Testing for nonparametric identification of treatment effects in the presence of a quasi-instrument*

Xavier de Luna[a], Per Johansson[b]
[a]Umeå University, [b]Uppsala University and [b]IFAU, Uppsala
xavier.deluna@stat.umu.se
per.johansson@ifau.uu.se

September 1, 2010

Abstract

The identification of average causal effects of a treatment in observational studies is typically based either on the unconfoundedness assumption or on the availability of an instrument. When available, instruments may also used to test for the unconfoundedness assumption (exogeneity of the treatment). In this paper, we define variables which we call quasi-instruments because they allow us to test for the unconfoundedness assumption although they do not necessarily yield nonparametric identification of the average causal effect. A quasi-instrument is loosely an instrument whose relation to the treatment is allowed to be confounded by unobservables. We propose a test for the unconfoundedness assumption based on a quasi-instrument, and give conditions under which the test has power. We perform a simulation study and apply the results to an observational study of the effect of job practice on employment. Quasi-instrument assumptions are weaker than instrument assumptions, and the former should therefore be available more often in applications.

Working paper: preliminary version.

*The first author acknowledges the financial support of the Swedish Research Council through the Swedish Initiative for Research on Microdata in the Social and Medical Sciences (SIMSAM), the Ageing and Living Condition Program and grant 70246501.

1 Introduction

Identification of the causal effect of T (treatment) on an outcome Y in observational studies is typically based either on the unconfounded assumption (also called selection on observables, exogeneity, ignorability, see, e.g., de Luna and Johansson, 2006) or on the availability of an instrument. The unconfoundedness assumption says loosely that all the variables affecting both the treatment T and the outcome Y are observed (we call them covariates) and can be controlled for. An instrument is a variable affecting the treatment T and related to the outcome Y only through T (and possibly the observed covariates). Moreover, the effect of the instrument on the treatment must be unconfounded (i.e. given the observed covariates the instrument can be considered as randomized). When available instruments can be used to identify causal effects in parametric situations and hence also to test the unconfoundedness assumption. Such test is typically performed by comparing the estimates of the causal effects obtained both under the unconfoundedness assumption, and by using the instrument. Nonparametric identification is also possible with the help of instruments and, for instance, Angrist, Imbens and Rubin (1996) develop a theory for the nonparametric identification and estimation of local average causal effects. Frölich (2007) extended their results to the situation where the observed covariates are related to the instrument. Battistin and Retore (2008) and Dias, Ichimura and van den Berg (2007) use (fuzzy) discontinuity designs to obtain identification which can be considered as an instrumentation.

In this paper, we consider quasi-instruments, i.e. variables with the properties of an instrument except that their effect on the treatment T is allowed to be confounded. Therefore, in general quasi-instruments will not yield identification of a causal effect when the the unconfoundedness assumption does not hold, with the exception of parametric linear systems (Pearl, 2000, p.

248). On the other hand, we show that the availability of quasi-instruments (which are based on weaker assumptions than instruments) allows us to gain evidence for the validity of the unconfoundedness assumption, and we propose a test statistic for this purpose. The procedure introduced can be interpreted as testing whether the quasi-instrument (which must be dichotomized if not binary) has no effect on the outcome. Indeed, under the quasi-instrumental assumptions made, we show that such effect must be null if there are no unobserved confounders. The proposed test is related to the use of two control groups to test the unconfoundedness assumption, an idea previously used, e.g., in Rosenbaum (1987), de Luna and Johansson (2006) and Dias, Ichimura and van den Berg (2007). Rosenbaum (1987) was probably first to formalize the idea that two control groups provide information on the unconfoundedness assumption and described actual observational studies where different controls groups where available. Our main contribution in this context is to introduce mild assumptions (defining quasi-instruments) under which one control group can be split into two to test the unconfoundedness assumption non-parametrically.

The paper is organized as follows. Section 2 presents the model, defines quasi-instruments, and develops the theoretical results which allow us to then introduce a test of the unconfoundedness assumption. Section 3 presents a simulation study and Section 4 an illustrative case study on the estimation of the effect of job practice for unemployed on employment.

2 Theory and method

2.1 Model

Assume that the variables $(\mathbf{X}, Z, T, Y)$ are observed for all the units in an observational study, where Y is an outcome of interest, T is a treatment (a variable one may intervene upon), the vector $\mathbf{X}$ is a set of pre-treatment variables (typically characteristics describing the units and their context before treatment), and Z is not post-treatment.

