Doubly-Robust Identification

1

Omer Erhan Erbis

Universität Bonn

2022-12-06

[1] Arkhangelsky, Imbens (2022)

Section 1

General Treatment Effect Models

The Non-parametric True Model

$$Y_{it} = g_t\left(W_{it}, U_i, X_{it}, \varepsilon_{it}\right) \tag{1}$$

Where:

- Y_{it} denoting the outcome of interest
- W_{it} an indicator for the treatment
- U_i the unobserved confounder
- X_{it} the observed attributes/confounders/covariates
- ε_{it} an idiosyncratic error term

The possibility that U_i may be correlated with W_{it} even after controlling for observed confounders prevents us from estimating the average effect of W_{it} on the outcome by comparing covariate-adjusted treated and control outcomes.

The General Two-Way Fixed Effects Model

$$g_t(w, u, x, e) = \alpha(u) + \lambda_t + w\tau + x^\top \beta + e$$

in combination with mean independence $\mathbb{E}\left[\varepsilon_{it} \mid \{(W_{il}, X_{il})\}_{l=1}^{T}\right] = 0$

Properties:

- Linear
- Separable
- Additive

However, this model makes strong implicit assumptions, particularly on τ

Omitted variable bias implies that the bias from an unobserved confounder U_i comes from the combination of its correlation with the outcome Y_{it} and its correlation with W_{it}.

The Modelling Restriction

To remove endogeneity, the strategy of conditioning on cluster-specific statistic is proposed

- $\underline{W}_i$ be the T -vector of assignments with typical element W_{it} from the support $\mathbf{W}$.
- where S_i is a known function of W_i

$$\underline{W}_i \perp U_i \mid S_i \tag{2}$$

- e.g. $S_i = \bar{W}_i = \sum_t W_{it}/T$

The strategy is to leverage the role of design assumptions to control for endogeneity in subpopulations, not to control or restrict the dependence of outcomes on unobserved confounders

The Proposed Estimator(TWFE estimator equivalence by De Chaisemartin and d'Haultfoeuille, 2020):

$$\hat{\tau} = \frac{1}{NT} \sum_{i,t} \hat{\gamma}_{it} Y_{it} \tag{3}$$

- $\hat{\gamma}_{it}$ are weights
- $\hat{\gamma}_{it} = \hat{\gamma}_{it}(\underline{W}_i)$ and $\hat{\gamma}_{it} = \hat{\gamma}_{it}(\underline{W}_i, X_i)$ also allowed
- $\hat{\gamma}_{it} \perp Y_{it}$

Researcher explicitly select weights by solving quadratic optimization problem

Section 2

Binary Treatment Effect Model without Covariates

Setup

- $\underline{w}^t \equiv (w_1, w_2, \ldots, w_t)$ sequence of treatment exposures up to time t
- $\underline{w}^T \equiv \underline{w}$ – Treatment allowed to switch on and off:
  e.g. $\underline{w}^3 = (1, 0, 1)$ for $t = 3$

- $Y_{it}\left(\underline{w}^t\right) \equiv Y_{it}\left(w_1, w_2, \ldots, w_t\right)$ potential outcome for unit i at time t given treatment history up to time t

Assumptions

Assumption B.1 - No Dynamic Treatment Effects

For arbitrary t-component assignment vectors $\underline{w}$ and $\underline{w'}$ such that the period t assignment is the same, $w_t = w_t'$ the potential outcomes in period t are the same:

$$Y_{it}(\underline{w}) = Y_{it}\left(\underline{w'}\right)$$

This reduces the potential outcomes to history-independent setting:

$$Y_{it}(\underline{w}) = Y_{it}\left(w_t\right).$$

And thus, the following definition and reduction is possible:

$$\underline{Y}_i(\underline{w}) \equiv \left(Y_{i1}\left(\underline{w}^1\right), \ldots, Y_{it}\left(\underline{w}^T\right)\right) \equiv \left(Y_{i1}\left(w_1\right), \ldots, Y_{it}\left(w_T\right)\right)$$

