Endogeneity in Panel and Network Models: Identification without IV

Andrei Zeleneev (University College London and CeMMAP)

August 2024

Econometrics in Rio

Setup

Panel (or network, etc.) regression with endogeneity:

$$Y_{it} = X_{it}\beta + U_{it}$$

- Observe outcomes Y_{it} and covariates X_{it} for $i \in \{1, \dots, N\}$, $t \in \{1, \dots, T\}$
- X_{it} is endogenous, i.e., it can be (arbitrarily?) correlated with U_{it}
  - Straightforward to allow for multiple endogenous variables and additional controls

Main questions:

1. What are potential sources of endogeneity in panel and network models?
2. Can we identify "causal effect" β in large panels and networks without IVs and etc.?

Decomposition of unobservables

We decompose U_{it} into a low-dimensional part and noise

$$U_{it} = g(\alpha_i, \gamma_t) + \varepsilon_{it}, \quad U = G + \mathcal{E}$$

- α_i and γ_t are unobserved fixed effects (FE) and $g(\cdot, \cdot)$ is unknown
  - Flexible model covering two-way fixed effects, linear factor models, etc.
- $g(\alpha_i, \gamma_t)$ captures unobserved heterogeneity
- ε_{it} are mean zero and independent across (i, t) conditional on FE
  - Allowing for weak dependence is straightforward

Can be motivated using exchangeability (e.g., Aldous, 1981; Hoover, 1979)

$$U_{it} = f(\alpha_i, \gamma_t, \eta_{it}), \quad g(\alpha_i, \gamma_t) := \mathbb{E}\left[f(\alpha_i, \gamma_t, \eta_{it}) | \alpha_i, \gamma_t\right], \quad \varepsilon_{it} := U_{it} - g(\alpha_i, \gamma_t)$$

- α_i, γ_t and η_{it} are independent draws from P_α, P_γ, P_η

Decomposition of unobservables

We decompose U_{it} into a low-dimensional part and noise

$$U_{it} = g(\alpha_i, \gamma_t) + \varepsilon_{it}, \quad U = G + \mathcal{E}$$

- α_i and γ_t are unobserved fixed effects (FE) and $g(\cdot, \cdot)$ is unknown
  - Flexible model covering two-way fixed effects, linear factor models, etc.
- $g(\alpha_i, \gamma_t)$ captures unobserved heterogeneity
- ε_{it} are mean zero and independent across (i, t) conditional on FE
  - Allowing for weak dependence is straightforward

Can be motivated using exchangeability (e.g., Aldous, 1981; Hoover, 1979)

$$U_{it} = f(\alpha_i, \gamma_t, \eta_{it}), \quad g(\alpha_i, \gamma_t) := \mathbb{E}\left[f(\alpha_i, \gamma_t, \eta_{it}) | \alpha_i, \gamma_t\right], \quad \varepsilon_{it} := U_{it} - g(\alpha_i, \gamma_t)$$

  - α_i, γ_t and η_{it} are independent draws from P_α, P_γ, P_η

Two sources of endogeneity

$$U_{it} = g(\alpha_i, \gamma_t) + \varepsilon_{it}, \qquad U = G + \mathcal{E}$$

Similarly, we decompose X_{it} into low-dimensional and noise parts

$$X_{it} = h(\alpha_i^x, \gamma_t^x) + v_{it}, \qquad X = X^* + V$$

X_{it} can be endogenous for two reasons

1. **Low-dimensional endogeneity**: correlation between $g(\alpha_i, \gamma_t)$ and $h(\alpha_i^x, \gamma_t^x)$
    - Latent confounding factors, aggregate shocks, etc.
2. **Idiosyncratic endogeneity**: correlation between ε_{it} and v_{it}
    - Idiosyncratic confounders, measurement error, etc.

Low-dimensional and idiosyncratic components cannot be (strongly) correlated

- Immediately follows from joint exchangeability of X_{it} and U_{it}
- Also can be inspected using random matrix theory
- If X_{it} is low-dimensional, no need to worry about idiosyncratic endogeneity

Two sources of endogeneity

$$U_{it} = g(\alpha_i, \gamma_t) + \varepsilon_{it}, \qquad U = G + \mathcal{E}$$

Similarly, we decompose X_{it} into low-dimensional and noise parts

$$X_{it} = h(\alpha_i^x, \gamma_t^x) + v_{it}, \qquad X = X^* + V$$

X_{it} can be endogenous for two reasons

1. **Low-dimensional endogeneity**: correlation between $g(\alpha_i, \gamma_t)$ and $h(\alpha_i^x, \gamma_t^x)$
   - Latent confounding factors, aggregate shocks, etc.
2. **Idiosyncratic endogeneity**: correlation between ε_{it} and v_{it}
   - Idiosyncratic confounders, measurement error, etc.

