

Identification of Dynamic Treatment Effects by Instrumental Variables

Ruth Miquel

**Swiss Institute for International Economics and Applied Economic Research (SIAW),
University of St. Gallen**

First version: May 2002

Revised version: January 2003

Date this version has been printed: 16 January 2003

Address for correspondence

Ruth Miquel

Swiss Institute for International Economics and Applied Economic Research (SIAW)

University of St. Gallen

Dufourstr. 48, CH-9000 St. Gallen, Switzerland

Ruth.Miquel@unisg.ch

www.siaw.unisg.ch/lechner/miquel

Abstract:

This paper considers the problem of the identification of causal effects using instrumental variables. We are interested in the effects of some treatments on certain outcomes. First, we consider that a participation in a treatment or a program is only possible one time but we have the choice between more than one program. Under a monotonicity condition and an exclusion restriction, pair-wise Local Average Treatment Effects are identifiable. Second, we consider the case where only one program is available but more than one participation is possible, leading to a comparison of sequences of participations (or sequences of programs). In this framework a problem of endogeneity appears: the outcome after one period, affected by the participation in this period, can influence the participation in the next period. Under different versions of the monotonicity condition and the exclusion restriction, identification of the causal effects of sequences of programs are investigated. The introduction of a second period implies a loss of identification for some effects of interest even without any endogeneity problem.

Keywords: Compliers, Local Average Treatment Effect, dynamic treatment regimes, nonparametric identification, instruments

JEL Classification : C40

I am grateful to Michael Lechner, Markus Frölich, Conny Wunsch, Dragana Djurdjevic and Gauthier Lanot for very helpful comments. A previous version of this paper has been presented at the “7th Colloque des jeunes économètres”, Mont st. Odile. All remaining errors are mine.

1) Introduction

The estimation and identification of causal effects is often encountered in conjunction with the evaluation of some treatments or programs, as for example vocational training programs. The literature dealing with this issue increased a lot since the last twenty years. The aim is to see if the application of a program is beneficial to a particular population. A summary of the principal methods and problems associated with program evaluation can be found in Angrist and Krueger (1999), Heckman, LaLonde and Smith (1999) and Lechner and Pfeiffer (2001). The evaluation problem is usually examined with Rubin's framework of potential outcomes (Roy 1951, Rubin 1974). Assuming that two possible states of nature exist, the outcome of an individual is described by two potential outcomes, each depending on or "defined" by one state of nature. The observed or realized outcome is one of these potential outcomes according to the realization of the state of nature or, in other words, according to the program applied. Then the effect of the program is computed as the difference between the potential outcomes. For example, for the evaluation of training programs we want to compute the effect of participating in the program compared to not participating. Thus, we need to compute the difference of the two potential outcomes defined in the state of nature "participation" and in the state of nature "no participation". As already mentioned, only one outcome is observable. Therefore, we have to state some hypotheses to be able to construct this difference and to estimate the causal effect of the program.

One possible approach is to use instrumental variables (if available). In this framework different effects are estimated depending on the assumptions made. Imbens and Angrist (1991, 1994) and Angrist, Imbens and Rubin (1996) have proved that the effect identified with instrumental variable restrictions is the Local Average Treatment Effect (LATE), an observable effect for an unobservable population: the compliers or the individuals for whom a change in the instrument's value implies a change in the treatment status. A second approach using instrumental variables is the Local Instrumental Variables method (LIV)¹ introduced by Heckman and Vytlacil (Heckman 1997, Heckman and Vytlacil 1999, 2000 a-b). With this method they extend the linear IV analysis and they are able to estimate a wider class of treatment parameters. A huge literature applies or extends these methods, among others Abadie (2000, 2002), Abadie, Angrist and Imbens (1998) Heckman, Tobias and Vytlacil (2000), Manski and Pepper (2000), Vytlacil (2002). Although these authors present very interesting results, for example that LATE allows the estimation of quantiles, LATE and LIV

are equivalent, all these papers work in a static framework. They do not study the impact of multiple participations over time and the influence of previous program participations. In fact, the effects of the programs attended by the individuals in the past have an influence on the choice of the next program and this influence cannot be handled in such a framework.

Although the dynamic character of the problem can be treated in a parametric framework (panel data models like duration models, van den Berg, 2000, Arellano and Honoré, 2002), few work has been done in a nonparametric framework. The origin can be found in the series of papers from Robins (1986, 1989, 1997) and Robins, Greenland and Hu (1999) in Biometrics. These papers use counterfactual outcomes in a dynamic framework using some sequential randomization to identify the effects of interest. A more complete work on the identification problem of average treatment effects in a counterfactual dynamic framework can be found in Lechner and Miquel (2001). The effects of sequences of programs are identified only for certain populations of participants when the dynamic aspect of the problem is taken into account (the outcome of one period, itself influenced by the participation in this period, affects the participation in the next period). But until now, no work has been done in the area of IV methods in a dynamic framework.

This paper presents an extension of the static case (one-period case) where now many programs or treatments are available (instead of only one program) and presents the necessary conditions to identify the effects of sequences of programs (many periods). The identified effects look like some Local Average Treatment Effects. In the static framework, the LATE is the effect of a program for the compliers, a population that changes its participation with a change in the value of an instrumental variable. In the multi-period framework we need to redefine which persons are compliers and some other similar populations. With these definitions, we establish the assumptions needed for the identification of the LATEs. Unfortunately, we are not able to identify all effects of interest whether we are in a nondynamic case where the outcome of the previous period does not affect the participation in the second period, or not. The paper is restricted to three periods. The more general case (with T periods) will generate no other problems or results as those presented here but will be more difficult to handle due to the large number of sequences of programs and possible instruments.

The next section of the paper introduces the notation used in the paper. Section 3 recalls the results of the one-period static case. Section 4 extends the one-period case with one program to a one-period case with multiple programs. Section 5 provides the assumptions and

¹ This method directly estimates the Marginal Treatment Effect.

identification results in a multi-period framework with one instrument available. Section 6 presents different sets of assumptions to identify the effects in a multi-period case with two instruments available. These sets of conditions are differentiated depending on the sequentiality and dynamics introduced in the process of the participation decision. The last section concludes.

2) Notation

One of the difficulties arising when we work in a dynamic framework is a trade-off between the degree of complexity of the notation required to correctly describe the problem with all its facets and the need of clarity. The notation used in this paper is based on the one introduced in Lechner and Miquel (2001). Without imposing any restrictions on the problem we can simplify it to the case of three discrete time periods denoted by the subscript $t = 0, 1, 2$. For each period only one training program exists. So each period the choice is between participating or not participating in this program. (This general framework will be modified a little in the next section.) Three different types of variables are available.

The first type consists of instrumental variables. These variables fulfill particular exclusion restrictions, which we will discuss later. These exclusion restrictions vary depending on the dynamic framework used. We limit us to the case of binary variables, i.e. to the case where the instrument variables are dummy variables. Moreover, the number of available instruments varies depending on the case studied. We can dispose of one or two instruments (and in the remainder of this section we treat the two-instruments case). We relate those variables to the time period. The first instrument will have the subscript 1 and the second will have the subscript 2 because the second one is available only in the second period. The random vector of the instrumental variables in all periods is represented by $\underline{Z}_2 = (Z_1, Z_2)$, $Z_t \in \{0, 1\}$. A particular value of the instrument at time t is denoted z_t .²

The second type of variables relates to the participation in the program or treatment. $\underline{S}_2(Z) = (0, S_1(Z_1), S_2(Z_2))$ represents the history of potential participations. In each period the participation indicator takes either of two values: 0, when the person does not participate in the program and 1, when she does, $S_t(Z_t) \in \{0, 1\}$. In the first period nobody participates in

² For all variables a value is represented by a lower case letter and the random variable by an upper case letter.

the program, hence $S_0 = 0$. We can represent the dependence of the participation indicator on the instrument using such notation because we make an additional assumption. Three different dependencies are possible and will be investigated³: firstly, the participation in both periods depends only on the first instrument (only one instrument is available); secondly, each participation indicator depends on one instrument (the participation in the first period depends on the first instrument and the participation in the second period depends only on the second instrument), and finally, the participation in the second period depends on both instruments. In the second type of dependencies, the participation in the second period depends only on the second instrumental variable, Z_2 , although the first instrument is still available. The first period participation indicator depends only on the first instrument because in the first period only this instrument is available. Due to the dependence of the participation indicator on the instrument, $S_t(Z_t)$ represents a potential participation, the potentiality being defined in terms of the instrument. $S_t(z_t)$ is still a random variable in the special world in which the instrumental variable takes the particular value z at time t . The observed participation indicator in one period is denoted by S_t . Sequences of programs are simply denoted by a sub-vector of the history: $\underline{S}_1(\underline{Z}_1) = (0, S_1(Z_1))$, $\underline{S}_2(\underline{Z}_2) = (0, S_1(Z_1), S_2(Z_2))$. We have to notice that the notation used for the instrumental variable $\underline{Z}_t$ is a misuse of the notation. The “underline notation” depicts the history of a variable up to a certain time. Using $\underline{S}_t(\underline{Z}_t)$ we are not saying that the dependence applies to all the history of Z up to period t , but that each element of this history has an influence on a corresponding element of the history of the participations.⁴ We also neglect the subscript for individuals. The participation indicator for an individual depends only on the value of her own instrumental variable. The instrument for the other individuals has no influence on her participation: we implicitly assume that the SUTVA assumption from Rubin (1974) holds.⁵

The third category of variables represents outcomes. The impact of the treatments is modeled using the concept of potential variables. Each sequence of treatments defines a different world. In each of these worlds a random outcome exists. Thus, each of these outcomes is regarded as a potential outcome, as only one of the states of the world is realized. In the world characterized by a particular value of the sequence $\underline{z}_t$, the potential outcome at time τ ($\tau \geq t$)

³ The second type of dependences is used for the presentation of the notation.

⁴ Note that for the participation indicator, $\underline{S}_t(\cdot)$ represents the history of participations up to period t .

is $Y_\tau^{\underline{s}}$ (for a detailed discussion of the choice of the time for the sequence and the outcome see Lechner and Miquel, 2001). The observed outcome at τ is represented by Y_τ .

In our three-time period framework, the following sequences of treatments are possible: $\underline{s}_1 \in \{(0,1), (0,0)\}$; $\underline{s}_2 \in \{(0,1,1), (0,0,1), (0,1,0), (0,0,0)\}$, implying six potential outcomes: $Y_\tau^{(0,1)}, Y_\tau^{(0,0)}, Y_\tau^{(0,1,1)}, Y_\tau^{(0,0,1)}, Y_\tau^{(0,1,0)}, Y_\tau^{(0,0,0)}$.

