Unifying DiD Estimators and TWFE Regression via Network Flows for Arbitrary Treatment Patterns and Heterogeneous Effects*

Yudong Chen[1], Xumei Xi[2], and Christina Lee Yu[2]

[1]University of Wisconsin-Madison, [2]Cornell University

October 19, 2024

Abstract

Two-Way Fixed Effect (TWFE) regressions are commonly used to analyze panel data for the purpose of estimating an average treatment effect. While TWFE regressions are well understood for simple treatment patterns such as staggered adoption, there is still a gap in our understanding when it comes to arbitrary treatment patterns. Many of the guarantees assume homogeneous treatment effects, which is unrealistic in practice. In this work, we show that for any *arbitrary treatment pattern*, when the outcomes under both treatment and control satisfy an additive fixed effects model, then an unbiased estimate for the outcome of a unit-time pair (i, t) under treatment or control can be constructed by weighting observations according to any (i, t)-unit flow in the graphs associated to the treatment and control observations respectively. This introduces a family of estimators parameterized by flows over networks, where flows restricted to length-3 paths correspond to the difference-in-differences (DiD) estimator. We show that the variance minimizing flow is the electrical flow, and it is equivalent to a variation of the TWFE regression estimator for *heterogeneous treatment effects*. The connectivity of the networks associated to the treatment pattern and control pattern govern the statistical properties of the electrical flow estimator (EFE) through the effective resistance, which gives a fine-grained and intuitive understanding of the impact of the given treatment pattern on estimation. This estimator is locally minimax optimal and is also the uniform minimum variance unbiased estimator for each individual treatment effect.

1 Introduction

Standard estimators for panel data often assume simple block or staggered treatment patterns [1, 2], whether directly in the construction of the estimator or in the analysis of their performance. However, for experiments arising from online platforms, the set of observed data and the treatment patterns may be highly irregular and unstructured. Another common practice is to focus on estimating average effects, which overlooks the underlying heterogeneity in the data. A recent line of work [7, 3, 11, 5] allows heterogeneous treatment effects and assumes certain structures of the effects. In this paper, we address the challenge of arbitrary treatment patterns and heterogeneous treatment effects, uncovering a surprising yet intuitive connection between DiD estimators and TWFE regression. There has been a flurry of work [13, 7, 10] drawing connections to better understand TWFE regression. However, we are the first to interpret these estimators using network flows, which carries physical meaning and paves the way for future exploration.

*For a complete version of this paper with proofs see [4].

2 Model Setup

Consider a population of $[n]$ units and $[T]$ time periods. Let $\Omega \subseteq [n] \times [T]$ be the set of entries that are observed. X_{it} denotes the treatment variable where 1 indicates treatment and 0 indicates control. For pairs $(i,t) \in \Omega$, we observe the outcome Y_{it}, which is a function of the treatment X_{it}. We assume a *heterogeneous additive two-way fixed effects model*, in which outcomes Y_{it} can be described as

$$Y_{it} = \alpha_i + \gamma_t + \beta_{it} X_{it} + E_{it} \quad \text{and} \quad \beta_{it} = \mu_i + \nu_t, \tag{1}$$

where E_{it} represent errors that have mean zero and are independent of X_{it}. Note that the outcomes under both treatment and control satisfy the parallel trends assumption, as they can be decomposed into an additive unit fixed effect and time fixed effect. Our model allows the treatment effect to be heterogeneous as long as it satisfies the additive fixed effect structure. The outcomes observed under control can be viewed as noisy observations of the matrix $F^*_{it} = \alpha_i + \gamma_t$, where the observation pattern is given by $\Omega^0 = \Omega \cap \{(i,t) : X_{it} = 0\}$. Similarly, the outcomes under treatment can be viewed as noisy observations of the matrix $G^*_{it} = \alpha_i + \mu_i + \gamma_t + \nu_t$, where the observation pattern is given by $\Omega^1 = \Omega \cap \{(i,t) : X_{it} = 1\}$. Estimating the individual causal effect $\beta_{it} = \mu_i + \nu_t = G^*_{it} - F^*_{it}$ is equivalent to the task of matrix estimation to recover F^* and G^*, when F^* and G^* are each given by the sum of row fixed effects and column fixed effects.

To introduce the estimator and theoretical results, we consider the general matrix estimation task for a matrix satisfying the additive model: We are given a noisy and partially observed matrix $M = \Omega \circ (M^* + E)$, where $E \in \mathbb{R}^{n \times m}$ is additive noise, $\Omega \in \{0,1\}^{n \times m}$ is the observation matrix, and $M^*_{ij} = a^*_i + b^*_j$ for latent fixed effects a^* and b^*. Our results will be conditioned on Ω, thus Ω can be correlated with a^* and b^*.