Assumptions

The Assumption B.1 makes it possible to define the observed outcomes $Y_{it} \in \mathbf{Y}$ where $\mathbf{Y} \in \mathbb{R}^{\mathbb{N} \times \mathbb{T}}$ in potential outcomes notation as:

$$Y_{it} = W_{it} Y_{it}(1) + (1 - W_{it}) Y_{it}(0)$$

Assumptions

Assumption B.2 - Latent Unconfoundedness

There exist a random variable $U_i \in \mathbb{R}^d$ such that the following conditional independence holds:

$$\underline{W}_i \perp \{\underline{Y}_i(\underline{w})\}_{\underline{w}} \mid U_i$$

Personal Note: *This assumption in particular restricts (more than stressed) the domain for the applications, as certain ones are unable to fulfill this due to the nature of relation between* W_i *and* $\{\underline{Y}_i(\underline{w})\}_{\underline{w}}$. *(e.g.* W_{it} *as a choice that is used for direct optimization of* Y_{it} *)*

Setting up the Stage

- $\pi_k \equiv \mathrm{pr}\left(\underline{W}_i = \mathbf{W}_k\right) = \mathbb{E}\left[\mathbf{1}_{\underline{W}} = \mathbf{W}_k\right]$ represents the (positive) probabilities of occurence of elements of assignment path.
- Let $K \leq 2^T$ be the number of rows in support $\mathbf{W}$
- Let $k(i)$ be the row $\mathbf{W}_k$ of the support matrix $\mathbf{W}$ such that $\mathbf{W}_{k(i)} = \underline{W}_i$

Recall that the proposed estimator has the following form:

$$\hat{\tau}(\gamma) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_{it}Y_{it} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_t(\underline{W}_i)Y_{it}$$

where proposed weights γ_{it} should satisfy the restrictions:
-$\frac{1}{NT}\sum_{i,t} W_{it}\gamma_{it} = 1$ -$\frac{1}{NT}\sum_{i,t}\left(1 - W_{it}\right)\gamma_{it} = -1$

Identification: An Example

Distribution for the Example

k	$\mathbf{W}_k$	π_k
1	$(0, 0, 0)$	0.09
2	$(1, 0, 0)$	0.04
3	$(0, 1, 0)$	0.11
4	$(1, 1, 0)$	0.14
5	$(0, 0, 1)$	0.07
6	$(1, 0, 1)$	0.08
7	$(0, 1, 1)$	0.15
8	$(1, 1, 1)$	0.32

Identification: An Example: Outcome Model

Suppose that in fact the potential outcomes $Y_{it}(w)$ satisfy a two-way-fixed-effect structure and that the treatment effect is constant across time and units:

$$
\begin{aligned}
&Y_{it}(w) = \alpha\left(U_i\right) + \lambda_t + \tau w + \varepsilon_{it}, \\
&\mathbb{E}\left[\varepsilon_{it} \mid \underline{W}_i, U_i\right] = 0
\end{aligned}
$$

Identification: An Example - Outcome Model

The Two-way Fixed Effects Estimator

The fixed-effect estimators uses least squares with two-way fixed effects to 'estimate' τ in population. This procedure leads in large samples to a particular set of weights $\gamma_t^{(\mathrm{fe})}(\underline{W}_i)$, and then to the following fixed-effect estimand:

$$\tau^{\mathrm{fe}} = \mathbb{E}\left[\frac{1}{T}\sum_t Y_{it}\gamma_t^{(\mathrm{fe})}(\underline{W}_i)\right]$$

The expectation here is taken over the $U_i, \{\varepsilon_{it}\}_{t=1}^{T}$ and the assignment path $\underline{W}_i$

Identification: An Example - Outcome Model

For the given distribution, the weights implied by the fixed-effect estimator are presented in the last three columns:

k	$\mathbf{W}_k$	π_k	$\gamma_{k1}^{(fe)}$	$\gamma_{k2}^{(fe)}$	$\gamma_{k3}^{(fe)}$
1	$(0,0,0)$	0.09	0.46	-0.64	0.18
2	$(1,0,0)$	0.04	5.70	-3.26	-2.44
3	$(0,1,0)$	0.11	-2.16	4.60	-2.44
4	$(1,1,0)$	0.14	3.08	1.98	-5.07
5	$(0,0,1)$	0.07	-2.16	-3.26	5.42
6	$(1,0,1)$	0.08	3.08	-5.88	2.80
7	$(0,1,1)$	0.15	-4.78	1.98	2.80
8	$(1,1,1)$	0.32	0.46	-0.64	0.18