Low-dimensional and idiosyncratic components cannot be (strongly) correlated

- Immediately follows from joint exchangeability of X_{it} and U_{it}
- Also can be inspected using random matrix theory
- If X_{it} is low-dimensional, no need to worry about idiosyncratic endogeneity

Two sources of endogeneity

$$U_{it} = g(\alpha_i, \gamma_t) + \varepsilon_{it}, \qquad U = G + \mathcal{E}$$

Similarly, we decompose X_{it} into low-dimensional and noise parts

$$X_{it} = h(\alpha_i^x, \gamma_t^x) + v_{it}, \qquad X = X^* + V$$

X_{it} can be endogenous for two reasons

1. **Low-dimensional endogeneity**: correlation between $g(\alpha_i, \gamma_t)$ and $h(\alpha_i^x, \gamma_t^x)$
    - Latent confounding factors, aggregate shocks, etc.
2. **Idiosyncratic endogeneity**: correlation between ε_{it} and v_{it}
    - Idiosyncratic confounders, measurement error, etc.

Low-dimensional and idiosyncratic components cannot be (strongly) correlated

- Immediately follows from joint exchangeability of X_{it} and U_{it}
- Also can be inspected using random matrix theory
- If X_{it} is low-dimensional, no need to worry about idiosyncratic endogeneity

Two sources of endogeneity

$$U_{it} = g(\alpha_i, \gamma_t) + \varepsilon_{it}, \qquad U = G + \mathcal{E}$$

Similarly, we decompose X_{it} into low-dimensional and noise parts

$$X_{it} = h(\alpha_i^x, \gamma_t^x) + v_{it}, \qquad X = X^* + V$$

X_{it} can be endogenous for two reasons

1. **Low-dimensional endogeneity**: correlation between $g(\alpha_i, \gamma_t)$ and $h(\alpha_i^x, \gamma_t^x)$
   - Latent confounding factors, aggregate shocks, etc.
2. **Idiosyncratic endogeneity**: correlation between ε_{it} and v_{it}
   - Idiosyncratic confounders, measurement error, etc.

Low-dimensional and idiosyncratic components cannot be (strongly) correlated

- Immediately follows from joint exchangeability of X_{it} and U_{it}
- Also can be inspected using random matrix theory
- If X_{it} is low-dimensional, no need to worry about idiosyncratic endogeneity

Setup: Recap

$$Y_{it} = X_{it}\beta + g(\alpha_i, \gamma_t) + \varepsilon_{it}, \qquad X_{it} = h(\alpha_i^x, \gamma_t^x) + v_{it}$$

- Observe Y_{it}, X_{it} for $i \in \{1, \ldots, N\}$, $t \in \{1, \ldots, T\}$
- FE (α_i, α_i^x) and (γ_t, γ_t^x) are independent draws from $P_{\mathcal{A}}$ and P_{Γ}
- $(\varepsilon_{it}, v_{it})$ are mean zero and independent across (i, t) conditional on FE
- Functions $g(\cdot, \cdot)$ and $h(\cdot, \cdot)$ are unknown and the dimensions of FE are not specified
- Straightforward to allow for weak-dependence of γ_t and $(\varepsilon_{it}, v_{it})$

Question: Can we identify and consistently estimate β in large panels $(N, T \to \infty)$ without additional information (e.g., IV) or assumptions?

