Diff-In-Diff Estimator in Dynamic Models: Recent Developments

Zhiyuan Chen

Empirical Methods
Renmin Business School

May 2025

Dynamic Economic Variables

- Many economic issues are dynamic in nature: *the present value depends on its past values.*
  - Examples: R&D, investment, demand, productivity, wage, etc.
- The autoregressive model is a common way to model dynamic variables:

$$
\begin{aligned}
Y_{it} =& \rho Y_{it-1} + \boldsymbol{X}'_{it}\alpha + \lambda_t + \mu_i + \varepsilon_{it} && (1)\\
\Rightarrow \quad Y_{it} =& \rho^t Y_{i0} + \sum_{s=0}^{t} \rho^s \left(\boldsymbol{X}'_{it-s}\alpha + \lambda_{t-s} + \mu_i + \varepsilon_{it-s} \right) && (2)
\end{aligned}
$$

  - Example of logged Cobb-Douglas production function:

$$
\begin{aligned}
y_{it} &= \beta_k k_{it} + \beta_l l_{it} + \gamma_t + \overbrace{\eta_i + v_{it}}^{TFP} + \varepsilon_{it} \quad \text{(production function)} && (3)\\
v_{it} &= \rho v_{it-1} + e_{it} \quad \text{(autoregressive productivity shock)} && (4)\\
\Rightarrow \quad y_{it} &= \rho y_{it-1} + (\beta_k k_{it} - \rho \beta_k k_{it-1}) + (\beta_l l_{it} - \rho \beta_l l_{it-1}) + (\gamma_t - \rho \gamma_{t-1})\\
&\quad + (1-\rho)\eta_i + \underbrace{(\varepsilon_{it} - \rho \varepsilon_{it-1}) + e_{it}}_{\xi_{it}} && (5)
\end{aligned}
$$

Example: Ashenfelter (1978)

- People looking to improve their earnings by participating in a training program have suffered some setbacks in the past
- Ashenfelter (1978) find that training participants typically have a preprogram dip in earnings:
  - Past earnings is a time-varying counfounder that cannot be subsumed in the fixed effects
- The conditional independence assumption would be:

$$E[Y_{it}^0|D_{it},\{Y_{it-\ell}\}_{\ell=1}^L,\boldsymbol{X}_{it}] = E[Y_{it}^0|\{Y_{it-\ell}\}_{\ell=1}^L,\boldsymbol{X}_{it}] \tag{6}$$

- In DID design, the parallel trend assumption is:

$$E[Y_{it}^0 - Y_{it-1}^0|D_{it},\{Y_{it-\ell}\}_{\ell=1}^L,\boldsymbol{X}_{it}] = E[Y_{it}^0 - Y_{it-1}^0|\{Y_{it-\ell}\}_{\ell=1}^L,\boldsymbol{X}_{it}] \tag{7}$$

We can specify a two-way fixed effects model with lagged dependent variables:

$$Y_{it} = \sum_{\ell=1}^{L} \rho_\ell Y_{it-\ell} + \delta D_{it} + \alpha' \boldsymbol{X}_{it} + \lambda_t + \mu_i + \varepsilon_{it} \tag{8}$$

We may difference the model:

$$\Delta Y_{it} = \sum_{\ell=1}^{L} \rho_\ell \Delta Y_{it-\ell} + \delta \Delta D_{it} + \alpha' \Delta \boldsymbol{X}_{it} + \Delta \varepsilon_{it} \tag{9}$$

- OLS estimate is not consistent because both ΔY_{it-1} and $\Delta\varepsilon_{it}$ are a function of ε_{it-1}.

We can specify a two-way fixed effects model with lagged dependent variables:

$$Y_{it} = \sum_{\ell=1}^{L} \rho_\ell Y_{it-\ell} + \delta D_{it} + \alpha' \boldsymbol{X}_{it} + \lambda_t + \mu_i + \varepsilon_{it} \tag{8}$$

We may difference the model:

$$\Delta Y_{it} = \sum_{\ell=1}^{L} \rho_\ell \Delta Y_{it-\ell} + \delta \Delta D_{it} + \alpha' \Delta \boldsymbol{X}_{it} + \Delta \varepsilon_{it} \tag{9}$$

