Lagged Explanatory Variables and the Estimation of Causal Effects

Bellemare, Marc F. and Masaki, Takaaki and Pepinsky, Thomas B.

University of Minnesota, Cornell University, Cornell University

23 February 2015

Lagged Explanatory Variables and the Estimation of Causal Effects*

Marc F. Bellemare† Takaaki Masaki‡ Thomas B. Pepinsky§

February 23, 2015

Abstract

Across the social sciences, lagged explanatory variables are a common strategy to confront challenges to causal identification using observational data. We show that "lag identification"—the use of lagged explanatory variables to solve endogeneity problems—is an illusion: lagging independent variables merely moves the channel through which endogeneity biases causal estimates, replacing a "selection on observables" assumption with an equally untestable "no dynamics among unobservables" assumption. We build our argument intuitively using directed acyclic graphs, then provide analytical results on the bias resulting from lag identification in a simple linear regression framework. We then present simulation results that characterize how, even under favorable conditions, lag identification leads to incorrect inferences. These findings have important implications for current practice among applied researchers in political science, economics, and related disciplines. We conclude by specifying the conditions under which lagged explanatory variables are appropriate for identifying causal effects.

Keywords: Causal Identification, Treatment Effects, Lagged Variables

JEL Classification Codes: C13, C15, C21

*Preliminary draft. We thank Metin Çakir, Bryce Corrigan, Peter Enns, Paul Glewwe, Yu Na Lee, Andrew Little, and Joe Ritter for helpful discussions. All remaining errors are ours.

†Assistant Professor, Department of Applied Economics, University of Minnesota, Saint Paul, MN 55108, mbellema@umn.edu.

‡Ph.D. Candidate, Department of Government, Cornell University, Ithaca, NY 14853, tm427@cornell.edu.

§Associate Professor, Department of Government, Cornell University, Ithaca, NY 14853, pepinsky@cornell.edu.

1 Introduction

It is common for researchers using observational data in the applied social sciences to lag explanatory variables in an effort to purge their estimates of endogeneity, i.e., to eliminate the correlation between the explanatory variables and the error term, a problem which prevents teasing out causal relationships from mere correlations. The lagged independent variable strategy—what we refer to as "lag identification" throughout this paper—is attractive because it purports to alleviate threats to causal identification without requiring any other data than that available in the dataset. This approach, however, is grounded in a pre-Credibility Revolution understanding of the problem of endogeneity (cf. Angrist and Pischke 2009, 2010, and 2014), one that is rooted in the work of the Cowles Commission on simultaneous equations in the middle of the 20th century (Christ, 1994).

In this paper we demonstrate that lag identification is almost never a solution to endogeneity problems in observational data, and that rather than allowing for the identification of causal relationships, lag identification merely moves the channel through which endogeneity biases estimates of causal effects. Specifically, we characterize precisely the conditions under which lagging an explanatory variable can achieve causal identification: these are (i) serial correlation in the potentially endogenous explanatory variable, and (ii) no serial correlation among the unobserved sources of endogeneity. This replaces the selection on observables assumption that motivates the regression with a new identification assumption of "no dynamics among unobservables." This assumption is intuitively problematic, because it requires substantive restrictions on the properties of a variable that is not observed. Put differently, lagging an explanatory variable to obtain an estimate of a causal parameter assumes the existence of temporal dynamics in the explanatory variable, but the same temporal dynamics must not characterize the unobservables. Our main conclusion is that the use of lagged explanatory variables is almost never justified on identification grounds, and so it does not buy causal identification on the cheap. The central identification assumption has simply been moved to a different point in the data generat-

ing process, and that new identification assumption is unlikely to be more defensible than the selection on observables assumption that motivates the regression.

This argument is most closely related to concurrent research by Reed (2014), who also studies the use of lagged explanatory variables for causal inference but focuses on simultaneity bias and proposes the use of lagged explanatory variables as instruments for endogenous explanatory variables.[1] In contrast, our work focuses on more general forms of endogeneity, and our results imply that Reed's recommendations are unlikely to represent a valid solution to the identification problem. Our work is also related to Blackwell and Glynn (2014), who are broadly concerned with establishing theoretical results about causal inference using time-series cross-sectional (i.e., large-T and large-N panel) data. All of our arguments are consistent with theirs. Our contribution is more focused, and designed to identify a specific practice in applied social science research whose consequences are not properly understood.

This paper is motivated by the same concern for credible statistical techniques for the estimation of causal effects that has motivated recent advances in randomized controlled trials (Duflo et al. 2007, Glennerster and Takavarasha 2013), field experiments (Harrison and List 2004, Gerber and Green 2012), instrumental variables (Angrist et al. 1996, Sovey and Green 2011, Imbens 2014), regression discontinuity (Imbens and Lemieux 2008), and differences-in-differences estimation (Bertrand et al. 2004) in the social sciences. The common theme uniting this literature is the critical importance of research design; in the words of Sekhon (2009), "without an experiment, a natural experiment, a discontinuity, or some other strong design, no amount of econometric or statistical modeling can make the move from correlation to causation persuasive." Lag identification is a response to an imperfect research design that relies on a simple—much too simple, it turns out—statistical fix to strengthen the argument that correlations are causal. Our results demonstrate that this fix does not work.

[1] Villas-Boas and Winer (1999), for example, use lagged prices as instruments for endogenous contemporaneous prices.

The rest of this paper is organized as follows. In section 2, we discuss the general problem posed by the use of lagged variables as regressors using directed acyclic graphs (Pearl 2009), and present an overview of recent articles in the top economics and political science journals which rely on lagged explanatory variables as a source of exogenous variation. Section 3 derives analytical results for the biases of lag identification in the context of an ordinary least squares (OLS) regression, providing a formal result for the "no dynamics among unobservables" condition that allows for conservative estimates of causal effects using lagged explanatory variables in the presence of endogeneity. Section 4 presents Monte Carlo results showing that the use of lagged explanatory variables can worsen the identification problem, with consequences for inference that are often worse than simply ignoring endogeneity. Section 5 concludes by summarizing our argument, offering recommendations for applied work and for future research, and outlining a set of guidelines for researchers to follow when using lagged explanatory variables to identify causal effects.

2 Problem Definition

There are three reasons why a lagged value of an independent variable might appear on the right hand side of a regression.

1. *Theoretical*: In some contexts, there are clear theoretical reason to expect that the effect of an explanatory variable only operates with a one-period lag. Such is the case, for example, when estimating Euler equations in order to study intertemporal substitution behaviors, or when considering the efficient market hypothesis in its random walk version, wherein p_t, the price of an asset today, is a function $p_t = p_{t-1} + e_t$ of the price of the same asset yesterday, p_{t-1}, and an iid error term e_t. It could also be the case that the analysis is directly interested in lagged effects conditional on contemporaneous effects, in which both current and lagged values of the independent variable would appear on the right hand side of a regression.

2. *Statistical*: In other contexts, lagged independent variables serve a statistical function. Examples include dynamic panel data analysis (Arellano and Bond 1991) as well as distributed lag, error correction, and related families of dynamic statistical models (see De Boef and Keele 2008).

3. *Causal Identification*: Frequently, applied researchers propose to use a lagged value of an explanatory variable X in order to "exogenize" it when estimating the causal effect of X on Y. Since Y_t cannot possibly cause X_{t-1}, the argument goes, replacing X_t with X_{t-1} obviates concerns that X is endogenous to Y.

Our focus in this paper is on the use of lagged explanatory variables for causal identification, i.e., the use of such variables to mitigate inferential threats that arise from endogeneity. None of our critiques of lag identification apply to theoretical or statistical motivations for including lagged values of independent variables on the right hand side of a regression, although we will touch briefly on both of these in our Monte Carlo analysis in section 4.

How common is the practice of lagging explanatory variables for identification purposes? To answer this question, we examined all articles published in the top general journals in political science, economics, and sociology, as well as several top journals in the political science subfields of comparative politics and international relations (see Table 1). We identified articles that used lagged explanatory variables by searching the full text of each for the word "lag," and then discarding articles that used lags purely for the purposes of forecasting, or that used the word "lag" in some other context, including articles that lagged only their dependent variable, or included only spatial lags.

The resulting count of articles in Table 1 suggests that this practice is much more common in political science relative to economics or sociology. The low number for the *American Political Science Review* in 2014 is also not typical for that journal: we uncovered twenty-three articles between 2010 and 2014. We also looked closely at the justifications that authors provided for including lagged explanatory variables. Articles in economics journals

Table 1: Journals Reviewed

Journal Name	Discipline	Lag Articles	Lag "Identified"
American Political Science Review	Political Science	3	1
American Journal of Political Science	Political Science	10	6
Journal of Politics	Political Science	10	8
Comparative Political Studies	Political Science	14	7
International Organization	Political Science	8	8
International Studies Quarterly	Political Science	15	10
World Politics	Political Science	7	6
American Economic Review	Economics	4	2
Econometrica	Economics	1	1
Journal of Political Economy	Economics	1	1
Quarterly Journal of Economics	Economics	2	0
Review of Economic Studies	Economics	1	1
Review of Economics and Statistics	Economics	8	6
American Sociological Review	Sociology	1	1
American Journal of Sociology	Sociology	0	0
European Sociological Review	Sociology	1	1

Notes: Lag Articles is a raw count of the number of articles published in 2014 that employed a lagged explanatory variable. Lag"Identified" is the number of Lag Articles that either involved endogeneity as a justification for lagging an explanatory variable, or contained no justification at all for lagging an explanatory variable.

frequently invoked theoretical concerns, and rarely justified their lag choices on endogeneity grounds.[2] However, articles in political science journals frequently invoked "simultaneity" or "reverse causality" explicitly as the sole motivation for lagging explanatory variables.[3] Somewhat more concerning, a substantial minority of articles that we identified in this survey contained no justification whatsoever for their lag choice. We did identify a number of cases where authors employed lagged explanatory variables as part of an

[2] An example is Kellogg (2014:1710), who justifies a three-month lag between his main predictor of interest (expected oil price volatility) and his outcome of interest (drilling an oil well) based on interviews with "industry participants" who suggested that this is how long it takes "between the decision to drill and the commencement of drilling."

[3] Some examples are as follows: Baccini and Urpelainen (2014:205) write "Most of these variables are lagged by one year to avoid endogeneity problems." Lehoucq and Perez-Linan (2014:1113) write "We lag both economic variables one year to minimize problems of endogeneity." Steinberg and Malhotra (2014:513) write "All independent and control variables are lagged by one year to mitigate the possibility of simultaneity or reverse causality bias."

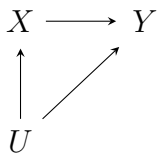
error correction or distributed lag model, but these remain the minority of the articles that we identified. As Table 1 shows, in 2014, across a range of journals, more than half of the articles that employed lagged exogenous variables either explicitly invoked endogeneity, or contained no justification at all.

