Australian Bureau of Statistics

1352.0.55.128

Research Paper

Propensity Score Matching: An Application using the ABS Business Characteristics Survey

New Issue

Research Paper

Propensity Score Matching: An Application using the ABS Business Characteristics Survey

Cristian Rotaru, Sezim Dzhumasheva and Franklin Soriano

Analytical Services Branch

Methodology Advisory Committee

30 November 2012, Canberra

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) TUES 26 FEB 2013

ABS Catalogue no. 1352.0.55.128

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper.
For further information, please contact Mr Ruel Abello, Analytical Services Branch on Canberra (02) 6252 6307 or email <analytical.services@abs.gov.au>.

PROPENSITY SCORE MATCHING: AN APPLICATION USING THE ABS BUSINESS CHARACTERISTICS SURVEY

Cristian Rotaru, Sezim Dzhumasheva and Franklin Soriano
Analytical Services Branch

QUESTIONS FOR THE COMMITTEE

1. In the literature, most researchers use the PSM to create matched samples that are used to calculate treatment effects for already available outcomes. There does not seem to be clear guidance around the use of a matched sample for regression modelling, particularly when the potential outcomes are rather unmeasured values (i.e. binary outcomes in this case). What are the theoretical and pragmatic implications of PSM in this case? Are there any implications on the current study?

2. What are the comments and/or ideas of the MAC members on the potential uses of PSM at ABS in the future?

3. In future studies the ABS might be interested in using the PSM for analyses regarding the whole population using survey weights. However, there does not seem to be clear guidance in the literature on how to go about them. What are the views of the MAC members regarding the implementation of PSM in this case? In particular, how should these weights be computed and applied given that only a portion of the sample is retained after matching?

4. Do the MAC members have any comments on the interpretation of the random effects component in the probit mixed model for paired samples? Should this be linked to the quality of matching?

5. The covariates used in estimating the propensity scores were also used in the probit modelling for innovation together with other variables. Although very limited, the studies which followed a similar approach focused mainly on the impact of the variable of interest (government assistance in this case) and did not say much about the other covariates. What are the views of the MAC members on this?

6. In assessing the effects of treatment on the outcome of interest one is often interested in computing marginal effects. If standard errors are to be computed for the marginal effects, what is the most effective way of doing this? In the case of bootstrapping, what are the implications of having matched the sample before computing the estimates?

CONTENTS

ABSTRACT 1

1. INTRODUCTION 2

2. BACKGROUND AND DATA 4
    - 2.1 Introduction to propensity score matching 4
    - 2.2 PSM related studies 5
    - 2.3 Data 6

3. PROPENSITY SCORE MATCHING 7
    - 3.1 Assumptions 7
    - 3.2 Matching algorithms 8

4. PSM APPLICATION 11
    - 4.1 Diagnostics 12
    - 4.2 Cautionary notes 17
    - 4.3 Summary 18

5. MODEL APPLICATION 19

6. CONCLUDING REMARKS 24

ACKNOWLEDGEMENTS 24

REFERENCES 25

APPENDIXES

A. DATA COMPILATION 28

B. OTHER PSM DIAGNOSTIC RESULTS 32

PROPENSITY SCORE MATCHING: AN APPLICATION USING THE ABS BUSINESS CHARACTERISTICS SURVEY

Cristian Rotaru, Sezim Dzhumasheva and Franklin Soriano
Analytical Services Branch

ABSTRACT

This study applies the propensity score matching (PSM), as suggested in Rosenbaum and Rubin (1983), in the context of causal modelling using the ABS *Business Characteristics Survey* (BCS). In particular, the study uses the PSM to match the firms which received government assistance to those which did not receive government assistance. In studying the effects of government assistance, such matching is important in order to account for the systematic differences between the treated (assisted) and control (non-assisted) firms. If not accounted for, there will be uncertainty about whether the difference in the outcome of interest between the two groups is caused by the effect of the treatment (government assistance) or because of the pre-treatment differences between the two groups. One could not simply assume that the government assistance is the only factor that differentiates the outcomes of the businesses.

The study examines different matching algorithms, conducts tests to evaluate the quality of matching, and applies a selected algorithm to a specific case study – analysing the effect of government assistance on the firm's propensity to innovate. In order to address the correlation within matched pairs, a random effects model for binary matched pairs is tested following the approach outlined in Agresti (2002) – in this case, a probit generalised linear mixed model (probit GLMM).

1. INTRODUCTION

It is often the case that a researcher or policy analyst is interested in assessing the effects of an intervention, such as that of a treatment, policy change, new drug, or a certain malady. One of the major issues faced in such an investigation is what is often referred to in the literature as the fundamental problem of causal inference (Holland, 1986), namely, that in the context of a treatment, one cannot observe both the response to treatment and non-treatment for the same subject, at one time. For example, in the case of a drug experiment, one cannot detect both the effects of a drug on a patient and the counterfactual effects, i.e. what would have happened in the absence of treatment. Given this problem, some analysts turn to the non-participating units for information about the missing data and for the estimation of the counterfactual outcomes. This is where statistical matching comes into play and where the idea is to obtain the required information about the missing data by matching the treated units to non-treated or control units on the basis of similar characteristics or similar covariate distributions.

Note that a random experiment ensures that the treated and control groups are only randomly different with respect to the covariates of interest. However, in the case of a non-random experiment or in the case of an observational study – where the analyst lacks the control over the randomisation of the outcomes – the units are generally not randomly assigned to treatment and there is the potential of selection bias (see Rosenbaum, 2002). One way of dealing with this is by using statistical matching techniques, such as the propensity score matching (PSM) suggested by Rosenbaum and Rubin (1983), which has become popular in policy evaluation studies (see Heinrich *et al.*, 2010). The attraction of PSM is in its simplicity, as it matches the treated and control units on a single dimension, the propensity score, which is defined as the conditional probability of receiving treatment given a set of observed covariates. Other alternatives include regression analyses, which incorporate the treatment selection process in the model.

It is in this setting of observational studies that this paper applies the propensity score matching in the context of the Australian Bureau of Statistics (ABS) Business Characteristics Survey (BCS). The focus is on matching the firms which received government assistance to those which did not receive government support. The paper begins with a methodological focus and examines different matching algorithms, such as the Nearest Neighbour (NN), the Caliper, and the 5 to 1 Digit Matching. The aim is to construct a new sample with balanced treated and control units so as to control for the selection bias. Hereafter, the new sample will be referred to as the 'matched sample' and the full sample as the 'unmatched sample'. In order to assess the quality of matching, different tests are conducted, including the popular chi-square and the standardised bias tests.

In the second part, the paper applies the PSM to a specific case study – analysing the effect of government assistance on the firm's propensity to innovate. Government assistance in this study refers to grants, funding, subsidies, tax concessions, or rebates. Using the *matched sample,* a random effects probit model is employed, with the random effects part controlling for the correlation within the matched pairs.

The paper is organised as follows. Section 2 provides a short discussion of the PSM framework, followed by a brief background of the existing PSM literature and a short description of the data. Section 3 covers the underlying PSM assumptions and methodology. Section 4 presents the PSM application and some diagnostics. Section 5 uses the matched sample for probit generalised linear mixed modelling and related analyses. The concluding remarks are given in Section 6.

2. BACKGROUND AND DATA

2.1 Introduction to propensity score matching

One challenge with analysing treatment effects in observational studies is dealing with the non-random allocation of treatment, which if ignored can lead to biased results. One way to address this problem is to make adjustments to the initial sample by balancing the treated and control units on the basis of selected observed covariates.

This balancing however, might lead to the curse of dimensionality associated with trying to match on a large number of covariates. To address this problem Rosenbaum and Rubin (1983) developed a widely used method, where the units are matched on propensity scores, hence the name of the method (PSM). The said probabilities, i.e. propensity scores, summarise all the relevant information contained in the set of covariates. For a more thorough understanding of the method see Rosenbaum and Rubin (1983).

The PSM Procedures

Caliendo and Kopeinig (2008) provides practical guidance for the implementation of propensity score matching. Heinrich *et al.* (2010) also presents a primer tailored for practitioners. As described in the mentioned studies, generally, PSM is implemented in four steps.

i. Estimating the propensity scores

Two important choices need to be made in estimating the propensity scores. The first relates to the correct specification of the model used to estimate the propensity scores, and the second to the identification of the covariates included in the model. For the specification of the model, most applications use either a logit or a probit model. For the selection of variables, Heinrich *et al.* (2010) notes that one should consider the existing criteria used in determining the treatment participation.

ii. Choosing a matching algorithm

Although there are many matching algorithms in the literature there is no clear indication as to the preferred one. According to Caliendo and Kopeinig (2008) the choice depends on the context and aim of the analysis. Coca-Perraillon (2006) and Heinrich *et al.* (2010) point out that all techniques share common elements, which include:

- an operational (or standard) definition of similarity (or distance) between propensity scores;
- a decision regarding the number of controls to be matched to each treated unit;
- whether the matching should be done with or without replacement; and
- whether one should use weights or not.