- OLS estimate is not consistent because both ΔY_{it-1} and $\Delta\varepsilon_{it}$ are a function of ε_{it-1}.
- Further lagged earnings are also correlated with ε_{it-1}

We can specify a two-way fixed effects model with lagged dependent variables:

$$Y_{it} = \sum_{\ell=1}^{L} \rho_\ell Y_{it-\ell} + \delta D_{it} + \alpha' \boldsymbol{X}_{it} + \lambda_t + \mu_i + \varepsilon_{it} \tag{8}$$

We may difference the model:

$$\Delta Y_{it} = \sum_{\ell=1}^{L} \rho_\ell \Delta Y_{it-\ell} + \delta \Delta D_{it} + \alpha' \Delta \boldsymbol{X}_{it} + \Delta \varepsilon_{it} \tag{9}$$

- OLS estimate is not consistent because both ΔY_{it-1} and $\Delta\varepsilon_{it}$ are a function of ε_{it-1}.
- Further lagged earnings are also correlated with ε_{it-1}
- No consistent estimator if ε_{it} is serially correlated

We can specify a two-way fixed effects model with lagged dependent variables:

$$Y_{it} = \sum_{\ell=1}^{L} \rho_\ell Y_{it-\ell} + \delta D_{it} + \alpha' \boldsymbol{X}_{it} + \lambda_t + \mu_i + \varepsilon_{it} \tag{8}$$

We may difference the model:

$$\Delta Y_{it} = \sum_{\ell=1}^{L} \rho_\ell \Delta Y_{it-\ell} + \delta \Delta D_{it} + \alpha' \Delta \boldsymbol{X}_{it} + \Delta \varepsilon_{it} \tag{9}$$

- OLS estimate is not consistent because both ΔY_{it-1} and $\Delta\varepsilon_{it}$ are a function of ε_{it-1}.
- Further lagged earnings are also correlated with ε_{it-1}
- No consistent estimator if ε_{it} is serially correlated
- Estimates of positive treatment effects may be too big!

Definition of Treatment Effects: One-time Policy Change

With lagged dependent variables, the treatment effect will be *carried over to the future periods*, posing challenge to correctly defining the treatment effect.

- Imposing $L \equiv 1$, F periods after the treatment (enacted in period t):

$$Y_{it+F} = \rho_1^F Y_{it-1} + \delta \sum_{s=0}^{F} \rho_1^{F-s} D_{it+s} + \sum_{s=0}^{F} \rho_1^{F-s} (\alpha' \boldsymbol{X}_{it} + \lambda_t + \mu_i + \varepsilon_{it}) \tag{10}$$

- With $D_{it+s} = 1$ for $s \geq 1$, we have

$$E(Y_{it+F} | \underbrace{D_{it} = 1}_{(D_{it+s})=1}, Y_{it-1}, \boldsymbol{X}_{it}, \lambda_t, \mu_i) - E(Y_{it+F} | \underbrace{D_{it} = 0}_{(D_{it+s})=0}, Y_{it-1}, \boldsymbol{X}_{it}, \lambda_t, \mu_i)$$

$$= \delta \times \frac{1-\rho_1^F}{1-\rho_1} \tag{11}$$

Dynamic Treatment Effects

The *immediate* treatment effect is δ, *medium-run* treatment effect is $\delta \times \frac{1-\rho_1^F}{1-\rho_1}$, *long-run* treatment effect is $\frac{\delta}{1-\rho_1}$.

Definition of Treatment Effects: Multiple Policy Changes

$\boxed{\text{What is the treatment?}}$

Individuals	$t = 1$	$t = 2$	$t = 3$	$t = 4$	
$i = 1$	0	1	1	1	*(stable policy change)*
$i = 2$	0	1	0	1	*(unstable policy change I)*
$i = 3$	0	1	1	0	*(unstable policy change II)*
$i = 4$	0	0	1	0	*(unstable policy change III)*
$i = 4$	0	0	0	0	*(no policy change)*

1. Force it to be absorbing treatment: treat units who have ever been treated as treated
2. Treatment effect (treatment entry) vs. Reversal treatment effect (treatment exit)
3. Unstable policy changes may have second-order effects: policy uncertainty, policy learning, etc.

