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
January 2019
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 Schultzberg[1]

*PhD-student, Department of Statistics, Uppsala University, Sweden*
*Email: marten.schultzberg@statistics.uu.se*

September 27, 2018

[1]Acknowledgement: I would like to thank Per Johansson for discussions and suggestions to this paper, and Bengt Muthén and Mattias Nordin 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 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.

Keywords: DSEM, intensive longitudinal data, matching, treatment effects, two-level time series

# 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), a weighted set of pre-treatment measurements of outcomes from a pool of non-treated groups 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 Average Treatment Effect of the Treated (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 the proposed 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 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 reminder 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 returns to the empirical example presented in the introduction, and the proposed strategies are applied. Section 5 discusses the results and gives concluding remarks.

## 2 Setup

Consider an observational study setting where observational units, $i = 1, ..., N$, are measured repeatedly for $T$ consecutive time periods. Time periods $t = 1, 2, ..., t_1 - 1$ and $t = t_1, t_1 + 1, ..., T$ refer to time periods pre and post assignment of treatment, respectively. Let $D_{it}$ be an indicator function, taking the value 0 if not treated and 1 if treated, at time period $t$ for unit $i$. Once treated, $D_{it} = 1$ throughout the study. Furthermore, let $Y_{it}(d)$ be the potential outcome (Splawa-Neyman et al., 1990 translating Neyman, 1923, Rubin, 1973) at time period $t$ where $d$ is set to 1 if the observational unit $i$ 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

$$\text{ATET}_t = E[Y_{it}(1) - Y_{it}(0)|D_{it} = 1], \ \ \forall \ t = t_1, ..., T. \tag{1}$$

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, the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980) and the assumption of overlap are fulfilled. The third assumption, known as the weak unconfoundedness assumption (UA), is that the distributions of the potential outcomes under non-treatment are the same for the treated and non-treated (controls) given a set, $\mathbf{Z}_i^*$, of time-invariant covariates and time-varying covariates observed at specific time periods. The UA is defined formally as

$$Y_{it_1}(0) \perp\!\!\!\perp D_{it_1}|\mathbf{Z}_i^*, \ \ \forall \ t = t_1, ..., T. \tag{2}$$

If $\mathbf{Z}_i^*$ is known and observed, the ATET can be consistently estimated under the UA by controlling for $\mathbf{Z}_i^*$, however, in general $\mathbf{Z}_i^*$ is not known. The question of interest is what part of $\mathbf{Z}_i^*$ can be derived from pre-treatment outcome data for identification and estimation of the ATET.

Let $\mathbf{W}_i$ be the set of all possible time-invariant covariates and $\mathbf{X}_i$ be the set of all possible time-varying covariates. Let the data generating process of the outcome under no treatment at time period $t_1$ be given by

$$Y_{it_1}(0) = f_1\left(\mathbf{X}_{it_1,p_Y}^{(1)}, \mathbf{W}_i^{(1)}, \nu_{i,t_1}\right), \tag{3}$$

where $f_1$ is some function, $\mathbf{W}_i^{(1)}$ and $\mathbf{X}_i^{(1)}$ are subsets of $\mathbf{W}$ and $\mathbf{X}$, and $\nu_{i,t_1}$ is independently and identically distributed (iid) for all $t$ for each observational unit $i$, and independent of $\mathbf{X}_{it_1,p_Y}^{(1)}$ and $\mathbf{W}_i^{(1)}$, where $\mathbf{X}_{it_1,p_Y}^{(1)} = (\mathbf{X}_{it_1}^{(1)}, \mathbf{X}_{it_1-1}^{(1)}, ..., \mathbf{X}_{it_1-p_Y}^{(1)})$.

Correspondingly, let the treatment assignment mechanism at time period $t_1$ be given by

$$D_{it_1} = g\left(\mathbf{Y}_{it_1-1,k_D}(0), \mathbf{X}^{(2)}_{it_1-1,p_D}, \mathbf{W}^{(2)}_i, \xi_{i,t_1}\right), \tag{4}$$

where $g$ is some function, $\mathbf{W}^{(2)}_i$ and $\mathbf{X}^{(2)}_i$ are subsets of $\mathbf{W}$ and $\mathbf{X}$, and $\xi_{i,t_1}$ is iid for all $t$ for each observational unit $i$, and independent of $\mathbf{Y}_{it_1-1,k_D}(0), \mathbf{X}^{(2)}_{it_1-1,p_D}, \mathbf{W}^{(2)}_i$, where $\mathbf{Y}_{it_1-1,k_D}(0) = (y_{it_1-1}, y_{it_1-2}, ..., y_{it_1-k_D})$ and $\mathbf{X}^{(2)}_{it_1-1,p_D} = (\mathbf{X}^{(2)}_{it_1-1}, \mathbf{X}^{(2)}_{it_1-2}, ..., \mathbf{X}^{(2)}_{it_1-p_D})$.

This implies that the necessary set, $\mathbf{Z}^*_i$, for which the UA is satisfied, is given directly by

$$\mathbf{Z}^*_i = \left\{\mathbf{W}^{(1)}_i \cap \mathbf{W}^{(2)}_i, \mathbf{X}^{(1)}_{it_1,p^*} \cap \mathbf{X}^{(2)}_{it_1,p^*}\right\}, \tag{5}$$

where $p^*$ includes the lags of $\mathbf{X}^{(1)}_i \cap \mathbf{X}^{(2)}_i$ affecting both $Y_{it_1}(0)$ and $D_{it_1}(0)$. Equation 5 means that it is *necessary* to condition on the variables, and measurements thereof, that are common to the outcome and the treatment assignment at time period $t_1$.

Let $\mathbf{X}^{(obs)}_i$ be the set of observed time-varying covariates.

**Assumption 1** *The outcome at time period $t$ under no treatment for unit $i$ is given by*

$$Y_{it}(0) = f_2\left(\mathbf{X}^{(obs)}_{it,p_Y}, \mathbf{W}^{(1)}_i, \eta_{it}\right), \ \forall\ t = 1, ..., T,$$

*such that $\eta_{it}$ is independent of $\{\mathbf{X}^{(obs)}_{it,p_Y}, \mathbf{W}^{(1)}_i\}$, and $D_{it_1}$ for all $i$ and $t = t_1, ..., T$.*

Assumption 1 says that the outcome process under no treatment is the same during the full study period. Furthermore, the error term is independent of $D_{it_1}$ for all $t \geq t_1$, regardless of if $\mathbf{X}^{(obs)}_{it_1,p_Y}$ is a subset or a superset of $\mathbf{X}^{(1)}_{it_1,p_Y}$, which means that $\mathbf{X}^{(obs)}_{it_1,p_Y}$ in Assumption 1 always fulfills $\mathbf{X}^{(obs)}_{it_1,p_Y} \supseteq \mathbf{X}^{(1)}_{it_1-1,p^*} \cap \mathbf{X}^{(2)}_{it_1-1,p^*}$.

