Appl. Math. J. Chinese Univ.
2020, 35(1): 57-83

Statistical Analysis and Evaluation of Macroeconomic Policies: A Selective Review

LIU Ze-qin[1] CAI Zong-wu[2] FANG Ying[1,3] LIN Ming[1,3,*]

Abstract. In this paper, we highlight some recent developments of a new route to evaluate macroeconomic policy effects, which are investigated under the framework with potential outcomes. First, this paper begins with a brief introduction of the basic model setup in modern econometric analysis of program evaluation. Secondly, primary attention goes to the focus on causal effect estimation of macroeconomic policy with single time series data together with some extensions to multiple time series data. Furthermore, we examine the connection of this new approach to traditional macroeconomic models for policy analysis and evaluation. Finally, we conclude by addressing some possible future research directions in statistics and econometrics.

§1 Introduction

It is a common practice for both policymakers and scholars to perform ex post assessment of the causal effects of social-economic programs or policies or public interventions. The statistical econometric analysis and evaluation of such causal effects have attracted considerable attention during the last two decades and by now undergone substantial development and progress. Some scholars, for example, Imbens and Wooldridge (2009) and Cerulli (2015), provided comprehensive surveys of the existing literature. However, most of these surveys mainly focus on microeconomic policy or social program evaluations, which might not be appropriate for macroeconomic policy or intervention evaluation.

There are at least three reasons on why one should pay a special attention to macroeconomic policy analysis and evaluation. First, in some macroeconomic policy evaluations, the main

Received: 2019-03-14. Revised: 2019-08-29.
MR Subject Classification: 62F02, 62G02.
Keywords: Impulse response function, Macroeconomic casual inferences, Macroeconomic policy evaluation, Multiple time series data, Potential outcomes, Treatment effect.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-020-3775-1.
Supported by the National Natural Science Foundation of China (71631004, Key Project), the National Science Fund for Distinguished Young Scholars (71625001), the Basic Scientific Center Project of National Science Foundation of China: Econometrics and Quantitative Policy Evaluation (71988101), the Science Foundation of Ministry of Education of China (19YJA910003) and China Scholarship Council Funded Project (201806315045).
*Correspondence author.

The original version of this article was revised due to a retrospective Open Access order.

interest is to estimate the causal effect of a policy intervention on a single country or region, rather than the average treatment effect in microeconomic policy or social program evaluations. For example, to estimate the effect of political and economic integration of Hong Kong with mainland China on Hong Kong's economic growth; see, for example, Hsiao, Ching and Wan (2012, henceforth HCW), there is only one unit, like Hong Kong, in the treated group. Therefore, in this example, only an individual treatment effect is concerned. Secondly, for macroeconomic policy evaluation, it may be more difficult to find a well-matched control group for the treatment unit. For example, it is not easy to match China with other countries to assess the impact of a Chinese policy on its economy, because China has a unique economic and/or political system. Finally, in a macroeconomic policy evaluation setting, individual observations in different periods are usually serially autocorrelated and treatment in one period often has impacts on the following periods.

Based on the above reasons, the macroeconomic policy evaluation framework should deserve a particular consideration and develope separately. Although there are many other methods, such as dynamic stochastic general equilibrium modeling (abbreviated as DSGE) and structural vector autoregressive (SVAR) models, which are popular to be adopted to analyze macroeconomic policy effects, most of these methods rely heavily on some structural specifications of the entire economic system. In such a way, the validity of their results relies on how precise the assumed economic models are. Therefore, conducting this new type of macroeconomic policy evaluations leads to some precise estimates which can best provide insights to further develop relevant structural models for policy simulations.

The main aim of this paper is to survey recent developments of an alternative route to make causal inferences for macroeconomic policy effects. The key problem in this literature is to estimate statistically the treatment effect of a macroeconomic policy program in a non-experimental setup. Similar to microeconomic policy evaluation, the fundamental challenge is the missing observation problem. That is, at most one outcome in different treatment levels can be observed because the unit can be exposed to only one level of the treatment. This problem is usually addressed by using the so-called counterfactual approach in the statistics literature; see, for example, Rubin (1974, 1977) for details, which is also called the Rubin causal model.

The rest of this paper is organized as follows. Section 2 briefly introduces the basic framework of the Rubin causal model. Section 3 reviews the estimation methods for estimating macroeconomic causal effects using single time series data. Section 4 focuses on the methods when multiple time series data are available. Section 5 concludes with discussions on some open and interesting research questions.

§2 Basic Model Setup

From the early 1970s, Rubin (1973a, 1973b, 1974, 1978, 1979), in a series of his papers, mapped the now dominant approach to the evaluation problem, labeled as Rubin causal model (RCM) by Holland (1986). There are two essential ingredients in RCM: potential outcomes and assignment mechanism. The potential outcomes framework is the hallmark of modern econometric analysis of treatment effect. This section gives a concise introduction to the basic model setup under the RCM framework, which can be used to analyze a macroeconomic policy's simultaneous treatment effects under certain assumptions.

2.1 Potential Outcomes

Suppose that we wish to estimate the effect of a macroeconomic policy on economic growth using observations on T periods, indexed by $t = 1, \ldots, T$. The policy is carried out in some of these periods, not in others. Let Y_t denote economic growth and D_t indicate whether a policy is implemented at period t, with $D_t = 1$ if it is implemented and $D_t = 0$ if it is not. D_t can be a binary variable, a multiple variable (see, e.g., Angrist and Imbens, 1995), or a continuous variable (Imbens, 2000, and Hirano and Imbens, 2004). If not otherwise specified, we assume D_t is binary in this paper.

Definition 2.1 (Potential Outcomes)**.** *Potential outcomes are pairs of outcomes for the same period given different levels of treatment, denoted by $Y_t(0)$ and $Y_t(1)$. $Y_t(0)$ denotes the outcome that would be realized at period t if the policy is not implemented. Similarly, $Y_t(1)$ denotes the outcome that would be realized at period t if the policy is implemented.*

Rubin (1973a, 1973b) defined novelly the causal effect as the difference between this pair of outcomes, i.e., $Y_t(1) - Y_t(0)$, the treatment effect of the policy at period t. This framework counters a fundamental missed observation problem, since at period t, the policy can either be implemented or not be implemented, but not both. Thus, only one of these two potential outcomes can be realized. However, prior to the assignment being determined, both are potentially observable and hence labeled as potential outcomes. If the policy is carried out at period t, $Y_t(1)$ is realized, and $Y_t(0)$ is a counterfactual outcome. On the other hand, if the policy is not carried out at period t, $Y_t(0)$ is realized, and $Y_t(1)$ is a counterfactual outcome. In casual inference, it is usual to make few assumptions other than the stable unit treatment value assumption (SUTVA), which is that one unit's outcomes are unaffected by another unit's treatment assignment.

The potential outcomes $Y_t(0)$ and $Y_t(1)$ have the following relationship with the realized outcome Y_t:

$$Y_t = \begin{cases} Y_t(0), & D_t = 0; \\ Y_t(1), & D_t = 1. \end{cases} \quad \text{Equivalently,} \quad Y_t = D_t Y_t(1) + (1 - D_t) Y_t(0), \tag{1}$$

which is the so-called potential outcome model (POM), and it is the fundamental relation linking unobservable and observable outcomes. The potential outcomes framework has five main advantages over the traditonal models defined by realized outcomes; see Imbens and Wooldridge (2009) for more details.

2.2 Parameters of Interest

Average treatment effect (ATE) and average treatment effect on the treated (ATET) are two prominent parameters of interest in the literature. If Y_t and D_t are strictly stationary, then

$$\text{ATE} = E[Y_t(1) - Y_t(0)], \quad \text{and} \quad \text{ATET} = E[Y_t(1) - Y_t(0)|D_t = 1].$$

Average treatment effect is defined over the entire period, and average treatment effect on the treated is averaged over the subset of treated periods. Another popular parameter is conditional average treatment effect (CATE):

$$\text{ATE}(\mathbf{x}) = E[Y_t(1) - Y_t(0)|\mathbf{X}_t = \mathbf{x}], \quad \text{and} \quad \text{ATE} = E_{\mathbf{x}}[\text{ATE}(\mathbf{x})],$$

where $\mathbf{X}_t$ is a vector of conditional variables (strictly stationary, and not affected by the intervention). Recently, a quantile treatment effect (QTE) has been popularly studied and extensively applied in the economics literature and real applications; see Abadie, Angrist and Imbens (2002), Chernozhukov and Hansen (2005), and Koenker and Bassett (1978). For a given $0 < q < 1$, QTE is defined as

$$\text{QTE}_q = F^{-1}_{Y(1)}(q) - F^{-1}_{Y(0)}(q), \tag{2}$$

where $F_{Y(d)}(\cdot)$ is the distribution of $Y_t(d)$ for $d = 1$ and 0. The conditional quantile treatment effect (CQTE) and conditional quantile treatment effect on the treated (CQTET) can be defined in the same way.

2.3 Assignment Mechanism

The assignment mechanism is the mechanism by which policies are carried out or not at period time t. It is expressed as the conditional probability of receiving the treatment given observed covariates and potential outcomes, i.e., $P(D_t = 1|\mathbf{X_t}, Y_t(0), Y_t(1))$, which is a function of potential outcomes and observed covariates. According to the relationship between potential outcomes and the assignment mechanism, three classes of assignment mechanism are distinguished commonly as: randomized experiments, selection on observables, and selection on unobservables, which are described next in detail.

The first class of assignment mechanisms is randomized experiments. In randomized experiments, $D_t \perp\!\!\!\perp \big(Y(0), Y(1)\big)$, i.e., the probability of assignment to treatment, a known function of covariates, is independent with potential outcomes. A typical example is a completely randomized experiment where we randomly choose $T_1 < T$ periods for treated and $T_0 = T - T_1$ for untreated, and each period has equal probability of being in the control group or the treatment group. Under the assumptions of iid (independently and identically distributed) and SUTVA, the well-known difference-in-means (DIM) estimator in classical statistics defined in (3) can be applied to recover the average treatment effect:

$$\widehat{\text{DIM}} = \frac{1}{T_1}\sum_{t=1}^{T} D_t Y_t - \frac{1}{T_0}\sum_{t=1}^{T}(1 - D_t)Y_t. \tag{3}$$

In reality, the decision-making process for a macroeconomic policy is complex. To carry out a policy, the authority would gather detailed and well-documented data and information, outline the problems, list goals and objectives, reveal potential gains and losses, etc. Thus, it is impossible for the authority to implement a macroeconomic policy randomly.

The second class of assignment mechanisms is selection on observables, which is also referred to as unconfounded assignment, presented by Rosenbaum and Rubin (1983). This mechanism assumes that factors determining whether a policy should be put into effect are observable. In this setting, $D_t \perp\!\!\!\perp \big(Y_t(0), Y_t(1)\big)|\mathbf{X}_t$, i.e., given $\mathbf{X}_t$, the assignment probability does not depend on the potential outcomes. Different from randomized experiments, the assignment probability is no longer assumed to be a known function of covariates.