- Conditioning on the same $L-period$ history of treatment status, the treatment effect and reversal treatment effect in period F after the policy change is

$$\begin{aligned} ATT(F,L) =& E\Big[\underbrace{Y_{it+F}(D_{it}=1, D_{it-1}=0, \{D_{i,t-\ell}\}_{\ell=2}^{L})}_{Y_{it+F}^{01}(L)} \\ & - \underbrace{Y_{it+F}(D_{it}=0, D_{it-1}=0, \{D_{i,t-\ell}\}_{\ell=2}^{L}}_{Y_{it+F}^{00}(L)} | D_{it}=1, D_{it-1}=0\Big] \\ ART(F,L) =& E\Big[\underbrace{Y_{it+F}(D_{it}=0, D_{it-1}=1, \{D_{i,t-\ell}\}_{\ell=2}^{L})}_{Y_{it+F}^{10}(L)} \\ & - \underbrace{Y_{it+F}(D_{it}=1, D_{it-1}=1, \{D_{i,t-\ell}\}_{\ell=2}^{L}}_{Y_{it+F}^{11}(L)} | D_{it}=1, D_{it-1}=0\Big] \end{aligned}$$

  - The set $\{D_{i,t-\ell}\}_{\ell=2}^{L}$ is the realized history of treatment status: $D_{i,t-\ell}=1$ is allowed
  - $ATT(F,L)$ is the average causal effect in period F assuming that the P.O. ONLY depends on L-period treatment history prior to the treatment.
- Imai, Kim and Wang (2021, AJPS) proposes using matching to

The Choices of F and L

How to choose F and L?

- A large value of L:

The Choices of F and L

How to choose F and L?

- A large value of L:
  - Pros: improves the credibility of carryover effects because it allows a greater number of past treatments to affect the outcome

The Choices of F and L

How to choose F and L?

- A large value of L:
  - Pros: improves the credibility of carryover effects because it allows a greater number of past treatments to affect the outcome
  - Cons: Reduce the number of matches and yield less precise estimates

The Choices of F and L

How to choose F and L?

- A large value of L:
  - Pros: improves the credibility of carryover effects because it allows a greater number of past treatments to affect the outcome
  - Cons: Reduce the number of matches and yield less precise estimates
- A large value of F:

The Choices of F and L

How to choose F and L?

- A large value of L:
  - Pros: improves the credibility of carryover effects because it allows a greater number of past treatments to affect the outcome
  - Cons: Reduce the number of matches and yield less precise estimates
- A large value of F:
  - Pros: Speaks to the long-term causal effects

The Choices of F and L

How to choose F and L?

- A large value of L:
  - Pros: improves the credibility of carryover effects because it allows a greater number of past treatments to affect the outcome
  - Cons: Reduce the number of matches and yield less precise estimates
- A large value of F:
  - Pros: Speaks to the long-term causal effects
  - Cons: Difficult to interpret the causal effects if many units switch the treatment status

PTA and Matching

Parallel Trend Assumption I: ATT

$$
\begin{aligned}
&E\left[Y^{01}_{it+F} - Y^{0}_{it-1} | D_{it} = 1, D_{it-1} = 0, \{D_{it-\ell}, Y_{it-\ell}\}_{\ell=2}^{L}, \{\boldsymbol{X}_{it-\ell}\}_{\ell=0}^{L}\right] \\
&= E\left[Y^{00}_{it+F} - Y^{0}_{it-1} | D_{it} = 0, D_{it-1} = 0, \{D_{it-\ell}, Y_{it-\ell}\}_{\ell=2}^{L}, \{\boldsymbol{X}_{it-\ell}\}_{\ell=0}^{L}\right]
\end{aligned}
$$

Parallel Trend Assumption II: ART

$$
\begin{aligned}
&E\left[Y^{10}_{it+F} - Y^{0}_{it-1} | D_{it} = 0, D_{it-1} = 1, \{D_{it-\ell}, Y_{it-\ell}\}_{\ell=2}^{L}, \{\boldsymbol{X}_{it-\ell}\}_{\ell=0}^{L}\right] \\
&= E\left[Y^{11}_{it+F} - Y^{0}_{it-1} | D_{it} = 1, D_{it-1} = 1, \{D_{it-\ell}, Y_{it-\ell}\}_{\ell=2}^{L}, \{\boldsymbol{X}_{it-\ell}\}_{\ell=0}^{L}\right]
\end{aligned}
$$

