VATT-TUTKIMUKSIA
106
VATT RESEARCH REPORTS

Takis Venetoklis *

AN EVALUATION OF WAGE SUBSIDY PROGRAMS TO SMEs UTILISING PROPENSITY SCORE MATCHING

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Helsinki 2004

* I would like to thank Aki Kangasharju and Jaakko Kiander for valuable comments and suggestions.

ISBN 951-561-480-5

ISSN 0788-5008

Valtion taloudellinen tutkimuskeskus

Government Institute for Economic Research

Arkadiankatu 7, 00100 Helsinki, Finland

Email: etunimi.sukunimi@vatt.fi

Oy Nord Print Ab

Helsinki, January 2004

VENETOKLIS, TAKIS: AN EVALUATION OF WAGE SUBSIDY PROGRAMS TO SMEs UTILISING PROPENSITY SCORE MATCHING. Helsinki, VATT, Valtion taloudellinen tutkimuskeskus, Government Institute for Economic Research, 2004, (B, ISSN 0788-5008, No 106). ISBN 951-561-480-5.

Abstract: We evaluate direct wage subsidy programs to Finnish SMEs utilising as the unit of interest the firm and its characteristics. In estimating Average Treatment effects on the Treated firms (ATT) we apply recent techniques of propensity score matching algorithms developed by Becker and Ichino (2002) together with Difference in Differences (DiD) estimations. We define our outcome indicator as the net salary costs growth of the treated firms for up to two years after the first receipt of the wage subsidy. We find that, on average, the subsidised firms compared to the non-subsidised ones have higher payroll costs during the first year after the receipt of wage subsidies. Also, the first post-subsidies-year average net payroll growth comes out larger than the estimated average wages spent for the subsidised workers only. This may indicate that firms employ the subsidised workers for at least one year after the subsidised period and perhaps employ, in addition, extra staff. Although still positive, the net aggregate salary costs growth measured two years after the receipt of subsidies diminishes and seems not to cover all of the subsidised workers' salary costs. One could interpret these results as initial signs of lay-offs of the subsidised workers and/or of the extra staff employed the previous period. From a pure economic efficiency as well as a distributive justice point of view this two year employment spell for the previously unemployed could be considered as an adequate program effect. On the contrary, the diminishing salary costs growth already during the second year after the subsidies receipt might indicate non-sustainable positive wage subsidies effects even on a short term basis.

Key words: Evaluation, employment, wage subsidies, estimation methods
JEL: J23, C8

Tiivistelmä: Tutkimuksessa arvioidaan pk-yrityksille annettujen suorien työllistämistukien vaikuttavuutta Suomessa. Tuen keskimääräisen vaikutuksen (ATT) arvioimisessa sovelletaan Beckerin ja Ichinon (2002) kehittämää Propensity Score Matching -tekniikkaa yhdessä Difference in Differences (DiD) estimoinnin kanssa. Tukien vaikuttavuutta tarkastellaan palkan kehityksen suhteen yhden ja kahden vuoden aikaperiodilla tuen saamisesta. Tutkimuksessa havaitaan, että tuettujen yritysten palkkakustannusten kasvu on suurempi ensimmäisenä vuonna tuen saamisen jälkeen kuin yritysten, jotka eivät ole saaneet tukea. Havaitun palkkasumman kasvun voidaan arvioida myös ylittävän tuetuille työntekijöille maksettujen keskimääräisten palkkojen määrän. Tätä tulosta voidaan tulkita siten, että tuetut työpaikat säilyvät keskimäärin vähintään vuoden ajan ja tämän lisäksi tukea saaneet yritykset työllistävät ensimmäisen vuoden aikana muuta henkilöstöä. Vaikka ero palkkakustannusten kasvussa on edelleen positiivinen, se kuitenkin pienenee toisena vuonna tuen saamisesta eikä enää kata tuettujen työpaikkojen kaikkia palkkakustannuksia. Tämä tulos saattaa viitata tuettujen työntekijöiden ja/tai ensimmäisenä vuonna työllistetyn lisähenkilöstön irtisanomiseen. Tukien havaittu väliaikainen työllisyysvaikutus voi olla sekä taloudellisesta että sosiaalisen oikeudenmukaisuuden näkökulmasta riittävä puoltamaan tukien käyttöä työllisyyspolitiikan keinona. Toisaalta, palkkakustannusten kasvun pieneneminen jo toisena vuotenna osoittanee, että työllistämistukien positiivinen vaikutus on vain väliaikainen.

Asiasanat: Evaluaatio, työllisyys, työllistämistuet, estimointimenetelmät

Contents

1. Introduction

The evaluation of government program interventions towards private firms has the last few years gradually become an important activity integrated in the whole program process. Legislature emphasizes the importance for such actions in order to make sure that public funds have been spent wisely and have achieved what they were initially designed to do. In English speaking countries evaluation of such programs has been implemented for several decades. In mainland Europe however, only during the last 15 years we note a boost of such activities, especially with the creation of supranational financial instruments. For example, the European Union structural funds allocate approximately 2% of their total appropriations for evaluation purposes.

With the obligation to evaluate we see in parallel, new developments in the analytical methods applied when evaluating such programs. In particular when it comes to social programs towards the unemployed (be it training, advice for re-employment, direct wages subsidies, subsidies for starting own private business, etc) pioneer analytical techniques have been devised to try and measure quantitatively with the least possible bias the true impacts of such programs. In the majority of cases the unit of interest has been the individual person and measurements are based on variables which describe his characteristics and his activities before and after the intervention of the program of interest.

One such method of analysis recently developed is the so called propensity score matching which its proponents claim produces as good estimates as other more well known techniques (Ordinary Least Squares, Instrumental Variables, Difference in Difference, etc.). It was introduced by Rubin and Robenbaum (1983) in connection to epidemiological observational studies and medical experiments[1].

During the 1990s attempts were made to transfer the logic of propensity score matching techniques to social program evaluation and the result was a number of seminal papers dealing with Active Labour Market Programs (ALMP) evaluations (e.g. Dehejia and Wabba (1999, 2002), Heckman, Ichimura, and Todd (1997, 1998), Heckman, Lalonde and Smith (1999), Sianesi (2002), Smith and Todd (2003), Lechner (2002), Kluve, Lehmann and Schmidt (2002). In all these papers, in addition to the method applied, the other common feature was that the unit of interest was the individual person and his characteristics.

Nevertheless, we have not found many studies which evaluate government programs to firms, apply propensity matching techniques, and at the same time use at the unit of interest the firm itself and her observable characteristics. One

[1] In medical related research there is plethora of papers using matching techniques. See for example a bibliography compiled by Love (2002).

rare example is the paper by Czarnitzki and Fier (2002). The purpose of this paper is to fill this gap. We evaluate direct employment (wage) subsidy programs to Finnish Small and Medium size Enterprises (SMEs) utilising as the unit of interest the firm and her characteristics.

To assist researchers in applying propensity score matching techniques for policy evaluations, several statistical tools have been developed in recent years. For example, Bergstrahl et al. (1996) have developed such code for the SAS statistical software programme; Abadie et al. (2003), Becker and Ichino (2002) and Sianesi (2003) are authors of add-on functions used with STATA statistical software. In this paper we analyse our data by applying Becker and Ichino's "regression like" commands `pscore` and `att*`. In addition, because of the functional form of our outcome variable our final estimate is also based on Difference in Differences (DiD) measurements.

We find that, on average, the subsidised firms compared to the non-subsidised ones have higher payroll costs during the first year after the receipt of wage subsidies. This first post-subsidies-year average net payroll growth comes out larger compared to the estimated average wages spent for the subsidised workers only. This in turn may indicate that firms employ the subsidised workers for at least one year after the subsidised period and perhaps hire, in addition, extra staff. On the other hand, the aggregate salary costs growth measured two years after the receipt of subsidies diminishes and seems not to cover all of the subsidised workers' salary costs. They are nonetheless still positive. We interpret these latter results as initial signs of lay offs of the subsidised workers and/or of those that were recently employed the previous period.

We do not know whether explicit quantitative targets for employment have been imposed beforehand. From a pure economic efficiency point of view this two year employment spell may be an adequate program effect. Also, we can not disregard the distributive justice dimension of the wage subsidies programs. One could argue that society as a whole might be better off in the long run paying for such programs regardless of their whatever direct economic (in)efficiency. And indeed the subsidised workers themselves may evaluate these programs in a positive manner.

However the decrease of the aggregate salary costs already during the second year after the receipt of subsidies may be a problem on the rise. Wage subsidies are theoretically designed not only to introduce previously unemployed workers once again to the labour market, but at the same time create the conditions and opportunities so that these workers continue to be employed for a prolonged period of time, perhaps within the firm which hired them in the first place.

The paper proceeds as follows. The following section, describes briefly the public programs of direct wage subsidies towards firms that are in force in

Finland nowadays. Section 3 introduces the theoretical basis of our empirical analysis. We initially discuss theoretical aspects of program evaluation and then proceed by introducing the estimation of Average Treatment effects on the Treated units (ATT) based on propensity score matching techniques. Section 4 describes the data set analysed and discusses the logic behind the different dependent and independent variables used in our constructed models. Section 5 proceeds in estimating program impacts using propensity matching and other estimation techniques, comments on the results and concludes.

