Bias in Local Projections

Edward P. Herbst and Benjamin K. Johannsen*

December 1, 2020

Abstract

Local projections (LPs) are a popular tool in applied macroeconomic research. We survey the related literature and find that LPs are often used with very small samples in the time dimension. With small sample sizes, given the high degree of persistence in most macroeconomic data, impulse responses estimated by LPs can be severely biased. This is true even if the right-hand-side variable in the LP is *iid*, or if the data set includes a large cross-section (i.e., panel data). We derive simple expressions for this bias in a variety of settings. As a byproduct, we propose a way to bias-correct LPs. We also show that, in small samples, autocorrelation-robust standard errors can dramatically understate the uncertainty surrounding LP estimators, even when appropriate. Our results suggest, in LP settings like the ones we study, researchers should avoid them. Using U.S. macroeconomic data and identified monetary policy shocks, we demonstrate that the bias in point estimates can be economically meaningful and the bias in standard errors can affect inference.

1 Introduction

We show that if a time series is persistent—as is generally the case when researchers are interested in impulse responses—then estimators of impulse responses by local projections (LPs)

*edward.p.herbst@frb.gov; benjamin.k.johannsen@frb.gov. This does not reflect the views of the Federal Reserve Board or the Federal Reserve System. We thank Todd Clark, Marco Del Negro, Eric Ghysels, Òscar Jòrda, Frank Schorfheide, and Minchul Shin, as well as seminar particpants, for helpful comments and discussions. We thank Nicholas Zevanove for excellent research assistance. Any errors are our own.

can be severely biased in sample sizes commonly found in the empirical macroeconomics literature.

Starting with Jorda (2005), LPs have been used by researchers as an alternative to other time series methods, such as vector autoregressions (VARs). We survey the literature and find that, over the past 15 years, LPs have been applied in a variety of settings that are notably different than the setting studied in Jorda (2005). In particular, we find that sample sizes in the time dimension are typically much smaller than the sample sizes studied in Jorda (2005) and that LPs have become increasingly prevalent when researchers also have a cross section of data (i.e., panel data.) Additionally, researchers often approach LPs with identified structural shocks in hand, rather than identifying those shocks as a part of the estimation.[1] We focus on this idealized case in this paper, as it is a natural benchmark for understanding the methodology.

Using Monte Carlo analysis, we demonstrate that the magnitude of the bias in LPs can be large when sample sizes in the time dimension are similar to those typically found in the empirical macroeconomics literature. Our Monte Carlo simulations use simple, linear data generating processes. While researchers may be drawn to LPs because they invoke fewer parametric restrictions than other methods, an important standard for this methodology is that it performs well in simple scenarios. Notably, we show that the bias in LPs persists even when the shock on the right-hand-side of the regressions in our Monte Carlo analysis is *iid*, as is often the case when researchers have access to a time series of identified structural shocks.

We analyze the small-sample bias in LPs using a higher-order expansion of the LP estimator, building on the related work of Kendall (1954), Rilstone et al. (1996), Anatolyev (2005), and Bao and Ullah (2007). We show that the bias of the LP estimator at horizon h is a function—specifically, a weighted sum—of the (population) impulse response function at other horizons. As a result, if LP estimators across horizons have the same sign (as is the case for hump-shaped impulse responses), then the least-squares estimators are biased toward zero at every horizon. Additionally, our analysis highlights that the small-sample estimates from LPs are not "local" because the small-sample biases of those estimates depend on the true impulse responses at other horizons.

We use the higher-order expansion of the LP estimator to develop a simple bias correction

[1] In what follows, we always refer to the regressor associated with the LP coefficient as the "shock."

that can added to the ordinary least squared (OLS) estimator. In Monte Carlo simulations, on average our bias-corrected estimators are markedly closer to the true values of the impulse responses.[2] In addition, using coverage probabilities as a metric, we show that our estimator performs relatively well when compared to the OLS estimator.

We extend our analysis to settings with panel data or with instrumental variables and show that the bias we document persists. For both the case of instrumental variables and the case of panel data, we provide formulas for bias correcting the OLS estimator. Additionally, we show that increasing the number of entities in panel data cannot eliminate the bias.

Finally, we also study the downward bias in the standard errors of LP estimators. We show that, like the LPs themselves, commonly-used standard errors for LPs that are heteroskedasticity- and autocorrelation-robust (HAR) are also severely downward biased when sample sizes are similar to those commonly found in the literature. We show that the reason for this downward bias is that, while the regression errors of the LP are autocorrelated, the relevant regression scores—the regression residuals times the regressors—for constructing the standard errors are not. We show that, in finite samples, the estimator of the autocorrelation of the regression score is biased downwards. This downward bias is large enough that, in LP settings like the ones we study, estimates of the autocorrelation are, on average, negative in small samples, even when the true autocorrelation is zero or positive. Popular HAR estimators, such as the Newey and West (1987) estimator, rely on these biased estimators of the autocorrelation of the regression score. The possibility of the bias leading to an incorrectly signed estimate of autocorrelations suggests researchers may prefer standard errors that are heteroskedasticity-consistent, but not autocorrelation-robust, such as Huber-White standard errors. In fact, in our empirical examples, switching from Newey-West to Huber-White estimators generally increases the estimates of standard errors, sometimes dramatically so. In the case that researchers need to use autocorrelation-robust standard errors, our small sample analysis suggests that they should be aware that these standard errors may be dramatically biased.

Our analysis of standard errors is related to recent work by Olea and Plagborg-Møller (2020), who suggest that researchers using LPs should use heteroskedasticity-consistent, but

[2]Because our bias correction does not completely eliminate small sample bias, in some settings researchers may prefer methods, such as VARs, that estimate the same impulse responses as LPs (see Plagborg-Møller and Wolf (2019)) and have well-understood, effective methods for bias correction (see Kilian (1998)).

not autocorrelation-robust, standard errors.[3] Our focus on finite sample issues leaves us at similar conclusions, but for different reasons. First, when researchers have the shocks in hand, we argue that under the null hypothesis that is typically of interest, HAR standared errors are not necessary. We provide an analytical characterization of the true autocorrelation function of the regression score in a simple setting. In this case, using HAR standard errors is not conservative; in fact, it will typically understate uncertainty, relative to a standard error calculated without attempting to take autocorrelation into account. Second, we show that even in cases where there is autocorrelation in the regression score, the downward bias is large enough to still lead researchers to avoid HAR estimators.

We analyze bias in three examples drawn from the empirical monetary economics literature. We show that the bias in point estimates can be economically meaningful. In these examples, consistent with our simulations and analytical results, the use of HAR estimators typically leads to smaller standard error estimates than under the Huber-White estimator; in many cases this difference is large enough to affect inference. This is true even when our estimated bias in point estimates is small.

Our paper is related to work by Kilian and Kim (2011), who study the coverage probabilities for confidence intervals for LP estimators using bootstrap methods. Their work focuses on the case when shocks are identified as a part of the LP estimation and uses the block bootstrap to approximate the finite sample distribution of the OLS estimate. By contrast, our paper considers the case when a time series of identified shocks is available to the researcher, so right-hand-side variables are *iid*. Our analysis relies on higher-order expansions of the OLS estimator, which illustrate the reasons that the LP estimator is biased and provide a natural bias correction without bootstrapping. In addition, we extend the analysis to panel data and instrumental variables settings, which are common settings for LPs in practice. In Appendix F, we compare the bias correction from the block bootstrap used in Kilian and Kim (2011) to our proposed bias correction, and use insights from our higher-order expansion of the OLS estimator to explain why the bootstrap performs relatively poorly unless block lengths are relatively long. More generally, our paper is related to work on bias in least-squares estimators of autocorrelation (such as Kendall (1954) and Shaman and Stine (1988)), in dynamic panel data settings (such as Nickell (1981) and Hahn and Kuersteiner (2002)), and in generalized method of moments systems (Rilstone et al. (1996), Anatolyev

[3]We analyze the effects of lag augmentation in our setup in Appendix G.

(2005), Bao and Ullah (2007)). We extend this work to our LP setting.

2 Some evidence on the use of LPs

To get a sense of how LPs are used in the literature, we examine the 100 "most relevant" papers citing Jorda (2005) on Google Scholar.[4] Google scholar's relevance ranking weights the text of the document, the authors, the source of the publication, and the number of citations. Of these 100 papers, 71 employed LPs in an empirical project (rather than merely citing but not applying LP).[5]

The focus of this paper is parameter bias associated with short time series, so for each of the studies we recorded the length of the time series, T, in the main LP in each of these papers. About two-thirds of the papers surveyed employed panel data. As mentioned in the introduction and discussed later, with entity-specific fixed effects, the time dimension is still the relevant component of the sample size for determining the LP bias. Because many of the panel data sets are unbalanced, constructing a single summary T is challenging. For unbalanced panels, we summarize the size of the time dimension using the mean T across entities, when readily available, or using the largest value of T across entities. In general, our assessment of T is extremely conservative in the sense that it overestimates the time series dimension of the data for many of the LP applications. It is not unusual, for example, to see unbalanced panels that have an average T that is less than half of the time-series dimension of the entire panel, or to see robustness exercises that use a small fraction of the data series. In these cases, we use the entire time series dimension of the panel, which biases our estimates of T up.

Figure 1 displays a histogram of the sample of 71 Ts collected in our literature review. The median T (the red dash dotted line) is around 95. These sample sizes are notably less than those typically used in many empirical macroeconometrics papers, as most of the papers surveyed here use the increasingly popular strategy of using observed shocks, such as the monetary policy shocks of Romer and Romer (2004), rather than identified shocks from a VAR, as in Jorda (2005). Constructing these observed shocks is often difficult and costly,

[4] We conducted this search in October 2019. See Appendix I for the list of citations.

[5] If a paper appeared as both a working paper and a published paper, we excluded the working paper version from our analysis.

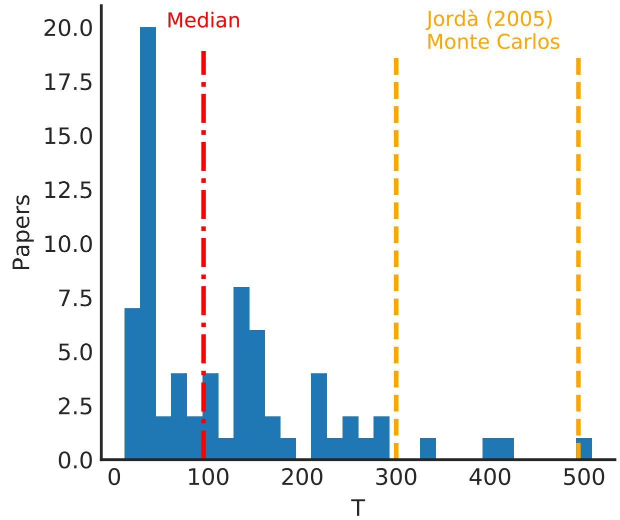


Figure 1: T is small in the literature using LPs.

and so the time series typically have short length.

The application of LPs to such short time series does not seem to have been anticipated in the early literature on LPs. In fact, the Monte Carlo study in Jorda (2005) used $T = 300$ and $T = 496$ (the orange, dashed lines in Figure 1). Less than 6 percent of the surveyed studies use sample sizes at least that large. Of course, it is difficult to fault Jorda (2005) for not anticipating how researchers would subsequentyly apply LP meathods. While many studies in our survey use monthly or quarterly data, Jorda (2005) used monthly data. In general, increasing T by using monthly data rather than quarterly or annual data will not avoid the issue of small-sample bias in LPs because the monthly series are likely to be more persistent, and the bias in LPs is more-severe when the data are more persistent.

3 Bias in LPs

In this section, we demonstrate that LPs can be severely biased with sample sizes that are similar to those documented in Section 2. We explore this bias using both Monte Carlo evidence and a new analytic approximation of the bias. The analytic approximation yields insights into the bias associated with LPs at different horizons, and suggests a correction of the bias.

3.1 Bias in LPs using an AR(1) example

To demonstrate that LPs can be severely biased in small samples, we first consider Monte Carlo analysis using an AR(1) data generating process. Our objective is to study the accuracy of estimated impulse responses via LPs for various sample sizes. For a given T, we simulate $N_{mc} = 10,000$ time series, $\{y_t\}_{t=1}^{T}$, for the data generating process:

$$y_t = \rho y_{t-1} + \varepsilon_t + \nu_t \quad \text{and} \tag{1}$$

Here, ε_t and ν_t are *iid* standard normal random variables. In these Monte Carlo simulations, to be consistent with the high persistence of macroeconomic data, we set $\rho = 0.95$.[6] We use the AR(1) time series model because LPs were designed to capture the dynamics of a wide range of data generating processes and one would hope that they would perform well in the simplest examples.

We assume that the researcher does not know the true data generating process, but is otherwise in a near-ideal setting for estimating the impulse response function of y_t using LPs. The researcher observes $\{y_t, \varepsilon_t\}_{t=1}^{T}$; that is, the researcher directly observes the shock ε_t which is independent over time and uncorrelated with past values of y_t. In addition, the researcher may like to control for other variables, denoted by the vector c_t. When we include such controls, we assume $c_t = y_{t-1}$. We stress that our regressions with controls are ideal in the sense that no useful additional information from earlier periods could be added to the regressors and we include the correct number of lags of y_t as controls.

The LP model is the set of regression models, indexed by the impulse response horizon h,

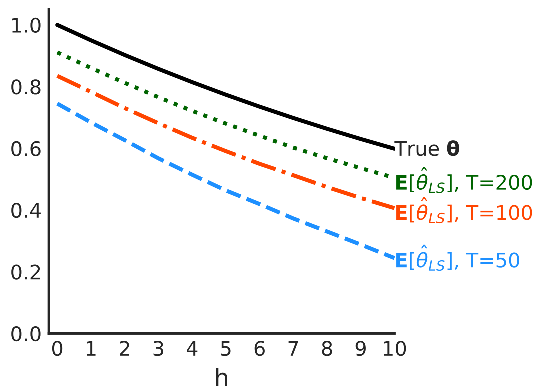
$$y_{t+h} = \alpha_h + \beta_h' x_t + u_{t,h}, \quad h = 0, \ldots, H. \tag{2}$$

where $x_t \equiv [\varepsilon_t, c_t']'$.[7] Thus, the first elements of the coefficient vectors $\{\beta_h\}_{h=0}^{H}$ trace out the impulse response of interest. We denote the $H+1$ vector describing the impulse response by θ with elements θ_h for $h = 0, \ldots, H$. As in the empirical macroeconomics literature, we estimate each β_h using least squares. We denote the estimator of the β_h by $\widehat{\beta}_{h,LS}$ and the estimator of the impulse response by $\widehat{\theta}_{LS}$.

[6] In Appendix C, we provide results for alternative values of ρ. In each simulation, we initialize y_0 at a draw from the unconditional distribution of y_t.

[7] When we do not include controls, $x_t = \varepsilon_t$.

(a) Without Controls



(b) With Controls

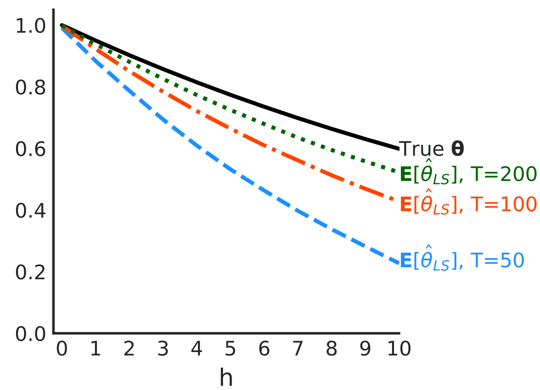


Figure 2: LP estimators are biased in empirically-relevant samples when y_t is an AR(1) with $\rho = 0.95$.

Using Monte Carlo simulations, we can compute, for any T, the finite sample expectation of the least-squares estimator, $\mathbb{E}[\widehat{\theta}_{LS}]$. Figure 2 displays the expectation for the LP estimators with and without controls for the AR(1) data generating process with $T \in \{50, 100, 200\}$. Recall that about half of the surveyed literature uses T less than 100.

When controls are not included in the LP (the left panel), the estimator is biased even at short horizons. This is true even for moderately long time series—i.e., $T = 200$. As the horizon of the impulse response increases, the bias becomes worse. When y_{t-1} is included as a control (the right panel), the bias diminishes substantially at short horizons. Intuitively, adding controls makes the least-squares error terms less correlated at short horizons. However, even for the impulse response only 10 periods ahead (2.5 years with quarterly data), the controls alleviate only a small fraction of the bias. The reason that controls are less effective at reducing the bias as h increases is that they are less effective at forecasting y_{t+h}. Note also that, at these longer horizons, the overlapping nature of the left-hand-side variables in the LP implies that the error terms are autocorrelated.[8]

[8] Following Jorda (2005), for each regression of horizon h, researchers typically use all available data, meaning that the regression error term is autocorrelated for at least $h - 1$ periods. With autocorrelated regression error terms, the generalized least-squares estimator asymptotically performs better than the ordinary least-squares estimator. However, researchers typically use the ordinary least-squares estimator because of the small-sample shortcomings of the feasible generalized least-squares estimator. That said, recent work by Lusompa (2019) suggests that well-behaved small sample GLS estimators may be obtained for LPs.

3.2 Approximating the small sample bias in LPs

In this subsection we derive an analytic approximation to the bias of the least-squares estimators, $\widehat{\theta}_{LS}$, given by $\mathbb{E}[\widehat{\theta}_{LS}]-\theta$. The expressions we derive are, to the best of our knowledge, new to the literature, and highlight the interdependence of the bias in LPs at different horizons. In order to illustrate the point clearly (and to avoid tedious matrix algebra), we first focus on LPs without controls. We generalize our analytic approximation to the case when controls are included in the LP, and the intuition for the bias is similar (we provide the associated derivations in our Appendix A).

In all of our derivations, we make the following assumptions. Let $w_t = [y_t, \varepsilon_t, c_t']'$.

Assumption 1. *The data $\{w_t\}$ is stationary and ergodic. The demeaned series $\{w_t - \mu_w\}$, where $\mu_w = E[w_t]$, is purely non-deterministic with absolutely summable Wold decomposition coefficients.*

Under Assumption 1, the Wold representation can be inverted and the data has a VAR(∞) representation. If $\{w_t\}$ is jointly Gaussian, the data satisfies the Assumption 1.

Assumption 2. $\mathbb{E}[\varepsilon_t|\{w_\tau\}_{\tau<t}, \{\varepsilon_\tau\}_{t<\tau}] = E[\varepsilon_t] = \mu_\varepsilon$ *and* $\mathbb{E}[(\varepsilon_t - \mu_\varepsilon)^2|\{w_\tau\}_{\tau<t}, \{\varepsilon_\tau\}_{t<\tau}] = \mathbb{E}[(\varepsilon_t - \mu_\varepsilon)^2] = \sigma_\varepsilon^2 > 0.$

This assumption encompasses idealized case where the researcher has *iid* shocks in hand. Such a stark assumption is consistent with the emerging practice of "constructing" such shocks with desirable statistical properties. It allows us to obtain sharp analytical results, a necessary first step in understanding finite sample issues in LPs.

Recall that in our framework, the researcher observes $\{y_t, \varepsilon_t\}_{t=1}^T$. For a given h, the ordinary least-squares estimator of θ_h can be written as

$$\widehat{\theta}_{h,LS} = \frac{\frac{1}{T-h}\sum_{t=1}^{T-h}\varepsilon_t y_{t+h} - \frac{1}{(T-h)^2}\left(\sum_{t=1}^{T-h}\varepsilon_t\right)\left(\sum_{t=1}^{T-h}y_{t+h}\right)}{\frac{1}{T-h}\sum_{t=1}^{T-h}\varepsilon_t^2 - \frac{1}{(T-h)^2}\left(\sum_{t=1}^{T-h}\varepsilon_t\right)^2} = \frac{\widehat{\text{cov}}[\varepsilon_t, y_{t+h}]}{\widehat{\text{var}}[\varepsilon_t]}. \tag{3}$$

Here, $\widehat{\text{cov}}$ and $\widehat{\text{var}}$ are the sample covariance and variance, respectively.[9] Equation (3) makes it clear that θ_h in population is the scaled covariance between y_{t+h} and ε_t.

[9]Note that for the variance to be consistent with the least-squares estimator, we use only the $T-h$ observations of ε_t.

