# Modeling Treatment Effect Dynamics in Panel Instrumental Variable Regression Models with Varying-intensity Treatments

Dakyung Seong[1] Shu Shen[2] Jaerim Choi[3]

[1]School of Economics, University of Sydney

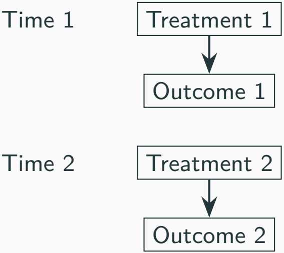
[2]Department of Economics, University of California, Davis

[3]Department of Economics, University of Hawaii at Manoa

# Treatment Effect Dynamics in Panel Data

- Panel settings with "repeated endogenous treatments" and "time-varying external instruments" are popular in applied literature
  - Card (2009): the impact of immigration on the wage gap b/w natives with different skill
  - Autor, Dorn and Hansen (2013; ADH): the impact of Chinese import exposure on the labor market structure
  - Acemoglu and Restrepo (2022), Bourchardi et al. (2019, 2020)
- The current paper concerns the dynamic treatment effect, represented by carryover effect & contemporaneous effect, in a short panel setting.

# Treatment Effect Dynamics in Panel Data




- The previous literature studies the treatment effect *separately* for each time $t$.
- Ignore a dynamic effect of the treatment at time 1 to the treatment effect at time 2.
- Initial treatment period is not informative.
- The effects in time 1 and 2 are often assumed to be identical.

- The following has been often employed:

$$\Delta Y_{it} = \beta_0 + \beta_1 \Delta X_{it} + \{\text{explanatory variables}\} + e_{it}$$

- E.g., Autor, Dorn and Hansen (2013, AER)
  - $\Delta Y_{it}$: the decadal change in the manufacturing employment share of the working-age population.
  - $\Delta X_{it}$: the decadal change in Chinese import exposure.
  - Time: 1990-2000, 2000-2007

# Treatment Effect Dynamics in Panel Data

- Repeated treatments but the same effect?
  - Estimating the model separately for each $t$ ignores a possible dynamic structure of the treatment effect.
  - Need to distinguish the contemporaneous effect from the total effect.

    Instrument $X_{it}$ to $Z_{it}$;

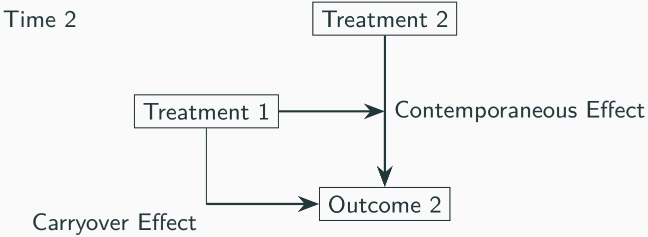
$$\Delta Y_{it} = \beta_0 + \beta_1 \Delta X_{it} + \{\text{explanatory variables}\} + e_{it}$$

- A possible effect of an initial treatment on a later period treatment effect
  - The initial treatment may change the labor mkt. situation at a later period.
  - E.g., in immigration literature, the dependent variable is native labor market outcome and the treatment variable is the local share of immigrants (see, e.g., Card, 2009).
  - Immigrants tend to settle down to regions known to favorable to them

    $\rightarrow$ possible impact of the contemporaneous treatment effect at $t = 2$

    $\rightarrow$ possible carryover effect at $t = 2$.
- Theoretically, ignoring the dynamic treatment effect will end up with the omitted variable bias.

$$Y_{i1} = \beta_1^0 X_{i1} + \epsilon_{i1}$$
$$Y_{i2} = \beta_2^0 X_{i2} + \beta_1^1 X_{i1} + \varepsilon_{i2} = \beta_2^0 X_{i2} + u_{i2}$$




- Suppose that the initial treatment $T_{i1}$ is either 1 or 0.
- $\beta(T_{i1})$: contemporaneous effect at $t = 2$
  - $|\beta(1)| > |\beta(0)|$: accelerated treatment effects
  - $\beta(1) = \beta(0)$: no effect of initial treatment
  - $|\beta(1)| < |\beta(0)|$: decelerated treatment effects
- In ADH's example, if $\beta(1) < \beta(0)$;

  a large increase in Chinese import exposure at $t = 1$

  $\rightarrow \downarrow$ in the share of manufacturing industry in that area

  $\rightarrow$ an industrial structure change

  $\rightarrow$ makes the region be robust to the change in Chinese import exposure at a later period

# Treatment Effect Dynamics in Panel Data

- Our benchmark model is given by

$$Y_{it} = \beta^1_{t-1}(X_{it-1}) + \beta^0_t(X_{it-1})X_{it} + \widetilde{H}'_{it}\gamma_t + \varepsilon_{it} \tag{1}$$

  - $X_{it}$: (possibly endogenous) treatment at time $t$
  - $\widetilde{H}_{it}$: exogenous variables with constant coefficients
  - $\beta^1_{t-1}(.)$: carryover effect
  - $\beta^0_t(.)$: contemporaneous treatment effect

- E.g., binary treatment with $T = 2$:

$$\begin{aligned} Y_{i1} &= \beta^0_1 X_{i1} + \varepsilon_{i1}, \\ Y_{i2} &= \beta^1_1 X_{i1} + \beta^{0,0}_2 X_{i2} 1\{X_{i1} = 0\} + \beta^{0,1}_2 X_{i2} 1\{X_{i1} = 1\} + \varepsilon_{i2} \end{aligned}$$

- The stacked two period model is a special case of our benchmark model.
- Econometric Literature on the dynamic treatment effects
  - Heckman and Navarro (2007); Heckman et al. (2016)
  - Hsu and Shen (2023) Cellini et al. (2010)
  - Sun and Abraham (2021), Callaway et al. (2021)

Back to the benchmark model:

$$Y_{it} = \beta^1_{t-1}(X_{it-1}) + \beta^0_t(X_{it-1})X_{it} + \widetilde{H}'_{it}\gamma_t + \varepsilon_{it} \tag{2}$$

*From a potential outcome framework:*

- $Y_{i1}(x_1)$, $Y_{i2}(x_1, x_2)$: potential outcomes given the treatments $x_1$ and $x_2$.
- $Y_{i1}(X_{i1})$, $Y_{i2}(X_{i1}, X_{i2})$: observed outcomes
- $\widetilde{H}_{it}$: exogenous variable with coefficient of time-varying but constant across $i$.

