Bias in Local Projections

Edward P. Herbst and Benjamin K. Johannsen*

November 20, 2022

Abstract

Local projections (LPs) are a popular tool in macroeconomic research. We show that LPs are often used with very small samples in the time dimension and, consequently, that LP point estimates can be severely biased. Under regularity conditions, we derive simple expressions to approximate this bias and propose a way to bias correct LPs. Using a medium-scale macroeconomic time-series model, we demonstrate that the bias in point estimates can be economically meaningful. We also show that the same small-sample bias issue can also lead some autocorrelation-robust standard errors to understate sampling uncertainty.

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

Starting with Jordà (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 Jordà (2005). In particular, we find that sample sizes in the time dimension are typically much smaller than the sample sizes studied in Jordà (2005) and that LPs have become increasingly prevalent when researchers also have a cross section of data (i.e., panel

*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 Jordà, Lutz Kilian, Mikkel Plagborg-Møller, 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.

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.

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), Nagar (1959), Rilstone et al. (1996), Anatolyev (2005), and Bao and Ullah (2007). We show that the approximate bias of the LP estimator at horizon h is a function—specifically, a weighted sum—of the (population) impulse response function at horizons up to h. As a result, if the data are positively autocorrelated and 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.

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. We conduct our Monte Carlo simulations using simple, linear data generating processes and an estimated medium-scale macroeconomic VAR. 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 these simple scenarios.

The expression for the approximate bias that we derive can be used to bias correct LP estimators. In Monte Carlo simulations, on average our bias-corrected estimators are markedly closer to the true values of the impulse responses.[2] We discuss the tradeoff that researchers face between reducing bias and potentially increasing the mean squared error (MSE) of the estimator. We note that the bias correction does not uniformly increase or decrease MSE in our Monte Carlo simulations.

We extend our analysis to settings using panel data and show that—when using fixed effects—the bias we document persists. We derive formulas to approximate the bias, which could be used to bias-correct LP estimators. We show that increasing the number of entities in the panel will not eliminate the bias.

We also study the downward bias in the standard errors of LP estimators. Our analysis of standard errors is related to recent work by Montiel Olea and Plagborg-Møller (2021), who suggest

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

[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)).

that researchers using LPs should use heteroskedasticity-consistent, but not autocorrelation-robust, standard errors. Our focus on finite sample issues leads us to similar conclusions, but for different reasons. We show that small-sample bias can be an important consideration when using Newey and West (1987) (NW) standard errors. These standard errors rely on estimators of the autocorrelation of the regression score. For similar reasons to the LP point estimators themselves, in finite samples, the estimators of these autocorrelations will be biased. In important, empirically-realistic settings the bias will be downward, yielding smaller estimates of standard errors. This result suggests researchers may prefer standard errors that are heteroskedasticity-consistent, but not autocorrelation-robust, such as those studied in Huber (1967) and White (1980) (HW). Alternatively, researchers may want to consider alternative HAR standard erros, like those discussed in Müller (2014), Sun (2014), and Lazarus et al. (2018).

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 least-squares estimate. By contrast, our analysis relies on higher-order expansions of the least-squares estimator, which illustrates the reasons that the least-squares estimator is biased and provides a natural bias correction without bootstrapping.[3] 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 VARs (such as in Nicholls and Pope (1988) and Pope (1990)), 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 (2005), and Bao and Ullah (2007)). We apply this work in 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 Jordà (2005) on Google Scholar.[4] Google Scholar's relevance ranking weighs the text of the document, the authors, the source of the publication, and the number of citations. Of these 100

[3] The bias correction procedure detailed here cannot be used within a block bootstrap procedure because the all value of θ_j are needed for $0 \leq j < h$ and the blocking effectively destroys the dependent structure of the data that would make such calculations possible.

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

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 we show in Section 4, 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 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.

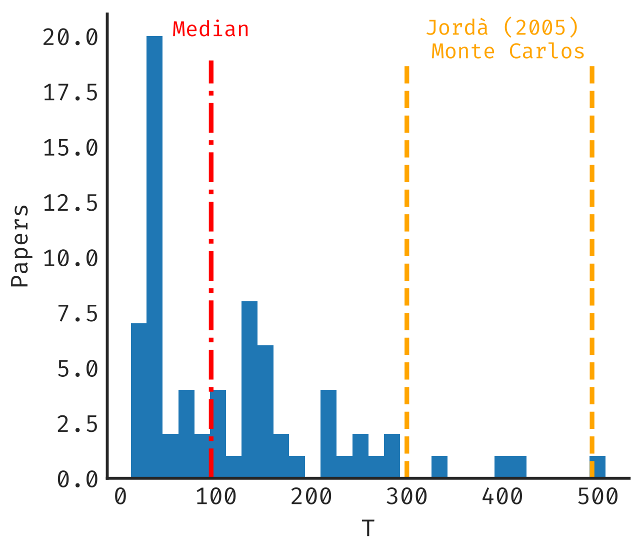


Figure 1: T is small in the literature using LPs.

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 significantly smaller than those typically used in empirical macroeconometrics papers, as most of the papers surveyed here use the increasingly popular strategy of using observed shocks, such as the monetary policy

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

shocks of Romer and Romer (2004), rather than identified shocks from a VAR, as in Jordà (2005). Constructing these observed shocks is often difficult and costly, 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 studies in Jordà (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. While many studies in our survey use annual or quarterly data, Jordà (2005) used monthly data. In general, however, increasing T by using monthly data rather than quarterly or annual data will not eliminate 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 analyze the bias in LPs using a Nagar (1959) expansion. We focus on LPs that estimate the impulse response of a macroeconomic variable, y_t, to a structural shock, ε_t. As mentioned in the introduction, the structural shock is observed and inference is conducted using linear, least-squares regression. The LP model is the set of regression models indexed by the impulse response horizon, h,

$$y_{t+h} = \alpha_h + \beta_h' x_t + u_{h,t}, \quad h = 0, \ldots, H. \tag{1}$$

where $x_t = [\varepsilon_t, c_{t-1}']'$ contains the structural shock and (time $t-1$) controls. For ease of exposition, we assume the control vector, c_{t-1}, is not empty—in practice, researchers use c_{t-1} to condition inference on information available at time $t-1$. 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 ordinary 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]

[6]It is also common in the LP literature to use $y_{t+h} - y_{t-1}$ as the left hand side variable. If $y_t - y_{t-1}$ can be included in w_t as defined in Assumption 1, and so long as our assumptions still hold, then Analytic Result 1 holds. Analogous results to other results in the paper can be derived using straightforward modifications to our analysis.

3.1 The approximate bias of the least-squares estimator

To ease notation and without loss of generality, throughout the paper we assume that the data we consider have mean zero. To facilitate the derivation of the bias we make the following assumption about the time series properites of y_t, ε_t and c_{t-1}.

Assumption 1. *The series $w_t = [y_t, \varepsilon_t, c_t']'$ is strictly stationary and ergodic. The series has a purely nondeterministic Wold representation with innovations $\omega_t = [\varepsilon_t, \nu_t']'$,*

- *ε_t is independent of ε_{t+j} for all $j \neq 0$,*
- *ε_t is independent of ν_{t+j} for all j,*
- *$\mathbb{E}[x_t x_t']$ is invertible and $\mathbb{E}[||x_t x_t'||]$ is finite, where $||\cdot||$ denotes the 2-norm.*

Our assumption about the properties of ε_t are meant to formalize what is meant by a structural shock. The linearity of w_t in all shocks is a stark assumption that facilitates the derivation of closed form expressions for the bias in LPs. Linearity in the shock of interest, ε_t, also represents an idealized case where "direct causal inference"—as in Nakamura and Steinsson (2018)—is possible. Notably, in the case where w_t is gaussian and ε_t is *iid*, the conditions of Assumption 1 are satisfied. We emphasize that violations of Assumption 1 do not imply that LPs are not biased in small samples. Rather, violations of Assumption 1 make it difficult to characterize the small sample bias. In addition, we make the following assumption about the regression errors from the LP.

Assumption 2. *The regression error, $u_{h,t}$, is independent of ε_{t-j} and ν_{t-j-1} for all $j \geq 0$.*

This assumption formalizes what we mean by conditioning inference on information available at time $t-1$ by including c_{t-1} in the LP. This guarantees that the regression error will be at most an $MA(h)$ in ν_t and an $MA(h-1)$ in ε_t, representing an idealized case where the controls perfectly control for time $t-1$ information. This effectively truncates the terms in the approximate bias, simplifying the subsequent calculations considerably. As with Assumption 1, a violation of Assumption 2 does imply a lack of bias; we discuss implications of departures from this assumption later in the paper. We can now state our main analytical result regarding the approximate bias in LPs.

Analytical Result 1 (Expression for the bias in LP)**.** *Under Assumptions 1-2, the approximate bias for the LP in (1) is given by*

$$\mathbb{E}\left[\widehat{\theta}_{h,LS}\right] - \theta_h = B_{h,LP} + O\left(T^{-3/2}\right), \tag{2}$$

where

$$B_{h,LP} = -\frac{1}{T-h}\sum_{j=1}^{h}\left(1 + tr\left\{\Sigma_{c,0}^{-1}\Sigma_{c,j}\right\}\right)\theta_{h-j} \text{ for } h > 0, \tag{3}$$

$\Sigma_{c,j} = E\left[c_{t-j}c_t'\right]$, *and* $tr\{\cdot\}$ *is the trace operator. Additionally,* $B_{0,LP} = 0$.

The derivation of equations (2) and (3) relies on the methodology of Bao and Ullah (2007).[7] A detailed derivation is in Appendix A. We refer to $B_{h,LP}$ as the approximate bias. Several remarks regarding Analytic Result 1 are in order. First, $|B_{h,LP}|$ is a decreasing function of T; for fixed h, the least-squares estimator is consistent. Second, $B_{h,LP}$ is a function of the impulse response coefficients at *all horizons up to h*. Intuitively, the data generating process affects the bias at similar horizons in similar ways. Conditioning inference on information available at time $t-1$, however, truncates the terms that contribute to the bias by eliminating the autocorrelation in the regression errors at sufficiently large lags.[8] Third, the contribution of θ_{h-j} to $B_{h,LP}$ is scaled by $1 + \text{tr}\left\{\Sigma_{c,0}^{-1}\Sigma_{c,j}\right\}$. When controls are relatively persistent this scaling factor is relatively large. Fourth, intuitively $B_{0,LP} = 0$ because the structural shock, ε_t, is *iid* and $u_{0,t}$ is independent over time.

It is useful to parse the contribution to $B_{h,LS}$ that is due to estimating α_h. To this end, we derive the approximate bias when α_h is known.

Analytical Result 2 (Expression for the bias in LP when α_h is known)**.** *Under Assumptions 1-2, and under the additional assumption that* α_h *is known, the expression for the bias in (1) is given*

[7] Some additional technical remarks regarding our analytical framework are in order. The least-squares estimator can be cast as a k-class estimator—see Theil (1961). The characteristics of k-class estimators are important for the study of parameter bias (both finite sample and asymptotic) in simultaneous equations models, in particular for the study of (weak) instrumental variables. Sawa (1972) shows that for many k-class estimators the first moment may not even exist, rendering the approximations in this paper inaccurate—see also Srinivasan (1970). We acknowledge this limitation, but note that with the additional assumption of normality—following Sawa (1972)—one can guarantee the existence of the first moment of $\widehat{\theta}_{LS}$. Additionally, one can think of the stochastic terms used in the procedure of Bao and Ullah (2007) as offering an approximation to the true finite sample distribution of $\widehat{\theta}_{LS}$. This approximation may have finite moments even when the exact distribution of $\widehat{\theta}_{LS}$ does not.