Thus far, a majority of statistical/econometric methods for evaluation have been developed in such a setup under the assumptions of iid and SUTVA. The most popular techniques in this vein include, but not limited to, regression-adjustment (RA) in Heckman, Ichimura, and Todd (1997, 1998), matching in Abadie and Imbens (2006, 2008, 2011), matching on propensity-score in Abadie and Imbens (2016), Dehejia and Wahba (2002), and Caliendo and Kopeinig (2010), re-weighting on propensity-score inverse probability (IPW) in Rosenbaum and Rubin (1983), Dehejia and Wahba (1999), Hahn (1998), Hirano, Imbens and Ridder (2003), Brunell and DiNardo (2004), Wooldridge (2010), Imbens (2004), and Lunceford and Davidian (2004), and double-roust estimator (DR) in Robins and Rotnitzky (1995), Robins, Rotnitzky and Zhao (1995), van der Lann and Robins (2003), and Wooldridge (2007), and the references therein.

Finally, the third one is selection on unobservables, containing cases apart from the random experiment and selection on observables. In this setting, whether a policy is implemented depends on not only observable factors but also unobservable factors. Thus, the assignment probability has some dependence on potential outcomes. Therefore, given $\mathbf{X}_t$, we cannot identify the treatment effects, and we require further assumptions for the analysis. There is no general solution for this situation.

§3 Evaluation Methods with Single Time Series Data

The model setup in section 2 is borrowed from microeconomic policy evaluation but is still suitable for assessing a macroeconomic policy's simultaneous effects on outcome variables with some additional assumptions. However, subsequent movements of the outcome variables responding to the policy are needed for a careful consideration in a macroeconomic policy evaluation framework. Angrist and Kuersteiner (2011), Angrist, Jordà and Kuersteiner (2018), Bojinov and Shephard (2019), Jordà and Taylor (2016), and Kuersteiner, Phillips and Villamizar-Villegas (2018) extended the modern statistical/econometric analysis of treatment effects to a time series context for macroeconomic data.

3.1 A Road Map

In cross-sectional analysis, we have data usually consisting of multiple units in one period, some being treated but others not. Since we cannot observe the pairs of potential outcomes simultaneously, the individual treatment effect cannot be identified. By comparing the outcomes of treatment group with control group, we obtain the consistent estimation of average treatment effect. However, in time series analysis, we only have data of a single unit in multiple periods. Instead of the average treatment effect, the concepts of potential outcome paths introduced by Robins (1986) and the dynamic potential outcomes introduced by Angrist and Kuersteiner (2011) and Angrist et al. (2018) become the heart of this literature.

3.1.1 Potential Outcome Paths

Let $D_{1:t} = (D_1, \cdots, D_t)$ denote the random "treatment path" and $d_{1:t}$ be a realization of $D_{1:t}$. By assuming that D_t is binary, we have 2^t treatment paths at time t. Bojinov and Shephard (2019) defined potential outcomes on the treatment path $D_{1:t}$ as follows.

Definition 3.1 (Potential Outcome Paths)**.** *The set of 2^t potential outcomes at time t is:*

$$Y_t(\cdot) = \{Y_t(d_{1:t}) : d_{1:t} \in \{0,1\}^t\},$$

and the potential path for the treatment path $d_{1:t}$ is

$$Y_{1:t}(d_{1:t}) = \{Y_1(d_{1:1}), Y_2(d_{1:2}), \cdots, Y_t(d_{1:t})\}.$$

Therefore, the potential outcome paths have the following relationship with observed outcome path:

$$Y_{1:t} = \sum_{d\in\{0,1\}^t} 1\{D_{1:t} = d\}Y_{1:t}(d), \quad t = 1, \cdots, T. \tag{4}$$

To see the detailed paths of potential outcome, let us consider the case that $T = 3$ with the potential outcome paths indicated in Figure 1. When $t = 1$, there are $2^1 = 2$ potential

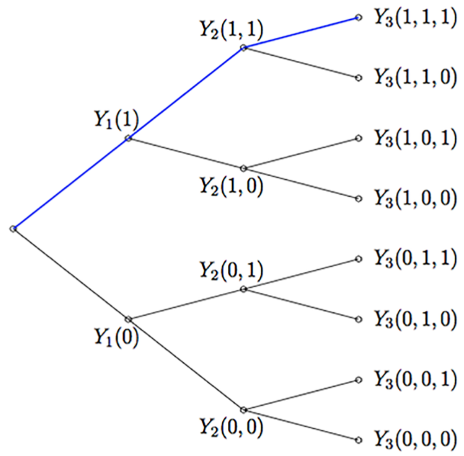

Figure 1: Potential outcome paths when T=3

outcomes $Y_1(0)$ and $Y_1(1)$. When $t = 2$, there are $2^2 = 4$ potential outcomes $Y_2(1,1)$, $Y_2(1,0)$,

$Y_2(0,1)$, and $Y_2(0,0)$. Assuming that the realized treatment path is $d_{1:3} = (1,1,1)$, then, the potential path for $d_{1:3}$ is $Y_{1:3}(d_{1:3}) = \{Y_1(1), Y_2(1,1), Y_3(1,1,1)\}$. According to (4), the observed outcome path is just the potential outcome path of observed treatment path $d_{1:t}$, when t=3, $Y_{1:t} = Y_{1:t}(d_{1:t}) = \{Y_1(1), Y_2(1,1), Y_3(1,1,1)\}$.

Given the potential outcome paths, any comparison of potential outcome paths at fixed period t, has a causal interpretation. Therefore, we can define a large number of treatment effects; see, for example, Bojinov and Shephard (2019).

Definition 3.2 (General Treatment Effect)**.** *For paths $d_{1:t}$ and $d'_{1:t}$, the t-th treatment effect is:*

$$\tau_t(d_{1:t}, d'_{1:t}) = Y_t(d_{1:t}) - Y_t(d'_{1:t}),$$

and the average treatment effect of the paths $d_{1:t}$ and $d'_{1:t}$ is:

$$\bar{\tau}(d_{1:T}, d'_{1:T}) = \frac{1}{T}\sum_{t=1}^{T} \tau_t(d_{1:t}, d'_{1:t}).$$

3.1.2 Dynamic Potential Outcomes

In the DSGE and SVAR frameworks, one of the main focuses is the impulse response function (IRF), which is the contemporaneous and subsequent movements of outcome variables responding to an exogenous policy shock at time t. Angrist and Kuersteiner (2011) and Angrist et al. (2018) proposed the new concept of dynamic treatment effects as the corresponding part of the impulse response function.

Definition 3.3 (Dynamic Potential Outcomes)**.** *Given t, l, and ψ, potential outcomes $\{Y^{\psi}_{t,l}(d); d \in \mathcal{D}\}$ are defined as the set of values the observed outcome variable Y_{t+l} would take on if $D_t = d$, with $d \in \mathcal{D} = \{0, \cdots, j, \cdots, J\}$. ψ denotes the policy regime, which takes values in a parameter space Ψ.*

Let $Y_{t:L} = (Y_t, Y_{t+1}, \cdots, Y_{t+L})'$ is the path of the observed outcome from t to $t+L$, $Y^{\psi}_{t:L}(d) = (Y^{\psi}_{t,0}(d), Y^{\psi}_{t,1}(d), \cdots, Y^{\psi}_{t,L}(d))'$ is the path of potential outcome from t to $t+L$. The relationship of $Y_{t:L}$ and $Y^{\psi}_{t:L}(d)$ is:

$$Y_{t:L} = \sum_{d \in \mathcal{D}} Y^{\psi}_{t:L}(d) 1\{D_t = d\}.$$

Based on the definition above, the causal effect of policy choice j is $Y^{\psi}_{t,l}(j) - Y^{\psi}_{t,l}(0)$, where $d = 0$ is a benchmark policy. Since individual causal effects can never be observed, Angrist and Kuersteiner (2011) and Angrist et al. (2018) thus defined the average treatment effects, labelled as the dynamic treatment effects in this paper.

Definition 3.4 (Dynamic Treatment Effects)**.** *Given l, the expected response of Y_{t+l} to policy choice j is:*

$$\theta_{l,j} = E\left[Y^{\psi}_{t,l}(j) - Y^{\psi}_{t,l}(0)\right], \tag{5}$$

and the collection of treatment effects from t to $t+L$ is:

$$\theta_j = E\left[Y^{\psi}_{t:L}(j) - Y^{\psi}_{t:L}(0)\right], \tag{6}$$

where $\theta_j = (\theta_{0,j}, \theta_{1,j}, \cdots, \theta_{L,j})$.

Similar to the traditional impulse response analysis in macroeconomics, the definition for θ focuses on a single policy shock on contemporaneous and subsequent outcomes. Indeed, it compares two different paths, by controlling the path of D_t before t, and assuming that $d_{t+1:t+L|t}$ is constant:

$$\theta_{l,j}(j,0) = E\left[Y^{\psi}_{t,l}(d_{1:t-1}, d_t = j, d_{t+1}, \cdots, d_{t+l}) - Y^{\psi}_{t,l}(d_{1:t-1}, d_t = 0, d_{t+1}, \cdots, d_{t+l})\right],$$

and

$$\theta_{j}(j,0) = E\left[Y^{\psi}_{t:L}(d_{1:t-1}, d_t = j, d_{t+1}, \cdots, d_{t+L}) - Y^{\psi}_{t:L}(d_{1:t-1}, d_t = 0, d_{t+1}, \cdots, d_{t+L})\right].$$

Note that both the timing of policy adoption and the horizon matter for $Y^{\psi}_{t,l}(j)$. For example, $Y^{\psi}_{t,l}(j)$ may be different from $Y^{\psi}_{t+1,l-1}(j)$, even though both describe outcomes in period $t+l$, because $Y^{\psi}_{t,l}(j)$ measures the effect of $D_t = j$ on the outcome at time $t+l$, $Y^{\psi}_{t+1,l-1}(j)$ measures the effect of $D_{t+1} = j$ on the outcome at time $t+l$, and for $Y^{\psi}_{t,l}(j)$, we do not limit $D_{t+1} = j$.

