Article

Not peer-reviewed version

The Differential Error Components Models under Treatment Effect Heterogeneity

Yongtao Jin *

Posted Date: 8 July 2024

doi: 10.20944/preprints202407.0569.v1

Keywords: Generalized least square; Error component; Fixed effects; Random effects; Treatment effect heterogeneity

Article

The Differential Error Components Models under Treatment Effect Heterogeneity

Yongtao Jin

School of Management and Economics, Beijing Institute of Technology, Beijing, 100081, China; yongtao.jin@bit.edu.cn

Abstract: The fixed and random effects models are popular methods for identifying the causal effects of treatment. First, I introduce the two-way fixed effects model and show a class of the estimators for the random effects models. Misidentification may occur in these models due to the treatment effect heterogeneity. Then, I propose the differential error component models to address such problems while the models exclude the fixed effects of units and time periods and consider the random effects of units and time periods. Meanwhile, I present a simulation of a staggered design and revisit the application in (De Chaisemartin and D'Haultfoeuille, 2020). The results demonstrate that one additional newspaper increases the 0.31% average effect of the presidential turnout and the 1.05% average effect when the newspapers' changes are non-negative, where the proposed estimators are robust and efficient.

Keywords: causal effects; generalized least square; fixed effects; random effects; error component; heterogeneity

1. Introduction

Identification of the treatment effects usually needs to exclude the fixed effects, the covariates effects, and the disturbances. The treatments may be policies, programs, or activities. The Fixed effects (FE) models and the random effects (RE) models are popular and efficient methods to identify the treatment effects, which are used to exclude the effects of units or groups and time periods. Additional information about samples is one way to obtain more precise treatment effects. The alternative is to suggest more precise estimators for treatment effects.

The two-way fixed effects (TWFE) model is an identification method with the effects of units and time periods are fixed. There has been abundant research on the TWFE model in recent years. (Goodman-Bacon, 2021) shows the negative weighted problem because the treated time periods of units in different groups are different in the TWFE model. He gives the Goodman-Bacon decomposition theorem to address such problems. De Chaisemartin and D'Haultfoeuille (2020) show an estimator to assess the robustness of the estimators for the TWFE model. They also propose an estimator that is a weighted sum of the treatment effects to avoid the negative weighted problem. (de Chaisemartin and D'Haultfoeuille, 2022) review the recent studies of the TWFE model with binary or non-binary treatment, staggered or not staggered design, and continuous or discrete time periods. The TWFE model is a popular method to identify the treatment effects of multiple time periods, and its extensions are also growing.

The one-way or two-way random effects models are the identification methods in which the effects of units or time periods are random. (Wallace and Hussain, 1969) and (Maddala, 1971) show some generalized least square (GLS) estimators of the multivariate linear regression model with three error components such as the Within estimator, the Between units estimator, and the time periods estimator. They also give the consistency, the unbiasedness, and the asymptotic normality of the estimators. (Swamy and Arora, 1972) show a class of estimators for the TWRE model, including two-stage estimators. (Baltagi, 1981) compares estimators of the OLS, the LSDV, the GLS, and six other two-stage estimators for the TWRE model by Monte-Carlo simulation. Baltagi also shows the advantages and disadvantages of these estimators in terms of both theoretical properties and application.

The question is, which is the optimal model of the above models for our daily research? Many papers have introduced the two-way effects models and their extensions. Baltagi, B. H. (2008) specifically shows the two-way effects models and discusses when we choose the TWFE or the TWRE model. (Hausman, 1978) proposes the tests for choosing the fixed e or the random effects model. However, the test result may mislead researchers into choosing an inefficient model for some applications. (Mundlak, 1978), (Chamberlain, 1984), (Hausman and Taylor, 1981), (Metcalf, 1996), and other papers show that one should test the restrictions of the fixed or random effects models before applying them to application. The RE models request that the effects of units or time periods are independent of the treatment variable and the covariates, while the models have heteroscedasticity. The FE models have endogeneity between the units, the time periods, the treatment variable, and the covariates in regression. Moreover, the FE and the RE models both have treatment effect heterogeneity. The differential models (DM) are one way to exclude heterogeneity, such as the first-difference (FD) model and the second-difference (SD) model.

Based on the above analysis, I show how the treatment effect heterogeneity causes the misidentification of the above models. Furthermore, I propose the differential error component (DRE) models that exclude the fixed effects of units and time periods and consider the random effects of units and time periods. Besides, I present a simulation of the staggered design with five groups and five time periods and revisit the application of the effects of newspapers on presidential turnout to test whether the proposed estimators are feasible and efficient.

The remainder of the paper is structured as follows. Section 2 introduces the TWFE and shows a class of estimators for the TWRE model. Section 3 proposes the DRE models. The section also shows the advantages and disadvantages of the above models. Section 4 presents a simulation and an application to compare the estimators mentioned in Section 2 and Section 3. Finally, Section 5 summarizes this paper.

2. Setups and the Misidentification in the FE and RE Models

2.1. Setups

Firstly, we show the TWFE regression model

$$Y_{i,t} = \gamma_t + \eta_i + \beta_{fe} D_{i,t} + \boldsymbol{W}'_{i,t}\boldsymbol{\beta}_w + \varepsilon_{i,t}, t \in \{1, \cdots, T\}, i \in \{1, \dots, n\} \qquad (2.1)$$

where γ_t is the time fixed effects and η_i is the unit fixed effects. β_{fe} represents the treatment effect coefficient. $D_{i,t}$ is a multilevel treatment statue. $\boldsymbol{W}$ is the covariate matrix. $\boldsymbol{\beta}_w$ is the coefficient vector of $\boldsymbol{W}$. $\varepsilon_{i,t}$ represents the random disturbance term. The TWFE model is a popular case of linear regression models. In this TWFE regression model, the treatments in Equation (2.1) to units may exit in some time periods. The units may be treated again in the next periods. Furthermore, we should consider the heterogeneity of units and time periods in Equation (2.1) when we analyze multiple time period data, such as panel data and repeated cross-section data. Wallace and Hussain (1969) first proposed the Within estimator of the TWFE model. Liu and Sun (2019) and de Chaisemartin and D'Haultfoeuille (2020) briefly introduce the Within method.

The one-way random effect (OWRE) model: The OWRE model is a simple case of the RE models that take the units or the time period effects as the random effects. Here, I introduce the OWRE model that takes the units' effects as random effects. Defining $\boldsymbol{I}_N$ is the $N \times N$ identify matrix. $\mathbf{1}_N$ is a $N \times 1$ vector that equals for all elements. $\boldsymbol{J}_{NT}$ equals to $\frac{1}{NT}\mathbf{1}_N\mathbf{1}'_N$. We have the multiple linear regression model with two error components follows

$$\boldsymbol{Y} = \mathbf{1}\beta_0 + \boldsymbol{D}\beta_{OWREU} + \boldsymbol{\gamma} + \boldsymbol{u}_\eta = \boldsymbol{X}\boldsymbol{\beta}_{OWRE} + \boldsymbol{u}_\eta \qquad (2.2)$$

where $\boldsymbol{Y} = (Y_{11}, \dots, Y_{1T}, \dots, Y_{N1}, \dots, Y_{NT}), \boldsymbol{D} = (D_{11}, \dots, D_{1T}, \dots, D_{N1}, \dots, D_{NT}), \beta_{OWREU}$ is the OWRE for unit. $\boldsymbol{X} = (\mathbf{1}_{NT}, \boldsymbol{D}, \mathbf{1}_N \otimes \boldsymbol{I}_T)$, $\boldsymbol{\beta}_{OWRE} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_{OWREU}, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_T)$, $\boldsymbol{u}_{\boldsymbol{\eta}} = (\boldsymbol{I}_N \otimes \mathbf{1}_T)\boldsymbol{\eta} + \boldsymbol{\varepsilon}$. $(\boldsymbol{\eta}, \boldsymbol{\varepsilon})$ are independent with each other. $E[\boldsymbol{\eta}] = \boldsymbol{\eta}_0, \boldsymbol{\eta}_0$ is a constant vector. $Cov(\boldsymbol{\eta}) = \sigma_\eta^2 \boldsymbol{I}_N, \sigma_\eta^2$ is the variance of η_i. $E[\boldsymbol{\varepsilon}] = \boldsymbol{0}, Cov(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \boldsymbol{I}_{NT}, \sigma_\varepsilon^2$ is the variance of $\varepsilon_{i,t}$. It is easy to show that the variance matrix of the disturbances $\boldsymbol{u}_\eta$ is

$$\text{Cov}(\boldsymbol{u}_\eta) = (T\sigma_\eta^2 + \sigma_\varepsilon^2)(\boldsymbol{P}_1 - \boldsymbol{J}_{NT}) + \sigma_\varepsilon^2(\boldsymbol{I}_{NT} - \boldsymbol{P}_1) + (T\sigma_\eta^2 + \sigma_\varepsilon^2)\boldsymbol{J}_{NT} = \sigma_1^2 \boldsymbol{Q}_1 + \sigma_3^2 \boldsymbol{Q}_{3'} + \sigma_1^2 \boldsymbol{Q}_4.$$

where $\boldsymbol{Q}_1 = \boldsymbol{P}_1 - \boldsymbol{J}_{NT}, \boldsymbol{Q}_{3'} = \boldsymbol{I}_{NT} - \boldsymbol{P}_1, \boldsymbol{Q}_4 = \boldsymbol{J}_{NT}, \boldsymbol{P}_1 = \boldsymbol{I}_N \otimes \boldsymbol{J}_T$.

Proposition 1 Let $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I})$ and $(\sigma_\eta^2, \sigma_\varepsilon^2)$ are given, we have the GLS estimators of Equation (2.2) as $\widetilde{\boldsymbol{\beta}}_{OWREU(1)} = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{X}}{\sigma_3^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_1^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{Y}}{\sigma_3^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_1^2}\right), \widetilde{\boldsymbol{\beta}}_{OWREU(2)} = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{X}}{\sigma_3^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{Y}}{\sigma_3^2}\right), \widetilde{\boldsymbol{\beta}}_{OWREU(3)} = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{X}}{\sigma_3^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_1^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{Y}}{\sigma_3^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_1^2}\right), \widetilde{\boldsymbol{\beta}}_{OWREU(4)} = (\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X} + \boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y} + \boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X})$, And

$\widehat{\boldsymbol{\beta}}_{OWU_{between}} = (\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y})$, and $\widehat{\boldsymbol{\beta}}_{OWU\_Within} = (\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{Y})$.

