

Sequential Causal Models for the Evaluation of Labor Market Programs

Michael Lechner

SIAW

Swiss Institute for International Economics
and Applied Economic Research



University of St.Gallen

First version: October 2002

Revised: August 2007

Date this version has been printed: 07 August 2007

ABSTRACT

This paper reviews inverse selection probability weighting to estimate dynamic causal effects. A distribution theory based on sequential GMM estimation is proposed and the method is applied to a re-evaluation of some parts of the Swiss active labor market policy to obtain new results and discuss several issues about the implementation of the estimation procedure.

Keywords: Dynamic treatment effects, nonparametric identification, causal effects, sequential randomization, program evaluation, panel data

JEL classification: C21, C31

Addresses for correspondence

Michael Lechner

Professor of Econometrics

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

University of St. Gallen, Bodanstr. 8, CH-9000 St. Gallen, Switzerland

Michael.Lechner@unisg.ch, www.siaw.unisg.ch/lechner

1 INTRODUCTION

There is a substantial literature about the estimation of average ‘causal effects’ of policy interventions using large micro data. Examples are numerous studies concerned with employment effects for unemployed who participate in publicly financed training programs. In this microeconomic framework, a causal effect is defined as a comparison of what would have happened in two (or more) different potential states of the world. This binary comparison is based on those observations that are actually observed in the two different states. If the allocation to a specific state is random, then the difference of the means of the outcome variables of the respective participants in the different states is a valid estimator for the average causal effect. If participation is not random, then selection problems arise and weighted means are required. In its static version, the selection problem received considerable attention in the microeconomics and statistics literature, resulting in numerous suggestions about how to derive the required weights. Angrist and Krueger (1999), Heckman, LaLonde, and Smith (1999), and Imbens (2004) provide surveys of various parts of this vast literature.

However, the static model may not be able to address all relevant selection issues that may occur. Consider the evaluation of training programs: Suppose that interest is not in the effect of one particular training course, but in a sequence of courses. However, if the first course in such a sequence is effective, some unemployed individuals take up a job before the second course starts. If interest is not in the effect of the first course, but in the effect of the sequence of two courses, such behavior creates a selection problem that cannot be addressed in the static model. For example, solely controlling for pre-training variables does not work for obvious reasons. One important selection variable, namely the outcome of the first participation, is missing. However, controlling for variables realized after the first training course that influence selection into the second course bears the potential problem that they may be

influenced by the first part of the training sequence. Thus, they are ‘endogenous’ and ruled out as control variables in standard econometric approaches. Finally, it would be hard to find an instrument that is valid in that particular case.

Robins (1986) suggests an explicitly dynamic causal framework that allows defining causal effects of dynamic interventions and systematically addressing this type of selection problem. His approach is based on the key assumption that it is possible to control for the variables jointly influencing outcomes and selection at each particular selection step of the sequential selection process. Robins' (1986) approach was applied subsequently in epidemiology and biostatistics. Gill and Robins (2001) extend this approach to continuous treatments. Estimation strategies proposed for this model by Robins and various co-authors usually make parametric or semiparametric modeling assumptions about the relation of the potential outcomes to the confounding variables, like in g-estimation proposed by Robins (1986). G-estimation is based on recursively modeling the dependence of the outcome variables on the histories of confounders and treatments. To do so more parsimoniously, so-called structural nested models (SNM) have been proposed by Robins (1993, 1994). However, estimation (typically by some form of maximum likelihood) of such models, which are necessarily complex, is very tedious. An alternative to this class of models are the so-called marginal structural models (MSM) again proposed by Robins (1998a, b). MSM are more tractable and probably more robust because they model the dependence of the potential outcomes only on those confounders that are time constant. Since controlling for those variables does not remove selection bias, weighting by the inverse of the treatment probabilities is used in addition. Robins (1999) discusses the advantages and disadvantages of MSM and SNM. Various types of MSM are discussed in Robins, Greenland, and Hu (1999), Robins, Hernan, and Brumback (1999), Hernan, Brumback, and Robins (2001), as well as Joffe, Ten Have, Feldman, and Kimmel (2004). Miller, Ten Have, Reboussin, Lohmann, and Rejeski (2001) relate MSN to

the literature on missing data problems. Estimation of MSM typically appears to be done by the (weighted) method of generalized estimating equations (GEE, Liang and Zeger, 1986), which has conceptual similarities to the generalized methods of moments (Hansen, 1982) that is more common in econometrics. So far, there seem to be no applications of dynamic potential outcome models to economic problems.

Lechner and Miquel (2001, LM01 further on) adapt Robins' (1986) framework to comparisons of more general sequences, different parameters and selection processes, and reformulate the identifying assumptions. The assumptions used by Robins and coauthors and LM01 bear considerable similarity to the selection-on-observables or conditional independence assumption (CIA) that is prominent in the static evaluation literature. When the CIA holds, inverse selection probability weighting estimators (IPW) as well as matching estimators are typically used (see Imbens, 2004). Both classes of estimators have the advantage that the relation of the outcome variables to the confounders does not need to be specified. They require only the specification of the relation of the confounders to the selection process. In typical labor market evaluation studies, this feature is considered an important advantage, because applied researchers may have better knowledge about the selection process than about the outcome process. Although inverse probability weighting estimators are in principle easy to implement and efficient under suitable conditions (e.g. Hirano, Imbens, and Ridder, 2003), they are known to have problematic finite sample properties (e.g. Frölich, 2004) when very small or very large probabilities occur which could lead to too much importance (weight) for a single (or a few) observations. However, although matching estimators may be more robust (but less efficient) in this sense, for many important versions of these estimators the asymptotic distribution has not yet been established (see Abadie and Imbens, 2006).

This paper proposes estimators for the models of Robins (1986) in the formulation suggested by LM01 that retain the flexible and convenient properties of the aforementioned estimators,

although some increase in complexity is unavoidable. Although sequential propensity score matching estimators could be applied in this context as well (and are discussed in an earlier discussion paper version of this paper that is available on the internet), this paper focuses on IPW estimation of the dynamic causal models. As mentioned above, it is an advantage that both methods do not require modeling the dependence of the outcomes and the confounders. The required specification of the relation between selection and confounders is done sequentially, i.e. each selection step over time is considered explicitly, which allows a detailed and flexible specification of this relation. Although these advantages are shared by both, matching and IPW estimators, it is easy to derive its asymptotic distribution only for IPW, though. For the case of a parametric modeling of the relation of the selection process to confounders, in each period, which is the common case in static evaluation studies, the paper provides the asymptotic distribution of the IPW estimators based on the theory of sequential GMM estimation as proposed, for example, by Newey (1984). As a further addition to the literature in epidemiology, this paper considers not only the estimation of the effects for some general population (the average treatment effect as it would be called in the static evaluation literature), but also some effect heterogeneity by treatment status (average treatment effects on the treated). Moreover, it discusses issues of common support that arise this context with IPW estimation in some detail. An empirical analysis of different components of the Swiss active labor market policy illustrates its application and addresses a couple of practical issues.

The paper proceeds as follows: Section 2 outlines the dynamic causal framework using the notation suggested by LM01 and details the IPW estimator. In Section 3, the practical application of the estimator is illustrated using Swiss administrative data. Section 4 concludes. Appendix A addresses issues that arise with multiple treatments as used in the empirical part. Appendix B shows the derivation of the sequential inverse probability weighted estimator and its asymptotic distribution.

2 THE DYNAMIC CAUSAL MODEL: DEFINITION OF THE MODEL AND IDENTIFICATION

2.1 *Notation and Causal Effects*

This section briefly repeats the definition of the dynamic selection model as well as the identification results as presented by LM01 using the terminology of the econometric evaluation literature. Since a three-periods-two-treatments model is sufficient to discuss most relevant issues that distinguish the dynamic from the static model and comes with a much lighter notational burden than the general model, this section shows the results only for this basic version.

Suppose there is an initial period '0', in which everybody is in the same treatment state (e.g. everybody in the population of interest is unemployed), followed by two periods in which different treatment states could be realized (for example, unemployed may participate in some training courses). Periods are indexed by t or τ ($t, \tau \in \{0, 1, 2\}$). For this population, the complete treatment is described by a vector of random variables (RV), $S = (S_1, S_2)$. A particular realization of S_t is denoted by $s_t \in \{0, 1\}$. Denote the history of variables up to period t by a bar below a variable, e.g. $\underline{s}_2 = (s_1, s_2)$. To differentiate between different sequences, sometimes a letter (e.g. j) is used to index a sequence like \underline{s}^j . As a further notational convention, capital letters usually denote random variables, whereas small letters denote specific values of the random variable. In the first period, a unit is observed in one of two treatments (participating in training, $s_1 = 1$, or not participating in training, $s_1 = 0$). In period two, she participates in one of four treatment sequences $\{(0, 0), (1, 0), (0, 1), (1, 1)\}$, depending on what happened in period 1. Therefore, every unit can participate in one sequence defined by s_1 and another sequence defined by the values s_1 and s_2 . To sum up, in the two (plus one)-periods-two-treatments example, we consider six different overlapping potential outcomes corresponding to

two mutually exclusive states defined by treatment status in period 1, plus four mutually exclusive states defined by treatment status in periods 1 and 2.