We use the Neyman-Rubin model for causal inference. Assume that T is binary[1], i.e. taking values in $\mathcal{T} = \{0,1\}$. Let us define $Y(t)$, $t \in \mathcal{T}$, such that $Y = \sum_{t\in\mathcal{T}} Y(t)\mathcal{I}(T = t)$, where $\mathcal{I}(\cdot)$ is the indicator function. If T is a variable that one can intervene upon (thereby called treatment herein), $Y(t)$ (called potential outcome) is typically interpreted as the the outcome result of the intervention $T = t$. For a given t, we consider $(\mathbf{X}, Z, T, Y(t))$ as a random vector variable with a given joint distribution (possibly one may consider $Y(t)$ as fixed), from which a random sample is drawn. Parameters that may be identified are in this context population parameters and, for instance, the average treatment effect $E(Y(1) - Y(0))$ or the average treatment effect on the treated $E(Y(1) - Y(0) \mid T = 1)$ are parameters often targeted in applications.

In observational studies where treatment T is not randomized by definition, an identifying assumption (e.g., Rosenbaum and Rubin, 1983, Imbens, 2004) for the average treatment effect is:

(A.1) $\forall t \in \mathcal{T}$,

$$\begin{array}{ll} T \perp\!\!\!\perp Y(t)|\mathbf{X} & \text{(unconfoundedness)}, \\ \Pr(T = t \mid \mathbf{X}) > 0 & \text{(common support)}. \end{array}$$

The unconfoundedness assumption is often considered as realistic in situations where the set of characteristics $\mathbf{X}$ is rich enough, and when there is subject-matter theory to support the assumption. However, this indentifiability assumption is untestable without further assumptions and/or information.

[1]The results of this paper are straightforward to generalize to situation where T takes a finite set of values.

2.2 Quasi-instrument, test and power

Let us now consider situations where we have more information than just the unconfoundedness assumption discussed above. Let us thus assume that

(A.2) $\forall t \in \mathcal{T}$,

$$\begin{array}{l} Z \perp\!\!\!\perp Y(t)|\mathbf{X}, \\ \Pr(Z = t \mid \mathbf{X}) > 0). \end{array}$$

Assumption (A.2) prohibits, e.g., a direct effect from Z to Y, i.e. an effect not going through T, or unobserved variables affecting both Z and Y.

We further use a stability (Pearl, 2000, also called faithfulness in the graphical model literature) assumption:

(A.3) (A.1-2) $\Longrightarrow$ $(Z, T) \perp\!\!\!\perp Y(t)|\mathbf{X}, \quad t \in \mathcal{T}$.

Because $(Z, T) \perp\!\!\!\perp Y(t)|\mathbf{X}$ implies (A.1) and (A.2), (A.3) says that (A.1-2) hold if and only if $(Z, T) \perp\!\!\!\perp Y(t)|\mathbf{X}$ hold. Stability/faithfulness assumptions are typically made to be able to perform inference on graphical models.

Situations where (A.3) does not hold are, for instance, such that there is a deterministic relationship between some of the variables involved.

Let us now state a main result of this paper.

Proposition 1 *Assume (A.1-3), then*

$$Y(t) \perp\!\!\!\perp Z|T, \mathbf{X}, \quad t \in \mathcal{T}. \tag{1}$$

Proof. Direct consequence of Lemma ?? in Dawid (1979)/weak union property. ■

The conditional independence statement obtained in Proposition 1 is testable from the data given $T = t$ (see next section). Finding evidence in the data against (1) will imply that we reject the assumptions of the proposition. Thus, evidence against (1) can be interpreted as evidence against the unconfoundedness assumption (A.1) if (A.2) is known to hold from theoretical considerations (without the latter the test is rejecting either (A.1) and/or (A.2)).

For the test proposed below to have power we further need to assume:

(A.4) Z and T are dependent conditional on $\mathbf{X}$.

This assumption is typically also made for instrumental variables to be useful for identification. Here we have the following result.

Proposition 2 *Assume (A.2-3). Then,*

$$\{(1) \implies (A.1)\} \implies (A.4).$$

Thus, (A.4) is a necessary condition for the test to have power. Sufficient conditions (expressed graphically; see Lauritzen, 1996) are given in Figure 2.2.

Identification of the effect of T on Y is guaranteed under (A.2) and (A.3-4) with linear models, see, e.g., Pearl (2000, p. 248), in which case Z may be called instrument. This, however, is not true in general (even for non-parametric identification of local average treatment effects, Angrist, Imbens and Rubin, 1996) and we therefore call Z for which (A.2-4) hold a *quasi*-instrument.