3 Estimator

We introduce a family of unbiased estimators parameterized by flows on an undirected bipartite graph $\mathcal{G}(\Omega) = (\mathcal{V}, \Omega)$, where the vertex set $\mathcal{V} = \{u_i\}_{i \in [n]} \cup \{v_j\}_{j \in [m]}$ represent the n rows and m columns, and the edge set is given by Ω. When applied to the panel data setting, we construct a treatment graph associated to the observation pattern under treatment as given by Ω^1, and a control graph associated to the observation pattern Ω^0. Figure 1 shows the construction of the treatment and control graphs from a given treatment pattern.

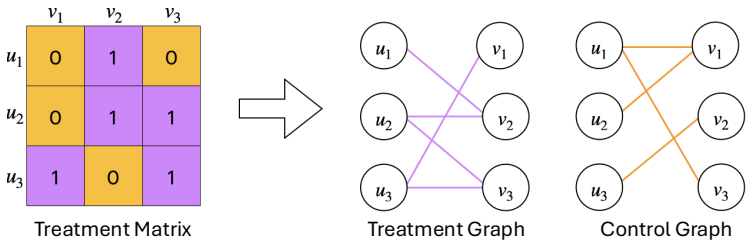

Figure 1: Example construction of treatment graph and control graph for a given treatment pattern.

Consider a concrete example as depicted in Figure 2(a), where we want to estimate entry $(1,1)$ in a 3-by-3 matrix, with the given Ω. There is a simple path connecting u_1 and v_1 in the corresponding bipartite graph. An unbiased estimate of M^*_{11} can be constructed by alternating between adding and subtracting the observations on this path: $\hat{M}_{11} = M_{12} - M_{22} + M_{23} - M_{33} + M_{31}$, its expectation being $(a^*_1 + b^*_2) - (a^*_2 + b^*_2) + (a^*_2 + b^*_3) - (a^*_3 + b^*_3) + (a^*_3 + b^*_1) = a^*_1 + b^*_1 = M^*_{11}$. When there are

multiple paths in the graph, it is natural to aggregate the estimates obtained from each path to make the most out of available information. We can achieve that by considering network flows, as any unit flow on the graph from u_i to v_j can be decomposed as a convex combination of paths from u_i to v_j. Consequently, we can construct an unbiased estimator for a unit flow by weighting the individual path estimators accordingly.

Let $f \in \mathbb{R}^{n_e}$ be a (u_i, v_j) unit flow on $\mathcal{G}(\Omega)$. The *unit flow estimator* corresponding to f is given by

$$\hat{M}_{ij}^{f} = \sum_{(k,\ell)\in[n]\times[m]} f_{u_k,v_\ell}\Omega_{k\ell}M_{k\ell}, \tag{2}$$

which computes a weighted sum of the observations according to f. Within the class of unbiased flow estimators, we can find the flow that minimizes the variance, for which there is a surprisingly simple answer. If we view the observation graph as an electrical network as depicted in Figure 2(b), then the variance of the flow estimator is proportional to the electrical energy of that flow. Therefore, the electrical flow estimator (EFE), obtained by letting f be equal to the unit electrical flow, is the optimal variance-minimizing flow estimator as it is the unique unit flow that minimizes the electrical energy by Thomson's principle. The variance of the resulting EFE is proportional to the effective resistance[1], which measures the connectivity between u_i and v_j [12, 9]. This connects to a rich literature in physics and computer science that studies properties of network flows, electrical flows, and effective resistance as a function of the network structure [6].

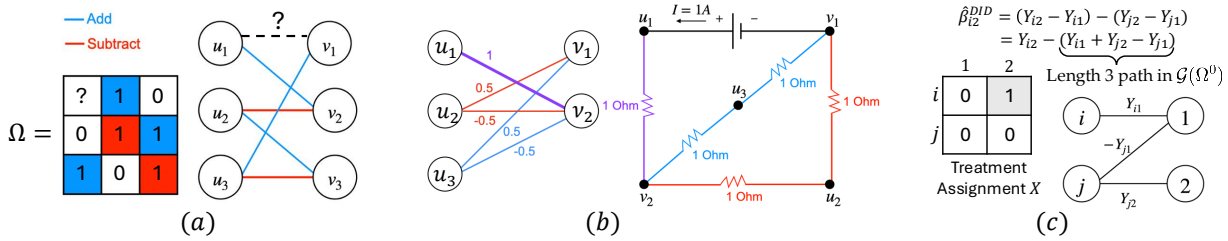

Figure 2: *(a) An example path estimator. (b) The (u_1, v_1) electrical flow sends a unit of current from u_1 to v_1 on the electrical network constructed by treating each edge in the graph as a unit resistor, putting higher weight on edges that appear on multiple paths. (c) Equivalence of DiD estimators to flow estimators constrained to length 3 paths.*