Variables used to measure the effects of the treatment, i.e. the potential outcomes (like earnings), are indexed by treatments, and denoted by $Y_t^{s_t}$. Potential outcomes are measured at the end of (or just after) each period, whereas treatment status is measured at the beginning of each period. For each length of a sequence (1 or 2 periods), one of the potential outcomes is observable and denoted by Y_t . The resulting two observation rules (called consistency condition by Robins, 1986, or stable unit value treatment assumption, SUTVA, by Rubin, 1980) are defined in the following equation :

$$Y_t = S_1 Y_t^1 + (1 - S_1) Y_t^0 = S_1 S_2 Y_t^{11} + S_1 (1 - S_2) Y_t^{10} + (1 - S_1) S_2 Y_t^{01} + (1 - S_1) (1 - S_2) Y_t^{00}; \quad t = 0, 1, 2.$$

Finally, variables that may influence treatment selection and (or) potential outcomes are denoted by X . The K -dimensional vector X_t , which may contain functions of Y_t , is observable at the same time as Y_t .

Next, the average causal effect (for period t) of a sequence of treatments defined up to period 1 or 2 (τ) compared to an alternative sequence ($\theta_t^{s_t^k; s_t^l}$) is defined below:

$$\theta_t^{s_t^k; s_t^l}(s_{\tilde{\tau}}^j) := E(Y_t^{s_t^k} | \underline{S}_{\tilde{\tau}} = s_{\tilde{\tau}}^j) - E(Y_t^{s_t^l} | \underline{S}_{\tilde{\tau}} = s_{\tilde{\tau}}^j),$$

$$0 \leq \tilde{\tau} \leq 2, \quad 1 \leq \tau \leq 2, \quad \tilde{\tau} \leq \tau, \quad k, l \in (1, \dots, 2^\tau), \quad j \in (1, \dots, 2^{\tilde{\tau}}).$$

Analogous to the static treatment literature, LM01 call $\theta_t^{s_t^k; s_t^l}$ the dynamic average treatment effect (DATE). Accordingly, $\theta_t^{s_t^k; s_t^l}(s_{\tau}^k)$, as well as $\theta_t^{s_t^k; s_t^l}(s_{\tau}^l)$ are termed DATE on the treated (DATET) and DATE on the nontreated, respectively. There are cases in-between, like $\theta_t^{s_t^k; s_t^l}(s_1^l)$ for which the conditioning set is defined by a sequence shorter than the ones defin-

ing the causal contrast, denoted by DATE(T). Note that the effects are symmetric for the same population ($\theta_i^{s_r^k:s_r^l}(s_r^j) = -\theta_i^{s_r^l:s_r^k}(s_r^j)$), which, however, does not exclude effect heterogeneity ($\theta_i^{s_r^k:s_r^l}(s_r^k) \neq -\theta_i^{s_r^l:s_r^k}(s_r^l)$, $s_2^j \neq s_2^l$).

2.2 Identification

Assume that a large random sample, $\{s_{1i}, s_{2i}, x_{0i}, x_{1i}, x_{2i}, y_{1i}, y_{2i}\}_{i=1:N}$, is available. It is drawn from the large population of interest (with the same value of s_0) and characterized by the random variables $(S_1, S_2, X_0, X_1, X_2, Y_1, Y_2)$. Furthermore, all conditional expectations of all random variables that are of interest shall exist. Under these conditions, Robins (1986) and LM01 show that, if we can observe the variables that jointly influence selection and outcomes at each stage, then particular average treatment effects are identified. This assumption is called the weak conditional independence assumption (W-DCIA) by LM01.

Weak dynamic conditional independence assumption (W-DCIA)

- a) $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_1 \mid X_0 = x_0$;
- b) $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_2 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1$;
- c) $1 > P(S_1 = 1 \mid X_0 = x_0) > 0$, $1 > P(S_2 = 1 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1) > 0$; $\forall \underline{x}_1 \in \underline{\mathcal{X}}_1$, $\forall s_1 : s_1 \in \{0, 1\}$.

$A \perp\!\!\!\perp B \mid C = c$ means that *each element* of the vector of random variables B is independent of the random variable A conditional on the random variable C taking a value of c in the sense of Dawid (1979). Part a) of W-DCIA states that, conditional on X_0 , potential outcomes are independent of treatment choice in period 1 (S_1). This is the standard version of the static CIA. Part b) states that conditional on the treatment in period 1, on observable outcomes (which may be part of X_1) and on confounding variables of period 0 and 1 (\underline{X}_1), potential outcomes are independent of participation in period 2 (S_2).

To see whether such an assumption is plausible in a particular application, we have to ask which variables influence *changes* in treatment status as well as outcomes and whether they are observable. If the answer to that question is yes, and if there is common support (defined in part c) of W-DCIA), then we have identification, even if some or all conditioning variables in period 2 are influenced by the outcome of the treatment in period 1. To be more precise, the above mentioned literature shows that, for example, quantities like $E(Y_2^{11})$, $E(Y_2^{11} | S_1 = 0)$, $E(Y_2^{11} | S_1 = 1)$, or $E[Y_2^{11} | \underline{S}_2 = (1,0)]$ are identified, but that $E[Y_2^{11} | \underline{S}_2 = (0,0)]$ or $E[Y_2^{11} | \underline{S}_2 = (0,1)]$ are not identified. Thus, $\theta_2^{s_2^k; s_2^l}$, $\theta_2^{s_2^k; s_2^l}(s_1^j)$ are identified, $\forall s_1^k, s_2^k, s_1^l, s_2^l, s_1^j, s_2^j \in \{0,1\}$, but $\theta_2^{s_2^k; s_2^l}(\underline{s}_2^j)$ are not identified if $s_1^l \neq s_1^k$. This result states that pair-wise comparisons of all sequences are identified, but only for groups of individuals defined by their treatment status in period 0 or 1. The reason is that while in the first period treatment choice is random conditional on exogenous variables, which is the result of the initial condition $S_0 = 0$ for all observations, in the second period, randomization into these treatments is conditional on variables already influenced by the first period treatment. W-DCIA has an appeal for applied work as a natural extension of the static framework.

LM01 show that to identify all treatment parameters, W-DCIA must be strengthened by imposing that the confounding variables used to control selection into the treatment of the second period are not influenced by the selection into the first treatment. Thus, the dynamic selection problem collapses into a static selection problem with multiple treatments as discussed by Imbens (2000), and Lechner (2001a).

Appendix A contains some brief considerations for the dynamic case with multiple treatments. The extension to more than 2 periods is direct.

2.3 Estimation

All relevant estimators of the causal effects of the treatment sequences are based on differences of weighted means of the observable outcomes of those actually receiving the two treatments of interest:

$$\hat{\theta}_\tau^{s_2^k, s_2^l}(s_\tau^j) = \frac{1}{\sum_{i \in \underline{s}_2^k} \hat{w}_i^{s_2^k, s_\tau^j}} \sum_{i \in \underline{s}_2^k} \hat{w}_i^{s_2^k, s_\tau^j} y_{t,i} - \frac{1}{\sum_{i \in \underline{s}_2^l} \hat{w}_i^{s_2^l, s_\tau^j}} \sum_{i \in \underline{s}_2^l} \hat{w}_i^{s_2^l, s_\tau^j} y_{t,i}; \quad \hat{w}_i^{s_2^l, s_\tau^j}, \hat{w}_i^{s_2^k, s_\tau^j} \geq 0; \tau \in \{0,1\}, j \in \{k,l\}.$$

$\hat{w}_i^{s_2^k, s_\tau^j}$ and $\hat{w}_i^{s_2^l, s_\tau^j}$ are the appropriate weights (' $\hat{\cdot}$ ' indicates that they are estimated).

An obvious way to reweight the observations of a specific treatment population towards the target population (s_τ^j) is to use the inverse of their respective selection probabilities of being observed in that treatment as opposed to the target population, following the ideas of Horvitz and Thomson (1952). Recently, the idea of directly reweighting the observations based on the choice probabilities has found renewed interest (e.g. Hahn, 1998, Hernan, Brumback, and Robins, 2001, Hirano and Imbens, 2001, Hirano, Imbens, and Ridder 2003, Nevo, 2003, Robins, Hernan, and Brumback 2000, Robins and Rotnitzky, 1995, Robins, Rotnitzky, and Zhao, 1995, and Wooldridge, 2002, 2004). The reason is probably two-fold. Firstly, as advocated for example by Robins, Hernan, and Brumback (2000), they are easy to use, if a parametric model is used to estimate the choice probabilities. Secondly, Imbens, Hirano, and Ridder (2003) show for the static model, that for a suitably chosen nonparametric estimator of the choice probabilities, the inverse-probability estimator is efficient.

For $s_\tau^j = s_0$, Robins and coauthors showed in various papers that the following weights lead to consistent estimators (empirical equivalent to the g-formula):

$$\hat{w}_i^{s_2^k} = \frac{1}{\hat{p}_2^{s_2^k | s_1^k}(\underline{x}_{1,i}) \hat{p}_1^{s_1^k}(x_{0,i})} \quad \forall i \in \underline{s}_2 = \underline{s}_2^k; \forall k.$$

$\hat{p}_2^{s_2^k | s_1^k}(\underline{x}_{1,i})$ denotes a consistent estimate of the transition probability

$P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_{1,i})$ and $\hat{p}_1^{s_1^k}(x_{0,i})$ is a consistent estimate of $P(S_1 = s_1^k | X_0 = x_{0,i})$. A

similar result can be derived for the DATE(T), namely:

$$\hat{w}_i^{s_2^k, s_1^j} = \frac{\hat{p}_1^{s_1^j}(x_{0,i})}{\hat{p}_2^{s_2^k | s_1^k}(\underline{x}_{1,i}) \hat{p}_1^{s_1^k}(x_{0,i})} \quad \forall i \in \underline{S}_2 = \underline{s}_2^k; \forall k, j.$$

Appendix B shows that such weights solve the selection problem.

It is so far common practice in the static evaluation literature to estimate such selection probabilities with parametric binary or multiple response models (for each selection or transition).