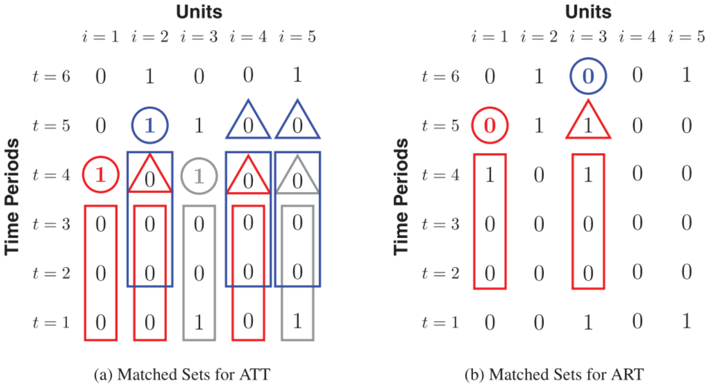
With the same treatment status history, the matched set for observation (i, t) is

$$
\mathscr{M}^{ATT}_{it} = \{(m, t) | m \neq i, D_{mt} = 0, (D_{m,t-\ell})_{\ell=1}^{L} \equiv (D_{i,t-\ell})_{\ell=1}^{L}\} \tag{12}
$$

$$
\mathscr{M}^{ART}_{it} = \{(n, t) | n \neq i, D_{nt} = 1, (D_{n,t-\ell})_{\ell=1}^{L} \equiv (D_{i,t-\ell})_{\ell=1}^{L}\} \tag{13}
$$

An Example of Panel Matching

FIGURE 2 An Example of Matched Sets with Five Units and Six Time Periods



Note: Panels (a) and (b) illustrate how matched sets are chosen for the ATT (as defined in Equation (11)) and the ART (see footnote 5), respectively, when $L = 3$. For each treated observation (coloured circles), we select a set of control observations from other units in the same time period (triangles with the same colour) that have an identical treatment history (rectangles with the same colour).

Matching-DID Estimator

The matching estimators are not so different from the cross-sectional distance-based matching estimators, but focusing on matched treatment-control observations:

- Mahalanobis Distance (Rubin, 2006)
- Propensity Score Matching (Rosenbaum and Rubin, 1983)

The ATT estimator is given by

$$\widehat{ATT(F,L)} = \frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T-F} G_{it}} \sum_{i=1}^{N}\sum_{t=L+1}^{T-F} G_{it} \left[Y_{it+F}^{01} - Y_{it-1}^{0} - \sum_{m\in\mathscr{M}_{it}} w_{it}^{m}\left(Y_{m,t+F} - Y_{m,t-1}\right)\right] \tag{14}$$

- $G_{it} = (D_{it} - D_{it-1})\cdot \boldsymbol{I}\{\|\mathscr{M}_{it}\| > 1\}$: equals 1 if the treatment switches on and has matched control units

Application: Identifying Treatment Effects on Productivity

based on Chen, Liao, and Schurter (2024, working paper)

Motivation: Background

- Firm-level productivity estimates are widely used in empirical analysis

Motivation: Background

- Firm-level productivity estimates are widely used in empirical analysis
- Researchers are interested in **drivers of productivity**
  - Industrial organization (R&D, digitalization), trade & development (export, import), political economies (state ownership), environment economics (pollution), public economics (tax cuts)......

Motivation: Background

- Firm-level productivity estimates are widely used in empirical analysis
- Researchers are interested in **drivers of productivity**
  - Industrial organization (R&D, digitalization), trade & development (export, import), political economies (state ownership), environment economics (pollution), public economics (tax cuts)......
- Firm-level productivity estimation relies on structural assumptions on firm behavior
  - Markovian productivity process
  - Key issue: structural assumptions often ignored

Motivation: Current Method 1

Ex-post Regression has been popular in empirical studies:

- First step: Run productivity estimation ignoring the existence of treatment
- Second step: Treat recovered productivity as truth and run causal inference on D_{it}

Systematic Issue:

- 2nd step implies existence of policy effect;
- 1st step ignore the existence of policy effect
- contradiction

Motivation: Ex-post Method

Ex-post Regression can still be seen in top journals in recent years (2015+).

Issue pointed out very early (De Locker, 2011), but why still used?

Motivation: Ex-post Method

Ex-post Regression can still be seen in top journals in recent years (2015+).

