The Problem With Historical Instrumental Variables

Alternative title: Identification of Bi-Directional Two-Variable System With Time-Invariant Instrument

John T.H. Wong

Illustrating the issue with a DAG

- Many historical IV papers use the following strategy.
  - e.g., AJR use settlers mortality (z_i) to instrument for constraints on the government's executive (d_{it}), and then estimate the latter's effect on output growth (y_{it}).

$$
\begin{tikzcd}
z_i \arrow[r, bend left] & d_{it} \arrow[d] \\
& y_{it}
\end{tikzcd}
$$

Illustrating the issue with a DAG

- Many historical IV papers use the following strategy.
  - e.g., AJR use settlers mortality (z_i) to instrument for constraints on the government's executive (d_{it}), and then estimate the latter's effect on output growth (y_{it}).

$$
\begin{tikzcd}
z_i \arrow[r, bend left] & d_{it} \arrow[d] \\
 & y_{it}
\end{tikzcd}
$$

- z_i must affect y_{it} only through d_{it}.

Illustrating the issue with a DAG

- Many historical IV papers use the following strategy.
  - e.g., AJR use settlers mortality (z_i) to instrument for constraints on the government's executive (d_{it}), and then estimate the latter's effect on output growth (y_{it}).

$$
\begin{tikzcd}
z_i \arrow[r, bend left] & d_{it} \arrow[d] \\
 & y_{it}
\end{tikzcd}
$$

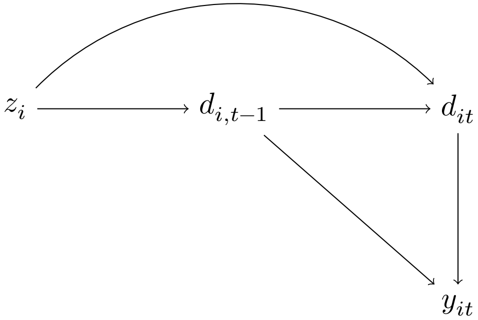
- z_i must affect y_{it} only through d_{it}.
- But note that z_i is time-invariant, whereas d_{it} is time-variant.

Illustrating the issue with a DAG

$$
\begin{tikzcd}
z_i \arrow[r] \arrow[rr, bend left] & d_{i,t-1} \arrow[r] \arrow[dr] & d_{it} \arrow[d] \\
 & & y_{it}
\end{tikzcd}
$$

- If z_i affects d_{it}, then it must also also affect $d_{i,t-1}$.

Illustrating the issue with a DAG



- If z_i affects d_{it}, then it must also also affect $d_{i,t-1}$.
- We can add $d_{i,t-2}$, $d_{i,t-3}$, and so forth.

Theoretical model
Second-stage equation

$$y_{it} = \beta_0 d_{it} + \beta_1 d_{i,t-1} + \epsilon_{y,it}.$$

Theoretical model
Second-stage equation

$$y_{it} = \beta_0 d_{it} + \beta_1 d_{i,t-1} + \epsilon_{y,it}.$$

- Note that this is a dynamic panel data model (DPDM).
- This is quite similar to an autoregressive distributed lag (ADL) setup.

Theoretical model

Second-stage equation

$$y_{it} = \beta_0 d_{it} + \beta_1 d_{i,t-1} + \epsilon_{y,it}.$$

- Note that this is a dynamic panel data model (DPDM).
- This is quite similar to an autoregressive distributed lag (ADL) setup.

First-stage equation

$$d_{it} = \delta z_i + \alpha_0 d_{i,t-1} + \epsilon_{d,it}.$$

Theoretical model

Second-stage equation

$$y_{it} = \beta_0 d_{it} + \beta_1 d_{i,t-1} + \epsilon_{y,it}.$$

- Note that this is a dynamic panel data model (DPDM).
- This is quite similar to an autoregressive distributed lag (ADL) setup.

First-stage equation

$$d_{it} = \delta z_i + \alpha_0 d_{i,t-1} + \epsilon_{d,it}.$$

What happens when we omit d_{t-1}?

- Obtain the particular solution of the first-stage equation:

$$d_{it} = (\delta \sum_{j=0}^{\infty} a_1^j) z_i + \underbrace{\sum_{j=0}^{\infty} a_1^j \epsilon_{i,t-j}}_{\text{Not iid!}}.$$

Let me prove it to you

- I simulated a panel with 50 units, each with 1000 observations (to show the misspecified model is inconsistent).

Let me prove it to you

- I simulated a panel with 50 units, each with 1000 observations (to show the misspecified model is inconsistent).