For the TWFE model (1), the individual treatment effect can be estimated by subtracting the EFEs obtained from the treatment graph $\mathcal{G}(\Omega^1)$ and the control $\mathcal{G}(\Omega^0)$: $\hat{\beta}_{it}^{\text{EFE}} = \hat{G}_{it}^{\text{EFE}} - \hat{F}_{it}^{\text{EFE}}$. Figure 2(c) illustrates an example where the estimator is the difference between the observed treatment outcome and the estimated control outcome using a length-3 path. Interestingly, this is equivalent to the DiD estimator. The general class of flow estimators offers more flexibility when the treatment and observation pattern may be unstructured. When there does not exist connecting paths of length 3 or shorter, the DiD estimator is unable to provide any estimate, while the flow estimator shows that the heterogeneous treatment effect β_{it} is identifiable as long as i and t are connected in both the treatment and control graphs.

[1]The effective resistance between s and t in an electrical network is equal to the potential difference that appears across s and t when a unit current source is applied between them. The effective resistance is small when there are many short paths between s and t, and it is large when there are few paths between s and t that tend to be long.

4 Theoretical Results

For additive models, under the assumption that E_{ij}'s are i.i.d. $\mathcal{N}(0, \sigma^2)$, we provide an error upper bound for the EFE in Theorem 1, which matches the local minimax lower bound presented in Theorem 2. This implies the EFE is optimal for every instance, which is stronger than the usual worst case minimax lower bound. The theorems are stated for a single matrix estimation task under an additive model with a fixed observation pattern Ω. $R_\Omega(i, j)$ denotes the effective resistance between i and j in the network represented by Ω. We will subsequently apply these theorems to the TWFE models by estimating separately on the treatment and control datasets corresponding to the networks Ω^0 and Ω^1 respectively.

Theorem 1. *With probability at least* $1-\delta$*, we have* $\left(\hat{M}_{ij}^{\mathrm{EFE}} - M_{ij}^*\right)^2 \le 2\sigma^2 R_\Omega(u_i, v_j) \log(2nm/\delta), \forall (i, j)$.

Theorem 2. *For a fixed observation* Ω*, a connected pair* (i, j)*, and additive models* Q^*, Q'*,*

$$\sup_{Q'} \inf_{\hat{M}} \max_{M^* \in \{Q^*, Q'\}} \mathbb{E}\left[(\hat{M}_{ij} - M_{ij}^*)^2\right] \ge \tfrac{2}{27}\sigma^2 R_\Omega(u_i, v_j). \tag{3}$$

Additionally, we can show that EFE is equivalent to the least squares estimator (LSE), which is given by $\hat{M}_{ij}^{\mathrm{LSE}} = \hat{a}_i + \hat{b}_j$ for $(\hat{a}, \hat{b}) = \operatorname{argmin}_{(a,b)\in\mathbb{R}^{nv}} \; f(a, b) := \left\|\Omega \circ (a\mathbf{1}_n^\top + \mathbf{1}_m b^\top - M)\right\|_F^2$.

Theorem 3. *If* u_i *and* v_j *are connected in* $\mathcal{G}(\Omega)$ *for* $(i, j) \in [n] \times [m]$*, we have* $\hat{M}_{ij}^{\mathrm{LSE}} = \hat{M}_{ij}^{\mathrm{EFE}}$.

Theorem 3 implies that EFE minimizes variance among both all flow estimators and all unbiased estimators. Applying Theorem 1 to the TWFE model yields

$$\left(\hat{\beta}_{it}^{\mathrm{EFE}} - \beta_{it}\right)^2 \le C\sigma^2 \left[R_{\Omega^0}(u_i, v_t) + R_{\Omega^1}(u_i, v_t)\right] \log(NT/\delta),$$

with probability at least $1-\delta$. The equivalence between EFE and LSE implies that the EFE matches the modified TWFE regression, allowing for heterogeneous treatment effects (with additive fixed effects),

$$(\hat{\mu}, \hat{\nu}) = \operatorname{argmin}_{\mu,\nu,\alpha,\gamma} \; \textstyle\sum_{i,t} \Omega_{it}^0 (\alpha_i + \gamma_t - Y_{it})^2 + \sum_{i,t} \Omega_{it}^1 (\alpha_i + \gamma_t + \mu_i + \nu_t - Y_{it})^2. \tag{4}$$

