Exploration of Data-Matching as an Evaluation Technique

Merging of 'Designated Relationship Managed' Companies into the ONS Annual Respondents Database and Econometric Analysis

Final Report to SQW Consulting for
Scottish Enterprise

By

Richard Harris

June 2009

[this page is blank to allow double-sided printing  *Save Paper]*

Contents

[this page is blank to allow double-sided printing  *Save Paper]*

Executive Summary

Introduction

E.1 Scottish Enterprise (SE) commissioned SQW Consulting to consider the feasibility of undertaking an econometric analysis of the Account and Client Management programme by combining SE administrative records on Account and Client Managed businesses with business performance data (such as turnover, employment, value add, etc) from the ONS database on company characteristics and performance known as the Annual Respondents Database (ARD).

E.2 Under Account and Client Management, SE provides an especially intensive form of support to companies that are considered to be capable of benefiting from a high level of attention. Account and Client Management are now together commonly referred to by SE as 'Direct Relationship Management' (DRM) and this is the term that will be used for the remainder of this report.

E.3 SQW worked on this assignment with Professor Richard Harris of Glasgow University. Professor Harris undertook the detailed econometric analysis.

E.4 The main thrust of this research is to complement more traditional evaluation survey approaches with econometric techniques, in order to go some way to determining the causal impact of DRM interventions on individual business performance and the Scottish economy. In addition to the (limited) econometric analysis of the DRM programme presented here, it is intended that this research will also establish a framework for future analyses of this type.

E.5 The research involved two phases. The first phase set out to answer the following questions:

- is there scope to perform econometric analysis of DRM support by combining SE records with business data contained in the ARD?
- what are the limitations - present and future - of utilising this type of analysis for evaluation of the DRM programme?

E.6 The second phase of the analysis (the undertaking of econometric work to consider whether DRM status has impacted on firm performance) was dependent on the outcome of phase one. Despite issues of the size of the linked DRM-ARD database (see section 3), it was agreed to undertake some exploratory work. Thus the key questions addressed in phase two of the research are:

- how representative is the sample of DRM firms (to be used in econometric modelling of firm performance) of all DRM firms?
- is the full DRM sample linked into the ARD a representative sample of all Scottish firms (i.e. how does the sample differ in terms of such important

characteristics as industry composition, age, size, R&D, capital-intensity, etc.)?

- after matching the DRM sample with a 'control group' of similar firms operating in the North East and North West of England, is there any evidence that DRM status led to any differences in firm performance?

Data linking and matching[1]

E.7 **Section 2** of the report reviews the DRM data supplied by SE. It reports that some 2,311 firms now have DRM status (for which SE products have been made available), but only 1,742 had been DRM assigned by 2005 (the remainder become DRM post-2005). Ultimately, this is the sub-group of firms we need to match into the ONS ARD for subsequent econometric analysis, since the ARD currently only contains information for the 1997-2005 period. DRM firms that entered *after* 2005 should be excluded from any analysis based on 2005 ARD data as there will be no information from ARD on their performance.

E.8 **Section 3** then reports on how many of the DRM firms were able to be linked into the ARD, before providing some analysis of the characteristics of the linked dataset. Table E1 summarises the problems of data attrition experienced when first linking the DRM data into the IDBR (the Inter-Departmental Business Register) and then linking into the ARD (ARD is a sub-set of IDBR).

Table E1: Data attrition when linking

Linking to IDBR	
2,311[(a)]	Account or Client Managed firms in SE database supplied
-686	*Cannot be found in IDBR*
1,625	Linked by Scottish Government (SG) statisticians to IDBR
-25	*Duplicate records*
1,613	Uniquely linked by SG statisticians to IDBR
-291	*Firms who apparently received no products*
1,322	DRM firms linked to IDBR for linking to ARD
-189	*Entered DRM status post-2005*
1,133	DRM firms entered by 2005 linked to IDBR for linking to ARD
Linking to ARD	
1,322	DRM firms linked to IDBR for linking to ARD
-353	*Not found in ARD (see para. 3.6) – probably mostly not covered in ARD*
969	Linked to ARD
-115	*Entered into DRM status post-2005*
854[(b)]	DRM-ARD matched database
-589	*Those without information on when first assisted by SE*
265[(c)]	**DRM-ARD database available for econometric analysis**

[(a)] of these 1,742 were DRM assigned by 2005
[(b)] of these, 614 have full financial information for use in section 4
[(c)] of these 195 have full financial information for use in section 4

[1] To avoid confusion, the term 'linking' is used to discuss the merging of different datasets comprising a cohort of firms; 'matching' refers to comparing firms that have undergone some form of 'treatment' (such as receiving assistance as DRM firms) with a 'control' group of firms with similar characteristics (the only major difference being that the control group did not receive the treatment).

E.9 Starting from the 1,742 firms that needed matching, some 1,133 (or 65%) were assigned IDBR local unit reference codes. (The fact that not all 1,133 firms could then be located in the ARD is less of an issue, as their absence is most likely due to their being allocated to industries not covered in the ARD and/or they set up in business after 2005).

E.10 However, since at least 45% of the relevant 1,133 DRM firms were not found in the ARD, this presents problems for constructing a sample of matched firms *in Scotland* that were not DRM-assisted but which are a valid comparison or 'control' group. It is more than likely that when constructing the matched sample, firms that did have DRM status, but which could not be identified in the ARD, would be erroneously allocated to the control group of firms against which the identified DRM firms are compared.

E.11 From this we conclude that any comparison of DRM firms should involve a control group drawn from another UK region where the type of products offered to DRM firms in Scotland are not generally available. Based on our knowledge of the North East and North West regions of England, it is likely that firms operating in these regions are much more likely to provide an adequate control group[2].

E.12 We have ARD data for many of the DRM firms for the period 1997-2005. However, according to the information provided by SE, the great majority of DRM firms only took on this status at the beginning of 2005, at which point SE had changed its approach to the programme. That being the case, since we do not have ARD data on firm performance *beyond* 2005, few conclusions could in theory be drawn about the relationship between receiving DRM support and changes to business performance.

E.13 However, after discussion with SE, it became apparent that many of the businesses which seemed to have had entered DRM in 2005 had in fact been DRM for some time before and had thus been receiving DRM-type products before 2005. Reference to the database created by ekos/gen for a recent evaluation of the DRM programme revealed pre-2005 DRM start data for 265 firms.

E.14 These formed the main group for the econometric analysis as we have information on when they were first assisted by SE prior to 2005. Knowing when firms were first assisted is essential in order to benchmark the pre- and post- performance of firms against the impact of receiving the products associated with the DRM scheme. If the date when assistance was first provided is unknown, it is not possible to accurately measure the performance of firms in relation to DRM over time.

[2] SQW undertook an analysis of the likelihood that firms in these regions might have benefited from assistance from organisations such as Regional Development Agencies or Business Link. The conclusion was that the possibility of any firms in the control group having received DRM-type assistance, and thus 'contaminating' the analysis, was very small. The analysis is contained in Appendix B.

E.15 Given these issues of data linking, **Section 4** undertakes some *exploratory* econometric work related to measuring the performance of DRM firms against linked control groups of firms drawn from the North East and North West regions of England, to test if DRM status 'made a difference'. The key questions addressed by the research are:

a. how representative is the sample of DRM firms (to be used in the econometric modelling of firm performance) of all DRM firms?

b. is the full DRM sample linked into the ARD a representative sample of all Scottish firms?

c. after matching the DRM sample with a 'control' group of similar firms operating in the North East and North West of England, is there any evidence that DRM status led to any differences in firm performance?

E.16 With regard to the first question (par. E.15.a) statistical analysis shows that the two sub-groups are very similar in terms of their overall characteristics and thus that the sample of DRM firms to be used for analysis is indeed a representative sample of all DRM firms.

E.17 When we consider if the full DRM sample linked into the ARD is a representative sample of all Scottish firms (cf. par. E.15.b.), we find evidence that shows there is a significant difference between DRM firms and non-DRM firms. DRM firms are significantly older, and have higher productivity, capital- and intermediate-inputs intensity, involvement in R&D, employment size, and foreign-ownership. They are less likely to be single-plant enterprises and more likely to operate in assisted areas. The local authority areas they are located in are less likely to be diversified (although the difference is small), but they have higher levels of agglomeration and population density. There are also significantly different patterns across industries.

'Control group' approach

E.18 These results are to be expected as DRM firms are the high growth enterprises obtaining assistance from SE; but such higher growth characteristics might also suggest that even without assistance they would likely do better than other non-assisted firms, which means that any assessment of the impact of DRM status must 'control' for such potential 'sample selection' effects.

E.19 This is an important approach to use as it allows a comparison of performance of assisted firms taking into account the counter-factual position (i.e. comparing performance with what would have happened if there had not been any assistance from SE, and thus whether such assistance made a difference). Simply comparing the performance of assisted firms with non-assisted firms is open to what is termed sample-selectivity bias – i.e. that such firms have characteristics that give them a higher growth potential that they would realise irrespective of whether assistance is available or not. Thus, any differences between DRM and non-DRM firms are likely to be a biased indicator of whether SE help really did have an impact, unless we control for sample-selectivity by comparing the treatment group (those receiving assistance) with a

'control group' (i.e., firms with strictly similar characteristics to the treated group but with the defining difference that they did *not* receive assistance).

Results

E.20 The results from the econometric analysis matching Scottish DRM firms with a control group of firms from the North East and North West of England (who are similar but did not receive DRM-type assistance) suggest that we can be (better than 95%) confident that Scottish DRM firms certainly did not do better and in fact the statistical evidence points to them having done worse in terms of *total factor productivity* (TFP), the key variable that has been used for the analysis.

E.21 Using matched data for 195 Scottish DRM firms and the control group from the North East, the results show that that there is some (weak) evidence that before being assisted, DRM firms had higher productivity; in the year when assistance started productivity significantly declined by around 9% with further (insignificant) falls until years t+5 and t+6. The latter summarises the whole post-assistance period average effect showing that from the 6th year onwards there was a significant overall decline in TFP of nearly 16% in Scottish DRM firms in comparison with the control group of similar firms in the North East.

E.22 With regard to the control group from the North West, the results show that there is again some (weak) evidence that several years before being assisted, Scottish DRM plants had higher productivity and then productivity fell in both t–2 and the year when assistance started (with declines of around 30% and 15%, respectively). After assistance, there were further (insignificant) falls until years t+5 and t+6. The latter summarises the whole post-assistance period average effect showing that from the 6th year onwards there was a significant overall decline in TFP of 33% in Scottish DRM firms in comparison with the control group of similar firms in the North West.

E.23 Any interpretation of these data should, however, take account of two important features of the analysis:

a. it relates to forms of intervention delivered by SE between five and ten years ago

b. it was essentially an exploratory exercise; a more full analysis covering later years, with an improved level of linking between DRM company data and ARD data, might produce different results.

Looking forward – repeating the analysis

E.24 The results from this study also show what is needed to undertake this type of econometric analysis (and therefore the lessons to be learned); if this exercise is to be repeated in the future (when more and better data are available) then it needs to take account of the following:

a. DRM firms can be matched to the IDBR (and subsequently the ARD) with much greater accuracy if their VAT registration number is collected from them. It is also useful to have them provide their company registration

number (CRN) in case the VAT number is missing (e.g. for firms below the VAT threshold) and possibly their PAYE code. All three numbers are required information that the companies must use for VAT, Companies House and Revenue & Customs purposes. All three codes (especially VAT codes) are available in the IDBR.

b. DRM company records should be updated if and when a company changes status (e.g. it is taken-over or merges).

c. In order to undertake an analysis of whether the firm experienced a change in performance as a result of SE assistance, it is necessary to have (accurate) information on when the firm first received assistance from SE (as well as information on what type of assistance was received). Otherwise, it is not possible to measure pre and post-assistance performance.

d. The DRM database should ideally also contain information on some of the most important characteristics of the firm – such as its industry SIC code and numbers of employees. Neither of these was available in the DRM database provided for this study.

e. If in the future SE wish to know if the reorganisation of DRM firms that occurred post-2004 has had a positive impact on performance, then they will need to wait until a longer-in-time ARD sample is available (currently data up to 2005 is only available). It is likely that data for both 2006 and 2007 ARD will be available within one year of this report being written.

[this page is blank to allow double-sided printing  *Save Paper]*

1. Overview

1.1 Scottish Enterprise (SE) commissioned SQW Consulting to consider the feasibility of undertaking an econometric analysis of the Account and Client Management programme by combining SE administrative records on Account and Client Managed businesses with business performance data (such as turnover, employment, value add, etc) from the ONS database on company characteristics and performance known as the Annual Respondents Database (ARD).

1.2 Under Account and Client Management, SE provides an especially intensive form of support to companies that are considered to be capable of benefiting from a high level of attention. Account and Client Management are now together commonly referred to by SE as 'Direct Relationship Management' (DRM) and this is the term that will be used for the remainder of this report.

1.3 SQW worked on this assignment with Professor Richard Harris of Glasgow University. Professor Harris undertook the detailed econometric analysis.

1.4 The main thrust of this research is to complement more traditional evaluation survey approaches with econometric techniques, in order to go some way to determining the causal impact of DRM interventions on individual business performance and the Scottish economy. In addition to the (limited) econometric analysis of the DRM programme presented here, it is intended that this research will also establish a framework for future analyses of this type.

1.5 The research involved two phases. The first phase set out to answer the following questions:

- is there scope to perform econometric analysis of DRM support by combining SE records with business data contained in the ARD?
- what are the limitations - present and future - of utilising this type of analysis for evaluation of the DRM programme?

1.6 The second phase of the analysis (the undertaking of econometric work to consider whether DRM status has impacted on firm performance) was dependent on the outcome of phase one. Despite issues of the size of the linked DRM-ARD database (see section 3), it was agreed to undertake some exploratory work. Thus the key questions addressed in phase two of the research are[3]:

[3] The original set of questions for the econometric analysis were: (i) what are the performance dynamics of account and client managed business?; (ii) how does this compare to non-account and client managed businesses?; (iii) does performance of account and client managed businesses vary by type of business (e.g. by sector, size, location, age of business, previous growth performance, type of Scottish Enterprise intervention/product received, etc); and (iv) does the available data allow an assessment of the impact of account and client management on business performance and the Scottish

- how representative is the sample of DRM firms (to be used in econometric modelling of firm performance) of all DRM firms?

- is the full DRM sample linked into the ARD a representative sample of all Scottish firms (i.e. how does the sample differ in terms of such important characteristics as industry composition, age, size, R&D, capital-intensity, etc.)?

- after matching the DRM sample with a 'control group' of similar firms operating in the North East and North West of England, is there any evidence that DRM status led to any differences in firm performance?

1.7 Lastly information is also provided, in the section comprising the econometric analysis, on the methodology used and in particular the approaches that can be used to test for the impact of policy when it is likely that firms receiving assistance have certain characteristics which (if not controlled for) will bias any results obtained. The approach used here is known as 'matching' (specifically propensity score matching was used)[4], which is fast becoming the standard approach used in this type of analysis. We discuss its strengths and weaknesses, as it is important to understand these when using the technique.

economy? If so, what is this impact? Because of data limitations, it was agreed at the Interim Report stage to amend the research to cover the questions set out in the main text.

[4] To avoid confusion, the term 'linking' will be used throughout this report to discuss the merging of different datasets comprising a cohort of firms; 'matching' refers to comparing firms that have undergone some form of 'treatment' (such as receiving assistance as DRM firms) with a 'control' group of firms with similar characteristics (the only major difference being that the control group did not receive the treatment).

2. Review of SE DRM data

Product-based analysis

2.1 The database supplied by SE comprised 13,636 observations on Account or Client Managed Growth firms, of which 11,085 had a product associated with them delivered at a specific date (thus 2,551 did not have either a product or a delivery date). Of those with no delivery date/product, the majority left 'Designated Relationship Managed' (DRM) status without ever being associated with a product (70% of the 2,551); while the remaining firms were nearly all DRM but were still waiting to receive a product.[5]

2.1 Thus, for present purposes we should concentrate on the 11,085 observations (or products) delivered and ignore the 2,551 without products since they have not (yet) received assistance from SE.[6]

Table 2.1: Year entered DRM and when product was delivered

Year entered DRM	Year in which product was delivered						
	2003	2004	2005	2006	2007	2008	Total
2003	3	1	0	0	0	0	4
2004	0	0	3	14	2	0	19
2005	0	0	2107	4125	2361	278	8871
2006	0	0	0	761	716	71	1548
2007	0	0	0	0	501	111	612
2008	0	0	0	0	0	31	31
Total	3	1	2110	4900	3580	491	11085

Source: own calculations based on DRM database

2.2 Firms have DRM status when they receive their product(s). However, the delivery date can be some time after entering DRM status, as Table 2.1 shows. The majority of observations were assigned as DRM in 2005 (when SE's programme of support to DRM companies was substantially overhauled, suggesting that some/most of the firms represented were probably client companies beforehand, and had received previous help[7]), but only some 24% of these were product-assisted in that year with 47% (i.e. 4125 out of 8871)

[5] There would appear to be 4 observations wrongly coded as 'Non Relationship Managed' NRM and still awaiting a product.