To simplify the exposition the effects are summarized in the following notation:

$$\theta_{\tau}^{\underline{s}, \tilde{s}}(M = m) = E(Y_\tau^{\underline{s}} - Y_\tau^{\tilde{s}} | M = m)$$

where M represents functions of the participation indicators.⁶

The relations between the observable outcome, Y_τ and the potential outcomes defined in Lechner and Miquel (2001) are valid in our framework. The following equations hold in the case of three time periods. To simplify the notation we suppress the first period participation indicator (it is zero for everybody) in the remainder of the paper.

$$\begin{aligned} Y_\tau &= S_1(z_1)Y_\tau^1 + (1 - S_1(z_1))Y_\tau^0 \\ &= S_1(z_1)S_2(z_2)Y_\tau^{11} + S_1(z_1)(1 - S_2(z_2))Y_\tau^{10} + (1 - S_1(z_1))(1 - S_2(z_2))Y_\tau^{00} + (1 - S_1(z_1))S_2(z_2)Y_\tau^{01}. \end{aligned}$$

In the following we will denote period $t = 0$ “start period” and period $t = 1$ “first period”. In the start period nothing happens and in the first and second period participation becomes possible.

3) The static case using the multi-period notation

This part restates the “static case” using the multi-period notation. We consider only the first period, hence, the participation in the program is only specified for this period. The participation for the second period remains unspecified. To identify the LATE we need the following assumptions using Dawid’s (1979) notation of independence $\perp\!\!\!\perp$:

⁵ We use a modified version (for the multi-period case) of the Stable Unit Treatment Value Assumption: a) if $\underline{Z}_{2i} = \underline{Z}_{2i}'$ then $\underline{S}_{2i}(Z) = \underline{S}_{2i}(Z')$ and b) if $\underline{Z}_{2i} = \underline{Z}_{2i}'$ and $\underline{S}_{2i} = \underline{S}_{2i}'$ then $Y_{ti}^{\underline{s}_2} = Y_{ti}^{\underline{s}_2'}$.

⁶ For example $M = \underline{S}_2$ and $m = (0,1,1)$.

ASSUMPTION 1: (STATIC INDEPENDENCE CONDITION)

$$(S_1(z_1), Y_\tau^1, Y_\tau^0), (S_1(0), S_1(1)) \perp\!\!\!\perp Z_1, \forall z_1 = 1, 0.$$

ASSUMPTION 2: (ONE PERIOD MONOTONICITY CONDITION):

$$P\left[S_1(z_1) \geq S_1(z_1')\right] = 1 \text{ or } P\left[S_1(z_1) \leq S_1(z_1')\right] = 1; z_1, z_1' \in \{0, 1\}.$$

These assumptions are those stated by Imbens and Angrist (1994). The variables are indexed only with a time subscript. The static independence condition (SIC), which is the first part of condition 1 in Imbens and Angrist (1994), defines an exclusion restriction. The joint distribution of the potential outcome and the potential participation indicator is independent of the instrument in the first period. Also, the joint distribution of the potential participations is independent of the instrument in the first period. The monotonicity condition (OMC), which is the second condition in the above mentioned paper, ensures a monotonic influence of the instrument on the participation. If it is more likely to participate when $Z_t = z_t$ than when $Z_t = z_t'$, then each person who participates when $Z_t = z_t'$ should also participate when $Z_t = z_t$. In the terminology of Imbens and Angrist this hypothesis excludes the subpopulation of defiers (if the instrument corresponds to an assignment in a program, these individuals will do the opposite of it). Both assumptions are untestable.

Theorem 1: UNDER ASSUMPTION 1 (SIC) AND 2 (OMC), THE EFFECT $\theta_t^{1,0} (S_1(1) - S_1(0) = 1)$ IS IDENTIFIED.

The proof can be found in Imbens and Angrist. Theorem 1 states, that the local average treatment effect for the compliers in the first period is identified. If we assume the instrument to be “should participate”, when taking the value 1, and “should not participate”, when taking the value 0, then the “compliers in the first period” are the subpopulation that follows the “assignment” of the instrument in the first period. The participation indicator S_1 will be equal to 1 and 0 respectively when Z_1 is equal to 1 and 0 respectively. Note that we do not specify the behavior of the compliers in the second period. We only want to differentiate the population on the basis of their behavior in the first period.

⁷ In Imbens and Angrist (1994) this condition is presented in the following form $S_t(z_t) \geq S_t(z_t')$ or $S_t(z_t) \leq S_t(z_t')$; $z_t, z_t' \in \{0, 1\}, z_t \neq z_t'$.

4) The static case with multiple programs

Let us stay in this one-period case with only one instrument and look at multiple alternatives faced by the individuals. Now, they have the choice to participate in one among several programs. As an example, consider only for this section that four different programs are available. The participation indicator S_1 equals zero when the individual does not participate in any program, and it equals one, two or three depending on the program attended, $S_1 \in \{0, 1, 2, 3\}$.⁸ The potential participation indicators, $S_1(1), S_1(0)$, can also take four different values and also four potential outcomes, Y^0, Y^1, Y^2, Y^3 , are now available. In this context identification can only be obtained if we place us again in a binary world. We have to restrict ourselves to a case where we know that the choice is only between two alternatives. To place ourselves once again in a binary world, where only two programs are considered, we introduce new binary variables, $\tilde{S}_{kl}(1), \tilde{S}_{kl}(0), \tilde{S}_{kl}$, $k, l = 0, 1, 2, 3$. The first two groups are potential variables (they depend on the value of the instrument Z_1) and the last one is observed. When the participation choice is restricted to be between programs k and l , these variables equal 1 when $S_1(1), S_1(0), S_1$ respectively take the value k and equal 0 when $S_1(1), S_1(0), S_1$ respectively take the value l . When we compare pairs of programs, they indicate the participation in one of the programs of the pair. As we only compare two programs we can use the terminology of the one-program case and rename one of our choices “participation” and the other one “nonparticipation” in the pair considered. In the following, these variables are called “pair participation indicators” and when we mention a “pair participation” we talk about the participation in one program instead of in a second program in a pair-wise comparison (only two programs are available for the choice).

ASSUMPTION 3: (CONDITIONAL - STATIC INDEPENDENCE CONDITION)

$$(\tilde{S}_{kl}(z_1), Y_{\tau}^k, Y_{\tau}^l), (\tilde{S}_{kl}(1), \tilde{S}_{kl}(0))) \perp\!\!\!\perp Z_1 \mid S_1 = \{k, l\}, \forall z_1, k, l.$$

The hypotheses needed for the identification of effects are still an independence and a monotonicity condition. The first one, the conditional static independence condition (C-SIC), states the exclusion restriction which defines the instrument. The instrument Z_1 should have no influence on the joint distribution of one potential binary variable indicating a pair

participation and the two potential outcomes corresponding to this pair (they are indexed by the programs in the pair), given that a participation in one of those programs considered in the pair is actually observed. It is exactly the same condition as the one defined by Angrist and Imbens. The second condition, the conditional one-period monotonicity condition (C-OMC), affects the pair participation indicators and not the participation indicators. Nevertheless, the same interpretation can be made in terms of pair participation. Given that the choice is between participating in program k and participating in program l , the instrument has a monotonic influence on participation.

ASSUMPTION 4: (CONDITIONAL-ONE PERIOD MONOTONICITY CONDITION):

$$P\left[\tilde{S}_{kl}(z_1) \geq \tilde{S}_{kl}(z'_1) \mid S_1 = \{k, l\}\right] = 1 \text{ or } P\left[\tilde{S}_{kl}(z_1) \leq \tilde{S}_{kl}(z'_1) \mid S_1 = \{k, l\}\right] = 1; \quad z_1, z'_1 \in \{0, 1\}.$$

Under these assumptions all the effects comparing any pair of programs are identified for the compliers of the subpopulation participating in one of the programs considered in the pair. The compliers population is defined with respect to the pair participation and not with respect to all possible participations. These results are presented in theorem 2.

Theorem 2: UNDER ASSUMPTION 3 (C- SIC) AND 4 (C – OMC), THE EFFECTS

$$\theta_{\tau}^{k,l}(\tilde{S}_{kl}(1) - \tilde{S}_{kl}(0) = 1, S_1 = \{k, l\}), \quad \forall k, l \quad \text{IS IDENTIFIED.}$$

The proof is given in appendix A.1. Here, identification is obtained as easily as in the case where we have the choice to participate in only one program. Indeed, the conditioning set restores a “binary approach” to the problem and the same mechanisms apply. The results are similar to those found by Imbens (2000) and Lechner (2001) for the identification of treatment effects under the conditional independence assumption when more than one treatment is available. Identification is satisfied for subpopulations that have a restricted choice of programs: they can only participate in one of two programs.

5) The multi-period, one-instrument case

Now, we consider the first type of dependence between the participation indicators and the instrument. The behavior of the individuals is investigated for both periods, but we only

⁸ The four sequences of programs defined in the notation section can be considered as four different programs, for example the sequence $(1,1)$ can be defined as program 1.

observe the first instrument. Or equivalently, the instruments are perfectly correlated, $Z_1 = Z_2$. As an example of such a case, consider that we want to investigate the return to schooling on the wage. Thus, our participation variable indicates if an additional year is spend at school ($S_1 = 1$ if the individual goes to school in period 1 and is zero otherwise and $S_2 = 1$ if the individual goes to school in period 2 and equals zero if not). As instrument for education, the education of the parents is often used (Griliches, 1979, Ashenfelter and Rouse, 1998 and Maluccio, 1998) arguing that family background variables can influence the schooling process of the child (for example, the education of the father can give some information on the permanent income of the family or can reflect the preferences of the parents). Hence, the same instrument (we can consider that the education of the parents is constant over the periods of study) influences participation in both periods.

Moreover, we only focus on the effects of the treatment on the outcome of period 2 (for the effect on the outcome of period 1, the same principle as in the previous sections applies). Within this context, how can we define the subpopulation of compliers? Does a population of compliers exist? More than one such subpopulation can be defined. For example, some individuals could comply only in one period and thus form a subpopulation of compliers for this period. Nevertheless, we will use the term “compliers” without reference to a time period but for the subpopulation of persons who comply in both periods. Thus, more than one effect can be defined depending on the subpopulation considered. Nonetheless, identification will be possible only for some of these subpopulations. With one instrument, two different classes of effects are identified. The assumptions needed comprise an independence condition (or an exclusion restriction) and a monotonicity condition. Although the first period variables can have an influence on the second period variables (e.g. the second participation depends on the first one), the structure of the problem is very similar to the one-period, multiple-program case of the previous section. The conditions used are also in a conditional form. Instead of conditioning on a restricted participation choice (only between two programs), we condition on the behavior in one period. We can choose to fix the behavior in the first period or in the second period depending on the effects we are interested in. For example, we can look at the population that participates in the first period.

