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:

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

where α_i is now the plant specific effect of treatment.[3] The essential problem is that using observations from the whole population may give a selection of non-treated plants that does not actually provide a valid comparison group for those who participate and prevent us from consistently estimating the average effect of treatment on the treated (ATT).

Secondly, we are interested in looking at differences in α_i across different subsample of the data; for example for firms in different industries or in different size classes. In the results section we present estimates of α_i across these different groups.

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

3. Data

This study uses data from the UK Office for National Statistics (ONS) Annual Respondents Database (ARD) in combination with administrative records from the SAMIS database. The merging process was described in more detail in the Stage 1 report of this project. Details on the ARD can be found in publications such as Criscuolo et al (2003). Here we focus on the construction of derived variables.

3.1. Labour productivity

Labour productivity for *i* at time *t* is constructed as

$$\text{Labour Productivity}_{\text{it}} = \frac{VA_{it}}{L_{it}}$$

where *VA* is value added and *L* is employment. To make value added comparable across time we deflating nominal values using 2 digit sectoral producer price indices.

3.2. Regression based TFP

Labour productivity might be a misleading measure of firm performance if some firms substitute other inputs for labour inputs. For example a firm's labour productivity might simply look higher because they buy in more of their intermediate inputs, rather than using labour to produce them in house. Economists' preferred productivity measure is therefore one that compares a firm's output with an index of all its production factors. Such a measure is referred to as Total Factor Productivity (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 on capital stock per employee, material inputs per employee and employment:

$$\ln\frac{GO_{it}}{L_{it}} = \beta_K \ln\frac{K_{it}}{L_{it}} + \beta_M \ln\frac{M_{it}}{L_{it}} + \beta_L \ln L_{it} + TFP_{it} \qquad (3)$$

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:

$$T\hat{F}P = \ln\frac{GO_{it}}{L_{it}} - \hat{\beta}_k \ln\frac{K_{it}}{L_{it}} - \hat{\beta}_M \ln\frac{M_{it}}{L_{it}} - \hat{\beta}_L \ln \mathrm{L}_{it}$$

where the hat over TFP indicates that it has been estimated from equation (8). 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.

$$TFP_{it} = \alpha RSA_{it} + \varepsilon_{it}$$

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

$$\ln\frac{GO_{it}}{L_{it}} = \beta_k \ln\frac{K_{it}}{L_{it}} + \beta_M \ln\frac{M_{it}}{L_{it}} + \beta_L \ln \mathrm{L}_{it} + \alpha RSA_{it} + \varepsilon_{it} \qquad (4)$$

These two procedures are equivalent if RSA_{it} is not correlated with any of the production factor variables. It is very unlikely that this condition is met, however, because treatment may both shift TFP and lead to adjustments of the factor 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.

3.3. Relative factor share TFP

We also use a factor share based TFP measure that is calculated relative to the 3-digit sectoral median in every year. Such a measure addresses a number of concerns:

- 2-digit price indices (in services the aggregation is even higher) might be inadequate.
- Endogeneity problems.
- Increased flexibility of the underlying production function.

Formally the measure we use is defined as follows:

$$\begin{aligned} TFP_{it}^{REL} = & \left(\ln GO_{it} - \ln GO_{Median,t}\right) \\ & - \bar{s}_{it}^{L}\left(\ln L_{it} - \ln L_{Median,t}\right) - \bar{s}_{it}^{M}\left(\ln M_{it} - \ln M_{Median,t}\right) \\ & - \left(1 - \bar{s}_{it}^{L} - \bar{s}_{it}^{M}\right)\left(\ln K_{it} - \ln K_{Median,t}\right) \end{aligned}$$

where

$$\bar{s}_{it}^{X} = \frac{s_{it}^{X} + s_{Median,t}^{X}}{2} \text{ for } X \in \{L, M\}$$

and s_{it}^{X} is the revenue share of a factor. For more details see Bailey et al. (1992)

3.4. Data cleaning

The sample we use in our different analytical approaches was subject to some basic cleaning. This included the dropping of observations with the following features:

- Negative value added.
- Gross output smaller than wage and material costs combined.
- Zero values for material inputs or employment.

3.5. Characteristics of recipients, RSA

Table 1 reports descriptive statistics distinguishing between firms which do not participate in the RSA scheme, the "non-treated" sample, and RSA recipients, the "treated" sample. For this latter group it reports the characteristics pre treatment.

Rows 1 and 2 report statistics on employment for non treated and treated firms. Treated firms are always larger than non treated firms on average (column 1) and along the whole distribution (columns 4 to 6). This is equally true for log employment – where logs are used to smooth away the presence of outliers – or if we look at employment relative to the geographical area where the firm is located (row 3). When looking at employment in log terms the difference between treated and non-treated firms becomes statistically significant as indicated by the stars in column 2. Row 4 shows also that, pre-treatment, treated firms are also growing more than non treated firms. Rows 5 and 7 and 8 show that both in terms of labour productivity, measured as value added per employee (VA/L), TFP and TFP relative to the local area average, treated firms are significantly less productive to start with. This is true along the whole distribution as shown in columns 6 to 8 reporting the 25th, 50th and 75th percentile of the distribution. Row 6 shows that in terms of (labour) productivity growth there is no difference between non-treated and treated firms.

The Table also reports figures for gross output in row 12. Although on average treated firms are smaller this results is driven by few outliers at the top of the distribution. In fact columns 4 to 6 show that at the median and at 1st and 3rd quartile of the distribution treated firms are always larger.
Looking at investment (in row 13) shows that even though at the mean non-participants appear to invest no less than participating firms, in the pre-program periods this result is driven by a few outliers at the top of the non-treated distribution. Looking at the medians shows that on average participating firms had larger investments. Row 10 looks at capital intensity and treated firms seem to be less capital intensive than non treated firms.

Table 1: Descriptive results, manufacturing, RSA

		1	2	3	4	5	6	7
Variable	**sample**	**mean**	**sig**	**sd**	**p25**	**p50**	**p75**	**Obs.**
1 Employment	non-treated	411.04		1667.9	47.62	122.00	331.30	6559
	treated	431.00		790.7	82.00	186.36	429.66	1076
2 ln(Employment)	non-treated	4.89		1.34	3.85	4.78	5.78	6559
	treated	5.26	***	1.20	4.38	5.19	6.03	1076
3 ln(EMPLOYMENT) Relative to TTWA	non-treated	-0.09		1.08	-0.87	-0.16	0.57	5985
	treated	0.08	**	0.97	-0.61	0.07	0.76	981
4 EMP growth	non-treated	0.01		0.19	-0.07	0.00	0.07	4373
	treated	0.04	***	0.20	-0.04	0.02	0.10	766
5 VA/EMP	non-treated	35.43		47.24	18.76	26.29	38.92	6254
	treated	27.45	***	16.27	17.07	23.78	32.24	945
6 VA/EMP growth	non-treated	0.03		0.33	-0.13	0.03	0.19	4119
	treated	0.03		0.30	-0.13	0.02	0.17	657
7 TFP	non-treated	0.01		0.33	-0.16	0.01	0.17	6038
	treated	-0.03	***	0.30	-0.18	-0.03	0.14	1029
8 TFP relative to TTWA	non-treated	0.02		0.31	-0.15	0.01	0.17	5473
	treated	-0.01	**	0.29	-0.18	-0.02	0.15	937
9 Materials/EMP	non-treated	78.45		158.78	23.66	42.14	78.00	6254
	treated	61.15	*	99.19	22.70	36.44	68.77	945
10 K/EMP	non-treated	114.05		244.61	26.09	53.35	110.65	2910
	treated	68.76	***	74.45	29.50	49.50	89.90	438
11 VA/GO	non-treated	0.40		0.16	0.29	0.39	0.51	6559
	treated	0.40		0.15	0.30	0.40	0.49	1076
12 Gross Output	non-treated	68089		339291	2833	9164	31207	6254
	treated	43238	*	93850	4025	12711	39017	945
13 Investment	non-treated	2890.89		18451	39.06	195.90	992.67	6559
	treated	1807.10		5139	102.33	341.72	1279.77	1076
14 Age	non-treated	13.71		8.06	7.00	14.00	19.00	6559
	treated	13.18		7.22	7.00	14.00	18.00	1076

Notes: Column 1, mean, reports the mean of the variables of interest separately for non treated firms and for the group of treated for the pre-treatment period. Column 2, sig, reports the significance of a t-test of equality between the values for treated firms relative to the group of non treated firms. Column 3, sd, reports standard deviations, while columns 4 to 6 describe the distribution (25th; median and 75th percentile respectively), of the variables for non treated and treated pre-treatment. Finally, column 7 reports the number of observations for each cell.