Analytical Result 1 (Expression for the bias in an LP without controls)**.** *Under Assumptions 1 and 2, the bias of the LP estimator in (3) is given by:*

$$\mathbb{E}\left[\widehat{\theta}_{h,LS}\right] - \theta_h = -\frac{1}{T-h}\sum_{j=1}^{T-h-1}\left(1-\frac{j}{T-h}\right)(\theta_{h+j}+\theta_{h-j}) + O\left(T^{-3/2}\right). \tag{4}$$

A proof of this claim is given in Appendix A, but we provide some discussion here. To derive an expression for the approximate bias of $\widehat{\theta}_{h,LS}$, we need to compute $\mathbb{E}\left[\widehat{\theta}_{h,LS}\right]$. In such a simple example, one can calculate higher-order expansions of (3). This is a widely adopted approach to these kinds of problems. We apply the methodology of Bao and Ullah (2007), which does not require, for example, that the shocks be normally distributed. Finally, note that Equation (4) gives an approximate expression for the bias of $\widehat{\theta}_{h,LS}$ that consists of only the impulse response coefficients at different horizons.[10]

To gain intuition for equation (4), it is useful to consider the case when ε_t is symmetric around the origin.[11] In this case,

$$\mathbb{E}\left[\widehat{\theta}_{h,LS}\right] = \frac{\mathbb{E}[\widehat{\text{cov}}[\varepsilon_t, y_{t+h}]]}{E[\widehat{\text{var}}[\varepsilon_t]]} + O\left(T^{-3/2}\right). \tag{5}$$

The numerator of equation (5) is given by

$$\begin{aligned}\mathbb{E}[\widehat{\text{cov}}[y_{t+h}, \varepsilon_t]] &= \left(1-\frac{1}{T-h}\right)\text{cov}[y_{t+h}, \varepsilon_t] \\ &\quad - \frac{1}{T-h}\sum_{j=1}^{T-h-1}\left(1-\frac{j}{T-h}\right)(\text{cov}[y_{t+h+j}, \varepsilon_t] + \text{cov}[y_{t+h-j}, \varepsilon_t]). \end{aligned} \tag{6}$$

Here, cov is the true covariance. The summation in equation (6) comes from the plug-in estimator for the mean in $\widehat{\text{cov}}$. The denominator of equation (5) is

$$\mathbb{E}[\widehat{\text{var}}[\varepsilon_t]] = \left(1-\frac{1}{T-h}\right)\text{var}[\varepsilon_t]. \tag{7}$$

Here, var is the true variance. Noting that $\theta_{h+j} = \text{cov}[\varepsilon_t, y_{t+h+j}]/\text{var}[\varepsilon_t]$, we obtain equation (4).[12] From this derivation, it is clear that the bias arrises because of the need to estimate the means of the variables in the OLS calculations.

[10]Given our maintained assumption that ε_t is uncorrelated with past values of y_t, $\theta_{-j} = 0$ for $j > 0$.

[11]In this case, the derivation of our expression for the bias of $\widehat{\theta}_{h,LS}$ is similar to the derivation in Kendall (1954) for the well-known bias of estimators of autocorrelation—see also Shaman and Stine (1988).

[12]In this derivation, the expression for the bias would be divided by $T-h-1$ rather than $T-h$, but that difference is absorbed in the term $O(T^{-3/2})$.

Several remarks regarding equation (4) are in order. First, as expected, the bias is a decreasing function of T; for fixed h, the least-squares estimator is consistent. Second, the bias of $\widehat{\theta}_{h,LS}$ is a function of the impulse response *all* other horizons. Intuitively, the data generating process affects the bias of OLS estimators at similar horizons in similar ways. The interdependence of LP estimates across h highlights that, in finite samples, LPs are not "local." Third, the contribution of the horizon $h+j$ impulse response coefficients to the bias in the least-squares estimate of the h impulse response coefficient decreases only at linear rate as j increases or decreases. In practice, this means that the bias in the portion of the impulse response of interest—typically, say, the first 20 periods in quarterly macroeconomic applications—can be meaningfully affected by the impulse response at much longer horizons. This is especially true for extremely persistent time series (like many macroeconomic series).

We next consider the case when a researcher includes controls in the LP. We maintain the following assumptions.

Assumption 3. *$\alpha_h + x_t'\beta_h$ is an optimal linear forecast for y_{t+h}.*

The assumption that the LP produces an optimal linear forecast implies that, using the true parameters, forecast errors are MA(h+1) processes.

Assumption 4. *The matrix $\mathbb{E}[||x_t x_t'||]$ is full rank.*

This assumption ensures our estimator is consistent and that our setup satisfies the assumptions of Rilstone et al. (1996). With these additional assumptions, we can state our next analytical result.

Analytical Result 2 (Expression for the bias in an LP with controls)**.** *Under Assumptions 1-4, the bias for the LP with controls is given by:*

$$\mathbb{E}\left[\widehat{\theta}_{h,LS}\right] - \theta_h = -\frac{1}{T-h}\sum_{j=1}^{h}\left(1 - \frac{j}{T-h}\right)\left(1 + trace\left\{\Sigma_{c,0}^{-1}\Sigma_{c,j}\right\}\right)\theta_{h-j} + O\left(T^{-3/2}\right), \quad (8)$$

where $\Sigma_{c,j} \equiv E\left[\left(c_{t-j} - \mu_c\right)\left(c_t - \mu_c\right)'\right]$.

The claim again relies on the results of Bao and Ullah (2007). A detailed derivation is in Appendix A. Several remarks regarding equation (6.1) are in order. First, as in the case without controls, the bias is a decreasing function of T; for fixed h, the least-squares estimator is consistent. Second, the bias of $\widehat{\theta}_{h,LS}$ is a function of the impulse response *all*

(a) Without Controls

(b) With Controls

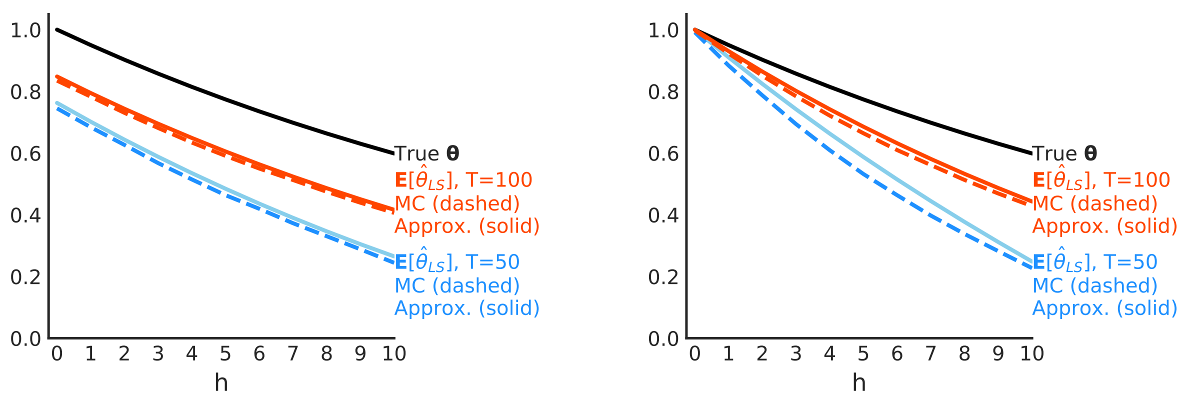


Figure 3: The bias approximation is accurate in our LPs.

horizons up to h. Intuitively, the data generating process affects the bias of OLS estimators at similar horizons in similar ways, however controlling for past data truncates the terms that are important for the bias by truncating the correlation in the regression errors. Third, the contribution of the horizon $h-j$ impulse response coefficients to the bias in the least-squares estimate of the h impulse response coefficient is scaled by $1+\text{trace}\left\{\Sigma_{c,0}^{-1}\Sigma_{c,j}\right\}$. As a result, when controls are persistent, or when unneeded persistent controls are added to the LP, the bias increases.

Using our AR(1) example, figure 3 shows $\mathbb{E}\left[\widehat{\theta}_{LS}\right]$ calculated using the approximation in equation (5), assuming that the true values of θ_h are known. The figure also shows the exact finite sample value from Monte Carlo simulations. Notably, the approximation works quite well in population. For the no-controls case, the analytic approximation is nearly exact for $T \in \{50, 100, 200\}$. With controls, the analytic approximation to the impulse response is somewhat above the true finite-sample expectation, though it still captures most of the bias associated with the least-squares estimator.

3.3 A bias-corrected estimator

Equations 4 and 6.1 lend themselves to constructing bias corrected estimators for θ_h, using plug-in estimators for θ_j and $\Sigma_{c,j}$. In the case of no controls, the bias depends on the values of θ_j for all $j \neq h$ and $|j| \leq T$. Given the inability to estimate all of these parameters with a sample size of T, in practice a researcher could truncate the horizon of the coefficients

(a) $T = 50$ (b) $T = 100$

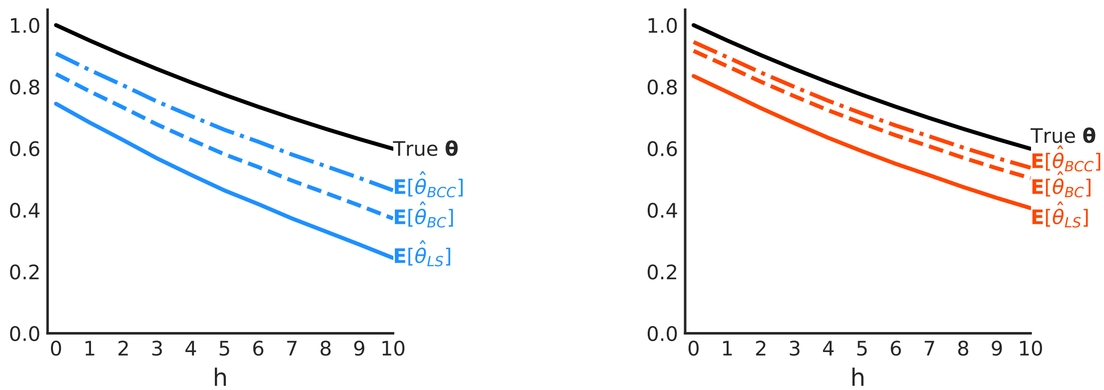


Figure 4: $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are closer than $\widehat{\theta}_{LS}$ to θ, on average, in our LPs without controls when y_t is an AR(1) with $\rho = 0.95$.

used in the bias correction. Without theory on how to optimally pick the maximum horizon used in the bias correction, H, we set H to 20, 25, and 50 for T equal to 50, 100, and 200, respectively. In the case when controls are included in the LP, all of the needed values of θ_j and $\Sigma_{c,j}$ are easily computed.

Notably, the researcher could bias correct the coefficients using the OLS estimates of θ_j. When we construct bias-corrected estimator in this way, we denote the estimator as $\widehat{\theta}_{BC,h}$. Alternatively, the researcher could iterate the bias correction on all values of θ_j. When we construct the bias-corrected estimator in this way, we denote the estimator as $\widehat{\theta}_{BCC,h}$.

For the case of an LP without controls and an LP with controls, respectively, figures 4 and 5 show the average value of $\widehat{\theta}_{LS}$, $\widehat{\theta}_{BC}$, and $\widehat{\theta}_{BCC}$ over our Monte Carlo simulations when y_t follows an AR(1) with $\rho = 0.95$. Clearly, our bias correction does not completely correct for the bias in $\widehat{\theta}_{LS}$ in either case, indicating that our bias-corrected estimator is not a panacea for bias in LPs. Nevertheless, $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are markedly closer than $\widehat{\theta}_{LS}$ to θ on average.

4 Extensions to instrumental variables and panel data

In this section, we extend our analysis of bias in LPs to settings where researchers use instrumental variables or a cross section of data (panel data). In both of these settings, we

(a) $T = 50$ (b) $T = 100$

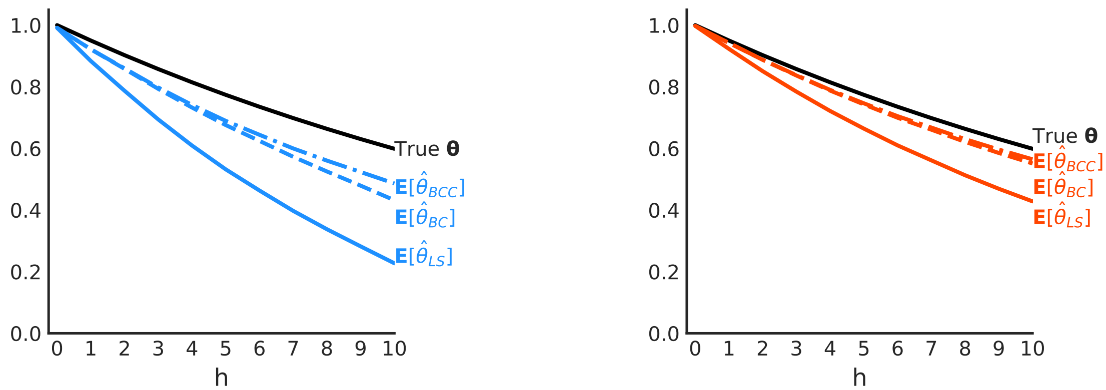


Figure 5: $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are closer than $\widehat{\theta}_{LS}$ to θ, on average, in our LPs with controls when y_t is an AR(1) with $\rho = 0.95$.

find that the bias we document in the previous section persists.

4.1 Bias in LPs with instrumental variables

A number of recent papers have employed LPs with instrumental variables (see, for example, Jordà et al. (2015), Ramey and Zubairy (2018), and Stock and Watson (2018)). Here, we derive the bias for LPs with instrumental variables.[13] Throughout this section, we maintain the assumptions from section 3.2. We show that the bias we analyzed in the case without instrumntal variables persists when instrumental variables are used. Additionally, Equations (11) and (22), which we derive, can be used to bias correct the instrumental-variables estimator of θ_h in an analogous way to our bias corrections of the OLS estimator.

4.1.1 With controls

In the case of instrumental variables with controls, a researcher brings a vector of instruments, z_t, to the estimation. We assume that z_t is a vector of valid instruments that is of the same length as x_t. The researcher estimates β_h by solving a generalized method of moments system

[13] We do not consider the case of weak instruments. See Ganics et al. (2019) for a discussion of asymptotic bias in the case of weak instruments.

where the moment conditions are given by

$$E\left[\begin{array}{c} y_{t+h}-\alpha_h-x_t'\beta_h \\ z_t\left(y_{t+h}-\alpha_h-x_t'\beta_h\right) \end{array}\right]=0. \tag{9}$$

We denote the instrumental variables estimator of θ_h as $\widehat{\theta}_{h,IV}$. In Appendix A.3, we show that

$$\mathbb{E}\left[\widehat{\theta}_{h,IV}\right]-\theta_h=-\frac{1}{T-h}\sum_{j=1}^{h}\left(1-\frac{j}{T-h}\right)\left(1+\text{trace}\left\{\Sigma_{dc,0}^{-1}\Sigma_{dc,j}\right\}\right)\theta_{h-j}+O\left(T^{-3/2}\right), \tag{10}$$

where $\Sigma_{dc,j}\equiv E\left[\left(d_{t-j}-\mu_d\right)\left(c_t'-\mu_c'\right)\right]$. This expression is analogous to equation (6.1). As a result, the bias of LPs with controls that we analyzed earlier persists in settings with instrumental variables.

4.1.2 Without controls

Here, we consider the case where a researcher runs an LP without controls using instrumental variables. Following the methodology used with controls, it follows that, for a given h, the finite-sample properties of the instrumental-variables estimator of θ_h can be approximated as

$$\mathbb{E}\left[\widehat{\theta}_{h,IV}\right]-\theta_h=-\frac{1}{T-h}\sum_{j=1}^{T-h-1}\left(1-\frac{j}{T-h}\right)\left(\theta_{h+j}+\theta_{h-j}\right)+O\left(T^{-3/2}\right), \tag{11}$$

which is identical to equation (4). As a result, the bias of LPs withough controls that we analyzed earlier persists in settings with instrumental variables.

4.2 Panel data

In this subsection, we demonstrate that LPs can be severely biased with sample sizes in the time dimension commonly found in the empirical macroeconomic literature even when researchers have access to a large cross-section (i.e. panel data). Of course, parameter bias in dynamic panel data models has been studied since at least Nickell (1981). We illustrate the bias in LPs using the expansion of the OLS estimator that is similar to one in the previous section. As before, the bias for an LP at horizon h is linked directly to the LP population coefficients at other horizons. In all of our derivations, we maintain the assumptions from the previou section, and, for algebraic simplicity, we assume that the panel is balanced.

4.2.1 Bias in LPs with panel data using the AR(1) example

To demonstrate that LPs can be severely biased in small samples with panel data, we generate data, $\{y_{i,t}\}_{t=1}^{T}$ for each entity $i = 1, \ldots, I$, using the data generating process specified in equations (1). For simplicity, we assume that all the data are independent across entities, but our derivations for the approximate bias do not depend on this assumption. We show results for panels containing $I = 10, 25,$ and 50 entities. As in the previous section we assume $\rho = 0.95$.

In the panel settings, the LP model is the set of regression models, indexed by the impulse response horizon h,

$$y_{i,t+h} = \alpha_{i,h} + \beta_h' x_{i,t} + u_{i,t,h}, \quad h = 0, \ldots, H. \tag{12}$$

where $x_{i,t} \equiv [\varepsilon_{i,t}, c_{i,t}']'$.[7] The first element of the coefficient vectors $\{\beta_h\}_{h=0}^{H}$ trace out the impulse response of interest, which we denote $\{\theta_h\}_{h=1}^{H}$.

Figure 6 displays the Monte-Carlo mean of LP estimators with and without controls for $T = 100$ and different numbers of entitites in the panels. As was the case without panel data, the LP estimates of the impulse responses are severely biased. Notably, the bias does not approach zero as the number of entities in the panel grows large (see Nickell (1981)).

As in the non-panel setting, the inclusion of controls is less effective at reducing the bias in LP estimators as h increases. The reason that controls are less effective at reducing the bias as h increases is that they are less effective at forecasting $y_{i,t+h}$. Thus, for large h, the bias is similar to the LP estimator without controls. Papers like Acemoglu et al. (2019) have argued that T as small as 40 should make the bias in panel LP estimators relatively small. While the bias documented by Nickell (1981) is small at $h = 1$ when controls are included, that bias can be large at $h = 10$ even when T is relatively large. In general, impulse responses are most of interest at moderate-to-large values of h.

4.2.2 Understanding bias in LPs with panel data

In this subsection we derive an approximate bias function for the LP estimator in the context of panel data with entity fixed effects. We do our expansions under the assumption that the time series is growing, but that the number of entities in the panel is constant. This seems like the most relevant setting for macroeconomic applications, where panels generally

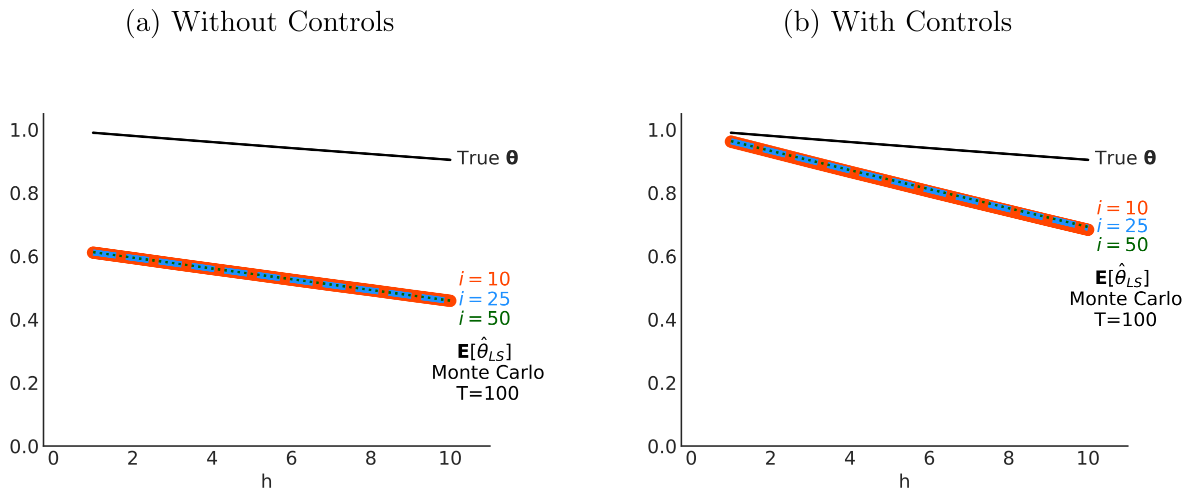


Figure 6: LP estimators without controls are biased in empirically-relevant samples when $y_{i,t}$ is an AR(1) with $\rho = 0.99$.

consists of countries, states, or even counties. We analyze the implications of having a larger number of entities in a panel for the size of the bias.

Throughout this sub-section, we maintain the following assumption.

Assumption 5. *Either* $\varepsilon_{i,t} \perp \varepsilon_{j,t}$ *for* $j \neq i$ *or if* $\varepsilon_{i,t}$ *and* $\varepsilon_{j,t}$ *are correlated then* $E[\varepsilon_{i,t}(y_{j,t+h} - \alpha_{j,h} - \beta_h \varepsilon_{j,t})] = 0$.

This assumption says that either the shocks are independent or they are valid instruments for one another.

Analytical Result 3 (Bias in Panel LPs without controls.)**.** *Under the maintained assumptions from this and the previous sections, the series expansion of the OLS estimator without controls implies*

$$\mathbb{E}\left[\widehat{\theta}_{h,LS}\right] - \theta_h = -\frac{1}{T-h}\sum_{j=1}^{T-h-1}\left(1 - \frac{j}{T-h}\right)(\theta_{h+j} + \theta_{h-j}) + O\left(T^{-3/2}\right). \tag{13}$$

Equation (13) is identical to (4), meaning that the expression for the bias in a panel setting is identical to the expression without panel data.[14] As a result, without controls the bias-corrected estimator from the previous section could be applied to panel data setup we analyze here, where T is used as the number of relevant observations rather than $I \times T$.

