

published in H. Engelhardt; H.-P.Kohler, and Alexia Fürnkranz-Prskawetz (eds.),
Causal Analysis in Population Studies, Concepts, Methods, Applications,
ISBN: 978-1-4020-9966-3, Chapter 3, 31-57

SEQUENTIAL POTENTIAL OUTCOME MODELS TO ANALYZE THE EFFECTS OF FERTILITY ON LABOR MARKET OUTCOMES

Michael Lechner

SIAW

Swiss Institute for International Economics
and Applied Economic Research



University of St.Gallen

First draft: November 2006

Revised: October 2007

Date this version has been printed: 16 October 2007

Abstract

This paper proposes to use dynamic treatment models to analyze the effects of fertility on labor market interactions when large data set are available. The main advantages are (i) its flexibility due to its nonparametric nature, (ii) its potential of allowing careful consideration of the selection issues coming from the dynamic interaction between fertility and labor market outcomes; and (iii) the possibility of defining relevant parameters of interest in a precise and detailed way. Based on artificial data that mimic important features of real data sets, the approach is implemented and issues that come up in any practical application of this approach are discussed.

Keywords: Dynamic treatment regimes, nonparametric identification, causal effects, sequential randomization, treatment effects, dynamic matching.

Address for correspondence

Michael Lechner

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

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

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

1 Introduction^{*}

This paper proposes to use dynamic treatment models to analyze the effects of fertility on labor market interactions. It argues that when large data sets are available the dynamic potential outcome model is an interesting modeling framework because it allows the careful consideration of the selection issues coming from the interaction of fertility and labor market decisions at different ages. It allows explicitly considering their dependence on the labor market and fertility history realized up to that period. There is no need to collapse the 'endogeneity' problem into a static setting since the dynamic nature and timing of the interaction can be explicitly addressed. Furthermore, the paper argues that this approach allows defining relevant parameters of interest in a more precise way. Based on artificial data, the approach is implemented and issues that may come up in practical applications of this approach are discussed.

The literature on the effect of fertility on labor market outcomes can be organized along the dimensions of model structure and time. The first strand, the so-called structural approach, uses fully structural behavioral models typically based on some sort of utility maximization subject to time and budget constraints. Usually, these models are fully parametrically specified. An early example of this approach is Moffitt (1984). There is in fact a considerable literature which is based on rather sophisticated structural modeling of the individual decision problems. These parametric models are combined with econometric modeling of the uncertainties in the model or the data. The resulting moment conditions or likelihood functions are then the basis for a parametric estimation of the parameters that are of particular interest in the specific application. This literature is surveyed in Arroyo and Zhang (1997) and Hotz, Klerman, and Willis (1997), for example. Recent papers based on this approach are Del

^{*} I have further affiliations with ZEW, Mannheim, CEPR and PSI, London, IZA, Bonn, and IAB, Nuremberg. The paper has been prepared for the conference "Causal Analysis in Population Studies: Concepts, Methods, Applications" in Vienna, 2006.

Bocca and Sauer (2005), Francesconi (2002) and Klepinger, Lundberg, and Plotnick (1999), among others.

An alternative approach to structural modeling is reduced form modeling, i.e. deriving the equation to estimate not directly from some mathematically formulated theoretical model, but instead specifying the empirical model to be estimated by appealing to some general properties of various theoretical models of interest in a more or less ad-hoc way. A prominent paper based on this approach analyzing the timing and spacing of births in Sweden is Heckman and Walker (1990) who estimate duration models. Recent papers by Adsera (2005a, b) are based on a fairly similar approach.

A related approach that is more explicit on the various endogeneity problems that almost naturally occur in the previously mentioned group of papers models labor market outcomes and fertility decisions jointly based on selection type models. Examples for this approach are Hotz and Miller (1988), Di Tommaso (1999), and Troske and Voicu (2004).

However, even when the different decisions are analyzed within a joint modeling framework, it remains questionable whether the different effects can really be identified independently of each other. Therefore, Angrist and Evans (1998) propose an instrumental variable approach that allows analyzing the labor supply reaction when family size is varied exogenously. The latter variation comes from the observation that parents are more likely to get a third child, if the first two children have the same sex. Using an explicit non- or semiparametric causal framework, like in the analyses of post-unification fertility in East and West Germany by Lechner (1998, 2001a), is a rather unusual approach in that literature.

Reduced form models appear to be typical as well for the demographic literature. For example, Rosenfeld (1996) and Connelly (1996) describe the work-fertility interaction without strict behavioral models, be it mathematical or not, and without a comprehensive empirical

analysis. Note that none of the mentioned econometric or demographic reduced form analyses uses a dynamic framework.

The virtue of the structural approach is that the behavioral assumptions of the model are very clear. The drawback is however that the tight parametric functions are restrictive and usually rejected by the data, at least when the sample is large enough. In contrast, the virtue of the reduced form approach, particularly in its nonparametric causal version, is that it usually does not impose more than the required just identifying restrictions on the data. The drawback may be that those just identifying conditions are not explicitly derived from a formal mathematical model of utility maximization. Therefore, the conditions for their validity sometimes may be less clear than in the case of structural models.

The second dimension 'time' is straightforward. Some paper collapse individual fertility and labour market histories into one observation at a specific point in time, while other empirical approaches follow the realization of fertility and labor market outcomes over time. The virtue of the former approach is clearly its simplicity, while the latter approach, which requires more and better data, allows taking into account the important time dimension of fertility and labor market events.

Next, consider recent advances in econometric methodology. One of the recent important developments in econometrics is the increased emphasis on discovering causal as opposed to associational relations from the data and clarifying the conditions required for the causal interpretation of the estimators used. Only causal relations are useful for policy advice, because they contain the reaction of the economic variables of interest.¹ Econometrics developed two different ways to define what a causal effect is. One concept originated in time series econometrics. The other concept comes from the sphere of microeconometrics and statistics.

¹ See the excellent account of the historical developments in econometrics by Heckman (2000).

The concept used in time series econometrics is due to Wiener (1956), Granger (1969), and Sims (1972) (see the review article by Geweke, 1984, for an overview). Their basic idea is that (non-) causality is very similar, if not the same, than (non-) predictability. Therefore, they consider one variable not to cause another variable, if the current value of the causing variable does not help to predict future values of the variables that might capture the effects of this cause. This statement is conditional on the information set available at each point in time. This concept is in principle (technically) applicable if one cross-sectional unit (e.g. a country) is observed for a sufficiently long period.

The alternative concept popular in microeconometrics, particularly and most explicitly in the program evaluation literature (e.g. Heckman, LaLonde, and Smith, 1999), is based on the idea that the relevant comparison is between different states of the world, each of which relates to a value of the causing variable. If causation is absent, then the outcomes that would have been realized if those potential states had actually been occurred would be the same (in some probabilistic sense). To relate this concept of different states of the world to data, it is necessary to observe different sample units in the different states. Then, so-called identifying assumptions are employed to relate the observed data to the distribution of the potential outcome variables, so that causal effects can be inferred from the 'real world' that is reflected in the data. The statistical formulation of the resulting inference problem is probably due to Neyman (1923) and was extended and popularized by Rubin (1974). Recently, dynamic versions of the potential outcome approach were suggested by Robins (1986) and extended by Lechner and Miquel (2001) that will be explained below.² Lechner (2004) proposes matching type estimators for this model.

² Pötter and Blossfeld (2001) discuss concepts of causality and the use of the time dimension in sociology. However, they do not consider the research potential of dynamic potential outcome models.

The paper proceeds as follows: Section 2 outlines the dynamic causal framework. The notation is introduced and the basic identification conditions are restated. The estimation problem is explained in Section 3 and sequential matching estimation is reviewed. Section 4 relates the model explicitly to the substantive questions at stake in the pseudo-empirical example. Section 5 details the artificial data. Section 6 presents the 'empirical' examples and Section 7 concludes.

2 The dynamic causal model - notation, effects, and identification

2.1 Introduction

Robins (1986) first suggested an explicitly dynamic causal framework based on potential outcomes that allows the definitions of causal effects of dynamic interventions and systematically addresses this type of selection problem. His approach was subsequently applied in epidemiology and biostatistics (e.g. Robins, 1989, 1997, 1999, Robins, Greenland, and Hu, 1999, for discrete treatments; Gill and Robins, 2001, for continuous treatments) to define the effect of treatments in discrete time. Identification is achieved by sequential randomization assumptions (see the very comprehensible summary by Abbring, 2003). The effects are frequently estimated using parametric models.