Under the standard regularity conditions of GMM estimation that are usually considered to be weak (see Hansen, 1982, or for this special case Newey, 1984), such an estimator is \sqrt{N} -consistent with an asymptotically normal distribution. Appendix B shows its asymptotic covariance matrix.

3 AN EMPIRICAL APPLICATION

In this section, the IPW estimator is applied to an evaluation of active labor market programs in Switzerland to show that it can provide interesting new insights and to discuss issues of implementation that may arise.

3.1 *Setting the Scene*

The study by Gerfin and Lechner (2002, GL02 henceforth) serves as an example for this exercise. They are interested in the effects of different components of the Swiss active labor market policy on subsequent labor market outcomes using a rich administrative database coming from the merged records of the Swiss unemployment insurance system and the public pension system. They consider several types of programs and argue extensively that the data are infor-

mative enough to make the conditional independence assumption plausible. It contains very good information about variables explaining program participation as well as potential outcomes, such as long employment histories, regional, socio-demographic and occupational information, as well as subjective information on the unemployed provided by the caseworkers who are in charge of program assignment. For more details about the programs evaluated and the data used, the reader is referred to GL02. This application relies on the same population as GL02, but uses the extended version of the sample described in the follow-up study by Gerfin, Lechner, and Steiger (2005).

GL02 estimate the effects of starting the first program in an unemployment spell. They focus on the *beginning* of the *first* program, because in a static model there are two endogeneity problems. Firstly, any timing of the outcome measurement related to the end of the program may already incorporate some program effects. Secondly, subsequent program participation may depend on the effect of the first program. The dynamic causal framework allows going further and accounting for the endogeneity of program duration and subsequent participation.

Here, interest is in the effects of being two periods in different states for individuals entering unemployment in the last quarter of 1997. The different states used are unemployment (U), participating in training courses (C) or employment programs (E), and receiving temporary wage subsidies (T). A *period* is defined as an interval of two months. Given the usual duration of Swiss programs, a bimonthly period appears to be appropriate to capture the essential dynamics. Any bimonthly period in which an individual participates at least two weeks in a program (C, E or T) is considered as program participation. In the rare event of participation in two different programs in the same period, the individual is defined to participate in the longest of those programs. Treatment occurs between January and April 1998. For example, considering sequences like EE compared to CC allows assessing the effects of four months of employment programs compared to four months of courses. Obviously, this approach allows

investigating the effects of program combinations as well, for example by comparing CE to UU or EE. As another example, it is possible to check the effect of *waiting* for future participation by, for example, comparing UC (or UCC) to CC. For the sake of brevity, only effects of the types CC, EE, TT and UU are considered. Since W-DCIA is assumed to hold, we consider only effects for subpopulations defined by the treatment in the first period.

The outcome variables of interest are the probability of finding unsubsidized employment as well as monthly earnings between May 1998 and the end of the observation period in December 1999. If we believe in the validity of the identification arguments of GL02 for the beginning of the first program, then the part of W-DCIA relating to the first period is no problem. When the intermediate outcomes that determine the selection in March are observable, then the part of W-DCIA relating to the second selection step is plausible as well.

Table 1 gives descriptive statistics on some important control variables related to W-DCIA (see Table 2 for a full list of control variables used), treatments, intermediate as well as final outcomes for the different subsamples of interest. The control variables (X_0) considered in this table are the monthly earnings in the previous job, the subjective valuation of the caseworkers as well as the number of sanction days without benefit (imposed by the caseworkers on unemployed individuals violating the rules). The subjective valuation gives an assessment of the employability of the unemployed. This assessment is a time varying variable that the caseworker may change at the monthly interview of the unemployed (it is observed for every month). Comparing descriptive statistics across subsamples defined by treatment status reveals that participants in employment programs are having the worst a priori chances in the labor market. Participants in temporary wage subsidies appear to be those with the best a priori chances.

----- Table 1 about here -----

The lower panel in Table 1 shows statistics for the four intermediate outcomes (X_1), namely employment, earnings, the subjective valuation of the caseworkers, and sanction days (i.e. reduced benefits if the unemployed violates the rules). Comparing the variables across treatments confirms the previous conclusion. However, it also shows that participants leaving the sequence are different from participants who finish the sequence.

3.2 The IPW Estimator in Practice

3.2.1 Estimation of the Conditional Participation Probabilities

Conditional probabilities are estimated at each step in each subsample using binary probits, which are subjected to specification tests. Table 2 presents the six binary probits that are required to estimate all selection equations that are of interest for the effects of the treatments defined above. The specifications broadly follow the specifications of GL02 and Gerfin, Lechner, and Steiger (2005) adjusted for the specifics of the dynamic framework, the different aggregation levels, and sample sizes. Since the outcome variables after period 1, X_1 , do not influence S_1 conditional on X_0 , they are omitted from the first period probits (their entries are denoted as 'X'). Furthermore, the sample sizes and the variation of the dependent and independent variables in some cases require omitting specific variables from the specification (their entries are denoted as '0'). The latter problem is particularly pronounced for the transition from E to EE with only 19 observations leaving the sequence at this stage.

----- Table 2 about here -----

3.2.2 Common Support and the Sensitivity of the IPW Estimator to Extreme Probabilities

It may not be possible to obtain reliable estimates of the mean effects of the target populations as a whole, when some groups of the target population have characteristics that make them very unlikely to appear in one (or both) of the sequences under consideration. If it is impossi-

ble for such groups to appear in one of the treatments, then the support assumption of W-DCIA is violated, which means that the data are uninformative about this part of the target population (other than by extrapolation from other parts of the population). The common remedy is to remove this subpopulation from the target sample and redefine the causal effect appropriately. Even if their probability of occurrence is positive but small, IPW estimators may have bad properties (e.g. Frölich, 2004, Robins, Hernan, and Brumback, 2000, and the Monte Carlo study in the discussion paper version of this paper). Therefore, even in this case the target population needs to be adjusted. Although, there is some discussion of support issues in the matching literature (e.g. Smith and Todd, 2005, Lechner, 2001b), the literature on IPW appears to be silent on that topic. The typical IPW solution to the problem is to remove (or rescale) observations with very large weights and check the sensitivity of the results with respect to the trimming rules used.

Trimming observations, however, changes the target distribution (shrinking large weights changes the target population as well, but in a way that is very hard to describe empirically). This change of the target distribution can be monitored by computing the same weights for the target population as for the treated population (i.e. the weights they would have obtained had they participated in such treatment). Then the subpopulation with values of those probabilities that lead to a removal in one of the treatment samples is descriptively compared to the subpopulation with no such extreme values.

Since two potential outcomes are estimated for the same target population, trimming has to be done in both treated subsamples. However, when the trimming is done independently for both subsamples, the estimates of the two mean potential outcomes may end up being valid for different target populations. Therefore, the same 'types' of observations should be removed. Here, this idea is implemented in a simple way: Denote the weights used to weight observations in sequence 1 by $w_{i \in 1}^1$ and those for observations in sequence 0 by $w_{i \in 0}^0$. Compute both

weights for the target population ($w_{i \in s^j}^0, w_{i \in s^j}^1$) and both treatment populations ($w_{i \in 0}^1, w_{i \in 1}^0$) as well. Define the desired cut-off level for each treatment as \bar{w}^1 and \bar{w}^0 . Reduce \bar{w}^1 (\bar{w}^0) to the largest value of $w_{i \in 0}^1$ ($w_{i \in 1}^0$) if it is below the original \bar{w}^1 (\bar{w}^0). This procedure could be iterated, but this was not necessary in the application. Based on \bar{w}^1 and \bar{w}^0 the target population is adjusted by considering only observations with values of $w_{i \in s^j}^0$ and $w_{i \in s^j}^1$ below \bar{w}^1 and \bar{w}^0 .

Comparing descriptive statistics of confounders in the retained versus the deleted target sample shows how this reduction affects the interpretation of the parameters estimated. Furthermore, predicting the means of the confounders for the reduced sample in the same way as the mean potential outcomes is a second way to characterize the adjusted target population.

In the application, we compare the results of four different trimming rules for the various sequences and subpopulations. The first rule does not trim any observation at all; the second to fourth ones trim a given share of the observations with the largest weights (applied in both treated subsamples separately). The shares that are used are 0.1%, 1%, and 5%, respectively.

3.2.3 Results for the Different Estimators

Table 3 contains four blocs, each bloc corresponding to an IPW estimator using one of the four cut-off rules. For each estimator, Table 3 contains a part related to common support issues (columns 3 to 6, and 12 to 15) and the weights (columns 7, 8, and 16, 17), as well as a part showing the estimates of the potential outcomes (monthly earnings in CHF in December 1999; columns 5, 6 and 18, 19) and the effects of the sequences (11, 20).

The part about the common support begins with the share of observations that would be deleted by adjusting the reference distribution (3, 12). The deleted share increases with the decreasing cut-off points for the weights and it varies with the different sequences investigated.

Concerning trimming, using the 5% trimming leads to a loss of between 30% and 55% of the observations, whereas a 0.1% trimming rule leads to considerably smaller losses between 3% and 18%. Comparing different sequences, the largest losses occur when comparing participants in employment programs (EE) to recipients of temporary wage subsidies (TT). This is expected, because they are the most dissimilar groups.

----- Table 3 about here -----

The next column (4, 13) gives the mean of the earnings variable in the last job before treatment. This variable cannot be influenced by the treatment and is thus exogenous. Comparing its mean in the adjusted reference sample with its mean in the original reference sample (same as those for the IPW without trimming) shows the direction of the adjustment. With the exception of some comparisons based on the 5% rule, adjusting the reference distribution does not change the mean of this lagged earnings measure very much. Columns (5, 14) and (6, 15) contain the predictions of this variable based on the two treatment samples using the same weights as used for computing the mean potential outcomes. Since the effect for this variable is zero, those estimates should be similar (if the probability model is correctly specified) and both values should be similar to its mean in the adjusted reference population (4, 13). Taking into account the sampling variability, those conditions appear to be fulfilled. Thus, those comparisons, which are like balancing tests used with static matching, give no indication of a misspecification of the selection model used.