Without loss of generality, the function $f_2$ can be parameterized by the vector $\boldsymbol{\theta}_i = h(\mathbf{W}^{(1)}_i)$, where $h$ is a vector of functions all allowed to be non-injective. The non-injectiveness, i.e. $h^{-1}(\boldsymbol{\theta}_i) \neq \mathbf{W}^{(1)}_i$ imply that the function $h$ are allowed to reduce the information in $\mathbf{W}^{(1)}_i$ arbitrarily. The meaning of this reparametrization is illustated in the next subsection. Under Assumption 1 it follows directly from $\boldsymbol{\theta}_i = h(\mathbf{W}^{(1)}_i)$ that

$$Y_{it_1}(0) \perp\!\!\!\perp D_{it_1} \Big| \left\{\boldsymbol{\theta_i}, \mathbf{X}^{(obs)}_{it_1,p_Y}\right\}. \tag{6}$$

To identify the ATET for all post-treatment time periods an additional assumption is needed. Since time-varying covariates must be controlled for also after treatment to identify the ATET the statement

$$Y_{it}(0) \perp\!\!\!\perp D_{it_1} \Big| \left\{ \boldsymbol{\theta_i}, \mathbf{X}^{(obs)}_{it_1,p_Y}, \mathbf{X}^{(obs)}_{it_1+1,p_Y}, ..., , \mathbf{X}^{(obs)}_{iT,p_Y} \right\} \ \forall\ t = t_1, ..., T, \tag{7}$$

only implies that the ATET is identified for all $t = t_1, ..., T$ if $\mathbf{X}^{(obs)}$ is exogenous. In this context, that $\mathbf{X}^{(obs)}$ is exogenous means that $\mathbf{X}^{(obs)}$ cannot be causally affected by the treatment. That is, the potential values for $\mathbf{X}^{(obs)}_{it}$ must be the same under treatment and control for all post-treatment time periods. To clarify further, let $\mathbf{X}^{(obs)}_{it}(d)$ be the potential value of $\mathbf{X}^{(obs)}_{it}$ for unit $i$ at time $t$ given treatment $d = 0$ or $d = 1$. The ATET for any $t \geq t_1$ is then given by

$$E(Y_t(1) - Y_t(0)|D = 1) = E\left( f_1(\mathbf{X}^{(1)}_{t,p_Y}(1), \mathbf{W}^{(1)}_i, \eta_{it}) - f_1(\mathbf{X}^{(1)}_{t,p_Y}(0), \mathbf{W}^{(1)}_i, \eta_{it})|D = 1 \right). \tag{8}$$

From equation 8 it is clear that the ATET cannot be identified if there are covariates in the minimal sufficient set $\mathbf{X}^{(1)} \cap \mathbf{X}^{(2)}$ that are affected by the treatment.

In summary, to identify the ATET for time periods $t = t_1, ..., T$, under Assumption 1, it is sufficient to condition on the parameters $\boldsymbol{\theta}_i$ describing the relation between the outcome and its previous values and the time-varying covariates in $\mathbf{X}^{(obs)}$, where the covariates in $\mathbf{X}^{(obs)}$ have to be exogenous.

## 2.1 Illustration of Assumption 1 and its implications

The purpose of this section is to give a better understanding of the notation and the intuition of Assumption 1. In addition, to illustrate that this strategy is robust against the functional forms of $h$, arbitrary, purposely complex relations are chosen.

Denote by $X_{q,i,t}$ covariate $q$ of unit $i$ at time $t$. Let the outcome process under no treatment be given by

$$Y_{i,t}(0) = \theta_{1,i} + \theta_{2,i} X_{1,i,t-2} + \theta_{3,i} X_{2,i,t} + \theta_{4,i} X_{3,i,t-1} + \nu_{i,t}, \tag{9}$$

where $\nu_{i,t}$ is iid with $Var(\nu_{i,t}) = \theta_{5,i}$, for $t = 1, ..., T$, where the parameters are given by

$$\begin{aligned}
\theta_{1,i} &= \beta_{1,1} + \beta_{1,2} W_{i,1} W_{i,2} \\
\theta_{2,i} &= \beta_{2,1} + \beta_{2,2} W_{i,2}^2 \\
\theta_{3,i} &= c \\
\theta_{4,i} &= \beta_{4,2} |W_{i,3}| \\
\log(\theta_{5,i}) &= \beta_{5,1} + \beta_{5,2} W_{i,1}^2
\end{aligned}$$

where $W_{i,j}$ are elements of $\mathbf{W}_i$. Further assume that the treatment assignment process is given by

$$D_{t_1} = l(\gamma_{1,i} + \gamma_{2,i} X_{1,i,t_1-2} + \xi_{i,t_1}),$$

where $\xi_{i,t}$ is iid with $Var(\xi_{i,t}) = \gamma_{3,i}$, $l$ is some suitable link function, and the parameters are given by

$$\begin{aligned}
\gamma_{1,i} &= \zeta_{1,1} + \zeta_{1,2} W_{i,2}^2 + \zeta_{1,3} W_{i,4}^2 \\
\gamma_{2,i} &= \zeta_{2,1} + \zeta_{2,2} e^{W_{i,4}} \\
\log(\gamma_{3,i}) &= \zeta_{3,2} W_{i,1}.
\end{aligned}$$

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 6, in which elements of $\mathbf{W}_i^{(1)}$ play no part. In other words, $\boldsymbol{\theta}_i$ can replace $\mathbf{W}_i^{(1)}$ regardless of the information about $\mathbf{W}_i^{(1)}$ it contains. For this example, $\mathbf{W}_i^{(1)} = \{W_1, W_2, W_3\}$, $\mathbf{X}_{i,t_1,p_Y}^{(1)} = \mathbf{X}_{i,t_1,2}^{(1)} = \{X_{1,i,t}, X_{1,i,t-1}, X_{1,i,t-2}, X_{2,i,t}, X_{2,i,t-1}, X_{2,i,t-2}, X_{3,i,t}, X_{3,i,t-1}, X_{3,i,t-2}\}$, $\boldsymbol{\theta}_i = \{\theta_{1,i}, \theta_{2,i}, \theta_{3,i}, \theta_{4,i}, \theta_{5,i}\}$. Moreover, $\mathbf{W}_i^{(2)} = \{W_{i,1}, W_{i,2}, W_{i,4}\}$, $\mathbf{X}_{i,t_1-1,p_D}^{(2)} = \mathbf{X}_{i,t_1-1,1}^{(2)} = \{X_{1,i,t_1-1}, X_{1,i,t_1-2}\}$, and $\boldsymbol{\gamma}_i = \{\gamma_{1,i}, \gamma_{2,i}, \gamma_{3,i}\}$. Hence, $\mathbf{W}_i^{(1)} \cap \mathbf{W}_i^{(2)} = \{W_1, W_2\}$, and $\mathbf{X}_{i,t_1,p^*}^{(1)} \cap \mathbf{X}_{i,t_1,p^*}^{(2)} = \{X_{1,i,t_1-2}\}$, and the minimal sufficient set is given by $\mathbf{Z}_i^* = \{W_{i,1}, W_{i,2}, X_{1,i,t_1-2}\}$.