⁷ A further exception is Ding and Lehrer (2003) who use this framework and related work by Miquel (2002, 2003) to evaluate a sequentially randomized class size study using difference-in-difference-type estimation methods.

Recently, Lechner and Miquel (2001, LM01 further on) extend Robins' (1986) framework to comparisons of more general sequences, different parameters, and different selection processes. Focusing on the case when all elements that influence selection and outcomes at each stage of the sequence are observable, LM01 discuss different identification conditions required for particular dynamic causal effects. Since the assumptions used in LM01 bear enough similarity to the selection on observables or conditional independence assumption (CIA) that is prominent in the static evaluation literature, Lechner (2004, L04 further on) proposed matching and inverse probability weighting estimators that are dynamic extensions of similar estimators used in the static model. These estimators retain most of the flexible and convenient properties of the static methods that made them the workhorse in empirical econometric evaluation studies (see the excellent survey by Imbens, 2004). Lechner (2008b) discusses some operational characteristics of this approach in the context of the evaluation of labor market programs. Although this approach has been applied by Lechner and Wiehler (2007) to analyze the effects of the timing and order of Austrian active labor market programs, generally applications of this explicit dynamic causal framework based on potential outcomes are however rare in econometrics so far.⁷

The following sections briefly repeat the definitions of the dynamic causal model as well as the identification results derived by Lechner and Miquel (2001) for the case of sequential selection on observables. To ease the notational burden, I use a three-period-two-treatments model to discuss the most relevant issues that distinguish the dynamic from the static model, although in the application more periods and more treatments are considered. As usual in the econometric evaluation literature, I use the standard statistics terminology based on treatments (fertility) and potential outcomes (labor market indicators) to define causal effects.

2.1 Basic structure of the model

For the fertility example, it is appropriate to treat age (in years) as the relevant time dimension, and consider calendar time as an individual exogenous attribute. Suppose that there is an initial period in which everybody is in the same treatment, plus two subsequent periods in which different treatment states are realized. The initial period is tied to the subpopulation of interest. Suppose further that interest is in the effects of births on the women aged 25 who did not give birth so far. In this case, the initial period would relate to age 25 and the population of interest (which then leads to the sample used in the estimation) would be the women of age 25 who did not give birth so far. However, no restrictions will be imposed on the population that relate to the periods after age 25. Periods are indexed by t or τ ($t, \tau \in \{0, 1, 2\}$). The treatment is defined over all periods. It is described by a vector of Bernoulli random variables (RV), $S = (S_1, S_2)$. S_t measures the occurrence of a birth in period t . For notational convenience, the treatment of the initial period ($S_0 = 0$) is sometimes not mentioned explicitly. 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 that variable, i.e. $\underline{s}_t = (s_1, s_2)$.⁸ Since effect heterogeneity is not restricted over time, it makes sense to define potential outcomes in terms of sequences of potential states of the world. Thus, in period one, a woman is observed in exactly one of two treatments. In period two, the treatment will be described by two potential outcomes depending on what happened in period 1. Therefore, she is part of one of four treatments defined by the sequences (0,0), (1,0), (0,1), and (1,1). Thus, every individual is observed in exactly one sequence defined by s_1 and another sequence defined by the same value s_1 and a value s_2 . To sum up, in the two (plus one)-period-two-treatments example there are six different overlap-

⁸ To differentiate between different sequences, sometimes a letter (e.g. j) is used to index a sequence, as in \underline{s}_t^j .

As a further convention, capital letters usually denote random variables, whereas small letters denote specific

ping potential outcomes corresponding to two mutually exclusive states defined by treatment status in period 1 only, plus four mutually exclusive states defined by treatment status in periods 1 and 2 together. Such states could be characterized for example by one birth in period 1, zero births in period 2, or by no births at all, or any other sequence of births.

Variables used to measure the effects of the treatment in period t , i.e. the potential outcomes, are indexed by treatments and denoted by $Y_t^{s_1}$ ($t \geq 1$) or $Y_t^{s_2}$ ($t \geq 2$). They are measured at the end of each period, whereas treatment status is measured in the beginning of each period. For each sequence length (1 or 2 periods), one of the potential outcomes is observable and denoted by Y_t . Here, the potential outcomes measure individual labor market status, like labor market participation or earnings. To link the potential outcomes of the causal model to the data, the following observation rules are defined in equation (1):

$$\begin{aligned}
 Y_1 &= S_1 Y_1^1 + (1 - S_1) Y_1^0; \\
 Y_2 &= S_1 Y_2^1 + (1 - S_1) Y_2^0 = S_1 S_2 Y_2^{11} + (1 - S_1) S_2 Y_2^{01} + S_1 (1 - S_2) Y_2^{10} + (1 - S_1) (1 - S_2) Y_2^{00}. \quad (1)
 \end{aligned}$$

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

2.2 Defining the estimand: average causal effects

Although it has already been stated that interest is in finding the effects of fertility on post-fertility labor market outcomes, it remains to formulate this estimand in terms of the dynamic causal model. First, assume for the sake of the simplified example that interest is in the effects of births at age 26 and 27 on the labor market outcome at age 28 (or any later age). It is

values of the random variables. When I deviate from this convention, the intended meaning should be obvious.

important to note at this stage that this analytical framework allows for complete effect heterogeneity, i.e. different individuals may react differently to the same treatment sequence. Therefore, I will define different average treatment effects for different subpopulations based on the comparison of the same treatment sequences. Those subpopulations may be characterized by exogenous characteristics or more importantly by the treatment sequences themselves, i.e. the fertility sequences the women actually experienced in those two years.

As in the static model, the potential outcomes are used to define several average causal effects. Equation (2) defines the causal effect (for period t) of a sequence of treatments up to period 1 or 2 (τ, τ') compared to an alternative sequence of the same or a different length for a population defined by one of those sequences or a third sequence:

$$\theta_t^{\underline{s}_t^k, \underline{s}_t^l}(\underline{s}_t^j) = E(Y_t^{\underline{s}_t^k} | \underline{S}_t = \underline{s}_t^j) - E(Y_t^{\underline{s}_t^l} | \underline{S}_t = \underline{s}_t^j),$$

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

The treatment sequences indexed by k, l , and j may correspond to (0) or (1) if τ (or τ') denotes period 1, or to the longer sequences (0,0), (0,1), (1,0), or (1,1) if τ (or τ') equals two. LM01 call $\theta_t^{\underline{s}_t^k, \underline{s}_t^l}$ the dynamic average treatment effect (DATE). Accordingly, $\theta_t^{\underline{s}_t^k, \underline{s}_t^l}(\underline{s}_t^k)$, as well as $\theta_t^{\underline{s}_t^k, \underline{s}_t^l}(\underline{s}_t^l)$ are termed DATE on the treated (DATET) and DATE on the nontreated.

There are cases in-between, like $\theta_t^{\underline{s}_t^k, \underline{s}_t^l}(\underline{s}_t^l)$, for which the conditioning set is defined by a sequence shorter than the one defining the causal contrast. Note that the effects are by definition symmetric for the same population ($\theta_t^{\underline{s}_t^k, \underline{s}_t^l}(\underline{s}_t^k) = -\theta_t^{\underline{s}_t^l, \underline{s}_t^k}(\underline{s}_t^k)$). This feature, however, does not restrict effect heterogeneity across individuals ($\theta_t^{\underline{s}_t^k, \underline{s}_t^l}(\underline{s}_t^k) \neq \theta_t^{\underline{s}_t^k, \underline{s}_t^l}(\underline{s}_t^l)$).

2.3 Identification

Having analytically defined the causal framework and the objects of interest, assumptions that are plausible in our application are required to find consistent estimators for those estimands, i.e. the estimands have to be linked to the data.