Suppose that

$$Y_{i1}(x_1) = Y_{i1}(0) + \beta^0_1(x_1)$$
$$Y_{i2}(x_1, x_2) = Y_{i2}(x_1, 0) + (Y_{i2}(x_1, x_2) - Y_{i2}(x_1, 0))$$

where $Y_{i1}(0) = \widetilde{H}'_{i1}\gamma^H_1 + \varepsilon_{i1}$,

$$Y_{i2}(x_1, 0) = \rho Y_{i1}(x_1) + \eta(x_1) + \widetilde{H}'_{i2}\gamma^H_2 + \varepsilon_{i2}$$
$$Y_{i2}(x_1, x_2) - Y_{i2}(x_1, 0) = \beta^0_2(x_1)x_2$$

Hence, the model converts to the benchmark model if

$$\beta^1_1(.) = \rho\beta^0_1(.) + \eta(.) \text{ and } \gamma_2 = \gamma^H_1\rho + \gamma^H_2.$$

# Extensions

A. $Y_{it} = \beta_t^0(X_{it-1})X_{it} + H_{it}'\gamma_t + \varepsilon_{it}$

- The constant term is soaked into $\gamma_t$.

B. $Y_{it} = \beta_{t-1}^1(Y_{it-1}) + \beta_t^0(Y_{it-1})X_{it} + \widetilde{H}_{it}'\gamma_t + \varepsilon_{it}$

- More likely to be relevant in empirical setup; in ADH's example, a changed industrial structure affects the treatment effect at a later period.
- If there's no path dependency, the model reduces to the dynamic model in Andrews and Lu (2001).

C. $Y_{it} = \beta_{t-1}^1(X_{it-2})X_{it-1} + \beta_t^0(X_{it-1})X_{it} + \widetilde{H}_{it}'\gamma_t + \varepsilon_{it}$

- At $t = 3$, $Y_{i3} = \beta_2^1(X_{i1})X_{i2} + \beta_3^0(X_{i2})X_{i3} + \widetilde{H}_{i3}'\gamma_3 + \varepsilon_{i3}$

# Parametric Estimation

- Standard IV approach may be (asymptotically) fine *if the parametric form of the dynamic effect is known.*
- Example 1, $Y_{i2} = \beta_1^1 X_{i1} + \beta_2^{0,1} X_{i1} X_{i2} + \beta_2^{0,2} X_{i2} + \varepsilon_{i2}$
- Example 1, $Y_{i2} = \beta_1^{1,1} 1\{X_{i1} \geq 0\} + \beta_1^{1,2} 1\{X_{i1} < 0\} + \beta_2^0 X_{i2} + \varepsilon_{i2}$
- With a proper instrument, e.g, $(Z_{i1}, Z_{i2}, Z_{i1} Z_{i2})$, we can show the consistency of the IV estimator
- This parametric specification, of course, is not always correct and its correctness should be checked by researchers.
- Inconsistency when a parametric function of $\beta_2^0(.)$ is misspecified.

  $\Rightarrow$ Semi or Non-parametric approach?

## Highlights

- Challenges in estimating the benchmark model:
  - Possible *endogeneity* of $X_{it}$ and $X_{it-1}$
  - Unknown structure of the the functional coefficients.
- Alongside our paper titled *"Path-dependent Effects in the Repeated Treatment Setting: Theory and the China Syndrome Application"*, we study *identification conditions* for the dynamic treatment effects.
  - Illustrate a practical pitfall of parametric approaches
  - Answer to the question on testing the existence of path-dependent carryover or contemporaneous effect.
- This paper discusses identification and semi-parametric estimation procedures based on the control function approach.
- Establishes the asymptotic properties of the semi-parametric estimator.
- Application to Acemouglu et al. (2016)

# Identification and Estimation

# Identification

Back to the benchmark model:

$$Y_{i2} = \beta_1^1(X_{i1}) + \beta_2^0(X_{i1})X_{i2} + \widetilde{H}'_{i2}\gamma_2 + \varepsilon_{it} \tag{3}$$

- Approach 1: projecting $X_{i2}$ on $\{Z_{i2}, \widetilde{H}_{i2}\}$ and replace it by the projected values conditional on $X_{i1}$
    - Requires the sequential exogeneity condition, i.e., $\mathbb{E}[\varepsilon_{i2}|Z_{i2}, \widetilde{H}_{i2}, X_{i1}] = 0 \rightarrow$ Not easy to verify
    - $\epsilon_{i2}$: technology shock & $X_{i1}$: changes in import exposure
- Approach 2: the conditional mean independence assumption such that

$$\mathbb{E}[\varepsilon_{i2}|Z_{i2}, \widetilde{H}_{i2}, X_{i1}] = \mathbb{E}[\varepsilon_{i2}|X_{i1}] \tag{4}$$

For $G_{i2} = (Z_{i2}, \widetilde{H}_{i2})$ and $g_2(x) = \beta_1^1(x) + \mathbb{E}[\varepsilon_{i2}|X_{i1} = x]$,

$$\begin{aligned}
&\mathbb{E}\left[G_{i2}\left(Y_{i2} - \left(g_2(x) + \beta_2^0(x)X_{i2} + \tilde{H}'_{i2}\gamma_2\right)\right)|X_{i1} = x\right] \\
=&\mathbb{E}\left[G_{i2}\left(\varepsilon_{i2} - \mathbb{E}[\varepsilon_{i2}|X_{i1} = x]\right)|X_{i1} = x\right] \\
=&\mathbb{E}\left[G_{i2}\left(\mathbb{E}\left[\varepsilon_{i2}|Z_{i2}, \tilde{H}_{i2}, X_{i1} = x\right] - \mathbb{E}\left[\varepsilon_{i2}|X_{i1} = x\right]\right)|X_{i1} = x\right] \\
=&0.
\end{aligned} \tag{5}$$

$\Rightarrow$ Cannot identify the carryover effect

- Approach 3: Control function restriction

$$\mathbb{E}[\varepsilon_{i2}|Z_{i2}, \widetilde{H}_{i2}, e_{i2}, X_{i1}] = \mathbb{E}[\varepsilon_{i2}|e_{i2}, X_{i1}] \tag{6}$$

where $e_{i2} = X_{i2} - \mathbb{L}_{G_{i2}}(X_{i2}|X_{i1})$
$\mathbb{L}_{G_{i2}}(A) = \mathbb{E}[AG'_{i2}|X_{i1}](\mathbb{E}[G_{i2}G'_{i2}|X_{i1}])^{-1}G_{it}$ for a random variable $A$.