ASSUMPTION 5: (CONDITIONAL-STATIC INDEPENDENCE CONDITION):

$$(S_t(1), Y_2^{s_2^k}, Y_2^{s_2^l}), (S_t(0), Y_2^{s_2^k}, Y_2^{s_2^l}), (S_t(1), S_t(0))) \perp\!\!\!\perp Z_1 \mid S_t = s_t, \forall k, l, s_t, t \neq t'.$$

The condition requires that for all pair of sequences (k, l) , the independence restriction holds. This strong condition is needed to identify the effects of all possible comparisons between the sequences. Nevertheless, for certain comparisons, this assumption should hold only for the pair of sequences compared.

ASSUMPTION 6: (CONDITIONAL-ONE PERIOD MONOTONICITY CONDITION):

$$P[S_t(1) \geq S_t(0) | S_{t'} = s_{t'}] = 1 \text{ or } P[S_t(1) \leq S_t(0) | S_{t'} = s_{t'}] = 1; \quad t \neq t'.$$

Part A of theorem 3 states that the first identifiable effects are those for the subpopulations whose participation in the first period is the same whatever the value of the instrument. The proof is given in appendix A.2. In this case, the first period ($t=1$) plays the role of the start period ($t=0$) for the subpopulations defined by their participation status in the first period. The identifiable effects compare sequences of treatments which have an identical participation in the first period. Thus, in the subpopulations defined by the participation or by the nonparticipation in the first period, the local average treatment effect (participation versus no participation in the second period) is identified for the second period compliers.

*Theorem 3: A) UNDER ASSUMPTIONS 5 (C-SIC) AND 6 (C-OMC) WITH $t=2$ AND $t'=1$,
THE EFFECTS $\theta_2^{11,10}(S_1=1, S_2(1) - S_2(0)=1)$ AND $\theta_2^{01,00}(S_1=0, S_2(1) - S_2(0)=1)$
ARE IDENTIFIED.*

*B) UNDER ASSUMPTIONS 5 (C-SIC) AND 6 (C-OMC) WITH $t=1$ AND $t'=2$,
THE EFFECTS $\theta_2^{11,01}(S_2=1, S_1(1) - S_1(0)=1)$ AND $\theta_2^{10,00}(S_2=0, S_1(1) - S_1(0)=1)$
ARE IDENTIFIED.*

We get a similar result when we consider the subpopulation who complies in the first period and always or never participates in the second one. This time, the conditions affect the first period potential participation indicators given the participation in the second period. We compare sequences that have the same participation in the second period. The effect induced by the different sequences comes from the comparison of participation and nonparticipation in the first period. However, the assumptions used are not very intuitive because we condition on a variable from the future. Nevertheless, we can identify the same effects under a set of assumptions which are a little different but more intuitive.

ASSUMPTION 7: (STATIC INDEPENDENCE CONDITION-BIS):

$$(S_1(1), S_2(1), Y_2^{s_2^k}, Y_2^{s_2^l}), (S_1(0), S_2(0), Y_2^{s_2^k}, Y_2^{s_2^l}) \perp\!\!\!\perp Z_1, \text{ and } (S_2(0), S_2(1)) \perp\!\!\!\perp Z_1 \mid S_1 = s_1,$$

$$\text{or } (S_1(0), S_1(1)) \perp\!\!\!\perp Z_1 \mid S_2 = s_2$$

The price to pay for the suppression of the conditioning set is a strengthening of the independence condition. Now it deals with the joint distribution of the potential outcome and the participation indicators in both periods. The results of theorem 3 are restated in theorem 4 and proved in appendix A.3.

Theorem 4: A) UNDER ASSUMPTIONS 7 (SIC_BIS), AND 6 (OMC) WITH $t = 2$ AND

ASSUMING THAT $P[S_1(1) = S_1(0)] = 1$, THE EFFECTS $\theta_2^{11,10}(S_1 = 1, S_2(1) - S_2(0) = 1)$

AND $\theta_2^{01,00}(S_1 = 0, S_2(1) - S_2(0) = 1)$ ARE IDENTIFIED.

B) UNDER ASSUMPTIONS 7 (SIC_BIS), AND 6 (OMC) WITH $t = 1$ AND ASSUMING

THAT $P[S_2(1) = S_2(0)] = 1$, THE EFFECTS $\theta_2^{11,01}(S_2 = 1, S_1(1) - S_1(0) = 1)$ AND

$\theta_2^{10,00}(S_2 = 0, S_1(1) - S_1(0) = 1)$ ARE IDENTIFIED.

The most important change is without doubt the third assumption which excludes the defiers and the compliers of the first or the second period. In fact, it imposes the independence between the participation indicator in period 1 and the instrument for part A and the independence between the participation indicator in period 2 and the instrument for part B. Nevertheless, the instrument is still valid because it stays correlated with one of the participation indicators, the one for the period in which we compare a participation versus a nonparticipation. Of course this restriction is undesirable. Moreover, the number of sequences we can compare in this context stays very small. With only one instrument available, some effects can be identified if we step back into a “one period context”. The possible comparisons of sequences are restricted due to the fact that only sequences with the same participation in one period can be considered to obtain identification.

6) The multi-period, two-instruments case

Nevertheless, in this two-period framework more than one instrument is available. This additional information will allow to identify more effects. As already mentioned in the notation section, we consider different dependencies between the participation indicator in the

second period and both instruments. In this part we assume that the participation in the second period is only influenced by the second instrument. (The second instrument does not include the information provided by the first instrument.) and we still concentrate on the effects at the end of period 2. We can as well use the previous example on returns to schooling in a modified form to illustrate this kind of relation between the participation variables and the instruments. Let us keep the same choice for the participation but change the instrument we can observe. If we observed it, the distance to the nearest school is used as an instrument (Kane and Rouse, 1993, Card, 1995 and Maluccio, 1997). It is often justified by arguing that the distance to the nearest school is a proxy for the cost of education. Thus, let Z_1 equal 1 if the individual lives near a college in period 1 and equal 0 if not, and let Z_2 equal 1 if the individual lives near a college in period 2 and equal 0 otherwise. It is obvious that each period only the distance to the college in this period influences the participation.

6.1) *Static case*

Although we have more than one period, we are still in a nondynamic framework, because we consider that the first period outcome has no effect on the participation and on the outcome in the second period. As in the previous sections, we need an independence and a monotonicity assumption to obtain identification. The first part of the assumption looks similar to the previous static independence conditions. The joint distribution of the potential outcomes and the potential participation indicators in periods 1 and 2 is independent of both instruments. Now, the difference is that both instruments have to fulfill this independence condition.

ASSUMPTION 8 (STATIC INDEPENDENCE CONDITION-TER):

$(S_1(z_1), S_2(z_2), Y_2^{s_1^k}, Y_2^{s_2^l}) \perp\!\!\!\perp (Z_1, Z_2)$, for all possible combinations of z_1, z_2 and for all pairs k, l .

$(S_1(1), S_1(0), S_2(0), S_2(1)) \perp\!\!\!\perp Z_1, (S_2(0), S_2(1)) \perp\!\!\!\perp Z_2, (S_2(0), S_2(1)) \perp\!\!\!\perp Z_2 \mid S_1 = s_1, Z_1 = z_1$.

The second part of the assumptions (the last three conditions) is necessary to identify the probability to be a complier in both periods. The first condition states that the joint distribution of all the potential participation indicators are independent of the first period instrument. The second condition is similar to the condition needed in the one participation case but it involves variables in the second period. Then, the last condition looks like the condition imposed in the previous section (the one instrument case) for the variables in the

second period with the conditioning set augmented by the instrument of the first period. Until now the monotonicity condition has affected only one period. Therefore, we need also to exclude the defiers in the second period. The new monotonicity condition is an extension of the one-period monotonicity condition to a two-period monotonicity condition.

ASSUMPTION 9 (TWO-PERIOD MONOTONICITY CONDITION):

$$P[S_1(1) \geq S_1(0)] = 1 \text{ or } P[S_1(1) \leq S_1(0)] = 1, \quad P[S_2(1) \geq S_2(0)] = 1 \text{ or } P[S_2(1) \leq S_2(0)] = 1$$

and $P[S_1(1) \geq S_1(0), S_2(1) \geq S_2(0)] = 1$ or $P[S_1(1) \leq S_1(0), S_2(1) \geq S_2(0)] = 1$, or

$$P[S_1(1) \geq S_1(0), S_2(1) \leq S_2(0)] = 1 \text{ or } P[S_1(1) \leq S_1(0), S_2(1) \leq S_2(0)] = 1.$$

In the following, we will arbitrarily consider the case where $S_1(1) \geq S_1(0)$ and $S_2(1) \geq S_2(0)$. Then, under those assumptions, the following fourteen effects can be identified:

Theorem 5: UNDER ASSUMPTIONS 8 (SIC-TER) AND 9 (TMC),

A) ALL LATE – EFFECTS OF TYPE $\theta_2^{ij,kl}((S_2(1) - S_2(0))(S_1(1) - S_1(0)) = 1)$, ARE IDENTIFIED.

B) THE FOLLOWING 8 EFFECTS ARE IDENTIFIED: $\theta_2^{11,01}(S_2(0)(S_1(1) - S_1(0)) = 1)$,

$\theta_2^{11,01}(S_2(1)(S_1(1) - S_1(0)) = 1)$, $\theta_2^{11,10}(S_1(0)(S_2(1) - S_2(0)) = 1)$, $\theta_2^{11,10}(S_1(1)(S_2(1) - S_2(0)) = 1)$,

$\theta_2^{10,00}((1 - S_2(0))(S_1(1) - S_1(0)) = 1)$, $\theta_2^{10,00}((1 - S_2(1))(S_1(1) - S_1(0)) = 1)$,

$\theta_2^{01,00}((1 - S_1(0))(S_2(1) - S_2(0)) = 1)$, $\theta_2^{01,00}((1 - S_1(1))(S_2(1) - S_2(0)) = 1)$.