2. Direct wage subsidy schemes in Finland[2]

Wage subsidies in Finland are delivered on the basis of job specific applications submitted by firms. The Finnish legislation related to subsidies is rather vague in defining which firms should be eligible, but basically it stipulates that the potential recipients should be profitable or should have the prerequisites to become so. Subsidised jobs are designed for the unemployed who cannot find themselves a job or labour market training through the local (Ministry of) Labour office, who are long term unemployed or are facing the threat, or are under 25 year of age. Subsidies are distributed through local Labour offices that appoint unemployed workers to the subsidised jobs. Also, unemployed graduates of adult educational centres are eligible for such subsidised jobs after they complete their obligatory not-payable on the job training period. The purpose of wage subsidies is to improve the human resources development of the unemployed work force as well as to encourage firms to increase employment. In other words, wage-subsidies are directed to firms who employ the kind of unemployed whose productivity and qualifications are lower than the levels needed in active labour markets. These workers are not easily employable with the prevailing minimum wage level of the sector in question. Wage subsidies are used to fill the gap between wages that firms are willing to pay to these people and the prevailing wage level. The subsidies are grants, in that the recipient firm is not obliged to pay the money received back to the distributor.

The wage subsidies are based on an amount of up to approx. €770 per month for up to 10 months (in 2002). On average, however, the length of the subsidised period is 6 months. The level of the worker's human capital in the subsidised job partly determines the exact amount of subsidy. Also, the longer the worker has been unemployed prior the subsidy, the higher the subsidy. Similarly, a lower level of education increases the subsidy.

Workers in subsidised jobs are usually paid according to the prevailing wage rate. As the typical subsidised jobs are for cleaners, clerks, secretaries, office workers, unskilled manufacturing workers and salesmen, we have estimated that, on average, firms pay 60 per cent of the employment payroll of a worker in a subsidised job. That is, for each euro received as subsidy, the firm must on average put in € 1.5 of her own money when creating a subsidised job (1.5/(1+1.5)=0.60).[3] This estimation is based on the centralised union wage agreement.

---

[2] This section is based on Kangasharju and Venetoklis (2003).

[3] According to the Ministry of Labor, the average wage subsidy was €620 a month in 1999. The average gross monthly wage in subsidized jobs was €1,560.

Apart from their subsidised status, these jobs in private firms have exactly the same specifications as the non-subsidised ones. When receiving a wage subsidy the firm must be able to demonstrate that the job is new, the worker has a permanent contract for at least 6 months, and the firm has not laid off workers from similar jobs prior to the subsidy period. However, the subsidised firm is not bind to keep the worker any longer than the 6 month period, nor does she promise to avoid laying off other employees in the future. From the point of view of our analysis, wage subsidies directly affect the payroll and the numbers of personnel, as the subsidies are part of the total payroll firms pay during a financial year.

3. Framework of empirical analysis

3.1 Theoretical aspects of program evaluation

In this paper we attempt to estimate the Average Effect of Treatment on the Treated (ATT). In our current framework the "Treatment" is wage subsidies, the "Treated" are private firms (Finnish SMEs) and the "Effect" is chosen to be the growth of salary costs during the same year of the treatment or within a certain period (subsequent years) after the receipt of the treatment.

In formal notation the ATT is denoted as follows. Let the treatment status of a firm be represented as a Dummy variable D, taking value 1 if the firm receives employment subsidies and 0 if she does not. Let Y_1 and Y_0 denote potential outcomes where 1 denotes treatment and 0 non-treatment. The observed outcome of an individual firm is $Y=DY_1 + (1-D)Y_0$. Thus, what we are attempting to find is actually the effect of the Treatment on the Treated compared to that of not being Treated or

$$E(Y_1 - Y_0) \mid D=1) = E(Y_1|D=1) - E(Y_0|D=1) \qquad [1]$$

The problem is of course that the counterfactual outcome of a treated firm (the $E(Y_0|D=1)$ in *[1]*) is unobservable because a firm can only be treated or non-treated at a specific point in time. This is what Holland (1986) calls the fundamental problem of evaluation. Consequently we need to impose certain assumptions in *[1]* so that ATT can be estimated. One way is to substitute the Expected outcome of the treated firms were they not treated (the $E(Y_0|D=1)$), with the expected outcome of the firms that were indeed not treated or $E(Y_0|D=0)$. However since the distribution of wage subsidies as mentioned earlier is not conducted randomly but follows certain steps and selection procedures, we can not assume that replacing $E(Y_0|D=1)$ *with* $E(Y_0|D=0)$ would give us unbiased estimates because it is very improbable that $E(Y_0|D=1) = E(Y_0|D=0)$.

Nevertheless, if we take into account certain observable characteristics of the selection process as well as characteristics that potentially influence the outcome of interest in the treated firms (both types of characteristics denoted as a vector of X characteristics) then we can rewrite *[1]*] as

$$E(Y_1 - Y_0) \mid D=1, X) = E(Y_1|D=1, X) - E(Y_0|D=0, X) \qquad [2]$$

Thus, one can easily see that in order to get an unbiased estimate of the treatment on the treated one has to identify the non-treated firms to be as similar as possible to the treated ones in terms of their general and behavioural characteristics captured by *X*.

Many methods have been proposed to identify well such counterfactual group which would then produce unbiased estimated of the ATT including Ordinary Least Squares (OLS), Instrumental Variables (IV), and Difference in Differences (DiD). In this paper, for our main estimation method we adopt a set of techniques known as "Matching on observables"; specifically, we apply Propensity Score Matching (PSM) techniques using STATA functions developed recently by Becker and Ichino (2002).

3.2 Estimation based on the matching rationale

3.2.1 Simple matching[4]

With the simple matching technique we take the observed characteristics of the treated units as a base and attempt to find the units in the untreated group which have the *exact* same characteristics as the ones defined in the treated group. The method is initially appealing but there are some practical problems that can prove the attempt fruitless. The reason is simple. The more characteristics one uses as a base, the more untreated observations one needs to be able to find and match them against the treated group.

Thus, as the observable characteristics of the treated group grow in number (horizontally – across) and in sub-categories/strata (vertical – within), the probability of finding an equivalent observation in the non-treated group diminishes quickly even if one has "rich" data in abundance. Just in the general case of binary factors, the number of control units is 2^p, where p is the number of binary characteristics of the treatment units. As the number of observable characteristics in the treated group increases linearly, the number of observations in the control group increases exponentially.

There are also problems in the number of observations within the clusters of equal characteristics. That is, one might be confronted with the following situation, depicted in Figure 1.

[4] This section is based on Venetoklis (2002).

Figure 1. *Simple matching technique: matching combinations between observations of treatment and control groups*

Group 1 (Treated)	Group 2 (Untreated/controls)
Number of unit observations in treated group with certain characteristics	Number of unit observations in untreated group with the same set of characteristics as in Group 1
1	1
>1	>1
	None

Attempting to match the observations of group 1 with the ones of group 2 entails some decisions on the selection criteria for matching. We have the following 6 possibilities:

(1,1), (1, >1), (1,None), (>1,1), (>1,>1), (>1, None).

We must disregard the cases where there are no equivalent units found in group 2; that is *(1,None)* and *(>1,None)*. In the three cases where there are more than one observations in either or both groups *(1, >1), (>1,1), (>1,>1)*, we would take the average of those units for what ever treatment and impact indicator we measure. The most straight forward case of course is the one to one matching *(1,1)*.

Finally, until now we have assumed that the matched variables are categorical. However, a common problem is that the perfect matching conditions just mentioned are even more difficult to achieve when the matched variables are continuous. It is (relatively) easier to find, for example, firms who are in the same sector(s) than find firms who have the same sales figure at a given financial year. We could try and solve this, by defining a margin, say, 10% over or under the respective sales figures of the treated group. Nonetheless, when it comes to firms and individuals the heterogeneity is so vast, that we are most likely to fail unless we limit ourselves to very few matching conditions and characteristics.

3.2.2 Propensity Score Matching[5]

A technique to deal with the aforementioned problems is called Propensity Score Matching (PSM). The propensity score is an indicator (a number) which depicts *the conditional probability of being assigned (or not) to a particular treatment*. By conditional we refer to a set of characteristics *X* that can predict such an

[5] This section is based on Dehejia and Wahba (1998).

assignment. In formal notation the propensity score could be depicted as $P(X) = P(D=1|X)$.

The propensity score replaces the collection of X characteristics in the observational study with just one number based on these characteristics. There after we can use the propensity score just as if it were the only material characteristic of interest.

Rosenbaum and Rubin (1983) prove that in this case, in *[2]* above, X can be substituted for $P(X)$ so that

$$E(Y_1 - Y_0) \mid D=1, P(X)) = E(Y1|D=1, P(X)) - E(Y_0|D=0, P(X)) \qquad [3]$$

This is very important because in practice it is much easier to condition on just one number (the probability of being treated) than on a vector of X characteristics, of which some might be continuous. Put it differently, the collection of X characteristics (creating multiple conditional dimensionality) is collapsed into a single composite number.

To ensure a so called "common support", i.e. that there are both treated and non-treated units (firms) for each characteristic in X for which we want to compare, we must assume that $0<(P(X)<1$. If the common support is not satisfied in the treatment group (i.e. some firms have characteristics that are only found in the treated firms or $P(X)=1$) then these firms are dropped and the ATT is estimated only for those firms where $P(X)<1$.