Proposition 2 Let $\boldsymbol{\eta} - \boldsymbol{\eta}_0 \xrightarrow{L} N(\boldsymbol{0}, \sigma_\eta^2 \boldsymbol{I}), \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_0 \xrightarrow{L} N(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I})$, The two-stage estimators of Equation (2.2) are

$$\widehat{\boldsymbol{\beta}}_{OWREU(1)} = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{s_{RE(1)}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{X}}{s_{RE(3')}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{s_{RE(1)}^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{s_{RE(1)}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{Y}}{s_{RE(3')}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{s_{RE(1)}^2}\right),$$

$$\widehat{\boldsymbol{\beta}}_{OWREU(2)} = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{s_{RE(1)}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{X}}{s_{RE(3')}^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{s_{RE(1)}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{Y}}{s_{RE(3')}^2}\right),$$

$$\widehat{\boldsymbol{\beta}}_{OWREU(3)} = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{X}}{s_{RE(3')}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{s_{RE(1)}^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{Y}}{s_{RE(3')}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{s_{RE(1)}^2}\right),$$

where $s_{RE(3')}^2 = \frac{\boldsymbol{Y}'\left(\boldsymbol{Q}_{3'} - \boldsymbol{Q}_{3'}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_{3'}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_{3'}\right)\boldsymbol{Y}}{m_{3'}}$.

The TWRE model: The TWRE model takes the effects of units and time periods as the random effects into the disturbance terms in the regression. Maddala (1971), Wallace and Hussain (1969), and Baltagi B. H. (2008) have introduced this linear model specifically. I will show a class of least square estimators for the TWRE model. The multiple linear regression model with three error components is

$$\boldsymbol{Y} = \boldsymbol{1}\beta_0 + \boldsymbol{X}\boldsymbol{\beta}_{TWRE} + \boldsymbol{u} \tag{2.3}$$

where $\boldsymbol{Y} = (Y_{11}, \dots, Y_{1T}, \dots, Y_{N1}, \dots, Y_{NT}), \boldsymbol{X} = (D_{11}, \dots, D_{1T}, \dots, D_{N1}, \dots, D_{NT}), \boldsymbol{\beta}_{TWRE}$ represents the TWRE vector. $\boldsymbol{u} = (\boldsymbol{I}_N \otimes \boldsymbol{1}_T)\boldsymbol{\eta} + (\boldsymbol{1}_N \otimes \boldsymbol{I}_T)\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} = (\varepsilon_{11}, \dots, \varepsilon_{1T}, \dots, \varepsilon_{N1}, \dots, \varepsilon_{NT})$. It is easy to show the covariance matrix of the disturbances $\boldsymbol{u}$ is

$$Cov(\boldsymbol{u}) = T\sigma_\eta^2(\boldsymbol{I}_N \otimes \boldsymbol{J}_T) + N\sigma_\gamma^2(\boldsymbol{J}_N \otimes \boldsymbol{I}_T) + \sigma_\varepsilon^2 \boldsymbol{I}_{NT}$$

with $E[\boldsymbol{\eta}] = \boldsymbol{\eta}_0, Cov(\boldsymbol{\eta}) = \sigma_\eta^2 \boldsymbol{I}_N. E[\boldsymbol{\gamma}] = \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_0$ is the constant vector. $Cov(\boldsymbol{\gamma}) = \sigma_\gamma^2 \boldsymbol{I}_T, \sigma_\gamma^2$ is the variance of γ_i. $(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\varepsilon})$ are independent with each other. For convenient of the estimation, let

$$Cov(\boldsymbol{u}) = (T\sigma_\eta^2 + \sigma_\varepsilon^2)(\boldsymbol{P}_1 - \boldsymbol{J}_{NT}) + (N\sigma_\gamma^2 + \sigma_\varepsilon^2)(\boldsymbol{P}_2 - \boldsymbol{J}_{NT}) + \sigma_\varepsilon^2(\boldsymbol{I}_{NT} + \boldsymbol{J}_{NT} - \boldsymbol{P}_1 - \boldsymbol{P}_2)$$

$$+ (T\sigma_\eta^2 + N\sigma_\gamma^2 + \sigma_\varepsilon^2)\boldsymbol{J}_{NT} = \sigma_1^2 \boldsymbol{Q}_1 + \sigma_2^2 \boldsymbol{Q}_2 + \sigma_3^2 \boldsymbol{Q}_3 + \sigma_4^2 \boldsymbol{Q}_4.$$

where $\boldsymbol{Q}_2 = \boldsymbol{P}_2 - \boldsymbol{J}_{NT}, \boldsymbol{Q}_3 = \boldsymbol{I}_{NT} + \boldsymbol{J}_{NT} - \boldsymbol{P}_1 - \boldsymbol{P}_2, \boldsymbol{P}_2 = \boldsymbol{J}_N \otimes \boldsymbol{I}_T$.

Proposition 3 Let $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I})$ and $(\sigma_\eta^2, \sigma_\gamma^2, \sigma_\varepsilon^2)$ are given. We have the GLS estimators of Equation (2.3) as

$$\widetilde{\boldsymbol{\beta}}_{TWRE}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right),$$

$$\widetilde{\boldsymbol{\beta}}_{TWRE(1)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}\right),\cdots,\widetilde{\boldsymbol{\beta}}_{TWRE(10)}=$$
$$\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right),\text{ and }\widehat{\boldsymbol{\beta}}_{TWRES(i)}=(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{Y}).\,(i=1,2,3,4)$$

Proposition 3 shows all GLS estimators of Equation (2.3). The ellipses of the estimators are similar to Theorem 1.

Proposition 4 Let $\boldsymbol{\eta}-\boldsymbol{\eta}_0\xrightarrow{L}N(\boldsymbol{0},\sigma_\eta^2\boldsymbol{I}),\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\xrightarrow{L}N(\boldsymbol{0},\sigma_\gamma^2\boldsymbol{I}),\boldsymbol{\varepsilon}\xrightarrow{L}N(\boldsymbol{0},\sigma_\varepsilon^2\boldsymbol{I})$, the two-stage estimators of Equation (2.3) are

$$\widehat{\boldsymbol{\beta}}_{TWRE}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{s_{RE(1)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{s_{RE(2)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{s_{RE(3)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{s_{RE(4)}^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{s_{RE(1)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{s_{RE(2)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{s_{RE(3)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{s_{RE(4)}^2}\right)$$

and

$$\widehat{\boldsymbol{\beta}}_{TWRE(1)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{s_{RE(1)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{s_{RE(2)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{s_{RE(3)}^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{s_{RE(1)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{s_{RE(2)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{s_{RE(3)}^2}\right),\cdots,\widehat{\boldsymbol{\beta}}_{TWRE(10)}=$$
$$\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{s_{RE(3)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{s_{RE(4)}^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{s_{RE(3)}^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{s_{RE(4)}^2}\right).\,(i=1,2,\cdots,10)$$

where $s_{RE(4)}^2=s_{RE(1)}^2+s_{RE(2)}^2-s_{RE(3)}^2$, $s_{RE(j)}^2$ is the estimator of $\sigma_j^2,(j=1,2,3,4)$

In the above, we show many estimators of the TWRE model, including $\widehat{\boldsymbol{\beta}}_{TWRE}$, $\widehat{\boldsymbol{\beta}}_{TWRES(j)}$, $\widehat{\boldsymbol{\beta}}_{TWRE(i)}$, $\widetilde{\boldsymbol{\beta}}_{TWRE}$ and $\widetilde{\boldsymbol{\beta}}_{TWRE(i)}$ $(i=1,2,\cdots,10.j=1,2,3,4)$. Firstly, I exclude the estimators when $\sigma_1^2=\sigma_2^2=\sigma_3^2=\sigma_4^2$. These estimators are accurate only if their covariance matrix equals to the true error covariance matrix of the data. Therefore, the estimators $\widetilde{\boldsymbol{\beta}}_{TWRE}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ and $\widetilde{\boldsymbol{\beta}}_{TWRE(i)}$ $(i=1,2,\cdots,10)$ are omitted here. Secondly, I exclude estimators of the combination of $(\boldsymbol{Q}_1,\boldsymbol{Q}_2,\boldsymbol{Q}_4)$ that contain insufficient information about units or time periods. $\widehat{\boldsymbol{\beta}}_{TWRES(j)}$ and $\widehat{\boldsymbol{\beta}}_{TWRE(i)}$ $(j=1,2,4.i=4,5,7,9)$ are omitted. As a result, we have $\widehat{\boldsymbol{\beta}}_{TWRES(3)}$, $\widehat{\boldsymbol{\beta}}_{TWRE}$ and $\widehat{\boldsymbol{\beta}}_{TWRE(i)}$ $(i=1,2,3,6,8,10)$ as ideal estimators for the TWRE model. These estimators are consistent in distribution because the covariance matrix strongly relates to the true covariance matrix of the data. Similarly, the above analysis is also adjusted to the differential error component models introduced in the next sections.

2.2. The Misidentification in the FE and RE Models

Misidentification may occur in FE and RE models due to the treatment effect heterogeneity. I introduce the negative weighted problem before showing heterogeneous treatment effects in these models. The negative weighted problem comes from the treatment effect heterogeneity between units or time periods. Goodman-Bacon (2021) demonstrates the negative weighted problem by a staggered design with three groups and three time periods. He shows the Goodman-Bacon decomposition theorem that considers the treatment effect heterogeneity over time periods. Figure 1 demonstrates that the negative weighted problem is caused by Case 4, which is a two-by-two DID with Group 1 and Group 2. This is the other view of the treatment effect heterogeneity compared to the one in Goodman-Bacon (2021). The heterogeneity of this 2x2 DID come from the difference in the treatment effects over time periods in the early treated Group 2. De Chaisemartin and D'Haultfoeuille (2020) show the negative weighted problem in the TWFE model. They proposed an estimator of a weighted sum of the treatment effects when the treatment status changes. The estimator is used to avoid the treatment effect heterogeneity over time periods.