This review of recent scholarship reveals that the practice of lagging explanatory variables for identification purposes remains common in the most influential and widely cited political science journals. We now turn to closer examination of the conceptual problems that this strategy creates.

2.1 Directed Acyclic Graphs

Following Pearl (2009), we begin with an intuitive discussion of the problem which relies on directed acyclic graph (DAGs). The DAG in Figure 1 shows the fundamental identification problem in observational data: the identification of the causal relationship flowing from X to Y is compromised by the presence of unobservable factors U which are correlated with both X and Y.

Figure 1: The Identification Problem



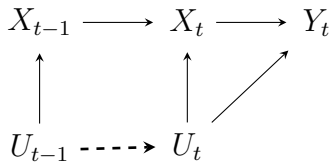
Notes: This is a representation of a causal relationship from X to Y where identification is compromised by unobservables U.

Figure 2, wherein we add subscripts t to clarify temporal ordering, illustrates the lag identification strategy that we study in this paper. Lag identification means replacing X_t with its lagged value, X_{t-1} in a regression of X on Y. The DAG representation in Figure 2 clarifies the logic behind this strategy. It must be the case that there is a causal pathway from $X_{t-1} \rightarrow X_t$, or else X_{t-1} could not be unrelated to Y. However, the fact that there is no direct causal link running from U_t to X_{t-1} means that there is no possibility that this

particular unobserved confounder U_t threatens causal identification.

But Figure 2 also shows that replacing X_t with X_{t-1} merely moves the endogeneity problem back one time period. It is true that X_{t-1} is unaffected by U_t, but it *is* affected by U_{t-1} for the same reason that $U_t \rightarrow X_t$. As a result, if there are any temporal dynamics in the unobservables, then the causal pathways $U_{t-1} \rightarrow U_t \rightarrow Y_t$ and $U_{t-1} \rightarrow X_{t-1} \rightarrow X_t \rightarrow Y_t$ prevent causal identification using X_{t-1}. The critical identification assumption in lag identification, therefore, is that there are no temporal dynamics among the unobservables. This assumption is not testable: doing so would require observing U, the unobservable confounder that motivates lagging X on identification grounds.[4]

Figure 2: Lagged Independent Variable as a Solution?



Notes: This is a representation of the causal relationship from X to Y that is implied when using a lagged value of X to overcome the identification problem in figure 1. The dashed line represents the causal relation among unobservables in time t and $t-1$ that must be zero.

Our discussion thus far has focused on endogeneity in the form of unobserved heterogeneity. In many applications, however, lag identification is justified on "reverse causality" grounds rather than unobserved heterogeneity grounds. The argument that temporal ordering prevents current realizations of the dependent variable from affecting past values of a causal variable of interest appears more reasonable as a defense against simultaneous or reverse causation. However, this perspective is misplaced. From a conceptual standpoint, we can reformulate most cases of reverse or simultaneous causation as problems of unobserved heterogeneity, in which a latent variable representing the "likelihood" or "propensity" of Y is an unobserved confounder that causes both Y and X.

[4]There also might be cases where $U_t \rightarrow X_t$ but $U_{t-1} \not\rightarrow X_{t-1}$. This would be a case of "time-varying endogeneity," and could yield identification even if there are dynamics in the unobservables. We are not aware of any case where this assumption has been invoked, much less been made explicit. At any rate, even if it were to be made, such an assumption would be unlikely to hold.

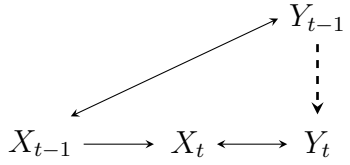
Because this idea may not be intuitive, we illustrate this argument using two concrete examples. Kelley and Simmons (forthcoming) study "the effect of monitoring and ranking on state behavior" (8), arguing that U.S. human rights reports shame countries into criminalizing human trafficking. They model their dependent variable Y (a dummy for "whether countries criminalize human trafficking in their domestic legislation") as a function of several key explanatory variables, including whether a country is named in the U.S. annual Trafficking in Persons Report. They are explicitly concerned about reverse causality: "All explanatory and control variables are lagged to help address reverse causality and selection issues" (8), and ask "Does the United States strategically shame countries that are likely to criminalize anyway?" (9). This articulation of the inferential threat facing their analysis is illuminating: the identification problem is not that criminalizing human trafficking causes countries to be named in the Trafficking in Persons Report, which would be a case of reverse causality. Rather, it is that strategic dynamics not captured in the observables determine both criminalization and being included in the report. In this case, the unobservable confounder U can be understood as whatever unobserved propensity to criminalize human trafficking is not captured in the explanatory or control variables, but which also drives U.S. scrutiny of a country's human trafficking problem. Substantively, this may be something like activism and political pressure by D.C.-linked activists in trafficking countries. The methodological point is that the inferential threats of "reverse causality" can be expressed as threats from unobserved heterogeneity: $U_{t-1} \rightarrow \text{Shaming}_{t-1}$ and $U_{t-1} \rightarrow U_t \rightarrow \text{Criminalization}_t$.

Warren (2014) offers another example. This article tests the hypothesis that "states with high levels of media accessibility will be less likely to experience the onset of civil war" (123). The independent variable of interest is a media density index. Identification is a problem, however: "to guard against spurious results due to reverse causation, all independent variables are lagged by one year" (126). In this case, it is theoretically possible that the onset of war directly reduces the density of countries' media markets.

But a more general formulation of the inferential problem is that there is a latent, unobserved probability of civil conflict that both leads to civil war onsets and that hampers media development even when a civil war onset does not actually occur. In this case, U is the latent probability of civil war, the past values of which affect past values of media density as well as the current onset of civil war: $\mathrm{P(Conflict)}_{t-1} \rightarrow \mathrm{MediaDensity}_{t-1}$ and $\mathrm{P(Conflict)}_{t-1} \rightarrow \mathrm{P(Conflict)}_{t} \rightarrow \mathrm{Conflict}_{t}$.

We note that we are not conjuring these endogeneity problems ourselves. Instead, we are articulating them on behalf of authors who have proposed statistical models that explicitly recognize that these challenges exist. Our point in highlighting them is to show that it is easy to reformulate problems of reverse causality as problems of unobserved heterogeneity.[5] For this reason, our formal analysis in the next section will represent endogeneity as unobserved heterogeneity, which we can capture as an omitted variable. Nevertheless, it is useful to highlight that our argument will also travel to contexts with "pure" reverse or simultaneous causation between X and Y. A classic example of this form of simultaneous causation is Haavelmo's (1943) treatment of the joint determination of consumption and investment. This causal process is depicted in figure 3, which shows that if Y_t causes X_t, Y_{t-1} also causes X_{t-1}.

Figure 3: Lagged Independent Variable with Pure Reverse Causality



Notes: This is a representation of pure simultaneous causation with no unobservables. The dashed line represents the causal relation among dependent variables in time t and $t-1$ that must be zero.

The identification assumption is now that there are dynamics in X but not Y. If $Y_{t-1} \rightarrow X_{t-1}$ and $Y_{t-1} \rightarrow Y_t$, then substituting X_{t-1} for X_t does not avoid the identification prob-

[5]See Pearl (2009:145-149) for a related argument on the observational equivalence of structural equation models. His argument begins as follows: "if we regard each bidirected arc $X \leftarrow\rightarrow Y$ as representing a latent common cause $X \leftarrow L \rightarrow Y$..."

lem. We will analyze a system of this sort in section 4.3.3 below.

3 Analytical Results

The DAGs in the preceding section are useful for clarifying the intuition behind lagged independent variables, and also for demonstrating why they are unlikely to sidestep problems of endogeneity. To characterize precisely the inferential problems that arise from lagged independent variables in the context of endogeneity, in this section we analyze formally the consequences of lag identification in a bivariate OLS regression setup. Here, assume that the OLS regression framework is the correct functional form for the estimation of the causal effect of X on Y. If the correct functional form is unknown, then a non-parametric approach such as those offered by Pearl or Rubin, as well as precise assumptions about counterfactual outcomes, are necessary to define estimators that estimate causal effects.

Consider the model

$$Y_{it} = \beta X_{it} + \delta U_{it} + \epsilon_{it} \tag{1}$$

$$X_{it} = \rho X_{it-1} + \kappa U_{it} + \eta_{it} \tag{2}$$

$$U_{it} = W_{it} + \phi W_{it-1} + \nu_{it} \tag{3}$$

$$\begin{matrix} \eta \\ W \end{matrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\eta^2 & \sigma_{\eta,W} \\ \sigma_{\eta,W} & \sigma_W^2 \end{pmatrix} \right) \tag{4}$$

where i and t index units and time, respectively; $0 \leq \rho < 1$; and $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$, $\eta_{it} \sim N(0, \sigma_\eta^2)$. Dropping i for the remainder of this section (it will reappear in the next section), it is well known that if we estimate

$$Y_t = bX_t + e_t \tag{5}$$

then the resulting estimate of β is biased because the unobserved confounder U is omitted.[6] The magnitude of the bias is a function of the variances and covariances of X and U as well as magnitude of the causal effect of the unobserved confounder:

$$\mathrm{E}[\hat{b}_{X_t}] = \beta + \delta \cdot \frac{\mathrm{Cov}(X, U)}{\mathrm{V}(X)} \tag{6}$$

If either δ or $\mathrm{Cov}(X, U) = 0$—if U has no effect on Y, or if U is uncorrelated with X—then endogeneity is not a problem, and $\mathrm{E}[\hat{b}_{X_t}] = \beta$.[7]

The system of equations in (1 – 4) allows for two distinct channels that can produce endogeneity. The first channel is the straightforward case of $U \rightarrow X$, in which the unobservable confounder has a causal relationship with the endogenous variable. The size of that causal effect of U on X is the parameter κ. But we also allow for a second source of endogeneity that captures *any other reason why X and U might be correlated*. That is the correlation parameter $\sigma_{\eta,W}$ in (4). This term will capture, for example, a still more deeper set of causal relations where another unobserved confounder causes both X and U. We include the term $\sigma_{\eta,W}$ to emphasize that our derivation captures *any* such form of endogeneity between X and U.

Now consider a regression that replaces X with X_{t-1}, but which otherwise remains subject to the same endogeneity problems as in (1).[8] This means estimating the following equation:

$$Y_t = bX_{t-1} + e_t \tag{7}$$

While this is plainly not an unbiased estimate of β,[9] one hope is that lag identification will

[6] We use Greek letters for population coefficients and Latin letters for sample coefficients.