3.2.2.1 Propensity score generation and data classification in blocks

In practice what is done is that a normal logit/logistic/probit model is estimated having as predictor variables those that for the specific government program (wage subsidies) influence theoretically the selection process as well as the impact outcome of interest. These predictor variables should have a time restriction in that they should be constant across time and/or should be measured prior to the treatment implemented. The dependent variable is a dummy *(1,0)* representing respectively the treated and non-treated or control units of our sample.

After this score is produced for each unit (firm), subgroups of firms from the unsubsidised group are identified which happen to have *a similar propensity score* as the firms that were subsidised. This is achieved in three steps:

1. The data at hand is sorted according to the estimated propensity score of each unit, ranking the observations from lowest to highest.

2. All observations are stratified into several groups so that the respective scores within each stratum (group) between treated and control units are close enough (do not differ significantly at a given significance level). Initially the sample is split into k equally spaced blocks of propensity score and the default is $k=5$ [6].

3. Within each block a test is made to determine whether the average propensity scores of treated and non-treated units do not differ in a statistically significant way. If they do differ, then that specific interval is split into 2 halves and testing is conducted recursively in those new intervals.

Once the intervals have finally been determined then another test is conducted. For each block $i=1..k$ of propensity score, each predictor variable $j=1..m$ used in the logit model is tested so that it does not differ significantly between the treated and non-treated groups[7]. Here we have three possibilities.

4. If the tests for any predictor within any interval turn out significantly different between treated and non-treated units (an x characteristic is not balanced) the user can attempt to modify the logit model by adding, for example, interaction terms (*factor x * factor z*) or higher order terms of *factor x* (i.e. x^2) and conduct the comparisons again (recursively) from step 1 above.

5. If after these modifications a factor still can not be balanced, then the user must specify a less parsimonious model for predicting the propensity score.

6. If on the other hand all the tests for each predictor within each interval come out successful ($j*k$ statistically non-significant differences are found) then the final blocks are defined and we proceed with the second set of commands described below.

3.2.2.2 Estimation of the ATT using different search and matching algorithms

Once we have defined the blocks of our data, we can use different matching algorithms for the ATT estimation calculations. One, called *Stratification matching* calculates the difference between the average outcome of the treated

[6] Why five blocks in particular? Cochran (1968) has calculated that comparisons using 5 or 6 subclasses (blocks) will typically remove 90% or more of the bias present in the raw comparisons between the Treated and Control groups.

[7] In this case we have k*m tests/comparisons done.

units and the outcome of the control units within each block and the ATT is reported as being the average of these average differences.

Another algorithm approach is more complicated. It does not use averages for each block, but within each block/interval the algorithm examines each treated observation, one at a time. Then based on that unit's propensity score, it searches for the propensity score of the respective units of the control group in the same block. Because the propensity score is a continuous variable it is very unlikely that an exact match will be found. Thus, what is normally done is that search algorithms are applied to find the respective control unit's propensity score which is "closest" or "close enough" to the one of the treated unit.

Two examples of such matching algorithms are the *Nearest Neighbour Matching* (NNM) and *Kernel Matching* (KM). NNM is usually applied with replacement, which means that the *same* control unit can be used as a match to a treatment unit. This in turn means that we normally end up not having one to one matching but in many cases the control units chosen are less than the treated. After this identification is done, as with the stratification matching, the difference between the outcome of treated units and the outcome of the matched control units is calculated and the ATT is generated by averaging these differences. Note that with NNM all treated units find a matched control.

With KM, all treated units are matched with a weighted average of all controls with weights which are inversely proportional to the distance between the propensity scores of treated and controls. The formal notations and formulas of the aforementioned algorithms are found in the Appendix[8].

In these techniques the classic Stable Unit Treatment Value Assumption (SUTVA) applies requiring that potential outcomes for each unit examined are not related (are independent) to the treatment status of the other units given the observable characteristics *X*, or $Y_1, Y_0 \amalg D \mid X$. In our context, this would imply that the outcomes of treated firms would not depend on the outcomes of other treated and non treated firms; putting it differently, cross and general equilibrium effects are excluded from our impact estimations.

3.2.2.3 Pros and cons in using Propensity Score Matching

Advantages

According to one of its creators (Rubin, (1997)) the PSM based on observable characteristics is a superior approach over other estimation methods, such as linear regression models, for a couple of reasons.

[8] These are only a few of many other existing matching algorithms such as caliper, and radius matching.

First if the treatment and control groups do not adequately overlap on one or more of their common observable characteristics the balancing process (see section 3.2.2.1) will identify it. In contrast there is nothing in the standard output of any regression modelling software that will display this critical factor because they do not include the careful analysis of the joint distribution of the utilised regressors. Thus if the balancing is not adequate, one can not make valid causal inferences about the average net effects of the treatment on the treated.

Second, even if the balancing process indeed succeeds, regression models depend on the specific form of the model and its variables (linear, log-linear, etc.). Matching techniques on the other hand, do not rely on any particular functioning form (e.g. linearity for the relationship between the outcome and the characteristics within each treatment and control group/block).

Limitations

As in all observational (non-randomised) studies, causal questions using PSM methods suffer from the fact that we can only control the observable characteristics at hand and not the unobservable ones. By using a combination of PSM and Difference in Difference methods we might account for the fixed unobservable characteristics of our data, but the unobserved and variable through time characteristics of individual units (which may influence both the selection process and the outcome) can not be accounted for.

Another limitation is a very pragmatic one. To be able to utilise the PSM methods one needs large databases with detailed and relevant observable characteristics, depicting the selection process and potential influence to the outcome. With simple matching we have a higher probability of balancing the relevant characteristics of the treatment group within a very large sample of controls than within a small one (see section 3.2.1). With PSM although the data requirements fall dramatically due to the collapse of the set of characteristics to just one number, we still need an adequate amount of data. However, in many observational studies where the data collected is not designed before hand having a specific estimation method in mind, this is not feasible.

Finally, a third possible limitation of PSM is its handling of prognostically weak characteristics (predictors) included in the propensity score estimation via the logit model. A characteristic related to treatment assignment but not to the outcome, is treated the same as a characteristic not related to the selection process but strongly related to the outcome of interest[9]. The ideal case would be to have observable characteristics which are strongly related both to the selection process and to the outcome of interest. But as mentioned above, data from observational studies frequently do not offer such possibilities.

[9] See more on this issue in section 5.

4. Description of data and variables

The data analyzed comprised of financial statements of business firms which submitted tax declarations to the Finnish taxation authorities between 1995 and 2001. To this data we linked the amounts of wage subsidies that some of these aforementioned firms received on a yearly basis during the same 7-year period.

After some basic data handling to exclude outliers and because we needed a complete set of variables pre and post intervention[10], the final data set analyzed comprised of 25 152 firms of which 815 received wage subsidies for the first time during 1999. Some of these firms received subsidies just that one year, but some others later as well (Table 1). In the same table we report the average amount of wage subsidies per year for the same sub-groups. They ranged from € 2990 to € 10760.

Table 1. Number of firms and average wage subsidies received; subsidies received for the first time in 1999

Firms received subsidies	**1999**	**2000**	**2001**
in 1999 only			
# of firms	471		
average subsidy (EURO)	3330		
in 1999 and 2000			
# of firms	220	220	
average subsidy (EURO)	5120	2990	
in 1999 and 2001			
# of firms	33		33
average subsidy (EURO)	3830		10760
in 1999, 2000 and 2001			
# of firms	91	91	91
average subsidy (EURO)	3830	5360	2650
Total treated firms			
# of firms	**815**	**311**	**124**
average subsidy (EURO)	**3490**	**3640**	**4820**
Total controls			
# of firms	**24337**	**24337**	**24337**
average subsidy (EURO)	**0**	**0**	**0**

[10] The reason for such restrictions is two-fold: First we needed to create a complete panel and follow the pre-Treatment historical performance of firms. The same applied for post-Treatment observations. Thus we restricted our sample to firms which recorded salary costs for each of the 7 year period examined. We also trimmed the salary costs at 5% level. That is, we dropped observations whose salary cost levels were in the 1st and last 40 -percentile for each year (see also sections on "Functional form of predictors" and "Other transformations of impact indicator" below).

4.1 Outcome variables

In general, the goal of wage subsidy programs is to create longer term employment opportunities for the subsidised workers. In theory wage subsidies help them get into the job market, with the hope that they will continue to be employed by the subsidised (treated) firm, even after their wage is no longer subsidised.

To evaluate the efficacy of wage subsidies we used as the outcome of interest the estimated *net* average growth of salary costs (positive or negative) spent by the treated firms compared to the respective salary growth amounts of the non-treated firms (ATT). In other words, we measured the growth of salary costs as the *difference* between the salary levels before and after the receipt of wage subsidies taking under consideration the salary growth amounts for both the treated and non-treated firms. The salary growth amounts for the non-treated firms served as a proxy[11] for the salary growth levels of the treated firms, had they not received wage subsidies. The ATT was measured during the three periods shown below, up to three years after the receipt of subsidies:

- *Period duration 1 year: the year of treatment ($1999_{end} – 1999_{beginning}$)*
- *Period duration 2 years: till after one year from treatment ($2000_{end} – 1999_{beginning}$)*
- *Period duration 3 years: till after two years from treatment ($2001_{end} – 1999_{beginning}$)*

We hypothesised that for the program to show some efficacy, the ATT would remain in positive and stable (non-decreasing) levels throughout the periods examined.