2.3 Method

For the sake of simplicity, we consider the situation where the parameter of interest is the average treatment effect on the treated, $\theta = E(Y(1) - Y(0)|T = 1)$. Hence, assumptions (A.1-3) need to hold only for $t = 0$. Different strategies may be adopted to test the null hypothesis defined by the conditional independence statement of Proposition 1 with $t = 0$, i.e.

$$H_0 : \ Y(0) \perp\!\!\!\perp Z|T = 0, \mathbf{X}$$

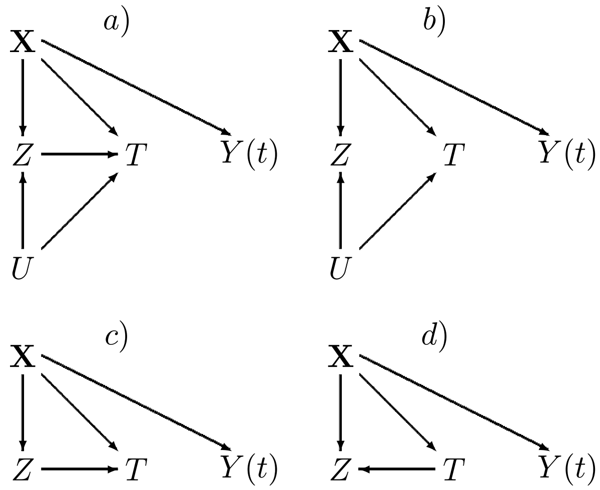

Figure 1: All four examples are such that (A.4) holds although only cases a), b) and c) are such that the test will have power, i.e. $Y(t) \perp\!\!\!\perp Z \mid T, \mathbf{X}$ does not hold if (A.1) does not hold.

One strategy could be to use the concept of two independent control groups (Rosenbaum, 1987). Under H_0 we can use Z to obtain two independent control groups (one defined by $Z = 1$ and one by $Z = 0$)[2] for estimating θ, yielding $\widehat{\theta}^{z=0}$ and $\hat{\theta}^{z=1}$. Under H_0 the difference $\widehat{\theta}^{z=0} - \hat{\theta}^{z=1}$ should have mean zero and this is the base for a test statistic. However, since we need to compute two non-parametric estimators of θ, the resulting statistic has poor finite sample properties, for instance, when the covariates have different support in the two control groups created.[3]

In this paper we propose a testing strategy based on fact that under H_0 we have $\delta(\mathbf{X}) = 0$, for all $\mathbf{X}$, where

$$\delta(\mathbf{X}) = E(y(0) \mid T = 0, \mathbf{X}, Z = 1) - E(y(0) \mid T = 0, \mathbf{X}, Z = 0).$$

[2]If Z is continuous then it may be made binary using a threshold. As noted by Rosenbaum (1987), the availability of two control groups provides information on the unconfoundedness assumption. In our particular case, this was shown in the previous section.

[3]This has been confirmed in simulation experiments not presented here.

Consider a non-parametric estimator for $\delta = E(\delta(\mathbf{X}))$[4]

$$\hat{\delta} = \sum_{i:Z=1} [y_i(0) - \hat{y}_i(0)] + \sum_{i:Z=0} [y_i(0) - \tilde{y}_i(0)],$$

where $\hat{y}_i(0)$ is a non-parametric estimator of $E(y_i(0) \mid T_i = 0, \mathbf{X}_i, Z_i = 0)$ and $\tilde{y}_i(0)$ is a non-parametric estimator of $E(y_i(0) \mid T_i = 0, \mathbf{X}_i, Z_i = 1)$. The two latter estimate may be obtained by nearest neighbour matching, or any other smoothing technique. Since $\delta = 0$ under H_0, the test statistic

$$C = \frac{\hat{\delta}}{s} \tag{2}$$

will, under the necessary regularity conditions, be normally distributed with mean zero and variance one, where s is the standard error of $\hat{\delta}$. For instance, if nearest neighbour matching estimator are used, then s was given in Abadie and Imbens (2006, Theorems 6 and 7). A subsampling estimator is also available in this case in de Luna, Johansson and Sjöstedt-de Luna (2010). We call in the sequel the test based on W a Chow test since it is related in nature to the tests developed in Chow (1960) and de Luna and Johansson (2006).

3 Monte Carlo study

We use a Monte Carlo study to investigate the finite sample properties (empirical size and power) of the Chow test (2). As a benchmark we also implement a parametric Durbin-Wu-Hausman (DWH) test where we first regress T on X and Z and then add the residuals from this fit as a covariate into the outcome equation for Y (using correctly specified models). The test for the unconfoundedness assumption is then a Wald test on the parameter for the included residual covariate (see, e.g., Wooldridge, 2003, Chap. 6). We use a robust covariance matrix (White, 1982).

[4] δ may be interpreted as the causal effect of Z on Y given $T = 0$.