Assume that a large sample $\{s_{1i}, s_{2i}, x_{0i}, x_{1i}, y_{1i}, y_{2i}\}_{i=1:N}$ of size N is available, randomly drawn from a large population defined by $S_0 = 0$. The latter is characterized by the corresponding random variables $(S_1, S_2, X_0, X_1, Y_1, Y_2)$.⁹ Furthermore, assume that all conditional expectations that are of interest in the remainder of this paper exist. To ease notation further, assume that interest is in the effects of sequences of length two only. If the variables that jointly influence selection at each stage as well as the outcomes are observable, some average treatment effects are identified (weak conditional independence assumptions):

*Weak dynamic conditional independence assumption (W-DCIA)*¹¹

- 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) $1 > P(S_1 = 1 \mid X_0 = x_0) > 0, 1 > P(S_2 = 1 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1) > 0; \forall \underline{x}_1 \in \underline{\mathcal{X}}_1, \forall s_1 : s_1 \in \{0, 1\}.$

$\underline{\mathcal{X}}_1 = (\mathcal{X}_0, \mathcal{X}_1)$ denotes the support of X_0 and X_1 . Part a) of W-DCIA states that the potential outcomes are independent of treatment choice in period 1 (S_1) conditional on X_0 . This is the standard version of the static CIA (e.g. Rubin, 1974). Part b) states that conditional on the

⁹ To simplify the notation further, I consider period 2 as the only period relevant for the outcome of interest. However, for all what follows Y_2 should be considered as measured at some point in time after treatment 2 occurred. The exact timing is determined by the substantive interest of the researcher conducting the empirical study. Here, it could be any labor market outcome after the last day of age 27.

¹¹ $A \perp\!\!\!\perp B \mid C = c$ 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).

treatment in period 1, on observable outcomes of period 1 (which may be part of X_1), and on the confounding variables from periods 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 this application, the question is which variables influence births as well as subsequent labor market outcomes and whether such variables are observable. If the answer to the latter question is yes, and if there is common support (defined in part c) of W-DCIA), i.e. there are individuals with the same observable characteristics that are observed in both treatment sequences of interest, then there is identification, even if some or all conditioning variables in period 2 are influenced by the labor market and fertility outcomes of period 1. LM01 show 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, $\theta_2^{s_2^k : s_2^l}$ and $\theta_2^{s_2^k : s_2^l}(s_1^j)$ 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 $\theta_2^{s_2^k : s_2^l}(s_2^j)$ is not identified if $s_1^l \neq s_1^k$, or $s_1^l \neq s_1^j$, or $s_1^k \neq s_1^j$. This result states that pair-wise comparisons of all birth sequences are identified, but only for groups of women defined by their birth status in periods 0 or periods 0 and 1 together. The relevant distinction between the populations defined by fertility states in period 1 and subsequent periods is that in period 1, treatment choice is random conditional on exogenous variables, which is the result of the initial condition stating that $S_0 = 0$ holds for everybody. However, in the second period, randomization into these treatments is conditional on variables already influenced by the first part of the treatment. W-DCIA has an appeal for applied work as a natural extension of the static framework. However, W-DCIA does not identify the classical treatment effects on the treated which would define the population of interest using one of the complete sequences (for all three periods), if the sequences of interest differ in period 1.

LM01 show that to identify all treatment parameters, W-DCIA must be strengthened by essentially 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-period treatment. This can be summarized by an independence condition like $Y_2^{\Delta} \perp\!\!\!\perp \underline{S}_2 \mid \underline{X}_1$ (LM01 call this the strong conditional dynamic independence assumption, *S-DCIA*). Note that the conditioning set includes the outcome variables from the first period. This is the usual conditional independence assumption used in the multiple treatment framework (with four treatments; see Imbens, 2000, and Lechner, 2001). In other words, when the control variables are not influenced by the previous treatments, the dynamic problem collapses to a static problem of four treatments with selection on observables. This result shows that by treating the dynamic selection process explicitly, the identifying assumption can be relaxed and still interesting effects are identified. In the example, the strong conditional independence assumption amounts to assuming that (intermediate) labor market outcomes at age 25 and 26 are not influenced by the contemporaneous birth events, which clearly is unrealistic.¹²

To sum up, note that the dynamic concept based on sequential conditional independence assumptions allows for weaker conditions on the selection process than using a static model: whereas the static model requires a selection on observables assumption to hold for all elements of the sequence at once, the dynamic model still works fine in most dimensions if the selection on observables assumption is only valid in each period for the next selection step conditional on the past selections steps, past outcomes, and other past confounders. However, both strategies will break down if there are unobservable characteristics that influence the selection steps as well as the potential outcomes. In this case, IV methods, like the one suggested by Miquel (2003) become relevant.

¹² Note that if this assumption does not hold, then the conditioning variable X_1 would become endogenous and thus change the meaning of the estimand in some way that is hard to interpret (see Lechner, 2008a).

Any attempts of nonparametrically estimating these effects face the problem that distributional adjustments based on a potentially high-dimensional vector of characteristics and intermediate outcomes (X) are required (details below). Therefore, in the applied static matching literature balancing scores are a popular device to reduce the dimension of the estimation problem (see Rosenbaum and Rubin, 1983). Similar properties hold for the dynamic model as well:

Balancing score property for W-DCIA

If the conditions of W-DCIA hold, then:

a) $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_1 \mid b_1(X_0) = b_1(x_0)$ holds for all $b_1(x_0)$ such that

$$E[p^{s_1}(x_0) \mid b_1(X_0) = b_1(x_0)] = p^{s_1}(x_0); \quad p^{s_1}(x_0) := P(S_1 = s_1 \mid X_0 = x_0).$$

b) $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_2 \mid b_2(\underline{X}_1, S_1) = b_2(\underline{x}_1, s_1)$ holds for all $b_2(\underline{x}_1, s_1)$ such that

$$E[p^{s_2|s_1}(\underline{x}_1) \mid b_2(\underline{X}_1, S_1) = b_2(\underline{x}_1, s_1)] = p^{s_2|s_1}(\underline{x}_1); \quad p^{s_2|s_1}(\underline{x}_1) := P(S_2 = s_2 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1).$$

A low-dimensional choice for balancing scores consists of conditional transition probabilities in combination with the variable indicating the treatment in the previous period (which of course can be ignored in the first period): $b_1(x_0) = p^{s_1}(x_0)$, $b_2(\underline{x}_1, s_1) = [p^{s_2|s_1}(\underline{x}_1), s_1]$.

3 Estimation

3.1 Structure of sequential estimators

Lechner (2004) shows that these scores are convenient for constructing sequential propensity score matching estimators to correct for selection bias under W-DCIA. I focus on this particular estimator because of its simplicity and because it is the workhorse of empirical evaluation studies. Other static matching-type estimators can be adapted to the dynamic context in a similar way (see Imbens, 2004, for an overview of available estimators). I refrain from

discussing estimation based on the S-DCIA explicitly, because that assumption is not relevant here.

Using the balancing scores suggested above, the following estimand results for quantities identified under W-DCIA:

$$E(Y_2^{s_2^k} | S_1 = s_1^j) = \underset{p^{s_1^k}(X_0)}{E} \left\{ \underset{p^{s_2^k|s_1^k}(\underline{X}_1)}{E} [E(Y_2 | \underline{S}_2 = \underline{s}_2^k, \underline{p}^{s_2^k|s_1^k, s_1^k}(\underline{X}_1)) | S_1 = s_1^k, p^{s_1^k}(X_0)] | S_1 = s_1^j \right\},$$

$$\underline{p}^{s_2^k|s_1^k, s_1^j}(\underline{X}_1) := [p^{s_2^k|s_1^k}(\underline{X}_1), p^{s_1^j}(X_0)], \quad s_1^k, s_2^k, s_1^j, s_1 \in \{0,1\} \quad (3)$$

To learn the counterfactual outcome for the population participating in s_1^j (the target population) had they participated in sequence \underline{s}_2^k , women with \underline{s}_2^k must be reweighted to make them comparable to the women in the target population (s_1^j). The dynamic, sequential structure of the causal model restricts the possible ways to do so. Intuitively, for the members of the target population, women in the first element of the sequence of interest (s_1^k) should be reweighted such that they have the same distribution of $p^{s_1^k}(X_0)$ as the target population. Call this artificially created group comparison group 1. Yet, to estimate the effect of the full sequence, the outcomes of women in \underline{s}_2^k instead of s_1^k are required. Thus, an artificial subpopulation of women in \underline{s}_2^k that has the same distribution of characteristics of $p^{s_1^k}(X_0)$ and $p^{s_2^k|s_1^k}(\underline{X}_1)$ as the artificially created comparison group 1 is required. The same principle applies for -dynamic average treatment effects in the population (DATE).