4. Results

4.1. Linear regression results, RSA

4.1.1. The results for the whole sample

We start by reporting the results from estimating equation (4) using simple OLS regressions for each of the following outcome variables: employment; investment; labor productivity and total factor productivity. Our key 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. In the vector of explanatory variables we include 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 entered before 1980 to control for left censoring of the age variable, a full set of four-digit industry dummies and time dummies. Note that in all the regressions involving our instruments we allow for clustering at the area level to take into account that the instruments' variation is at this level.

We first turn to analyzing employment in Table 2. The first four columns do not include fixed effects and 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 significant with a coefficient that indicates that RSA participation is associated with about a 16.4%[4] increase in employment. Column (2) reports the reduced form where we regress firm employment on our policy instruments: a set of dummy variables capturing if a firm is located in an area eligible for RSA. The two dummies account for different levels of support (<=20%, >20%). The omitted reference category is composed of firms that do not have any plants in areas eligible for investment subsidies. The policy dummies are positive and jointly and individually significant. 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. A caveat with the regressions discussed so far is that the results could be driven by un-controlled for heterogeneity between programme participants and other firms. For example, if larger firms tend to be located in areas that are eligible to RSA then the results in columns 3 to 5 could emerge even if there was no causal impact of the RSA itself. With panel data we can account for this by introducing fixed effects for each firm in the sample (See section 2.5.1) which implies that we identify treatment effects exclusively from changes within firms. Column (5) reports the first such specification where the coefficient on the RSA dummy is very similar in magnitude and significance to that of column (1). Columns (6) and (7) report the reduced form and first stage respectively. The policy instruments lose almost all of their explanatory power, except for the highest categories of RSA subsidies in column (6). This implies that there is simply not enough variation in eligibility over time so

[4] This was calculated as: exp(0.152)-1.

that our instruments are too weak for this kind of analysis. The weakness of our instruments is also shown in the last row of the table where we report the F-statistics for the excluded instruments which is just above 1.[5] Therefore it comes as no surprise that in column (8) the IV results are larger than the fixed effects estimates in column (5) and are very imprecisely estimated and therefore not statistically significantly different from zero.
What does this imply for our conclusions? As discussed in section 2.5.2, a fixed effects specification as reported in column 5 adequately controls for non time varying heterogeneity between firms. The objective of our strategy involving instruments was to additionally take account of un-observed time varying factors that determine both programme participation and firm level outcomes. Given the weakness of the instruments we cannot rule out that our results are partly driven by such time varying factors. A re-assuring piece of evidence is however that in our earlier analysis of the RSA for the UK as a whole where our instrumental variable approach could be successfully implemented (Criscuolo et al 2007) the effects typically became stronger relative to the simple fixed effects specification from column 5. Table 13 reproduces a result table from Criscuolo et al (2007) which demonstrates this point. The layout of the table is equal to tables 2 to 8. Looking at columns 6 and 7 we see that both the reduced form and the first stage fixed effects regression confirm that the instruments have strong explanatory power.[6] Comparing then columns 5 and 8 we see that in the instrumented regression in column 8 the coefficient value increases markedly by almost a factor 5.

Table 13 also suggests that the impact of RSA in Wales is very similar to the impact of RSA in the UK as a whole: compare the point estimate of 0.167 in column 5 of table 2 with the 0.168 found in table 13.

The discrete dummy for program participation does not take into account the varying levels of support to different firms. We simply estimate an average effect on the treated. Obviously this discards some useful information on the intensity of the treatment. In Table 9 we report the estimates obtained when we use the actual RSA grant intensity awarded to the firms calculated using the ratio between the grant amount and the total costs of the investment project. Therefore, Table 3 repeats the same order of specifications as Table 2 but uses as treatment variable the proportion of projects costs covered by RSA rather than a simple indicator variable for whether the firm has received treatment or not. Relative to Table 2 we note the following:

- Firstly the coefficients reported in the Table are semi-elasticities; therefore to get back to elasticities we need to multiply the coefficients by the average treatment intensity – in our sample 14%. Therefore the estimated semi-elasticity of 1.558 in column 1 corresponds to an elasticity of 0.218 for the average treated firm; i.e. programme participation increases employment by 21.8% in the average participating firm.
- Secondly, now the difference between the OLS estimates in Column 1 and the fixed effects estimates in Column 5 are much starker: in fact the fixed effects estimates are a third of the ones presented in column 1. This can potentially be explained by larger firms receiving relatively higher support levels.

[5] In general the rule of thumb used to exclude the problem of weak instruments in an IV regression is that the F-statistics of excluded instruments must be at least above 10.
[6] In those earlier results we are representing support levels using 5 different categories whereas in the new results in this paper we use only 2 categories. We checked however that this is not the cause for the poor performance of the instruments in this study. In fact the reduction in the number of categories was an unsuccessful effort to improve the performance of the instruments.

- Thirdly, as shown by the single coefficients on the eligibility instruments and by the higher F-statistics of exclusion restrictions from the first stage equation in column (7) the instruments are less weak (although still well below the threshold value of 10) and this leads to a more significant IV estimates in column 8.

Table 4 repeats the same order of specifications as Table 2 but uses investment as dependent variable. The broad pattern of results is similar to that for employment. First, the magnitude of all RSA effects remains unchanged when we control for fixed effects (the difference between the OLS RSA coefficient in column (1) and the Fixed Effects coefficient in column (5) is negligible). Second, the policy instruments are not informative. In fact they are insignificant in both the reduced form and first stage when we control for firm fixed effects. Third, and most importantly, the IV results are not significantly different from zero.

Similarly, Table 5 repeats the same order of specifications as Table 3 using the continuous treatment indicator in conjunction with investment as the dependent variable. The patterns are very similar to those showed in Table 3 but now the IV coefficient reported in Table 8 is like in Table 4 too imprecisely estimated and not significantly different from zero. Again this is likely due to the lack of power of our instruments as shown by the F-statistics of exclusion restrictions reported at the bottom of column 7.

Tables 6, 7, 8 and 9 report estimates of the impact of RSA on productivity. In all tables the dependent variable is labour productivity measured as real gross output over employment. Tables 8 and 9 include as explanatory variable material inputs over employment and capital over employment. The results capture consequently the effect of RSA on total factor productivity (TFP). The coefficient on the RSA variable is always negative – although insignificant - in the OLS regressions of column (1). When we control for fixed effects, however, the coefficient becomes positive - and is significant in Tables 7 and 9. The instrumental variables estimates suffer from the same problems of weak instruments highlighted above.

In summary, our basic results suggest that RSA has a positive effect on employment and investment of participating firms; but not a robust impact on productivity. We need to rely on fixed effects estimates as our preferred specification since the instruments available are too weak to help us identify the causal impact of RSA on the outcome variables considered.

4.1.2. Results for different industries

Table 10 reports separate estimates of the treatment effect for employment by splitting the sample according to the technological intensity of different sectors. We use an OECD classification for that purpose which is detailed in the Appendix. We differentiate between four levels: low tech, medium-low, medium-high and high-tech. Looking at the point estimates in column 1 would suggest that differences between sectors are rather high and that effects are generally stronger in low tech sectors. However, a caveat is that the number of treated firms varies significantly between different sectoral classes (see Table 9 in the Stage 1 report of this project). In particular, the number of treated firms in the high tech sector is very small.