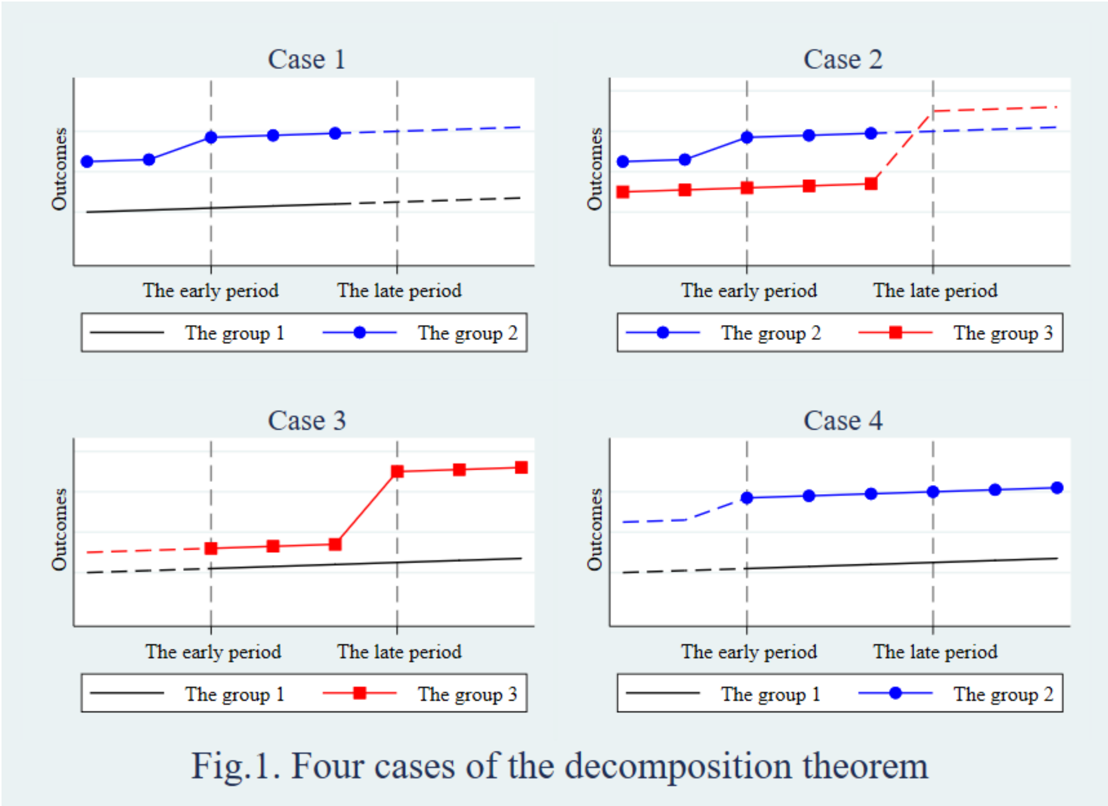

Figure 1. Four cases of the decomposition theorem.

The RE models take the heterogeneity of treatment effects as part of the error components. These are misleading to the estimation of the treatment effects because the distribution of the treatment effect heterogeneity is not given or pre-estimated as normality. Therefore, the estimates of the TWRE are not consistent but may be more robust than the TWFE model. The OWRE model that takes the time period effects as fixed and the unit effects as random is even worse than the TWFE or TWRE models. Accordingly, the estimation results of the OWRE in Section 4 are overestimated or negative.

3. The Differential Error Component Model

The OWRE or the TWRE models take unit effects or time period effects as random effects in Section 2. It is reasonable that the disturbances of a large panel have asymptotic normality in these models. Meanwhile, the regression models need intercept terms to avoid fixed effects on the treatment effect heterogeneity. However, it is hard to find such fitted terms because of the differences in fixed effects between units or time periods. Here, I introduce the differential models that exclude the fixed effects in regression. The FD model gives the trends of change in the treatment effects. The SD model, also called FD of FD's, gives the characteristic of change in treatment effects. The FD is similar to speed in Physics, while the SD is similar to acceleration. Next, I take the RE models into estimation to avoid random effects to the treatment effects. This three-stage estimation excludes the fixed effects in the first step and the random effects in the last two steps. The differential error component models that combine the differential models and the error component models are also called as the differential random effects (DRE) model.

3.1. The First Differential Random Effects (FDRE) Models

Firstly, we have the FD model

$$\boldsymbol{Y}_{t+1} - \boldsymbol{Y}_t = (\boldsymbol{X}_{t+1} - \boldsymbol{X}_t)\boldsymbol{\beta}_{FDRE} + \boldsymbol{u}. \ (t = 1, 2, \cdots, T-2\,) \tag{3.1}$$

with $E[\boldsymbol{u}] = \boldsymbol{0}$. The covariance matrix of the disturbances $\boldsymbol{u}$ does not discussed here if we combine Equation (2.1) and Equation (3.1) as the FD model. I show no assumptions about the disturbances in

 doi:10.20944/preprints202407.0569.v1

Equation (2.1). Without loss of generality, I assume the disturbances in Equation (3.1) satisfy the following assumption.

Assumption 3.1 The disturbances in Equation (3.1) have the error components as $(\boldsymbol{\eta}, \boldsymbol{\gamma}', \boldsymbol{\varepsilon}')$.

The first differential two-way random effects (FDTW) model: The TWRE model mentioned in Section 2 under Assumption 3.1 follows

$$\boldsymbol{Y}_{t+1} - \boldsymbol{Y}_t = (\boldsymbol{X}_{t+1} - \boldsymbol{X}_t)\boldsymbol{\beta}_{FDTW} + \boldsymbol{u}_{FDTW} \quad (3.2)$$

where $\boldsymbol{u}_{FDTW} = (\boldsymbol{I}_N \otimes \boldsymbol{1}_{T-1})\boldsymbol{\eta} + (\boldsymbol{1}_N \otimes \boldsymbol{I}_{T-1})\boldsymbol{\gamma}' + \boldsymbol{\varepsilon}'$, $\boldsymbol{\gamma}' = (\gamma_1, \dots, \gamma_{T-1}), \boldsymbol{\varepsilon}' = (\varepsilon_{11}, \dots, \varepsilon_{1(T-1)}, \dots, \varepsilon_{N1}, \dots, \varepsilon_{N(T-1)})$.

The covariance matrix of $\boldsymbol{u}_{FDTW}$ in Equation (3.2) is

$$\text{Cov}(\boldsymbol{u}_{FDTW}) = (T-1)\sigma_{\boldsymbol{\eta}}^2(\boldsymbol{I}_N \otimes \boldsymbol{J}_{T-1}) + N\sigma_{\boldsymbol{\gamma}'}^2(\boldsymbol{J}_N \otimes \boldsymbol{I}_{T-1}) + \sigma_{\boldsymbol{\varepsilon}'}^2 \boldsymbol{I}_{N(T-1)}.$$

with $E[\boldsymbol{\eta}] = \boldsymbol{\eta}_0, Cov(\boldsymbol{\eta}) = \sigma_{\boldsymbol{\eta}}^2 \boldsymbol{I}_N, E[\boldsymbol{\gamma}'] = \boldsymbol{\gamma}'_0, Cov(\boldsymbol{\gamma}') = \sigma_{\boldsymbol{\gamma}'}^2 \boldsymbol{I}_{T-1}, E[\boldsymbol{\varepsilon}'] = \boldsymbol{0}, Cov(\boldsymbol{\varepsilon}') = \sigma_{\boldsymbol{\varepsilon}'}^2 \boldsymbol{I}_{N(T-1)}$. $(\boldsymbol{\eta}, \boldsymbol{\gamma}', \boldsymbol{\varepsilon}')$ are independent with each other. Furthermore, we have

$$\text{Cov}(\boldsymbol{u}_{FDTW}) = \sigma_1^2 \boldsymbol{Q}_1 + \sigma_2^2 \boldsymbol{Q}_2 + \sigma_3^2 \boldsymbol{Q}_3 + \sigma_4^2 \boldsymbol{Q}_4.$$

where $\sigma_1^2 = (T-1)\sigma_{\boldsymbol{\eta}}^2 + \sigma_{\boldsymbol{\varepsilon}'}^2$, $\sigma_2^2 = N\sigma_{\boldsymbol{\gamma}'}^2 + \sigma_{\boldsymbol{\varepsilon}'}^2$, $\sigma_3^2 = \sigma_{\boldsymbol{\varepsilon}'}^2$, $\sigma_4^2 = (T-1)\sigma_{\boldsymbol{\eta}}^2 + N\sigma_{\boldsymbol{\gamma}'}^2 + \sigma_{\boldsymbol{\varepsilon}'}^2$, $\boldsymbol{Q}_1 = \boldsymbol{P}_1 - \boldsymbol{J}_{N(T-1)}$, $\boldsymbol{Q}_2 = \boldsymbol{P}_2 - \boldsymbol{J}_{N(T-1)}$, $\boldsymbol{Q}_3 = \boldsymbol{I}_{N(T-1)} + \boldsymbol{J}_{N(T-1)} - \boldsymbol{P}_1 - \boldsymbol{P}_2$, $\boldsymbol{Q}_4 = \boldsymbol{J}_{N(T-1)}$, $\boldsymbol{P}_1 = \boldsymbol{I}_N \otimes \boldsymbol{J}_{(T-1)}$, $\boldsymbol{P}_2 = \boldsymbol{J}_N \otimes \boldsymbol{I}_{T-1}$.

Lemma 1 $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{Q}_3$ and $\boldsymbol{Q}_4$ are symmetric and idempotent. $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{Q}_3$ and $\boldsymbol{Q}_4$ are orthogonal with each other. And $\boldsymbol{P}_1\boldsymbol{J}_{NT} = \boldsymbol{P}_2\boldsymbol{J}_{NT} = \boldsymbol{P}_1\boldsymbol{P}_2 = \boldsymbol{J}_{NT}$.

Lemma 2 The rank of $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{Q}_3$ and $\boldsymbol{Q}_4$ are $N, T, N-1, T-1, (N-1)(T-1)$ and 1. And $rk(\boldsymbol{X}) = rk(\boldsymbol{Q}_1\boldsymbol{X}) = rk(\boldsymbol{Q}_2\boldsymbol{X}) = rk(\boldsymbol{Q}_3\boldsymbol{X}) = k$.