[7] Based on the DGPs in Equations 1-3, one can also derive $\mathrm{Cov}(X, U)$. The derivation is presented in Appendix 1.

[8] For purposes of clarity we do not consider here more complicated models that condition on past values of Y. We will show in our simulation analysis and in the conclusion that lagging the dependent variable in addition to the endogenous explanatory variable does not avoid endogeneity problems either.

[9] Indeed, even when equation (2) is such that $X_{it} = X_{it-1} + \eta_i t$, β suffers from attenuation bias given that

estimate a function of β and the autocorrelation in X, or ρ—a moderated, or "conservative," estimate of β. Indeed, by expressions (1 – 4), lag identification implies estimating the following population parameter:

$$Y_t = \beta(\rho X_{t-1} + \kappa U_t + \eta_t) + \delta U_t + \epsilon_t \tag{8}$$

$$= \beta\rho X_{t-1} + \beta\kappa U_t + \beta\eta_t + \delta U_t + \epsilon_t \tag{9}$$

This immediately makes clear why lag identification fails, for the error term e_t in (7) now contains $(\beta\kappa+\delta)U_t+\beta\eta_t+\epsilon_t$. Therefore, $\hat{b}_{X_{t-1}}$ from (7) is not a consistent estimate of either β or the conservative $\beta\rho$. To see why, recall that $\hat{b}_{X_{t-1}} = \frac{Cov(X_{t-1},Y_t)}{V(X_t-1)}$. We may write

$$\text{plim}_{n\to\infty}\,\hat{b}_{X_{t-1}} = \beta\rho + \frac{\text{Cov}(X_{t-1}, (\beta\kappa+\delta)U_t + \beta\eta_t + \epsilon_t)}{\text{V}(X_{t-1})} \tag{10}$$

$$= \beta\rho + \frac{\text{Cov}(\rho X_{t-2} + \kappa U_{t-1} + \eta_{t-1}, (\beta\kappa+\delta)U_t + \beta\eta_t + \epsilon_t)}{\text{V}(X_{t-1})} \tag{11}$$

$$= \beta\rho + \frac{\rho(\beta\kappa+\delta)\text{Cov}(X_{t-2}, U_t) + \rho\beta\text{Cov}(X_{t-2}, \eta_t) + \rho\text{Cov}(X_{t-2}, \epsilon_t)}{\text{V}(X_{t-1})}$$
$$+ \frac{\kappa(\beta\kappa+\delta)\text{Cov}(U_{t-1}, U_t) + \kappa\beta\text{Cov}(U_{t-1}, \eta_t) + \kappa\text{Cov}(U_{t-1}, \epsilon_t)}{\text{V}(X_{t-1})}$$
$$+ \frac{(\beta\kappa+\delta)\text{Cov}(\eta_{t-1}, U_t) + \beta\text{Cov}(\eta_{t-1}, \eta_t) + \text{Cov}(\eta_{t-1}, \epsilon_t)}{\text{V}(X_{t-1})} \tag{12}$$

We know that by design, given expressions (1 – 4), that $\text{Cov}(U_{t-1}, U_t) = \phi\sigma^2_W$ and $\text{Cov}(\eta_{t-1}, U_t) = \phi\sigma_{W,\eta}$ with all the other covariances set at zero. Thus, equation (12) reduces to

$$\text{plim}_{n\to\infty}\,\hat{b}_{X_{t-1}} = \beta\rho + \frac{\phi(\beta\kappa+\delta)(\kappa\sigma^2_W + \sigma_{\eta,W})}{\text{V}(X_{t-1})} \tag{13}$$

Contrasting lag identification bias in (13) with the standard result for omitted variable bias in (6) usefully highlights the troublesome properties of lagged independent variables

X_{it} is simply X_{it-1} measured with error.

from a causal identification perspective. As $\sigma_{\eta,W}$, κ, and δ grow larger, $\hat{b}_{X_{t-1}}$ grows larger as well. And critically, in the cases where (7) does *not* yield a biased estimate of β, (13) does. Those are the cases where $\delta = 0$ but $\phi \neq 0$ (or there exists dynamics in the unobservable variable U_t). More substantively, lagging X_t and using it as a regressor can open up a "backward channel" through $U_{t-1} \rightarrow X_{t-1}$ and $U_{t-1} \rightarrow U_t \rightarrow Y_t$.

In fact, expression (13) confirms that *either one* of the following conditions should hold for lag identification to produce a consistent estimate of $\beta\rho$ (which is a "conservative" estimate of the effect of X on Y, attenuated by ρ).

1. No serial autocorrelation in U ($\phi = 0$), i.e., no dynamics among unobservables.

2. There is no endogeneity of any type, which means that $\kappa = \sigma_{\eta,W} = 0$.

The former condition is precisely the condition identified in Section 2.1 above. In that case, the second term reduces to zero, and $\hat{b}_{X_{t-1}} = \beta\rho$.

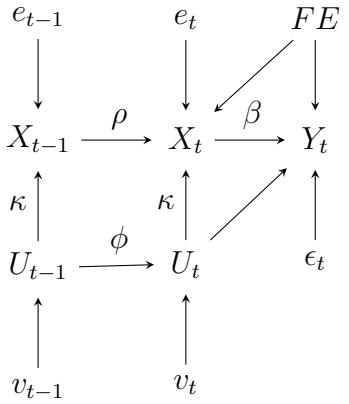
4 Monte Carlo Analysis

We have argued so far that when researchers believe that endogeneity threatens their ability to estimate causal effects, lagging independent variables does not alleviate these concerns, it simply moves the identification assumption to a different point in the data-generating process. We have also characterized analytically the magnitude of the bias in a lagged independent variables framework in a simple OLS regression setup. In this section, we use Monte Carlo experiments to demonstrate the consequences of using lagged independent variables in empirical research.

4.1 Setup

Our task is to estimate β, the causal effect of X on Y. Figure 4 is an extension of our earlier analysis which parameterizes the causal relations of interest. As above, the source

Figure 4: Monte Carlo Simulations



Notes: This is a schematic representation of our Monte Carlo simulations, with Greek letters representing the parameters that we vary in our simulations. X is the causal variable of interest, represented here as a function of a random variable e and its own past value as well as unit fixed effects FE. U is an unbserved source of endogeneity, and is itself a function of a random variable v and its own past value. Y is the dependent variable, and is a function of observed X, unobserved U, fixed effects FE, and a random error term ϵ. β is the causal parameter to be estimated, κ measures the size of the endogeneity problem, and ρ and ϕ capture dynamics in X and U, respectively.

of endogeneity bias is the unobserved confounder U, which is correlated with both X and Y. In all simulations, we set the direct effect of U on Y (which we called δ above) equal to 1, and explore the consequences of endogeneity bias by varying κ, the causal pathway that makes X endogenous to Y by forcing $\mathrm{Cov}(X, U) \neq 0$. The remaining two parameters are the autocorrelation parameters ρ and ϕ, which capture serial correlation in X and U, respectively. When either of the autocorrelation parameters is zero, then the value of each variable is statistically independent of its own lag. These are the four parameters that we vary across simulations; a summary appears in Table 2.

Table 2: Simulation Parameters

Parameter	Causal Pathway	Simulation Values
β	$X_t \rightarrow Y_t$	0, 2
κ	$U_t \rightarrow X_t$, $U_{t-1} \rightarrow X_{t-1}$	0, .5, 3
ρ	$X_{t-1} \rightarrow X_t$	0, .5, .9
ϕ	$U_{t-1} \rightarrow U_t$	0, .5, .9

For each simulation, we generate a panel with $N = 100$ units and $T = 50$ periods, for a total of 5,000 unit-period observations. To replicate a standard panel data problem, we also include time-invariant unit fixed effects FE_i, which we do not observe, and which affect both Y and X. Altogether, then, we simulate the following system of equations.[10]

$$Y_{it} = \beta X_{it} + 1 \cdot U_{it} + 1 \cdot FE_i + \epsilon_{it} \tag{14}$$

$$X_{it} = \rho X_{it-1} + \kappa U_{it} + 1 \cdot FE_i + e_{it} \tag{15}$$

$$U_{it} = \phi U_{it-1} + v_{it} \tag{16}$$

where

$$\begin{matrix} FE_i \\ \epsilon_{it} \\ e_{it} \\ v_{it} \end{matrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right) \tag{17}$$

We replicate each combination of parameter values in Table 2 a total of 100 times, and then test the performance of three estimators in estimating β: (i) the "naïve" estimator that regresses Y on X and ignores endogeneity, (ii) the "lag explanatory variables" estimator that regresses Y on X_{t-1} in an attempt to avoid endogeneity problems, and (iii) a "true" estimator that conditions on the unobservable U.[11] The "true" estimator is, of course, counterfactual: we presume that the researcher does not observe U, else she would condition on it. The estimates obtained from a regression model that correctly follows the data generating process, however, will serve as our empirical benchmark against which to

[10] We set the variance of FE and ϵ at 5 in order to allow for a realistic amount of model uncertainty. Most estimates from our simulations have an overall R^2 between 0.05 and 0.1, which is comparable to R^2 measures in much applied work. Note also that the covariance matrix in (17) reflects an assumption that $\sigma_{\eta,W} = 0$, so the only form of endogeneity that we model in our simulations is where $U \rightarrow X$. Our results, of course, generalize to other forms of endogeneity reflected in (4) as well.

[11] Each estimator is a fixed effects regression, which is necessary given our data generating process. We estimate these models using the "within" estimator implemented in the `plm` library in R.

gauge the performance of the other two estimators.

We emphasize that our causal model has many moving parts, but is still simple in terms of the dynamics that it allows. Among many other simplifications, we assume that there are no dynamic causal relationships among unobservables and observables. For example, lagged omitted variables U are not direct causes of current values of Y—they only affect Y through the pathway $U_{t-1} \rightarrow U_t \rightarrow Y_t$. Moreover, there are no complex temporal dynamics in X or U, just simple one-period autocorrelation, and the fixed effects FE are independent of U. We view this relatively straightforward setup as a conservative way to show just how difficult it is to justify lagged independent variables as sources of exogeneity in even the most favorable cases.

We evaluate the consequences of lag identification according to three criteria: (i) bias, (ii) root mean squared error (RMSE), and (iii) the likelihood of Type 1 or Type 2 error. The last of these is perhaps the most important from the perspective of applied researchers, as it tells us the extent to which researchers will make faulty inferences—rejecting true null hypotheses that $\beta = 0$, or failing to reject the null hypothesis when the true $\beta > 0$—when using lagged independent variables to sidestep problems of endogeneity.

