



UPPSALA
UNIVERSITET

Working Paper 2018:1

Department of Statistics

Using High Frequency Pre-treatment Outcomes To Identify Causal Effects In Non-experimental Data

Mårten Schultzberg





Working Paper 2018:1
June 2018
Department of Statistics
Uppsala University
Box 513
SE-751 20 UPPSALA
SWEDEN

Working papers can be downloaded from www.statistics.uu.se

Title: Using high frequency pre-treatment outcomes to identify causal effects in non-experimental data

Author: Mårten Schultzberg

E-mail: marten.schultzberg@statistik.uu.se



Using high frequency pre-treatment outcomes to identify
causal effects in non-experimental data

Mårten Schultzberg¹

June 27, 2018

¹I would like to thank Per Johansson for valuable discussions and suggestions to this paper, and Bengt Muthén for feedback on earlier drafts.

Abstract

In observational studies it is common to use matching strategies to consistently estimate the average treatment effect of the treated (ATET) under the unconfoundedness assumption of the outcome and the treatment assignment mechanism. Matching is often based on a set of time invariant covariates together with one or a few pre-treatment measurements of the outcome. This paper proposes estimation strategies using a large number of pre-treatment measurements of the outcome to consistently estimate the average treatment effect of the treated (ATET). The assumptions under which these approaches are valid are given. It is shown when and how the strategies can be used to replace, or add to, time-invariant covariates to identify and consistently estimate the ATET. The theoretical results and estimation strategies are illustrated by a study of electricity consumption.

1 Introduction

This paper proposes to use pre-treatment time series measurements of the outcome to identify casual effects in observational studies. The motivation for this paper comes from the study of electricity-consumption behaviour of firms, where large amounts of repeated measurements of the consumption are available for all firms, both before and after treatment. The treatment in this case is change of tariff where a flat fee is replaced with a peak-based dynamic tariff. The treatment effect of interest is the change in the average electricity consumption compared to keeping the flat fee. Because treatment is not randomly assigned, causal effects can be identified from this type of data only if confounders can be controlled for. Very few covariates, however, are typically observed besides the pre-treatment outcome. On the other hand, the pre-treatment outcomes are arguably the most important covariates to control for which indicates that rich pre-treatment outcome time series data should be highly useful. By using novel multilevel time series models, the heterogeneity among a large number of long time series can easily be characterized by a small set of random effects. This paper investigates how this information can be used to help identify causal effects by controlling for pre-treatment differences between firms derived from pre-treatment time series of the outcomes. The proposed methods can be used to help identify causal effects in many electricity studies where this type of data is commonly available. However, the methodology is generally applicable to all types of intensive longitudinal data (ILD), including data obtained by recent sampling strategies such as experience sampling methods (ESM), ecological momentary assessment (EMA), ambulatory assessment, to name a few see; e.g., Trull and Ebner-Priemer (2009) for an overview. ILD are likely to become more prevalent since the technological development of personal electronic devices like smart phones, smart watches, fitness trackers, and the Internet of Things, have substantially simplified the collection of high frequency repeated measurements data.

The development of the formal assumptions and tools for drawing causal inference from observational studies started with the influential papers by Rubin (1973, 1974). Since then, matching has become part of the standard tool kit for observational studies and is still an area with intense ongoing research, see; e.g., Stuart (2010) for an overview. The importance and utility of repeated measurements for drawing casual inference in observational studies have been discussed extensively in the causal literature; for recent papers on this topic see e.g. Chabé-Ferret (2015) and O’Neill et al. (2016). However, the main focus of the matching literature has been on how to efficiently

utilize, and/or select among a large set of time-invariant covariates and/or time-varying covariates observed right before treatment. As a consequence, the discussion about if and how pre-treatment measurements should be included in matching is usually based on the presumption that only one, or possibly a few, pre-treatment measurements are available. With one or a few pre-treatment measurements it might be difficult to separate, e.g., temporal fluctuations due to external shocks from central tendencies stable over time. In such situations, one has to be careful about using the pre-treatment measurements of the outcome in matching due to phenomena like the Ashenfelter’s dip (Ashenfelter, 1978), where observational units with similar values of the outcome going into the treatment may still be qualitatively different. The case when the available set of covariates is small but the number of pre-treatment measurements of the outcome is large, is seldom discussed. This is likely due to the fact that such data have been rare and difficult to obtain until recently. However, if the observed subset of covariates is small and is not likely to contain all confounders, the classical unconfoundedness assumption becomes unrealistic. If, on the other hand, large numbers of pre-treatment measurements of the outcome are available such that central tendencies can be separated from temporary fluctuations, then such pre-treatment measurements are highly useful for identifying casual effects.

One available framework utilizing pre-treatment measurements of the outcome is the synthetic control framework (Abadie et al., 2003, 2010). This literature is mainly focused on the case where only group-level aggregated data are available, i.e. in what could be viewed as a classical Difference-in-Difference setting. In Abadie et al. (2010), the pre-treatment measurements of the outcome are used to construct the control group and evaluate the success of this construction. This is similar to the strategies proposed in the present paper in that the pre-treatment data are extensively utilized. However, the present paper focuses on the case where observational unit-specific time series are observed and a non-synthetic control group can be found from the pools of controls, such that there is balance between the chosen control and treatment groups in the outcome under no treatment. More importantly, the strategies presented in this paper can be used to non-parametrically estimate the ATET and are not restricted to any particular inference. Based on the idea of synthetic controls, other strategies have been developed in papers like Brodersen et al. (2015) where the counterfactual of a synthetic control group is predicted by a Bayesian times series model, chosen by a data driven model selection tool. This is similar to the present paper in the sense that parametric time series models are fitted on pre-treatment measurements of the outcome. An important difference is that

the effect estimation in our strategy is independent of this time series model, whereas in Brodersen et al. (2015) it is used to predict a counterfactual time series. In the present papers, the time series models are used only as tools to non-parametrically identify the ATET.

There is a rich pool of time series clustering and matching strategies available see; e.g., Fu (2011) and Aghabozorgi et al. (2015) for overviews. The focus of this literature is mainly on data mining, and finding similar sequences for predictive purposes. However, to the best of our knowledge, this literature has not been utilizing time-series matching to identify causal effects. The theoretical results derived in this paper can essentially be combined with any time series matching strategy. The model based time-series characterization strategy proposed in this paper, using two-level time series models, is chosen for several reasons; it is an intuitive, easy to use, and, for most purposes, a sufficiently sophisticated method to achieve closeness in the outcome under no treatment between the treatment and control group. In addition, the model-based time series strategy enables the evaluation of the important overlap assumption, discussed in detail in the following section.

In summary, this paper specifies the assumptions under which high frequency pre-treatment measurements of the outcome can be utilized, instead of or in addition to observed covariates, to identify the Average Treatment Effect of the Treated (ATET). It is shown that when the pre-treatment data are rich enough and these assumptions hold, the heterogeneity across observational units in unobserved covariates affecting the outcome under no treatment can be derived from the pre-treatment data. The remainder of this paper is structured as follows. Section 2 presents the notation and gives the theoretical results. Section 3 presents estimation strategies utilizing the theoretical results. Section 4 gives an empirical example illustrating the theoretical results and estimation strategies. Section 5 discusses the results and gives concluding remarks.

2 Setup

Consider an observational study setting where observational units are measured repeatedly for T consecutive time periods. Time periods $t = 1, 2, \dots, t_1 - 1$ and $t = t_1, t_1 + 1, \dots, T$ refer to time periods pre and post assignment of treatment, respectively. Let D_t be an indicator function, taking the value 0 if not treated and 1 if treated, at time period t . Once treated, $D_t = 1$ throughout the study. Furthermore, let $Y_t(d)$ be the potential outcome (Neyman, 1923; Splawa-Neyman et al., 1990; Rubin, 1973) at time period t when d is set to 1 if the observational unit is treated at time t and 0 if not treated at time t . The focus of this paper is to identify the ATET which for time period t is given by

$$\text{ATET}_t = E[Y_t(1) - Y_t(0) | D_t = 1] \quad (1)$$

There are three assumptions commonly used to identify the ATET from observational data. Throughout this paper, if nothing else is stated, it is assumed that two of these, Stable Unit Treatment Variable Assumption (SUTVA) (Rubin, 1980) and the assumption of overlap are fulfilled. The third assumption, which is the focus of this paper, is the unconfoundedness assumption (UA). The UA for identifying the ATET for time period t_1 is given by Assumption 1.

Assumption 1

$$Y_{it_1}(0) \perp\!\!\!\perp D_{it_1-1} | \mathbf{Z}_i^*$$

for all observational units in a population $i=1, \dots, N$, where \mathbf{Z}_i^* is a set of time-invariant covariates and time-varying covariates observed at specific time periods.