Let the set of observed time-varying covariates be given be, e.g., $\mathbf{X}_i^{(obs)} = \{X_1, X_2\}$, which clearly fulfils $\mathbf{X}_i^{(obs)} \supseteq \mathbf{X}_i^{(1)} \cap \mathbf{X}_i^{(2)}$. By the results in the previous section, this implies that under Assumption 1, e.g., the folloWing statements[1] hold

$$\{Y_{i,t_1}(0), Y_{i,t_1+1}(0)\} \perp\!\!\!\perp D_{i,t_1} | W_{i,1}, W_{i,2}, X_{1,i,t_1-2}, X_{1,i,t_1-1} \tag{10}$$

$$\{Y_{i,t_1}(0), Y_{i,t_1+1}(0)\} \perp\!\!\!\perp D_{i,t_1} | \boldsymbol{\theta}_i, \mathbf{X}_{i,t_1,2}^{(obs)}, \mathbf{X}_{i,t_1+1,2}^{(obs)} \tag{11}$$

$$\{Y_{i,t_1}(0), Y_{i,t_1+1}(0)\} \perp\!\!\!\perp D_{i,t_1} | \mathbf{W}_i^{(1)}, \mathbf{X}_{i,t_1,2}^{(1),}, \mathbf{X}_{i,t_1+1,2}^{(1),}, \tag{12}$$

[1] Any superset of the minimal sufficient set will fulfill the conditional independence.

where only two post-treatment periods are considered to make it easier to follow. The proof of any of the statements follows directly from substituting $Y(0)$ and $D$ With corresponding functions. For example, at time period $t_1$, Equation 11 equals

$$\theta_{1,i} + \theta_{2,i} X_{1it_1-2} + \theta_{3,i} X_{2it_1} + \theta_{4,i} X_{3it_1-1} + \nu_{it_1} \perp\!\!\!\perp D_{it_1} \Big| \begin{Bmatrix} \theta_{1,i}, \theta_{2,i}, \theta_{3,i}, \theta_{4,i}, \theta_{5,i}, \\ X_{1,i,t_1}, X_{1,i,t_1-1}, X_{1,i,t_1-2}, \\ X_{2,i,t_1}, X_{2,i,t_1-1}, X_{2,i,t_1-2} \end{Bmatrix}$$

from which conditional independence follows immediately since $X_3$ and $\eta$ are independent of the treatment mechanism by assumption, and everything else in the outcome process are conditioned on. The exact same arguments applies for any other post-treatment period up to $T$.

Clearly, all but the first statement conditions on more information than necessary, however, Equation 11 and 12 are particularly useful for estimation as will be shown. The following sections extend Assumption 1 in two steps.

## 2.2 Assumption enabling estimation

Let $\mathbf{Y}_{it_1,k_Y}(0) = (y_{it_1-1}, y_{it_1-2}, ..., y_{it_1-k_Y})$.

**Assumption 2** *The distributions of the covariates in* $\mathbf{X}^{(obs)}$ *are the same under treatment and no treatment for each* $t = 1, ..., T$. *Furthermore, the outcome process under no treatment can be expressed as*

$$Y_{it}(0) = f^* \left( \mathbf{Y}_{it,k_Y}(0), \mathbf{X}^{(obs)}_{it,p_Y}, \boldsymbol{\theta}^*_i, \epsilon_{it} \right) \; \forall \, t = 1, ..., T, i =, ..., N$$

*such that*

$$\epsilon_{it} \perp\!\!\!\perp D_{it_1} \forall \, t = 1, ..., T$$

*and*

$$Y_{it}(0) \perp\!\!\!\perp D_{it_1} \Big| \left\{ \boldsymbol{\theta}^*_i, \mathbf{X}^{(obs)}_{it_1,p_Y}, \mathbf{X}^{(obs)}_{it_1+1,p_Y}, ..., , \mathbf{X}^{(obs)}_{iT,p_Y} \right\} \; \forall \; t = 1, ..., T.$$

Assumption 2 states that the outcome can be modelled as a time series process of lags and observed time varying covariates with error terms independent of the treatment assignment at all $t$. Furthermore, treatment assignment is unconfounded given the derived $\boldsymbol{\theta}^*_i$, and observed time-varying covariates. An intuition for the assumption is that, lags of the outcome are sufficient proxies for the part of $\mathbf{X}^{(1)}$ that is not in $\mathbf{X}^{(obs)}$ or in $\mathbf{X}^{(1)} \cup \mathbf{X}^{(2)}$, and although those X's are not necessary

to condition on to fulfil the UA, they might be necessary to estimate $\boldsymbol{\theta}_i^*$ such that Assumption 2 holds.

The assumption $\epsilon_{it} \perp\!\!\!\perp D_{it_1} \forall t = 1, ..., T$ is strong. However, if the number of parameters estimated to derive the distribution of $\boldsymbol{\theta}_i^*$ is less than the number of pre-treatment observations, sensitivity analysis of this assumption is possible. Given the final sample of treated and controls (based on similarity in $\mathbf{X}_{it_1,p_Y}^{(obs)}$ and estimates of $\boldsymbol{\theta}_i^*$, using e.g. matching) the distributions of the error terms, $e_{it}^1 = Y_{it} - \overline{Y}_t^1$ $i = 1, ..., n_1$ and $e_{jt}^0 = Y_{jt} - \overline{Y}_t^0, j = 1, ..., n_0$ for the treated (1) and controls (0) can be studied for all time periods $t = 1, .., t_1 - 1$. Substantial differences in the distributions at any pre-treatment time period period falsifies Assumption 2. A simplified sensitivity analysis is to restrict the evaluation to the differences in means, i.e, study the balance of $\overline{Y}_t^1$ and $\overline{Y}_t^0$ for all time periods (illustrated in the empirical example in Section 4). If the estimates of $\boldsymbol{\theta}_i^*$ do not contain at least as much information about the heterogeneity in the processes as $\boldsymbol{\theta}_i$, it is implausible that balance would be obtained for all pre-treatment time periods.

Note that if there are no time-varying covariates affecting the outcome or the treatment assignment (i.e. $(\mathbf{X}_{it}^{(1)} \cap \mathbf{X}_{it}^{(2)} = \emptyset)$, Assumption 2 implies that it is sufficient to control for *only* the pre-treatment measurements of the outcome. This holds even if $\mathbf{X}_{it}^{(1)} \neq \emptyset$, since by the assumption, the relation between $Y$ and $\mathbf{X}^{(1)}$ can be rewritten in terms of lags of the outcome in this case.

# 3 Estimation strategies

Several considerations have to be made in the estimation of $\boldsymbol{\theta}_i^*$. The most obvious challenge is to condense the information in the pre-treatment measurements. For example, the naive estimator $\hat{\boldsymbol{\theta}}_i^* = \{y_{it}\}_1^{t_1-1}$, would obviously bring the curse of dimensionality into the ATET estimation, and can be disregarded directly with intensive longitudinal data. This paper instead suggests approximating $f^*$ with a parametric time series model for each observational unit. Once a time series is fitted to an observational unit's 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 parameter estimates might give a sufficiently detailed description of the heterogeneity in processes to achieve balance in the matching.