Purging idiosyncratic endogeneity

We can rewrite model as

$$Y_{it} = \underbrace{h(\alpha_i^x, \gamma_t^x)\beta + g(\alpha_i, \gamma_t)}_{\text{low-dimensional signal}} + \underbrace{\varepsilon_{it} + v_{it}\beta}_{\text{noise}}, \quad Y = \underbrace{X^*\beta + G}_{:= Y^*} + \underbrace{\mathcal{E} + V\beta}_{\text{noise}}$$

Idea: we can denoise Y_{it} and X_{it} to purge idiosyncratic endogeneity

- Zeleneev (2019) provides an estimator consistent in the max norm for Y^* and X^*
  - Or one can use standard tools from matrix denoising/completion literature (e.g., PCA)

For identification, we can treat Y_{it}^* and X_{it}^* as effectively observed

$$Y_{it}^* = X_{it}^*\beta + g(\alpha_i, \gamma_t)$$

Purging idiosyncratic endogeneity

We can rewrite model as

$$Y_{it} = \underbrace{h(\alpha_i^x, \gamma_t^x)\beta + g(\alpha_i, \gamma_t)}_{\text{low-dimensional signal}} + \underbrace{\varepsilon_{it} + v_{it}\beta}_{\text{noise}}, \quad Y = \underbrace{X^*\beta + G}_{:=Y^*} + \underbrace{\mathcal{E} + V\beta}_{\text{noise}}$$

Idea: we can denoise Y_{it} and X_{it} to purge idiosyncratic endogeneity

- Zeleneev (2019) provides an estimator consistent in the max norm for Y^* and X^*
    - Or one can use standard tools from matrix denoising/completion literature (e.g., PCA)

For identification, we can treat Y_{it}^* and X_{it}^* as effectively observed

$$Y_{it}^* = X_{it}^*\beta + g(\alpha_i, \gamma_t)$$

Purging idiosyncratic endogeneity

We can rewrite model as

$$Y_{it} = \underbrace{h(\alpha_i^x, \gamma_t^x)\beta + g(\alpha_i, \gamma_t)}_{\text{low-dimensional signal}} + \underbrace{\varepsilon_{it} + v_{it}\beta}_{\text{noise}}, \quad Y = \underbrace{X^*\beta + G}_{:=Y^*} + \underbrace{\mathcal{E} + V\beta}_{\text{noise}}$$

Idea: we can denoise Y_{it} and X_{it} to purge idiosyncratic endogeneity

- Zeleneev (2019) provides an estimator consistent in the max norm for Y^* and X^*
  - Or one can use standard tools from matrix denoising/completion literature (e.g., PCA)

For identification, we can treat Y_{it}^* and X_{it}^* as effectively observed

$$Y_{it}^* = X_{it}^*\beta + g(\alpha_i, \gamma_t)$$

Disentangling low-dimensional endogeneity

We purged idiosyncratic endogeneity by denoising Y_{it} and X_{it}

$$Y_{it}^* = X_{it}^* \beta + g(\alpha_i, \gamma_t), \qquad X_{it}^* = h(\alpha_i^x, \gamma_t^x)$$

Question: can we identify β from Y_{it}^* and X_{it}^*?

Challenge: both regressor X_{it}^* and unobserved $g(\alpha_i, \gamma_t)$ are (approximately) low rank

- Most methods in the panel literature rely on "full rank" exogenous variation in X_{it}
  - Some limited results (Moon and Weidner, 2017, 2019) in linear factor models
  - No general identification/consistency results with (approximately) low rank regressors

Good news: we can build on the general results of Zeleneev (2019)

Disentangling low-dimensional endogeneity

We purged idiosyncratic endogeneity by denoising Y_{it} and X_{it}

$$Y^*_{it} = X^*_{it}\beta + g(\alpha_i, \gamma_t), \qquad X^*_{it} = h(\alpha^x_i, \gamma^x_t)$$

Question: can we identify β from Y^*_{it} and X^*_{it}?

Challenge: both regressor X^*_{it} and unobserved $g(\alpha_i, \gamma_t)$ are (approximately) low rank

- Most methods in the panel literature rely on "full rank" exogenous variation in X_{it}
  - Some limited results (Moon and Weidner, 2017, 2019) in linear factor models
  - No general identification/consistency results with (approximately) low rank regressors

Good news: we can build on the general results of Zeleneev (2019)

Disentangling low-dimensional endogeneity

We purged idiosyncratic endogeneity by denoising Y_{it} and X_{it}

$$Y_{it}^* = X_{it}^* \beta + g(\alpha_i, \gamma_t), \qquad X_{it}^* = h(\alpha_i^x, \gamma_t^x)$$

Question: can we identify β from Y_{it}^* and X_{it}^*?

