

1. Introduction

Using the lag of an endogenous treatment as an instrument for its current value is attractive because it requires no external instrument and persistence often generates a strong first stage. Neither fact establishes that the resulting instrumental-variable (IV) coefficient identifies the contemporaneous causal effect assigned to it. The relevant question is target-specific: does the population coefficient selected by the own-lag IV procedure equal the contemporaneous direct causal effect of treatment on the outcome?

This question matters in applied work. Own-lag instruments have been used for policy wages and test scores (Rutledge et al. 2026; Delaney and Devereux 2025), while recent dynamic-panel work uses variables' own lagged levels and differences as internal instruments under restrictions on feedback and serial dependence (Djankov et al. 2025). Such uses need not be invalid; they show why identification must rest on the maintained dynamic structure rather than temporal ordering alone.

The concern is longstanding. Fair (1970) examined the estimation of simultaneous-equation models with lagged endogenous variables and serially correlated disturbances, emphasizing the restrictions required for consistent estimation. Angrist and Krueger (2001) subsequently cautioned against lag instruments whose validity rests on difficult-to-assess dynamic assumptions. Reed (2015) distinguishes replacing an endogenous regressor with its lag from using that lag as an instrument, noting that the latter requires both relevance and exclusion from the outcome equation. Bellemare, Masaki, and Pepinsky (2017) show that lagging a regressor does not generally remove endogeneity. Wang and Bellemare (2020) compare the bias and root mean squared error (RMSE) of own-lag IV and naive OLS under particular dynamic confounding processes. The present note instead asks when the population own-lag IV estimand equals a prespecified contemporaneous direct causal effect and provides a graphical characterization of the dynamic pathways that prevent equivalence. Tsao et al. (2026) study a related estimand–target mismatch arising instead from aggregation.

This note makes three contributions relative to this literature. It does not propose a new IV orthogonality or general graphical criterion. Instead, it makes the estimand–target distinction explicit for a contemporaneous direct effect, specializes established graphical IV logic to dynamic own-lag pathways, and derives a closed-form estimand–target gap for the stationary persistent-confounding specification of Wang and Bellemare (2020), showing why more information can sharpen inference around a mismatched estimand.

Lag instruments are central to dynamic-panel econometrics (Arellano and Bond 1991; Baltagi 2026), where their validity follows from maintained moment restrictions for specified equations. While dynamic-panel GMM relies on model-specific moment restrictions to identify parameters of dynamic equations, the present note asks whether a particular own-lag IV estimand coincides with a prespecified contemporaneous direct causal effect.

2. Target and population own-lag IV estimand

Let units be indexed by i and time by t . Current treatment is D_{it} , current outcome is Y_{it} , the maintained conditioning set is W_{it} , and the candidate instrument is

$$Z_{it} = D_{it-1}. \quad (1)$$

The specified structural target is the contemporaneous direct causal effect β in

$$Y_{it} = \beta D_{it} + V_{it}, \quad (2)$$

where V_{it} collects determinants of current outcome outside the channel $D_{it} \rightarrow Y_{it}$. Under the maintained additive linear structural model, β is the contemporaneous structural direct effect associated with the treatment-to-outcome edge: the effect of an intervention that changes D_{it} by one unit while holding fixed the other arguments of the outcome equation. Because the effect is constant, its unit-level and population-average values coincide. The linear benchmark is appropriate because it corresponds directly to the population 2SLS interpretation used in the motivating applications; heterogeneous or nonlinear effects would require redefining both the target and the estimator-specific estimand.