This implies that the point estimates for these sectors are considerably more imprecise. This leads to two conclusions. Firstly, table 10 makes explicit that RSA works in those sectors where the bulk of the RSA activity is taking place. Secondly, rather than looking at point estimates it might be more relevant to look at confidence intervals. This is being done in columns 2 and 3. Notice that the upper boundary of the confidence interval for the high tech sector is 0.281 which is well above the point estimate for the low tech sector. Therefore, we do not have strong enough evidence to conclude that there are significant differences between sectors.

4.1.3. Results across different size classes

Table 11 reports results from our fixed effects specification for single establishment firms only and separately for different size classes of companies. Consider first column 1 showing results for single establishment firms. Using either a simple treatment dummy (upper panel) or our measure of support intensity (NGE) we find a positive and significant effect on employment of RSA treatment. Quantitatively the results are somewhat stronger than for the sample of all firms (compare columns 5 of tables 2 and 3). This is what we would expect as looking only looking at single establishment firms avoids measurement error from ambiguous matches between administrative records and ARD data.
Columns 2 to 5 split the sample into small and large firms. In columns 2 and 3 the split point is a firm size of 100 in columns 4 and 5 the split point is 250 employees. To avoid selection into groups to be endogenous to treatment we use employment size in the first year a firm is observed for classification into size bands. The results in the upper panel where we look at a simple treatment dummy suggest that the impact of treatment is larger for smaller firms. The difference in effects is particularly pronounced when using the 100 employee split point. Using the continuous treatment indicator (NGE) in the lower panel we still find a stronger impact for smaller firms. However, with the 250 employee split point the ranking of the treatment effects reverses, although the difference is not statistically significant.
Finding a stronger effect for smaller firms is very much in line with prior expectations. The use of log employment as dependent variable implies that the estimated dummy variable coefficient can be interpreted as proportional effect (i.e. a treatment dummy equals to 1 is associated with a X% increase in employment). For smaller firms it is more likely that an RSA grants amounts to a larger share of their overall investment and therefore has a stronger proportional effect on their employment. The following can provide an explanation for the reversal of effects between discrete and continuous treatment indicator in columns 4 and 5: The result could emerge if larger firms would get higher rates of support in their larger investment projects.

Does the heterogeneity in performance imply that there is scope to improve the efficiency of the programme; i.e. could the government achieve a higher amount of job creation at equal or lower levels of subsidy given out? To assess this we need to look not only at the relative programme impact but at the amount of money spent per job. Table 12 does this for average levels across different size classes. For each size class column 1 reports the average employment, column 2 the average value of the grant given out and column 3 the coefficient values from Table 11. Using the coefficient value and average employment we work out the number of jobs created in an average firms as

$$\text{Jobs Created}=[\exp(\text{coefficient value})-1] \times \text{Average}(L)$$

Column 5 then reports the cost per job as average grant value over Jobs Created. This leads to the following findings: row 1 and 2 report results for the 100 employees split. They show that even though smaller firms have a much higher impact coefficient – almost twice as large as for firms with more than 100 employees– the difference in the cost per job values across the two groups are very small, just £1,407 (£14,532-£13,125). Indeed when using the 250 employees split (rows 3 and 4) the ranking reverses: larger firms have a lower cost per job. Thus, while larger firms have a smaller proportional impact of the programme, the cost per job for large firms is virtually the same to or even lower than for small firms. This is because the absolute number of jobs created is much higher than then absolute increase in grant value.

As always a note of caution has to be attached to such numbers. At best they can provide a lower bound of the costs as overheads and tax distortions from raising the funds are not considered. Nor are we taking into account that creation of a job in a specific establishment might lead to displacement elsewhere in the economy. The numbers are also estimated figures and therefore likely to be subject to high standard errors. Nevertheless it is encouraging to find figures of similar orders of magnitude across different categories, suggesting that the government is fairly efficient in allocating the grants.

Table 2: Linear regressions, employment, manufacturing, RSA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**Ln(EMP)**	**ln(EMP)**	**NGE RSA**	**ln(EMP)**	**ln(EMP)**	**ln(EMP)**	**NGE RSA**	**ln(EMP)**
	OLS	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
RSA	0.152***			2.356***	0.167***			1.566
	(0.050)			(0.859)	(0.035)			(0.995)
age	-0.006	-0.012	0.019***	-0.061***	0.037***	0.039***	0.022***	0.004
	(0.008)	(0.008)	(0.003)	(0.014)	(0.005)	(0.006)	(0.005)	(0.027)
Age squared	0.001***	0.001***	-0.000***	0.002***	-0.001***	-0.001***	-0.000	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age dummy	0.277***	0.257***	-0.083***	0.531***				
	(0.079)	(0.091)	(0.024)	(0.145)				
Foreign owned	1.174***	1.141***	-0.004	1.317***	0.038	0.039	0.012	0.021
	(0.061)	(0.083)	(0.022)	(0.098)	(0.029)	(0.038)	(0.027)	(0.052)
Domestic group	0.987***	0.959***	-0.072***	1.285***	0.034	0.033	-0.004	0.040
	(0.049)	(0.072)	(0.013)	(0.085)	(0.022)	(0.020)	(0.016)	(0.026)
NGE<=20%		0.268***	0.057***			0.017	0.003	
		(0.057)	(0.015)			(0.011)	(0.015)	
NGE>20%		0.192***	0.093***			0.033***	0.020	
		(0.058)	(0.019)			(0.013)	(0.013)	
Observations	9578	9578	9578	9578	9578	9578	9578	9405
Number of firms	1806	1806	1806	1806	1806	1806	1806	1806
Fixed effects	NO	NO	NO	NO	YES	YES	YES	YES
F-stats							1.15	

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 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 heteroscedacity 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 also include a full set of firm dummies. Time period is 1985-2004.

Table 3: Linear regressions, employment, manufacturing, money spent RSA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**Ln(EMP)**	**ln(EMP)**	**NGE RSA**	**ln(EMP)**	**Ln(EMP)**	**ln(EMP)**	**NGE RSA**	**ln(EMP)**
	OLS	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
NGE RSA	1.558***			13.185**	0.601***			7.221**
	(0.341)			(5.142)	(0.197)			(3.481)
foreign group	1.175***	1.141***	-0.002	1.364***	0.039	0.039	0.001	0.031
	(0.060)	(0.083)	(0.003)	(0.081)	(0.030)	(0.038)	(0.006)	(0.049)
Domestic group	0.981***	0.959***	-0.004*	1.169***	0.032	0.033	0.002	0.021
	(0.049)	(0.072)	(0.002)	(0.073)	(0.022)	(0.020)	(0.004)	(0.032)
NGE<=20%		0.268***	0.011***			0.017	0.003**	
		(0.057)	(0.002)			(0.011)	(0.001)	
NGE>20%		0.192***	0.018***			0.033***	0.004**	
		(0.058)	(0.003)			(0.013)	(0.002)	
Observations	9578	9578	9578	9578	9578	9578	9578	9405
Number of Firms	1806	1806	1806	1806	1806	1806	1806	1806
Fixed effects	NO	NO	NO	NO	YES	YES	YES	YES
F-stat							4.87	

Notes: NGE_RSA is the investment subsidy received by the firm (total payment divided by total investment cost). 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.

Table 4: Linear regressions, investment, manufacturing, DUMMY RSA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**Ln(Investment)**	**Ln(Investment)**	**RSA**	**Ln(Investment)**	**Ln(Investment)**	**Ln(Investment)**	**RSA**	**Ln(Investment)**
	OLS	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
RSA	0.463***			3.223**	0.473***			-2.144
	(0.081)			(1.391)	(0.074)			(1.784)
foreign group	1.574***	1.534***	-0.008	2.028***	-0.029	-0.019	0.016	0.015
	(0.094)	(0.124)	(0.023)	(0.150)	(0.065)	(0.064)	(0.029)	(0.104)
Domestic group	1.141***	1.085***	-0.079***	1.527***	0.047	0.046	-0.004	0.038
	(0.072)	(0.084)	(0.014)	(0.176)	(0.060)	(0.047)	(0.018)	(0.071)
NGE<=20%		0.313***	0.056***			-0.026	0.001	
		(0.089)	(0.015)			(0.039)	(0.015)	
NGE>20%		0.246***	0.091***			-0.045	0.019	
		(0.090)	(0.020)			(0.034)	(0.014)	
Observations	8723	8723	8723	8723	8723	8723	8723	8438
Number of firms	1634	1634	1634	1634	1634	1634	1634	1634
Fixed effects	NO	NO	NO	NO	YES	YES	YES	YES
							0.97	

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 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.