If the opposite occurred, this would indicate that treated firms do not employ those subsidised workers for a sustainable period of time. It might also indicate a substitution effect occurring within the treated firms. This would materialise by laying off other non-subsidised workers during the same period and could be identified by lower than previously salary growth levels.

We assumed that, had the firm employed the subsidised worker after the subsidised period with her own funds, the period would be identified as the year after the subsidised one. Again, in reality firms may employ subsidised workers for period which carry over from one year to another. In our data however, since the particular month when wage subsidies were received was not identified, we assumed that if a firm reported subsidy receipts any one year, this referred to a 6-month period on average, during that *same* year. Also, following Sianesi (2002, p.44, footnote 35), we assumed that the ATT measured was caused by the *initial*

[11] This is the counterfactual situation which can not be observed.

receipt of wage subsidies in 1999. Any subsequent wage subsidies received in 2000 and/or in 2001 are viewed as outcomes of that first 1999 receipt.

Three versions of the outcome variable were used. The first (denoted (A) in Tables 5 and 6 – see below) calculated the net salary growth without any adjustments.

- $A1 = salary\ 1999_{end} - salary\ 1999_{beginning}$
- $A2 = salary\ 2000_{end} - salary\ 1999_{beginning}$
- $A3 = salary\ 2001_{end} - salary\ 1999_{beginning}$

As mentioned above we wanted to estimate the average length of employment of the subsidised worker within the respective firm and similarly test potential substitution behavioural effect of the same firm. Hence, in the other two version of the dependent variable (B, C), we progressively deducted certain amounts of salaries from the unadjusted net salary growth amount for each firm. These amounts, as will be shown below, depicted the total salary costs that the firm would have spent had she employed the subsidised worker for a certain period after the wage subsidy received.

The second version (B) was created by deducting from (A) two amounts: (a) the amount of subsidies that the treated firms received during any one year and (b) the obligatory salary contributions of the treated firms for the subsidised employees; that is wage *subsidy + wage subsidy*1.5*[12]. With (B) we tested whether the firm kept the subsidised worker for the 6 month average obligatory period and at the same time did not fire anyone else instead, ceteris paribus.

- $B1 = A1 - subsidy_{1999} - subsidy_{1999} *1.5$
- $B2 = A2 - subsidy_{1999} - subsidy_{1999} *1.5 - subsidy_{2000} - subsidy_{2000} *1.5$
- $B3 = A3 - subsidy_{1999} - subsidy_{1999} *1.5 - subsidy_{2000} - subsidy_{2000} *1.5 - subsidy_{2001} - subsidy_{2001} *1.5$

In version (C) we assumed that the firm after the 6-month subsidised period retained the previously subsidised employee for up to two years after the receipt of the wage subsidies, but now covering all his salary costs from her own funds. That meant that from (A) we deducted the respective total salary amount (*subsidy + subsidy*1.5*) for the obligatory 6-month subsidised period twice. With (C2) we tested whether the subsidised workers were still employed by the beneficiary firms one year after the subsidised period (at the end of the year 2000); with (C3)

[12] Wage subsidy*1.5 equals 60% of the total salary cost for each subsidised employee (see also section 2).

we tested whether the subsidised workers were still with the firm two years after the subsidised period (at the end of the year 2001).

- $C2 = A2 - 2*subsidy_{1999} - 2*subsidy_{1999}*1.5 - subsidy_{2000} - subsidy_{2000}*1.5$

- $C3 = A3 - 2*subsidy_{1999} - 2*subsidy_{1999}*1.5 - 2*subsidy_{2000} - 2*subsidy_{2000}*1.5 - subsidy_{2001} - subsidy_{2001}*1.5$

Note that for subsidies received during 2000, the salaries deducted when calculating (B2) and (C2) did *not* exceed the amount of salary costs for a 6-month period. Respectively, for the salaries received during 2001 we assumed again, that on average, the worker was employed for 6 months, thus an estimated salary cost for just 6 months was deducted from (B3) and (C3). Of course in all cases (B1, B2, B3, C2, C3) the respective salary amounts of non-treated firms (controls) remained unchanged and were the same as in (A1), (A2) and (A3).

The functional form of the outcome variable (net growth of salaries) enabled us to combine a Difference in Differences (DiD) estimation with matching, and thus account for any fixed unobservable characteristics within the two groups. Had we used the traditional transformation of logging the salary amounts this would have not been possible.

4.2 Independent variables

According to matching theory, the logit model via which the propensity score is generated should include predictor variables that influence the selection procedure and the outcome of interest. Since in our data we had information solely on financial information of firms and on a few other of their characteristics (industrial sector and geographical location), we were somewhat limited in our attempts to model explicitly the selection procedure. The decision to employ unskilled workers via wage subsidies may not be a matter reflected directly in the financial figures of firms as, for example, in the case of investment subsidies. There, the decision to give such subsidies depends in theory at least on the financial condition of the applicant firm and on the project in question; and these conditions can be accessed to some extent by looking directly at the financial statements of firms. With wage subsidies what is important is the willingness of the firm to employ the worker initially for the 6-month period, the activity and initiative of the worker to find such a firm, and the existence and activity of "satellite" organisations whose responsibility is to find firms who are willing to hire these people. For example as mention in section 2, there are adult education centres who actively send their pupils to a list of pre-identified firms to be wage subsidised.

Nonetheless, the financial conditions of the firm may be an indirect factor identifying her as a potential employer of a subsidised worker. Financially sound firms may be more willing to employee such workers, the workers themselves maybe more prone to go to a "better" firm if they have a choice and the satellite organisations may also "shortlist" such firms as more reliable employers for the current as well as for future unemployed workers.

In terms of observable non-financial characteristics (Tables 2 and 3) we had to settle for the firms' industrial sector using the standard SIC95 two digit 18-group classification. Furthermore, the firms' geographical location was based on the 5-group classification used in Kangasharju and Venetoklis (2002).

Table 2. Industrial 2-digit sector based on SIC 95 classification

No	Letter code	Description
01	A	Agriculture, hunting and forestry
02	B	Fishing
03	C	Mining and quarrying
04	D	Manufacturing
05	E	Electricity, gas and water supply
06	F	Construction
07	G	Wholesale and retail trade; repair of motor vehicles, motorcycles personal and household goods
08	H	Hotels and restaurants
09	I	Transport, storage and communication
10	J	Financial intermediation
11	K	Real estate
12	L	Public administration and defence; compulsory social security
13	M	Education
14	N	Health and social work
15	O	Other community, social and personal service activities
16	P	Private households with employed persons
17	Q	Extra territorial organizations and bodies
18	X	Industry unknown

We assumed that firms in certain industrial sectors and in certain geographical areas are more prone to hire such subsidised employees. For example, manufacturing firms might be more prone to hire unskilled workers than, say, financial institutions. Also the geographical location might play a role as well since hiring through wage subsidy programs might differ according to where a firm is located (e.g. in the Country side or in a Large University centre).

Table 3. Regional sub-regions by regional group

Capital Region (1)		Other Large University Centres (2)		Other Provincial Centres (3)		Intermediate Industrial Centres (4)		Country-side (5)			
011	Helsinki	131	Jyväskylä	201	Porvoo	103	Savonlinna	094	Kärkikunnat	068	Lounais-Pirkanma
		023	Turku	081	Kouvola	052	Riihimäki	146	Järviseutu	053	Forssa
		064	Tampere	071	Lahti	082	Kotka-Hamina	153	Sydösterbottens kustregion	024	Vakka-Suomi
		171	Oulu	043	Pori	013	Tammisaari	124	Keski-Karjala	066	Koillis-Pirkanma
				211	Mariehamn	154	Jakobstadsregionen	111	Ylä-Savo	177	Ylivieska
				101	Mikkeli	022	Salo	115	Sisä-Savo	197	Pohjois-Lappi
				182	Kajaani	093	Imatra	176	Nivala-Haapajärvi	196	Tunturi-La
				122	Joensuu	135	Äänekoski	141	Suupohja	194	Koillis-Lap
				051	Hämeenlinna	134	Jämsä	144	Kuusiokunnat	181	Kehys-Kainuu
				191	Rovaniemi	063	Etelä-Pirkanmaa	172	Lakeus	178	Koillismaa
				162	Kokkola	041	Rauma	044	Pohjois-Satakunta	193	Torniolaak
				142	Pohjoiset seinänaapurit	012	Lohja	143	Eteläiset seinänaapurit	123	Ilomantsi
				152	Vaasa	114	Varkaus	025	Loimaa	175	Siikalatva
				091	Lappeenranta	174	Raahe	121	Outokumpu	173	Ii
				112	Kuopio	192	Kemi-Tornio	062	Kaakkois-Pirkanmaa	212	Föglö
								067	Pohjois-Pirkanmaa	125	Pielisen Karjala
								065	Itä-Pirkanmaa	137	Viitasaari
								021	Åboland-Turunmaa	161	Kaustinen
								042	Kaakkois-Satakunta	102	Juva
								145	Härmänmaa	136	Saarijärvi
								202	Loviisa	113	Koillis-Sav
								105	Pieksämäki	104	Joroinen
								151	Kyrönmaa	132	Kaakkoine Keski-Suo
								061	Luoteis-Pirkanmaa	133	Keuruu
								072	Itä-Häme	092	Länsi-Saimaa

As far as other predictors utilised in the logit models are concerned, we attempted to find simple financial indicators which would reveal in aggregate terms the course through time (the short term history) and the financial strength of the firms, *prior* to the receipt of wage subsidies. We assumed that such variables might be good indicators of the behaviour of the firm in the post wage-subsidy receipt period had she not been subsidised. Also, as mentioned above, the financial strength of the firm depicted in these indicators may have after all, influenced the selection and hiring of unemployed workers. And this, either because (a) the workers themselves wanted to go to a place where they have a chance of finding more permanent employment (the better the firm the better chances they have) or (b) these favourable characteristics were recognised by the wage subsidy distributing officers of the local Labour office and/or the respective officers of the adult education training centres responsible for the appointment of these positions of their graduates.