3.2 Connection to Classic Macroeconomic Models

To understand how the dynamic treatment effect defined above corresponds to a nonlinear impulse response function, we illustrate their connection using the following model:

$$D_t = D(z_t, \psi, u_t), \quad \text{and} \quad Y_{t+l} = F_{t,l}(z_t, D_t, \eta_{t,l}),$$

where z_t is a vector, including covariate $\mathbf{X}_t$, and some lags of Y_t as well as D_t, u_t is unobservable idiosyncratic information, which is assumed to be independent of potential outcomes, and $\eta_{t,l}$ is white noise. Considering a perturbation ϵ_t of D_t at time t, the nonlinear impulse response of Y_{t+l} is defined as:

$$\delta_l(z_t, D_t, \epsilon_t) = F_{t,l}(z_t, D_t + \epsilon_t, \eta_{t+l}) - F_{t,l}(z_t, D_t, \eta_{t+l}). \tag{7}$$

If $D_t + \epsilon_t = j$, and $D_t = 0$, it is clear that (7) is just the dynamic treatment effect as defined in (5). Specifically, Angrist and Kuersteiner (2011) made this strong link to linear impulse response function in their paper. Suppose in a SVAR model,

$$\Gamma_0 \begin{bmatrix} Y_t \\ D_t \\ \mathbf{X}_t \end{bmatrix} = -\Gamma(L) \begin{bmatrix} Y_t \\ D_t \\ \mathbf{X}_t \end{bmatrix} + \begin{bmatrix} \eta_t \\ \epsilon_t \\ \zeta_t \end{bmatrix},$$

where L is the lag operator, ϵ_t represents the policy shocks in which we are interested, η_t and ζ_t represents other innovations. Assume $C(L) = (\Gamma_0 + \Gamma(L))^{-1}$ exists, then,

$$\begin{bmatrix} Y_t \\ D_t \\ \mathbf{X}_t \end{bmatrix} = C(L) \begin{bmatrix} \eta_t \\ \epsilon_t \\ \zeta_t \end{bmatrix}.$$

We further suppose that the outcome is determined by $Y_t = \phi_1 D_t + \phi_2 \eta_t$, and the policy rule is given by $D_t = \psi Y_{t-1} + \epsilon_t$, where Y_t has a moving average representation:

$$Y_t = \sum_{l=0}^{\infty} \rho_l \epsilon_{t-l} + \sum_{l=0}^{\infty} \gamma_l \eta_{t-l}.$$

Here, ρ_l is the impulse response function for output in response to the interested policy shocks. In this setup, potential outcome for $D_t = j$ is defined as:

$$Y_{t,l}^{\psi}(j) = \rho_l(j - \psi Y_{t-1}) + \sum_{i=0, i\neq l}^{\infty} \rho_i \epsilon_{t+l-i} + \sum_{i=0}^{\infty} \gamma_i \eta_{t+l-i}.$$

The associated causal effect of this policy change is $Y_{t,l}^{\psi}(j) - Y_{t,l}^{\psi}(0) = \rho_l j$, which means how the outcome Y_{t+j} would change if the policy changing from 0 to j, by assuming that everything else is unchanged. Obviously, this is precisely the idea of the impulse response function. Moreover, compared to SVAR and DSGE models, the identification and estimation of dynamic treatment effects require no need to specify the structural process of Y_t and only focuses on the policymaking process, alleviating the crucial model misspecification problem faced in these main macroeconomic models and providing a more flexible tool for the analysis and evaluation of macroeconomic causal relationships.

3.3 Inferences with Selection on Observables

Selection on observables is a natural starting point for identification and estimation of the causal inferences. Angrist and Kuersteiner (2011) and Angrist et al. (2018) defined the assumption of selection on observables in time series framework as follows:

Assumption 3.1 (Selection on Observables)**.** *Given covariate z_t, $Y_{t,l}^{\psi}(d) \perp\!\!\!\perp D_t \,|\, z_t$, for all $l \geq 0$, $d \in \mathcal{D}$, with ψ fixed, and $\psi \in \Psi$.*

This assumption is also known as conditional independence assumption (CI). It also can be expressed as the following form:

$$P(D_t = d | D_{1:t-1} = d_{1:t-1}, Y_{1:T}^{\psi}(\cdot), z_t) = P(D_t = d | D_{1:t-1} = d_{1:t-1}, Y_{1:t-1}^{\psi}(d_{1:t-1}), z_t). \quad (8)$$

In view of (8), it is clear see that the assumption means that our treatments only depend on past observables of potential outcomes and are not influenced by the corresponding period value or by future values of potential outcomes. This assumption is simply the "sequential randomization" assumption as in Robins (1994), Robins, Greenland and Hu (1999), Abbring and van den Berg (2003), and Lok (2008). Under this assumption, we have the following relationship:

$$\theta_{l,j} = E\Bigg[E\Big[Y_{t,l}^{\psi}(j) - Y_{t,l}^{\psi}(0) | z_t\Big]\Bigg] = E\Bigg[E\Big[Y_{t,l} | D_t = j, z_t\Big] - E\Big[Y_{t,l} | D_t = 0, z_t\Big]\Bigg]. \quad (9)$$

Then, $\theta_{l,j}$ can be identified, since (9) is cast in terms of observable conditional means. Therefore, to estimate $\theta_{l,j}$, we can draw lessons from the iid case. Indeed, Angrist et al. (2018) applied the IPW estimator to time series data to estimate the causal effect of monetary policy on macroeconomic aggregates, while Jordà and Taylor (2016) adopted the doubly robust estimator to time series data to evaluate the fiscal policy. They are elaborated in detail in the next subsections.

3.3.1 Estimation: IPW Method

Define $p^j(z_t,\psi)=P(D_t=j|z_t)$ as the policy propensity score, with $p^j(z_t,\psi)$ being a flexible parametric model with parameter ψ determined by the policy regime, where z_t is a covariate. Further, assume that $0<\underline{p}\le p^j(z_t,\psi)\le\overline{p}<1$, which is a common assumption in the literature. Then, the selection on observables assumption implies that

$$E\left[Y_{t,l}1\{D_t=j\}|z_t\right]=E\left[Y_{t,l}^{\psi}(j)|z_t\right]p^j(z_t,\psi),$$

which leads to

$$\theta_{l,j}^{IPW}=E\left[Y_{t,l}\left(\frac{1\{D_t=j\}}{p^j(z_t,\psi)}-\frac{1\{D_t=0\}}{p^0(z_t,\psi)}\right)\right]. \tag{10}$$

This re-weighting scheme was first proposed by Horvitz and Thompson (1952) and adapted by Hanh (1998) and Hirano et al. (2003) for treatment effect estimation in cross-sectional cases.

Denote

$$w_{t,j}(\psi)=\frac{1\{D_t=j\}}{p^j(z_t,\psi)}-\frac{1\{D_t=0\}}{p^0(z_t,\psi)}.$$

Then, (10) can be written as $\theta_{l,j}^{IPW}=E[w_{t,j}(\psi)Y_{t,l}]$, a weighted expectation of $Y_{t,l}$, and clearly, $\theta_{l,j}^{IPW}$ can be estimated by a two-step procedure. First, estimate the parameter ψ in $p^j(z_t,\psi)$. Secondly, take $\hat{\psi}$ to $w_{t,j}(\hat{\psi})$ and then obtain the estimation of $\theta_{l,j}^{IPW}$. In a correctly specified model, the weight $w_{t,j}(\psi)$ has a mean zero and is uncorrelated with z_t. To ensure this condition is true, Angrist et al. (2018) suggested weight $Y_{t,l}$ by $\ddot{w}_{t,j}$, where $\ddot{w}_{t,j}=w_{t,j}(\hat{\psi})-E[w_{t,j}(\hat{\psi})|z_t]$, is the residual from a regression of $w_{t,j}(\hat{\psi})$ on z_t and a constant. Therefore, the estimators of $\theta_{l,j}^{IPW}$ and θ_j^{IPW} are given by

$$\hat{\theta}_{l,j}^{IPW}=\frac{1}{T}\sum_{t=1}^{T}\ddot{w}_{t,j}Y_{t+l},\quad\text{and}\quad\hat{\theta}_j^{IPW}=\frac{1}{T}\sum_{t=1}^{T}\ddot{w}_{t,j}Y_{t:L},$$

respectively. In practice, when $p^j(z_t,\hat{\psi})$ is too small, the weight would be extremely large. Then, the estimation of $\theta_{l,j}^{IPW}$ is dominated by these outliers. To overcome this difficulty, the trimmed method should be used by dropping observations with $p^j(z_t,\hat{\psi})<0.025$ when $1\{D_t=j\}=1$. Note that when the dimension of z_t is not very high, the nonparametric estimate of $p^j(z_t)$ can be applied to avoid a possible misspecification of parametric form $p^j(z_t,\psi)$.

3.3.2 Estimation: Doubly Robust Approch

It is well documented in the literature that if the parametric form of $p^j(z_t,\psi)$ is correctly specified, the IPW estimators $\hat{\theta}_{l,j}^{IPW}$ and $\hat{\theta}_j^{IPW}$ are consistent. To circumvent the possible misspecification of $p^j(z_t,\psi)$, one can let $p^j(z_t)=P(D_t=j|z_t)$ be unspecified as a nonparametric form so that the IPW estimators are consistent. However, the dimension of z_t is high in practice, the nonparametric estimate of $p^j(\cdot)$ may suffer from the "curse of dimensionality". To overcome these difficulties, Robins and Rotnitzky (1995) proposed the so-called doubly robust estimator (DR).

Jordà and Taylor (2016) adopted the doubly robust estimator to assess the impact of fiscal austerity on subsequent economic growth. The doubly robust estimator has an attractive property that consistency of the estimated average treatment effect only requires either the propensity score model or the regression model of the potential outcome to be correctly specified.

The doubly robust estimator used by Jordà and Tapylor (2016) is indeed an AIPW (augmented IPW), which has the smallest asymptotic variance within the doubly robust class as in Robins and Rotnitzky (1995), Robins et al. (1995), and Lunceford and Davidian (2004). It is defined by

$$\theta_{l,j}^{AIPW} = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1\{D_t=j\}}{p^j(z_t,\hat{\psi})}Y_{t+l} + \left(1-\frac{1\{D_t=j\}}{p^j(z_t,\hat{\psi})}\right)\mu_j(z_t,\hat{\beta}_j)\right] - \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1\{D_t=0\}}{p^0(z_t,\hat{\psi})}Y_{t+l} + \left(1-\frac{1\{D_t=0\}}{p^0(z_t,\hat{\psi})}\right)\mu_0(z_t,\hat{\beta}_0)\right],$$

where $\mu_j(z_t,\hat{\beta}_j) = E(Y_{t+l}|D_t = j, z_t)$ and $\mu_0(z_t,\hat{\beta}_j) = E(Y_{t+l}|D_t = 0, z_t)$, which are two regression estimators of potential outcomes $Y_{t,l}^{\psi}(j)$ and $Y_{t,l}^{\psi}(0)$, respectively.

The AIPW estimator is the basic IPW estimator plus an adjustment of the weighted average of the two regression estimators. The adjustment term stabilizes the estimator when the propensity scores become close to zero or one (Glynn and Quinn, 2010). In the cross-sectional context, when the propensity score and the regression function are modeled correctly, the AIPW can achieve the semiparametric efficiency bound. If the propensity score $p^j(\cdot)$ is modeled correctly, the AIPW estimator has the asymptotic variance smaller than or equal to that for the simple IPW estimator. If the regression model $\mu_j(\cdot)$ is modeled correctly, the AIPW estimator should have the asymptotic variance greater than or equal to that for the simple regression estimator, but it gives protection in the event that $\mu_j(\cdot)$ is misspecified.

3.3.3 Estimation: IPW Combined with Machine Learning

Liu, Cai and Fang (2019) combined machine learning with IPW to investigate the effects of monetary policy and macro-prudential policy on financial stability, price stability and economic growth in China. Liu et al. (2019) adopted a data-driven method to deal with estimation problems with a large set of covariates. One of the policy propensity score models they defined was:

$$p^j(z_t,\psi) = \frac{e^{\psi_j z_t}}{\sum_{k=0}^{J} e^{\psi_k z_t}}, \qquad j = 0, 1, \cdots, J,$$

where z_t contained over 700 variables, including lages of $\mathbf{X}_t$, Y_t and D_t. By adding L_1 type or L_2 type penalty, $\hat{\psi}$ was estimated to minimize the following objective function:

$$L_q(\psi,\lambda) = -\frac{1}{T}\left[\sum_{t=1}^{T}\sum_{j=0}^{J} 1\{D_t=j\}\left(\psi_j z_t - \log\sum_{k=0}^{J} e^{\psi_k z_t}\right)\right] + \lambda_1\|\psi\|_q^q,$$

where $\|\psi\|_q$ was the L_q normal with, say, $q = 1$ or 2. In addition, they also adopted support vector machine (SVM) and random forests (RF) to fit policy propensity score model. Through utilizing the information from a large number of series and eliminating the arbitrary reliance on a small number of predefined variables, Liu et al. (2019) found that machine learning methods could help predict the Chinese government's policy decision well.