	True	TSLS
Intercept	0.00	0.00
		(0.01)
d_t	0.30	-0.10^{***}
		(0.02)
Ld_t	-0.40	
Num. obs.		49950

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Two-Stage Least Squares Results With Omitted Treatment Lag

Monte Carlo Results

- These results are consistently biased across samples. (Each unit has 100 observations.)

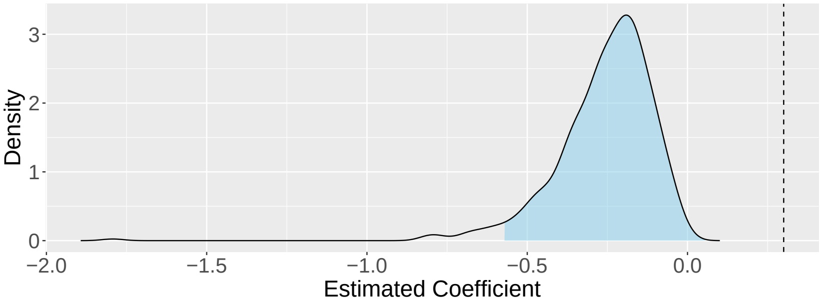


Figure 1: Monte Carlo Results, Omitted Treatment Lag (500 iterations; $\pm$ 2 SD shaded)

Solution

- Including Ld_{it} in both stages of the equation leads to a consistent estimator on all variables.

	True	TSLS	TSLS With Lag
Intercept	0.00	0.00	0.00
		(0.01)	(0.00)
d_t	0.30	-0.10^{***}	0.30^{***}
		(0.02)	(0.03)
Ld_t	-0.40		-0.40^{***}
			(0.01)
Num. obs.		49950	49950

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: Two-stage least squares results with treatment lag

Monte Carlo results

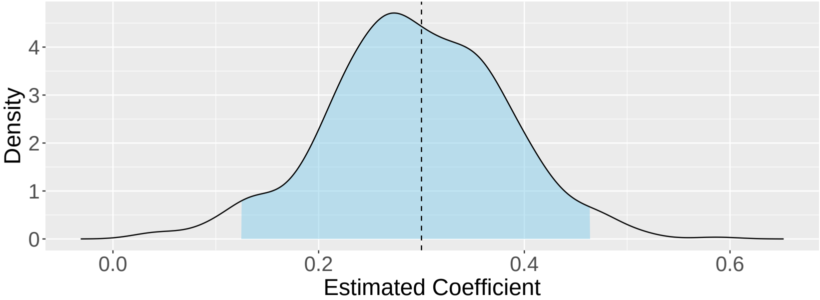


Figure 2: Monte Carlo results, with treatment lag (500 iterations; 50 units; 100 observations per unit; ± 2 SD shaded)

Generalize to bi-directional Granger causation

- What if y_{t-1} feeds into y_{it} *and* d_{it}?

$$
\begin{tikzcd}
z_i \arrow[r] \arrow[rr, bend left] & d_{i,t-1} \arrow[r] \arrow[d] \arrow[dr] & d_{it} \arrow[d] \\
& y_{i,t-1} \arrow[r] \arrow[ur] & y_{it}
\end{tikzcd}
$$

For example, Solow model

$$
\begin{aligned}
&(1)\ \Delta k = k_t - k_{t-1} = sy_{t-1} - \delta k_{t-1} \\
&\implies k_t = sy_{t-1} + (1-\delta)k_{t-1} \\
&(2)\ y_t = f(k_t)
\end{aligned}
$$

Estimation

We simulate then estimate the following equations:

$$
\begin{aligned}
y_{it} &= \beta d_{it} + \alpha_{11} y_{i,t-1} + \alpha_{12} d_{i,t-1} + \epsilon_{y,it} \\
d_{it} &= \alpha_{11} y_{i,t-1} + \alpha_{12} d_{i,t-1} + \delta z_i + \epsilon_{it}.
\end{aligned}
$$

Estimation

We simulate then estimate the following equations:

$$
\begin{aligned}
y_{it} &= \beta d_{it} + \alpha_{11} y_{i,t-1} + \alpha_{12} d_{i,t-1} + \epsilon_{y,it} \\
d_{it} &= \alpha_{11} y_{i,t-1} + \alpha_{12} d_{i,t-1} + \delta z_i + \epsilon_{it}.
\end{aligned}
$$