[6] That is, each observation/product refers to one product per company per year. Thus it can be seen the average firm received 4.8 products over 2005-2008, given there were 2,311 firms receiving in total 11,085 products.

[7] Information from a recent survey of DRM firms on when firms first received help from SE was supplied to the Scottish Government statisticians that linked the DRM database to the Inter-departmental Business Register (IDBR) – see next section – and while this only covers some 30% of all firms in the database it does suggest a significant number had received assistance well before receiving DRM status.

assisted in 2006 and 27% assisted in 2007. In 2006 some 1,550 additional observations were assigned as DRM, falling to 612 in 2007.

2.3 Table 2.2 shows when an observation was assigned as DRM and when it left DRM status. It shows that overall some 75% still have DRM status (although it is unclear what is meant by assigning a date for leaving DRM that is beyond the current date[8]).

Table 2.2: Year entered and left DRM status

Year entered DRM	Year left DRM status						
	2005	2006	2007	2008	>2008	Not left	Total
2003	4	0	0	0	0	0	4
2004	1	0	1	0	0	17	19
2005	22	1008	1114	168	32	6527	8871
2006	0	76	237	39	3	1193	1548
2007	0	0	36	16	15	545	612
2008	0	0	0	0	0	31	31
Total	27	1084	1388	223	50	8313	11085

Source: own calculations based on DRM database

Table 2.3: Product Group by year product was delivered

Product Group	Year product delivered				
	2005	2006	2007	2008	Total
BUSDEVELOP	120	633	489	76	1318
INNOVATION	60	170	101	12	343
INVESTMENT	7	6	2	0	15
MARKETDEV	439	1104	1089	178	2814
STARTUP	571	1470	527	20	2588
STRATEGY	473	649	372	53	1547
WORKFORCE	321	776	965	151	2213
Non-SE funded	27	42	35	1	105
Products no longer used	78	50	0	0	128
Total	2096	4900	3580	491	11067[a]

[a] 18 products had no information on their product group

Source: own calculations based on DRM database

2.4 Table 2.3 shows the distribution of products by when they were delivered. Note, products are delivered in categories (with a range of products available under each category) many of which were specific to individual LECs. Some categories are not SE-funded or no longer used, although information has been supplied by SE that will allow these categories to be recoded to other sub-groups. Note, only firms that had DRM status and who received STARTUP and STRATEGY products are included here (the latter products were also supplied

[8] Dates include the following years: 2037, 2099, 2100 and 2999.

to a much larger group of non-DRM firms by SE during the period covered). Hence, products associated with market development, workforce development (training, etc.) and business development are also important (compared to the overall portfolio of help provided by SE to all firms).

Firm-based analysis

2.5 The 11,085 observations (or products) were delivered to **2,311 DRM firms**. Some 600 firms (or 26%) received only one product, the median firm received three, and nearly 10% of firms received ten or more products (one firm received 84 products).

2.6 Tables 2.4 – 2.6 reproduce the earlier Tables 2.1 – 2.3 but aggregated to the level of the firm. Table 2.4 confirms that the majority of firms (75%) who received products were assigned as DRM in 2005, but 43% of these received assistance after 2005.

Table 2.4: Year firm first entered DRM and when product was delivered

Year first entered DRM	Year in which product was first delivered						
	2003	2004	2005	2006	2007	2008	Total
2003	1	0	0	0	0	0	1
2004	0	0	3	0	1	0	4
2005	0	0	988	618	124	7	1737
2006	0	0	0	234	123	4	361
2007	0	0	0	0	174	26	200
2008	0	0	0	0	0	8	8
Total	1	0	991	852	422	45	2311

Source: own calculations based on DRM database

Table 2.5: Year firm first entered and year left DRM status

Year first entered DRM	Year firm left DRM status						
	2005	2006	2007	2008	>2008	Not left	Total
2003	0	0	0	0	0	1	1
2004	0	0	1	0	0	3	4
2005	9	267	277	37	3	1144	1737
2006	0	30	57	9	3	262	361
2007	0	0	9	5	9	177	200
2008	0	0	0	0	0	8	8
Total	9	297	344	51	15	1595	2311

Source: own calculations based on DRM database

2.7 Table 2.5 confirms that overall 69% of firms still have DRM status. Lastly, Table 2.6 confirms that the present data only includes the STARTUP and STRATEGY products for only DRM (and not NRM) firms, and thus these sub-groups are significantly less important compared to the overall portfolio of assistance provided by SE for the period (which is dominated by assistance for

start-ups).[9] Note, when firms had more than one product the first product group listed in Table 2.6 was chosen as the category to be assigned.

Table 2.6: Product Group by year product was delivered

Product Group	Year in which product was first delivered				
	2005	2006	2007	2008	Total
BUSDEVELOP	372	290	95	7	764
INNOVATION	79	57	17	0	153
INVESTMENT	3	1	1	0	5
MARKETDEV	250	230	147	15	643
NONFRPART	11	7	6	0	24
NONFRSE	25	7	0	0	32
STARTUP	131	125	59	4	319
STRATEGY	87	93	58	13	251
WORKFORCE	31	42	39	6	118
Total	**989**	**852**	**422**	**45**	**2308**[a]

[a] Three firms had no information on their product group

Source: own calculations based on DRM database

Summary and conclusions

2.8 Some 2,311 firms had DRM status (for which products were made available), but only 1,742 had been DRM assigned by 2005 (the remainder become DRM post-2005 – see Table 2.4).10 Ultimately, this is the sub-group of firms we need to match into the ONS Annual Respondents Database (ARD) for subsequent econometric analysis, since the ARD currently only contains information for the 1997-2005 period. *DRM firms that entered the programme after 2005 should be excluded from any analysis based on 2005 ARD data as there will be no information from ARD on their performance*.

2.9 However, as the ARD becomes available post-2005, we shall want to match more than 1,742 firms, and therefore the next section considers matching all the DRM data available into the ARD.

[9] Originally SE provided a DRM/NRM database much bigger than the database analysed here. Since it contained all the NRM firms (e.g., helped through Business Gateway), the database was dominated by STARTUP (and to a lesser extent STRATEGY) products.

[10] There is an issue of whether we should limit the analysis to just the 991 firms had received assistance by 2005 (Table 2.4), rather than those who were DRM but had yet to receive a product(s). It may be that not all products were included in the DRM database, the products were provided over a period of time culminating in the date recorded, and/or the firm received more general assistance once DRM assigned.

3. Linking DRM data into IDBR and ARD

Linking DRM into the IDBR

3.1 In 2006 SE asked Scottish Government (SG) statisticians to find the Inter-departmental Business Register (IDBR) local unit reference codes[11] for the set of DRM firms that they supplied to SG. Information on company name and address was primarily used to find matches in the IDBR. In all, some 1,625 unique links were made for the companies in the DRM database (although 2,170 matches were obtained indicating that 545 companies had multiple local unit links, and it was not possible for the SG statisticians to say which link was the most appropriate one). Note, Table 3.1 summarises this linking process, including the implications of linking into the ARD (the next section).

Table 3.1 Data attrition when linking and matching

Linking to IDBR	
2,311[(a)]	Account or Client Managed firms in SE database supplied
-686	*Cannot be found in IDBR*
1,625	Linked by SG statisticians to IDBR
-25	*Duplicate records*
1,613	Uniquely linked by SG statisticians to IDBR
-291	*Firms who apparently received no products*
1,322	DRM firms linked to IDBR for linking to ARD
-189	*Entered DRM status post-2005*
1,133	DRM firms entered by 2005 linked to IDBR for linking to ARD
Linking to ARD	
1,322	DRM firms linked to IDBR for linking to ARD
-353	*Not found in ARD (see par. 3.6) – probably mostly not covered in ARD*
969	Linked to ARD
-115	*Entered into DRM status post-2005*
854[(b)]	DRM-ARD matched database
-589	*Those without information on when first assisted by SE*
265[(c)]	**DRM-ARD database available for econometric analysis**

[(a)] of these 1,742 were DRM assigned by 2005
[(b)] of these 614 have full financial information for use in section 4
[(c)] of these 195 have full financial information for use in section 4

3.2 A 'look-up' table comprising the 2,170 companies with IDBR local-unit codes was deposited by SG into the ONS virtual micro-laboratory (VML).[12] The VML is a secure site that is only accessible to 'approved researchers'; no information can enter or leave the VML without clearance by ONS staff, using strict rules related to disclosure.

[11] Note, the IDBR contains four sets of reference codes: company codes at the highest company level (e.g. the consolidated company level); company codes for separate companies or divisions within the overall company; reporting unit codes (RU's are the units typically surveyed by government for statistical purposes and for multi-plant firms can contain information for a number of plants – or local units – located in more than one region); and local unit codes (which equate to a single location).
[12] The dataset also included the year that the company first received help from SE, based on a survey of firms undertaken for SE. This additional field only contains information on some 30% of all DRM companies with a linked IDBR code.

3.3 Thus, of the 2,311 DRM firms covered in section 2 (above), at most only 1,625 can (in theory) be linked to the IDBR and those databases associated with the IDBR (e.g. the ARD).

3.4 On checking the dataset deposited in the VML, we found 24 of the IDBR local-unit codes listed to be duplicates; thus there are potentially only 1,613 firms that can be linked into the ARD. Further, after linking the deposited SG 'look-up' table to the data on 2,311 DRM firms considered in section 2 (which was also – separately – deposited into the VML), it was also found that 291 of the 1,613 firms uniquely linked to the IDBR had no products associated with them (i.e. they either exited DRM status without receiving assistance, or were still waiting for assistance in 2008). Therefore, the final number of potential links with the ARD was 1,322.

3.5 In the next sub-section we shall try to locate these 1,322 firms in the ARD; however, of these, only 1,133 (or 86 %) had entered DRM status by 2005; the remainder entered into DRM status post-2005. Thus these are the firms that can be potentially linked and used in subsequent econometric analysis (since currently we only have data up to 2005 for the ARD).

Linking DRM into the ARD

3.6 The ARD comprises annual data from the Annual Business Inquiry (ABI), which from 1997 has been carried out for most market-based sectors of the UK economy.[13] It is based on information collected by the ABI (for a stratified sample of reporting units) which is then added to the non-surveyed reporting units, allowing weighting of the collected information to take place that can then produce nationally representative data for the UK. A discussion of the ARD (and its uses) is available in Oulton (1997), Harris (2002, 2005). The version of the ARD assembled in the VML by the present author comprises plant-level panel data covering 1997-2005 for all market-based sectors covered in the ARD.

3.7 In attempting to find the 1,322 DRM companies which can be potentially linked into the ARD, using the local-unit IDBR codes contained in both databases, it was possible to locate 969 (or 73%) of them when using ARD data covering 1997-2005. Thus, 353 firms with IDBR local unit codes were not found in the ARD. Presumably firms that could not be found are either:

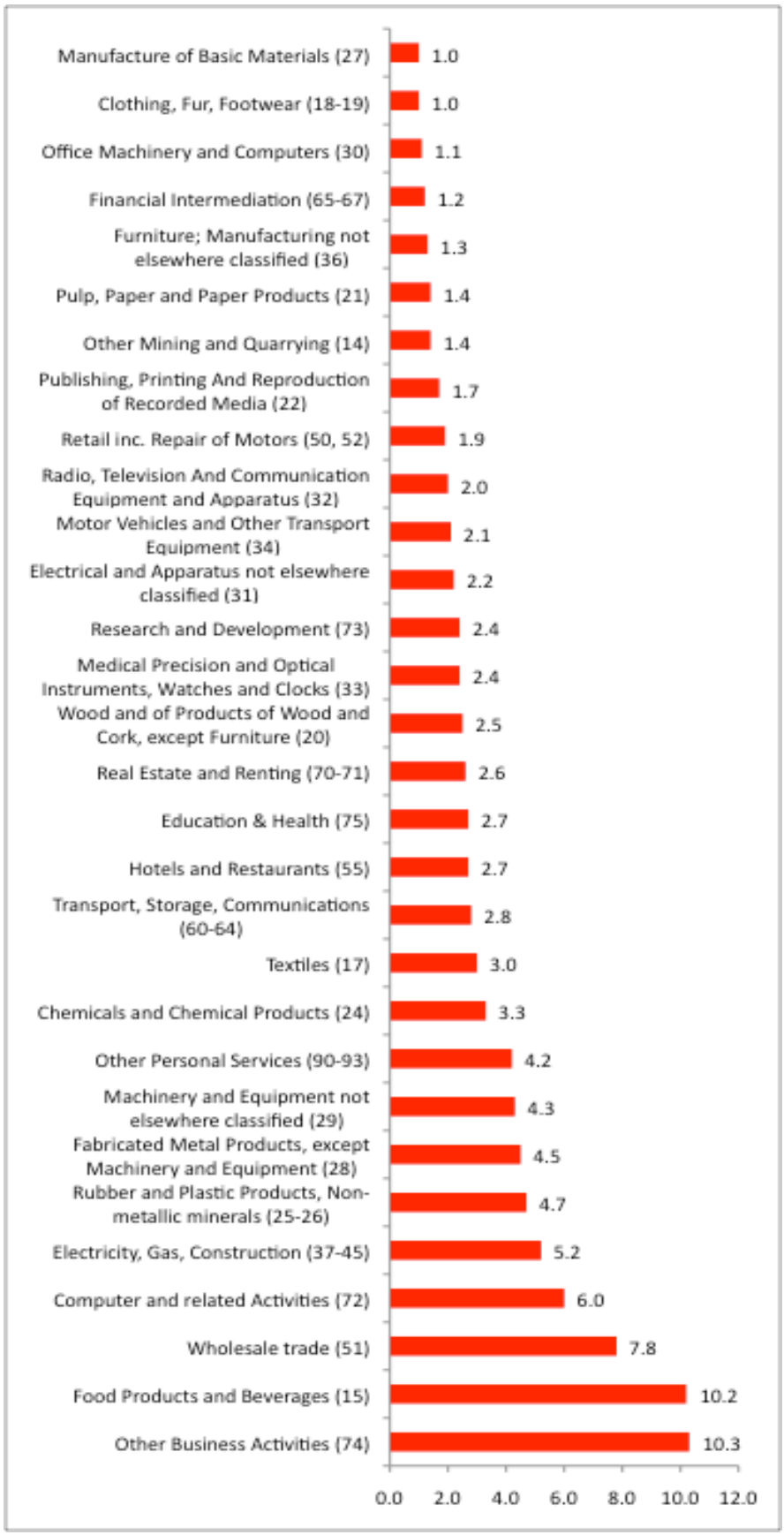
a) in sectors not covered by the ARD – see last footnote; or

b) they opened (and thus entered DRM status) after 2005.

3.8 We can confirm a) only if we have information on the Standard Industrial Classification (SIC) codes of the DRM firms (this data was not included in the dataset provided by SE); to check on b) would require us to investigate if the IDBR codes linked by SG statisticians appeared only after 2005. The latter cannot be done using the ARD (as it only covers the period up to 2005), but in

[13] Certain sectors are omitted (apart from non-market based sectors such as most of public administration and defence, and non-market based activities in health and education) such as agriculture and most of financial services (particularly banking and those areas regulated by the FSA). Note, manufacturing has been covered since 1970, and other industries in the production sector have been covered in various years prior to 1997 (Harris et. al., 2006, provide details).

Figure 3.1: DRM linked firms by industry (1992 SIC) - percentages



Source: linked DRM-ARD database

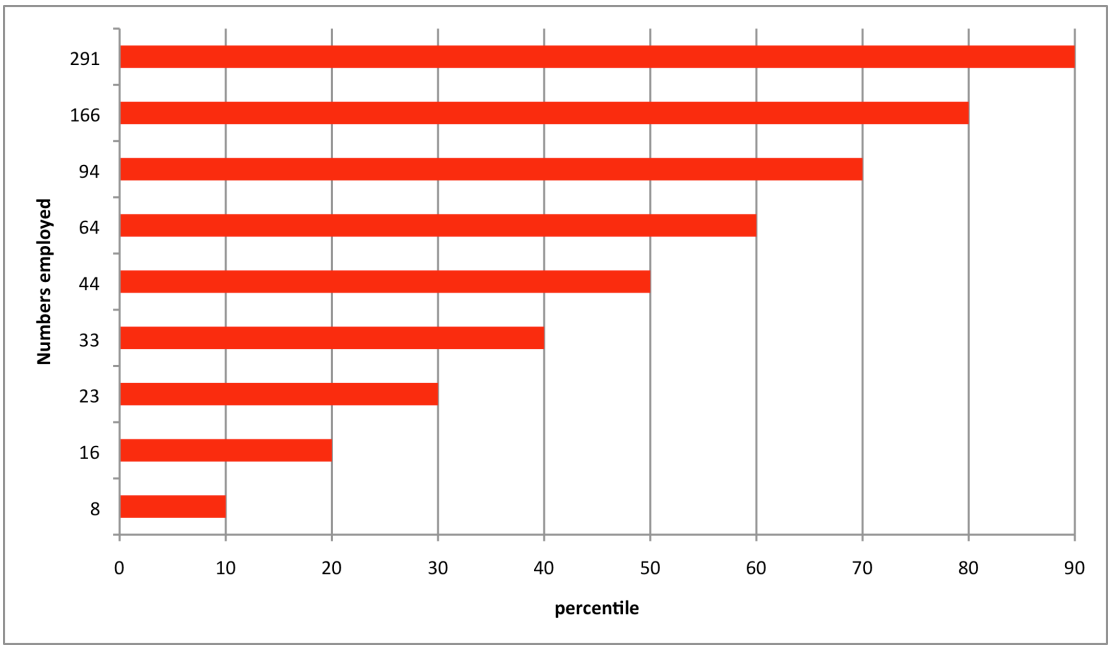
principle we could use the Business Structure Database available in the VML, which currently contains annual snap-shots of the IDBR up until 2007. This is a major exercise to undertake, and while it may confirm that either a) or b) are responsible for non-matching into the ARD, it will not result in any of the 353 non-matches being of further use in any subsequent analysis.

