

Appendix to "Estimation of time-varying average treatment effects using panel data when unobserved fixed effects affect potential outcomes differently"

Shosei Sakaguchi* †

July 10, 2016

Abstract

Section 1 provides a representation of the proposed estimator as a GMM estimator along with the regularity conditions. Section 2 provides the proof of Proposition 1. Section 3 concerns the proposed approach to the common trend assumption. Section 4 constructs the proposed approach in the repeated cross-section data setting.

1 Representation of the proposed estimator as a GMM estimator

Let $\theta_t = (\beta_1^{0'}, \beta_t^{0'}, \beta_t^{1'}, \beta_T^{1'}, \tau_t^{ate})'$ be a vector of parameters, Θ_t be a parameter space of θ_t , $\theta_{t,o} = (\beta_{1,o}^{0'}, \beta_{t,o}^{0'}, \beta_{t,o}^{1'}, \beta_{T,o}^{1'}, \tau_{t,o}^{ate})'$ be a vector of true parameters¹, and $\hat{\theta}_t = (\hat{\beta}_1^{0'}, \hat{\beta}_t^{0'}, \hat{\beta}_t^{1'}, \hat{\beta}_T^{1'}, \hat{\tau}_t^{ate})'$ be a vector of estimators derived in section 3 of the main text of the paper. $\hat{\theta}_t$ is a GMM estimator with the following $4K + 1$ vector of stacked moment functions

$$m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_t) = \begin{bmatrix} (1 - D_{it}) \cdot \begin{pmatrix} X_{it} \\ -X_{i1} \end{pmatrix} (Y_{it} - Y_{i1} - X'_{it}\beta_t^0 + X'_{i1}\beta_1^0) \\ D_{it} \cdot \begin{pmatrix} X_{it} \\ -X_{iT} \end{pmatrix} (Y_{it} - Y_{iT} - X'_{it}\beta_t^1 + X'_{iT}\beta_T^1) \\ (X'_{it}\beta_t^1 - X'_{iT}\beta_T^1 + Y_{iT}) - (X'_{it}\beta_t^0 - X'_{i1}\beta_1^0 + Y_{i1}) - \tau_t^{ate} \end{bmatrix},$$

*Graduate School of Economics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan. Email: sakaguchi.shosei.75x@st.kyoto-u.ac.jp.

†Research Fellow, Japan Society for the Promotion of Science.

¹While in the main text of the paper the true parameters are denoted without the subscript o , we denote the true parameters with the subscript o here.

and its weighting matrix is an identity matrix. The first and subsequent $2K$ elements of $m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_t)$ correspond to the moment functions for OLS in (4) and (6), respectively. The last one corresponds to the moment function obtained by combining (5), (7), and (8). Since we consider the just-identified case, the choice of weighting matrix is irrelevant.

We suppose the following regularity conditions.

Assumption A.1.

(i) Θ_t is compact. (ii) $E[\sup_{\theta_t \in \Theta_t} \|m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_t)\|] < \infty$.

2 Proof of Proposition 1

First, I prove the consistency result. Under Assumptions 1, 2, 3, 4, and A.1, $E[(1 - D_{it}) \cdot (X'_{it}, -X'_{i1})'(Y_{it} - Y_{i1} - X'_{it}\beta_{t,o}^0 + X'_{i1}\beta_{1,o}^0)] = 0$ and $E[D_{it} \cdot (X'_{it}, -X'_{iT})'(Y_{it} - Y_{iT} - X'_{it}\beta_{t,o}^1 + X'_{iT}\beta_{T,o}^1)] = 0$ uniquely hold with

$$\begin{pmatrix} \beta_{t,o}^0 \\ \beta_{1,o}^0 \end{pmatrix} = E \left[(1 - D_{it}) \cdot \begin{pmatrix} X_{it} \\ -X_{i1} \end{pmatrix} \begin{pmatrix} X_{it} \\ -X_{i1} \end{pmatrix}' \right]^{-1} E \left[(1 - D_{it}) \cdot \begin{pmatrix} X_{it} \\ -X_{i1} \end{pmatrix} (Y_{it} - Y_{i1}) \right]$$

and

$$\begin{pmatrix} \beta_{t,o}^1 \\ \beta_{T,o}^1 \end{pmatrix} = E \left[D_{it} \cdot \begin{pmatrix} X_{it} \\ -X_{iT} \end{pmatrix} \begin{pmatrix} X_{it} \\ -X_{iT} \end{pmatrix}' \right]^{-1} E \left[D_{it} \cdot \begin{pmatrix} X_{it} \\ -X_{iT} \end{pmatrix} (Y_{it} - Y_{iT}) \right],$$

respectively. Under Assumptions 1, 2, and 3:

$$\begin{aligned} & E[\{X'_{it}\beta_{t,o}^1 - X'_{iT}\beta_{T,o}^1 + Y_{iT}\} - \{X'_{it}\beta_{t,o}^0 - X'_{i1}\beta_{1,o}^0 + Y_{i1}\}] - \tau_{t,o}^{ate} \\ &= (E[Y_{it}(1) - Y_{iT}(1)] + E[Y_{it}(1)]) - (E[Y_{it}(0) - Y_{i1}(0)] + E[Y_{i1}(0)]) - \tau_{t,o}^{ate} \\ &= E[Y_{it}(1)] - E[Y_{it}(0)] - \tau_{t,o}^{ate} \\ &= 0. \end{aligned}$$

Therefore, $E[m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_{t,o})] = 0$ and $E[m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_t)] \neq 0$ for $\theta_t \neq \theta_{t,o}$, so $\theta_{t,o}$ is identified. Combining this identification result and Assumptions 1, 2, 3, 4, and A.1, the consistency of $\hat{\tau}_t^{ate}$ follows Theorem 2.6 in Newey and MacFadden (1994).

Next, I prove the asymptotic normality of $\hat{\tau}_t^{ate}$. Since $\hat{\theta}_t$ is a GMM estimator with a

vector of stacked moment functions $m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_t)$ and we consider the just-identified case, as following Theorem 6.1 in Newey and MacFadden (1994):

$$\sqrt{N}(\hat{\theta}_t - \theta_{t,o}) \xrightarrow{d} N(0, M_t^{-1} E[m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_{t,o}) m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_{t,o})'] M_t'^{-1}),$$

where

$$M_t = E[\partial m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_{t,o}) / \partial \theta_t']$$

$$= E \begin{bmatrix} -(1 - D_{it}) \cdot \begin{pmatrix} X_{it} \\ -X_{i1} \end{pmatrix} \begin{pmatrix} X_{it} \\ -X_{i1} \end{pmatrix}' & 0 & 0 \\ 0 & -D_{it} \cdot \begin{pmatrix} X_{it} \\ -X_{iT} \end{pmatrix} \begin{pmatrix} X_{it} \\ -X_{iT} \end{pmatrix}' & 0 \\ -\begin{pmatrix} X_{it} \\ -X_{i1} \end{pmatrix}' & \begin{pmatrix} X_{it} \\ -X_{iT} \end{pmatrix}' & -1 \end{bmatrix}.$$

Since the asymptotic variance of $\sqrt{N}(\hat{\tau}_t^{ate} - \tau_{t,o}^{ate})$ is the $(4K + 1) \times (4K + 1)$ element of $M_t^{-1} E[m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_{t,o}) m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_{t,o})'] M_t'^{-1}$, this is the mean squared value of the $(4K + 1) \times 1$ element of $M_t^{-1} m_t(X_{i1}, X_{it}, X_{iT}, D_{it}; \theta_{t,o})$ which is given by:

$$\begin{aligned} & \tau_{t,o}^{ate} - (X_{it}' \beta_{t,o}^1 - X_{iT}' \beta_{T,o}^1 + Y_{iT}) + (X_{it}' \beta_{t,o}^0 - X_{i1}' \beta_{1,o}^0 + Y_{i1}) \\ & + E \begin{bmatrix} X_{it} \\ -X_{i1} \end{bmatrix} E \left[(1 - D_{it}) \cdot \begin{pmatrix} X_{it} \\ -X_{i1} \end{pmatrix} \begin{pmatrix} X_{it} \\ -X_{i1} \end{pmatrix}' \right]^{-1} \left[(1 - D_{it}) \cdot \begin{pmatrix} X_{it} \\ -X_{i1} \end{pmatrix} (u_{it}^0 - u_{i1}^0) \right] \\ & - E \begin{bmatrix} X_{it} \\ -X_{iT} \end{bmatrix} E \left[D_{it} \cdot \begin{pmatrix} X_{it} \\ -X_{iT} \end{pmatrix} \begin{pmatrix} X_{it} \\ -X_{iT} \end{pmatrix}' \right]^{-1} \left[D_{it} \cdot \begin{pmatrix} X_{it} \\ -X_{iT} \end{pmatrix} (u_{it}^1 - u_{iT}^1) \right]. \end{aligned}$$