3.1 Design

We use the following data generating process for unit i

$$\begin{aligned} Z_i &= I(U_{1i} + \varepsilon_{Zi} > 0), \\ T_i &= I(\beta_0 + X_i + 0.5Z_i + \beta_1 U_{1i} + \beta_2 U_{2i} + \varepsilon_{Ti} > 0), \end{aligned}$$

and

$$Y_i = 1 + X_i + \delta X_i^2 + \theta_i T_i + U_{2i} + \varepsilon_{Yi}.$$

We let ε_{Yi}, ε_{Zi}, ε_{Ti}, U_{1i}, U_{2i} be independently distributed as $N(0, 0.25)$. Moreover, we let also let $X_i \sim N(0, 2)$ and consider three cases for θ_i: $\theta_i = 1$, $\theta_i \sim N(1, 0.25)$, and $\theta_i = 1 + X_i$. Paramaters are varied in the study as follows: $\beta_0 \in \{-1.5, -1\}$, $\beta_1 \in \{0, 0.5\}$, $\beta_2 \in \{0, 0.1, 0.2, 0.3, 0.6, 0.9, 1.5, 2\}$, and $\delta \in \{0, 1/3\}$. We consider sample sizes $N = 500, 1500$ and 3000. The number of replicates are set throughout to 10000.

The treatment is confounded when $\beta_1 \neq 0$. Note that Z is an instrument when $\beta_1 = 0$, and only then is the nonparametric instrumental estimator suggested by Frölich (2007) consistent for the estimation of the local average treatment effect (Angrist et al., 1996).

3.2 Results

The results from the Monte Carlo study are displayed in Figures 2 to 5. Each figure displays the size ($\beta_2 = 0$) and power. We start by discussing the results for the "linear" model ($\delta = 0$). Figures 2 and 3 display the results for the situation when Z is an instrument ($\beta_1 = 0$) and a quasi-instrument ($\beta_1 = 0.5$), respectively. We can see the same pattern in both figures: both tests have the correct size when the heterogeneous treatment effect is not depending on the covariate. Power for both tests is increasing with the fraction of treated (β_0). The power of the DWH test is, as expected, larger than for the nonparametric Chow test, but the latter performs well considering its nonparametric nature. The results when the treatment effects depend on X are displayed in the top panels. We see that the DWH test has too large size.

Figures (4) and (5) display the same results for $\delta = 1/3$. Note that here, the DWH test is based on the estimation of the same model as above, that is we erroneously assume that response Y is linear in X. Again the pattern when Z is an instrument (4) and a quasi instrument (5) are similar. Moreover, from both figures we can see that the DWH test has too large size, and power decreasing with β_2. The nonparametric Chow test has the correct size in all situations. The power of the test is lower compared to the linear setup.

4 Application: Effect of job practice

We consider a case study where the interest lies in estimating the effect of job practice for unemployed on employment status. Job practice (JP) was offered within two separate labor market training (LMT) programs in Sweden in 1998. One program was run by the regular program provider in Sweden; the Swedish National Labor Market Board (AMV). The other program was offered by the Federation of Swedish Industries (Swit).

To be eligible to the programs the unemployed individuals must be at least 20 years of age and enrolled at the public employment service. The weekly cost for the Swit and the AMV programs were on average €273 and €289, respectively (Näringsdepartementet, 1999; the Swit-yrkesutbildning, 2000). This cost does not include unemployment benefits which are received when an individual is openly unemployed or in a LMT program. There was no difference in benefits for the two groups of trainees.

The fundamental idea with the Swit program was to increase the contacts between the unemployed individual and employers by providing job practice.