When controls are included, the bias calculation becomes more complicated, although the intuition about the size and scope of bias is essentially unchanged. Define covariance of the εs between panelists as

$$\sigma_{j,k} = E\left[\left(\varepsilon_{j,t} - \mu_{j,\varepsilon}\right)\left(\varepsilon_{k,t} - \mu_{k,\varepsilon}\right)\right].$$

Next, write the covariance of $t-u$ and t controls of panelists j and k as

$$\Sigma_{j,k,u} = E\left[\left(c_{j,t-u} - \mu_{c,j}\right)\left(c'_{k,t} - \mu'_{k,z}\right)\right].$$

Then the average of the *variance* of the controls is given by

$$\bar{\Sigma}_0 = \frac{1}{I}\sum_{j=1}^{I} \Sigma_{j,j,c,0}.$$

Analytical Result 4 (Bias in Panel LPs with controls.)**.** *Under the maintained assumptions from this and the previous sections, the series expansion of the OLS estimator with controls implies*

$$E\left[\widehat{\theta}_{h,LS}\right] - \theta_h = -\frac{1}{T-h}\sum_{u=1}^{h} b_u \left[1 + \frac{1}{I}\sum_{i=1}^{I}\frac{1}{I}\sum_{k=1}^{I} trace\left\{\bar{\Sigma}_0^{-1}\Sigma_{k,i,c,u}\right\}\frac{\sigma_{i,k,\varepsilon}}{\frac{1}{I}\sum_{j=1}^{I}\sigma_{j,j,\varepsilon}}\right]\theta_{h-u} + O\left(T^{-3/2}\right) \tag{14}$$

where

$$b_u \equiv 1 - \frac{u}{T-h} \text{ and } \Sigma_{j,k,c,u} \equiv E\left[\left(c_{j,t-u} - \mu_{c,j}\right)\left(c'_{k,t} - \mu'_{k,z}\right)\right].$$

[14] The moment conditions with and without controls are given by

$$E\begin{bmatrix} \left(y_{1,t+h} - \alpha_{1,t} - \beta x_{1,t}\right) \\ \left(y_{2,t+h} - \alpha_{2,t} - \beta x_{2,t}\right) \\ \vdots \\ \left(y_{I,t+h} - \alpha_{I,t} - \beta x_{I,t}\right) \\ \frac{1}{I}\sum_{i=1}^{I} x_{i,t}\left(y_{i,t+h} - \alpha_{i,t} - \beta x_{i,t}\right) \end{bmatrix} = 0.$$

A few comments are in order regarding equation (14). First, as was the case without panel data and because the OLS estimator is consistent, the bias goes to zero as the sample size goes to infinity. Second, the cross-autocovariance of the control variables plays a role in the bias. Notably, if the controls are not correlated, the bias is smaller than if they are correlated. Third, as was the case without panel data, if the controls are positively autocorrelated, or if unnecessary positively autocorrelated controls are include in the LP, the bias is larger. Fourth, even with controls that are independent across entities or over time, the bias does not go to zero as the number of panelists increases.

5 Beyond point estimation: bias in standard errors

Since Jorda (2005), the conventional wisdom has been that heteroskedasticity and autocorrelation robust (HAR) standard errors are necessary because the regression residuals of LPs are autocorrelated. That is the reason that most practioners use the HAR standard errors of Newey-West or more recent ones detailed in Sun (2014) and Lazarus et al. (2018). However, under Assumption 2 in the LP with controls the *regression score*—the product of the ε_t and the regression residuals—is serially uncorrelated.[15] Thus, in large samples HAR standard errors are not necessary; instead, Huber-White (heteroskedasticity-robust) standard errors are valid.

In an LP without controls, under the AR(1) DGP, one can show that the autocovariance function of the regression score, $r_t = \varepsilon_t(y_{t+h} - \theta_h \varepsilon_t)$, is given by

$$\operatorname{cov}[r_t, r_{t-\tau}] = \begin{cases} \rho^{2h} E[\varepsilon_t^2]^2 & \tau = 0, \ldots, h \\ 0 & \tau > h. \end{cases} \tag{15}$$

Thus, unless ε_t is uncorrelated with y_t, the regression score will be serially correlated, as suggested by the early LP literature. However, in an LP without controls, researchers are generally interested in rejecting the null hypothesis that θ_h is zero. Under the null hypothesis that the $\theta_h = 0$ for all h, the regression score is uncorrelated in population, and Huber-White standard errors are asymptotically valid for the purposes of hypothesis testing.

[15] Olea and Plagborg-Møller (2020) use lag augmentation to achieve (population) residualized regressors, whereas our setup does not require this step because of Assumption 2.

HAR standard errors are often justified by researchers as being somehow more conservative than standard errors that do not control for autocorrelation. While HAR standard errors are asymptotically valid even when Huber-White will do, using popular HAR estimators, like the Newey-West estimator, can dramatically *reduce* the size of the estimated standard errors. The reason is that these estimators rely on estimators of the long-run variance (LRV) of the regression score that are functions of the autocovariance of the regression score. In a linear regression like an LP, standard errors are typically a function of the LRV and an estimator of $E[x_t x_t']$. Because we assume that $\varepsilon_t \perp c_t$, it is enough to focus on the estimator of the LRV to analyze the size of the standard errors. In finite samples, estimates of the autocovariance of the regression score are *downward biased*. When the actual autocorrelation function is zero, the bias in the autocorrelation function makes it *negative*, on average, in finite samples. We derive a simple expression for this downward bias in the case of an LP without controls in Appendix B.

In the case of LP with controls, where the autocovariance function of the regression score is zero in population, the downward bias means that increasing the bandwidth of a Newey-West estimator, on average, *reduces* the size of the estimated LRV. In the case of LP without controls, the downward bias means increasing the bandwidth reduces the size of the standard errors, on average, even though the regression score is autocorrelated. Thus, HAR standard errors based on estimated autocovariances of the regression score will tend to *underestimate the uncertainty* associated with $\widehat{\theta}_{h,LS}$.

To explore the effects of the downward bias in standard errors, we consider Monte Carlo evidence using our AR(1) example. Figure 7 shows the true values of the LRV for the regression score in an LP without controls (panels (a) and (b)) and in an LP with controls (panels (c) and (d)). The horizon of the LP is $h = 5$. The figures also show the Monte Carlo average of Newey-West estimators that use a bandwidth u. The left panels (panels (a) and (c)) show the estimators in the case when $\theta_h = 0$ for all h—that is, the shock is noise uncorrelated with y_t—which is typically the maintained null hypothesis to reject. The right panels (panels (b) and (d)) show the estimators in the case when the shock is correlated with y_t, so $\theta_h = \rho^h$.[16]

Several features of these figures are woth noting. First, in all cases the Newey-West estimators dramatically under-estimate the long-run variance. Second, increasing the band-

[16]We keep the variance of y_t fixed across all of our specifications.

(a) Without controls, $\theta_h = 0$

(b) Without controls, $\theta_h = \rho^h$

(c) With controls, $\theta_h = 0$

(d) With controls, $\theta_h = \rho^h$

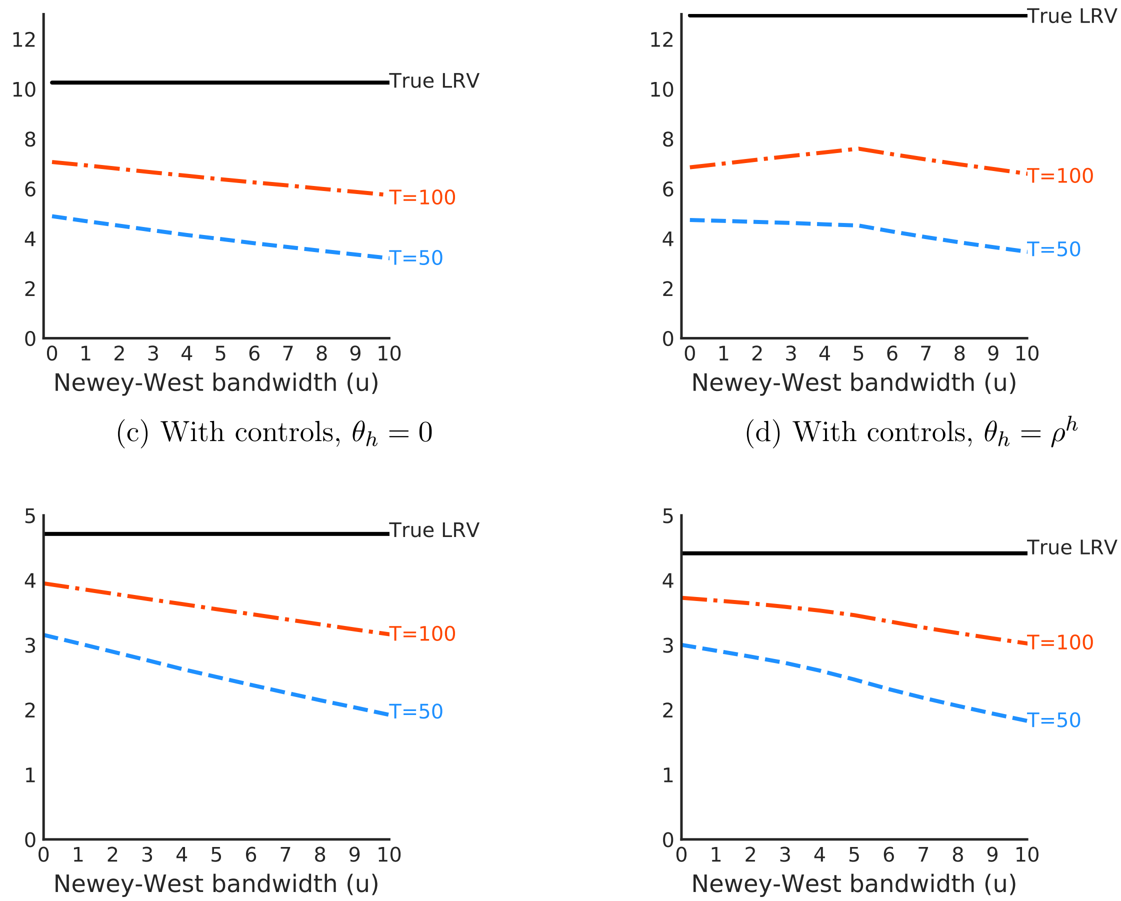


Figure 7: Estimators of standard errors in an LP with $h = 5$ are biased in empirically-relevant samples when y_t is an AR(1) with $\rho = 0.95$.

width of the Newey-West estimator tends to reduce the estimate of the LRV. Third, even when there is autocorrelation in the population regression score (panel (b)), in empirically relevant sample sizes, increasing the bandwidth of the the Newey-West estimator does little to improve the estimate of the long-run variance and can reduce that estimate. Fourth, the Huber-White estimators, which are given by the Newey-West estimator with $u = 0$, are also downward biased. However, the Huber-White estimator is not further attenated by the estimates of the autocovariances of the regression score included in the Newey-West estimator.

Often, standard errors are interpreted as providing a credible region around the estimated impulse response without reference to a null hypothesis. To this end, it is also useful to consider the implications of bias in point estimates and standard errors under the maintained assumption that $\theta_h \neq 0$. We analyze this issue using Monte Carlo analysis and an AR(1) generating process for y_t. The OLS estimator of the impulse response function is biased down and typical estimators of standard errors are also biased down. Both sources of bias may make confidence intervals fail to contain the true impulse response function.

Using Huber-White and Newey-West standard errors, Table 1 displays the percentage of confidence sets with 95% nominal coverage probability (based on the asymptotic normal approximation) that contain the true impulse response function at horizon h in a Monte Carlo simulation of an LP without controls where y_t is an AR(1) with $\rho = 0.95$ and a sample size of $T = 50$.[17] Several features of these results are worth noting. First, implementing our proposed bias correction and using controls can increase coverage probabilities. Second, even with our bias correction and controls, the coverage probabilities are markedly lower than 95%. Third, the Huber-White standard errors appear to perform better than the Newey-West standard errors.

Overall, our analysis of standard errors indicates that bias in LPs is an important problem to address when conducting inference. In small samples, Huber-White standard errors are preferable to HAR standard errors when using critical values from the normal asymptotic limiting distribution. When researchers are interested in credible regions around point estimates, implementing our bias adjustment to $\widehat{\theta}_{h,LS}$ can improve the coverage probabilities.

6 Application to monetary policy shocks

In this section we provide empirical examples of bias in LPs and also the estimation of standard errors. We highlight three examples from the literature on the effect of monetary

[17]When constructing the Newey-West estimator, we use a bandwidth of $0.7\,(T-h)^{1/3}$. We use this bandwidth as the "textbook" choice from Lazarus et al. (2018). When one uses the fixed-b critical values suggested by Lazarus et al. (2018), the performance of the Newey-West estimator improves somewhat (see our Appendix), though the coverage probabilities are not uniformly better than those from the Huber-White standard errors. Fixed-b asymptotics involve using a larger bandwidth for the Newey-West estimator and larger critical values than those implied by the asymptotic normal approximation. It turns out that the bandwidth is not that much larger in sample sizes typically found in the literature. As a result, it is not surprising that using a larger critical value improves the coverage probabilities given that the confidence intervals are too small.

Table 1: Coverage probability of different estimators of standard errors for $\widehat{\theta}_h$ in LP when y_t is an AR(1) with $\rho = 0.95$ and $T = 50$

	$\widehat{\theta}_{h,LS}$, no controls		$\widehat{\theta}_{h,BCC}$, no controls		$\widehat{\theta}_{h,BCC}$, controls	
h	Huber-White	Newey-West	Huber-White	Newey-West	Huber-White	Newey-West
0	0.87	0.82	0.86	0.82	0.92	0.91
1	0.83	0.80	0.82	0.80	0.90	0.88
2	0.80	0.77	0.79	0.78	0.87	0.86
3	0.78	0.75	0.76	0.75	0.85	0.83
4	0.76	0.73	0.75	0.73	0.83	0.81
5	0.75	0.72	0.74	0.72	0.81	0.79
6	0.75	0.72	0.73	0.71	0.80	0.78
7	0.74	0.70	0.73	0.70	0.78	0.76
8	0.74	0.71	0.73	0.70	0.77	0.75
9	0.74	0.71	0.73	0.70	0.76	0.74
10	0.74	0.71	0.73	0.70	0.75	0.74

policy shocks on the macroeconomy. These shocks are constructed using either the narrative approach of Romer and Romer (2004) or through asset prices as pioneered by Kuttner (2001).

6.1 The effects of monetary policy shocks on output and inflation

Using a setup similar to Gorodnichenko and Lee (2019), we estimate the effects of Romer and Romer (2004) monetary policy shocks on output and inflation.[18] The data sample runs from 1969:Q1-2008:Q4. We estimate LPs of the form in equation (2) on real output growth and annualized inflation. We include controls consisting of four lags of real output growth, inflation, the federal funds rate, and the monetary policy innovation.[19]

[18]The shock series was extended to 2008 by Coibion et al. (2017).

[19]The LPs here are slightly different from the ones in Gorodnichenko and Lee (2019) in two ways. First, we use y_{t+h} rather than $y_{t+h} - y_{t-1}$. The bias discussed in this paper is still present under the latter formula. Second, we omit TFP innovations because the objective here is not to study relative variance contributions. Taken together, these differences lead to only minor changes in the estimated LPs.

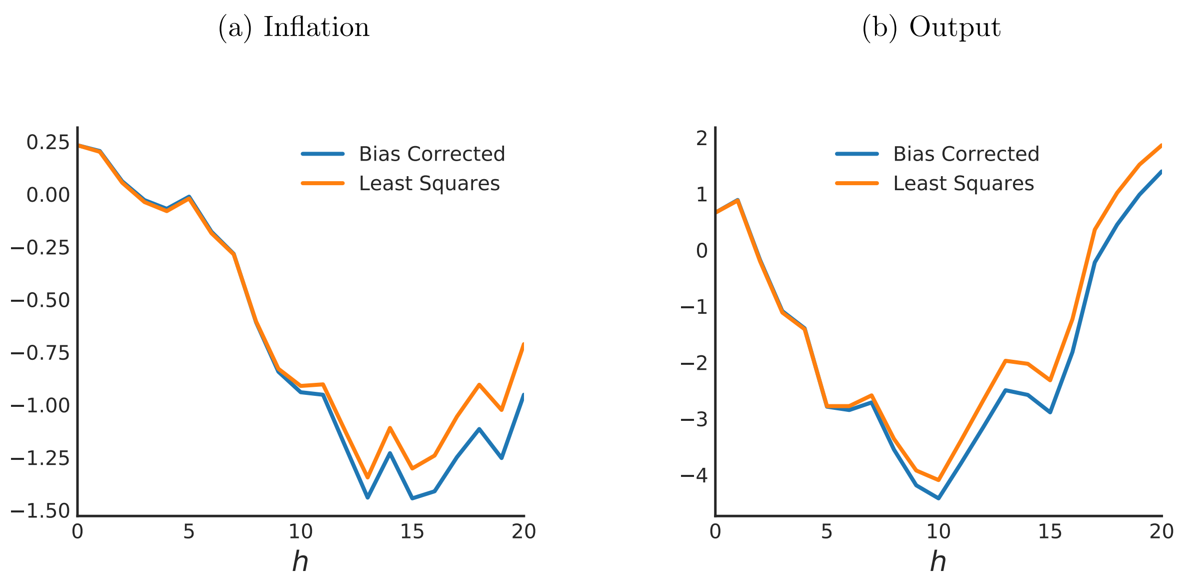


Figure 8: The effect of monetary policy shocks.

The estimated impulse responses of inflation and output to a monetary policy shock are displayed in Figure 8. As in Gorodnichenko and Lee (2019), we cumulate the impulse response of output growth. Figure 8 also shows the bias-corrected estimate of the impulse response. To focus attention on the difference between the two impulse responses in this illustrative example, we omit confidence bands.

The estimated inflation impulse response roughly accords with Gorodnichenko and Lee (2019): A contractionary 100 basis-point monetary policy shock causes inflation to be little changed for the first few periods after the shock and then eventually decline persistently. The bias-corrected impulse response indicates that inflation responds somewhat more to a monetary policy shock than under the conventional estimates. On average, the response of inflation is about 15 basis points lower in the bias-corrected impulse response, a moderate but nontrivial difference. The bias-corrected estimator is lower than the least-squares LP estimator because the estimated least-squares impulse response is negative for almost horizons, and the bias-correction is a weighted average of the impulse responses at all previous horizons.

The estimated output impulse response also broadly accords with the results in Gorodnichenko and Lee (2019): a contractionary 100 basis-point monetary policy shock causes the level of output to contract by about 4 percent after two and half years, after which effects of the shock slowly dissipate. Notably, the bias-corrected estimator impliees the output decline is about 1/2 percentage point larger. Essentially, this is because the bias-correction at say, horizon $h = 15$ is influenced by the LP coefficients at previous horizons.

For both inflation and output, the corrections are moderated but economically meaningfully. The corrections are larger are longer horizon, a consequence of the single-sided nature of our bias correction, given in equation ().[20] It is perhaps not surprising that the bias correction is moderate given that the sample size used here is markedly larger than those typically found in the LP literature.

6.2 State dependence in the effects of monetary policy shocks

One of the key advantages of LPs, relative to other popular methods in macroeconomics, are their ability to handle nonlinearities. Tenreyro and Thwaites (2016) use a "smoothly transitioning local projection model" to investigate whether monetary policy has larger effects during recessions or expansions. The LP specification is given by

$$y_{t+h} = \tau t + F(z_t)(\alpha_h^b + \theta_h^b \varepsilon_t + \gamma^{b'} c_t) + (1 - F(z_t))(\alpha_h^r + \theta_h^r \varepsilon_t + \gamma^{r'} c_t) + u_t. \tag{16}$$

Here y_{t+h} is the endogenous variable of interest (output), c_t is a vector of controls and $F(z_t)$ is a smooth increasing function of an indicator of the state of the economy z_t. The value of $F(z_t)$ is exogenous from the perspective of the regression. The shock ε_t is once again a variant of the Romer and Romer (2004) measure. The coefficients of interest are θ_h^b and θ_h^r, the effects of a monetary policy shock at horizon h in a boom (b) and recession (r), respectively.

Since $F(z_t)$ is exogenous, we can estimate this LP model by OLS. We proceed by first removing the trend and estimate (16) and then omit the linear trend in the regression specification.[21] We estimate the model on quarterly data using a sample that runs from 1969:Q1 to 2002:Q4. Following, Tenreyro and Thwaites (2016), the control variables c_t are one lag of detrended output and the federal funds rate.

Table 2 shows the LS and BC estimates of with β^b and β^r, along with with Newey-West and Huber-White standard errors in parenthesis for $h = 9, \ldots, 12$. The coefficient estimates

[20] An earlier working paper version of this paper presented a slightly difference formula for the bias correction in the case of LP with controls. While both are valid, the correction in the previous version for horizon h was influenced by both estimates of impulse responses at all horizons. The current formula depends only on horizons $j \leq h$. This is much easier implement in practice.

[21] Results are similar under alternative ways of detrending output.