Theorem 1 Let $\boldsymbol{\varepsilon}' \sim N(\boldsymbol{0}, \sigma_{\boldsymbol{\varepsilon}'}^2 \boldsymbol{I})$ and $\left(\sigma_{\boldsymbol{\eta}}^2, \sigma_{\boldsymbol{\gamma}'}^2, \sigma_{\boldsymbol{\varepsilon}'}^2\right)$ are given, we have the GLS estimators of Equation (3.2) as

$$\widetilde{\boldsymbol{\beta}}_{FDTW} = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1} \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right),$$

$$\widetilde{\boldsymbol{\beta}}_{FDTW(1)} = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}\right)^{-1} \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}\right), \cdots, \widetilde{\boldsymbol{\beta}}_{FDTW(10)} = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1} \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right), \text{ and}$$

$$\widehat{\boldsymbol{\beta}}_{FDTWS(i)} = (\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{Y}). (i = 1, 2, 3, 4)$$

Corollary 1 $E[\widetilde{\boldsymbol{\beta}}_{FDTW}] = E[\widetilde{\boldsymbol{\beta}}_{FDTW(i)}] = E[\widehat{\boldsymbol{\beta}}_{FDTWS(j)}] = \boldsymbol{\beta}_{FDTW}, (i = 1, 2, \cdots, 10. j = 1, 2, 3, 4)$. $Var(\widetilde{\boldsymbol{\beta}}_{FDTW}) = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}$, $Var(\widetilde{\boldsymbol{\beta}}_{FDTW(1)}) = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}\right)^{-1}$, $\cdots$, $Var(\widetilde{\boldsymbol{\beta}}_{FDTW(10)}) = \left(\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}$ $(i = 1, 2, \cdots, 10)$. $Var(\widehat{\boldsymbol{\beta}}_{FDTWS(j)}) = (\boldsymbol{X}'\boldsymbol{Q}_j\boldsymbol{X})^{-1}$ $(j = 1, 2, 3, 4)$.

Corollary 2 Let $\boldsymbol{\eta} - \boldsymbol{\eta}_0 \xrightarrow{L} N(\boldsymbol{0}, \sigma_{\boldsymbol{\eta}}^2 \boldsymbol{I}), \boldsymbol{\gamma}' - \boldsymbol{\gamma}'_0 \xrightarrow{L} N\left(\boldsymbol{0}, \sigma_{\boldsymbol{\gamma}'}^2 \boldsymbol{I}\right), \boldsymbol{\varepsilon}' \xrightarrow{L} N(\boldsymbol{0}, \sigma_{\boldsymbol{\varepsilon}'}^2 \boldsymbol{I})$, then $\widetilde{\boldsymbol{\beta}}_{FDTW}$, $\widetilde{\boldsymbol{\beta}}_{FDTW(i)}$ $(i = 1, \cdots, 10)$, and $\widehat{\boldsymbol{\beta}}_{FDTWS(j)}$ $(j = 1, 2, 3, 4)$ are asymptotic normal with the expectation $\boldsymbol{\beta}_{FDTW}$ and the variance matrix corresponding to Corollary 1.

In fact, we cannot get the true $\sigma_{\boldsymbol{\eta}}^2, \sigma_{\boldsymbol{\gamma}'}^2$ and $\sigma_{\boldsymbol{\varepsilon}'}^2$ while Theorem 1 assumes the variance matrix of the disturbances are given. On the one hand, we can give these variances by empirical method. One simple way is to let $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$. We have $\sigma_{\boldsymbol{\eta}}^2 = 0$ or $\sigma_{\boldsymbol{\gamma}'}^2 = 0$ for the most estimators in Theorem 1 though the estimators are unbiased and asymptotic normal. These estimators are simplified as $\widetilde{\boldsymbol{\beta}}_{FDTW} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ and $\widetilde{\boldsymbol{\beta}}_{FDTW(i)}$ $(i = 1, 2, \cdots, 10)$ with $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$. On the other hand, we can estimate these variances by methods like the two-stage estimation, see also Baltagi, B. H. (2008). We have the following estimators of the disturbances for Equation (3.2)

$$Q_i\hat{\boldsymbol{u}} = Q_i(Y - X\hat{\beta}) = (Q_i - Q_iX(X'Q_iX)^{-1}X'Q_i)Y = Q'Y, i = 1,2,3.$$

Defining the estimator of $\sigma^2_{RE(i)}$ $(i = 1,2,3)$ as

$$s^2_{RE(i)} = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{m_i} = \frac{(Y - X\hat{\beta})'Q_i(Y - X\hat{\beta})}{m_i} = \frac{Y'Q'Y}{m_i}.$$

where $m_i = rk(Q_i) - rk(Q_iX), Q'$ is symmetric and idempotent.

Corollary 3 $s^2_{RE(i)}$ $(i = 1,2,3)$ are unbiased. And let conditions in Corollary 2 hold, we have

$$\frac{m_i s^2_{RE(i)}}{\sigma^2_{(i)}} \xrightarrow{L} \chi_{m_i}\ i = 1,2,3.$$

Corollary 4 Let conditions in Corollary 2 hold, we have $\boldsymbol{\beta}_{FDTWS(i)}$ $(i = 1,2,3)$ are independent with $\hat{s}^2_{RE(j)}$ $(j = 1,2,3)$, and $\hat{s}^2_{RE(j)}$ $(j = 1,2,3)$ are independent with each other.

Therefore, under Corollaries 3-4, we can estimate $\sigma^2_{\boldsymbol{\eta}}, \sigma^2_{\boldsymbol{\gamma}'}$ and $\sigma^2_{\boldsymbol{\varepsilon}'}$ for the estimators with the unknown true variances of the error components in Theorem 1.

Theorem 2 Let $\boldsymbol{\eta} - \boldsymbol{\eta}_0 \xrightarrow{L} N(\boldsymbol{0}, \sigma^2_{\boldsymbol{\eta}}\boldsymbol{I}), \boldsymbol{\gamma}' - \boldsymbol{\gamma}'_0 \xrightarrow{L} N(\boldsymbol{0}, \sigma^2_{\boldsymbol{\gamma}'}\boldsymbol{I}), \boldsymbol{\varepsilon}' \xrightarrow{L} N(\boldsymbol{0}, \sigma^2_{\boldsymbol{\varepsilon}'}\boldsymbol{I})$, the two-stage estimators of Equation (3.2) are

$$\widehat{\boldsymbol{\beta}}_{FDTW} = \left(\frac{X'Q_1X}{s^2_{RE(1)}} + \frac{X'Q_2X}{s^2_{RE(2)}} + \frac{X'Q_3X}{s^2_{RE(3)}} + \frac{X'Q_4X}{s^2_{RE(4)}}\right)^{-1}\left(\frac{X'Q_1Y}{s^2_{RE(1)}} + \frac{X'Q_2Y}{s^2_{RE(2)}} + \frac{X'Q_3Y}{s^2_{RE(3)}} + \frac{X'Q_4Y}{s^2_{RE(4)}}\right)$$

and

$$\widehat{\boldsymbol{\beta}}_{FDTW(1)} = \left(\frac{X'Q_1X}{s^2_{RE(1)}} + \frac{X'Q_2X}{s^2_{RE(2)}} + \frac{X'Q_3X}{s^2_{RE(3)}}\right)^{-1}\left(\frac{X'Q_1Y}{s^2_{RE(1)}} + \frac{X'Q_2Y}{s^2_{RE(2)}} + \frac{X'Q_3Y}{s^2_{RE(3)}}\right), \cdots, \widehat{\boldsymbol{\beta}}_{FDTW(10)} = \left(\frac{X'Q_3X}{s^2_{RE(3)}} + \frac{X'Q_4X}{s^2_{RE(4)}}\right)^{-1}\left(\frac{X'Q_3Y}{s^2_{RE(3)}} + \frac{X'Q_4Y}{s^2_{RE(4)}}\right). (\ i = 1,2,\cdots,10)$$

where $s^2_{RE(4)} = s^2_{RE(1)} + s^2_{RE(2)} - s^2_{RE(3)}$.

These estimators, characterized by units or time periods, focus on different information in estimation as estimators in Theorem 1 do. The unbiasedness of the above estimators follows

Corollary 5 Let conditions in Corollary 2 hold, then $\widehat{\boldsymbol{\beta}}_{FDTW}$ and $\widehat{\boldsymbol{\beta}}_{FDTW(i)}$ $(i = 1,2,\cdots,10)$ are unbiased estimators of $\boldsymbol{\beta}_{FDTW}$.

The first differential one-way random effects (FDOW) model

Assumption 3.2 The disturbances in Equation (3.1) have the error components as $(\boldsymbol{\eta}, \boldsymbol{\varepsilon}')$.

The OWRE model mentioned in Section 2 under Assumption 3.2 follows

$$Y_{t+1} - Y_t = (X_{t+1} - X_t)\boldsymbol{\beta}_{FDOWU} + \boldsymbol{u}_{FDOWU} \quad (3.3)$$

where $\boldsymbol{u}_{FDOWU} = (\boldsymbol{I}_N \otimes \boldsymbol{1}_{T-1})\boldsymbol{\eta} + \boldsymbol{\varepsilon}'$. The covariance matrix of $\boldsymbol{u}_{FDOWU}$ in Equation 3.3) is $Cov(\boldsymbol{u}_{FDOWU}) = (T-1)\sigma^2_\eta(\boldsymbol{I}_N \otimes \boldsymbol{J}_{T-1}) + \sigma^2_{\varepsilon'}\boldsymbol{I}_{N(T-1)}$. Defining $\sigma^2_1 = (T-1)\sigma^2_{\boldsymbol{\eta}} + \sigma^2_\varepsilon$, $\sigma^2_3 = \sigma^2_{\boldsymbol{\varepsilon}'}$, $\boldsymbol{Q}_1 = \boldsymbol{P}_1 - \boldsymbol{J}_{N(T-1)}, \boldsymbol{Q}_{3'} = \boldsymbol{I}_{N(T-1)} - \boldsymbol{P}_1$, $\boldsymbol{Q}_4 = \boldsymbol{J}_{N(T-1)}$. We have

$$Cov(\boldsymbol{u}_{FDOWU}) = \sigma^2_1\boldsymbol{Q}_1 + \sigma^2_3\boldsymbol{Q}_{3'} + \sigma^2_1\boldsymbol{Q}_4.$$

Theorem 2 Let conditions hold in Proposition 3 and Assumption 3.2, the estimators and their asymptotic properties of the FDOW model in Equation (3.3) are similar to Propositions 1-2 and Corollaries 1-5.

