The Problem With Historical Instrumental Variables

John T.H. Wong

I argue in this essay that many papers that use historical (i.e., time-invariant) are misspecified. This is because if the instrument influences the treatment, it must also influence the lags of treatment. This paper makes two contributions. First, using Monte Carlo methods, I prove that two-stage least squares (2SLS) estimates are not only biased and but also inconsistent. Second, I show that a consistent 2SLS estimator of the *instantaneous* treatment effect can be obtained by simply including lagged treatment *and*/or lagged outcome as control variables (a.k.a. exogenous regressors). The simplicity stands in contrast to [bib]'s nonlinear two-parameter estimator, which requires discretion in choosing the initial shock year and requires the delta method for standard error estimates, only to recover a long-run treatment effect. Unlike proxy SVARs (see [bib] for a detailed explainer), the proposed setup also does not require a time-variant instrument—which is not available when attempting to identify a treatment over long periods of time. However, the proposed setup does require a somewhat long panel to obtain unbiased estimates.

This essay is structured as follows. **?@sec-adl** demonstrates, in the simple case of a treatment that enters the outcome equation as an AR(1) sequence, how the 2SLS estimator is inconsistent and the proposed correction. **?@sec-var** extends the analysis to a case of bivariate Granger causality. **?@sec-consistency** uses Monte Carlo to infer the necessary panel length and width under which the correction is robust. **?@sec-conclude** concludes.

1 The AR(1) Treatment Case

Figure 1 illustrates how most instrumental variable papers violate exclusion restriction. The exclusion restriction requires that z_i affect y_{it} only through d_{it}. Take [bib] for example, who use settlers mortality (z_i) in former colonies to identify the effect of expropriation risk in ($d_{it}, t = [1985, 1995]$) on log output per capita in ($y_{it}, t = 1995$). For a thorough list of papers that use a similar design, see [bib].

The problem with this identification strategy is that if z_i is correlated with d_{it}, then it should also be correlated $d_{i,t-1}$, or $d_{i,t-j}$ for up till the j lags of d_{it} that enter into the function for y_{it}.

Let us stipulate that only the first lag of the treatment determines the outcome. More formally, suppose y_{it} and d_{it} are respectively determined by the following equations:

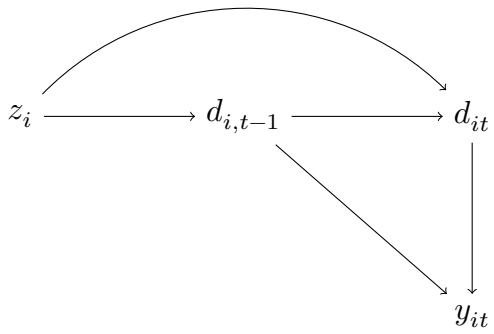

Figure 1: Directed acyclic graph: lagged treatment determines treatment and outcome.

$$y_{it} = \beta_0 d_{it} + \beta_1 d_{i,t-1} + \epsilon_{y,it}$$
$$d_{it} = \delta z_i + \alpha_0 d_{i,t-1} + \epsilon_{d,it},$$

where $\epsilon_{y,it}$ and $\epsilon_{d,it}$ are iid. I set intercepts in both equations to zero without loss of generality. I assume both equations are stationary.

The consequence of omitting $d_{i,t-1}$ can be analyzed by solving the d_{it} by iteration, which yields:

$$d_{it} = (\delta \sum_{j=0}^{\infty} a_1^j) z_i + \sum_{j=0}^{\infty} a_1^j \epsilon_{i,t-j}.$$

There are two potential misspecifications here. First, the parameter on z_i will be biased by $\delta \sum_{j=0}^{\infty} a_1^j - \delta = \delta \sum_{j=1}^{\infty} a_1^j$, though this is perhaps not an issue as the first-stage results are not necessarily of interest to researchers. The second issue is that error term potentially violates white noise assumptions required for consistent estimates. If each unit i is only observed for one period, then $d_{it} \to d_i$, and the term is still independent across units. However, if a panel of d_{it} is used, then the error of more recent observations will be a sum of past errors, violating the assumption of white noise.

Perhaps more concerining is that our estimator of β_0 in the second-stage equation will be biased. For this I can analyze the 2SLS estimator:

$$\beta_{\text{2SLS}} = \frac{\text{Cov}(y_{it}, z_i)}{\text{Cov}(d_{it}, z_i)} = \beta_0 + \beta_1 \frac{\text{Cov}(d_{i,t-1}, z_i)}{\text{Cov}(d_{it}, z_i)}.$$

Notice that if there are a long number of periods between t and whenever z_i is determined, $\text{Cov}(d_{it}, z_i) = \text{Cov}(d_{i,t-1}, z_i) = \delta/(1 - \alpha_1)$. In this case, the estimator yields the biased result:

$$\beta_{\text{2SLS}} = \beta_0 + \beta_1.$$

I can further demonstrate this point by simulating a panel of observations that are determined by the true relationships. The simulated panel has 50 units, each with 1000 periods. Column 1 of Table 1 shows the true parameters. When estimated with a naive 2SLS setup, the parameters are biased (Column 2).