Table 5: Linear regressions, investment, manufacturing, money spent RSA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**Ln(Investment)**	**Ln(Investment)**	**NGE RSA**	**Ln(Investment)**	**Ln(Investment)**	**Ln(Investment)**	**NGE RSA**	**Ln(Investment)**
	OLS	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
NGE RSA	3.220***			17.381**	1.815***			-10.677
	(0.541)			(7.945)	(0.404)			(8.883)
foreign group	1.577***	1.534***	-0.003	2.085***	-0.024	-0.019	0.002	-0.002
	(0.094)	(0.124)	(0.003)	(0.141)	(0.065)	(0.064)	(0.006)	(0.091)
Domestic group	1.117***	1.085***	-0.004**	1.355***	0.041	0.046	0.002	0.068
	(0.071)	(0.084)	(0.002)	(0.133)	(0.060)	(0.047)	(0.004)	(0.068)
NGE<=20%		0.313***	0.011***			-0.026	0.003**	
		(0.089)	(0.002)			(0.039)	(0.001)	
NGE>20%		0.246***	0.018***			-0.045	0.004*	
		(0.090)	(0.003)			(0.034)	(0.002)	
Observations	8723	8723	8723	8723	8723	8723	8723	8438
Number of firms	1634	1634	1634	1634	1634	1634	1634	1634
Fixed effects	NO	NO	NO	NO	YES	YES	YES	YES
F-stats							4.05	

Notes: NGE_RSA is the investment subsidy received by the firm (total payment divided by total investment cost). 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.

Table 6: Linear regressions, labour productivity, manufacturing, dummy RSA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**Ln(GO/EMP)**	**Ln(GO/EMP)**	**RSA**	**Ln(GO/EMP)**	**Ln(GO/EMP)**	**Ln(GO/EMP)**	**RSA**	**Ln(GO/EMP)**
	OLS	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
RSA	-0.022**			0.059	0.006			-0.392
	(0.011)			(0.161)	(0.017)			(0.529)
age	-0.003	-0.003**	0.019***	-0.004	0.017***	0.017***	0.019***	0.024**
	(0.002)	(0.001)	(0.003)	(0.004)	(0.003)	(0.002)	(0.005)	(0.010)
age2	0.000*	0.000**	-0.000***	0.000	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
age dummy	-0.022	-0.020	-0.094***	0.038				
	(0.015)	(0.021)	(0.026)	(0.032)				
foreign group	0.062***	0.062***	-0.023	0.088***	0.034***	0.034**	-0.002	0.033**
	(0.013)	(0.011)	(0.021)	(0.016)	(0.013)	(0.014)	(0.027)	(0.016)
Domestic group	0.027***	0.029***	-0.089***	0.038**	0.031***	0.031**	-0.005	0.029**
	(0.010)	(0.010)	(0.014)	(0.019)	(0.010)	(0.014)	(0.015)	(0.014)
ln(Mat/EMP)	0.678***	0.678***	-0.011	0.725***	0.497***	0.498***	0.026	0.508***
	(0.011)	(0.011)	(0.011)	(0.016)	(0.021)	(0.019)	(0.017)	(0.025)
ln(EMP)	0.010**	0.009*	0.020*	-0.002	-0.142***	-0.142***	0.091***	-0.106*
	(0.004)	(0.005)	(0.010)	(0.007)	(0.015)	(0.019)	(0.023)	(0.061)
NGE<=20%		0.002	0.055***			0.001	0.001	
		(0.007)	(0.017)			(0.007)	(0.014)	
NGE>20%		0.003	0.093***			-0.005	0.015	
		(0.010)	(0.019)			(0.006)	(0.013)	
Observations	9123	9123	9123	9123	9123	9123	9123	8921
Number of firms	1775	1775	1775	1775	1775	1775	1775	1775
Fixed effects	NO	NO	NO	NO	YES	YES	YES	YES
F-stat							0.65	

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 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 heteroscedacity 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 also include a full set of firm dummies. Time period is 1985-2004.

Table 7: Linear regressions, labour productivity, manufacturing, money spent RSA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**Ln(GO/EMP)**	**Ln(GO/EMP)**	**NGE RSA**	**Ln(GO/EMP)**	**Ln(GO/EMP)**	**Ln(GO/EMP)**	**NGE RSA**	**Ln(GO/EMP)**
	OLS	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
NGE RSA	-0.024			0.332	0.200**			-0.486
	(0.076)			(0.934)	(0.088)			(1.763)
foreign group	0.063***	0.062***	-0.006**	0.090***	0.034***	0.034**	-0.000	0.034**
	(0.013)	(0.011)	(0.003)	(0.015)	(0.013)	(0.014)	(0.006)	(0.013)
Domestic group	0.029***	0.029***	-0.007***	0.035**	0.031***	0.031**	0.001	0.031**
	(0.010)	(0.010)	(0.002)	(0.014)	(0.010)	(0.014)	(0.003)	(0.013)
ln(Mat/EMP)	0.678***	0.678***	-0.002	0.724***	0.497***	0.498***	0.001	0.498***
	(0.011)	(0.011)	(0.001)	(0.015)	(0.020)	(0.019)	(0.002)	(0.019)
ln(EMP)	0.009**	0.009*	0.004***	-0.001	-0.144***	-0.142***	0.009***	-0.137***
	(0.004)	(0.005)	(0.001)	(0.007)	(0.015)	(0.019)	(0.003)	(0.025)
NGE<=20%		0.002	0.010***			0.001	0.003**	
		(0.007)	(0.002)			(0.007)	(0.001)	
NGE>20%		0.003	0.018***			-0.005	0.003	
		(0.010)	(0.003)			(0.006)	(0.002)	
Observations	9123	9123	9123	9123	9123	9123	9123	8921
Number of firms	1775	1775	1775	1775	1775	1775	1775	1775
Fixed effects	NO	NO	NO	NO	YES	YES	YES	YES
F-stat4.24								

Notes: NGE_RSA is the investment subsidy received by the firm (total payment divided by total investment cost). 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.

Table 8: Linear regressions,multifactor productivity, manufacturing, dummy RSA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**Ln(GO/EMP)**	**Ln(GO/EMP)**	**RSA**	**Ln(GO/EMP)**	**Ln(GO/EMP)**	**Ln(GO/EMP)**	**RSA**	**Ln(GO/EMP)**
	OLS	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
RSA	-0.032***			0.030	0.003			-0.232
	(0.010)			(0.139)	(0.017)			(0.368)
age	-0.002	-0.003**	0.021***	-0.003	0.016***	0.016***	0.020***	0.021***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.005)	(0.007)
age2	0.000*	0.000**	-0.000***	0.000	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
age dummy	-0.019	-0.017	-0.088***	0.034				
	(0.015)	(0.020)	(0.027)	(0.029)				
foreign group	0.063***	0.063***	-0.022	0.084***	0.034***	0.035**	-0.003	0.034**
	(0.013)	(0.011)	(0.021)	(0.014)	(0.013)	(0.014)	(0.028)	(0.014)
Domestic group	0.032***	0.034***	-0.085***	0.043***	0.031***	0.031**	-0.006	0.030**
	(0.010)	(0.010)	(0.013)	(0.016)	(0.010)	(0.014)	(0.016)	(0.014)
ln(Mat/EMP)	0.670***	0.671***	-0.029**	0.708***	0.502***	0.502***	0.020	0.507***
	(0.012)	(0.011)	(0.012)	(0.016)	(0.021)	(0.019)	(0.017)	(0.021)
ln(K/EMP)	0.027***	0.026***	0.033***	0.040***	0.008***	0.008***	0.013***	0.011**
	(0.003)	(0.003)	(0.004)	(0.005)	(0.002)	(0.002)	(0.003)	(0.005)
ln(EMP)	0.005	0.005	0.013	-0.005	-0.138***	-0.138***	0.090***	-0.117***
	(0.004)	(0.005)	(0.010)	(0.006)	(0.015)	(0.019)	(0.024)	(0.045)
NGE<=20%		0.000	0.055***			-0.001	0.002	
		(0.007)	(0.016)			(0.007)	(0.014)	
NGE>20%		0.002	0.091***			-0.004	0.018	
		(0.009)	(0.019)			(0.006)	(0.013)	
Observations	8815	8815	8815	8815	8815	8815	8815	8579
Number of firms	1699	1699	1699	1699	1699	1699	1699	1699
Fixed effects	NO	NO	NO	NO	YES	YES	YES	YES
F-stats							0.87	

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 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 heteroscedacity 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 also include a full set of firm dummies. Time period is 1985-2004.