The specific time periods of the time-varying¹ covariates in \mathbf{Z}_i^* are specified with subscripts, explained in detail below. Assumption 1 implies that \mathbf{Z}_i^* is the set of all covariates confounding the outcome under no treatment and the assignment process, and therefore, \mathbf{Z}_i^* needs to be conditioned on in order to identify the ATET. \mathbf{Z}_i^* usually contains an unknown and/or partly unobserved set of covariates which makes identifying causal effects challenging in observational studies. If \mathbf{Z}_i^* is observed, the ATET can be consistently estimated under the UA by controlling for \mathbf{Z}_i^* . Below it is investigated how, and under what additional assumption, the ATET can be consistently estimated

¹The term 'time-invariant' is used throughout this paper to mean constant during the study period, i.e. for $t=1, \dots, T$.

under the UA by replacing some, or all, time-invariant covariates in \mathbf{Z}_i^* with information from the pre-treatment measurements of the outcome.

Let \mathbf{W} be the set of all possible time-invariant covariates and \mathbf{X} be the set of all possible time-varying covariates.

Assumption 2 *Let $i = 1, \dots, n$ be the number of observational units in a sample, and $\mathbf{W}_i^{(1)}$ and $\mathbf{X}_i^{(1)}$ be subsets of \mathbf{W} and \mathbf{X} respectively. Let the outcome at time period t_1 under no treatment for observational unit i be given by*

$$Y_{it_1}(0) = f\left(\mathbf{Y}_{it_1, k_Y}(0), \mathbf{X}_{it_1, p_Y}^{(1)}, \mathbf{W}_i^{(1)}, \epsilon_{it}\right),$$

such that ϵ_{it} is independent of $\{\mathbf{Y}_{it_1, k_Y}(0), \mathbf{X}_{it_1, p_Y}^{(1)}, \mathbf{W}_i^{(1)}\}$, and D_{it_1-1} .

Assumption 2 says that the outcome of observational unit i under no treatment at time period t_1 is a function of k_Y previous outcome measurements, a set of time-varying covariates containing p_Y lags, and a set of time-invariant covariates, i.e.

$$\begin{aligned} \mathbf{Y}_{it, k_Y}(0) &= (Y_{it-1}(0), Y_{it-2}(0), \dots, Y_{it-k_Y}(0)), \quad k_Y \in (1, \dots, t-1) \\ \mathbf{X}_{it, p_Y} &= (X_{it}, X_{it-1}, \dots, X_{it-p_Y}), \quad p_Y \in (0, 1, \dots, t). \end{aligned} \tag{2}$$

The obvious implication of Assumption 2 is that it is sufficient to condition on $\mathbf{Y}_{it_1, k_Y}(0), \mathbf{X}_{it_1, p_Y}^{(1)}, \mathbf{W}_i^{(1)}$ to fulfill the UA assumption.

To specify the minimal sufficient (that is, necessary) set of covariates and measurements, additional assumptions on the assignment mechanism are required². To illustrate, assume that the treatment assignment mechanism at the time of the treatment assignment is given by, e.g.,

$$D_{it_1-1} = g\left(\mathbf{Y}_{it_1-1, k_D}(0), \mathbf{X}_{it_1-1, p_D}^{(2)}, \mathbf{W}_i^{(2)}, \xi_{i, t_1-1}\right),$$

where $\mathbf{W}_i^{(2)}$ and $\mathbf{X}_i^{(2)}$ are subsets of \mathbf{W} and \mathbf{X} , and ξ_{i, t_1-1} is independently and identically distributed for all t for each observational unit i , and independent of $\mathbf{Y}_{it_1-1, k_D}(0), \mathbf{X}_{it_1-1, p_D}^{(2)}, \mathbf{W}_i^{(2)}$. This means that the minimal sufficient set, \mathbf{Z}_i^* in the UA, is given directly by

$$Y_{it_1}(0) \perp\!\!\!\perp D_{it_1-1} \mid \left\{ \mathbf{W}_i^{(1)} \cap \mathbf{W}_i^{(2)}, \mathbf{Y}_{it_1, k^*}(0), \mathbf{X}_{it_1, p^*}^{(1)} \cap \mathbf{X}_{it_1, p^*}^{(2)} \right\} \Rightarrow \epsilon_{it} \perp\!\!\!\perp \xi_{i, t_1-1}, \tag{3}$$

²The minimal sufficient set is not the focus of this paper, however, it is included here as it might help to provide a better understanding of the implications of the unconfoundedness assumption.

where k^* and p^* include lags of $Y(0)$ and $\mathbf{X}^{(1)} \cap \mathbf{X}^{(2)}$, respectively, affecting both $Y_{it_1}(0)$ and $D_{it_1-1}(0)$, see the example below for details. Equation 3 means that it is *necessary* to condition on the variables, and measurements thereof, that are common to the outcome at time period t_1 and the treatment assignment at $t_1 - 1$. The outcome and assignment mechanism do not have to be functions of previous measurements of the outcome and/or time-varying covariates, but can instead be determined solely by, e.g., $\mathbf{W}_i^{(1)}$ and $\mathbf{W}_i^{(2)}$, respectively.

Assumption 3.a and 3.b presents important extensions of Assumption 2 that enables the estimation strategy proposed in Section 3. Assumption 3.b is discussed in Section 2.2. Assumption 3.a is motivated by the reasoning behind classical two-level (i.e., random effects) modelling, where, the variation in model parameters (random coefficients) across units are assumed to be some functions of unit-specific covariates. Assumption 3.a, is needed for later when this reasoning is reversed; if the covariates explains the variation in the random coefficients, then the random coefficients should reflect the variation in the covariates; at least the part of the variation in the covariates that affects the coefficients. Or, even more specifically, the variation in the random coefficients should reflect at least the variation in the functions of the covariates that affects the coefficients.

Assumption 3.a *In addition to Assumption 2, the function f can be parameterized by the vector $\boldsymbol{\theta}_i = h(\mathbf{W}_i^{(1)})$, where h is a vector of functions all allowed to be non-injective.*

On a more technical note, Assumption 3.a adds that the outcome at time period t can be parametrized by the vector $\boldsymbol{\theta}_i = h(\mathbf{W}_i^{(1)})$, where h is a vector of functions, all allowed to be non-injective, i.e. $h^{-1}(\boldsymbol{\theta}_i) \neq \mathbf{W}_i^{(1)}$. The non-injectiveness is explicitly allowed to emphasize that it is not important for our purposes to be able to find $\mathbf{W}_i^{(1)}$ from $\boldsymbol{\theta}_i$. In other words, functions in h are allowed to reduce the information in $\mathbf{W}_i^{(1)}$ arbitrarily. That $\boldsymbol{\theta}_i$ is only a function of time-invariant covariates follows from that it is assumed to be constant over time. The parameter vector $\boldsymbol{\theta}_i$ is governing the relation between the outcome of observational unit i at time t and previous measurements of the outcome and time-varying covariates. If the function h in fact is non-injective $\boldsymbol{\theta}_i$ can be viewed as a dimension reduction of the information in \mathbf{W} that governs the outcome process. This further implies that the function f can be rewritten in terms of $\boldsymbol{\theta}$, i.e. $Y_{it}(0) = f\left(\mathbf{Y}_{it,k_Y}(0), \mathbf{X}_{it,p_Y}^{(1)}, \boldsymbol{\theta}_i, \epsilon_{it}\right)$, and it follows directly that under Assumption 3.a

$$Y_{it_1}(0) \perp\!\!\!\perp D_{it_1-1} \mid \left\{ \boldsymbol{\theta}_i, \mathbf{Y}_{it_1,k_Y}(0), \mathbf{X}_{it_1,p_Y}^{(1)} \right\} \Leftrightarrow \epsilon_{it} \perp\!\!\!\perp D_{it_1-1} \quad (4)$$

In summary, under Assumption 3.a, it is sufficient to condition on (i) the parameters describing

the relation between the outcome and its previous values and the time-varying covariates in $\mathbf{X}^{(1)}$, and (ii) the pre-treatment measurements of $\mathbf{Y}(0)$ and $\mathbf{X}^{(1)}$ affecting the first outcome measurement after treatment assignment.