Table 2: State-Dependent Effects of Monetary Policy Shocks

Horizon		$\hat{\beta}^b_{LS}$	$\hat{\beta}^r_{LS}$	Difference	$\hat{\beta}^b_{BC}$	$\hat{\beta}^r_{BC}$	Difference
	Point Estimate	-1.27	-0.70	-0.58	-1.36	-0.80	-0.56
$h = 9$	NW SE.	(0.39)	(0.29)	(0.59)	(0.39)	(0.29)	(0.58)
	HW SE.	(0.45)	(0.64)	(0.89)	(0.46)	(0.65)	(0.90)
	Point Estimate	-1.49	-0.41	-1.08	-1.61	-0.52	-1.09
$h = 10$	NW SE.	(0.37)	(0.25)	(0.53)	(0.37)	(0.28)	(0.55)
	HW SE.	(0.50)	(0.74)	(1.01)	(0.51)	(0.75)	(1.02)
	Point Estimate	-1.10	-0.38	-0.73	-1.26	-0.49	-0.77
$h = 11$	NW SE.	(0.38)	(0.34)	(0.64)	(0.39)	(0.39)	(0.67)
	HW SE.	(0.60)	(0.72)	(1.05)	(0.60)	(0.73)	(1.06)
	Point Estimate	-0.64	-0.47	-0.17	-0.81	-0.58	-0.23
$h = 12$	NW SE.	(0.45)	(0.44)	(0.80)	(0.46)	(0.49)	(0.85)
	HW SE.	(0.65)	(0.62)	(1.01)	(0.64)	(0.64)	(1.02)

Notes: Table shows the LS and BC estimates of with β^b and β^r with Newey-West and Huber-White standard errors.

can be interpreted as the percentage point effect on real output in period $t + h$ of a 100 point monetary policy shock at time t when the economy is definitely in either a boom or recession, respectively. The point estimates under both LS and BC indicate that positive shocks occuring during booms have are more contractionary than identically-sized shocks occuring during recessions, consistent with Tenreyro and Thwaites (2016). Relative to the LS estimates, the bias-corrected estimates are more negative for both β^b_h and β^r_h, consistent with the earlier simulation study.

The difference between the coefficients for booms and recessions is larger when using the bias-corrected estimators. That is, there is moderately more evidence for differential effects of monetary policy shocks depending on the state of the business cycle. As in Tenreyro and Thwaites (2016), the uncertainty surrounding these estimates is large. Notably, the Huber-White standard errors are markedly larger than the Newey-West standard errors, consistent with our analysis in Section 5. For a number of the entries in Table 2, conclusions about statistical significance would be different using the Huber-White standard errors as compared to the Newey-West standard errors.

6.3 Time dependence in the effect of monetary policy shocks

Our final example assesses the evidence for a change in transmission of monetary policy over the early part of the 2000s. Lunsford (2020) identifies two monetary policy shocks from high frequency movements in assets in prices. The first shock is to the level of current federal funds rate and the second shock, the *forward guidance shock* is to the market's expected path of federal funds rate beyond the current month. Lunsford (2020) argues that the responses of asset prices and macroeconomic aggregates to forward guidance shocks changed in 2003, as the policy statement issued by the Federal Open Market Committee (FOMC) following monetary policy decisions shifted to emphasize future policy inclinations rather than risks to the economic outlook. Using the identified federal funds rate (ε_t^{FFR}) and forward guidance ε_t^{FG} shocks, Lunsford (2020) effectively estimates LPs of the form

$$y_{t+h} - y_t = \alpha_h + \theta_h^{FFR}\varepsilon_t^{FFR} + \theta_h^{FG}\varepsilon_t^{FG} + u_{t,t+h}, \quad h = 1, \dots, H. \tag{17}$$

where y_{t+h} is a macroeconomic aggregate. The focus here will be on estimates of the coefficient θ_h^{FG}, and whether they have changed over two samples of monthly data from February to June 2003 and August 2003 to May 2006. Following Lunsford (2020), the horizon of interest in $h = 12$, so θ_{12}^{FG} measures the effect of a forward guidance shock one year after its realization (after netting the contemporaneous effect). The size of the two sample deserves emphasis as they are 28 and 23 observations, respectively. These are extremely small—about half of the smallest sample considered in the Monte Carlo simulations in this paper.[22]

Before describing the regression results, we note two things about the LP model in (17). First, subtracting y_t from the dependent variables tends to the reduce the parameter bias substantially. The model in (17) is essentially an LP without controls. One can see, either from the trajectories in Figure~2 or the expression in (4), that in an LP without controls a

[22]For ease of exposition, the paper assumes that the econometrician observes T total observations of both y_t and ε_t (and potentially controls.) So there are $T-h$ observations for the hth horizon LP. In some practical applications, the econometrician is limited only by the observations of shock. That is, they observe a sample of size $T+H$ of y_t and a sample of size T of ε_t, where H is the maximum horizon considered. In this case, there are T observations for each of the hth horizon LPs–this is the case in Lunsford (2020). Using a constant rather than shrinking sample size results in extremely minor modifications of the analytic expressions and essentially no changes to the Monte Carlo simulations in this paper.

sizable portion of the bias associated with least squares estimate of θ_h at horizons greater than zero is accounted for the bias in θ_0. Informally, subtracting y_t from y_{t+h} out removes this portion of the bias.

Table 3 displays $\hat{\theta}^{FG}_{12,LS}$ from the LP model in (17) along with HAC and HW standard errors over the two samples for three dependent variables: the growth in real personal consumption expenditures (PCE), unemployment rate changes, and the growth in industrial production (IP). As in 3, the HAC standard errors are computed using $u = 10$ lags. The point estimates, as mentioned above, are computed using least squares, as the bias correction would likely be minor when using $y_{t+h} - y_t$ as the dependent variable.

The point estimates indicate that while contractionary–i.e, positive—forward guidance shocks were associated with increases in real PCE growth and a fall in the unemployment rate in the first sample, they were associated with a fall in consumption growth and an increase in unemployment over the second sample. The point estimates associated with IP are both negative. Consistent with standard practice, Lunsford (2020) uses HAC standard errors, which are replicated in Table 3; when using the fixed-b asymptotic critical values from Sun (2014), which 3 employs, the coefficients associated PCE growth and the unemployment rate are moderately statistically significant. Table 3 also displays the HW standard errors. With one exception—the unemployment rate in the August to May 2006 sample—these are larger than the HAC standard errors, sometimes substantially so. The standard error estimates associated with PCE growth and IP growth in the second sample increases from 3.82 to 6.15 and 4.37 to 10.90, respectively.[23]

As argued in Section 5, in small samples, even when appropriate, HAC standard errors tend to underestimate the long-run variance relative to the Huber-White estimator. Note that this true even when the estimate of the coefficient of interest exhibits little bias itself, as we have argued is plausible in this case. The Huber-White standard errors based on the indicate that the uncertainty surrounding these regression coefficients is considerably larger than implied the Newey-West standard errors.

[23]In this case, whether the critical values used to assess statistical significance should change depending on the construction of the standard errors is complicated and requires an explicit statement of the null hypothesis. In any event, given such small sample sizes, relying on critical values associated with limiting distributions could be problematic.

Table 3: Response to Forward Guidance Shock

	Feb. 2000 to Jun. 2003	Aug. 2003 to May 2006
PCE Growth Results		
Point Estimate	2.88	-10.51
NW SE	(2.26)	(3.82)
HW SE	(2.61)	(6.15)
Unemployment Results		
Point Estimate	-4.01	3.49
NW SE	(1.13)	(1.24)
HW SE	(2.45)	(1.04)
IP Growth Results		
Point Estimate	-9.87	-10.10
NW SE	(8.53)	(4.37)
HW SE	(12.03)	(10.90)

Notes: The Table shows $\hat{\theta}^{FG}_{12,LS}$ from the LP model in (17) along with Newey-West (NW) and Huber-White (HW) standard errors. The NW standard errors are computed using a bandwidth of $u = 10$.

7 Conclusion

We have shown that LPs can be severely biased in sample sizes commonly found in the related literature. We derived an approximate bias function that shows that LPs are intimately linked across horizons in small samples. The bias of LPs persists even when researchers have access to a large cross-section (panel data) or when researchers use instrumental variables.

We used our approximate bias function to bias correct LPs. In Monte Carlo analysis our bias correction does not completely correct for the bias in LPs. These results suggest that other time series models with well-understood, effective methods for bias correction (such as VARs) may be better alternatives for estimated impulse responses if researchers have data samples in the time dimension that are similar to those typically found in empirical macroeconomic research. In particular, specifying time series models that are generative for the time series of interest would allow researchers to use likelihood methods.

We also analyzed bias in standard errors computed for estimated impulse response functions from LPs. We showed that, in small samples, standard errors that rely on estimated autocovariances of the regression score, like the Newey-West estimator, typically understate the amount of uncertainty surrounding the estimated impulse response functions. We argued that, in most cases, researchers should prefer standard errors that are heteroskedasticity consistent, but not autocorrelation robust.

Recent work on standard errors in time series regression has focused on limiting distributions other than the normal distribution (see Sun (2014) and Lazarus et al. (2018)). However, with samples typically found in the LP literature, it is difficult to appeal to limiting critical values as accurate approximations. As a result, if researchers are going to use HAR standard errors, they may want to check to see if Huber-White standard errors would lead to different conclusions. If the Huber-White standard errors are larger than the HAR standard errors, researchers should consider what might lead to the apparent negative autocovariance in the regression score. Without other reasonable theory, it may be that the negative estimates of the autocovariance of the regression score are the result of small sample bias, which could result in spurious inference.

References

Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (2019): "Democracy does cause growth," *Journal of Political Economy*, 127, 47–100.

Anatolyev, S. (2005): "GMM, GEL, Serial Correlation, and Asymptotic Bias," *Econometrica*, 73, 983–1002.

Bao, Y. and A. Ullah (2007): "The second-order bias and mean squared error of estimators in time-series models," *Journal of Econometrics*, 140, 650 – 669.

Coibion, O., Y. Gorodnichenko, L. Kueng, and J. Silvia (2017): "Innocent Bystanders? Monetary Policy and Inequality," *Journal of Monetary Economics*, 88, 70–89.

Ganics, G., A. Inoue, and B. Rossi (2019): "Confidence Intervals for Bias and Size Distortion in IV and Local Projections-IV Models," *Journal of Business & Economic Statistics*, 0, 1–18.

Gorodnichenko, Y. and B. Lee (2019): "Forecast Error Variance Decompositions with Local Projections," *Journal of Business & Economic Statistics*, 0, 1–24.

Hahn, J. and G. Kuersteiner (2002): "Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large," *Econometrica*, 70, 1639–1657.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*, Springer New York.

Hall, P., J. L. Horowitz, and B.-Y. Jing (1995): "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika*, 82, 561–574.

Jorda, O. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161?182.

Jordà, Ò., M. Schularick, and A. M. Taylor (2015): "Betting the house," *Journal of International Economics*, 96, S2–S18.

Kendall, M. G. (1954): "Note on bias in the estimation of autocorrelation," *Biometrika*, 41, 403–404.

Kilian, L. (1998): "Small-sample confidence intervals for impulse response functions," *Review of economics and statistics*, 80, 218–230.

Kilian, L. and Y. J. Kim (2011): "How reliable are local projection estimators of impulse responses?" *Review of Economics and Statistics*, 93, 1460–1466.

Kuttner, K. N. (2001): "Monetary Policy Surprises and Interest Rates: Evidence From the Federal Funds Futures Market," *Journal of Monetary Economics*, 47, 523–544.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018): "HAR Inference: Recommendations for Practice," *Journal of Business & Economic Statistics*, 36, 541–559.

Lunsford, K. G. (2020): "Policy Language and Information Effects in the Early Days of Federal Reserve Forward Guidance," *American Economic Review*, 110, 2899–2934.

Lusompa, A. (2019): "Local Projections, Autocorrelations, and Efficiency," *Unpublished Manuscript, UC-Irvine.*

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation," *Econometrica*, 55, 703–708.

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

Olea, J. L. M. and M. Plagborg-Møller (2020): "Local Projection Inference is Simpler and More Robust Than You Think," .

Plagborg-Møller, M. and C. K. Wolf (2019): "Local Projections and VARs Estimate the Same Impulse Responses," *Unpublished Manuscript, Princeton University.*

Ramey, V. A. and S. Zubairy (2018): "Government spending multipliers in good times and in bad: evidence from US historical data," *Journal of Political Economy*, 126, 850–901.

Rilstone, P., V. Srivastava, and A. Ullah (1996): "The second-order bias and mean squared error of nonlinear estimators," *Journal of Econometrics*, 75, 369 – 395.

Romer, C. D. and D. H. Romer (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.

Shaman, P. and R. A. Stine (1988): "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association*, 83, 842–848.

Stock, J. H. and M. W. Watson (2018): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *Economic Journal*, 128, 917–948.

Sun, Y. (2014): "Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference," *Journal of Econometrics*, 178, 659–677.

Tenreyro, S. and G. Thwaites (2016): "Pushing on a string: US monetary policy is less powerful in recessions," *American Economic Journal: Macroeconomics*, 8, 43–74.

A Derivations of approximate bias

In this appendix, we derive our expressions for the bias of the LP estimators studied in our paper. To do so, we employ the framework proposed by Rilstone et al. (1996) and extended to time series models by Bao and Ullah (2007). These papers derive expressions for finite-sample moments for a wide class of estimators via an approximation of an estimator $\hat{\beta}$ of the form:

$$\hat{\beta} - \beta = a_{-1/2} + a_{-1} + a_{-3/2} + o_p(T^{-3/2}). \tag{18}$$

It can be verified that under Assumption 1, combined with the least squares estimation framework, satisfies the necessary assumptions of Rilstone et al. (1996). Assumptions 2 and 3 allow one to obtain tractable expressions.

In our derivation, we use the notation of Bao and Ullah (2007) where possible. For each derivation, we will cast the OLS estimator as a GMM problem with moment conditions given by $q(\beta; w_t$, where the data vector $w_t = [y_t, x_t']'$ for the LP models with and without controls. For the IV and Panel models, the data vector is expanded to include the additional observables. The objective function is thus given by:

$$\psi_{T-h}(\beta; W_{1:T}) = \frac{1}{T-h} \sum_{t=1}^{T-h} q(\beta; w_t).$$

Let $\nabla^i A(\beta)$ be the matrix of ith order partial derivatives of A with respect to β. In what follows, write $\psi_{T-h}(\beta; W_{1:T})$ as ψ_{T-h} and $q(\beta; w_t)$ as q_t. Define the series of matrices

$$H_i = \nabla^i \psi_{T-h} \text{ and } \overline{H}_i = \mathbb{E}[H_i] \text{ with } Q = \overline{H}_1^{-1}, V = H_1 - \overline{H}_1 \text{ and } W = H_2 - \overline{H}_2.$$

Bao and Ullah (2007) show that the expressions for the terms in (18) are given by:

$$a_{-1/2} = -Q\psi_{T-h}, \quad a_{-1} = -QVa_{-1/2} - \frac{1}{2}Q\overline{H}_2\left[a_{-1/2} \otimes a_{-1/2}\right] \quad \text{and}$$

$$a_{-3/2} = -QVa_{-1} - \frac{1}{2}QW\left[a_{-1/2} \otimes a_{-1/2}\right] - \frac{1}{2}Q\overline{H}_2\left(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}\right) - \frac{1}{6}Q\overline{H}_3\left(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}\right).$$

We are interested in computed in the bias, that is $\mathbb{E}\left[\hat{\beta} - \beta\right]$. It is obvious that $\mathbb{E}\left[a_{-1/2}\right] = 0$. Moreover, because the moment conditions associated with the LP estimator are linear in β, we have $H_j = \bar{H}_j = 0$ for $j \geq 2$. It can be verified that the expression for the bias simplifies considerably to

$$\mathbb{E}\left[\hat{\beta} - \beta\right] = \mathbb{E}\left[QH_1Q\psi_{T-h}\right] := B. \tag{19}$$

Our objective is to derive the element of the vector B that is associated with the shock of interest ε for the LP model with and without controls. Before proceeding with introduce notation to define first and second moments of the shocks, ε_t, and controls, c_t:

$$\mu_\varepsilon = \mathbb{E}[\varepsilon_t], \quad \sigma_\varepsilon^2 = \mathbb{E}\left[(\varepsilon_t - \mu_\varepsilon)^2\right], \quad \mu_c = \mathbb{E}[c_t], \quad \text{and } \Sigma_c = \mathbb{E}\left[(c_t - \mu_c)(c_t - \mu_c)'\right].$$

A.1 LP without controls

For the LP without controls, the moment conditions associated with the OLS estimator are defined as

$$q_t \equiv \begin{bmatrix} y_{t+h} - \alpha_h - \theta_h \varepsilon_t \\ \varepsilon_t(y_{t+h} - \alpha_h - \theta_h \varepsilon_t) \end{bmatrix},$$

where we have departed slightly from the notation of Bao and Ullah (2007) in defining the parameter vector. Before constructing the matrices of derivatives, note that we can rearrange the moment conditions to deduce:

$$\theta_h = \frac{\mathbb{E}\left[(\varepsilon_t - \mu_\varepsilon)(y_{t+h} - \alpha_h)\right]}{\mathbb{E}\left[(\varepsilon_t - \mu_\varepsilon)^2\right]} = \frac{\mathbb{E}\left[(\varepsilon_t - \mu_\varepsilon)(y_{t+h} - \alpha_h)\right]}{\sigma_\varepsilon^2}$$

Additionally for $s < t$, we have

$$\theta_h = \frac{\mathbb{E}\left[(\varepsilon_t - \mu_\varepsilon)(y_{t+h} - \mathbb{E}_s[y_{t+h}] + \mathbb{E}_s[y_{t+h}] - \alpha_h)\right]}{\sigma^2} = \frac{\mathbb{E}\left[(\varepsilon_t - \mu_\varepsilon)(y_{t+h} - \mathbb{E}_s[y_{t+h}] - \alpha_h)\right]}{\sigma^2} \tag{20}$$

Where the second equality follows from the fact that $\mathbb{E}_s[\varepsilon_t] = \mu$ for all $s < t$. It is easy to see that:

$$H_1 = \frac{1}{T-h}\sum_{t=1}^{T-h}\begin{bmatrix} -1 & -\varepsilon_t \\ -\varepsilon_t & -\varepsilon_t^2 \end{bmatrix}, \quad \overline{H}_1 = \begin{bmatrix} -1 & -\mu_\varepsilon \\ -\mu_\varepsilon & -\left(\sigma_\varepsilon^2 + \mu_\varepsilon^2\right) \end{bmatrix}, \quad \text{and } Q = \begin{bmatrix} -\left(1 + \frac{\mu_\varepsilon^2}{\sigma_\varepsilon^2}\right) & \frac{\mu_\varepsilon}{\sigma_\varepsilon^2} \\ \frac{\mu_\varepsilon}{\sigma_\varepsilon^2} & -\frac{1}{\sigma_\varepsilon^2} \end{bmatrix}.$$

Tedious algebra yields:

$$QH_1Q = \frac{1}{T-h}\sum_{t=1}^{T-h}\begin{bmatrix} -\left(1 - 2\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}(\varepsilon_t - \mu_\varepsilon) + \left(\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}\right)^2(\varepsilon_t - \mu_\varepsilon)^2\right) & -\left(\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t - \mu_\varepsilon) - \frac{\mu_\varepsilon}{(\sigma_\varepsilon^2)^2}(\varepsilon_t - \mu_\varepsilon)^2\right) \\ -\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t - \mu_\varepsilon) + \frac{\mu_\varepsilon}{(\sigma_\varepsilon^2)^2}(\varepsilon_t - \mu_\varepsilon)^2 & -\frac{1}{(\sigma_\varepsilon^2)^2}(\varepsilon_t - \mu_\varepsilon)^2 \end{bmatrix}.$$

Recall that we only need to obtain the second element of B which corresponds with the bias associated with $\hat{\theta}_h$. Thus, we only need to calculate $[QH_1Q]_{2,\cdot}\psi_{T-h}$. This is given by:

$$\begin{aligned}
[QH_1Q]_{2,\cdot}\psi_{T-h} &= \frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\left(-\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t - \mu_\varepsilon) + \frac{\mu_\varepsilon}{(\sigma_\varepsilon^2)^2}(\varepsilon_t - \mu_\varepsilon)^2\right)\sum_{t=1}^{T-h}(y_{t+h} - \alpha_h - \theta_h\varepsilon_t) \\
&\quad - \frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\frac{1}{(\sigma_\varepsilon^2)^2}(\varepsilon_t - \mu_\varepsilon)^2\sum_{t=1}^{T-h}\varepsilon_t(y_{t+h} - \alpha_h - \theta_h\varepsilon_t) \\
&= \frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h} -\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t - \mu_\varepsilon)(y_{s+h} - \alpha_h - \theta_h\varepsilon_s) - \frac{1}{(\sigma_\varepsilon^2)^2}(\varepsilon_t - \mu_\varepsilon)^2(\varepsilon_s - \mu_\varepsilon)(y_{s+h} - \alpha_h - \theta_h\varepsilon_s) \\
&= \frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}(\phi_I(t,s) + \phi_{II}(t,s))
\end{aligned}$$

Consider the expectation of the term ϕ_I. We have:

$$\mathbb{E}[\phi_I(t,s)] = \begin{cases} 0 & \text{if } t = s \\ -\theta_{h+s-t} & \text{otherwise .} \end{cases}$$

Now consider the expectation of the term ϕ_{II},

$$\begin{aligned}
\mathbb{E}[\phi_{II}(t,s)] &= \frac{1}{(\sigma_\varepsilon^2)^2}\mathbb{E}[(\varepsilon_t - \mu_\varepsilon)^2(\varepsilon_s - \mu_\varepsilon)(y_{s+h} - \alpha_h - \theta_h\varepsilon_s)] \\
&= \frac{1}{(\sigma_\varepsilon^2)^2}\mathbb{E}[(\varepsilon_t - \mu_\varepsilon)^2\,\mathbb{E}[(\varepsilon_s - \mu_\varepsilon)(y_{s+h} - \alpha_h - \theta_h\varepsilon_s)\,|\varepsilon_t]] \\
&= \frac{1}{(\sigma_\varepsilon^2)^2}\mathbb{E}\left[(\varepsilon_t - \mu_\varepsilon)^2\,\mathbb{E}\left[(\varepsilon_s - \mu_\varepsilon)\left(\underbrace{y_{s+h} - \alpha_{h-(t-s)} - \theta_{h-(t-s)}\varepsilon_t}_{\delta_I(t,s)} + \underbrace{\alpha_{h-(t-s)} + \theta_{h-(t-s)}\varepsilon_t - \alpha_h - \theta_h\varepsilon_s}_{\delta_{II}(t,s)}\right)|\varepsilon_t\right]\right].
\end{aligned}$$

By Assumptions I, ε_t does not enter into $\delta_I(t,s)$. Moreover, under Assumption 2, conditioning on it does not convey any useful information about ε_s or other components of y_{t+s}.

$$\mathbb{E}\left[(\varepsilon_s-\mu_\varepsilon)\delta_I(t,s)|\varepsilon_t\right]=\mathbb{E}\left[(\varepsilon_s-\mu_\varepsilon)\delta_I(t,s)\right]. \tag{21}$$

Direct calculation yields:

$$\mathbb{E}\left[(\varepsilon_s-\mu_\varepsilon)\delta_I(t,s)\right]=\begin{cases}0 & \text{if } t=s\\ \theta_h\sigma_\varepsilon^2 & \text{otherwise.}\end{cases}$$

Consider now the term involving $\delta_{II}(t,s)$. By direct calculation:

$$\mathbb{E}\left[(\varepsilon_s-\mu_\varepsilon)\delta_{II}(t,s)|\varepsilon_t\right]=\begin{cases}0 & \text{if } t=s\\ -\theta_h\sigma_\varepsilon^2 & \text{otherwise.}\end{cases}$$

Thus $\mathbb{E}\left[\phi_{II}(t,s)\right]=0$ for all t and s. Combining these results, we have:

$$\mathbb{E}\left[[QH_1Q]_{2,\cdot}\psi_{T-h}\right]=\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}-\theta_{h+s-t}\mathbf{1}_{\{t\neq s\}}.$$

Tedious arithmetic confirms that:

$$B=-\frac{1}{T-h}\sum_{j=1}^{T-h-1}\left(1-\frac{j}{T-h}\right)\left(\theta_{h+j}+\theta_{h-j}\right).$$

This delivers equation (4).