In VAR terms:

$$
\begin{bmatrix} 1 & -\beta \\ {\color{red}0} & 1 \end{bmatrix} \begin{bmatrix} y_{it} \\ d_{it} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} y_{i,t-1} \\ d_{i,t-1} \end{bmatrix} + \begin{bmatrix} {\color{red}0} \\ \delta \end{bmatrix} z_i + \begin{bmatrix} \epsilon_{y,it} \\ \epsilon_{d,it} \end{bmatrix}.
$$

Estimation

We simulate then estimate the following equations:

$$\begin{aligned} y_{it} &= \beta d_{it} + \alpha_{11} y_{i,t-1} + \alpha_{12} d_{i,t-1} + \epsilon_{y,it} \\ d_{it} &= \alpha_{11} y_{i,t-1} + \alpha_{12} d_{i,t-1} + \delta z_i + \epsilon_{it}. \end{aligned}$$

In VAR terms:

$$\begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{it} \\ d_{it} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} y_{i,t-1} \\ d_{i,t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \delta \end{bmatrix} z_i + \begin{bmatrix} \epsilon_{y,it} \\ \epsilon_{d,it} \end{bmatrix}.$$

- Note that lagged outcome enters the first-stage equation.
- Our procedure is analogous to a Cholesky decomposition.

Estimation

We simulate then estimate the following equations:

$$
\begin{aligned}
y_{it} &= \beta d_{it} + \alpha_{11} y_{i,t-1} + \alpha_{12} d_{i,t-1} + \epsilon_{y,it} \\
d_{it} &= \alpha_{11} y_{i,t-1} + \alpha_{12} d_{i,t-1} + \delta z_i + \epsilon_{it}.
\end{aligned}
$$

In VAR terms:

$$
\begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{it} \\ d_{it} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} y_{i,t-1} \\ d_{i,t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \delta \end{bmatrix} z_i + \begin{bmatrix} \epsilon_{y,it} \\ \epsilon_{d,it} \end{bmatrix}.
$$

- Note that lagged outcome enters the first-stage equation.
- Our procedure is analogous to a Cholesky decomposition.

What most historical IV papers are doing, in VAR terms

$$
\begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{it} \\ d_{it} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{i,t-1} \\ d_{i,t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \delta \end{bmatrix} z_i + \begin{bmatrix} \epsilon_{y,it} \\ \epsilon_{d,it} \end{bmatrix}.
$$

Results

	1st-Stage (d_t)			2nd-Stage (y_t)		
	True	(a)	(b)	True	(c)	(d)
Intercept	0.00	0.02	0.00	0.00	0.01	0.01
		(0.01)	(0.00)		(0.01)	(0.00)
z_t	0.20	0.29^{***}	0.20^{***}			
		(0.01)	(0.00)			
Ld_t	0.50		0.50^{***}	-0.40		-0.40^{***}
			(0.00)			(0.01)
Ly_t	0.70		0.70^{***}	0.60		0.60^{***}
			(0.00)			(0.02)
d_t				0.30	-0.24^{***}	0.30^{***}
					(0.03)	(0.03)
Num. obs.		49950	49950		49950	49950

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Two-stage least squares results, comparison

Monte Carlo results

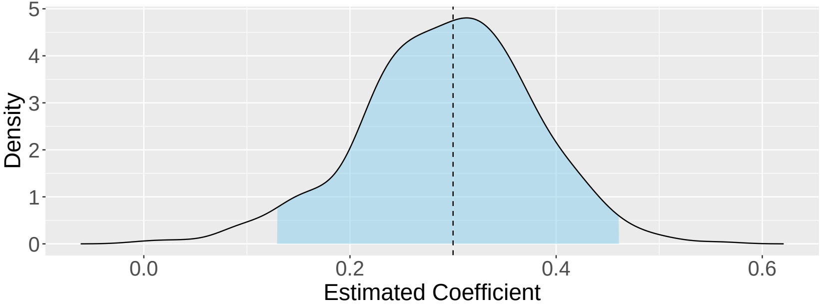


Figure 3: Monte Carlo results, with treatment and outcome lag (500 iterations; 50 units; 100 observations per unit; $\pm$ 2 SD shaded)