2.1 Illustration of Assumption 3.a and its implications

The purpose of Assumption 3.a is to ensure that the information in the set of time-invariant covariates affecting the outcome, $\mathbf{W}_i^{(1)}$ can be substituted by information that potentially can be extracted from the outcome. To better understand the implications of Assumption 3.a with respect to the UA, consider the following example. Let $Y_{it}(0)$ and $\mathbf{X}_i^{(1)}$ be measured for $i = 1, \dots, n$ observational unit for all T time periods. Assume that the outcome under no treatment is an autoregressive process

$$Y_{it}(0) = \theta_i^{(1)} + \theta_i^{(2)}Y_{it-1} + \theta_i^{(3)}Y_{it-2} + \theta_i^{(4)}X_{it-2}^{(1)} + \epsilon_{it}, \quad (5)$$

where ϵ_{it} are independent and identically distributed with means zero and variances $\theta_i^{(5)}$, and the parameters are given by

$$\begin{aligned} \theta_i^{(1)} &= \beta_{11} + \beta_{12}w_{i1}w_{i2} \\ \theta_i^{(2)} &= \beta_{21} + \beta_{22}w_{i2}^2 \\ \theta_i^{(3)} &= c \\ \theta_i^{(4)} &= \beta_{42}|w_{i3}| \\ \log(\theta_i^{(5)}) &= \beta_{51} + \beta_{52}w_{i1}^2 \end{aligned}$$

where w_{ij} are elements of \mathbf{W}_i , and $\theta_i^{(3)} = c$ implies that the relation between Y_t and Y_{t-2} is independent of \mathbf{W} and does not vary across observational units. The purposely complex relations between $\boldsymbol{\theta}$ and \mathbf{W} are chosen to illustrate that this strategy is robust against the functional forms of h in Assumption 3.a. Further assume that the treatment assignment process is given by the process

$$D_{t_1-1} = l(\gamma_i^{(1)} + \gamma_i^{(2)}X_{it_1-2}^{(1)} + \xi_{it_1-1})$$

where ξ_{it_1-1} are independently and identically distributed with mean zero and variance $\gamma_i^{(3)}$, l is some suitable link function, and the parameters are given by

$$\begin{aligned}\gamma_i^{(1)} &= \zeta_{11} + \zeta_{12}w_{i2}^2 + \zeta_{13}w_{i4}^2 \\ \gamma_i^{(2)} &= \zeta_{21} + \zeta_{22}e^{w_{i4}} \\ \log(\gamma_i^{(3)}) &= \zeta_{32}w_{i1}\end{aligned}$$

Note that some functions of w 's, e.g. w_{i2}^2 , are non-injective as discussed above. That this is unproblematic follows directly from Equation 4, in which elements of $\mathbf{W}_i^{(1)}$ play no part. In other words, θ_i can replace $\mathbf{W}_i^{(1)}$ regardless of the information about $\mathbf{W}_i^{(1)}$ it contains. For this example, $\mathbf{W}_i^{(1)} = \{w_1, w_2, w_3\}$, $\mathbf{X}_{i,t_1,p_Y}^{(1)} = \mathbf{X}_{i,t_1,2}^{(1)} = \{X_{i,t_1-1}^{(1)}, X_{i,t_1-2}^{(1)}\}$, $\theta_i = \{\theta_i^{(1)}, \theta_i^{(2)}, \theta_i^{(3)}, \theta_i^{(4)}, \theta_i^{(5)}\}$, and $\mathbf{Y}_{i,t_1,k_Y}(0) = \mathbf{Y}_{i,t_1,2}(0) = \{Y_{i,t_1-1}(0), Y_{i,t_1-2}(0)\}$. Moreover, $\mathbf{W}_i^{(2)} = \{w_{i1}, w_{i2}, w_{i4}\}$, $\mathbf{X}_{i,t_1-1,p_D}^{(2)} = \mathbf{X}_{i,t_1-1,2}^{(2)} = \{X_{i,t_1-2}^{(1)}\}$, $\gamma_i = \{\gamma_i^{(1)}, \gamma_i^{(2)}, \gamma_i^{(3)}\}$, and $\mathbf{Y}_{i,t_1,k_D}(0) = \emptyset$. Hence, $\mathbf{W}_i^{(1)} \cap \mathbf{W}_i^{(2)} = \{w_1, w_2\}$, $\mathbf{Y}_{i,t_1,k^*}(0) = \emptyset$, and $\mathbf{X}_{i,t_1,p^*}^{(1)} \cap \mathbf{X}_{i,t_1,p^*}^{(2)} = \mathbf{X}_{i,t_1,2}^{(1)} \cap \mathbf{X}_{i,t_1,2}^{(2)} = \{X_{i,t_1-2}^{(1)}\}$. Thus, $\mathbf{Z}^* = \{w_{i1}, w_{i2}, X_{i,t_1-2}^{(1)}\}$. By the results in the previous section, this implies that all of the following statements holds:

$$\{Y_{it}(0)\}_{t_1}^T \perp\!\!\!\perp D_{it_1-1} | w_{i1}, w_{i2}, X_{it_1-2}^{(1)} \quad (6)$$

$$\{Y_{it}(0)\}_{t_1}^T \perp\!\!\!\perp D_{it_1-1} | \mathbf{W}_i^{(1)}, X_{it_1-2}^{(1)} \quad (7)$$

$$\{Y_{it}(0)\}_{t_1}^T \perp\!\!\!\perp D_{it_1-1} | \theta_i, X_{it_1-2}^{(1)} \quad (8)$$

$$\{Y_{it}(0)\}_{t_1}^T \perp\!\!\!\perp D_{it_1-1} | \theta_i, \mathbf{X}_{i,t_1,2}^{(1)}, \mathbf{Y}_{i,t_1,2}(0) \quad (9)$$

The proof of any of the statements follows directly from substituting $Y(0)$ and D with corresponding functions. For example, Equation 9 at time period t_1 equals

$$\theta_i^{(1)} + \theta_i^{(2)}Y_{it_1-1} + \theta_i^{(3)}Y_{it_1-2} + \theta_i^{(4)}X_{it_1-2}^{(1)} + \epsilon_{it_1} \perp\!\!\!\perp D_{it_1-1} | \theta_i, X_{it_1-2}^{(1)}, Y_{it_1-1}(0), Y_{it_1-2}(0)$$

from which conditional independence follows immediately. Clearly, all but the first statement conditions on more information than necessary, however, particularly the last statement (Equation 9) is useful for estimation as will be shown. The following sections presents a strategy for estimating θ_i from pre-treatment data.

2.2 Assumption relating the identification and the estimation

In this section an updated version of Assumption 3.a is presented where all observed time periods of the outcome are included. This version of Assumption 3.a allows the pre-treatment data to be

used in the estimation.

Assumption 3.b Let $i = 1, \dots, n$ be the number of observational units in a sample, and $\mathbf{W}_i^{(1)}$ and $\mathbf{X}_i^{(1)}$ be subsets of \mathbf{W} and \mathbf{X} respectively. Let the outcome process at time period t under no treatment for observational unit i be given by

$$Y_{it}(0) = f\left(\mathbf{Y}_{it,k_Y}(0), \mathbf{X}_{it,p_Y}^{(1)}, \mathbf{W}_i^{(1)}, \epsilon_{it}\right) \forall t = 1, \dots, T,$$

where ϵ_{it} is independent of $\{\mathbf{Y}_{it,k_Y}(0), \mathbf{X}_{it,p_Y}^{(1)}, \mathbf{W}_i^{(1)}\}$ and D_{it_1-1} for all $t = t_1, \dots, T$. The function f can be parametrized by $\boldsymbol{\theta}_i$, i.e.

$$Y_{it}(0) = f\left(\mathbf{Y}_{it,k_Y}(0), \mathbf{X}_{it,p_Y}^{(1)}, \mathbf{W}_i^{(1)}, \epsilon_{it}\right) = f\left(\mathbf{Y}_{it,k_Y}(0), \mathbf{X}_{it,p_Y}^{(1)}, \boldsymbol{\theta}_i, \epsilon_{it}\right) \forall t = 1, \dots, T,$$

where $\boldsymbol{\theta}_i = h(\mathbf{W}_i^{(1)})$ and h is a vector of functions all allowed to non-injective, such that

$$Y_{it}(0) \perp\!\!\!\perp D_{it_1-1} | \mathbf{Y}_{it,k_Y}(0), \mathbf{X}_{it,p_Y}^{(1)}, \boldsymbol{\theta}_i \quad \forall t = 1, \dots, T,$$

Assumption 3.b is the key assumption of this paper connecting the identification with the estimation. It contains two parts: First, the outcomes are assumed to follow same processes over time that can be parametrized by finite parameter vectors; secondly, the processes are not changing during the window of observation unless any treatment is applied. This means that if the pre-treatment period is long enough, the differences and similarities across observational units in the parameter vectors of the pre-treatment processes reflect the heterogeneity in the outcome during the period when the treatment effect(s) is estimated. That is, under Assumption 3.b, the ATET is identified for all post-treatment time periods $t = t_1, \dots, T$.