Table 9: Linear regressions,multifactor productivity, manufacturing, money spent RSA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	**Ln(GO)**	**Ln(GO)**	**NGE RSA**	**Ln(GO)**	**Ln(GO)**	**Ln(GO)**	**NGE RSA**	**Ln(GO)**
	OLS	Reduced Form	First Stage	IV	FE	Reduced Form	First Stage	IV
NGE RSA	-0.063			0.160	0.195**			-0.664
	(0.074)			(0.788)	(0.088)			(1.588)
foreign group	0.063***	0.063***	-0.006**	0.084***	0.035***	0.035**	-0.001	0.034**
	(0.013)	(0.011)	(0.003)	(0.014)	(0.013)	(0.014)	(0.006)	(0.013)
Domestic group	0.034***	0.034***	-0.007***	0.042***	0.031***	0.031**	0.000	0.032**
	(0.009)	(0.010)	(0.002)	(0.013)	(0.010)	(0.014)	(0.003)	(0.014)
ln(Mat/EMP)	0.671***	0.671***	-0.004**	0.708***	0.502***	0.502***	0.000	0.502***
	(0.012)	(0.011)	(0.002)	(0.015)	(0.021)	(0.019)	(0.002)	(0.019)
ln(K/EMP)	0.027***	0.026***	0.004***	0.040***	0.008***	0.008***	0.001*	0.009***
	(0.003)	(0.003)	(0.001)	(0.004)	(0.002)	(0.002)	(0.001)	(0.003)
ln(EMP)	0.005	0.005	0.004**	-0.005	-0.140***	-0.138***	0.009***	-0.132***
	(0.004)	(0.005)	(0.001)	(0.006)	(0.015)	(0.019)	(0.003)	(0.023)
NGE<=20%		0.000	0.011***			-0.001	0.003***	
		(0.007)	(0.002)			(0.007)	(0.001)	
NGE>20%		0.002	0.018***			-0.004	0.003*	
		(0.009)	(0.003)			(0.006)	(0.002)	
Observations	8815	8815	8815	8815	8815	8815	8815	8579
Number of firms	1699	1699	1699	1699	1699	1699	1699	1699
Fixed effects	NO	NO	NO	NO	YES	YES	YES	YES
F-stat							4.7	

Notes: NGE_RSA is the investment subsidy received by the firm (total payment divided by total investment cost). 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.

Table 10: Linear regressions, employment different industries, manufacturing, RSA

	(1)	(2)	(3)	(4)
	RSA	lower	upper	Obs
LOW-TECH sectors	0.221***	0.109	0.333	3443
	(0.057)			
MEDIUM-LOW TECH sectors	0.111*	-0.009	0.231	2838
	(0.061)			
MEDIUM-HIGH TECH sectors	0.213***	0.093	0.333	2366
	(0.061)			
HIGH-TECH sectors	0.040	-0.201	0.281	931
	(0.123)			

Notes: Column 1 report estimates of RSA which equals unity for all the periods in and after a firm has participated in the program and zero otherwise. Standard errors below coefficients are robust to heteroscedacity and arbitrary serial correlation and are clustered by firms. Columns 2 and 3 report the lower and upper boundary of a 5% confidence interval around the point estimate. Different rows report results for industries of different technology level (see the Glossary for a detailed description of the classification). Time period is 1985-2004.

Table 11: Linear regressions employment different sizes , fixed effects estimates, manufacturing, RSA

	(1)	(2)	(3)	(4)	(5)
Dep. Variable	**Ln(EMP)**	**ln(EMP)**	**Ln(EMP)**	**ln(EMP)**	**ln(EMP)**
	single plant	less than 100	more than 100	less than 250	more than 250
RSA	0.193***	0.258***	0.134***	0.197***	0.128*
	(0.036)	(0.045)	(0.046)	(0.036)	(0.066)
age	0.043***	0.047***	0.019**	0.043***	-0.008
	(0.006)	(0.006)	(0.009)	(0.005)	(0.014)
age2	-0.001***	-0.001***	-0.001**	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
foreign group	0.007	0.027	0.033	0.054	0.015
	(0.035)	(0.047)	(0.037)	(0.034)	(0.054)
Domestic group	0.043*	0.036	0.011	0.019	0.012
	(0.024)	(0.031)	(0.029)	(0.024)	(0.049)
Obs	5718	4291	5287	6540	3038
payX	0.788***	1.135***	0.523**	0.572**	0.692**
	(0.249)	(0.283)	(0.221)	(0.239)	(0.312)
age	0.046***	0.051***	0.021**	0.046***	-0.007
	(0.006)	(0.007)	(0.009)	(0.005)	(0.013)
age2	-0.001***	-0.001***	-0.001**	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
foreign group	0.010	0.027	0.033	0.059*	0.014
	(0.035)	(0.048)	(0.037)	(0.035)	(0.054)
Domestic group	0.039	0.038	0.007	0.017	0.010
	(0.025)	(0.031)	(0.029)	(0.024)	(0.049)
Obs.	5718	4291	5287	6540	3038

Notes: The top panel of the table report estimates of regressions that controls for RSA using a dummy that equals unity for all the periods in and after a firm has participated in the program and zero otherwise The bottom panel of the table report estimates of regressions that controls for RSA using the variable NGE_RSA ito measure the investment subsidy received by the firm (total payment divided by total investment cost). All columns include a full set of firm dummies and time dummies. Standard errors below coefficients are robust to heteroskedacity and arbitrary serial correlation (they are clustered by firm. Time period is 1985-2004.

Table 12: Cost per job calculations

	(1)	(2)	(3)	(4)	(5)
size	average employment	average grant value	coefficient value	number of jobs created	cost per job
<=100	54	209,266	0.258	16	13,125
>100	581	1,211,022	0.134	83	14,532
<=250	93	375,934	0.197	20	18,554

Table 13: Results for the whole UK

	(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	0.343***			4.988***	0.168***			0.801***
	(0.022)			(0.712)	(0.014)			(0.160)
NGE = 10%		0.272*	0.042			-0.032	0.022	
		(0.156)	(0.057)			(0.034)	(0.020)	
NGE = 15%		0.374***	0.123***			-0.015	0.057***	
		(0.059)	(0.028)			(0.020)	(0.016)	
NGE = 20%		0.398***	0.048***			0.030***	0.014***	
		(0.020)	(0.009)			(0.004)	(0.005)	
NGE = 30%		0.415***	0.087***			0.045***	0.036***	
		(0.030)	(0.012)			(0.009)	(0.006)	
NGE = 35%		0.257***	0.190***			0.028	0.090***	
		(0.052)	(0.031)			(0.021)	(0.017)	
Observations				105346				
Number of Firms				24549				
F-stats for excluded instruments			20.11				14.12	
Fixed effects	NO	NO	NO	NO	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 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.

5. Conclusion

This report summarises the results of an econometric evaluation of the Regional Selective Assistance programme for Wales. This follows and earlier similar study for the UK as a whole.