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 to improve the description of the heterogeneity across units. Returning to the electricity consumption example described in the introduction an example of such variables is 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 $\boldsymbol{\theta}_i^*$. Including important time-varying covariates also makes Assumption 2 more realistic as less information is approximated by lags of $Y$. Appendix B gives a simple illustration of the potential gain of including parameters describing the dynamic aspects of the outcome processes in the matching.

An important aspect in the estimation of $\boldsymbol{\theta}_i^*$ is that, without parametric assumptions, the identification of the parameters is limited by the available pre-treatment data. If, e.g., some parts of $\boldsymbol{\theta}_i^*$ describe the relation between a time-varying covariate and the outcome, the estimated $\hat{\boldsymbol{\theta}}_i^*$ will only describe this relation for the span of variation in the time-varying covariate that was observed during the pre-treatment period. This means that, if the time-varying covariate takes very different values in the post-treatment period as compared to the pre-treatment period, the estimate might not accurately describe the relation during the post-treatment period. This indicates the importance of having a long enough pre-treatment period that reflects the 'typical' variation in the time-varying covariates that affects the outcome. In the electricity example, one important time-varying covariate is temperature. Therefore, it is important to not only have a pre-treatment period during, e.g., the summer if the post-treatment period contains winter etc. And even if the pre-treatment periods contains all seasons it is good practise to make sure the pre-treatment period was not very atypical, if such information is available.

## 3.1 Multivariate two-level time series approach

In this section an estimation strategy to consistently estimate the ATET under Assumption 2 is proposed. This paper suggests 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., 2018) 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. Bayesian MCMC estimation is employed to accommodate many random effects. Non-informative priors are used. 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 auto-regressive coefficients and residual variances. Since zero estimates are allowed, 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. Combined, this means that DSEM helps to reduce the amount of information in the pre-treatment by allowing for similarity while conserving the heterogeneity across units in a natural way. Appendix A presents a simple example illustrating 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 be be too similar, to a degree where the heterogeneity is underestimated, it will be apparent.

## 3.2 Effect estimation using matching

In this paper, the following estimator is proposed. Fit a two-level time series model to the pre-treatment data and match on $\boldsymbol{\theta}_i^*$ and any observed time-invariant covariates, say $\mathbf{W}_i^{(obs)}$, 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 coefficients $\boldsymbol{\theta}_i^*$ and the observed covariates $\mathbf{W}^{(obs)}$. To estimate ATET consistently, $\mathbf{X}^{(obs)}$ must be conditioned on as well. As $t_1$ is supposedly large, it is generally not possible to match on the full pre-treatment $\mathbf{X}^{(obs)}$. If all covariates in $\mathbf{X}^{(obs)}$ do not vary across units (i.e. $\mathbf{X}_{it}^{(obs)} = \mathbf{X}_{t}^{(obs)} \, \forall \, t$) the time-varying covariates can be excluded in the estimation right away as they are independent of the treatment assignment by definition. However, all time-varying covariates in $\mathbf{X}^{(obs)}$ that vary across units must, in addition to the matching on $\boldsymbol{\theta}_i^*$, be controlled for explicitly in estimation, by assuming a functional form between the outcome and the covariates guided by the estimates from the two-level model.

# 4 Empirical example with electricity consumption data

In the motivating 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.

This example uses 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. 157 firms in Sandviken and Sundsvall were excluded from the sample due to lack of variation or too large amount of pre-treatment periods with zero consumption. In total, 1110 firms from both cities were included in the matching. In the final effect estimation, 184 of the firms in Sandviken were included and the matched control group consisted of 140 firms from Sundsvall. 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).

There are three important aspects to account for when comparing electricity consumption of firms. The first, most obvious aspect is outdoor temperature. It is important since heating is one of the largest sources of electricity consumption. The second aspect is the price for electricity. The treatment in this case concerns the price of the grid, but the price for the watt hours is still free. Finally, the third aspect that must be accounted for is the market, including recessions and changes in demand etc. The city of Sundsvall was chosen as a control group to Sandviken with these aspects in mind. The temperature difference is small since the cities are closely located (200km)[2]. Regarding the market, the population size of Sundvall is around two times as large as Sandviken, however, the two cities have similar industry structure, are part of the same region, and have access to the same electricity market. Figure 7 displays the average electricity consumption of the supplied firms in the two cities the year before treatment.

[Figure 1 about here.]

From Figure 7 it is clear that the seasonality pattern of electricity consumption is similar 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 is 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 it has to be assumed that a firm located in Sandviken could have been located in Sundsvall, given all possible values of all potential confounders.

Assumption 2 is somewhat simplified in this case as the firms cannot choose to be treated, however since the location and the treatment are exactly dependent, any time-varying covariates that may vary between these two locations must be included. The cities are chosen as they are close to each other and share most of the environmental factors such as electricity price and recessions.

[2]The small remaining difference is addressed in the estimation section below.

The only thing that might be slightly different is the outdoor temperature in the two cities. As temperature is not independent of treatment-assignment (location) and temperature affects the outcome, $\mathbf{X}_{it}^{(1)} \cap \mathbf{X}_{it}^{(2)} = Temp_{jt}$ where $j =$(Sandviken, Sundsvall). The outdoor temperature cannot be causally affected by the treatment assignment which implies that $Temp_{jt}(0) = Temp_{jt}(1)$ for all $j$ and $t$. This means that under Assumption 2, in line with Equation 6, it is necessary to condition on $\mathbf{W}^{(1)} \cap \mathbf{W}^{(2)}$ and the temperature lags to identify the ATET for all post-treatment time periods. It is, however, sufficient to condition on $\boldsymbol{\theta}_i^*$ and the temperature lags.

To consistently estimate the ATET under Assumption 2 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 (Amp50 and Amp60) and the estimated difference in cost if the firm does not change their behaviour under the dynamic tariff given by

$$\Delta_{cost,i} = \frac{\text{Yearly distribution tariff cost during the pre-treatment period|Demand-based tariff(new)}}{\text{Yearly distribution tariff cost during the pre-treatment period|Energy-based tariff(old)}}. \tag{13}$$

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 $\boldsymbol{\theta}_i^*$.