	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 1: Two-stage least squares, second-stage results

I then use Monte Carlo methods to repeat the simulation for 500 iterations. Each iteration still has 50 units, but now with only 100 periods each. Figure 2 shows that the estimated coefficients are consistently biased from the true value, in the direction our analysis predicts.

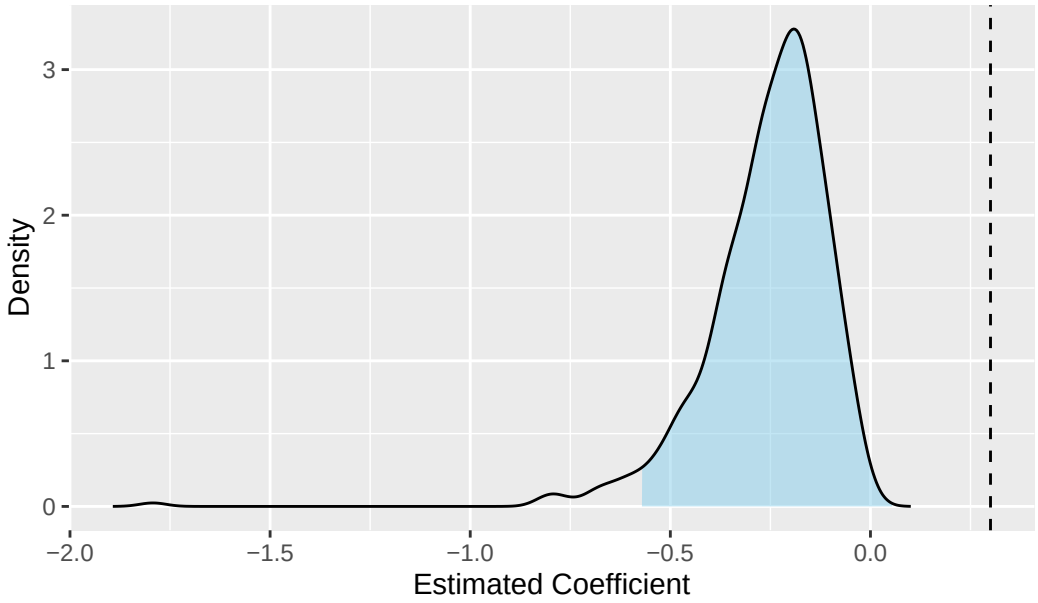

Figure 2: Monte Carlo Results, omitted treatment lag (500 iterations; 50 units; 100 observations per unit; ± 2 SD shaded; black dotted line indicates true mean)

I propose a simple correction method: we include d_{t-1} as a control variable. The results are indicated in Table 1, Column 3. Note that this setup is able to recover the true parameters, even though I did not need to use an additional instrument to identify the lagged treatment.

Using Monte Carlo, I can show that this specification robustly recovers the true parameter value of β_0 across random samples (Figure 3).

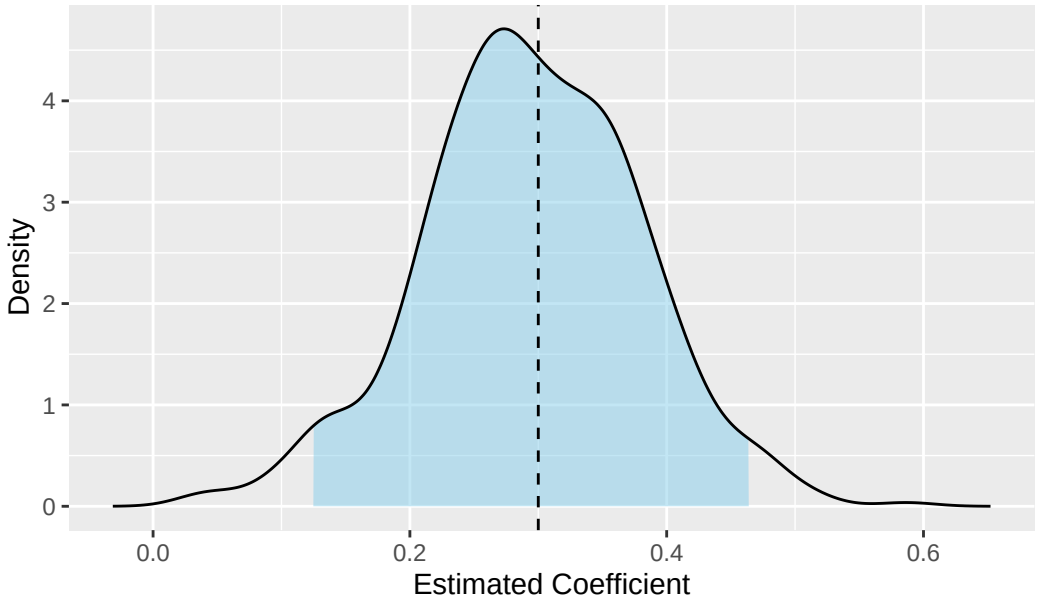

Figure 3: Monte Carlo Results, with treatment lag (500 iterations; 50 units; 100 observations per unit; ± 2 SD shaded; black dotted line indicates true mean)