An important implication of Assumption 3.b is that an observational unit has the same outcome processes under no treatment during the full observed period, both before and after the treatment assignment. This implies that if closeness between the groups is successfully obtained conditional on $\boldsymbol{\theta}$, $\mathbf{X}_{t_1 k_1}^{(1)}$, and $\mathbf{Y}_{it_1 k_1}(0)$, balance in $Y_t(0)$ is obtained for all pre-treatment time periods, i.e. $E(Y_t(0)|D=1) = E(Y_t(0)|D=0) \forall t = 1, \dots, t_1 - 1$. Thus, if $Y(0)$ and $\mathbf{X}^{(1)}$ are measured repeatedly for a long time before treatment assignment ($t_1 - 1$ is large), $\boldsymbol{\theta}$ can be estimated from the pre-treatment measurements. The model specification and estimation of $\boldsymbol{\theta}$ can be performed in several ways iteratively, in a machine-learning fashion, until balance in pre-treatment $Y(0)$ is achieved (if

possible). After estimation, $\hat{\boldsymbol{\theta}}$, $\mathbf{Y}_{it_1k_1}(0)$, and $\mathbf{X}_{t_1k_1}^{(1)}$ can be controlled for to consistently estimate the ATET. The full pre-treatment vectors of $\mathbf{Y}(0)$ and $\mathbf{X}^{(1)}$ are first used to estimate the relation between the outcome and the covariates, after which the last measurements before treatment, $\mathbf{X}_{t_1k_1}^{(1)}$ and $\mathbf{Y}_{it_1k_1}(0)$, are controlled for together with the estimate $\hat{\boldsymbol{\theta}}_i$ in the estimation of the ATET. Assumption 3.b is implicitly made in cross sectional matching cases where $Y_{t_1-1}(0)$ is used to check balance. Put differently, the point of using $Y_{t_1-1}(0)$ to check balance builds on the idea that the outcome at two consecutive measurements are similar. This is exactly what Assumption 3.b makes explicit.

The ATET can be consistently estimated by controlling for *only* the pre-treatment measurements of the outcome if Assumption 3.b holds and there are no time-varying covariates affecting the outcome under no treatment. This follows since in that case $\boldsymbol{\theta}$ can be estimated from pre-treatment measurements of the outcome alone. The following section suggests a strategy for estimating $\boldsymbol{\theta}$ encompassing a large set of possible outcome processes f .

3 Estimation strategies

Several considerations have to be made in the estimation of $\boldsymbol{\theta}$. The most obvious challenge is to condense the information in the pre-treatment measurements. For example, the naive estimator $\hat{\boldsymbol{\theta}}_i = (\{y_{it}\}_1^{t_1-1}, \{\mathbf{x}_{it}\}_1^{t_1-1})$, would obviously bring the curse of dimensionality into the ATET estimation, and can be disregarded directly with ILD. We suggest approximating f with a parametric time series model for each observational unit. Once a time series is fitted to an observational units' pre-treatment data, the usual time series tools to evaluate fit can be used to guide model selection for the next iteration. Even if the fit is poor for some observational units, the estimates might give a sufficiently detailed description of the heterogeneity in processes to achieve balance in the matching. This is important as it means that the model does not have to describe the true underlying outcome processes of each observational unit under no treatment, but only describe the pre-treatment information in some sufficiently good way to fulfill Equation 4. A parametric time series model is able to handle also the second challenge of the estimation, namely the inclusion of time-varying covariates. The time series model can be specified to include time-varying covariates observed at the observational unit level. However, perhaps more interestingly, variables that do not vary across observational units but only over time can also be included, e.g. temperature common

to observational units located in the same city. The relation between the outcome of interest and any type of time-varying covariate can of course be estimated. However, in addition, a time series model can be made multivariate, where the time series process of observational unit-specific time-varying covariates can be estimated including cross-lags between the covariates and the outcome of interest. That is, time-varying covariates can also be viewed as dependent variables by estimating the multivariate time series model of both the covariates and the outcome of interest. This means that any covariates believed to affect the outcome process can, and should, be included, to improve the estimate of θ . In general if $\mathbf{X}^{(1)}$ is not an empty set, it might not be possible to estimate θ sufficiently well without including $\mathbf{X}^{(1)}$, parts of it, or at least some proxy variables.

3.1 Multivariate two-level time series approach

In this section an estimation strategy to consistently estimate the ATET under Assumption 3.b is proposed. We suggest starting by estimating a general multivariate two-level time series model.

In the application below the Dynamic Structural Equation Modelling (DSEM) framework (Asparouhov et al., 2017) available in Mplus version 8 (Muthén and Muthén, 2017) has been utilized. The DSEM framework is a general multivariate two-level time series modelling framework with time on level 1 and observational units on level 2. For the purposes of this paper, some especially practical DSEM features are the ability to include time-varying covariates with measurement frequencies different from that of the outcome, being able to fit random effects of time-varying covariates that do not vary across observational units, and being able to allow for observational unit-specific residual variances. DSEM can essentially encompass observational units with various different orders of VARMA models in one estimation since all restricted models nested in the general model are encompassed. In Appendix A a simple example illustrates how a slightly over fitted two-level time series model can distinguish between observational units with different orders of AR outcome processes.

One potential problem with the two-level time series approach is that the parametric assumptions of the distributions of the random coefficients pre-supposes similarity that is imposed on the estimates. Although this might be true, at least for short time series, this will in such cases show up as lack of balance in the pre-treatment outcome. That is, if the time series are so short that the distributional assumption of the random coefficients forces the estimates to be too similar, to a degree where the heterogeneity is underestimated, it will be apparent. The two-level time series

models are efficiently estimated using Bayesian MCMC with non-informative prior distributions.

3.2 Matching

We propose using a matching estimator, i.e. match on θ to consistently estimate the ATET under Assumption 3.b. After fitting the two-level time series model, matching can be based on θ and any observed covariates, say $\mathbf{W}^{(obs)}$, using for example propensity score (PS) matching (Rosenbaum and Rubin, 1983), calliper matching (Rosenbaum and Rubin, 1985) as illustrated in the empirical example below, or recent novel methods such as suggested in Zubizarreta (2012). Propensity score based matching strategies gives the benefit of being able to evaluate overlap. If PS based methods are used, the PS can be directly estimated in the level-2 model, as a function of the random coefficient and the observed covariates $\mathbf{W}^{(obs)}$. Using matching, the balance can be easily checked by simply comparing the pre-treatment outcome of the treated group and matched control group. If the matching achieves balance in the parameter estimates but not the pre-treatment outcomes, the conclusion is that the model was not able to estimate θ well enough. Appendix B gives a simple illustration of the potential gain of including parameters describing the dynamic aspects of the outcome processes in the matching.

4 Empirical example with electricity consumption data

In the motivation example for this paper, taken from Öhrlund et al. (2018), the focus is on the effects of a dynamic tariff in contrast to a flat-fee price tariff on electricity grid fee for firms. The aim of a dynamic price tariff instead of a flat-fee is to lower the peaks in the grid. High peaks are associated with high costs for the company supplying the grid. This specific dynamic tariff has costs proportional to the customers' highest peak of consumption during each month, whereas the flat fee tariff is based on the total kWh usage each month. By reducing each firm's highest peaks, the grid-supplying company aims to lower the cumulative peaks in the grid.

In this example, we make use of data from a specific company supplying electricity grid to firms in the cities of Sandviken and Sundsvall. All the firms had the flat fee tariff in 2014. The dynamic tariff was introduced to all firms supplied in Sandviken in 2015 but a flat-fee remained in Sundsvall. There are 212 firms in Sandviken and 1055 in Sundsvall, which means that 16.7% of the sample is in the treated group. The daily electricity consumptions is observed for all firms from 2014 up to

and including 2016, three years in total. All the details of this study can be found in Öhrlund et al. (2018).

The population size of Sundvall is around two times as large as Sandviken, however, the two cities have similar industry structure and the distance between them is only 200 km. Figure 1 displays the average electricity consumption of the supplied firms in the two cities the year before treatment. From Figure 1 it is clear that the seasonality pattern of electricity consumption is similar

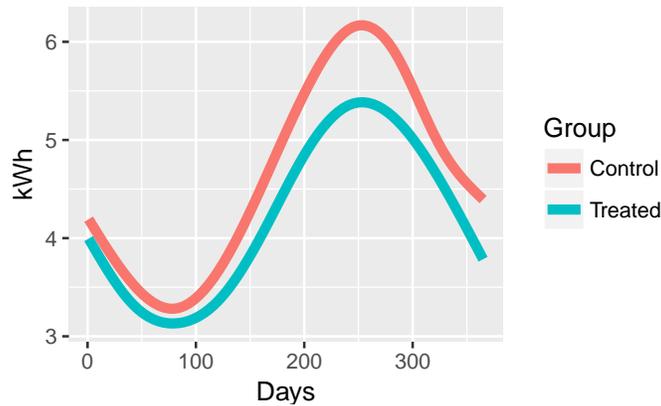


Figure 1: Smoothed pre-treatment electricity consumption average for the cities of Sandviken (treated) and Sundvall (control) the year before treatment.

for the two cities but also indicates that the firms in Sundsvall consume more electricity on average. This means that a naive comparison of the electricity consumption during 2015 and 2016 would most likely provide a biased estimate of the effects from the dynamic tariff.

4.1 Identification and Estimation

The parameter of interest is the ATET. The SUTVA should hold as there is only one form of the treatment, and, given that firms are cost minimizing it is not likely that the firms' electricity consumption are affected by other firms' treatment status. With regards to overlap, this evaluation is a special case as the assignment is at the city level with no overlap in location. This means that we need to assume that a firm located in Sandviken could have been located in Sundsvall, given all possible values of all potential confounders.