A.2 LP with controls

In the case of controls, we have $x_t=[\varepsilon_t,c_t]'$, so the moment conditions associated with the OLS estimator are defined so that

$$q_t\equiv\begin{bmatrix} y_{t+h}-\alpha_h-x_t'\beta_h\\ x_t\left(y_{t+h}-\alpha_h-x_t'\beta_h\right)\end{bmatrix}=\begin{bmatrix} y_{t+h}-\alpha_h-x_t'\beta_h\\ \varepsilon_t\left(y_{t+h}-\alpha_h-x_t'\beta_h\right)\\ c_t\left(y_{t+h}-\alpha_h-x_t'\beta_h\right)\end{bmatrix}.$$

Recall the object of interest, θ_h, is the first element of β_h. We have that

$$H_1=\frac{1}{T-h}\sum_{t=1}^{T-h}\begin{bmatrix}-1 & -\varepsilon_t & -c_{t-1}'\\ -\varepsilon_t & -\varepsilon_t^2 & -\varepsilon_t c_{t-1}'\\ -c_{t-1} & -\varepsilon_t c_{t-1} & -c_{t-1}c_{t-1}'\end{bmatrix},\quad \overline{H}_1=\begin{bmatrix}-1 & -\mu_\varepsilon & -\mu_c'\\ -\mu_\varepsilon & -\left(\sigma_\varepsilon^2+\mu_\varepsilon^2\right) & -\mu_\varepsilon\mu_c'\\ -\mu_c & -\mu_\varepsilon\mu_c & -\left(\Sigma_c+\mu_c\mu_c'\right)\end{bmatrix},$$

$$\text{and } Q=\begin{bmatrix}-\left(1+\frac{\mu_\varepsilon^2}{\sigma_\varepsilon^2}+\mu_c'\Sigma_c^{-1}\mu_c\right) & \frac{\mu_\varepsilon}{\sigma_\varepsilon^2} & \mu_c'\Sigma_c^{-1}\\ \frac{\mu_\varepsilon}{\sigma_\varepsilon^2} & -\frac{1}{\sigma_\varepsilon^2} & 0\\ \Sigma_c^{-1}\mu_c & 0 & -\Sigma_c^{-1}\end{bmatrix}.$$

As with previous deriviation, we need only calculate the second row of QH_1Q. Direct calculation yields

$$[QH_1Q]_{2,\cdot} = \frac{1}{T-h}\sum_{t=1}^{T-h}\frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)\begin{bmatrix} 1-\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)-\mu_c'\Sigma_c^{-1}\left(c_{t-1}-\mu_c\right) \\ \frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right) \\ \Sigma_c^{-1}\left(c_{t-1}-\mu_c\right)\end{bmatrix}'.$$

Then the second element of $QH_1Q\psi_{T-h}$ is given by

$$\begin{aligned}[QH_1Q\psi_{T-h}]_2 =&\frac{1}{T-h}\sum_{t=1}^{T-h}\frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)\begin{bmatrix} 1-\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)-\mu_c'\Sigma_c^{-1}\left(c_{t-1}-\mu_c\right) \\ \frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right) \\ \Sigma_c^{-1}\left(c_{t-1}-\mu_c\right)\end{bmatrix}' \\ &\times\frac{1}{T-h}\sum_{s=1}^{T-h}\begin{bmatrix} y_{s+h}-\alpha_h-x_s'\beta_h \\ \varepsilon_s\left(y_{s+h}-\alpha_h-x_s'\beta_h\right) \\ c_s\left(y_{s+h}-\alpha_h-x_s'\beta_h\right)\end{bmatrix}.\end{aligned}$$

Explicit calculation of this object yields:

$$\begin{aligned}[QH_1Q\psi_{T-h}]_2 =& -\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)\left(y_{s+h}-\alpha_h-x_s'\beta_h\right) \\ &-\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\left[\frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)\right]^2\left(\varepsilon_s-\mu_\varepsilon\right)\left(y_{s+h}-\alpha_h-x_s'\beta_h\right) \\ &-\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)\left(c_{t-1}-\mu_c\right)'\Sigma_c^{-1}\left(c_{s-1}-\mu_c\right)\left(y_{s+h}-\alpha_h-x_s'\beta_h\right) \\ =& -\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\phi_I(t,s)+\phi_{II}(t,s)+\phi_{III}(t,s).\end{aligned}$$

Consider first $\mathbb{E}[\phi_I(t,s)]$. Direct calculation yields:

$$\mathbb{E}[\phi_I(t,s)] = \begin{cases}\theta_{h-(t-s)} & \text{if } s<t\leq s+h \\ 0 & \text{otherwise.}\end{cases}$$

Note, unlike the no controls case, for $t<s$, $\mathbb{E}[\phi_I(t,s)]=0$ as a direct consequence of Assumption 3. By similar argument to the previous section, $\mathbb{E}[\phi_{II}(t,s)]=0$ for all t and s. Consider the expection of $\phi_{III}(t,s)$:

$$\mathbb{E}\left[\phi_{III}(t,s)\right] = \mathbb{E}\left[\frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)\left(c_{t-1}-\mu_c\right)'\Sigma_c^{-1}\left(c_{s-1}-\mu_c\right)\left(y_{s+h}-\alpha_h-x_s'\beta_h\right)\right]$$

To calculate the expectation, first fix t and s. Then the expression is

$$E\left[\left\{\frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)+\left[\frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)\right]^2\left(\varepsilon_s-\mu_{\tilde{\varepsilon}}\right)+\frac{1}{\sigma_\varepsilon^2}\left(\varepsilon_t-\mu_\varepsilon\right)\left(c_{t-1}-\mu_c\right)'\Sigma_c^{-1}\left(c_{s-1}-\mu_c\right)\right\}\left(y_{s+h}-\alpha_h-x_s'\beta_h\right)\right]$$

If $s \geq t$ or $s < t-h$, then this expectation is zero. If $t-h \leq s < t$,

$$E\left[\left\{\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)+\left[\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)\right]^2(\varepsilon_s-\mu_\varepsilon)+\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)(c_{t-1}-\mu_c)'\Sigma_c^{-1}(c_{s-1}-\mu_c)\right\}(y_{s+h}-\alpha_h-x_s'\beta_h)\right]=$$

$$E\left[\left\{\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)+\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)(c_{t-1}-\mu_c)'\Sigma_c^{-1}(c_{s-1}-\mu_c)\right\}(y_{s+h}-\alpha_h-x_s'\beta_h)\right]=$$

$$\theta_{h-(t-s)}+E\left[(c_{t-1}-\mu_c)'E\left[\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)(y_{s+h}-\alpha_h-x_s'\beta_h)|x_s,c_{t-1}\right]\Sigma_c^{-1}(c_{s-1}-\mu_c)\right]=$$

$$\theta_{h-(t-s)}\left(1+E\left\{(c_{t-1}-\mu_c)'\Sigma_c^{-1}(c_{s-1}-\mu_c)\right\}\right)$$

Plugging these expectations into the expression for B delivers equation (6.1).

A.3 LP with instrumental variables

In the case of instrumental variables with controls, the moment conditions associated with the OLS esitmator are defined so that

$$q_t \equiv \begin{bmatrix} y_{t+h}-\alpha_h-x_t'\beta_h \\ z_t(y_{t+h}-\alpha_h-x_t'\beta_h) \end{bmatrix}.$$

where $z_t \equiv [\omega_t, d_{t-1}]'$ is a vector of valid instruments of the same length as x_t. Define

$$\begin{aligned}&\mu_\omega \equiv E[\omega_t], \quad \mu_d \equiv E[d_t], \quad \sigma_{\omega\varepsilon} \equiv E[(\omega_t-\mu_\omega)(\varepsilon_t-\mu_\varepsilon)],\\ &\sigma_\omega^2 \equiv E\left[(\omega_t-\mu_\omega)^2\right], \quad \Sigma_d \equiv E\left[(d_t-\mu_d)(d_t-\mu_d)'\right], \quad \text{and } \Sigma_{dc} \equiv E\left[(d_t-\mu_d)(c_t-\mu_c)'\right].\end{aligned}$$

We have that:

$$H_1=\frac{1}{T-h}\sum_{t=1}^{T-h}\begin{bmatrix}-1 & -\varepsilon_t & -c_{t-1}' \\ -\omega_t & -\omega_t\varepsilon_t & -\omega_t c_{t-1}' \\ -d_{t-1} & -\omega_t c_{t-1} & -d_{t-1}c_{t-1}'\end{bmatrix}, \quad \overline{H}_1=\begin{bmatrix}-1 & -\mu_\varepsilon & -\mu_c' \\ -\mu_\varepsilon & -(\sigma_{\omega\varepsilon}+\mu_\omega\mu_\varepsilon) & -\mu_\omega\mu_c' \\ -\mu_d & -\mu_\varepsilon\mu_c & -(\Sigma_{dc}+\mu_d\mu_c')\end{bmatrix},$$

$$\text{and } Q=\begin{bmatrix}-\left(1+\frac{\mu_\omega\mu_\varepsilon}{\sigma_{\omega\varepsilon}}+\mu_c'\Sigma_{dc}^{-1}\mu_d\right) & \frac{\mu_\varepsilon}{\sigma_{\omega\varepsilon}} & \mu_c'\Sigma_{dc}^{-1} \\ \frac{\mu_\omega}{\sigma_{\omega\varepsilon}} & -\frac{1}{\sigma_{\omega\varepsilon}} & 0 \\ \Sigma_{dc}^{-1}\mu_d & 0 & -\Sigma_{dc}^{-1}\end{bmatrix}.$$

As with the previous derivations, we only need the second row of QH_1Q. Tedious arithmetic yields

$$[QH_1Q]_{2,\cdot}=-\frac{1}{T-h}\sum_{t=1}^{T-h}\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)\begin{bmatrix}1-\frac{\mu_\omega}{\sigma_{\omega\varepsilon}}(\varepsilon_t-\mu_\varepsilon)-\left(c_{t-1}'-\mu_c'\right)\Sigma_{dc}^{-1}\mu_d \\ \frac{1}{\sigma_{\omega\varepsilon}}(\varepsilon_t-\mu_\varepsilon) \\ \left(c_{t-1}'-\mu_c'\right)\Sigma_{dc}^{-1}\end{bmatrix}'.$$

So, the second element of $QH_1Q\psi_{T-h}$ is given by

$$\begin{aligned}[QH_1Q\psi_{T-h}]_2=&-\frac{1}{T-h}\sum_{t=1}^{T-h}\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)\begin{bmatrix}1-\frac{\mu_\omega}{\sigma_{\omega\varepsilon}}(\varepsilon_t-\mu_\varepsilon)-\left(c_{t-1}'-\mu_c'\right)\Sigma_{dc}^{-1}\mu_d \\ \frac{1}{\sigma_{\omega\varepsilon}}(\varepsilon_t-\mu_\varepsilon) \\ \left(c_{t-1}'-\mu_c'\right)\Sigma_{dc}^{-1}\end{bmatrix}' \\ &\times\frac{1}{T-h}\sum_{s=1}^{T-h}\begin{bmatrix}y_{s+h}-\alpha_h-x_s'\beta_h \\ \varepsilon_s(y_{s+h}-\alpha_h-x_s'\beta_h) \\ c_s(y_{s+h}-\alpha_h-x_s'\beta_h)\end{bmatrix}\end{aligned}$$

To calculate the expectation, first fix t and s. Then the expression of interest is

$$\mathbb{E}\left[\left\{\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)+\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)\frac{1}{\sigma_{\omega\varepsilon}}(\varepsilon_t-\mu_\varepsilon)(\omega_s-\mu_\omega)+\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)(c_{t-1}-\mu_c)'\Sigma_{dc}^{-1}(d_{s-1}-\mu_d)\right\}\right.$$
$$\left.\times\left(y_{s+h}-\alpha_h-x_s'\beta_h\right)\right]$$

If $s \geq t$ or $s < t-h$, then this expectation is zero. If $t-h \leq s < t$,

$$\mathbb{E}\left[\left\{\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)+\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)\frac{1}{\sigma_{\omega\varepsilon}}(\varepsilon_t-\mu_\varepsilon)(\omega_s-\mu_\omega)+\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)(c_{t-1}-\mu_c)'\Sigma_{dc}^{-1}(d_{s-1}-\mu_d)\right\}\right.$$
$$\left.\times\left(y_{s+h}-\alpha_h-x_s'\beta_h\right)\right]=$$
$$\mathbb{E}\left[\left\{\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)+\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)(c_{t-1}-\mu_c)'\Sigma_{dc}^{-1}(d_{s-1}-\mu_d)\right\}\left(y_{s+h}-\alpha_h-x_s'\beta_h\right)\right]=$$
$$\theta_{h-(t-s)}+E\left[(c_{t-1}-\mu_c)'\Sigma_c^{-1}(d_{s-1}-\mu_d)\,E\left[\frac{1}{\sigma_{\omega\varepsilon}}(\omega_t-\mu_\omega)\left(y_{s+h}-\alpha_h-x_s'\beta_h\right)|x_s,z_s,c_{t-1}\right]\right]=$$
$$\theta_{h-(t-s)}\left(1+E\left[(c_{t-1}-\mu_c)'\Sigma_{dc}^{-1}(d_{s-1}-\mu_d)\right]\right)=$$
$$\theta_{h-(t-s)}\left(1+\text{trace}\left\{E\left[(d_{s-1}-\mu_d)(c_{t-1}-\mu_c)'\right]\Sigma_{dc}^{-1}\right\}\right)$$

$$\mathbb{E}\left[\widehat{\theta}_{h,IV}\right]-\theta_h=-\frac{1}{T-h}\sum_{j=1}^{h}\left(1-\frac{j}{T-h}\right)\left(1+\text{trace}\left\{\Sigma_{dc,0}^{-1}\Sigma_{dc,j}\right\}\right)\theta_{h-j}+O\left(T^{-3/2}\right). \tag{22}$$

where $\Sigma_{dc,j}\equiv\mathbb{E}\left[(d_{t-j}-\mu_d)(c_t'-\mu_c')\right]$. This expression is analogous to equation (6.1).

B Derivation of bias in standard errors

We are going to consider a case where a researcher uses an LP without controls and wants to estimate the standard error of the estimator. To do so, the researcher needs to compute the variance of the regression score, which is composed of terms of the form

$$\gamma_{h,u}\equiv\mathbb{E}\left[\;\varepsilon_t\left(y_{t+h}-\alpha_h-\varepsilon_t\beta_h\right)\left(y_{t+h-u}-\alpha_h-\varepsilon_{t-u}\beta_h\right)\varepsilon_{t-u}\;\right].$$

The researcher wants to report standard errors under the null hypothesis that the coefficient on the shock of interest is zero. If this hypothesis is maintained at all horizons, then the researcher's maintained assumption is that

$$\varepsilon_t\perp y_s$$

for all t and s. We will use this maintained assumption. In this section, for ease of exposition, we assume $E[\varepsilon_t]=0$, but this is without loss of generality. Notably, typical HAC estimators of the standard error of $\widehat{\beta}_{h,LS}$ are functions only of $\gamma_{h,u}$, the sample size, and a bandwidth parameter.

When researchers calcluate $\gamma_{h,u}$ (for example, when computing Newey-West standard errors), they typically use plug-in estimators derived from the empirical regression scores. To understand how this procedure affects the small-sample properties of $\widehat{\gamma}_{h,u}$, it is useful to think about estimating the regression coefficients and and $\gamma_{h,u}$ jointly. In this case

$$q_t = \begin{bmatrix} (y_{t+h} - \alpha_h - \varepsilon_t \beta_h) \\ \varepsilon_t (y_{t+h} - \alpha_h - \varepsilon_t \beta_h) \\ \varepsilon_t (y_{t+h} - \alpha_h - \varepsilon_t \beta_h) \varepsilon_{t-u} (y_{t+h-u} - \alpha_h - \varepsilon_{t-u} \beta_h) - \gamma_{h,u} \end{bmatrix}.$$

Unlike our previous derivations, because of the $\beta_h \beta_h'$ term in q_t, the matrix H_2 is not a matrix of zeros.

$$E(\nabla q_t) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -\sigma_\varepsilon^2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

$$Q = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -\frac{1}{\sigma_\varepsilon^2} & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

$$Q\psi_{T-h-u} = -\frac{1}{T-h-u} \sum_{t=u+1}^{T-h} \begin{bmatrix} (y_{t+h} - \alpha_h - \varepsilon_t \beta_h) \\ \frac{\varepsilon_t}{\sigma_\varepsilon^2} (y_{t+h} - \alpha_h - \varepsilon_t \beta_h) \\ \varepsilon_t (y_{t+h} - \alpha_h - \varepsilon_t \beta_h) \varepsilon_{t-u} (y_{t+h-u} - \alpha_h - \varepsilon_{t-u} \beta_h) - \gamma_{h,u} \end{bmatrix}$$

$$\begin{aligned} QVQ\psi_{T-h-u} = & \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sigma_\varepsilon^2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \frac{1}{T-h-u} \sum_{t=1+u}^{T-h} \nabla q_t \\ & \times \frac{1}{T-h-u} \sum_{s=1+u}^{T-h} \begin{bmatrix} (y_{s+h} - \alpha_h - \varepsilon_s \beta_h) \\ \frac{\varepsilon_s}{\sigma_\varepsilon^2} (y_{s+h} - \alpha_h - \varepsilon_s \beta_h) \\ \varepsilon_s (y_{s+h} - \alpha_h - \varepsilon_s \beta_h) \varepsilon_{s-u} (y_{s+h-u} - \alpha_h - \varepsilon_{s-u} \beta_h) - \gamma_{h,u} \end{bmatrix} \end{aligned}$$

We care about the third row of $QVQ\psi_{T-h-u}$. Fix s and t, the terms in the third row are of the form

$$\begin{aligned} & [-\varepsilon_t \varepsilon_{t-u} (y_{t+h} - \alpha_h - \varepsilon_t \beta_h) - \varepsilon_t \varepsilon_{t-u} (y_{t+h-u} - \alpha_h - \varepsilon_{t-u} \beta_h)] (y_{s+h} - \alpha_h - \varepsilon_s \beta_h) \\ & + \left[-\varepsilon_t \varepsilon_{t-u}^2 (y_{t+h} - \alpha_h - \varepsilon_t \beta_h) - \varepsilon_t^2 \varepsilon_{t-u} (y_{t+h-u} - \alpha_h - \varepsilon_{t-u} \beta_h)\right] \frac{\varepsilon_s}{\sigma_\varepsilon^2} (y_{s+h} - \alpha_h - \varepsilon_s \beta_h) \\ & - \varepsilon_s (y_{s+h} - \alpha_h - \varepsilon_s \beta_h) \varepsilon_{s-u} (y_{s+h-u} - \alpha_h - \varepsilon_{s-u} \beta_h) + \gamma_{h,u} \end{aligned}$$