We selected the following five variables (accounts) from the financial statements of firms: *Turnover, Tangible Assets, Profit /Loss, Salary costs and Own Capital.* Since we had a panel of such financial information, for each of these amounts we calculated their growth rates from year to year for three period prior to the year of the receipt of subsidies. That is, since we chose firms that received wage subsidies for the first time in 1999, we calculated the growth rates for each of the aforementioned variables between 1995–1996, 1996–1997 and 1997–1998. That meant that we ended up having 15 different predictors depicting historical performance of the firms in our sample. The initial growth rate for each variable was based on the simple formula

(variable value at Year T+1$_{end}$ *– variable value at Year T* $_{end}$*)/ variable value at Year T* $_{end}$.

These growth rates were of course calculated for both subsidised and non-subsidised firms. Finally, to account for size and taking under consideration of the outcome variable of interest (the average net salary cost growth of the treated firms attributed to the receipt of wage subsidies) we used the *level values* of sales and salary costs of the year *prior* to the first receipt year of subsidies. In our case it was sales and salary cost figures for 1998. All in all, the RHS of our logit models comprised of 17 predictors (34 if you count the dummy variables generated see Table 4 later on); in our robust OLS regressions (see below) we added one more binary variable (1,0) denoting treatment.

Functional forms of predictors

The final functional form of the predictors used in the logit models was not in their original form (as simple growth rates). Firms are very heterogeneous organisations and although growth rates of predictors have less variability than their respective level amounts, their variability was still quite high[13]. We attempted to drop all outlier values but the process proved futile since we ended up losing too many observations. Hence we initially left the predictors intact and used those in the logit models in an effort to balance the treatment and control groups for each variable within each block (see section 3.2.2.1). This approach was unsuccessful. In case of such imbalances, Rubin (1997) suggests adding square transformations or even interaction terms for those predictors that do not balance. We followed this, using different specifications but the models still did not balance. We thus decided instead to use the predictors' 100th percentiles. In this way we created similar predictors as in their original form but with somewhat wider ranges. We transformed all continuous predictors (17 in total) to their percentile versions. Finally, for the two categorical predictors (2-digit SIC95 Industrial classification and Geographical location) we created n -1 strata dummies, (13 and 4 dummies respectively). With these transformations the logit models balanced and we were able to produce the propensity scores per firm, used for the ATT estimations.

Additional estimation methods conducted

The estimations methods applied comprised of the three propensity score matching algorithms described in section 3.2.2.2. and in the Appendix (Nearest Neighbour, Kernel and Stratification Matching). To test for robustness and sensitivity we also run ordinary least square (OLS) regressions with two similar model specifications; one (OLS cat.) utilizing the exact list of predictors used in the respective logit models and another (OLS cont.) where the 100th percentile variables were replaced with their respective continuous counterparts (original

[13] To give a idea of the variability for the predictor variables in our dataset, even in their growth rate format, we list below some descriptive statistics for each one of them.

Variable	Obs	Mean	Std. Dev.	Min	Max
Sales98_97	35148	.7532346	33.35657	-.9999222	4543.041
Sales97_96	33989	1.114087	43.85345	-.9999018	5592.582
Sales96_95	33250	5.279354	634.8665	-.9998553	113991
Salary98_97	36411	.1718259	.5076854	-.7937619	5.626699
Salary97_96	36411	.2131669	.6224734	-.8100489	6.083219
Salary96_95	36411	.2972546	.8279375	-.8746377	7.247317
T.Assets98_97	34307	7.605602	897.8302	-.9998555	156200
T.Assets97_96	34172	9.688509	1465.146	-.9999782	270400
T.Assets96_95	33867	1.213845	62.15755	-.9999933	10079
OwnCap98_97	35390	7.725263	3016.01	-341458.7	440071.9
OwnCap97_96	35326	.350484	43.44511	-2400.463	6519.696
OwnCap96_95	35199	-14.89542	2897.8	-543577	6832.5
Pr/Loss98_97	35844	63.56889	12749.94	-146788.8	2408482
Pr/Loss97_96	35855	2.663792	794.4731	-81243.2	70693.46
Pr/Loss 96_95	**35810**	**14.3585**	**2113.98**	**-130835**	**366054**

growth rates). With this we wanted to check the sensitivity of the initial logit specification indirectly; that is, whether the estimate generated would deviate considerably had we been able to use the latter alternative model specification (which as stated earlier was not possible due to the inability to balance for all continuous predictors)[14]. For each OLS model we applied the Huber/White/sandwich estimator of variance. We finally run basic T-tests, that is, simply compared the difference in outcome between the treated and control groups without controlling for any observable characteristics.

Other transformations on the impact indicator

We restricted our sample to those firms that had non-missing values for the impact indicator (salary costs) through out the period examined (1995–2001). Of the records left, we eventually measured a 5% trimmed distribution to reduce potential outlier effects. Finally, to control for inflationary effects which might have biased our ATT average net salary cost growth estimates, we deflated all salary costs amounts for each year in our panel (1995 to 2001) using 1995 as our base year. In turn, the same was done for the wage subsidies. The deflation was done for each firm separately based on her SIC 95 industrial sector (at 2-digit level) using the respective deflation indices from Statistics Finland.

[14] The list of predictor variables used in both of the OLS models are found in the logit models in Tables 4 and 5. The functional form of these predictors is in the case of OLS_cat their 100th percentiles and in the case of OLS_cont their original continuous form (see previous footnote). In the RHS of the equation we also added a binary variable indicating whether the firm received wage subsidies or not. The binary variable's coefficient and t-value are reported in Tables 5 and 6 under the respective OLS models. The dependent variables are the same as in the matching models.

5. Results and discussion

The results are analysed in three sections. In section 5.1 we briefly discuss some considerations that derive from the logit models built to generate the propensity scores. In section 5.2 we examine thoroughly the ATT estimates produced and then compare the results among themselves using different criteria. We conclude in section 5.3.

5.1 Examination of logit model

Table 4 lists the logit (probit) model based on which the propensity scores estimated for the set of firms receiving subsidies for the first time in 1999.The Pseudo R-squared is low (less than 10%) which denotes that the model, although overall significant, manages to account for only a small part of the variability of the binary treatment variable. This is in line with our previous comments where we warn that our data set does not contain the best possible predictors of treatment selection. If one examines the predictors themselves it is found that indeed some of them do not come statistically significant. However this does not necessarily mean that they should not be included in our models. Rubin and Thomas ((1996), cited in Shadish et al, (2002, p. 162)) argue that

> *"…unless a variable can be excluded because there is a consensus that it is unrelated to outcome or is not a proper covariate, it is advisable to include it in the propensity score model even if it is not statistically significant"* .

The same is reported by Augursky and Schmidt ((2000), cited in Kluve et al., (2002)). Based on a simulation study they conclude that it is more important to achieve balance on the relevant covariates/characteristics than painstakingly try to model the selection process.