3.9 The rest of this sub-section will consider certain characteristics of the 969 DRM firms located in the ARD, in order to provide a check on the consistency of the matching process. Figure 3.1 shows how the 969 firms are distributed by industry. Other business activities (legal, accounting, advertising, etc.), followed by the manufacture of food & beverages, have the largest shares while manufacturing of basic materials, clothing etc., and office machinery & computers have the lowest shares. Note, these figures are based on the number of firms (not the employment or output they represent), and also note financial intermediation only covers those enterprises included in the ARD.

3.10 Of the 969 companies linked into the ARD, 417 (or 43%) were single-plant enterprises; thus these were operating on one site (and therefore by definition only in Scotland).

3.11 In terms of the employment size of the companies, firms in the top percentile employed 291 or more employees, while firms in the lowest decile employed eight or fewer. The median plant employed 44 workers (Figure 3.2).

Figure 3.2: DRM linked firms by employment size



Source: linked DRM-ARD database

3.12 Since the ARD is a panel dataset[14], information is available on when the firm was first observed (starting in 1997); Table 3.2 shows that of the 854 firms that entered DRM status in 2005, 513 were first observed in the ARD in 1997 (the first year for which we have data for the ARD). The other 341 firms were observed first (and thus began operating) between 1998 and 2005.

Table 3.2: DRM linked firms by when first observed in both ARD and DRM

	Year first entered DRM status		
Year first observed in ARD	2005	2006-2007	Total
1997	513	59	572
1998-2005	341	56	397
Total	854	115	969

Source: linked DRM-ARD database

3.13 This confirms that the majority of DRM firms have existed for some time before achieving DRM status; it also means that we have information on their performance for a number of years in advance of receiving assistance as DRM firms. These 854 firms comprise the set of linked DRM-ARD firms that can be considered in subsequent analyses, since for the other 115 linked firms in Table 3.2, while they are observed in the ARD, they did not receive DRM assistance until after 2005. Thus, in total this suggests that of the 1,133 firms that entered DRM status in 2005 that were potentially available for linking into the ARD (see para. 3.5), 75% (i.e. 854) can be used in further econometric analyses.

3.14 There are, however, two additional reasons for expecting that fewer than 854 firms can be subsequently used; firstly, 26 linked firms had supposedly closed by 2005 (as they are not observed in the ARD in that year). This suggests that the matching carried out by the SG statisticians did not link these 26 to the correct IDBR local unit code (since the firms should have been operating in 2005 to have obtained DRM status). It seems prudent therefore to leave out of any subsequent analysis those firms that were not operating in 2005 (according to the ARD).

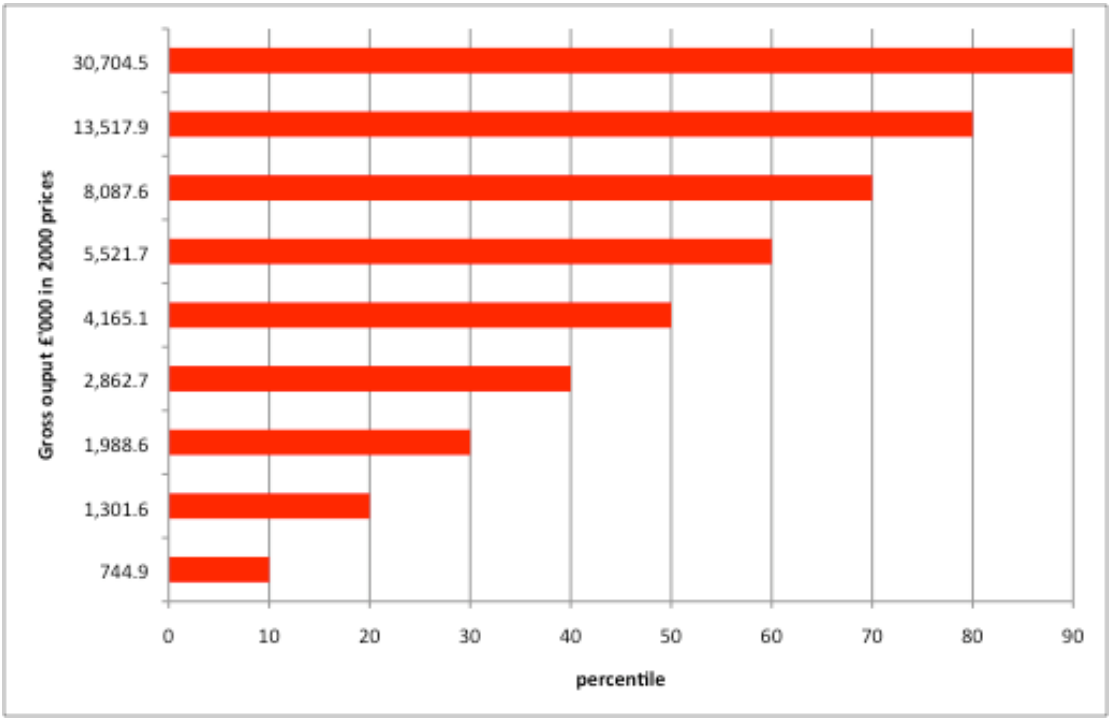
3.15 Some 14 of the 854 firms that achieved DRM status in 2005 state that they were first assisted *post-2005*. These firms should be left out of subsequent analysis if it is believed they have accurately recalled the date when they first were helped by SE; however, they should remain in the analysis if it is believed the respondent to the question provided inaccurate information. It is highly likely that the survey respondent has given the wrong data, so we believe these firms should remain in the dataset for subsequent analysis.

3.16 Finally, as financial information in the ARD is obtained from the ABI (which is based on a stratified sample of all firms), we will not have financial data (e.g. on

[14] That is plants can be tracked over time, so the dataset comprises a cross-section of plants with observations in different time periods.

turnover and gross value-added) for firms that were not surveyed. In fact, of the firms operating in 2005, 670 were surveyed and provided financial data. The other firms are in the ARD, but we only have information on their employment (plus other characteristics such as location, ownership, and age). Figure 3.3 shows the relative (real) gross output (which is almost equivalent to turnover) for these 670 firms, by deciles. Thus, the largest 10% of firms had gross output of £30.7m or greater (in 2000 prices), while the bottom decile comprised firms that only had gross output of £0.7m or less. Presumably achieving the goals set by SE with respect to DRM firms[15] should, other things being equal, be easier for the largest firms.

Figure 3.3: DRM linked firms by (real) gross output in 2005



Source: linked DRM-ARD database

Growth characteristics of matched DRM-ARD firms

3.17 As a final check of the DRM firms for which we have data on (real) gross output from the ARD, this sub-section compares the growth of these companies against the growth achieved by non-DRM firms in the ARD. Table 3.3 shows average (across firms) per annum output growth for 1997-2005 and 2001-2005 for the two sub-sets

[15] Note, SE define Account Managed businesses as those firms which have the potential to increase turnover, with assistance, by at least £800,000 over three years. Client Managed businesses are those firms which have the potential to increase turnover, with assistance, by at least £400,000 over three years (and these businesses receive less intensive support).

Table 3.3: Real growth p.a. of gross output in DRM and non-DRM firms in the matched DRM-ARD dataset (all market-based sectors covered)

	1997-2005	2001-2005
Non-DRM		
Mean	-1.7	-1.1
Median	-1.9	-1.2
DRM		
Mean	3.2	3.0
Median	1.6	2.7

Source: linked DRM-ARD database

3.18 Table 3.3 shows that DRM firms had a significantly higher growth rate than non-assisted firms, although this is based on a comparison of the average values for each sub-group (i.e. mean and median scores). It can be argued that comparing mean growth values is not as strong a test as considering whether the distribution of growth for one sub-group (e.g. DRM firms) dominates the distribution of a different sub-group (e.g. non-DRM firms).[16] Thus it is necessary to test the rank ordering of the growth distribution of firms that differ by whether they received assistance from SE or not.

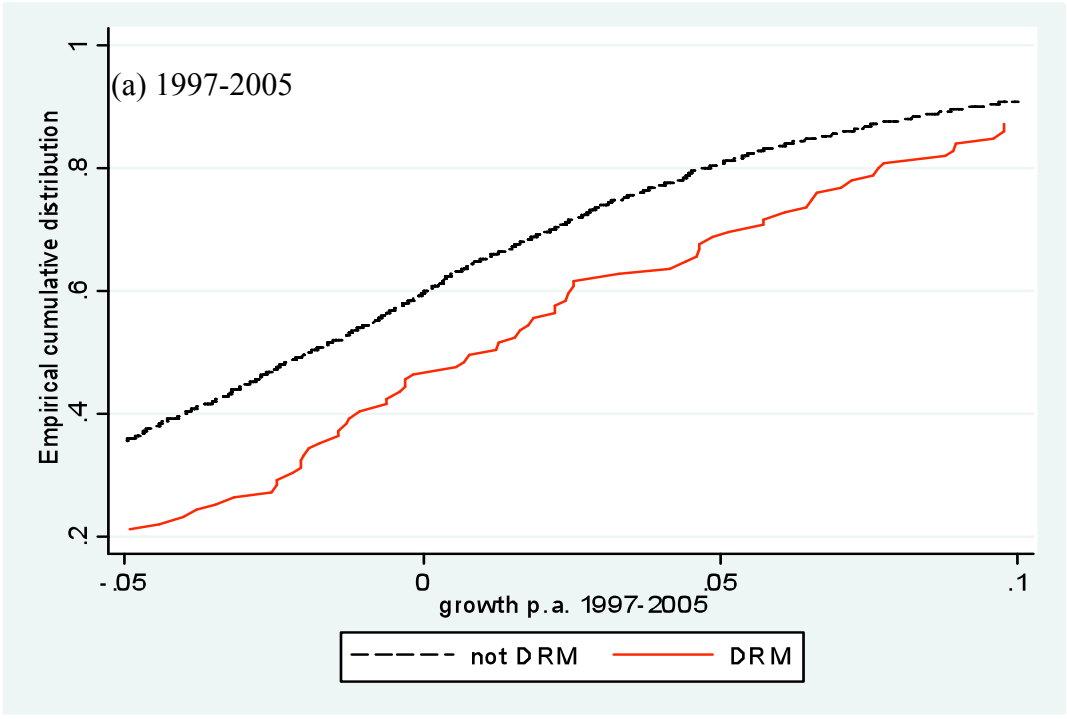
3.19 Figure 3.4 presents the (cumulative) distribution of growth rates across firms for (a) 1997-2005 and (b) 2001-2005, confirming that in both cases the distribution of growth rates for DRM firms were always to the right of the distribution of non-DRM firms (i.e. they had higher growth at all points in the distribution), both at the bottom and the top of the distribution of growth rates (although there is some evidence that growth rates converge for the highest growing firms in both sub-groups).

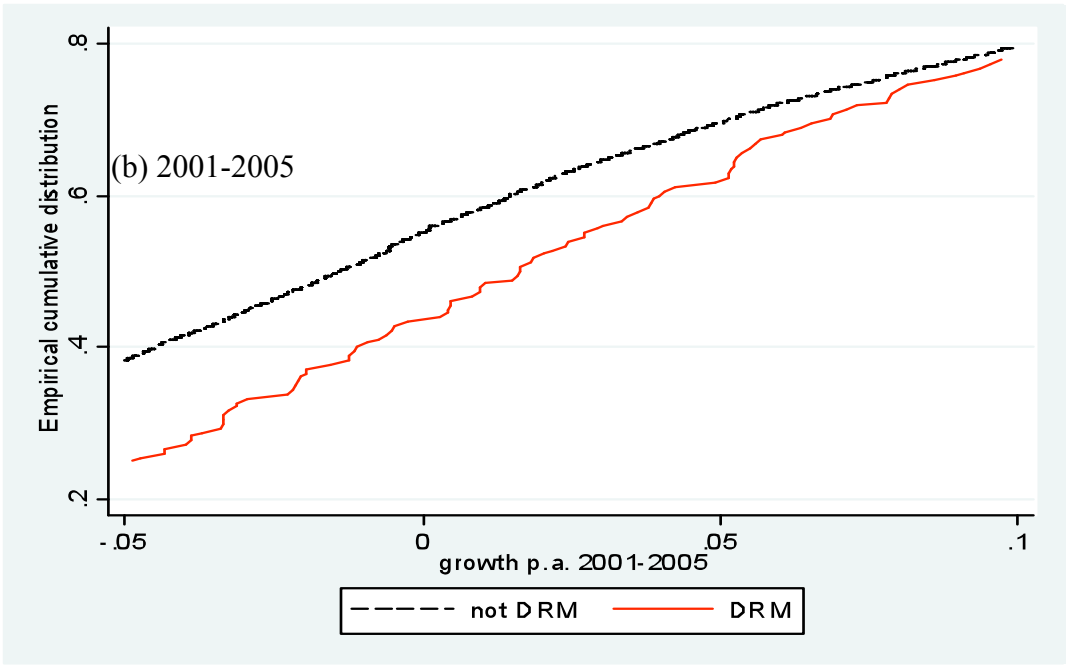
3.20 Calculating a two-sided Kolmogorov-Smirnov statistic, it is possible to *test* whether the growth distribution of one sub-group of firms lies to the right of another sub-group. If so, there is shown to be first-order stochastic dominance between such (random) variables, which is a stricter test than simply comparing average growth rates levels across sub-groups.

3.21 Table 3.4 presents the results obtained when applying the Kolmogorov-Smirnov test to our data on growth rates. Note, for groups tested we test the null hypothesis that the difference between the two distributions is favourable to one sub-group over the other, and thus being able to reject this null for one sub-group (e.g. non-DRM firms) would suggest that the other sub-group (e.g., DRM firms) have a distribution to the right of the rejected sub-group. Note, the values reported in Table 3.4 measure the greatest difference between the two sub-groups, and a positive value means that a sub-group lies to the *left* of the opposing sub-group (by definition of the way differences are calculated).

[16] That is, comparing means can be biased through outliers – only a single statistic (which is open to distortion if there are outliers in the data) is used; comparing the whole distribution uses more information and is therefore not likely to be biased by outliers.

Figure 3.4: Real gross output growth differences between DRM and non-DRM firms





Source: linked DRM-ARD database

Table 3.4: Two-Sample Kolmogorov-Smirnov Tests on the distribution of growth rates by DRM and non-DRM firms

	Difference favourable to:	
Period	Not-DRM firms	DRM firms
1997-2005	0.207***	-0.005
2001-2005	0.151***	-0.055

*** significantly different at the 1% or better significance level Source: linked DRM-ARD database

3.22 Thus we can confirm that in both periods examined firms that received DRM assistance from SE have a distribution that lies significantly to the right of non-DRM firms, and the difference between the two distributions was between 0.15 and 0.21. It is therefore confirmed that SE were assisting those firms in Scotland with the highest growth rates. However, whether this higher growth was due (at least in part) to such assistance, or whether these firms would have achieved such higher growth on their own can only be analysed using appropriate (matching) statistical techniques, that control for the characteristics of these firms. This is the subject of section 4 below.

3.23 One final issue with the linked data is that we do have some limited information collected from the database created by ekos/gen for a recent evaluation of the DRM programme[17] that indicates when some companies believed that they were first in receipt of assistance from SE.[18] From this source, we can identify 265 linked firms in the DRM-ARD database for which we have information on when they were first assisted.

3.24 Last, it is worth summarising the problems of data attrition experienced when first linking the DRM data into the IDBR; then into the ARD; and then taking account of the number of businesses about which we can be confident about knowing when they first started as DRM assisted. As noted above, Table 3.1 summarises this process.

Summary and conclusions

3.25 In section 2, it was noted that 1,742 firms had received DRM status by 2005, and thus these were the firms that needed to be linked to the ARD for

[17] GEN, Hayton Research and Research Resource (2009), Economic Impact Evaluation of Scottish Enterprise's Interventions with Account and Client Managed Companies

[18] Of the 969 linked firms in the DRM-ARD database, only 294 have information on when they were first assisted by SE based on the ekos/gen survey. Of these, 25% stated they first were assisted between 1970 and 1997; 68% were first assisted between 1998 and 2005; and 7% were first assisted after 2005. If those assisted after 2005 are omitted, and if we further exclude any of the 294 who (according to SE records) entered into DRM status after 2005, we are left with 265 linked firms in the DRM-ARD database who can provide information on when they were first assisted.

subsequent analyses.[19] However, only 1,133 DRM firms that had entered into DRM status by 2005 were uniquely allocated IDBR reference codes by SG statisticians. Of these 1,133 firms, some 75% (i.e. 854) were located in the ARD, and can be used in further econometric analyses (although some 26 of these firms could be excluded on the grounds that according to the ARD they had closed by 2005). If financial data is needed in any subsequent analyses, then information is available on 670 firms.