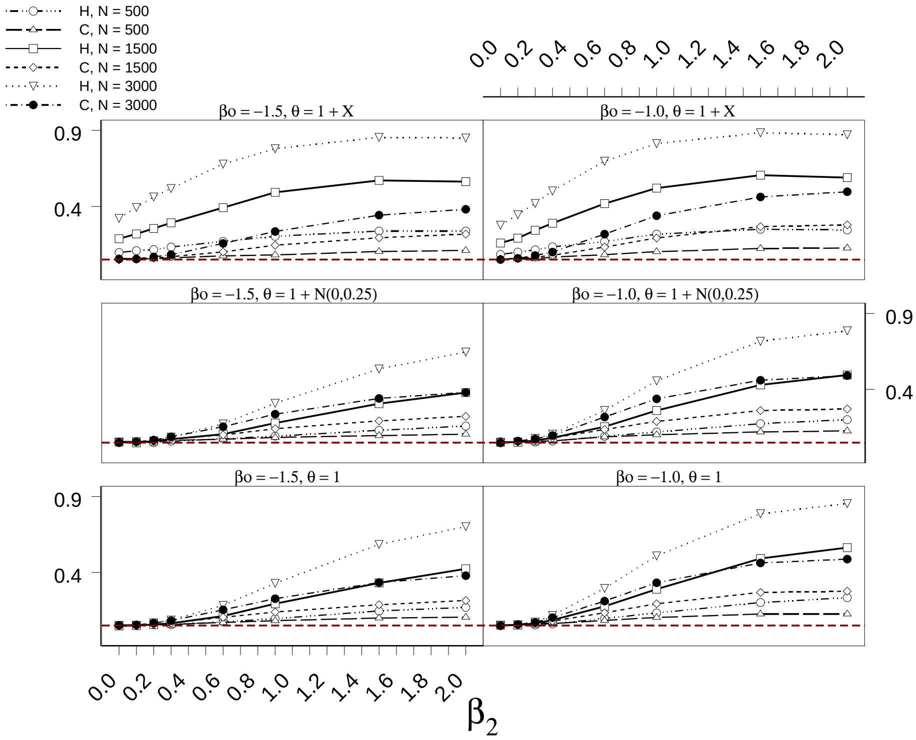

Figure 2: Size and power for conditional Chow test (C) and Hausman (H) test (based on robust covariance matrix) when Z is an instrument (i.e. $\beta_1 = 0$) and with a DGP that is linear in X (i.e. $\delta = 0$).

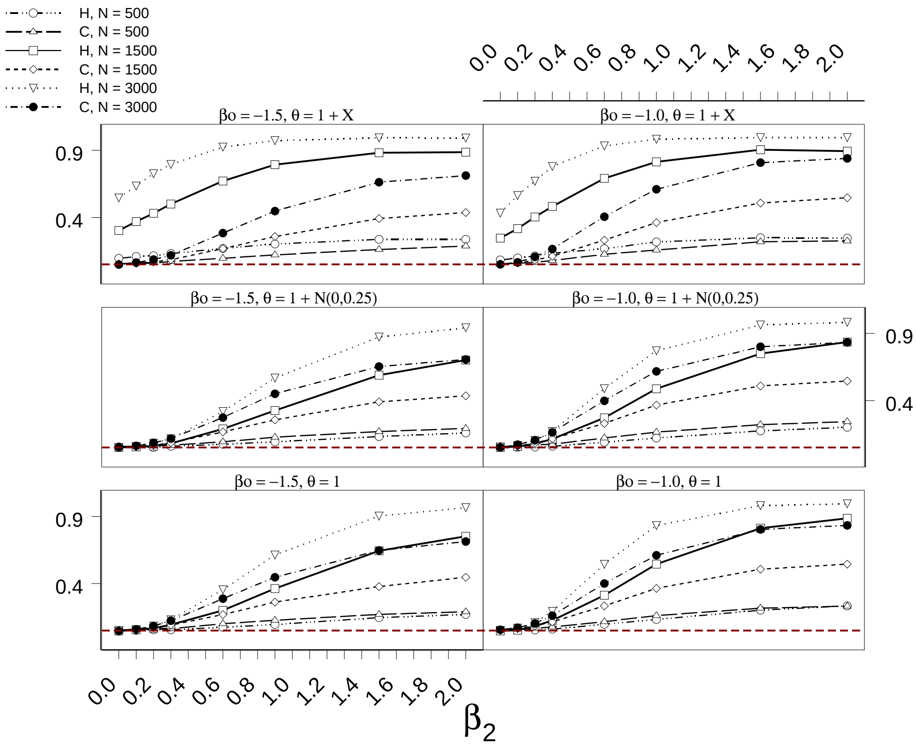

Figure 3: Size and power for conditional Chow test (C) and Hausman (H) test (based on robust covariance matrix) when Z is a quasi instrument (i.e. $\beta_1 = 0.5$) and with a DGP that is linear in X (i.e. $\delta = 0$).