Clearly, $E[-\varepsilon_s (y_{s+h} - \alpha_h - \varepsilon_s \beta_h) \varepsilon_{s-u} (y_{s+h-u} - \alpha_h - \varepsilon_{s-u} \beta_h) + \gamma_{h,u}] = 0$ for all s. Consider

$$-[\varepsilon_t \varepsilon_{t-u} (y_{t+h} - \alpha_h - \varepsilon_t \beta_h) + \varepsilon_t \varepsilon_{t-u} (y_{t+h-u} - \alpha_h - \varepsilon_{t-u} \beta_h)] (y_{s+h} - \alpha_h - \varepsilon_s \beta_h)$$

If $u > 0$ and $\varepsilon_t \perp y_s$, then this is zero in expectation for all t and s. If $u = 0$, then the expectation is

$$-2\sigma_\varepsilon^2 E\left[(y_t - \mu_y)^2\right]$$

Consider

$$-\left[\varepsilon_t \varepsilon_{t-u}^2 (y_{t+h} - \alpha_h - \varepsilon_t \beta_h) + \varepsilon_t^2 \varepsilon_{t-u} (y_{t+h-u} - \alpha_h - \varepsilon_{t-u} \beta_h)\right] \frac{\varepsilon_s}{\sigma_\varepsilon^2} (y_{s+h} - \alpha_h - \varepsilon_s \beta_h)$$

If $u > 0$ and $\varepsilon_t \perp y_s$, then this is zero in expectation if $s \neq t$ and $s \neq t - u$. If $s = t$ or $s = t - u$, then this equals

$$-\sigma_\varepsilon^2 E\left(y_t - \mu_y\right).$$

If $u = 0$, then the expectation is

$$-E\left\{\left[2\varepsilon_t^3\left(y_{t+h} - \alpha_h - \varepsilon_t\beta_h\right)\right]\frac{\varepsilon_s}{\sigma_\varepsilon^2}\left(y_{s+h} - \alpha_h - \varepsilon_s\beta_h\right)\right\}.$$

If $s \neq t$ and $\varepsilon_t \perp y_s$, then the expectation is zero. If $s = t$, the the expectation is

$$-2\frac{E\left(\varepsilon_t^4\right)}{\sigma_\varepsilon^2}E\left(y_t - \mu_y\right).$$

Note that

$$\nabla^2 q_t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2\varepsilon_t\varepsilon_{t-u} & \varepsilon_t\varepsilon_{t-u}^2 + \varepsilon_t^2\varepsilon_{t-u} & 0 & \varepsilon_t^2\varepsilon_{t-u} + \varepsilon_{t-u}^2\varepsilon_t & 2\varepsilon_t^2\varepsilon_{t-u}^2 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\overline{H}_2 = E\left(\nabla^2 q_t\right) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\left(\sigma_\varepsilon^2\right)^2 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$-\frac{1}{2}Q\overline{H}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \left(\sigma_\varepsilon^2\right)^2 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

To compute

$$-\frac{1}{2}Q\overline{H}_2 E\left[Q\psi_{T-h-u} \otimes Q\psi_{T-h-u}\right]$$

we need to evaluate

$$E\left[\frac{1}{T-h-u}\sum_{t=1+u}^{T-h}\frac{\varepsilon_t}{\sigma_\varepsilon^2}\left(y_{t+h} - \alpha_h - \varepsilon_t\beta_h\right)\frac{1}{T-h-u}\sum_{s=1+u}^{T-h}\frac{\varepsilon_s}{\sigma_\varepsilon^2}\left(y_{s+h} - \alpha_h - \varepsilon_s\beta_h\right)\right].$$

If $t \neq s$ and $\varepsilon_t \perp y_s$, then

$$E\left[\frac{\varepsilon_t}{\sigma_\varepsilon^2}\left(y_{t+h} - \alpha_h - \varepsilon_t\beta_h\right)\frac{\varepsilon_s}{\sigma_\varepsilon^2}\left(y_{s+h} - \alpha_h - \varepsilon_s\beta_h\right)\right] = 0.$$

If $t = s$ and $\varepsilon_t \perp y_s$, then

$$E\left[\frac{\varepsilon_t^2}{\left(\sigma_\varepsilon^2\right)^2}\left(y_{t+h} - \alpha_h\right)^2\right] = E\left[\frac{1}{\sigma_\varepsilon^2}\left(y_{t+h} - \alpha_h\right)^2\right].$$

Putting all of this together, if $u > 0$, the bias is

$$-\frac{1}{T-h-u}\frac{T-h-2u}{T-h-u}\sigma_\varepsilon^2 E\left[\left(y_t - \mu_y\right)^2\right].$$

If $u = 0$, the bias is

$$-\frac{1}{T-h-u}2\frac{E\left(\varepsilon_t^4\right)}{\sigma_\varepsilon^2}E\left(y_t - \mu_y\right) - \frac{1}{T-h-u}\sigma_\varepsilon^2 E\left[\left(y_t - \mu_y\right)^2\right]$$

In the case of normal variation, this is

$$-\frac{1}{T-h-u}7\sigma_\varepsilon^2 E\left(y_t-\mu_y\right)$$

Notice that under the null hypothesis,

$$\gamma_{h,0}=\sigma_\varepsilon^2 E\left[\left(y_t-\mu_y\right)^2\right].$$

Under the null hypothesis that $\varepsilon_t \perp y_s$ and under normal variation,

$$E\left(\widehat{\gamma}_{h,0}\right)=\gamma_{h,0}\left(1-\frac{7}{T-h}\right)+O\left(T^{-3/2}\right)$$
$$E\left(\widehat{\gamma}_{h,u}\right)=\gamma_{h,u}-\frac{1}{T-h}\gamma_{h,0}+O\left(T^{-3/2}\right).$$

Clearly, when T is small, these distortions can be substantial, which explains why increasing the bandwidth of a Newey-West estimator makes the standard errors even smaller in expectation.

C Additional figures for the AR(1) example

In this appendix, we show figures analogous to those in the text for different values of ρ in the AR(1) example. In the main text, we set $\rho = 0.95$. Here, we consider $\rho = 0.9$ and $\rho = 0.99$.

C.1 $\rho = 0.9$

(a) Without Controls

(b) With Controls

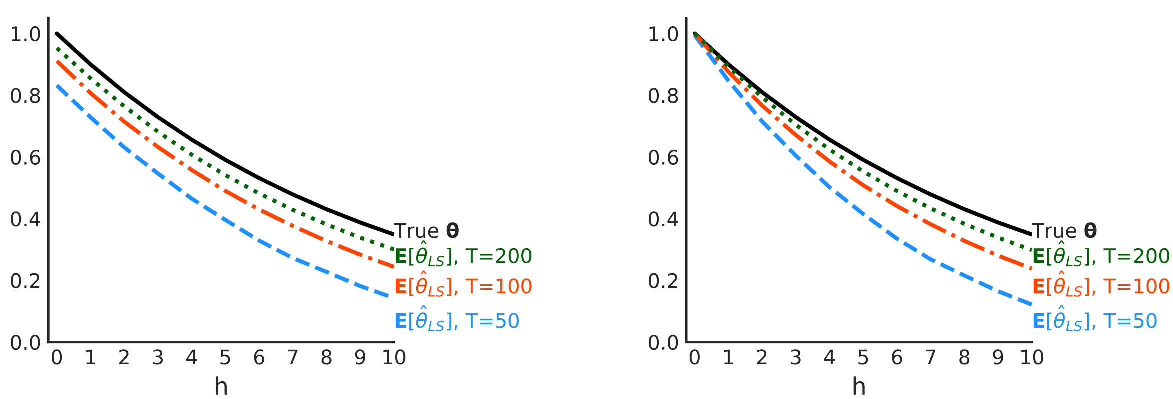


Figure 1: LP estimators are biased in empirically-relevant samples when y_t is an AR(1) with $\rho = 0.90$.

(a) Without Controls (b) With Controls

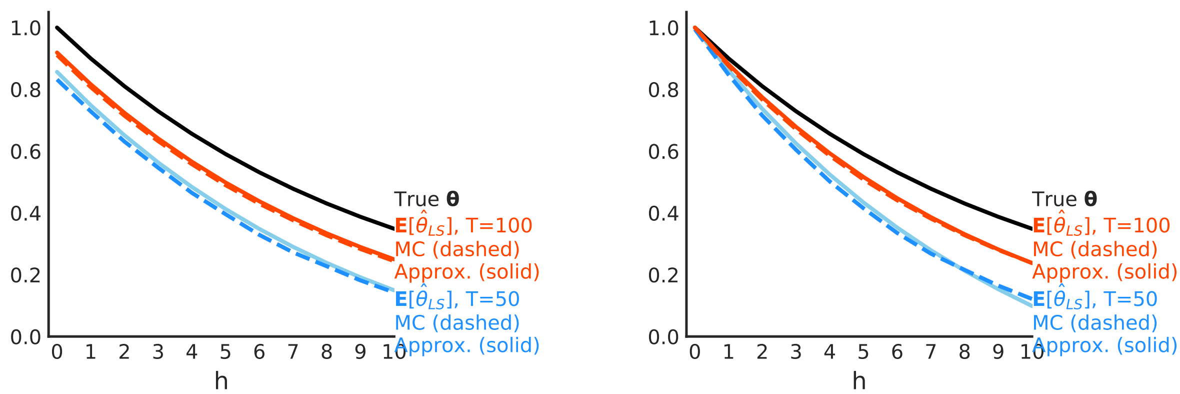


Figure 2: The bias approximation is accurate in our LPs.

(a) $T = 50$ (b) $T = 100$

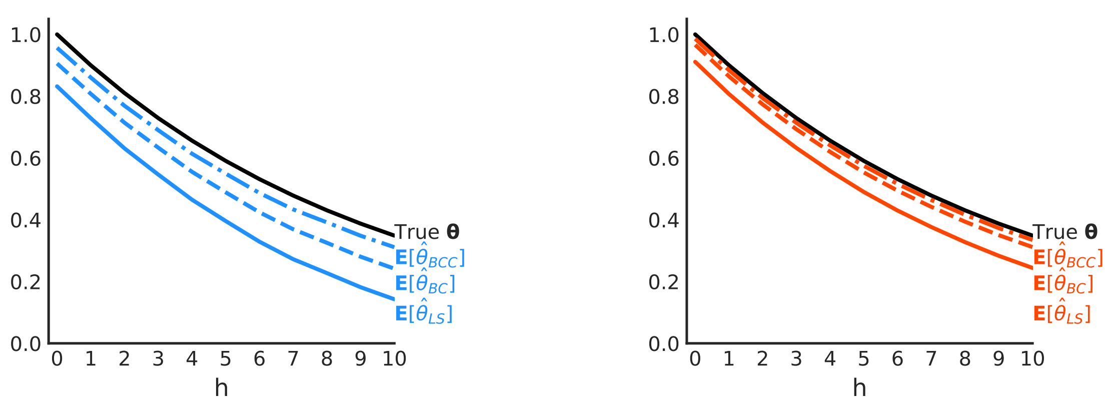


Figure 3: $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are closer than $\widehat{\theta}_{LS}$ to θ, on average, in our LPs without controls when y_t is an AR(1) with $\rho = 0.90$.

(a) $T = 50$ (b) $T = 100$

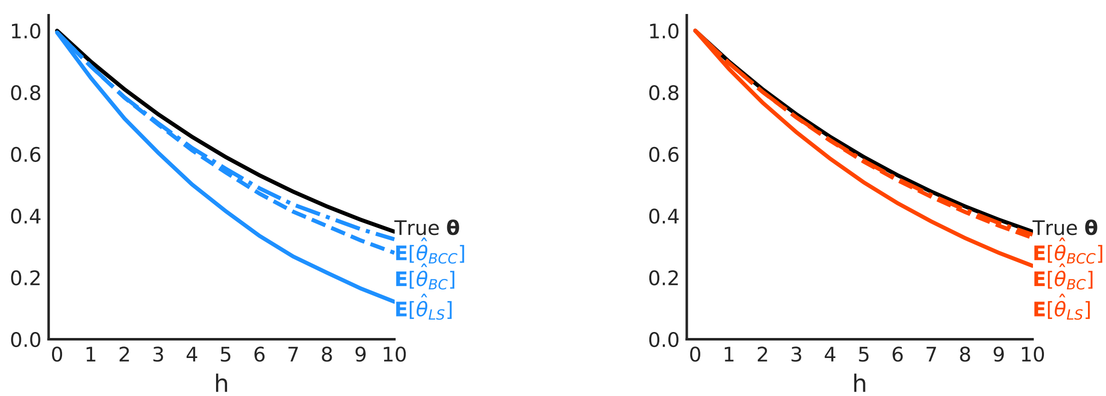


Figure 4: $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are closer than $\widehat{\theta}_{LS}$ to θ, on average, in our LPs with controls when y_t is an AR(1) with $\rho = 0.90$.

C.2 $\rho = 0.99$

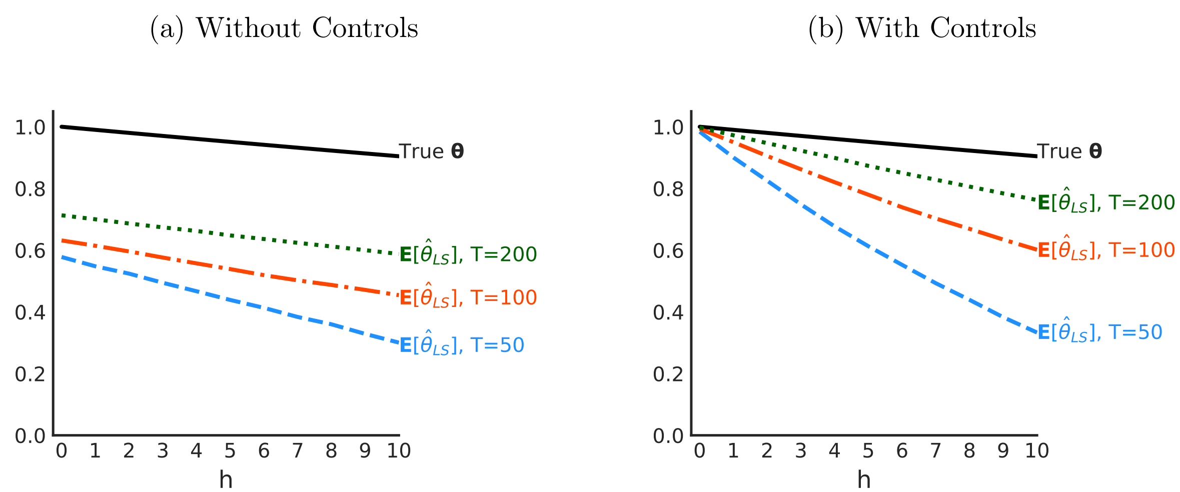


Figure 5: LP estimators are biased in empirically-relevant samples when y_t is an AR(1) with $\rho = 0.99$.

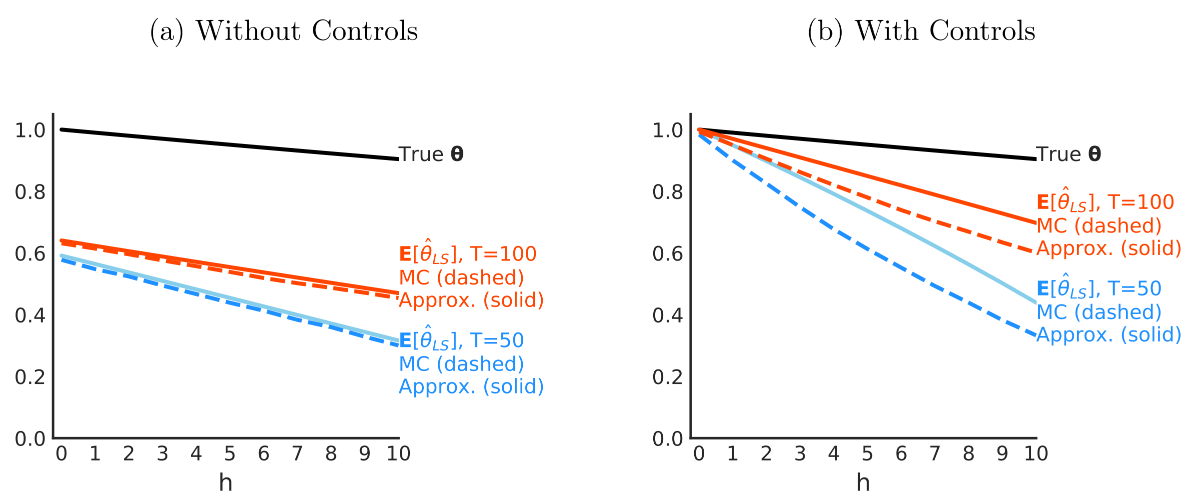


Figure 6: The bias approximation is accurate in our LPs.

(a) $T = 50$ (b) $T = 100$

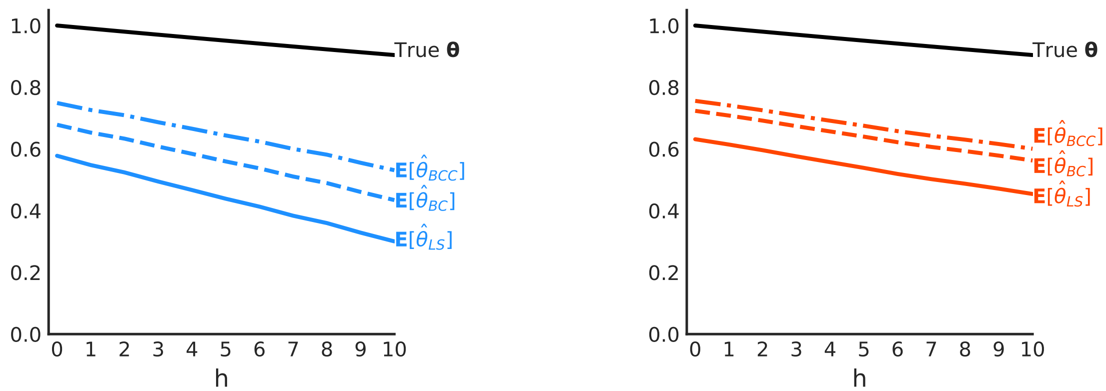


Figure 7: $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are closer than $\widehat{\theta}_{LS}$ to θ, on average, in our LPs without controls when y_t is an AR(1) with $\rho = 0.99$.

(a) $T = 50$ (b) $T = 100$

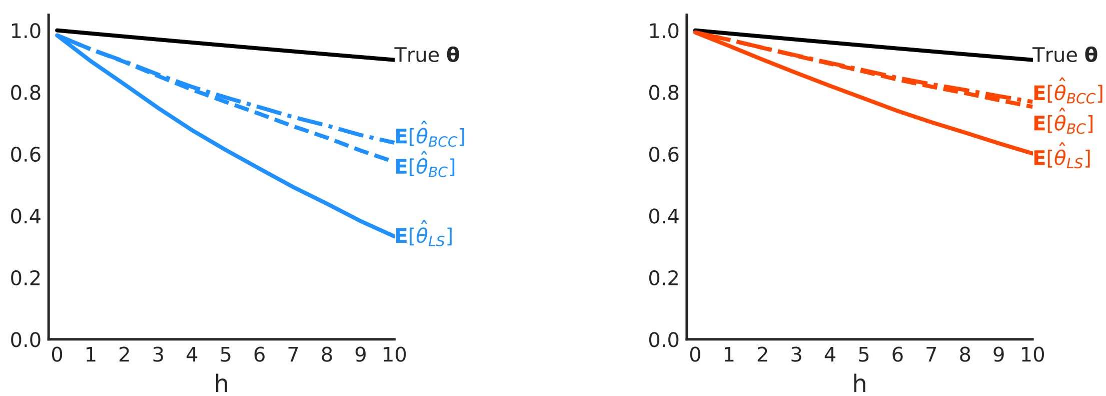


Figure 8: $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are closer than $\widehat{\theta}_{LS}$ to θ, on average, in our LPs with controls when y_t is an AR(1) with $\rho = 0.99$.

D An AR(2) example

In this appendix, we show figures analogous to those in the text for an AR(2) example. We specify the example so that

$$y_t = (\rho + \psi)y_{t-1} - \psi\rho y_{t-2} + \varepsilon_t + \nu_t.$$

This process delivers hump-shaped impulse response functions. We set $\rho = 0.95$ and $\psi = 0.4$. When we include controls, we set $c_t = [y_{t-1}, y_{t-2}]'$.

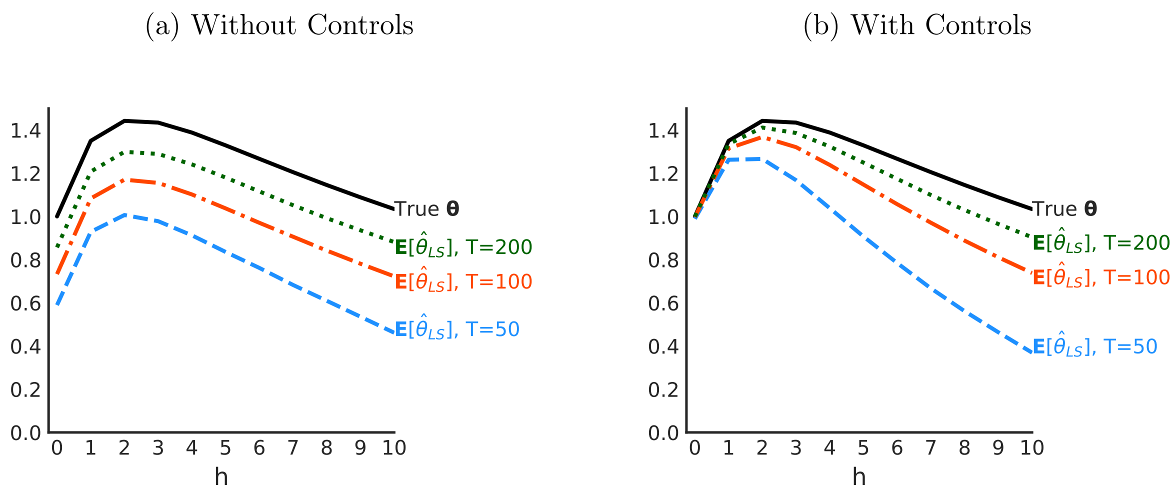


Figure 9: LP estimators are biased in empirically-relevant samples when y_t is an AR(2) with $\rho = 0.95$ and $\psi = 0.4$.