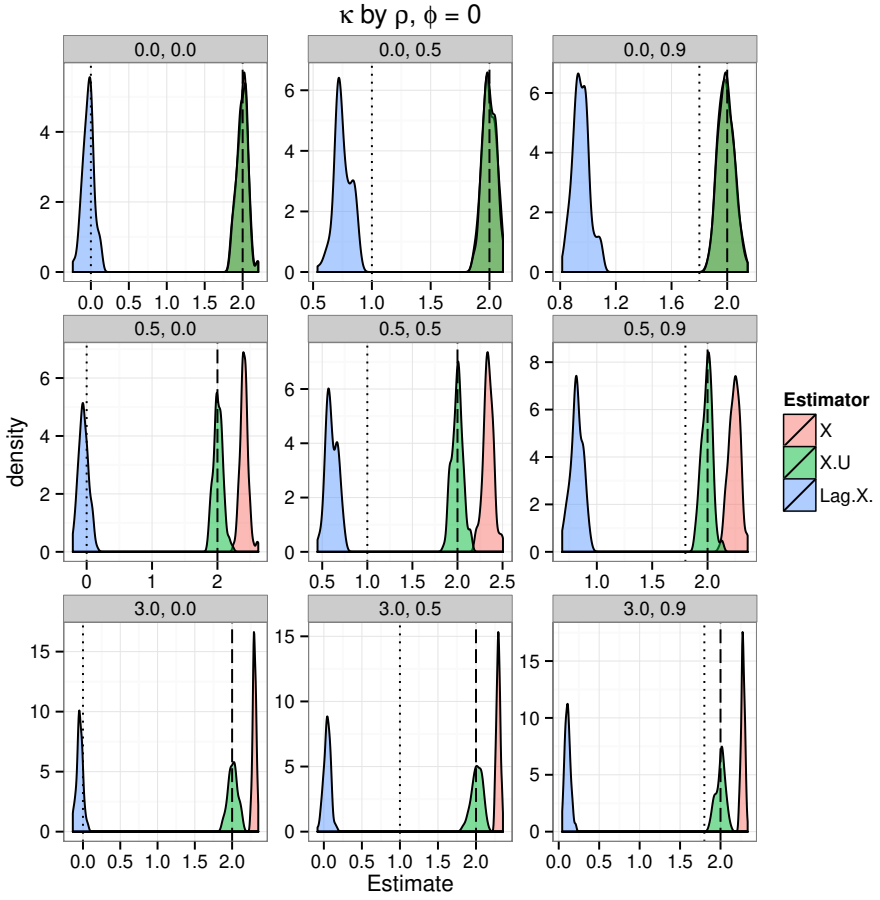
4.2 Results

We begin by comparing bias across the three estimators. For each combination of parameter values, we save the estimated parameter $\hat{\beta}$ from each of the 100 simulations, and the plot the distribution of estimates along with the true value of β from the data generating process. Figure 5 summarizes our main results for the case where the true causal effect β is equal to 2 and the autocorrelation in the unobservables ϕ is set equal to 0, and Figure 6 shows there results when $\phi = 0.5$.[12]

The results from these simulations are clear. For any value of ϕ, and regardless of

[12]Our results are comparable when $\phi = 0.9$, but we do not report them here to save space. They are available upon request.

Figure 5: Bias in Lag Explanatory Variable Regressions, $\phi = 0$

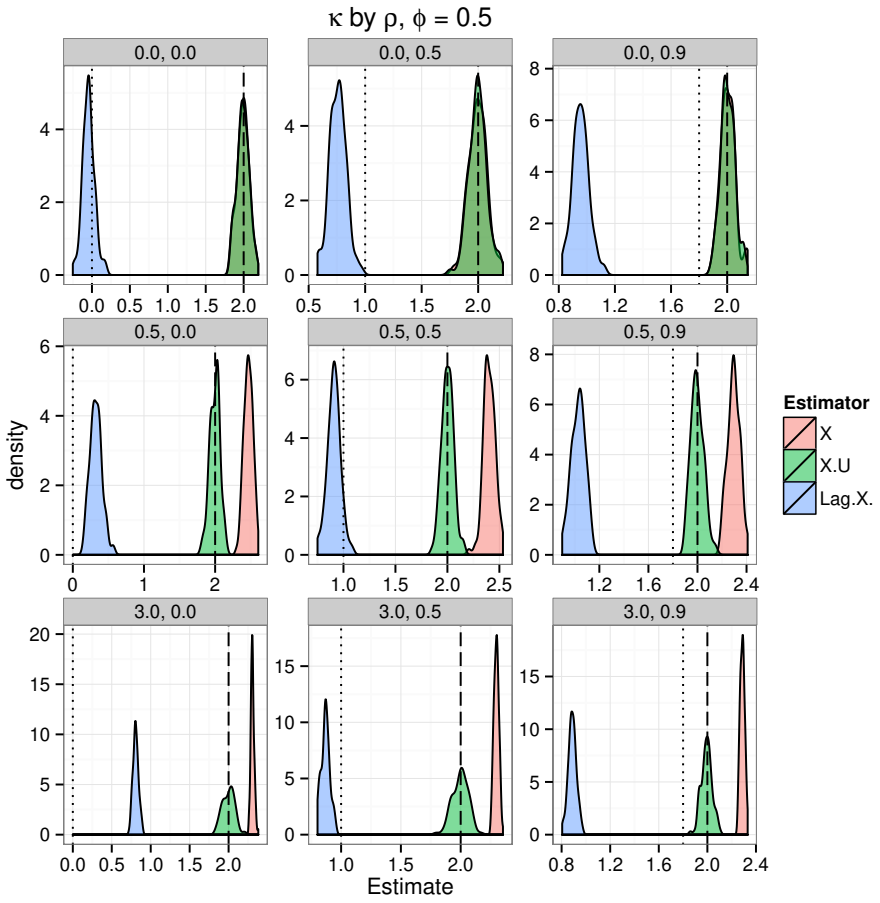


Notes: This figure plots the empirical density of estimated coefficients $\hat{\beta}$ from 100 simulations of the data generating process in Figure 4. The dashed line corresponds to the true value of $\beta = 2$. The dotted line is $\beta \times \rho$.

whether the data generating process makes X endogenous to Y, lagged explanatory variable estimates are biased towards zero. This result is intuitive: X_{t-1} is a proxy for X_t which measures the latter with error, producing attenuation bias in β even without endogeneity. Moreover, and as expected, when there is no endogeneity (i.e., $\kappa = 0$), then regressing Y_t on X_t suffices to identify β. However, when $\kappa > 0$, we find that neither the "naïve" nor the "lag explanatory variable" estimator identifies β. The amount of bias in both is a function of ρ: the higher the correlation between X_t and X_{t-1}, the less the bias.

In Table 3 we calculate the root mean squared error (RMSE) of each estimator for each combination of parameters. We also add two additional estimators in which we condition

Figure 6: Bias in Lag Explanatory Variable Regressions, $\phi = .50$



Notes: This figure plots the empirical density of estimated coefficients $\hat{\beta}$ from 100 simulations of the data generating process in Figure 4. The dashed line corresponds to the true value of $\beta = 2$. The dotted line is $\beta \times \rho$.

on Y_{t-1} in an attempt to account for dynamics.

The results show that the RMSE of the lag explanatory variable estimator is far larger than that of the naïve estimator, which in turn is much larger than the estimator that conditions on U. The same is true even when conditioning on lagged values of Y. The only exceptions are purely incidental: when the true value of $\beta = 0$ (not reported in Table 3), there are a small number of parameter combinations in which the lower variance of the lag explanatory variable estimator yields a lower RMSE than that of the estimator that conditions on U. Purely on RMSE grounds, our results indicate that the lag explanatory variable estimator is almost always worse than the naïve estimator which simply ignores

Table 3: Root Mean Squared Error

$X_t \vert U_t$	X_t	X_{t-1}	$X_t \vert Y_{t-1}$	$X_{t-1} \vert Y_{t-1}$	β	ϕ	κ	ρ
0.071	0.072	2.035	0.073	1.994	2	0.0	0.0	0.0
0.072	0.413	2.048	0.412	2.000	2	0.0	0.5	0.0
0.064	0.300	2.044	0.300	1.996	2	0.0	3.0	0.0
0.055	0.057	1.252	0.058	1.210	2	0.0	0.0	0.5
0.061	0.348	1.395	0.365	1.313	2	0.0	0.5	0.5
0.071	0.295	1.957	0.295	1.877	2	0.0	3.0	0.5
0.056	0.057	1.051	0.061	1.010	2	0.0	0.0	0.9
0.048	0.248	1.183	0.272	1.093	2	0.0	0.5	0.9
0.058	0.278	1.890	0.280	1.793	2	0.0	3.0	0.9
0.079	0.079	2.046	0.078	2.044	2	0.5	0.0	0.0
0.071	0.472	1.677	0.470	1.700	2	0.5	0.5	0.0
0.077	0.308	1.193	0.310	1.167	2	0.5	3.0	0.0
0.080	0.082	1.244	0.085	1.236	2	0.5	0.0	0.5
0.057	0.407	1.093	0.412	1.084	2	0.5	0.5	0.5
0.066	0.299	1.131	0.305	1.085	2	0.5	3.0	0.5
0.050	0.053	1.045	0.053	1.040	2	0.5	0.0	0.9
0.052	0.300	0.974	0.309	0.954	2	0.5	0.5	0.9
0.045	0.289	1.112	0.296	1.057	2	0.5	3.0	0.9
0.063	0.066	2.041	0.066	2.067	2	0.9	0.0	0.0
0.073	0.613	1.452	0.602	1.525	2	0.9	0.5	0.0
0.066	0.313	0.952	0.314	0.935	2	0.9	3.0	0.0
0.066	0.069	1.250	0.070	1.271	2	0.9	0.0	0.5
0.070	0.539	0.938	0.531	0.961	2	0.9	0.5	0.5
0.063	0.310	0.919	0.314	0.890	2	0.9	3.0	0.5
0.059	0.061	1.050	0.062	1.073	2	0.9	0.0	0.9
0.060	0.394	0.876	0.390	0.895	2	0.9	0.5	0.9
0.049	0.300	0.903	0.305	0.861	2	0.9	3.0	0.9

endogeneity entirely.