The parameter vectors $\boldsymbol{\theta}_i^*$ are estimated using the two-level time series approach discussed in Section 3.1. After fitting and refitting, 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

$$Y_{it} = \mu_i + \phi_{kWh,i} Y_{it-1} + \phi_{Temp,i} Temp_{j,t} + \phi_{Temp-1} Temp_{j,t-1} + \boldsymbol{\gamma}'_{i,month} \boldsymbol{M}_t + \epsilon_{it}, \tag{14}$$

for $t = 1, ...t_1 - 1$, where $\epsilon_{it} \sim N(0, \sigma_i^2)$, $\mathbf{M_t}$ 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 $j$ = (Sandviken, Sundsvall). The final level-2 model is given by

$$\begin{aligned}
\mu_i &\sim N(\mu_\mu, \sigma^2_\mu) \\
\phi_{kWh,i} &\sim N(\mu_{\phi_{kWh}}, \sigma^2_{\phi_{kWh}}) \\
\phi_{Temp,i} &\sim N(\mu_{\phi_{Te}}, \sigma^2_{\phi_{Te}}) \\
\phi_{Temp-1,i} &\sim N(\mu_{\phi_{Te-1}}, \sigma^2_{\phi_{Te-1}}) \\
\log(\sigma^2_i) &\sim N(\mu_{\log\sigma^2}, \sigma^2_{\log\sigma^2}) \\
\boldsymbol{\gamma_{ik}} &\sim N(\mu_{\gamma_k}, \sigma^2_{\gamma_k}), k = 1, ..., 11 \\
\text{Probit}(D_i) &= \beta_0 + \beta_1\mu_i + \beta_2\phi_{kWh,i} + \beta_3\phi_{Temp,i} + \\
&\quad \beta_4 \log(\sigma^2_i) + \boldsymbol{\gamma}'_i\boldsymbol{\beta} + \beta_{16}\Delta_{\text{cost},i} + \beta_{17}\text{Amp50}_{\ i} + \beta_{18}\text{Amp60}_i,
\end{aligned} \tag{15}$$

where $\boldsymbol{\gamma}$ is the vector of the month-dummy estimates and $\boldsymbol{\beta} = (\beta_5, ..., \beta_{15})^T$. Since PS-based calliper matching is used, the PS is estimated directly using the probit link for the treatment assignment, 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 the month. The large number of parameters capturing potential heterogeneity in the temperature dependency is motivated by the often large importance of outdoor temperature.

Table 1 shows some key estimates. The parameters correspond to the parameters in Equation 15. 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.

[Table 1 about here.]

Figure 8 displays the overlap in the propensity of being located in Sandviken. Clearly there are firms in Sundvall 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 $\hat{\boldsymbol{\theta}}^*_i$, the overlap will be fulfilled as firms in Sundsvall with non-overlapping propensities will be excluded.

[Figure 2 about here.]

The final model displayed in Equations 14 and 15 was obtained by studying balance in the pre-treatment outcome using a balance measure, Mean Difference standardized (MDS), suggested (Imbens and Rubin, 2015) for continuous outcomes given by

$$MDS_t = \frac{\bar{y}_{t,d=1} - \bar{y}_{t,d=0}}{\sqrt{\frac{s^2_{t,d=0} + s^2_{t,d=1}}{2}}}. \tag{16}$$

Figure 9 displays the pre-treatment balance for the final model. $MDS_t$ is clearly smaller than the rule of thumb 0.25 (Imbens and Wooldridge, 2009) for all pre-treatment $t$, indicating balance.

[Figure 3 about here.]

In this application, balance in the pre-treatment outcomes was obtained after trying out only three different model specifications, formulated in consultation with researchers from within the electricity field. This indicates that the model is a sufficiently good description of the outcome processes to capture the important heterogeneities, and makes it highly implausible that the pre-treatment outcome balance is obtained by chance rather than successful matching. In other words, it is not plausible that matching on the parameter estimates from a model specification that does not describe the outcome processes obtains balance in pre-treatment outcome for the full 365 pre-treatment time period, in the third attempt.

For convenience in comparing the result before and after matching, Figure 10a repeats Figure 7. Figure 10b 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 10a. The curve of the control is updated to including only the firms in the control group (Sundsvall) matched to some firm in the treated group.

[Figure 4 about here.]

In this particular case, successful matching implies that matched firms have similar response to temperature shifts, and similar level of consumption, although there are small differences in temperature between the cities. $\boldsymbol{\theta}_i^*$ is clearly sufficiently well estimated to capture the non-constant temperature difference between the cities over the pre-treatment year. This is reassuring for the quality of $\hat{\boldsymbol{\theta}}_i^*$, however, it is important to keep in mind that the parts of $\hat{\boldsymbol{\theta}}_i^*$ that describe the temperature dependency might only describe the relation between the outcome and the temperatures

within the temperature span that were observed during the pre-treatment year. If the temperatures or temperature differences are very different from the pre-treatment year during the two post-treatment years, $\hat{\boldsymbol{\theta}}_i^*$ might not describe the relation between the consumption and the temperature during this period. This potential problem can be addressed by consulting the temperature data for all years. The temperature difference between the cities in this example is small overall, and the temperatures and temperature differences are similar for all three years.

After the matching was deemed 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 estimated effect see Öhrlund et al. (2018). Figure 11 displays the smoothed mean difference between the treatment and matched control firms before and after the treatment. The point estimate of the ATET, estimated with a panel regression model controlling for temperature, was significantly different from zero and found to be -0.32 kWh averaged over the two post-treatment years, which is about 7.4% of the average pre-treatment consumption. For completeness, Figure 12 displays the continuation of the smoothed means of the matched groups.

[Figure 5 about here.]

[Figure 6 about here.]

# 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 and residual variance, balance in the variance only implies balance in the residual variance if the autoregressive parameters are balanced. 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, some important pros and cons of the suggested strategy are discussed in this paper. The matching estimation strategy suggested in this paper, uses a parametric two-level time series model to characterize each units outcome process. However, 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 can be utilized in the final step of the effect evaluation. If there are no necessary time-varying covariates, it is even possible to non-parametrically estimate the ATET.

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 the proposed 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.

In conclusion, this paper has extended to a time series setting the notation and assumptions used in cross sectional observational studies to identify the ATET. 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 Shirkhorshidi, 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. (2018). 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

$$Y_{jit} = \mu_j + \phi_{1ij} Y_{it-1} + \phi_{2ij} Y_{it-2} + \epsilon_{ijt} \tag{17}$$

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

$$\begin{aligned}
\mu_{i1}, \mu_{i2} &\sim N(1, 0.2) \\
\mu_{i3}, \mu_{i4} &\sim N(3, 0.1) \\
\phi_{1i1}, \phi_{1i2} &\sim N(0.5, 0.3) \\
\phi_{1i3}, \phi_{1i4} &\sim N(-0.2, 0.3) \\
\phi_{2i1}, \phi_{2i4} &= 0 \\
\phi_{2i2}, \phi_{2i3} &\sim N(-0.4, 0.3) \\
\epsilon_{i1t} &\sim N(0, 1) \\
\epsilon_{i2t} &\sim N(0, 3) \\
\epsilon_{i3t} &\sim N(0, 0.5) \\
\epsilon_{i4t} &\sim N(0, 0.1)
\end{aligned}$$

Figure 1 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

$$Y_{jit} = \mu_j + \phi_{1ij} Y_{it-1} + \phi_{2ij} Y_{it-2} + \phi_{3ij} Y_{it-3} + \epsilon_{ijt}. \tag{18}$$