3.26 The main issue is whether the matching process has been sufficiently successful to warrant further econometric analysis. To consider whether DRM status has had any impact on firm performance, it is necessary to compare the performance of such assisted firms with a control group of matched non-assisted firms that, in all other respects, have similar characteristics as the assisted firms (in terms of such factors as size, age, industry, and spatial attributes).

3.27 Starting from the 1,742 firms that needed matching, some 1,133 (or 65%) were assigned IDBR local unit reference codes. (The fact that not all 1,133 firms could then be located in the ARD is less of an issue, as their absence is most likely due to their being allocated to industries not covered in the ARD and/or they opened after 2005 – see para. 3.6 above). However, since at least 45% of the relevant DRM firms were *not* found in the ARD, this presents problems for constructing a sample of matched firms in Scotland that were not DRM but which are a valid comparison group. It is more than likely that when constructing the matched sample, firms that did have DRM status, but which could not be identified in the ARD, would be erroneously allocated to the control group of firms against which the identified DRM firms are compared.

3.28 From this we conclude that any comparison of DRM firms should involve a comparison (i.e. control) group drawn from another region where the type of products offered to DRM firms in Scotland are not generally available. Based on our knowledge of the types of assistance available in the North East and North West regions of England, it is likely that firms operating in these regions are much more likely to provide an adequate control group. Therefore, it was agreed at the Interim Report stage that in order to proceed to stage 2 the control group should be based on firms with similar characteristics but which are located outwith Scotland where they were unlikely to receive the type of assistance that was available from SE[20].

3.29 There is one further problem. We have ARD data for many of the DRM firms for the period 1997-2005. However, according to the information provided by SE, the great majority of DRM firms only took on this status at the beginning of 2005, at which point SE had changed its approach to the programme. That

[19] The appendix to this section provides a table summarizing how we move from the original 2,311 firms in the Account and Client Managed database to the number of firms available for analysis in section 4.

[20] SQW undertook an analysis of the likelihood that firms in these regions might have benefited from assistance from organisations such as Regional Development Agencies or Business Link. The conclusion was that the possibility of any firms in the control group having received DRM-type assistance, and thus 'contaminating' the analysis, was very small. The analysis is contained in Appendix B.

being the case, since we have no ARD data on firm performance *beyond* 2005, few conclusions could in theory be drawn about the relationship between receiving DRM support and changes to business performance.

3.30 However, after discussion with SE, it became apparent that many of the businesses which appear to have had entered DRM in 2005 had in fact been DRM for some time before and had thus been receiving DRM-type products before 2005. Reference to the database created by ekos/gen for the recent evaluation of the DRM programme revealed pre-2005 DRM start data for **265** firms.

3.31 These have therefore formed the main group for the econometric analysis, as we have information on when they were first assisted by SE prior to 2005 and can look at their performance while they were DRM using the ARD data. Knowing when firms were first assisted is essential in order to benchmark their pre-DRM and post-DRM performance. If the date when assistance was first provided is unknown, it is not possible to accurately measure the performance of firms in relation to DRM over time. However, we do have this data for 265.

3.32 The 'treatment' group needs to be limited to these 265 matched firms for which we have the date when they were first assisted by SE. This comprises a rather small sub-group of all the original 1,742 DRM-assisted firms; and thus there is an issue of whether such a sample would be atypical in terms of their characteristics, and as a consequence whether any results obtained would be biased. Hence there is a need to compare the characteristics of the 265 sample of DRM firms chosen for econometric analysis with the full sample of linked DRM firms (and Scottish firms in general).

3.33 Lastly, there are some important lessons that have emerged while undertaking this first stage of the project, related specifically to how the DRM database can be linked to the IDBR and thus ARD. Certain information (such as Company House registration codes and/or VAT codes) if supplied to the SG statisticians would make the task of linking the data to the IDBR much simpler (and more accurate).

[this page is blank to allow double-sided printing  *Save Paper]*

4. Exploratory Analysis of linked DRM-ARD data

4.1 Despite issues of reliability with the data, as outlined in section 3, it was agreed that some exploratory analysis be undertaken to ensure that the study offered a full exploration of how SE could use data-matching as an evaluation technique. This involves econometric work related to measuring the performance of DRM firms against a control group of firms drawn from the North East and North West regions of England, to test if DRM status 'made a difference'. Thus the key questions addressed by the research are:

- how representative is the sample of DRM firms (to be used in econometric modelling of firm performance) of all DRM firms?
- is the full DRM sample linked into the ARD a representative sample of all Scottish firms (i.e., how does the sample differ in terms of such important characteristics as industry composition, age, size, R&D, capital-intensity, etc.)?
- after matching the DRM sample with a 'control' group of similar firms operating in the North East and North West of England, is there any evidence that DRM status led to any differences in firm performance?

4.2 This section begins with a comparison of the 265 DRM firms (which matched into the ARD and which have information on when they first received assistance prior to 2005) with those firms comprising a wider set of linked DRM-ARD data. Following this analysis, we discuss the econometric issues surrounding how to measure the relative performance of the 'treated' (DRM) group, taking account of what is referred to as potential 'sample selection bias'.

4.3 Finally, we estimate models comprising both the 265 treated DRM firms and matched control groups of firms with similar characteristics drawn from the tow English regions. Two regions were used to show how sensitive the procedure is with respect to matching, and thus the potentially different results that can be obtained from using different samples of 'treated' and 'control' groups. This highlights both the stringent requirements of the 'matching' approach used, juxtaposed against the limitations inherent in the DRM-ARD dataset that is available.

Representativeness of DRM sample

4.4 As set out in section 3, there are 854 linked firms in the DRM-ARD database (with linking having taken place at the plant and not firm level); of these 265 have information on when they first received assistance from SE prior to 2005. There are a further 172,000 plants in the Scottish ARD (operating in the 1997-2005 period) which were not linked to DRM status; of these, 25,000 (or 14.5%

of the ARD) were sampled by the ONS in order to provide information on financial performance.[21]

4.5 Table 4.1 sets out the variables available in the ARD that can be used when undertaking comparisons across plants (and for econometric analysis). Note, the dataset is a panel, so that variables refer to plant *i* in time *t*.

Table 4.1: Variable definitions used in ARD panel dataset for 1997-2005

Variable	Definitions
Real gross output	Plant level gross output data deflated by 2-digit ONS producer price (output) indices. Data are in £'000 (2000 prices)
Real intermediate inputs	Plant level intermediate inputs (gross output minus GVA) deflated by 2-digit ONS producer price (input) indices (non-manufacturing only has a single PPI). Data are in £'000 (2000 prices)
Employment	Number of employees in plant. (Labour productivity measured as real gross output divided by employment)
Capital	Plant & machinery capital stock (£m 1980 prices for manufacturing; £m 1995 prices for non-manufacturing) plus real value of plant and machinery hires (deflated by producer price index) in plant. Source: Harris and Drinkwater (2000, updated) and Harris (2005b).
Wage rate	Plant level total labour costs divided by employment (deflated by PPI when real values used)
Age	Age of plant in years based on year of entry (source: IDBR)
Single-plant	Dummy coded 1 when plant comprises a single-plant enterprise
US-owned	Dummy coded 1 if US-owned
EU-owned	Dummy coded 1 if EU-owned
SE-Asian-owned	Dummy coded 1 if owned by Malaysia, Taiwan, Japan, Singapore, HK
Other foreign-owned	Dummy coded 1 if other foreign-owned
Density	Population per hectare for local authority in which plant is located (source 2001 Census of Population)
R&D spending	Percentage undertaking intramural and/or extramural R&D
Industry agglomeration*	% of industry output (at 5-digit SIC level) located in local authority district in which plant is located – MAR-spillovers.
Diversification	% of 5-digit industries (from over 650) located in local authority district in which plant is located – Jacobian spillovers
Assisted Areas	Dummy variable = 1 if plant located in assisted area
Region	Dummy variable =1 if plant located in particular Government Office region
Industry	1992 SIC of plant (used at 2-digit level in most analyses).

* This is a particularly interesting factor which has been subject to some research. This is summarised in Appendix C.

4.6 We begin by considering whether the 265 DRM firms (with information on when they first obtained assistance from SE) that have been linked into the ARD are a representative sample of the 854 linked DRM-ARD dataset. Table 4.2 produces the mean values (across plants and years) for the 854 firms (first column of data) and the 265 firms (second column); the last column provides a univariate *t*-test of whether there is a significant difference between the means values for each variable.

[21] Recall, the ARD comprises the population of plants operating in the market-based economy (although it largely excludes some industries – such as agriculture and most of financial services – and it omits those companies with no employees). Of these only a (stratified) sample are required to provide financial information when they are included in the Annual Business Inquiry. Details are provided in Harris et. al. (2006).

4.7 There is only a statistically significant difference between the mean values (at around the 5% level or better) for:

- labour productivity (the DRM firms with dates for first receiving SE help were slightly less productive)
- the 18-34 employment size-band (DRM firms with dates were more represented in this sub-group)
- population density where the plant was located (DRM firms with dates operated in slightly lower density areas) and
- wholesale distribution (DRM firms with dates were less likely to belong to this industry).

4.8 However, the overall picture suggests the two sub-groups are very similar.

Table 4.2: Mean values[a,b] (1997-2005) of certain variables for DRM firms in DRM-ARD dataset

	No date	With date	*t*-test
ln AGE	2.154	2.195	-0.53
ln real labour productivity*	4.241	4.079	1.94
ln real capital-labour ratio*	-4.915	-5.253	1.28
l*n* real intermediate inputs-labour ratio*	3.681	3.522	1.60
ln real wage rate*	2.911	2.818	1.84
R&D spending	27.7	28.3	-0.19
Employed 1-17	22.4	18.1	1.56
Employed 18-34	17.5	24.2	-2.17[†]
Employed 35-65	19.5	19.3	0.09
Employed 66-168	20.2	20.4	-0.06
Employed 169+	20.4	18.1	0.77
Foreign-owned	15.6	17.0	-0.50
EU-owned	6.1	6.0	0.04
US-owned	6.3	8.7	-1.27
Other-FO	3.2	2.3	0.77
Single plant enterprise	42.1	43.4	-0.35
Assisted Area	65.5	63.8	0.50
ln Diversification	-0.462	-0.462	-0.02
ln Industry agglomeration	-0.187	-0.070	-1.17
ln Density	1.229	0.991	1.97[†]
Manufacture of Food Products and Beverages (15)	10.5	11.3	-0.34
Textiles, clothing, footwear (17-19)	4.2	4.5	-0.19
Wood & wood products (20)	2.4	3.8	-1.05
Manufacture of Pulp, Paper and Paper Products (21)	2.0	0.4	1.83
Publishing, Printing And Reproduction of Recorded Media (22)	1.4	1.9	-0.58
Manufacture of Chemicals and Chemical Products (24)	3.2	3.0	0.16
Manufacture of Rubber and Plastic Products (25)	4.1	3.8	0.21
Manufacture of other Non-Metallic Mineral Products (26)	0.9	1.1	-0.38
Manufacture of Basic Materials (27)	1.0	1.5	-0.57

	No date	With date	t-test
Manufacture of Fabricated Metal Products (28)	5.1	4.9	0.12
Manufacture of Machinery and Equipment n.e.c. (29)	4.1	4.9	-0.55
Manufacture of Office Machinery and Computers (30)	1.5	0.8	0.93
Manufacture of Electrical and Apparatus n.e.c. (31)	2.2	1.9	0.30
Manufacture of Radio, TV And Communication Equipment (32)	1.9	2.6	-0.73
Manufacture of Precision Instruments, Watches and Clocks (33)	2.0	3.0	-0.82
Manufacture of motors & other transport equipment (34-35)	1.9	2.6	-0.68
Manufacture of Furniture; Manufacturing n.e.c; recycling (36-37)	1.9	0.8	1.23
Construction (45)	5.6	3.8	1.13
Wholesale distribution (50-51)	9.9	4.9	2.42†
Retail distribution (52)	1.0	1.5	-0.62
Hotels & catering (55)	1.0	1.5	0.59
Land, Water & Air transport (60-62)	0.5	1.5	-1.24
Support of transport (63)	1.2	0.8	0.57
Financial intermediation (65-67)	1.5	0.4	1.45
Real Estate Activities (70)	1.7	2.6	-0.91
Computer and related Activities (72)	6.1	6.0	0.04
Research and Development (73)	2.0	3.8	-1.33
Other Business Activities (74)	9.2	10.6	-0.62
Recreational, Cultural and Sporting Activities (92)	3.1	1.1	1.68
Other Service Activities (93)	1.0	1.9	-1.04
Sample size N	589	265	
* Sample size N (financial variables)	419	195	

[a] All variables are in percentages except logged variables.
[b] The data includes only those with no missing values across all the variables in the table. Some firms have missing data (e.g., on capital stock information) and we lose them – in all about 6% of the data is lost hence the industry totals sum to <100%.
† Significant at 5% level (or better).

4.9 To confirm whether the sample of DRM firms (to be used in econometric modelling of firm performance) is representative of *all* DRM firms, regression analysis was undertaken. With the dependent variable classified as DRM firms with a date coded 1 (DRM firms with no date were coded 0), all the variables included in Table 4.2 were allowed to enter in a stepwise procedure in the probit model; the results of significant variables entering the model are presented in Table 4.3. Note two models are estimated: the first (presented in the top half of the table) included financial variables (thus cutting the sample size from 854 to 556[22]); the other omitted such information.

[22] This is less than the 614 firms overall that have financial information (Table 4.2) as some 58 firms had missing data for one or more of the financial variables.

Table 4.3: Probit regression of whether DRM firm has date of first assistance

Variable	$\partial\hat{p}/\partial x$	z-value	$\overline{X}$
Including financial variables			
ln real wage rate	-0.091	-2.46	2.878
Employed 18-34	0.126	2.05	0.160
Employed 169+	-0.140	-2.97	0.272
US-owned	0.222	2.68	0.085
Assisted Area	0.090	1.90	0.664
ln Industry agglomeration	0.049	2.80	0.092
ln Density	-0.049	-3.34	1.014
Manufacture of Pulp, Paper and Paper Products	-0.218	-2.20	0.020
Manufacture of motors & other transport equipment	0.260	1.94	0.027
Land, Water & Air transport	0.451	2.72	0.011
Research and Development	0.402	3.24	0.025
Other Business Activities	0.176	1.92	0.065
Other Service Activities	0.531	2.96	0.007
N	556		
Pseudo-R^2	0.08		
Excluding financial variables			
Employed 18-34	0.084	2.02	0.196
US-owned	0.110	1.59	0.070
ln Density	-0.018	-1.81	1.155
Manufacture of Pulp, Paper and Paper Products	-0.242	-3.55	0.015
Wholesale distribution	-0.146	-3.06	0.083
Financial intermediation	-0.228	-2.81	0.012
Recreational, Cultural and Sporting Activities	-0.165	-2.12	0.025
N	854		
Pseudo-R^2	0.03		

4.10 Table 4.3 (top half) shows that of all the potential financial variables included in Table 4.2, only the (logged) real wage rate is significant in determining whether there are differences between the DRM sub-groups. When financial variables (collected in the ABI) are included, the resulting sample of firms is known to be biased towards larger plants (see Harris, 2002). The data omitting these variables is, however, representative of the population and therefore the results presented in the lower part of Table 4.3 are more reliable. Firms that employed between 18-34 workers were 8.4% more likely to be in the DRM sample with a known date of first assistance from SE; they were also 11% more likely to be US-owned, 24% less likely to belong to the manufacture of pulp & paper industry, 15% less likely in wholesale distribution, 23% less likely in financial intermediation and 17% less likely to belong to recreational services.

4.11 While there are these (small) differences between the DRM sample of 265 firms with start dates with SE, and the whole sample of DRM-ARD firms, it still

remains the case that this sample (to be used in the econometric modelling of firm performance) does seem to be broadly representative of all DRM firms.[23]

4.12 Even when we limit the sample to just those 195 firms with financial information, there appear to be no significant differences that might suggest that this sample would result in a downward bias in the likely financial performance of DRM firms; none of the financial variables (such as productivity, R&D and the use of factor inputs) other than real wages were significant, and if anything, certain variables likely to be associated with stronger company performance (such as being US-owned, and being more likely to belong to the R&D and other business sectors) are more prevalent in the DRM sample of companies.

4.13 Lastly in this sub-section, we consider how far the full DRM sample linked into the ARD is a representative sample of all Scottish firms. Table 4.4 produces the mean values (across plants and years) for the 172,000 plants in the non-DRM Scottish ARD (first column of data) and the 854 DRM-ARD firms (second column); the last column provides a univariate *t*-test of whether there is a significant difference between the means values for each variable.