For any scalar A_{it} , let $\tilde{A}_{it} = A_{it} - \Pi_A' W_{it}$ denote its residual from the population linear projection on W_{it} , including an intercept. The just-identified population IV coefficient solves

$$E[\tilde{Z}_{it}(\tilde{Y}_{it} - b\tilde{D}_{it})] = 0. \quad (3)$$

Under relevance, $E(\tilde{Z}_{it}\tilde{D}_{it}) \neq 0$, its solution is

$$\beta_{IV}^* = \frac{E(\tilde{Z}_{it}\tilde{Y}_{it})}{E(\tilde{Z}_{it}\tilde{D}_{it})} = \beta + \frac{E(\tilde{Z}_{it}\tilde{V}_{it})}{E(\tilde{Z}_{it}\tilde{D}_{it})}. \quad (4)$$

Proposition 1 (estimand-target equivalence). Under finite second moments and relevance, the population own-lag IV estimand identifies the contemporaneous direct effect β if and only if

$$E(\tilde{Z}_{it}\tilde{V}_{it}) = 0. \quad (5)$$

Proof. Equation (4) follows by residualizing (2), substituting into (3), and solving. Under relevance, $\beta_{IV}^* = \beta$ exactly when its numerator adjustment is zero. ▀

Equation (5) is elementary IV algebra, but it disciplines the causal claim. Relevance determines whether the ratio exists; it does not eliminate a path from past treatment to the target-specific residual. For example, if $V_{it} = \xi D_{it-1} + R_{it}$, then

$$\beta_{IV}^* = \beta + \xi \frac{\text{Var}(\tilde{Z}_{it})}{\text{Cov}(\tilde{Z}_{it}, \tilde{D}_{it})} + \frac{\text{Cov}(\tilde{Z}_{it}, \tilde{R}_{it})}{\text{Cov}(\tilde{Z}_{it}, \tilde{D}_{it})}. \quad (6)$$

First-stage relevance does not eliminate the carryover component: unless it is offset by the remaining residual covariance, the own-lag IV estimand differs from the contemporaneous direct causal effect.

3. Graphical content of the moment

Represent the system by a directed acyclic graph (DAG). To assess β , delete $D_{it} \rightarrow Y_{it}$ before latent projection, leaving other equations unchanged. In the resulting acyclic directed mixed graph (ADMG), m-separation is equivalent to d-separation in the underlying DAG for observed variables

(Richardson and Spirtes 2002; Pearl 2009). This applies Pearl's (2012, Theorem 2) graphical IV logic to own-lag instruments.

Proposition 2 (dynamic specialization of the graphical IV criterion). Suppose the maintained causal system has finite second moments, is linear and Markov with respect to its underlying DAG, the conditioning set contains only nondescendants of the current outcome, and the conditional expectations used in residualization are linear. If Z_{it} and Y_{it} are m-separated given W_{it} in the latent projection of the DAG after deleting $D_{it} \rightarrow Y_{it}$, then (5) holds.

Proof. In the edge-deleted system, m-separation and the global Markov property imply conditional independence of the lag and edge-deleted outcome (the structural residual) given W_{it} . The lag predates the current outcome, and the controls are restricted to its nondescendants; consequently, neither changes when the target edge is deleted. Under linear conditional expectations, FWL residuals are conditional-mean residuals; iterated expectations then imply (5). Relevance remains the separate denominator condition in Proposition 1. ▀

This condition is sufficient, not necessary for numerical orthogonality: open paths may cancel at exceptional parameter values. It nevertheless distinguishes structural mechanisms. In the baseline path

$$D_{it-1} \rightarrow D_{it} \leftrightarrow Y_{it}, \quad (7)$$

D_{it} is a collider after the target edge is removed. Without conditioning on current treatment, the path is blocked. Contemporaneous endogeneity of D_{it} therefore need not invalidate the lag.

By contrast, each of the following can remain open after target-edge deletion:

- direct carryover, $D_{it-1} \rightarrow Y_{it}$;
- outcome dynamics, $D_{it-1} \rightarrow Y_{it-1} \rightarrow Y_{it}$;
- persistent latent confounding, represented after projection by $D_{it-1} \leftrightarrow Y_{it}$;
- persistent omitted aggregate shocks that induce the same observed-variable bidirected edge.

The ADMG records the identifying dependence, not its economic origin. Substantive knowledge is required to decide which underlying mechanism is plausible. It is also misleading to regress Y_{it} on D_{it-1} while controlling for D_{it} and interpret the lag coefficient as an exclusion test: conditioning on the collider in (7) can create association even when (5) holds.