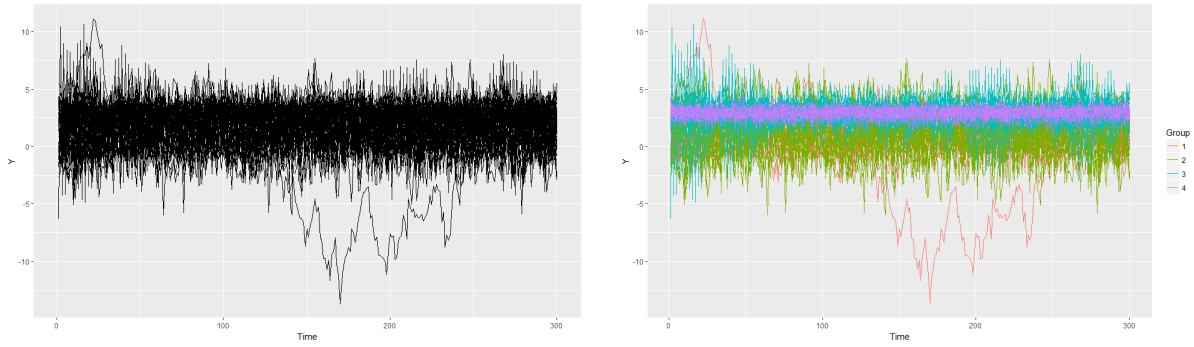



Figure 1: The time series of the 60 observational units. In the right panel the time series are coloured according to group.

The parameter estimates divided by groups are displayed in Figure 2. 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 3 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.

# 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

$$Y_{it} = \phi_i Y_{it-1} + \epsilon_{it} \tag{19}$$

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 funtion, $Z_i \sim N(0, 0.1)$, and

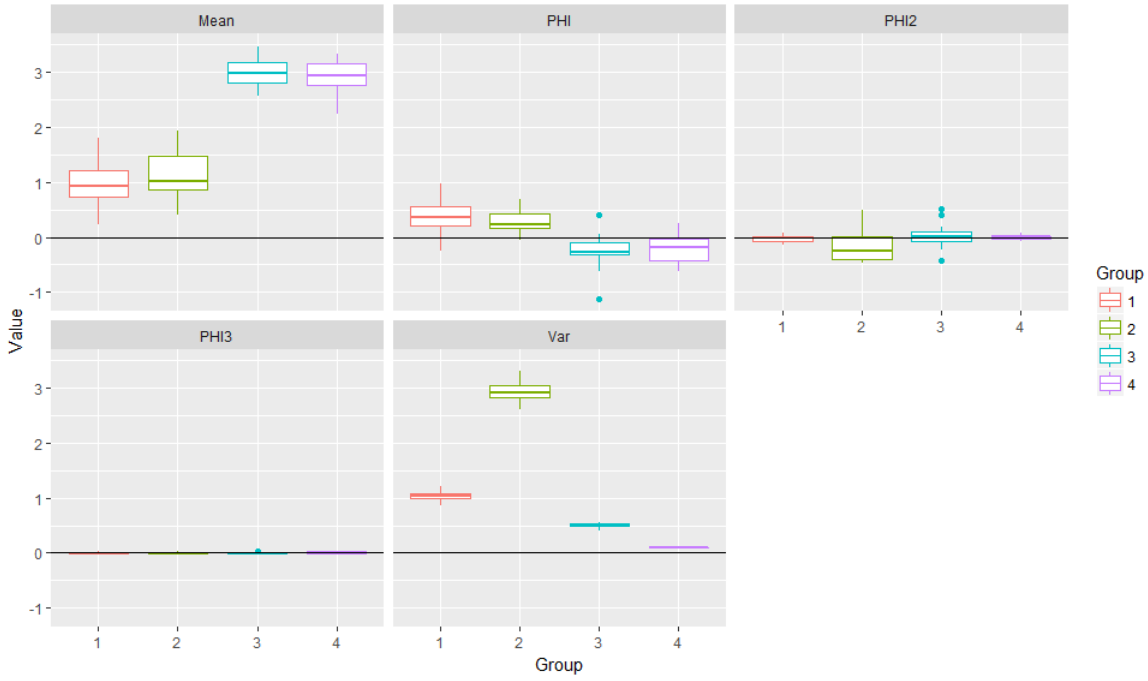



Figure 2: Grouped parameter estimates for all parameters of the model.

$p_i \sim Ber(0.5)$. 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 at a time, in the sample before matching is given in Figure 4. 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 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 5 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 6 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

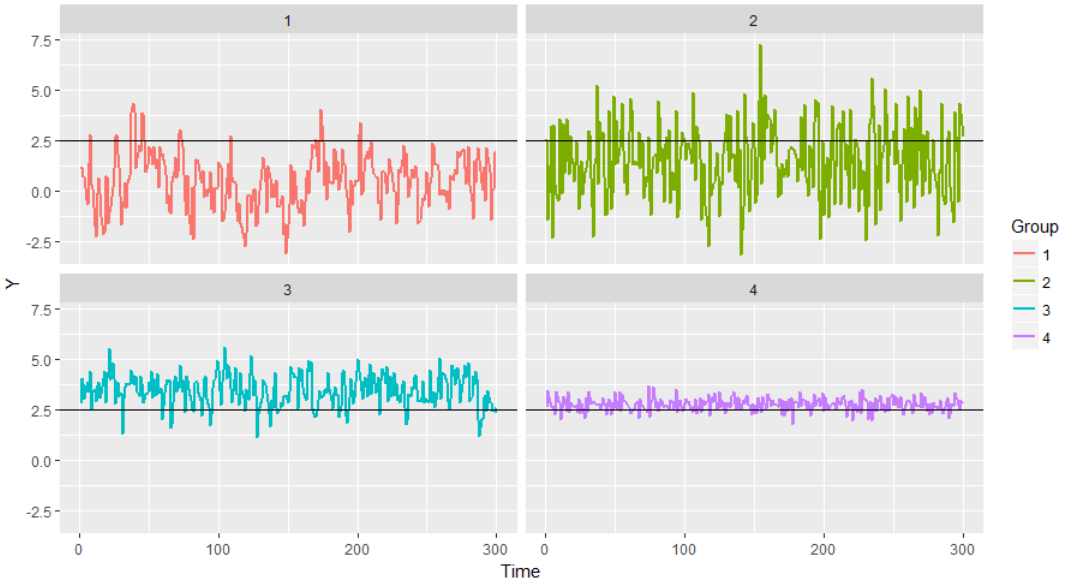



Figure 3: One example process from each of the four groups. The solid line is given as a level reference point.