4.14 There is a statistically significant difference between the mean values in Table 4.4 (at around the 5% level or better) for nearly every variable; linked DRM firms are significantly older, and have higher:

- productivity
- capital- and intermediate-inputs intensity
- involvement in R&D
- employment size
- foreign-ownership.

4.15 They are less likely to be single-plant enterprises, and more likely to operate in assisted areas. The local authority areas they are located in are less likely to be diversified (although the difference is small), but they have higher levels of agglomeration and population density. There are also significantly different patterns across industries[24]

[23] Note overall, a very low pseudo-R^2 value for both models estimated in Table 4.3 confirms that there is little difference between the between the DRM sample of 265 firms with start dates with SE, and the whole DRM sample in the DRM-ARD database – based on the variables included in the table.
[24] Note the percentages across industries do not sum to 100 since certain industries are omitted; these include all of agriculture and mining, tobacco, coke & petroleum industries, utilities, post & telecoms, and renting of equipment. These sectors are omitted because the number of linked DRM companies is smaller than the number required by the ONS to ensure there is no potential risk of disclosure of information.

Table 4.4: Mean values[a] (1997-2005) of certain variables for DRM and non-DRM firms in linked ARD dataset

	ARD	DRM-ARD	*t*-test
ln AGE	1.524	2.166	-20.17†
ln real labour productivity*	4.041	4.190	-5.53†
ln real capital-labour ratio*	-5.248	-5.022	-6.06†
l*n* real intermediate inputs-labour ratio*	3.443	3.631	-4.32†
ln real wage rate*	2.568	2.881	-13.32†
R&D spending	0.6	27.9	-77.26†
Employed 1-17	87.8	21.1	44.14†
Employed 18-34	6.2	19.6	-11.28†
Employed 35-65	3.1	19.4	-11.97†
Employed 66-168	2.1	20.3	-14.17†
Employed 169+	0.9	19.7	-11.95†
Foreign-owned	2.4	16.0	-6.40†
EU-owned	1.2	6.1	-3.98†
US-owned	0.7	7.0	-4.56†
Other-FO	0.6	2.9	-1.73
Single plant enterprise	66.7	42.5	7.49†
Assisted Area	57.4	65.0	-4.59†
ln Diversification	-0.448	-0.462	2.71†
ln Industry agglomeration	-1.039	-0.151	-12.78†
ln Density	0.874	1.155	-4.35†
Manufacture of Food Products and Beverages (15)	0.7	10.8	-9.34†
Textiles, clothing, footwear (17-19)	0.4	4.3	-5.00†
Wood & wood products (20)	0.4	2.8	-3.86†
Manufacture of Pulp, Paper and Paper Products (21)	0.1	1.5	-3.28†
Publishing, Printing And Reproduction of Recorded Media (22)	0.8	1.5	-2.17†
Manufacture of Chemicals and Chemical Products (24)	0.1	3.2	-5.41†
Manufacture of Rubber and Plastic Products (25)	0.2	4.0	-5.33†
Manufacture of other Non-Metallic Mineral Products (26)	0.3	0.9	-2.42†
Manufacture of Basic Materials (27)	0.1	1.2	-2.00†
Manufacture of Fabricated Metal Products (28)	1.0	5.0	-5.21†
Manufacture of Machinery and Equipment n.e.c. (29)	0.5	4.3	-5.52†
Manufacture of Office Machinery and Computers (30)	0.1	1.3	-3.03†
Manufacture of Electrical and Apparatus n.e.c. (31)	0.2	2.1	-4.19†
Manufacture of Radio, TV And Communication Equipment (32)	0.1	2.1	-3.86†
Manufacture of Precision Instruments, Watches and Clocks (33)	0.2	2.3	-4.34†
Manufacture of motors & other transport equipment (34-35)	0.2	2.1	-4.38†
Manufacture of Furniture; Manufacturing n.e.c; recycling (36-37)	0.6	1.5	-2.98†
Construction (45)	8.7	5.0	5.43†
Wholesale distribution (50-51)	7.6	8.3	-2.57†
Retail distribution (52)	15.1	1.2	36.81†
Hotels & catering (55)	8.7	1.2	11.92†
Land, Water & Air transport (60-62)	2.3	0.8	4.33†
Support of transport (63)	1.2	1.1	-0.97
Financial intermediation (65-67)	2.5	1.2	3.43†

	ARD	DRM-ARD	t-test
Real Estate Activities (70)	3.9	2.00	5.14†
Computer and related Activities (72)	3.3	6.1	-3.58†
Research and Development (73)	0.2	2.6	-4.82†
Other Business Activities (74)	14.5	9.6	4.13†
Recreational, Cultural and Sporting Activities (92)	3.8	2.5	2.26†
Other Service Activities (93)	3.1	1.3	4.84†
Sample size N	172024	854	
* Sample size N (financial variables)	25160	614	

[a] all variables are in percentages except logged variables. † Significant at 5% level (or better).

4.16 We again confirm whether the DRM firms are representative of all Scottish market-based firms, using a probit regression model. With the dependent variable classified as a DRM firm being coded 1 (non-DRM firms were coded 0), all the variables included in Table 4.4 were allowed to enter the model using a stepwise procedure. The results of significant variables entering the model are presented in Table 4.5. As before, two models are estimated: the first (presented in the top half of the table) included financial variables (thus cutting the sample size from nearly 173,000 to just under 21,000); the other omitted these variables.

4.17 Restricting the model to include financial variables, the results confirm that DRM companies are older, have higher intermediate inputs, pay higher wages do more R&D and are larger. The coefficients presented are marginal effects (i.e., the probability of the plant being DRM when there is a change in the variable of interest) and they take low (although significant) values since only 0.5% of firms are DRM. Hence, plants that employ 169+ workers are nearly 8% more likely to be DRM, which is very large given the relatively small numbers of DRM firms in this employment size-band.

4.18 Interestingly, after controlling for the effect of the other covariates entering the model (especially size), being US-owned is less associated with DRM status while being a single-plant enterprise is more associated with being a DRM company.

4.19 The results in the top half of the table are biased towards the largest plants in the ARD; those in the lower half are not. However, they provide a similar picture of the significantly different characteristics of the DRM sub-group of firms (e.g. older, larger, more likely to do R&D, operate in assisted areas and where agglomeration and population density is higher).

Table 4.5: Probit regression of whether Scottish plant is a DRM firm, 1997-2005

Variable	$\partial\hat{p}/\partial x$	z-value	$\overline{X}$
Including financial variables			
ln AGE	0.0003	2.28	1.753
l*n* real intermediate inputs-labour ratio	0.0003	1.97	3.579
ln real wage rate	0.0005	1.98	2.548
R&D spending	0.0038	2.22	0.011
Employed 18-34	0.0085	3.90	0.133
Employed 35-65	0.0193	4.36	0.076
Employed 66-168	0.0375	5.19	0.060
Employed 169+	0.0784	5.46	0.033
US-owned	-0.0005	-2.00	0.052
Single plant enterprise	0.0035	3.32	0.131
ln Diversification	-0.0026	-3.32	-0.431
Manufacture of Food Products and Beverages (15)	0.0067	2.51	0.017
Manufacture of other Non-Metallic Mineral Products (26)	0.0029	0.81	0.003
Retail distribution (52)	-0.0021	-4.36	0.347
Hotels & catering (55)	-0.0005	-1.88	0.106
Land, Water & Air transport (60-62)	-0.0008	-3.39	0.020
N	20981		
Pseudo-R^2	0.41		
Excluding financial variables			
ln AGE	0.0001	1.68	1.528
R&D spending	0.0117	6.01	0.007
Employed 18-34	0.0143	10.52	0.063
Employed 35-65	0.0264	9.84	0.032
Employed 66-168	0.0391	9.67	0.021
Employed 169+	0.0652	8.30	0.009
Single plant enterprise	0.0008	7.29	0.666
Assisted Area	0.0002	1.97	0.575
ln Diversification	-0.0024	-6.49	-0.448
ln Industry agglomeration	0.0002	4.54	-1.036
ln Density	0.0001	1.84	0.875
Manufacture of Food Products and Beverages (15)	0.0390	6.23	0.007
Textiles, clothing, footwear (17-19)	0.0230	3.75	0.004
Wood & wood products (20)	0.0279	3.62	0.004
Manufacture of Pulp, Paper and Paper Products (21)	0.0257	2.28	0.001
Publishing, Printing And Reproduction of Recorded Media (22)	0.0109	2.88	0.008
Manufacture of Chemicals and Chemical Products (24)	0.0519	3.77	0.002
Manufacture of Rubber and Plastic Products (25)	0.0508	4.15	0.002
Manufacture of other Non-Metallic Mineral Products (26)	0.0146	2.18	0.003
Manufacture of Basic Materials (27)	0.0229	1.66	0.001
Manufacture of Fabricated Metal Products (28)	0.0189	4.61	0.010
Manufacture of Machinery and Equipment n.e.c. (29)	0.0260	4.28	0.005
Manufacture of Office Machinery and Computers (30)	0.0481	2.22	0.001
Manufacture of Electrical and Apparatus n.e.c. (31)	0.0384	3.21	0.002
Manufacture of Radio, TV And Communication Equipment (32)	0.0474	2.75	0.001
Manufacture of Precision Instruments, Watches and Clocks (33)	0.0242	3.00	0.002
Manufacture of motors & other transport equipment (34-35)	0.0454	3.53	0.002

Variable	$\partial\hat{p}/\partial x$	z-value	$\overline{X}$
Manufacture of Furniture; Manufacturing n.e.c; recycling (36-37)	0.0202	3.37	0.006
Construction (45)	0.0024	3.71	0.087
Wholesale distribution (50-51)	0.0083	6.58	0.077
Land, Water & Air transport (60-62)	0.0075	2.85	0.011
Support of transport (63)	0.0015	1.68	0.025
Financial intermediation (65-67)	0.0038	2.86	0.039
Real Estate Activities (70)	0.0180	6.05	0.034
Computer and related Activities (72)	0.0330	3.56	0.002
Research and Development (73)	0.0040	6.00	0.145
Other Business Activities (74)	0.0042	3.46	0.038
Recreational, Cultural and Sporting Activities (92)	0.0043	2.69	0.030
N	172858		
Pseudo-R^2	0.34		

4.19 These results are to be expected as DRM firms are the high growth enterprises obtaining assistance from SE; but such higher growth characteristics might also suggest that *even without assistance* they would likely do better than other non-assisted firms, which means that any assessment of the impact of DRM status must 'control' for such potential 'sample selection' effects.

4.20 A detailed discussion of sample selection issues is provided in Appendix A. Given the availability of the different approaches to the selectivity problem and the data available to us, we have chosen to test for the relationship between DRM status and *total factor productivity* (TFP) using a matching approach (based on the propensity scores obtained from the probability of achieving DRM status as set out in equations A4.4 and A4.5 in Appendix A).

4.21 TFP is the preferred variable.[25] It is measured as the level of output that is *not* attributable to factor inputs (employment, intermediate inputs and capital). Rather, TFP measures the contribution to output of all other influences, capturing such determinants as technological progress and/or changes in efficiency (where the latter also captures the under-utilising of factor inputs unless this is taken into account when measuring these inputs). [26]

Modelling the performance of the DRM sample

4.22 Initially the North East of England was chosen as the region from which we would obtain the 'control' group of matched plants. The ARD covers all regions of the UK, so all the variables considered above (see. Table 4.1) were available for the analysis. The 195 Scottish DRM firms with financial data (see Table 4.2) and the ARD data for the North East were merged, with Table 4.6 presenting the outcome of estimating equation (A4.4) using a stepwise probit model.

[25] Appendix D provides a more general discussion of the measurement of TFP

[26] Thus, TFP is equivalent to a combination of the residual εit from (4.1) and the time trend, t, which represents technological change. Harris (2005) provides a detailed explanation of how this approach is preferable to other estimators of productivity.

4.23 Larger, single-plant enterprise firms operating in relatively more diversified and agglomerated areas, with lower levels of population density, in particular industries, were the attributes that defined the likelihood of achieving DRM status. Thus, firms in the North East with similar (overlapping) propensity scores were chosen as the control group (using equation A4.5).

Table 4.6: North East probit regression of whether plant is a DRM firm, 1997-2005*

Variable	$\hat{\beta}$	*z*-value
ln real gross output	0.094***	3.07
Single plant enterprise	0.419***	3.82
ln Diversification	2.841***	7.39
ln Industry agglomeration	0.123***	2.81
ln Density	-0.415***	-12.65
Manufacture of Electrical and Apparatus n.e.c. (31)	0.791***	2.32
Manufacture of Radio, TV And Communication Equipment (32)	0.861***	2.39
Manufacture of Precision Instruments, Watches and Clocks (33)	0.761***	4.19
Manufacture of motors & other transport equipment (34-35)	0.612***	2.52
Construction (45)	0.036	0.23
Retail distribution (52)	-0.340	-1.06
Land, Water & Air transport (60-62)	0.625***	3.58
Computer and related Activities (72)	1.325***	5.14
Other Business Activities (74)	-0.234	-0.67
Constant	-1.488***	-4.29
N	78444	
Psuedo-R^2	0.34	

* matched sample of Scottish DRM (with dates) and NE control group

***/**/* significant at the 1%/5%/10% level

4.24 We then used the matched data comprising Scottish DRM firms (with start dates of when they first received assistance) in a model of the determinants of TFP. This is based on estimating a dynamic-form of the Cobb-Douglas production function using (unbalanced) panel-data for the 1997-2005 period:

$$\ln Y_{it} = \beta_0 + \sum_{j=1}^{4} \pi_{1j} x_{jit} + \sum_{j=1}^{4} \pi_{2j} x_{ji,t-1} + \pi_3 \ln Y_{i,t-1} + \phi_x X_{it} + \sum_{s=-4}^{6} \gamma_s D_{i,t-s} + \eta_i + t_t + (1-\rho) e_{it} \quad (4.1)$$

where the subscripts *i* and *t* represent the *i*-th plant and the *t*-th year of observation, respectively

Y represents real gross output (in £'000 2000 prices)

x_1 represents the logarithm of intermediate inputs (in £'000 2000 prices), *m*

x_2 represents the logarithm of tangible assets (in £'000 2000 prices), *k*

x_3 represents the logarithm of total employment, *e*

x_4 represents a time trend to take account of technical progress, *t*

X is a vector of variables determining TFP (comprising most of the other variables in Table 4.1), and includes industry dummies

D_t is a dummy variable taking on a value of 1 in the year when the plant first entered DRM status;[27] and

the composite error term has three elements with the fixed-effect term η_i affecting all observations for the cross-section plant *i*; t_t affects all plants for time period *t*; and e_{it} affects only plant *i* during period *t*.[28]

4.25 To allow for potential endogeneity of factor inputs and output, equation (4.1) was estimated using the Generalised Method of Moments (GMM) systems approach available in STATA 9.2 (Arellano and Bond, 1998). This is sufficiently flexible to allow for both endogenous regressors (through the use of appropriate instruments involving lagged values – in levels and first differences – of the potentially endogenous variables in the model[29]) and a first-order autoregressive error term.[30]

4.26 Table 4.7 presents the results from estimating equation (4.1) using matched data for the 195 Scottish DRM firms and the control group from the North East.[31] The key result of interest is the time profile of TFP associated with a firm entering DRM status at time *t*. This is tracked for the whole period before and after *t* (with the parameter estimate for D_{t-4} covering the period 4+ years before being assisted, while D_{t+6} covers the period 6+ years after being assisted).

4.27 The results show that that there is some (weak) evidence that before being assisted plants had higher TFP; in the year when assistance started, productivity significantly declined by around 9%[32] with further (insignificant) falls until years *t*+5 and *t*+6. The latter summarises the whole post-assistance period average effect from year *t*+6, showing that from the 6th year onwards there was

[27] Note, D_t enters contemporaneously and with leading and lagged terms, to consider the time profile of productivity for DRM firms leading up to and post being first assisted. The longest lead and lag are set to capture all other time periods covered in the dataset. Missing values for these terms are assigned a value of 0 (given the unbalanced nature of the panel dataset).

[28] Note, if e_{it} is serially correlated such that $e_{it} = \rho e_{it-1} + u_{it}$ then u_{it} is uncorrelated with any other part of the model, and $|\rho|<1$ ensures the model converges to a long-run equilibrium (i.e. the variables in the model are cointegrated).

[29] As well as output, intermediate inputs, labour and capital, whether the plant undertook R&D was also treated as endogenous.

[30] Using the GMM systems approach the model is estimated in both levels and first-differences. This is important, since Blundell and Bond (1999) argue that including both lagged levels and lagged first-differenced instruments leads to significant reductions in finite sample bias as a result of exploiting the additional moment conditions inherent from taking their system approach.

[31] Note, the model estimated passes diagnostic tests for autocorrelation and the Hansen test that the over-identifying restrictions are valid.