Assumption 3.b is somewhat simplified in this case as the firms cannot choose to be treated,

which implies that the consumption and other time-varying covariate-measurements cannot affect the treatment assignment, i.e. $\mathbf{Y}_{it_1, k^*}(0) = \mathbf{X}_{it, p^*}^{(1)} \cap \mathbf{X}_{it, p^*}^{(2)} = \emptyset$. This means that under Assumption 3.b in line with Equation 4, it is necessary to condition on $\mathbf{W}^{(1)} \cap \mathbf{W}^{(2)}$ to identify the ATET, however, it is sufficient to condition on $\boldsymbol{\theta}$ only.

To consistently estimate the ATET under Assumption 3.b and assert the overlap assumption, a propensity-based calliper matching estimator is used. In this example, this means that similar firms in Sundsvall and Sandviken are identified, and the electricity consumption of this matched sample and Sandviken are compared. The observed firm characteristics are few: The ampere dimension of the power subscription (Amp35 and Amp50) and the estimated difference in cost if the firm does not change their behaviour under the dynamic tariff given by

$$\Delta_{cost, i} = \frac{\text{Yearly distribution tariff cost during the pre-treatment period} | \text{Demand-based tariff}(\text{new})}{\text{Yearly distribution tariff cost during the pre-treatment period} | \text{Energy-based tariff}(\text{old})}. \quad (10)$$

This is clearly an insufficient amount of characteristics to base matching on; it lacks important factors for electricity consumption such as machine park, size of facilities, number of employees, heating system, etc.. However, the daily electricity consumption and temperature data, for the pre-treatment year of the study, are available for all firms in the study which makes it possible to estimate $\boldsymbol{\theta}$.

The parameter vectors $\boldsymbol{\theta}_i$ are estimated using the two-level time series approach discussed in Section 3.1. After fitting and refitting until balance in outcome was achieved in all pre-treatment time periods, the final within-firm time series model used for the electricity consumption under no treatment was given by

$$Y_{it} = \mu_i + \phi_{kWh, i} Y_{it-1} + \phi_{Temp, i} Temp_{j, t} + \phi_{Temp-1} Temp_{j, t-1} + \gamma'_{i, month} \mathbf{Z}_{it} + \epsilon_{it}, \quad (11)$$

for $t = 1, \dots, t_1 - 1$, where $\epsilon_{it} \sim N(0, \sigma_i^2)$, \mathbf{Z}_{it} is a vector of time-varying dummies for month, Y is the observed daily kWh consumption of electricity under the flat free, and $Temp_t$ is the outdoor

temperature at day t in city $j = (\text{Sandviken}, \text{Sundsvall})$. The final level-2 model is given by

$$\begin{aligned}
\mu_i &\sim N(\mu_\mu, \sigma_\mu^2) \\
\phi_{kWh,i} &\sim N(\mu_{\phi_{kWh}}, \sigma_{\phi_{kWh}}^2) \\
\phi_{Temp,i} &\sim N(\mu_{\phi_{Te}}, \sigma_{\phi_{Te}}^2) \\
\phi_{Temp-1,i} &\sim N(\mu_{\phi_{Te-1}}, \sigma_{\phi_{Te-1}}^2) \\
\log(\sigma_i^2) &\sim N(\mu_{\log \sigma^2}, \sigma_{\log \sigma^2}^2) \\
\gamma_{ik} &\sim N(\mu_{\gamma_k}, \sigma_{\gamma_k}^2), k = 2, \dots, 11 \\
\text{Probit}(D_i) &= \beta_0 + \beta_1 \mu_i + \beta_2 \phi_{kWh,i} + \beta_3 \phi_{Temp,i} + \\
&\quad \beta_4 \log(\sigma_i^2) + \gamma'_i \beta + \beta_{15} \Delta_{\text{cost},i} + \beta_{16} \text{Amp35}_i + \beta_{17} \text{Amp50}_i,
\end{aligned} \tag{12}$$

where γ is the vector of the month-dummy estimates. Since PS-based calliper matching is used, the PS is estimated directly using the probit link for the treatment assignment, $\text{Probit}(D_i)$. The model says that, the electricity consumption at day t is modelled as a function of the consumption of the previous day, the temperature that day, and what month it is.

Table 1 shows some key estimates. The parameters correspond to the parameters in Equation 12. Starting from the bottom of the table, it is clear that there is, e.g., a large variation in the residual variance term and a number of other parameters including the mean level consumption, all indicating heterogeneity across firms in consumption processes. Some firms may have a low mean level but a large variance due to high autocorrelation and residual variance, whereas other might be more dependent on the temperature and the season. Indeed, several of these unit-specific parameters also have significant effects on the treatment assignment indicating the importance of including these in the balancing procedure.

Figure 2 displays the overlap in the propensity of being located in Sandviken. Clearly there are firms in Sundsvall with similar propensity as those in Sandviken. However, the opposite is not true which is not a problem since the estimand of interest is the ATET. By using calliper matching based on the estimated propensity score and $\hat{\theta}$, the overlap will be fulfilled as firms in Sundsvall with non-overlapping propensities will be excluded.

The final model displayed in Equations 11 and 12 was obtained by studying balance in the pre-treatment outcome using a balance measure Δ_t suggested (Imbens and Rubin, 2015) for continuous

Parameter	Estimate	CI(low)	CI(high)
β_1	-0.008	-0.054	0.045
β_2	0.266	-0.2	0.713
β_3	1.878	-0.387	4.314
β_4	-0.072	-0.183	0.036
γ_8	0.655	0.043	1.328
γ_9	-0.958	-1.674	-0.308
γ_{10}	-0.366	-1.18	0.381
γ_{11}	2.049	0.888	3.126
γ_{12}	-0.764	-1.561	0.025
β_{15}	0.38	-0.168	0.942
β_{16}	0.176	-0.1	0.432
β_{17}	0.333	0.033	0.647
σ_μ^2	10.387	9.532	11.274
$\sigma_{\phi_{kWh}}^2$	0.057	0.052	0.062
$\sigma_{\phi_{Te}}^2$	0.003	0.003	0.004
$\sigma_{\gamma_8}^2$	0.085	0.075	0.096
$\sigma_{\gamma_9}^2$	0.087	0.076	0.1
$\sigma_{\gamma_{10}}^2$	0.059	0.051	0.068
$\sigma_{\gamma_{11}}^2$	0.043	0.038	0.05
$\sigma_{\gamma_{12}}^2$	0.046	0.04	0.053
$\sigma_{\log \sigma^2}^2$	1.813	1.669	1.976

Table 1: Shows some key estimates from the Y and D regressions.

outcomes given by

$$\Delta_t = \frac{\bar{y}_{t,d=1} - \bar{y}_{t,d=0}}{\sqrt{\frac{s_{t,d=0}^2 + s_{t,d=1}^2}{2}}}. \quad (13)$$

Figure 3 displays the pre-treatment balance for the final model. Δ_t is clearly smaller than the rule of thumb 0.25 (Imbens and Wooldridge, 2009) for all pre-treatment t , indicating balance. For convenience in comparing the result before and after matching Figure 4a repeats Figure 1. Figure 4b

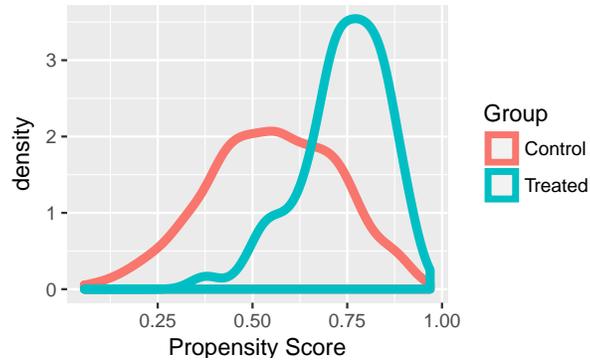


Figure 2: Overlap in firms' propensity of being located in Sandviken (being in the treated group).

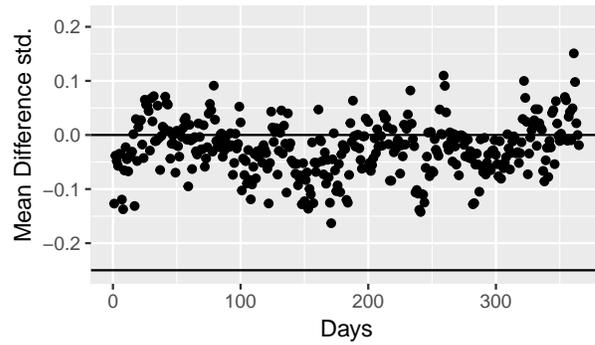


Figure 3: Δ_t for all pre-treatment time periods.

displays the smoothed average after matching based on the final model, i.e. the curve of the treated (Sandviken) is identical to that of Figure 4a. The curve of the control is updated to including only the firms in the control group (Sundsvall) matched to some company in the treated group.