Of course, the finding that an incorrect regression specification generates biased parameter estimates is not surprising. In fact, for most applied researchers, attenuation bias does not matter because—all too commonly, in our view—the size of the estimate of β is not of direct interest, but rather its p-value. That is, scholars are less interested in whether their estimate $\hat{\beta} = 2$ or $\hat{\beta} = 0.02$, but whether the associated p-value from their t-test leads them to reject the null that $\beta = 0$ at some level of significance, which supports the presence of a causal relationship. Some readers may even believe that the attenuation bias of

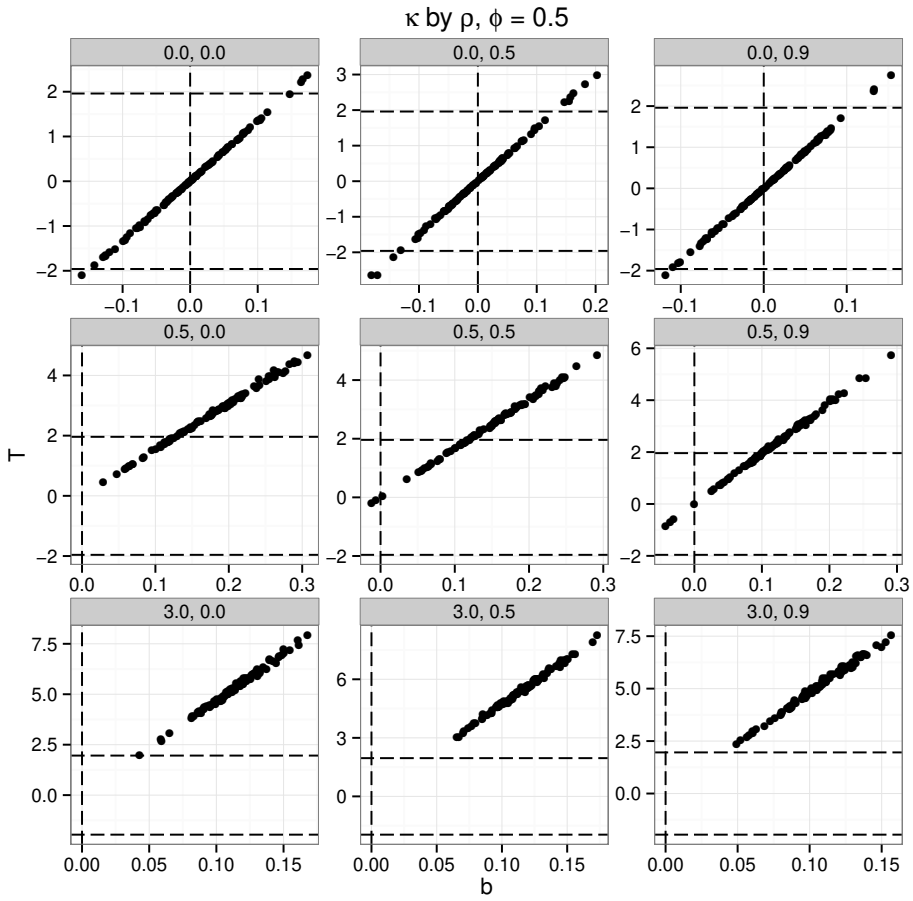
the lag explanatory variable estimator is one of its strengths relative to the naïve estimator that ignores endogeneity altogether because attenuation bias leads to more conservative hypothesis tests, although our discussion in the previous section should dispel that notion. We think that the overwhelming focus placed on statistical significance in applied work is a major problem, but this is nonetheless an accurate description of current practice. And so we ask what would happen if an applied researcher were to use a lagged independent variable in the standard fashion to test the hypothesis that X causes Y when the true $\beta = 0$. In Figure (7), we plot the estimates of β against their t-statistics for models where $\phi = 0.5$.

These results are troubling, yet consistent with the analytical results previously described. When there is no endogeneity problem ($\kappa = 0$, as in the top three panels) then Type 1 error is rare, corresponding to just about 95% of models when the level of statistical significance is $\alpha = .05$. But the likelihood of Type 1 error increases dramatically when $\kappa > 0$, as in the bottom six panels. The reason for this is apparent in expression (13), which shows that $\hat{b}_{X_{t-1}}$ is a function of the causal effect of the unobserved confounder U, δ, as well as $\mathrm{Cov}(X, U)$. Unless both are exactly zero, lag identification will produce non-zero estimates of β even when $\beta = 0$.

Substantively, this means that lagging independent variables in response to concerns about endogeneity will lead analysts working within the mainstream approach to hypothesis testing to reject null hypotheses that are true, and to find too many estimates of causal effects that are spurious. As suggested in the formal analysis, the direction of the error depends on the correlations between X and U and Y and U. When both are positive, lag explanatory variable models estimate $\hat{\beta} > 0$; when both are negative, lag explanatory variable models estimate $\hat{\beta} < 0$; and when one is positive and the other is negative the sign of the $\hat{\beta}$ depends on their relative absolute size.[13] When $\phi = .9$ (not reported), in-

[13] In such cases there is a range of parameter values in which lag explanatory variable models do not generate Type 1 errors, but there is also a range of parameter values for which lag explanatory variable models generate Type 2 errors when the true value of $\beta \neq 0$.

Figure 7: Type I Error in Lag Explanatory Variable Regressions, $\phi = 0.5$



Notes: This figure compares estimated coefficients $\hat{\beta}$ and t-statistics from 100 simulations of the data generating process in Figure 4. The vertical dotted line corresponds to the true value of $\beta = 0$, and the horizontal dotted lines denote the 95% confidence region from -1.96 to +1.96.

dicating even stronger autocorrelation among the unobservables U, Type 1 error (given our parameters) is almost certain with any degree of endogeneity and regardless of the autocorrelation in X.

The summary message from these Monte Carlo simulations is unambiguous. Under favorable conditions, lagging independent variables generates estimates that are more biased, and with higher RMSE, than simply ignoring endogeneity altogether. Worst of all, such estimates are more likely to produce Type 1 error when endogeneity actually does threaten causal identification and the true causal effect of X on Y is zero. One worry-

ing implication of this last result is that lagged independent variables may be a popular identification strategy under conditions of endogeneity precisely because they generate statistically significant results.

4.3 Extensions

In this section we entertain several potential objections to our simulation results, focusing on temporal sequencing of causal effects, dynamic panel data estimators, and pure cases of simultaneous causality.

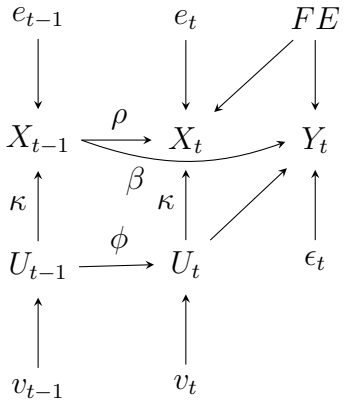
4.3.1 Lagged Causal Effects

One criticism of our baseline results is that they do not realistically reflect the kinds of data generating processes that scholars mean to capture using lag identification to avoid endogeneity problems. If theory suggests that causal effects operate with a one-period time lag, for example, then lag identification is not just a way to avoid endogeneity, it is also the natural way to estimate the correct causal parameter β. Such an objection might suggest that the disturbing results in the previous subsection are simply a consequence of proposing a different data generating process than the one that might justify the lag identification strategy.

Attuned to such concerns, in Figure 8 we propose a different causal model for Monte Carlo analysis. Here, as before X is endogenous to Y through U, but the causal effect of interest β is the one-period lagged effect of X on Y. We therefore assume that the causal effect operates with a one-period lag, that the empirical specification is designed to estimate that quantity, and also that the contempaneous casual effect of X_t on Y_t is exactly zero. This reflects perhaps the most favorable case for lag identification, one in which causal effects operate over time and in which there is no direct causal pathway that runs from the unobserved source of U_t to the causal variable X_{t-1}.

In Figure 9 we compare estimates of $\hat{\beta}$ from the lag explanatory variable estimator, an

Figure 8: Monte Carlo Simulations: X_{t-1} as the Causal Variable

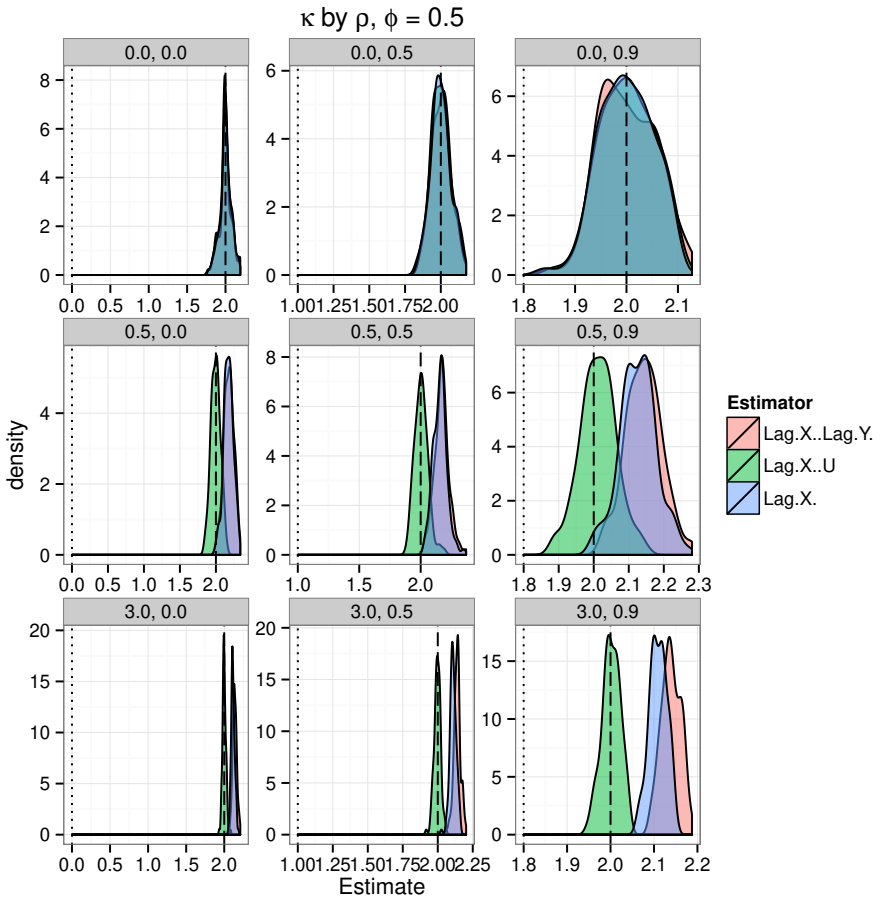


Notes: This is a schematic representation of our Monte Carlo simulations where X_{t-1} is the true causal variable (the causal effect of X_t is, by assumption, 0). It is otherwise identical to Figure 4.

extended version of the lag explanatory variable estimator that also conditions on Y_{t-1} in an attempt to capture temporal dynamics in the unobservables, and the "true" model that conditions on U_t, once again as an empirical benchmark against which to judge the others. Our results show that even under the favorable assumption that X_{t-1} is the causal variable of interest, lagged independent variables do not alleviate endogeneity bias. As above, when there is no endogeneity ($\kappa = 0$), then the three models are equivalent. When $\kappa > 0$, however, lagging independent variables generates biased estimates of β whose variance narrows as endogeneity grows larger. The results in Figure 9 also highlight that when X_{t-1} is the true causal variable, lag identification is not even "conservative," for estimates are *further away* from zero than β.