With policy propensity score estimated by machine learning methods, Liu et al. (2019)

estimated the dynamic treatment effects of monetary policy and macro-prudential policy on financial stability, price stability and economic growth in China according to (10). The findings were that easy monetary policy alone could reduce financial systematic risk, while tight monetary policy alone could reduce CPI/PPI, and slow down the economic growth rate; macro-prudential policy alone had little effect on financial and economic outcomes;when monetary and macro-prudential policies were employed simultaneously, macro-prudential policy would offset the effect of monetary policy on systematic financial risk; tight macro-prudential policy could neutralize the negative effect of tight monetary policy on economic growth; tight macro-prudential policy would strengthen tight monetary policys effect on PPI, but offset the impact of tight monetary policy on CPI; see Liu et al. (2019) for more details.

3.3.4 Testing for The Selection on Observables Assumption

As one can see from (9) that the selection on observables assumption plays a vital role in identifying the parameters of interest $\theta_{l,j}$. However, this assumption may be violated in practice if there exist unobserved confounders which affect both the potential outcomes $Y_{t,l}^{\psi}(j)$ and the treatment variable D_t. If this assumption does not hold, both the IPW and doubly robust estimators discussed above will deliver inconsistent estimation in general for the parameters of interest $\theta_{l,j}$. Thus, it is desirable to formally test whether the selection on observables assumption holds or not. Recently, Cai, Fang, Lin and Tang (2019) proposed a new method to test the selection on observables assumption for both cross-sectional and time series data settings. Specifically, their method relies on the existence of an auxiliary variable which is correlated to potential outcomes but is independent of the treatment variable given on potential outcomes and some other observable covariates. Formally, their procedure requires the available auxiliary variable ω_t satisfying the following assumption.

Assumption 3.2. *There exists an auxiliary variable ω_t such that $\omega_t \perp\!\!\!\perp D_t \mid \left(Y_{t,l}^{\psi}(j), z_t\right)$ for all $l \geq 0, d$, with ψ fixed, and $\psi \in \Psi$.*

Under this assumption, Cai et al. (2019) showed that the selection on observables assumption implied that $E(D_t|\omega_t, z_t) = E(D_t|z_t)$, so that testing for the selection on observables assumption could be transformed into testing the insignificance of the auxiliary variable ω_t, which could be formulated as the following testing hypothesis

$$H_0 : E(D_t|\omega_t, z_t) = E(D_t|z_t)$$

a.s. (almost surely) versus

$$H_1 : E(D_t|\omega_t, z_t) \neq E(D_t|z_t) \text{ on a set with positive measure.}$$

Let $\varpi_t = (z_t', \omega_t')' \in \mathbb{R}^p$, where $p = q + r$ with q being the dimension of z_t and r being the dimension of ω_t. Define $\varepsilon_t = D_t - E(D_t|z_t)$. Then, the null hypothesis can be rewritten as

$$H_0 : E(\varepsilon_t|\varpi_t) = 0 \text{ a.s.}$$

and the alternative hypothesis is

$$H_1 : E(\varepsilon_t|\varpi_t) \neq 0 \text{ on a set with positive measure.}$$

Furthermore, based on the following conditional moment

$$S = E\Big[\varepsilon_t f(z_t) \cdot E\big(\varepsilon_t f(z_t)|\varpi_t\big) \cdot f_\varpi(\varpi_t)\Big],$$

where $f(z_t)$ and $f_\varpi(\varpi_t)$ are the density functions of z_t and ϖ_t, respectively, Cai et al. (2019) proposed the following test statistic

$$S_T = \frac{1}{T(T-1)h^p}\sum_{t=1}^{T}\sum_{s\neq t}\Big(\hat{\varepsilon}_t\hat{f}(z_t)\cdot\big(\hat{\varepsilon}_s\hat{f}(z_s)\big)\Big)K\Big(\frac{\varpi_s-\varpi_t}{h}\Big),$$

where $\hat{\varepsilon}_t = D_t - \widehat{D}_t$,

$$\widehat{D}_t = \frac{1}{(T-1)h_1^q}\sum_{s\neq t}K_1\Big(\frac{z_s-z_t}{h_1}\Big)D_s/\hat{f}(z_t),$$

with $\hat{f}(z_t)$ being the kernel density estimator of $f(z_t)$, $K(\cdot)$ being a p-dimensional product kernel function, $K_1(\cdot)$ being a q-dimensional product kernel function , and both bandwidths h and h_1 being the smoothing parameters.

Finally, under some assumptions, Cai et al. (2019) obtained the following asymptotic results, under H_0,

$$\widetilde{S}_T = \frac{Th^{p/2}S_T}{\sqrt{2}\widehat{\sigma}_T} \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\widehat{\sigma}_T^2 = \frac{1}{T(T-1)h^p}\sum_{t=1}^{T}\sum_{s\neq t}\Big(\hat{\varepsilon}_t\hat{f}(z_t)\Big)^2\cdot\Big(\hat{\varepsilon}_s\hat{f}(z_s)\Big)^2K_{ts}^2,$$

being a consistent estimator of σ_T^2 given by

$$\sigma_T^2 = E\Big[f^4(z_t)f_\varpi(\varpi_t)\sigma^4(\varpi_t)\Big]\cdot\Big(\int K^2(u)du\Big),$$

with $\sigma^2(\varpi_t) = E\big(\varepsilon_t^2|\varpi_t\big)$, and under H_1, $P(\widetilde{S}_T > Q_T) \to 1$ for any non-stochastic sequence $Q_T = o(Th^{p/2})$. Therefore, the decision rule is that H_0 is rejected at the significance level α_0 if $\widetilde{S}_T > c$, where c is the upper α_0-percentile of the standard normal distribution.

3.4 RDD in Time Series

Regression discontinuity design (RDD) is another popular framework to identify treatment effects; see, for example, Thistlethwaite and Campbell (1960), Klaauw (2008), Lee and Card (2006), Lee (2008), Angrist and Lavy (1999), Ludwig and Miller (2007), and Hahn, Todd, and Klaauw (2001) for details. Kuersteiner et al. (2018) extended the RDD approach to a time-series environment. Let D_t follow a fixed rule $D_t = 1\{R_t > c\}$, where the running variable R_t is a continuous, nonrandom function of z_t such that $R_t = g(z_t)$ for some function $g(\cdot)$, z_t is a covariate, and c is a known threshold. Assume the potential outcome functions are continuous at the cutoff point c, i.e., $\eta_{t,l}$ in $Y_{t+l} = F_{t,l}(z_t, D_t, \eta_{t,l})$ satisfying $E[\eta_{t,l}|R_t = c]$ is a.s. continuous at c. Then, around the critical point c, the treatment effect is given by:

$$\theta_l(c) = \lim_{\Delta\downarrow 0}E[Y_{t+l}|R_t = c+\Delta] - \lim_{\Delta\uparrow 0}E[Y_{t+l}|R_t = c-\Delta].$$

The estimator $\hat{\theta}_l(c)$ is obtained by solving the problem:

$$(\hat{a}, \hat{b}, \hat{\gamma}, \hat{\theta}_l(c)) = arg \min_{a,b,\gamma,\theta_l} \sum_{t=1}^{T-l} (y_{t+l} - a_l - b_l(R_t - c) - \theta_l D_t - \gamma_l(R_t - c)D_t)^2 K\left(\frac{R_t - c}{h}\right),$$

where $K(\cdot)$ is a kernel function and h is a bandwidth. To illustrate the importance of RDD in practice, we use the example in Kuersteiner et al. (2018) to investigate how to use RDD in macroeconomic policy evaluation.

Example 3.1. *In October 1999, the Central Bank of Colombia (CBoC henceforth) adopted an inflation-targeting regime with a floating exchange rate. Meanwhile, to control the volatility of exchange rate, CBoC carried out rule-based currency interventions. The mechanics of rule-based interventions were as follows: at the close of any business day, whenever the average exchange rate for the entire day* e_t*, appreciated or depreciated (vis-a-vis its last 20-day moving average* $\bar{e}_t$*) at a rate faster than a cutoff* r_t*, the rule would be triggered, and call or put options on pesos would be issued. Options expired one month after the issued day, and could be exercised on days that the rule was triggered. There are two necessary conditions for CBoC to issue options:*

(1) The appreciating or depreciating rate of e_t *exceeds the cutoff* r_t*;*

(2) There are no outstanding options from a previous auction.

Thus, the running variable can be defined as:

$$C_t = \frac{1}{r_t}\frac{e_t - \bar{e}_t}{\bar{e}_t}(1 - OC_t), \quad \text{and} \quad P_t = \frac{1}{r_t}\frac{e_t - \bar{e}_t}{\bar{e}_t}(1 - OP_t),$$

where C_t *is the running variable for issuing call options on pesos,* P_t *is the running variable for issuing put options on pesos, and* OC_t *(*OP_t*) is a dummy variable denoting whether call (put) options from a previous auction remain outstanding at date t. The policy intervention dummies* $(D_{C,t}, D_{P,t})$ *are defined as,* $D_{C,t} = \mathbf{1}\{C_t \geq 1\}$ *with* $R_t = C_t$ *and* $c = 1$*, and* $D_{P,t} = \mathbf{1}\{P_t \leq -1\}$ *with* $R_t = P_t$ *and* $c = -1$*. During the time period from January 2002 to February 2012, the rule was trigged* 231 *times,* 38 *auctions were issued, and options were exercised in* 75 *cases. The aim is to evaluate the dynamic treatment effects of the rule-based interventions on exchange rates. Here,* Y_t *is change in log exchange rate, i.e.* $Y_t = \log e_{t+l} - \log e_t$*, where* $l > 0$*.*

In this example, the authors used the RDD method to measure the dynamic treatment effects. Consider the put options, for example, by assuming that $E[\eta_{t,l}|P_t = -1]$ was continuous, where $\eta_{t,l}$ was the unobservable shocks of $Y_{t,l}$ as defined above. Then, if P_t was very close to -1, based on the facts that traders could neither manipulate the running variable nor predict with certainty whether the rule was triggered or not, they concluded that any movement in P_t at this point was as good as random noise. While the small movement of P_t had a small effect on the average $\eta_{t,j}$, the movement could move $D_{P,t}$ from 0 to 1 or vice versa. Therefore, the movement when P_t was local to the cutoff was as good as if they could randomly change $D_{P,t}$, and they obtained a local random assignment mechanism. Thus, the local dynamic treatment effect of put options could be estimated by:

$$\theta_l = \lim_{p\downarrow(-1)} E[Y_{t+l}|P_t = p] - \lim_{p\uparrow(-1)} E[Y_{t+l}|P_t = p].$$

Define the dynamic treatment effect for horizons $1, \cdots L$ as:

$$\theta = (\theta_1, \cdots, \theta_L)'.$$

Then, $\hat{\theta}$ could be obtained by local linear method:

$$(\hat{a}, \hat{b}, \hat{\gamma}, \hat{\theta}) = arg \min_{a,b,\gamma,\theta} \sum_{l=1}^{L} \sum_{t=1}^{T-L} (y_{t+l} - a_l - b_l(P_t + 1) - \theta_l D_{P,t} - \gamma_l(P_t + 1)D_{P,t})^2 K \left(\frac{P_t + 1}{h} \right),$$

with $c = -1$.