After the matching is successful, the post-treatment data were consulted to estimate the ATET. In this paper the result is limited to a graphical illustration to save space, for all details of the

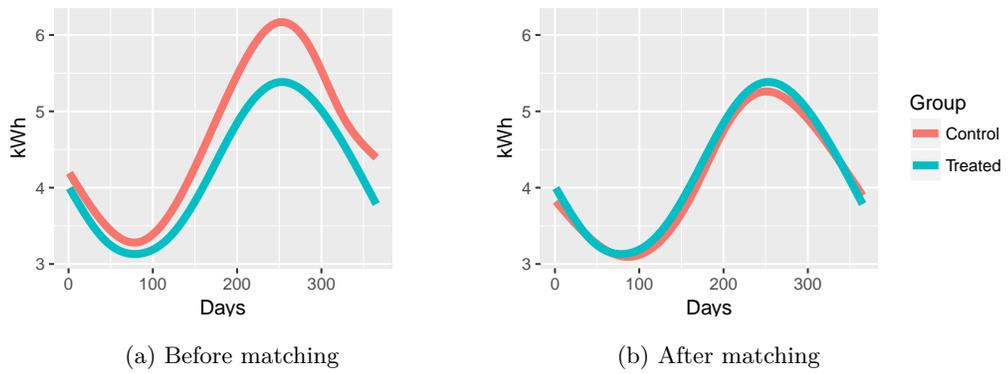


Figure 4: Grouped smoothed pre-treatment electricity consumption before and after matching. In the left figure all firms in Sundvall are included in the control-group, whereas in the right figure only the matched control firms are included.

estimated effect see Öhrlund et al. (2018). Figure 5 displays the smoothed mean difference between the matched treatment and control firms before and after the treatment. The point estimate of the ATET, averaged over the two post-treatment years, was significantly different from zero and found to be -0.32 kWh, which is about 7.4% of the average pre-treatment consumption. For completeness, Figure 6 displays the continuation of the smoothed means of the matched groups.

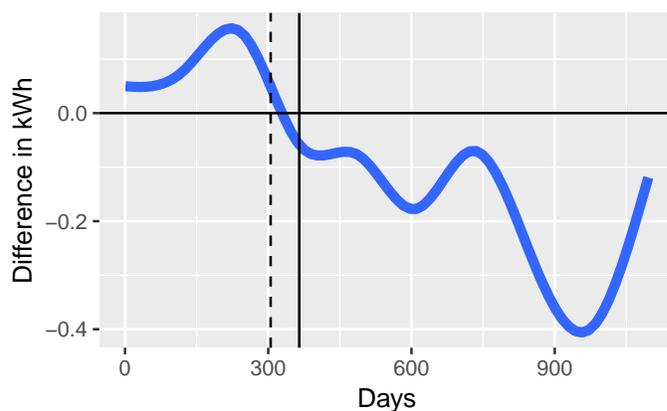


Figure 5: Smoothed mean for the difference between matched treated and control firms before and after treatment. The dashed and solid lines indicate the time periods where the firms in the treatment group were informed about the treatment and when the new tariff was applied, respectively.

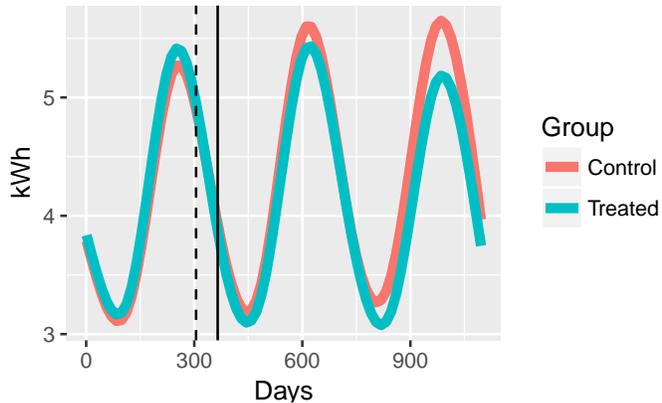


Figure 6: The pre and post-treatment smoothed means of the treated group and the matched control group. The dashed and solid lines indicate the time periods where the firms in the treatment group were informed about the treatment and when the new tariff was applied, respectively.

5 Discussion

This paper has proposed a method for identifying causal effects from non-experimental data by utilizing time series measurements of the pre-treatment outcome. The results in this paper show under which assumptions, and how, pre-treatment measurements of the outcome can be utilized in the identification and the estimation of the ATET. The theoretical results and estimation methods can be used in addition to non-complete sets of observed confounders, and in some situations completely substitute for time-invariant covariates.

The suggested strategies build on being able to characterize a high frequency pre-treatment outcome process of an observational unit by a small set of informative statistics. These statistics can then be used to identify causal effects by e.g. adjustment or matching. One might question why

observational units are not simply matched on, e.g., within-unit pre-treatment means. However, the results in this study show that the within-unit short time dynamics can hold important information about the heterogeneity across units. In other words, units with similar average levels may have different processes around their average level implying that there might be time periods for which the groups are not balanced even if they are balanced on average over time. A simple example illustrating this was given in appendix B. One trivial theoretical example illustrating why it might be important to balance dynamic aspects in order to get similar distributions in $Y(0)$ between the treatment groups, is if the outcome follows heterogeneous AR(1) processes. Since the variance of the marginal distribution of the outcome of each observational unit is a function of the autocorrelation, balance in the variance implies balance in the autoregressive parameters. This indicates that to make the outcome distribution of the treatment and control groups as similar as possible the groups should be balanced also with respect to the within-unit dynamics.

Given that there many time-series characterization and clustering/matching strategies and strategies for modelling the counterfactuals using time-series models, we discuss some important pros and cons of the suggested strategy. The matching estimation strategy suggested in this paper, uses a parametric two-level time series model to characterize each units outcome process. Matching on the observational unit specific parameter estimates makes it possible to non-parametrically estimate the ATET. The matching strategy, evaluated on the pre-treatment outcome, can be used in combination with any type of ATET estimator and inference. This means that even though complex Bayesian time series analysis is used to achieved balance, any classical effect estimator, or non-parametric tests can be utilized in the final step of the effect evaluation.

Comparing the proposed time-series characterizing strategy to non-model based strategies, the greatest difference is the ability in the proposed strategy to include time-varying covariates that do not vary across observational units. More specifically, the relation between such covariates and the outcome of each unit can in a natural way be utilized in the estimation. This is a major contribution of our strategy as the relation between time-varying covariates that do not vary across observational units and the outcome of each unit might reveal important differences in outcome processes, as was illustrated by the inclusion of temperature and season in the empirical example. In addition, the proposed strategy has, in our opinion, one great advantage in a causal inference setting, namely that it provides the possibility to check the overlap assumption. The model-based approach quantifies the heterogeneity in terms of parameter estimates, estimates that can in turn be used as independent

variables in a PS estimation used to check and fulfill the essential overlap assumption. This means that even though the non model-based time series matching strategies might give the better matches in some contexts (Aghabozorgi et al., 2015), the model based matching is more suitable in a causal identification context.

It is important to understand that time-varying covariates in $\mathbf{X}^{(1)}$ are only needed as long as they are also affecting the treatment mechanism. From Equation 4, it is clear that it is only the variation in Y which is related to the treatment assignment mechanism process that needs to be controlled for. This further implies that any proxies of covariates in X fulfilling Equation 4 can be used as substitutes. If e.g. X =temperature is unobserved and affects the outcome but not the assignment, it might be enough to add e.g. monthly dummies to the model to ensure that $\hat{\theta}$ fulfills Assumption 1. By the possibility to iterate the time series model specification, proxies can be added and removed until balance is achieved. We stress here again that, balance in $\hat{\theta}$ depends on the model specification and estimation, whereas balance in the pre-treatment measurements of $Y(0)$ does not. This means that the researcher does not risk being fooled by a close match on an incorrect estimate of θ as long as the evaluation of the group closeness is based on pre-treatment measurements of the outcome.

In conclusion, this paper has extended the notation and assumptions used in cross sectional observational studies to identify the ATET to a time series setting. The time series setting opens up the possibility to alter the assumptions utilizing the pre-treatment data more extensively in the identification and estimation. This development should enable more studies to be able to identify the ATET, even in fields where covariates are difficult or expensive to collect.