Columns (7, 16) and (8, 17) contain the 10% concentration ratio of the weights in the two subsamples of treated observations (after adjusting for common support). This ratio shows the share of the target population covered by those 10% of the treatment population that have the largest weights. In the case of random assignment, this ratio would be about 10. However, if only a single observation would be used to predict outcomes for all members of the target population, this ratio would be 100. The results show systematically higher ratios when small

treated groups are used to predict outcomes for large (and diverse) target populations, but overall do not indicate that the estimates are dominated by very few observations.

Columns (9, 18) and (10, 19) show the predicted counterfactual mean earnings of the sequences (1) for the respective target populations (2). Comparing the exogenous lagged earnings of both treatment groups with those of the target group using the lower panel of Table 1 gives an idea whether we expect IPW to adjust mean observed post-treatment earnings upwards or downwards. Indeed these adjustments appear to move the estimators in the expected direction in almost all cases.

Comparing the stability of the results across estimators (and thus across different common supports) shows qualitatively similar results, although the 5% trimming rule leads to a fairly different magnitude of the effect for the comparison of UU to CC for population U. In this case, the latter estimator finds considerably lower mean potential earnings for participating in CC compared to the other estimators. This goes along with the fact that when using the weights for the subpopulation in CC, the predicted mean lagged earnings are lower than the sample mean for the 5%-rule, but are higher for the other estimators with more extreme cut-off values. This difference is reflected in the effect estimates as well. However, even though these differences exist, standard errors for that mean counterfactual are rather large. Therefore, it is probably not possible to reject the hypothesis that the estimators give the same results.

Finally, it should be pointed out that for several comparisons the number of observations ending up in the sequences is too small to lead to significant results, although the estimated effects are not too small (like for the comparison of UU to EE or EE to TT).

3.2.4 Effects of the Program Sequences on Labor Market Outcomes

In addition to the result for December 1999, Table 4 presents results for outcome variables 1, 6, 17, 20 months after the end of the sequence of interest to shed some light on the development of the effects over time. Furthermore, since it is the ‘official’ objective for the active labor market policy to increase reemployment chances, in addition the earnings variable that is shown in the upper panel of Table 4, an outcome variable measuring whether an individual is employed or not is presented in the lower panel of Table 4.

----- Table 4 about here -----

To summarize, the results suggest that it is clearly better to participate in a course or receive a wage subsidy for two periods compared to remaining unemployed over that period. However, for employment programs it does not matter whether an unemployed participates in such a program or remains unemployed. Of course, since the latter is much cheaper, employment programs appear to be ineffective. Looking at the dynamics of the program effects, the reason seems to be a large lock-in effect (individuals in the program search less intensively than those outside the program), since the typical duration of an employment program would be six months, which is longer than for the other measures. The comparison between wage subsidies and employment programs indicates the superiority of the former when both have at least a duration of two periods, but this effect is not precisely enough estimated for a firm conclusion.

3.2.5 Standard Errors

Table 5 shows the results of five different estimators for the standard errors. The first two procedures shown in columns (3, 8, 13) and (4, 9, 14) are agnostic about the specific structure of the first-step estimator. They exploit the fact that the mean potential outcomes are weighted

means of the observed outcomes in the respective sequence. To ease notation, define normal-

ized weights as $\tilde{w}_i^{s_2^k, s_\tau^j} = \frac{\hat{w}_i^{s_2^k, s_\tau^j}}{\sum_{i \in s_2^k} \hat{w}_i^{s_2^k, s_\tau^j}}$. \hat{V}^1 estimates the variance conditional on the weights:

$$\hat{V}^1 = \widehat{Var}\left(\sum_{i \in s_2^k} \tilde{w}_i^{s_2^k, s_\tau^j} y_{t,i} \mid \hat{w}_i^{s_2^k, s_\tau^j}\right) = \sum_{i \in s_2^k} (\tilde{w}_i^{s_2^k, s_\tau^j})^2 \widehat{Var}(y_{t,i} \mid \hat{w}_i^{s_2^k, s_\tau^j}); \quad j \in \{k, l\}.$$

The conditional variance of the outcome appearing in this term is estimated by a k-nearest-neighbor method. The number of neighbors is set to $2\sqrt{N}$. It is robust to rather large changes in that smoothing factor, indicating that heteroscedasticity conditional on the weights may not be a big issue in this application. Note that inference based on \hat{V}^1 is informative about the effects defined for the sample, not the population (see Abadie and Imbens, 2006, for the distinction between population and sample treatment effects).

The second version acknowledges the randomness of the weights ($Var(A) = E(Var(A/B)) + Var(E(A/B))$). Therefore, the empirical variance of $\tilde{w}_i^{s_2^k, s_\tau^j} \hat{E}(y_{t,i} \mid \hat{w}_i^{s_2^k, s_\tau^j})$ in the respective subsample is added to \hat{V}^1 , again using nearest-neighbor estimation to estimate the conditional expectation. Furthermore, since the observations in the different sequences are independent, the variance of the effect is computed as the sum of the variances of the potential outcomes. These estimators are in the same spirit as proposals by Abadie and Imbens (2006).

Since the parametric IPW estimators can be interpreted as sequential GMM estimators as, for example, analyzed by Newey (1984), the GMM approach is used to compute standard errors as well. Columns (5, 10, 15) show the version which ignores the estimation of the probit coefficients, whereas the results in columns (6, 11, 16) takes full account of all random elements in the estimation process. All details on the GMM variances are presented in Appendix B.2.

Finally, columns (7, 12, 17) present naïve bootstrap estimates by taking the empirical standard deviation over the effects estimated in 1000 random draws of 100%-subsamples. Therefore,

the bootstrap estimator is the only estimator that explicitly takes into account the variation that comes from the trimming rules.

Estimates are presented for the effects (13-17) as well as for the potential outcomes (3-7, 8-12). Note that the estimation of the weights may induce a correlation between the outcomes, so that the variance of the effect does not necessarily equal the sum of the variances of the mean potential outcomes.

Comparing the estimates across the board, it is obvious that for the larger treatment samples all estimates are close together (level of UU for target population U, T, or E), whereas for the smaller samples they deviate much more from each other. It may be suspected that this results from less precise estimation of the variances in the smaller samples. In general, it is hard to detect any systematic trends, other than \hat{V}^2 being systematically larger than \hat{V}^1 (which must be the case by construction of the estimators). Furthermore, in many cases \hat{V}^1 gives the smallest variance. Comparing the two GMM estimators with each other, we see that, with 5 exceptions, taking into account the fact that the selection probabilities are estimated reduces the variance. This feature of IPW estimation has been frequently observed in the literature (see for example Hirano, Imbens, and Ridder, 2003, Lunceford and Davidian, 2004, Robins, Rotnitzky, and Zhao, 1995, or Wooldridge, 2002). Finally, the bootstrapped standard errors appear to be at the lower end of the values presented, which is somewhat surprising. However, it might be conjectured that the rule brings in some non-smoothness that leads to a deviation from the validity of the bootstrap. However, it is beyond the scope of this paper to investigate the issue of precise variance estimation of the IPW estimator in depth. Here, it will suffice to see that the different estimators lead to similar statistical decisions for most cases.

4 CONCLUSION

This paper discusses sequential inverse selection probability weighted estimators to estimate mean causal effects defined within the dynamic causal model introduced by Robins (1986). GMM theory is used to provide a distribution theory for the case when the selection probabilities are estimated nonparametrically. Using an evaluation study of the Swiss active labor market policy, some implementational issues, like the so-called common support problem, are discussed and results are provided. The application indicates that the dynamic approach allows investigating interesting questions about causal effects of different regimes of active labor market policies. It also indicates that precise answers to these new questions require sufficiently large samples.

ACKNOWLEDGEMENTS

I am also affiliated with CEPR, London, ZEW, Mannheim, and IZA, Bonn, and PSI, London. Financial support from the Swiss National Science Foundation (projects 4043-058311 and 4045-050673) is gratefully acknowledged. Part of the data originated from a database generated for the evaluation of the Swiss active labor market policy together with Michael Gerfin. I am grateful to the Swiss State Secretariat for Economic Affairs (*seco; Arbeitsmarktstatistik*) and the Swiss Federal Office for Social Security (*BSV*) for providing the data and to Dragana Djurdjevic, Michael Gerfin, and Heidi Steiger for their substantial input in preparing them. I thank Ruth Miquel and Blaise Melly for helpful comments on several issues raised in the paper. I thank Conny Wunsch and Stephan Wiehler for careful proof reading. I presented the paper in seminars at Cambridge University, Mannheim University, Tinbergen Institute, Amsterdam, and the ESRC meeting in Bristol. I thank participants for helpful comments. Furthermore, I thank an associate editor of this journal as well as anonymous referees for helpful comments on earlier versions of the paper. The usual disclaimer applies.