$\hat{\theta}$ obtained by RDD method can be cast in both a local nonlinear impulse response function framework and a potential outcome framework, which again illustrates the link between dynamic treatment effects defined by potential outcomes and classic macroeconomic model. Consider a perturbation $\epsilon(\delta)$ of z_t such that the function of running variable $R_t = g(z_t)$ satisfying $g(z_t + \epsilon) - g(z_t) = \delta$ with $\delta > 0$. For δ fixed and $\epsilon = \epsilon(\delta)$, the nonlinear impulse response function of y_{t+j} is:

$$\theta_l(\epsilon, z_t) = F_{t,l}(D(z_t + \epsilon), z_t + \epsilon) - F_{t,l}(D(z_t), z_t).$$

Since $F_{t,l}(\cdot)$ is continuous, we have $\lim_{\delta \to 0} \theta_l(\epsilon(\delta), z_t) = 0$, when $g(z_t)$ is far away from the point c, and we have $\lim_{\delta \to 0} \theta_l(\epsilon(\delta), z_t) = \theta_l^{RDD}(z_t) = F_{t,l}(1, z_t) - F_{t,l}(0, z_t)$, when $g(z_t)$ is around the point c. According to the definition of dynamic potential outcomes given by Angrist and Kuersteiner (2011) and Angrist et al. (2018), we know that $F_{t,l}(1, z_t) = Y_{t,l}(1)$, and $F_{t,l}(0, z_t) = Y_{t,l}(0)$. Therefore, the local nonlinear impulse response function is simply the treatment effect defined by potential outcomes.

From the equation $\theta_l^{RDD}(z_t) = F_{t,l}(1, z_t) - F_{t,l}(0, z_t)$, we can check the different conditions among traditional impulse response function, selection on observables, and RDD. Write the observed outcome Y_{t+l} as:

$$Y_{t+l} = F_{t,l}(0, z_t) + \theta_l^{RDD}(z_t) D_t. \tag{11}$$

When assuming $F_{t,l}(0, z_t)$ is linear in z_t, $\theta_l^{RDD}(z_t)$ is a constant, and $E[D_t|z_t, F_{t,l}(\cdot)] = E[D_t|z_t]$ (selection on observables), (11) is a linear regression function, and $\theta_l^{RDD}(z_t)$ is the traditional impulse response coefficient. Without assuming that $F_{t,l}(0, z_t)$ is linear in z_t and $\theta_l^{RDD}(z_t)$ is a constant, the framework of selection on observables introduced in Section 3.3.1 can be applied. When the assumption of selection on observables fails, if a policy is implemented based on a clear rule, we can use the RDD method to estimate the dynamic treatment effects under the assumption that the potential outcomes function $F_{t,l}(z_t, D_t, \eta_{t,l})$ is a.s. continuous at c.

§4 Evaluation Methods with Multiple Time Series Data

In this section, we survey the existing approaches for estimation of macroeconomic policy effects with multiple time series data. We first discuss synthetic control method (SCM) and then examine the panel data approach proposed by Hsiao et al. (2012). In this setting, we have observations of multiple periods on multiple units, indexed by $i = 1, \cdots, N$, $t = 1, \cdots, T$. Before T_1 $(1 \leq T_1 < T)$, neither of units receives the policy intervention. after T_1, without loss of generality, we assume that only the first unit is exposed to the policy intervention.

4.1 Synthetic Control Method

A way to assess treatment effects for the multiple time series setting is the synthetic control method (SCM), which is proposed by Abadie and Gardeazabal (2003), and Abadie, Diamond and Hainmueller (2010). The main idea behind SCM is using a linear combination of all control units to form an artificial control unit that is more similar to the treatment unit in the pre-intervention periods than any of control units on their own. The counterfactual outcomes after the implementation periods for the treatment unit is estimated as outcomes of the artificial control unit in the same periods, which are weighted average ones of the control units. The SCM method is illustrated in detail using the example in Abadie and Gardeazabal (2003).

Example 4.1. *Assume there are N regions, $i = 1, \cdots N$. The first unit (i=1) is Basque Country, and the others are other Spanish regions. The regions were observed at periods from 1955 to 1997, i.e., $t = 1, \cdots T$, $t = 1$ denotes the year 1955, and $t = T$ denotes the year 1997. From year $T_1 + 1$ ($T_1 = 1967$), Euskadi Ta Askatasuna (ETA), a Basque terrorist organization, claimed its first victim in Basque Country and continued afterwards. ETA's terrorist activity was no more than two victims per year before 1973, increased to almost 16 victims per year on average in the period 1974-1977, and peaked to a total of 235 victims in the years of 1978-1980, after which it decreased gradually. In September 1998, ETA declared a total and indefinite cease fire. Since ETA's violent activity was concentrated in the Basque Country, other areas were deemed to not require such treatment. Let D_{it} denote a binary treatment indicator, Thus, we have the data $\{(D_{it}, Y_{it}, \mathbf{X}_i)\}$, where $D_{it} = 0$ for $i = 1 \cdots N$ and $t = 1, \cdots, T_1$, $D_{1t} = 1$ for $t = T_1 + 1, \cdots, T$, and $D_{it} = 0$ for $i = 2, \cdots, N$ and $t = T_1 + 1, \cdots, T$, and Y_{it} is per capita GDP for region i at time t, $\mathbf{X}_i$ is a vector of economic growth predictors for region i with the dimension of K (not affected by the intervention).*

The main target of the analysis in Abadie and Gardeazabal (2003) is to evaluate the impact that terrorism has had on economic growth for the Basque Country. Since there is no intervention before T_1, it is clear that $Y_{it} = Y_{it}(0)$ for $i = 1, \cdots, N, t = 1, \cdots, T_1$. After T_1, only the first unit receives the treatment. Then, $Y_{1t} = Y_{1t}(1)$ for $t = T_1 + 1, \cdots, T$ and $Y_{it} = Y_{it}(0)$ for $i = 2, \cdots, N$ and $t = T_1 + 1, \cdots, T$. To recover $Y_{1t}(0)$ for periods $t = T_1 + 1, \cdots, T$, Abadie and Gardeazabal (2003) conceptualized a weighted average of other Spanish regions as a "synthetic" Basque Country without terrorism; that was,

$$Y_{1t}(0) = \sum_{i=2}^{N} w_i \cdot Y_{it}, \quad t = T_1 + 1, \cdots, T,$$

with w_i ($i = 2, \cdots, N$) being the weight of region i in the synthetic Basque Country, $w_i \geq 0$, and $\sum_{i=2}^{N} w_i = 1$. Let $W = (w_2, \cdots, w_N)'$, W is obtained by minimizing

$$(Z_1 - Z_0 W)' V (Z_1 - Z_0 W)',$$

with $w_i \geq 0$ and $\sum_{i=2}^{N} w_i = 1$. Here, $Z_1 = (Y_{11}, \cdots, Y_{1T_0}, \mathbf{X}_1')'$ is a $(T_1 + K) \times 1$ vector of outcomes before T_1 and K covariates for Basque Country, and Z_0 is a $(T_1 + K) \times (N - 1)$ matrix of the same values of other N-1 regions. Further, V is a $(T_1 + K) \times (T_1 + K)$ positive-definite matrix, with the diagonal elements v_{ii} reflecting the relative importance of the different variables in determining Y. The choice of V could be subjective. After obtaining the optimal weights w_i^*, the treatment effect can be estimated by:

$$\hat{\Delta}_{1t} = Y_{1t} - \sum_{i=2}^{N} w_i^* Y_{it}.$$

SCM method can be regarded as a generalization of the differnce-in-differences (DID) model. A standard SCM model is supposed as:

$$Y_{it}(0) = \delta_t + b_i' f_t + \beta_t' \mathbf{X}_i + \epsilon_{it}, \quad \text{and} \quad Y_{it}(1) = Y_{it}(0) + \Delta_{it} D_{it}, \tag{12}$$

where δ_t is an unknown common factor with constant factor loadings across units, f_t is a vector of unknown common factors with varying factor loadings b_i, $\mathbf{X}_i$ is a vector of covariates (independent from intervention and time-invariant), β_t is a vector of unknown parameters with possible time-varying, and ϵ_{it} is unobserved shock with zero mean for all unit i. If we impose that f_t is constant for all t in SCM:

$$Y_{it}(0) = \delta_t + f b_i + \beta_t' \mathbf{X}_i + \epsilon_{it},$$

the DID model is obtained with $b_i = g_i$, where g_i is the group unit i belonged to, and $g_i \in \{0, 1, \cdots, G\}$. Thus, the SCM model allows for the time-varying effects of unobserved common factors, but the DID model restricts the effects of these unobservable common factors to be constant in time.

4.2 HCW Method

Hsiao et al. (2012) proposed a flexible and simple-to-implement panel data methodology to analyze treatment effects. This method exploited the correlations among cross sectional units to construct the counterfactuals. Hsiao et al. (2012) argued the cross-sectional dependence was due to the presence of some unobserved common factors. Based on this point, they developed a straightforward way to construct counterfactuals with observed data. In this section, we discuss the basic model and some extensions of their approach.

4.2.1 Model Setup

The setting is the same as that in SCM. That is, $Y_{1t} = Y_{1t}(0)$ for $t = 1, \cdots, T_1$, $Y_{1t} = Y_{1t}(1)$, for $t = T_1 + 1, \cdots, T$, and $Y_{it} = Y_{it}(0)$ for $i = 2, \cdots, N$ and $t = 1, \cdots, T$. To recover $Y_{1t}(0)$ for periods $T_1 + 1, \cdots, T$, the HCW method assumes that $Y_{it}(0)$ has the following factor model:

$$Y_{it}(0) = \alpha_i + b_i' f_t + \epsilon_{it}, \;\; i = 1, \cdots N; \;\; t = 1, \cdots T,$$

where α_i is the fixed individual-specific effect, f_t is a $K \times 1$ vector of unobserved common factors that vary over time, b_i is a $K \times 1$ vector of factor loadings that vary across i, ϵ_{it} is the unobservable random idiosyncratic component with $E[\epsilon_{it}] = 0$. The HCW model can be expressed in terms of matrix form as:

$$Y_t(0) = \alpha + B f_t + \epsilon_t,$$

where $Y_t(0) = (Y_{1t}(0), \ldots, Y_{Nt}(0))'$, $\alpha = (\alpha_1, \ldots \alpha_N)'$, $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$, B is the $N \times K$ factor loading matrix $B = (b_1, \ldots, b_N)'$, with $\|b_i\| = c < \infty$ for all i, ϵ_t is $I(0)$ such that $E(\epsilon_t) = 0$

and $E(\epsilon_t \epsilon_t') = V$, where V is a diagonal constant matrix, $E(\epsilon_t f_t') = 0$, and $RANK(B) = K$.