References

- Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505.
- Abadie, A., Gardeazabal, J., American, T., Review, E., and Mar, N. (2003). The Economic Costs of Conflict : A Case Study of the Basque Country The Economic Costs of Conflict : A Case Study of the Basque Country. *American Economic Review*, 93(1):113–132.
- Aghabozorgi, S., Seyed Shirkorshidi, A., and Ying Wah, T. (2015). Time-series clustering - A decade review. *Information Systems*, 53:16–38.
- Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earnings. *The Review of Economics and Statistics*, 60(1):47–57.
- Asparouhov, T., Hamaker, E. L., and Muthén, B. (2017). Dynamic Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3):1–30.
- Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics*, 9(1):247–274.
- Chabé-Ferret, S. (2015). Analysis of the bias of Matching and Difference-in-Difference under alternative earnings and selection processes. *Journal of Econometrics*, 185(1):110–123.
- Fu, T. C. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1):164–181.
- Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press.
- Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86.
- Muthén, L. K. and Muthén, B. O. (2017). *Mplus User ’ s Guide. Eighth Edition*. Los Angeles, CA: Muthén & Muthén.

- Neyman, J. (1923). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. *Annals of Agricultural Sciences*.
- Öhrlund, I., Schultzberg, M., and Bartusch, C. (2018). The issue of identifying and estimating demand response.
- O’Neill, S., Kreif, N., Grieve, R., Sutton, M., and Sekhon, J. S. (2016). Estimating causal effects: considering three alternatives to difference-in-differences estimation. *Health Services and Outcomes Research Methodology*, 16(1-2):1–21.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, 39(1):33–38.
- Rubin, D. B. (1973). Matching to Remove Bias in Observational Studies. *Biometrics*, 29(1):159–183.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.
- Rubin, D. B. (1980). Bias Reduction Using Mahalanobis-Metric Matching. *Biometrics*, 36(2):293–298.
- Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. P. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science : a review journal of the Institute of Mathematical Statistics*, 25(1):1–21.
- Trull, T. J. and Ebner-Priemer, U. W. (2009). Using Experience Sampling Methods/Ecological Momentary Assessment (ESM/EMA) in Clinical Assessment and Clinical Research: Introduction to the Special Section. *Psychological Assessment*, 21(4):457–462.

Zubizarreta, J. R. (2012). Using mixed integer programming for matching in an observational study of kidney failure after surgery. *Journal of the American Statistical Association*, 107(500):1360–1371.

A Illustration of the dimension reduction using two-level time series analysis

In this section a small simulation study illustrates how general multi-level time series models can be used to conveniently characterize a large number of long time series.

A.1 Part 1

Here $N=60$, $T=300$. To illustrate the utility of the models, 4 groups with outcome processes following different AR-processes are constructed. The data are generated as

$$Y_{jit} = \mu_j + \phi_{1ij}Y_{it-1} + \phi_{2ij}Y_{it-2} + \epsilon_{ijt} \quad (14)$$

where $j = 1, 2, 3, 4$. refers to the groups. Furthermore,

$$\mu_{i1}, \mu_{i2} \sim N(1, 0.2)$$

$$\mu_{i3}, \mu_{i4} \sim N(3, 0.1)$$

$$\phi_{1i1}, \phi_{1i2} \sim N(0.5, 0.3)$$

$$\phi_{1i3}, \phi_{1i4} \sim N(-0.2, 0.3)$$

$$\phi_{2i1}, \phi_{2i4} = 0$$

$$\phi_{2i2}, \phi_{2i3} \sim N(-0.4, 0.3)$$

$$\epsilon_{i1t} \sim N(0, 1)$$

$$\epsilon_{i2t} \sim N(0, 3)$$

$$\epsilon_{i3t} \sim N(0, 0.5)$$

$$\epsilon_{i4t} \sim N(0, 0.1)$$

Figure 7 displays the time series of all 60 observational units for 300 time periods. From the left panel it is difficult to distinguish between the different groups. The following model is fitted to this data

$$Y_{jit} = \mu_j + \phi_{1ij}Y_{it-1} + \phi_{2ij}Y_{it-2} + \phi_{3ij}Y_{it-3} + \epsilon_{ijt}. \quad (15)$$

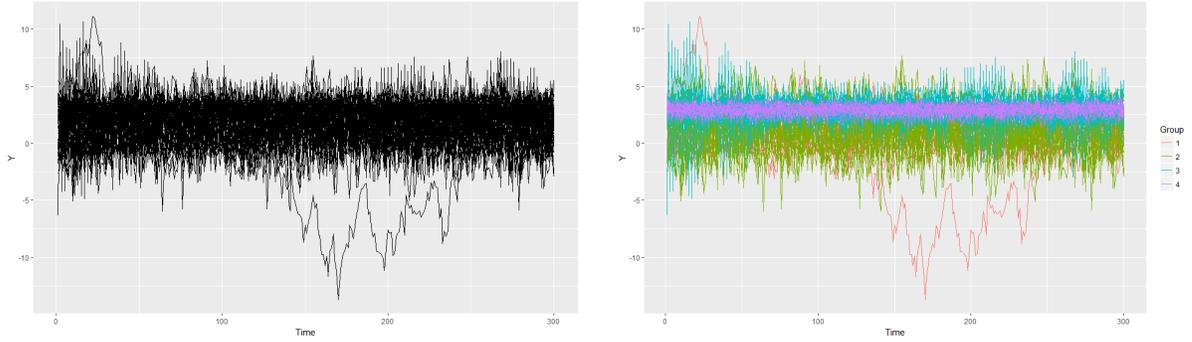


Figure 7: The time series of the 60 observational units. In the right panel the time series are coloured according to group.

The parameter estimates divided by groups are displayed in Figure 8. By looking at several parameters these four groups can easily be identified. For example, group 1 and 2 cannot be distinguished from based on the mean only, however, using the mean and the variance the separation is clear. Although not all parameters have perfect estimates, the model successfully captures and distinguishes between the four different groups, including two different orders of AR processes, in one estimation. Figure 9 displays an example from each of the four groups. It is clear that the levels and variances are different, however, the difference in the order and strength of the autocorrelation is less obvious from the plots.

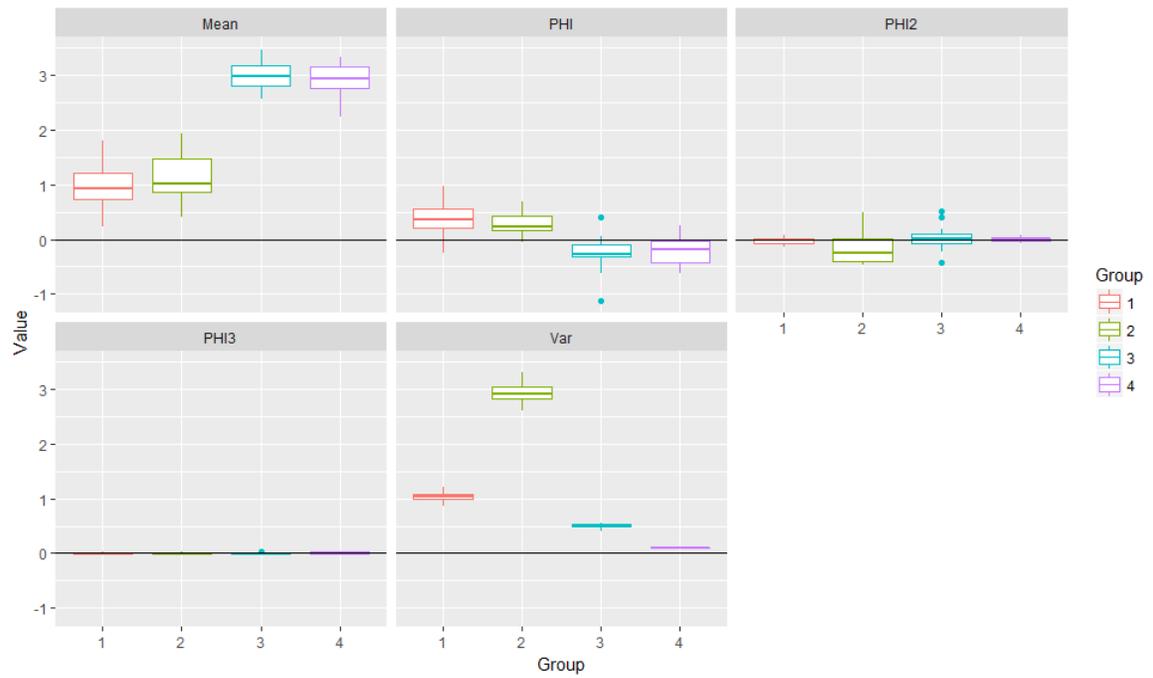


Figure 8: Grouped parameter estimates for all parameters of the model.