All proposed estimators in L04 have the same structure: They are computed as weighted means of the outcome variables observed in subsample $\underline{S}_2 = \underline{s}_2^k$. The weights, $w(\cdot)$ depend on the specific effects of interest and are functions of the balancing scores.

$$\widehat{E(Y_2^{s_2^k} | S_1 = s_1^j)} = \sum_{i \in \underline{S}_2^k} w_i^{s_2^k, s_1^j} (p^{s_2^k | s_1^k, s_1^j}(\underline{x}_{1,i}, s_1^k) y_i); \quad w_i^{s_2^k, s_1^j} \geq 0; \quad \sum_{i \in \underline{S}_2^k} w_i^{s_2^k, s_1^j} = 1; \quad (4)$$

$$\widehat{E(Y_2^{s_2^k})} = \sum_{i \in \underline{S}_2^k} w_i^{s_2^k} (p^{s_2^k | s_1^k}(\underline{x}_{1,i}, s_1^k) y_i); \quad w_i^{s_2^k} \geq 0; \quad \sum_{i \in \underline{S}_2^k} w_i^{s_2^k} = 1. \quad (5)$$

Note that in the case of more than two treatments, the balancing scores for (4) and (5) will differ with respect to the participation probability for the first period. For equation (4), the required quantity is $P(S_1 = s_1^k | X_0 = x_0, S_1 \in \{s_1^k, s_1^l\})$, whereas in equation (5), in which all of the population is the target, $P(S_1 = s_1^k | X_0 = x_0)$ is appropriate.

3.2 Sequential matching estimators (SM)

Lechner (2004) propose to extend the simple pair-matching estimators that are highly popular in applied studies to the dynamic context. The idea is to perform the required adjustments by sequentially choosing close pairs of observations in the various steps, so as to mimic the sequential conditional expectations appearing in expressions (4) and (5). The first step is the same for both effects and consists in finding for every women in $S_1 = s_1^k$ a women in $\underline{S}_2 = \underline{s}_2^k$ with very similar (the same) values of $p^{s_2^k | s_1^k}(\underline{x}_{1,i})$ and $p^{s_1^k}(x_{0,i})$. Note that matching must be with replacement, because the target population may be larger than the treatment population. In the second step, every women in $S_1 = s_1^j$ (equation (4)) or $S_0 = 0$ (equation (5)) is to be paired with a women observed with $S_1 = s_1^k$ with very similar (same) values of $p^{s_1^k}(x_{0,i})$. The positive weights that are attached to some or all women in $\underline{S}_2 = \underline{s}_2^k$ coming from step 1 are then updated depending on how often a women in $\underline{S}_2 = \underline{s}_2^k$ is matched to a women of the target population via the intermediate matching step. This procedure leads to the following weights:

$$w_i^{s_2^k, s_1^j} = \frac{1}{N^{s_1^j}} \sum_{n \in s_1^j} \sum_{m \in s_1^k} v_1[p^{s_1^k}(x_{0,n}), p^{s_1^k}(x_{0,m}); \cdot] v_2[\underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,m}), \underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,i}), s_1^k; \cdot]; \quad \forall i \in \underline{S}_2 = \underline{s}_2^k; \quad (6)$$

$$w_i^{s_2^k} = \frac{1}{N} \sum_{n=1}^N \sum_{m \in s_1^k} v_1[p^{s_1^k}(x_{0,n}), p^{s_1^k}(x_{0,m}); \cdot] v_2[\underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,m}), \underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,i}), s_1^k; \cdot]; \quad \forall i \in \underline{S}_2 = \underline{s}_2^k. \quad (7)$$

$N^{s_1^j}$ denotes the number of observations for which $S_1 = s_1^j$. The function $v_1[p^{s_1^k}(x_{0,n}), p^{s_1^k}(x_{0,m}); \cdot]$ is defined to be one if $p^{s_1^k}(x_{0,m})$ is closest to $p^{s_1^k}(x_{0,n})$ among all observations belonging to the subsample defined by $S_1 = s_1^k$, and zero otherwise. Similarly, $v_2[\underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,m}), \underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,i}), s_1^k; \cdot]$ is one if observation i is closest to observation m (with $s_{1,m} = s_1^k$) in terms of $\underline{p}^{s_2^k | s_1^k}(\underline{x}_{1,i})$ and $p^{s_1^k}(x_{0,i})$, and zero otherwise. The Mahalanobis metric (a quadratic form of the variables defining the distance weighted by the inverse of their sample covariance matrix) is a frequently used measure for similarity. Note that the weight of observation i is 0, if it is not matched to any member of the target population. On the other extreme, if observation i is matched to every member of the target population its weight would be 1. A specific variant of this estimator is shown in Table 1 for the example of estimating $\theta_t^{s_2^1, s_2^0}(s_1^1)$.

Some remarks about this protocol that are already contained in L04 are worth repeating: First, matching is with replacement. Every step of the matching sequence is essentially the same as for matching in a static framework. However, sequential propensity score matching involves several probabilities in the second period matching step. Second, some issues arise from the sequential nature of matching. By choosing observations as matches with *similar* values of the probabilities instead of the *same* values (because such observations may not be available), it may happen that the probabilities attached to observations in earlier matching steps (relating to transitions in early periods) change over different sequential matching steps due to imprecise matching. To prevent this from happening, every matched comparison observation in

period 2 is recorded with the values $\hat{p}_i^{s_1^l}$ of the observation it was matched to in period 1, instead of its own (\hat{p} denotes a consistent estimate of p). Hence, the ‘history’ of the match, or, in other words, the characteristics of the reference distribution, does not change when the next match occurs in the subsequent period.

Third, to compute $E(Y^{s_2^k} | S_1 = s_1^l)$ the only information that is needed for the $N^{s_1^l}$ participants in s_1^l is $\hat{p}_i^{s_1^k}$. Similarly, for participants in s_2^k , all probabilities of the type $\hat{p}_i^{s_2^k | s_1^k, s_1^l}$ are required. For participants in s_1^k , but not in s_2^k , only $\hat{p}_i^{s_1^k}$ is needed, and so on. To estimate $E(Y^{s_2^l} | S_1 = s_1^l)$ instead of $E(Y^{s_2^k} | S_1 = s_1^l)$, the only change in the matching protocol is that the initial matching step on $\hat{p}_i^{s_1^l}$ is redundant. When interest is in the average effect in the population ($E(Y^{s_2^k})$), then the whole population plays the role of the first reference group (instead of s_1^l). In this case, in the matching step based on $\hat{p}_i^{s_1^k}$, all participants in s_1^k are matched to themselves. In addition selected participants in s_1^k are matched to participants in the remaining treatments in the first period.

Table 1: A sequential matching estimator for $\theta_t^{s_2^1, s_2^0}(s_1^1)$ based on propensity scores

Step 0: Sample reduction		Delete all observations not belonging to s_1^1 , s_2^1 , or s_2^0
Step A: Match $s_2^0 = (s_1^0, s_2^0)$ to s_1^1 $(E(Y_t^{s_2^0} S_1 = s_1^1))$	A.1.0	Define a weight $w_i^{s_2^0} = 0$ for every observation in s_2^0 .
	A.1.P	Estimate a probit for $P(S_1 = s_1^1 X_0 = x_0) \rightarrow p_i^{s_1^1}(x_{0,i}) = p_i^{s_1^1}$
	A.1.CS	Delete all obs. of s_1^1 with lower or higher values of $p_i^{s_1^1}$ than obs. in s_2^0
	A.1.M	For every obs. in s_1^1 not deleted in A.1.CS find the obs. in s_2^0 that is closest in terms of $p_i^{s_1^1}$ (a match).
	A.1.C	For the matched obs. keep the value of $p_i^{s_1^1}$ of the obs. in s_1^1 they have been matched to. Some obs. in s_2^0 may appear many times in this matched sample.
	A.2.R	Define a sample of obs. in s_1^1 .
	A.2.P	Estimate a probit for $P(S_2 = s_2^0 S_1 = s_1^0, X_1 = x_1) \rightarrow p_i^{s_2^0 s_1^0}(x_{1,i}) =: p_i^{s_2^0 s_1^0}$
	A.2.CS	Delete all obs. of the matched comparison sample of s_1^1 (defined in A.1.C) (as well as the corresponding elements of the target population s_1^1) with lower or higher values of $p_i^{s_1^1}$ and $p_i^{s_2^0 s_1^0}$ than obs. in s_2^0 .
	A.2.M	For every obs. in the matched comparison sample of s_1^1 not deleted in A.2.CS find an obs. in s_2^0 that is closest in terms of $p_i^{s_2^0 s_1^0}$ and $p_i^{s_1^1}$ using the Mahalanobis metric (covariance computed in s_1^1). Every time an obs. in s_2^0 is matched, its weight $w_i^{s_2^0}$ is increased by 1.
	Step B: Match $s_2^1 = (s_1^1, s_2^1)$ to s_1^1 $(E(Y_t^{s_2^1} S_1 = s_1^1))$	B.1.0
B.2.R		Reduce sample to participants in s_1^1 still in the common support.
B.2.P		Estimate a probit for $P(S_2 = s_2^1 S_1 = s_1^1, X_1 = x_1) \rightarrow p_i^{s_2^1 s_1^1}(x_{1,i}) =: p_i^{s_2^1 s_1^1}$
B.2.CS		Delete all obs. of s_1^1 with lower or higher values of $p_i^{s_2^1 s_1^1}$ than obs. in s_2^1 .
B.2.M		For every obs. in s_1^1 not deleted in B.2.CS find the member of s_2^1 that is closest in terms of $p_i^{s_2^1 s_1^1}$. Every time an obs. in s_2^1 is matched, its weight $w_i^{s_2^1}$ is increased by 1.
Step C: Joint common support	C.1	Reduce $w_i^{s_2^1}$ by 1 for every obs. i matched to an obs. in s_1^1 deleted in A.1.CS or A.2.CS.
	C.2	Reduce $w_i^{s_2^0}$ by 1 for every obs. i matched to an obs. in s_1^1 deleted in B.2.CS.
Step D: Estimation of $\theta_t^{s_2^1, s_2^0}(s_1^1)$	D.1	$\hat{\theta}_t^{s_2^1, s_2^0}(s_1^1) = \frac{1}{\sum_{i \in s_2^1} w_i^{s_2^1}} \sum_{i \in s_2^1} w_i^{s_2^1} y_i - \frac{1}{\sum_{i \in s_2^0} w_i^{s_2^0}} \sum_{i \in s_2^0} w_i^{s_2^0} y_i$