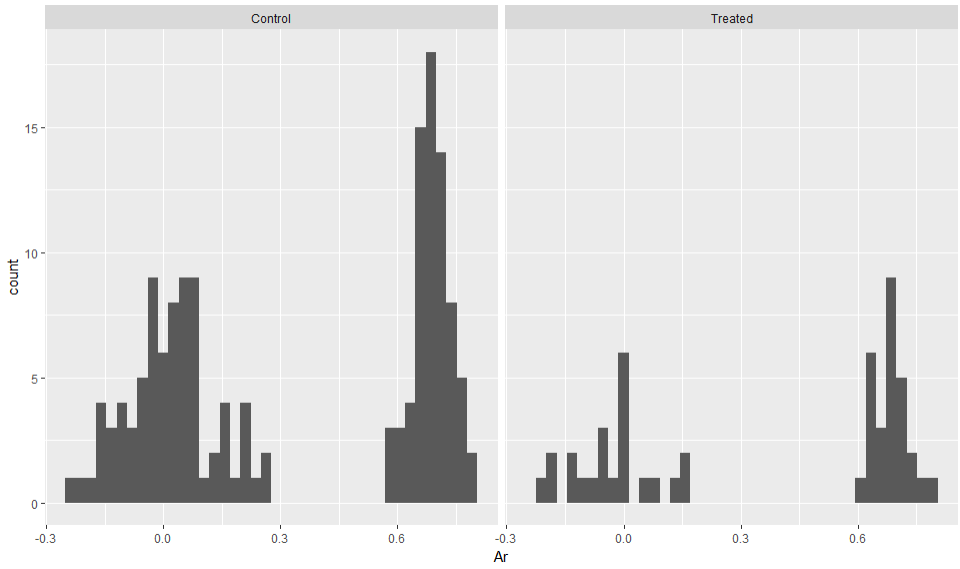



Figure 4: Distribution of autoregressive coefficients in the treatment and control pool.

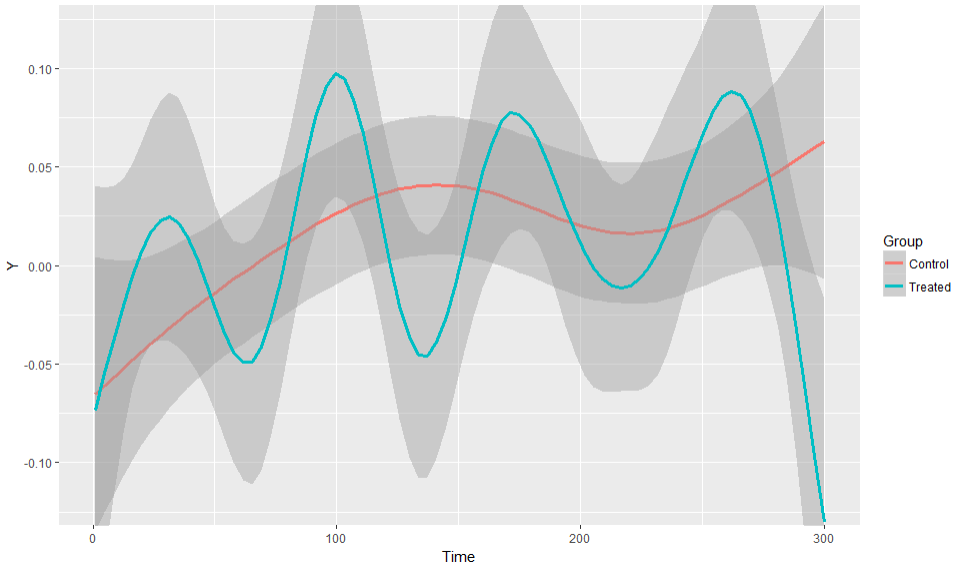



Figure 5: Balance checked against the pre-treatment within-observational unit mean only.

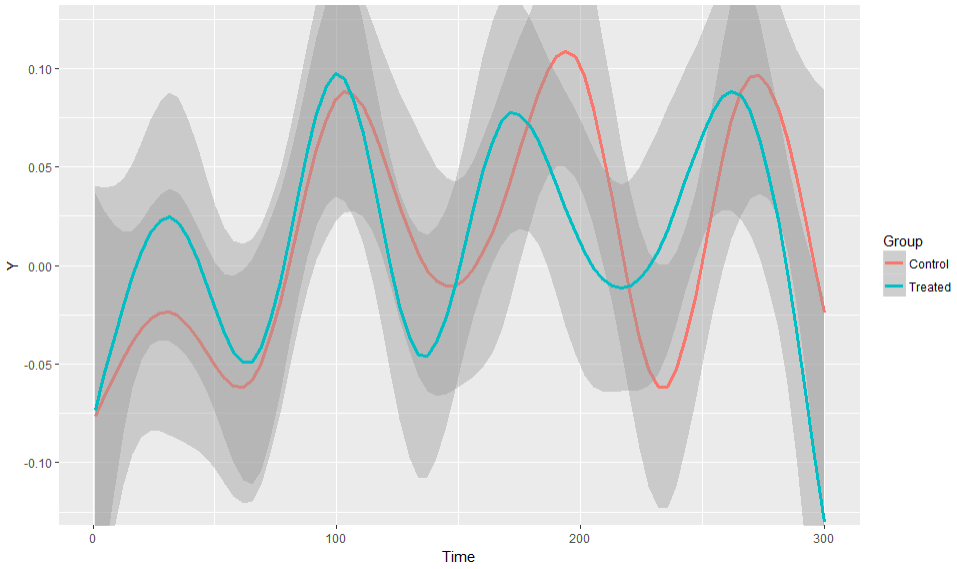



Figure 6: Balance checked against the estimated pre-treatment autoregressive coefficient and within-observational unit mean.

pre-treatment processes drastically. The results illustrate the potential importance of utilizing the pre-treatment data fully.

# List of Figures

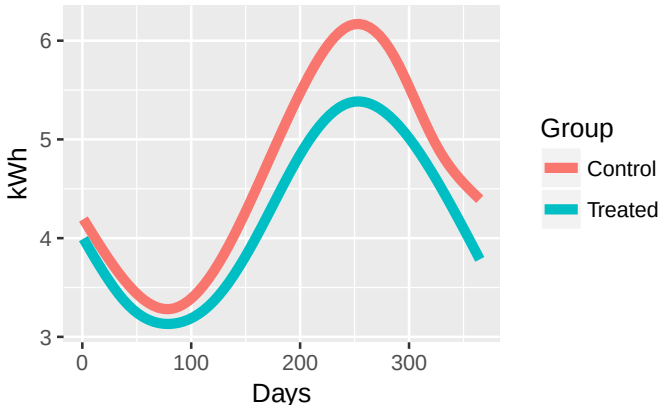



Figure 7: Smoothed pre-treatment electricity consumption average for the cities of Sandviken (treated) and Sundvall (control) the year before treatment. The time series starts May 1 and ends April 30 the following year.

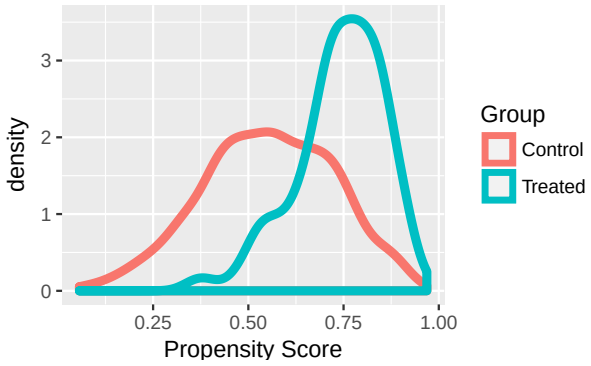



Figure 8: Overlap in firms' propensity of being located in Sandviken (being in the treated group).