[8] An earlier version of the paper examined the bias in the alternative polar case: LPs without controls. This LP violates assumption 2, and requires a slightly different derivation. In the AR(1) model here, the bias is given by: $(1-\rho^{-(T-h)})(\rho^T - \rho^{h+1})/(1-\rho)$. This can be smaller or larger than the approximate bias in Equation 6 depending on ρ, h, and T. Herbst and Johannsen (2021) contains more details on the bias in LPs without controls.

by

$$\mathbb{E}\left[\widehat{\theta}_{h,LS}\right] - \theta_h = -\frac{1}{T-h}\sum_{j=1}^{h} tr\left\{\Sigma_{c,0}^{-1}\Sigma_{c,j}\right\}\theta_{h-j} + O\left(T^{-3/2}\right), \tag{4}$$

where $\Sigma_{c,j}$ *and* $tr\{\cdot\}$ *are defined as in Analytic Result 1. Additionally,* $\mathbb{E}\left[\widehat{\theta}_{h,LS}\right] - \theta_h = O(T^{-3/2})$.

From equation (4), it is immediately apparent that estimating α_h adds the terms $-(T-h)^{-1}\sum_{j=1}^{h}\theta_h$ to the approximate bias. When $\theta_h > 0$ and $\mathrm{tr}\left\{\Sigma_{c,0}^{-1}\Sigma_{c,j}\right\} > 0$, as is the case in many macroeconomic settings, estimating α_h increases the magnitude of the (negative) approximate bias.

3.2 An AR(1) example

Here we examine the approximate bias in LPs using a canonical AR(1) model for y_t. In particular, we assume that

$$y_t = \alpha + \theta_0\varepsilon_t + \rho y_{t-1} + \nu_t \quad \text{with} \quad \varepsilon_t \overset{iid}{\sim} N(0,\sigma_\varepsilon^2) \text{ and } \nu_t \overset{iid}{\sim} N(0,\sigma_\nu^2). \tag{5}$$

Further, we assume that $\rho \in (0,1)$. The researcher observes $\{(y_t,\varepsilon_t)\}_{t=1}^{T}$ and estimates the LP defined in equation (1) with $c_{t-1} = y_{t-1}$ using ordinary least squares. This example satisfies Assumptions 1 and 2.

3.2.1 Explicit approximate bias

Noting that $\theta_h = \theta_0\rho^h$, it is easy to see from Analytic Result 1 that the approximate bias of the LP is given by

$$B_{h,LP} = -\frac{1}{T-h}\sum_{j=1}^{h}\left(1+\rho^j\right)\theta_0\rho^{h-j} = -\frac{\theta_0}{T-h}\left(\frac{1-\rho^h}{1-\rho} + h\rho^h\right). \tag{6}$$

The term $(1-\rho^h)/(1-\rho)$ accounts for the bias arising from estimating the constant, α_h, and the term $h\rho^h$ accounts for the bias arising from the persistence of the control, ρ. If $\theta_0 > 0$ (a normalization), $B_{h,LP} \le 0$ for all T and $0 < h < T$. That is, the bias is always downward. It is not straightforward to say at which horizon the bias is largest: $\mathrm{argmax}_{0<h<T}|B_{h,LP}|$ is a complicated function of ρ and T because the component of the bias associated with estimating the constant is generally decreasing in magnitude, whereas the component associated with persistent regressors in increasing in magnitude.

3.2.2 Quality of the approximation

Here we analyze how well $B_{h,LP}$ approximates the small-sample bias of $\widehat{\theta}_{h,LS}$ by comparing the approximate bias to the exact bias calculated using Monte Carlo simulations—using equation 5—for various ρ and T. For this this exercise, we set $\sigma_\varepsilon = \sigma_\nu = 1$ and $\theta_0 = 1$. The variance parameters do not appear in $B_{h,LP}$.[9] Figure 2 shows $B_{h,LP}$ along with the Monte Carlo estimate. The figure shows results for $\rho \in \{0.90, 0.95, 0.99\}$ and $T \in \{50, 100, 200\}$.

For $\rho = 0.9$ and $\rho = 0.95$, $B_{h,LP}$ is a good approximation to the exact small-sample bias in $\widehat{\theta}_{h,LS}$ for all h shown. The quality of the approximation improves somewhat as T increases. Clearly, when $\rho = 0.99$, $B_{h,LP}$ is not as good of an approximation as it is for smaller values of ρ. Nevertheless, even with $\rho = 0.99$ and $T = 50$ $B_{h,LP}$ captures salient features of the small-sample bias, including that it is growing in magnitude in h over the values of h shown. We conclude that for empirically relevant sample sizes, $B_{h,LP}$ offers a reasonable approximation to the bias in LPs, though the quality of the approximation is somewhat worse for smaller values of T and larger values of ρ. In Appendix C, we show that similar results hold for an AR(2) model with hump-shaped impulse responses.

3.2.3 Comparison to parametric approach

LP estimators are often compared to estimates of impulse responses from VARs. These discussions are often centered around the different asymptotic bias and variance trade offs associated with the (relatively flexible) LP and (tightly constrained) VAR estimators. A natural question is how the small-sample bias of LP estimators compares to the small-sample bias of VAR estimators.[10] To address this, we compare $B_{h,LP}$ to the small-sample bias arrising from estimating the coefficients α, θ_0, and ρ in equation (5) using ordinary least squares and computing the h-period impulse response as $\widehat{\theta}_{0,LS} \left(\widehat{\rho}_{LS}\right)^h$. This approach, which we call the "AR approach," is analogous to the estimation of impulse responses using VARs when the researcher has the time series data for the structural shock of interest.

To compute the approximate bias of the AR approach, we again use the methodology of Bao

[9]In Appendix C, we show that setting $\sigma_\varepsilon = 10\sigma_\nu = 1$ or $10\sigma_\varepsilon = \sigma_\nu = 1$ has little effect on the quality of the approximation offered by $B_{h,LP}$. In Appendix C, we also analyze $B_{h,LP}$ in the context results from a Monte Carlo exercise under the assumption that the means of the data are known.

[10]The small-sample bias of VAR estimators has been explored by Nicholls and Pope (1988), Pope (1990), and others.

and Ullah (2007). The resulting approximation is

$$\mathbb{E}\left[\widehat{\theta}_{0,LS}\widehat{\rho}_{LS}^{h}-\theta_0\rho^h\right]=B_{h,AR}+O(T^{-3/2}) \tag{7}$$

where

$$B_{h,AR}=-\frac{\theta_0\gamma}{T-1}\left[h\rho^{h-1}\left(1+3\rho\right)-\frac{1}{2}h\left(h-1\right)\rho^{h-2}\left(1-\rho^2\right)\right] \text{ and } \gamma=\frac{\sigma_\nu^2}{\sigma_\nu^2+\theta_0^2\sigma_\varepsilon^2}. \tag{8}$$

The details of the derivation are in Appendix B. To understand the expression for $B_{h,AR}$, it is useful to first analyze γ. Note that $\gamma \in [0,1]$ is the ratio of the variance of $u_{0,t}$ to the one-step-ahead variance of y_t. Larger values of γ increase the magnitude of $|B_{h,AR}|$. That is, when ε_t explains relatively little of the variation in y_t, γ and thus $|B_{h,AR}|$ are larger. Conversley, when instead ν_t explains relatively little of the variation in y_t, γ and thus $|B_{h,AR}|$ are smaller. As the variance of the regression error collapses, the AR approach is less biased.

Several remarks about the relationship between $B_{h,AR}$ and $B_{h,LP}$ are in order. First, $B_{0,AR}=B_{0,LP}=0$. Second, in general, the values of $B_{h,AR}$ and $B_{h,LP}$ will not coincide for $h>0$. Third, for $\rho\in(0,1)$, $B_{h,AR}$ cannot be signed for $h>0$, unlike $B_{h,LP}$ which is never positive. Fourth, γ affects $B_{h,AR}$, but does not enter $B_{h,LP}$. That is, the approximate bias in the LP is not effected by the amount of variation in y_t that is explained by ε_t. While the magnitude and sign of $B_{h,AR}-B_{h,LP}$ depend on parameters of the AR(1) process and h, it is useful to consider a particular value of $\gamma=\frac{1+\rho}{1+3\rho}$, which sets $B_{1,AR}=B_{1,LP}$. Then, for $\rho\in(0,1)$ and $h>1$,

$$B_{h,AR}=-\frac{1}{T-1}\left(h\rho^h+h\rho^{h-1}\right)+\frac{1}{T-1}\frac{1}{2}h\left(h-1\right)\rho^{h-2}\frac{1+\rho}{1+3\rho}\left(1-\rho^2\right)>B_{h,LP}. \tag{9}$$

So, for any $\gamma<\frac{1+\rho}{1+3\rho}$, $B_{h,AR}-B_{h,LP}<0$ for all $h>1$. This result illustrates that there are parameter values for which the AR approach can be uniformly less negative than the LP approach. However, it is difficult to know whether these conditions hold without knowledge of the data generating process. Above all, the AR(1) example illustrates that the the small-sample bias in LPs is different than the small-sample bias in VARs.[11]

3.3 A Bias Corrected Estimator

Analytic Result 1 suggests a bias-corrected estimator for θ_h using plug-in estimators for θ_j and $\Sigma_{c,j}$. With enough data to calculate $\widehat{\theta}_{h,LS}$, all of the needed estimates of θ_j and $\Sigma_{c,j}$ are easily computed.

[11]In VARs, researchers commonly employ the parametric bootstrap to correct for bias (see, for example, Kilian (1998)). However, in an LP, researchers would need to use, for example, a block bootstrap. With sample sizes typically seen the literature that uses LPs, these bootstrapping methods are likely to perform poorly.

Notably, the researcher could bias correct the coefficients using the least-squares estimates of θ_j. We denote the estimator of θ_h constructed in this way as $\widehat{\theta}_{h,BC}$. Alternatively, the researcher could iterate on this bias correction, effectively jointly correcting the impulse response estimate. We denote this estimator as $\widehat{\theta}_{h,BCC}$. If θ_j all have the same sign, then the bias correction used to construct $\widehat{\theta}_{h,BCC}$ will be larger than the bias correction used to construct $\widehat{\theta}_{h,BC}$.

Here, we analyze how well $\widehat{\theta}_{h,BCC}$ performs in the context of our AR(1) example data generating process. The left column of Figure 3 shows the Monte Carlo estimate $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$, for different values of ρ and T, as a function of h. The bias corrected estimator does eliminate the bias entirely, particularly for low T and large ρ. That said, it reduces the bias relative to the least-squares estiamtor. For example, when $\rho = 0.99$ and $T = 50$, $\widehat{\theta}_{h,BCC}$ still exhibits a bias of around -0.3 for moderate h, the corresponding bias for $\widehat{\theta}_{h,LS}$ is around -0.6 (see Figure 2). So $\widehat{\theta}_{h,BCC}$ represents a substantial improvement.