- If we let $(\alpha_2^Z(x)\ \alpha_2^H(x))' = (\mathbb{E}[G_{i2}G'_{i2}|X_{i1} = x])^{-1}\mathbb{E}[G_{i2}X_{i2}|X_{i1} = x]$, the first stage reduced form equation can be written as follows:

$$X_{i2} = Z_{i2}\alpha_t^Z(X_{i1}) + H'_{i2}\alpha_2^H(X_{i1}) + e_{i2}, \tag{7}$$

with satisfying $\mathbb{E}[G_{it}e_{it}|X_{i(t-1)}] = 0$

# Identification

- The function $\mathbb{E}[\varepsilon_{i2}|e_{i2}, X_{i1}] = \beta_2^e(X_{i1})e_{i2}$

$$
\begin{aligned}
&\mathbb{E}[Y_{i2}|X_{i2}, Z_{i2}, e_{i2}, \widetilde{H}_{i2}, X_{i1}] \\
&= \mathbb{E}[Y_{i2}|\mathbb{L}_{G_{i2}}(X_{i2}|X_{i1}) + e_{i2}, Z_{i2}, e_{i2}, \widetilde{H}_{i2}, X_{i1}] \\
&= \beta_1^1(X_{i1}) + \beta_2^0(X_{i1})X_{i2} + \widetilde{H}'_{i2}\gamma_2 + \mathbb{E}[\varepsilon_{i2}|e_{i2}, X_{i1}] \\
&= \beta_1^1(X_{i1}) + \beta_2^0(X_{i1})X_{i2} + \widetilde{H}'_{i2}\gamma_2 + \beta_2^e(X_{i1})e_{i2},
\end{aligned}
$$

- Can identify both the carryover effect and the contemporaneous effect $\Rightarrow$ Back out the total effect
- In the first stage, $X_{i1}$ needs not to be exogenous.
  $\rightarrow$ wider applicability in practice.
- Differentiate this paper from the functional coefficient models, e.g., Fan and Huang (2005), Cai et al (2019), etc.

The benchmark model can be rewritten as

$$Y_{i2} = \beta_1^1(X_{i1}) + \beta_2^0(X_{i1})X_{i2} + \widetilde{H}_{i2}'\gamma_2 + \beta_2^e(X_{i1})e_{i2} + v_{i2}, \tag{8}$$

where $\mathbb{E}[v_{i2}|X_{i1}, X_{i2}, Z_{i2}, \widetilde{H}_{i2}, e_{i2}] = 0$ and $e_{i2} = e_{i2}(X_{i1})$ satisfying $\mathbb{E}[e_{i2}|X_{i1}] = 0$.

The model can be extended to general $t \geq 2$.

In the special case where $\beta_2^e(X_{i1}) = \beta_2^e$, the partial effect of $X_{i1}$ on $Y_{i2}$:

$$\frac{\partial Y_{i2}}{\partial X_{i1}} = \frac{\partial \beta_1^1(X_{i1})}{\partial X_{i1}} + \frac{\partial \beta_2^0(X_{i1})}{\partial X_{i1}} X_{i2} + \beta_2^0(X_{i1})\frac{\partial X_{i2}}{\partial X_{i1}} + \beta_2^e \frac{\partial e_{i2}(X_{i1})}{\partial X_{i1}}.$$

If $X_{i2} = \rho(X_{i1}) + Z_{i2}\pi + e_{i2}$,

$$\mathbb{E}\left[\frac{\partial Y_{i2}}{\partial X_{i1}}|X_{i1} = x\right] = \left.\frac{\partial \beta_1^1(X_{i1})}{\partial X_{i1}}\right|_{X_{i1}=x} + \left.\frac{\partial \beta_2^0(X_{i1})}{\partial X_{i1}}\right|_{X_{i1}=x} \mathbb{E}[X_{i2}|X_{i1} = x] + \beta_2^0(x)\left.\frac{\partial \rho(X_{i1})}{\partial X_{i1}}\right|_{X_{i1}=x}$$

If $X_{i2} = \rho + Z_{i2}\pi + e_{i2}$,

$$\mathbb{E}\left[\frac{\partial Y_{i2}}{\partial X_{i1}}|X_{i1} = x\right] = \left.\frac{\partial \beta_1^1(X_{i1})}{\partial X_{i1}}\right|_{X_{i1}=x} + \left.\frac{\partial \beta_2^0(X_{i1})}{\partial X_{i1}}\right|_{X_{i1}=x} \mathbb{E}[X_{i2}|X_{i1} = x].$$

# Estimation Procedure

- Three-step procedure: (i) estimate the control function (ii) estimate the constant coefficients and (iii) estimate the functional coefficients
- Start with the local constant estimator.
- $\beta(x) = (\beta_1^1(x), \beta_2^0(x), \beta_2^e(x)) \approx \beta(x_0)$ for $x$ in a neighborhood of $x_0$.
- $Y_{i2}^* = Y_{i2} - \widetilde{H}_{i2}'\gamma_2$
- $\kappa_h(.) = \kappa(./h)/h$ is a kernel function $\kappa(.)$ and bandwidth $h$

Solve

$$\widetilde{\beta}(x) = \arg\min_{\beta(x)} \sum_i \left(Y_{i2}^* - ((1\ X_{i2}\ e_{i2})\beta(x))\right)^2 \kappa_h(X_{i1} - x), \tag{9}$$

## Estimation Procedure

- If the control function is available,

$$\widetilde{\beta}(x) = \left(\mathbf{D}'(x)\mathbf{K}(x)\mathbf{D}(x)\right)^{-1}\mathbf{D}'(x)\mathbf{K}(x)(\mathbf{Y}_2 - \widetilde{\mathbf{H}}_2\widetilde{\gamma}_2), \qquad (10)$$

  where $\mathbf{D}(x)$ is a matrix whose $i$ row is $(1, X_{i2}, e_{i2})$ and $\mathbf{K}(x)$ is a diagonal matrix consisting of $\{\kappa_h(X_{i1} - x)\}_{i=1}^{N}$.

- Given the estimates, the estimator of $\gamma_2$ is

$$\widetilde{\gamma} = (\widetilde{\mathbf{H}}_2'(\mathbf{I}-\mathbf{S})'(\mathbf{I}-\mathbf{S})\widetilde{\mathbf{H}}_2)^{-1}\widetilde{\mathbf{H}}_2'(\mathbf{I}-\mathbf{S})'(\mathbf{I}-\mathbf{S})\mathbf{Y}_2, \qquad (11)$$

- $\mathbf{S}$ is a $N \times N$ matrix whose $i$th row is

$$(1 \; X_{i2} \; e_{i2})\left(\mathbf{D}'(X_{i1})\mathbf{K}(X_{i1})\mathbf{D}(X_{i1})\right)^{-1}\mathbf{D}'(X_{i1})\mathbf{K}(X_{i1}).$$

- In practice, the control function $e_{i2}$ should be replaced by its estimates such that

$$\widehat{e_{i2}^{x}} = X_{i2} - G_{i2}'(\sum_i G_{i2}G_{i2}'\kappa_h(X_{i1} - x))^{-1}(\sum_i G_{i2}X_{i2}'\kappa_h(X_{i1} - x)). \qquad (12)$$

- Feasible estimators: $\hat{\beta}(.)$ and $\hat{\gamma}_2$.