4. Persistent confounding: a stationary solution

Wang and Bellemare's (2020) first scenario isolates persistent confounding without a direct effect of lagged treatment on the current outcome. In our notation, consider the stationary system

$$Y_{it} = \beta D_{it} + \delta U_{it} + \varepsilon_{it}, \quad (8)$$

$$D_{it} = \rho D_{it-1} + \kappa U_{it} + \eta_{it}, \quad (9)$$

$$U_{it} = \phi U_{it-1} + v_{it}, \quad (10)$$

where the innovations are mean zero, mutually independent across equations and time, and have finite second moments, $|\rho| < 1$, $|\phi| < 1$, and $\sigma_U^2 = \sigma_v^2/(1 - \phi^2)$. Define $c_0 = \text{Cov}(D_{it}, U_{it})$ and $c_1 = \text{Cov}(D_{it-1}, U_{it})$. Stationarity gives

$$c_0 = \frac{\kappa\sigma_U^2}{1 - \rho\phi}, \quad c_1 = \phi c_0 = \frac{\phi\kappa\sigma_U^2}{1 - \rho\phi}. \quad (11)$$

Let $q = \text{Var}(D_{it})$ and $s = \text{Cov}(D_{it-1}, D_{it}) = \rho q + \kappa c_1$. For $s \neq 0$, substitution into the IV ratio yields

$$\beta_{IV}^* = \beta + \delta \frac{\phi\kappa\sigma_U^2}{(1 - \rho\phi)\rho q + \phi\kappa^2\sigma_U^2}. \quad (12)$$

Proposition 3 (persistent-confounding mismatch). If $s \neq 0$ and $\delta\phi\kappa \neq 0$, and the latent confounder has strictly positive variance, the own-lag IV estimand in (12) differs from β . No direct effect of D_{it-1} on Y_{it} or U_{it} is required.

The target-edge-deleted path is $D_{it-1} \leftarrow U_{it-1} \rightarrow U_{it} \rightarrow Y_{it}$. Persistence carries contemporaneous confounding forward by one period. Under stationarity and ergodicity, sample moments converge to their population counterparts, so the own-lag IV estimator converges to β_{IV}^* , provided the denominator is nonzero. Increasing T does not force c_1 to zero or eliminate the estimand–target gap. Mean reversion produces stable population moments; it does not restore target identification.

5. Monte Carlo illustration

The simulations separate convergence to an estimand from identification of the target. Across three designs, $\beta = 1$, $N = 100$, a 100-period burn-in is discarded, and 400 replications are generated. Design A contains contemporaneous treatment endogeneity only and satisfies (5). Design B adds outcome carryover, generating a population IV estimand of 1.4934. Design C implements (8)-(10), generating an estimand of 1.4509. Table I reports the benchmark at $T = 50$; Figure 1 shows results over $T \in \{10, 25, 50, 100, 200\}$.

The Monte Carlo means closely track the analytical estimands. In the invalid designs, RMSE relative to the estimand approaches zero as T grows, while RMSE relative to the causal target remains near the population mismatch. More observations can therefore produce a more persuasive estimate of the wrong causal quantity.

6. Implications

An own-lag IV design should be defended in the following order. First, define the causal effect being claimed. Second, write the corresponding target-specific residual. Third, list plausible dynamic channels linking past treatment to current outcome. Fourth, delete the target edge and inspect whether any bypassing path remains open given the maintained conditioning set. Fifth, translate that graphical assessment into the estimator’s population moment. Sixth, distinguish first-stage strength and sampling precision from equality between estimand and target.

This reasoning does not invalidate dynamic-panel GMM. Arellano-Bond instruments are justified by explicit restrictions such as $E(Y_{it-s}\Delta\varepsilon_{it}) = 0$ for specified lag depths. They are valid because those restrictions follow from a maintained dynamic model, not merely because the variables are lagged. Dynamic-panel GMM asks a different question: whether parameters of a maintained dynamic equation are identified by transformation- and lag-specific moments. Here the question is whether the particular coefficient selected by an own lag equals the stated contemporaneous target. The two questions can overlap, but they are not interchangeable.