Table 1 to be continued.

Table 1: Continued

Step E: Estimation of $\text{Var}[\hat{\theta}_t^{s_2^1, s_2^0}(s_1^1)]$	D.2	$\widehat{\text{Var}}[\hat{\theta}_t^{s_2^1, s_2^0}(s_1^1)] = \frac{\sum_{i \in \mathcal{S}_2^1} [(w_i^{s_2^1})^2 \widehat{\text{Var}}(Y_i S = s_2^1, w = w_i^{s_2^1})]}{\sum_{i \in \mathcal{S}_2^1} (w_i^{s_2^1})^2} + \frac{\sum_{i \in \mathcal{S}_2^0} [(w_i^{s_2^0})^2 \widehat{\text{Var}}(Y_i S = s_2^0, w = w_i^{s_2^0})]}{\sum_{i \in \mathcal{S}_2^0} (w_i^{s_2^0})^2}$ $\widehat{\text{Var}}(Y_i S = s_2^j, w = w_i^{s_2^j}) = \frac{1}{N^{i \in I(s_2^j, w_i^{s_2^j})}} \sum_{i \in I(s_2^j, w_i^{s_2^j})} \left[y_i - \bar{y}^{I(s_2^j, w_i^{s_2^j})} \right]^2,$ $N^{i \in I(s_2^j, w_i^{s_2^j})} = \sum_{i \in I(s_2^j, w_i^{s_2^j})} 1, \quad \bar{y}^{I(s_2^j, w_i^{s_2^j})} = \frac{1}{N^{I(s_2^j, w_i^{s_2^j})}} \sum_{i \in I(s_2^j, w_i^{s_2^j})} y_i.$ <p>The set $I(s_2^j, w_i^{s_2^j})$ is determined as the $2\sqrt{N}$ closest neighbors w.r.t. the value of $w_i^{s_2^j}$ of observations with same observed treatment as observation 'i'.</p>
--	-----	--

Note: $t > 1$. Changes required for the case $t=1$ are obvious. The number of neighbors in the k-NN estimation of the conditional variance is the one used in L04. In the empirical application below, the results appear not to be sensitive to small deviations from this value.

When matching is on the propensity score instead of directly on the confounding variables, there is the issue of selecting a probability model. It seems that, so far, even in the static model the literature has not addressed this thoroughly. So far the consensus seems to be that a flexibly specified (and extensively tested) parametric model is sufficiently rich and that the choice of the model does not really matter (e.g., see the Monte Carlo results by Zhao, 2004). Similarly, the suggestions in the literature to guide the specification choice by the ability to achieve balancing of the respective covariates (e.g. Rosenbaum and Rubin, 1984, or Rubin, 2004) can be applied here as well (in each step).

Next, there is the issue of consistent estimation of the standard errors that is not yet resolved for the static matching literature. Based on the simulation results presented in L04, the standard errors are computed conditional on the weights. In other words, the fact that the weights are estimated quantities is ignored. Furthermore, the outcomes may show heteroscedasticity. However, heteroscedasticity is only relevant in this context if related to the weights. Therefore, a simple k-nearest neighbor estimator is used as in L04 to adjust for any such heterosce-

dasticity. Although such an estimator performed well in L04, there is potential for improvement.

The final remark about the matching protocol concerns the common support. The region of common support - defined on the reference distribution for which the effect is desired - has to be adjusted period by period with respect to the conditioning variables of that period. The matching estimator makes it easy to trace back the impact of this procedure on the reference distribution.

3.3 *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 take account of 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. $Y_2^{s_2^k} \mathbb{I}(S_2 = s_2^k) | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1$, and $Y_2^{s_1^k} \mathbb{I}(S_1 = s_1^k) | S_1 \in \{s_1^j, s_1^k\}, X_0 = x_0$ (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.¹⁴ Hence, as already noted $P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1)$ and $P[S_1 = s_1^k | X_0 = x_0, S_1 \in \{s_1^j, s_1^k\}]$ may be used in the matching step in period 1. The multiple treatment feature of the problem does not add to the dimension of the propensity scores.

¹⁴ Imbens (2000) and Lechner (2001) develop the same argument to show that in static multiple treatment models conditioning on appropriate one-dimensional scores is sufficient.

4 Specifying causal parameters of interest

4.1 General issues

Since the causal model is formulated in discrete time, the first issue that arises concerns the concept of time to be used and the related question about the length of a period.

An important distinction is between process time, i.e. the clock starts running at some specific event that is related to the object of interest, or calendar time. The choice between the two concepts depends on the application, i.e. whether interesting causal effects are naturally defined in process time (like age specific birth patterns) or more naturally defined in calendar time. If one of those concepts is chosen, usually the other dimension of the problem will be controlled for in the estimation. For example, if process time is given by age, than any analysis based on calendar time will control for age effects and vice versa. In our current example concerning the labor market impacts of differential fertility behavior over the life-cycle, it appears natural to specify the model in terms of process time, here age. Doing so, allows specifying meaningful sequences as well as maximum flexibility in modeling relevant selection processes.

In an ideal world with very large sample sizes and high frequency data, the type of selection problem expected is one of the key determinants of the desirable length of a period. If it is assumed that short term events have important influences on the fertility decisions of the respective women, then shorter time periods allow more flexibility than longer periods. The price to pay for such flexibility is a loss in the precision of the estimates, because if the sequences cover only very specific events, then not many observations will be observed in any such sequence. In addition, it will be hard to interpret the (noisy) effects of sequences that have a very similar economic interpretation (like a birth in the first or second month after the 20th birthday). Therefore, more parsimonious models leading to more precise estimates can be obtained by using longer time periods. Since the systematic factors of fertility decisions are

based on longer term considerations, the application considers two years as one period and counts the number of births within such a period. With smaller samples, further aggregation of periods may be advantageous.

Another issue concerns the length of the sequence specified. The longer the sequences, i.e. the treatments are fixed over a longer time horizon, the more precise is the specification of the contrasts of interest, which could be important in some situation. However, as before, the price to pay will be a larger number of parameters to be estimated with a smaller number of observations.

4.2 *Number of children*

First, consider the issue of estimating the effects of the number of children on labor market outcomes. For example, if interest is in the effect of one child compared to no child at all, one may want to consider sequences that cover (almost) all of the fertile ages of women. The contrasts involve one sequence with zeros everywhere compared to an alternative sequence with one child in some period. Since the timing of birth is not relevant in the above formulated causal question, one may want to aggregate the effects for all possible sequences describing one birth based on some weighting scheme. A plausible weighting scheme could for example be based on the number of observations in each such one-birth-sequence.

If this question is only interesting for example for births up to the age of 30, then the specified one-birth and zero-birth sequences would only cover the ages up to age 30. The same principles apply when comparing two to zero or two to one birth, and so on. Section 6 presents concrete examples.