As a guideline, matching with replacement is recommended when the size of the control group is small or when there is little overlap in the propensity score distributions of the two groups. Some of the most commonly employed matching algorithms include nearest neighbour, caliper, radius, kernel or local linear matching, stratification, and interval matching. More information about these algorithms and running them in SAS can be found in Parsons (2001, 2004) and Coca-Perraillon (2006, 2007).

iii. Performing diagnostics so as to evaluate the assumptions and the quality of matching

To ensure the validity of the PSM, it is important to verify the key assumptions, namely, the conditional independence and the common support conditions. These assumptions are further discussed in Section 3. Assessment of the quality of matching can be undertaken by using standard test procedures like the standardised bias test, the t-test, the joint significance test, and the pseudo R-squared test. Caliendo and Kopeinig (2008) and Heinrich *et al.* (2010) provide an elaborate description of these procedures.

iv. Estimating the treatment effect

Once the diagnostic tests are conducted, the analyst proceeds to the estimation of the treatment effects and its associated standard errors. One common way to estimate the treatment effect is by averaging the differences in outcomes between each paired observations. The standard errors are conventionally calculated using bootstrapping methods.

2.2 PSM related studies

Propensity score matching has been applied in a wide variety of studies. A paper by Stuart (2010) provides a good overview of the evolution of the method and its uses in various fields. However, in spite of this, its application to evaluating the relationships between various forms of government assistance and innovation has been limited. One example is Almus and Czarnitzki (2003), where the authors looked at the effects of public R&D subsidies on the firms' innovation activities (for Eastern Germany). Other examples include Heijs and Herrera (2004) and Herrera and Nieto (2006) where the authors conducted similar analyses for Spain. Perhaps the most relevant to this paper is the UK study by Foreman-Peck (2010) which made use of PSM to determine if the government support increased the likelihood of innovation by small and medium sized enterprises (SMEs) in the manufacturing and services industries.

There are several Australian studies that applied propensity score matching. Two examples are Dockery (2005) and Houssard *et al.* (2010). The first assessed the value of additional years of schooling for non-academically inclined, and the second the impact of HECS debt on socioeconomic inequality and transition to adulthood outcomes.

To the best of the authors' knowledge this paper is the first to methodologically examine the technique and apply it to ABS business survey micro data.

2.3 Data

This study utilises firm level data for Australian businesses covered by the 2009–2010 wave of the Business Characteristics Survey (BCS). The BCS is an annual survey that provides population estimates for a range of business topics and themes. The BCS collects detailed information on business demographics, innovation activities, use of information and communication technology (ICT) and related practices and influences. The survey asks information on whether businesses received government assistance in the form of grants, funding, subsidies, tax concessions, or rebates. The 2009–10 survey shows that, overall, 18.2% of Australian businesses received some form of government financial assistance (ABS, 2011b). The details of the compilation of the variables used in the analysis are described in Appendix A.

3. PROPENSITY SCORE MATCHING

As mentioned in the previous section, propensity score matching aims to create a matched dataset with balanced observed covariates. By this, the intention is to adjust for the selection bias and attribute the differences between the outcomes of the treated and non-treated units to the treatment alone and therefore control for the effects coming from the observed covariates. In other words, the PSM attempts to create an analogue of a sample coming from a pure randomised experiment, where the treated and non-treated units in the matched sample are considered random after controlling for observed covariates. (See Rubin, 2006, for a good coverage of the PSM methodology.)

The PSM method, which is a neat and simple concept, comes with a set of assumptions and it has its own complexities when applied in practice. As an understanding of these assumptions is necessary for any PSM application, this section starts with a brief explanation of two important assumptions followed by a discussion of the different algorithms that were used for matching the pairs.

3.1 Assumptions

In order to address the selection bias it is important to ensure that two central assumptions are satisfied: the conditional independence assumption and the common support condition.

The Conditional Independence Assumption

The assumption states that after controlling for the observable covariates (denoted by vector X), the potential outcomes for receiving or not receiving treatment (denoted by Y_1 and Y_0, respectively) are independent of the treatment assignment (denoted by T). This can be written as

$$\left(Y_0, Y_1\right) \perp T \middle| X .$$

The assumption implies that the researcher observes all the covariates that influence the treatment assignment and the potential outcomes simultaneously.

Heinrich *et al.* (2010) identifies several requirements that are important for justifying this assumption. These include the availability of a large set of covariates, the need for both the control and treatment sets to belong to the same source, consistency in handling the missing data, and the need for a large pool of control units with corresponding characteristics to the treatment units.

The Common Support condition

The common support assumption states that for each value of *X* there is a positive probability of both receiving and not receiving treatment (Heinrich *et al*., 2010). In mathematical terms, this means that

$$P\left(T=1\middle|X\right) \in (0,1).$$

This requirement ensures that there is sufficient overlap, or common support, in the characteristics of the treated and control units. To check for this condition, this paper followed the approach suggested by Heinrich *et al.* (2010) and Caliendo and Kopeinig (2008) and visually inspected the density distributions for the two groups.

3.2 Matching algorithms

Once the two assumptions are verified and the propensity scores are estimated the analyst needs to select an algorithm to match the estimated propensity scores. Although there are no clear guidelines regarding the most optimal matching algorithms, some considerations are important when choosing between them. The first is in regards to the desired measure of proximity between matched units, where the analyst might be interested in imposing a restriction on the maximum distance between the propensity scores of a matched pair. The second concerns the weighting function which is to be assigned to the units or to the neighbourhood of units. (See Essama-Nssah, 2006 for more details about the first two considerations.) The third is whether the matching should be done with or without replacement.

This paper considered three common matching methods:

- the Nearest Neighbour (NN),
- the Caliper, and
- the $d_1 \to d_0$ Digit Matching (which in the rest of the paper is abbreviated as $DM_{d_1 \to d_0}$).

Note that all three algorithms impose a weight of one to the nearest neighbour and zero to the others and that they were applied without replacement. A brief description of each algorithm is given below.

The Nearest Neighbour (NN)

The NN is the simplest of the three methods and it matches the treated units to the control units based on the closest propensity score. In mathematical terms, this can be defined as

$$c\left(p_i,\tau_i\right) := \min_j \left\| p_i - p_j \right\|, i \in S_p,\ j \in S_n,\ \tau_i \in \{0,1\}.$$

Note that in the above notation S_p denotes the set of participants in the treatment group, S_n the set of non-participants, p_i and p_j the propensity scores of unit i and j, respectively, and $c(p_i, \tau_i)$ denotes the neighbourhood of the participating unit i. In the expression, τ_i denotes whether the matching was done with replacement $(\tau_i = 0)$ or without replacement $(\tau_i = 1)$. See Essama-Nssah (2006) and Todd (2006) for more details.

The Caliper

The caliper matching is similar to the nearest neighbour but it assigns a 'caliper' or maximum distance between the propensity scores of a matched pair. As such, the algorithm aims to correct for the bad matches that might result from the implementation of the nearest neighbour algorithm. In practice, the caliper is usually set to 0.2 or 0.25 standard deviations of the propensity score (see Rosenbaum and Rubin, 1985).

In mathematical terms the algorithm can be defined as

$$c(p_i, \tau_i) := \min \left\| p_i - p_j \right\| \text{ subject to } \left\| p_i - p_j \right\| < \delta,$$

where, as before, $i \in S_p$, $j \in S_n$ and $\tau_i \in \{0,1\}$, and δ is the 'caliper'.

Note that the above expression is similar to that of the Nearest Neighbour algorithm with the difference that it imposes a restriction on the maximum distance between propensity scores.

See Essama-Nssah (2006) and Todd (2006) for more details.

$d_1 \rightarrow d_0$ *Digit Matching*

The digit matching algorithm can be thought of as a modified version of the nearest neighbour algorithm in that it matches the units of the two groups in terms of the closest propensity scores. The algorithm can also be considered as a special kind of Caliper algorithm as it imposes an implicit restriction (or caliper) on the distances between propensity scores at each digit level. Developed by Parsons (2004), the algorithm performs the matching in a number of stages. This number depends on the difference between the initial (d_1) and final (d_0) number of decimal digits that are considered for matching. After sorting the treated and control units based on their unrounded propensity scores, the algorithm first matches only those units which have exactly the same first d_1 digits of their rounded propensity scores (rounded to the first d_1 digits). For those that did not match, the algorithm then matches on the first $d_1 - 1$ digits, then on $d_1 - 2$, and so on until the matching is done on the first d_0 decimal digits. When this is reached, the algorithm stops considering any more matchings. So, for example if $d_1 = 5$ and $d_0 = 1$, which is the case considered in this paper[1], the first stage would match those pairs that have the same first five decimal digits of their rounded propensity scores. For those that did not match, the algorithm would continue matching on the first four digits, then on three, two, and finally on one digit.