The theorem is proved in the appendix A.4. All effects for the compliers are identified. For other subpopulations fewer effects are identified. We can only compare sequences which have the same participation in one period. The persons concerned comply in the period when the sequences differ. In the other period only the participation or the treatment under a special value of the instrument is specified. When the instrument equals 1 we cannot differentiate between the compliers and the always-takers. When it equals 0, due to the monotonicity assumption the defiers are excluded, the effect affects the never-takers. These effects are identical to those identified in section 5, with one exception. The populations for which the effects are investigated are not based on always-takers and never-takers in one period as in the previous case. In the period when they do not comply, the individuals are selected on a particular value of one potential participation indicator (the participation given one of the

values of the regressor). Evidently, those potential indicators are not observed. Thus, although the sequences compared are the same as those in the multi-period, one-instrument case, the populations concerned differ. In the one-instrument case the populations are observable in one period. In the two-instrument case they are not. From the proof, the effects are expressible as a function of probabilities and expectations, which can be estimated with the available data,

$$\text{e.g. } \theta^{11,01}((S_2(1) - S_2(0))(S_1(1) - S_1(0)) = 1) = (E[Y_2 S_2 | Z_1 = 1, Z_2 = 1] - E[Y_2 S_2 | Z_1 = 0, Z_2 = 1] - E[Y_2 S_2 | Z_1 = 1, Z_2 = 0] + E[Y_2 S_2 | Z_1 = 0, Z_2 = 0]) / (P((S_2(1) - S_2(0))(S_1(1) - S_1(0)) = 1)).$$

And the denominator equals $E(S_2 | Z_2 = 1, Z_1 = 1) - E(S_2 | Z_2 = 0, Z_1 = 1) - E(S_2 | Z_2 = 1, Z_1 = 1, S_1 = 0)(1 - E(S_1 | Z_1 = 1)) + E(S_2 | Z_2 = 0, Z_1 = 1, S_1 = 0)(1 - E(S_1 | Z_1 = 1)) - E(S_2 | Z_2 = 1, Z_1 = 0, S_1 = 1)E(S_1 | Z_1 = 0) + E(S_2 | Z_2 = 0, Z_1 = 0, S_1 = 1)E(S_1 | Z_1 = 0)$. This last expression must be different from 0. We can rewrite this effect in terms of covariances, but the link between the effect and the “classical” instrumental variable estimator is not as obvious as in the one-period case:

$$\left(\frac{\text{cov}(Y_2 S_2, Z_2 | Z_1 = 1)}{P(Z_2 = 1 | Z_1 = 1)(1 - P(Z_2 = 1 | Z_1 = 1))} - \frac{\text{cov}(Y_2 S_2, Z_2 | Z_1 = 0)}{P(Z_2 = 1 | Z_1 = 0)(1 - P(Z_2 = 1 | Z_1 = 0))} \right) / \left(\frac{\text{cov}(S_2, Z_2 | Z_1 = 1)}{P(Z_2 = 1 | Z_1 = 1)(1 - P(Z_2 = 1 | Z_1 = 1))} - \frac{\text{cov}(S_2, Z_2 | Z_1 = 1, S_1 = 0)P(S_1 = 0 | Z_1 = 1)}{P(Z_2 = 1 | Z_1 = 1, S_1 = 0)(1 - P(Z_2 = 1 | Z_1 = 1, S_1 = 0))} - \frac{\text{cov}(S_2, Z_2 | Z_1 = 0, S_1 = 1)P(S_1 = 1 | Z_1 = 0)}{P(Z_2 = 1 | Z_1 = 0, S_1 = 1)(1 - P(Z_2 = 1 | Z_1 = 0, S_1 = 1))} \right).$$

Is it possible to identify more effects or other effects with some additional conditions? Unfortunately, even if we limit the relation between the periods by introducing some sort of independence the number of identified effects does not increase. These additional assumptions will only simplify the identification of the probability to be in the population of interest (the one for which we want to compute the effect, e.g. the probability to be a complier in both periods). Nevertheless, some effects can be bounded, if the outcomes, the potential outcomes and their expectations are themselves bounded. Some of the bounds are informative and some of them are informative depending on the value of different probabilities (conditional probabilities involving the realizations of the potential indicators). The results are available upon request.

6.2) Sequentiality

It is more realistic to assume that participation is decided on at the beginning of each period, because it is often the case that a participation in a particular program is conditioned on a previous participation in another particular program (eg. before they participate in a training program, the trainees should have experienced a job search assistance). Therefore, we now consider a sequential process. The participation decision is made on the basis of the observed past participations. This sequentiality emerges in the independence assumption.

ASSUMPTION 10: (SEQUENTIAL STATIC INDEPENDENCE CONDITION):

$(S_1(z_1), S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \perp\!\!\!\perp Z_1, (S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \perp\!\!\!\perp Z_2 \mid Z_1 = z_1, S_1 = s_1$, for all possible combinations of z_1, z_2 and for all pairs k, l , $S_1 \perp\!\!\!\perp Z_2 \mid Z_1 = z_1$,
 $(S_1(0), S_1(1), S_2(0), S_2(1)) \perp\!\!\!\perp Z_1, (S_2(0), S_2(1)) \perp\!\!\!\perp Z_2 \mid Z_1 = z_1, S_1 = s_1$.

The independence condition concerning the first-period instrument remains unchanged. The second-period instrument has to fulfil three conditions: one already appears in the previous section and concerns the identification of the probability to be in the population of interest, one is a modified form of the independence between the joint distribution of the potential outcomes and the potential participation in period 2, and the last one is a new hypothesis. The independence condition dealing with the potential outcomes only says that for the second period the potential variables are independent of the instrument given the past instrument and the past participation. Except from the conditioning, this assumption is the same as the one in the one period case. The new condition imposes the independence between the participation in the first period and the instrument in the second period given the instrument in the first period. This condition is a testable restriction. Indeed, the sequentiality in the condition introduces a cross period relation. To identify the effects, we need to restrict this cross section relation.

Lemma 1: ASSUMPTION 8 (SIC-TER) AND ASSUMPTION 10 (SSIC) ARE EQUIVALENT.

Surprisingly, assumptions 8 and 10 are equivalent. This result is stated in lemma 1 and proved in appendix A.5. The equivalence depends principally on the fact that in assumption 8 a sequential condition already appears and that the testable condition of assumption 10 has to be valid. In fact, in assumption 8, we treat the participation indicators as symmetric variables and impose for each of them an independence condition with both instruments. In assumption 10,

the symmetry is broken and the independence between the participation in the first period and the instrument in the second period is retrieved by this additional testable assumption.⁹ Theorem 6 presents the effects identified under this conditional approach. The proof is direct due to lemma 1.

Theorem 6: UNDER ASSUMPTIONS 10 (SSIC) AND 9 (TMC), THE SAME EFFECTS AS THOSE STATED IN THEOREM 5 ARE IDENTIFIED.

We encounter no loss of identification, but the cost is to impose the testable restriction on the relation between the instrument in the second period and the participation in the first period. If the participation in the first period has no influence on the value of the instrument in the second period for individuals with the same realization of the instrument in the first period, then considering the sequentiality in the participation process or not plays no role for the computation of the effects.

6.3) Dynamics

Suppose now, the interactions between the periods are complicated. It seems credible that in addition to the influence of the realized participation in the first period, the realized outcome also influences the decision to participate in the second period. For example, if a person is still unemployed after a certain period of time after the end of the first training program, it is more likely that she participates in another program. Again, this new hypothesis affects only the independence assumption. We are not in a static framework anymore, because the endogenous outcome of the first period (or some functions of it) influences the second period variables. Hence, we have to modify the conditions of the sequential independence assumption concerning the independence of the instrument in the second period and the other variables. The conditioning set formed by the instrument and the participation indicator in period 1 is completed by the outcome of period 1. Indeed, knowing that the participation decision is influenced by the outcome of the first period, the part of the potential participation indicators in the second period not explained by the participation in the first period and by the outcome in the first period has to be independent of the instrument in the second period. A similar reasoning is applied to the relation between the instrument in the second period and the potential outcomes. As in the sequential case, we have to impose an additional restriction

⁹ If we want to prove theorem 6 without using the equivalence result of lemma 1, it becomes obvious that without this testable assumption no identification is achieved.

on the cross period relations, in the sense that the outcome of the first period and the instrument have to be independent given the instrument and the participation in the first period. It is a similar assumption to the one imposed between the participation in the first period and the instrument in the second period in assumption 10.

ASSUMPTION 11: (DYNAMIC INDEPENDENCE CONDITION):

$(S_1(z_1), S_2(z_2), Y_2^{\underline{s}_2^k}, Y_2^{\underline{s}_2^l}) \perp\!\!\!\perp Z_1, (S_2(z_2), Y_2^{\underline{s}_2^k}, Y_2^{\underline{s}_2^l}) \perp\!\!\!\perp Z_2 \mid Z_1 = z_1, S_1 = s_1, Y_1 = y_1$, for all possible combinations of z_1, z_2 and for all pairs k, l , $S_1 \perp\!\!\!\perp Z_2 \mid Z_1 = z_1, Y_1 \perp\!\!\!\perp Z_2 \mid Z_1 = z_1, S_1 = s_1$, $(S_1(0), S_1(1), S_2(0), S_2(1)) \perp\!\!\!\perp Z_1, (S_2(0), S_2(1)) \perp\!\!\!\perp Z_2 \mid Z_1 = z_1, S_1 = s_1, Y_1 = y_1$.

Except for the presence of the outcome of the first period, the conditions look very similar to the conditions met in the sequential case. Indeed, if assumption 11 is fulfilled then assumption 10 will be valid. We state this result in lemma 2 and prove it in appendix A.6.

Lemma 2: ASSUMPTION 11 (DIC) IMPLIES ASSUMPTION 10 (SIC-TER).

Here, the key condition for the implication (and also for the identification) is the condition relating the outcome in period 1 and the instrument in period 2. For example, when investigating the effect of participating in a training program on the probability to be unemployed, the region of residence or the labor office which the individuals are assigned to could be good candidates for the instruments. They can influence the participation in a training program, but have no influence (if we consider two neighboring regions with similar labor markets) on the probability to be unemployed, the outcome. If we suppose that this choice is valid to fulfil the assumption 10, we need to check further the validity of the instruments before considering that the effects we identify correspond to the dynamic effects we are looking for. If the individuals still unemployed at the end of period 1, who participated in the program in period 1, falsely think that they have more chances to find a job in the neighbor region and move to it, then the key condition is not fulfilled. Although the assumption to identify the effects in a case of sequential participation decision is valid, no dynamic effect can be identified.

As in the sequential case the monotonicity assumption 9 (two period monotonicity condition) is still required unchanged. Then, theorem 7 states the identification results. There is no loss of identification due to the dynamic nature of the problem.

Theorem 7: UNDER ASSUMPTIONS 11 (DIC) AND 9 (TMC), ALL THE EFFECTS OF THEOREM 5 ARE IDENTIFIED.