(a) Without Controls

(b) With Controls

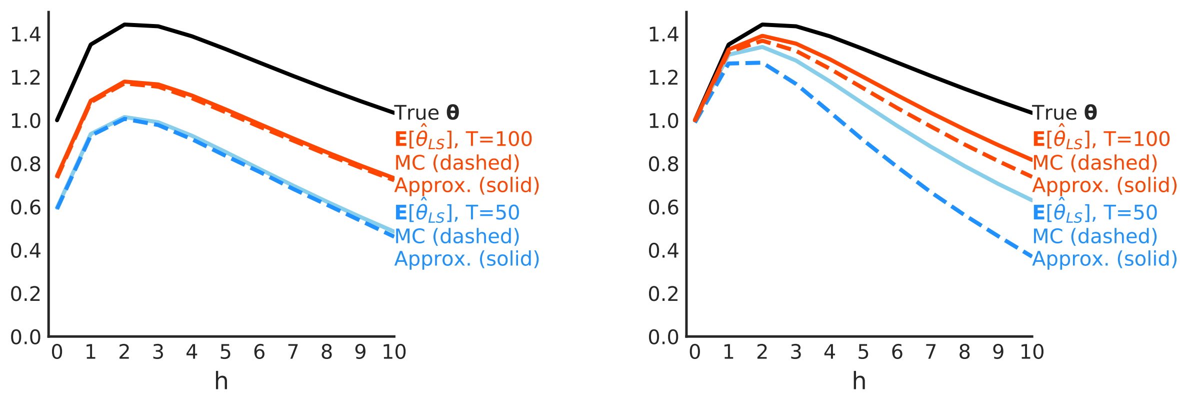


Figure 10: The bias approximation is accurate in our LPs.

(a) $T = 50$

(b) $T = 100$

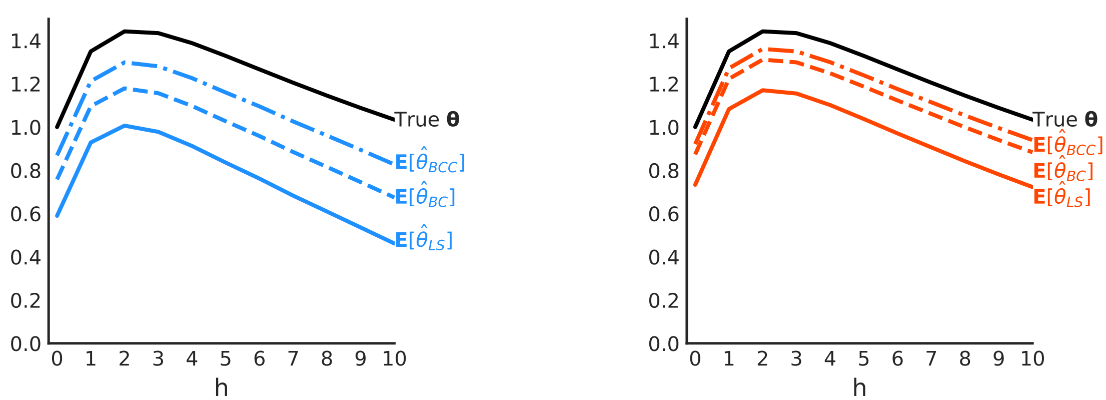


Figure 11: $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are closer than $\widehat{\theta}_{LS}$ to θ, on average, in our LPs without controls when y_t is an AR(1) with $\rho = 0.95$ and $\psi = 0.4$.

(a) $T = 50$

(b) $T = 100$

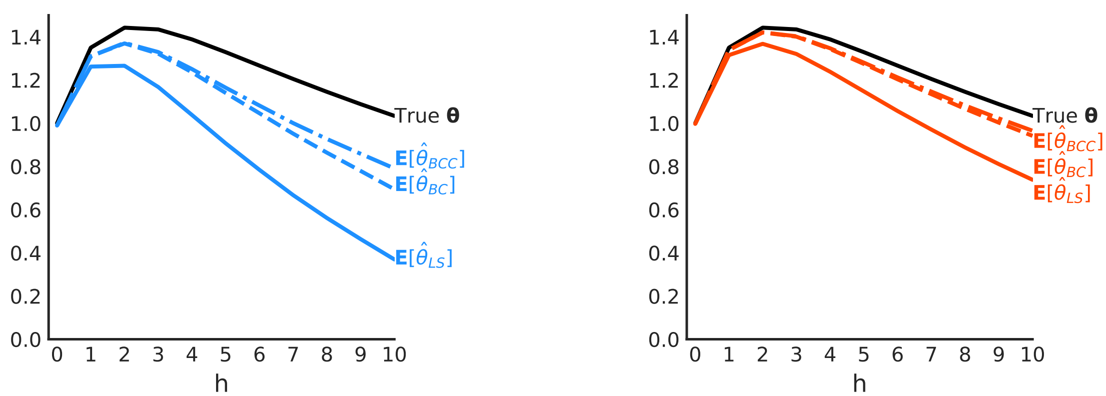


Figure 12: $\widehat{\theta}_{BC}$ and $\widehat{\theta}_{BCC}$ are closer than $\widehat{\theta}_{LS}$ to θ, on average, in our LPs with controls when y_t is an AR(2) with $\rho = 0.90$ and $\psi = 0.4$.

E Newey-West standard errors and fixed-b asymptotics

Sun (2014) and Lazarus et al. (2018) suggest that researchers should use fixed-b asymptotics when conducting inference using HAR estimators. For the Newey-West estimator, they suggest using a bandwidth of $1.3\sqrt{T-h}$. The asymptotic limiting distribution of the test statistic is not standard.

Table 4 shows results that are analogous to the results shown in Table 1, but using the bandwidth for the Newey-West estimator suggested by Lazarus et al. (2018) and the non-standard critical values to construct confidence sets using the Newey-West standard error. When one uses the fixed-b critical values, the performance of the Newey-West estimator improves somewhat, though the coverage probabilities are not uniformly better than those from the Huber-White standard errors.

Fixed-b asymptotics involve using a larger bandwidth for the Newey-West estimator and larger critical values than those implied by the asymptotic normal approximation. It turns out that the bandwidth is not that much larger in sample sizes typically found in the literature. As a result, the bias is not that much larger when the larger bandwidth is used. It is then not surprising that using a larger critical value improves the coverage probabilities given that the confidence intervals are too small.

Table 4: Coverage probability of different estimators of standard errors for $\widehat{\theta}_h$ in LP without controls when y_t is an AR(1) with $\rho = 0.95$ and $T = 50$ using fixed-b critical values

	$\widehat{\theta}_{h,LS}$, no controls		$\widehat{\theta}_{h,BCC}$, no controls		$\widehat{\theta}_{h,BCC}$, controls	
h	Huber-White	Newey-West	Huber-White	Newey-West	Huber-White	Newey-West
0	0.87	0.82	0.86	0.85	0.92	0.93
1	0.83	0.81	0.82	0.83	0.90	0.90
2	0.80	0.80	0.79	0.82	0.87	0.89
3	0.78	0.77	0.76	0.79	0.85	0.86
4	0.76	0.76	0.75	0.78	0.83	0.85
5	0.75	0.75	0.74	0.78	0.81	0.83
6	0.75	0.75	0.73	0.77	0.80	0.82
7	0.74	0.73	0.73	0.76	0.78	0.82
8	0.74	0.74	0.73	0.77	0.77	0.82
9	0.74	0.75	0.73	0.77	0.76	0.81
10	0.74	0.75	0.73	0.78	0.75	0.81

F Bias and the block bootstrap

An alternative approach to achieve bias correction in LPs is through bootstrapping.[24] Bootstrap methods construct approximation to the distribution of an estimator (for example), by resampling observables or the errors from a parametric model. In a time series context, where the dynamic relationship between observables or errors is important to preserve, block bootstrapping techniques—in which the resampling scheme seeks to preserve some of the correlation in the original data set—are typically used, as in Kilian and Kim (2011). We revisit the Monte Carlo simulations in Section 3.1 and attempt to correct for the finite-sample bias in LPs using the block bootstrap.[25] Figure 13 displays the bias correction from the block bootstrap when y_t is our AR(1) example.

Similar to results reported by Kilian and Kim (2011), the block bootstrap offers little in the way of LP bias correction. As in Kilian and Kim (2011), we set the block length to 4.[26] Given equation (4), it is not surprising that with a short block length the block bootstrap does little to bias correct. The block bootstrap works by maintaining the autocorrelation structure of the data in LPs within a given block, but by destroying the autocorrelation across blocks. By destroying the autocorrelation across blocks, the block bootstrap destroys some of the autocorrelation information needed to adjust the estimates. As a result, the block bootstrap underestimates the bias in LPs, rendering bias correction based on the block bootstrap relatively ineffective.

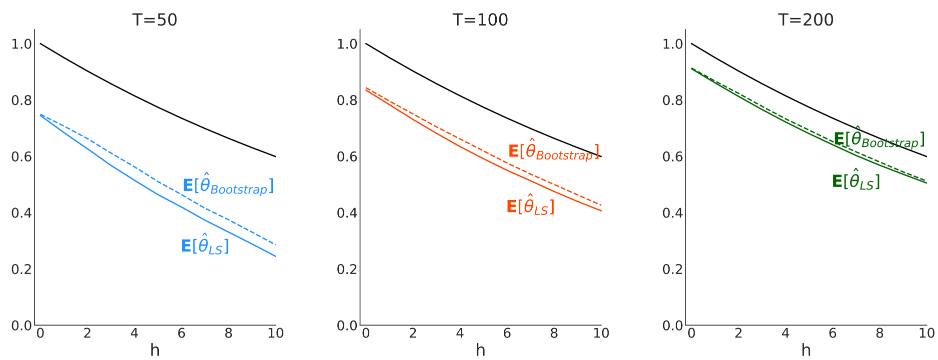
Longer block lengths make the block bootstrap more effective at bias correcting, but they reduce the number of non-overlapping blocks in a dataset. Kilian and Kim (2011) report that longer block lengths lead to worse coverage probabilities.

[24] See Hall (1992) for a textbook treatment.

[25] We use the same block length as Kilian and Kim (2011).

[26] Results are similar when we set the block length to $(T-h)^{1/3}$, as suggested by Hall et al. (1995).

(a) Without Controls



(b) With Controls

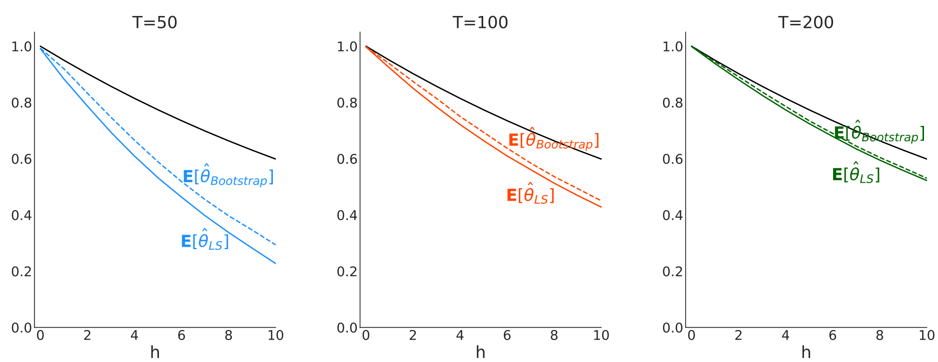


Figure 13: $\widehat{\beta}_{Bootstrap}$ provides little bias correction in our LPs y_t is an AR(1) with $\rho = 0.95$.

G Lag augmentation in the AR(1) example

In this section, we consider the effects of lag augmentation in our AR(1) example. Olea and Plagborg-Møller (2020) argue that lag augmentation improves the performance of LPs. There are some important differences between the LP setting we consider and the setting in Olea and Plagborg-Møller (2020). First, we consider an LP in which a constant must be estimated, while Olea and Plagborg-Møller (2020) do not. Second, we consider LP estimates of impulse responses to identified shocks that that the researcher brings to the LP, rather than identifying the shock as a part of the LP system. Here, we consider the effects of lag augmentation in our setup. The main example in Olea and Plagborg-Møller (2020) is an AR(1) without a constant, so the example from our paper has few other differences than the two important differences identified above.

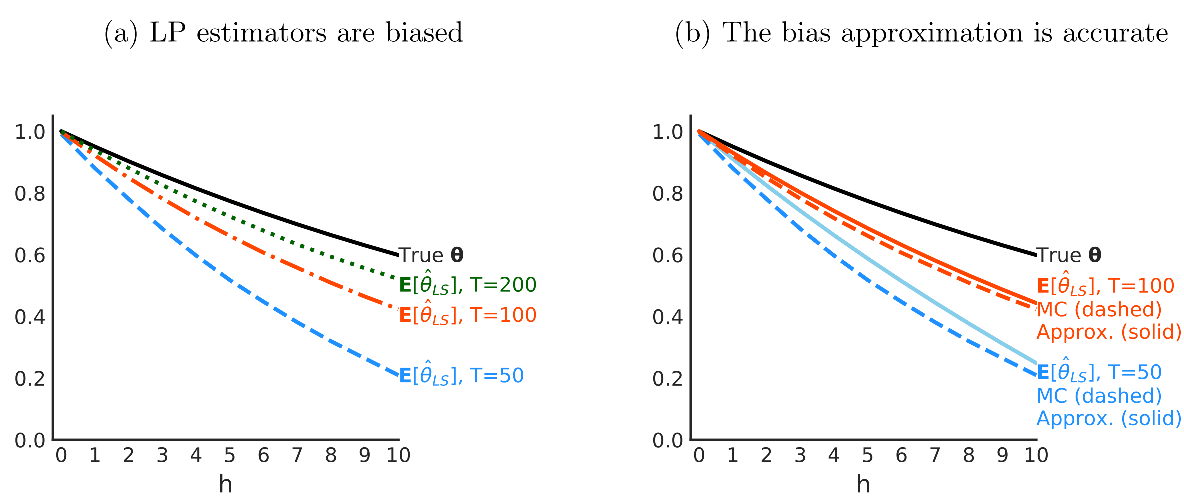


Figure 14: LP with lag augmentation when y_t is an AR(1) with $\rho = 0.95$.

Figure 14 shows the Monte Carlo mean of $\widehat{\theta}_{LS}$ in our AR(1) example when $\rho = 0.95$. The bias in the impulse response estimator is essentially the same as without lag augmentation.

Figure 15 shows the Monte Carlo mean of $\widehat{\theta}_{LS}$, $\widehat{\theta}_{BC}$, and $\widehat{\theta}_{BCC}$ in our AR(1) example when $\rho = 0.95$. The mean bias correction is essentially the same as without lag augmentation.

Table 5 shows the coverage probabilities for confidence intervals constructed using Huber-White and Newey-West standard errors in our AR(1) example with lag agumentation when $\rho = 0.95$. The coverage probabilities are similar to their values without lag augmentation.

(a) $T = 50$ (b) $T = 100$

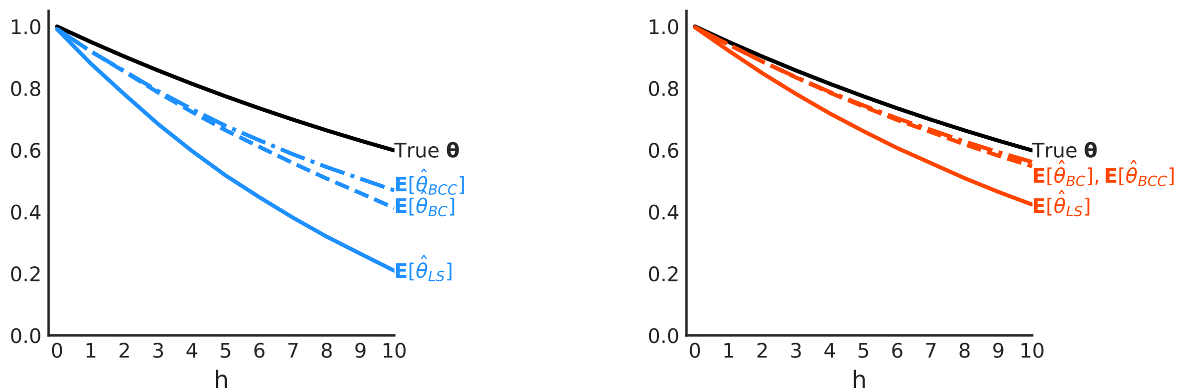


Figure 15: Bias correction in an LP with lag augmentation when y_t is an AR(1) with $\rho = 0.95$.

Table 5: Coverage probability of different estimators of standard errors for $\widehat{\theta}_h$ in LP with lag augmentation when y_t is an AR(1) with $\rho = 0.95$ and $T = 50$

	$\widehat{\theta}_{h,LS}$		$\widehat{\theta}_{h,BCC}$	
h	Huber-White	Newey-West	Huber-White	Newey-West
0	0.92	0.91	0.92	0.91
1	0.89	0.87	0.90	0.88
2	0.85	0.83	0.87	0.85
3	0.83	0.80	0.84	0.82
4	0.80	0.77	0.82	0.80
5	0.77	0.74	0.80	0.78
6	0.76	0.73	0.78	0.76
7	0.74	0.71	0.77	0.75
8	0.73	0.69	0.76	0.74
9	0.71	0.68	0.75	0.73
10	0.70	0.67	0.74	0.72

H Derivation of bias in an LP without controls using differences in y_t

Some authors use $y_{t+h} - y_{t-1}$ or $y_{t+h} - y_t$ as the left hand side variable for their LP. If controls are included and the researcher uses $y_{t+h} - y_{t-1}$ as the left-hand side variable, it has no effect on our earlier derivations. If $\theta_0 = 0$ and controls are included, it will also have no effect on our earlier derivation if the researcher uses $y_{t+h} - y_t$ as the left-hand-side variable. If controls are not included in the regression, then using the difference of y_t reduces the bias. Intuitively, using the difference of y_t removes much of the persistence in the dependent variable, and for near-unit-root processes is almost well-specified so that the regression errors are nearly MA(h+1) processes. Lunsford (2020) uses an LP without controls and $y_{t+h} - y_t$, so we focus on this setup. Given that Lunsford (2020) works with high-frequency monetary policy and forward guidance shocks, along with monthly macroeconomic data, it is reasonable to assume that $\theta_0 = 0$. The algebra for the case when $y_{t+h} - y_{t-1}$ is the left-hand-side variable is similar.