Therefore, we have $\sqrt{N}(\hat{\tau}_t^{ate} - \tau_{t,o}^{ate}) \xrightarrow{d} N(0, V_t)$ where

$$\begin{aligned} V_t = & E\{[(X_{it}' \beta_{t,o}^1 - X_{iT}' \beta_{T,o}^1 + Y_{iT}) - (X_{it}' \beta_{t,o}^0 - X_{i1}' \beta_{1,o}^0 + Y_{i1}) - \tau_{t,o}^{ate} \\ & - E[X_{it}', -X_{i1}']' E \left[(1 - D_{it}) \cdot (X_{it}', -X_{i1}')' (X_{it}', -X_{i1}') \right]^{-1} [(1 - D_{it}) \cdot (X_{it}', -X_{i1}')' (u_{it}^0 - u_{i1}^0)] \\ & + E[X_{it}', -X_{iT}']' E \left[D_{it} \cdot (X_{it}', -X_{iT}')' (X_{it}', -X_{iT}') \right]^{-1} [D_{it} \cdot (X_{it}', -X_{iT}')' (u_{it}^1 - u_{iT}^1)]\}^2. \end{aligned}$$

□

3 Relation to the common trend assumption

Under Assumption 2, the potential outcome models (1) and (2) satisfy the following common trend assumptions due to the absence of interaction between $g^j(C_i)$ and X_{it} for $j = 0, 1$:

$$\begin{aligned} & E[Y_{it}(0) \mid D_{it} = 1, X_{i1}, \dots, X_{iT}, C_i] - E[Y_{i1}(0) \mid D_{it} = 1, X_{i1}, \dots, X_{iT}, C_i] \\ &= E[Y_{it}(0) \mid D_{it} = 0, X_{i1}, \dots, X_{iT}, C_i] - E[Y_{i1}(0) \mid D_{it} = 0, X_{i1}, \dots, X_{iT}, C_i] \quad (\text{A.1}) \\ &= E[Y_{it}(0) \mid X_{it}, C_i] - E[Y_{i1}(0) \mid X_{i1}, C_i] \end{aligned}$$

and

$$\begin{aligned} & E[Y_{it}(1) \mid D_{it} = 1, X_{i1}, \dots, X_{iT}, C_i] - E[Y_{iT}(1) \mid D_{it} = 1, X_{i1}, \dots, X_{iT}, C_i] \\ &= E[Y_{it}(1) \mid D_{it} = 0, X_{i1}, \dots, X_{iT}, C_i] - E[Y_{iT}(1) \mid D_{it} = 0, X_{i1}, \dots, X_{iT}, C_i] \quad (\text{A.2}) \\ &= E[Y_{it}(1) \mid X_{it}, C_i] - E[Y_{iT}(1) \mid X_{iT}, C_i], \end{aligned}$$

where the exogeneity assumption also holds. Equation (A.1) is a usual common trend assumption, which is the defining assumption for the DID approach, and states that the difference in the expected potential outcome under no treatment between periods t and 1 is unrelated to the treated or control group. On the other hand, Equation (A.2) is an unusual common trend assumption and states that the difference in the expected potential outcome under treatment between periods t and T is unrelated to the treated or control group.

The proposed method is based on these common trend assumptions. As seen in the population analogue of the discussion in section 3 of the main text, the proposed method identifies the average potential outcome under no treatment for the entire population, $E[Y_{it}(0)]$, based on (A.1) and the average potential outcome under treatment for the entire population, $E[Y_{it}(1)]$, based on (A.2). Therefore, the proposed method identifies the ATE for the entire population based on (A.1) and (A.2).

4 Repeated cross-section data

The discussion in the main text is based on the panel data setting. However, the proposed method can also be constructed in the repeated cross-section data setting wherein the advantage of the proposed method is to identify the ATE on the entire population rather than on the treated only, unlike the standard DID approach.

In the repeated cross-section data setting, we suppose that the observed data sample $\{Y_{it}, D_{it}, X_{it}\}$ consists of T independent cross-sections at different points in time. The $t = 1, \dots, T$ denotes the cross-section, while $i = 1, \dots, N_t$ indexes the units in cross-section t . Units in different cross-sections are not same. Y_{it} denotes an observed outcome. X_{it} denotes observed covariates that many include a constant term and observed confounders. $D_{it} \in \{0, 1\}$ is a group dummy such that $D_{it} = 1$ indicates that unit i in cross-section t is treated.

We consider the following potential outcome models:

$$Y_{it}(0) = X'_{it}\beta_t^0 + \gamma^0 D_{it} + u_{it}^0, \quad (\text{A.3})$$

$$Y_{it}(1) = X'_{it}\beta_t^1 + \gamma^1 D_{it} + u_{it}^1, \quad (\text{A.4})$$

where u_{it}^j is a mean-zero error term defined as $u_{it}^j = Y_{it}(j) - E[Y_{it}(j) | X_{it}, D_{it}]$ for $j = 0, 1$. We replace unobserved unit fixed effects components $g^0(C_i)$ and $g^1(C_i)$ in the panel data setting with group fixed effects components $\gamma^0 D_{it}$ and $\gamma^1 D_{it}$, respectively, in the repeated cross-section data setting.