# Asymptotic properties of the estimator

Under the mild conditions, Assumptions we have

- The infeasible estimator $\widetilde{\gamma}$ and the first two rows of $\widetilde{\beta}(.)$, denoted $[\widetilde{\beta}(.)]_{1:2}$, satisfy the asymptotic properties that
  - $\sqrt{N}(\widetilde{\gamma} - \gamma) \rightarrow_d \mathcal{N}(0, \widetilde{\mathbf{\Sigma}}^{-1}\widetilde{\mathbf{\Psi}}\widetilde{\mathbf{\Sigma}}^{-1})$,
  - $\sqrt{Nh}([\widetilde{\beta}(x)]_{1:2} - [\beta(x)]_{1:2}) \rightarrow_d \mathcal{N}(0, \mathbf{\Omega}^{-1}(x)\widetilde{\mathbf{\Phi}}(x)\mathbf{\Omega}^{-1}(x))$.
- The feasible estimator $\hat{\gamma}$ and the first two rows of $\hat{\beta}(.)$, denoted $[\hat{\beta}(.)]_{1:2}$, satisfy the asymptotic properties of the infeasible estimators.
- Here, $\widetilde{\mathbf{\Sigma}} = \mathbb{E}[\widetilde{H}_{i2}\widetilde{H}'_{i2}] - \mathbb{E}[\mathbb{E}[\ddot{X}_{i2}\widetilde{H}'_{i2}|X_{i1}](\mathbb{E}[\widetilde{H}_{i2}\widetilde{H}'_{i2}|X_{i1}])^{-1}\mathbb{E}[\ddot{X}_{i2}\widetilde{H}'_{i2}]]$, and $\widetilde{\mathbf{\Phi}}(x)$ can be written by

$$\int \kappa^2(u) du \mathtt{f}_{X_1}(x)\mathbf{A}'(x)\mathbb{E}[(\epsilon_{i2}^2 + e_{i2}^2(\beta_2^e(X_{i1}))^2)G_{i2}G'_{i2}|X_{i1} = x]\mathbf{A}(x).$$

  for a matrix $\mathbf{A}(x)$.

# Simulation Study

# Simulation Study

- Consider the following DGP:

$$\begin{aligned} Y_i &= \widetilde{H}_i + \beta_1(X_{i1}) + \beta_2(X_{i1})X_{i2} + e_{i2} + 0.6\epsilon_i, \\ X_{i2} &= \widetilde{H}_i + \pi(X_{i1})Z_i + e_{i2}, \\ e_{i2} &= 0.4X_{i1}e_{i1}. \end{aligned} \tag{13}$$

- $X_{i1} \sim_{iid} U[-1, 1]$
- $e_{i1}$, $\epsilon_i$, $Z_i$ and $\widetilde{H}_i \sim_{iid} N(0, 1)$
- For $\{d_j\}_{j=1}^2 \sim \text{Unif}[0, 3]$ and $\{c_j\}_{j=1}^4 \sim \text{Unif}[0, 1]$, we let

$$\beta_{2,1}(x) = d_1\phi(x) + d_2, \tag{14}$$

  where $\phi(.)$ is the standard normal density function, and

$$\beta_2(x) = (2c_1 + c_2x)\exp(-2c_3(x - c_4)^2). \tag{15}$$

- $\pi(.)$ is the logistic distribution pdf multiplied by 2.

**Table 1:** RMSE estimates for $\beta(.)$

| $c_h$ | | | $N = 250$ | | | | $N = 500$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Mean | MED | MAD | IQR | Mean | MED | MAD | IQR |
| 2.5 | $\beta_1(.)$ | LC | 0.095 | 0.094 | 0.017 | 0.035 | 0.073 | 0.072 | 0.011 | 0.023 |
| | | LL | 0.105 | 0.103 | 0.018 | 0.036 | 0.079 | 0.078 | 0.013 | 0.025 |
| | $\beta_2(.)$ | LC | 0.123 | 0.117 | 0.026 | 0.053 | 0.09 | 0.085 | 0.017 | 0.035 |
| | | LL | 0.125 | 0.117 | 0.026 | 0.054 | 0.092 | 0.087 | 0.017 | 0.035 |
| 5 | $\beta_1(.)$ | LC | 0.071 | 0.068 | 0.017 | 0.036 | 0.054 | 0.051 | 0.012 | 0.024 |
| | | LL | 0.076 | 0.073 | 0.018 | 0.037 | 0.058 | 0.057 | 0.013 | 0.026 |
| | $\beta_2(.)$ | LC | 0.195 | 0.192 | 0.03 | 0.06 | 0.134 | 0.133 | 0.019 | 0.039 |
| | | LL | 0.112 | 0.104 | 0.028 | 0.057 | 0.083 | 0.078 | 0.018 | 0.038 |

- Bandwidth: $c_h N^{-0.3}$
- RMSE is computed as

$$\text{RMSE} = \left( \frac{1}{N} \sum_{i=1}^{N} (\bar{\beta}_s(X_{i1}) - \beta_s(X_{i1})) \right)^{1/2},$$

**Figure 1:** Estimates of $\beta_1(.)$

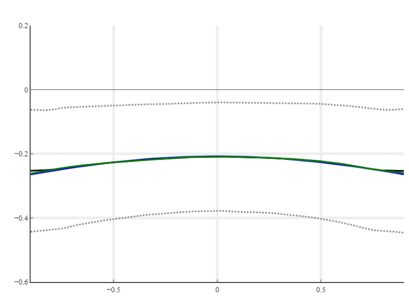
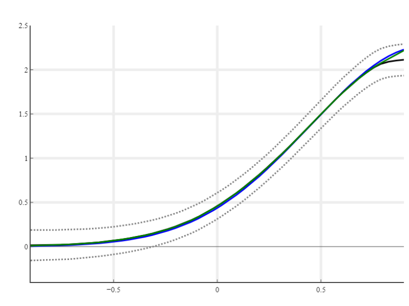



**Figure 2:** Estimates of $\beta_2(.)$




Functional coefficient estimates computed with the LC (black) and LL (green). The true coefficients are in blue solid line.

# Extension:

## Linear First Stage without nonlinearity w.r.t. $X_{it-1}$

## Extension: Semi-parametric estimation without dynamic structure of $X_{it}$

- So far, it has been assumed that $X_{i2} = \mathbb{L}_{G_{i2}}(X_{i2}|X_{i1}) + e_{i2}(X_{i1})$.
- The endogeneity of $X_{i2}$ comes from the endogenous adjustment to the initial treatment $X_{i1}$.

  e.g., Chinese import exposure at time 1 $\rightarrow$ endogenous change in industries $\rightarrow$ induces technical shock at time 2.
- In practice, technology shocks are often driven by unobservable factors that are not included in the model and those are often "persistent".
- For example, suppose that

$$\varepsilon_{i1} = \rho_1^0 e_{i1} + \eta_{i1} \text{ and } \varepsilon_{i2} = \rho_1 \varepsilon_{i1} + \rho_2^0 e_{i2} + \eta_{i2} \tag{16}$$

  where $e_{it} = X_{it} - \mathbb{L}_{G_{it}}(X_{it})$.
- Having only $e_{i2}$ as a control function is not enough and the above approach needs a slight modification.