Challenge: both regressor X_{it}^* and unobserved $g(\alpha_i, \gamma_t)$ are (approximately) low rank

- Most methods in the panel literature rely on "full rank" exogenous variation in X_{it}
  - Some limited results (Moon and Weidner, 2017, 2019) in linear factor models
  - No general identification/consistency results with (approximately) low rank regressors

Good news: we can build on the general results of Zeleneev (2019)

Identification of β

$$Y^*_{it} = X^*_{it}\beta + g(\alpha_i, \gamma_t)$$

- Suppose we find two units i and j with $\alpha_i = \alpha_j$:

$$Y^*_{it} - Y^*_{jt} = (X^*_{it} - X^*_{jt})\beta$$

- Can identify β by regressing $Y^*_{it} - Y^*_{jt}$ on $X^*_{it} - X^*_{jt} = h(\alpha^x_i, \gamma^x_t) - h(\alpha^x_j, \gamma^x_t)$
- Need variation in $X^*_{it} - X^*_{jt}$ conditional on $\alpha_i = \alpha_j$: need $\alpha^x_i \neq \alpha^x_j$

Question: how do we match such units?

Finding units with the same values of α

$$Y_{it}^* = X_{it}^*\beta + g(\alpha_i, \gamma_t)$$

Consider the following pseudo-distance between units i and j

$$\begin{aligned} d_{ij}^2 &:= \min_{b \in \mathcal{B}} \mathbb{E}\left[(Y_{it}^* - Y_{jt}^* - (X_{it}^* - X_{jt}^*)b)^2 | \alpha_i^x, \alpha_i, \alpha_j^x, \alpha_j\right] \\ &= \min_{b \in \mathcal{B}} \mathbb{E}\left[\underbrace{(g(\alpha_i, \gamma_t) - g(\alpha_j, \gamma_t)}_{=0, \text{ when } \alpha_i = \alpha_j} - (X_{it}^* - X_{jt}^*)(b - \beta))^2 | \alpha_i^x, \alpha_i, \alpha_j^x, \alpha_j\right] \end{aligned}$$

- Expectation over (γ_t^x, γ_t)
- Identified from Y^* and X^*

Under a standard rank condition, $d_{ij}^2 = 0$ implies $\alpha_i = \alpha_j$ and $b^* = \beta$ a.s.

Finding units with the same values of α

$$Y_{it}^* = X_{it}^* \beta + g(\alpha_i, \gamma_t)$$

Consider the following pseudo-distance between units i and j

$$\begin{aligned} d_{ij}^2 &:= \min_{b \in \mathcal{B}} \mathbb{E}\left[(Y_{it}^* - Y_{jt}^* - (X_{it}^* - X_{jt}^*)b)^2 | \alpha_i^x, \alpha_i, \alpha_j^x, \alpha_j \right] \\ &= \min_{b \in \mathcal{B}} \mathbb{E}\left[\underbrace{(g(\alpha_i, \gamma_t) - g(\alpha_j, \gamma_t)}_{=0, \text{ when } \alpha_i = \alpha_j} - (X_{it}^* - X_{jt}^*)(b - \beta))^2 | \alpha_i^x, \alpha_i, \alpha_j^x, \alpha_j \right] \end{aligned}$$

- Expectation over (γ_t^x, γ_t)
- Identified from Y^* and X^*

Under a standard rank condition, $d_{ij}^2 = 0$ implies $\alpha_i = \alpha_j$ and $b^* = \beta$ a.s.

Finding units with the same values of α

$$Y_{it}^* = X_{it}^*\beta + g(\alpha_i, \gamma_t)$$

Consider the following pseudo-distance between units i and j

$$\begin{aligned} d_{ij}^2 &:= \min_{b \in \mathcal{B}} \mathbb{E}\left[(Y_{it}^* - Y_{jt}^* - (X_{it}^* - X_{jt}^*)b)^2 | \alpha_i^x, \alpha_i, \alpha_j^x, \alpha_j\right] \\ &= \min_{b \in \mathcal{B}} \mathbb{E}\left[\underbrace{(g(\alpha_i, \gamma_t) - g(\alpha_j, \gamma_t)}_{=0, \text{ when } \alpha_i = \alpha_j} - (X_{it}^* - X_{jt}^*)(b - \beta))^2 | \alpha_i^x, \alpha_i, \alpha_j^x, \alpha_j\right] \end{aligned}$$

- Expectation over (γ_t^x, γ_t)
- Identified from Y^* and X^*

Under a standard rank condition, $d_{ij}^2 = 0$ implies $\alpha_i = \alpha_j$ and $b^* = \beta$ a.s.