The proof is simple due to lemma 2 and lemma 1. The introduction of the dynamic aspect of the problem does not lead to a loss of identification. As shown in the proof of theorem 5, the effects are functions of some observable conditional expectations (an example is presented in section 6.1, the estimable form of the effects being the same in the dynamic case and in the other case). The conditioning events are all combinations of the instruments: $\{(Z_1 = 1, Z_2 = 1), (Z_1 = 1, Z_2 = 0), (Z_1 = 0, Z_2 = 1), (Z_1 = 0, Z_2 = 0)\}$. When the realizations of the instruments are correlated but not perfectly correlated, $0 < P(Z_1 \neq Z_2) < 1$, our results still hold. If the instruments are perfectly correlated, we can easily extend the results of section 5. Of course, we only investigate the case for the participation in the second period given that the participation in the first period is known (the other way makes no sense in terms of dynamics). Therefore, no more than two effects can be identified in this context. Introduce in the conditioning set of assumption 5 the outcome of period 1. Following the same steps as those we would use if we proved theorem 7 directly (without referring to the result of the lemmas), we obtain

$$E_{Y_1|S_1, Z_1=1} \left\{ E \left[S_2(1)Y_2^{11} + (1-S_2(1))Y_2^{10} \mid S_1 = 1 \right] \right\} - E_{Y_1|S_1, Z_1=0} \left\{ E \left[S_2(0)Y_2^{11} + (1-S_2(0))Y_2^{10} \mid S_1 = 1 \right] \right\}.$$

Thus, to be able to continue the proof and achieve identification, we need the additional assumption that $Y_1 \perp\!\!\!\perp Z_1 \mid S_1 = s_1$. This assumption is more intuitive than the one added in the two instrument case. If we had imposed the usual independence condition also in the first period, this additional assumption would be trivially fulfilled. It is part of the definition of the quality of the variable Z as instrument.

6.4) Robustness

Until now, we made an implicit exclusion restriction when we assumed the potential participation indicators in the second period to depend only on one instrument, namely the second one. But is this really necessary? If there is no such restriction, four potential participation indicators exist in the second period: $S_2(1,1), S_2(0,1), S_2(1,0), S_2(0,0)$, where the

first number is the value of the instrument in the first period (or the first instrument) and the second one the value of the instrument in the second period. To illustrate this dependence, combine the previous return to schooling examples which clarify the first two kinds of dependence of the instruments and the participation indicators. Suppose that the participation variables stay the same and that in the first period the distance to the school is not observable. Therefore, in period 1 only the education of the parents can be used as instrument. Then, the quality of the data has been improved in the second period: for the second period, we also observe the distance to the nearest college. For the second period two instruments are available and influence the participation in the second period, the education of the parents and the distance to the nearest school.

Now, it is extremely difficult to divide the population in the four categories of compliers, defiers, never-takers and always-takers in the second period. If we treat the instrument as an assignment to a program, its value in the first period corresponds to the assignment to a program for this period and its value in the second period corresponds to the assignment for this period. Thus, we can possibly regard the second-period compliers as those who participate if the instrument is 1 in the second period and who do not participate if the instrument is 0 in this period (independently of the value of the instrument in the first period). Another possibility is to look at a sequence of participations and a sequence of assignments. A complier would be a person who follows her sequence of assignments in all the periods. For a complier, the following pairs of events have to be true: $(S_1(1)=1, S_2(1,1)=1), (S_1(0)=0, S_2(0,1)=1), (S_1(1)=1, S_2(1,0)=0)$ and $(S_1(0)=0, S_2(0,0)=0)$. With this new dependence, we need to restate the monotonicity condition.¹⁰

ASSUMPTION 12: (ONE PERIOD MONOTONICITY CONDITION-II):

$$P[S_2(z_1,1) \geq S_2(z_1,0)] = 1 \text{ or } P[S_2(z_1,1) \leq S_2(z_1,0)] = 1, \forall z_1.$$

Under this condition and one of the independence conditions previously made, only few effects are identified, namely the effects that compare sequences with the same treatment in the first period. Nonetheless, it is not true for all subpopulations. Only the following effects are identified:¹¹ $\theta_2^{11,10}(S_1(1)(S_2(1,1) - S_2(1,0)) = 1)$, $\theta_2^{11,10}(S_1(0)(S_2(0,1) - S_2(0,0)) = 1)$, $\theta_2^{01,00}((1 - S_1(1))(S_2(1,1) - S_2(1,0)) = 1)$, $\theta_2^{01,00}((1 - S_1(0))(S_2(0,1) - S_2(0,0)) = 1)$. It follows that

¹⁰ Even if we have a monotonicity property for the first period variables, we do not obtain identification of all the effects. We do not need such a condition for the possible identifications.

¹¹ The proof is presented in appendix A.7.

our previous results are not robust to this change in the relation between the participation indicator in the second period and the instruments. Thus, the implicit exclusion restriction imposed on this relation is a necessary condition to obtain identification.

Before coming to the conclusions we present in table 1 a summary of the hypotheses and results obtained. The last set of results presents only one combination of the monotonicity condition and an independence assumption, but other combinations are possible.

Table 1

7) Conclusion

In this article we examine the identifiability of different effects of sequences of programs using instrumental variables. More than one LATE can be defined. We consider different sets of assumptions and their implications on the identification of the parameters of interest. In a one-period framework with more than one program, identification is obtained for subpopulations defined by a pair-wise participation, i.e. they can only choose between two programs. Unfortunately, in a multi-period framework we cannot identify all possible effects even in a case with some independence condition between the periods. Nevertheless, the effects for individuals who comply each period are identified whatever the assumption set is. These sets of assumptions principally comprise a monotonicity hypothesis and some exclusion restrictions. When only one instrument is available in both periods, only few effects can be identified. These compare sequences which have the same participation in one of the two periods. In this period, the affected populations for which we want to compute the effects are observed to have the same participation and these populations are compliers in the other period. The effects for such individuals are not identifiable anymore when more than one instrument is available. Nonetheless, we can identify a lot of other different effects. The introduction of endogeneity in the dynamic case does not lead to a loss of identification, but the conditions needed can be more difficult to fulfil. This paper covers a part of the problem of identification of dynamic treatment effects using instrumental variables. The LIV approach of Heckman and Vytlacil (1999) is not at all discussed and needs also to be translated into a dynamic context. Moreover, the IV approach allows the identification of more than some means. In a static framework the distribution (quantiles) of some treatment effects are also identified. This point is not yet investigated in the dynamic framework.

BIBLIOGRAPHY:

- Abadie A. (2000): "Semiparametric Instrumental Variable Estimation of Treatment Response Models", National Bureau of Economic Research Technical Working Paper n° 260.
- Abadie A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models", Journal of American Statistical Association, Vol. 97, 284-292.
- Abadie A., J. D. Angrist and G. W. Imbens (2002): "Instrumental Variables estimates of the Effect of Subsidized training on the Quantiles of Trainee Earnings", Econometrica, Vol.70, n°1, 91-117.
- Angrist J. D., G. W. Imbens and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables", Journal of American Statistical Association, Vol. 91, 444-458.
- Angrist J. D. and A. B. Krueger (1999): "Empirical Strategies in Labor Economics," in O. Ashenfelter and D. Card (eds.), Handbook of Labor Economics, Vol. III A, chapter 23, 1277-1366.
- Angrist J. D. and G. W. Imbens (1991): "Sources of Identifying Information in Evaluation Models", NBER Technical Working Paper, n° 117.
- Arellano M. and B. Honoré (2002): "Panel Data Models: Some Recent Developments", in J.J. Heckman and E. Leamer (eds.): Handbook of Econometrics, Vol. V, ch.53.
- Ashenfelter O. and C. E. Rouse (1998): "Income, Schooling and Ability: Evidence from a New Sample of Identical Twins", Quarterly Journal of Economics, 113, 253-284.
- Card D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling", in L. N. Christofides, E. K. Grant and R. Swidinsky (eds): Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp, University of Toronto Press, Toronto; Canada, 201-222.
- Dawid A.P. (1979): "Conditional Independence in Statistical Theory", Journal of the Royal Statistical Society Series B, 41, 1-31.
- Dawid A.P. (1980): "Conditional Independence for Statistical Operations", the Annals of Statistics 8, 598-617.
- Gourieroux C. and A. Monfort (1989): Statistique et Modèles Econometriques, Economica.
- Griliches Z. (1979): "Sibling Models and Data in Economics: Beginnings of a Survey", Journal of Political Economy, 87, 37-65.
- Heckman J. J. (1997): "Instrumental Variables, A Study of Implicit Behavioral Assumptions", Journal of Human Resources, 32(3), 441-462.

- Heckman J. J., R. J. LaLonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs", in O. Ashenfelter and D. Card (eds.): *Handbook of Labor Economics*, Vol. III A, 1865-2097, North-Holland, Amsterdam.
- Heckman J. J., J. L. Tobias and E. Vytlačil (2000): "Simple Estimators for Treatment Parameters in a Latent Variables Framework with an Application to Estimating the Returns to Schooling", National Bureau of Economic Research Working Paper 7950.
- Heckman J. J. and E. Vytlačil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects", *Proceeding of the National Academy of Sciences* 96, 4730-4734.
- Heckman J. J. and E. Vytlačil (2000a): "The Relationship between Treatment Parameters within a Latent Variable Framework", *Economics Letters*, 66, 33-39.
- Heckman J. J. and E. Vytlačil (2000b): "Local Instrumental Variables", in Hsiao, c., k. Morimune, and J. Powell (eds), *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, Cambridge: Cambridge Univrsity Press.
- Imbens G. W. and J. D. Angrist (March 1994): " Identification and Estimation of Local Average Treatment Effects", *Econometrica*, vol. 62, n°2, 467 – 475.
- Imbens G. W. (2000): "The Role of Propensity Score in Estimating Dose-Response Functions", *Biometrika*, vol. 87, n°3, 706-710.
- Kane T.J. and C. E. Rouse (1993): "Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and do Degrees Matter?", NBER Working Paper n°4268.
- Lechner M. (2001): "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", in: Lechner, M., Pfeiffer, F. (eds), *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physica/Springer, 43-58.
- Lechner M. and F. Pfeiffer, (eds., 2001), *Econometric Evaluation of Active Labour Market Policies*, Heidelberg: Physica.
- Lechner M. and R. Miquel (2001): "A Potential Outcome Approach to Dynamic Programme Evaluation: Nonparametric Identification", Working Paper n° 2001-07, University of St. Gallen.
- Maluccio J. (1997): "Endogeneity of Schooling in the Wage Function", Unpublished Manuscript, Department of Economics, Yale University.
- Maluccio J. (1998): "Endogeneity of Schooling in the Wage Function: Evidence from the Rural Philippines", FCND Discussion Paper n°54.