We also find similar results for Type 1 error, which appear in Figure 10. As before, these results indicate that with any amount of endogeneity, t-statistics associated with the $\hat{\beta}$ from a lagged explanatory variable estimate are likely to lead applied researchers to reject the null that $\beta = 0$ when the null is true. The implication of this analysis is that even if a strong theory dictates that the causal process linking X to Y operates with exactly and exclusively a one-period lag, lagged independent variables do not avoid problems of endogeneity.

Figure 9: Bias in Lag Explanatory Variable Regressions, $\phi > 0$, X_{t-1} is the Causal Variable



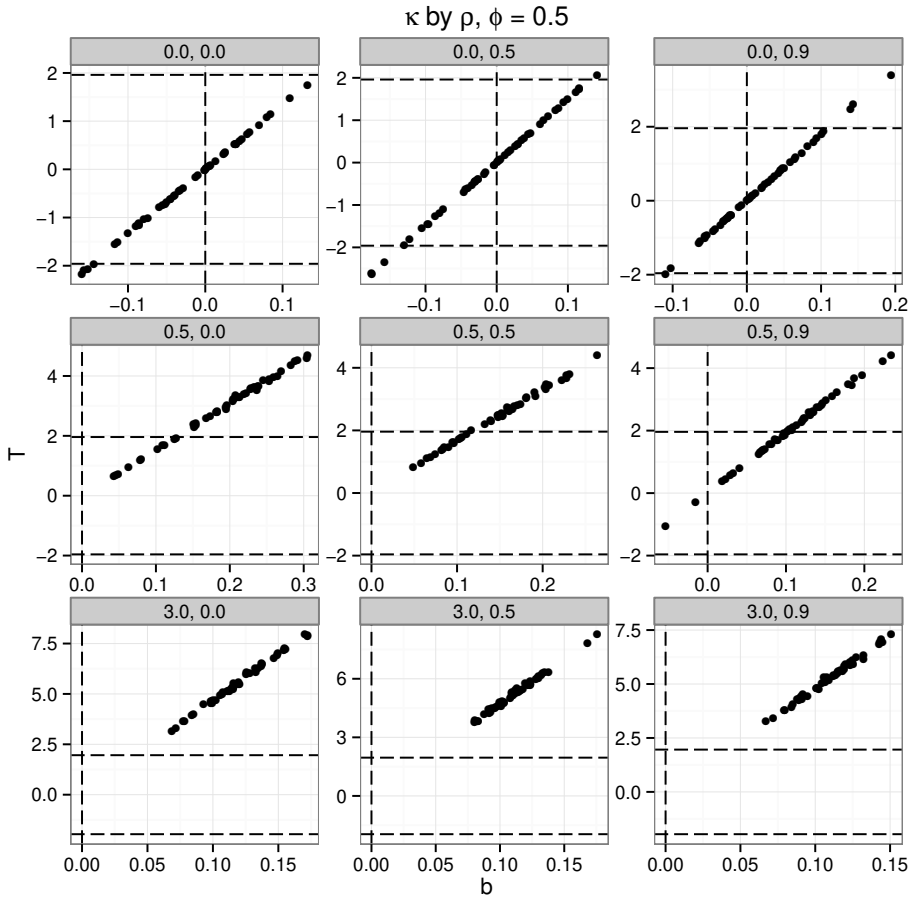
Notes: This figure plots the empirical density of estimated coefficients $\hat{\beta}$ from 100 simulations of the data generating process in Figure 8. The dashed line corresponds to the true value of $\beta = 2$. The dotted line is $\beta \times \rho$.

4.3.2 GMM Estimation

Another possible interpretation of our results is that standard panel data techniques are inappropriate for the dynamic causal processes that we have proposed. Specifically, when we include Y_{t-1} as a regressor in an attempt to account for the dynamics in U, we generate biased estimates of the coefficient on Y_{t-1} that might in turn bias our estimates of β, at least in finite samples. See Nickell (1981) for a fuller treatment.

We explore whether standard dynamic panel data models (Arellano and Bond 1991, Blundell and Bond 1998), which use higher order lags and differences of both X and Y

Figure 10: Type I Error in Lag Explanatory Variable Regressions, $\phi = 0.5$, X_{t-1} is the Causal Variable



Notes: This figure compares estimated coefficients $\hat{\beta}$ and t-statistics from 100 simulations of the data generating process in Figure 8. The vertical dotted line corresponds to the true value of $\beta = 0$, and the horizontal dotted lines denote the 95% confidence region from -1.96 to +1.96.

as instruments for X, X_{t-t}, and Y_{t-1}, yield better results.[14] The results of these analyses are available upon request, but the summary finding is straightforward: GMM estimation does a better job of recovering the causal effect of X_{t-1} on Y when the data generating process follows Figure 8 than does a simple lag explanatory variable strategy, but results remain biased away from zero and Type 1 errors remain very likely. Relative to the "true"

[14]We estimate these models using the `pgmm` estimator implemented in the `plm` library in R. We estimate two families of models, using the second and third lags of X and Y as instruments in the full panel of $T = 50$, and using all available lags of X and Y as instruments in a short panel of $T = 10$. We also vary the lag structure, estimating models in which we condition on X_{t-1} only as well as models where we condition on both X and X_{t-1}.

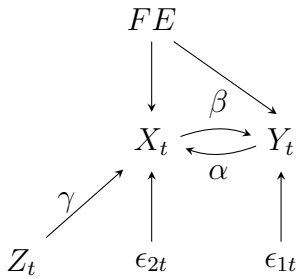
estimates obtained by conditioning on the confounder U, moreover, GMM estimates are much less efficient.

4.3.3 True Simultaneous Causality

Our final extension returns to the problem of simultaneous causality. Above, we argued that most instances of simultaneous or reverse causality can be reformulated in terms of an unobserved confounder U. Here, we entertain the possibility that true simultaneous causality has different implications for the estimation of causal effects.

Specifically, we consider the causal model in Figure 11, which is an extension of Haavelmo's (1943) classic treatment of simultaneous equations and the problem of causal inference (see also Pearl forthcoming). We incorporate into this model both unit-of-observation fixed effects and a (possible) instrument for X, denoted Z. Endogeneity in Figure 11 is not

Figure 11: Monte Carlo Simulations: Pure Reverse Causality



Notes: This is a schematic representation of our Monte Carlo simulations where X and Y are truly "simultaneous" equations. Z serves as an instrument for X whenever $\gamma > 0$. ϵ_{2t} follows an autoregressive process in our simulations: $\epsilon_{2t} = \phi\epsilon_{2t-1} + \eta$.

a function of unobserved confounders, but rather of a simultaneous causal relationship in which Y and X directly cause one another. We simulate using the following system of

equations to represent this causal structure:

$$Y_{it} = \beta X_{it} + \delta^Y FE_i + \epsilon_{1it} \tag{18}$$

$$X_{it} = \alpha Y_{it} + \delta^X FE_i + \gamma Z_{it} + \epsilon_{2it} \tag{19}$$

$$\epsilon_{2it} = \phi \epsilon_{2it-1} + \eta_{it} \tag{20}$$

where

$$\begin{matrix} FE_i \\ \epsilon_{1it} \\ \epsilon_{2it} \\ \eta_{it} \end{matrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right) \tag{21}$$

We introduce dynamics in X into the system by allowing for autocorrelation in ϵ_2, as in (20). If $\phi = 0$, meaning there is no autocorrelation in ϵ_2, then X_t and X_{t-1} are also uncorrelated. We note here that by substituting equations (18) and (20) into (19), we can express X solely in terms of model parameters and errors:

$$X_{it} = \frac{\alpha \delta^Y FE_i + \alpha \epsilon_{1it} + \delta^X FE_i + \eta_{it} + \phi \epsilon_{2t-1}}{1 - \alpha\beta} \tag{22}$$

This expression reveals the magnitude of endogeneity bias when regressing Y_t on X_t.

Because there are no unobserved variables in the data generating process represented in Figure 11, there is no identification strategy available—even theoretically—that involves conditioning on an unobservable, as there was in our prior simulations. Identification requires an instrumental variable Z. Throughout this subsection, we maintain the assumption in Figure 11 that Z is a *valid* instrument for X, and vary only the *relevance* (or "strength") of Z as an instrument for X, which we capture with the parameter γ. The

parameters for this final set of simulations are summarized in Table 4:

Table 4: Simulation Parameters: Simultaneous Causality

Parameter	Causal Pathway	Simulation Values
β	$X \rightarrow Y$	0, 2
α	$Y \rightarrow X$	0, 1, 3
ϕ	$\epsilon_{2t-1} \rightarrow \epsilon_{2t}$	0, .5, .9
γ	$Z_t \rightarrow X_t$	0, 1, 10

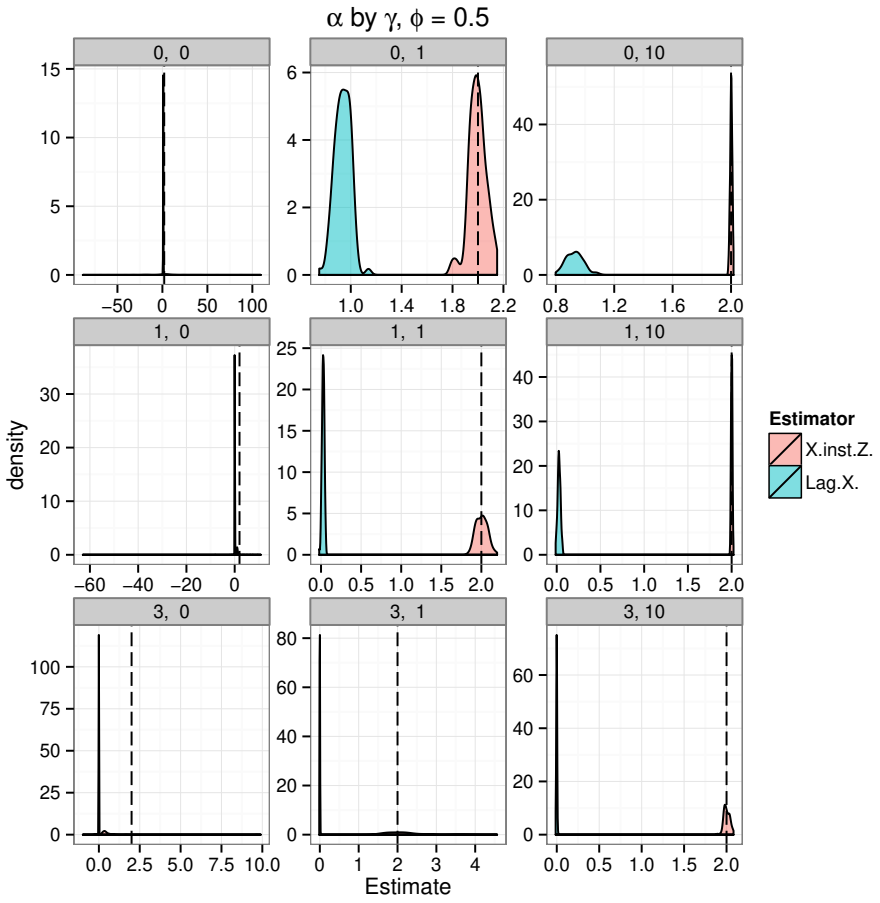
Our main results appear in Figure 12. Relative to other figures, these results are slightly more difficult to interpret. When $\gamma = 0$, meaning that the instrumental variable Z is unrelated to X, instrumental variables regression is completely uninformative, as reflected in the essentially flat density plot. When $\gamma > 0$, instrumental variables uncover the proper estimate of β, and the larger γ is, the better the instrumental variables estimator performs (recall that the larger γ is relative to its standard error, the stronger Z is as an instrument; see Bound et al. 1995). And regardless of the value of γ, the lag explanatory variable estimator always returns an estimate of β that is biased towards zero when $\beta = 2$.[15] The size of this lag explanatory variable estimator bias is increasing in α, the degree of endogeneity.