4.3 *The effects of timing and spacing*

A frequent question that arises is whether early births lead to different labor market outcomes than late births. This can be easily analyzed comparing sequences with the same total number

of total births occurring at different ages. Varying the ages would also allow getting an estimate of the effect of an incremental postponement. Similarly, the spacing of the births, i.e. the time between the first and the second birth may be analyzed by comparing sequences with first births occurring at the same age in both sequences, but with the second birth occurring at a different age. Whereas such sequences could cover (most) of the fertile ages, one may be interested in the effects of 'starting' later or earlier without necessarily keeping the overall number of births fixed, but rather considering it as being determined by the early or late start. In this case, the sequences to be specified are much shorter. They would cover only the ages until the first birth of interest. This age can be varied to understand the full pattern of this effect.

Again, if interest is in a combination of different sequences, all effects for birth sequences with the same distance between births may be aggregated.

4.4 *The role of confounding variables*

Compared to the static potential outcome model, the dynamic potential outcome model is a powerful tool if the selection process into (and the dropout from) the sequences is determined by time varying variables that are related to the outcome of interest. Those variables, termed intermediate outcomes in the previous sections, play an important role in addition to any selection variables that are constant over the time window considered. Therefore, the data available must be informative about such variables. In the second part of section 5, the selection processes into and out of some selected sequences are documented.

5 Data

To illustrate the methods, I generated a rather large sample of 100000 'women' with 'yearly' information on labor market status, fertility, and some background characteristics that show similarity to typical variables measuring education, vocational degree and labor demand.

These women are observed from the age of 16 to the age 45. For simplicity the data do not contain calendar time effects.¹⁵ Before discussing the 'data' in more detail, note that they are simulated for the purpose of illustrating some of the potentials of the suggested dynamic causal methods. They are neither meant to reflect any real-life data nor are they supposed to reflect some specific insights about the connection between fertility and socio-economic variables. Quite to the contrary, when investigating some of the descriptive statistics for specific subsamples to be presented below, it is obvious that some of the statistics show rather extreme properties that are however not detrimental to the purpose intended in this paper. It should also be pointed out that although the actual sample size is rather large compared to real life data, the suggested approach can be used with smaller sample sizes as well.

In the simulated data, the number of births per year is modeled by a latent index model of the probit type. Choices depend on observables that also appear in the outcome processes as well as on normally distributed unobservables that are independent of observables and unobservables appearing in the outcome equations. Note obviously that normality or an index model assumption is not required for W-DCIA to hold, but they are convenient choices for a simulation exercise. Such observables are schooling, vocational degree, regional indicators, as well as lagged labor market indicators and the birth history. The labor market indicators, such as earnings and employment, are modeled as dynamic processes influenced by the exogenous variables as well as by their own past and current and past fertility behavior. All selection processes fulfill W-DCIA, but not D-CIA.

Table 2 contains descriptive statistics as well as a characterization of the type of variable for the most important time-dependent and time-independent variables. They are the usual types of variables with typical codings, means, and standard errors. A full set of statistics for all

¹⁵ Since space constraints do not allow reproducing the data generating process explicitly, a Gauss 8.0 program is available from the author on request. Furthermore, a Gauss 8.0 program performing the estimation is available as well.

variables are available on request from the author. All descriptive statistics are shown for particular subsamples defined by selected treatment sequences. Since these are simulated data, I refrain from any interpretation of the descriptive statistics. However, it is important that the variation of the values of the covariates and intermediate outcomes across sequences reveals considerable selection effects as well as considerable differences in the outcome variables.

Table 2: Descriptive statistics for selected variables and selected subsamples

Variable	Type	Subsamples defined by fertility status between age 21 up to age 42												
		all		1		0		101		000000110		110000000		
		mean	std.	mean	std.	mean	std.	mean	std.	mean	std.	mean	std.	
Monthly earnings in EUR (0 for non-workers)														
Age	25	C	1133	1078	190	596	1393	1034	35	242	1846	450	241	699
	35	C	1108	1172	362	839	1314	1166	248	674	1040	905	962	1124
	45	C	1239	1445	409	970	1465	1475	259	708	1511	1068	949	1257
Out of labor force in %														
Age	25	I	38	-	83	-	25	-	93	-	.1	-	75	-
	35	I	42	-	77	-	32	-	82	-	13	-	48	-
	45	I	40	-	75	-	31	-	80	-	14	-	48	-
Employment in %														
Age	25	I	57	-	10	-	71	-	2	-	99	-	12	-
	35	I	53	-	17	-	63	-	13	-	61	-	46	-
	45	I	54	-	18	-	64	-	13	-	75	-	43	-
Total number of children per women														
Age	25	D	.7	.9	1.9	.5	.3	.6	1.7	.4	0	-	2	-
	35	D	1.5	1.2	2.6	.7	1.1	1.2	2.6	.6	.5	.5	2	-
	45	D	1.7	1.3	2.9	.8	1.5	1.2	2.8	.6	2.1	.3	2.1	.3
Other variables														
Schooling (8-12)		D	10.1	1.7	10.5	1.2	9.9	1.3	9.9	1.1	10.2	1.2	10	1.2
Vocational degree (0,1, 2)		D	.9	.4	.9	.6	.9	.6	.7	.5	.8	.6	.7	.6
Regional share of service sector		C	56	15	45	12	59	14	43	10	65	9.2	50	13
Regional share of production sector		C	30	11	33	12	29	10	33	12	80	8.9	32	11
Sectoral UE rate		C	12	4	12	4	12	4	12	4	12	4	12	4
Occupational UE rate		C	12	5	12	5	12	5	12	5	12	5	12	5
Observations			85843		16856		67327		2908		1409		2526	

Note: *I*: Binary indicator variable (0, 1); *D*: Discrete variable; *C*: Continuous variable. For indicator variables the share of ones is given in %. The number of observations in the full sample is 100,000. The table is based on the subsample of women who had no birth before the age of 21. Descriptive statistics for the variables and subsamples not included are available on request from the author.

The selection process along a sequence is modeled by a binary probit model. The actual probit models underlying the results presented in the next sections are slightly misspecified, but in

ways that remain largely undetected by conventional specification tests. The misspecification relates to the functional form as well as to the omission of some covariates that are highly correlated with covariates included in the sample. In this respect as well, the artificial data seem to exhibit similar problems and questions as real data sets usually do. The results for selected sequences are given in Table 3.

Table 3: Estimated coefficients of sequential probit models (first part of a selected sequence)

Variable	0 vs. 1	00 if 0	001 if 00	0011 if 001	00110 if 0011	001100 if 00110	0011000 if 00110
Schooling (8-12)	<i>-.33</i>	<i>-.31</i>	<i>.24</i>	<i>.26</i>	<i>-.44</i>	<i>-.43</i>	<i>-.44</i>
Vocational degree (0,1, 2)	<i>-.31</i>	<i>-.22</i>	<i>.12</i>	<i>.14</i>	<i>-.45</i>	<i>-.31</i>	<i>-.34</i>
Regional share of service sector	<i>.005</i>	<i>.009</i>	<i>-.01</i>	<i>-.01</i>	<i>.003</i>	<i>.01</i>	<i>.007</i>
Regional share of production sector	<i>-.001</i>	<i>-.001</i>	<i>-.000</i>	<i>0.002</i>	<i>-.003</i>	<i>-.004</i>	<i>-.004</i>
Sectoral UE rate	<i>-.002</i>	<i>-.000</i>	<i>-.001</i>	<i>.002</i>	<i>.005</i>	<i>.006</i>	<i>.01</i>
Regional UE rate	<i>.001</i>	<i>-.000</i>	<i>.002</i>	<i>.002</i>	<i>.005</i>	<i>.008</i>	<i>.01</i>
Employed at age 20	<i>.42</i>	0	0	0	0	0	0
Out-of labor force at age 20	<i>-.24</i>	0	0	0	0	0	0
Earnings at age 20	<i>.0007</i>	0	0	0	0	0	0
Employed at age 22	X	<i>.49</i>	0	0	0	0	0
Earnings at age 22	X	<i>.0008</i>	0	0	<i>-.000</i>	<i>-.0002</i>	<i>.000</i>
Employed at age 24	X	X	<i>-.50</i>	0	0	0	0
Earnings at age 24	X	X	<i>-.0007</i>	0	0	0	0
Out of labor force at age 26	X	X	X	<i>.62</i>	0	0	0
Earnings at age 26	X	X	X	<i>-.0007</i>	0	0	0
Out of labor force at age 28	X	X	X	X	<i>-.73</i>	0	0
Out of labor force at age 30	X	X	X	X	X	<i>-.73</i>	0
Out of labor force at age 32	X	X	X	X	X	X	<i>-1.12</i>
Subsample	1 and 0	0	00	001	0011	00110	001100
Number of observations in subsample	84183	67327	55131	6504	4787	4260	3933
Dependent variable	0	00	001	0011	00110	001100	0011000
Mean of dep. variable in subsample	0.8	.8	.14	.74	.89	.93	.86

Note: Binary probit model estimated on the respective subsample. All specifications include an intercept. If not stated otherwise, all information in the variables relates to age 20. *Exclusion restrictions*: 0: Variables omitted from specification. X: Variable not temporarily prior to the 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.