# Extension: Semi-parametric estimation without dynamic structure of $X_{it}$

- Information on instruments for time $t = 1$ is needed.
- $f_{it}$: collection of instruments at time $t$, e.g., $f_{i1} = Z_{i1}, f_{i2} = (Z_{i1}\ Z_{i2})$.
- Assume the standard linear first stage: for $t = 1, 2$,

$$X_{it} = f_{it}'\alpha_t + H_{it}'\rho_t + e_{it}$$

  where $\mathbb{E}[e_{it}|f_{it}, \tilde{H}_{it}] = 0$ and the matrix $\mathbb{E}[(f_{it}'\ H_{it}')'(f_{it}'\ H_{it}')]$ is of full rank.

- Modified control function assumption:

$$\mathbb{E}[\varepsilon_{i2}|f_{i2}, \widetilde{H}_{i2}, e_{i1}, e_{i2}] = \mathbb{E}[\varepsilon_{i2}|e_{i1}, e_{i2}] \tag{17}$$

- Identification from the above:

$$\begin{aligned}&\mathbb{E}[\varepsilon_{i2}|X_{i1}, X_{i2}, Z_{i1}, Z_{i2}, \tilde{H}_{i2}, e_{i1}, e_{i2}]\\ =&\mathbb{E}[\varepsilon_{i2}|Z_{i1}, Z_{i2}, \tilde{H}_{i2}, e_{i1}, e_{i2}] = \mathbb{E}[\varepsilon_{i2}|e_{i1}, e_{i2}]\end{aligned}$$

- Assume $\mathbb{E}[\varepsilon_{i2}|e_{i1}, e_{i2}]$ is linear in $e_{i1}$ and $e_{i2}$.

# Extension: Estimation procedure

- $\beta(.) = (\beta_1^1(.), \beta_2^0(.))'$ and $\gamma_{-1} = (\gamma_2', \gamma_e')'$
- $\ddot{X}_{i2} = (1 \; X_{i2})'$ and $\ddot{H}_{i2} = (\widetilde{H}_{i2}' \; e_{i1} \; e_{i2})'$.
- $\mathbf{D}(x)$ be $N \times 4$ matrix whose $i$th row is $(\ddot{X}_{i2}' \; \frac{X_{i1}-x}{h}\ddot{X}_{i2})$.

$$\tilde{\gamma} = \left(\ddot{\mathbf{H}}_2'(\mathbf{I}-\mathbf{S})'(\mathbf{I}-\mathbf{S})\ddot{\mathbf{H}}_2\right)^{-1}\ddot{\mathbf{H}}_2'(\mathbf{I}-\mathbf{S})'(\mathbf{I}-\mathbf{S})\mathbf{Y}_2,$$
$$\tilde{\beta}(x) = \left[(\mathbf{D}'(x)\mathbf{K}(x)\mathbf{D}(x))^{-1}\mathbf{D}'(x)\mathbf{K}(x)\left(\mathbf{Y}_2 - \ddot{\mathbf{H}}_2\tilde{\gamma}\right)\right]_{[1:2]}, \text{ for all } x \in \mathcal{X}_1,$$

where $\mathbf{S}$ is a $N \times N$ smoothing matrix defined by

$$\mathbf{S} = \begin{bmatrix} (\ddot{X}_{12}' \; \mathbf{0}_2')(\mathbf{D}'(X_{11})\mathbf{K}(X_{11})\mathbf{D}(X_{11}))^{-1}\mathbf{D}'(X_{11})\mathbf{K}(X_{11}) \\ \vdots \\ (\ddot{X}_{N2}' \; \mathbf{0}_2')(\mathbf{D}'(X_{N1})\mathbf{K}(X_{N1})\mathbf{D}(X_{N1}))^{-1}\mathbf{D}'(X_{N1})\mathbf{K}(X_{N1}) \end{bmatrix}.$$

- $\hat{\gamma}$ and $\hat{\beta}(.)$ are the feasible version of $\widetilde{\gamma}$ and $\widetilde{\beta}(.)$ obtained using the estimated control functions.

## Extension: Asymptotic Properties

- The infeasible estimators $\tilde{\gamma}$ and $\tilde{\beta}(.)$ with known first-stage control functions satisfy that
  - $\sqrt{N}(\tilde{\gamma} - \gamma) \to_d \mathcal{N}(\mathbf{0}_{d_{h2}+1}, \mathbf{\Sigma}_1^{-1}\mathbf{\Psi}_2\mathbf{\Sigma}_1^{-1})$,
  - $\sqrt{Nh}(\tilde{\beta}(x) - \beta(x)) \to_d \mathcal{N}(\mathbf{0}_2, \mathbf{Q}^{-1}(x)\mathbf{\Phi}(x)\mathbf{Q}^{-1}(x))$, for all $x \in \mathcal{X}_1$.
- The feasible estimators with unknown first-stage control functions satisfy that
  - $\sqrt{N}(\widehat{\gamma} - \gamma) \to_d \mathcal{N}(\mathbf{0}_{d_{h2}+1}, \mathbf{\Sigma}_1^{-1}(\mathbf{\Psi}_1 + \mathbf{\Psi}_2)\mathbf{\Sigma}_1^{-1})$,
  - $\sqrt{Nh}(\widehat{\beta}(x) - \beta(x)) \to_d \mathcal{N}(\mathbf{0}_2, \mathbf{Q}^{-1}(x)\mathbf{\Phi}(x)\mathbf{Q}^{-1}(x))$, for all $x \in \mathcal{X}_1$.
- Note that $\mathbf{\Psi}_2 = \mathbb{E}[\epsilon_{i2}^2 \ddot{H}_{i2}^{\perp}(X_i)\ddot{H}_{i2}^{\perp}(X_i)']$ and $\mathbf{\Psi}_1 = \mathbb{E}[\varpi_j \gamma_e \gamma_e' \varpi_j']$ where

$$\varpi_{jt} = \mathbb{E}[(\ddot{H}_{i2} - \mathbf{P}'(X_{i1})\mathbf{Q}^{-1}(X_{i1})\ddot{X}_{i2})G_{it}'](\mathbb{E}[G_{it}G_{it}'])^{-1}G_{jt}e_{jt},$$

- In contrast with the previous case, the estimation error comes into the variance of the functional coefficients.