We conclude by studying Type 1 and Type 2 error for the pure simultaneous causality case. First consider the case where $\beta = 0$. We discover in Figure 13 that lag explanatory variable estimators are more likely to generate statistically significant negative estimates of β when the data generating process is characterized by true simultaneous causation as in Figure 11 and where there is any degree of endogeneity ($\alpha > 0$). However, additional complications arise where $\beta = 2$. Figure 14 reveals that here, Type 1 error—failing to reject a null that $\beta = 0$ when the null is actually false—is common when there is any amount of endogeneity.

The results of this analysis once again demonstrate the pitfalls of using lagged independent variables to achieve causal identification. Even under true simultaneous causality, where there are no unobserved confounders that generate endogeneity, lagged indepen-

[15] This result also holds regardless of the existence or strength of autocorrelation in ϵ_2 (the parameter ϕ).

Figure 12: Bias in Lag Explanatory Variable Regressions, $\phi = 0.5$, Pure Reverse Causality



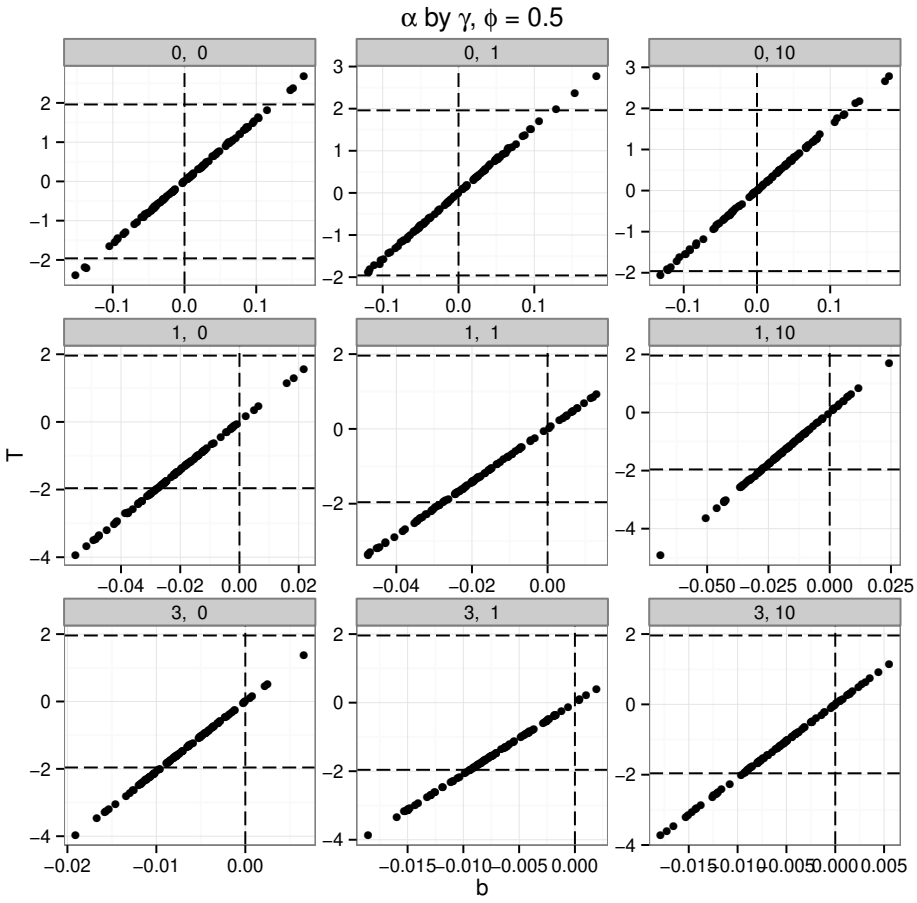
Notes: This figure plots the empirical density of estimated coefficients $\hat{\beta}$ from 100 simulations of the data generating process in Figure 11. The dotted line corresponds to the true value of $\beta = 2$.

dent variables lead to greater rates of both Type 1 *and* Type 2 error. Instrumental variables fare better, but only so long as Z is a relevant instrument for X. When relevant instruments are unavailable, there is no strategy for identifying statistically the causal effect of X on Y.

5 Summary and Conclusions

The genesis of this paper was a conversation among ourselves in which we commiserated about the frequency with which we reviewed manuscripts that used lagged explana-

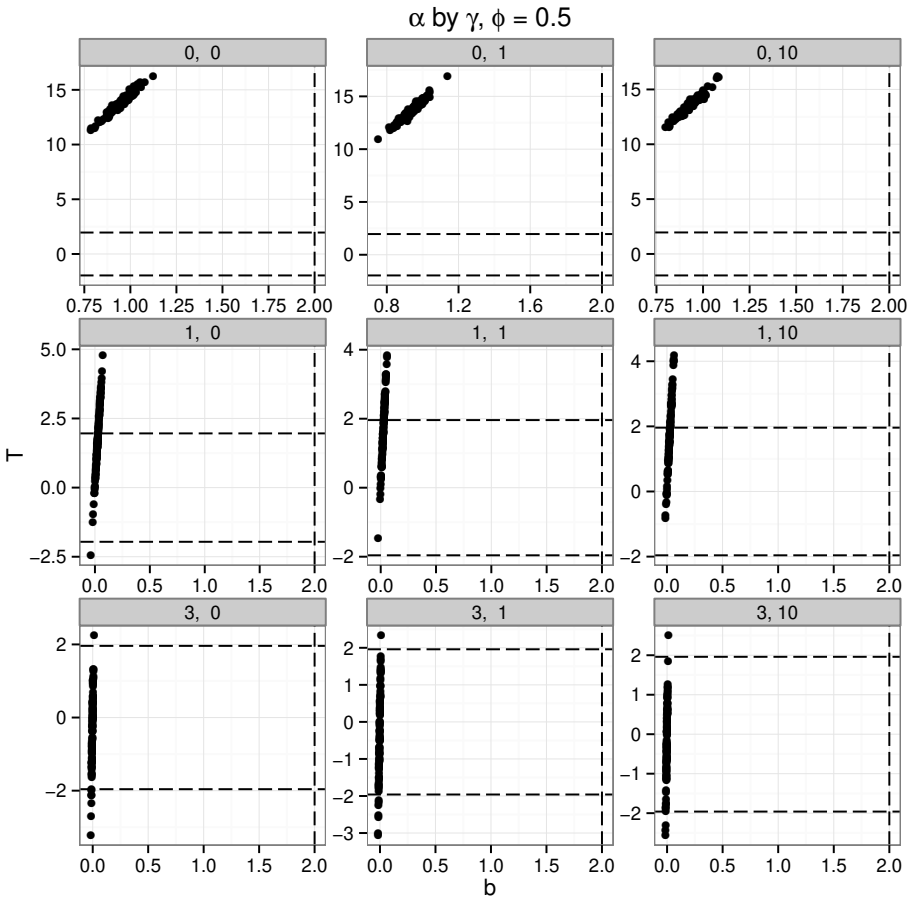
Figure 13: Type I Error in Lag Explanatory Variable Regressions, $\phi = 0.5$, Pure Simultaneous Causality



Notes: This figure compares estimated coefficients $\hat{\beta}$ and t-statistics from 100 simulations of the data generating process in Figure 11. The vertical dotted line corresponds to the true value of $\beta = 0$, and the horizontal dotted lines denote the 95% confidence region from -1.96 to +1.96.

tory variables to address endogeneity concerns. Many social scientists trained since the Credibility Revolution (Angrist and Pischke 2010) have an intuitive sense that this identification strategy is problematic, yet we suspect that few are able to precisely articulate the reasons why that is. In this paper, we showed the extent of the problem by showing how common lag identification is in political science relative to the cognate disciplines of economics and sociology. We then provided a simple treatment of the nature of the problem using directed acyclic graphs to uncover the "no dynamics among unobservables"

Figure 14: Type I Error in Lag Explanatory Variable Regressions, $\phi = 0.5$, Pure Reverse Causality



Notes: This figure compares estimated coefficients $\hat{\beta}$ and t-statistics from 100 simulations of the data generating process in Figure 11. The vertical dotted line corresponds to the true value of $\beta = 2$, and the horizontal dotted lines denote the 95% confidence region from -1.96 to +1.96.

assumption necessary to justify lagging independent variables as an identification strategy. We then showed analytically how lag identification introduces bias, even relative to a naïve identification strategy wherein one chooses to ignore the endogeneity of X_t altogether, and derived a formal result for the no dynamics among unobservables assumption in a simple OLS regression setup. To explore the consequences of lag identification in practice, we then presented the results of Monte Carlo simulations that show how the practice of lag identification not only fails to avoid the identification problem, it will lead to faulty

inferences under the null hypothesis significance testing paradigm.

Reed (2014), commenting on lagged explanatory variables and simultaneity bias, suggests that lagged values of X might be more defensible as *instrumental variables* (rather than proxies) for current values of X. He warns, however, that "is only an effective estimation strategy if the lagged values do not themselves belong in the respective estimating equation, and if they are sufficiently correlated with the simultaneously-determined explanatory variable" (8). Our analysis shows that even these conditions are too generous: in the "pure simultaneity case" consistent IV estimation will require that there are no dynamics in Y (cf. Figure 3), and in the more general endogeneity case it will still require no dynamics in U (cf. Figure 2). This is the case even if the lagged values can be excluded from the main estimating equation.