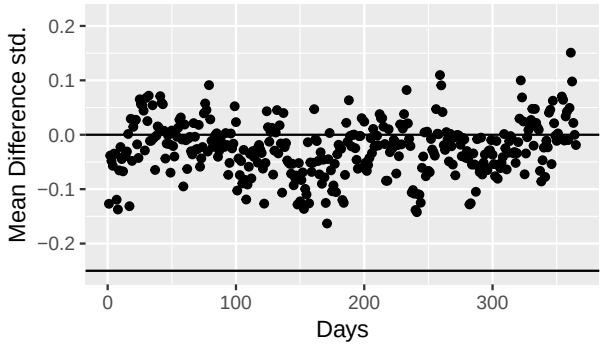
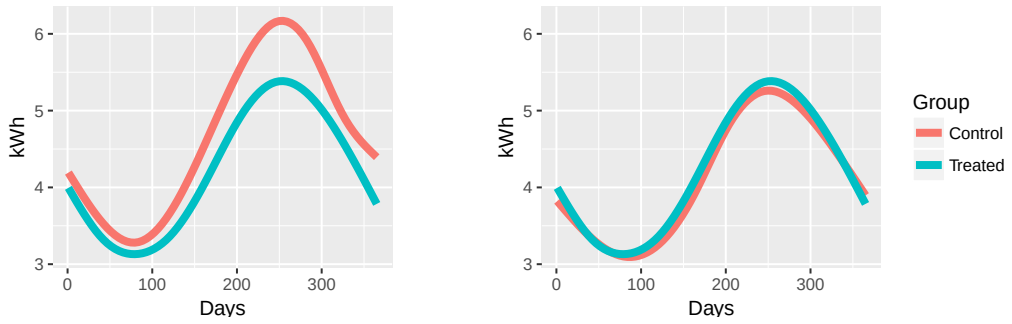



Figure 9: $MDS_t$ for all pre-treatment time periods.




(a) Before matching (b) After matching

Figure 10: 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. The time series starts May 1 and ends April 30 the following year.

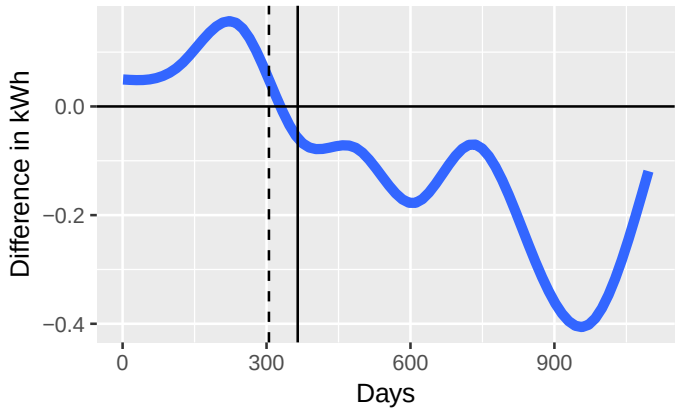



Figure 11: Smoothed mean for the difference between matched treated and control firms before and after treatment. The dashed and solid lines indicate the time periods where the firms in the treatment group were informed about the treatment and when the new tariff was applied, respectively.

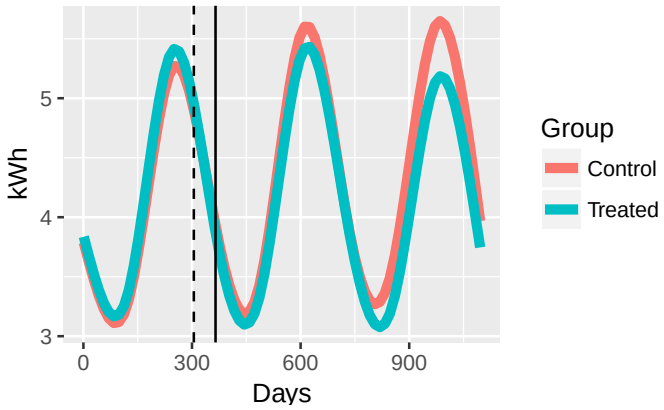



Figure 12: The pre and post-treatment smoothed means of the treated group and the matched control group. The dashed and solid lines indicate the time periods where the firms in the treatment group were informed about the treatment and when the new tariff was applied, respectively.

# List of Tables

| Parameter | Estimate | CI(low) | CI(high) |
|---|---|---|---|
| $\beta_1$ | -0.008 | -0.054 | 0.045 |
| $\beta_2$ | 0.266 | -0.2 | 0.713 |
| $\beta_3$ | 1.878 | -0.387 | 4.314 |
| $\beta_4$ | -0.072 | -0.183 | 0.036 |
| $\beta_8$ | 0.655 | 0.043 | 1.328 |
| $\beta_9$ | -0.958 | -1.674 | -0.308 |
| $\beta_{10}$ | -0.366 | -1.18 | 0.381 |
| $\beta_{11}$ | 2.049 | 0.888 | 3.126 |
| $\beta_{12}$ | -0.764 | -1.561 | 0.025 |
| $\beta_{16}$ | 0.38 | -0.168 | 0.942 |
| $\beta_{17}$ | 0.176 | -0.1 | 0.432 |
| $\beta_{18}$ | 0.333 | 0.033 | 0.647 |
| | | | |
| $\mu_\mu$ | 4.597 | 4.413 | 4.787 |
| $\mu_{\phi_{kWh}}$ | 0.447 | 0.433 | 0.460 |
| $\mu_{\phi_{Te}}$ | -0.042 | -0.046 | -0.039 |
| $\mu_{\gamma_8}$ | 0.405 | 0.383 | 0.426 |
| $\mu_{\gamma_9}$ | 0.368 | 0.345 | 0.391 |
| $\mu_{\gamma_{10}}$ | 0.363 | 0.343 | 0.383 |
| $\mu_{\gamma_{11}}$ | 0.176 | 0.158 | 0.193 |
| $\mu_{\gamma_{12}}$ | 0.054 | 0.037 | 0.070 |
| $\mu_{\log \sigma^2}$ | -0.807 | -0.888 | -0.728 |
| | | | |
| $\sigma^2_\mu$ | 10.387 | 9.532 | 11.274 |
| $\sigma^2_{\phi_{kWh}}$ | 0.057 | 0.052 | 0.062 |
| $\sigma^2_{\phi_{Te}}$ | 0.003 | 0.003 | 0.004 |
| $\sigma^2_{\gamma_8}$ | 0.085 | 0.075 | 0.096 |
| $\sigma^2_{\gamma_9}$ | 0.087 | 0.076 | 0.1 |
| $\sigma^2_{\gamma_{10}}$ | 0.059 | 0.051 | 0.068 |
| $\sigma^2_{\gamma_{11}}$ | 0.043 | 0.038 | 0.05 |
| $\sigma^2_{\gamma_{12}}$ | 0.046 | 0.04 | 0.053 |
| $\sigma^2_{\log \sigma^2}$ | 1.813 | 1.669 | 1.976 |

Table 1: Estimates and 95% credibility intervals for key parameters of the Y and D regressions.