Identification: Recap

1. Identification of Y_{it}^* and X_{it}^*
    - Based on identification of units with the same values of both α and α^x (and γ and γ^x)

$$Y_{it}^* = X_{it}^* \beta + g(\alpha_i, \gamma_t)$$

2. Identification of units with same values of α but different values of α^x
3. Identification of β

$$Y_{it}^* - Y_{jt}^* = \underbrace{(h(\alpha_i^x, \gamma_t^x) - h(\alpha_j^x, \gamma_t^x))}_{X_{it}^* - X_{jt}^*} \beta$$

Identification: Recap

1. Identification of Y_{it}^* and X_{it}^*
    - Based on identification of units with the same values of both α and α^x (and γ and γ^x)

$$Y_{it}^* = X_{it}^* \beta + g(\alpha_i, \gamma_t)$$

2. Identification of units with same values of α but different values of α^x

3. Identification of β

$$Y_{it}^* - Y_{jt}^* = \underbrace{(h(\alpha_i^x, \gamma_t^x) - h(\alpha_j^x, \gamma_t^x))}_{X_{it}^* - X_{jt}^*} \beta$$

Identification: Recap

1. Identification of Y_{it}^* and X_{it}^*
   - Based on identification of units with the same values of both α and α^x (and γ and γ^x)

$$Y_{it}^* = X_{it}^* \beta + g(\alpha_i, \gamma_t)$$

2. Identification of units with same values of α but different values of α^x

3. Identification of β

$$Y_{it}^* - Y_{jt}^* = \underbrace{(h(\alpha_i^x, \gamma_t^x) - h(\alpha_j^x, \gamma_t^x))}_{X_{it}^* - X_{jt}^*}\beta$$

Identification: Discussion

$$Y_{it} = X_{it}\beta + g(\alpha_i, \gamma_t) + \varepsilon_{it}$$

To identify β, we rely on exogenous low-dimensional variation in X_{it}

Main identifying assumptions (all testable!):

- X_{it} has a non-degenerate low-dimensional variation in X_{it}^*
- We have variation in $X_{it}^* - X_{jt}^*$ conditional on $\alpha_i = \alpha_j$, i.e., we have $\alpha_i^x \neq \alpha_j^x$
- Variation in $X_{it}^* - X_{jt}^*$ is not collinear with variation in $g(\alpha_i, \gamma_t) - g(\alpha_j, \gamma_t)$
    - Otherwise, we will have $d_{ij}^2 = 0$ achieved with different values of $b^* \neq \beta$

Estimation

1. Estimation of $\hat{Y}^*_{it}$ and $\hat{X}^*_{it}$ by denoising/completing matrices Y and X
   - Zeleneev (2019) provided $\hat{Y}^*$ and demonstrated $\|\hat{Y}^* - Y^*\|_{\max} \to_p 0$
   - Or one can use some other standard denoising technique such as PCA, USVT, etc., ...

2. Construction of $\hat{d}^2_{ij}$

$$\hat{d}^2_{ij} := \min_{b \in \mathcal{B}} \frac{1}{T} \sum_{t=1}^{T} (\hat{Y}^*_{it} - \hat{Y}^*_{jt} - (\hat{X}^*_{it} - \hat{X}^*_{jt})b)^2$$

3. Matching agents based on $\hat{d}^2_{ij}$ and estimation of β

$$\hat{\beta} := \frac{\sum_{i<j} K\left(\frac{\hat{d}^2_{ij}}{h_n^2}\right) \sum_{t=1}^{T} (\hat{X}^*_{it} - \hat{X}^*_{jt})(\hat{Y}^*_{it} - \hat{Y}^*_{jt})}{\sum_{i<j} K\left(\frac{\hat{d}^2_{ij}}{h_n^2}\right) \sum_{t=1}^{t} (\hat{X}^*_{it} - \hat{X}^*_{jt})^2}$$

   - K and h_n are some kernel and bandwidth
   - Combines regressions of $\hat{Y}_{it} - \hat{Y}_{jt}$ on $\hat{X}^*_{it} - \hat{X}^*_{jt}$ based on kernel weights
   - Additional gains are possible if we also match along the t dimension

Estimation

1. Estimation of $\hat{Y}^*_{it}$ and $\hat{X}^*_{it}$ by denoising/completing matrices Y and X
   - Zeleneev (2019) provided $\hat{Y}^*$ and demonstrated $\|\hat{Y}^* - Y^*\|_{\max} \to_p 0$
   - Or one can use some other standard denoising technique such as PCA, USVT, etc., ...