Following the specification over time, it becomes obvious that all probits include the same set of time constant variables, but differ with respect to time varying variables. According to the theory outlined so far, the number of time varying variables to be included in the probit should increase along the sequences, since each estimation step should include all variables previously included as well as those variables directly prior to the current treatment. Doing so would however lead to multicollinearity problems due to overparametrization. This problem

becomes more binding over time, as the number of women still in the sequence decreases. Since it is not the purpose of this estimation step to obtain consistent estimates of probit coefficients, but of the respective probabilities instead, deleting variables because of their almost perfect correlation with other variables already in the model does not harm consistent estimation of the treatment effects.

6 Estimation results

6.1 Quantity

Based on the considerations of Section 4, Table 4 presents some example on how to estimate the labor market effects of additional children. In the left part of the table two different ways of modeling the effect of one additional child are shown: The first comparison is for a birth at the age 21 / 22 and no birth for the next four years (sequence 100) compared to no births over six years from age 21 to 26 (000). Since it may be true that the first child has a different effect than the second one, a second specification is added which compares the one birth at 21/22 case to another sequence with an additional birth at 23/24 (110). Note that the way the effect is specified it is left open whether there will be or will not be more births after the age of 26. In this sense it is a minimal specification with respect to time periods that has been chosen because of sample size considerations (sequences with one child only are very rare in the simulated data). If there are enough observations in the individual sequences one may want to compare three children to two children as well as consider different timings of births. Furthermore, one may consider longer time periods without birth as is done in the final specifications contained in that table. The latter compare two or three early children with no children. The coverage of these sequences is now from age 21 to 40 and thus covers almost all of women's fertile period.

The particular target population and its size after imposing the common support criterion is shown in the line just below the sequences.¹⁶ The effects are estimated for different subpopulations that are defined by the first diverging element in each of the sequences that are compared. These effects are identified by the W-DCIA. A more stringent specification of those subpopulations (i.e. longer sequences for the target population) would require the S-DCIA to hold instead of the weaker W-DCIA, which is however not true with this simulated data, and usually not plausible in such applications as well.

Table 4: Labor market effects of birth sequences at age 45 (quantity)

Outcome	Employment	Earnings	Employment	Earnings	Employment	Earnings	Employment	Earnings
	<i>1 additional birth</i>				<i>2 births</i>		<i>3 births</i>	
Sequences $\underline{s}^1 - \underline{s}^0$	100 – 000		110-100		1100000000 – 0000000000		1110000000 – 0000000000	
Target population s (obs. after common support, N_s)	1 (15563)		11 (9634)		1 (14847)		1 (14484)	
Estimated outcome: $E(Y_t^{\underline{s}^1} S = s)$	28 (2)	377 (39)	16 (.4)	348 (10)	25 (2)	547 (51)	13 (2)	308 (51)
Estimated outcome: $E(Y_t^{\underline{s}^0} S = s)$	23 (3)	520 (89)	15 (2)	312 (43)	90 (15)	1490 (537)	68 (22)	1196 (837)
Effect: $\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	-5 (-3)	-142 (98)	1 (2)	35 (45)	<i>-64</i> (15)	<i>-943</i> (540)	<i>-56</i> (23)	<i>-889</i> (838)
$s(N_s)$	0 (64696)		10 (5068)		0 (60467)		0 (46092)	
$E(Y_t^{\underline{s}^1} S = s)$	57 (2)	1207 (54)	24 (1)	489 (22)	57 (4)	1144 (96)	24 (11)	508 (298)
$E(Y_t^{\underline{s}^0} S = s)$	64 (.4)	1354 (9)	27 (1)	565 (18)	90 (4)	1740 (106)	74 (8)	1436 (173)
$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	<i>-6</i> (2)	<i>-147</i> (55)	<i>-3</i> (1)	<i>-75</i> (29)	<i>-33</i> (5)	<i>-596</i> (143)	<i>-50</i> (13)	<i>-928</i> (345)

Note: Each element of a sequence covers two years. The first element of each sequence refers to the age group 21-22. All sequences are defined such that there are no births before age 21. Outcome variables are measured at age 45. Standard errors are in parentheses. **Bold and italics:** Effect is significant at 1% level. **Bold:** Effect is significant at 5% level. *Italics:* Effect significant at 10% level. N_s is the sample size of the target population after imposing common support.

The results of the estimations are given below the line for the target population. They show the estimated values for the (counterfactual) potential outcomes in the target population as

¹⁶ Members of the target population, for whom no comparable observations exist for any step in any of the two sequences, are removed. They violate W-DCIA. This procedure changes the estimand, but is common practice in static matching estimation.

well as their difference, which is the desired causal effect of the comparison of the two sequences. Standard errors are in brackets.

With respect to the results, all effects have the expected sign. For the larger target population of those women without a birth at age 21/22 all effects appear to be significant at conventional levels, whereas for the smaller subpopulation precision becomes a problem for the one-child comparisons, but not so for the comparisons involving more than one child. Comparing the effects across populations reveals some indication for possible effect heterogeneity.

6.2 Timing

Table 5 shows 8 different comparisons that all relate to timing issues, like when to have the birth, the time between first and second birth, and when to start fertility. The first 6 comparisons are based on rather long sequences (20 years if sample sizes permit) to investigate the various dimensions of early versus late kids. Therefore, the overall number of kids is kept constant over those 20 years. The remaining two comparisons investigate the effects of timing the first birth only ('starting early' vs. 'starting late'). Thus, the remaining birth history is not specified and considered as being part of the effect.

Table 5: Labor market effects of birth sequences at age 45 (timing)

Outcome	Employment	Earnings	Employment	Earnings	Employment	Earnings	Employment	Earnings
	<i>Early vs. late births</i>				<i>Time between births</i>			
Sequences $\underline{s}^1 - \underline{s}^0$	1100000000-001100000		1100000000-0000000110		0011000000-0000000110		200000-101000	
Target population s (obs. after common support, N_s)	1 (11589)		1 (4381)		001 (2879)		2 (1076)	
Estimated outcome: $E(Y_t^{\underline{s}^1} S = s)$	25 (1)	494 (25)	51 (2)	1064 (47)	57 (2)	1202 (51)	21 (4)	518 (93)
Estimated outcome: $E(Y_t^{\underline{s}^0} S = s)$	23 (2)	467 (44)	63 (6)	1258 (130)	64 (3)	1283 (61)	21 (5)	564 (147)
Effect: $\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	2 (2)	27 (51)	-12 (6)	-194 (138)	-7 (3)	-80 (79)	1 (6)	-46 (173)
$s(N_s)$	0 (56259)		0 (42337)		000 (39526)		1 (13914)	
$E(Y_t^{\underline{s}^1} S = s)$	56 (4)	1122 (98)	69 (5)	1386 (120)	80 (4)	1667 (107)	17 (2)	345 (45)
$E(Y_t^{\underline{s}^0} S = s)$	60 (3)	1243 (64)	81 (1)	1584 (36)	82 (1)	1630 (35)	19 (2)	408 (55)
$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	-4 (4)	-120 (117)	-11 (5)	-198 (128)	-2 (4)	36 (113)	-2 (3)	-63 (71)
	<i>Time between births</i>				<i>Early vs. late start</i>			
$\underline{s}^1 - \underline{s}^0$	1100000000-1001000000		1-01		1-0001		1-0000000001	
$s(N_s)$	11 (7092)		1 (16739)		1 (11954)		1 (3766)	
$E(Y_t^{\underline{s}^1} S = s)$	21 (1)	437 (21)	18 (.3)	411 (8)	23 (.4)	472 (8)	50 (1)	1033 (18)
$E(Y_t^{\underline{s}^0} S = s)$	24 (2)	480 (56)	19 (.7)	427 (20)	24 (6)	500 (145)	53 (5)	1058 (105)
$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	-3 (3)	-43 (60)	-1 (1)	-17 (21)	-1 (6)	-28 (145)	-2 (5)	-25 (106)
$s(N_s)$	10 (4457)		0 (64821)		0 (62376)		0 (46846)	
$E(Y_t^{\underline{s}^1} S = s)$	28 (2)	569 (31)	57 (1)	1180 (36)	59 (1)	1214 (37)	71 (2)	1462 (49)
$E(Y_t^{\underline{s}^0} S = s)$	32 (2)	635 (53)	60 (1)	1275 (27)	63 (2)	1329 (46)	77 (1)	1515 (39)
$\theta_t^{\underline{s}^0, \underline{s}^1}(s)$	-4 (3)	-66 (61)	-3 (2)	-95 (45)	-4 (2)	-114 (59)	-5 (2)	-52 (63)

See note below Table 3.