Table 4. Logit model generating propensity score. Firms received wage subsidies for the first time in 1999

Treatment 1,0	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Base SIC95 category 1						
3	-.3762873	.4276207	-0.88	0.379	-1.214408	.4618338
4	.2815814	.1444249	1.95	0.051	-.0014861	.5646489
5	-.1770899	.3375593	-0.52	0.600	-.8386939	.4845142
6	.2727578	.1417463	1.92	0.054	-.0050598	.5505754
7	.2705878	.1393823	1.94	0.052	-.0025965	.5437722
8	.3870742	.1631356	2.37	0.018	.0673343	.7068142
9	-.0069341	.1504345	-0.05	0.963	-.3017803	.2879121
10	-.2104531	.3698702	-0.57	0.569	-.9353854	.5144792
11	.3227289	.1424636	2.27	0.023	.0435055	.6019524
12	1.052966	.6024427	1.75	0.080	-.1278004	2.233732
13	.3763724	.217426	1.73	0.083	-.0497748	.8025197
14	.2084586	.1721459	1.21	0.226	-.1289411	.5458583
15	.4922723	.1626071	3.03	0.002	.1735681	.8109764
Base geographical location category 1						
2	-.0765968	.0850879	-0.90	0.368	-.243366	.0901724
3	-.1894863	.0864938	-2.19	0.028	-.359011	-.0199617
4	-.2241432	.0773632	-2.90	0.004	-.3757723	-.0725142
5	-.6599598	.0830664	-7.94	0.000	-.8227669	-.4971527
salary 1999	.005161	.0015148	3.41	0.001	.0021921	.0081299
sales 1999	.0049296	.001726	2.86	0.004	.0015468	.0083125
OwnCap98_97	-.0003088	.0007084	-0.44	0.663	-.0016972	.0010795
OwnCap97_96	.0001239	.0006885	0.18	0.857	-.0012256	.0014733
OwnCap96_95	-.0004436	.0006742	-0.66	0.511	-.001765	.0008779
pr/loss98_97	-.0013687	.0007289	-1.88	0.060	-.0027974	.00006
pr/loss97_96	-.00106	.0007221	-1.47	0.142	-.0024753	.0003553
pr/loss96_95	-.0002445	.0007366	-0.33	0.740	-.0016883	.0011992
sales98_97	.0026758	.0009375	2.85	0.004	.0008383	.0045133
sales97_96	.0007415	.0009389	0.79	0.430	-.0010987	.0025818
sales96_95	.0002167	.0009126	0.24	0.812	-.0015721	.0020054
salary98_97	-.0009343	.0008906	-1.05	0.294	-.0026798	.0008112
salary97_96	.0013247	.0008974	1.48	0.140	-.0004341	.0030835
salary96_95	.0007995	.000872	0.92	0.359	-.0009096	.0025085
T.assets98_97	.000977	.0006929	1.41	0.159	-.0003811	.0023351
T.assets97_96	.0013801	.0006928	1.99	0.046	.0000223	.0027379
T.assets96_95	.0012799	.0006928	1.85	0.065	-.000078	.0026379
_cons	**-2.656272**	**.1768901**	**-15.02**	**0.000**	**-3.00297**	**-2.309574**

Table 4. Logit model generating propensity score. Firms received wage subsidies for the first time in 1999 (continued)

No of obs per block to balance

Inferior of block of prop. score (with common support)	**Controls**	**Treated**	**Totals**
.0015175	2,852	18	2,870
.0104167	4,009	63	4,072
.0208333	5,848	173	6,021
.0416667	4,127	269	4,396
.0833333	808	82	890
.1666667	8	3	11
Total	**17,652**	**608**	**18,260**
Inferior of block of prop. score (without common support)	**Controls**	**Treated**	**Totals**
.0003271	9,537	225	9,762
.0104167	4,009	63	4,072
.0208333	5,848	173	6,021
.0416667	4,127	269	4,396
.0833333	808	82	890
.1666667	8	3	11
Total	**24,337**	**815**	**25,152**

5.2 Examination of ATT estimates

By definition, in retrospective observational studies, we cannot test directly the goodness of our estimations since identifying a counterfactual population resembling closely the treated group will always be limited to the availability of existing data. Matching estimation techniques depend on the richness of the data at hand in depicting the selection process, the data's relevance to the outcome measured and their availability in terms of historical performance/behaviour of the units of interest.

In this paper, the ATT estimates are generated utilising six different estimation methods; three using propensity score search algorithms, two using OLS estimators and one by conducting a simple T-test between the treatment and control groups in question.

The basis of the analysis evolves around three indicators:
(a) the magnitude of ATT point estimates
(b) their sign and
(c) whether they come out statistically significant.

For the latter case we look at the level of the t-value next to each ATT estimate. We choose a 5% significance level as our cut off point. Thus, taking into consideration the number of observations and degrees of freedom per model, t-values of approximately over 1.96 denote statistical significance of the ATT estimates. Put differently, if the t-value were over 1.96, the respective ATT amount estimated would denote a statistically significant difference in favour of or against[15] the Treated firms. If the opposite were the case ($t < 1.96$) the result would be due to chance sample variability.

The three aforementioned indicators can be examined not only independently, but in a comparative fashion as well. We compare the ATT estimates among the six estimation method used. Furthermore, we check the ATT levels, signs and t-values generated from the two samples of firms where the common support restriction has (Table 5) and has not been imposed (Table 6).

Comparison of estimation methods

Looking at the different estimation methods we notice that no method generates result patterns which differ exceptionally from the others.

The three matching algorithms produce similar results in terms of statistical significance. Their point estimates however differ somewhat, especially when looking at the Nearest Neighbour Matching (NNM) versus Kernel (KM) or Stratification Matching (SM). This is of no surprise if one looks at the number of observations analysed. NNM uses a more restricted and smaller sample than the other two estimation methods.

The specification used in the "OLS_cont." produces somewhat larger ATT estimates compared to the specification used with the "OLS_cat.". These consistent results may lead us to assume that had we been able to use the continuous versions of the predictors in our logit models, the ATT results produced by the propensity score methods might have been slightly higher than their current levels but not in any significant way. Furthermore, as mentioned earlier, in the OLS regression specifications we included three versions of lagged

[15] In favour, if the sign is positive and against if the sign is negative. This is because in all six methods ATT is calculated as the net difference of the average outcome indicator for the *treated less* the average outcome indicator for the *non treated* firms.

variables for each financial statement variable. We thus suspected problems with multicollinearity. Nevertheless, VIF diagnostic tests showed no such problems[16].

16

	Table 5 models		Table 6 models	
	VIF mean	**VIF max**	**VIF mean**	**VIF max**
OLS-cat	2,65	9,97	2,61	9,73
OLS_cont	2,37	9,55	2,34	9,55

For multicollinearity to be a problem the mean VIF should be over 10 and the VIF max multiple of 1.

Table 5. Estimation of ATT on salary costs of firms which received employment subsidies for the first time in 1999; Common support restriction

Logit model and propensity score (ps) information

LL:	-2507,4
Pseudo R2:	0,06
Model obs	18303
Common Support Region	
Lower ps bound	0,001
Higher ps bound	0,212
Blocks	6

Estimation method		**NNM**		**KM**		**SM**		**OLS_cat.**		**OLS_con.**		**T-Test**	
Treated firms		608		608		608		608		608		608	
Control firms		587		17652		17652		17652		17652		17652	
Totals		1195		18260		18260		18260		18260		18260	
Dep. variable	**Period**	**ATT**	**t-values**	**ATT**	**t-values***	**ATT**	**t-values**	**ATT**	**t-values****	**ATT**	**t-values****	**ATT**	**t-values**
A. Net salary growth without subsidised employment adjustments													
A1	1.1.99 – 31.12-99	12027	1,216	15843		14501	1,563	14728	1,610	17362	1,700	16022	1,746
A2	1.1.99 - 31.12.00	36497	2,550	38179		33806	2,589	33841	2,590	36700	2,570	38612	2,917
A3	1.1.99 - 31.12.01	28424	1,962	24555		17660	1,340	18357	1,450	17488	1,310	25165	1,974
B. A less subsidies and less 60% firm's obligatory salary contribution Test for obligatory 6 – month employment													
B1	1.1.99 – 31.12-99	3110	0,314	6925	0,720***	5584	0,600	5813	0,640	8437	0,409	7104	0,771
B2	1.1.99 - 31.12.00	24219	1,686	25733		21360	1,638	21479	1,640	24310	1,700	26228	1,981
B3	1.1.99 - 31.12.01	13861	0,956	9988		3110	0,236	3882	0,310	2955	0,220	10620	0,832
C: A less salary costs for 12 months employment of subsidised person from firm's own funds Test for employment of at least 12 months (C2). Test for employment for at least 24 months (C3)													
C2	1.1.99 - 31.12.00	15264	1,062	16819	1,445***	12456	0,953	12564	0,960	15385	1,070	17310	1,305
C3	1.1.99 - 31.12.01	1477	0,101	-2388	-0,198***	-9267	-0,698	-8480	0,660	-9434	-0,700	-1764	-0,137

* in some cases analytical standard errors and t-values not computed;
** robust standard errors
*** t-values based on bootstrapped standard errors (50 repetitions)

Table 6. *Estimation of ATT on salary costs of firms which received employment subsidies for the first time in 1999; No common support restriction*

Logit model and propensity score (ps) information: As in Table 5

Estimation method		**NNM**		**KM***		**SM**		**OLS_cat.****		**OLS_con.****		**T-Test**	
Treated firms		815		815		608						815	
Control firms		7229		24337		17695						24337	
Totals		8044		25152		18303		18352		18352		25152	
Dep. variable	**Period**	**ATT**	**t-values**	**ATT**	**t-values***	**ATT**	**t-values**	**ATT**	**t-values****	**ATT**	**t-values****	**ATT**	**t-values**
A. Net salary growth without subsidised employment adjustments													
A1	1.1.99 – 31.12-99	12060	0,927	15848		14502	1,563	14752	1,602	17359	1,700	16803	2,249
A2	1.1.99 - 31.12.00	36665	2,594	38179		33806	2,589	33877	2,590	36696	2,570	35139	3,297
A3	1.1.99 - 31.12.01	28424	1,894	24555		17660	1,430	18428	1,460	17491	1,310	30019	2,649
B. A less subsidies and less 60% firm's obligatory salary contribution Test for obligatory 6 – month employment													
B1	1.1.99 – 31.12-99	3124	0,240	6930		5584	0,600	5837	0,640	8434	0,820	8021	1,071
B2	1.1.99 - 31.12.00	24219	1,716	25733		21360	1,638	21515	1,640	24307	1,700	22867	2,147
B3	1.1.99 - 31.12.01	13917	0,923	10008		3110	0,236	3951	0,310	2958	0,220	15911	1,407
C: A less salary costs for 12 months employment of subsidised person from firm's own funds Test for employment of at least 12 months (C2). Test for employment for at least 24 months (C3)													
C2	1.1.99 - 31.12.00	15313	1,083	16819		12457	0,953	12600	0,960	15381	1,070	14085	1,320
C3	1.1.99 - 31.12.01	1522	0,101	-2371		-9267	-0,698	-8410	-0,660	-9431	-0,700	3640	0,320

* analytical standard errors and t-values not computed;
** robust standard errors

It is also interesting to note that the T-test estimation method produces similar ATT estimates as the other methods, albeit the fact that it uses “gross” comparisons (without controlling for any characteristics within the two groups as the other methods). In only a few cases do the results have different statistical significance compared to the other methods (e.g. compare the ATT generated with the T-test in (A1) and (A3) with the ATT of the other respective models).