Issue pointed out very early (De Locker, 2011), but why still used?

- Easy to implement: existing code to estimate productivity
- Wide toolbox in treatment analysis: Diff-in-Diff, RD design, PSM
- Good interpretation of treatment effect (ATE/ATT)

Motivation: Current Method 2

There has been attempt to endogenize the treatment:

- Adding D_{it} into the productivity Markov Process
- Solve the issues of ex-post method in some context;
- Doraszelski and Jaumandreu (2013); Aw et al. (2011);

Motivation: Current Method 2

There has been attempt to endogenize the treatment:

- Adding D_{it} into the productivity Markov Process
- Solve the issues of ex-post method in some context;
- Doraszelski and Jaumandreu (2013); Aw et al. (2011);

Some Issues:

- Just adding D_{it} into productivity Markov Process is not general enough;
- Loss the treatment effect toolbox
- Not very appealing policy interpretation/evaluation

This Paper: New Econometric Framework

Structural Model: Firm Dynamic Behavior

$+$

Reduced Form: Potential Outcome on Productivity

Why this new framework?

- Mid-approach in this paper:
  1. No complicated firm behavioral assumptions (✓Potential Outcome)
  2. Interpret: causal inference language (✓ATE/ATT Potential Outcome)
  3. Counterfactual analysis on productivity (✓structural model)
- Minor cost:
  1. Slightly complicated identification
  2. Estimation requires slightly more computation

This Paper: Detail

- **A dynamic firm model: a** binary treatment that affects firm-level productivity

This Paper: Detail

- **A dynamic firm model:** **a** binary treatment that affects firm-level productivity

- **Identification of Productivity:** Characterize the moment conditions

This Paper: Detail

- **A dynamic firm model:** **a** binary treatment that affects firm-level productivity
- **Identification of Productivity:** Characterize the moment conditions
- **Identification of Treatment Effects on Productivity:** Dynamic ATT:
  - *Conditional parallel trend (New Assumption)*

Position in the Literature

- **Production function estimation approaches:** OP (1996); LP (2003); ACF(2015); GNR(2020)
- **Empirical policy effect on productivity:** Pavcnik (2003); Amiti and Konings (2007); De Loecker (2007); Doraszelski and Jaumandreu (2013); Braguinsky et al. (2015); Brandt et al. (2017); Chen et al. (2021)
- **Dynamic Treatment:** Heckman and Navarro (2007); Abbring and Heckman (2007); Vikstrom et al. (2018); Abraham and Sun (2020).

Insights Beyond Productivity

Major results are on productivity analysis, but broadly:

- Think of

  'Structural' (Prod function and evolution params β)

  $$+$$

  'Potential Outcome' (Treat Effect/ ATT θ)

- Influence is two-way:
  1. Adding PO into Structural $\Rightarrow$ Correct moments to identify β (avoid mis-specification)
  2. Adding Structural into PO $\Rightarrow$ Adapt exciting tools (DiD/RD/PSM adaption) to identify θ.

Production and Treatment

- Two potential productivity outcomes $\left(\omega_{it}^0, \omega_{it}^1\right)$

$$\omega_{it} = \underbrace{\omega_{it}^1 D_{it}}_{treated} + \underbrace{\omega_{it}^0 (1 - D_{it})}_{non-treated}$$

  - Generally, $\omega_{it}^1 \neq \omega_{it}^0$
- Firm i in period t has the production function

$$Q_{it} = e^{\omega_{it}} F(K_{it}, L_{it}, M_{it}, D_{it}; \beta)$$

  - $D_{it} \in \{0,1\}$, $\beta \in \mathbb{R}^{\infty}$ (non-parametric)
  - ω_{it} is the **realized productivity**

Productivity Process

Define $G_{it} = D_{it} - D_{it-1}$, the productivity process is Markovian:

$$\begin{pmatrix} \omega_{it}^0 \\ \omega_{it}^1 \end{pmatrix} = \mathbb{I}(G_{it} = 0)\underbrace{\bar{h}\begin{pmatrix} \omega_{it-1}^0 \\ \omega_{it-1}^1 \end{pmatrix}}_{\textit{No Switch}} + \mathbb{I}(G_{it} = 1)\underbrace{h^{+}\begin{pmatrix} \omega_{it-1}^0 \\ \omega_{it-1}^1 \end{pmatrix}}_{\textit{Positive Switch}}$$