3.2. The Second Differential Random Effects (SDRE) Models

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0569.v1

The SD model gives the characteristic of change in treatment effects. It also excludes the fixed effects of units and time periods. We have the SD model as follows.

$$\Delta \boldsymbol{Y}_{t+1} - \Delta \boldsymbol{Y}_t = (\Delta \boldsymbol{X}_{t+1} - \Delta \boldsymbol{X}_t)\boldsymbol{\beta}_{TWFE} + \boldsymbol{u}. (t = 1, 2, \cdots, T-2) \tag{3.4}$$

where $\Delta \boldsymbol{Y}_t = \boldsymbol{Y}_t - \boldsymbol{Y}_{t-1}, \Delta \boldsymbol{X}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t-1}, \Delta \boldsymbol{\varepsilon}_t = \boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_{t-1}$, and with $E[\boldsymbol{u}] = \boldsymbol{0}$. The covariance matrix of the disturbances $\boldsymbol{u}$ does not discuss here if we combine Equation (2.1) and Equation (3.4) as the SD model. Without loss of generality, I assume the disturbances in Equation (3.4) satisfy the following assumption.

Assumption 3.3 The disturbances in Equation (3.4) have the error components as $(\boldsymbol{\eta}, \boldsymbol{\gamma}'', \boldsymbol{\varepsilon}'')$.

The second differential two-way random effects (SDTW) model: The TWRE model mentioned in Section 2 under Assumption 3.3 follows

$$\Delta \boldsymbol{Y}_t - \Delta \boldsymbol{Y}_{t-1} = (\Delta \boldsymbol{X}_t - \Delta \boldsymbol{X}_{t-1})\boldsymbol{\beta}_{SDTW} + \boldsymbol{u}_{SDTW} \tag{3.5}$$

where $\boldsymbol{u}_{SDTW} = (\boldsymbol{I}_N \otimes \boldsymbol{1}_{T-2})\boldsymbol{\eta} + (\boldsymbol{1}_N \otimes \boldsymbol{I}_{T-2})\boldsymbol{\gamma}'' + \boldsymbol{\varepsilon}''$, $\boldsymbol{\gamma}'' = (\gamma_1, \dots, \gamma_{T-2}), \boldsymbol{\varepsilon}'' = (\varepsilon_{11}, \dots, \varepsilon_{1(T-2)}, \dots, \varepsilon_{N1}, \dots, \varepsilon_{N(T-2)})$. The covariance matrix of $\boldsymbol{u}_{SDTW}$ in Equation (3.5) is

$$\mathrm{Cov}(\boldsymbol{u}_{SDTW}) = (T-2)\sigma_\eta^2(\boldsymbol{I}_N \otimes \boldsymbol{J}_{T-2}) + N\sigma_{\gamma''}^2(\boldsymbol{J}_N \otimes \boldsymbol{I}_{T-2}) + \sigma_{\varepsilon''}^2 \boldsymbol{I}_{N(T-2)}.$$

with $E[\boldsymbol{\eta}] = \boldsymbol{\eta}_0, Cov(\boldsymbol{\eta}) = \sigma_{\boldsymbol{\eta}}^2, E[\boldsymbol{\gamma}''] = \boldsymbol{\gamma}_0'', Cov(\boldsymbol{\gamma}'') = \sigma_{\boldsymbol{\gamma}''}^2 \boldsymbol{I}_{T-2}, E[\boldsymbol{\varepsilon}''] = \boldsymbol{0}, Cov(\boldsymbol{\varepsilon}'') = \sigma_{\boldsymbol{\varepsilon}''}^2 \boldsymbol{I}_{N(T-2)}$. $(\boldsymbol{\eta}, \boldsymbol{\gamma}'', \boldsymbol{\varepsilon}'')$ are independent with each other. Furthermore, defining $\sigma_1^2 = (T-2)\sigma_{\boldsymbol{\eta}}^2 + \sigma_{\boldsymbol{\varepsilon}''}^2$, $\sigma_2^2 = N\sigma_{\boldsymbol{\gamma}''}^2 + \sigma_{\boldsymbol{\varepsilon}''}^2$, $\sigma_3^2 = \sigma_{\boldsymbol{\varepsilon}''}^2$, $\sigma_4^2 = (T-2)\sigma_{\boldsymbol{\eta}}^2 + N\sigma_{\boldsymbol{\gamma}''}^2 + \sigma_{\boldsymbol{\varepsilon}''}^2$, $\boldsymbol{Q}_1 = \boldsymbol{P}_1 - \boldsymbol{J}_{N(T-2)}$, $\boldsymbol{Q}_2 = \boldsymbol{P}_2 - \boldsymbol{J}_{N(T-2)}$, $\boldsymbol{Q}_3 = \boldsymbol{I}_{N(T-2)} + \boldsymbol{J}_{N(T-2)} - \boldsymbol{P}_1 - \boldsymbol{P}_2$, $\boldsymbol{Q}_4 = \boldsymbol{J}_{N(T-2)}$, $\boldsymbol{P}_1 = \boldsymbol{I}_N \otimes \boldsymbol{J}_{(T-2)}$, $\boldsymbol{P}_2 = \boldsymbol{J}_N \otimes \boldsymbol{I}_{T-2}$. We have

$$Cov(\boldsymbol{u}_{SDTW}) = \sigma_1^2 \boldsymbol{Q}_1 + \sigma_2^2 \boldsymbol{Q}_2 + \sigma_3^2 \boldsymbol{Q}_3 + \sigma_4^2 \boldsymbol{Q}_4.$$

Corollary 6 Let conditions hold in Corollary 2 and Assumption 3.3, the estimators and their asymptotic properties of the SDTW model in Equation (3.5) are similar to Theorem 1 and Corollaries 1-5.

The second differential one-way random effects (SDOW) model

Assumption 3.4 The disturbances in Equation (3.4) have the error components as $(\boldsymbol{\eta}, \boldsymbol{\varepsilon}'')$.

The OWRE model mentioned in Section 2 under Assumption 3.4 follows

$$\Delta \boldsymbol{Y}_t - \Delta \boldsymbol{Y}_{t-1} = (\Delta \boldsymbol{X}_t - \Delta \boldsymbol{X}_{t-1})\boldsymbol{\beta}_{SDOWU} + \boldsymbol{u}_{SDOWU} \tag{3.6}$$

where $\boldsymbol{u}_{SDOWU} = (\boldsymbol{I}_N \otimes \boldsymbol{1}_{T-2})\boldsymbol{\eta} + \boldsymbol{\varepsilon}''$. The covariance matrix of $\boldsymbol{u}_{SDOWU}$ in Equation (3.6) is $Cov(\boldsymbol{u}_{SDOWU}) = (T-2)\sigma_{\boldsymbol{\eta}}^2(\boldsymbol{I}_N \otimes \boldsymbol{J}_{T-2}) + \sigma_{\boldsymbol{\varepsilon}''}^2 \boldsymbol{I}_{N(T-2)}$. Defining $\sigma_1^2 = (T-2)\sigma_{\boldsymbol{\eta}}^2 + \sigma_{\boldsymbol{\varepsilon}''}^2$, $\sigma_3^2 = \sigma_{\boldsymbol{\varepsilon}''}^2$, $\boldsymbol{Q}_1 = \boldsymbol{P}_1 - \boldsymbol{J}_{N(T-2)}, \boldsymbol{Q}_{3'} = \boldsymbol{I}_{N(T-2)} - \boldsymbol{P}_1$, $\boldsymbol{Q}_4 = \boldsymbol{J}_{N(T-2)}$. We have

$$\mathrm{Cov}(\boldsymbol{u}_{SDOWU}) = \sigma_1^2 \boldsymbol{Q}_1 + \sigma_3^2 \boldsymbol{Q}_{3'} + \sigma_1^2 \boldsymbol{Q}_4.$$

Corollary 7 Let conditions hold in Proposition 2 and Assumption 3.4, the estimators and their asymptotic properties of the SDOW model in Equation (3.6) are similar to Propositions 1-2 and Corollaries 1-5.

The Higer-order DRE models: The SDRE and FDRE models exclude the fixed effects in regression. They apply to different practical samples. However, the third or other higher-order differences for the error component model are unsuitable for identifying causal effects. These models need strong restrictions on data and will waste a lot of information in the sample. We omitted them here.

The estimations of the DRE models exclude fixed effects and consider random effects. It is more robust than the TWFE, the Differential, or the RE models. The models or estimators which we adopt in estimation depend on the data. The following section will show the optimal model for a given panel.

3.3. Comparisons of the Above Identifications

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0569.v1

We have discussed the advantages and disadvantages of the FE, the RE, and the DRE models in the above. The TWFE model has endogeneity between units, time periods, and regressors, assuming that the disturbances have asymptotic normality. There is no need for intercept terms in regression. However, the TWFE model may have some negative weights in estimation, which causes the treatment effects weights to be negative. Secondly, the RE models need intercept terms so that the exceptions of error components are equal to 0. The RE model requests independence between units, time periods, and regressors. Moreover, the DRE models exclude fixed effects and consider random effects. The DRE models are better than the RE models or the FE models in large samples.

Here, I show the optimal model for different panels in our daily research. I also transform the unbalanced data to the balanced data by methods like interpolation. These data are characterized by the size of samples and the length of time periods. Table 1 shows the optimal model for different types of data.

Table 1. The optimal model for different panels.