Practically, then, when scholars suspect that endogeneity may bias their estimates, the solution cannot arrive at the analysis stage. It must come earlier in the research process, at the study design stage. As Sekhon (2009) notes in a critical discussion of matching and causal inference, "for causal inference, issues of design are of utmost importance; a lot more is needed than just an algorithm. Like other methods, matching algorithms can always be used, and they usually are, even when design issues are ignored in order to obtain a nonparametric estimate from the data. Of course, in such cases, what exactly has been estimated is unclear." This point—invoked in a discussion of matching—applies equally to our critique lagged explanatory variables. As we have argued, lag identification replaces the assumption of "selection on observables" with the assumption of "no dynamics among unobservables." For causal inference to be credible, that assumption must be invoked and defended explicitly, not relied upon implicitly, as is currently the dominant practice in empirical political science. We stress that the assumption of no dynamics among unobservables could *in principle* be defensible. Unfortunately, however, we think it unlikely that many authors will be comfortable conceding that unobservable confounders exist, yet insisting that unobserved realizations of those confounders are temporally indepen-

dent from one another. For this reason, we conclude that without careful arguments on substantive grounds, lagged explanatory variables should never be used for identification purposes.

This argument, we emphasize, *does not imply that lagged explanatory variables are always and everywhere inappropriate*. We therefore conclude by providing some simple guidelines for researchers seeking to use them. There are several kinds of data generating processes in which lagged explanatory variables are appropriate:

1. In the context of unobserved confounding, in which case we have shown that the following two auxiliary assumptions are necessary:
   (a) No dynamics among unobservables U.
   (b) The lagged endogenous variable X is a stationary autoregressive process.
2. In the context of no unobserved confounding, in which case one of the following DGPs must be assumed:
   (a) There is no reverse causality ($Y \not\rightarrow X$) and the causal effect operates with a one period lag only ($X_{t-1} \rightarrow Y$ but $X_t \not\rightarrow Y_t$)
   (b) There is reverse causality ($Y \rightarrow X$), but reverse causality is contemporaneous only, and the causal effect of X on Y operates with a one period lag only (see Appendix 2 Figure A1 for one example).
   (c) There is reverse causality, and the causal effect of X on Y is contemporaneous, there are no dynamics in Y ($Y_{t-1} \not\rightarrow Y_t$), but there are dynamics in X ($X_{t-1} \rightarrow X_t$) (see the Appendix Figure A2 for one example).

Our focus in this manuscript has been on showing why Scenario 1 is almost always indefensible. But Scenarios 2(a), 2(b), and 2(c) are valid justifications for lagging explanatory variables.[16]

[16]The causal effect of X_{t-1} is identified under Scenarios 2(a) and 2(b) only. See Appendix Figure A2 for a discussion of why Scenario 2(c) does not allow for identification of causal effects of either X_t or X_{t-1} on Y.

Invoking 2(a), 2(b), or 2(c) requires that endogeneity takes the form of reverse or simultaneous causality, and also requires assumptions about the temporal dynamics in the data generating process. As assumptions, these are ultimately untestable. However, Scenarios 2(b) and 2(c) do suggest falsification tests that can be used to *rule out* each. Because each assumes selection on observables, each requires only data that is already available to the analyst.

Those tests are as follows:

1. Under Scenario 2(b), it must be the case that there there is no contemporary correlation between X and Y. If a regression of $Y_t = b_1 X_t + b_2 X_{t-1}$ uncovers a non-zero coefficient on b_1, then the data reject Scenario 2(b) as the data generating process.

2. Under Scenario 2(c), it must be the case that there are no dynamics in Y. If a regression of $Y_t = bX_t + \lambda Y_{t-1}$ uncovers a non-zero coefficient on λ, then the data reject Scenario 2(c) as the data generating process.

When employing lagged explanatory variables in the context of endogeneity, following these guidelines will help researchers to make explicit the data generating process that underlies their identification strategy. Doing so, in turn, will ensure that research designs with lagged explanatory variables are credible.

References

[1] Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association* 91(434): 444-455.

[2] Angrist, Joshua D., and Jörn-Steffen Pischke (2009), *Mostly Harmless Econometrics*, Princeton: Princeton University Press.

[3] Angrist, Joshua D., and Jörn-Steffen Pischke (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives* 24(2): 3-30.

[4] Angrist, Joshua D., and Jörn-Steffen Pischke (2014), *Mastering 'Metrics*, Princeton: Princeton University Press.

[5] Arellano, Manuel, and Stephen Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, " *Review of Economic Studies* 58(2): 277-297.

[6] Baccini, Leonardo, and Johannes Urpelainen (2014). "International Institutions and Domestic Politics: Can Preferential Trading Agreements Help Leaders Promote Economic Reform?, " *Journal of Politics* 76(1): 195-214.

[7] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics* 119(1): 249-275.

[8] Blackwell, Matthew, and Adam Glynn (2014), "How to Make Causal Inferences with Time-Series Cross-Sectional Data," Working Paper, Harvard University.

[9] Blundell, Richard, and Stephen Bond (1998), "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics* 87(1): 115-143.

[10] Bound, John, David Jaeger, and Regina Baker (1995), " Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association* 90(430): 443-450.

[11] Christ, Carl F. (1994), "The Cowles Commission Contributions to Econometrics at Chicago: 1939–1955," *Journal of Economic Literature* 32(1): 30-59.

[12] De Boef, Suzanna, and Luke Keele (2008), "Taking Time Seriously," *American Journal of Political Science* 52(1): 184-200.

[13] Duflo, Esther, Rachel Glennerster, and Michael Kremer (2007), "Using Randomization in Development Economics Research: A Toolkit," in T. Paul Schultz and John A. Strauss, *Handbook of Development Economics, volume 4*, Amsterdam: Elsevier-North Holland.

[14] Fitzgerald, Jennifer, David Lebland, and Jessica C. Teets (2014), "Defying the Law of Gravity: The Political Economy of International Migration," *World Politics* 66(3): 406-445.

[15] Gerber, Alan S., and Donald P. Green (2012), *Field Experiments: Design, Analysis, and Interpretation*, New York: W.W. Norton & Company.

[16] Glennerster, Rachel, and Kudzai Takavarasha (2013), *Running Randomized Evaluations: A Practical Guide*, Princeton, NJ: Princeton University Press.

[17] Haavelmo, Trygve (1943), "The Statistical Implications of a System of Simultaneous Equations," *Econometrica* 11(1): 1-12.

[18] Harrison, Glenn W., and John A. List (2004), "Field Experiments," *Journal of Economic Literature* 42(4): 1009-1055.

[19] Imbens, Guido W. (2014), "Instrumental Variables: An Econometrician's Perspective," Working Paper, National Bureau of Economic Research.

[20] Imbens, Guido W., and Thomas Lemieux (2008), "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics* 142(2): 615-635.

[21] Kelley, Judith G., and Beth A. Simmons (forthcoming), "Politics by Number: Indicators as Social Pressure in International Relations," *American Journal of Political Science*.

[22] Kellogg, Ryan (2014), "The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling," *American Economic Review* 104(6): 1698-1734.

[23] Lehoucq, Fabrice, and Anibal Pérez-Liñán (2014), " Breaking Out of the Coup Trap: Political Competition and Military Coups in Latin America," *Comparative Political Studies* 47(8): 1105-1129.

[24] Nickell, Stephen (1981), "Biases in Dynamic Models with Fixed Effects," *Econometrica* 49(6): 1417-1426.

[25] Pearl, Judea (2009), *Causality, 2nd edition*, Cambridge: Cambridge University Press.

[26] Pearl, Judea (forthcoming), "Trygve Haavelmo and the Emergence of Causal Calculus," *Econometric Theory*.

[27] Reed, W. Robert (2014), "A Note on the Practice of Lagging Variables to Avoid Simultaneity,"Ẇorking Paper, University of Canterbury.

[28] Sekhon, Jasjeet S. (2009), "Opiates for the Matches: Matching Methods for Causal Inference," *Annual Review of Political Science* 12:487–508.

[29] Steinberg, David A., and Krishna Malhotra (2014), "The Effect of Authoritarian Regime Type on Exchange Rate Policy," *World Politics* 66(3): 491-529.

[30] Sovey, Allison J., and Donald P. Green (2011), "Instrumental Variables Estimation in Political Science: A Readers' Guide," *American Journal of Political Science* 55(1): 188-200.

[31] Villas-Boas, J. Miguel, and Russell S. Winer (1999), "Endogeneity in Brand Choice Models," *Marketing Science* 45(10): 1324-1338.

[32] Warren, T. Camber (2014), " Not by the Sword Alone: Soft Power, Mass Media, and the Production of State Sovereignty," *International Organization* 68(1): 111-141.

Appendix 1 Bias in b as Estimated in Equation

In this section, we derive bias in our estimate of β in Model 7 based on the DGPs for Y, X, and U defined in expressions (1 – 4). The derivation is presented below:

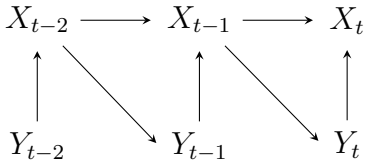
$$\text{plim}_{n\to\infty}\,\hat{b}_{X_t} = \beta + \frac{\text{Cov}(X_t, U_t)}{\text{V}(X_t)} \tag{23}$$

$$= \beta + \delta\frac{\text{Cov}(\rho X_{t-1} + \kappa(W_{it} + \phi W_{t-1} + v_{it}) + \eta_t, W_{it} + \phi W_{t-1} + v_{it})}{\text{V}(X_t)} \tag{24}$$

$$= \beta + \delta\frac{(\rho\phi^2 + \phi^2 + 1)\kappa\sigma_W^2 + \kappa\sigma_v^2 + \sigma_{\eta W}}{\text{V}(X_t)} \tag{25}$$

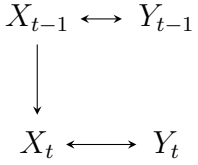
Appendix 2 Valid Data Generating Processes

Figure A1: Scenario 2(b)



Notes: This is a representation of a data generating process in which reverse causality exists ($Y_t \rightarrow X_t$) but the causal effect of X_{t-1} on Y is identified because $X_t \not\rightarrow Y_t$.

Figure A2: Scenario 2(c)



Notes: This is a representation of a data generating process in which reverse causality exists ($Y_t \rightarrow X_t$) but a "conservative" effect of X_{t-1} on Y is identified because $Y_{t-1} \not\rightarrow Y_t$. To be clear, however, *neither* the direct effect of X_t on Y_t *nor* the total effect of X_{t-1} on Y_t is identified because of the simultaneous relationship between X and Y.