To show that using $y_{t+h} - y_t$ as the left-hand-side variable reduces bias, consider that the moment conditions associated with the OLS esitmator are

$$E[q_t] = 0$$

where

$$q_t \equiv \begin{bmatrix} y_{t+h} - y_t - \alpha_h - \beta_h \varepsilon_t \\ \varepsilon_t \left(y_{t+h} - y_t - \alpha_h - \beta_h \varepsilon_t \right) \end{bmatrix}.$$

We have that

$$\nabla^1 q_t = \begin{bmatrix} -1 & -\varepsilon_t \\ -\varepsilon_t & -\varepsilon_t^2 \end{bmatrix}$$

$$H_1 = \frac{1}{T-h} \sum_{t=1}^{T-h} \begin{bmatrix} -1 & -\varepsilon_t \\ -\varepsilon_t & -\varepsilon_t^2 \end{bmatrix}$$

$$\overline{H}_1 = \begin{bmatrix} -1 & -\mu_\varepsilon \\ -\mu_\varepsilon & -\left(\sigma_\varepsilon^2 + \mu_\varepsilon^2\right) \end{bmatrix}$$

$$\nabla^2 q_t = H_2 = \overline{H}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$V = \frac{1}{T-h} \sum_{t=1}^{T-h} \begin{bmatrix} 0 & -\left(\varepsilon_t - \mu_\varepsilon\right) \\ -\left(\varepsilon_t - \mu_\varepsilon\right) & -\left(\varepsilon_t^2 - \left(\sigma_\varepsilon^2 + \mu_\varepsilon^2\right)\right) \end{bmatrix}$$

$$Q = \begin{bmatrix} -\left(1 + \frac{\mu_\varepsilon^2}{\sigma_\varepsilon^2}\right) & \frac{\mu_\varepsilon}{\sigma_\varepsilon^2} \\ \frac{\mu_\varepsilon}{\sigma_\varepsilon^2} & -\frac{1}{\sigma_\varepsilon^2} \end{bmatrix}$$

Because $\overline{H}_2$ is a matrix of zeros,

$$B = E\left\{ QVQ\psi_{T-h} \right\}.$$

Note that

$$B = E\left\{Q\left(H_1 - \overline{H}_1\right)Q\psi_{T-h}\right\} = E\left\{QH_1Q\psi_{T-h}\right\}.$$

Also,

$$\begin{aligned} QH_1 &= \frac{1}{T-h}\sum_{t=1}^{T-h}\begin{bmatrix} -\left(1+\frac{\mu_\varepsilon^2}{\sigma_\varepsilon^2}\right) & \frac{\mu_\varepsilon}{\sigma_\varepsilon^2} \\ \frac{\mu_\varepsilon}{\sigma_\varepsilon^2} & -\frac{1}{\sigma_\varepsilon^2}\end{bmatrix}\begin{bmatrix} -1 & -\varepsilon_t \\ -\varepsilon_t & -\varepsilon_t^2\end{bmatrix} \\ &= \frac{1}{T-h}\sum_{t=1}^{T-h}\begin{bmatrix} \left(1-\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)\right) & \left(1-\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)\right)\varepsilon_t \\ \frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon) & \frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)\varepsilon_t\end{bmatrix}\end{aligned}$$

and

$$\begin{aligned} QH_1Q &= \frac{1}{T-h}\sum_{t=1}^{T-h}\begin{bmatrix} \left(1-\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)\right) & \left(1-\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)\right)\varepsilon_t \\ \frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon) & \frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)\varepsilon_t\end{bmatrix}\begin{bmatrix} -\left(1+\frac{\mu_\varepsilon^2}{\sigma_\varepsilon^2}\right) & \frac{\mu_\varepsilon}{\sigma_\varepsilon^2} \\ \frac{\mu_\varepsilon}{\sigma_\varepsilon^2} & -\frac{1}{\sigma_\varepsilon^2}\end{bmatrix} \\ &= \frac{1}{T-h}\sum_{t=1}^{T-h}\begin{bmatrix} -\left(1-2\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)+\left(\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}\right)^2(\varepsilon_t-\mu_\varepsilon)^2\right) & -\left(\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)-\frac{\mu_\varepsilon}{(\sigma_\varepsilon^2)^2}(\varepsilon_t-\mu_\varepsilon)^2\right) \\ -\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)+\frac{\mu_\varepsilon}{(\sigma_\varepsilon^2)^2}(\varepsilon_t-\mu_\varepsilon)^2 & -\frac{1}{(\sigma_\varepsilon^2)^2}(\varepsilon_t-\mu_\varepsilon)^2\end{bmatrix}\end{aligned}$$

Then

$$\begin{aligned} QH_1Q\psi_{T-h} =& \frac{1}{T-h}\sum_{t=1}^{T-h}\begin{bmatrix} -\left(1-2\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)+\left(\frac{\mu_\varepsilon}{\sigma_\varepsilon^2}\right)^2(\varepsilon_t-\mu_\varepsilon)^2\right) & -\left(\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)-\frac{\mu_\varepsilon}{(\sigma_\varepsilon^2)^2}(\varepsilon_t-\mu_\varepsilon)^2\right) \\ -\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)+\frac{\mu_\varepsilon}{(\sigma_\varepsilon^2)^2}(\varepsilon_t-\mu_\varepsilon)^2 & -\frac{1}{(\sigma_\varepsilon^2)^2}(\varepsilon_t-\mu_\varepsilon)^2\end{bmatrix} \\ &\times\frac{1}{T-h}\sum_{s=1}^{T-h}\begin{bmatrix} y_{s+h}-y_s-\alpha_h-\beta_h\varepsilon_s \\ \varepsilon_s\left(y_{s+h}-y_s-\alpha_h-\beta_h\varepsilon_s\right)\end{bmatrix}\end{aligned}$$

We only need the expectation of the second row of this expression. To calculate it, fix t and consider

$$\begin{aligned} E\left\{-\left(\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)-\frac{\mu_\varepsilon}{\left(\sigma_\varepsilon^2\right)^2}(\varepsilon_t-\mu_\varepsilon)^2\right)(y_{s+h}-y_s-\alpha_h-\beta_h\varepsilon_s)-\frac{1}{\left(\sigma_\varepsilon^2\right)^2}(\varepsilon_t-\mu_\varepsilon)^2\varepsilon_s(y_{s+h}-y_s-\alpha_h-\beta_h\varepsilon_s)\right\} &= \\ E\left\{-\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)(y_{s+h}-y_s-\alpha_h-\beta_h\varepsilon_s)-\frac{1}{\left(\sigma_\varepsilon^2\right)^2}(\varepsilon_t-\mu_\varepsilon)^2(\varepsilon_s-\mu_\varepsilon)(y_{s+h}-y_s-\alpha_h-\beta_h\varepsilon_s)\right\} &= \\ E\left\{-\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)(y_{s+h}-y_s-\alpha_h-\beta_h\varepsilon_s)\right\} &= \\ E\left\{-\frac{1}{\sigma_\varepsilon^2}(\varepsilon_t-\mu_\varepsilon)(y_{s+h}-y_t+y_t-y_s-\alpha_h-\beta_h\varepsilon_s)\right\}& \end{aligned}$$

If $s=t$, or $s<t-h$, then this expression is zero. If $t-h\le s<t$, then this expression is equal to $-\beta_{s+h-t}$. If $t<s$, this expression is equal to

$$-\beta_{s+h-t}+\beta_{s-t}$$

The existence of the second term mitigates the bias relative to the case where the difference is not taken.

I Sources for meta study

Abiad, M. A., D. Furceri, and P. Topalova (2015): *The macroeconomic effects of public investment: evidence from advanced economies*, 15-95, International Monetary Fund.

Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (2019): "Democracy does cause growth," *Journal of Political Economy*, 127, 47–100.

ADB, A. A., D. Furceri, and P. T. IMF (2016): "The macroeconomic effects of public investment: Evidence from advanced economies," *Journal of Macroeconomics*, 50, 224–240.

Adler, G., M. R. A. Duval, D. Furceri, K. Sinem, K. Koloskova, M. Poplawski-Ribeiro, et al. (2017): *Gone with the headwinds: Global productivity*, International Monetary Fund.

Alesina, A., O. Barbiero, C. Favero, F. Giavazzi, and M. Paradisi (2015a): "Austerity in 2009–13," *Economic Policy*, 30, 383–437.

——— (2017): "The effects of fiscal consolidations: Theory and evidence," Tech. rep., National Bureau of Economic Research.

Alesina, A., C. Favero, and F. Giavazzi (2015b): "The output effect of fiscal consolidation plans," *Journal of International Economics*, 96, S19–S42.

Amior, M. and A. Manning (2018): "The persistence of local joblessness," *American Economic Review*, 108, 1942–70.

Andreasen, M. M., J. Fernández-Villaverde, and J. F. Rubio-Ramírez (2017): "The pruned state-space system for non-linear DSGE models: Theory and empirical applications," *The Review of Economic Studies*, 85, 1–49.

Angrist, J. D., Ò. Jordà, and G. M. Kuersteiner (2018): "Semiparametric estimates of monetary policy effects: string theory revisited," *Journal of Business & Economic Statistics*, 36, 371–387.

Arezki, R., V. A. Ramey, and L. Sheng (2017): "News shocks in open economies: Evidence from giant oil discoveries," *The quarterly journal of economics*, 132, 103–155.

Arregui, N., J. Beneš, I. Krznar, and S. Mitra (2013): "Evaluating the net benefits of macroprudential policy: A cookbook," .

Auerbach, A. J. and Y. Gorodnichenko (2012): "Fiscal multipliers in recession and expansion," in *Fiscal policy after the financial crisis*, University of Chicago Press, 63–98.

——— (2017): "Fiscal multipliers in Japan," *Research in Economics*, 71, 411–421.

Ball, L. M., D. Furceri, M. D. Leigh, and M. P. Loungani (2013): *The distributional effects of fiscal consolidation*, 13-151, International Monetary Fund.

Banerjee, R., M. B. Devereux, and G. Lombardo (2016): "Self-oriented monetary policy, global financial markets and excess volatility of international capital flows," *Journal of International Money and Finance*, 68, 275–297.

Banerjee, R. N. and H. Mio (2018): "The impact of liquidity regulation on banks," *Journal of Financial intermediation*, 35, 30–44.

Barnichon, R. and C. Brownlees (2019): "Impulse response estimation by smooth local projections," *Review of Economics and Statistics*, 101, 522–530.

Basher, S. A., A. A. Haug, and P. Sadorsky (2012): "Oil prices, exchange rates and emerging stock markets," *Energy Economics*, 34, 227–240.

Baumeister, C. and L. Kilian (2016): "Lower oil prices and the US economy: Is this time different?" *Brookings Papers on Economic Activity*, 2016, 287–357.

Bayer, C., R. Lütticke, L. Pham-Dao, and V. Tjaden (2019): "Precautionary savings, illiquid assets, and the aggregate consequences of shocks to household income risk," *Econometrica*, 87, 255–290.

Ben Zeev, N. and E. Pappa (2017): "Chronicle of a war foretold: The macroeconomic effects of anticipated defence spending shocks," *The Economic Journal*, 127, 1568–1597.

Berton, F., S. Mocetti, A. F. Presbitero, and M. Richiardi (2018): "Banks, firms, and jobs," *The Review of Financial Studies*, 31, 2113–2156.

Bhattacharya, U., N. Galpin, R. Ray, and X. Yu (2009): "The role of the media in the internet IPO bubble," *Journal of Financial and Quantitative Analysis*, 44, 657–682.

Bluedorn, J. C. and C. Bowdler (2011): "The open economy consequences of US monetary policy," *Journal of International Money and Finance*, 30, 309–336.

Borio, C., M. Drehmann, and K. Tsatsaronis (2014): "Stress-testing macro stress testing: does it live up to expectations?" *Journal of Financial Stability*, 12, 3–15.

Borio, C. E., E. Kharroubi, C. Upper, and F. Zampolli (2016): "Labour reallocation and productivity dynamics: financial causes, real consequences," .

Born, B., G. J. Müller, and J. Pfeifer (2014): "Does austerity pay off?" *Review of Economics and Statistics*, 1–45.

Boss, M., G. Fenz, J. Pann, C. Puhr, M. Schneider, E. Ubl, et al. (2009): "modeling credit risk through the Austrian business cycle: An update of the OeNb model," *Financial Stability Report*, 17, 85–101.

Brady, R. R. (2011): "Measuring the diffusion of housing prices across space and over time," *Journal of Applied Econometrics*, 26, 213–231.

Butt, N., R. Churm, M. F. McMahon, A. Morotz, and J. F. Schanz (2014): "QE and the bank lending channel in the United Kingdom," .

Caggiano, G., E. Castelnuovo, V. Colombo, and G. Nodari (2015): "Estimating fiscal multipliers: News from a non-linear world," *The Economic Journal*, 125, 746–776.

Caldara, D. and E. Herbst (2019): "Monetary policy, real activity, and credit spreads: Evidence from bayesian proxy svars," *American Economic Journal: Macroeconomics*, 11, 157–92.

Caselli, F. G. and A. Roitman (2016): "Nonlinear exchange-rate pass-through in emerging markets," *International Finance*.

Chiţu, L., B. Eichengreen, and A. Mehl (2014): "When did the dollar overtake sterling as the leading international currency? Evidence from the bond markets," *Journal of Development Economics*, 111, 225–245.

Chodorow-Reich, G. and L. Karabarbounis (2016): “The limited macroeconomic effects of unemployment benefit extensions,” Tech. rep., National Bureau of Economic Research.

Chong, Y., Ò. Jordà, and A. M. Taylor (2012): “The Harrod–Balassa–Samuelson Hypothesis: Real Exchange Rates And Their Long-Run Equilibrium,” *International Economic Review*, 53, 609–634.

Cloyne, J. and P. Hürtgen (2016): “The macroeconomic effects of monetary policy: a new measure for the United Kingdom,” *American Economic Journal: Macroeconomics*, 8, 75–102.

Coibion, O. and Y. Gorodnichenko (2012): “What can survey forecasts tell us about information rigidities?” *Journal of Political Economy*, 120, 116–159.

Coibion, O., Y. Gorodnichenko, L. Kueng, and J. Silvia (2012): “Innocent bystanders? Monetary policy and inequality in the US,” Tech. rep., National Bureau of Economic Research.

——— (2017): “Innocent Bystanders? Monetary policy and inequality,” *Journal of Monetary Economics*, 88, 70–89.

Dabla-Norris, M. E., M. S. Guo, M. V. Haksar, M. Kim, M. K. Kochhar, K. Wiseman, and A. Zdzienicka (2015): *The new normal: A sector-level perspective on productivity trends in advanced economies*, International Monetary Fund.

De Cos, P. H. and E. Moral-Benito (2016): “Fiscal multipliers in turbulent times: the case of Spain,” *Empirical Economics*, 50, 1589–1625.

Dessaint, O. and A. Matray (2017): “Do managers overreact to salient risks? Evidence from hurricane strikes,” *Journal of Financial Economics*, 126, 97–121.

Dupor, B., J. Han, and Y.-C. Tsai (2009): “What do technology shocks tell us about the New Keynesian paradigm?” *Journal of Monetary Economics*, 56, 560–569.

Eichler, M. (2013): “Causal inference with multiple time series: principles and problems,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371, 20110613.

Favara, G. and J. Imbs (2015): "Credit supply and the price of housing," *American Economic Review*, 105, 958–92.

Fazzari, S. M., J. Morley, and I. Panovska (2015): "State-dependent effects of fiscal policy," *Studies in Nonlinear Dynamics & Econometrics*, 19, 285–315.

Francis, N., M. T. Owyang, J. E. Roush, and R. DiCecio (2014): "A flexible finite-horizon alternative to long-run restrictions with an application to technology shocks," *Review of Economics and Statistics*, 96, 638–647.

Funke, M., M. Schularick, and C. Trebesch (2016): "Going to extremes: Politics after financial crises, 1870–2014," *European Economic Review*, 88, 227–260.

Furceri, D., M. L. E. Bernal-Verdugo, and M. D. M. Guillaume (2012a): *Crises, labor market policy, and unemployment*, 12-65, International Monetary Fund.

Furceri, D., S. Guichard, and E. Rusticelli (2012b): "The effect of episodes of large capital inflows on domestic credit," *The North American Journal of Economics and Finance*, 23, 325–344.

Furceri, D., P. Loungani, and A. Zdzienicka (2018): "The effects of monetary policy shocks on inequality," *Journal of International Money and Finance*, 85, 168–186.

Furceri, D. and A. Zdzienicka (2011): "How costly are debt crises?" *IMF working papers*, 1–29.

——— (2012): "How costly are debt crises?" *Journal of International Money and Finance*, 31, 726–742.

Gal, P. N. and A. Hijzen (2016): *The short-term impact of product market reforms: A cross-country firm-level analysis*, International Monetary Fund.

Gertler, M. and S. Gilchrist (2018): "What happened: Financial factors in the great recession," *Journal of Economic Perspectives*, 32, 3–30.

Hall, A. R., A. Inoue, J. M. Nason, and B. Rossi (2012): "Information criteria for impulse response function matching estimation of DSGE models," *Journal of Econometrics*, 170, 499–518.

HAMILTON, J. D. (2011): "Nonlinearities and the macroeconomic effects of oil prices," *Macroeconomic dynamics*, 15, 364–378.

HAUTSCH, N. AND R. HUANG (2012): "The market impact of a limit order," *Journal of Economic Dynamics and Control*, 36, 501–522.

HOLLO, D., M. KREMER, AND M. LO DUCA (2012): "CISS-a composite indicator of systemic stress in the financial system," .

JORDÀ, Ò. (2009): "Simultaneous confidence regions for impulse responses," *The Review of Economics and Statistics*, 91, 629–647.

JORDÀ, Ò. AND M. MARCELLINO (2010): "Path forecast evaluation," *Journal of Applied Econometrics*, 25, 635–662.

JORDÀ, Ò., M. SCHULARICK, AND A. M. TAYLOR (2013): "When credit bites back," *Journal of Money, Credit and Banking*, 45, 3–28.

——— (2015a): "Betting the house," *Journal of International Economics*, 96, S2–S18.

——— (2015b): "Leveraged bubbles," *Journal of Monetary Economics*, 76, S1–S20.

——— (2016): "Sovereigns versus banks: credit, crises, and consequences," *Journal of the European Economic Association*, 14, 45–79.

JORDÀ, Ò. AND A. M. TAYLOR (2016): "The time for austerity: estimating the average treatment effect of fiscal policy," *The Economic Journal*, 126, 219–255.

KILIAN, L. AND Y. J. KIM (2011): "How reliable are local projection estimators of impulse responses?" *Review of Economics and Statistics*, 93, 1460–1466.

KILIAN, L. AND R. J. VIGFUSSON (2011): "Nonlinearities in the oil price–output relationship," *Macroeconomic Dynamics*, 15, 337–363.

——— (2017): "The role of oil price shocks in causing US recessions," *Journal of Money, Credit and Banking*, 49, 1747–1776.

KRAAY, A. (2014): "Government spending multipliers in developing countries: evidence from lending by official creditors," *American Economic Journal: Macroeconomics*, 6, 170–208.

Krishnamurthy, A. and T. Muir (2017): "How credit cycles across a financial crisis," Tech. rep., National Bureau of Economic Research.

Leduc, S. and D. Wilson (2013): "Roads to prosperity or bridges to nowhere? Theory and evidence on the impact of public infrastructure investment," *NBER Macroeconomics Annual*, 27, 89–142.

Leigh, M. D., W. Lian, M. Poplawski-Ribeiro, R. Szymanski, V. Tsyrennikov, and H. Yang (2017): *Exchange rates and trade: A disconnect?*, International Monetary Fund.

Listorti, G. and R. Esposti (2012): "Horizontal price transmission in agricultural markets: fundamental concepts and open empirical issues," *Bio-based and Applied Economics Journal*, 1, 81–108.

Luetticke, R. (2018): "Transmission of monetary policy with heterogeneity in household portfolios," .

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2016): "Currency value," *The Review of Financial Studies*, 30, 416–441.

Mertens, K. and J. L. Montiel Olea (2018): "Marginal tax rates and income: New time series evidence," *The Quarterly Journal of Economics*, 133, 1803–1884.

Mian, A., A. Sufi, and E. Verner (2017): "Household debt and business cycles worldwide," *The Quarterly Journal of Economics*, 132, 1755–1817.

Miyamoto, W., T. L. Nguyen, and D. Sergeyev (2018): "Government spending multipliers under the zero lower bound: Evidence from Japan," *American Economic Journal: Macroeconomics*, 10, 247–77.

Moretti, E. and D. J. Wilson (2017): "The effect of state taxes on the geographical location of top earners: evidence from star scientists," *American Economic Review*, 107, 1858–1903.

Nakamura, E. and J. Steinsson (2018): "Identification in macroeconomics," *Journal of Economic Perspectives*, 32, 59–86.

Ottonello, P. and T. Winberry (2018): "Financial heterogeneity and the investment channel of monetary policy," Tech. rep., National Bureau of Economic Research.

Owyang, M. T., V. A. Ramey, and S. Zubairy (2013): "Are government spending multipliers greater during periods of slack? Evidence from twentieth-century historical data," *American Economic Review*, 103, 129–34.

Ramey, V. A. (2016): "Macroeconomic shocks and their propagation," in *Handbook of macroeconomics*, Elsevier, vol. 2, 71–162.

Ramey, V. A. and S. Zubairy (2018): "Government spending multipliers in good times and in bad: evidence from US historical data," *Journal of Political Economy*, 126, 850–901.

Riera-Crichton, D., C. A. Vegh, and G. Vuletin (2015): "Procyclical and countercyclical fiscal multipliers: Evidence from OECD countries," *Journal of International Money and Finance*, 52, 15–31.

Romer, C. D. and D. H. Romer (2015): "New evidence on the impact of financial crises in advanced countries," Tech. rep., National Bureau of Economic Research.

——— (2017): "New evidence on the aftermath of financial crises in advanced countries," *American Economic Review*, 107, 3072–3118.

Salomons, A. et al. (2018): "Is automation labor-displacing? Productivity growth, employment, and the labor share," Tech. rep., National Bureau of Economic Research.

Santoro, E., I. Petrella, D. Pfajfar, and E. Gaffeo (2014): "Loss aversion and the asymmetric transmission of monetary policy," *Journal of Monetary Economics*, 68, 19–36.

Schaller, J. (2013): "For richer, if not for poorer? Marriage and divorce over the business cycle," *Journal of Population Economics*, 26, 1007–1033.

Stock, J. H. and M. W. Watson (2018): "Identification and estimation of dynamic causal effects in macroeconomics using external instruments," *The Economic Journal*, 128, 917–948.

Stoyanov, A. and N. Zubanov (2012): "Productivity spillovers across firms through worker mobility," *American Economic Journal: Applied Economics*, 4, 168–98.

Taylor, A. M. (2015): "Credit, financial stability, and the macroeconomy," *Annu. Rev. Econ.*, 7, 309–339.

Tenreyro, S. and G. Thwaites (2016): "Pushing on a string: US monetary policy is less powerful in recessions," *American Economic Journal: Macroeconomics*, 8, 43–74.

Teulings, C. N. and N. Zubanov (2014): "Is economic recovery a myth? Robust estimation of impulse responses," *Journal of Applied Econometrics*, 29, 497–514.

Yang, J., H. Guo, and Z. Wang (2006): "International transmission of inflation among G-7 countries: A data-determined VAR analysis," *Journal of Banking & Finance*, 30, 2681–2700.

Zidar, O. (2019): "Tax cuts for whom? Heterogeneous effects of income tax changes on growth and employment," *Journal of Political Economy*, 127, 1437–1472.