	Small sample	Large sample
Short time periods	TWFE/DM	TWFE/DRE
Long time periods	TWFE/DRE	TWRE/DRE

4. Simulation and Application

Here, I show some classes of estimators for identifying the treatment effects. These estimators mentioned in Sections 2 and 3 are classified by the models of the FE, the RE, and the DRE. The simulation data is a staggered design between five groups over five time periods as depicted in Figure 2. There are four groups that are treated at different time periods and one untreated group. The outcomes matrix and the treatment status matrix are

$$\boldsymbol{Y} = \begin{pmatrix} 20 & 30 & 10 & 25 & 15 \\ 32 & 32 & 12 & 27 & 17 \\ 34 & 49 & 14 & 29 & 19 \\ 36 & 51 & 21 & 31 & 21 \\ 38 & 53 & 23 & 45.5 & 23 \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

In this staggered design, the treatment effects of four groups are (10, 15, 5, 12.5). The ATT is 10.625 without the heterogeneity over time periods. On the other hand, the treatment effects of 10 2x2 DIDs are (10, 10, 10, 10, 15, 15, 15, 5, 5, 12.5) under the Goodman-Bacon decomposition theorem. The ATT is 10.75 under the theorem. I use these two ATTs as the best values to compare with estimates of the methods mentioned in Section 2 and Section 3. Tables 2–4 demonstrate that the estimates of the RE models are not ideal because of the ignorance of the fixed effects, especially in such a small sample. However, the estimates of the DRE models are more accurate by excluding the fixed effects at the first step.

The other data comes from Gentzkow and Shapiro (2011) about the effects of the entries and exits of the newspapers on the presidential turnout from 1872 to 1928. Firstly, I use interpolation for the default values where the lead value and lag value are not default. The balanced data are expanded from 847 groups to 905 groups. Secondly, there are negative values in the changes in the newspapers over time periods. I estimate the treatment effects of newspaper changes and the treatment effects of the newspapers' positive changes. Gentzkow and Shapiro (2011) use the FD method to regress the country in which the number of newspapers is positive, while this paper highlights the changes in the newspapers are non-negative. Therefore, we have the straight and positive effects of the newspaper changes on presidential turnout.

In Tables 2–4, the TWFE and the FD method estimates are not significant for the heteroscedasticity in regression, while the SD method is significant. Tables 2–4 also demonstrate that the OWRE method estimates are insignificant. The SDOW method is accurate, and the average effect of the treatment is about 0.25%. The estimates of the TWRE method are significant, but the TWRE method may not be accurate for the treatment effect heterogeneity. The estimates of the FDTW

 doi:10.20944/preprints202407.0569.v1

method are inconsistent, while the estimates of the SDTW method are significant and consistent. In a word, the results in Tables 2–4 demonstrate that one additional newspaper increases the 0.31% average effect of the presidential turnout. Furthermore, there are 12% negative changes in the newspapers in the data. These negative changes may cause the underestimation of the treatment effects. Here, I drop such negative changes and estimate the rest in the newspapers. The TWFE, the OWRE, the TWRE, and the FDTW estimates are about 1.32%, 1.26%, 1.23%, and 1.05%, while the last three methods are significant. The FDTW method is better than others for excluding the fixed and the random effects. Hence, one additional newspaper increases the 1.05% average effect of the presidential turnout when the newspapers' changes are all non-negative.

Table 2. The estimations of the TWFE, the FD, and the SD.

	Staggered design	Newspapers for all changes	Newspapers for positive changes
	10.50		
TWFE		0.0032	0.0132
	(0.1095)	(0.0613)	(0.2130)
	12.63		
FD		0.0019	-0.0065
	(0.0000)	(0.0785)	(0.9644)
SD	10.42	0.0025	0.0036
	(0.0000)	(0.0346)	(0.4448)

Table 3. The estimations of the OWRE, the FDOW and the SDOW.

	$\hat{\beta}_{OWRES(2)}$	$\hat{\beta}_{OWRE}$	$\hat{\beta}_{OWRE(2)}$	$\hat{\beta}_{OWRE(3)}$
Staggered design				
OWRE	15.50	16.35	15.52	16.34
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
FDOW	10.63	12.04	10.63	12.09
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
SDOW	10.31	11.13	11.13	10.31
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Newspapers for all changes				
OWRE	-0.0241	0.0018	0.0018	0.0032
	(1.0000)	(0.0399)	(0.0475)	(0.0021)

 doi:10.20944/preprints202407.0569.v1

FDOW	0.0026	0.0005	0.0039	-0.0010
	(0.0292)	(0.3425)	(0.0018)	(0.7595)
SDOW	0.0024	0.0027	0.0024	0.0028
	(0.0414)	(0.0226)	(0.0420)	(0.0217)
Newspapers for positive changes				
OWRE	-0.0936	0.0126	0.0126	0.0115
	(1.0000)	(0.0021)	(0.0026)	(0.0049)
FDOW	-0.0088	-0.023	-0.0031	-0.0285
	(0.8542)	(0.9996)	(0.6504)	(1.0000)
SDOW	0.0011	0.0031	0.0055	-0.0014
	(0.4523)	(0.3606)	(0.2673)	(0.5601)

Table 4. The estimations of TWRE, FDTW and SDTW.

	$\hat{\beta}_{TWRES(3)}$	$\hat{\beta}_{TWRE}$	$\hat{\beta}_{TWRE(1)}$	$\hat{\beta}_{TWRE(2)}$	$\hat{\beta}_{TWRE(3)}$	$\hat{\beta}_{TWRE(6)}$	$\hat{\beta}_{TWRE(8)}$	$\hat{\beta}_{TWRE(10)}$
Staggered design								
TWRE	10.50	18.26	18.26	10.56	10.56	18.26	18.26	10.50
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
FDTW	10.63	11.50	10.63	11.50	10.63	10.63	10.63	11.53
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Newspapers for all changes								
TWRE	0.0032	0.0022	0.0022	0.0023	0.0032	0.0023	0.0032	0.0032
	(0.0014)	(0.0137)	(0.0137)	(0.0113)	(0.0018)	(0.0113)	(0.0018)	(0.0014)
FDTW	0.0026	0.0037	0.0037	0.0037	0.0026	0.0037	0.0026	0.0026
	(0.0179)	(0.0012)	(0.0012)	(0.0012)	(0.0181)	(0.0012)	(0.0179)	(0.0181)
SDTW	0.0031	0.0030	0.0030	0.0031	0.0031	0.0030	0.0031	0.0031
	(0.0079)	(0.0084)	(0.0084)	(0.0084)	(0.0080)	(0.0084)	(0.0080)	(0.0078)
Newspapers for positive changes								
TWRE	0.0115	0.0043	0.0043	0.0123	0.0033	0.0123	0.0033	0.0115
	(0.0215)	(0.2128)	(0.2128)	(0.0140)	(0.2739)	(0.0140)	(0.2739)	(0.0215)
FDTW	0.0075	0.0102	0.0105	0.0107	0.0066	0.0110	0.0069	0.0071
	(0.1606)	(0.0786)	(0.0732)	(0.0691)	(0.1885)	(0.0642)	(0.1783)	(0.1701)
SDTW	0.0050	0.0084	0.0084	0.0085	0.0050	0.0085	0.0050	0.0051
	(0.2694)	(0.1475)	(0.1475)	(0.1446)	(0.2739)	(0.1445)	(0.2737)	(0.2695)

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0569.v1

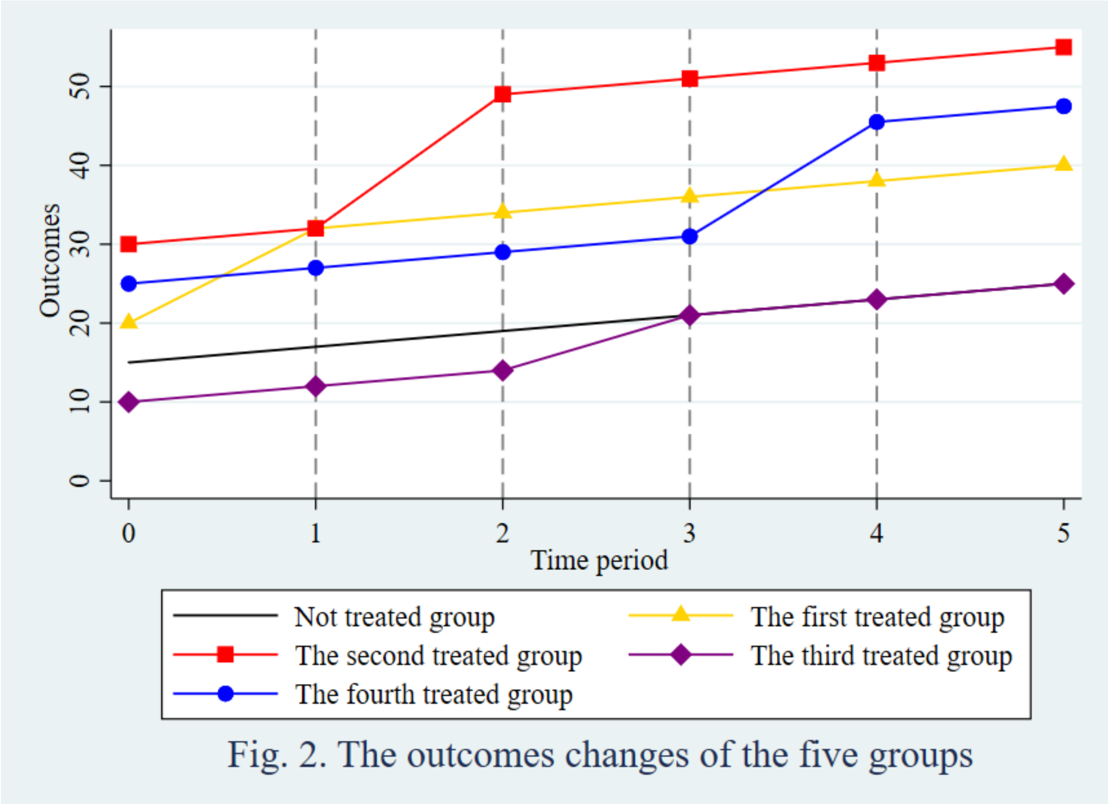

Figure 2. The outcomes changes of the five groups.