Identification: An Example - Assignment Model

Suppose that in fact the potential outcomes $Y_{it}(w)$ satisfy a two-way-fixed-effect structure and that the treatment effect is constant across time and units: suppose that DGP for the assignment mechanism $\underline{W_i}$ has the following form (which is consistent with the probabilities):

$$\forall\left(t, t'\right) : W_{it} \perp W_{it'} \mid U_i, \quad \mathbb{E}\left[W_{it} \mid U_i\right] = \frac{\exp\left(\alpha\left(U_i\right) + \lambda_t\right)}{1 + \exp\left(\alpha\left(U_i\right) + \lambda_t\right)}.$$

One should note that this pair of restrictions implies a conditional independence restriction in the applications (The design choices authors make require S_i to incorporate $\bar{W}_i$):

$$\underline{W}_i \perp \left\{\underline{Y}_i(\underline{w})\right\}_{\underline{w}} \mid \bar{W}_i$$

where $\bar{W}_i \equiv \sum_{t=1}^{T} W_{it}/T$ is the fraction of treated periods for unit i.

Identification: An Example - (Weak) Exclusivity of Models

Suppose assignment model is correctly specified and outcome model is misspecified. The outcome model estimand τ^{fe} may still be equal to the treatment effect $\left(\tau^{\mathrm{fe}} = \tau\right)$ if $\tau_{it} = \tau$ and the following condition on the weights is satisfied for:

$$\mathbb{E}\left[\gamma_t^{\mathrm{fe}}\left(\underline{W}_i\right) \mid \bar{W}_i\right] = 0 \; \forall \; t, \bar{W}_i$$

Identification: An Example - (Weak) Exclusivity of Models

Suppose outcome model is correctly specified and assignment model is misspecified. Generally, the following is the case:

$$\frac{1}{T}\sum_{t=1}^{T}\gamma_t^{(IP)}\left(\underline{W}_i\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{W_{it}}{\mathbb{E}\left[W_{it} \mid \bar{W}_i\right]} - \frac{1-W_{it}}{\mathbb{E}\left[(1-W_{it}) \mid \bar{W}_i\right]}\right) \neq 0$$

thus not balancing individual fixed effects.

Identification: An Example - (Weak) Exclusivity of Models

One combine both strategies to form a double robust estimator that returns true treatment effect τ if either model is correctly specified that returns the following properties simultaneously: row sums to be 0 and $\mathbb{E}\left[\gamma_t^{\mathrm{fe}}\left(\underline{W}_i\right) \mid \bar{W}_i\right] = 0 \; \forall \; t, \bar{W}_i$

(W_1, W_2, W_3)	π_k	$\gamma_1^{(dr)}\left(\underline{W}_k\right)$	$\gamma_2^{(dr)}\left(\underline{W}_k\right)$	$\gamma_3^{(dr)}\left(\underline{W}_k\right)$
$(0, 0, 0)$	0.09	0.00	0.00	0.00
$(1, 0, 0)$	0.04	6.59	−3.95	−2.64
$(0, 1, 0)$	0.11	−1.46	4.10	−2.64
$(1, 1, 0)$	0.14	3.24	1.66	−4.90
$(0, 0, 1)$	0.07	−1.46	−3.95	5.42
$(1, 0, 1)$	0.08	3.24	−6.39	3.15
$(0, 1, 1)$	0.15	−4.81	1.66	3.15
$(1, 1, 1)$	0.32	0.00	0.00	0.00

Identification through Outcome Model

The Assumption I.1

The potential outcomes satisfy:

$$\mathbb{E}\left[Y_{it}(w) \mid U_i\right] = \alpha\left(U_i\right) + \lambda_t + \tau_t\left(U_i\right) w \tag{4}$$