[32] Since the dependent variable is logged, the marginal effect is exp ($\hat{\beta}$)-1.

a significant overall decline in TFP of 15.6% in comparison with the control group.[33]

Table 4.7: Systems GMM production function, matched DRM-NE England data,1997-2005[a] (equation 4.9)

	$\hat{\beta}$	*z*-value
ln real gross output$_{t-1}$	0.684***	9.65
ln employment$_t$	0.170***	3.14
ln employment$_{t-1}$	-0.107*	-1.75
ln capital$_t$	0.085**	2.80
ln capital$_{t-1}$	-0.052	-1.52
ln intermediate inputs$_t$	0.600***	6.49
ln intermediate inputs$_{t-1}$	-0.348**	-2.51
ln Industry agglomeration$_t$	0.010	0.30
ln Diversification$_t$	0.174	1.10
ln AGE$_t$	0.023	0.50
Single plant enterprise$_t$	-0.009	-0.20
EU-owned$_t$	-0.044	-1.06
US-owned$_t$	-0.018	-0.50
Other foreign-owned$_t$	0.037	0.38
R&D spending$_t$	-0.014	-0.28
Assisted Area$_t$	-0.031	-1.07
t	0.000	-0.09
D_{t-4}	0.170	1.58
D_{t-3}	0.072	0.67
D_{t-2}	-0.019	-0.29
D_{t-1}	0.072	1.18
D_t	-0.095**	-1.97
D_{t+1}	-0.037	-0.37
D_{t+2}	0.069	1.12
D_{t+3}	-0.093	-1.35
D_{t+4}	-0.041	-0.59
D_{t+5}	-0.111	-1.89
D_{t+6}	-0.170***	-2.68
Constant	0.544*	1.79
Industry dummies	yes	
AR(1) z-statistic	-2.79***	0.01
AR(2) z-statistic	1.67	0.10
Hansen test $\chi^2(90)$	104.33	0.14
No. of Obs.	822	
No. of groups	320	

[a] Note the 2-step GMM system estimator in STATA9.2 is used (i.e. "xtabond2")
***/**/* significant at the 1%/5%/10% level

[33] Note, these are the short-run impacts; in the long-run (when equilibrium is achieved) the impact of assistance is measured by dividing the parameter estimates by (1– *ln* real gross output$_{t-1}$). Thus the long-run impact of assistance in years *t*, *t*+5 and from *t*+6 onwards are -25.9%, -29.7% and -41.6%, respectively. All parameters are significant at the 10% level (or better).

4.28 As to the other parameter estimates reported in Table 4.7, the long-run elasticities of output with respect to labour, capital, and intermediate inputs are 0.2, 0.1 and 0.8, respectively (and all significant at the 1% level). Summing these three terms results in a value greater than 1, indicating technology operated with increasing returns-to-scale. All of the other parameter estimates in the model were insignificant.

4.29 Turning to the results using linked Scottish DRM and North West of England data, Table 4.8 presents the outcome of estimating equation (A4.4) using a stepwise probit model. As with the results based on using the North East as the control group, larger, single-plant enterprise firms operating in relatively more diversified and agglomerated areas, with lower levels of population density, in particular industries, were the attributes that defined the likelihood of achieving DRM status. Thus, firms in the North West with similar (overlapping) propensity scores were chosen as the control group (using equation A4.5).

Table 4.8: North West probit regression of whether plant is a DRM firm, 1997-2005*

Variable	$\hat{\beta}$	*z*-value
ln real gross output	0.104***	4.17
Single plant enterprise	0.418***	4.47
ln Diversification	1.347***	3.55
ln Industry agglomeration	0.145***	3.90
ln Density	-0.372***	-10.23
Manufacture of Radio, TV And Communication Equipment (32)	0.954***	3.28
Manufacture of Precision Instruments, Watches and Clocks (33)	0.610*	1.85
Manufacture of motors & other transport equipment (34-35)	0.899***	5.82
Manufacture of Furniture; Manufacturing n.e.c; recycling (36-37)	0.638***	3.17
Wholesale distribution (50-51)	-0.054	-0.39
Hotels & catering (55)	-0.390	-1.41
Support of transport (63)	0.475***	3.14
Research and Development (73)	1.251	5.92
Recreational, Cultural and Sporting Activities (92)	0.107	-0.35
Constant	-2.930***	-9.77
N	212672	
Psuedo-R^2	0.26	

* matched sample of Scottish DRM (with dates) and NW control group

***/**/* significant at the 1%/5%/10% level

4.30 Table 4.9 presents the results from estimating equation (4.1) using matched data for the 195 Scottish DRM firms and the control group from the North West.[34] Again, the key result of interest is the time profile of TFP associated with a firm entering DRM status at time *t*; the results show that when compared to the North West control group there is some (weak) evidence that several years before being assisted plants had higher productivity and then productivity fell in

[34] Note, the model estimated passes diagnostic tests for autocorrelation and the Hansen test that the over-identifying restrictions are valid (although weakly for the Hansen test).

both *t*–2 and the year when assistance started (with declines of around 30% and 15%, respectively). After assistance there were further (insignificant) falls until years *t*+5 and *t*+6. The latter summarises the whole post-assistance period average effect from year *t*+6, showing that from the 6th year onwards there was a significant overall decline in TFP of 33% in comparison with the control group.[35]

Table 4.9: Systems GMM production function, matched DRM-NW England data,1997-2005[a] (equation 4.9)

	$\hat{\beta}$	*z*-value
ln real gross output$_{t-1}$	0.307**	2.35
ln employment$_t$	0.151**	1.97
ln employment$_{t-1}$	-0.071	-1.01
ln capital$_t$	0.067**	2.04
ln capital$_{t-1}$	-0.017	-1.26
ln intermediate inputs$_t$	0.802***	10.15
ln intermediate inputs$_{t-1}$	-0.213*	-1.90
ln Industry agglomeration$_t$	0.050	0.87
ln Diversification$_t$	0.872	1.47
ln AGE$_t$	0.042	0.54
Single plant enterprise$_t$	0.054	0.61
EU-owned$_t$	-0.214**	-2.29
US-owned$_t$	0.117*	1.93
Other foreign-owned$_t$	-0.237***	-3.52
R&D spending$_t$	-0.035	-0.36
Assisted Area$_t$	-0.060	-1.20
t	0.011	0.89
D_{t-4}	–	–
D_{t-3}	0.175	1.04
D_{t-2}	-0.314***	-2.78
D_{t-1}	-0.058	-0.42
D_t	-0.167*	-1.71
D_{t+1}	-0.166	-1.09
D_{t+2}	-0.144	-1.28
D_{t+3}	-0.184	-1.41
D_{t+4}	-0.183	-1.42
D_{t+5}	-0.267*	-1.83
D_{t+6}	-0.402**	-2.22
Constant	1.693**	1.96
Industry dummies	yes	
AR(1) z-statistic	-2.15***	0.01
AR(2) z-statistic	1.03	0.30
Hansen test $\chi^2(90)$	108.01	0.09
No. of Obs.	805	
No. of groups	317	

[a] Note the 2-step GMM system estimator in STATA9.2 is used (i.e. "xtabond2")
***/**/* significant at the 1%/5%/10% level

[35] The long-run impact of assistance in years *t*–2, *t*, *t*+5 and from *t*+6 onwards are -36.5%, -21.4%, -32.0% and -44.0%, respectively. All parameters are significant at the 5% level (or better).

4.31 As to the other parameter estimates reported in Table 4.9, the long-run elasticities of output with respect to labour, capital, and intermediate inputs are 0.11, 0.07 and 0.85, respectively (and all significant at the 5% level). Summing these three terms results in a value greater than 1, indicating technology operated with increasing returns-to-scale. Other significant parameter estimates in the model were obtained for those plants that were foreign-owned; in the long-run EU-, US- and other foreign-owned plants had significantly different TFP of -27%, 18% and -29%, respectively.[36]

4.32 These results when using the North West as the control group are similar, but not the same, to those obtained when using data drawn from the North East. The results are different because the control groups are different; i.e. based on regions of different sizes, industrial compositions, etc. *and* reflecting the fact that the matching process is dependent on the quality of the data available from which to obtain matched plants sharing very similar characteristics (except that one sub-group – the Scottish DRM firms – receive 'treatment').

Summary and conclusions

4.33 This section reports on some *exploratory* work that is designed to measure the performance of those (small numbers of) DRM firms for which we have data on their first year of assistance from SE. Because of potential contamination of the non-DRM Scottish data identified in the ARD (resulting from only being able to link around 55% of the known DRM companies into the ARD), it was agreed to measure the performance of the DRM firms against a control group of firms from the North East and North West regions of England. The use of two comparator regions reflects the known sensitivity of the 'matching' approach to the quality of the data available that is used to match firms with (as far as possible) very similar characteristics.

4.34 The first task before undertaking matching was to test the representativeness of the DRM sample of 265 Scottish firms for which there is information on when they first received assistance from SE. When tested against the linked DRM-ARD dataset comprising 854 firms, to see if the sample of 265 is representative of the larger sub-group, the overall result suggests that the two sub-groups are very similar in composition. Thus, there is reason to believe that using the sample (where start dates are available) should not bias any subsequent econometric testing of whether DRM status had any impact on the financial performance of DRM firms.

4.35 We also test if the full DRM sample (of 854 firms) was representative of all Scottish market-sector firms operating across 1997-2005. They clearly are not, tending to be older, have higher productivity, capital and intermediate inputs intensity, involvement in R&D, employment size, and foreign-ownership. This is to be expected, as DRM firms are the high growth enterprises supported by SE. However, such higher growth characteristics might also suggest that even

[36] The better performance of US-owned but not other foreign-owned is in line with the evidence presented in Harris and Robinson (2003).

without assistance they would likely do better than other non-assisted firms, which means that any assessment of the impact of DRM status must control for such potential sample selection effects.

4.36 Thus, sample selection techniques were discussed, with a particular emphasis on the 'matching' approach to overcoming potential sample selection bias. Essentially, under the matching assumption DRM- and non-DRM firms have the same (observable) attributes that impact on productivity except that one sub-group receives assistance and the other does not; put another way, *the outcome that would result in the absence of SE assistance is the same in both cases*. Thus the non-DRM matched sub-group constitutes the correct counterfactual for the missing information on the outcomes that DRM firms would have experienced, on average, if they had not received help.

4.37 There are a number of issues with this matching process, including the need for a rich dataset set that includes all relevant variables that impact on productivity and all variables that impact on whether the firm receives assistance or not. Matching is done on these sets of variables, so that any selection on unobservables is assumed to be trivial and does not affect outcomes in the absence of assistance. Of course, if the data available is inadequate for obtaining 'good' matches (i.e., if we lack information on the key determinants of which firms achieve DRM status), then matching will not be able to eliminate the full impact of any potential sample selection bias.

4.38 Here, we have adopted the propensity score matching approach where we first estimate a model to identify the probability of a firm achieving DRM-status (i.e. the propensity score) using a probit model, and based on similar propensity scores we then chose the 'control' sub-group of firms that were used alongside the Scottish DRM firms.

4.39 The propensity score models used with the North East and the North West of England as the comparator groups gave similar results, suggesting that we were obtaining a consistent set of matched plants from each region. Using this matched data, we then estimated production functions (first using Scottish DRM and North East control group firms, and then substituting North West firms as the control group), in order to test whether achieving DRM status in time t had any impact on productivity either before time t and/or after obtaining assistance.

4.40 Using North East England control group data, the results showed that that there is some (weak) evidence that before being assisted plants had higher productivity; in the year when assistance started productivity significantly declined by around 9% with further (insignificant) falls until years t+5 and t+6. The latter summarises the whole post-assistance period average effect from year t+6, showing that from the 6th year onwards there was a significant overall decline in TFP of 15.6% in comparison with the control group.

4.41 Using matched data for the 195 Scottish DRM firms and the control group from the North West, again the results show that there is some (weak) evidence that several years before being assisted plants had higher productivity and then productivity fell in both t–2 and the year when assistance started (with declines of around 27% and 15%, respectively). After assistance there were further

(insignificant) falls until years *t*+5 and *t*+6. The latter summarises the whole post-assistance period average effect from year *t*+6, showing that from the 6th year onwards there was a significant overall decline in TFP of 33% in comparison with the control group.

4.42 The results when using the North West as the control group are similar, but not the same, to those obtained when using data drawn from the North East of England. The results are different because the control groups are different; i.e., based on regions of different sizes, industrial compositions, etc., *and* reflecting the fact that the matching process is dependent on the quality of the data available from which to obtain matched plants sharing very similar characteristics (except that one sub-group – the Scottish firms – receive 'treatment').

4.43 Any interpretation of these data should take account of two important features of the analysis:

a. it relates to forms of intervention delivered by SE between five and ten years ago

b. it was essentially an exploratory exercise; a more full analysis covering later years, with an improved level of linking between DRM data and ARD data, might produce different results.

5. Future Steps

5.1 This study has demonstrated the extent to which it has been possible to merge data on DRM firms into the ARD for use in econometric analysis that sets out to determine whether assisted firms have a different productivity profile to non-assisted firms.

5.2 This is an important approach as it allows a comparison of performance of assisted firms taking into account the counter-factual position (i.e., comparing performance with what would have happened if there had not been any assistance from SE, and thus whether such assistance made a difference). Simply comparing the performance of assisted firms with non-assisted firms is open to what is termed sample-selectivity bias – i.e., that such firms have characteristics that give them a higher growth potential that they would realise irrespective of whether assistance is available or not. Thus, any differences between DRM- and non-DRM firms are likely to be a biased indicator of whether SE help really did have an impact, unless we control for sample-selectivity by comparing the treatment group (those receiving assistance) with a control group (i.e., firms with similar characteristics to the treated group but with the defining difference that they not receive assistance).

5.3 The results from this study can be considered both in terms of what is needed to undertake this type of analysis (and therefore the lessons to be learned), and also whether the results obtained are indicative of the impact that SE has had via its DRM approach to helping firms. With regards to the latter, this study has had to consider various data issues that ultimately lead to a rather small sample being available for econometric analysis (see Table 3.4).

5.4 However, the testing carried out in section 4 suggests that the sub-group of firms, with information on when they were first assisted by SE, are a representative sample of all DRM firms and therefore subsequent statistical analysis of their performance is providing SE with useful information on whether DRM status made a difference. The results suggest that we can be (better than 95%) confident that when compared to the control groups of firms in the North East and North West of England (who are similar but did not receive DRM-type assistance), Scottish DRM firms certainly did not do better and in fact the statistical evidence points to them having done worse in terms of total factor productivity, the key variable used in the analysis.

5.5 That said, interpretation of these data should take account of two important features of the analysis:

a. it relates to forms of intervention delivered by SE between five and ten years ago

b. it was essentially an exploratory exercise; a more full analysis covering later years, with an improved level of linking between DRM data and ARD data, might produce different results.

5.6 With regards to the lessons that can be learnt about what is needed to undertake this type of analysis in the future (when more and better data will be available), it needs to take account of the following:

a. DRM firms can be matched to the IDBR (and subsequently the ARD) with much greater accuracy if their VAT registration number is collected. It is also useful to have them provide their company registration number (CRN) in case the VAT number is missing (e.g. for firms below the VAT threshold) and possibly their PAYE code. All three numbers are required information that the companies must use for VAT, Companies House and Revenue & Customs purposes. All three codes (especially VAT codes) are available in the IDBR.

b. The records listed above should be updated if and when a company changes status (e.g. it is taken-over or merges).

c. In order to undertake an analysis of whether the firm experienced a change in performance as a result of SE assistance, it is necessary to have (accurate) information on when the firm first received it (as well as information on what type of assistance was available). Otherwise, it is not possible to measure pre- and post-assistance performance.

d. The DRM database should ideally also contain information on some of the most important characteristics of the firm – such as its industry SIC code and numbers of employees. Neither of these was available in the DRM database provided for this study.

e. If in the future SE wish to know whether the reorganisation of DRM firms that occurred post-2004 has had a positive impact on performance, then they will need to wait until a longer in time ARD sample is available (currently data up to 2005 is only available). It is likely that data for both 2006 and 2007 ARD will be available within the next year.