In mathematical terms, the authors formulated the algorithm as shown below, where the initial step matches the propensity score on the first d_1 digits, the second on $d_1 - 1$ digits, the third on $d_1 - 2$, and so on (as denoted by the k-th step) until the final step matches on the first d_0 digits.

Initial Step:
$$c_{d_1}(p_i, \tau_i) := \min_j \left\| p_i - p_j \right\|,$$
$$\text{subject to } \operatorname{nint}\left(10^{d_1} p_i\right) - \operatorname{nint}\left(10^{d_1} p_j\right) = 0;$$

k-th Step:
$$c_{d_1-(k-1)}(p_i, \tau_i) := \min_j \left\| p_i - p_j \right\|,$$
$$\text{subject to } \operatorname{nint}\left(10^{d_1-(k-1)} p_i\right) - \operatorname{nint}\left(10^{d_1-(k-1)} p_j\right) = 0;$$

$$\vdots$$

Final Step:
$$c_{d_0}(p_i, \tau_i) := \min_j \left\| p_i - p_j \right\|,$$
$$\text{subject to } \operatorname{nint}\left(10^{d_0} p_i\right) - \operatorname{nint}\left(10^{d_0} p_j\right) = 0;$$

where, as before, $i \in S_p$, $j \in S_n$, $\tau_i \in \{0,1\}$ and k takes on all the integer values from 1 through $d_1 - d_0 + 1$.

Note that $nint(.)$ is the nearest integer function which rounds the number to the closest integer.

1 Note that this study also considered some other variations of the digits, such as 8→1, 7→1 and 6→1. The results however were similar.

4. PSM APPLICATION

This section presents the application of the PSM to the ABS Business Characteristics Survey where the firms that received government assistance are matched to those that did not receive government support. The aim is to control for the selection bias by balancing the sample, at least with respect to the observed covariates. This sample will be used later in the modelling in Section 5.

Before proceeding to the PSM procedure, the study checked if the matching is required. The sample was closely examined and tests were performed to investigate the balance of the sample across a set of key variables, which were likely to have an effect on the firm's propensity to receive government support. Table 4.1, second column, presents the differences in means (proportions in this case since the variables are categorical) of the selected characteristics between both groups. The groups were significantly and substantially different across all selected variables. This indicates that the granting of government assistance may be predisposed to selection bias and that the selection process is unlikely to be random. The implementation of PSM could play an important role in addressing this issue.

4.1 Chi-square values (p-values) on differences

			After matching					
Firm characteristics	*Before matching*		*Nearest Neighbour*		*Caliper*		*5 to 1 digit matching*	
Market competition	11.80	(0.0081)	1.26	(0.7396)	4.03	(0.2578)	3.22	(0.3585)
R & D agreement	238.64	(<0.0001)	64.07	(<0.0001)	0.37	(0.5455)	0.31	(0.5762)
Export activity	156.77	(<0.0001)	12.33	(0.0004)	0.74	(0.3899)	0.17	(0.6790)
Other finance	183.58	(<0.0001)	8.66	(0.0032)	0.001	(0.9738)	0.02	(0.8954)
Foreign ownership	85.83	(<0.0001)	4.77	(0.0923)	2.47	(0.2904)	1.06	(0.5891)
ICT intensity	481.17	(<0.0001)	40.65	(<0.0001)	2.14	(0.7102)	2.91	(0.5722)
Number of employees	1188.67	(<0.0001)	150.67	(<0.0001)	1.60	(0.6598)	0.39	(0.9433)
Industry division	498.32	(<0.0001)	159.01	(<0.0001)	15.96	(0.4559)	9.98	(0.8677)

In order to estimate the propensity scores, an ordinary binary probit model with the government assistance as the dependent variable was used. The selection of covariates was based on previous similar studies (see Heinrich *et al*., 2010; Caliendo and Kopeinig, 2008), the program eligibility criteria, institutional factors, as well as theoretical and pragmatic considerations. The following business characteristics were included in the model: business size; industry of operation; a variable describing whether the firm has cooperative R&D; degree of competition; degree of foreign ownership; whether financing options were used; exporting activity; and a variable capturing the firm's information and communication technology (ICT) intensity. The descriptions of the variables are provided in Appendix A.

The regression results of the estimation of propensity scores are shown in Appendix B.1. Most of the coefficients were significant ($p<0.05$) with the exception of competition, several industry categories and one of the ICT intensity and foreign ownership categories.

When the chi-square tests were re-run on the matched pairs – to check whether the differences between participating and non-participating units across the selected variables persist after matching – the results for the nearest neighbour (NN), included in table 4.1, were poorer than those of the other two algorithms. This poor performance can be attributed to the fact that since the NN algorithm matched more pairs (as the algorithm does not impose a maximum tolerance level) some of these additional matches were poor, bringing down the overall test results.

4.1 Diagnostics

This subsection further explains the performance of the PSM and of the three algorithms used. The section firstly addresses the mentioned PSM assumptions, presents the standardised bias diagnostic results, and finally presents some micro assessment results meant to canvass the performance of the matching.

Verifying the assumptions

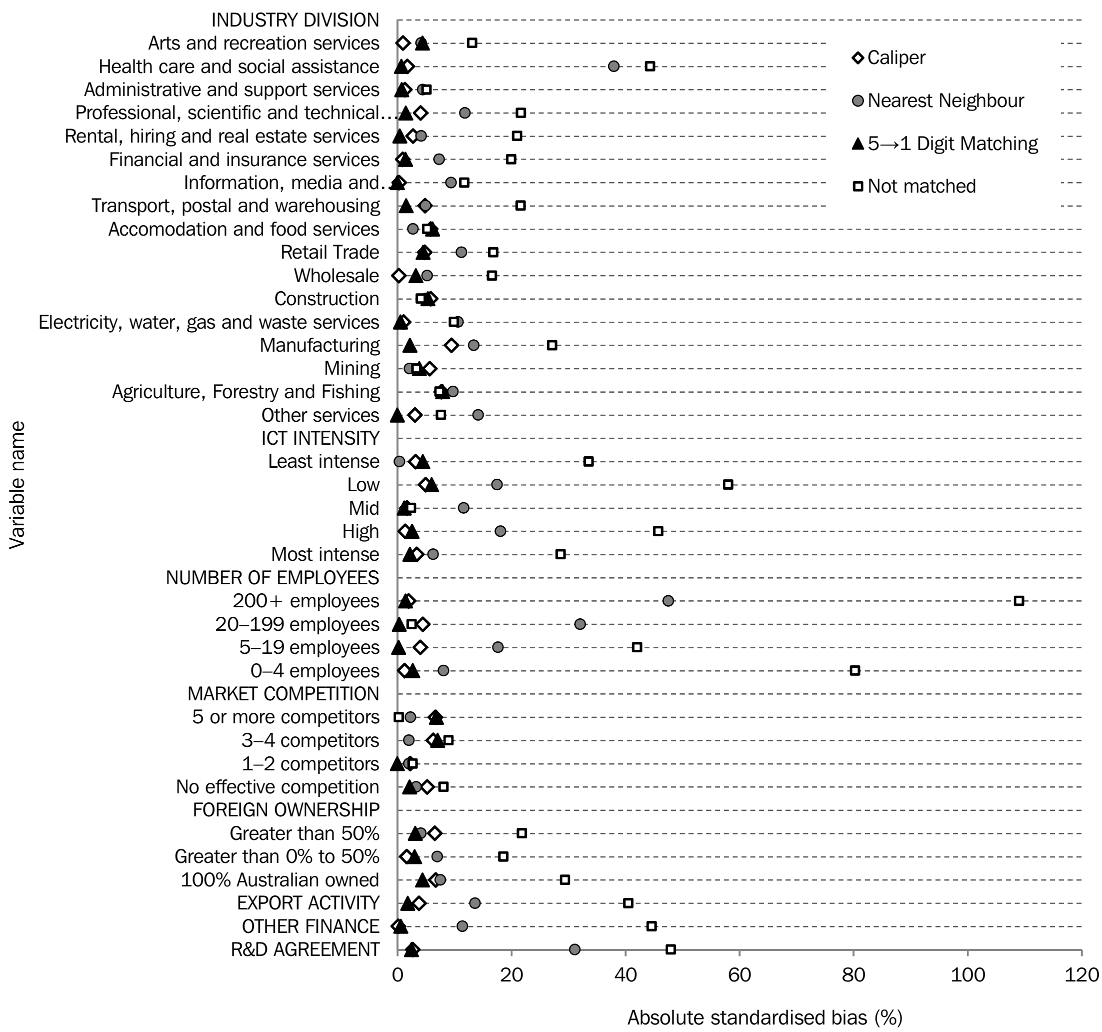
As discussed in Section 3.1, there are two important conditions that are important in the implementation of the PSM. Regarding the first, the conditional independence assumption, the authors paid close attention in including the relevant variables in estimating the propensity scores and in correctly specifying the model. (See Almus and Czarnitzki, 2003; Heijs and Herrera, 2004; Herrera and Nieto, 2006; and Foreman-Peck, 2010 for more details regarding the requirements of the independence condition.) The study also satisfied the requirements mentioned in Heinrich *et al.* (2010): the availability of a rich dataset with a large number of observations and covariates; the same source for the control and treatment sets; the consistency in handling the missing data; and the availability of a large pool of control units with corresponding characteristics to the treatment units.