APPENDIX A: MULTIPLE TREATMENTS AND MANY PERIODS

The main issue concerns the specification of the propensity scores: For example, when specifying the probability of participating in s_2^k conditional on participating in s_1^k , is it necessary to account for the fact that not participating in s_2^k implies a range of possible other states in period 2? The answer is no, because in each step the independence assumption relates only to a binary comparison, e.g. $Y_2^{s_2^k} \prod \mathbb{1}(S_2 = s_2^k) | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1$, and

$Y_2^{s_1^k} \prod \mathbb{1}(S_1 = s_1^k) | S_1 \in \{s_1^j, s_1^k\}, X_0 = x_0$ (s_1^j being the target population as before). Therefore, the conditional probabilities of not participating in the event of interest conditional on the history are sufficient (Imbens (2000) and Lechner (2001a) develop the same argument to show that in the static multiple treatment models conditioning on appropriate one-dimensional scores is sufficient). Hence, $P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1)$, and

$P[S_1 = s_1^k | X_0 = x_0, S_1 \in \{s_1^l, s_1^k\}]$ may be used to construct the weights. The multiple treatment feature of the problem does not add to the dimension of the propensity scores.

APPENDIX B: THE SEQUENTIAL INVERSE PROBABILITY WEIGHTED ESTIMATORS

B.1 Nonparametric Identification

In this appendix, two examples show that the SIPW estimators proposed in Section 2.3 are indeed estimating the desired quantities. For simplicity, it is assumed that the probabilities are estimated consistently and with enough regularity such that the following exposition based on true probabilities holds asymptotically with estimated probabilities as well.

a) The SIPW estimator for $E(Y_2^{s_2^k})$ is given in detail:

$$E \frac{1}{N} \sum_{i=1}^N \frac{y_{2,i} \mathbb{1}(s_{1,i} = s_1^k) \mathbb{1}(s_{2,i} = s_2^k)}{P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_{1,i}) P(S_1 = s_1^k | X_0 = x_{0,i})} =$$

$$= E \frac{Y_2^{s_2^k} \mathbb{1}(S_1 = s_1^k) \mathbb{1}(S_2 = s_2^k)}{P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1) P(S_1 = s_1^k | X_0 = x_0)} =$$

$$\begin{aligned}
&= E_{X_0} \frac{E\left[\frac{Y_2^{s_2^k} \mathbb{1}(S_1 = s_1^k) \mathbb{1}(S_2 = s_2^k)}{P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1)} \mid X_0 = x_0\right]}{P(S_1 = s_1^k | X_0 = x_0)} = \\
&= E_{X_0} \frac{E\left[\frac{Y_2^{s_2^k} \mathbb{1}(S_2 = s_2^k)}{P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1)} \mid S_1 = s_1^k, X_0 = x_0\right] P(S_1 = s_1^k | X_0 = x_0)}{P(S_1 = s_1^k | X_0 = x_0)} = \\
&= E_{X_0} E_{X_1 | X_0, S_1} \frac{E[Y_2^{s_2^k} \mathbb{1}(S_2 = s_2^k) \mid S_1 = s_1^k, \underline{X}_1 = \underline{x}_1]}{P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1)} = \\
&= E_{X_0} E_{X_1 | X_0, S_1} \frac{E[Y_2^{s_2^k} \mid \underline{S}_2 = s_2^k, \underline{X}_1 = \underline{x}_1] P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1)}{P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1)} = \\
&= E_{X_0} E_{X_1 | X_0, S_1} E[Y_2^{s_2^k} \mid \underline{S}_2 = s_2^k, \underline{X}_1 = \underline{x}_1] = \\
&= E_{X_0} E_{X_1 | X_0, S_1} E[Y_2^{s_2^k} \mid S_1 = s_1^k, \underline{X}_1 = \underline{x}_1] = E_{X_0} E[Y_2^{s_2^k} \mid S_1 = s_1^k, X_0 = x_0] = E_{X_0} E[Y_2^{s_2^k} \mid X_0 = x_0] = E(Y_2^{s_2^k}).
\end{aligned}$$

b) The SIPW estimator for $E(Y_2^{s_2^k} \mid S_1 = s_1^j)$ (of which $E(Y_2^{s_2^k} \mid S_1 = s_1^k)$ is a special case) is obtained by

using the same steps as in a) and noting that $\frac{N}{N^{s_1^j}}$ and $\frac{N}{N^{s_2^j}}$ converge to $\frac{1}{P(S_1 = s_1^j)}$ and

$\frac{1}{P(\underline{S}_2 = s_2^j)}$. Furthermore, Bayes rule gives the connection between the conditional and unconditional

counterfactuals. For $E(Y_2^{s_2^k} \mid S_1 = s_1^j)$, for example, using $f(x_0) = \frac{f(x_0 \mid S_1 = s_1^j) P(S_1 = s_1^j)}{P(S_1 = s_1^j \mid X_0 = x_0)}$ leads to

the desired result:

$$\begin{aligned}
&E_{X_0} \left\{ E[Y_2^{s_2^k} \mid S_1 = s_1^k, X_0 = x_0] \frac{P(S_1 = s_1^j \mid X_0 = x_0)}{P(S_1 = s_1^j)} \right\} = \\
&= E_{X_0} \left\{ E[Y_2^{s_2^k} \mid S_1 = s_1^j, X_0 = x_0] \frac{P(S_1 = s_1^j \mid X_0 = x_0)}{P(S_1 = s_1^j)} \right\} = \\
&= E_{X_0 | S_1 = s_1^j} \left\{ E[Y_2^{s_2^k} \mid S_1 = s_1^j, X_0 = x_0] \frac{P(S_1 = s_1^j \mid X_0 = x_0)}{P(S_1 = s_1^j)} \frac{P(S_1 = s_1^j)}{P(S_1 = s_1^j \mid X_0 = x_0)} \right\} = \\
&= E_{X_0 | S_1 = s_1^j} E[Y_2^{s_2^k} \mid S_1 = s_1^j, X_0 = x_0] = E[Y_2^{s_2^k} \mid S_1 = s_1^j].
\end{aligned}$$

B.2 Distribution of Two-step Estimator With Parametric First Step

The previous section showed that the suggested sequential inverse probability weighted estimator is non-parametrically identified. Here, the asymptotic distribution of the specific estimator used in the application is presented. It consists of estimating the weights in a first step without considering the outcomes. In the second step, the estimated weights are used to obtain the reweighted mean of the outcomes in the respective subsample. The general structure of the estimator is as follows:

$$\hat{\theta}_{\underline{s}_t^k; \underline{s}_t^l}(\underline{s}_t^j) = \frac{1}{N} \sum_{i=1}^N w_i^{\underline{s}_t^k; \underline{s}_t^l; \underline{s}_t^j}(\underline{x}_{1,t} \underline{s}_{2,i}, \hat{\beta}) y_{t,i}.$$

For simplicity, reduce the notation, so that $\hat{\theta} = \frac{1}{N} \sum_{i=1}^N w_i(\underline{x}_{1,t} \underline{s}_{2,i}, \hat{\beta}) y_{t,i}$. The weights are known functions of

the data and the estimated transition probabilities. Assume that these probabilities are known up to a finite number of unknown coefficients, β . Furthermore, assume that there is a consistent, asymptotically normal estimator of these coefficients, $\hat{\beta}$, obtained by solving a number of moment conditions of the same dimension as β .

These moment conditions are denoted by $g(\underline{x}_{1,i}, \underline{s}_{2,i}; \beta)$ and have expectation zero at the true value of β . The estimator for choosing the unknown values of β fulfils the following condition:

$$\frac{1}{N} \sum_{i=1}^N g(\underline{x}_{1,i}, \underline{s}_{2,i}; \hat{\beta}) = 0.$$

For example, this would be the case if a probit model is used for the estimation of the various probabilities. In this case, the moment conditions correspond to the score of the likelihood function. Note that this moment condition is a collection of all moment conditions that are required to compute all necessary transition probabilities that enter the weight function. Thus, β is a collection of all parameters necessary to characterize the transition probabilities required for a particular weight function.

To ease notation, rewrite the estimator of the effect in terms of a moment condition:

$$h(y_{t,i}, \underline{x}_{1,t} \underline{s}_{2,i}, \beta, \theta) := \theta - w_i(\underline{x}_{1,t} \underline{s}_{2,i}, \beta) y_{t,i} ;$$

$$\frac{1}{N} \sum_{i=1}^N h(y_{t,i}, \underline{x}_{1,t} \underline{s}_{2,i}, \hat{\beta}, \hat{\theta}) = 0.$$

These equations constitute a standard parametric generalized method of moments (GMM) estimation problem as analyzed in the classical paper by Hansen (1982) as well as by Newey and McFadden (1994). Newey (1984) considered the special case that is of relevance here, namely that some of the moment conditions depend only on a subset of the parameters (here the moment conditions coming from the first stage selection steps do not depend on the causal effects of the treatments). Newey (1984) established the regularity conditions (conditions (i) to (iv) on page 202) required for consistency and asymptotic normality of this type of estimator. These conditions require that the data come from stationary and ergodic processes, that first and second moment functions and their respective derivatives exist, are measurable, continuous, and the parameters are finite and not on the boundary of the parameter space. Furthermore, the derivatives of the moment conditions with respect the parameters have full rank. The sample moments should converge to their population counterparts with diminishing variances and lead to uniquely identified solutions for the unknown coefficients (consistent estimators).