Identification through Outcome Model - Restrictions

To identify a convex combination of $\tau_t(U_i)$ (we need to construct final treatment effect) we consider the weights γ_{kt} that satisfy the following four restrictions:

$$\frac{1}{T}\sum_{k=1}^{K}\sum_{t=1}^{T}\pi_k\gamma_{kt}\mathbf{W}_{kt} = 1 \tag{i}$$

$$\forall k, \sum_{t}\gamma_{kt} = 0, \tag{ii}$$

$$\forall t, \sum_{k=1}^{K}\pi_k\gamma_{kt} = 0 \tag{iii}$$

$$\forall (t,k), \gamma_{kt}\mathbf{W}_{kt} \geq 0 \tag{iv}$$

$i - iv$ are natural given the outcome model described above. The *i* & *iv* ensure that we focus on a convex combination of treatment effects. The *ii* & *iii* guarantee that weights balance out the systematic variation in the baseline outcomes $Y_{it}(0)$.

Identification through Outcome Model

Let $\mathbb{W}_{\text{outc}}$ be the set of weights $\{\gamma_{kt}\}_{k,t}$ which satisfy previously discussed restrictions

For any generic element $\gamma \in \mathbb{W}_{\text{outc}}$ define $\gamma_t\left(\underline{W}_i, \gamma\right)$(total weight of $\underline{W}_i$ instance) to pick out the period t weight for a unit with assignment path $\underline{W}_i$:

$$\gamma_t\left(\underline{W}_i, \gamma\right) \equiv \sum_{k=1}^{K} \gamma_{kt} \mathbf{1}_{\underline{W}_i = \mathbf{w}_k}$$

Using these weights we define the following estimand:

$$\tau(\gamma) = \mathbb{E}\left[\frac{1}{T} \sum_{t=1}^{T} Y_{it} \gamma_t\left(\underline{W}_i, \gamma\right)\right]$$

Identification through Outcome Model

Proposition P.I.1.

Suppose Assumptions B.1, B.2, and I.1 hold, and that $\gamma \in \mathbb{W}_{\text{outc.}}$. Then $\tau(\gamma)$ is a convex combination of $\tau_t\left(U_i\right)$ (over i and t)

Identification through Assignment Model (Design)

Propensity Scores

For some random variable $S_i \equiv S\left(\underline{W}_i\right)$ let $r(\underline{w}, s)$ be the generalized propensity score:

$$r(\underline{w}, s) \equiv \operatorname{pr}\left(\underline{W}_i = \underline{w} \mid S_i = s\right).$$

The Assumption I.2

There exists a known $\underline{W_i}$-measurable sufficient statistic $S_i \in \mathbb{S}$ and a subset $\mathbb{A} \subset \mathbb{S}$ such that:

$$\underline{W}_i \perp U_i \mid S_i = s,$$

and for all $s \in \mathbb{A}$:

$$\max_{\underline{w}} \{r(\underline{w}, s)\} < 1.$$

Identification through Assignment Model (Design)

Proposition P.I.2. (Weak Unconfoundedness)

Suppose Assumptions B.1, B.2, and I.2 hold. Then for any w:

$$\underline{\mathbf{1}}_{\underline{W}_i=\underline{w}} \perp \underline{Y}_i(\underline{w}) \mid S_i.$$

Identification through Assignment Model (Design)

Identification based on design only impose B.1 and B.2 on potential outcomes. With this, one can identify a convex combination of individual treatment effects $\tau_t\left(U_i\right)$ using the weights γ_{kt} that satisfy the following restrictions:

$$\frac{1}{T}\sum_{tk}\pi_k\gamma_{kt}\mathbf{W}_{kt} = 1\;\forall\{k,s,t\} \tag{v}$$

$$\sum_{k:\mathbf{W}_k\in\mathbf{W}^s}\pi_k\gamma_{kt} = 0\;\forall\{k,s,t\} \tag{vi}$$

$$\sum_{k:\mathbf{W}_k\in\mathbf{W}^s}\pi_k\gamma_{kt}\mathbf{W}_{kt} \geq 0\;\forall\{k,s,t\} \tag{vii}$$

Identification through Assignment Model (Design)

Let $\mathbb{W}_{\text{design}}$ be the set of weights $\{\gamma_{tk}\}_{t,k}$ that satisfy previously discussed assignment-model restrictions.