The main findings are that participation in the RSA has a significant positive impact on employment and investment of participating firms. We cannot find any effects on productivity of firms. The quantitative values are very similar to the results found earlier for the whole of the UK. We employ a number of specifications. Our most general specification controls for unobserved time unvarying heterogeneity between firms. However, we find that the available instruments do not have enough explanatory power to lead to meaningful results. It is reassuring, however, that for the UK as a whole the instrumentation strategy works well and typically leads to estimates that suggest stronger treatment effects.

We examine heterogeneity of the treatment effects by dividing the sample between low and high tech sectors. This is stretching the data somewhat as we have only a limited number of observations and does not produce any evidence of significant differences between different sectors.

We finally look at treatment heterogeneity between different size classes. It appears that estimated treatment effects are stronger for smaller firms. We also look at the "cost per job", the amount of subsidy spent to create one additional job according to our estimation results. Interestingly, despite the larger treatment effect smaller firms do not seem to have much lower cost per job values than larger firms.

We would like to close this report with a suggestion that would make similar studies in the future easier and more reliable. The quality of evaluation analysis as provided in this paper depends crucially on the quality and success of the match from administrative data on RSA applicants to independent performance data. Matching for this report was based on name and address data. A simple change to ensure better matching rates would be to request IDBR reference numbers from firms (where available) as part of the application process. If this is not possible, it is clearly important to retain the address data.

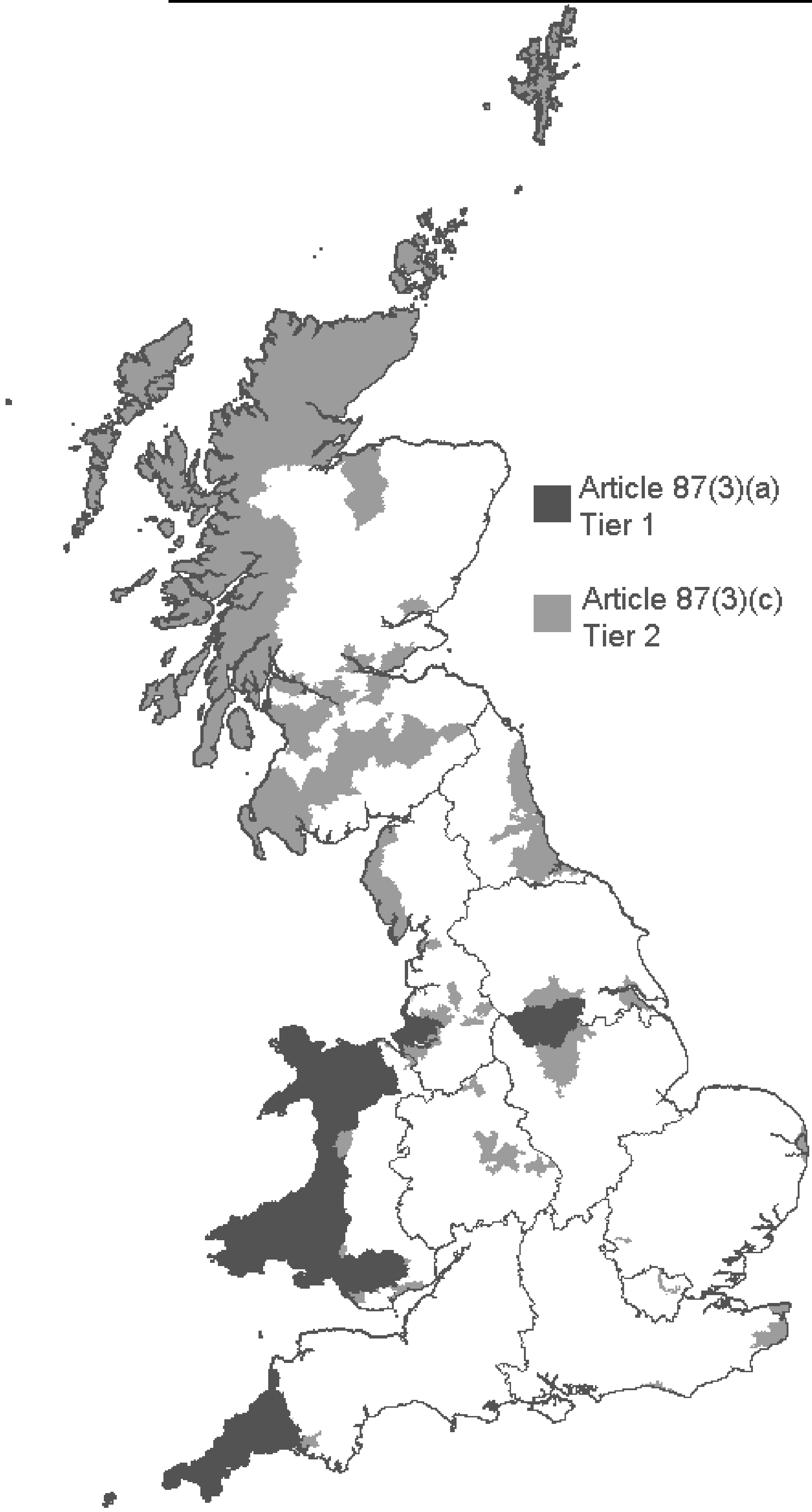
Figure 1: Assisted Areas Map prior to August 1st 1993

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

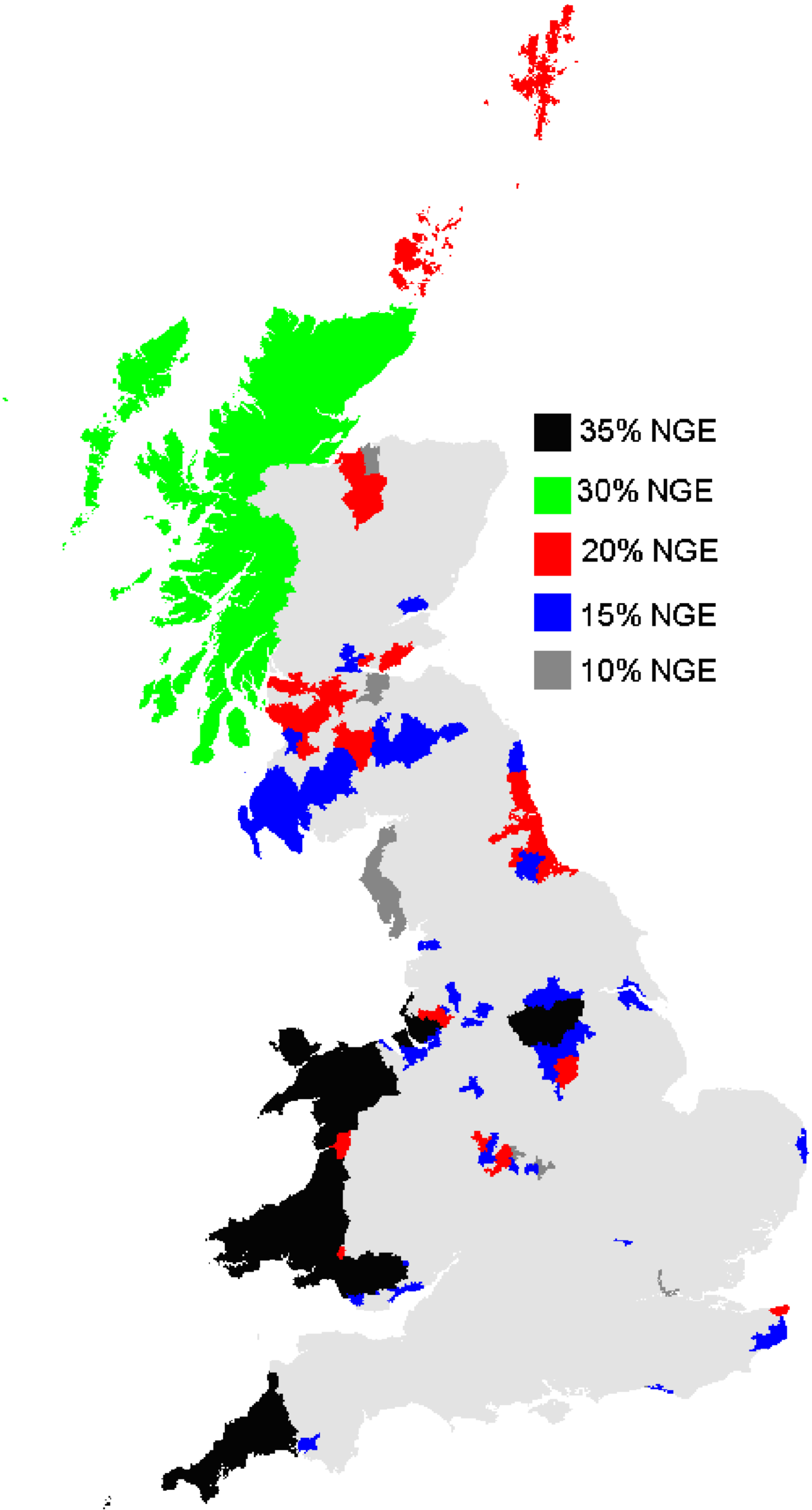
Notes: The shaded areas are those which are eligible for some Regional Selective Assistance. The dark shaded areas are the very deprived areas eligible for an investment subsidy of up to 30% NGE (Net Grant Equivalence). The light shaded areas are eligible for up to 20% NGE.
Source: Department of Trade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