# Simulation Study

$$Y_{i2} = 0.5 + \beta_2(X_{i1})X_{i2} + e'_{i,1:2}\gamma_2^e + \pi_2\epsilon_{2t},$$
$$X_{i1} = Z_{i1}\rho_1^0 + e_{i1},$$
$$X_{i2} = Z_{i1}\rho_1^1 + Z_{i2}\rho_2^0 + e_{i2},$$

$e_{it}$, $Z_{it}$ are randomly drawn from Unif[-0.5,0.5] and Unif[-1,1]

$\epsilon_{i2} \sim_{iid} N(0,1)$, $\gamma_2^e = (0.3, 0.4, 0.4)$

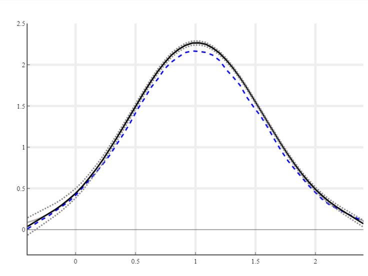
- CMI: estimator computed with conditional mean independence assumption
- CF: estimator computed with the control function approach

**Table 2:** RMSE estimates for $\beta_2(.)$

| | | $N = 250$ | | | | $N = 500$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $c_h$ | | Mean | MED | MAD | IQR | Mean | MED | MAD | IQR |
| 2.5 | CF | 0.087 | 0.081 | 0.017 | 0.035 | 0.069 | 0.066 | 0.011 | 0.022 |
| | CMI | 0.447 | 0.427 | 0.070 | 0.142 | 0.367 | 0.359 | 0.044 | 0.089 |
| 5 | CF | 0.076 | 0.068 | 0.015 | 0.034 | 0.051 | 0.047 | 0.011 | 0.023 |
| | CMI | 0.218 | 0.214 | 0.039 | 0.078 | 0.186 | 0.184 | 0.028 | 0.055 |

**Figure 3:** Estimated functional coefficient $\beta_2(.)$

**(a)** $c_h = 2.5$




**(b)** $c_h = 5$

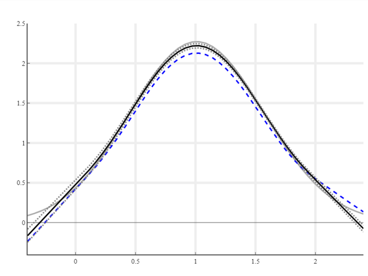
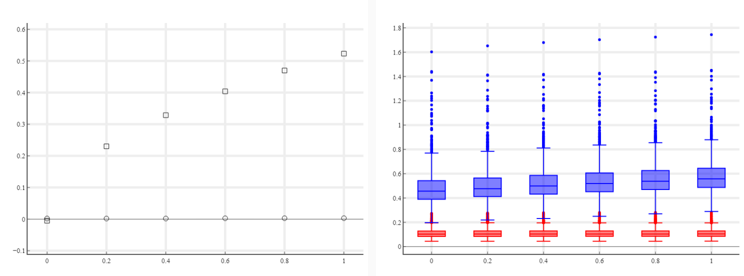


**Figure 5:** Estimation results at different $\lambda$

**(a)** Average Bias of $\gamma_{2,1}$ Estimates **(b)** Box plots of RMSEs




Notes: Figure 6a reports the average bias of the constant coefficient computed with our approach (sqaure) and the approach based on the CMI assumption (circle) is reported. Figure 6b reports the boxplots of the RMSEs of the CF estimates (red) and the CMI estimates (blue).

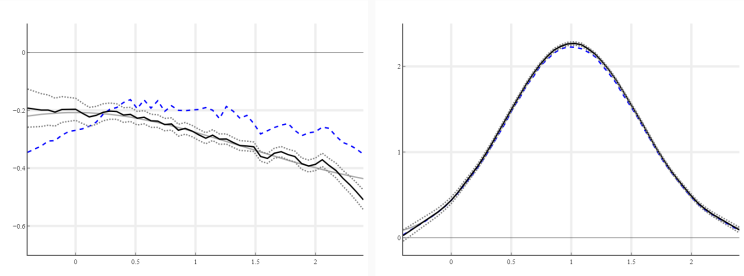
$$\frac{\gamma_{21}^{e2}\mathbb{E}[e_{i,1}^2]}{\gamma_2^e\mathbb{E}[e_{i,1:2}e_{i,1:2}']\gamma_2^e} = \lambda \quad \text{and} \quad \frac{\lambda\gamma_2^{e\prime}\mathbb{E}[e_{i,1:2}e_{i,1:2}']\gamma_2^e}{\lambda\gamma_2^{e\prime}\mathbb{E}[e_{i,1:2}e_{i,1:2}']\gamma_2^e + \pi_2^2} = 0.6,$$

for $\lambda \in \{0, 0.2, \ldots, 1\}$.

When the carryover effect is also path dependent:

**Figure 7:** Estimated functional coefficients

**(a)** $\beta_1(.)$ **(b)** $\beta_2(.)$



# Empirical Application

## Application to Acemoglu et al. (2016)

- $\Delta Y_{it} = \beta_1^1(\Delta X_{it-1}) + \beta_2^0(\Delta X_{it-1})X_{it} + \widetilde{H}_{it}'\gamma_t + \epsilon_{it}$
- Using the industry-level data of Acemoglu et al. (2016)
- 392 four-digit manufacturing industries over 2 time periods
- $Y_t$: log employment change over 1991-1999 ($t = 1$) and over 1999-2011 ($t = 2$).
- $X_t$: change in U.S. exposure to Chinese import over the same two periods.
- $Z_t$: change in exposure to Chinese import in a set of comparable countries over the same two periods.
- $\widetilde{H}_{it}$ includes computer investment, high-tech investment, changes in log real wage, production employment share and so on.
- The exogeneous variables are standardized

# Application to Acemoglu et al. (2016)

**Table 3:** Estimation Results

| x2 | Computer Investment | High-Tech Investment | log (Avr.Wage) | Change in real wage | Production employment | Capital/value added | Change in industry share | control 1 | control 2 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Naive TSLS | | | | | |
| -0.3445 | -0.1228 | -0.1064 | 1.1675*** | -0.6300** | -0.3615 | -0.1481 | 0.3951* | | |
| (0.2419) | (0.2810) | (0.5247) | (0.4033) | (0.2627) | (0.3283) | (0.2577) | (0.2236) | | |
| | | | | Dynamic Treatment | | | | | |
| | -0.5060* | 0.0410 | 1.7529*** | -0.6658 | -0.6193* | -0.0059 | 0.6311*** | | |
| | (0.2632) | (0.2029) | (0.6623) | (0.4922) | (0.3279) | (0.2249) | (0.2093) | | |
| | | | | Static Treatment | | | | | |
| | -0.4573* | 0.2116 | 1.4795* | -0.6750 | -0.5344* | -0.0589 | 0.6398*** | 1.4301 | 0.1665 |
| | (0.2739) | (0.2362) | (0.8017) | (0.5090) | (0.3103) | (0.2288) | (0.2128) | (3.2122) | (0.5156) |