With respect to the common support or overlap condition, this study visually inspected the propensity score distributions for the two groups before and after matching. These are plotted in Appendix B.2. The propensity score distributions are considerably more similar after matching, a case when the plots also reveal a good overlap.

The standardised bias test

Chi-square test results were included earlier (see table 4.1) and they were used to assess the performance of the three matching algorithms. In what follows, another popular PSM validation test, suggested by Rosenbaum and Rubin (1985), was conducted, namely the standardised bias test. For each covariate, the standardised bias is calculated by dividing the difference between the means of the treated and matched control subsamples by the square root of the average of the variances in both groups (Caliendo and Kopeinig, 2008). Note that as with the chi-square test, one of the aims of the investigation is to evaluate the sample before and after matching and to check for any imbalances that remain between the groups. For successful matching there should be a decrease in the selection bias due to the increase in the balance between the treatment and control groups (see Heinrich *et al*., 2010). The standardised bias test has a limitation in that it does not have a clear threshold for acceptable balance, although most studies consider that a standardised bias of less than 3% or 5% suffices.

4.2 Standardised bias graph before and after matching



From the results, shown in figure 4.2, a few observations can be made. First, before matching, the standardised bias was considerably larger for most of the variables. This is consistent with the previous findings of pre-matching imbalances between the two groups across the selected variables. Second, in line with the previous chi-square results, the Caliper and $DM_{5\rightarrow1}$ results show significant improvements in addressing the selection bias. Third, consistent with the previous results, the NN matching results are inferior to those of the other two algorithms.[2]

Micro assessment of the matched pairs

In order to visually inspect the results and the quality of matching in more detail, the authors conducted micro assessments on the matched sample with respect to two key variables, firm size and industry. Table 4.3 presents the assessment results for size, table 4.4 for industry, and table 4.5 for both industry and size.

4.3 Breakdown of the matched pairs by business size* (numbers and percentages)

	Nearest Neighbour		*Caliper*		$DM_{5\rightarrow1}$	
	Number of firms	*%*	*Number of firms*	*%*	*Number of firms*	*%*
Correctly matched						
micro to micro	221	8.3	220	10.7	220	10.7
small to small	255	9.5	263	12.7	270	13.2
medium to medium	450	16.8	455	22.0	488	23.8
large to large	846	31.6	680	33.0	690	33.6
Total	1,772	66.3	1,618	78.4	1,668	81.3
Not correctly matched						
micro to small	34	1.3	31	1.5	51	2.5
micro to medium	42	1.6	37	1.8	27	1.3
micro to large	47	1.8	16	0.8	13	0.6
small to medium	119	4.5	88	4.3	78	3.8
small to large	176	6.6	51	2.5	46	2.2
medium to large	484	18.1	223	10.8	169	8.2
Total	902	33.7	446	21.6	384	18.7

* micro (0–4 employees); small (5–19 employees); medium (20–199 employees); large (200+ employees)

From table 4.3 it can be noted that the $DM_{5\rightarrow1}$ algorithm outperformed the NN method and was just slightly better than the Caliper algorithm. The $DM_{5\rightarrow1}$ correctly matched over 81 per cent of the government non-assisted firms to similarly sized government assisted firms. Of the non-correctly matched cases, almost half were because medium sized firms were matched to large firms, and around 20 per cent

2 This refers to the fact that the NN was not successful in balancing the sample. Note that although the NN matched more pairs than the other two algorithms, and was therefore more successful in terms of retaining more sample observations, its performance was much poorer in terms of balancing the two groups of firms.

because small firms were matched to medium sized firms. The most serious mismatches occurred when micro firms were matched to medium or large ones, or when small firms were matched to large firms, which from the table seems to have occurred in just over 4 per cent of the cases for $DM_{5\to1}$.

4.4 Breakdown of the matched pairs by industry (numbers and percentages)

	Nearest Neighbour		*Caliper*		$DM_{5\to1}$	
Industry	*Number of firms*	*% correctly matched*	*Number of firms*	*% correctly matched*	*Number of firms*	*% correctly matched*
Agriculture, forestry and fishing	251	45.4	198	65.7	203	65.0
Mining	277	23.8	204	37.3	198	44.4
Manufacturing	817	41.4	605	58.5	585	69.1
Electricity, water, gas and waste services	122	19.7	70	34.3	69	43.5
Construction	440	46.8	358	64.8	364	70.3
Wholesale	336	45.8	295	56.3	286	61.5
Retail trade	301	52.5	263	68.4	262	74.0
Accommodation and food services	308	49.4	269	64.7	266	68.4
Transport, postal and warehousing	452	37.6	323	54.5	340	69.4
Information, media & telecommunications	228	37.7	189	49.7	188	54.3
Financial and insurance services	114	42.1	94	57.4	89	60.7
Rental, hiring and real estate services	149	51.0	129	60.5	129	63.6
Professional, scientific & technical services	295	41.4	246	54.5	245	59.6
Administrative and support services	361	44.9	319	65.8	325	72.6
Health care and social assistance	394	26.9	169	54.4	168	71.4
Arts and recreation services	228	50.0	183	64.5	177	68.9
Other services	275	56.0	214	73.8	210	74.3

Just like in table 4.3, the results in table 4.4 indicate that the $DM_{5\to1}$ algorithm performed better in matching the firms across the industry categories. Further examination of the results reveals that the algorithm outperformed the Caliper and the Nearest Neighbour across all industries with the exception of the Agriculture industry. In the case of Agriculture, the percentage of correctly matched firms is slightly lower (although the number of matches is higher) than that of the Caliper. Apart from two exceptions, namely the Mining and the Electricity, Water, Gas and Waste Services industries, the $DM_{5\to1}$ algorithm correctly matched more than 50 per cent of the firms in each industry.

In order to evaluate further the $DM_{5\to1}$ performance, the authors also partitioned the data to even finer levels and investigated the percentage of correctly matched firms for each industry at each digit level.

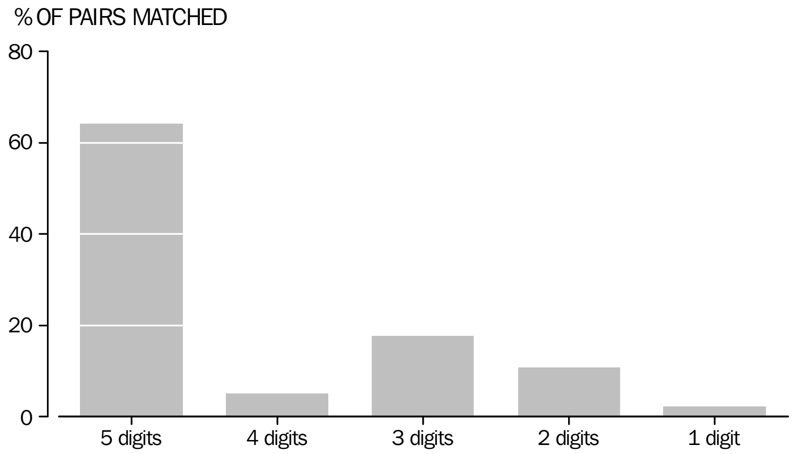
In line with the previous findings, the results in table 4.5 point to the $DM_{5\to1}$ algorithm as the favoured method with respect to both size and industry. In particular, of the firms matched, more than 65 per cent were correctly matched with respect to both size (which included four subcategories) and industry (which in this analysis included 17 subcategories). The most serious mismatches occurred in just under 18 per cent of the cases, when the $DM_{5\rightarrow1}$ failed to correctly match any of the size and industry categories. The other algorithms performed worse, with the Caliper mismatching more than 20 per cent of the firms and the NN more than 30 per cent with respect to both size and industry.