Appendix A - Background on econometric evaluation methods

The objective of econometric evaluation techniques is to estimate the causal impact of a particular scheme on an outcome variable such as productivity; i.e. we try to answer the question: “what would productivity be like if the scheme had not taken place?”

Following the literature we define scheme participation very broadly as “treatment”. Ideally, to evaluate the effect of treatment we would want to observe what would happen to the “treated” firm had it not participated in the scheme. However, because this is not observable, we need to find an alternative approach that allows us to evaluate “the treatment effect on the treated”, i.e. the effect of the scheme on participating firms. In what follows we will describe the assumptions under which different methodologies consistently estimate the “treatment on the treated”.

Let us start from the case in which scheme participation is completely random. A random allocation of the treatment creates directly comparable treatment and control groups and allows researchers to estimate the treatment effect simply as the difference between the means of the outcome variable in the treatment and control groups.

However, it is very unlikely that scheme participation is random. Instead, participation is likely correlated with the expected benefits from the treatment. Since participation is not random, if firms who participate are simply compared with those who did not, the estimates will suffer from selection bias. In this case quasi-experimental methods need to be used to construct suitable control groups. The idea is to use observed data together with some appropriate identifying assumptions to “construct” the missing counterfactual using control groups: firms to whom the intervention is applied (the treatment group) are matched with an “equivalent” group from which the intervention is withheld and the average value of the outcome indicator for the treatment group is compared with the average of that for the constructed control group.

We will group the various estimators in two broad categories:

The first category includes all the methods that assume that participants and non-participants only differ in terms of observable characteristics which we can control for (what econometricians call “ignorability assumptions” or “selection on the observables”). In this category we include OLS and other regression methods, methods based on propensity score and other matching methods. The second category includes all the estimators based on the existence of an instrumental variable that helps explain participation to the program but has no direct effect on the outcome (in our case productivity).

Let us start by assuming that it is possible to control for all possible reasons why outcomes might differ between participants and non-participants and that there is a single homogenous effect of the scheme on participants. In this case one might try estimating the treatment effect using OLS. Multivariate regression analysis is used to control for observable characteristics that distinguish participants and non-participants. The treatment effect is estimated as the differences in the mean outcomes of the two groups, participants and non-participants, conditional on the set of variables that cause outcome and participation.

There are several potential sources of bias. First, there may be differences in the effect of the program across different firms. At best we can try and estimate the mean effect rather than the effect on each firm. Secondly, there may be omitted variables. Thirdly, there may be a lack of common support between the treatment and control group as signalled by large differences in the empirical distributions of observables between the two groups of firms. Matching estimators try to solve these sources of bias incurred by OLS, but still rely on the assumptions that there is only selection on observables (so do not deal with the second problem).

The idea of matching estimators is to identify among the non-treated firms a “control” group of firms with similar observable characteristics (common support) to the treatment group. In this case, the difference in the average outcome between the treated and the matched non-treated firms (the control group) consistently estimates the effect of the treatment on the treatment group.

However, we face further problems if both observable and *unobservable* firm characteristics drive participation outcomes. In this case, matching estimates, although more flexible than OLS estimates, are still affected by bias due to these un-observables.

One possible solution to this bias is provided by the “difference-in-difference” estimator. This method compares a treatment and a control group (first difference) before and after the intervention (second difference). Once the mean difference between the “after” and “before” values of the outcome variable for each of the treatment and control groups is calculated, the difference between these two mean differences is calculated. This difference in difference is the estimate of the impact of the program.

The main drawback of this approach is the need to identify the control group: a group that is unaffected by the program but who would have responded identically to changes in the environment as would the treatment group. There are many possible strategies for constructing the control group such as using geography – some areas are eligible for the program (e.g. pilots) and some are not and a strategy is to compare the firms in pilot areas with those in non-pilot areas. A variant of this is to use firms on either side of a boundary. Another possibility is size – firms just below a size threshold are eligible whereas those just above are not eligible. Having multiple possible dimensions enables us to test for the identification assumptions underlying the difference in difference approach.

Formally, this method estimates the effect of the treatment consistently if we assume that the un-observables that affect participation decision and outcome are separable into (i) an individual-specific effect (constant over time); (ii) a common macroeconomic effect, which is the same across all firms (common trends assumption) and (iii) an idiosyncratic shock that is not correlated with participation and the outcome of interest. Note that matching can be combined with difference in differences. We might select a sub-sample of the non-pilot areas for example that are more closely matched on the observables with the treatment group[7].

If we are not willing to accept the assumption of selection on observables nor assumptions (i)-(iii) on un-observables, we must turn to our second category of estimators which relies on the use of Instrumental Variables (IV). Instrumental variables are chosen so that they determine program participation, but do not affect outcomes given participation. This identifies the exogenous variation in outcomes attributable to the program, acknowledging that the treatment may be non-random. The "instrumental variables" are first used to predict program participation; then one sees how the outcome indicator varies with the predicted values. Instruments might be constructed on the basis of variations in scheme availability in different geographical areas or over time. However, in general, finding a suitable instrument is not that easy since it must satisfy the criteria of being correlated with the participation choice while being correctly excluded from the productivity equation.[8]

In the results sections of this report we are discussing for each of the schemes which instruments might be available or if only the more basic evaluation techniques can be applied in a possible Stage 2 of the project.

[14] See Blundell et al. (2004).

[8] The control function approach is similar in nature to the IV estimator. We do not discuss it here for brevity.

Appendix B - Glossary

Annual Respondents Database (ARD): Provides the performance data for firms. It is the data underlying the Annual Census of Production in the UK. Collected by the Office for National Statistics (ONS), the ARD is an extremely rich data set which contains information on a sample of UK establishments that account for about 80% of total UK employment.

Coefficient: Measures the impact of changes in an explanatory variable on a dependent variable. See also explanatory variables and dependent variables.

Coefficient estimates: Our “best guess” at the true value of a coefficient based on a statistical analysis of the underlying data.

Confidence intervals: Used to demonstrate the uncertainty about the measured impact of an explanatory variable on performance. In our study, used to demonstrate the degree of uncertainty about the time profile of treatment effects. If the confidence interval for a treatment effect in a given year includes zero, then the estimated treatment effect is not significantly different from zero in that year (see also Significance)

Control group: A reference group of firms that have not participated in the support scheme.

Control variables: See Regression analysis.

Dependent variables: See Regression analysis.

Difference-in-difference: A method to examine the impact of a support scheme by looking at the change in a treated firms performance before and after treatment and comparing it to the change experienced by a firm that did not take part in the scheme.