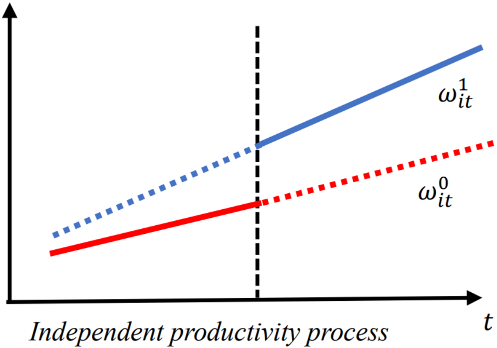
$$+\mathbb{I}(G_{it} = -1)\underbrace{h^{-}\begin{pmatrix} \omega_{it-1}^0 \\ \omega_{it-1}^1 \end{pmatrix}}_{\textit{Negative Switch}} + \underbrace{\begin{pmatrix} \varepsilon_{it}^0 \\ \varepsilon_{it}^1 \end{pmatrix}}_{\textit{Shocks}},$$

Economic Meaning of the Productivity Process

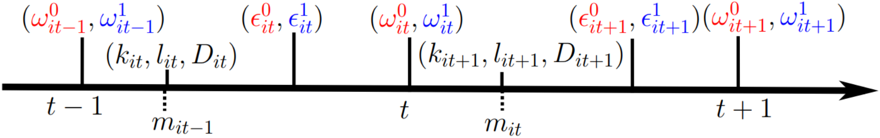
Key Example (Independent Productivity Processes)

Imposing $\bar{\boldsymbol{h}}$, $\boldsymbol{h}^{+}$, $\boldsymbol{h}^{-}$ to be diagonal in $\left(\omega_{it}^{0}, \omega_{it}^{1}\right)$.

- Green technology vs. Polluting technology



Timing



- m_{it} is static choice, while (k_{it}, l_{it}, D_{it}) are dynamic choices

Implication of PO on Structural Model

Change of Moments with PO

- When there is no treatment $\omega_{it} = \omega_{it}^0$:
$$\mathbb{E}[\omega_{it}(\beta) - h(\omega_{it-1}(\beta))|\{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}] = 0 \quad \forall t = 1, ..., T \tag{15}$$

- $\omega_{it}(\beta) = q_{it} - f(k_{it}, l_{it}, m_{it}, D_{it} = 0; \beta)$

Implication of PO on Structural Model

Change of Moments with PO

- When there is no treatment $\omega_{it} = \omega_{it}^0$:

$$\mathbb{E}[\omega_{it}(\beta) - h(\omega_{it-1}(\beta))|\{k_{it}, l_{it}, k_{it-1}, l_{it-1}, m_{it-1}\}] = 0 \quad \forall t = 1, ..., T \tag{15}$$

- $\omega_{it}(\beta) = q_{it} - f(k_{it}, l_{it}, m_{it}, D_{it} = 0; \beta)$
- When there is treatment change and with potential productivity:

$$\mathbb{E}[\omega_{it}^0(\beta) - \bar{h}_0(\omega_{it-1}^0(\beta))|\mathscr{Z}_{it}, D_{it} = D_{it-1} = 0] = 0 \tag{16}$$

$$\mathbb{E}[\omega_{it}^1(\beta) - \bar{h}_1(\omega_{it-1}^1(\beta))|\mathscr{Z}_{it}, D_{it} = D_{it-1} = 1] = 0 \tag{17}$$

- $\omega_{it}^1(\beta) = q_{it} - f(k_{it,}, l_{it}, m_{it}, D_{it} = 1; \beta)$

Compare to the Endogenous Productivity Process

- **Endogenous Productivity Process** uses the following transition process

$$\omega_{it} = h(\omega_{it-1}, D_{it}) + \varepsilon_{it}$$

- Their method at the transition period is equivalent to:

$$\text{Impose:} \quad E[\omega_{it}^1(\beta) - \bar{h}_1(\omega_{it-1}^0(\beta)) | \mathscr{Z}_{it}, D_{it} = 1, D_{it-1} = 0] = 0$$
$$\text{Real:} \quad E[\omega_{it}^1(\beta) - \bar{h}^+(\omega_{it-1}^0(\beta), \omega_{it-1}^1(\beta)) | \mathscr{Z}_{it}, D_{it} = 1, D_{it-1} = 0] = 0$$

The Econometrician's Information

Corollary (Potential Productivity Identification)

We can recover the unobserved potential productivity ω_{it}^{d} for firms such that $D_{it} = d$.