4.5 Breakdown of the matched pairs by industry and business size (numbers and percentages)

	Nearest Neighbour		*Caliper*		$DM_{5\rightarrow1}$	
	Number of firms	*%*	*Number of firms*	*%*	*Number of firms*	*%*
Correctly matched	1,064	39.8	1,197	58.0	1,339	65.3
Only industry is the same	61	2.3	28	1.4	19	0.9
Only the size is the same	708	26.5	421	20.4	329	16.0
Both size and industry differ	841	31.5	418	20.3	365	17.8
Total pairs	2,674		2,064		2,052	

The authors also examined the performance of the $DM_{5\rightarrow1}$ algorithm by looking at the proportion of matches at each digit level. Intuitively, a successful matching would imply that as many matches as possible would occur at the highest digit level, since at this level the algorithm would be more precise. A breakdown of the proportion of matches at each digit level is shown in figure 4.6. It can be noted that by far most of the matches were performed at the highest digit level and that only very few firms were sieved through to the final digit level. This is a positive sign as it indicates that most of the matches were done at the highest specified level of precision.

4.6 Distribution of the matched sample at each digit level (5 to 1 Digit Matching)



In addition to this evaluation, the assessment of matching with respect to size and industry was repeated at each digit level. The aim was to obtain an even finer picture about how successful the algorithm was at matching firms with respect to these two categories at each digit level. The results in Appendixes B.3–4 show that while at the five-digit level the proportion of correct matches by size and by size and industry is higher than 98 per cent, the results deteriorate at the lower digit levels.

4.2 Cautionary notes

Before concluding the section, it is worth listing a few cautionary notes regarding the PSM and its implementation in this study. First, as stated in Rubin (2006), propensity score matching stochastically balances the observed variables only, without adjusting for the effects of the unobserved covariates. Although the authors took all necessary steps to ensure that this was the case, there is no guarantee that some relevant variables could have been omitted. It is worth noting that the data used in the study was based on a survey where not all possible firm characteristic questions were asked to avoid provider overload.

Second, this paper has made a trade-off between including more firms in the matched sample and achieving matching precision. As already mentioned, based on the evaluation results, the $DM_{5\to1}$ was deemed most successful for this case study and although the algorithm performed pretty well at the five digit level some poorer matches were done at the lower levels. The study could have used the most precise matches but then this would significantly limit the sample size. (In fact only around 64% of the sample will be retained.) As such a trade-off was made to maximise precision conditional on also having a good sample size and coverage, as usually done in practice.

Third, there are some limitations regarding the actual matching of the propensity scores. This paper considered a one-to-one matching, which means that some units, and therefore some of the useful information from the non-assisted firms, could have been excluded in the analysis. An alternative is to perform a one-to-many matching. However, using more control units comes with the cost of poor matches (Stuart, 2010). Under the same point, another extension would be to use matching with replacement. Note however that as demonstrated by Dehejia and Wahba (2002), matching with replacement is particularly recommended in cases where the number of treated units is larger than that of the control units – this is not the case in this study. This could be investigated in any future study.

4.3 Summary

In this section, propensity score matching was implemented for the purpose of creating a matched dataset by balancing the distributions of the observed covariates across the two groups of firms, i.e. those which received government assistance and those that did not. Once the propensity scores were estimated, three matching algorithms were considered for constructing pairs, namely the Nearest Neighbour (NN), the Caliper, and the sequential 5 to 1 Digit Matching ($DM_{5\to1}$). Two important conditions for PSM were verified and different diagnostics were performed to investigate the performance of the three algorithms and of the propensity score matching. From these investigations some conclusions and remarks are worth noting.

First, the paper found that among the three algorithms applied, the $DM_{5\to1}$ was most successful. This was supported by all the diagnostic tests performed as well as by the micro assessments of the matched pairs. In the light of this, the matched sample from the $DM_{5\to1}$ algorithm was used in the model application explained in the following section.

Second, based on the results of the investigations conducted (i.e. with respect to the $DM_{5\to1}$ algorithm), the paper found that PSM was successful, at least for the scope of this analysis, in balancing the observed covariates distributions across the two groups of government assisted and non-assisted firms. This was reflected in the chi-square results which indicated that after matching, the differences between the two groups, with respect to the selected covariates, were not significantly different from zero. The standardised bias test and the micro assessments also supported this finding.

Third, it is worth noting the importance of conducting micro assessments on the matched pairs. The assessment could be considered as an alternative tool of visually inspecting the results of the PSM with respect to covariates of key importance to the analyst. For example, in this case study, the authors checked, amongst others, for the presence of serious mismatches on size and industry, two variables which were considered important for the application. The authors also used the micro assessment to get a better understanding of what was happening at each digit level. It was interesting to note how the trade-off between precision and the inclusion of a larger matched sample unfolded at each digit. As in any econometric application, visualisation of data before and after matching should play an important role.

5. MODEL APPLICATION

This section uses the matched sample (derived using the $DM_{5\to 1}$ algorithm) to model the impact of government assistance on innovation. Note that as it has been mentioned the sample was matched in order to control for the selection bias of receiving government assistance.

The definition of "innovation" follows the *Oslo Manual* as

> "… the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation or external relations." (OECD, 2005, p. 46)

In addition to government assistance the following business characteristics were included as explanatory variables: ICT intensity, number of employees, industry division, market competition, foreign ownership, other finance, R&D agreement, exporting activity, and flexible working arrangements.

In estimating the innovation model, this study followed the approach outlined in Agresti (2002) and implemented a probit regression model with a random effects component on the matched pairs – which is an example of a generalised linear mixed model (GLMM) for binary matched pairs. The random effects component was included to control for the pair effect, as businesses within matched pairs are expected to be highly correlated. Although, in the case of a one-to-one match, the random effect is expected to be insignificant, its importance is more evident when dealing with large pairs, which is the case when there is a one-to-many matching. To assess the impact of matching on the regression results, a separate probit model was run on the unmatched sample, whose results were then compared to those coming from the model run on the matched sample.

This section is organised in three parts. The first briefly outlines the theoretical framework of the random effects probit model; the second presents and discusses the results; and the last explains some of the limitations of the empirical application.

The Random Effects (RE) Probit Model

For a firm i, belonging to a matched pair m, where m ranges from 1 to M (the total number of pairs), the RE probit model has the form

$$\begin{aligned} P\left(y_{im}=1 \middle| X_{im},\alpha_m\right) &= P\left(y^*_{im}>0 \middle| X_{im},\alpha_m\right) \\ &= P\left(\alpha_m + X'_{im}\beta + \varepsilon_{im} > 0 \middle| X_{im},\alpha_m\right) \\ &= \Phi\left(\alpha_m + X'_{im}\beta\right) \end{aligned}$$

where the latent variable can be expressed as

$$y_{im}^{*} = \alpha_m + X_{im}'\beta + \varepsilon_{im} \quad , \forall m = 1,\ldots,M \; ; i = 1,2 \; ,$$

and where

y_{im}^{*} is an unobserved binary variable which corresponds to y_{im}, the observed dichotomous variable. In this study, y_{im} takes the value of 1 if the firm innovated and 0 otherwise. The relationship between y_{im} and y_{im}^{*} is shown by

$$y_{im} = \begin{cases} 1 & \text{if } y_{im}^{*} > 0 \\ 0 & \text{otherwise} \end{cases}$$

X_{im} is a vector of observed covariates including a constant term;

β is a vector of fixed, yet unknown, population parameters;

α_m stands for the random component of the matched pair m. As it is common custom, $\{\alpha_m\}$ were assumed to be normally distributed with mean zero and variance σ_α^2 and independent of the error term ε_{im};

ε_{im} is the error term which follows a normal distribution;

$\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

Modelling results

Table 5.1 presents the results of the probit models for innovation. The results of the random effects model are consistent with the expectation of the authors as it was expected that the large number of pairs and the small size of the pairing would lead to small or insignificant pair effects.

In both models, the coefficients for government assistance and three working arrangements variables are positive and highly significant. In line with the results of Todhunter and Abello (2011), the ICT categories are all highly significant. All the market competition categories are highly significant, which is consistent with the results of Soames *et al.* (2011).