To estimate $Y_{1t}(0)$ for periods $T_1 + 1, \cdots, T$, it is needed to further assume that

$$E[\epsilon_{js}|D_{it}] = 0 \text{ for } j \neq i. \tag{13}$$

The assumption in (13) means that other units are not affected by the policy implemented to the treatment unit.

Let $v = (1, -\gamma')'$ and $\gamma = (\gamma_2, \cdots \gamma_N)'$, such that $v'B = 0$, i.e., $v \in \mathcal{N}(B)$, where $\mathcal{N}(B)$ is the null space of B. Then,

$$v'Y_t(0) = v'\alpha + v'\epsilon_t.$$

By rearranging, we obtain

$$Y_{1t}(0) = \gamma_1 + \gamma'\tilde{Y}_t + \epsilon_{1t}^*,$$

where $\gamma_1 = v'\alpha$, $\tilde{Y}_t = (Y_{2t}, \cdots, Y_{Nt})'$, $\epsilon_{1t}^* = v'\epsilon_t = \epsilon_{1t} - \gamma'\tilde{\epsilon}_t$, and $\tilde{\epsilon}_t = (\epsilon_{2t}, \cdots, \epsilon_{Nt})'$. Since $\tilde{Y}_t$ is correlated with ϵ_{1t}^*, we decompose $\epsilon_{1t}^* = E(\epsilon_{1t}^*|\tilde{Y}_t) + \eta_{1t}$ and $E[\eta_{1t}|\tilde{Y}_t] = 0$. Then,

$$Y_{1t}(0) = \gamma_1 + \gamma'\tilde{Y}_t + E(\epsilon_{1t}^*|\tilde{Y}_t) + \eta_{1t}.$$

Furthermore, it is assumed that

$$E(\epsilon_{1t}^*|\tilde{Y}_t) = a_0 + b_0'\tilde{Y}_t. \tag{14}$$

Then,

$$Y_{1t}(0) = a + c'\tilde{Y}_t + \eta_{1t}, \quad t = 1, \cdots, T, \tag{15}$$

where $a = a_0 + \gamma_1$, and $c = \gamma + b_0$, which means that we can use observations of $Y_{2t}, \cdots, Y_{Nt}$ to predict $Y_{1t}(0)$. Later, Li and Bell (2017) showed that without the assumption in (14), (15) still held. By minimizing the following equation:

$$\frac{1}{T_1}(Y_1(0) - a'\mathbf{e} - c'Y_{-1})A(Y_1(0) - a'\mathbf{e} - c'Y_{-1}),$$

we can obtain a consistent ordinary least squares (OLS) estimator of a and c. Here, $Y_1(0) = (Y_{11}, \cdots, Y_{1T_1})'$, $\mathbf{e}$ is a $T_1 \times 1$ vector of 1's, Y_{-1} is a $T_1 \times (N-1)$ matrix of T_1 time series observations of $\tilde{Y}_t$, A is a $T_1 \times T_1$ positive definite matrix. So, $Y_{1t}(0)$ for periods $t = T_1 + 1, \cdots, T$ is:

$$\hat{Y}_{1t}(0) = \hat{a} + \hat{c}'\tilde{Y}_t, \quad t = T_1 + 1, \cdots, T.$$

Thus, the treatment effect for unit 1 is:

$$\hat{\Delta}_{1t} = Y_{1t} - \hat{Y}_{1t}(0), \quad t = T_1 + 1, \cdots, T.$$

The standard deviation of $\hat{Y}_{1t}(0)$, denoted by $\sigma_{Y_{1t}(0)}$, can be calculated by the standard formula for standard deviation. For example, when η_{1t} is iid, then

$$\sigma_{Y_{1t}(0)}^2 = \sigma_{\eta_1}^2[1 + (1, \tilde{Y}_t')(Y_{-1}'Y_{-1})^{-1}(1, \tilde{Y}_t')].$$

The confidence interval of Δ_{1t} is correspondingly as $\hat{\Delta}_{1t} \pm c\sigma_{Y_{1t}(0)}$. Furthermore, suppose Δ_{1t} follows an autoregressive moving average model (ARMA):

$$a(L)\Delta_{1t} = \mu + \theta(L)\eta_t,$$

where L is the lag operator, η_t is white noise, and the roots of $\theta(L) = 0$ lie outside the unit circle. If all roots of $a(L) = 0$ lie outside the unit circle, then the long-term treatment effect is

$$\Delta_1 = a(L)^{-1}\mu = \mu^*,$$

and Δ_1 can be estimated by taking the simple average of the treatment effect:

$$\hat{\Delta}_1 = \frac{1}{T_2} \sum_{t=T_1+1}^{T} \hat{\Delta}_{1t},$$

where $T_2 = T - T_1$. Indeed, Li and Bell (2017) derived the consistency and asymptotic distribution of $\hat{\Delta}_1$ by showing that

$$\hat{\Delta}_1 - \Delta_1 = O_p(T_1^{-1/2} + T_2^{-1/2}),$$

and

$$\sqrt{T_2}(\hat{\Delta}_1 - \Delta_1) \xrightarrow{d} N(0, \Sigma), \tag{16}$$

where $\Sigma = \Sigma_1 + \Sigma_2$, $\Sigma_1 = \eta E(\chi_t)' V E(\chi_t)$,
$\eta = \lim_{T_1, T_2 \to \infty} T_2/T_1$, $\chi_t = (1, \tilde{Y}_t')'$, $V = \lim_{T_1, T_2 \to \infty} \text{Var}(\sqrt{T_1}\hat{\beta})$, $\hat{\beta} = (\hat{a}, \hat{c}')'$,
and $\Sigma_2 = \lim_{T_1, T_2 \to \infty} \text{Var}\left(T_2^{-1/2} \sum_{t=T_1+1}^{T} (\Delta_{1t} - E(\Delta_{1t}) + \eta_{1s})\right)$.
Obviously, the asymptotic normality in (16) can be used for making inferences for Δ_1 provided that a consistent estimate of Σ in (16) is available.

The above is the main idea of the HCW method. By using this method, Hsiao et al. (2012) assessed the effect of political and economic integration of Hong Kong with mainland China after 1997 on Hong Kong's economic growth.

4.2.2 Selection of Control Units

In real applications, the question how to select control units arises. It is well known that when $\tilde{Y}_t$ is used to predict $Y_{1t}(0)$, given the size of time series, it is not always better to include all units in the donor pool to the prediction regression. To balance the within-sample fit and post-sample prediction error, two criteria must be met to select control units. One is that control units must display strong correlations with the treatment unit based on the pretreatment data. The other criterion is that control units should be independent of the treatment.

Hsiao et al. (2012) proposed a two-step way to choose the strongest correlated predictors. First, assume we can use j units to predict $Y_{1t}(0)$, $j = 1, \cdots, N-1$ and then use R^2 or likelihood values to select the best predictor for $Y_{1t}(0)$, denoted by $M(j)^*$. Secondly, choose the best $M(m)^*$ from $M(1)^*, M(2)^*, \cdots, M(N-1)^*$ in terms of a model selection criterion.

The essence of the procedure proposed by Hsiao et al. (2012) is to minimize the Akaike information criterion (AIC) or its extension (corrected version of AIC, AICC) or the Baysian information criterion (BIC). However, AIC and AICC are asymptotically inconsistent (Shao, 1993, 1996) because neither the probability of selecting the optimal model nor the probability of selecting the model with the best predictive ability converges to 1 as $T \to \infty$. To overcome this problem, Du and Zhang (2015) suggested using a leave-n_v-out cross-validation criterion (CV(n_v)) to choose the optimal control units, which had the consistent property. The CV(n_v) method first chooses the data from cross-sectional dimension, the chosen subset of $\{2, \cdots, N\}$ is denoted by S, and the components in S are $\tilde{Y}_{St}$. Then, split the data into two parts from the time dimension. That is, $\{Y_{1t}, \tilde{Y}_{St}'\}_{t=1}^{T_1}$ is split into $\{(Y_{1k}, \tilde{Y}_{Sk}'), k \in \kappa\}$, $\{Y_{1k}, \tilde{Y}_{Sk}'), k \in \kappa^c\}$, where κ is a subset of $\{1, \cdots, T_1\}$ with n_v elements and κ^c is its complement. After that, regression Y_{1k} on constant and $\tilde{Y}_{Sk}$ using the data indexed by κ^c, and obtain the estimation of $\hat{a}_{S,\kappa^c}$ and $\hat{c}_{S,\kappa^c}$. Calculate the average squared prediction error using the data indexed by κ, $\frac{1}{n_v}\sum_{k\in\kappa}(Y_{1k}-\hat{a}_{S,\kappa^c}-\hat{c}'_{S,\kappa^c}\tilde{Y}_{Sk})^2$. There are $C_{T_1}^{n_v}$ ways to divide $1,2,\cdots T_1$ into κ and κ^c. Therefore, M of them are randomly drawn, as $\{\kappa_j,\kappa_j^c\}_{j=1}^M$, and the objective function is defined as

$$\frac{1}{n_vM}\sum_{j=1}^{M}\sum_{k\in\kappa_j}(Y_{1k}-\hat{a}_{S,\kappa_j^c}-\hat{c}'_{S,\kappa_j^c}\tilde{Y}_{Sk})^2. \tag{17}$$

From all the possible subsets of $\{2,\cdots N\}$, Du and Zhang (2015) chose S^* to minimize the objective function defined in (17). They showed that the CV(n_v) method could produce the consistent estimate and give smaller out-of-sample prediction results than AIC and AICC via simulations.

Instead of using some classical model selection approaches, Li and Bell (2017) proposed using the least absolute shrinkage and selection operator (Lasso) method to select control units. The objective function of the Lasso method is

$$\sum_{t=1}^{T_1}(Y_{1t}-\chi_t'\beta)^2+\lambda\sum_{j=1}^{N}|\beta_j|,$$

where $\chi_t=(1,\tilde{Y}_t')'$, $\beta=(\hat{a},\hat{c}')'$, and λ is a tuning parameter. The larger λ is, the more penalty is imposed on nonzero β_j. To obtain a Lasso estimator of β, a value of λ must first be selected. Li and Bell (2017) adopted the leave-one-out cross validation method to select λ over a discrete set $\Lambda_L=\{\lambda_1,\lambda_2,\cdots,\lambda_L\}$. For each $\lambda\in\Lambda_L$, and each $t=1,\cdots,T_1$, by minimizing the following leave-one-out objective function:

$$\sum_{k=1,k\neq t}^{T_1}(Y_{1k}-\chi_k'\beta)^2+\lambda\sum_{l=1}^{N}|\beta_l|,$$

we get the leave-one-out estimator of $\beta_{-t,\lambda}$. Then, we compute the CV(λ),

$$\text{CV}(\lambda)=\frac{1}{T_1}\sum_{t=1}^{T_1}(Y_{1t}-\chi_t'\beta_{-t,\lambda})^2.$$

Finally, we choose λ to minimize CV(λ).