- Manski C. F. and J. V. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling", *Econometrica*, Vol. 68, n°4, 997-1010.
- Resnick S. I. (1999): *A Probability Path*, Birkhäuser Boston.
- Robins J. M. (1986): "A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect." *Mathematical Modelling*, 7:1393-1512, with 1987 Errata to "A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect." *Computers and Mathematics with Applications*, 14:917-921; 1987 Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect." *Computers and Mathematics with Applications*, 14:923-945; and 1987 Errata to "Addendum to 'A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect'." *Computers and Mathematics with Applications*, 18:477.
- Robins J. M. (1989): "The Analysis of Randomized and Nonrandomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies", Sechrest, L., H. Freeman, A. Mulley (eds.), *Health Service Research Methodology: A Focus on Aids*, 113-159, Washington, D.C.: Public Health Service, National Center for Health Services Research.
- Robins J. M. (1997): "Causal Inference from Complex Longitudinal Data. Latent Variable Modeling and Applications to Causality", in M. Berkane, ed., *Lecture Notes in Statistics* (120), NY: Springer Verlag, pp. 69-117.
- Robins J. M., S. Greenland and F. Hu (1999): "Estimation of the Causal Effect of a Time-varying Exposure on the Marginal Mean of a Repeated Binary Outcome", *Journal of the American Statistical Association*, 94, 687-700, with discussion.
- Roy A.D. (1951): "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, 3, 135 – 146.
- Rubin D.B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688 – 701.
- van den Berg G. J. (2001): "Duration Models: Specification, Identification, and Multiple Durations", in Heckman, James J., and Edward Leamer (eds.), *Handbook of Econometrics*, vol. V, Amsterdam: North-Holland.

Vytlacil E. (2002): “Independence, Monotonicity, and Latent Index Model: an Equivalence Result”, *Econometrica*, Vol 70, n°1, 331-341.

Appendix:

A.1) Proof of theorem 2:

We need to prove that $\theta_{\tau}^{k,l}(S_{kl}(1) - S_{kl}(0) = 1, S_1 = \{k, l\})$ is a function of observables.

$$\begin{aligned}
& E(Y_{\tau} | Z_1 = 1, S_1 = \{k, l\}) - E(Y_{\tau} | Z_1 = 0, S_1 = \{k, l\}) \\
&= E\left[\tilde{S}_{kl}(1)Y_{\tau}^k + (1 - \tilde{S}_{kl}(1))Y_{\tau}^l | Z_1 = 1, S_1 = \{k, l\}\right] - E\left[\tilde{S}_{kl}(0)Y_{\tau}^k + (1 - \tilde{S}_{kl}(0))Y_{\tau}^l | Z_1 = 0, S_1 = \{k, l\}\right] \\
&\stackrel{Ass.3}{=} E\left[\tilde{S}_{kl}(1)Y_{\tau}^k + (1 - S_{kl}(1))Y_{\tau}^l | S_1 = \{k, l\}\right] - E\left[\tilde{S}_{kl}(0)Y_{\tau}^k + (1 - \tilde{S}_{kl}(0))Y_{\tau}^l | S_1 = \{k, l\}\right] \\
&= E\left[(\tilde{S}_{kl}(1) - \tilde{S}_{kl}(0))(Y_{\tau}^k - Y_{\tau}^l) | S_1 = \{k, l\}\right] \\
&= P\left[\tilde{S}_{kl}(1) - \tilde{S}_{kl}(0) = 1 | S_1 = \{k, l\}\right] E\left[Y_{\tau}^k - Y_{\tau}^l | \tilde{S}_{kl}(1) - \tilde{S}_{kl}(0) = 1, S_1 = \{k, l\}\right] \\
&\quad - P\left[\tilde{S}_{kl}(1) - \tilde{S}_{kl}(0) = -1, S_1 = \{k, l\}\right] E\left[Y_{\tau}^k - Y_{\tau}^l | \tilde{S}_{kl}(1) - \tilde{S}_{kl}(0) = -1, S_1 = \{k, l\}\right] \\
&\stackrel{Ass.4}{=} P\left[\tilde{S}_{kl}(1) - \tilde{S}_{kl}(0) = 1 | S_1 = \{k, l\}\right] E\left[Y_{\tau}^k - Y_{\tau}^l | \tilde{S}_{kl}(1) - \tilde{S}_{kl}(0) = 1, S_1 = \{k, l\}\right] \\
&\stackrel{Ass.4}{=} \left\{E(\tilde{S}_{kl}(1) | Z_1 = 1, S_1 = \{k, l\}) - E(\tilde{S}_{kl}(0) | Z_1 = 0, S_1 = \{k, l\})\right\} E\left[Y_{\tau}^k - Y_{\tau}^l | \tilde{S}_{kl}(1) - \tilde{S}_{kl}(0) = 1, S_1 = \{k, l\}\right] \\
&\rightarrow E\left[Y_{\tau}^k - Y_{\tau}^l | \tilde{S}_{kl}(1) - \tilde{S}_{kl}(0) = 1, S_1 = \{k, l\}\right] = \frac{E(Y_{\tau} | Z_1 = 1, S_1 = \{k, l\}) - E(Y_{\tau} | Z_1 = 0, S_1 = \{k, l\})}{E(\tilde{S}_{kl} | Z_1 = 1, S_1 = \{k, l\}) - E(\tilde{S}_{kl} | Z_1 = 0, S_1 = \{k, l\})} \\
&E\left[Y_{\tau}^k - Y_{\tau}^l | \tilde{S}_{kl}(1) - \tilde{S}_{kl}(0) = 1, S_1 = \{k, l\}\right] = \frac{E(Y_{\tau} | Z_1 = 1, S_1 = \{k, l\}) - E(Y_{\tau} | Z_1 = 0, S_1 = \{k, l\})}{P(S_1 = k | Z_1 = 1, S_1 = \{k, l\}) - P(S_1 = k | Z_1 = 0, S_1 = \{k, l\})} \\
&= \frac{\frac{E(Y_{\tau} | Z_1 = 1, S_1 = \{k, l\}) - E(Y_{\tau} | Z_1 = 0, S_1 = \{k, l\})}{P(S_1 = k | Z_1 = 1)}}{\frac{P(S_1 = k | Z_1 = 1) + P(S_1 = l | Z_1 = 1)}{P(S_1 = k | Z_1 = 0) + P(S_1 = l | Z_1 = 0)}}.
\end{aligned}$$

The effect equals a function of estimable expectations and probabilities and thus is identified.

A.2) Proof of theorem 3:

We only prove part A of the theorem for the effect $\theta_2^{11,10}(S_1=1, S_2(1)-S_2(0)=1)$. The proof for the second effect, $\theta_2^{01,00}(S_1=0, S_2(1)-S_2(0)=1)$, follows exactly the same steps with $E(Y_2 | S_1=0, Z_1=z_1) = E(S_2(z_1)Y_2^{01} + (1-S_2(z_1))Y_2^{00} | S_1=0, Z_1=z_1)$ as a starting point.

Using the observation rule, $E(Y_2 | S_1=1, Z_1=1) - E(Y_2 | S_1=1, Z_1=0)$ is equal to $E[S_2(1)Y_2^{11} + (1-S_2(1))Y_2^{10} | S_1=1, Z_1=1] - E[S_2(0)Y_2^{11} + (1-S_2(0))Y_2^{10} | S_1=1, Z_1=0]$. Under assumption 5, this difference is independent of the instrument and equals $E[S_2(1)Y_2^{11} + (1-S_2(1))Y_2^{10} | S_1=1] - E[S_2(0)Y_2^{11} + (1-S_2(0))Y_2^{10} | S_1=1]$. Regrouping the terms and using the implication of assumption 6, we obtain $E(Y_2 | S_1=1, Z_1=1) - E(Y_2 | S_1=1, Z_1=0) = P[S_2(1) - S_2(0) = 1 | S_1=1] E[Y_2^{11} - Y_2^{10} | S_1=1, S_2(1) - S_2(0) = 1]$.

Therefore, to prove that the effect is identified, we have to prove that the probability to be a complier in the second period given participation in the first period is identified.

$$\begin{aligned}
 P[S_2(1) - S_2(0) = 1 | S_1 = 1] &= 1 - P[S_2(1) = S_2(0) = 0 | S_1 = 1] - P[S_2(1) = S_2(0) = 1 | S_1 = 1] \\
 &\stackrel{\text{Ass.5}}{=} 1 - P[S_2(1) = S_2(0) = 0 | S_1 = 1, Z_1 = 1] - P[S_2(1) = S_2(0) = 1 | S_1 = 1, Z_1 = 0] \\
 &= 1 - P[S_2 = 0 | S_1 = 1, Z_1 = 1] + P[S_2(1) = 0, S_2(0) = 1 | S_1 = 1, Z_1 = 1] \\
 &\quad - P[S_2 = 1 | S_1 = 1, Z_1 = 0] + P[S_2(1) = 0, S_2(0) = 1 | S_1 = 1, Z_1 = 0] \\
 &\stackrel{\text{Ass.6}}{=} 1 - P[S_2 = 0 | S_1 = 1, Z_1 = 1] - P[S_2 = 1 | S_1 = 1, Z_1 = 0] \\
 &= P[S_2 = 1 | S_1 = 1, Z_1 = 1] - P[S_2 = 1 | S_1 = 1, Z_1 = 0].
 \end{aligned}$$

$$\text{Thus, } \theta_2^{11,10}(S_1=1, S_2(1)-S_2(0)=1) = \frac{E(Y_2 | S_1=1, Z_1=1) - E(Y_2 | S_1=1, Z_1=0)}{E(S_2 | S_1=1, Z_1=1) - E(S_2 | S_1=1, Z_1=0)}.$$

Similarly, following the same steps, part B of the theorem is proved. We obtain

$$\theta_2^{11,01}(S_2=1, S_1(1)-S_1(0)=1) = \frac{E(Y_2 | S_2=1, Z_1=1) - E(Y_2 | S_2=1, Z_1=0)}{E(S_1 | S_2=1, Z_1=1) - E(S_1 | S_2=1, Z_1=0)}.$$

A.3) Proof of theorem 4:

A) We only prove that $\theta_2^{11,10}(S_1=1, S_2(1)-S_2(0)=1)$ is identified. The proof for the other effects follows exactly the same steps. $E(S_1Y_2 | Z_1=1) - E(S_1Y_2 | Z_1=0)$
 $= E[S_1(1)S_2(1)Y_2^{11} + S_1(1)(1-S_2(1))Y_2^{10} | Z_1=1] - E[S_1(0)S_2(0)Y_2^{11} + S_1(0)(1-S_2(0))Y_2^{10} | Z_1=0]$
 $\stackrel{Ass.7}{=} E[S_1S_2(1)Y_2^{11} + S_1(1-S_2(1))Y_2^{10}] - E[S_1S_2(0)Y_2^{11} + S_1(1-S_2(0))Y_2^{10}]$ (the potential form of the participation indicator can be deleted because $P(S_1(1)=S_1(0))=1$)
 $= E[S_1(S_2(1)-S_2(0))(Y_2^{11}-Y_2^{10})] \stackrel{Ass.6}{=} E[Y_2^{11}-Y_2^{10} | S_1(S_2(1)-S_2(0))=1] P[S_1(S_2(1)-S_2(0))=1]$
 $= E[Y_2^{11}-Y_2^{10} | S_1=1, S_2(1)-S_2(0)=1] P[S_2(1)-S_1(0)=1 | S_1=1] P(S_1=1).$

It remains to prove that the probability to be a complier in the second period given a participation in the first period can be estimated. However, the proof has already been presented in the proof of theorem 3 and will not be restated.