Appendix A

Sample selection issues

A.1 To illustrate, the standard evaluation problem presented in the literature will be briefly presented (cf. Heckman, 2000, and Heckman and Navarro-Lozano, 2004). The key issue is measuring without bias the outcome Y_i of the treatment effect on firms in terms of whether they receive the treatment D_i or not. That is:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \qquad (A4.1)$$

To measure the impact using equation (A4.1), we only have the following information:

$$E[Y_i^1|D_i = 1] - E[Y_i^0|D_i = 0] \qquad (A4.2)$$

that is, the difference between what participants ($D_i = 1$) receiving the treatment experience in terms of outcome (Y_i^1) and what non-participants ($D_i = 0$) not receiving the treatment experience (Y_i^0). What is not observed is the outcome for participants had they not participated (i.e. $E[Y_i^0|D_i = 1]$). The latter counterfactual can be used to expand (A4.2) to give the following:

$$E[Y_i^1 - Y_i^0|D_i = 1] + \{E[Y_i^0|D_i = 1] - E[Y_i^0|D_i = 0]\} \qquad (A4.3)$$

A.2 Equation (A4.3) shows that a comparison between treated and untreated firms (in terms of what is observable – cf. equation 4.2) equals the effect of 'treatment on the treated' (the first term in equation 4.3) plus a bias term (the second major term after the addition sign). As pointed out by Angrist et. al. (1999), this bias would be zero if treated firms were randomly assigned (or at least assigned to ensure independence between Di and Y_i^0).[37] So, for example, if firms enter DRM status independent of (say) the firm's potential productivity gain from receiving assistance, then the bias term would be zero. But this seems unrealistic because selection into DRM status is likely to be made taking account of the potential productivity gains from assistance, and it might be expected that those most likely to achieve higher growth will have a higher probability of breaking-down the barriers to above average performance. Put another way, and referring to the second term in equation (A4.3), bias occurs

[37] Note if D_i is also independent of Y_i^1 (as would be expected in a 'laboratory-type' experiment where firms were randomly assigned) then $E[Y_i^1 - Y_i^0|D_i = 1] = E[Y_i^1 - Y_i^0]$ and the 'treatment on the treated effect equals the unconditional average treatment effect (that is, the impact on a DRM firm drawn randomly from the population of firms).

because the characteristics of the DRM firms are such that they are likely to achieve better performance than non-assisted firms even when they do not receive assistance, and this 'better performance' is correlated with the decision to provide assistance. Thus, the essential problem at the core of the problem of evaluating the effect of DRM status is an attempt to estimate missing data, i.e. obtain an estimate of the unobserved counterfactual that is not biased because of any simultaneous relationship between the decision to provide a company with assistance and the potential gains from such assistance.

A.3 There are several approaches that attempt to eliminate the bias that arises from self-selection (cf. Blundell et. al., 2005). The first considered here is matching. Essentially, this involves matching every DRM firm with another firm that has (very) similar characteristics but does not receive assistance (firms not receiving help from SE that have non-similar characteristics to those who do are of course not included in any analysis of the impact on productivity of DRM status). Thus, under the matching assumption DRM- and non-DRM firms have the same (observable) attributes that impact on productivity except that one sub-group receives assistance and the other does not; put another way, the outcome that would result in the absence of SE assistance is the same in both cases. Thus the non-DRM, matched sub-group constitutes the correct counterfactual for the missing information on the outcomes that DRM firms would have experienced, on average, if they had not received help.38

A.4 Different approaches can be used to match firms, from using simple propensity score matching algorithms (Rosenbaum and Rubin, 1983), where such scores are obtained from a probit/logit regression approach, to covariate matching estimators (that use complicated algorithms to match DRM firms with non-DRM firms). There are a number of issues with this matching process, including the need for a rich dataset set that includes all relevant variables (Xi) that impact on productivity and all variables that impact on whether the firm receives assistance or not (Zi). Matching is done on the set of variables W = (X, Z), so that any selection on unobservables is assumed to be trivial and does not affect outcomes in the absence of assistance. As Heckman and Navarro-Lozano (2004) point out, this requirement can lead to problems since "…if the analyst has too much information about the decision of who takes treatment, so that P(W) =1 or 0, the method breaks down because people cannot be compared at a common W…(thus) methods for choosing W based on the fit of the model to data on D are potentially problematic".39, 40

---

[38] In terms of equation (A4.3), it is assumed: $E[Y_i^0 | D_i = 1] = E[Y_i^0 | D_i = 0]$. Thus matching assumes that Y_i^1 and Y_i^0 are independent of D_i.

[39] Typically DRM firms which are not 'supported' by firms from the non-DRM population are dropped, which can reduce significantly the size of the DRM sub-group included in any analysis. So where there is little common support between the treated and non-treated comparators, matching breaks down.

[40] Another issue is that by definition, matching assumes that the effect for the average DRM firm is the same as the effect for the marginal firm (the 'treatment on the treated' effect equals the unconditional average treatment effect). Heckman and Navarro-Lozano (*op. cit.*) argue that this is an unattractive implication.

A.5 In terms of the practical issues faced in any empirical design of matching firms Bryson et. al. (2002), Imbens (2004) and Zhao (2004) provide a detailed and useful discussion. Here, we have adopted the propensity score matching approach where we first estimate a model to identify the probability of a firm achieving DRM-status (i.e. the propensity score) using the following probit model:

$$P(DRM_{it} = 1) = \phi(X^1_{it-1}, X^2_{it}) \qquad \text{(A4.4)}$$

where *DRM* is coded 1 if the firm received assistance starting in *t*, during 1997-2005; X^1 includes variables that change over time for plant *i* (such as logged real gross output, logged age of the plant, logged capital intensity, logged intermediate inputs intensity, logged diversification and agglomeration indices); X^1 includes variables that do not change over time for plant *i* (such as ownership status, whether R&D was undertaken, whether the plant was located in an assisted area, the employment size-band to which it belonged). Following Girma *et. al.* (2004), if P_i is the propensity score of DRM status for firm *i* at time *t*, we then use the propensity score matching procedure available in STATA 9.2 to find the closest match (using the "nearest-neighbour" approach with common support) for each DRM firm in terms of the propensity scores from the sub-group of non-DRM firms, i.e.:

$$\left|P_i - P_j\right| = \min_{k \notin \{DRM_k = 0\}} \{P_i - P_j\} \qquad \text{(A4.5)}$$

Note, DRM firms with propensity scores P_i that do not have 'common support' (i.e. the scores are higher than the maximum or less than the minimum propensity score for the non-DRM group) are dropped.

A.6 Having obtained a matched sample of DRM and non-DRM firms, there are generally two ways to proceed: firstly, the outcome variable (e.g. total factor productivity - TFP) can be compared for each matched pair and the average value obtained as a measure of the impact of DRM status on TFP. In common with most studies in this field, we do not take this approach but rather estimate a multivariate model using the matched data to test hypotheses regarding the impact of DRM status. This combination of matching and parametric estimation is argued (e.g. Blundell and Costa Dias, 2000) to improve the results obtained from this type of non-experimental evaluation study, as other impacts on the outcome variable are explicitly controlled for.

A.7 A second approach to dealing with self-selection bias is instrumental variable (IV) estimation. If a variable(s) can be found (belonging to Zi) that affects whether a firm enters DRM status but does not affect outcomes (Yi) directly (i.e. Zi is not completely determined by Xi) then such a variable(s) can be used to instrument for Di and overcome the problem of self-selection.41 Put another way, such a variable(s) affects outcomes indirectly since it determines whether a firm is assisted (which is presumed to be correlated with productivity), but it does not need to enter the outcome equation directly (i.e. does not belong to Xi) and is consequently a source of exogenous influence that can be used to identify

[41] Note, the fact that D_i is dichotomous is not a problem according to Angrist (2001).

the causal impact of Di in the model.[42] The main issue with the approach is finding an appropriate instrument(s) that affects the acceptance of a firm into a DRM relationship but does not directly affect outcomes (other than through the firm being assisted). As Angrist and Krueger (2001) point out: "…good instruments often come from detailed knowledge of the economic mechanism and institutions determining the regressor of interest" (p. 73). Blundell et. al. (2005) note that natural candidates as instruments are time constant factors and/or "pre-treatment characteristics". However, in this study we do not have access to any valid instruments; all the variables that determine whether a firm receives assistance can validly enter the model determining productivity.

A.8 The last approach considered here for eliminating the bias that arises from self-selection is the difference-in-difference estimator. If information is available for a pre- and post-treatment period (denoted t′ and t, respectively), then measuring the impact of treatment can be achieved using an amended version of equation (A4.2):

$$\left\{E[Y_{it}^{1}|D_i=1]-E[Y_{it'}^{0}|D_i=1]\right\}-\left\{E[Y_{it}^{0}|D_i=0]-E[Y_{it'}^{0}|D_i=0]\right\} \qquad \text{(A4.6)}$$

where the first term represents the experience of firms who receive assistance between $(t-t')$ and the second term is the experience between $(t-t')$ of those not assisted. To justify this difference-in-difference estimator, it is assumed that (in terms of the counterfactual) what DRM firms would have experienced in the post-entry period, had they not received assistance, is the same as the experience of non-DRM firms, i.e.

$$\left\{E[Y_{it}^{0}|D_i=1]-E[Y_{it'}^{0}|D_i=1]\right\}=\left\{E[Y_{it}^{0}|D_i=0]-E[Y_{it'}^{0}|D_i=0]\right\} \qquad \text{(A4.7)}$$

A.9 The missing counterfactual is now known since rearranging (A4.7) gives:

$$E(Y_{it}^{0}|D_i=1)=E(Y_{it'}^{0}|D_i=1)+\left\{E[Y_{it}^{0}|D_i=0]-E[Y_{it'}^{0}|D_i=0]\right\} \qquad \text{(A4.8)}$$

that is, the outcome that DRM-firms would have experienced post-assistance, had they not received such assistance, equals their outcome effect before assistance takes place adjusted for what happens over the period to all non-DRM firms (the last major term in equation A4.8).

A.10 A major issue with this approach is the assumption underlying equation (A4.7), which is needed to justify the difference-in-differences estimator. Essentially it is assumed that the outcome effect for DRM firms would have been the same as that experienced by non-DRM firms in the absence of assistance; but this seems unlikely if DRM firms are a (self-)selected sub-group exhibiting characteristics that make it more likely they will do better in terms of productivity if they achieve DRM status.

[42] For example, a valid instrument is one that 'forces' a firm into DRM status but which is not correlated with the factors that determine total factor productivity, even though we suspect that DRM status is correlated with TFP.

Appendix B

Analysis of support for businesses in comparator regions

Aims

B.1 As part of the study, we wished to match the 265 relevant DRM businesses to a control group of similar businesses from a comparator region. This group needed to contain firms that were similar to the DRM firms in every respect except that they had received no DRM-type assistance. It was necessary to be reasonably sure that control groups from other regions would not be 'contaminated' in this way. The North East and North West of England were identified as possible areas. In order to determine whether or not they would provide suitable control groups, it was necessary to establish:

- the business interventions that were taking place in these areas during the relevant time period (approximately 2000 – 2005)
- whether or not these business interventions provided support at least as intensive as that received by DRM businesses.

Available evidence base

B.2 A number of difficulties were encountered in establishing the evidence base. Business support agencies' websites provide details about current support available but understandably very little in the way of information about past interventions. SQW colleagues in our North East and North West offices were, however, able to identify some previous initiatives. Evaluation evidence was also sourced from One NorthEast's Enterprise Support evaluation (York Consulting, 2001) and the Economic Impact Study of Business Link Local Service (University of Warwick, Aston Business School and Kingston University, 2007).

B.3 In addition, telephone discussions with One NorthEast (ONE) and North West Development Agency (NWDA) business development managers provided valuable information about start-dates and levels of intensity of support provided. More details about the individual interventions can be found in the following section.

B.4 Overall, this evidence base suggested that there was little available support during the relevant pre-2005 time period, apart from Business Link.

Business Link

B.5 Eighty nine Business Links were established in 1994[43]. They were reorganised in 1999, at which time the number of BLOs (Business Link Organisations) was

[43] Source: http://www.agma.gov.uk/ccm/cms-service/stream/asset/?asset_id=739017

reduced to 45[44]. A further reorganisation in 2007 saw the Regional Development Agencies (RDAs) take over the running of Business Link.

B.6 The University of Warwick Study distinguished 'intensively assisted' businesses and 'other' businesses supported by Business Link. It noted that 72% of the companies that were classified as "intensively assisted" received light touch support comprising a mixture of factual information and advice. Furthermore, over 70% only received a visit from a Business Link manager over a frequency of every three months or less.

B.7 The discussions with SQW colleagues and managers in the RDAs confirmed that this support pre-2007 was well below the level of DRM assistance and varied considerably across local providers.

B.8 Furthermore, Business Link appeared to operate in a different market to DRM. The Warwick study showed that companies supported by Business Link were much smaller than those assisted through DRM. The median size of businesses "intensively assisted" by Business Link was nine employees, compared to an average of nearly 100 employees for DRM assisted companies (ekos/gen, 2009).

B.9 From this we can conclude that if Business Link was the only available form of support in the North East and the North West pre-2005, it is highly unlikely that any companies in the control groups would have been supported in any way comparable to DRM.

Current and previous regional provision

B.10 However, as a cross check, we also looked at what is available in these regions *now*. Table B1 describes some of the main initiatives operating in the **North East** of England.

Table B1: Business support in the North East

Name of intervention	Provider	Time period	Description of support	Number of beneficiaries	Source
High Growth Business Support programme	ONE	2003-07	Successful, high growth companies are matched to a consultant who can help them meet their short-term business development goals. The consultancy support runs **for 6 – 10 days**, although the company and the consultant may agree to maintain the relationship on a wholly independent basis after this time. The programme was supported twice by ONE. *"The first, trial period was run under the auspices of Business Link Tyne and Wear, which paved the way for the establishment of The Alchemists (a private*	100	ONE evaluation of the Business Theme – Enterprise Support (York Consulting, 2008, p14)

[44] Source: University of Warwick (2005), Economic Impact Study of Business Link Local Services, Department for Business, Enterprise and Regulatory Reform

Name of intervention	Provider	Time period	Description of support	Number of beneficiaries	Source
			company) to continue the service." (York Consulting, 2008, p14)		
North East Strategic Account Management (SAM) programme	ONE	Post 2006 (the 2006 RES makes reference to developing a SAM programme)	The SAM programme aims to build long-standing relationships with strategic companies in order to understand, influence and support their medium-long term investment plans. A single point of contact is assigned to the company to help connect them into the wider business support network.	The programme expects to engage approximately 500 companies during its lifespan.	http://www.onenortheast.com/lib/liReport/13422/Corporate%20Plan%202008-13%20-%20Final%20version.pdf http://www.onenortheast.co.uk/lib/liReport/9653/Regional%20Economic%20Strategy%202006%20-2016.pdf

Source: SQW Consulting desk research and interviews

Table B2 outlines the business support currently available in the **North West**.

Table B2: Business support in the North West

Name of intervention	Provider	Time period	Description of support	Number of beneficiaries	Source
High Growth Programme	Business Link	2008 - present	The programme support includes: **"Intensive Coaching** from an experienced mentor. Established SMEs can receive up to 10 days coaching, pre starts and early stage ventures receive between 3 to 5 days **High Growth Workshops** - a series of half day workshops delivered across the North West. **Experience Sharing Networks** an opportunity to share experiences with other, similar, high growth businesses **High Growth Observatory** ... on demand access to a range of learning resources designed for entrepreneurs and those managing high growth businesses"	Target – 1,000 over programme lifetime	Business Link – Support for North West Businesses document. NWDA website. http://www.highgrowthprogramme.co.uk/the_programme/whats_on_offer/
Regional Cluster Organisations Support	NWDA		Regional cluster organisations e.g. aerospace, advance mf, food, environment, bio, media – each offers a range of sector specific support, typically arms length		SQW colleague, NWDA
Account managers	NWDA		A team of account managers are tasked with maintaining relationships with 'star' businesses across the region. However, this support is not as intensive as the high growth coaching and covers a very small proportion of the business population		SQW colleague, NWDA

Source: SQW Consulting desk research and interviews

B.11 This information suggests that the interventions that have been put in place in both regions post-2005 do not strictly compare with DRM, though some might begin to approach it in intensity.

B.12 In addition, NWDA's product briefing for the High Growth Programme noted that "while the region has many of the structures to nurture high growth e.g. business incubation facilities, it lacks a vehicle to build the capabilities to deliver high growth within 'growth wish' SMEs and entrepreneurs." Since this product was only introduced in 2008, it seems clear that there was very little in the way of intensive support for high growth businesses over the pre-2005 period being looked at in the econometric analysis. The recent introduction of the high growth coaching programme offers more intensive support and although this programme will not affect the data period being looked at for the current analysis, it will have implications if the exercise is repeated in future years.

B.13 It would appear that most of the high growth support programmes are recent interventions and were not operational prior to 2005. SQW colleagues and staff at the RDAs were in agreement that those initiatives that were operating earlier were nothing like the same order as DRM. The businesses in these areas did not receive such intensive levels of support as DRM companies did in Scotland, if they received anything at all. Overall, there is little likelihood that the control group would be contaminated in this way.

Documents reviewed for this Appendix

University of Warwick, Aston Business School and Kingston University (2007) *Economic Impact Study of Business Link Local Service,* Final Report, carried out for the Department for Business Enterprise and Regulatory Reform

ekos/gen (2009) *Economic Impact Evaluation of Scottish Enterprise's Interventions with Account and Client Managed Companies*.

NWDA (2008) *Support for Northwest Businesses,* available at: http://www.nwda.co.uk/PDF/NW%20Business%20Support_Nov%2008.pdf

NWDA (year unknown), Product Briefings: High Growth Programme, available at: http://www.nwautoalliance.com/files/documents/news_nwda_high_growth_programme.doc

PACEC (2006) *Mapping of Government Services for Small Business*, prepared for the Small Business Service.

