One-stage robust difference-in-differences regression

John Gardner
University of Mississippi

Southern Economic Association Annual Meeting
Nov., 2025

Introduction

- The canonical TWFE DD regression specification recovers difficult-to-interpret measures of average treatment effects when adoption is staggered and treatment effects are heterogeneous
  - de Chaisemartin and Haultfœuille, 2020; Sun Abraham, 2021; Goodman-Bacon, 2022; Borusyak, Jaravel and Spiess, 2024
- Several robust alternative estimators have been developed
  - Borusyak, et al., 2021; Callaway and Sant'Anna, 2021; de Chaisemartin and D'Haultfœuille, 2020; Dube, Jordà and Taylor, 2023; Gardner, 2021; Gardner, Thakral, Tô, and Yap, 2023; Liu, Wang and Xu, 2023; Sun and Abraham, 2021; Wooldridge, 2025; Deb, Norton, Wooldridge, and Zabel, 2024

- I develop an alternative approach that produces robust DD estimates, along with approximately valid standard errors, from a single regression
  - Motivated by the literature on matching and selection on observables
  - Extends to event-study analyses and testing parallel trends
- Another DD estimator?
  - No, point estimates are numerically identical to the two-stage difference in differences, imputation, or fixed-effects counterfactual estimators (Gardner, 2021; Gardner et al., 2023; Borusyak et al. 2021; Liu, Wang and Xu, 2023)
  - Wooldridge, 2025 (cf. Deb et al., 2024) also show how 2SDD/imputation estimates can be obtained by aggregating coefficients from an extended TWFE specification

- This paper develops an *alternative* way of accomplishing this, which doesn't require the aggregation step
  - Prior to ∼2018, people ran the regression

$$Y_{it} = \lambda_{c(i)} + \gamma_t + \beta D_{it} + \varepsilon_{it}$$

  - We now know that β from this specification is hard to interpret as an average causal effect when TEs are heterogeneous and adoption is staggered
  - My specification allows for

$$Y_{it} = \lambda_{c(i)} + \gamma_t + \beta D_{it} + \text{Additional terms} + \varepsilon_{it}$$

    where β is the *overall ATT* (the average across all treated groups and periods), the same as what 2SDD/imputation identifies
- An alternative approach, not a superior one (but may be useful to researchers using software that doesn't automate aggregation of group×time-specific coefficients)

- This paper develops an *alternative* way of accomplishing this, which doesn't require the aggregation step
  - Prior to ∼2018, people ran the regression

$$Y_{it} = \lambda_{c(i)} + \gamma_t + \beta D_{it} + \varepsilon_{it}$$

  - We now know that β from this specification is hard to interpret as an average causal effect when TEs are heterogeneous and adoption is staggered
  - My specification allows for

$$Y_{it} = \lambda_{c(i)} + \gamma_t + \beta D_{it} + \text{Additional terms} + \varepsilon_{it}$$

    where β is the *overall ATT* (the average across all treated groups and periods), the same as what 2SDD/imputation identifies

- An alternative approach, not a superior one (but may be useful to researchers using software that doesn't automate aggregation of group×time-specific coefficients)

- This paper develops an *alternative* way of accomplishing this, which doesn't require the aggregation step
  - Prior to ∼2018, people ran the regression

$$Y_{it} = \lambda_{c(i)} + \gamma_t + \beta D_{it} + \varepsilon_{it}$$

  - We now know that β from this specification is hard to interpret as an average causal effect when TEs are heterogeneous and adoption is staggered
  - My specification allows for

$$Y_{it} = \lambda_{c(i)} + \gamma_t + \beta D_{it} + \text{Additional terms} + \varepsilon_{it}$$

    where β is the *overall ATT* (the average across all treated groups and periods), the same as what 2SDD/imputation identifies

- An alternative approach, not a superior one (but may be useful to researchers using software that doesn't automate aggregation of group×time-specific coefficients)