Under relevance, an own lag identifies the contemporaneous direct causal effect only when the target-specific structural residual is orthogonal to the instrument after residualization on the maintained controls; carryover or persistent dependence can break that equivalence. Lagged instruments should therefore be neither accepted nor rejected categorically. The contribution is to characterize this target-specific dynamic moment, not to propose a new general IV criterion.

References

- Angrist, J.D. and A.B. Krueger (2001) “Instrumental variables and the search for identification: From supply and demand to natural experiments” *Journal of Economic Perspectives* 15, 69–85.
- Arellano, M. and S. Bond (1991) “Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations” *Review of Economic Studies* 58, 277–297.
- Baltagi, B.H. (2026) “A historically structured survey of panel data econometrics” *Econometrics and Statistics*, advance online publication. <https://doi.org/10.1016/j.ecosta.2026.08.004>.
- Bellemare, M.F., T. Masaki and T.B. Pepinsky (2017) “Lagged explanatory variables and the estimation of causal effect” *Journal of Politics* 79, 949–963.
- Delaney, J.M. and P.J. Devereux (2025) “Teacher bias and evaluation differences in test scores: Different methods for different questions” *Oxford Bulletin of Economics and Statistics* 87, 924–941.
- Djankov, S., A. Melcarne, G.B. Ramello and R. Spruk (2025) “Timely justice as a determinant of economic growth” *Journal of Economic Behavior and Organization* 238, 107235.
- Fair, R.C. (1970) “The estimation of simultaneous equation models with lagged endogenous variables and first order serially correlated errors” *Econometrica* 38, 507–516.
- Pearl, J. (2012) “The causal foundations of structural equation modeling” in *Handbook of Structural Equation Modeling*, R.H. Hoyle, ed., Guilford Press: New York, 68–91.
- Pearl, J. (2009) *Causality: Models, Reasoning, and Inference*, 2nd ed., Cambridge University Press: Cambridge.
- Reed, W.R. (2015) “On the practice of lagging variables to avoid simultaneity” *Oxford Bulletin of Economics and Statistics* 77, 897–905.
- Richardson, T.S. and P. Spirtes (2002) “Ancestral graph Markov models” *Annals of Statistics* 30, 962–1030.
- Rutledge, Z., M. Castillo, T.J. Richards and P. Martin (2026) “H-2A adverse effect wage rates and U.S. farm wages” *American Journal of Agricultural Economics* 108, 494–517.
- Tsao, D., K. Muandet, F. Eberhardt and E. Perković (2026) “Lost in aggregation: The causal interpretation of the IV estimand” arXiv:2601.12120, version 2.
- Wang, Y. and M.F. Bellemare (2020) “Lagged variables as instruments” Working paper, University of Minnesota.

Table I. Monte Carlo benchmark at $N = 100$ and $T = 50$

Design	Target β	Population IV β_{IV}^*	MC mean	95% Monte Carlo interval	RMSE: target	RMSE: estimand	Mean first stage
A: current endogeneity only	1.0000	1.0000	0.9997	[0.9738, 1.0274]	0.0143	0.0143	0.7000
B: dynamic outcome carryover	1.0000	1.4934	1.4944	[1.4684, 1.5214]	0.4946	0.0139	0.8002
C: persistent confounding	1.0000	1.4509	1.4509	[1.4011, 1.4984]	0.4516	0.0247	0.6450

Notes: 400 replications. The interval is the 2.5th-97.5th percentile range of the simulated IV coefficient. The first-stage entry is the mean coefficient from the regression of D_{it} on D_{it-1} . Design A sets treatment persistence to 0.70 and the contemporaneous innovation correlation to 0.60. Design B sets treatment persistence to 0.80 and outcome carryover to 0.30. Design C sets treatment persistence to 0.50, confounder persistence to 0.60, and the confounder coefficients in the treatment and outcome equations to 0.50 and 1.00, respectively.