Dummy variables: An indicator that takes value zero or one. It allows us to capture the impact of discrete firm characteristics on performance. For example, whether foreign firms perform differently to domestic firms or whether treated firms perform differently to non-treated firms. See also Regression analysis.

Elasticity: A scale neutral way of describing the effect of one variable on another. Defined as the percentage change in variable 1 for a percentage change in variable 2.

Endogeneity: A crucial assumption underlying linear regression is that the explanatory variable causes changes in the dependent variable and not-vice versa. For example, we need to assume that treatment causes changes to firm performance rather than changes to firm performance resulting in treatment. If the latter is the case, the regression specification suffers from endogeneity (because of the inclusion of an endogenous variable) and as a result we may overstate the impact of treatment. This problem can also occur if some unobserved characteristic of firms is related to both treatment and firm performance. For example, if a new manager both raises firm performance and applies to a support scheme we may overstate the impact of treatment.

Explanatory variables: See Regression analysis.

Fixed effects: A dummy variable that is specific to each firm but does not vary over time. They are introduced to control for time invariant unobserved characteristics of firms (e.g. the quality of firm management). It is important to control for these unobserved characteristics if they might affect both participation and performance.

Insignificance: See significance.

Instrumental variable (IV): When we are worried that an explanatory variable may be endogenous, one solution is to look for an alternative variable that we expect to be related to the endogenous explanatory variable but to be unrelated to firm performance. For example, if we are worried that treatment may be partly caused by changes in firm performance we can look for a variable (such as eligibility for the scheme) that determines the likelihood of treatment but not performance. We can then use these 'instrumental' variable to proxy for the endogenous explanatory variable.

Labour productivity (LP): A measure of the amount of output produced per worker (defined as value added per employee)

Local unit (LU): A business unit at single mailing address.

Regression analysis: In this instance, a methodology for estimating the impact that firm characteristics have on performance (productivity, size etc). The measure of performance may be referred to as the dependent variable (it depends on the firm characteristics). The firm characteristics that are thought to affect performance are often referred to as explanatory or control variables (because they help explain performance). Regression analysis allows us to estimate the impact of treatment on firm performance while holding constant (or controlling for) other characteristics of firms. These other characteristics are sometimes called control variables.

Reporting unit (RU): The fundamental unit of observation in the ARD. May be made up of several local units.

Significance: When undertaking regression analysis it is possible that we detect a relationship between an explanatory variable and a dependent variable just by chance. Significance tells us how likely it is that the relationship that we observe in the data could have arisen by chance. Generally, economists discount findings where there is a more than 10% probability that the results occurred by chance. In such cases, we refer to the results as insignificant.

Total factor productivity (TFP): A measure of how efficiently the firm uses all factors of production. We use various definitions of TFP that are defined in the text.

Treatment: Participation in a support scheme.

Treated: Firms that participate in a support scheme (i.e. that receive treatment)

Treatment effects: The impact of scheme participation on firm performance (usually reported after controlling for other firm characteristics that might also affect performance.

Technology level The table presents a classification of the manufacturing industries according to technology intensity using an ISIC Rev. 3 activity breakdown and the classification suggested by the OEC.

The OECD determines the division of manufacturing industries into high-technology, medium-high-technology, medium-low-technology, low-technology groups after ranking the industries according to their average over 1991 to 1997 of aggregate OECD R&D intensities. Industries classified to a superior category have a higher average OECD intensity for both indicators than industries in an inferior category. Also considered were: *i)* time stability: for adjacent years, industries classified to a superior category have a higher average OECD intensity than industries in an inferior category; and *ii)* country-median-stability: industries classified to a superior category have a higher median-intensity, than industries in an inferior category.

	ISIC Rev. 3
High-technology industries	
Aircraft and spacecraft	353
Pharmaceuticals	2423
Office, accounting and computing machinery	30
Radio, television and communication equipment	32
Medical, precision and optical instruments	33
Medium-high-technology industries	
Electrical machinery and apparatus, n.e.c.	31
Motor vehicles, trailers and semi-trailers	34
Chemicals excluding pharmaceuticals	24 excl. 2423
Railroad equipment and transport equipment, n.e.c.	352 + 359
Machinery and equipment, n.e.c.	29
Medium-low-technology industries	
Coke, refined petroleum products and nuclear fuel	23
Rubber and plastic products	25
Other non-metallic mineral products	26
Building and repairing of ships and boats	351
asic metals	27
Fabricated metal products, except machinery and equipment	28
Low-technology industries	
Manufacturing, n.e.c. and recycling	36-37
Wood, pulp, paper, paper products, printing and publishing	20-22
Food products, beverages and tobacco	15-16
Textiles, textile products, leather and footwear	17-19
Total manufacturing	15-37

References

Baily, M.N.; Hulten, C.and Campbell, D. (1992) “Productivity Dynamics in Manufacturing Plants”, *Brookings Papers: Microeconomics*, 187-267.

Blundell, R., Costa-Dias, M., Meghir, C. and Van Reenen, J. (2004) “Evaluating the employment effects of a mandatory job search assistance programme” *Journal of the European Economic Association*, 2(4) 569-606.

Criscuolo, C., Martin, R., Overman, H., van Reenen, J., ‘Longitudinal Micro Data Study of Selected DTI Business Support Programmes – Stage 1’, London, March 2006

Criscuolo, C., Martin, R., Overman, H., van Reenen, J., ‘Longitudinal Micro Data Study of Selected DTI Business Support Programmes – Stage 2’, London, February 2007.

Criscuolo, C., Haskel, J. & Martin, R. (2003), 'Building the evidence base for productivity policy using business data linking', *Economic Trends* 600, 39-51.

DuMouchel, W.H. and Duncan, G.J. (1983), “Using sample weights in multiple regression analysis of stratified samples.”, *Journal of the American Statistical Association* 78(383), 535-543.

Girma, S., Gorg, H. and Maioli, S. (2005), “Longitudinal Microdata Study of selected DTI Business Support Programmes”, DTI mimeo.

Harris, R. and Robinson, C. (2004), “Industrial policy in Great Britain and its effect on total factor productivity in manufacturing plants, 1990-1998”, *Scottish Journal of Political Economy*, Vol. 51, No. 4, September.

Harris, R. and Robinson, C. (2001), “An Analysis of current DTI Industry Support Patterns”, Technical report, Research Project on DTI Industrial Support Policies.

Heckman, J.J. (1976) “The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models.” *Annals of Economic and Social Measurement 5*, 475-492.

Heckman, J.J. and Navarro-Lozano, S. (2004) “Using matching, instrumental variables, and control functions to estimate economic choice models.” *Review of Economics and Statistics* 86, 30-57.

Heckman, J.J. and Robb, R. (1986) “Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes”, in: Wainer, H. (Ed.), Drawing Inferences from Self-Selected Samples. Springer-Verlag, New York, pp. 63-113.

Heckman, J.J., Tobias, J.L. and Vytlacil, E. (2003) “Simple estimators for treatment parameters in a latent-variable framework”. *Review of Economics and Statistics* 85, 748-755.

Heckman, J.J. and Vytlacil, E. (1998) “Instrumental variables methods for the correlated random coefficient model.” *Journal of Human Resources* 33, 974- 987.

Imbens, G. W. and Angrist, J. D. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62, pp. 467-75.

Jaffe, A.B. (2002), “Building program evaluation into the design of public research support programs”, *Oxford Review of Economic Policy* 18(1).

Perry, J. and Vale, S. (2003),'Dissemination of Information on Business Registers in the UK', paper prepared for the Joint UNECE/Eurostat Seminar on Business Registers.

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

Secretary of State for Trade and Industry, the First Minister of the Scottish Executive and the First Minister of the National Assembly of Wales (2002) “Industrial Development Act 1982: Annual Report”, London, The Stationary Office.

Vella, F. and Verbeek, M. (1999) “Estimating and interpreting models with endogenous treatment effects.” *Journal of Business and Economic Statistics* 17, 473-478.

Wooldridge, J. (2002) “*Econometric Analysis of Cross Section and Panel Data.*” Cambridge, MA: MIT Press, 2002.