The right column of Figure 3 shows the MSE ratio of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$. A number larger than unity indicates that the MSE of $\widehat{\theta}_{h,BCC}$ is larger than the MSE of $\widehat{\theta}_{h,LS}$. While $\widehat{\theta}_{h,BCC}$ does a better job, on average, of correcting for the bias when ρ is smaller, the relative MSE is smaller for larger ρ. Intuitively, when ρ is large, the bias of the LP estimator is also large, and the mean bias correction is enough to reduce the MSE, even though the bias correction procedure introduces volatility into the estimator. Overall, the MSE metric does not uniformly favor either $\widehat{\theta}_{h,LS}$ or $\widehat{\theta}_{h,BCC}$.

It ought to be mentioned that the potential for asymptotic bias from parametric models is a key reason that many investigators turn to LPs in the first place. It stands to reason that these researchers exhibit preferences that give more weight to the bias reduction than lower variance estimators. Thus, $\widehat{\theta}_{BCC}$, which reduces the finite sample component of this bias, may be attractive even if it exhibits a higher variance (and higher MSE) than its least-squares counterpart.

4 Extension to panel data

Here, we extend our analysis to the setting of panel data with a fixed number of entities, I. We let the subscript i indicate data for a particular panelist (for example, $y_{i,t}$). We consider the setting where the entities have common slope coefficients, but entity-specific intercepts, which is a common setup in the LP literature using panel data. That is,

$$y_{i,t+h} = \alpha_{i,h} + [\varepsilon_{i,t}, c'_{i,t-1}, c'_{t-1}]\beta_h + u_{i,h,t}. \tag{10}$$

The vector c_{t-1} is included to accommodate common regressors. We set up the estimation problem as stacked least squares, which is overwhelmingly the most common approach in the related LP literature that uses panel data. We refer to the estimator of the impulse response function at horizon h as $\widehat{\theta}_{h,LS,I}$.

4.1 The approximate bias of the least-squares estimator with panel data

To fix notation, let

$$x_{i,t} = [1(i=1), 1(i=2), \ldots, 1(i=I), \varepsilon_{i,t}, c_{i,t-1}, c_{t-1}]'. \tag{11}$$

$$w_t = [\varepsilon'_{1,t}, \varepsilon'_{2,t}, \ldots, \varepsilon'_{I,t}, c'_{1,t}, c'_{2,t}, \ldots, c'_{I,t}, c'_t]'. \tag{12}$$

We make the following assumption, which is analogous to Assumption 1.

Assumption 3. *The series w_t is strictly stationary and ergodic. The series has a purely nondeterministic Wold representation with innovations $\omega_t = [\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots, \varepsilon_{I,t}, \nu'_t]'$. Additionally, for all i,*

- *$\varepsilon_{i,t}$ is independent of $\varepsilon_{k,t+j}$ for all $j \neq 0$ and all k.*
- *$\varepsilon_{i,t}$ is independent of ν_{t+j} for all j.*
- *$\mathbb{E}[\sum_{i=1}^{I} x_{i,t}x'_{i,t}]$ is invertible, and $\mathbb{E}[||\sum_{i=1}^{I} x_{i,t}x'_{i,t}||^2]$ is finite.*

Our assumptions about $\varepsilon_{i,t}$ are meant to formalize what we mean by a structural shock.[12] Additionally, as was the case without panel data, the linearity of w_t facilitates the derivation of analytic expressions for the approximate bias. One consideration that is specific to panel data is the way that $\varepsilon_{j,t}$ affects $y_{i,t+h}$.

In addition, we make the following assumption, which is analogous to Assumption 2.

Assumption 4. *For every i, the regression error $u_{i,h,t}$ is independent of $\varepsilon_{k,t-j}$ and ν_{t-j-1} for all k and $j \geq 0$.*

This assumption formalizes what we mean by conditioning inference on information available at time $t-1$ by including $c_{i,t-1}$ and c_{t-1} in the LP. Additionally, Assumption 4 requires that if $\varepsilon_{j,t}$ affectes $y_{i,t+h}$, it only does so through correlation with $\varepsilon_{i,t}$. In Appendix A.3, we derive the following analytic result.

[12] In the event that a researcher wants to investigate the response of $y_{i,t+h}$ to a structural shock that is common to all panelists, ε_t, straightforward modification of our setup so that $\varepsilon_{i,t} = \varepsilon_t$ will accommodate.

Analytical Result 3 (Bias in panel LPs with controls.)**.** *Under Assumptions 3-4, the approximate bias of* $\widehat{\theta}_{h,LS,I}$ *is given by*

$$\mathbb{E}\left[\widehat{\theta}_{h,LS,I} - \theta_h\right] = B_{h,LP,I} + O\left(T^{-3/2}\right), \tag{13}$$

where

$$B_{h,LP,I} = -\frac{1}{T-h}\sum_{j=1}^{h}\left[1+\vartheta_j\right]\theta_{h-j} \text{ and } \vartheta_j = \frac{1}{I^2}\sum_{i=1}^{I}\sum_{k=1}^{I}\frac{tr\left\{\Sigma_{c,0,I}^{-1}\Sigma_{c,j,k,i}\right\}\sigma_{\varepsilon,i,k}}{\sigma_{\varepsilon,I}^2}. \tag{14}$$

Here, $\sigma_{\varepsilon,I}^2 = \frac{1}{I}\sum_{i=1}^{I}\sigma_{\varepsilon,i,i}$, $\sigma_{\varepsilon,i,k} = \mathbb{E}[\varepsilon_{i,t}\varepsilon_{k,t}]$, $\Sigma_{c,0,I} = \frac{1}{I}\sum_{i=1}^{I}\Sigma_{c,0,i,i}$, $\Sigma_{c,j,i,k} = \mathbb{E}[\tilde{c}_{i,t-j}\tilde{c}_{k,t}']$, *and* $\tilde{c}_{i,t} = [c_{i,t}', c_t']'$*. Additionally,* $B_{0,LP,I} = 0$*.*

A few comments are in order regarding equation (14). First, as was the case without panel data and because the least-squares estimator is consistent, the approximate bias goes to zero as the sample size goes to infinity. Second, the cross-autocovariance of the control variables plays a role in the approximate bias. Third, as was the case without panel data, if the IRF of interest is a persistent IRF and the controls are positively autocorrelated, 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, holding T fixed.

It is instructive to consider the implications of estimating the $\alpha_{i,h}$'s. In Appendix A.3, we show that the contribution to $B_{h,LP,I}$ from estimating the $\alpha_{i,h}$'s is given by

$$-\frac{1}{T-h}\sum_{j=1}^{T-h}\theta_{h-j}. \tag{15}$$

As a result, if the $\alpha_{i,h}$'s (means of the data) are known, increasing the number of panelists, I, can make $B_{h,LP,I}$ approach zero.

4.2 An AR(1) example with panel data

Here, we examine the approximate bias in panel LPs using independent AR(1) data generating processes for $y_{i,t}$. In particular, we assume that

$$y_{i,t} = \alpha_i + \theta_0\varepsilon_{i,t} + \rho y_{i,t-1} + \nu_{i,t} \quad \text{with} \quad \varepsilon_{i,t} \overset{iid}{\sim} N(0,\sigma_\varepsilon^2) \text{ and } \nu_{i,t} \overset{iid}{\sim} N(0,\sigma_\nu^2). \tag{16}$$

Further, we assume that $\rho \in (0,1)$. The researcher observes $\{(y_{i,t}, \varepsilon_{i,t})\}_{t=1}^{T}$ for $i = 1, 2, \ldots, I$, and estimates the impulse response of $y_{i,t+h}$ to $\varepsilon_{i,t}$ for $h = 0, \ldots, H$ using the LP defined in equation 10 with $c_{i,t-1} = y_{i,t-1}$ and $c_{t-1} = \emptyset$. This example satisfies Assumptions 3 and 4.

4.2.1 Explicit approximate bias with panel data

Noting that $\theta_h = \theta_0\rho^h$, it is easy to see from Analytic Result 3 that the approximate bias of the panel LP is given by

$$B_{h,LP,I} = -\frac{1}{T-h}\sum_{j=1}^{h}\left(1+\frac{1}{I}\rho^j\right)\theta_0\rho^{h-j} = -\frac{\theta_0}{T-h}\left(\frac{1-\rho^h}{1-\rho}+\frac{1}{I}h\rho^h\right). \tag{17}$$

Notice that the expression for $B_{h,LP,I}$ in equation (17) is similar to the expression for $B_{h,LP}$ in equation (6), except that the second term in parentheses is divided by I. This result reflects the common AR coefficient across $y_{i,t}$ and our assumption that $y_{i,t}$ is independent of $y_{j,t}$ for $i \neq j$. Equation (17) also illustrates that adding panelists with uncorrelated data can reduce the approximate bias. At the same time, $B_{h,LP,I}$ clearly does not approach zero as I increases. It is worth noting that when ρ is near unity, the first term in parentheses in equation (17) will be relatively large, and increasing I will have little effect on the approximate bias.

4.2.2 Quality of the approximation with panel data

Here we analyze how well $B_{h,LP,I}$ approximates the small-sample bias of $\widehat{\theta}_{h,LS,I}$. Figure 4 shows $B_{h,LP,I}$ along with the sample mean of a Monte Carlo exercise conducted using the data generating process in equation (16).[13] The figure shows results for different values of ρ and T and is constructed under the assumption that $I = 50$. This choice of I is reasonably representative in the context of the LP literature that uses panel data because many studies consider either state-level data or country-level data for advanced economies. Because all of the values of ρ shown imply that the data are persistent, the figure is very similar for smaller values of I.

Notably, Figure 4 is similar to Figure 2. As a result, we conclude that for empirically relevant sample sizes, $B_{h,LP,I}$ offers a reasonable approximation to the bias in LPs, though the quality of the approximation is somewhat worse for smaller values of T and larger values of ρ. Additionally, we can conclude that for ρ near unity, adding a panel dimension to the LP does not materially reduce small-sample bias.

4.3 A bias-corrected estimator with panel data

Analytic Result 3 lends itself to constructing bias-corrected estimators for θ_h using plug-in estimators for θ_j and $\Sigma_{c,i,j}$. However, the number of cross-autocovariances that are required to construct

[13]For the Monte Carlo exercises, we set $\sigma_\varepsilon = \sigma_\nu = 1$ and $\theta_0 = 1$.

the values of $\Sigma_{c,i,j}$ is large. Because the term of the approximate bias involving $\Sigma_{c,i,j}$ is multiplied by I^{-1}, in applications with reasonably large I, researchers could instead use

$$B_{h,LP,I} \approx -\frac{1}{T-h} \sum_{j=1}^{h} \theta_{h-j}. \tag{18}$$

While this expression is not an exact expression for the approximate bias to $O(T^{-1})$, it could be used to improve upon the least-squares estimator by picking up some (or potentially most) of the approximate bias. We denote the estimator of θ_h constructed in this way as $\widehat{\theta}_{h,BC,I}$. Alternatively, the researcher could iterate the bias correction on all values of θ_h. We denote this estimator as $\widehat{\theta}_{h,BCC,I}$.

Here, we analyze how well performs $\widehat{\theta}_{h,BCC,I}$ performs in the context of our AR(1) example data generating process from equation (16). The left column of Figure 5 shows the Monte Carlo average of $\widehat{\theta}_{h,BCC,I} - \theta_h$, for different values of ρ and T, as a function of h. Clearly, the bias-corrected estimator performs better for smaller values of ρ and larger values of T.