2. Construction of $\hat{d}^2_{ij}$

$$\hat{d}^2_{ij} := \min_{b \in \mathcal{B}} \frac{1}{T} \sum_{t=1}^{T} (\hat{Y}^*_{it} - \hat{Y}^*_{jt} - (\hat{X}^*_{it} - \hat{X}^*_{jt})b)^2$$

3. Matching agents based on $\hat{d}^2_{ij}$ and estimation of β

$$\hat{\beta} := \frac{\sum_{i<j} K\left(\frac{\hat{d}^2_{ij}}{h_n^2}\right) \sum_{t=1}^{T} (\hat{X}^*_{it} - \hat{X}^*_{jt})(\hat{Y}^*_{it} - \hat{Y}^*_{jt})}{\sum_{i<j} K\left(\frac{\hat{d}^2_{ij}}{h_n^2}\right) \sum_{t=1}^{t} (\hat{X}^*_{it} - \hat{X}^*_{jt})^2}$$

   - K and h_n are some kernel and bandwidth
   - Combines regressions of $\hat{Y}_{it} - \hat{Y}_{jt}$ on $\hat{X}^*_{it} - \hat{X}^*_{jt}$ based on kernel weights
   - Additional gains are possible if we also match along the t dimension

Estimation

1. Estimation of $\hat{Y}^*_{it}$ and $\hat{X}^*_{it}$ by denoising/completing matrices Y and X
   - Zeleneev (2019) provided $\hat{Y}^*$ and demonstrated $\|\hat{Y}^* - Y^*\|_{\max} \to_p 0$
   - Or one can use some other standard denoising technique such as PCA, USVT, etc., ...
2. Construction of $\hat{d}^2_{ij}$

$$\hat{d}^2_{ij} := \min_{b \in \mathcal{B}} \frac{1}{T} \sum_{t=1}^{T} (\hat{Y}^*_{it} - \hat{Y}^*_{jt} - (\hat{X}^*_{it} - \hat{X}^*_{jt})b)^2$$

3. Matching agents based on $\hat{d}^2_{ij}$ and estimation of β

$$\hat{\beta} := \frac{\sum_{i<j} K\left(\frac{\hat{d}^2_{ij}}{h_n^2}\right) \sum_{t=1}^{T} (\hat{X}^*_{it} - \hat{X}^*_{jt})(\hat{Y}^*_{it} - \hat{Y}^*_{jt})}{\sum_{i<j} K\left(\frac{\hat{d}^2_{ij}}{h_n^2}\right) \sum_{t=1}^{t} (\hat{X}^*_{it} - \hat{X}^*_{jt})^2}$$

   - K and h_n are some kernel and bandwidth
   - Combines regressions of $\hat{Y}_{it} - \hat{Y}_{jt}$ on $\hat{X}^*_{it} - \hat{X}^*_{jt}$ based on kernel weights
   - Additional gains are possible if we also match along the t dimension

Estimation

1. Estimation of $\hat{Y}^*_{it}$ and $\hat{X}^*_{it}$ by denoising/completing matrices Y and X
   - Zeleneev (2019) provided $\hat{Y}^*$ and demonstrated $\|\hat{Y}^* - Y^*\|_{\max} \to_p 0$
   - Or one can use some other standard denoising technique such as PCA, USVT, etc., ...
2. Construction of $\hat{d}^2_{ij}$

$$\hat{d}^2_{ij} := \min_{b \in \mathcal{B}} \frac{1}{T} \sum_{t=1}^{T} (\hat{Y}^*_{it} - \hat{Y}^*_{jt} - (\hat{X}^*_{it} - \hat{X}^*_{jt})b)^2$$

3. Matching agents based on $\hat{d}^2_{ij}$ and estimation of β

$$\hat{\beta} := \frac{\sum_{i<j} K\left(\frac{\hat{d}^2_{ij}}{h_n^2}\right) \sum_{t=1}^{T} (\hat{X}^*_{it} - \hat{X}^*_{jt})(\hat{Y}^*_{it} - \hat{Y}^*_{jt})}{\sum_{i<j} K\left(\frac{\hat{d}^2_{ij}}{h_n^2}\right) \sum_{t=1}^{t} (\hat{X}^*_{it} - \hat{X}^*_{jt})^2}$$