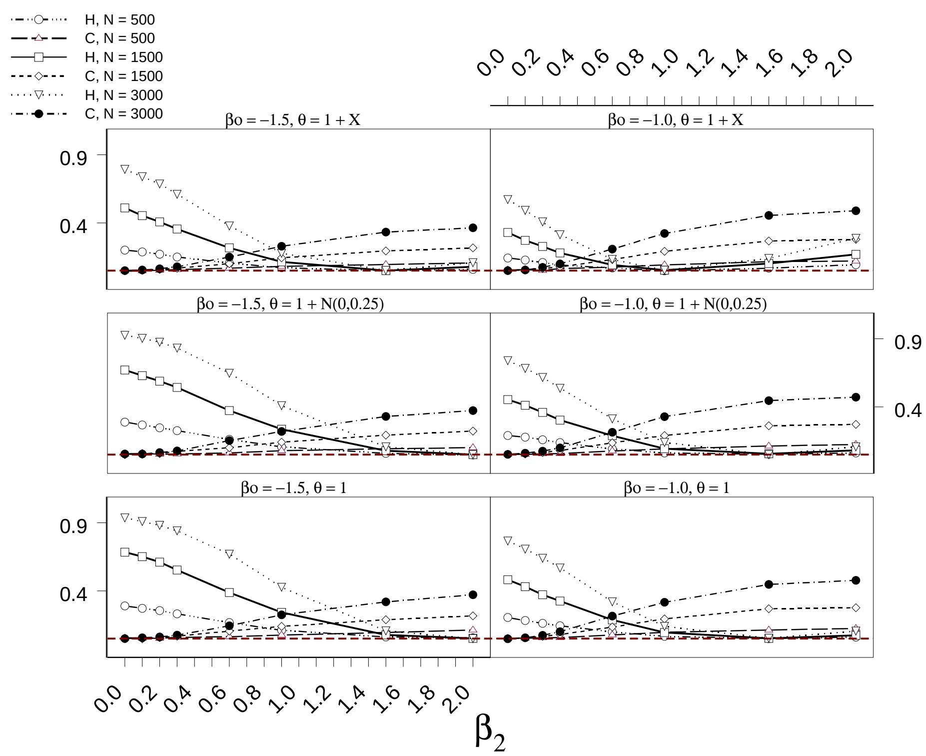

Figure 4: Size and power for conditional Chow test (C) and Hausman (H) test (based on robust covariance matrix) when Z is an instrument (i.e. $\beta_1 = 0.5$) and with a DGP that is non-linear in X (i.e. $\delta = 1/3$).

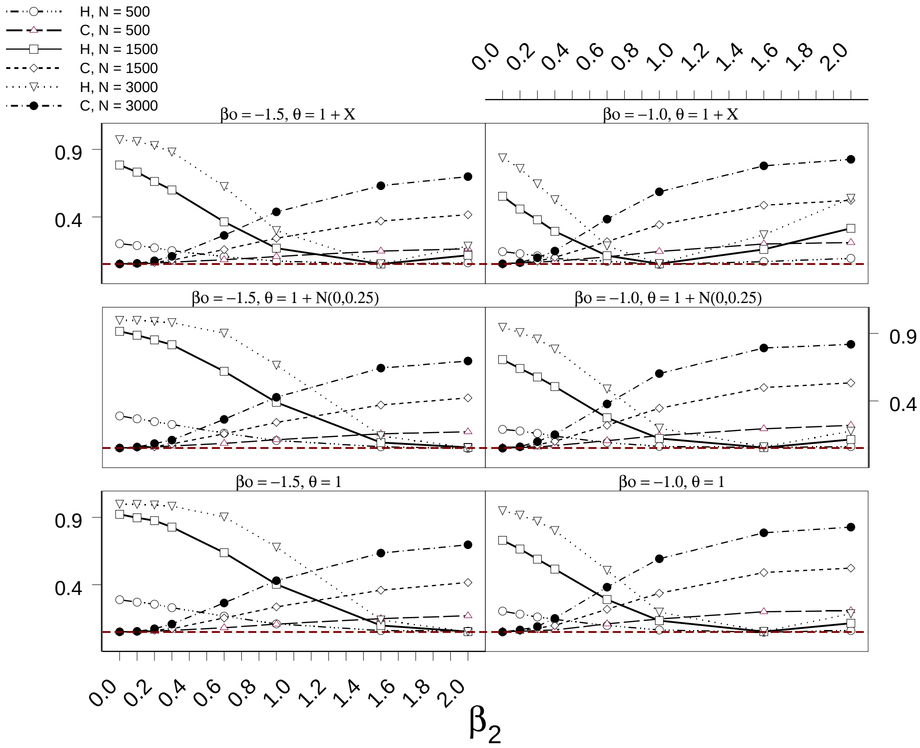

Figure 5: Size and power for conditional Chow test (C) and Hausman (H) test (based on robust covariance matrix) when Z is a quasi instrument (i.e. $\beta_1 = 0.5$) and with a DGP that is non-linear in X (i.e. $\delta = 1/3$).

Table 1: The frequency distribution of the courses within the two programs.

	AMV ($n = 796$)	Swit ($n = 794$)
Programmer	32	27
Computer technician	31	29
Application support	10	16
IT-pedagogue	2	6
IT-entrepreneur	1	3
Other	17	15
Missing	7	4

From a survey conducted in June 2000 on $1,000$ program participants from either program it can be seen that 69.5 percent of the Swit participants and 52 percent of the AMV participants stated that they participated in JP.[5] Except for the idea to provide more contacts with employers the two programs are similar. Both programs tested the individual's motivation and ability before being recruited to the programs by similar selection procedures (see Johansson, 2008, for a thorough description of the selection). The types of courses given within the Swit and the AMV programs are displayed in Table 1. The similarities of the two programs are apparent. Thus, despite the differences in procurement between the two organizations (the Swit and the AMV), there do not seem to be any large differences between the types of labor market training courses which suggest that they can be used as a quasi-instrument to test for if selection to job practice can be controlled for by observed covariates.