- This paper develops an *alternative* way of accomplishing this, which doesn't require the aggregation step
  - Prior to ∼2018, people ran the regression

$$Y_{it} = \lambda_{c(i)} + \gamma_t + \beta D_{it} + \varepsilon_{it}$$

  - We now know that β from this specification is hard to interpret as an average causal effect when TEs are heterogeneous and adoption is staggered
  - My specification allows for

$$Y_{it} = \lambda_{c(i)} + \gamma_t + \beta D_{it} + \text{Additional terms} + \varepsilon_{it}$$

    where β is the *overall ATT* (the average across all treated groups and periods), the same as what 2SDD/imputation identifies

- An alternative approach, not a superior one (but may be useful to researchers using software that doesn't automate aggregation of group×time-specific coefficients)

- This paper develops an *alternative* way of accomplishing this, which doesn't require the aggregation step
  - Prior to ∼2018, people ran the regression

$$Y_{it} = \lambda_{c(i)} + \gamma_t + \beta D_{it} + \varepsilon_{it}$$

  - We now know that β from this specification is hard to interpret as an average causal effect when TEs are heterogeneous and adoption is staggered
  - My specification allows for

$$Y_{it} = \lambda_{c(i)} + \gamma_t + \beta D_{it} + \text{Additional terms} + \varepsilon_{it}$$

    where β is the *overall ATT* (the average across all treated groups and periods), the same as what 2SDD/imputation identifies

- An alternative approach, not a superior one (but may be useful to researchers using software that doesn't automate aggregation of group×time-specific coefficients)

Setup

- We observe the random panel $\{Y_{it}, D_{it}, X_{it}\}$ for $t = 1, \ldots, T$ and $i = 1, \ldots, N$
- The treatment is irreversible, unanticipated, and does not affect the covariates
- $C_i \in \mathcal{C} = \{2, \ldots, T, \infty\}$ denotes i's treatment cohort and $C_i^j \in \{0, 1\}$ is an indicator for whether i belongs to treatment cohort $j \in \mathcal{C}$ (TEs not identified for first cohort, "∞" means never treated)
- (Y_{it}^0, Y_{it}^1) denote the counterfactual outcomes that i would receive in t given covariates x
  - $\beta_{it} = Y_{it}^1 - Y_{it}^0$ is the effect of the treatment for i
  - $\beta_{ct}(X_{it}) = E(\beta_{it}|C_i, X_{it})$ is the cohort-time-covariate ATT
- Untreated outcomes satisfy parallel trends:

$$E(Y_{i,t+1}^0 - Y_{it}^0|X, C^j) = E(Y_{i,t+1}^0 - Y_{it}^0|X, C^k) = \Delta\gamma_t + \Delta X_{it}'\delta$$

for all $j, k \in \mathcal{C}$

Review of 2SDD

- The overall average ATT $\beta = E[\beta_{ct}(X_{it})|D_{it} = 1]$ (over cohorts, time and covariates) can be estimated by:
  1. Estimating the model

$$Y_{it} = \lambda_c + \gamma_t + X_{it}'\delta + \varepsilon_{it}$$

     in the sample of untreated observations
  2. Regressing adjusted outcomes $Y_{it} - \hat{\lambda}_c - \hat{\gamma}_t - X_{it}'\hat{\delta}$ on treatment status
- Extends to individual fixed effects, event studies, and tests of parallel trends
- Standard errors can be adjusted for estimation of first-stage parameters (Kyle Butts' `did2s` packages handle this automagically)

Proposition

$\hat{\beta}^{2SDD}$ is consistent and asymptotically normal.