5. Conclusion

Identification of the treatment effects usually needs to exclude the fixed effects, the covariates effects, and the disturbances. The fixed and random effects models are popular methods for identifying the causal effects of treatment. Based on abundant theoretical and empirical research of the fixed or the random effects models, this paper introduces explicitly a class of the least square estimators of the OWRE, and the TWRE models. Then, the paper depicts how the treatment effect heterogeneity causes the misidentification of these models. Further, I propose the DRE models, such as the FDOW, the FDTW, the SDOW, and the SDTW models, to address such problems. These models exclude the fixed effects of units and time periods and consider the random effects of units and time periods. Besides, the estimators of the DRE models are consistent and unbiased. The asymptotic properties of these estimators are shown in the Appendix. Meanwhile, I show the staggered design with five groups and revisit the application about the effects of newspapers on presidential turnout. The results demonstrate that the estimates of the DRE models are more significant and robust than the FD, SD, FE, or RE models. I leave the best estimators of the above models for a given data as further research.

Appendix A. Proofs

A.1 Proof of Theorem 1

Firstly, we have the GLS estimator of Equation (3.2) under Lemma 1-2 follows

$$\widetilde{\boldsymbol{\beta}}_{FDTW} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$$

where $\boldsymbol{\Sigma}^{-1} = (\sigma_1^2\boldsymbol{Q}_1 + \sigma_2^2\boldsymbol{Q}_2 + \sigma_3^2\boldsymbol{Q}_3 + \sigma_4^2\boldsymbol{Q}_4)^{-1} = \frac{\boldsymbol{Q}_1}{\sigma_1^2} + \frac{\boldsymbol{Q}_2}{\sigma_2^2} + \frac{\boldsymbol{Q}_3}{\sigma_3^2} + \frac{\boldsymbol{Q}_4}{\sigma_4^2}$. $\sigma_\eta^2, \sigma_\gamma^2$ and σ_ε^2 are given.

Secondly, we have following four equations by multiplying Equation (3.2) with $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{Q}_3$ and $\boldsymbol{Q}_4$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0569.v1

$$\boldsymbol{Q}_j\boldsymbol{Y}=\boldsymbol{Q}_j\boldsymbol{X}\boldsymbol{\beta}_{FDTWS(j)}+\boldsymbol{Q}_j\boldsymbol{u}. \tag{A1}$$

where $\boldsymbol{Q}_j\boldsymbol{1}_{N(T-1)}=\boldsymbol{0},(j=1,2,3,4)$. Under Lemma 1-2, the GLS estimators of Equation (A1) are $\widehat{\boldsymbol{\beta}}_{FDTWS(i)}=(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{Y}),(i=1,2,3,4)$ where $\widehat{\boldsymbol{\beta}}_{FDTWS(1)}$ is the Between units estimator and $\widehat{\boldsymbol{\beta}}_{FDTWS(2)}$ is the Between time periods estimator. $\widehat{\boldsymbol{\beta}}_{FDTWS(4)}$ charactered nothing but constant is too large to become significant. These estimators are omitted for insufficient regression information. Furthermore, we obtain other 10 GLS estimators with combination of equations in Equation (10). Here I show the proof of $\boldsymbol{\beta}_{FDTW(1)}$ while the proof of other estimators is similar. We have partitioned matrix equation by combining first three equations in Equation (A2) as follow.

$$\begin{pmatrix}\boldsymbol{Q}_1\boldsymbol{Y}\\ \boldsymbol{Q}_2\boldsymbol{Y}\\ \boldsymbol{Q}_3\boldsymbol{Y}\end{pmatrix}_{3NT\times1}=\begin{pmatrix}\boldsymbol{Q}_1\boldsymbol{X}\\ \boldsymbol{Q}_2\boldsymbol{X}\\ \boldsymbol{Q}_3\boldsymbol{X}\end{pmatrix}_{3NT\times1}\beta_{FDTW(1)}+\begin{pmatrix}\boldsymbol{Q}_1\boldsymbol{u}\\ \boldsymbol{Q}_2\boldsymbol{u}\\ \boldsymbol{Q}_3\boldsymbol{u}\end{pmatrix}_{3NT\times1}$$

The GLS estimator of the partitioned matrix equation is

$$\widetilde{\boldsymbol{\beta}}_{FDTW(1)}=\left[\begin{pmatrix}\boldsymbol{Q}_1\boldsymbol{X}\\ \boldsymbol{Q}_2\boldsymbol{X}\\ \boldsymbol{Q}_3\boldsymbol{X}\end{pmatrix}'\boldsymbol{\Sigma}_{FDTW(1)}^{-1}\begin{pmatrix}\boldsymbol{Q}_1\boldsymbol{X}\\ \boldsymbol{Q}_2\boldsymbol{X}\\ \boldsymbol{Q}_3\boldsymbol{X}\end{pmatrix}\right]^{-1}\begin{pmatrix}\boldsymbol{Q}_1\boldsymbol{X}\\ \boldsymbol{Q}_2\boldsymbol{X}\\ \boldsymbol{Q}_3\boldsymbol{X}\end{pmatrix}'\boldsymbol{\Sigma}_{FDTW(1)}^{-1}\begin{pmatrix}\boldsymbol{Q}_1\boldsymbol{Y}\\ \boldsymbol{Q}_2\boldsymbol{Y}\\ \boldsymbol{Q}_3\boldsymbol{Y}\end{pmatrix}=$$

$$\left[(\boldsymbol{X}'\boldsymbol{Q}_1\quad \boldsymbol{X}'\boldsymbol{Q}_2\quad \boldsymbol{X}'\boldsymbol{Q}_3)\begin{pmatrix}\frac{1}{\sigma_1^2}\boldsymbol{Q}_1^- & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \frac{1}{\sigma_2^2}\boldsymbol{Q}_2^- & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \frac{1}{\sigma_3^2}\boldsymbol{Q}_3^-\end{pmatrix}\begin{pmatrix}\boldsymbol{Q}_1\boldsymbol{X}\\ \boldsymbol{Q}_2\boldsymbol{X}\\ \boldsymbol{Q}_3\boldsymbol{X}\end{pmatrix}\right]^{-1}(\boldsymbol{X}'\boldsymbol{Q}_1\quad \boldsymbol{X}'\boldsymbol{Q}_2\quad \boldsymbol{X}'\boldsymbol{Q}_3)\begin{pmatrix}\frac{1}{\sigma_1^2}\boldsymbol{Q}_1^- & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \frac{1}{\sigma_2^2}\boldsymbol{Q}_2^- & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \frac{1}{\sigma_3^2}\boldsymbol{Q}_3^-\end{pmatrix}\begin{pmatrix}\boldsymbol{Q}_1\boldsymbol{Y}\\ \boldsymbol{Q}_2\boldsymbol{Y}\\ \boldsymbol{Q}_3\boldsymbol{Y}\end{pmatrix}=$$

$$\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}\right),$$

where $\boldsymbol{Q}_1^-$ is generalized inverse of $\boldsymbol{Q}_1$. And

$$\boldsymbol{\Sigma}_{FDTW(1)}=Cov\begin{pmatrix}\boldsymbol{Q}_1\boldsymbol{u}\ \boldsymbol{Q}_1\boldsymbol{u}\\ \boldsymbol{Q}_2\boldsymbol{u},\boldsymbol{Q}_2\boldsymbol{u}\\ \boldsymbol{Q}_3\boldsymbol{u}\ \boldsymbol{Q}_3\boldsymbol{u}\end{pmatrix}=\begin{pmatrix}\boldsymbol{Q}_1\\ \boldsymbol{Q}_2\\ \boldsymbol{Q}_3\end{pmatrix}Cov(\boldsymbol{u},\boldsymbol{u})\begin{pmatrix}\boldsymbol{Q}_1\\ \boldsymbol{Q}_2\\ \boldsymbol{Q}_3\end{pmatrix}'=\begin{pmatrix}\sigma_1^2\boldsymbol{Q}_1 & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \sigma_2^2\boldsymbol{Q}_2 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \sigma_3^2\boldsymbol{Q}_3\end{pmatrix}.$$

Similarly, we have other GLS estimators for partitioned matrix equation of combination of equations in Equation (A2) as $\widetilde{\boldsymbol{\beta}}_{FDTW(2)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right)$, $\widetilde{\boldsymbol{\beta}}_{FDTW(3)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right)$, $\widetilde{\boldsymbol{\beta}}_{FDTW(4)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right)$, $\widetilde{\boldsymbol{\beta}}_{FDTW(5)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{D}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{D}}{\sigma_2^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2}\right)$, $\widetilde{\boldsymbol{\beta}}_{FDTW(6)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}\right)$, $\widetilde{\boldsymbol{\beta}}_{FDTW(7)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right)$, $\widetilde{\boldsymbol{\beta}}_{FDTW(8)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}\right)$, $\widetilde{\boldsymbol{\beta}}_{FDTW(9)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right)$, $\widetilde{\boldsymbol{\beta}}_{FDTW(10)}=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right)$. ■

A.2 Proof of Corollary 1

$$E[\tilde{\boldsymbol{\beta}}_{FDTW}] = E\left[\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{Y}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{Y}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{Y}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{Y}}{\sigma_4^2}\right)\right]$$

$$=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\boldsymbol{X}'\left(\frac{\boldsymbol{Q}_1}{\sigma_1^2}+\frac{\boldsymbol{Q}_2}{\sigma_2^2}+\frac{\boldsymbol{Q}_3}{\sigma_3^2}+\frac{\boldsymbol{Q}_4}{\sigma_4^2}\right)E(\boldsymbol{Y}) = \boldsymbol{\beta}_{FDTW}.$$

$$Var(\tilde{\boldsymbol{\beta}}_{FDTW}) = Var\left[\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\boldsymbol{X}'\left(\frac{\boldsymbol{Q}_1}{\sigma_1^2}+\frac{\boldsymbol{Q}_2}{\sigma_2^2}+\frac{\boldsymbol{Q}_3}{\sigma_3^2}+\frac{\boldsymbol{Q}_4}{\sigma_4^2}\right)\boldsymbol{Y}\right]$$

$$=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}\boldsymbol{X}'\left(\frac{\boldsymbol{Q}_1}{\sigma_1^2}+\frac{\boldsymbol{Q}_2}{\sigma_2^2}+\frac{\boldsymbol{Q}_3}{\sigma_3^2}+\frac{\boldsymbol{Q}_4}{\sigma_4^2}\right)Var(\boldsymbol{Y})\left(\frac{\boldsymbol{Q}_1}{\sigma_1^2}+\frac{\boldsymbol{Q}_2}{\sigma_2^2}+\frac{\boldsymbol{Q}_3}{\sigma_3^2}+\frac{\boldsymbol{Q}_4}{\sigma_4^2}\right)\boldsymbol{X}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}$$

$$=\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{\sigma_1^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{\sigma_2^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{\sigma_3^2}+\frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{\sigma_4^2}\right)^{-1}.$$