The right column of Figure 5 shows the MSEs of $\widehat{\theta}_{h,BCC,I}$ relative to the MSEs associated with $\widehat{\theta}_{h,LS,I}$. Note that a number larger than unity indicates that the MSE of $\widehat{\theta}_{h,BCC,I}$ is larger than the MSE of $\widehat{\theta}_{h,LS,I}$. Notably, $\widehat{\theta}_{h,BCC,I}$ leads to a smaller MSE than $\widehat{\theta}_{h,LS,I}$. Intuitively, the reason is that the values $\widehat{\theta}_{h,LS,I}$ has an normal asymptotic limiting distribution when multiplied by $\sqrt{TI}$, so in small samples it has relatively little variance when used as a plug in estimator for the approximate bias, which converges at rate $\sqrt{T}$. Of course, estimating the values of $\Sigma_{c,i,j}$ to capture the entire approximate bias would add additional variance.

5 Bias in the Context of a Medium-Scale Monetary Time Series

In this section, we conduct a Monte Carlo study using a larger time series model. To construct an empirically realistic data generating process, we follow Christiano et al. (2005)—subsequently referred to as CEE—and estimate a nine variable VAR(4). We focus on the dynamic effects of a monetary policy shock, identified recursively as in CEE. Figure 6 shows the impulse responses of real GDP, price level, and the federal funds rate to a monetary policy shock. The estimated VAR(4) serves as the data generating process for our Monte Carlo exercise.[14]

[14]The inflation responses are computed *ex post* as the (annualized) percent change in the price level. Note also that the VAR, at the MLE, is stationary. The magnitudes of the five largest eigenvalues are $[0.99, 0.97, 0.97, 0.95, 0.95]$. More details about the VAR are available in the Appendix.

Our Monte Carlo study examines the properties of both the standard least-squares estimator and our bias-corrected LP for estimating these three impulse responses. We assume that the econometrician observes sets of time series of length T of the dependent variable y_t—either the log level of real GDP, the log level of prices, or the federal funds rate—sets of controls c_t, and the monetary policy shock ε_t. In this section, we assume that the vector of controls is "full" in the sense of containing the four lags of *all* variables as in the VAR. Results under different assumptions about the conditioning set are broadly similar and are available in Appendix E. The sample sizes vary, as in the earlier simulations, with $T \in \{100, 200\}$. We simulate 2000 trajectories from the VAR for each sample size T. Each simulation is initialized from a random point in the stationary distribution.

Figure 7 displays Monte Carlo estimates of the bias of $\widehat{\theta}_{h,LS}$ and $\widehat{\theta}_{h,BCC}$ (the dashed and solid lines in left column, respectively) and the MSE ratio (right column) for output, the price level, and the federal funds (rows). Focusing first on the bias, we see that all of the estimated impulse responses exhibit bias, with largest bias associated with the impulse response of of real GDP. As expected the bias is decreasing in the sample size T with the red lines ($T = 200$) generally closer to zero than the green lines ($T = 100$). Interestingly, unlike for the univariate AR(1) example, this bias is not always—or even typically—downwards. In nearly all cases the bias of $\widehat{\theta}_{h,BCC}$ is smaller in magnitude than the bias of $\widehat{\theta}_{h,LS}$, indicating that the bias correction works well. But as with the univariate AR(1) example, the bias corrected estimator does not completely eliminate the bias.

Turning to the relative MSE, displayed the right column of Figure 7, we see that neither $\widehat{\theta}_{h,LS}$ nor $\widehat{\theta}_{h,BCC}$ is uniformly best. For example, for real GDP the MSE of $\widehat{\theta}_{h,BCC}$ is less than that of $\widehat{\theta}_{h,LS}$ for $h \in [4, 12]$ as the bias term dominates the MSE calculation at those horizons. When the bias associated with $\widehat{\theta}_{h,LS}$ is relatively small (here at small or large h), this estimator is preferred in an MSE sense. For the price level the bias switches signs around $h \approx 14$, so for these horizons $\widehat{\theta}_{h,BCC}$ has a relatively high MSE. Finally, for the federal funds rate, again neither estimator dominates across all h. Overall, the MSE criterion does not definitively select either $\widehat{\theta}_{h,LS}$ or $\widehat{\theta}_{h,BCC}$ as best.[15]

[15] In the Appendix, we compare the LP estimators to those of VAR(1) and VAR(4) models.

6 Some considerations related to standard errors

The asymptotic covariance matrix of the least-squares estimator of the parameters in equation (1) is given by

$$V_{T-h} = \left(\frac{1}{T-h} \sum_{t=1}^{T-h} \mathbb{E}\left[\tilde{x}_t \tilde{x}_t \right]' \right)^{-1} S_{T-h} \left(\frac{1}{T-h} \sum_{t=1}^{T-h} \mathbb{E}\left[\tilde{x}_t \tilde{x}_t \right]' \right)^{-1} \tag{19}$$

where $\tilde{x} = [1, x_t']'$ and

$$S_{T-h} = \frac{1}{T-h} \sum_{t=1}^{T-h} \sum_{s=1}^{T-h} \mathbb{E}\left[\tilde{x}_t \tilde{x}_s' u_{h,t} u_{h,s} \right]. \tag{20}$$

The challenging piece of the standard-error calculation is estimating S_{T-h}. Since Jordà (2005), the conventional wisdom has been that heteroskedasticity and autocorrelation robust (HAR) standard errors are necessary because the regression residuals of LPs are autocorrelated. A popular choise of estimator for S_{T-h} in the LP listerature is the estimator of Newey and West (1987) (NW). Because the NW estimator tends to understate sampling uncertainty, work by Müller (2014), Sun (2014), and Lazarus et al. (2018) have suggested alternatives. However, under Assumptions 1 and 2, the *LP regression score*—the product of the ε_t and the LP regression residuals—is serially uncorrelated.[16] Thus, in large samples the (heteroskedasticity-robust) estimator of Huber (1967) and White (1980) (HW) is valid. Here, we consider the small sample implications of using the NW—i.e., setting $m = 0$—estimator when HW will do.

Because we assume that ε_t is independent of c_{t-1}, to compute the standard error for $\widehat{\theta}_{h,LS}$ the only relevant element of S_{T-h} is the diagonal element in the same position as θ_h, which is given by

$$\frac{1}{T-h} \sum_{t=1}^{T-h} \sum_{s=1}^{T-h} \mathbb{E}\left[\varepsilon_t (y_{t+h} - \alpha_h - x_t'\beta) \varepsilon_s (y_{s+h} - \alpha_h - x_s'\beta) \right] = \frac{1}{T-h} \sum_{t=1}^{T-h} \sum_{s=1}^{T-h} \gamma_{h,|t-s|}. \tag{21}$$

The NW estimatorof this value with bandwidth m is given by

$$\widehat{\gamma}_{h,0} + \sum_{\ell=1}^{m} \left(1 - \frac{\ell}{m+1} \right) \frac{T-h-\ell}{T-h} 2\widehat{\gamma}_{h,\ell}. \tag{22}$$

where

$$\widehat{\gamma}_{h,\ell} = \frac{1}{T-h-\ell} \sum_{t=\ell+1}^{T-h} \varepsilon_t (y_{t+h} - \alpha_h - x_t'\beta) \varepsilon_{t-\ell} (y_{t-\ell+h} - \alpha_h - x_{t-\ell}'\beta). \tag{23}$$

[16] Montiel Olea and Plagborg-Møller (2021) use lag augmentation to achieve (population) residualized regressors, whereas our setup does not require this step because we assume the researcher has access to ε_t.

Note that $\widehat{\gamma}_{h,\ell}$ is the sample autocovariance of the regression score. Under Assumptions 1 and 2, $\gamma_{h,\ell} = 0$ for $\ell > 0$. As a result, setting $m > 0$ is not asymptotically necessary to conduct valid inference in our LP setup. However, using the NW estimator may have implications for inference in small samples because $\widehat{\gamma}_{h,\ell}$ may have small sample bias akin to the bias we have documented in LP estimators of impulse response functions.

6.1 Explicit approximate bias of the autocovariance of the regression score

To investigate the small-sample implications of using the NW estimator when HW will do, we focus attention on the case where $\theta_h = 0$ for all h. We choose this setup for two reasons. First, the null hypothesis that $\theta_h = 0$ is often of interest for researchers using LPs and our derivation would be correct under that hypothesis. Second, the assumption that $\theta_h = 0$ facilitates deriving expressions for the approximate bias in $\widehat{\gamma}_{h,\ell}$. In Appendix D we show that, under suitable regularity conditions, when α_h, β_h, and $\gamma_{h,\ell}$ are jointly estimated

$$\mathbb{E}\widehat{\gamma}_{h,\ell} - \gamma_{h,\ell} = -\frac{1}{T-h-\ell}\gamma_{h,0} + O\left(T^{-3/2}\right). \tag{24}$$

It is immediate that increasing m in the NW estimator is likely to decrease the value of the element of $\widehat{S}_{T-h}$ that is relevant to construct standard errors for the least-squares estimator in our LP setting decreases, which reduces the size standard error.[17]

6.2 Bias in SEs in the context of an AR(1) example

To explore the small sample bias of standard errors for LPs, we again use the AR(1) data generating process in equation (5) for Monte Carlo simulations. We consider two specifications. In the first specification, we assume that $\theta_0 = 0$, $\sigma_\varepsilon^2 = 1$, $\sigma_\nu^2 = 2$. This specification conforms to our assumption in the previous subsection that $\theta_h = 0$ for all h. In the second specification, we assume that $\theta_0 = 1$, $\sigma_\varepsilon^2 = 1$, $\sigma_\nu^2 = 1$. This specification has $\theta_h \neq 0$ for all h, but maintains the same variance of y_t as in the first specification. We set $c_{t-1} = y_{t-1}$.[18]

[17] In Appendix D, we show that, for $\ell > 0$, the approximate bias in $\widehat{\gamma}_{h,\ell}$ is the same in the case when α_h is known. Additionally, we derive the approximate bias in the case when controls are not included in the LP. The expression for the approximate bias in terms of $\gamma_{0,h}$ is unchanged.

[18] We focus on HW and NW estimators to isolate the effect of bias in the estimated regression score autocovariances and also because of their popularity in practice. In fact, here the errors are homoskedastic, so heteroskedasticity-robust estimators are unnecessary.

We fix the horizon of the LP to be $h = 10$ and estimate $\gamma_{10,\ell}$ using the fitted regression scores from the LPs.[19] Figure 8 shows the Monte Carlo mean of $\widehat{\gamma}_{10,\ell}$ for $\ell > 0$ in each of the two specifications. Several remarks are in order. First, when $\theta_h = 0$, the Monte Carlo mean of $\widehat{\gamma}_{10,\ell}$ is roughly constant for all ℓ shown, as implied by the analytic expression for the approximate bias derived in the previous subsection. Second, when $\theta_h > 0$, the Monte Carlo mean of $\widehat{\gamma}_{10,\ell}$ is also negative and for some h is more negative than when $\theta_h = 0$. Third, regardless of whether $\theta_h = 0$ or $\theta_h > 0$, the Monte Carlo mean of $\widehat{\gamma}_{10,\ell}$ is more negative for smaller values of T or larger values of ρ.