For any generic element $\gamma \in \mathbb{W}_{\text{design}}$ define the $\gamma_t(W_i, \gamma)$ in the same way as before:

$$\gamma_t\left(\underline{W}_i, \gamma\right) \equiv \sum_{k=1}^{K} \gamma_{kt} \mathbf{1}_{\underline{W}_i = \mathbf{w}_k}$$

Using these weights consider the following estimand:

$$\tau(\gamma) = \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} Y_{it}\gamma_t\left(\underline{W}_i, \gamma\right)\right]$$

Proposition P.I.3.

Suppose Assumptions B.1, B.2, and I.2 hold, and that $\gamma \in \mathbb{W}_{\text{design. Then}}$ $\tau(\gamma)$ is a convex combination of treatment effects $\tau_t\left(U_i\right)$ (over i and t)

Section 3

Double Robustness

Double Robust Identification

Let $\mathbb{W}_{\mathrm{dr}} = (\mathbb{W}_{\mathsf{outc}} \cap \mathbb{W}_{\mathsf{design}})$ and observe that, combining the restrictions *i-vii*, we get that any $\gamma \in \mathbb{W}_{\mathsf{dr}}$ satisfies the following restrictions:

$$\begin{aligned}
\text{Target} &: \frac{1}{T}\sum_{tk} \pi_k \gamma_{kt} \mathbf{W}_{kt} = 1, \\
\text{Within } - \text{ unit balance} &: \frac{1}{T}\sum_{t=1}^{T} \gamma_{kt} = 0, \\
\text{Within - period balance} &: \sum_{k:\mathbf{W}_k \in \mathbf{W}^s} \pi_k \gamma_{kt} = 0, \\
\text{Non - negativity} &: \gamma_{kt} \mathbf{W}_{kt} \geq 0.
\end{aligned}$$

Double Robust Identification

Theorem DR

Suppose Assumptions B.1, B.2 hold, and either I.1, or I.2, or both hold. Then for any $\gamma \in \mathbb{W}_{\mathrm{dr}}$, the estimand $\tau(\gamma)$ is a convex combination of treatment effects $\tau_t(U_i)$ (over i and t).

Section 4

Double Robustness with Covariates

Double Robust Algorithm

- Panel Data: $\{Y_{it}, W_{it}, X_i\}_{i,t}$ where X_i time-invariant
- Sufficient statistic S_i provided (assumed to be constructed)
- two p-dimensional functions of (X_i, S_i, t) and (X_i, S_i) for X_i:

$$\psi^{(1)}(X_i, S_i, t) \equiv \left(\psi_1^{(1)}(X_i, S_i, t), \ldots, \psi_p^{(1)}(X_i, S_i, t)\right)$$
$$\psi^{(2)}(X_i, t) \equiv \left(\psi_1^{(2)}(X_i, t), \ldots, \psi_p^{(2)}(X_i, t)\right)$$
$$\psi_t(X_i, S_i) \equiv \left(\psi^{(1)}(X_i, S_i, t), \psi^{(2)}(X_i, t)\right)$$

Double Robust Algorithm

The Proposed Estimator:

$$\hat{\tau} = \frac{1}{NT} \sum_{i,t} \hat{\gamma}_{it} Y_{it} \tag{5}$$

The weights: Try to minimize the variance of weights (minimize variance of $\hat{\tau}$)

$$\begin{aligned}
\{\hat{\gamma}_{it}\}_{it} = \underset{\{\gamma_{it}\}_{it}}{\arg\min} & \frac{1}{(NT)^2} \sum_{it} \gamma_{it}^2 \\
\text{subject to: } & \frac{1}{nT} \sum_{it} \gamma_{it} W_{it} = 1, \quad \frac{1}{T} \sum_{i} \gamma_{it} = 0, \\
& \frac{1}{N} \sum_{t} \gamma_{it} = 0, \quad \frac{1}{NT} \sum_{it} \gamma_{it} \psi_t (X_i, S_i) = 0, \\
& \gamma_{it} W_{it} \geq 0
\end{aligned}$$