When these standard GMM conditions (usually considered to be rather weak particularly in the case of data that are i.i.d. in the dimension 'I') hold, then the resulting (sequential) GMM estimator is consistent and asymptotically normal (for the proof see Hansen, 1982, Newey, 1984, or Newey and McFadden, 1994). For our special case, Newey (1984) devised an easy way to obtain its asymptotic covariance matrix:

$$asVar(\sqrt{N}\hat{\theta}) = H_{\theta}^{-1}(V_{hh} + H_{\beta}G_{\beta}^{-1}V_{gg}G_{\beta}^{-1}'H_{\beta}' - H_{\beta}G_{\beta}^{-1}V_{gh} - V_{hg}G_{\beta}^{-1}H_{\beta}')H_{\theta}^{-1}';$$

$$H_{\theta} := E \frac{\partial h(\cdot)}{\partial \theta} = 1; \quad H_{\beta} := E \frac{\partial h(\cdot)}{\partial \beta} = -E \frac{\partial w_i(\underline{x}_{1,i}, \underline{s}_{2,i}, \beta)}{\partial \beta} y_{i,i}; \quad G_{\beta} := E \frac{\partial g(\cdot)}{\partial \beta};$$

$$V_{hh} := E[h(\cdot)^2] = Var[w_i(\underline{x}_{1,i}, \underline{s}_{2,i}, \beta) y_{i,i}]; \quad V_{gg} := E[g(\cdot)g(\cdot)']; \quad V_{gh} := E[g(\cdot)h(\cdot)]; \quad V_{hg} := V_{gh}'.$$

Therefore, without specifying the model used to estimate the probabilities, we obtain the following specialization of the general variance formula:

$$asVar(\sqrt{N}\hat{\theta}) = Var[w_i(\underline{x}_{1,i}, \underline{s}_{2,i}, \beta) y_{i,i}] + \\ + E \left[\frac{\partial w_i(\cdot)}{\partial \beta} y_{i,i} \right] G_{\beta}^{-1} V_{gg} G_{\beta}^{-1}' E \left[\frac{\partial w_i(\cdot)}{\partial \beta} y_{i,i} \right] + E \left[\frac{\partial w_i(\cdot)}{\partial \beta} y_{i,i} \right] G_{\beta}^{-1} V_{gh} + V_{hg} G_{\beta}^{-1} E \left[\frac{\partial w_i(\cdot)}{\partial \beta} y_{i,i} \right].$$

Ignoring the estimation of the first step leads to $asVar(\sqrt{N}\hat{\theta}) = Var[w_i(\underline{x}_{1,i}, \underline{s}_{2,i}, \beta) y_{i,i}]$, because it would

imply setting $\frac{\partial w_i(\cdot)}{\partial \beta} = 0$. Whether this variance is actually smaller or larger than the full variance, depends on the

correlation of the moment conditions and on $\frac{\partial w_i(\cdot)}{\partial \beta}$. Under certain conditions, the full variance is smaller than the

variance ignoring that probabilities are estimated (see for example Hirano, Imbens, and Ridder, 2003, Lunceford and Davidian, 2004, Robins, Rotnitzky, and Zhao, 1995, or Wooldridge, 2002).

To understand the structure of the covariance matrix better, let us consider more special forms of the weights, like

those implied by $\hat{\theta}_i^{s_2^k; s_1^l}(s_1^k) = \frac{1}{N} \sum_{i=1}^N w_i^{s_2^k; s_1^l; s_1^k}(\underline{x}_{1,i}, \underline{s}_{2,i}, \hat{\beta}) y_{i,i}$:

$$w_i^{s_2^k; s_1^l; s_1^k}(\underline{x}_{1,i}, \underline{s}_{2,i}, \beta) = 1(\underline{s}_{2,i} = \underline{s}_2^k) w_i^{s_2^k; s_1^l; s_1^k}(\underline{x}_{1,i}, \underline{s}_{2,i}, \beta) - 1(\underline{s}_{2,i} = \underline{s}_2^l) w_i^{s_2^l; s_1^l; s_1^k}(\underline{x}_{1,i}, \underline{s}_{2,i}, \beta).$$

$$w_i^{s_2^k; s_1^l; s_1^k}(\cdot) = \frac{1}{P(S_2 = s_2^k | \underline{X}_1 = \underline{x}_{1,i}, S_1 = s_{1,i}^k)}; \quad w_i^{s_2^l; s_1^l; s_1^k}(\cdot) = \frac{P(S_1 = s_1^k | X_0 = x_{0,i})}{P(S_2 = s_2^l | \underline{X}_1 = \underline{x}_{1,i}, S_1 = s_{1,i}^l) P(S_1 = s_1^l | X_0 = x_{0,i})}.$$

Assume that the transition probabilities are estimated by a probit model and denote by $\Phi(a)$ the c.d.f. of the

standard normal distribution evaluated at a . Therefore, the equalities $P(S_1 = s_1^k | X_0 = x_{0,i}) = \Phi(x_{0,i} \beta^{s_1^k})$,

$$P(S_1 = s_1^l | X_0 = x_{0,i}) = \Phi(x_{0,i} \beta^{s_1^l}), \quad P(S_2 = s_2^k | \underline{X}_1 = \underline{x}_{1,i}, S_1 = s_{1,i}^k) = \Phi(\underline{x}_{1,i} \beta^{s_2^k}), \quad \text{and}$$

$$P(S_2 = s_2^l | \underline{X}_1 = \underline{x}_{1,i}, S_1 = s_{1,i}^l) = \Phi(\underline{x}_{1,i} \beta^{s_2^l})$$

lead to the following specialized weights:

$$w_i^{s_2^k; s_1^l; s_1^k}(\underline{x}_{1,i}, \underline{s}_{2,i}, \beta^{s_2^k}) = \frac{1}{\Phi(\underline{x}_{1,i} \beta^{s_2^k})}; \quad w_i^{s_2^l; s_1^l; s_1^k}(\underline{x}_{1,i}, \underline{s}_{2,i}, \beta^{s_1^k}, \beta^{s_2^l}, \beta^{s_1^l}) = \frac{\Phi(x_{0,i} \beta^{s_1^k})}{\Phi(\underline{x}_{1,i} \beta^{s_2^l}) \Phi(x_{0,i} \beta^{s_1^l})}.$$

Taking derivatives of those weights with respect to $\beta = (\beta^{s_1^k}, \beta^{s_2^k}, \beta^{s_1^l}, \beta^{s_2^l})'$ is straightforward and is omitted here for the sake of brevity.

As usual in GMM estimation, consistent estimators can be obtained by estimating the expectations with sample means and plugging in consistent estimators for the unknown coefficients.

REFERENCES

- Abadie, A., and Imbens, G. W. (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects", *Econometrica*, 74, 235-267.
- Angrist, J. D., and Krueger, A. B. (1999), "Empirical Strategies in Labor Economics", *Handbook of Labor Economics*, eds. Ashenfelter, O., and Card, D., Vol. III A, Amsterdam: North-Holland, 1277-1366.
- Dawid, A. P. (1979), "Conditional Independence in Statistical Theory", *Journal of the Royal Statistical Society B*, 41, 1-31.
- Frölich, M. (2004), "Finite-sample properties of propensity score matching and weighting estimators", *Review of Economics and Statistics*, 86, 77-90.
- Gerfin, M., and Lechner, M. (2002), "Microeconomic Evaluation of the Active Labour Market Policy in Switzerland," *The Economic Journal*, 112, 854-893.
- Gerfin, M., Lechner, M., and Steiger, H. (2005), "Does subsidised temporary employment get the unemployed back to work? An econometric analysis of two different schemes", *Labour Economics - An International Journal*, 12, 807-835.
- Gill, R. D., and Robins, J. M. (2001), "Causal Inference for Complex Longitudinal Data: the continuous case", *The Annals of Statistics*, 2001, 1-27.
- Hahn, J. (1998), "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects", *Econometrica*, 66, 315-331.
- Hansen, L. P. (1982), "Large Sample Properties of Generalized Methods of Moments Estimators", *Econometrica*, 50, 1029-1055.
- Heckman, J. J., LaLonde, R. J., and Smith, J. A. (1999), "The Economics and Econometrics of Active Labor Market Programs", *Handbook of Labor Economics*, eds. Ashenfelter, O., and Card, D., Vol. III A, Amsterdam: North-Holland, 1865-2097.
- Hernan, M. A., Brumback, B., and Robins, J. M. (2001), "Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Trials," *Journal of the American Statistical Association*, 96, 440-448.
- Hirano, K., and Imbens, G. W. (2001), "Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Heart Catheterization", *Health Services and Outcome Research Methodology*, 2, 259-278.

- Hirano, K., Imbens, G. W., and Ridder, G. (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score", *Econometrica*, 71, 1161-1189.
- Horvitz, J. L., and Thomson, D. (1952), "A Generalisation of Sampling without Replacement from a Finite Population", *Journal of the American Statistical Association*, 47, 663-685.
- Imbens, G. W. (2000), "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87, 706-710.
- Imbens, G. W. (2004), "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review", *Review of Economics and Statistics*, 86(1), 4-29.
- Joffe, M. M., Ten Have, T. R., Feldman, H. I., and Kimmel, S. (2004), "Model Selection, Confounder Control, and Marginal Structural Models", *The American Statistician*, November 2004, 58-4, 272-279.
- Lechner, M. (2001a): "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption", in M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Active Labour Market Policies*, Heidelberg: Physica, 43-58.
- (2001b), "A note on the common support problem in applied evaluation studies", *Discussion Paper 2001-01*, Department of Economics, University of St. Gallen.
- Lechner, M., and Miquel, R. (2001), "A Potential Outcome Approach to Dynamic Programme Evaluation – Part I: Identification", *Discussion Paper 2001-07*, Department of Economics, University of St. Gallen; revised 2005 (new title: "Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions").
- Liang, K.-Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models", *Biometrika*, 73, 13-22.
- Lunceford, J. K., and Davadian, M. (2004), "Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study", *Statistics in Medicine*, 23, 2937-2960.
- Miller, M. E., Ten Have, T. R., Reboussin, B. A., Lohman, K. K., and Rejeski, W. J. (2001), "A Marginal Model for Analysing Discrete Outcomes From Longitudinal Surveys With Outcomes Multiple-Cause Nonresponse", *Journal of the American Statistical Association*, 455, 844-857.
- Nevo, A. (2003), "Using Weights to Adjust for Sample Selection When Auxiliary Information is Available", *Journal of Business & Economic Statistics*, 21, 43-52.