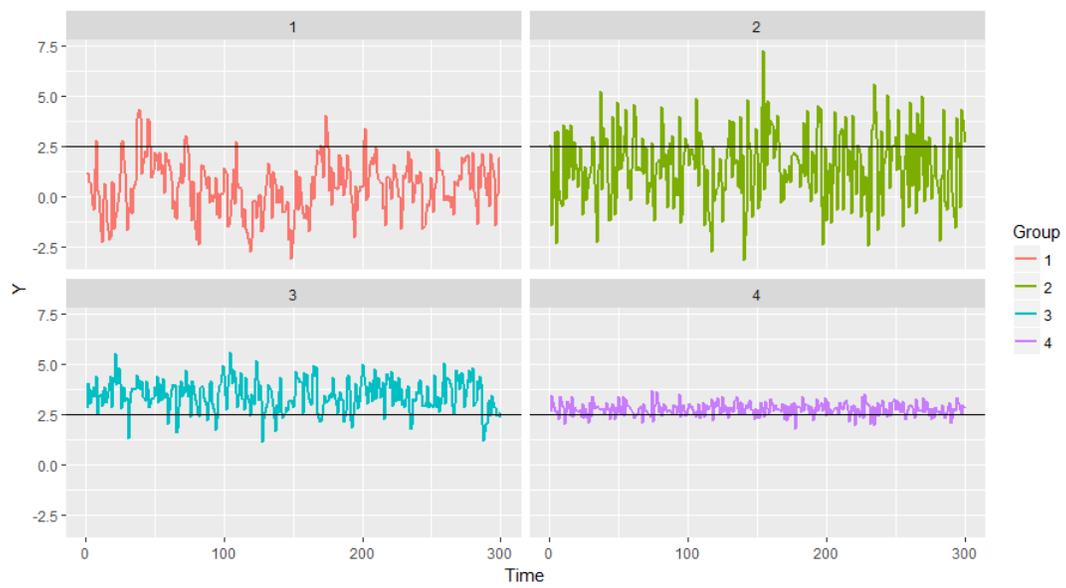


Figure 9: One example process from each of the four groups. The solid line is given as a level reference point.

B Possible gains of including dynamic parameters in matching

Here it is illustrated that balance evaluation based on pre-treatment within-observational unit means does not imply balance in the group means at each pre-treatment period, and, how this balance can be improved by evaluating the balance also in the autoregressive coefficients. Data are generated as

$$Y_{it} = \phi_i Y_{it-1} + \epsilon_{it} \quad (16)$$

where the error term is independently and identically distributed $\epsilon_{it} \sim N(0, 1)$ for all i and t . $\phi_i = I(p_i > 0.5)Z_i + (1 - I(p_i > 0.5))0.7$, where I is an indicator function, $Z_i \sim N(0, 0.1)$, and $p_i \sim Ber(0.5)$. Only one sample of 50 treatment and 150 controls is generated. 50 of the controls are randomly sampled without replacement from the pool of controls until there is balance on the respective criteria. The distribution of estimated AR-coefficients, estimated one time series

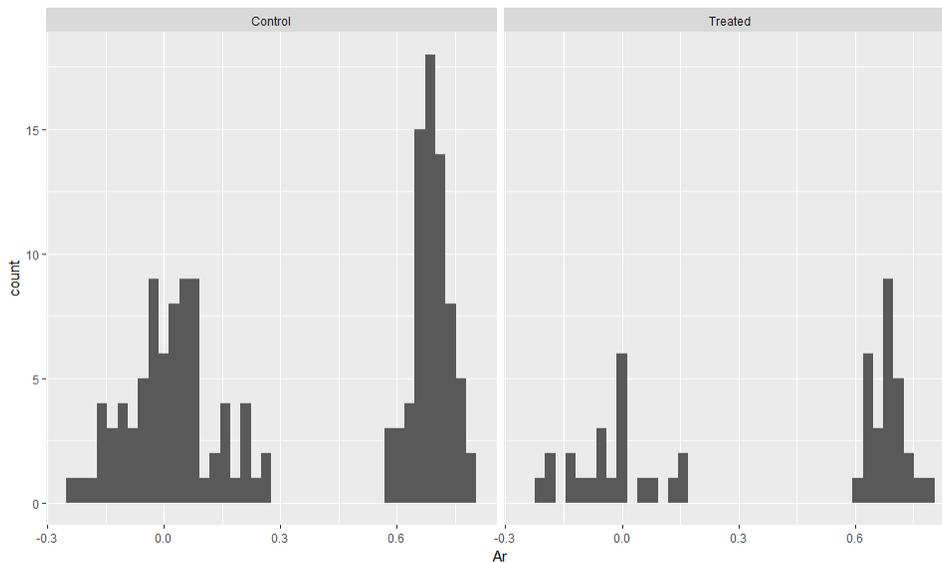


Figure 10: Distribution of autoregressive coefficients in the treatment and control pool.

at a time, in the sample before matching is given in Figure 10. The two groups with different AR-coefficients are clearly present in both the treatment group and the pool of eligible controls. This means that it is possible to find matches with similar AR-coefficients. To ensure that the

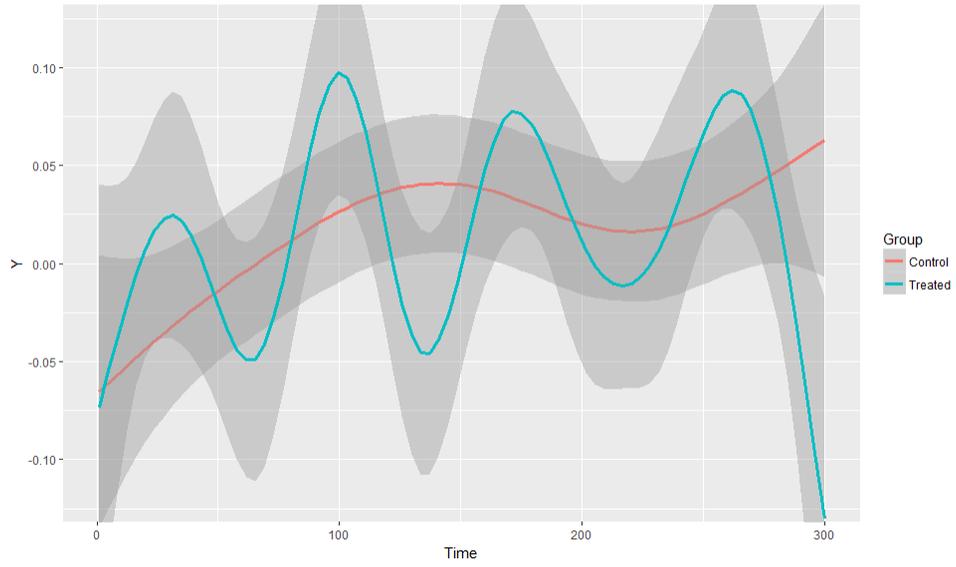


Figure 11: Balance checked against the pre-treatment within-observational unit mean only.

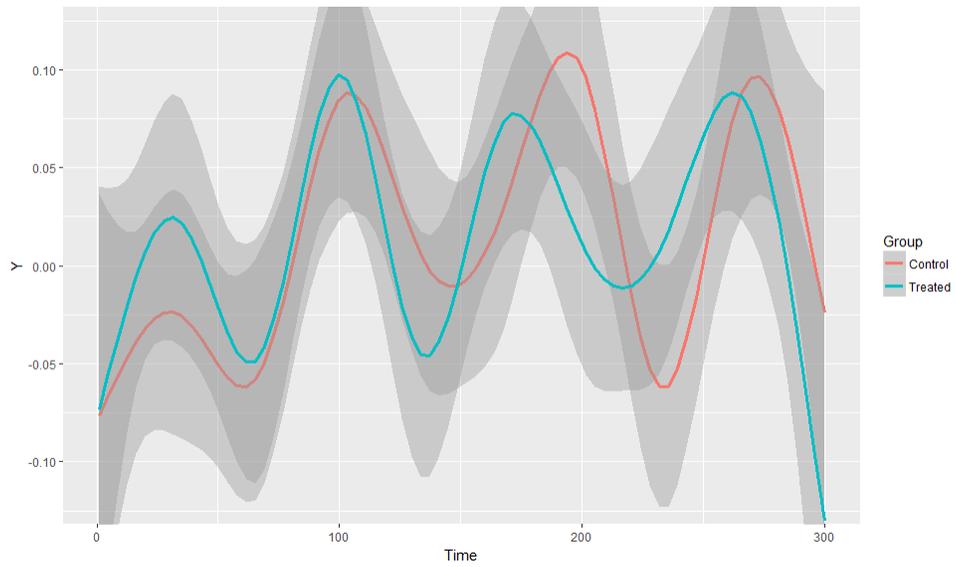


Figure 12: Balance checked against the estimated pre-treatment autoregressive coefficient and within-observational unit mean.

groups are balanced at all post-treatment time points of interest, the goal of the matching is to make the groups as similar as possible in the pre-treatment period. If the groups are comparable for all pre-treatment time points it should increase the likelihood for the balance to continue after this period unless any treatment is added. Figure 11 displays the pre-treatment smoothed group means after matching based on only the within-unit levels. It is clear that the mean level over the full pre-treatment period is approximately balanced. However, at several time points the groups differ substantially. Figure 12 displays the pre-treatment smoothed group means after matching based on the level and the estimated AR-coefficients. It is clear that this increases the similarity in pre-treatment processes drastically. The results illustrate the potential importance of utilizing the pre-treatment data fully.