To analyze the effect of using estimates of the autocovariance of the regression score to construct standard errors, we consider coverage probabilities from symmetric 95% confidence intervals constructed using the method of HW and NW. We also consider the equally-weighted cosine (EWC) estimator discussed in Lazarus et al. (2018), a frequency-domain-based alternative standard error estimator.[20] Figure 9 displays the coverage probabilities for both $\widehat{\theta}_{LS}$ (left column) and $\widehat{\theta}_{BCC}$ (right column) of confidence intervals constructed using the three standard error estimates, with the rows corresponding to different values of ρ. For all ρ, T the frequency-domain-based EWC delivers the best coverage, except perhaps at $h = 0$. The NW-based confidence interval uniformly provides the worst coverage. This is despite the fact that there is in fact some autocorrelation in the population regression scores. Finally, the CI using $\widehat{\theta}_{BCC}$ provide better coverage than those using $\widehat{\theta}_{LS}$ regardless of the standard errors used. Overall, our results point show the favorable performance of the $\widehat{\theta}_{BCC}$ and the poor performance of NW. While the EWC-based intervals are clearly superior, in practice some users may continue to prefer the conventional time-domain based estimators, in which case they should use HW.

6.3 Medium-Scale Model Revisited

Figure 10 displays the same coverage probabilities for the CEE real GDP (top), the price level (middle) and the federal funds rate (bottom) for $T = 100$. Once again, the EWC-based confidence

[19] In Appendix D, we derive the bias under the assumption that α_h, β_h, and $\gamma_{h,\ell}$ are estimated as a part of the same GMM system. As a result, those derivations work with only $T - h - \ell$ observations. For our Monte Carlo exercises in this sub-section, we instead calculate $\widehat{\gamma}_{h,\ell}$ using the $T - h$ values of the fitted regression score. So, α_h and β_h are estimated using $T - h$ observations for every ℓ.

[20] For EWC we use a bandwidth $\approx 0.41T^{2/3}$. One can also use the NW estimator under fixed-b asymptotics by appropriate choice of bandwidth ($\approx 1.3\sqrt{T-h}$). We found this performed poorly, and so omit it from the presentation.

interval performs the best, with the NW-based ones perform the worst. The HW-based intervals always perform better than the NW ones, suggesting again that among time-domain based estimators of standard errors, they have the best finite sample performance. The coverage for the price level is particularly poor at large h. $\widehat{\theta}_{h,BCC}$-based intervals are not uniformly better (or worse) than $\widehat{\theta}_{h,LS}$ ones.

7 Conclusion

We have shown that LPs can be severely biased in sample sizes commonly found in the related literature. Our analytic expression for the approximate bias have shown that LPs are intimately linked across horizons in small samples. We also have shown how researchers could used our expression for the approximate bias to bias correct LPs. When correcting for bias, researchers face a small sample tradeoff betwen bias and MSE. Using Monte Carlo analysis, we have demonstrated that the performance of our bias corrected estimator depends on the underlying data generating process and on the LP horizon of interest. In our Monte Carlo examples, our bias correction does not completely correct for the bias in LPs. This 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 have also analyzed bias in standard errors computed for estimated impulse response functions from LPs. In small samples, standard errors that rely on estimated autocovariances of the regression score, like the NW estimator, typically understate the amount of uncertainty surrounding the estimated impulse response functions. 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 NW standard errors, they may want to check to see if HW or fixed-b standard errors would lead to different conclusions. If the HW standard errors are larger than the NW standard errors, researchers should consider what might lead to the apparent negative autocovariance in the regression score. Without another reasonable theory, it may be that the negative estimates of the autocovariance of the regression score are the result of small sample bias.

References

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.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of Political Economy*, 113, 1–45.

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.

Herbst, E. and B. K. Johannsen (2021): "Bias in local projections," .

Huber, P. J. (1967): "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather Modification; University of California Press: Berkeley, CA, USA*.

Jordà, Ò. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161–182.

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.

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.

Montiel Olea, J. L. and M. Plagborg-Møller (2021): "Local Projection Inference Is Simpler and More Robust Than You Think," *Econometrica*, 89, 1789–1823.

Müller, U. K. (2014): "HAC Corrections for Strongly Autocorrelated Time Series," *Journal of Business & Economic Statistics*, 32, 311–322.

Nagar, A. L. (1959): "The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations," *Econometrica: Journal of the Econometric Society*, 575–595.

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

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

Nicholls, D. and A. L. Pope (1988): "Bias in the estimation of multivariate autoregressions," *Australian Journal of Statistics*, 30, 296–309.

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

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

Pope, A. L. (1990): "Biases of estimators in multivariate non-Gaussian autoregressions," *Journal of Time Series Analysis*, 11, 249–258.

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.

Sawa, T. (1972): "Finite-sample properties of the k-class estimators," *Econometrica*, 653–680.

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

Sims, C. A. and T. Zha (1999): "Error Bands for Impulse Responses," *Econometrica*, 67, 1113–1155.

Srinivasan, T. N. (1970): "Approximations to Finite Sample Moments of Estimators Whose Exact Sampling Distributions are Unknown," *Econometrica*, 38, 533–541.

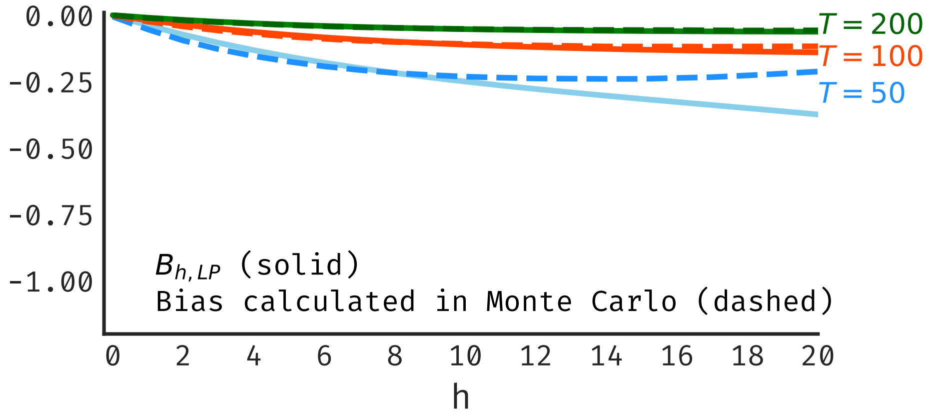
Sun, Y. (2014): "Let's Fix It: Fixed-b Asymptotics Versus Small-b Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference," *Journal of Econometrics*, 178, 659–677.

Theil, H. (1961): *Economic forecasts and policy*, North-Holland Pub. Co.

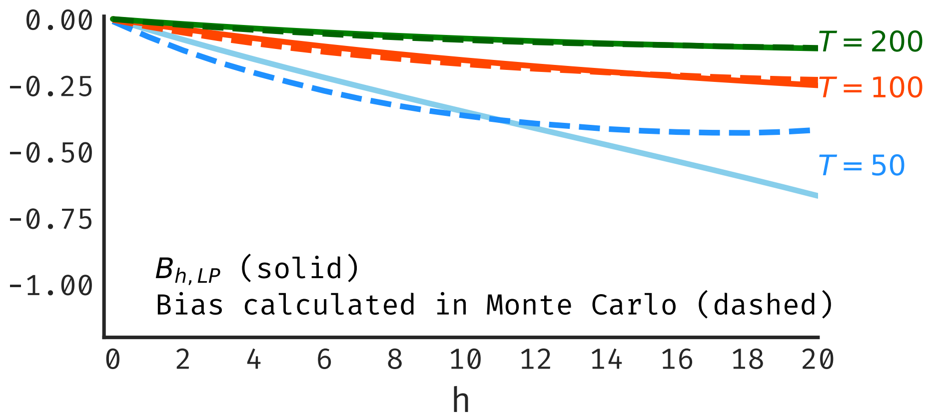
White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

Figure 2: $B_{h,LP}$ performs well in empirically-relevant samples when y_t is an AR(1).

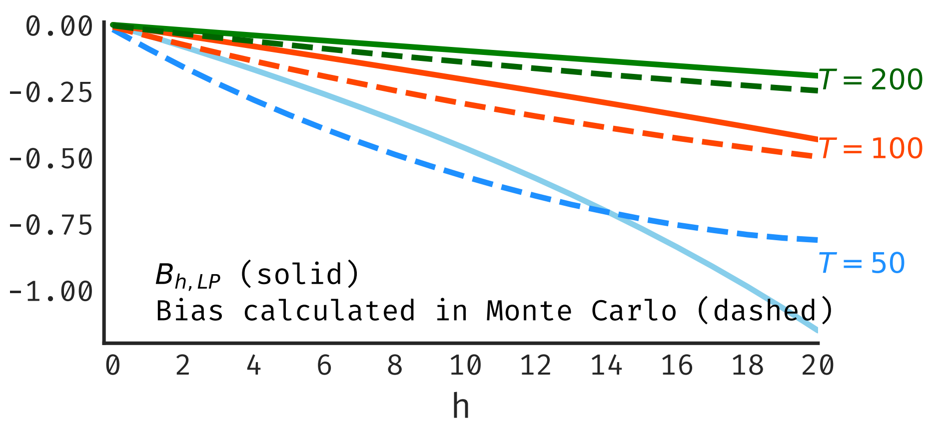
(a) $\rho = 0.90$



(b) $\rho = 0.95$



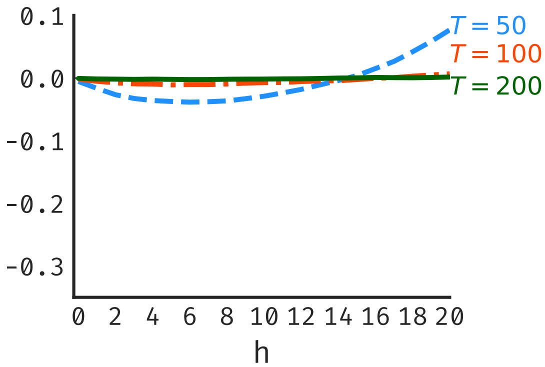
(c) $\rho = 0.99$



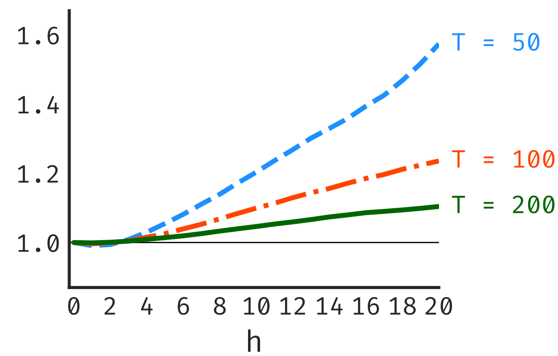
Note: the sub-figures show the value of $B_{h,LP}$ and the Monte Carlo means of $\widehat{\theta}_{LS}$ estimated on data simulated from equation (5). We use 1,000,000 Monte Carlo simulations. We set $\sigma_\varepsilon = \sigma_\nu = 1$ and $\theta_0 = 1$.

Figure 3: Performance of $B_{h,BCC}$ in an AR(1) example.