Comparison of the two common support regimes

Here we notice that, apart from the number of observations which in the case of no common support is naturally larger, the results are qualitatively and quantitatively rather similar. For the majority of the estimation methods, the point estimates for the ATT of the two common support regimes are very close to each other qualitatively (they have in all cases the same sign) and quantitatively. The only exception is found with the last method, the “gross” T-test, where of course no control variables are used in the estimations. There we notice that the point estimates of the no common support regime are larger than the case when the common support is implemented. The similarity of the above results is in line with the comments of Lechner ((2001), cited in Becker and Ichino (2002)) who mentions that the common support restriction is not necessarily better. If anything, the larger number of observations utilised without the common support restriction, makes it easier to make inferences to a wider population of firms.

5.3 Discussion

The point estimates of the ATT generated by the majority of the methods, types of dependent variables, and for the majority of the periods examined come out in favour of the Treated firms (positive) but statistically insignificant. In a few cases however, the estimates come positive *and* statistically significant. One such example are the ATT for period (A2) in both the common support and no-common support samples of firms. Another, is the ATT under the T-test for the “no common support” sample of firms (Table 6). The ATT estimates begin having a negative sign, but statistically insignificant t-values, in the last period examined of the (C) version of the dependent variable, the (C3); and this for both the common support and no-common support samples of firms.

In general, the ATT show an increase of net salary growth from the one year (A1) period[17] to the two-year (A2) estimation period and then a decrease for the three-year (A3) period. This means that, on average, the subsidised firms have higher net payroll costs during the first year after the receipt of wage subsidies compared to the non-subsidised firms. This first post-subsidies-year average net payroll growth comes out larger than the estimated average wages spent for the

[17] This one-year period is measured as the same year of the receipt of subsidies

subsidised workers (B2). This in turn may indicate that firms employ the subsidised workers for at least one year after the subsidised period and perhaps employ in addition extra staff. On the other hand, the salary costs net growth measured two years after the receipt of subsidies diminishes and seems not to cover all of the subsidised workers' salary costs since the ATT comes out negative (C3). Because these results come out statistically insignificant, one might argue that the negative signs are due to sample variability. Unfortunately, we do not have data to check the net payroll growth, say, three years after the receipt of subsidies up to 2002. Regardless of this, we interpret these latter results as initial signs of lay offs of the subsidised workers and/or of those that were recently employed the previous period[18].

Evaluation is a two phase process; measurement of potential impacts and giving a judgement on their worth. We have till now concentrated on the first phase. Before we proceed with the second, we should mention a caveat of the study that we feel is important for the reader to be reminded of. It refers to the sample analysed and the fact that it may not be the most representative of the total population of firms which have been receiving wage subsidies in Finland. The reason is the elimination from our original dataset of many observations which did not fill our criteria for analysis (e.g. they should have had non-missing historical financial data for the total 7-year (1995–2001) period examined and they should have received wage subsidies for the first time only in 1999). Thus the results' external validity is rather weak and general inferences should be made with care.

Is this two year period of employment attributed to a wage subsidies program an acceptable result? We do not know whether explicit quantitative targets for employment spells have been imposed beforehand by the policy makers, program designers and program implementers. From a pure economic efficiency (or a very simple cost-benefit) point of view however, and focusing only in the short three year period under scrutiny (1999–2001), the results are probably acceptable. Our estimations indicate that firms pay from their own money on aggregate more than five times the amount of wage subsidies received, by employing the worker for at least two years after the receipt of wage subsidies. In addition, we can not disregard the distributive justice dimension of the wage subsidies programs regardless of their whatever economic (in)efficiencies. Society as a whole might be better off in the long run paying for such programs. And indeed the subsidised workers themselves may evaluate these programs in a positive manner.

On the other hand the decrease of wage levels from the first to the second year is something that is not desirable since it is a probable sign of layoffs. Wage subsidies are theoretically designed not only to introduce previously unemployed

---

[18] Of course from our data we can not identify which exact workers are laid off, but the chances are that workers employed later with the firm would be the first to go.

workers once again to the labour market, but at the same time create the conditions and opportunities so that these workers continue to be employed for a *prolonged* period of time, perhaps within the firm which hired them in the first place. Thus if the decrease of salary growth were to continue in the next few years, policy makers should ponder whether such effects are indeed legitimate program targets or whether one needs to look for other tools to achieve longer, more sustainable employment for the previously unemployed.

References

Abadie, A., Drukker, D., Leber Herr, J. and Imbens, G.W. (2003): Implementing Matching Estimators for Average Treatment Effects in STATA. Draft (October 17,2002). in http://emlab.berkeley.edu/users/imbens/statamatch_02oct17.pdf as at 22.09.03 Download Software from http://emlab.berkeley.edu/users/imbens/estimators.shtml

Augurzky, B. and Schmidt, C.M. (2000): The propensity Score: A means to an End. Dept of Economics Discussion paper No. 334. University of Heidelberg.

Becker, S.O. and Ichino, A. (2002): Estimation of Average Treatment effects based on propensity score. Stata Journal 2 (4) 358-377. Download software from http://www.sobecker.de/pscore.html

Bergstrahl, E., Kosanke, J. and Jacobsen, S. (1996): Software for optimal matching in observational studies. Epidemiology, vol. 7, pp.331-332.

Cochran, W. G. (1968): The effectiveness of adjustment by sub-classification in removing bias in observational studies. Biometrics, 24 pp.295-313.

Czarnitzki, D. and Fier, A. (2002): Do Innovation Subsidies Crowd Out Private Investments? Evidence from the German Service Sector. Applied Economics Quarterly (Konjukturpolitik) Vol. 48:1, pp. 1-25.

Dehejia, R.H. and Wabba, S. (1999): Causal Effects in non experimental studies: Re-evaluating the evaluation of training programmes. Journal of the American Statistical Association, vol. 94 , pp.1053-1062.

Dehejia, R.H. and Wabba, S. (2002): Propensity Score Matching Methods for Non-Experimental Causal Studies. *Review of Economics and Statistics*, Volume 84 (February 2002), pp. 151-161.

Heckman, J.J., Ichimura, H. and Todd, P. (1997): Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. Review of Economic Studies, Oct 1997, Vol. 64 (4), pp.605-654.

Heckman, J.J., Ichimura, H. and Todd, P. (1998): Matching as an Economic Evaluation Estimator. Review of Economic Studies, April 1998, Vol. 65 (2), pp.261-294.

Heckman, J.J., Lalonde, R.L. and Smith, J. (1999): The Economics and Econometrics of ALMP. In Orley Ashenfelter and David Card (eds.): Handbook of labour economics, Vol. 3A. Amsterdam: North-Holland, pp.1865-2097.

Holland, P.W. (1986): Statistics and Causal Inference. Journal of the American Statistical Association 81:396, 945-970, with discussion.

Kangasharju A. and Venetoklis, T. (2002): Business subsidies and employment of firms: Overall Evaluation and Regional Extensions. Government Institute for Economic Research, Discussion papers No 268, Helsinki.

Kangasharju A. and Venetoklis, T. (2003): Do wage subsidies increase employment in firms? Government Institute for Economic Research, Discussion papers No 304, Helsinki.

Kluve, J., Lehmann, H. and Schmidt, C. (2002): Disentangling Treatment Effects of Polish Active Labour market Policies: Evidence from Matched Samples. IZA Institute for the study of Labour discussion paper 355, Bonn.

Lechner, M. (2001): A note on the common support problem in applied evaluation studies. Discussion paper 2001-01 Department of Economics, University of St.Gallen.

Lechner, M. (2002): Program heterogeneity and propensity score matching: An application to the evaluation of active labour market policies. The Review of Economics and Statistics, 84(2); 205-220.

Love, T.E. (2002): Propensity Scores for Selection Bias Adjustment in Observational Studies. Department of Operations, Weatherhead School of Management. In http://www.chrp.org/love/OPpropensityreferences.pdf as at 22.09.03.

Robin, D.B. and Thomas, N. (1996): Matching using estimated propensity scores: relating theory to practice. Biometrics, 52, pp. 249-264.