- Newey, W.K. (1984), "A Method of Moments Interpretation of Sequential Estimators", *Economics Letters*, 14, 201-206.
- Newey, W.K., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing", *Handbook of Econometrics*, Vol. 4, eds. Engle, R.F., and McFadden, D.L., 2113-2245, Amsterdam: North-Holland.
- 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.
- (1993), "Analytical methods for HIV treatment and cofactor effects", *AIDS Epidemiology – Methodological Issues*, eds. Ostrow, D. G., and Kessler, R., 213-290, New York: Plenum Publishing.
- (1994), "Correcting for non-compliance in randomized trials using structural nested mean models", *Communications in Statistics*, 23, 2379-2412.
- (1998a), "Testing and estimation of direct effects by reparametrizing direct acyclic graphs with structural nested models", *Computation, Causation, and Discovery*, eds. Glymour, C., and Cooper, G., Cambridge: MIT.
- (1998b), "Marginal structural models", *1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*, 1-10.
- (1999), "Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference", *Statistical Models in Epidemiology: The Environment and Clinical Trials*, eds. Halloran, M. E., and Berry, D., IMA Volume 116, 95-134, New York: Springer-Verlag.
- Robins, J. M., Greenland, S., and Hu, F.-C. (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 and rejoinder, 708-712.
- Robins, J. M., Hernan, M. A., and Brumback, B. (2000), "Marginal Structural Models and Causal Inference in Epidemiology", *Epidemiology*, 11 (5), 550-560.

- Robins, J. M., and Rotnitzky, A. (1995), "Semiparametric Efficiency in Multivariate Regression Models with Missing Data", *Journal of the American Statistical Association*, 90, 122-129.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995), "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data", *Journal of the American Statistical Association*, 90, 106-121.
- Rubin, D. B. (1980), "Discussion of 'Randomization Analysis of Experimental Data in the Fisher Randomization Test' by D. Basu," *Journal of the American Statistical Association*, 75, 591-593.
- Smith, J., and Todd, P. (2005), "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?", *Journal of Econometrics*, 125 (1-2), 305-353.
- Wooldridge, J. M. (2002), "Inverse Probability Weighted M-estimators for Sample Selection, Attrition, and Stratification", *Portuguese Economic Journal*, 1, 117-139.
- Wooldridge, J. M. (2004), "Inverse Probability Weighted Estimation for General Missing Data Problems", forthcoming in *Journal of Econometrics*.

TABLES

Table 1: Selected Descriptive Statistics of Control Variables and Intermediate Outcomes

Variables	Means / shares in % in subsamples							
	Subpopulations (treatment status Jan/Feb., March/April '98)							
	U	UU	C	CC	E	EE	T	TT
<i>Selected control variables (measured in Dec. 1997 or before)</i>								
Insured earnings in CHF	3943	3904	4008	3932	3592	3517	4023	4067
Chances to find a job (employability): no information	9	8	5	5	8	7	11	14
very easy	8	8	6	5	3	4	6	6
easy	17	16	19	18	10	12	18	18
medium	55	56	59	60	58	58	56	54
difficult	9	10	11	10	17	15	7	7
special case	1	1	1	1	4	4	1	2
Sanction days without benefits (current spell; until Dec. '97)	3.1	3.2	3.8	3.2	3.5	3.3	2.6	2.3
<i>Selected intermediate outcome variables (measured in February 1998) after 1st period</i>								
Employed	13	4	2	1	1	0	6	1
Earnings in CHF	509	164	1283	51	33	0	229	30
Chances to find a job (employability): no information	2	2	2	2	0	0	3	3
very easy (or some employment)	20	12	8	5	4	4	12	8
easy	16	16	17	16	13	14	20	21
medium	52	57	60	62	59	59	56	59
difficult	9	11	13	14	20	19	7	7
special case	1	2	1	1	3	4	1	1
Sanction days without benefits Jan. + Feb. '98	1.2	1.3	.9	.8	.5	.4	.5	.6
<i>Outcome variables</i>								
Gross monthly earnings in May 1998 (in CHF)	1648	1345	987	867	590	371	1919	1368
Gross monthly earnings in September 1998 (in CHF)	2151	1913	1809	1728	1356	1112	2407	2124
Gross monthly earnings in September 1999 (in CHF)	2792	2533	2805	2683	2243	2163	3046	2921
Gross monthly earnings in December 1999 (in CHF)	2420	2171	2758	2699	2105	2068	2713	2595
Employed in May 1998	28	22	17	14	12	8	29	22
Employed in September 1998	36	31	30	28	28	25	39	35
Employed in September 1999	49	44	49	46	40	39	52	52
Employed in December 1999	40	36	44	42	37	37	45	46
Sample size	7985	5122	573	316	118	99	790	382

NOTE: The sample is based on the selection criteria of Gerfin and Lechner (2002), but constrained to those entering unemployment in the 4th quarter of 1997. The sequences are defined on a bimonthly basis. The states in each period are abbreviated as U (unemployed), C (training course), E (employment program), and T (temporary wage subsidy). X (XX) means being unemployed at the end of 1997 and in state X (XX) for the first 2 (4) months in 1998 (X = U, C, E, T).

Table 2: Probit Specifications and Coefficient Estimates

Variable	U vs. C	U vs. E	U vs. T	UU if U	CC if C	EE if E	TT if T
Age in years / 10	-.04	-.00	.02	.06	-.00	-.44	-.00
Female	-.16	.26	-.22	-.06	.05	.02	.17
Mother tongue not German / French / Italian	-.20	.37	.14	.04	.23	0	.11
Mother tongue G / F / I, not main language in own canton	-.09	-.09	-.16	-.04	.14	.16	.19
French mother tongue	-.26	.18	-.12	-.07	.47	-.65	.25
Italian mother tongue	-.16	-.33	-.12	.09	.59	0	-.03
Nationality: foreign with yearly permit	-.03	.26	.08	.11	.02	.40	-.03
Foreign Languages: other Swiss language	.05	.26	.13	-.03	.30	-1.14	.06
Job position very low	.03	0	-.03	-.11	.16	.58	.02
Qualification level: skilled (highest)	-.17	-.13	-.03	-.04	-.00	1.11	-.22
unskilled (lowest)	.01	-.11	.07	.04	-.19	-.62	.00
Chances to find a job: no information	.25	-.01	-.03	-.00	-.06	0	.37
(reference category: medium) very easy	.11	.36	.06	-.23	.52	0	-.08
easy	.02	.28	-.01	-.06	.22	0	.10
difficult	.05	-.11	.13	-.00	-.05	0	.15
special case	.07	-.14	.07	-.02	0	0	0
Chances to find a job : no inform. Feb '98	X	X	X	.06	0	0	0
(reference category: medium) very easy Feb '98	X	X	X	.19	-.61	.32	.05
easy Feb '98	X	X	X	.00	-.31	.71	-.09
difficult Feb '98	X	X	X	-.01	-.02	.32	-.19
special c. Feb '98	X	X	X	.05	0	0	0
Looking for job as % of full-time job, Feb. '98 (0-1)	X	X	X	1.95	-.52	-.52	1.12
Unemployment-status: full-time unemployed	-.20	-.14	-.15	-.52	-1.00	-.71	-.16
part-time unemployed	.15	-.11	-.04	.13	-1.43	0	.38
No information on desired full or part-time job	.03	-.00	.11	1.54	-.67	0	1.14
Desired = previous occupation, 3-digit level, Feb. '98	X	X	X	-.09	.11	.44	.14
Previous occupation: metals	.26	.15	0	.06	-.76	0	0
painting, technical drawing	0	0	.12	0	0	0	-.33
entrepreneurs, senior officials, justice	0	0	.51	0	0	0	-.14
office and computer	-.24	.14	.27	-.04	.01	0	.06
retail trade	-.10	.44	.24	.08	-.38	0	.17
science	-.45	.16	0	-.02	-.07	0	0
Duration of UE spell (days / 10)	-.78	-.43	0	.12	-.13	-.05	0
Unemployment benefits / 1000 in 1996	-.01	.00	0	.07	-.01	0	0
in 1997	.07	-.07	.07	.00	-.00	.00	.01
Month of entry into social security system	-.46	-.02	.18	.09	.72	-1.07	-.36
Share of employment 1988-1997	-.35	.59	-.49	-.34	.57	0	.75
Average duration of UE spells 1988-1997 / 10	-.80	.73	1.21	-.01	3.95	0	-.31
Subsidized temporary job before Dec. 97	0	0	-.38	0	0	0	.19
Size of town (previous employment) < 30.000	-.03	-.05	-.03	-.14	0	0	0
Region of placement office (PO) in rural area	.00	.11	-.07	.07	.01	-.20	-.15
Inflow rate to long-term unemployment (by region of PO)	.53	.41	.56	.19	.37	-1.98	.25

Table 2 to be continued.