A robust one-stage approach

- Motivation comes from literature on cross-sectional matching and selection on observables
- Temporarily forget the DD context, and suppose that $(Y_0, Y_1) \perp\!\!\!\perp D|X$ and $E(Y_d|X) = X'\delta_d$, $d \in \{0, 1\}$
- Counterfactual mean outcomes functions can be estimated from treatment-status-specific regressions, or from the pooled regression

$$Y = X'\delta_0 + D \cdot X'\delta_1 + q$$

- The ATT is identified as

$$E[E(Y_1|X) - E(Y_0|X)|D = 1] = E(X|D = 1)'\delta_1$$

- The "aggregation" step can be avoided by estimating ATT as $\hat{\beta}$ from the regression (replacing $E(X|D = 1)$ with its sample analog; Wooldridge, 2010)

$$Y = X'\rho_0 + D[X - E(X|D = 1)]'\rho_1 + \beta D + r$$

Extending this approach to DD

- Let W_{it} be a vector of cohort indicators, time indicators, and time-varying controls
- Under parallel trends,

$$E(Y_{it}^0|W_{it}) = \lambda_c + \gamma_t + X_{it}'\delta \equiv W_{it}'\rho_0$$

and

$$E(Y_{it}^1|W_{it}) = \lambda_c + \gamma_t + X_{it}'\delta + \beta_{ct}(X_{it}) \equiv W_{it}'\rho_0 + \beta_{ct}(X_{it}).$$

- If treatment effects vary at the cohort$\times$time level, *treated outcomes are not linear in cohort and time*
- Take the linear projection of $\beta_{ct}(X_{it})$ onto W_{it} *in the treated population*:

$$\beta_{ct}(X_{it}) = \beta_c + \beta_t + X_{it}'\beta_x + \tilde{\beta}_{ct} = W_{it}'\rho_1 + \tilde{\beta}_{ct},$$

where $E(\tilde{\beta}_{ct}|D_{it} = 1) = 0$ by definition

- Now, on average among the treated,

$$E[(Y_{it}^1 - Y_{it}^0 | W_{it}) | D_{it} = 1] = E(\beta_c + \beta_t + X_{it}'\beta_x + \tilde{\beta}_{ct} | D_{it} = 1)$$
$$= E(W_{it} | D_{it} = 1)'\rho_1$$

- Hence, we could estimate the overall ATT (β) by estimating the regression

$$Y_{it} = W_{it}'\rho_0 + D_{it}W_{it}'\rho_1 + s_{it}$$

  and taking the average $\bar{W}^{1'}\hat{\rho}_1$, where
  $\bar{W}^1 = (\sum_{it} D_{it}W_{it}) / \sum_{it} D_{it}$
- Or we could directly estimate it as $\hat{\beta}$ from the specification

$$Y_{it} = W_{it}'\rho_0 + D_{it}(W_{it} - \bar{W}^1)'\rho_2 + \beta D_{it} + r_{it}$$

1SDD and its properties

- To summarize, the overall ATT can be estimated by regressing outcomes on
  1. Cohort and time-period indicators, as well as any time-varying controls (W_{it})
  2. Interactions between treatment status and deviations in cohort indicators, time indicators, and time-varying controls from their means among treated units $[D_{it}(W_{it} - \bar{W}^1)]$
  3. Treatment status (D_{it})
- Courtesy of the Frisch-Waugh-Lovell theorem, we also have:

Proposition

$\hat{\beta}^{1SDD} = \hat{\beta}^{2SDD}$.

- The equivalence result can be used to prove consistency (which can also be obtained directly by appealing to the preceding motivation)
- The regression can be implemented in any econometrics/statistics software package
- Will automatically produce approximately valid standard errors (up to the sampling error in $\bar{W}^1$)
- *A note on individual FEs*: The equivalence result still holds if cohort FEs are replaced with unit FEs, but clustered standard errors will be mechanically biased (because OLS FOCs require each unit's residuals to sum to zero in this case)

Dynamic and placebo effects

- Consider a regression of outcomes on
  1. Cohort indicators, time indicators, and time-varying controls (W_{it}),
  2. Interactions between treatment status and deviations in cohort indicators, time indicators, and covariates from *duration-specific* means [$D_{it}(W_{it} - \bar{W}_{it}^{r})$ for all $r \geq k$, for some $k \leq 0$]
  3. Leads and lags of treatment status (D_{it}^{r}, $r \geq k$)
- Coefficients on lags identify duration-specific ATTs
- Coefficients on leads identify placebo ATTs (i.e., pretending that the treatment happened k periods before it actually did)