Implication of Structural Model on Treatment Obj

Standard Diff-in-Diff with fixed treatment assignment at time e

- Focus: $ATT_0 = \mathbb{E}\left[\omega_{it}^1 - \omega_{it}^0 | t = e, D_{ie} = 1\right]$

Implication of Structural Model on Treatment Obj

Standard Diff-in-Diff with fixed treatment assignment at time e

- Focus: $ATT_0 = \mathbb{E}\left[\omega_{it}^1 - \omega_{it}^0 | t = e, D_{ie} = 1\right]$
- Simple Parallel Trend

$$\mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | D_{ie} = 1] = \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | D_{ie} = 0]$$

Implication of Structural Model on Treatment Obj

Standard Diff-in-Diff with fixed treatment assignment at time e

- Focus: $ATT_0 = \mathbb{E}\left[\omega_{it}^1 - \omega_{it}^0 | t = e, D_{ie} = 1\right]$
- Simple Parallel Trend

$$\mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | D_{ie} = 1] = \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | D_{ie} = 0]$$

- Simple Parallel Trend fails:

$$\begin{aligned} +\text{ switch: } & \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | D_{ie} = 1] = \mathbb{E}[h_0^+(\omega_{it-1}^1, \omega_{it-1}^0) - \omega_{it-1}^0 | D_{ie} = 1], \\ \text{non switch: } & \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | D_{ie} = 0] = \mathbb{E}[\bar{h}_0(\omega_{it-1}^0) - \omega_{it-1}^0 | D_{ie} = 0], \end{aligned}$$

Implication of Structural Model on Treatment Obj

Standard Diff-in-Diff with fixed treatment assignment at time e

- Focus: $ATT_0 = \mathbb{E}\left[\omega_{it}^1 - \omega_{it}^0 | t = e, D_{ie} = 1\right]$
- Simple Parallel Trend

$$\mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | D_{ie} = 1] = \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | D_{ie} = 0]$$

- Simple Parallel Trend fails:

$$\begin{aligned} + \text{ switch: } & \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | D_{ie} = 1] = \mathbb{E}[h_0^+(\omega_{it-1}^1, \omega_{it-1}^0) - \omega_{it-1}^0 | D_{ie} = 1], \\ \text{non switch: } & \mathbb{E}[\omega_{it}^0 - \omega_{it-1}^0 | D_{ie} = 0] = \mathbb{E}[\bar{h}_0(\omega_{it-1}^0) - \omega_{it-1}^0 | D_{ie} = 0], \end{aligned}$$

- Conditional Parallel Trend: $\bar{h}_0 = h_0^+$;

$$ATT_0 = \mathbb{E}[\omega_{it} - \bar{h}_0(\omega_{it-1}) | t = e, D_{ie} = 1]$$

Traditional DID Setting

Dynamic ATT: Sufficient Condition 2

Sufficient Condition 2 (Strongly Independent Productivity Shocks)

The ℓ-period-ahead shock independence $(\varepsilon_{ie}^{0}, ..., \varepsilon_{ie+l}^{0}) \perp D_{ie} | \omega_{ie-1}^{0}$.

The ℓ-period-ahead ATT is identified as

$$
\begin{aligned}
ATT_{\ell} &= \mathbb{E}_{\omega_{ie-1}} [\Delta(\ell, \omega_{ie-1})] \\
\Delta(\ell, \omega) &\equiv \mathbb{E}[\omega_{ie+\ell} | D_{it} = 1, \omega_{ie-1} = \omega] \\
&\quad - \mathbb{E}\left[\omega_{ig+\ell} | D_{it} = 0, \omega_{ie-1} = \omega\right]
\end{aligned}
$$

- This is a matching method: match using ω_{ie-1}

Additional Results

- Identification of Production functions with additional data limitations
  - Moments motivated by PO structure
- Identification of the dynamic treatment effect
  - Requires assumptions on structural model residuals
  - Typically more complicated than the ATT_0.