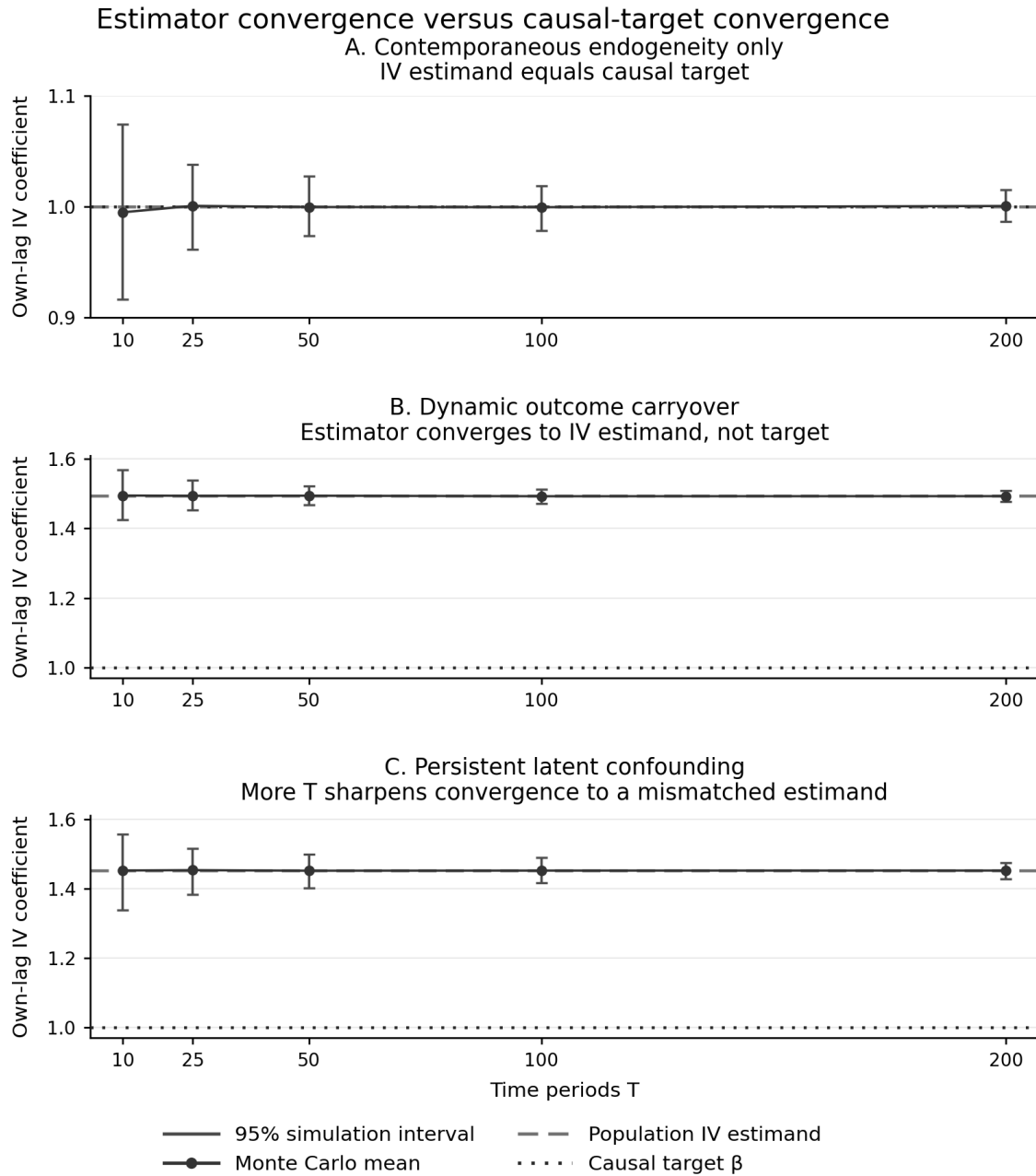


Figure 1. Estimator convergence versus causal-target convergence. Monte Carlo means and 95 percent simulation intervals as T increases with $N = 100$ fixed. In Design A the IV estimand coincides with the target. In Designs B and C, increasing T concentrates the estimator around β_{IV}^* rather than β . Vertical scales differ across panels.