Based on the survey one can see in Table 2 that there is a statistical significant 18.1 percentage points difference in employment six months after

[5] A thorough description of the survey can be found Johansson and Martinson (2000). The response rates were 79.4 and 79.6 percent for participants in the Swit and the AMV, respectively. The survey contained a total of 19 questions. These concerned i) the individual's background, ii) the individual's labor market training and iii) the individual's present labor market situation.

leaving the program (the two programs have same average length). In the table we have some individual background variables: (i) education, (ii) work handicap (see disabled), (iii) gender (1 if man and 0 if women) and (iv) immigration status (1 if immigrant 0 else). Finally we have information on the individuals region of residence. We have divided Sweden into four regions: Stockholm, Skåne, Västra Götaland and the rest of the country. Stockholm, Skåne and Västra Götaland are the tree regions with the largest population and with, in general, the best labor market opportunities.

We can see some average differences between the two samples. Those with job practice are: (i) less disabled and (ii) less likely to live in Stockholm. The level of education also differs: they have on average more compulsory and upper secondary education but also less college education than those with no JP. Based on these average differences, it is difficult to argue that those with JP have better labor market prospects without JP. The single factor suggesting the JP population have better labor opportunities without JP is that they are less likely disabled. In order to further study the selection into JP we used the covariates from the table and estimated a logit regression model. The results from this estimation (not displayed) are that individuals who are: (i) from Stockholm or Västra Götaland, and (ii) disabled, are less likely to receive JP. There is no statistical significant (5 percent) differences in education between the two groups. In Figure 1 we display the propensity score for the two samples, which give evidence for the common support assumption.

In order to estimate the average treatment effect on the treated (ATT) of JP we use the one to one matching (hence $M = 1$) estimator. The matching is performed by using the minimum Mahalanobis distance using the covariates displayed in Table 2 and the estimated propensity score. The ATT is estimated to 24.3 percentage points with both estimators. Based on estima-

Table 2: Descriptive statistics.

Job Practice	Yes	No	Diff	t-test
Employment	64.9	46.8	18.1	6.82
	Education			
Compulsory	5.1	7.6	-2.5	-1.9
Upper secondary	67.8	62.1	5.7	2.19
College	27.1	30.3	-3.2	-1.3
Disabled	7.5	11.5	-4.0	-2.5
Man	62.1	61.9	0.2	0.1
Immigrant	5.6	6.4	-0.9	-0.7
Stockholm	21.4	27.8	-6.5	-2.7
Skåne	10.6	8.3	2.3	1.5
Västra Götaland	13.8	16.5	-2.6	-1.3
Rest of the country	54.2	47.3	2.7	2.53
Sample size	969	528		

tor of the unconditional variance given in Abadie and Imbens (2006) we also find that this estimate is statistical significant for both estimators (t-ratio 4.9 and 4.8 using the covariates and propensity score, respectively). Hence, the estimated effect is larger after matching than before.

4.1 Testing the unconfoundedness assumption

Here we test for (A.1) used in the estimation of the ATT above. We start by displaying in Table 3 descriptive statistics where we have conditioned on the quasi-instrument. From this table we can see some differences with respect to Swit and AMV participants. The Swit participants with JP when compared to AMV: (i) have a 5.8 (67.2 - 61.4) percentage points higher employment; (ii) have more upper secondary education, (iii) have less individuals with a college degrees, (iii) are less likely an immigrant and (iv) are more (less) likely to live in Skåne and Västra Götaland (Stockholm). When we condition on No JP there are less statistical significant differences than when conditio-

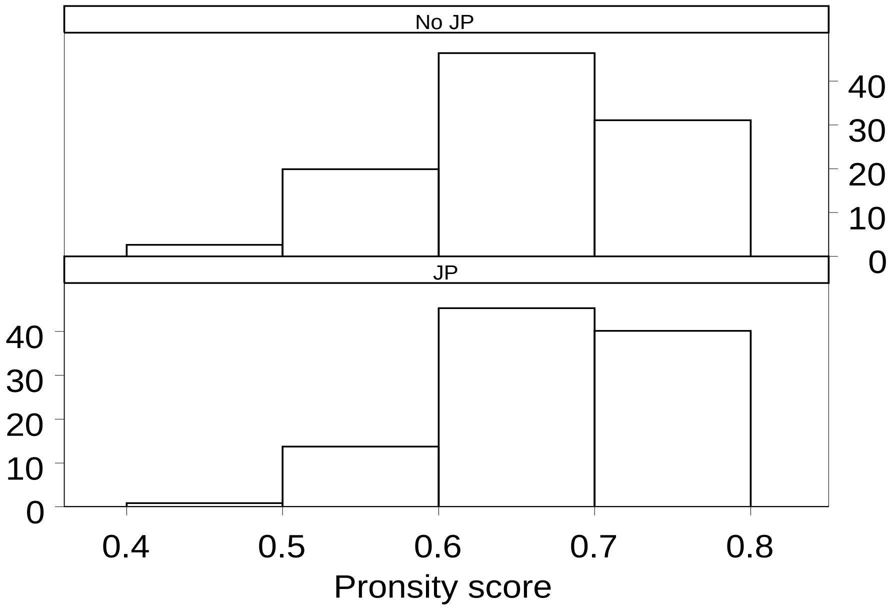

Figure 6: The distribution (percent) of the propensity scores for the treated (JP) and the non treated (No JP).