According to the data generating process, later births should have positive effects on labor market outcomes compared to earlier kids. Those effects should be somewhat smaller than the effects for different quantities of births. This result is generally confirmed by the data, although the precision of the estimates is not always enough to pin down the effects precisely.

7 Conclusions

This paper discusses to potential use of dynamic potential outcome models to analyze the effects of fertility on labor market interactions when large and informative data sets are available. The main advantages of such an approach are (i) its flexibility due to its non- or semiparametric nature, (ii) that it allows addressing the selection issues coming from the dynamic interaction between fertility and labor market decisions and realizations in a detailed way; and that (iii) it allows defining relevant parameters of interest in a more precise way than in any static approach. Based on artificial data, the approach is implemented and issues that come in any practical application of this approach are discussed.

References

- Abbring, J.H. (2003): "Dynamic Econometric Program Evaluation." IZA, Discussion Paper, 804.
- Adsera, A. (2005a): "Labor Market Performance and the Timing of Births", University of Chicago, Department of Economics.
- Adsera, A. (2005b): "Where Are the Babies Gone? Labor Market Conditions and Fertility in Europe", IZA Discussion paper 1576
- Angrist, J. D., and W. N. Evans (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size", *The American Economic Review*, 88, 450-477.
- Arroyo, C., and J. Zhang (1997): "Dynamic microeconomic models of fertility choice: A survey", *Journal of Population Economics*, 10, 23-65.
- Connelly, R. (1996): "Comments on the Fertility / Employment Interaction", *Population and Development Review*, 22, Supplement, 290-294.
- Dawid, A. P. (1979): "Conditional Independence in Statistical Theory." *Journal of the Royal Statistical Society B*, 41, 1-31.
- Del Boca, D., and R. Sauer (2005): "Life cycle employment and fertility across institutional environments", mimeo.
- Di Tommaso (1999): "A trivariate model of participation, fertility and wages: the Italian case", *Cambridge Journal of Economics*, 23, 623-640.
- Ding, W., and S.F. Lehrer (2003): "Estimating Dynamic Treatment Effects from Project STAR," mimeo.
- Francesconi, M. (2002): "A Joint Dynamic Model of Fertility and Work of Married Women", *Journal of Labor Economics*, 20, 336-380.
- Geweke, J. (1984): "Inference and Causality in Economic Time Series", in: Z. Griliches, M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2, Amsterdam: North-Holland, 1102-1144.

- Gill, R. D., and J. M. Robins, (2001): "Causal Inference for Complex Longitudinal Data: The Continuous Case." *The Annals of Statistics*, 2001, 1-27.
- Granger, C. W. J. (1969): "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", *Econometrica*, 37, 424-438.
- Granger, C. W. J. (1986): "Comment", *Journal of the American Statistical Association*, 81, 967-968.
- Heckman, J. J. (2000): "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective", *Quarterly Journal of Economics*, 115, 45-97.
- Heckman, J. J., and J. J. Walker (1990): "The Relationship between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data", *Econometrica*, 58, 1411-1441.
- Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in Ashenfelter, O., and D. Card (eds.), *Handbook of Labor Economics*, Vol. III A, North-Holland, Amsterdam, 1865--2097.
- Holland, P. W. (1986): "Statistics and Causal Inference", *Journal of the American Statistical Association*, 81, 945-970, with discussion.
- Hotz, J. V., and R. A. Miller (1988): "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply", *Econometrica*, 56, 91-118.
- Hotz, J. V., J. A. Klermann, and R. J. Willis (1997): "The Econometrics of Fertility in Developed Countries", in Rosenzweig and Stark (eds.), *Handbook of Population and Family Economics*, 276-347.
- 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.
- Klepinger, D., S. Lundberg, and R. Plotnick (1999): "How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women?", *The Journal of Human Resources*, 34, 421-448.
- Lechner, M. (1998): "Eine empirische Analyse der Geburtenentwicklung in den neuen Bundesländern aus der Sicht der neoklassischen Bevölkerungsökonomie," *Zeitschrift für Wirtschaft- und Sozialwissenschaften (ZWS)*, 118, 463-488.
- Lechner, M. (2001a): "The Empirical Analysis of East German Fertility after Unification: An Update", *European Journal of Population*, 17, 61-74, 2001.
- Lechner, M. (2001b): "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption," in Lechner, M., Pfeiffer, F. (eds.), *Econometric Evaluation of Active Labour Market Policies*, Physica, Heidelberg, 43-58.
- Lechner, M. (2004): "Sequential Matching Estimation of Dynamic Causal Models," University of St. Gallen, Discussion paper. Revised and shortened version forthcoming in the *Journal of Business & Economic Statistics*.
- Lechner, M. (2008a): "A Note on Endogenous Control Variables in Causal Studies," forthcoming in *Statistics and Probability Letters*.

- Lechner, M. (2008b): "Matching estimation of dynamic treatment models: Some practical issues", forthcoming in D. Millimet, J. Smith, and E. Vytlacil (eds.), *Advances in Econometrics, Volume 21, Modelling and Evaluating Treatment Effects in Econometrics*.
- 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.
- Lechner, M., and S. Wiehler (2007): "Does the Order and Timing of Active Labour Market Programmes Matter", Discussion paper 2007-38, Department of Economics, University of St. Gallen.
- Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9", translated in *Statistical Science* (with discussion), 1990, 5, 465-480.
- Miquel, R. (2002): "Identification of Dynamic Treatments Effects by Instrumental Variables," University of St. Gallen, Discussion paper, 2002-11.
- Miquel, R. (2003): "Identification of Effects of Dynamic Treatments with a Difference-in-Differences Approach." University of St. Gallen, Discussion paper, 2003-06.
- Moffitt, R. (1984): "Profiles of Fertility, Labour Supply and Wages of Married Women: A Complete Life-Cycle Model", *Review of Economic Studies*, 51, 263-278.
- Pötter, U., and H. Blossfeld (2001): "Causal Inference from Series of Events", *European Sociological Review*, 17 (1), 21-32.
- 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 Modeling*, 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," in Sechrest, L., Freeman, H., Mulley, A. (eds.), *Health Service Research Methodology: A Focus on Aids*, 113-159, Washington, D.C., Public Health Service, National Centre for Health Services Research.
- Robins, J.M. (1997): "Causal Inference from Complex Longitudinal Data. Latent Variable Modelling and Applications to Causality," in Berkane, M. (Ed.), *Lecture Notes in Statistics* (120), Springer, New York, 69-117.
- Robins, J. M. (1999): "Association, Causation, and Marginal Structural Models," *Synthese*, 121, 151-179.
- 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.

- Rosenbaum, P. R., and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, 41-50.
- Rosenbaum, P. R., and D. B. Rubin, (1984): "Reducing bias in observational studies using subclassification on the propensity score," *Journal of the American Statistical Association*, 79, 516-524.
- Rosenfeld, R. A. (1996): "Women Work Histories", *Population and Development Review*, 22, Supplement, 199-222.
- Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D. B. (2004): "On principles for modeling propensity scores in medical research," *Pharmacoepidemiology and Drug Safety*, 2004, 13, 855—857.
- Sims, Christopher A. (1972): "Money, Income, and Causality", *American Economic Review*, 540-552.
- Troske, K. R., and A. Voicu (2004): "Joint estimation of sequential labour force participation and fertility decisions using Markov chain Monte Carlo techniques", University of Missouri, Department of Economics.
- Wiener, N. (1956): "The Theory of Prediction", *Modern Mathematics for Engineers, Series 1* (edited by E. F. Beckenham), Chapter 8.
- Zhao, Z. (2004): "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metric and a Monte Carlo Study," *The Review of Economics and Statistics*, 86, 91-107.