Section 5

Inference

The Setup

Observe a random sample $\{(\underline{Y}_i, \underline{W}_i, X_i)\}_{i=1}^{N}$ where each $(\underline{Y}_i, \underline{W}_i, X_i)$ is distributed according to distribution $\mathcal{P}$. For each unit we construct a sufficient statistic $S_i \equiv S(\underline{W}_i, X_i)$, which includes $\bar{W}_i$ that by assumption fixes the distribution of U_i. In the analysis we focus on approximations with large N and fixed T.

Maintaining initial Assumptions B.1 and B.2, and the following assumptions restrict the outcome model to introduce covariates in a tractable way.

Assumption INF.1

Let δ be a residual unspecified function The following restriction holds for $t \in \{1, \ldots, T\}$:

$$\begin{aligned} & Y_{it}(0) = \alpha_t\left(U_i\right) + \psi^{(2)}\left(X_i, t\right)^{\top} \delta^{(2)} + v_{it}, \\ & \mathbb{E}\left[v_{it} \mid U_i, X_i\right] = 0. \end{aligned}$$

Assumption INF.2

If Assumption I.2 is satisfied, then the following holds

$$\alpha_t(U_i) = \beta_t + \psi^{(1)}(X_i, S_i, t)^\top \delta^{(1)} + \nu_{it},$$
$$\mathbb{E}[\nu_{it} \mid S_i, X_i] = 0.$$

Otherwise, the two-way model holds:

$$\alpha_t(U_i) = \beta_t + \alpha(U_i).$$

Define the overall error:

$$\varepsilon_{it} \equiv v_{it} + \nu_{it} \tag{6}$$

and easily observe that under B.2 and INF.2:

$$\mathbb{E}\left[\varepsilon_{it} \mid \underline{W_i}, X_i\right] = 0 \tag{7}$$

Now, it is possible to separate the proposed estimator

$$\hat{\tau} = \frac{1}{NT} \sum_{it} \hat{\gamma}_{it} Y_{it} \tag{8}$$

$$= \frac{1}{NT} \sum_{it} \hat{\gamma}_{it} \tau_{it} W_{it} + \frac{1}{NT} \sum_{it} \hat{\gamma}_{it} \varepsilon_{it} \tag{9}$$

And focus on the conditional weighted average treatment effect (which is a convex combination of τ_{it}):

$$\tau_{\text{cond}} = \frac{1}{NT} \sum_{it} \hat{\gamma}_{it} W_{it} \mathbb{E}\left[\tau_{it} \mid \underline{W}_i, X_i\right] \tag{10}$$

and define $u_{it} \equiv \tau_{it} - \mathbb{E}\left[\tau_{it} \mid \underline{W}_i, X_i\right]$ and observe that:

$$\hat{\tau} - \tau_{\text{cond}} = \frac{1}{NT} \sum_{it} \hat{\gamma}_{it} u_{it} W_{it} + \frac{1}{NT} \sum_{it} \hat{\gamma}_{it} \varepsilon_{it} \tag{11}$$

Inference

I do not want to get into the formal results where some additional assumptions on joint distribution of panel data variables, S_i and errors $\epsilon \equiv \varepsilon_{it} + W_{it} u_{it}$ are made for $\hat{\gamma}_{it}$ to be well-behaved that in turn guarantees the existence of the estimator, correct weight identification and asymptotic normality.

However, a key component for inference should be noted, that the variance estimator is constructed via conventional unit-level bootstrapping.

With all these, the asymptotically correct/conservative confidence intervals are constructed.