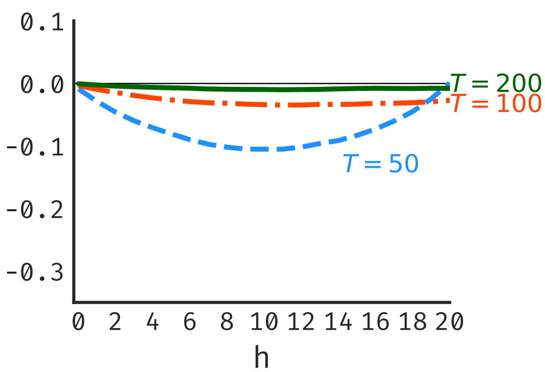
(a) $\rho = 0.90$, $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$



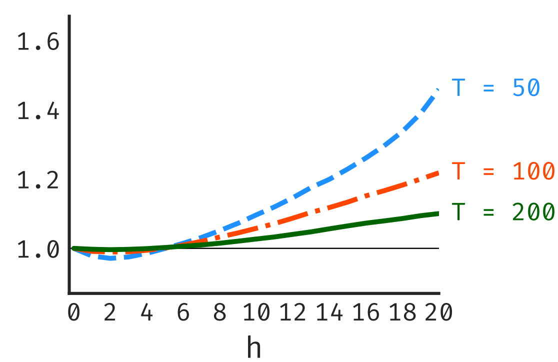
(b) $\rho = 0.90$, MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



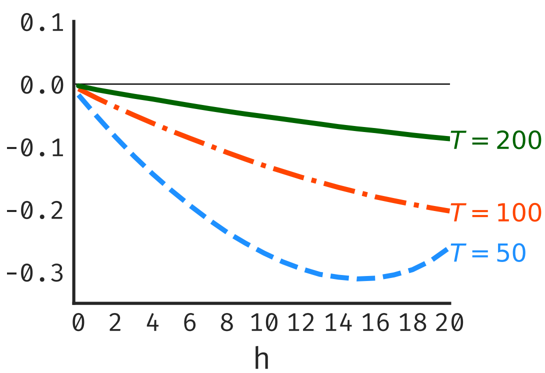
(c) $\rho = 0.95$, $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$



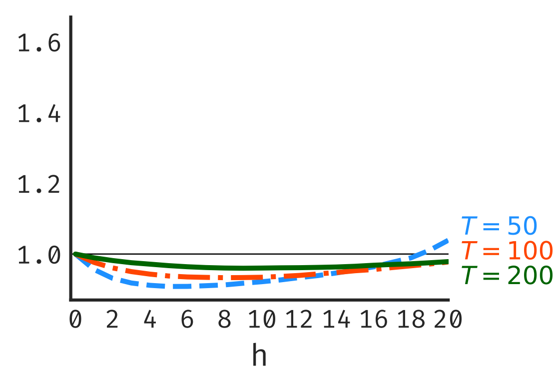
(d) $\rho = 0.95$, MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



(e) $\rho = 0.99$, $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$



(f) $\rho = 0.99$, MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



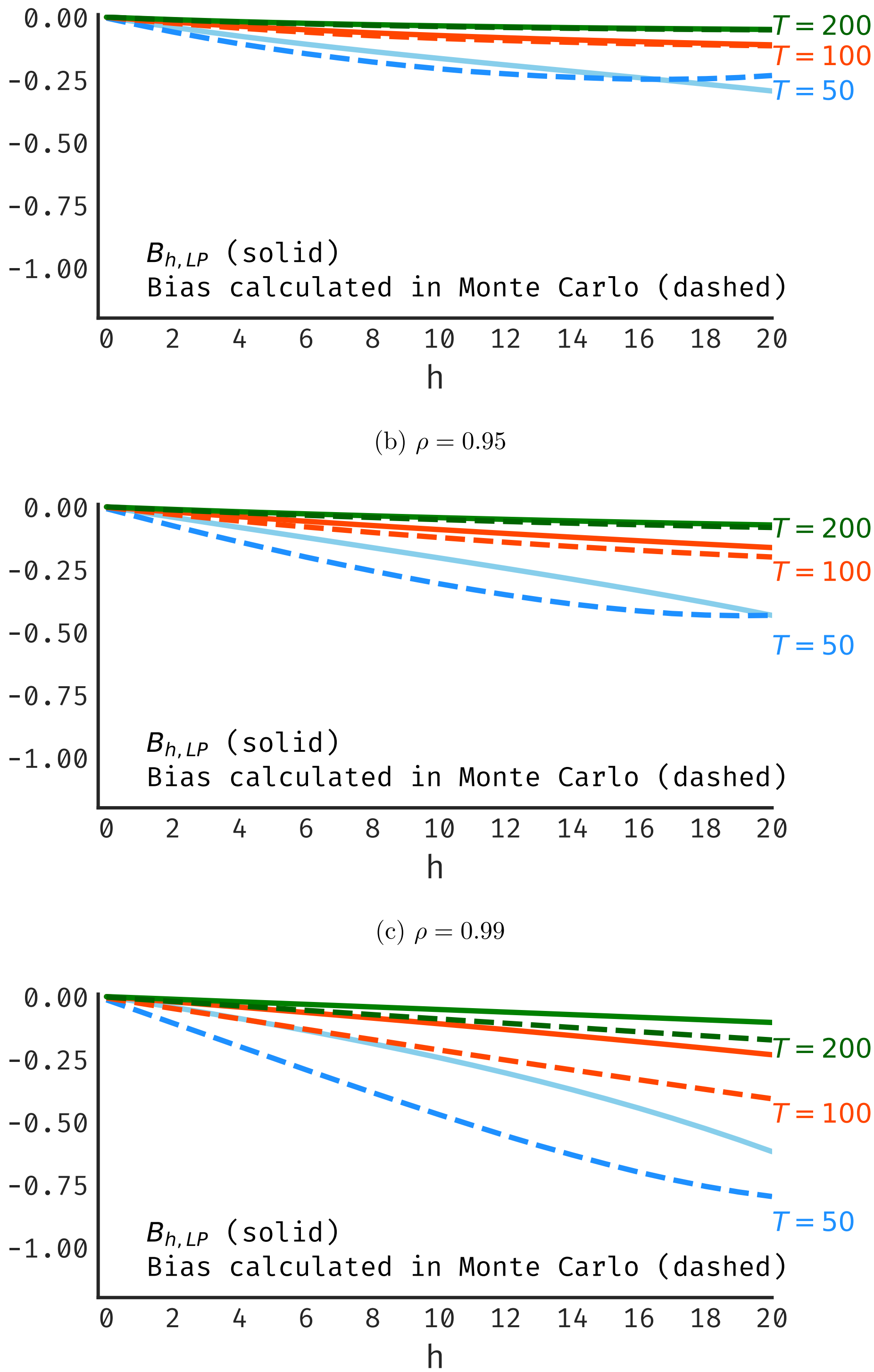
Note: the sub-figures on the left show the Monte Carlo means of $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ estimated on data simulated from equation (5). The sub-figures on the right show the Monte Carlo value of the MSE of $\widehat{\theta}_{h,BCC}$ relative to the MSE of $\widehat{\theta}_{h,LS}$. A number larger than unity indicates that the MSE of $\widehat{\theta}_{h,BCC}$ is larger than $\widehat{\theta}_{h,LS}$. We use 1,000,000 Monte Carlo simulations. We set $\sigma_\varepsilon = \sigma_\nu = 1$ and $\theta_0 = 1$.

Figure 4: $B_{h,LP,I}$ performs well in empirically-relevant samples when $y_{i,t}$ is an AR(1) and $I = 50$.

(a) $\rho = 0.90$

(b) $\rho = 0.95$

(c) $\rho = 0.99$



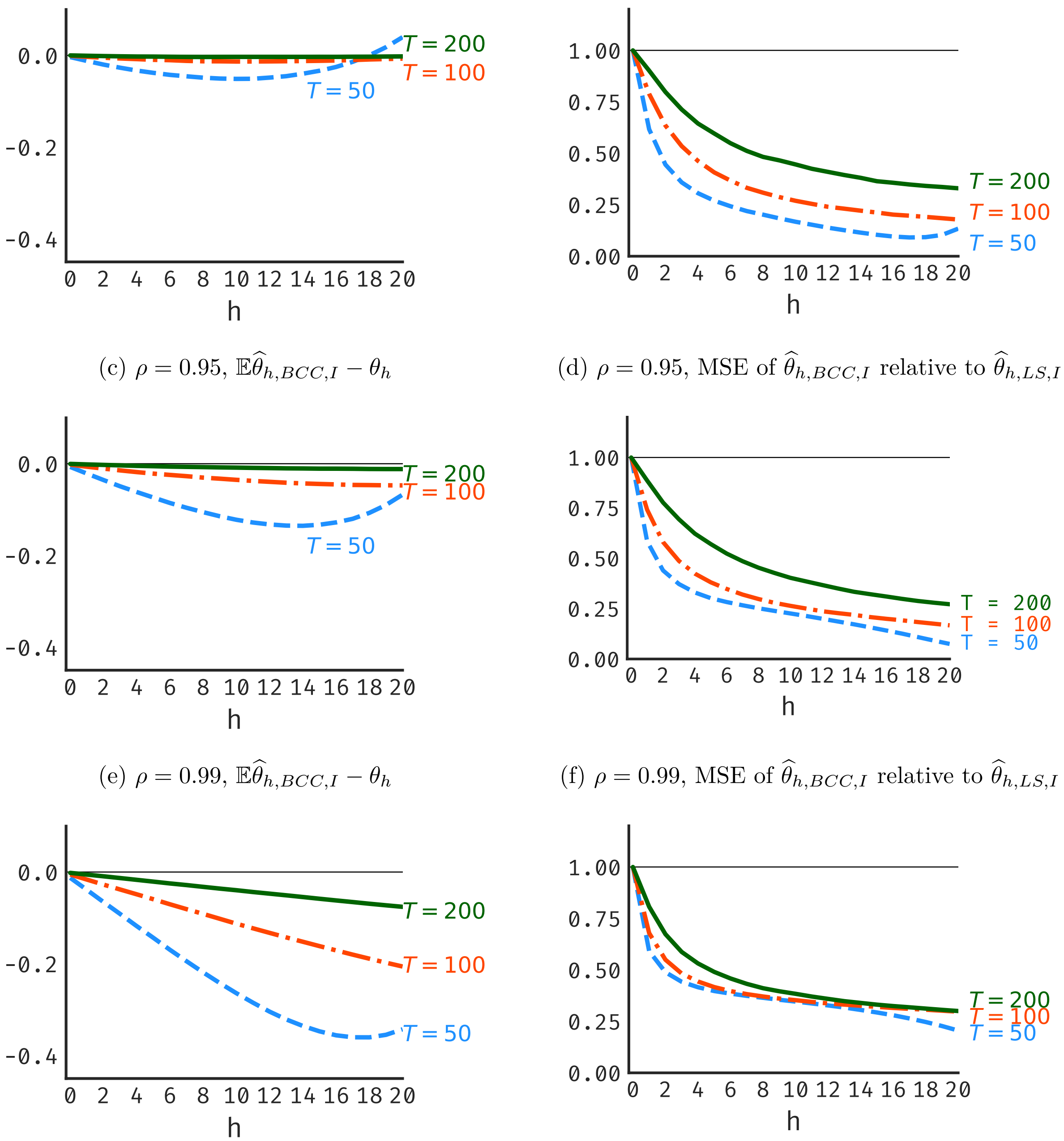
Note: the sub-figures show the value of $B_{h,LP,I}$ and the Monte Carlo means of $\widehat{\theta}_{h,LS,I}$ estimated on data simulated from equation (5). We use 1,000,000 Monte Carlo simulations. We set $\sigma_\varepsilon = \sigma_\nu = 1$ and $\theta_0 = 1$.

Figure 5: Performance of $B_{h,BCC,I}$ in an AR(1) example.