Rosenbaum, P. R., and Rubin, D.B. (1983): The central role of the Propensity Score in Observational Studies for causal Effects. Biometrica 70, 41-50.

Rubin, D.B. (1997): Estimation from non-randomized treatment comparisons using sub-classification on propensity scores. Proceeding of the international Conference on non-randomised comparative clinical studies in Heidelberg April 10-11, 1997 at http://www.symposion.com/nrccs/rubin.htm as at 20.08.2003. This article is a modification and expansion of an article in Annals of Internal Medicine, Vol. 127 (2), pp. 757-763.

Rubin, D.B. and Thomas, N. (1996): Matching using estimated propensity scores: Relating Theory to practice. Biometrics, Vol. 52, pp.249-264.

Shadish, W.R., Cook, T.D. and Campbell, D.T.(2002): Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin Company.

Sianesi, B. (2002): Essays on the evaluation of social programmes and educational qualifications. IFAU Institute for Labour Market Policy Evaluation. Dissertation Series No. 3, University College London, United Kingdom.

Sianesi (2003): Implementing Propensity Score Matching Estimators with Stata. In http://fmwww.bc.edu/RePEc/usug2001/psmatch.pdf as at 22.09.03. Download software from http://ideas.repec.org/c/boc/bocode/s432001.html

Smith, J. and Todd, P. (2001): Reconciling Conflicting Evidence on Performance of Propensity Score Matching Methods. American Economic Review Papers and Proceedings, May 2001, Vol. 91(2), pp.112-118.

Venetoklis, T. (2002): Public Policy Evaluation: Introduction to Quantitative Methodologies. Government Institute for Economic Research. Research Papers, No 90, Helsinki.

Appendix: Notation and formulas

Formal notation and formulas calculating ATT matching estimators used in this paper (cf. Becker and Ichino (2003))

Nearest Neighbour Matching (NNM)

We let T and C be the groups of treated and control /non-treated units and Y_i^T and Y_j^C j be the observed outcomes of interest of the T and C groups respectively. We denote $C_{(i)}$ the group of control units matched to the treated units i with an estimated value of the propensity score of $p_{(i)}$. Nearest Neighbour matching sets to minimise the absolute difference within this group of the propensity scores between the unit *i* of treated and unit *j* of the controls or

$$C_{(i)} = \min_{i} \ \| p_{(i)} - p_{(i)} \|$$

Here we denote the number of controls matched with observation $i \in T$ by N_i^C and define the weights $w_{ij} = \frac{1}{N_i^C}$ if $i \in C_{(i)}$ and $w_{ij} = 0$ otherwise.

The ATT using NNM is computed based on the formula

$$T^M = \frac{1}{N^T} \sum_{i \in T} \left[Y_i^T - \sum_{i \in C_{(i)}} w_{ij} Y_i^C \right]$$

where the number of units for the control group are denoted by N^T.

Stratification Matching (SM)

We let *q* index the blocks defined over the intervals of the propensity score.

Within each block the program computes

$$T_q^S = \frac{\sum i \in I(q) Y_i^T}{N_q^T} - \frac{\sum j \in I(q) Y_j^C}{N_q^C}$$

where I_q is the set of units in block *q* while N_q^T and N_q^C are the numbers of treated and control units in block *q*.

The ATT using SM is computed based on the formula

$$T^{S} = \sum_{q=1}^{Q} T_{q}^{S} \frac{\sum i \in I(q) D_{i}}{\sum \forall i D_{i}}$$

where the weight for each block is given by the corresponding fraction of treated units and Q is the number of blocks.

Kernel matching (KM)

The KM estimator is given by the formula

$$T^{K} = \frac{1}{N^{T}} \sum_{i \in T} \left[Y_{i}^{T} - \frac{\sum_{j \in C} Y_{j}^{C} G\left(\frac{p_{j} - p_{i}}{h_{n}}\right)}{\sum_{k \in C} G\left(\frac{p_{k} - p_{i}}{h_{n}}\right)} \right]$$

where the *G(.)* is a kernel function and h_n is a bandwidth parameter.

VATT-TUTKIMUKSIA -SARJASSA ILMESTYNEITÄ

PUBLISHED VATT RESEARCH REPORTS

74. Laine Veli – Uusitalo Roope: Kannustinloukku-uudistuksen vaikutukset työvoiman tarjontaan. Helsinki 2001.
75. Kemppi Heikki – Lehtilä Antti – Perrels Adriaan: Suomen kansallisen ilmasto-ohjelman taloudelliset vaikutukset. Vaiheen 2 loppuraportti. Helsinki 2001.
76. Milne David – Niskanen Esko – Verhoef Erik: Legal and Institutional Framework for Marginal Cost Pricing in Urban Transport in Europe. Helsinki 2001.
77. Ilmakunnas Seija – Romppanen Antti – Tuomala Juha: Työvoimapoliittisten toimenpiteiden vaikuttavuudesta ja ennakoinnista. Helsinki 2001.
78. Milne David – Niskanen Esko – Verhoef Erik: Acceptability of Fiscal and Financial Measures and Organisational Requirements for Demand Management. Helsinki 2001. (Not yet publiched).
79. Venetoklis Takis: Business Subsidies and Bureaucratic Behaviour. Helsinki 2001.
80. Riihelä Marja – Sullström Risto: Tuloerot ja eriarvoisuus suuralueilla pitkällä aikavälillä 1971-1998 ja erityisesti 1990-luvulla. Helsinki 2001.
81. Ruuskanen Petri: Sosiaalinen pääoma – käsitteet, suuntaukset ja mekanismit. Helsinki 2001.
82. Perrels Adriaan – Kemppi Heikki – Lehtilä Antti: Assessment of the Macro-economic Effects of Domestic Climate Policies for Finland. Helsinki 2001. Tulossa.
83. Venetoklis Takis: Business Subsidies and Bureaucratic Behaviour, A Revised Approach. Helsinki 2001.
84. Moisio Antti – Kangasharju Aki – Ahtonen Sanna-Mari: Menestyksen mitta? Vaihtoehtoisia mittareita aluetalouden kehitykselle. Helsinki 2001.
85. Tuomala Juha: Työvoimakoulutuksen vaikutus työttömien työllistymiseen. Helsinki 2002.
86. Ruotoistenmäki Riikka – Babygina Evgenia: The Actors and the Financial Affairs of the Northern Dimension. Helsinki 2002.
87. Kyyrä Tomi: Funktionaalinen tulonjako Suomessa. Helsinki 2002.
88. Räty Tarmo – Luoma Kalevi – Koskinen Ville – Järviö Maija-Liisa: Terveyskeskusten tuottavuus vuosina 1997 ja 1998 sekä tuottavuuseroja selittävät tekijät. Helsinki 2002.
89. Hakola Tuulia: Economic Incentives and Labour Market Transitions of the Aged Finnish Workforce. Helsinki 2002.
90. Venetoklis Takis: Public Policy Evaluation: Introduction to Quantitative Methodologies. Helsinki 2002.

91. Berghäll Elina – Heikkilä Tuomo – Hjerppe Reino – Kiander Jaakko – Kilponen Juha – Lavrac Vladimir – Stanovnik Peter: The Role of Science and Technology Policy in Small Economies. Helsinki 2002.

92. Räisänen Heikki (toim.): Rakenteellinen työttömyys. Tutkimusinventaari ja politiikkajohtopäätökset. Helsinki 2002.

93. Moisio Antti: Essays on Finnish Municipal Finance and Intergovernmental Grants. Helsinki 2002.

94. Parkkinen Pekka: Hoivapalvelut ja eläkemenot vuoteen 2050. Helsinki 2002.

95. Junka Teuvo: Maailman kilpailukykyisin maa? Tuottavuus ja investoinnit Suomessa 1975-2000. Helsinki 2003.

96. Cogan Josehp – McDevitt James: Science, Technology and Innovation Policies in Selected small European Countries. Helsinki 2003.

97. Perrels Adriaan – Kemppi Heikki: Liberalised Electricity Markets – Strengths and Weaknesses in Finland and Nordpool. Helsinki 2003.

98. Sarvimäki Matti: Euroopan Unionin itälaajentuminen ja maahanmuutto Suomeen. Helsinki 2003.

99. Räty Tarmo – Luoma Kalevi – Mäkinen Erkki – Vaarama Marja: The Factors Affecting the Use of Elderly Care and the Need for Resources by 2030 in Finland. Helsinki 2003.

100. van Beers Cees: The Role of Foreign Direct Investments on Small Countries' Competitive and Technological Position. Helsinki 2003.

101. Kangasharju Aki: Maksaako asumistuen saaja muita korkeampaa vuokraa? Helsinki 2003.

102. Honkatukia Juha – Forsström Juha – Tamminen Eero: Energiaverotuksen asema EU:n laajuisen päästökaupan yhteydessä. Loppuraportti. Helsinki 2003.

103. Simai Mihály (ed.): Practical Guide for Active National Policy Makers – what Science and Technology Policy Can and Cannot Do? Helsinki 2003.

104. Luoma Arto – Luoto Jani – Siivonen Erkki: Growth, Institutions and Productivity: An empirical analysis using the Bayesian approach. Helsinki 2003.

105. Montén Seppo – Tuomala Juha: Muuttoliike, työssäkäynti ja työvoimavarat Uudellamaalla. Helsinki 2003.