The potential outcome models (A.3) and (A.4) satisfy the following common trend assumptions:

$$\begin{aligned} & E[Y_{it}(0) | D_{it} = 1, X_{it}] - E[Y_{i1}(0) | D_{it} = 1, X_{i1}] \\ &= E[Y_{it}(0) | D_{it} = 0, X_{it}] - E[Y_{i1}(0) | D_{it} = 0, X_{i1}] \\ &= E[Y_{it}(0) | X_{it}] - E[Y_{i1}(0) | X_{i1}] \end{aligned}$$

and

$$\begin{aligned} & E[Y_{it}(1) | D_{it} = 1, X_{it}] - E[Y_{iT}(1) | D_{it} = 1, X_{iT}] \\ &= E[Y_{it}(1) | D_{it} = 0, X_{it}] - E[Y_{iT}(1) | D_{it} = 0, X_{iT}] \\ &= E[Y_{it}(1) | X_{it}] - E[Y_{iT}(1) | X_{iT}]. \end{aligned}$$

Then, under modifications of Assumptions 1, 3, 4, and A.1 for the repeated cross-section data setting, the ATE on the entire population, τ_t^{ate} , can be estimated using a method similar that discussed in section 3 of the main text as in the following three steps.

First step:

For the potential outcome under no treatment, we estimate β_t^0 and β_1^0 from the model (A.3) using the subsample of units with $D_{it} = 0$ in cross-section t and $D_{i1} = 0$ in cross-section 1 (this is the full sample in cross-section 1). We obtain the OLS estimators of β_t^0 and β_1^0 denoted by $\hat{\beta}_t^0$ and $\hat{\beta}_1^0$, respectively.

Then, using the full sample in cross-sections t and 1, we estimate $E[Y_{it}(0)]$ as follows:

$$E[\widehat{Y_{it}(0)}] = \hat{E}[X_{it}]' \hat{\beta}_t^0 - \hat{E}[X_{i1}]' \hat{\beta}_1^0 + \hat{E}[Y_{i1}],$$

where $\hat{E}[\cdot]$ denotes a sample mean.

Second step:

The second step is a symmetry of the first step. For the potential outcome under treatment, we estimate β_t^1 and β_T^1 from the model (A.4) using the subsample of units with $D_{it} = 1$ in cross-section t and $D_{iT} = 1$ in cross-section T (this is the full sample in cross-section T). We obtain the OLS estimators of β_t^1 and β_T^1 denoted by $\hat{\beta}_t^1$ and $\hat{\beta}_T^1$, respectively.

Then, using the full sample in cross-sections t and T , we estimate $E[Y_{it}(1)]$ as follows:

$$E[\widehat{Y_{it}(1)}] = \hat{E}[X_{it}]' \hat{\beta}_t^1 - \hat{E}[X_{iT}]' \hat{\beta}_T^1 + \hat{E}[Y_{iT}].$$

Third step:

Finally, τ_t^{ate} is estimated as follows:

$$\hat{\tau}_t^{ate} = E[\widehat{Y_{it}(1)}] - E[\widehat{Y_{it}(0)}].$$

The estimator is consistent and asymptotic normal, which can be proved in the same way as the proof of Proposition 1. The asymptotic variance has a similar form with (9) as the following:

$$\begin{aligned} V_{t,o} = & E[\{(X'_{it}\beta_{t,o}^1 - X'_{iT}\beta_{T,o}^1 + Y_{iT}) - (X'_{it}\beta_{t,o}^0 - X'_{i1}\beta_{1,o}^0 + Y_{i1}) - \tau_{t,o}^{ate} \\ & - E[X_{it}]' E[(1 - D_{it}) \cdot X_{it} X'_{it}]^{-1} [(1 - D_{it}) \cdot X'_{it} u_{it}^0] \\ & + E[X_{i1}]' E[(1 - D_{it}) \cdot X_{i1} X'_{i1}]^{-1} [(1 - D_{it}) \cdot X'_{i1} u_{i1}^0] \\ & + E[X_{it}]' E[D_{it} \cdot X_{it} X'_{it}]^{-1} [D_{it} \cdot X'_{it} u_{it}^1] \\ & - E[X_{iT}]' E[D_{it} \cdot X_{iT} X'_{iT}]^{-1} [D_{it} \cdot X'_{iT} u_{iT}^1]\}^2], \end{aligned}$$

where the expectation of the random variables is taken under their joint distribution.

References

Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing.
In: Handbook of Econometrics. vol. 4, pp.2111-2245.