(a) $\rho = 0.90$, $\mathbb{E}\widehat{\theta}_{h,BCC,I} - \theta_h$ (b) $\rho = 0.90$, MSE of $\widehat{\theta}_{h,BCC,I}$ relative to $\widehat{\theta}_{h,LS,I}$

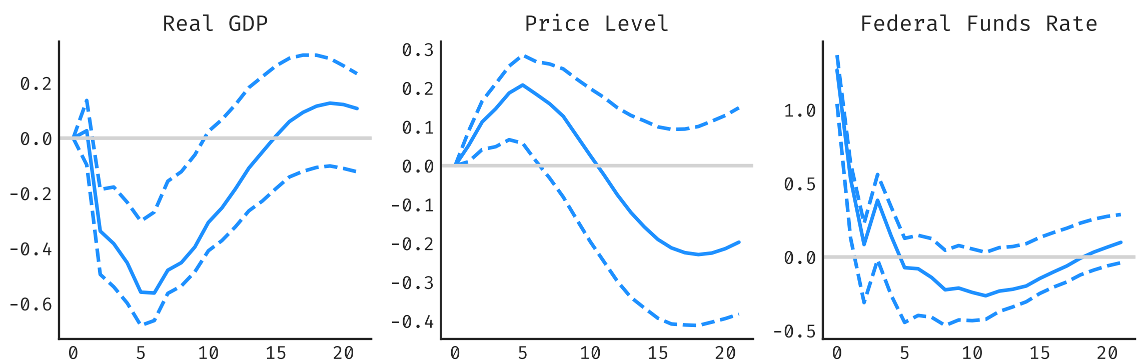
(c) $\rho = 0.95$, $\mathbb{E}\widehat{\theta}_{h,BCC,I} - \theta_h$ (d) $\rho = 0.95$, MSE of $\widehat{\theta}_{h,BCC,I}$ relative to $\widehat{\theta}_{h,LS,I}$

(e) $\rho = 0.99$, $\mathbb{E}\widehat{\theta}_{h,BCC,I} - \theta_h$ (f) $\rho = 0.99$, MSE of $\widehat{\theta}_{h,BCC,I}$ relative to $\widehat{\theta}_{h,LS,I}$



Note: the sub-figures on the left show the Monte Carlo means of $\mathbb{E}\widehat{\theta}_{h,BCC,I} - \theta_h$ estimated on data simulated from equation (16). The sub-figures on the right show the Monte Carlo value of the MSE of $\widehat{\theta}_{h,BCC,I}$ relative to the MSE of $\widehat{\theta}_{h,LS,I}$. A number larger than unity indicates that the MSE of $\widehat{\theta}_{h,BCC,I}$ is larger than $\widehat{\theta}_{h,LS,I}$. We use 1,000,000 Monte Carlo simulations. We set $\sigma_\varepsilon = \sigma_\nu = 1$ and $\theta_0 = 1$.

Figure 6: Impulse Response to Monetary Policy Shock in a CEE-style VAR

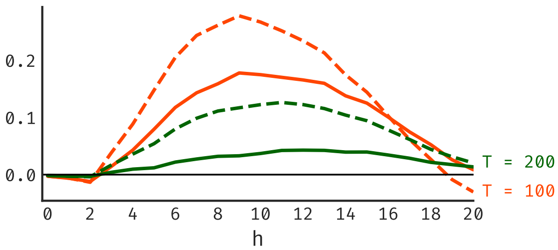


The figure displays impulse response of output (in percent), the price level (in percent), and the federal funds rate (in percentage points) to a one standard deviation increase in the monetary policy shock (identified recursively via the Cholesky factorization.) The solid lines display the median impulse responses and the dashed lines 90 percent confidence intervals computed using the method of Sims and Zha (1999).

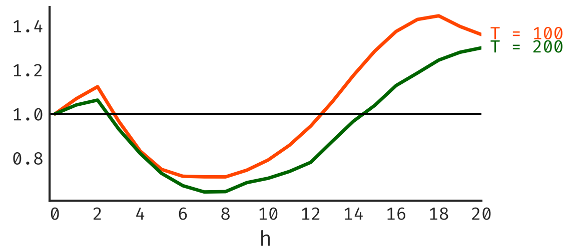
Figure 7: Bias and MSE under a CEE-type VAR Data Generating Process

Real GDP

(a) $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid)

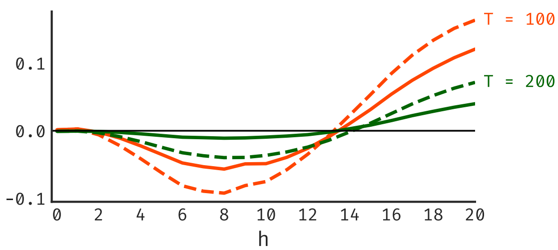


(b) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$

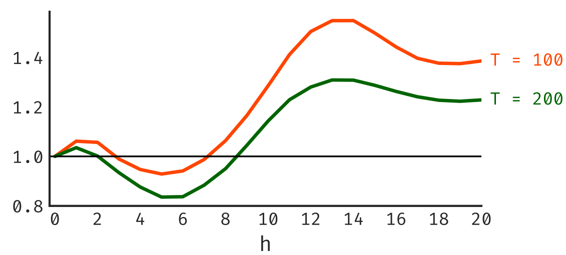


Price Level

(c) $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid)

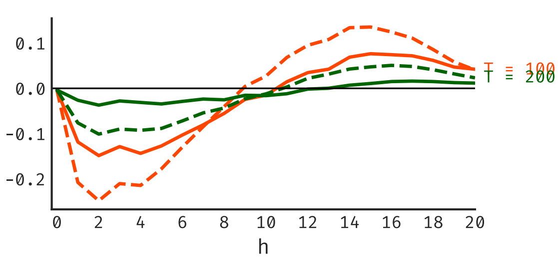


(d) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$

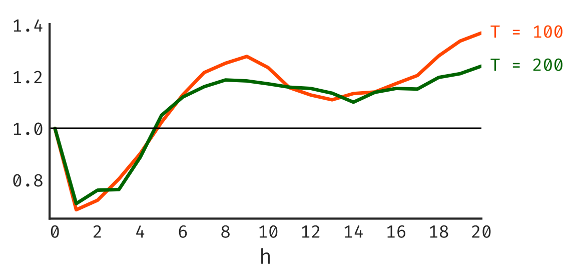


Federal Funds Rate

(e) $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid)

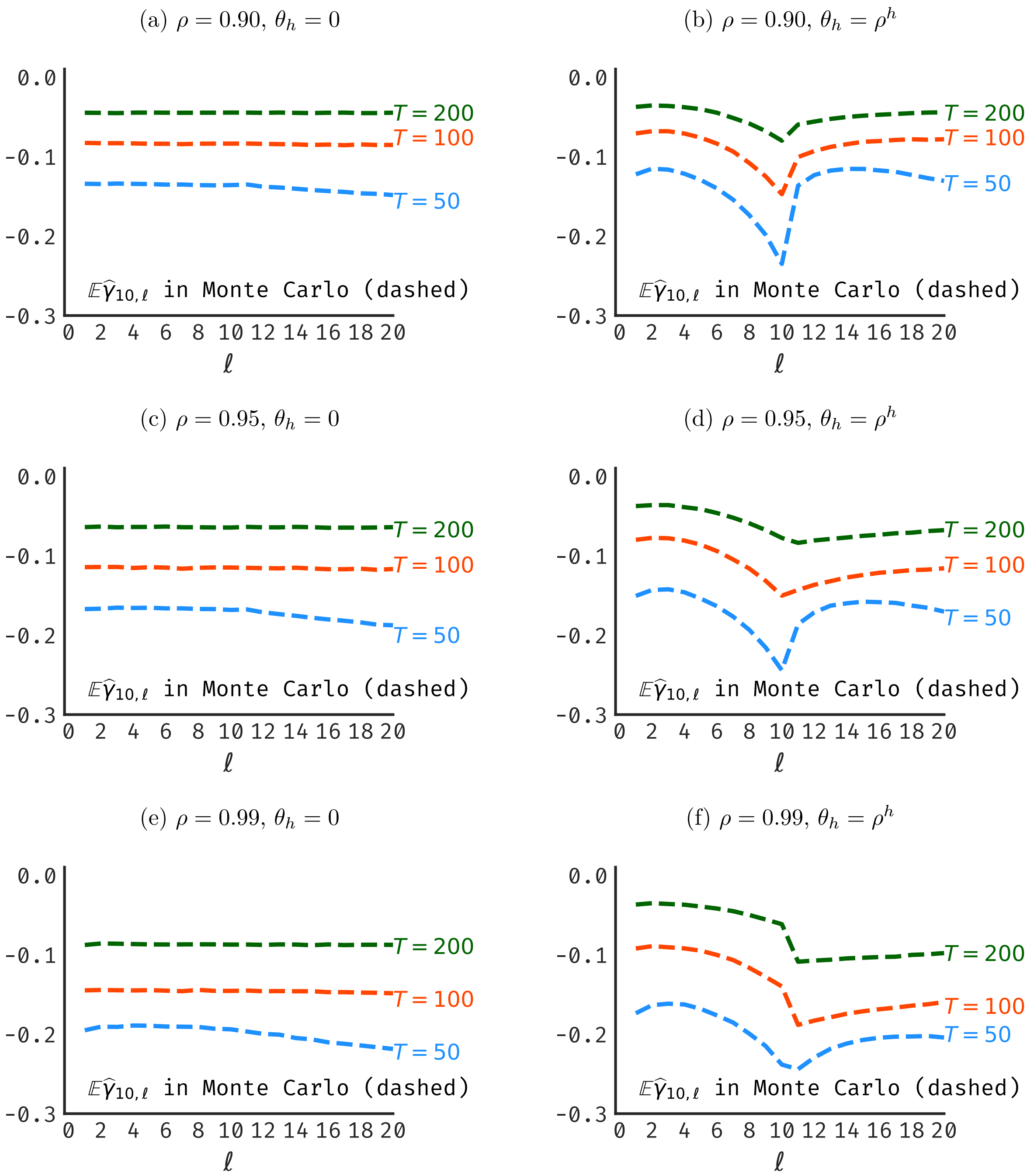


(f) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



The sub-figures in the left column show the $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed lines) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid lines) for $T = 100$ (red) and $T = 200$ (green). The sub-figures in the right column show the ratio of the MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$ for $T = 100$ (red) and $T = 200$ (green).

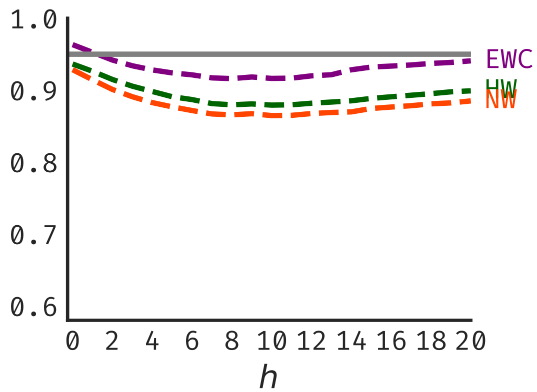
Figure 8: $\widehat{\gamma}_{10,\ell}$ is biased down in small samples when y_t is an AR(1).