   - K and h_n are some kernel and bandwidth
   - Combines regressions of $\hat{Y}_{it} - \hat{Y}_{jt}$ on $\hat{X}^*_{it} - \hat{X}^*_{jt}$ based on kernel weights
   - Additional gains are possible if we also match along the t dimension

Illustration: Nakamura and Steinsson (2014)

Researchers often construct IV leveraging differential exposures to aggregate shocks

$$Y_{it} = X_{it}\beta + \varepsilon_{it}, \qquad X_{it} = \pi_i Z_t + V_{it} \qquad \text{(plus TWFE)}$$

Nakamura and Steinsson (2014):

- Y_{it} and X_{it} represent growth in output and military procurement spending
- Z_t is the change in total government spending
- Fiscal multiplier β is identified using "exogenous" low-rank variation in $\pi_i Z_t$

Arkhangelsky and Korovkin (2019) revisit this application and develop an alternative estimator robust to the presence of aggregate unobserved shocks

$$Y_{it} = X_{it}\beta + \theta_i H_t + \varepsilon_{it}, \qquad X_{it} = \pi_i Z_t + \theta_i^x H_t + V_{it} \qquad \text{(plus TWFE)}$$

- The goal is to allow for low-rank endogeneity due to unobserved confounder H_t

Illustration: Nakamura and Steinsson (2014)

Researchers often construct IV leveraging differential exposures to aggregate shocks

$$Y_{it} = X_{it}\beta + \varepsilon_{it}, \qquad X_{it} = \pi_i Z_t + V_{it} \qquad \text{(plus TWFE)}$$

Nakamura and Steinsson (2014):

- Y_{it} and X_{it} represent growth in output and military procurement spending
- Z_t is the change in total government spending
- Fiscal multiplier β is identified using "exogenous" low-rank variation in $\pi_i Z_t$

Arkhangelsky and Korovkin (2019) revisit this application and develop an alternative estimator robust to the presence of aggregate unobserved shocks

$$Y_{it} = X_{it}\beta + \theta_i H_t + \varepsilon_{it}, \qquad X_{it} = \pi_i Z_t + \theta_i^x H_t + V_{it} \qquad \text{(plus TWFE)}$$

- The goal is to allow for low-rank endogeneity due to unobserved confounder H_t

Illustration: Nakamura and Steinsson (2014)

Both approaches rely on observing Z_t to extract “exogenous” low-rank variation $\pi_i Z_t$

$$Y_{it} = X_{it}\beta + \theta_i H_t + \varepsilon_{it}, \qquad X_{it} = \pi_i Z_t + \theta_i^x H_t + V_{it} \qquad \text{(plus TWFE)}$$

Fits into our framework with $\alpha_i = \theta_i$, $\gamma_t = H_t$, $\alpha_i^x = (\pi_i, \theta_i^x)$, and $\gamma_t^x = (Z_t, H_t)$

Our method: We don’t need to observe Z_t (or any other IV). Our method automatically extracts low-rank variation from X_{it} and uses it to identify β

- The identifying assumptions are satisfied provided that
  - π_i’s are heterogeneous
  - Z_t and H_t are not collinear
- These assumptions are already imposed in Arkhangelsky and Korovkin (2019)
- Additional efficiency gains are possible if other exogenous factors in X_{it} exist

Illustration: Nakamura and Steinsson (2014)

Both approaches rely on observing Z_t to extract "exogenous" low-rank variation $\pi_i Z_t$

$$Y_{it} = X_{it}\beta + \theta_i H_t + \varepsilon_{it}, \qquad X_{it} = \pi_i Z_t + \theta_i^x H_t + V_{it} \qquad \text{(plus TWFE)}$$

Fits into our framework with $\alpha_i = \theta_i$, $\gamma_t = H_t$, $\alpha_i^x = (\pi_i, \theta_i^x)$, and $\gamma_t^x = (Z_t, H_t)$

Our method: We don't need to observe Z_t (or any other IV). Our method automatically extracts low-rank variation from X_{it} and uses it to identify β