Other extensions

- Can replicate canonical TWFE event study by including (interacted versions of) all lags (D_{it}^r, $r \geq 0$) and all but one lead (D_{it}^r, $r < -1$)
  - Leads identify pre-treatment deviations from trend subject to a normalization
- Can be adapted to
  - Group averages
  - Time averages
  - Triple DD

Simulations

- Simulation setting:

$$Y_{it} = \lambda_i + \gamma_t + \beta_{it} D_{it} + \varepsilon_{it},$$

  with $\lambda_i \sim N(C_i, 1)$, $\gamma_t \sim N(0, 1)$, $\varepsilon_{it} \sim N(0, 3)$, $T = 5$, and the C_i are drawn from a discrete uniform distribution
- Four variations:
  1. No covariates: $\beta_{it} \sim N(0, 2)$
  2. No covariates, heterogeneous TEs: $\beta_{it} = t - C_i + 1$
  3. TE depends on covariate linearly: $X_{it} \sim N(1, 1)$ and $\beta_{it} = t - C_i + 1 + X_{it}/4$
  4. Covariates correlated within FEs and TE is multiplicative: $X_{it} \sim N(\lambda_i/25, 1)$ and $\beta_{it} = (t - C_i + 1) \cdot X_{it}/4$
- I estimate the treatment effect using 1SDD and 2SDD, as well as by "manually" aggregating $\bar{W}^{1\prime}\hat{\rho}_1$ (which is numerically equivalent)
- I draw samples of size $N \in \{100, 500, 1000\}$ for 1K simulations, and record the rejection rates

		Simulation 1			Simulation 2		
	N	100	500	1000	100	500	1000
2SDD	D	0.064	0.047	0.045	0.051	0.044	0.050
1SDD	D	0.061	0.046	0.045	0.051	0.044	0.050
Manual	D	0.061	0.046	0.045	0.051	0.044	0.050
		Simulation 3			Simulation 4		
	N	100	500	1000	100	500	1000
2SDD	D	0.057	0.050	0.048	0.058	0.052	0.047
1SDD	D	0.057	0.051	0.048	0.055	0.050	0.048
Manual	D	0.057	0.051	0.048	0.055	0.050	0.048

Only reporting overall ATT for simplicity $\diamond$ The rejection rates for 1SDD and 2SDD are similar, even in smaller samples

Empirical application: Stand your ground laws

- I also revisit Cheng and Hoekstra's (2013) analysis of the effects of "stand your ground" laws on violent crime (log violent crimes per 10,000 people)
- I estimate three specifications: Overall, dynamic, and dynamic with three leads (for placebo testing)
  - For 2SDD, I also implement an alternative test of PT (using all untreated obs. in first stage and including leads in second stage)

		One stage			Two stage		
		(1)	(2)	(3)	(4)	(5)	(6)
Overall ATT	D	0.0706**			0.0706**		
		(0.0279)			(0.0347)		
Dynamic effects	D^1		0.0852***	0.0811**		0.0852***	0.0811**
			(0.0267)	(0.0397)		(0.0291)	(0.0406)
	D^2		0.0727**	0.0657		0.0727*	0.0657
			(0.0304)	(0.0401)		(0.0386)	(0.0454)
	D^3		0.0658*	0.0497		0.0658	0.0497
			(0.0363)	(0.0456)		(0.0450)	(0.0520)
	D^4		0.0373	0.0145		0.0373	0.0145
			(0.0397)	(0.0478)		(0.0502)	(0.0579)
	D^5		0.126***	0.105**		0.126***	0.105**
			(0.0465)	(0.0463)		(0.0447)	(0.0437)
	N	550	550	550	550	550	550

The 1SDD SEs tend to be slightly smaller (although not in every case), although this doesn't change the practical conclusions of the exercise