A.4) Proof of theorem 5:

To simplify the presentation of the proof, we present intermediate results in lemma A1.

Lemma A1: under Assumption 8, the following equalities hold

- (1) $E[Y_2S_2 | Z_1=1, Z_2=1] - E[Y_2S_2 | Z_1=0, Z_2=1] = E[S_2(1)(S_1(1)-S_1(0))(Y_2^{11}-Y_2^{01})]$
- (2) $E[Y_2S_2 | Z_1=1, Z_2=0] - E[Y_2S_2 | Z_1=0, Z_2=0] = E[S_2(0)(S_1(1)-S_1(0))(Y_2^{11}-Y_2^{01})]$
- (3) $E[Y_2S_2 | Z_1=1, Z_2=1] - E[Y_2S_2 | Z_1=1, Z_2=0] = E[(Y_2^{11}-Y_2^{01})S_1(1)(S_2(1)-S_2(0))] + E[Y_2^{01}(S_2(1)-S_2(0))]$
- (4) $E[Y_2S_2 | Z_1=1, Z_2=1] - E[Y_2S_2 | Z_1=0, Z_2=0] = E[(Y_2^{11}-Y_2^{01})(S_1(1)S_2(1)-S_1(0)S_2(0))] + E[Y_2^{01}(S_2(1)-S_2(0))]$
- (5) $E[Y_2S_2 | Z_1=1, Z_2=0] - E[Y_2S_2 | Z_1=0, Z_2=1] = E[(Y_2^{11}-Y_2^{01})(S_1(1)S_2(0)-S_1(0)S_2(1))] - E[Y_2^{01}(S_2(1)-S_2(0))]$

Similar results are available for the conditional expectations of Y_2S_1 , $Y_2(1-S_1)$, $Y_2(1-S_2)$ and Y_2 . They are not presented here but are available upon request.

PROOF: AS DEFINED IN ABADIE (2002, PROOF OF LEMMA 2.1), ASSUMPTION 8 IMPLIES THAT $Y^{jk}S_i(m)$, $j, k, i, m = 0, 1$, ARE INDEPENDENT OF Z_i . THE RESULTS FOLLOW BY DIRECT CALCULATIONS.

We will prove the identification of only one effect of theorem 5. The proofs for the identification of the other five effects follow exactly the same steps using different combinations of the results of lemma A1 as starting point. To identify the effect

$\theta_2^{11,00}((S_2(1) - S_2(0))(S_1(1) - S_1(0)) = 1)$, subtract the sum of (2) and

$E[Y_2(1 - S_1) | Z_1 = 1, Z_2 = 1] - E[Y_2(1 - S_1) | Z_1 = 1, Z_2 = 0]$ from the sum of (1) and

$E[Y_2(1 - S_1) | Z_1 = 0, Z_2 = 1] - E[Y_2(1 - S_1) | Z_1 = 0, Z_2 = 0]$. After rearranging the terms we

obtain: $E[(S_2(1) - S_2(0))(S_1(1) - S_1(0))(Y_2^{11} - Y_2^{00})] = E[Y_2(S_2 + S_1 - 1) | Z_1 = 1, Z_2 = 1]$

$- E[Y_2(S_2 + S_1 - 1) | Z_1 = 0, Z_2 = 1] - E[Y_2(S_2 + S_1 - 1) | Z_1 = 1, Z_2 = 0] + E[Y_2(S_2 + S_1 - 1) | Z_1 = 0, Z_2 = 0]$.

Then, using assumption 9, $E[Y_2^{11} - Y_2^{00} | (S_2(1) - S_2(0))(S_1(1) - S_1(0)) = 1]$

$= \frac{E[Y_2(S_2 + S_1 - 1) | Z_1 = 1, Z_2 = 1] - E[Y_2(S_2 + S_1 - 1) | Z_1 = 0, Z_2 = 1] - E[Y_2(S_2 + S_1 - 1) | Z_1 = 1, Z_2 = 0]}{P[(S_2(1) - S_2(0))(S_1(1) - S_1(0)) = 1]}$

$+ \frac{E[Y_2(S_2 + S_1 - 1) | Z_1 = 0, Z_2 = 0]}{P[(S_2(1) - S_2(0))(S_1(1) - S_1(0)) = 1]}$. Thus, it remains to be proved that

$P[(S_2(1) - S_2(0))(S_1(1) - S_1(0)) = 1]$ is observable. Using assumption 9, $P[S_2(1) - S_2(0) = 1] =$

$= P[(S_2(1) - S_2(0))(S_1(1) - S_1(0)) = 1] + P[(S_2(1) - S_2(0)) = 1 | S_1(1) = S_1(0) = 1]P[S_1(1) = S_1(0) = 1]$

$+ P[(S_2(1) - S_2(0)) = 1 | S_1(1) = S_1(0) = 0]P[S_1(1) = S_1(0) = 0]$. Therefore, the probability to be

a complier equals $P[S_2(1) - S_2(0) = 1] - P[(S_2(1) - S_2(0)) = 1 | S_1(1) = S_1(0) = 1]P[S_1(1) = S_1(0) = 1]$

$- P[(S_2(1) - S_2(0)) = 1 | S_1(1) = S_1(0) = 0]P[S_1(1) = S_1(0) = 0]$ and we only need to proof that

each term is estimable.

a) $P[S_2(1) - S_2(0) = 1] \stackrel{\text{Ass.9}}{=} 1 - P[S_2(1) = S_2(0) = 0] - P[S_2(1) = S_2(0) = 1]$

$$\stackrel{\text{Ass.8}}{=} 1 - P[S_2(1) = S_2(0) = 0 | Z_2 = 1] - P[S_2(1) = S_2(0) = 1 | Z_2 = 0]$$

$$\stackrel{\text{Ass.9}}{=} 1 - P[S_2 = 0 | Z_2 = 1] - P[S_2 = 1 | Z_2 = 0] = P[S_2 = 1 | Z_2 = 1] - P[S_2 = 1 | Z_2 = 0]$$

$$\Rightarrow P[S_2(1) - S_2(0) = 1] = E[S_2 | Z_2 = 1] - E[S_2 | Z_2 = 0].$$

$$\text{b) } P[S_2(1) - S_2(0) = 1 | S_1(1) = S_1(0) = 1] \stackrel{\text{Ass.8}}{=} P[S_2(1) - S_2(0) = 1 | S_1(1) = S_1(0) = 1, Z_1 = 0]$$

$$\stackrel{\text{Ass.9}}{=} 1 - P[S_2(1) = S_2(0) = 0 | S_1 = 1, Z_1 = 0] - P[S_2(1) = S_2(0) = 1 | S_1 = 1, Z_1 = 0]$$

$$\begin{aligned} \Leftrightarrow P[S_2(1) - S_2(0) = 1 | S_1 = 1, Z_1 = 0] &\stackrel{\text{Ass.8}}{=} 1 - P[S_2(1) = S_2(0) = 0 | Z_2 = 1, S_1 = 1, Z_1 = 0] \\ &\quad - P[S_2(1) = S_2(0) = 1 | Z_2 = 0, S_1 = 1, Z_1 = 0] \end{aligned}$$

$$\stackrel{\text{Ass.9}}{=} 1 - P[S_2 = 0 | Z_2 = 1, S_1 = 1, Z_1 = 0] - P[S_2 = 1 | Z_2 = 0, S_1 = 1, Z_1 = 0]$$

$$= P[S_2 = 1 | Z_2 = 1, S_1 = 1, Z_1 = 0] - P[S_2 = 1 | Z_2 = 0, S_1 = 1, Z_1 = 0]$$

$$= E[S_2 | Z_2 = 1, S_1 = 1, Z_1 = 0] - E[S_2 | Z_2 = 0, S_1 = 1, Z_1 = 0].$$

Therefore,

$$P[S_2(1) - S_2(0) = 1 | S_1(1) = 1, S_1(0) = 1] = E[S_2 | Z_2 = 1, S_1 = 1, Z_1 = 0] - E[S_2 | Z_2 = 0, S_1 = 1, Z_1 = 0].$$

$$\text{c) Following the proof in b), we get: } P[S_2(1) - S_2(0) = 1 | S_1(1) = 0, S_1(0) = 0]$$

$$= E[S_2 | Z_2 = 1, S_1 = 0, Z_1 = 1] - E[S_2 | Z_2 = 0, S_1 = 0, Z_1 = 1].$$

$$\text{d) } P[S_1(1) = S_1(0) = 1] \stackrel{\text{Ass.8}}{=} P[S_1(1) = S_1(0) = 1 | Z_1 = 0] \stackrel{\text{Ass.9}}{=} P[S_1 = 1 | Z_1 = 0] = E[S_1 | Z_1 = 0].$$

$$\text{e) } P[S_1(1) = S_1(0) = 0] \stackrel{\text{Ass.8}}{=} P[S_1(1) = S_1(0) = 0 | Z_1 = 1] \stackrel{\text{Ass.9}}{=} P[S_1 = 0 | Z_1 = 1] = 1 - E[S_1 | Z_1 = 1].$$

All the expectations are observable in the sample, therefore our effect is identified.

A.5) Proof of lemma 1

Before starting the proof, we present two results that will be extensively used.

R1: Let A , B and C be three random variables. Then, $(A, B) \perp\!\!\!\perp C \Leftrightarrow A \perp\!\!\!\perp C$ and $B \perp\!\!\!\perp C | A$ (see for example Dawid, 1980 or Gouriéroux and Monfort, 1989)).

R2: Let A , B and C be three random variables. Then, $A \perp\!\!\!\perp B | C$ and $B \perp\!\!\!\perp C \Rightarrow A \perp\!\!\!\perp B$.
 $(f(A|B) = \int f(A|B, C) f(C|B) dC \underset{\text{assumptions}}{=} \int f(A|C) f(C) dC = f(A))$, where $f(\cdot)$ denotes a distribution function.)