- The identifying assumptions are satisfied provided that
  - π_i's are heterogeneous
  - Z_t and H_t are not collinear
- These assumptions are already imposed in Arkhangelsky and Korovkin (2019)
- Additional efficiency gains are possible if other exogenous factors in X_{it} exist

Numerical evidence

Motivated by Nakamura and Steinsson (2014) and Arkhangelsky and Korovkin (2019)

$$Y_{it} = X_{it}\beta + \theta_i H_t + \varepsilon_{it}, \quad X_{it} = \pi_i Z_t + \theta_i^x H_t + V_{it}, \quad \begin{pmatrix} \varepsilon_{it} \\ V_{it} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)$$

$$\begin{pmatrix} \theta_i \\ \theta_i^x \\ \pi_i \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 & 0.7 \\ 0.7 & 1 & 0.7 \\ 0.7 & 0.7 & 1 \end{pmatrix} \right), \quad \begin{pmatrix} H_t \\ Z_t \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \right)$$

Estimators:

- OLS
- TSLS using Z_t as in Nakamura and Steinsson (2014)
- Bai (2009)
- Bai and Ng (2010) using X_{it}^* as an IV
- Our method using Epanechnikov kernel and Silverman's rule of thumb to choose h_n^2

Numerical evidence

Motivated by Nakamura and Steinsson (2014) and Arkhangelsky and Korovkin (2019)

$$Y_{it} = X_{it}\beta + \theta_i H_t + \varepsilon_{it}, \quad X_{it} = \pi_i Z_t + \theta_i^x H_t + V_{it}, \quad \begin{pmatrix} \varepsilon_{it} \\ V_{it} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)$$

$$\begin{pmatrix} \theta_i \\ \theta_i^x \\ \pi_i \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 & 0.7 \\ 0.7 & 1 & 0.7 \\ 0.7 & 0.7 & 1 \end{pmatrix} \right), \quad \begin{pmatrix} H_t \\ Z_t \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \right)$$

Estimators:

- OLS
- TSLS using Z_t as in Nakamura and Steinsson (2014)
- Bai (2009)
- Bai and Ng (2010) using X_{it}^* as an IV
- Our method using Epanechnikov kernel and Silverman's rule of thumb to choose h_n^2

Numerical evidence

	$(N,T)=(50,30)$			$(N,T)=(100,50)$			$(N,T)=(200,100)$		
	bias	std	rmse	bias	std	rmse	bias	std	rmse
OLS	0.425	0.046	0.428	0.425	0.032	0.426	0.424	0.023	0.425
TSLS	0.337	0.061	0.342	0.337	0.043	0.340	0.336	0.030	0.338
Bai (2009)	0.334	0.042	0.337	0.329	0.026	0.330	0.327	0.018	0.327
Bai & Ng (2010)	0.406	0.058	0.410	0.403	0.042	0.405	0.400	0.030	0.401
This paper	0.033	0.124	0.128	-0.001	0.079	0.079	-0.013	0.044	0.045

Based on 5,000 replications

Numerical evidence

	$(N,T)=(50,30)$			$(N,T)=(100,50)$			$(N,T)=(200,100)$		
	bias	std	rmse	bias	std	rmse	bias	std	rmse
OLS	0.425	0.046	0.428	0.425	0.032	0.426	0.424	0.023	0.425
TSLS	0.337	0.061	0.342	0.337	0.043	0.340	0.336	0.030	0.338
Bai (2009)	0.334	0.042	0.337	0.329	0.026	0.330	0.327	0.018	0.327
Bai & Ng (2010)	0.406	0.058	0.410	0.403	0.042	0.405	0.400	0.030	0.401
This paper	0.033	0.124	0.128	-0.001	0.079	0.079	-0.013	0.044	0.045

Based on 5,000 replications

Conclusion

Establish identification of β in a class of panel and network models with endogeneity

- Do not specify the form of endogeneity and the role of unobservables
- Do not rely on observing auxiliary identifying variables such as IV

Our method automatically extracts exogenous “low rank” variation and use it to ID β

- The main identifying assumptions are transparent and falsifiable

Provide a consistent estimator of β

- Easy to implement
- Works well in Monte-Carlo experiments

Inference is not (yet) in the paper but...

- Seems possible for linear factor models with strong factors or...
- Maybe we can extend the approach Armstrong, Weidner, and Zeleneev (2022)