Section 6

Extension - Non-Binary Treatments

Non-Binary Treatments

Outcome Model

$$
\begin{aligned}
Y_{it}(w) &= \alpha(U_i) + \lambda_t + \tau_t(U_i) w + \epsilon_{it} \\
\mathbb{E}[\epsilon_{it} \mid U_i] &= 0, \qquad (12)
\end{aligned}
$$

Assignment Model

Consider a baseline distribution $f_0(w)$ that has the same support as W_{it}. Assume that the distribution of W_i conditional on U_i belongs to the following exponential family where $\psi_t(\cdot)$ is a known function:

$$
f(W_i \mid U_i) = \exp\left\{\sum_t \beta^\top(U_i)\psi_t(W_{it}) - \psi(U_i)\right\}\prod_t f_0(W_{it}) \qquad (13)
$$

Non-Binary Treatments

Exponential structure of the assignment model implies the general unconfoundedness condition where $S_i = \sum_t \psi_t(W_{it})$:

$$W_i \perp \{Y_i(w)\}_w \mid S_i,$$

Given S_i, TWFE regression within clusters/subpopulations and aggregation would identify the treatment effects:

$$Y_{it} = \alpha_i + \lambda_t + \tau_{it} W_{it} + \epsilon_{it},$$

Section 7

Applications

Empirical Results

Charles and Stephens (2013a, 2013b)

- **Y** - turnout at presidential elections at the county level, $\mathbb{R}^{n*T}$ matrix
- **W** - log income per capita at the county level, $\mathbb{R}^{n*T}$ matrix
- D_1 - indicator for medium (1) or large (2) importance of coal in the county, $n * 1$ vector
- D_2 - indicator for medium (1) or large (2) importance of gas in the county, $n * 1$ vector
- Z_1 - log national employment for oil and gas, $T * 1$ vector
- Z_2 - log national employment for coal, $T * 1$ vector

The stylized regressor is defined as such:

$$Y_{it} = \alpha_i + \lambda_t + \tau W_{it} + \epsilon_{it}$$

Empirical Results

The authors use IV for identification and the first-stage regression is:

$$\Delta W_{it} = \theta_t + \gamma_1^\top D_{1i} \Delta Z_{1t} + \gamma_2^\top D_{2i} \Delta Z_{2t} + v_{it}$$

Arkhangelsky and Imbens propose the following first-stage regression:

$$W_{it} = \beta_i + \theta_t + \gamma_{1i} Z_{1t} + \gamma_{2i} Z_{2t} + v_{it},$$

and assume that $(\beta_i, \gamma_{1i}, \gamma_{2i})$ are correlated with the potential outcomes, while $\{v_{it}\}_t$ are not, thus effectively saying $U_i = (\beta_i, \gamma_{1i}, \gamma_{2i})$

$$S_i \equiv \left(\sum_{t \leq T} W_{it}, \sum_{t \leq T} Z_{1t} W_{it}, \sum_{t \leq T} Z_{2t} W_{it} \right).$$

Empirical Results

	Estimate	Heterogeneity	s.e.
$\hat{\tau}_{FE}$	0.009	–	0.003
$\hat{\tau}_{DR}$	0.013	0.092	0.007

Simulations - Selection Bias

	$\rho(L)$	RMSE		Bias	
		DR	TW	DR	TW
Design $\eta = 0$	0.045	0.023	0.03	0.018	0.028
Design $\eta = 0.01$	0.186	0.067	0.126	0.06	0.126
Design $\eta = 0.03$	0.29	0.10	0.20	0.10	0.20
Design $\eta = 0.05$	0.36	0.13	0.25	0.12	0.25

Simulations - Different Cluster Sizes

$\eta = 0$	# cluster	$\rho(L)$	RMSE		Bias	
			DR	TW	DR	TW
Design 1	748	0.045	0.022	0.030	0.018	0.028
Design 2	1497	0.045	0.020	0.030	0.008	0.028
Design 3	374	0.045	0.037	0.030	0.035	0.028

Simulations - Different Cluster Sizes

$\eta = 0.01$	$\#$ cluster	$\rho(L)$	RMSE		Bias	
			DR	TW	DR	TW
Design 1	748	0.19	0.07	0.13	0.07	0.13
Design 2	1497	0.19	0.05	0.13	0.05	0.13
Design 3	374	0.19	0.09	0.13	0.09	0.13