5.1 Results of Probit models for innovation (matched and unmatched samples)

	Random effects (matched)		*Binary Probit (unmatched)*	
Variables	*Coefficient*	*Std error*	*Coefficient*	*Std error*
Intercept	-0.150	0.119	-0.089	0.084
Government assistance				
Not received government assistance				
Received government assistance	0.148 **	0.042	0.122 **	0.036
ICT intensity				
Most intense				
High	-0.260 **	0.059	-0.231 **	0.045
Mid	-0.297 **	0.063	-0.352 **	0.048
Low	-0.596 **	0.070	-0.700 **	0.048
Least intense	-0.823 **	0.133	-0.946 **	0.075
Number of employees				
0–4 employees	0.006	0.077	-0.049	0.045
5–19 employees				
20–199 employees	0.143 *	0.064	0.089 *	0.045
200+ employees	0.026	0.071	-0.037	0.052
Industry division				
Manufacturing				
Agriculture, forestry and fishing	-0.168	0.117	-0.111	0.095
Mining	-0.249 *	0.112	-0.251 **	0.080
Electricity, water, gas and waste services	-0.277	0.171	-0.260 *	0.113
Construction	-0.090	0.093	-0.142 *	0.071
Wholesale	-0.076	0.100	-0.004	0.070
Retail trade	0.047	0.105	-0.026	0.074
Accommodation and food services	-0.242 *	0.103	-0.158 *	0.074
Transport, postal and warehousing	-0.243 **	0.094	-0.293 **	0.075
Information, media and telecommunications	0.054	0.120	-0.121	0.083
Financial and insurance services	0.263	0.176	-0.021	0.097
Rental, hiring and real estate services	-0.321 *	0.131	-0.194 *	0.084
Professional, scientific and technical services	-0.081	0.107	-0.136	0.073
Administrative and support services	-0.186	0.099	-0.248 **	0.073
Health care and social assistance	0.041	0.124	-0.049	0.083
Arts and recreation services	-0.019	0.120	-0.015	0.079
Other services	-0.075	0.111	-0.137	0.081
Market competition				
No effective competition				
1–2 competitors	0.324 **	0.094	0.351 **	0.065
3–4 competitors	0.323 **	0.087	0.350 **	0.060
5 or more competitors	0.281 **	0.076	0.347 **	0.051
Foreign ownership				
100% Australian owned				
Foreign ownership > 0% to 50%	0.004	0.112	-0.058	0.088
Foreign ownership > 50%	0.045	0.075	0.068	0.058
Other finance				
No debt or equity finance				
Seek debt or equity finance	0.198 **	0.045	0.242 **	0.035
R&D agreement				
No joint R&D (co-operative) agreement				
Joint R&D (co-operative) agreement	0.395 **	0.084	0.416 **	0.063
Export activity				
Non-exporter				
Exporter	0.079	0.062	0.097 *	0.048
Flexible working hours arrangement				
No flexible work hours				
Flexible work hours	0.305 **	0.053	0.308 **	0.037
Flexible leave arrangement				
No flexible leave				
Flexible leave	0.062	0.054	0.075 *	0.038
Job sharing arrangement				
No job sharing				
Job sharing	0.161 **	0.054	0.188 **	0.041
Working from home arrangement				
No working from home				
Work from home	0.143 **	0.052	0.110 **	0.038
Sigma	0.002	0.017		
Log likelihood	-2441.4		-4799.2	
Percent of correctly predicted	70.3		73.4	
AIC	4956.9		9672.4	
Observations (n)	4,093		8,125	

**=significant at the 0.01 level; *=significant at the 0.05 level

Overall, the results of the regression run on the matched sample are not too different from those of the model run on the unmatched sample. Although similar, there are some changes in the magnitude of the coefficients for most of the key business characteristics. For example, the coefficients for most of the ICT intensity indicators and for all the business sizes categories are higher for the matched sample. There are also some changes in the sign, but mostly in the industry division categories. Changes in the significance are observed for a few industry division categories, flexible leave arrangement, and export activity. Also, the magnitude of the coefficient for government assistance, which is positive and highly significant, is relatively higher for the matched sample than for the unmatched sample. Note that the log likelihood shows an improvement in the model fitness for the matched sample, although it has marginally lower predicted power.

To complement the above analysis and to provide additional indication of the effects associated with government assistance, the authors estimated marginal effects (by number of employees) for a reference firm that is 100% Australian owned, belongs to the manufacturing sector, has low ICT intensity, with no debt or equity finance, no effective competition, no cooperative R&D, no flexible working arrangements and is non-exporting. The results in table 5.2 show that the absolute difference in the marginal effects between government assisted and non-assisted firms has increased by approximately 1 per cent following the implementation of the PSM.

5.2 Impact of receiving government assistance on the probability of innovation, by business size

	Binary Probit (unmatched)			*Random effects (matched)*		
	NG	*G*	*Difference*	*NG*	*G*	*Difference*
Business size						
1–4 employees	20.1%	23.7%	3.6%	23.0%	27.7%	4.7%
5–19 employees	21.5%	25.2%	3.7%	22.8%	27.5%	4.7%
20–199 employees	24.2%	28.2%	4.0%	27.3%	32.5%	5.1%
200+ employees	20.5%	24.1%	3.6%	23.6%	28.4%	4.8%

NG — No Government assistance
G — Received Government assistance

Summary

In this section, the study used the matched sample – so as to address the selection bias of receiving government assistance – to model the effect of government assistance on innovation while accounting for some other key business characteristics. This was done by running a probit model with a random effect component on the matched pairs – this component was included to control for the correlation within matched firms. An ordinary probit model was also considered for the unmatched sample. The following were noted: first, the random effect was not significant, indicating that the pair effects were small. Second, when comparing the results of the model run on the matched sample to those run on the unmatched sample, overall, the results were not too different, although there were some changes in the magnitudes and signs of the estimated coefficients for some key business characteristics.

When taking into account the above results, there are some limitations that need to be considered:

i. The estimated innovation model is based on a sample of businesses and therefore the results should not be generalised to the whole population of Australian businesses. A possible extension would be to apply survey weights to the models in order to obtain population estimates.

ii. The analysis is limited to the selected business characteristics available within the BCS framework. A possible extension would be to include other indicators coming from the administrative tax data that could be linked to the BCS results.

6. CONCLUDING REMARKS

This paper investigated the application of propensity score matching (PSM) to construct a matched sample data so as to control for the selection bias of receiving government assistance. The PSM was implemented by balancing the distributions of the observed covariates across two groups of firms, those which received government assistance and those that did not.

The paper found that among the three matching algorithms considered, namely the Nearest Neighbour (NN), the Caliper, and the sequential 5 to 1 Digit Matching ($DM_{5\rightarrow 1}$), the last method was the most successful. In addition, the paper has demonstrated the importance of investigating the quality of the matched results by implementing micro assessments as an alternative tool for visually inspecting the PSM results.

Once the matching was successfully implemented, the paper used the matched sample to model the impact of government assistance on the firm's propensity to innovate. This was achieved by using the generalised linear mixed model for binary matched pairs and the standard binary probit model. The modelling found a statistically significant and positive association between government assistance and innovation.

ACKNOWLEDGEMENTS

The authors would like to thank the important contributions of the following ABS staff: Ruel Abello, Diane Braskic, Melissa Gare and Dr. Siu-Ming Tam. Comments from the ABS Innovation and Technology Statistics Branch staff are also acknowledged. The authors are also grateful to Professors Ray Chambers and Robert Clark of the University of Wollongong for their comments and advice. Thanks also go to Peter Rossiter for his help in editing and formatting the paper. Any remaining errors and omissions are the responsibility of the authors.

REFERENCES

Agresti, A. (2002) *Categorical Data Analysis (2nd edition)*, John Wiley & Sons, Inc., New Jersey.

Almus, M. and Czarnitzki, D. (2003) "The Effects of Public R&D Subsidies on Firms' Innovation Activities: The Case of Eastern Germany", *Journal of Business and Economic Statistics*, 21(2), pp. 226–236.

Australian Bureau of Statistics (2011a) *Summary of IT Use and Innovation in Australian Business, 2009–10*, cat. no. 8166.0, ABS, Canberra.

—— (2011b) *Selected Characteristics of Australian Business, 2009–10*, cat. no. 8167.0, ABS, Canberra.

—— (2010) *Innovation in Australian Business, 2008–09*, cat. no. 8158.0, ABS, Canberra.

Caliendo, M. and Kopeinig, S. (2008) "Some Practical Guidance for the Implementation of Propensity Score Matching", *Journal of Economic Surveys*, 22(1), pp. 31–72.

Coca-Perraillon, M. (2007) "Local and Global Optimal Propensity Score Matching", *Proceedings of the SAS Global Forum 2007 Conference*, 16–19 April 2007, Paper 185-2007, Orlando, Florida.

Coca-Perraillon, M. (2006) *Matching with Propensity Scores to Reduce Bias in Observational Studies*, Paper presented at the 19th Annual NESUG Conference, 17–20 September 2006, Philadelphia, PA.

Dehejia, R. and Wahba, S. (2002) "Propensity Score-Matching Methods for Nonexperimental Causal Studies", *Review of Economics and Statistics*, 84(1), pp. 151–161.

Dockery, A. (2005) *Assessing the Value of Additional Years of Schooling for the Non Academically Inclined*, Longitudinal Surveys of Australian Youth, Research Report No. 38. (Last viewed on 12 February 2013) <http://research.acer.edu.au/lsay_research/42>.

Essana-Nssah, B. (2006) *Propensity Score Matching and Policy Impact Analysis*, World Bank Policy Research Working Paper 3877, The World Bank, Washington, D.C.