4.2.3 Semiparametric Model

Ouyang and Peng (2015) relaxed the linear conditional mean assumption in (14) and extended the HCW model to a semiparametric setting. Under semiparametric setting, (15) becomes:

$$Y_{1t}(0)=g(\tilde{Y}_t)+\eta_{1t},$$

where $g(\tilde{Y}_t)=\gamma_1+\gamma'\tilde{Y}_t+E(\epsilon_{1t}^*|\tilde{Y}_t)$, and $g(\tilde{Y}_t)$ can be estimated using a nonparametric method, such as the local constant kernel method:

$$\hat{g}(\tilde{Y}_t)=\frac{\sum_{s=1}^{T_1}Y_{1s}K_{2s,2t}}{\sum_{s=1}^{T_1}K_{2s,2t}},\quad t=T_1+1,\cdots,T,$$

where $K_{2s,2t}=\Pi_{j=2}^{N}k((Y_{js}-Y_{jt})/h_j)$ is the (N-1)-dimensional product kernel function of the univariate kernel $k(\cdot)$, and h_j is the bandwidth associated with covariate Y_{jt} for $j=2,\cdots,N$.

The average treatment effect can then be estimated by:

$$\hat{\Delta}_{1,NP} = \frac{1}{T_2} \sum_{t=T_1+1}^{T} (Y_{1t} - \hat{g}(\tilde{Y}_t)).$$

To avoid the so-called "curse of dimensionality" problem for large N, $g(\tilde{Y}_t)$ can be estimated semiparametrically:

$$g(\tilde{Y}_t) = \beta' z_{1t} + h(z_{2t}), \tag{18}$$

where $z_{1t} \cup z_{2t} = \tilde{Y}_t$, and $z_{1t} \cap z_{2t} = \emptyset$. If q, the dimension of z_{2t} is low, the "curse of dimensionality" problem can be greatly alleviated. $\hat{\beta}$ and $\hat{h}(\cdot)$ can be estimated by the profile least squares method using the pretreatment data. First, we treat β as if it was known and estimate $h(z_{2t})$ by

$$\tilde{h}(z_{2t}) = T_1^{-1} \sum_{s=1}^{T_1} (Y_{1s} - \beta' z_{1t}) K_{sh}/\hat{f}_t^* = B_{1t} - B_{2t}'\beta, \tag{19}$$

where $B_{1t} = T_1^{-1} \sum_{s=1}^{T_1} Y_{1s} K_{sh}/\hat{f}_t^*$, $B_{2t} = T_1^{-1} \sum_{s=1}^{T_1} z_{1s} K_{sh}/\hat{f}_t^*$, $\hat{f}_t^* = T_1^{-1} \sum_{s=1}^{T_1} K_{sh}$, and $K_{sh} = \Pi_{l=1}^{q} h_l^{-1} k((z_{2s,l} - z_{2t,l})/h_l)$. Since β is unknown, $\tilde{h}(z_{2t})$ is not feasible. Replacing $h(z_{2t})$ in (18) by $\tilde{h}(z_{2t})$, we obtain:

$$Y_{1t} - B_{1t} = (z_{1t} - B_{2t})'\beta + \eta_{1t}.$$

Using the OLS to estimate β, then,

$$\hat{\beta} = \left[\sum_{t=1}^{T_1} (z_{1t} - B_{2t})(z_{1t} - B_{2t})' \right]^{-1} \sum_{t=1}^{T_1} (z_{1t} - B_{2t})(Y_{1t} - B_{1t}).$$

Plugging $\hat{\beta}$ into (19) yields

$$\hat{h}(z_{2t}) = T_1^{-1} \sum_{s=1}^{T_1} (Y_{1s} - \hat{\beta}' z_{1t}) K_{sh}/\hat{f}_t^*.$$

4.2.4 Disentangling the Effects of Multiple Treatments

All the methods proposed above assume that there is only one treatment. However, in many cases, there may be several treatments working simultaneously. Fujiki and Hsiao (2015) extended the standard HCW to distinguish the effect of one treatment from the other when the units were exposed to both working treatments at the same time. We use their example of Great Hanshin-Awaji earthquake to introduce the idea behind this methodology.

Example 4.2. *The Great Hanshin-Awaji earthquake took place on January 17, 1995. After that, the economic growth of the Kobe region declined dramatically. However, at approximately the same time, there were fundamental structural changes occurring around the disaster area. On the one hand, the port of Kobe was met by a challenge from other lower-cost ports in Asia such as Pusan, Hong Kong or Singapore. On the other hand, traditional industries, which were heavily concentrated in the disaster area, were challenged by globalization. For example, the chemical shoe industry, one of the most important local industries, had to compete with the cheaper shoes from China and the expensive shoes from Italy and France. How much the Great*

Hanshin-Awaji earthquake should account for this economic decline was the issue explored by Fujiki and Hsiao (2015).

Real GDP data from the 47 *prefectures are available from 1955 to fiscal year 2009, i.e., we have observations of* $N = 47$ *units in* $T = 54$ *periods, and* T_1 *is the fiscal year 1994. Before* T_1, *no unit received treatment. From* $T_1 + 1$ *onwards, the first unit, the Hyogo prefecture (the Kobe region) received two treatments* D_{1t} *and* D_{2t}, *while the rest did not.* D_{1t} *is a binary indicator for the earthquake treatment, and* D_{2t} *is a binary indicator for the other treatment, such as structural economic changes. Suppose the net effect of natural disaster is transitory, then, from* $T_2 + 1$ $(T_2 > T_1)$, *the observed value no longer contains the earthquake effect. Here,* Y *denotes the real GDP. Define* $Y_{1t} = Y_{1t}(0, 0)$ *for* $t = 1, \cdots, T_1$, $Y_{1t} = Y_{1t}(1, 1)$ *for* $t = T_1 + 1, \cdots, T_2$, $Y_{1t} = Y_{1t}(0, 1)$ *for* $t = T_2 + 1, \cdots, T$, *and* $Y_{it} = Y_{it}(0, 0)$ *for* $i = 2, \cdots, N$ *and* $t = 1, \cdots, T$, *where* $Y_{it}(i, j)$ *denotes the potential outcomes of* Y, *when* $D_{1t} = i$ *and* $D_{2t} = j$ *for* $i = 0$ *and* 1, *and* $j = 0$ *and* 1.

To measure the net impact of the earthquake on the economy of Kobe region, Fujiki and Hsiao (2015) first estimated $\hat{Y}_{1t}(0, 0)$ after period T_1 using the HCW method. That is,

$$\hat{Y}_{1t}(0, 0) = E[Y_{1t}(0, 0)|\tilde{Y}_t] = \gamma_1 + \gamma' \tilde{Y}_t^*, \ \ t = T_1 + 1, \cdots, T,$$

where $\tilde{Y}_t^*$ is the chosen subset of $\tilde{Y}_t$ according to the two criteria to predict $Y_{1t}(0, 0)$. Using data from 1 to T_1, $\hat{\gamma}_1$ and $\hat{\gamma}$ are obtained and we have:

$$\hat{Y}_{1t}(0, 0) = \hat{\gamma}_1 + \hat{\gamma}' \tilde{Y}_t^*, \ \ t = T_1 + 1, \cdots, T.$$

Since during periods from $T_1 + 1$ to T_2, $Y_{1t} = Y_{1t}(1, 1)$, the treatment effects $Y_{1t} - \hat{Y}_{1t}(0, 0)$ are the combined effects of earthquake and structural change. To isolate the net earthquake effects, Fujiki and Hsiao (2015) noticed the fact that only the treatment of structural change worked after T_2. Using the same way, they constructed $\hat{Y}_{1t}(0, 1)$ before T_2,

$$\hat{Y}_{1t}(0, 1) = E[Y_{1t}(0, 1)|\tilde{Y}_t] = \delta_1 + \delta' \tilde{Y}_t^{**}, \ \ t = 1, \cdots, T_2,$$

with $\tilde{Y}_t^{**}$ being the subset of $\tilde{Y}_t$ to predict $Y_{1t}(0, 1)$. Using data from $T_2 + 1$ to T, $\hat{\delta}_1$ and $\hat{\delta}$ are obtained and the estimation of potential outcome $Y_{1t}(0, 1)$ before T_2 is backcasted as:

$$\hat{Y}_{1t}(0, 1) = \hat{\delta}_1 + \hat{\delta}' \tilde{Y}_t^{**}, \ \ t = 1, \cdots, T_2.$$

Then, the net earthquake effects from $T_1 + 1$ to T_2 are:

$$\hat{\Delta}_{1t}^e = Y_{1t} - \hat{Y}_{1t}(0, 1), \ \ t = T_1 + 1, \cdots, T_2,$$

and the net effects of structural change from $T_1 + 1$ to T_2 are estimated by:

$$\hat{\Delta}_{1t}^s = \hat{Y}_{1t}(0, 1) - \hat{Y}_{1t}(0, 0), \ \ t = T_1 + 1, \cdots, T_2.$$

4.2.5 Comparison of SCM and HCW

As mentioned above, the HCW and SCM models are both applicable to the setting that only one unit is exposed to the treatment after period T_1, while others are never exposed to the treatment. Moreover, when δ_t from (12) in the SCM model is incorporated into $b_i' f_t$, a general underlying factor model for $Y_{it}(0)$ as the following can encompass the two methods:

$$Y_{it}(0) = \alpha_i + b_i' f_t + \beta_t' \mathbf{X}_i + \epsilon_{it}, \ \ i = 1, \cdots N, \ \ t = 1, \cdots T,$$

where the individual-specific effect α_i is not considered in SCM, and the covariate effect $\beta_t'\mathbf{X}_i$ is not included in the original version of HCW, see, for example, Gardeazabal and Vega-Bayo (2016) for details.

However, there are still important differences between SCM and HCW. Wan, Xie and Hsiao (2018) explained these differences carefully and they argued that the main differences between SCM and HCW were the assumptions that each method was based on. HCW places restrictions on the control units by assuming that control units must have strong correlation with the treatment unit and should be independent of the treatment, while SCM assumes that the weights must be nonnegative and add up to one. If the HCW assumptions about control units hold for both approaches, the difference between SCM and HCW is that HCW is an unconstrained regression, while SCM is a constrained regression. When the constraints are valid, SCM is more efficient. However, when the constraints are invalid, SCM could lead to biased estimation. Therefore, through conducting simulation studies, Wan et al. (2018) found that HCW significantly dominated SCM in a majority of cases.

§5 Conclusion

Evaluating the effect of macroeconomic policies quantitatively is one of the central issues in economic studies and policy research in many applied fields. This paper provides a selective review of recent advances in macroeconomic policy evaluation in the framework of the Rubin causal model. Compared to other popular methods such as DSGE and SVAR models, the method proposed by Angrist et al. (2018) and Kuersteiner et al. (2018) does not need to specify the model of the whole economy. Moreover, the dynamic treatment effect exactly corresponds to the impulse function induced from a DSGE or SVAR model, which means that the dynamic treatment effect provides an alternative method under less restrictive assumptions. When panel data are available, Hsiao et al. (2012), Abadie and Gardeazabal (2003), and Abadie et al. (2010) provided new methods to estimate the individual treatment effect of macroeconomic policies. In summary, macroeconomic policy evaluation is still a very dynamic and challenging research area which deserves further studies. As expected, without doubt, this area will receive a great attention in the near future. For example, of importance and challenge is to consider the synthetic control method and/or the HCW model to the quantile treatment effect as in (2).

The macroeconomic policies are inherently different over time, in the sense that the size of the same type of policy shocks is different at different time periods and/or they are implemented under different economic scenarios (deep recession versus mild recession versus boom), and/or relevant policy regimes (e.g., consider fiscal policy with or without zero lower bound or generally accommodating monetary policy). Therefore, the part of this article discusses some potential approaches of identifying one policy impacts by controlling for other policies scenarios most. Of course, a future research on whether the time-varying policy impacts can be addressed in this type of framework is very interesting, and there would have some interesting applications.