Table 2: continued

Variable		U vs. C	U vs. E	U vs. T	UU if U	CC if C	EE if E	TT if T
Region (reference Central)	west	-.01	.08	-.20	-.20	-.57	0	.06
	east	0	.67	0	0	0	0	0
	Zurich	.18	.69	.12	.24	-.24	.29	-.47
	south-west	.55	.68	-.24	-.08	-.68	0	-.13
	north-west	.42	.43	.21	.14	-.74	-.38	-.03
	Ticino	.56	.82	.37	.20	-1.01	0	.19
	Additional regional effects by canton: Bern	.24	.02	.14	.37	.02	0	-.21
French main language in canton	.11	-.15	.39	.47	-.18	0	-.13	
Monthly insured earnings (in last job) in CHF / 1000	-.02	.09	-.02	-.04	-.01	-.26	.05	
Employed Feb. '02	X	X	X	-1.16	-.58	0	-1.62	
Current unemployment spell is first spell	-.02	.01	.12	.07	.31	.80	-.07	
Sanction days without benefits (current spell) / 10	.01	.02	.05	.03	-.03	-.09	-.05	
Sanction days without benefits / 10 Jan. + Feb. '98	X	X	X	.02	-.20	0	.17	
Subsample	U or C	U or E	U or T	U	C	E	T	
Number of observations in subsample	8512	8100	8772	7982	574	118	790	
Dependent variable	U	U	U	UU	CC	EE	TT	
Mean of dependent variable in subsample	.94	.98	.91	.64	.55	.83	.48	

NOTE: Binary probit model estimated on the respective subsample. All specifications include an intercept and are subjected to specification tests (omitted variables, non-normality). If not stated otherwise, all information in the variables relates to the last day in December 1997. *Exclusion of variables: 0*: Variables omitted from specification. *X*: Variable not temporarily prior to dependent variable. **Bold** letters in *italics* denote significance at the 1% level. **Bold** letters denote significance at the 5% level. *Italics* denote significance at the 10% level. See also note below Table 1.

Table 3: Estimation Results for Gross Monthly Earnings (in CHF) in December 1999

Sequences	Target population	% of target sample deleted	Common support					Effects on current earnings				Common support					Effects on current earnings				
			Mean of Y_0		Predicted value of $E(Y_0 s_1)$			Concentration ratio of weights (in %)		$E(Y_t^{s_2} s_1)$		% of target sample deleted	Mean of Y_0		Predicted value of $E(Y_0 s_1)$			Concentration ratio of weights (in %)		$E(Y_t^{s_2} s_1)$	
			s_1	s_2^1	s_2^0	s_2^1	s_2^0	s_2^1	s_2^0	s_2^1	s_2^0		$\theta_t^{s_2^1, s_2^0}(s_1)$	s_1	s_2^1	s_2^0	s_2^1	s_2^0	s_2^1	s_2^0	$\theta_t^{s_2^1, s_2^0}(s_1)$
$s_2^1 - s_2^0$	s_1	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)		
IPW without adjustment											IPW – 0.1% of largest weights removed										
UU-CC	U	0	3943	3912	4048	16	42	2232	3140	-907	15	3920	3911	4116	15	41	2227	3098	-871		
	C	0	4008	3971	4054	32	21	2316	2943	-626	3	4005	3972	4034	30	20	2271	2907	-636		
UU-EE	U	0	3943	3911	3912	17	46	2232	1971	260	18	3844	3822	3666	16	42	2207	2166	41		
	E	0	3592	3557	3598	42	18	2037	2076	-39	5	3532	3497	3558	42	18	2014	2118	-103		
UU-TT	U	0	3943	3911	3886	17	33	2234	2480	-246	13	3935	3925	3841	15	32	2194	2536	<i>-347</i>		
	T	0	4023	3969	4053	29	20	2341	2574	-233	8	4040	3996	4037	26	18	2261	2567	-306		
TT-EE	T	0	4023	4053	4165	19	48	2592	2047	545	16	3941	3978	3827	18	40	2500	2361	139		
	E	0	3590	3624	3598	45	18	2313	2076	237	10	3581	3604	3613	44	17	2249	2161	88		
IPW – 1% of largest weights removed											IPW – 5% of largest weights removed										
UU-CC	U	22	3925	3926	4065	14	37	2231	2991	-759	44	3966	3972	3904	13	30	2307	2692	<i>-385</i>		
	C	9	3985	3989	3987	31	21	2293	2848	-555	30	3895	3945	3914	29	19	2281	2731	-449		
UU-EE	U	24	3816	3801	3711	15	36	2185	2394	-209	50	3675	3674	3683	13	32	2077	1973	104		
	E	11	3572	3542	3596	39	17	2056	2146	-89	32	3602	3552	3603	34	15	2111	2201	-90		
UU-TT	U	17	3940	3947	3813	14	29	2179	2454	-274	30	3992	3998	4005	13	24	2198	2458	-260		
	T	14	4037	3998	4034	25	16	2257	2620	-362	30	4018	4026	4005	22	15	2284	2590	-305		
TT-EE	T	36	3827	3785	3669	16	39	2412	2196	215	55	3763	3730	3655	15	33	2479	2503	-23		
	E	25	3666	3652	3717	41	17	2271	2197	73	43	3637	3741	3657	35	16	2561	2263	298		

NOTE: *Sequences*: Treatments for which the effects are estimated. *Target population*: Population for which the effects of the treatments are estimated. *Common support*: Part of the target sample for which the effects could be estimated from observations in the two subsamples of treatment participants. In the *adjusted target sample* observations for whom the treatment sample could not be used to predict the effects are removed (the share of the removed part is shown in the column headed by % of target sample deleted). *Predicted value of ...* means that the mean of Y_0 is predicted using the same weights as used for the estimation of the effects. *Concentration ratio of weights*: Sum of the 10% largest weights divided by the sum of all weights *100. Y_0 denotes 'insured monthly earnings' in CHF (used by the unemployment insurance to calculate unemployment benefits). Since Y_0 is computed from the UE insurance data, they may not be perfectly comparable to the outcome variable that is calculated from the pension records. *Earnings* are coded as 0 if individuals receive a temporary wage subsidy, participate in a programme, or are unemployed. For the effects, **bold** letters *in italics* denote significance at the 1% level. **Bold** letters denote significance at the 5% level. *Italics* denote significance at the 10% level. Estimates for the same counterfactual mean outcome may differ across different comparisons because the common support is comparison specific. See also note below Table 1.

Table 4: Dynamics of the Average Potential Outcomes and Effects (IPW, 0.1% trimming)

Seq. $s_2^1 - s_2^0$	Target pop- ulation	$E(Y_t^{s_1^1} S_1 = s_1)$				$E(Y_t^{s_2^0} S_1 = s_1)$				$\theta_t^{s_1^1, s_2^0}(s_1)$			
		1998		1999		1998		1999		1998		1999	
		May	Sept.	Sept.	Dec.	May	Sept.	Sept.	Dec.	May	Sept.	Sept.	Dec.
<i>Gross monthly earnings (in CHF)</i>													
UU-CC	U	1373	1911	2539	2227	1002	1826	3126	3098	371	85	-587	-871
	C	1198	1780	2484	2271	880	1773	2906	2907	317	8	-421	-636
UU-EE	U	1375	1928	2561	2207	350	1346	2577	2166	1025	582	-15	41
	E	1128	1662	2214	2014	422	1165	2179	2118	705	496	34	-104
UU-TT	U	1365	1934	2564	2194	1143	1918	2808	2536	221	15	-243	-341
	T	1395	1997	2607	2261	1394	2136	2879	2567	1	-138	-271	-306
TT-EE	T	1402	2106	2840	2500	457	1319	2641	2361	944	787	199	139
	E	1218	1818	2628	2249	442	1221	2264	2161	776	596	364	88
<i>Employment (in %)</i>													
UU-CC	U	22	31	44	37	17	29	51	47	5	2	-7	-10
	C	18	29	43	36	14	28	49	46	5	0	-7	-10
UU-EE	U	22	32	45	37	7	30	51	39	16	2	-5	-2
	E	19	29	41	35	8	26	37	37	11	3	4	-1
UU-TT	U	23	32	44	36	19	31	52	46	4	1	-7	-11
	T	23	32	44	36	23	35	51	45	-1	-4	-7	-9
TT-EE	T	24	35	52	46	9	30	44	38	15	5	8	9
	E	22	33	51	44	9	28	40	37	13	6	12	7

NOTE: See notes below Tables 1 and 3.

Table 5: Asymptotic Standard Errors for Gross Monthly Earnings (IPW 0.1% trimming)

Seq. $s_2^1 - s_2^0$	Target pop.	$E(Y_t^{s_1^1} S_1 = s_1)$					$E(Y_t^{s_2^0} S_1 = s_1)$					$\theta_t^{s_1^1, s_2^0}(s_1)$				
		weights		GMM		boots- trap	weights		GMM		boots trap	weights		GMM		boots trap
		fixed	rnd.	fixed	full		fixed	rnd.	fixed	full		fixed	rnd.	fixed	full	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
UU-CC	U	37	39	39	39	42	234	339	361	350	230	237	341	362	352	231
	C	48	58	56	77	65	155	181	255	199	151	162	190	207	206	152
UU-EE	U	37	39	40	39	61	341	433	418	414	343	343	434	419	415	348
	E	52	62	60	127	90	228	248	235	234	237	233	256	243	262	231
UU-TT	U	35	37	37	37	50	160	195	194	190	150	164	198	198	194	155
	T	42	49	48	67	51	127	133	131	130	123	134	141	140	144	129
TT-EE	T	194	251	231	226	219	238	257	244	240	246	308	360	336	328	313
	E	137	145	141	135	148	381	480	472	426	370	405	501	493	447	394

NOTE: See notes below Tables 1 and 3. *Weights fixed*: Std. error of weighted mean estimator conditional on weights allowing for weight dependent heteroscedasticity. *Weights rnd.*: Std. error of weighted mean estimator allowing for weight dependent heteroscedasticity and accounting for randomness of weights. *GMM fixed*: Std. errors derived from GMM approach ignoring estimation of probabilities used to compute weights. *GMM full*: Std. errors derived from GMM accounting for the estimation of the probabilities used to compute weights. *Bootstrap*: Standard deviation of 1000 independent bootstrap estimates of the effects.