2 Adding Bi-Directional Granger Causality

Most dynamic systems present an additional challenge: the contemporaneous treatment is determined by the lags of the outcome. Figure 4 builds on Figure 1, with the additional relationships illustrated with dotted lines. Notice that $y_{i,t-1}$ feeds into both d_{it} and y_{it}. Note that I posit only unidirectional contemporaneous causality: d_{it} affects y_{it} (and $d_{i,t-1}$ affects $y_{i,t-1}$), but not the other way around. This type of treatment is found in the growth literature for example; capital in the Solow model depends on past capital and output, but not contemperaneous output.

We can more compact represent the implied sets of equations with a VAR system (even though we do not estimate it as one):

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

Several features are immediately apparent:

1. The unidirectional causality from d_{it} to y_{it} in Figure 4 is analogous to a Cholesky decomposition that stipulates the treatment as the more exogenous variable.
2. The omission of z_i in the second-stage is analogous to forcing the coefficient of the instrument in the first row to be zero in a VAR system.

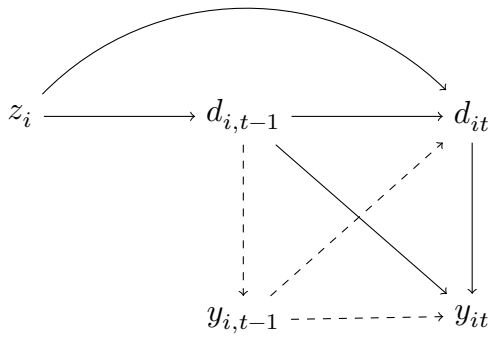

Figure 4: Directed acyclic graph: lagged treatment determines lagged outcome and present treatment and outcome. Lagged outcome determines present treatment and outcome.

3. The α_{ij} parameters allow for the estimation of Granger causality. Historical IV papers that omit the A matrix are essentially imposing $\alpha_{ij} = 0 \ \forall \ i, j$, which is arguably an onerous set of restrictions.

Again, I simulate a panel of observations, and then attempt to recover the true parameters. Table 2 displays the results. The left-most sub-columns show the true parameters. Columns (a) and (c) show results from a naive 2SLS procedure, whereas Columns (b) and (d) are results from a 2SLS procedure where lagged treatment and outcome are included in the second-stage *and the first-stage* equations.

	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 2: Two-stage least squares results, comparison

As before, the naive estimates are significantly biased, whereas the proposed specification can recover the true estimates. These results are robust in Monte Carlo (Figure 5).

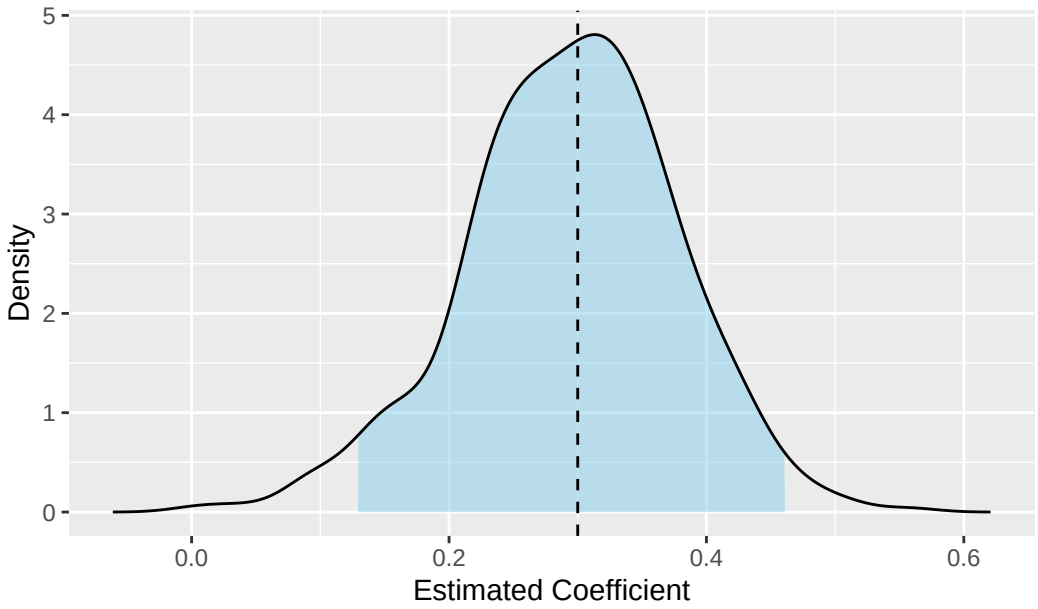

Figure 5: Monte Carlo results, with treatment and outcome lag (500 iterations; 50 units; 100 observations per unit; ± 2 SD shaded)

3 References