The unbiasedness and the variance matrix of $\tilde{\boldsymbol{\beta}}_{FDTW(i)}$ $(i = 1,\cdots,10)$ and $\hat{\boldsymbol{\beta}}_{FDTWS(j)}$ $(j = 1,2,3,4)$ are similar with the above proof and are omitted here. ■

A.3 Proof of Corollary 3

$$E\left(s_{RE(i)}^2\right) = E\left(\frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{m_i}\right) = E\left[\frac{\boldsymbol{Y}'(\boldsymbol{Q}_i-\boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i)\boldsymbol{Y}}{m_i}\right]$$
$$= \frac{1}{m_i}\{E(\boldsymbol{Y}')(\boldsymbol{Q}_i-\boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i)E(\boldsymbol{Y}) + Tr[(\boldsymbol{Q}_i-\boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i)Cov(\boldsymbol{Y})]\}$$
$$= \frac{\sigma_i^2}{m_i}\{\mathrm{tr}(\boldsymbol{Q}_i) - \mathrm{tr}(\boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i)\}$$
$$= \frac{\sigma_i^2}{m_i}[\mathrm{rk}(\boldsymbol{Q}_i) - \mathrm{rk}(\boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i)] = \sigma_i^2.$$

where $\mathrm{rk}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X}) = \mathrm{rk}[(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i] = \mathrm{rk}[\boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i] = k$. The third equality follows from the expectation of the matrix of a quadratic form. The fourth equality follows from the trace of

 doi:10.20944/preprints202407.0569.v1

the matrix. The fifth equality follows from the rank equals to the trace under a symmetric and idempotent matrix.

Secondly, we have

$$\frac{m_i s_{RE(i)}^2}{\sigma_{RE(i)}^2} = \frac{\boldsymbol{Y}'\boldsymbol{Q}_i(\boldsymbol{Q}_i - \boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i)\boldsymbol{Q}_i\boldsymbol{Y}}{\sigma_{RE(i)}^2}$$

$$= \boldsymbol{Y}'\boldsymbol{\Sigma}^{-1/2}\frac{\boldsymbol{\Sigma}^{1/2}(\boldsymbol{Q}_i - \boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i)\boldsymbol{\Sigma}^{1/2}}{\sigma_{RE(i)}^2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y},$$

where $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y} \sim N\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}\boldsymbol{\beta}_{FDTW}, \boldsymbol{I}\right)$, $\frac{\boldsymbol{\Sigma}^{1/2}\left(\boldsymbol{Q}_i - \boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i\right)\boldsymbol{\Sigma}^{1/2}}{\sigma_{RE(i)}^2}$ is a symmetric and idempotent matrix. The quadratic form with normal vectors has a chi-square distribution if its matrix is symmetric and idempotent. Hence,

$$\frac{m_i s_{RE(i)}^2}{\sigma_{RE(i)}^2} \xrightarrow{L} \chi_{m_i}, i = 1, 2, 3.$$

where $m_i = \mathrm{rk}\left[\frac{\boldsymbol{\Sigma}^{1/2}\left(\boldsymbol{Q}_i - \boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i\right)\boldsymbol{\Sigma}^{1/2}}{\sigma_{RE(i)}^2}\right] = \mathrm{rk}(\boldsymbol{Q}_i) - \mathrm{rk}(\boldsymbol{Q}_i\boldsymbol{X})$. ■

A.4 Proof of Corollary 4

Given that $m_i s_{RE(i)}^2$ has a quadratic form with normal vectors, It is easy to show,

$$(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{\Sigma}\frac{\boldsymbol{Q}_j - \boldsymbol{Q}_j\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{Q}_j\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Q}_j}{m_j} = 0, i = 1, 2, 3. j = 1, 2, 3$$

$$\frac{\boldsymbol{Q}_i - \boldsymbol{Q}_i\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{Q}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_i}{m_i}\boldsymbol{\Sigma}\frac{\boldsymbol{Q}_j - \boldsymbol{Q}_j\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{Q}_j\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Q}_j}{m_j} = 0. 1 \le i < j \le 3$$

Then $\widehat{\boldsymbol{\beta}}_{FDTWS(i)}$ $(i = 1, 2, 3)$ are independent with $\hat{s}_{RE(j)}^2$ $(j = 1, 2, 3)$ and $\hat{s}_{RE(j)}^2$ $(j = 1, 2, 3)$ are independent with each other under the properties of linear independence of the quadratic form with normal vectors. ■

A.5 Proof of Theorem 2

The proof of Theorem 2 can be obtained directly under Theorem 1 and Corollaries 3-4.

A.6 Proof of Corollary 5

$$E\left(\widehat{\boldsymbol{\beta}}_{FDTW} - \boldsymbol{\beta}_{FDTWS}\right) = E\left[E\left(\widehat{\boldsymbol{\beta}}_{FDTW} - \boldsymbol{\beta}_{FDTWS}\middle| s_{RE(1)}^2, s_{RE(2)}^2, s_{RE(3)}^2\right)\right]$$

$$= E\left[\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{X}}{s_{RE(1)}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{X}}{s_{RE(2)}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{X}}{s_{RE(3)}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{X}}{s_{RE(4)}^2}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{Q}_1\boldsymbol{u}}{s_{RE(1)}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_2\boldsymbol{u}}{s_{RE(2)}^2}\right.\right.$$

$$\left.\left. + \frac{\boldsymbol{X}'\boldsymbol{Q}_3\boldsymbol{u}}{s_{RE(3)}^2} + \frac{\boldsymbol{X}'\boldsymbol{Q}_4\boldsymbol{u}}{s_{RE(4)}^2}\right)\middle| s_{RE(1)}^2, s_{RE(2)}^2, s_{RE(3)}^2\right] = 0.$$

where the first equality follows from the iterated expectation theorem. The second equality is straightforward. The proof of unbiasedness of $\widehat{\boldsymbol{\beta}}_{FDTW(i)}$ $(i = 1, 2, \cdots, 10)$ are similar and omitted here. ■

Appendix B. Supplementary data

 doi:10.20944/preprints202407.0569.v1

Supplementary material related to this manuscript can be found in the Supplementary Material.zip.

References

1. Abadie, A. (2005), "Semiparametric difference-in-differences estimators," *The review of economic studies*, 72 (1), 1-19.
2. Angrist, J. D., and Imbens, G. W. (1995), "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of the American statistical Association*, 90 (430), 431-442.
3. Arkhangelsky, D., and Imbens, G. W. (2022), "Doubly robust identification for causal panel data models," *The Econometrics Journal*, 25 (3), 649-674.
4. Baltagi, B. H. (1981), "Pooling: An experimental study of alternative testing and estimation procedures in a two-way error component model," *Journal of Econometrics*, 17 (1), 21-49.
5. Baltagi, B. H., and Baltagi, B. H. (2008), *Econometric analysis of panel data* (Vol. 4): Springer.
6. Baltagi, B. H., Song, S. H., and Jung, B. C. (2001), "The unbalanced nested error component regression model," *Journal of Econometrics*, 101 (2), 357-381.
7. Baltagi, B. H., Song, S. H., and Jung, B. C. (2002), "A comparative study of alternative estimators for the unbalanced two-way error component regression model," *The Econometrics Journal*, 5 (2), 480-493.
8. Casella, G., and Berger, R. L. (2002), *Statistical Inference*: Thomson Learning.
9. Chamberlain, G. (1984), "Panel data," *Handbook of econometrics*, 2, 1247-1318.
10. de Chaisemartin, C., and d'Haultfoeuille, X. (2020), "Two-way fixed effects regressions with several treatments," *arXiv preprint arXiv:2012.10077*.
11. De Chaisemartin, C., and d'Haultfoeuille, X. (2020), "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 110 (9), 2964-2996.
12. --- (2023), "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey," *The Econometrics Journal*, 26 (3), C1-C30.
13. Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2011), "The effect of newspaper entry and exit on electoral politics," *American Economic Review*, 101 (7), 2980-3018.
14. Goodman-Bacon, A. (2021), "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225 (2), 254-277.
15. Hausman, J. A. (1978), "Specification tests in econometrics," *Econometrica: Journal of the econometric society*, 1251-1271.
16. Hausman, J. A., and Taylor, W. E. (1981), "Panel data and unobservable individual effects," *Econometrica: Journal of the econometric society*, 1377-1398.
17. Imai, K., and Kim, I. S. (2021), "On the use of two-way fixed effects regression models for causal inference with panel data," *Political Analysis*, 29 (3), 405-415.
18. Liu, C., and Sun, Y. (2019), "A simple and trustworthy asymptotic t test in difference-in-differences regressions," *Journal of Econometrics*, 210 (2), 327-362.
19. Maddala, G. S. (1971), "The use of variance components models in pooling cross section and time series data," *Econometrica: Journal of the econometric society*, 341-358.
20. Metcalf, G. E. (1996), "Specification testing in panel data with instrumental variables," *Journal of Econometrics*, 71 (1-2), 291-307.
21. Mundlak, Y. (1978), "On the pooling of time series and cross section data," *Econometrica: Journal of the econometric society*, 69-85.
22. Pereda-Fernández, S. (2021), "Copula-Based Random Effects Models for Clustered Data," *Journal of Business & Economic Statistics*, 39 (2), 575-588. DOI: 10.1080/07350015.2019.1688665.
23. Swamy, P., and Arora, S. S. (1972), "The exact finite sample properties of the estimators of coefficients in the error components regression models," *Econometrica: Journal of the econometric society*, 261-275.
24. Van der Vaart, A. W. (2000), *Asymptotic statistics* (Vol. 3): Cambridge university press.
25. Wallace, T. D., and Hussain, A. (1969), "The use of error components models in combining cross section with time series data," *Econometrica: Journal of the econometric society*, 55-72.
26. Wooldridge, J. M. (2021), "Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators," *Available at SSRN 3906345*.