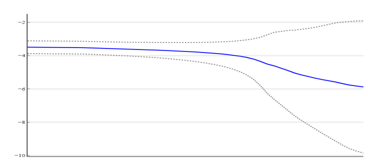
# Application to Acemoglu et al. (2016)

**Table 4:** Estimation Results (with $x_1 < 0.2$)

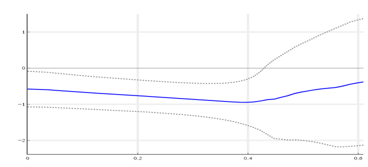
| x2 | Computer Investment | High-Tech Investment | log (Avr.Wage) | Change in real wage | Production employment | Capital/value added | Change in industry share | control 1 | control 2 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Naive TSLS | | | | | |
| -0.2659 | -0.4117$^{*}$ | -0.3604 | 0.5084$^{*}$ | -0.1133 | -0.9152$^{***}$ | -0.1477 | 0.4235$^{**}$ | | |
| (0.1884) | (0.2208) | (0.3844) | (0.2812) | (0.2007) | (0.2489) | (0.1912) | (0.1721) | | |
| | | | | Dynamic Treatment | | | | | |
| | -0.7455$^{***}$ | -0.1723 | 0.8295$^{***}$ | -0.1243 | -1.2104$^{***}$ | -0.0481 | 0.5832$^{**}$ | | |
| | (0.2622) | (0.1933) | (0.3027) | (0.2362) | (0.2458) | (0.2243) | (0.2605) | | |
| | | | | Static Treatment | | | | | |
| | -0.1007$^{*}$ | 0.0142 | 2.7308$^{**}$ | -0.0226 | -0.0751$^{***}$ | -0.1063 | 14.5023$^{**}$ | 27.4410 | -0.0503 |
| | (0.0571) | (0.0574) | (1.2683) | (0.0256) | (0.0259) | (0.4050) | (6.5019) | (23.1810) | (0.6722) |

**Figure 9:** Estimated path-dependent treatment effects

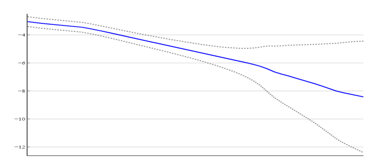
**(a)** carryover effect



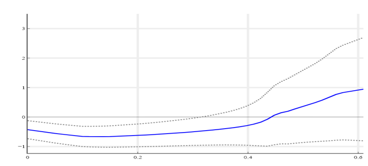

**(b)** contemporaneous effect




**(c)** carryover effect (static)




**(d)** contemporaneous effect (static)



# Conclusion

# Conclusion

- Propose a path-dependent dynamic treatment effect for a short panel setup
- Suggest a semi-parametric estimator based on the control function approach
- The control function approach is necessary due to the endogeneity of $X_{it}$ and to back out the total effect
- Various extensions: SSIV, static first stage, auto-regressive path dependency
- Still working on the empirical application.

Thank you !

## Extension: Inference

- For a specific parametric function $\beta(x, \mathfrak{b})$, test

$$\mathsf{H}_0 : \beta(x) = \beta(x, \mathfrak{b}) \quad \text{vs.} \quad \mathsf{H}_1 : \beta(x) \neq \beta(x, \mathfrak{b}).$$

- e.g., $\beta(x, \mathfrak{b}) = 0$ (signifiance testing), $\beta(x, \mathfrak{b}) = x'\mathfrak{b}$ (lienarity)
- A $\sqrt{N}$-consistenty estimator of $\mathfrak{b}$ is available.

$$\mathsf{F}_n = N \frac{\mathsf{RSS}_0 - \mathsf{RSS}_1}{\mathsf{RSS}_1}, \tag{18}$$

where $\mathsf{RSS}_s = N^{-1} \sum_{i=1}^{N} \mathfrak{R}_s(X_{i1})$ and

$$\mathfrak{R}_s(x) = \left( Y_2 - \mathbf{D}(x)\tilde{\beta}_s(x) - \hat{\ddot{\mathbf{H}}}_2 \widetilde{\gamma} \right)' \mathbf{K}(x) \left( Y_2 - \mathbf{D}(x)\tilde{\beta}_s(x) - \hat{\ddot{\mathbf{H}}}_2 \widetilde{\gamma} \right)$$

- $\widetilde{\beta}_s(.)$ estimators under the null.
- $\widetilde{\gamma}$ estimator of $\gamma$ under the null.

- Under the conditions for the limiting distribution of the functional coefficients, we have

$$\varrho^{-1}\left(\mathsf{F}_n - \frac{\mathbb{E}[\mathrm{tr}(\mathbf{Q}^{-1}(X_{i1})\mathbf{L}(X_{i1}))]g(0)}{h\mathbb{E}[\epsilon_{i2}^2 f_X(X_{i1})]}\right) \to_d \mathcal{N}(0,1),$$

where

$$\varrho^2 = \frac{2\mathbb{E}[\mathtt{f}_X(X_{i1})\mathrm{tr}(\mathbf{Q}^{-1}(X_{i1})\mathbf{L}(X_{i1})\mathbf{Q}^{-1}(X_{i1})\mathbf{L}(X_{i1}))]\int g^2(u)du}{h(\mathbb{E}[\epsilon_{i2}^2 f_X(X_{i1})])^2}$$

with $\mathbf{L}(X_{i1}) = \mathbb{E}[\epsilon_{i2}^2 \ddot{X}_{i2}\ddot{X}_{i2}' | X_{i1} = x]$, $\lambda_{22} = [\mathbf{\Lambda}]_{2,2} = \int u^2\kappa(u)du$ and $g(t) = \int_{\mathcal{X}_1} \kappa(u)\kappa(u-t) + \lambda_{22}^{-1}u(u-t)\kappa(u)\kappa(u-t)du$.

- In practice, size and power of the test depend on the bandwidth parameter

# Extension:

## Shift-Share Instrumental Variables

# Identification with Shift-share instrumental variables

- An important class of examples is the case with so-called "Shift-Share Instruments" (SSIV) that is given by

$$Z_{\ell t} = \sum_{i=1}^{N_I} \omega_{i\ell t} f_{it} \tag{19}$$

where $i$ denotes industry and $\ell$ is location at time $t$

- $\omega_{i\ell t}$: share of industry $i$ in location $\ell$ s.t. $\sum_i \omega_{i\ell t} = 1$
- $f_{it}$: exogenous shock given to the industry $i$
- In cross-sectional case, identification and consistency of the TSLS comes from the exogeneity of $f_{it}$ (Adao et al., 2020) or $\omega_{i\ell t}$ (Goldsmith-Pinkham et al., 2020)
- No discussion for the dynamic case.