Let all the conditions in assumption 8 be denoted by Roman numerals and all the conditions in assumption 10 by capital letters, I: $(S_1(z_1), S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \perp\!\!\!\perp (Z_1, Z_2)$, II: $(S_2(0), S_2(1)) \perp\!\!\!\perp Z_2$, III: $(S_2(0), S_2(1)) \perp\!\!\!\perp Z_2 | S_1 = s_1, Z_1 = z_1$, IV: $(S_1(1), S_1(0), S_2(0), S_2(1)) \perp\!\!\!\perp Z_1$ and A: $(S_1(z_1), S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \perp\!\!\!\perp Z_1$, B: $(S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \perp\!\!\!\perp Z_2 | Z_1 = z_1, S_1 = s_1$, C: $S_1 \perp\!\!\!\perp Z_2 | Z_1 = z_1$, D: $(S_1(0), S_1(1), S_2(0), S_2(1)) \perp\!\!\!\perp Z_1$, E: $(S_2(0), S_2(1)) \perp\!\!\!\perp Z_2 | Z_1 = z_1, S_1 = s_1$.

First, we prove that assumption 10 implies assumption 8.

Conditions III and IV are trivially valid under assumption 10 because they are the same conditions as E and D. To prove that II is also valid, we use some intermediate results. Condition D implies that $(S_2(0), S_2(1)) \perp\!\!\!\perp Z_1$ (D') and conditions E and C imply that $(S_2(0), S_2(1)) \perp\!\!\!\perp Z_2 | Z_1 = z_1$ (E+C)', due to R2. Using a similar argument with the results of D' and (E+C)', we have that $(S_2(0), S_2(1)) \perp\!\!\!\perp Z_2$ is valid. Finally, condition I is valid under assumption 10 due to the following results: condition C can be rewritten as $S_1(z_1) \perp\!\!\!\perp Z_2 | Z_1 = z_1$ (C'), condition B can be rewritten as $(S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \perp\!\!\!\perp Z_2 | Z_1 = z_1, S_1(z_1) = s_1$ (B'). Combining C' and B' (R1), we obtain $(S_1(z_1), S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \perp\!\!\!\perp Z_2 | Z_1 = z_1$ (C'+B')'. In the end, applying R1 with A and (C'+B')' condition I is implied.

Second, we prove that assumption 8 implies assumption 10.

Furthermore, conditions D and E are trivially fulfilled when assumption 8 is valid. Condition I implies that $S_1(z_1) \perp\!\!\!\perp (Z_1, Z_2)$. Then, from R1, this implies that $S_1(z_1) \perp\!\!\!\perp Z_1$ (I-1)' and $S_1(z_1) \perp\!\!\!\perp Z_2 | Z_1 = z_1$ (I-2)' are valid. But (I-2)' is equivalent to $S_1 \perp\!\!\!\perp Z_2 | Z_1 = z_1$, the condition C of assumption 10. Finally, it follows directly from I that condition A is fulfilled. Then, with R1, condition I implies $(S_1(z_1), S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \perp\!\!\!\perp Z_2 | Z_1 = z_1$ (B'). But (B)' is equivalent to

$(S_1, S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \amalg Z_2 \mid Z_1 = z_1$ and as C is valid with R1, we obtain $(S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \amalg Z_2 \mid Z_1 = z_1, S_1 = s_1$, the condition B.

A.6) Proof of lemma 2:

Let also, number the conditions of assumption 11 with Arabic numerals, 1:

$$\begin{aligned} (S_1(z_1), S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \amalg Z_1, \quad 2: \quad (S_2(z_2), Y_2^{s_2^k}, Y_2^{s_2^l}) \amalg Z_2 \mid Z_1 = z_1, S_1 = s_1, Y_1 = y_1, \quad 3: \\ S_1 \amalg Z_2 \mid Z_1 = z_1, \quad 4: \quad Y_1 \amalg Z_2 \mid Z_1 = z_1, S_1 = s_1, \quad 5: \quad (S_1(0), S_1(1), S_2(0), S_2(1)) \amalg Z_1, \quad 6: \\ (S_2(0), S_2(1)) \amalg Z_2 \mid Z_1 = z_1, S_1 = s_1, Y_1 = y_1. \end{aligned}$$

Conditions A, C and D of assumption 10 are trivially fulfilled because they are the same conditions as conditions 1, 3 and 5. Then, the result R2 applied to conditions 2 and 4 implies condition B and applied to conditions 6 and 4 implies condition E.

A.7) Proof of the results of section 6.4:

We only prove the result for the first effect under the static independence condition (assumption 8). Using the other independence conditions leads to the same structure for the proof, except the additional use of $E_{S_1|Z_1}(\cdot)$ and $E_{Y_1|Z_1, Z_2, S_1}(\cdot)$. The identification of the other effects

is proved in the same way replacing S_1 by $1 - S_1$ or replacing $Z_1 = 1$ by $Z_1 = 0$ or both.

$$\begin{aligned} E(Y_2 S_1 \mid Z_1 = 1, Z_2 = 1) - E(Y_2 S_1 \mid Z_1 = 1, Z_2 = 0) &= E(S_1(1) S_2(1, 1) Y_2^{11} + S_1(1)(1 - S_2(1, 1)) Y_2^{10} \mid Z_1 = 1, Z_2 = 1) \\ &\quad - E(S_1(1) S_2(1, 0) Y_2^{11} + S_1(1)(1 - S_2(1, 0)) Y_2^{10} \mid Z_1 = 1, Z_2 = 0) \\ &\stackrel{\text{Ass.8}}{=} E(S_1(1) S_2(1, 1) Y_2^{11} + S_1(1)(1 - S_2(1, 1)) Y_2^{10} - S_1(1) S_2(1, 0) Y_2^{11} - S_1(1)(1 - S_2(1, 0)) Y_2^{10}) \\ &= E(S_1(1)(S_2(1, 1) - S_2(1, 0))(Y_2^{11} - Y_2^{10})) \\ &\stackrel{\text{Ass.12}}{=} E(Y_2^{11} - Y_2^{10} \mid S_1(1)(S_2(1, 1) - S_2(1, 0)) = 1) P(S_1(1)(S_2(1, 1) - S_2(1, 0)) = 1) \end{aligned}$$

$$\Rightarrow E(Y_2^{11} - Y_2^{10} \mid S_1(1)(S_2(1, 1) - S_2(1, 0)) = 1) = \frac{E(Y_2 S_1 \mid Z_1 = 1, Z_2 = 1) - E(Y_2 S_1 \mid Z_1 = 1, Z_2 = 0)}{P(S_1(1)(S_2(1, 1) - S_2(1, 0)) = 1)}.$$

The identification of the denominator has already been proved in the other theorems. We just have to replace the monotonicity condition used by assumption 12.

Table 1: the multi-period two instrument case, a summary

Assumptions			Effects identified
$P[S_1(1) \geq S_1(0)] = 1$ or $P[S_1(1) \leq S_1(0)] = 1$, $P[S_2(1) \geq S_2(0)] = 1$ or $P[S_2(1) \leq S_2(0)] = 1$, $P[S_1(1) \geq S_1(0), S_2(1) \geq S_2(0)] = 1$ or $P[S_1(1) \leq S_1(0), S_2(1) \geq S_2(0)] = 1$, or $P[S_1(1) \geq S_1(0), S_2(1) \leq S_2(0)] = 1$ or $P[S_1(1) \leq S_1(0), S_2(1) \leq S_2(0)] = 1$			$\theta_2^{ij,kl} ((S_2(1) - S_2(0))(S_1(1) - S_1(0)) = 1)$, $\theta_2^{11,01} (S_2(0)(S_1(1) - S_1(0)) = 1)$,
$(S_1(z_1), S_2(z_2), Y_2^{\underline{s}_2^k}, Y_2^{\underline{s}_2^l}) \amalg (Z_1, Z_2)$, $(S_2(0), S_2(1)) \amalg Z_2$, $(S_1(1), S_1(0), S_2(0), S_2(1)) \amalg Z_1$, $(S_2(0), S_2(1)) \amalg Z_2 \mid S_1 = s_1, Z_1 = z_1$	$(S_1(z_1), S_2(z_2), Y_2^{\underline{s}_2^k}, Y_2^{\underline{s}_2^l}) \amalg Z_1$, $(S_2(z_2), Y_2^{\underline{s}_2^k}, Y_2^{\underline{s}_2^l}) \amalg Z_2 \mid Z_1 = z_1, S_1 = s_1$, $S_1 \amalg Z_2 \mid Z_1 = z_1$, $(S_1(0), S_1(1), S_2(0), S_2(1)) \amalg Z_1$, $(S_2(0), S_2(1)) \amalg Z_2 \mid Z_1 = z_1, S_1 = s_1$	$(S_1(z_1), S_2(z_2), Y_2^{\underline{s}_2^k}, Y_2^{\underline{s}_2^l}) \amalg Z_1$, $(S_2(z_2), Y_2^{\underline{s}_2^k}, Y_2^{\underline{s}_2^l}) \amalg Z_2 \mid Z_1 = z_1, S_1 = s_1, Y_1 = y_1$ $S_1 \amalg Z_2 \mid Z_1 = z_1$, $Y_1 \amalg Z_2 \mid Z_1 = z_1, S_1 = s_1$, $(S_1(0), S_1(1), S_2(0), S_2(1)) \amalg Z_1$ $(S_2(0), S_2(1)) \amalg Z_2 \mid Z_1 = z_1, S_1 = s_1, Y_1 = y_1$	$\theta_2^{11,01} (S_2(1)(S_1(1) - S_1(0)) = 1)$, $\theta_2^{11,10} (S_1(0)(S_2(1) - S_2(0)) = 1)$, $\theta_2^{11,10} (S_1(1)(S_2(1) - S_2(0)) = 1)$, $\theta_2^{10,00} ((1 - S_2(0))(S_1(1) - S_1(0)) = 1)$, $\theta_2^{10,00} ((1 - S_2(1))(S_1(1) - S_1(0)) = 1)$, $\theta_2^{01,00} ((1 - S_1(0))(S_2(1) - S_2(0)) = 1)$, $\theta_2^{01,00} ((1 - S_1(0))(S_2(1) - S_2(0)) = 1)$.
$P[S_2(z_1, 1) \geq S_2(z_1, 0)] = 1$ or $P[S_2(z_1, 1) \leq S_2(z_1, 0)] = 1$ + one of the three independence assumption			$\theta_2^{11,10} (S_1(1)(S_2(1, 1) - S_2(1, 0)) = 1)$, $\theta_2^{11,10} (S_1(0)(S_2(0, 1) - S_2(0, 0)) = 1)$, $\theta_2^{01,00} ((1 - S_1(1))(S_2(1, 1) - S_2(1, 0)) = 1)$, $\theta_2^{01,00} ((1 - S_1(0))(S_2(0, 1) - S_2(0, 0)) = 1)$.