York Consulting (2001) *IEF Impact evaluation of the Business Theme – Enterprise Support*, prepared for ONE Northeast

Appendix C

Spillovers and agglomeration economies

C.1 The first type of agglomeration economies is generally labelled localisation externalities and they are attributable to Marshall (1890), Arrow (1962), and Romer (1986) – hence the term MAR-spillovers. Such spillovers minimise transport and transaction costs for goods, people, or ideas, and thus to benefit from them suggests that firms within a specific industry locate near other firms along the supply chain (be they customers or suppliers); locate near other firms that use similar labour; and/or locate near other firms that might share knowledge (Ellison, *et. al*., 2007). MAR-spillovers are associated with industrial specialisation and are to a large extent an intra-industry phenomenon (where this covers firms belonging to a particular industry, or closely related industries).

C.2 Clearly firms locate in close proximity to reduce the costs of purchasing from suppliers, or shipping to downstream customers. Co-location is also likely if there is a large, common pool of labour. This maximises the 'fit' between productivity levels in firms and workers, since it allows (at lower cost) for labour sorting. It also facilitates workers acquiring industry-specific skills (human capital), since the risk of not being able to appropriate the returns from training are lower where there a large(r) number of potential employers. Again, reverse causality is a possibility because firms may be hiring the same type of workers, because they happen to already be located in the same geographical area. Lastly, firms may co-locate to obtain knowledge spillovers that occur when similar firms engage in, say, R&D to solve similar or related problems. Physical proximity (and density) speeds the flow of ideas, especially when a significant part of intangible knowledge is often tacit (and therefore difficult to codify), and (social) networks tend to be strong.

C.3 As well as MAR-spillovers leading to specialisation, spillovers can also result from urbanisation externalities due to the size and heterogeneity (or diversity) of an (urban) agglomeration. These are labelled Jacobian spillovers (Jacobs, 1970, 1986), and they result when different industries benefit from economies of scope (rather than scale). A greater range of activities (e.g. R&D, business services, cultural and lifestyle amenities, and the overall quality of the public infrastructure – cf Florida, 2002; Glaeser et. al., 2001) leads to inter-industry spillovers. (Larger) firms – and especially multinationals – tend to locate their head office management and R&D functions in urban agglomerations. Thus these agglomerations not only tend to generate more product innovations, but there is more likelihood of spin-offs and/or start-ups, which creates a thicker entrepreneurial culture.

Appendix D

Measurement of Total Factor Productivity[45]

D.1 It is useful to start with a standard production function approach such as:

$$y_{it} = \alpha_0 + \alpha_E e_{it} + \alpha_M m_{it} + \alpha_K k_{it} + \alpha_T t + \varepsilon_{it} \qquad (D.1)$$

where y, e, m and k refer to the logarithms of real gross output, employment, intermediate inputs and capital stock in plant i in time t. In order to calculate TFP, estimates of the elasticities of output with respect to inputs (α_E, α_M, and α_K) need to be obtained using either a growth accounting or production function approach (see below) and then TFP is measured as the level of output that is not attributable to factor inputs (employment, intermediate inputs and capital). Rather TFP measures the contribution to output of all other influences, capturing such determinants as technological progress and/or changes in efficiency (where the latter also captures the under-utilising of factor inputs unless this is taken into account when measuring these inputs). Thus, such a measure of TFP is equivalent to a combination of the residual ε_{it} from (D.1) and the time trend, t, which represents technological change. Hence, TFP is obtained from:

$$\ln \hat{TFP}_{it} \equiv y_{it} - \hat{\alpha}_E e_{it} - \hat{\alpha}_M m_{it} - \hat{\alpha}_K k_{it} = \hat{\alpha}_0 + \hat{\alpha}_T t + \hat{\varepsilon}_{it} \qquad (D.2)$$

D.2 In terms of labour productivity, a relationship can be obtained by subtracting the logarithm of employment from both sides of (D.1):

$$y - e = (\hat{\alpha}_E - 1)e + \hat{\alpha}_M m + \hat{\alpha}_K k + \ln TFP \qquad (D.3)$$

This shows that changes in labour productivity $(y - e)$ are negatively related to changes in employment [since $(\alpha_E - 1) < 0$], and positively related to changes in intermediate inputs, capital stock and TFP. Indeed, if over time there is an increase in capital deepening (*cet. par*. the K/E ratio rises as capital is substituted for labour perhaps due to greater automation) or outsourcing (*cet. par* the M/E ratio rises as less is made internally and more semi-finished and finished products, and services, are bought from suppliers), then labour productivity will increase as relatively less labour is used to produce output.[46] Thus, increases in labour productivity do not depend on just technological progress and/or gains in efficiency, since what happens with the other factors of production is also important. This is the major reason why TFP is preferred as it does not depend on factor substitution.

[45] This section is a shortened version of Harris et. al. (2006)

[46] If a value-added production function were used instead of a gross output function (with VA=Y–M), and constant returns-to-scale imposed with perfect competition in factor and output markets, then (2.4) simplifies to:

$$y - e = (1 - \hat{\alpha}_E)(k - e) + \ln TFP \qquad (D.3a)$$

which shows that labour productivity depends positively on capital deepening and TFP.

Growth accounting approach

D.3 The usual approach to obtain estimates of α_E, α_M, and α_K in equation (D.1) is not to estimate the production function but to use cost shares in total revenue for each factor input (i.e. the ratio of the cost of each input – such as the total wage bill – to total revenue). That is, if it is assumed that firms price goods at marginal cost (and factors are also paid at their marginal costs), then it can be shown that:

$$\alpha_x = \frac{p_x X}{p_Y Y}; \qquad x = E,M,K \qquad \alpha_K = 1 - \alpha_M - \alpha_E \qquad (D.4)$$

D.4 This approach is a useful proxy for obtaining estimates of TFP when data is limited and there is a wish to avoid any econometric estimation. Of course it comes at a price, since the major difficulty with this approach is that the underlying assumption that the sum of factor input shares in total revenue generated equals 1 (the so-called 'adding-up' condition) – which is only consistent with constant returns-to-scale technology and perfect competition in factor and output markets – is unlikely to hold for most industries.

D.5 Growth accounting (sometimes referred to as the Solow residual approach) is a descriptive method and so is most useful when the objective is to measure the relative contributions of input growth and 'other influences' on output growth. Using such an approach has a number of strengths and weaknesses that are summarised in the following table:

Strengths:	Weaknesses:
▪ Only requires data on real output, factor inputs, total revenue and the total cost of labour and intermediate goods – these data are often readily available. Cost shares (which proxy output elasticities) are allowed to vary flexibly across time.	▪ Neither perfect competition or CRTS are likely in many sectors leading to - Biased estimates of TFP (perfect competition underestimates; CRTS overestimates)
▪ No econometric estimation needed	▪ Method does not explain what determines TFP as it is a 'residual'
▪ Easy to construct and interpret	▪ Method does not include the full range of determinants of TFP leading to biased estimates of TFP (see par. D.12 below)
▪ Can expand inputs (and in principle adjust them for quality).	▪ Results obtained are sensitive to measurement errors in the data

Production function approach

D.6 In contrast, the production function approach to measuring TFP (cf. equation 4.1) can be based on the following augmented model:

$$y_t = \alpha_0 + \alpha_E e_t + \alpha_M m_t + \alpha_K k_t + \alpha_T t + \gamma X_t + \varepsilon_t \qquad \text{(D.5)}$$

which is equivalent to equation (D.1) except that we have included a vector of variables, X, that determine TFP to try to ensure that estimates of TFP are not biased because of omitted variables (see below).[47]

D.7 There is an issue about the likely endogeneity of inputs and outputs in (D.5) – given that profit maximisation is usually assumed when specifying this model. Therefore in practice we need to use an estimation approach that takes account of simultaneity (such as using an instrumental variables estimator). There are also econometric issues when using panel data, and the need to take account of fixed effects. However, recent advances in econometrics (such as the systems panel GMM estimator due to Arellano and Bond, 1998) mean that in principle it is often possible to tackle the econometric issues associated with using a production function approach to estimating TFP.

D.8 Summarising we obtain the following:

Strengths:	Weaknesses:
▪ Only requires data on real output, factor inputs, and the determinants of TFP	▪ Does require econometric estimation
▪ Does not impose perfect competition or CRTS	▪ Sensitive to functional form
▪ Flexible functional forms can be used	▪ Sensitive to taking account of endogeneity of output and inputs
▪ Can expand inputs (and in principle adjust them for quality)	▪ Results obtained are sensitive to measurement errors in the data

Explaining the causes of TFP

D.9 Probably the greatest strength of the growth accounting approach to productivity measurement is that it is a straightforward, and easy to understand, technique that identifies the relative importance of different proximate sources of growth. However, it has then to be extended if one wants to explore the underlying causes of growth, innovation and productivity change. (cf. par 2.5.1. OECD, 2001).

D.10 Thus a problem with using either the growth accounting approach, or estimating (D.1) to obtain output-elasticities, and then using these to obtain TFP using

[47] Hence TFP in this instance is defined as: $\ln TFP = \hat{\alpha}_0 + \hat{\alpha}_T + \hat{\gamma}X \equiv y - \hat{\alpha}_E e - \hat{\alpha}_M m - \hat{\alpha}_K k$

(D.2) is that in empirical studies that seek to understand what causes differences in TFP we would then need to model the determinants of TFP; that is, the TFP estimates from equation (D.2) would need to be regressed (using a second-stage model) against a set of determinants which do not feature when obtain the output-elasticities underlying equation (D.1) and yet which clearly are not random (even though in the production function approach they are captured in the random term ε_{it} in equation (D.1), where $\varepsilon_{it} \sim$ n.i.d(0, σ^2) is required for efficient and unbiased estimation of the model).[48]

D.11 It can be shown that using $\ln \hat{TFP}_{it}$ based on equation (D.2) in a second-stage model results in both (i) inefficient estimates (potentially inconsistent standard errors and hence inconsistent *t*-values) of the determinants of TFP, since a two-stage approach has been used (Newey and McFadden, 1999, section 6); and (ii) potentially biased estimates since by omitting factors from equation (D.1) that determine output, the estimates of the $\hat{\alpha}_i$ will suffer from an omitted variable problem and thus $\ln \hat{TFP}_{it}$ is incorrectly measured (Wang and Schmidt, 2002). In general, two-stage approaches are inefficient because they ignore any cross-equation restrictions; but even if there are no cross-equation restrictions, such an approach does not take account of the correlation of error terms across equations.[49]

D.12 The more serious problem is the omitted variable problem. The first-step (equation D.1) ignores other known determinants of output (which are subsequently shown to be statistically significant); here standard econometric theory says that the estimates of $\hat{\alpha}_i$ (and thus TFP) will be biased by such an omission.[50] Moreover, the estimates obtained in the second-step regression will also be biased (downward – see Wang and Schmidt, op. cit., section 2.3 for an explanation). This is true regardless of whether factor inputs and those variables that determine TFP are correlated (see previous footnote). Wang and Schmidt (op. cit.) show that in the case of two-step estimators of technical efficiency using the stochastic frontier production function approach, simulations indicate that bias due to the omitted variable problem is substantial. It is almost certain that their results extend to the present discussion of two-step estimation of the determinants of TFP.

D.13 Thus, if the problem under consideration is to understand the causes of TFP, the preferred approach is arguably to directly include the determinants of output (and thus TFP) into equation (D.1), since this avoids any problems of inefficiency and bias, and also allows one to directly test whether such determinants are statistically significant. That is, since TFP is defined as any

[48] The major reason why the two-stage approach has been popular in the literature is that estimates of the $\hat{\alpha}_i$ are often not obtained from estimating equation (D.1) but rather from using the growth accounting approach.

[49] Since TFP is likely to be endogenous, clearly on this front alone the error terms between stage 1 and 2 are correlated.

[50] Bias will be negligible only if the two sets of determinants of output (i.e. factor inputs and those variables that determine TFP) are uncorrelated. Since both sets of factors are firm specific, they are likely to be highly correlated.

change in output not due to changes in factor inputs, these determinants should be included directly into equation (D.1), resulting in equation (D.5).

Most appropriate methodology

D.14 As a means of describing the relative importance of TFP at the industry level, growth accounting is a useful approach to adopt. It is likely to be more appropriate for producing figures that can be published by Scottish Enterprise, that are complied using a well-know (and comparable) technique complying to OECD standards (OECD, 2001).

D.15 However, to understand the sources of TFP and its main determinants, and thus the reasons behind Scotland's performance in this area, then a production function approach is more appropriate.

D.16 Thus, both approaches have their place, and both can usefully be pursued subject to the availability of data. To undertake the growth accounting approach in its simplest form requires information on sales/turnover, intermediate inputs, labour input, the capital stock, plus total revenue and the shares of total labour and intermediate costs in total revenue. Such data are regularly supplied by firms in their annual returns made to Companies House.

References

Angrist, J.D. (2001) Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice. *Journal of Business and Economic Statistics*, 19(1), 2-16.

Angrist, J.D. and Imbens, G. (1995) Two-Stage Least Squares Estimates of Average Causal Response in Models with Variable Treatment Intensity. *Journal of the American Statistical Association*, 90, 431-442.

Angrist, J.D. and Krueger, A. (2001) Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, 15(4), 69-85.

Angrist, J.D., Imbens, G. and Krueger, A. (1999) Jackknife Instrumental Variables Estimation. *Journal of Applied Econometrics*, 14, 57-68.

Arellano, M. and Bond, S.R. (1998) Dynamic Panel Data Estimation using DPD98 for GAUSS: A Guide for Users. available at http://www.american.edu/academic.depts/cas/econ/gaussres/regress/dpd/dpd98.pdf

Arrow, K.J. (1962) The economic implications of learning by doing, *Review of Economic Studies*, 29, 155-173

Blundell, R. and Bond, S. (1999) GMM Estimation with Persistent Panel Data: An Application to Production Functions. IFS Working Paper Series No. W99/4.

Blundell, R. and Costa Dias, M. (2000) Evaluation Methods for Non Experimental Data. *Fiscal Studies*, 21(4), 427-468.

Blundell, R., Dearden, L. and Sianesi, B. (2005) Evaluating the Effect of Education on Earnings: Models, Methods and Results from the National Child Development Survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(3), 473-512.

Bryson, A., Dorsett, R., and Purdon, S. (2002) The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies. Policy Studies Institute and National Centre for Social Research Working Paper No. 4.

Ellison, G. *et. al*. (2007) What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns, *NBER Working Paper Series*, Working Paper 13068.

Florida, R. (2002) *The rise of the creative class and how its transforming work, leisure community and everyday life.* Basic Books, New York.

Glaeser, E. *et. al*. (2001) Consumer City, *Journal of Economic Geography*, 1, pp. 27-50.

Harris, R.I.D (2002) Foreign Ownership and Productivity in the United Kingdom - Some Issues When Using the *ARD* Establishment Level Data, *Scottish Journal of Political Economy*, 49, 318-335.

Harris, R.I.D and Robinson, C. (2003) Foreign Ownership and Productivity in the United Kingdom: Estimates for UK Manufacturing Using the *ARD*, *Review of Industrial Organisation*, 22, 207-223

Harris, R.I.D. (2005) Economics of the Workplace: Special Issue Editorial. *Scottish Journal of Political Economy,* 52(3): 323-343.

Harris, R.I.D. (2005b) Deriving Measures of Plant-level Capital Stock in UK Manufacturing, 1973-2001. Report to the DTI, London.

Harris, R.I.D. and Drinkwater, S. (2000) UK Plant and Machinery Capital Stocks and Plant Closures, *Oxford Bulletin of Economics and Statistics*, 62, 239-261.

Harris, R.I.D., O'Mahony, M., and C. Robinson (2006) Research on Scottish productivity. Reported submitted to the Scottish Executive Office of the Chief Economic Advisor. http://www.cppr.ac.uk/media/media_5535_en.pdf.

Heckman, J. (2000) Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *The Journal of Human Resources*, 32(3), 441-462.

Heckman, J. and Navarro-Lozano, S. (2004) Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models. *The Review of Economics and Statistics*, 86(1), 30-57.

Imbens, G.W. (2004) Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. *The Review of Economics and Statistics*, 86, 4-29.

Jacobs, J. (1970) *The Economy of Cities*, Vintage, New York, NY.

Jacobs, J. (1986) *Cities and the Wealth of Nations.* Vintage, New York, NY.

Marshall, A. (1890) *Principles of Economics*, Macmillan, London.

Moffitt, R.A. (2004) Introduction to the Symposium on the Econometrics of Matching. *The Review of Economics and Statistics*, 86(1), 1-3.

Newey, Whitney K. and Daniel McFadden (1999) "Large Sample Estimation and Hypothesis Testing", in Daniel L. McFadden and Robert F Engle (eds.) *Handbook of Econometrics Volume 4*, Amsterdam: North Holland.

OECD (2001) Measuring Productivity. Measurement of Aggregate and Industry Level Productivity. OECD Manual. OECD, Paris.

Oulton, N. (1997). 'The ABI respondents database: a new resource for industrial economics research', *Economic Trends*, Vol. 528, pp. 46-57.

Romer, P.M. (1986) Increasing returns and long-run growth, *Journal of Political Economy*, 94(5) pp.1002-1037.

Rosenbaum, P.R. and Rubin, D.B. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41-55.

Wang, Hung-Jen and Peter Schmidt (2002) "One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels", *Journal of Productivity Analysis,* 18, 129-144.

Zhao, Z. (2004) Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence. *The Review of Economics and Statistics*, 86, 91-107.