Foreman-Peck, J. (2010) *Effectiveness and Efficiency of SME Innovation Policy* (Last viewed on 12 February 2013) <http://www.isbe.org.uk/content/assets/C._Business_Support_Policy_and_Practice-_James_Foreman-Peck.pdf>.

Heijs, J. and Herrera, L. (2004) *The Distribution of R&D Subsidies and its Effect on the Fiscal Outcome of Innovation Policy*, Working Paper No. 46, Institute for Industrial and Financial Analysis, University Completense of Madrid, Madrid.

Heinrich, C.; Maffioli, A. and Vázquez, G. (2010) *A Primer for Applying Propensity-Score Matching*, Impact Evaluation Guidelines, Strategy Development Division, Technical Notes No. IDB-TN-161, Inter-American Development Bank, Washington, D.C.

Herrera, L. and Nieto, M. (2006) *The Regional Dimension of the Distribution and Effects of Public Incentives directed towards Innovation of Firms*, Working Paper 03-06, Departamento de Direccion y Economica de la Empresa, Universidad de Leon, Leon.

Holland, P.W. (1986) "Statistics and Causal Inference", *Journal of the American Statistical Association*, 81(396), pp. 945–960.

Houssard, C.; Sastro, A. and Hardy, S. (2010) *The Impact of HECS Debt on Socioeconomic Inequality and Transition to Adulthood Outcomes*, School of Economics Working Paper, University of New South Wales.
(Last viewed 12 February 2013)
<http://www.melbourneinstitute.com/downloads/hilda/Bibliography/Working+Discussion+Research_Papers/2010/Houssard_etal_impact_of_HECS_debt.pdf>.

OECD (2005) *The Measurement of Scientific and Technological Activities: Guidelines for Collecting and Interpreting Innovation Data: Oslo Manual*, Third Edition, prepared by the Working Party of the National Experts on Scientific and Technology Indicators, OECD, Paris.

Parsons, L.S. (2004) "Performing a 1:N Case Control Match on Propensity Score", *Proceedings of the Twenty-Ninth SAS Users Group International Conference*, Montreal, Canada.

Parsons, L.S. (2001) "Reducing Bias in a Propensity Score Matched-Pair Sample Using Greedy Matching Techniques", *Proceedings of the Twenty-Sixth Annual SAS Users Group International Conference*, Cary, NC.

Rosenbaum, P. (2002) *Observational Studies (Second edition)*, Springer-Verlag, New York.

Rosenbaum, P. and Rubin, D.B. (1985) "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score", *American Statistician*, 39(1), pp. 33–38.

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

Rubin, D.B. (2006) *Matched Sampling for Casual Effects*, Cambridge University Press, New York.

Soames, L.; Brunker, D. and Talgaswatta, T. (2011) "Competition, Innovation and Productivity in Australian Businesses", *Methodology Research Papers*, cat. no. 1351.0.55.035, Australian Bureau of Statistics and Productivity Commission, Canberra.

Stuart, E.A. (2010) "Matching Methods for Causal Inference: A Review and a Look Forward", *Statistical Science*, 25(1), pp. 1–21.

Todhunter, J. and Abello, R. (2011) "Business Innovation and the Use of Information and Communications Technology", *Methodology Research Papers*, cat. no. 1351.0.55.033, Australian Bureau of Statistics, Canberra.

Todd, P. (2006) "Matching Estimators", *Palgrave Dictionary of Economics*, Palgrave Macmillan, U.K..

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.

APPENDIXES

A. DATA COMPILATION

This section describes how government assistance, innovation and selected key business characteristics variables have been constructed for the analysis.

Government assistance

In this paper the term 'government assistance' refers to any financial assistance, in the form of grants, on-going funding, subsidies, tax concessions, rebates, as well as others, received by the business from Australian government organisations. The government organisations included federal, state, territory and local government. Financial assistance relating to: employment (e.g. apprenticeships); starting and expanding the business; R&D, innovation and/or exporting; and hardship (e.g. drought) are also included in the assistance.

A binary government assistance variable has been constructed as:

Description	*Range of values*
Government assistance (binary) *Firm received/not received any form of financial assistance (grants, on-going funding, tax concession, subsidies, rebates, other government financial assistance)*	0/1 dummy

Innovation

The scope of innovative activity, as measured by the BCS, follows the Oslo Manual (OECD, 2005) and covers four broad types of innovation:

- *Goods or services* – Any good or service or combination of these which is new to a business (or significantly improved). Its characteristics or intended uses differ significantly from those previously produced/offered.
- *Operational processes* – New or significantly improved methods of producing or delivering goods or services of a business (including significant change in techniques, equipment and/or software).
- *Organisational/managerial processes* – New or significantly improved strategies, structures or routines of a business which aim to improve performance.
- *Marketing methods* – New or significantly improved design, packaging or sales methods aimed to increase the appeal of goods or services of a business or to enter new markets.

There are three statuses of innovation, namely:

- *Introduced or implemented* – the business successfully introduced or implemented an innovation during the reference period (although the innovation does not need to have been commercially successful).
- *Still in development* – the business was in the process of developing, introducing or implementing an innovation during the reference period but work on the innovation was still in progress at the end of the period.
- *Abandoned* – the business abandoned the development and/or introduction of an innovation during the reference period (i.e. work on the innovation ceased without full introduction occurring).

A business is called 'innovation-active' if it engaged in any innovation activities that were implemented, still in development or abandoned during the period. Note that in the BCS, businesses could report more than one type of innovation.

The empirical application in Section 4 investigated the likelihood of a business to engage in any innovation activity, hence a binary variable was constructed as

Description	*Range of values*
Innovation (binary) *Firm engaged / not engaged in any types of innovation*	0/1 dummy

Selected key business characteristics

The study followed Todhunter and Abello (2011) for the inclusion and creation of the key business characteristics. There were five more variables included. The first is an equity/finance variable which indicates whether the firm sought debt or equity finance during the financial year. Debt finance includes any finance that the business must repay, while equity finance includes any finance which is provided in exchange for a share in the ownership of the business. The other four are dichotomous variables for flexible working hours, flexible leave arrangements, job sharing, and working from home.

The selected key business characteristics employed in the modelling are described below.

Description	*Range of values*
Number of employees *0–4 Employees* *5–19 Employees* *20–199 Employees* *200+ Employees*	0/1 dummy (each category)
Degree of competition in the market *No effective competition* *1–2 competitors* *3–4 competitors* *5 or more competitors*	0/1 dummy (each category)
Degree of foreign ownership *100% Australian owned* *> 0% to 50% foreign owned* *> 50% foreign owned*	0/1 dummy (each category)
Sought any debt and equity finance	0/1 dummy
Business was involved in co-operative arrangement for joint research and development (R&D)	0/1 dummy
Business received income from directly exporting goods and/or services	0/1 dummy
Industry division (Based on ANZSIC 2006) *Agriculture, forestry and fishing* *Mining* *Manufacturing* *Electricity, water, gas and waste services* *Construction* *Wholesale* *Retail trade* *Accommodation and food service* *Transport, postal and warehousing* *Information, media and telecommunications* *Financial and insurance services* *Rental, hiring and real estate services* *Professional, scientific and technical services* *Administrative and support services* *Health care and social assistance* *Arts and recreation services* *Other services*	0/1 dummy (each category)
ICT intensity *Most intense* – Business had broadband connection, web presence, places and receives orders via the internet *High* – Business had broadband connection, web presence, and only places orders via the internet *Mid* – Business had broadband connection, web presence, but does not place/receive orders via the internet *Low* – Business had broadband connection, but has no web presence *Least intense* – Business does not use broadband connection	0/1 dummy (each category)
Flexible working hours – Business offered flexible working arrangements regarding working hours (e.g. to enable employees to deal with non-work issues). Selection of own roster or shifts is also included	0/1 dummy

Flexible leave arrangement – Business offered flexible working arrangements regarding the use of leave, which include the employee's ability to buy extra annual leave, cash out annual leave, take leave without pay, access paid parental leave, and flexibility on the use of personal sick, unpaid or compassionate leave (e.g. to take care for other people who are sick)	0/1 dummy
Job sharing – Business allowed staff to share job	0/1 dummy
Working from home – Business allowed staff to work from home	0/1 dummy