References

[1] A Abadie, J D Angrist, G W Imbens. *Instrumental variable estimates of the effect of subsidized training on the quantile of trainee earnings*, Econometrica, 2002, 70(1): 91-117.

[2] A Abadie, A Diamond, J Hainmueller. *Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program*, J Amer Statist Assoc, 2010, 105 (2010): 493-505.

[3] A Abadie, J Gardeazabal. *The economic costs of conflict: a case study of the Basque country*, Am Econ Rev, 2003, 93(1): 113-132.

[4] A Abadie, G W Imbens. *Large sample properties of matching estimators for average treatment effects*, Econometrica, 2006, 74(1): 235-267.

[5] A Abadie, G W Imbens. *On the failure of the bootstrap for matching estimators*, Econometrica, 2008, 76(6): 1537-1557.

[6] A Abadie, G W Imbens. *Bias-corrected matching estimators for average treatment effects*, J Bus Econ Stat, 2011, 29(1): 1-11.

[7] A Abadie, G W Imbens. *Matching on the estimated propensity score*, Econometrica, 2016, 84(2): 781-807.

[8] J H Abbring, G J van den Berg. *The nonparametric identification of treatment effects in duration models*, Econometrica, 2003, 71(5): 1491-1517.

[9] J D Angrist, G W Imbens. *2-stage least-squares estimation of average causal effects in models with variable treatment intensity*, J Am Stat Assoc, 1995, 90(430): 431-442.

[10] J D Angrist, Ò Jordà, G M Kuersteiner. *Semiparametric estimates of monetary policy effects: string theory revisited*, J Bus Econ Stat, 2018, 36(3): 371-387.

[11] J D Angrist, G M Kuersteiner. *Causal effects of monetary shocks: semiparametric conditional independence tests with a multinomial propensity score*, Rev Econ Stat, 2011, 93(3): 725-747.

[12] J D Angrist , V Lavy. *Using maimonides' rule to estimate the effect of class size on scholastic achievement*, Q J Econ, 1999, 114(2): 533-575.

[13] I Bojinov, N Shephard. *Time series experiments and causal estimands: exact randomization tests and trading*, J Am Stat Assoc, 2019, Forthcoming.

[14] T L Brunell, J E DiNardo. *A propensity score reweighting approach to estimating the Partisan effects of full turnout in American presidential elections*, Polit Anal, 2004, 12(1): 28-45.

[15] Z Cai, Y Fang, M Lin, S Tang. *Testing conditional unconfoundedness using auxiliary variables*, *Working Paper*, Department of Economics, University of Kansas, 2019.

[16] M Caliendo, S Kopeinig. *Some practical guidance for the implementation of propensity score matching*, J Econ Surv, 2010, 22(1): 31-72.

[17] G Cerulli. *Econometric evaluation of socio-economic programs*, Springer, 2015.

[18] V Chernozhukov, C Hansen. *An IV model of quantile treatment effects*, Econometrica, 2005, 73(1): 245-261.

[19] R H Dehejia, S Wahba. *Causal effects in nonexperimental studies: reevaluating the evaluation of training programs*, J Am Stat Assoc, 1999, 94(448): 1053-1062.

[20] R H Dehejia, S Wahba. *Propensity score-matching methods for nonexperimental causal studies*, Rev Econ Stat, 2002, 84(1): 151-161.

[21] Z C Du, L Zhang. *Home-purchase restriction, property tax and housing price in China: A counterfactual analysis*, J Econometrics, 2015, 188(2): 558-568.

[22] H Fujiki, C Hsiao. *Disentangling the effects of multiple treatmentsMeasuring the net economic impact of the 1995 great Hanshin-Awaji earthquake*, J Econometrics, 2015, 186(1): 66-73.

[23] J Gardeazabal, A VegaBayo. *An empirical comparison between the synthetic control method and HSIAO et al.'s panel data approach to program evaluation*, J Appl Economet, 2017, 32(5): 983-1002.

[24] A N Glynn, K M Quinn. *An introduction to the augmented inverse propensity weighted estimator*, Polit Anal, 2010, 18(1): 36-56.

[25] J Hahn. *On the role of the propensity score in efficient semiparametric estimation of average treatment effects*, Econometrica, 1998, 66(2): 315-331.

[26] J Hahn, P Todd, W V D Klaauw. *Identification and estimation of treatment effects with a regression-discontinuity design*, Econometrica, 2001, 69(1): 201-209.

[27] J J Heckman, H Ichimura, P E Todd. *Matching as an econometric evaluation estimator: evidence from evaluating a job training programme*, Rev Econ Stud, 1997, 64(4): 605-654.

[28] J J Heckman, H Ichimura, P E Todd. *Matching as an econometric evaluation estimator*, Rev Econ Stud, 1998, 65(2): 261-294.

[29] K Hirano, G W Imbens. *The propensity score with continuous treatments*, in *Applied bayesian modeling and causal inference from incomplete-data perspectives*, edited by G Andrew and X L Meng, pp.73-84: Wiley-Blackwell, 2004.

[30] K Hirano , G W Imbens , G Ridder. *Efficient estimation of average treatment effects using the estimated propensity score*, Econometrica, 2003, 71(4): 1161-1189.

[31] P W Holland. *Statistics and causal inference*, J Am Stat Assoc, 1986, 81(396): 945-960.

[32] D G Horvitz, D J Thompson. *A Generalization of Sampling Without Replacement From a Finite Universe*, J Am Stat Assoc, 1952, 47(260): 663-685.

[33] C Hsiao, H S Ching, S K Wan. *A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with Mainland China*, J Appl Economet, 2012, 27(5): 705-740.

[34] G W Imbens. *The role of the propensity score in estimating dose-response functions*, Biometrika, 2000, 87(3): 706-710.

[35] G W Imbens. *Nonparametric estimation of average treatment effects under exogeneity: a review*, Rev Econ Stat, 2004, 86(1): 4-29.

[36] G W Imbens, J M Wooldridge. *Recent developments in the econometrics of program evaluation*, J Econ Lit, 2009, 47(1): 5-86.

[37] Ò Jordà, A M Taylor. *The time for austerity: estimating the average treatment effect of fiscal policy*, Econ J, 2016, 126(590): 219-255.

[38] W V D Klaauw. *Breaking the link between poverty and low student achievement: An evaluation of Title I* , J Econometrics, 2008, 142(2): 731-756.

[39] R Koenker, G Bassett. *Regression quantiles*, Econometrica, 1978, 46(1): 33-50.

[40] G M Kuersteiner, D C Phillips , M Villamizar-Villegas. *Effective sterilized foreign exchange intervention? Evidence from a rule-based policy*, J Int Econ, 2018, 113: 118-138.

[41] M J van der Lann, J M Robins. *Unified methods for censored longitudinal data and causality*, Springer, 2003.

[42] D S Lee. *Randomized experiments from non-random selection in U.S. house elections*, J Econometrics, 2008, 142(2): 675-697.

[43] D S Lee, D Card. *Regression discontinuity inference with specification error*, J Econometrics, 2006, 142(2): 655-674.

[44] K T Li, D R Bell. *Estimation of average treatment effects with panel data: Asymptotic theory and implementation*, J Econometrics, 2017, 197(1): 65-75.

[45] Z Liu, Z Cai, Y Fang. *Policy evaluation of monetary policy and macro-prudential policy in China* , *Working paper*, The Wang Yanan Institute for Studies in Economics, Xiamen University, 2019.

[46] J J Lok. *Statistical modeling of causal effects in continuous time*, Ann Stat, 2008, 36(3): 1464-1507.

[47] J Ludwig, D L Miller. *Does head start improve children's life chances? evidence from a regression discontinuity design*, Q J Econ, 2007, 122(1): 159-208.

[48] J K Lunceford, M Davidian. *Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study*, Stat Med, 2004, 23(19): 2937-2960.

[49] M Ouyang, Y Peng. *The treatment-effect estimation: a case study of the 2008 economic stimulus package of China*, J Econometrics, 2015, 188(2): 545-557.

[50] J M Robins. *A new approach to causal inference in mortality studies with a sustained exposure periodapplication to control of the healthy worker survivor effect*, Mathl Modelling, 1986, 7(9): 1393-1512.

[51] J M Robins. *Correcting for non-compliance in randomized trials using structural nested mean models*, Commun Stat-Theor M, 1994, 23(8): 2379-2412.

[52] J M Robins, S Greenland, F C Hu. *Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome*, J Am Stat Assoc, 1999, 94(447): 687-700.

[53] J M Robins, A Rotnitzky. *Semiparametric efficiency in multivariate regression models with missing data*, J Am Stat Assoc, 1995, 90(429): 122-129.

[54] J M Robins, A Rotnitzky, L P Zhao. *Analysis of semiparametric regression models for repeated outcomes in the presence of missing data*, J Am Stat Assoc, 1995, 90(429): 106-121.

[55] P R Rosenbaum, D B Rubin. *The central role of the propensity score in observational studies for causal effects*, Biometrika, 1983, 70(1): 41-55.

[56] D B Rubin. *Matching to remove bias in observational studies*, Biometrics, 1973a, 29(1): 159-183.

[57] D B Rubin. *The use of matched sampling and regression adjustment to remove bias in observational studies*, Biometrics, 1973b, 29(1): 185-203.

[58] D B Rubin. *Estimating causal effects of treatments in randomized and nonrandomized studies*, J Educ Psychol, 1974, 66(5): 688-701.

[59] D B Rubin. *Assignment to Treatment Group on the Basis of a Covariate*, J Educ Stat, 1977, 2(1): 1-26.

[60] D B Rubin. *Bayesian inference for causal effects: the role of randomization*, Ann Stat, 1978, 6(1): 34-58.

[61] D B Rubin. *Using multivariate matched sampling and regression adjustment to control bias in observational studies*, J Am Stat Assoc, 1979, 74(366): 318-328.

[62] J Shao. *Linear model selection by cross-validation*, J Am Stat Assoc, 1993, 88(422), 486-495.

[63] J Shao. *Bootstrap model selection*, J Am Stat Assoc, 1996, 91(434), 655-665.

[64] D L Thistlethwaite, D T Campbell. *Regression-discontinuity analysis: an alternative to the ex post facto experiment*, J Educ Psychol, 1960, 51(6): 309-317.

[65] S K Wan, Y Xie, C Hsiao. *Panel data approach vs synthetic control method*, Econ Lett, 2018, 164: 121-123.

[66] J M Wooldridge. *Inverse probability weighted estimation for general missing data problems*, J Econometrics, 2007, 141(2): 1281-1301.

[67] J M Wooldridge. *Econometric analysis of cross section and panel data*, MIT Press, 2010.

[1]The Wang Yanan Institute for Studies in Economics, MOE Key Laboratory of Econometrics (Xiamen University), Fujian Provincial Key Laboratory of Statistical Science, Xiamen University, Xiamen 361005, China.

E-mail: linming50@xmu.edu.cn

[2]Department of Economics, University of Kansas, Lawrence, KS 66045, USA.

[3]Department of Statistics, School of Economics, Xiamen University, Xiamen 361005, China.