Table 3: Descriptive statistics. Proportion in

	Job Practice			No Job Practice		
Swit (Z)	Yes	No	t-test	yes	no	t-test
Employment	67.2	61.4	2.9	54.9	42.8	5.1
	Education					
Compulsory	4.1	8.2	-1.6	6.3	1.8	-0.8
Upper secondary	66.5	69.7	4.0	56.0	65.2	1.8
College	29.3	37.7	-3.2	23.8	26.6	-1.2
Disabled	4.3	2.9	1.4	12.5	15.9	-1.7
Man	59.7	58.9	0.3	65.8	63.5	0.9
Immigrant	5.1	8.0	-2.1	6.3	5.7	0.5
Stockholm	24.7	46.3	-8.3	16.2	18.7	-1.2
Skåne	12.6	7.4	3.3	7.6	8.8	-0.8
Västra Götaland	8.0	4.0	3.3	22.7	22.7	0.0
Rest of the country	54.6	53.5	0.3	42.2	49.9	-1.6
Sample size	586	383		175	353	

ning on JP. A reason for this may be the smaller sample size. The general pattern concerning education and gender distribution is the same as when conditioning on JP however.

We use the selection into the two programs to test the unconfoundedness assumption using the test statistic (2), resulting in a p-value of 0.32. We also perform a DWH test by estimating a linear probability model with the discrete covariates displayed in table 3, yielding a p-value of 0.09. Thus, none of the test can reject the null hypothesis that the effect of job practice on employment is not confounded at the 5% level, although the DWH test by making stronger assumptions has a p-value under 10%.

5 References

Abadie, A. and Imbens, G. (2006). "On the Failure of the Bootstrap for Matching Estimators". *Econometrica* **76**, 1537-155

Angrist, D. Imbens, G. and Rubin W (1996). "Identification of Causal Effects Using Instrumental Variables". *JASA* 91, 444-455

Battistin, E. and Retore, E. (2008). "Ineligibles and Eligible Non-Participants as a Double Comparison Group in Regression Discontinuity Designs", *Journal of Econometrics*, **142**, 715-730.

Chow G. C. (1960). "Tests of Equality Between Sets of Coefficients in Two Linear Regressions". *Econometrica* **28**: 591–605.

Dawid, A. P. (1979). Conditional independence in statistical theory, *Journal of the Royal Statistical Society Ser. B* **41**, $1-31$.

de Luna, X. and Johansson P. (2006). "Exogeneity in Structural Equation Models", *Journal of Econometrics* **132** 527-543.

de Luna, X., Johansson P. and Sjöstedt-de Luna S. (2010). Bootstrap inference for K-nearest neighbour matching estimators. Working paper.

Dias M., Ichimura H. and van den Berg G. (2007). The matching method for treatment evaluation with selective participation and ineligibles. Working paper 2008:6, IFAU.

Frölich, M. (2007): "Nonparametric IV estimation of local average treatment effects with covariates" *Journal of Econometrics* **139**, 35-75.

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

Johansson, P. and Martinson, S. (2000). "Det nationella IT-programmet - en slutrapport om the Swit" *Forskningsrapporter* 2000:8. Institute of Labor market Policy Evaluation.

Johansson, P. (2008). "The importance of employer contacts: Evidence based on selection on observables and internal replication". *Labour Economics* **15**, 2008, 350-369.

Lauritzen, S. L. (1996). *Graphical Models.* Oxford: Oxford University Press.

Näringsdepartementet (1999), *Underlag om det Nationella Programmet för IT-utbildning (the Swit).* Arbetsmarknadsenheten

Swit-yrkesutbildning (2000). *En Presentation av Organisation, Arbetsmetodik och Gjorda Erfarenheter*

Pearl, J. (200). *Causality: Models, Reasoning, and Inference.* Cambridge University press Cambridge UK.

Rosenbaum (1987). "The Role of a Second Control Group in an Observational Study (with discussion)". Statistical Science **2** 292-316.

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

Wooldrige, J.(2001). *Econometric Analysis of Cross Section and Panel Data.* The MIT Press Cambridge, Massachusetts.

White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models", *Econometrica* **50,** $1 - 25$