Monte Carlo results

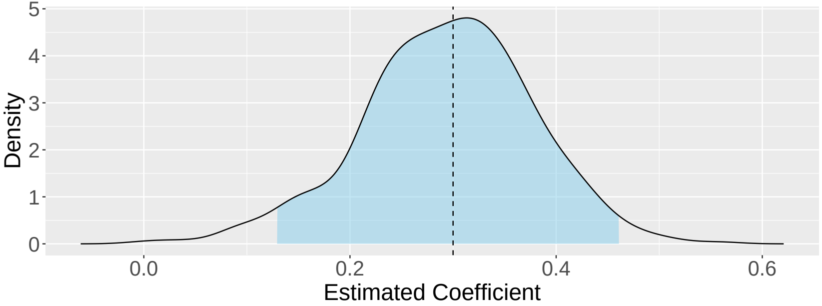


Figure 3: Monte Carlo results, with treatment and outcome lag (500 iterations; 50 units; 100 observations per unit; $\pm$ 2 SD shaded)

Note that

1. We did not need a second set of instruments.
2. Our instruments needn't be time-variant.
3. Once d_{it} is instrumented for, the coefficient on d_{t-1} is unbiased.

Results by sample size

(Work in progress...)

Discussion: is anyone talking about this?

- Most of the papers are in epidemiology (see Labrecque and Swanson 2018 in particular).
  - This is because they use genetic variants to predict disease's effect (e.g., smoking) on health outcome (e.g., life expectancy), i.e., Mendelian randomization.
  - Also, their theoretical derivations didn't hold up in my Monte Carlos.

Discussion: is anyone talking about this?

- Most of the papers are in epidemiology (see Labrecque and Swanson 2018 in particular).
  - This is because they use genetic variants to predict disease's effect (e.g., smoking) on health outcome (e.g., life expectancy), i.e., Mendelian randomization.
  - Also, their theoretical derivations didn't hold up in my Monte Carlos.
- There is one development econ paper which talks about this (Casey and Klemp 2021), but their solution is questionable.
  - They propose to estimate d_{it} and $d_{i,t-Q}$, and instrument for the latter with z_i, and use the resulting parameter to adjust the second-stage equation parameter.

Discussion: is anyone talking about this?

- Most of the papers are in epidemiology (see Labrecque and Swanson 2018 in particular).
  - This is because they use genetic variants to predict disease's effect (e.g., smoking) on health outcome (e.g., life expectancy), i.e., Mendelian randomization.
  - Also, their theoretical derivations didn't hold up in my Monte Carlos.
- There is one development econ paper which talks about this (Casey and Klemp 2021), but their solution is questionable.
  - They propose to estimate d_{it} and $d_{i,t-Q}$, and instrument for the latter with z_i, and use the resulting parameter to adjust the second-stage equation parameter.
  - The lag length is arbitrary. And even if it works, their method only recovers a "long-run" parameter, not the instantaneous parameter that is of policy interest.

Discussion: is anyone talking about this?

- Most of the papers are in epidemiology (see Labrecque and Swanson 2018 in particular).
  - This is because they use genetic variants to predict disease's effect (e.g., smoking) on health outcome (e.g., life expectancy), i.e., Mendelian randomization.
  - Also, their theoretical derivations didn't hold up in my Monte Carlos.
- There is one development econ paper which talks about this (Casey and Klemp 2021), but their solution is questionable.
  - They propose to estimate d_{it} and $d_{i,t-Q}$, and instrument for the latter with z_i, and use the resulting parameter to adjust the second-stage equation parameter.
  - The lag length is arbitrary. And even if it works, their method only recovers a "long-run" parameter, not the instantaneous parameter that is of policy interest.
- There are actually a lot of time series tools that can help analyze and solve the problem. But the Anderson-Rubin causal inference people don't talk to the time series people or something?

Discussion: how important is this result?

1. This does not require a time-variant treatment (unlike proxy SVARs, aka SVARs with external instruments).

Discussion: how important is this result?

1. This does not require a time-variant treatment (unlike proxy SVARs, aka SVARs with external instruments).
   - In lieu, this setup reverts back to a Cholesky identification (but this is not an issue, and in fact taken as given in the causal inference setting).

Discussion: how important is this result?

1. This does not require a time-variant treatment (unlike proxy SVARs, aka SVARs with external instruments).
   - In lieu, this setup reverts back to a Cholesky identification (but this is not an issue, and in fact taken as given in the causal inference setting).
2. The development economics literature has been misspecifying a lot of IV papers.

Discussion: how important is this result?

1. This does not require a time-variant treatment (unlike proxy SVARs, aka SVARs with external instruments).
   - In lieu, this setup reverts back to a Cholesky identification (but this is not an issue, and in fact taken as given in the causal inference setting).
2. The development economics literature has been misspecifying a lot of IV papers.
3. A contribution to the causal revolution-revolution, e.g., issues with TWFE DID estimator (Callaway & Sant'Anna 2020), geographical IVs (Mellon 2022).