Adding the estimates yields $\hat{\beta}_{it}^{\mathrm{LSE}} = \hat{\mu}_i + \hat{\nu}_t$. Since LSE is equivalent to EFE, we have $\hat{\beta}_{it}^{\mathrm{EFE}} = \hat{\beta}_{it}^{\mathrm{LSE}}$ and (4) can be solved by using the electrical flows from the treatment graph and the control graph. We can thus interpret TWFE regression estimates as a sophisticated method for combining paths–potentially longer and overlapping ones–not just those of length-3 as in DiD. Because of this flexibility, EFE (and LSE) can recover treatment effects for arbitrary treatment patterns, even when DiD fails due to missing length-3 paths. This offers a clean interpretation of the minimum variance by leveraging the notion of effective resistance in electrical networks. In an electrical network, effective resistance between two vertices is lower if there are multiple short paths connecting them, indicating greater connectivity. [8] has previously showed that graph connectivity as characterized by the spectral gap and overlap of direct neighbors is key to estimation accuracy of TWFE regression for fixed effect models. Our result further gives an explicit and exact equivalence of the variance of the best estimator with the effective resistance, providing fine-grained entry-specific results.

5 Experiments

We conduct synthetic experiments with a sparsified staircase treatment pattern as depicted in Figure 3(a). No length-3 path exists in the graph and as a result, DiD fails in this scenario. For

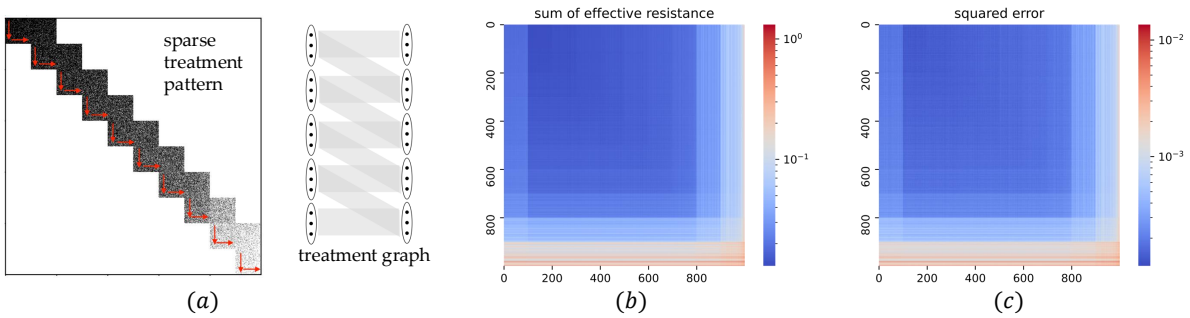

Figure 3: *(a) The treatment pattern has a staircase structure which gets progressively sparser to the bottom right; the corresponding treatment graph has a zig-zag pattern. (b) Heatmap of the sum of effective resistances in the graphs corresponding to treated and control observations. (c) Heatmap of the entrywise squared estimation error.*

the fixed treatment pattern, Figure 3(b) shows a heatmap of the sum of the effective resistances in the graphs corresponding to the treatment and control observations. We generate row and column fixed effects from a Gaussian distribution and set the variance of the observation noise as $\sigma^2 = 0.01$. For a given realization of the observations we compute the corresponding electrical flow estimates $\hat{\beta}_{it}^{\text{EFE}}$ for all pairs (i, t). We repeat the experiment 1000 times and compute the entrywise mean squared error for each entry. Figure 3(c) shows the entrywise squared error of the EFE. The results confirm that the estimation accuracy is proportional to the sum of the effective resistance from the treatment and control graphs as stated in Theorem 1.

References

[1] Susan Athey, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730, 2021.

[2] Susan Athey and Guido W Imbens. Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1):62–79, 2022.

[3] Brantly Callaway and Pedro HC Sant'Anna. Difference-in-differences with multiple time periods. *J of Econometrics*, 225(2), 2021.

[4] Yudong Chen, Xumei Xi, and Christina Lee Yu. Entry-specific matrix estimation under arbitrary sampling patterns through the lens of network flows. *arXiv preprint arXiv:2409.03980*, 2024.

[5] Clément De Chaisemartin and Xavier d'Haultfoeuille. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, 26(3):C1–C30, 2023.

[6] Arpita Ghosh, Stephen Boyd, and Amin Saberi. Minimizing effective resistance of a graph. *SIAM Review*, 50(1):37–66, 2008.

[7] Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. 225(2):254–277, 2021.

[8] Koen Jochmans and Martin Weidner. Fixed-effect regressions on network data. *Econometrica*, 87(5):1543–1560, 2019.

[9] D. J. Klein and M. Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1):81–95, 1993.

[10] Tobias Rüttenauer and Ozan Aksoy. When can we use two-way fixed-effects (twfe): A comparison of twfe and novel dynamic difference-in-differences estimators. *arXiv preprint arXiv:2402.09928*, 2024.

[11] Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021.

[12] Prasad Tetali. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4:101–109, 1991.

[13] Jeffrey M Wooldridge. Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*, 2021.