		One stage			Two stage		
		(1)	(2)	(3)	(4)	(5)	(6)
Placebo effects	D^0			-0.000905		0.00841	-0.000905
				(0.0360)		(0.0181)	(0.0378)
	D^{-1}			-0.0533		-0.0356**	-0.0533
				(0.0335)		(0.0176)	(0.0351)
	D^{-2}			-0.0165		0.00493	-0.0165
				(0.0333)		(0.0146)	(0.0320)
	D^{-3}			-0.00194		0.0176	-0.00194
				(0.0256)		(0.0185)	(0.0292)
	D^{-4}					-0.0174	
						(0.0201)	
	D^{-5}					0.0263	
						(0.0169)	
	D^{-6}					0.0464**	
						(0.0184)	
	D^{-7}					-0.0605	
						(0.0369)	
	D^{-8}					-0.153***	
						(0.0370)	
	D^{-9}					-0.252***	
						(0.0268)	
	N	550	550	550	550	550	550

The tests of parallel trends also agree (as do estimates in models with and without placebos Application 2

Conclusion

- 1SDD is a simple alternative way to obtain 2SDD (aka imputation or FEct) estimates
- Motivated by regression methods for matching under selection on observables
- Produces approximately valid SEs with standard statistical software
- Performs well in simulations and empirical application

Empirical application 2: Employment at will

Back

- I also revisit Autor's (2003) analysis of exceptions to the "employment at will" doctrine on employment in the temporary help services sector
- As before, I estimate the effect of the policy by 2SDD and 1SDD

		One stage			Two stage		
		(1)	(2)	(3)	(4)	(5)	(6)
Overall ATT	D	0.0628			0.0628		
		(0.175)			(0.171)		
Dynamic effects	D^1		0.0843	-0.0557		0.0843	-0.0557
			(0.0680)	(0.184)		(0.0650)	(0.170)
	D^2		0.0908	-0.0190		0.0908	-0.0190
			(0.0879)	(0.208)		(0.0802)	(0.187)
	D^3		0.144	0.0901		0.144	0.0901
			(0.120)	(0.239)		(0.110)	(0.219)
	D^4		0.0159	-0.0746		0.0159	-0.0746
			(0.141)	(0.270)		(0.131)	(0.245)
	D^5		0.0789	0.0278		0.0789	0.0278
			(0.169)	(0.293)		(0.155)	(0.264)
	D^6		0.123	0.128		0.123	0.128
			(0.204)	(0.324)		(0.190)	(0.294)
	D^7		0.0911	0.0633		0.0911	0.0633
			(0.222)	(0.343)		(0.203)	(0.310)
	D^8		0.0877	0.0564		0.0877	0.0564
			(0.252)	(0.362)		(0.231)	(0.326)
	D^9		0.0378	0.0467		0.0378	0.0467
			(0.264)	(0.372)		(0.241)	(0.335)
	D^{10}		-0.0730	-0.143		-0.0730	-0.143
			(0.276)	(0.417)		(0.252)	(0.373)
	N	714	714	476	714	714	476

		One stage			Two stage		
		(1)	(2)	(3)	(4)	(5)	(6)
Placebo effects	D^0			-0.0889		-0.0254	-0.0889
				(0.432)		(0.0311)	(0.394)
	D^{-1}			0.393		-0.0252	0.393
				(0.516)		(0.0291)	(0.455)
	D^{-2}			-0.133		-0.0338	-0.133
				(0.136)		(0.0437)	(0.125)
	D^{-3}			0.0145		0.0616	0.0145
				(0.0801)		(0.0391)	(0.0813)
	D^{-4}					0.00181	
						(0.0553)	
	D^{-5}					-0.0121	
						(0.0482)	
	D^{-6}					0.0718*	
						(0.0435)	
	D^{-7}					0.0983	
						(0.114)	
	D^{-8}					-0.151	
						(0.157)	
	N	714	714	476	714	714	476

In this case, 1SDD produces slightly larger SEs