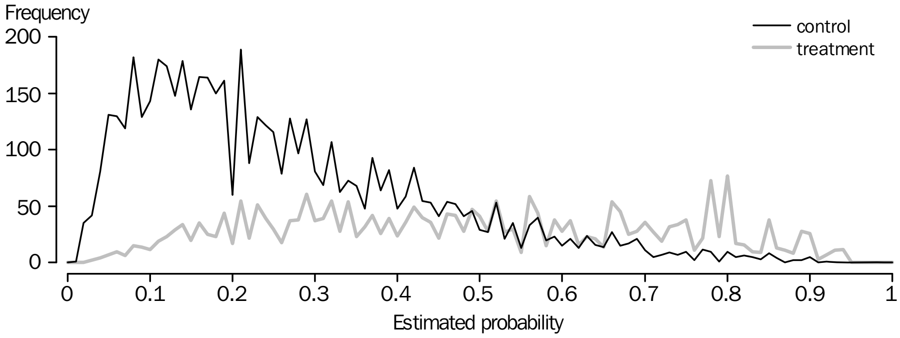
B. OTHER PSM DIAGNOSTIC RESULTS

B.1 Government probit model results for the PSM

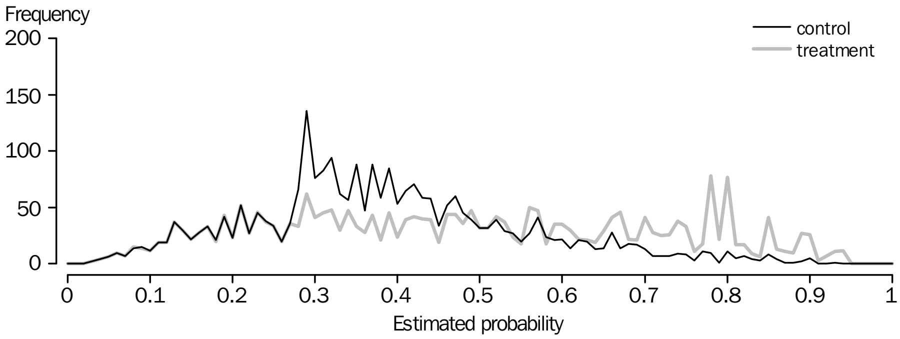
Variables	*Coefficient*	*Standard error*	*Pr > ChiSq*
Intercept	–0.547	0.084	<.0001
ICT intensity			
Most intense			
High	–0.005	0.045	0.906
Mid	–0.213	0.048	<.0001
Low	–0.329	0.052	<.0001
Least intense	–0.488	0.086	<.0001
Number of employees			
0–4 employees	–0.265	0.052	<.0001
5–19 employees			
20–199 employees	0.277	0.047	<.0001
200+ employees	0.896	0.051	<.0001
Industry division			
Manufacturing			
Agriculture, forestry and fishing	0.387	0.094	<.0001
Mining	–0.032	0.083	0.702
Electricity, water, gas and waste services	0.211	0.119	0.077
Construction	0.134	0.073	0.067
Wholesale	–0.444	0.072	<.0001
Retail trade	–0.468	0.078	<.0001
Accommodation and food services	–0.125	0.077	0.108
Transport, postal and warehousing	0.239	0.074	0.001
Information, media and telecommunications	–0.288	0.085	0.001
Financial and insurance services	–0.847	0.110	<.0001
Rental, hiring and real estate services	–0.418	0.092	<.0001
Professional, scientific and technical services	–0.579	0.076	<.0001
Administrative and support services	–0.358	0.076	<.0001
Health care and social assistance	0.492	0.085	<.0001
Arts and recreation services	–0.251	0.084	0.003
Other services	0.054	0.084	0.525
Market competition			
No effective competition			
1–2 competitors	–0.015	0.069	0.830
3–4 competitors	–0.013	0.063	0.837
5 or more competitors	–0.079	0.055	0.154
Foreign ownership			
100% Australian owned			
Foreign ownership > 0% to 50%	–0.097	0.085	0.251
Foreign ownership > 50%	–0.258	0.057	<.0001
Other finance			
No debt or equity finance			
Seek debt or equity finance	0.244	0.035	<.0001
R&D agreement			
No joint R&D (co-operative) agreement			
Joint R&D (co-operative) agreement	0.424	0.058	<.0001
Export activity			
Non-exporter			
Exporter	0.159	0.047	0.001
Log likelihood	-4231.9		
Pseudo R-squared	0.2197		
Percent correctly predicted	77.4		
AIC	8527.8		
Observations (n)	8,160		

B.2 Comparisons of propensity score distributions

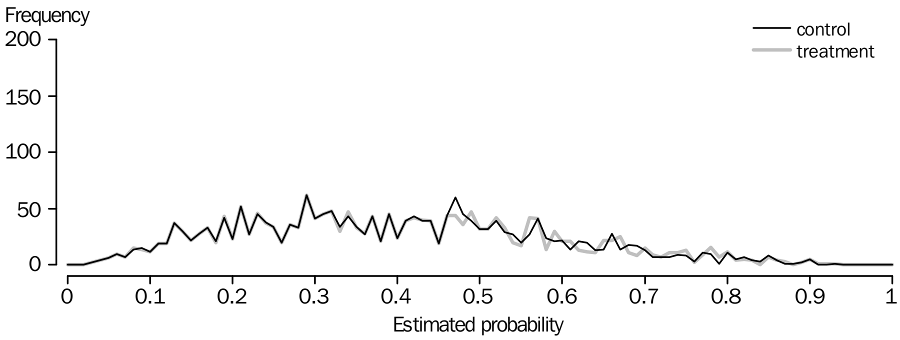
BEFORE MATCHING



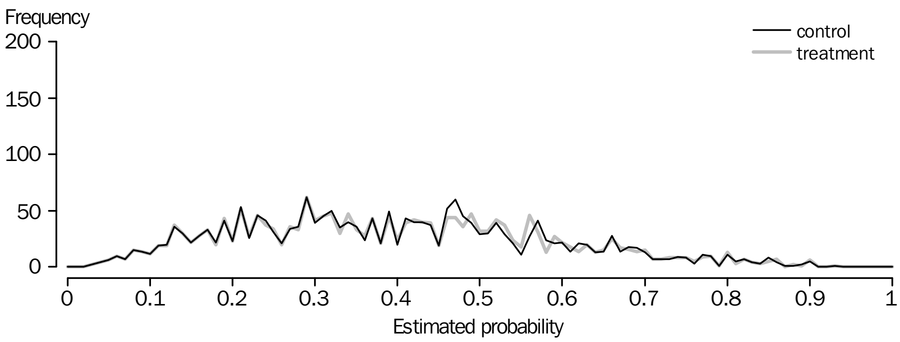
NEAREST NEIGHBOUR



CALIPER



5 TO 1 DIGIT MATCHING



B.3 Breakdown of the matched pairs for the 5 to 1 Digit Matching by business size*

	5 digits		*4 digits*		*3 digits*		*2 digits*		*1 digit*	
	Number of firms	*%*	*Number of firms*	*%*	*Number of firms*	*%*	*Number of firms*	*%*	*Number of firms*	*%*
Correctly matched										
micro to micro	201	15.3	5	4.9	13	3.6	1	0.5	0	0.0
small to small	244	18.5	8	7.8	16	4.4	2	0.9	0	0.0
medium to medium	409	31.0	9	8.7	42	11.6	25	11.2	3	6.7
large to large	448	34.0	21	20.4	101	27.9	91	40.6	29	64.4
Total	1,302	98.8	43	41.8	172	47.5	119	53.1	32	71.1
Not correctly matched										
micro to small	1	0.1	10	9.7	30	8.3	10	4.5	0	0.0
micro to medium	2	0.2	5	4.9	15	4.1	5	2.2	0	0.0
micro to large	1	0.1	3	2.9	7	1.9	1	0.5	1	2.2
small to medium	4	0.3	14	13.6	38	10.5	22	9.8	0	0.0
small to large	1	0.1	8	7.8	22	6.1	13	5.8	2	4.4
medium to large	7	0.5	20	19.4	78	21.6	54	24.1	10	22.2
Total	16	1.2	60	58.3	190	52.5	105	46.9	13	28.9

* micro (0–4 employees); small (5–19 employees); medium (20–199 employees); large (200+ employees)

B.4 Breakdown of the matched pairs for the 5 to 1 Digit Matching (Business size and Industry)

	5 digits		*4 digits*		*3 digits*		*2 digits*		*1 digit*	
	Number of firms	*%*	*Number of firms*	*%*	*Number of firms*	*%*	*Number of firms*	*%*	*Number of firms*	*%*
Correctly matched	1,292	98.0	2	1.9	24	6.6	14	6.3	7	15.6
Only the industry is the same	0	0.0	9	8.7	4	1.1	5	2.2	1	2.2
Only the size is the same	10	0.8	41	39.8	148	40.9	105	46.9	25	55.6
Both size and industry differ	16	1.2	51	49.5	186	51.4	100	44.6	12	26.7
Total pairs	1,318		103		362		224		45	

FOR MORE INFORMATION . . .

INTERNET **www.abs.gov.au** The ABS website is the best place for data from our publications and information about the ABS.

LIBRARY A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our website for a list of libraries.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website, or purchase a hard copy publication. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS www.abs.gov.au