# Identification with SSIV

- Back to our estimator:
- If we focus on the "contemporaneous treatment effect", the estimator is

$$\left(\widetilde{\mathbf{X}}_2'\mathbf{K}^{1/2}(x)\mathcal{P}_{\mathbf{K}^{1/2}\mathbf{Z}_2}(x)\mathbf{K}^{1/2}(x)\widetilde{\mathbf{X}}_2\right)^{-1}\widetilde{\mathbf{X}}_2'\mathbf{K}^{1/2}(x)\mathcal{P}_{\mathbf{K}^{1/2}\mathbf{Z}_2}(x)\mathbf{K}^{1/2}(x)\widetilde{\mathbf{Y}}_2.$$

  where for a random matrix $\mathbf{A}$ with $N_L$ rows, the matrices
  $\mathcal{P}_{\mathbf{K}^{1/2}\mathbf{A}}(x) = \mathbf{K}^{1/2}(x)\mathbf{A}(\mathbf{A}'\mathbf{K}(x)\mathbf{A})^{-1}\mathbf{A}'\mathbf{K}(x)$ and
  $\widetilde{\mathbf{A}} = (\mathbf{I} - \iota(\iota'\mathbf{K}(x)\iota)^{-1}\iota'\mathbf{K}(x))\mathbf{A}$.

- Given the invertibility of $\mathbf{Z}_2'\mathbf{K}(x)\widetilde{\mathbf{X}}_2$ and $\mathbf{Z}_2'\mathbf{K}(x)\mathbf{Z}_2$, it reduces to

$$\hat{\beta}_2^0(x) = (\mathbf{Z}_2'\mathbf{K}(x)\widetilde{\mathbf{X}}_2)^{-1}\mathbf{Z}_2'\mathbf{K}(x)\widetilde{\mathbf{Y}}_2,$$

# Identification with SSIV

$$\mathbf{Z}_2'\mathbf{K}(x)\widetilde{\mathbf{X}}_2 = \sum_\ell Z_{\ell 2}\widetilde{X}_{\ell 2}\kappa_h(X_{\ell 1} - x) = \sum_\ell \sum_i w_{i\ell 2} f_{i2}\widetilde{X}_{\ell 2}\kappa_h(X_{\ell 1} - x)$$
$$= \sum_i f_{i2}\Big(\sum_\ell w_{i\ell 2}\kappa_h(X_{\ell 1} - x)\Big)\sum_\ell \frac{w_{i\ell 2}\widetilde{X}_{\ell 2}\kappa_h(X_{\ell 1} - x)}{\sum_\ell w_{i\ell 2}\kappa_h(X_{\ell 1} - x)}$$
$$= \sum_i f_{i2}\bar{w}_{i2}(x)\bar{X}_{i2}(x)$$

**Proposition 1**

Suppose that $\beta_2^0(.)$ is continuously differentiable. Then, the SSIV estimator in (36) equals to the local constant IV estimator $\bar{\beta}_2^0(.)$ associated with the sample moment condition

$$N_I^{-1}\sum_i f_{i2}\bar{w}_{i2}(x)(\bar{Y}_{i2}(x) - \bar{X}_{i2}(x)\bar{\beta}_2^0(x)) = 0,$$

where $\bar{w}_{i2}(x) = \sum_\ell w_{\ell i2}\kappa_h(X_{\ell 1} - x)$ and $\bar{A}_{i2}(x) = \sum_\ell \frac{w_{\ell i2}\kappa_h(X_{\ell 1} - x)}{\sum_{\ell'} w_{\ell' i2}\kappa_h(X_{\ell' 1} - x)}\widetilde{A}_{\ell 2}$ for a random variable $A_{\ell 2}$.

## Identification with SSIV

In our dynamic case, the identification of the SSIV can be achieved from the following assumptions

- $f_{it} \perp (w_{\ell it}, \varepsilon_{it}) | X_{\ell(t-1)}$ for all $i = 1, \ldots, N_I$ and $\ell = 1, \ldots, N_L$;
- $\{f_{it}\}_{t=1}^{N_I}$ is i.i.d.;
- $\sum_i w_{\ell it} = 1$ for all $\ell = 1, \ldots, N_L$.

$$\Rightarrow \mathbb{E}[\sum_\ell Z_{\ell 2}\varepsilon_{\ell 2}|X_{\ell 1}] = \sum_\ell \mathbb{E}[f_{i2}|X_{\ell 1}]\mathbb{E}[\sum_i w_{\ell i2}\varepsilon_{\ell 2}|X_{\ell 1}] = \sum_\ell \mathbb{E}[f_{i2}|X_{\ell 1}]\mathbb{E}[\varepsilon_{\ell 2}|X_{\ell 1}],$$

and thus

$$\begin{aligned}
&\mathbb{E}[\sum_\ell Z_{\ell 2}(\varepsilon_{\ell 2} - \mathbb{E}[\varepsilon_{\ell 2}|X_{\ell 1} = x])|X_{\ell 1} = x] \\
&= \mathbb{E}[\sum_\ell Z_{\ell 2}\varepsilon_{\ell 2}|X_{\ell 1} = x] - \sum_\ell \mathbb{E}[\varepsilon_{\ell 2}|X_{\ell 1} = x]\mathbb{E}[\sum_i w_{\ell i2} f_{i2}|X_{\ell 1} = x] \\
&= \mathbb{E}[\sum_\ell Z_{\ell 2}\varepsilon_{\ell 2}|X_{\ell 1} = x] - \sum_\ell \mathbb{E}[\varepsilon_{\ell 2}|X_{\ell 1} = x]\mathbb{E}[f_{i2}|X_{\ell 1} = x] = 0,
\end{aligned}$$

Back

### Assumption 1

(A) *The density function of $X_{i1}$, denoted by $\mathtt{f}_{X_1}(\cdot)$, is Lipschitz continuous and bounded away from zero on its compact support $\mathcal{X}_1$.*

(B) *The kernel function $\kappa(\cdot)$ is a symmetric density function with a compact support.*

(C) *The function $\beta(\cdot) : \mathcal{X}_1 \to \mathbb{R}^3$ is continuously differentiable.*

(D) *There exists a constant $s > 2$ such that $\sup_{x\in\mathcal{X}_1} \mathbb{E}[\max\{\epsilon_{i2}^s, \|X_{i2}\|^{2s}, \|G_{i2}\|^{2s}\}|X_{i1} = x] < \infty$ and for some $\varsigma < 1 - s^{-1}$ satisfying $N^{2\varsigma-1}h \to \infty$.*

(E) $Nh^5 \to 0$ *and* $Nh^2/(\log N)^2 \to \infty$.