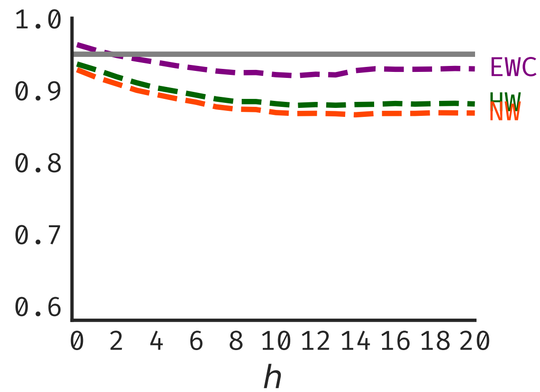
Note: the sub-figures on the left show the Monte Carlo means of $\widehat{\gamma}_{10,\ell}$ estimated on data simulated from equation (5) when $\theta_0 = 0$. The sub-figures on the right show analogous figures when $\theta_0 = 1$. When $\theta_0 = 1$, we set $\sigma_\varepsilon = \sigma_\nu = 1$. When $\theta_0 = 0$, we increase σ_ν so that the variance of y_t is unchanged. We use 1,000,000 Monte Carlo simulations.

Figure 9: Empirical Coverage of 95% Confidence Intervals

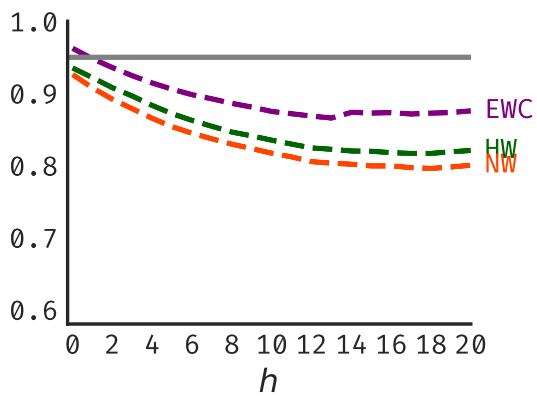
(a) $\rho = 0.90$, $T = 100$, Coverage of $\widehat{\theta}_{LS}$



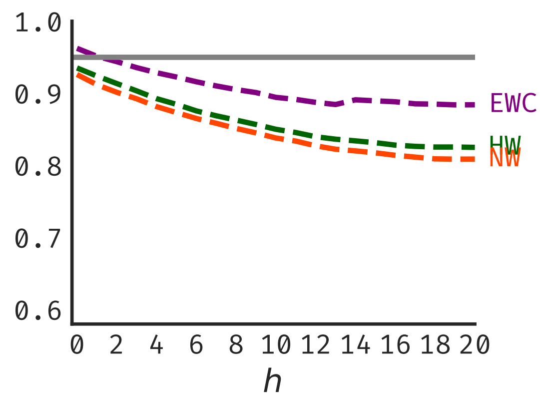
(b) $\rho = 0.90$, $T = 100$, Coverage of $\widehat{\theta}_{BCC}$



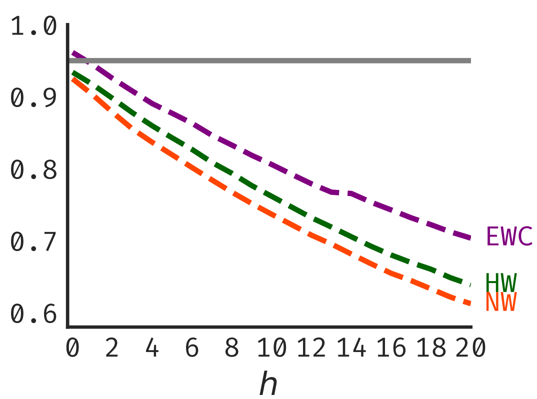
(c) $\rho = 0.95$, $T = 100$, Coverage of $\widehat{\theta}_{LS}$



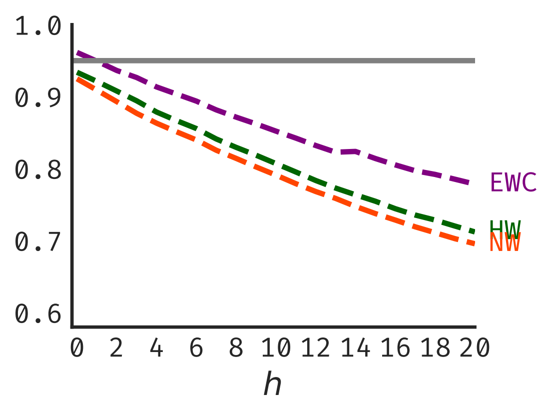
(d) $\rho = 0.95$, $T = 100$, Coverage of $\widehat{\theta}_{BCC}$



(e) $\rho = 0.99$, $T = 100$, Coverage of $\widehat{\theta}_{LS}$



(f) $\rho = 0.99$, $T = 100$, Coverage of $\widehat{\theta}_{BCC}$

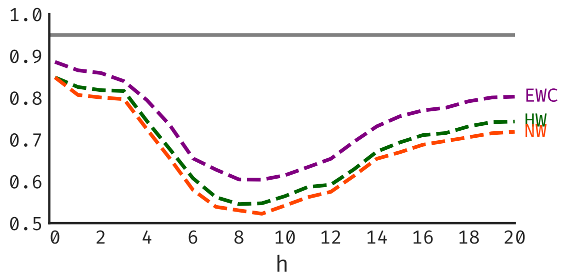


Note: the sub-figures on the left show the empirical coverage of symmetric 95% confidence intervals constructed using HW (green), NW (orange), and EWC (purple) based standard errors for $\widehat{\theta}_{LS}$ and $\widehat{\theta}_{BCC}$.

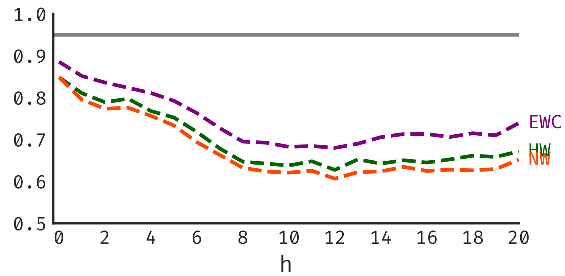
Figure 10: Coverage of 95% Confidence Intervals under a CEE-type VAR DGP

Real GDP

(a) Coverage of $\widehat{\theta}_{h,LS}$

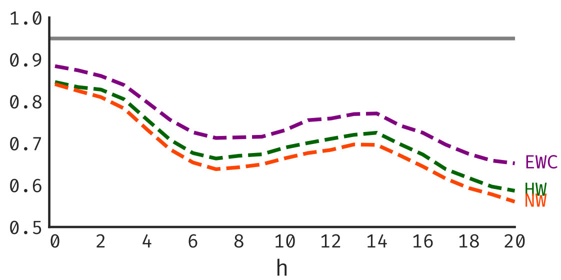


(b) Coverage of $\widehat{\theta}_{h,BCC}$

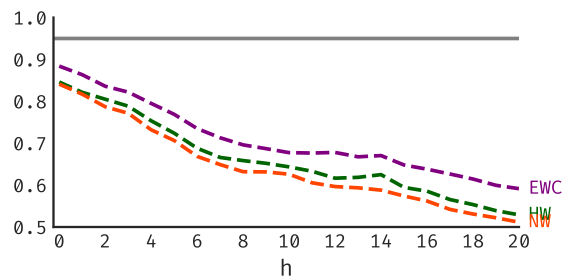


Price Level

(c) Coverage of $\widehat{\theta}_{h,LS}$

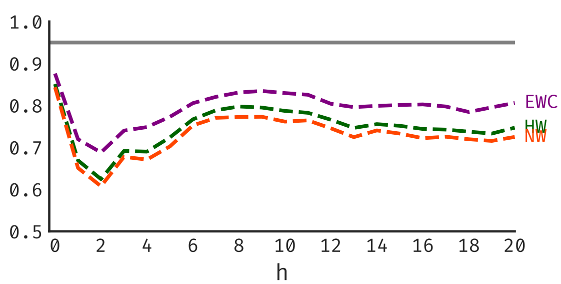


(d) Coverage of $\widehat{\theta}_{h,BCC}$

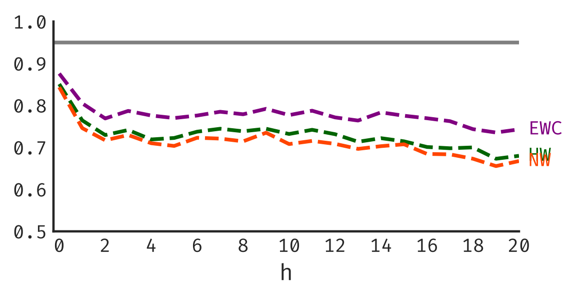


Federal Funds Rate

(e) Coverage of $\widehat{\theta}_{h,LS}$



(f) Coverage of $\widehat{\theta}_{h,BCC}$



Note: the sub-figures on the left show the empirical coverage of symmetric 95% confidence intervals constructed using HW (green), NW (orange), and EWC (purple) based standard errors for $\widehat{\theta}_{LS}$ and $\widehat{\theta}_{BCC}$ for $T = 100$.

Appendix for "Bias in Local Projections"

A Derivations of approximate bias

In this Appendix, we derive our expressions for the approximate 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:

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

Assumption 1, combined with the leaast-squares estimation framework satisfies the necessary assumptions of Rilstone et al. (1996). Assumption 2 facilitates the derivation of tractable expressions for the approximate bias.

In this Appendix, we use the notation of Bao and Ullah (2007) where possible. For each derivation, we will cast the least-squares estimator as a generalized method of moments (GMM) estimator with moment conditions given by $q_h(\beta; \xi_t)$, where $\beta = [\alpha_h, \beta_h']'$ and $\xi_t = [y_{t+h}, x_t']'$ is the data vector for the LP model. The GMM empirical moments are given by

$$\psi_{h,T-h}(\beta; \{\xi_t\}_{t=1}^{T-h}) = \frac{1}{T-h} \sum_{t=1}^{T-h} q_h(\beta; \xi_t). \tag{A.2}$$

Let $\nabla^i A(\beta)$ be the matrix of ith order partial derivatives of A with respect to β. In what follows, we write $\psi_{h,T-h}(\beta; \{\xi_t\}_{t=1}^{T-h})$ as $\psi_{h,T-h}$ and $q_h(\beta; \xi_t)$ as $q_{h,t}$. Define the series of matrices

$$H_i = \nabla^i \psi_{h,T-h} \text{ and } \overline{H}_i = \mathbb{E}\left[H_i\right] \text{ with } Q = \overline{H}_1^{-1}, V = H_1 - \overline{H}_1. \tag{A.3}$$

Bao and Ullah (2007) show that the expressions for the terms in (A.1) are given by:

$$a_{-1/2} = -Q\psi_{h,T-h} \quad \text{and} \quad a_{-1} = -QVa_{-1/2} - \frac{1}{2}Q\overline{H}_2\left[a_{-1/2} \otimes a_{-1/2}\right]. \tag{A.4}$$

We are interested in computing

$$B = \mathbb{E}[a_{-1/2} + a_{-1}] = \mathbb{E}\left\{QVQ\psi_{h,T-h}\right\} - \mathbb{E}\left\{\frac{1}{2}Q\overline{H}_2\left[(Q\psi_{h,T-h}) \otimes (Q\psi_{h,T-h})\right]\right\}, \tag{A.5}$$

which we refer to as the approximate bias. Throughout this Appendix, and without loss of generality, we assume all data have mean zero. Before proceeding, we introduce notation for second moments of the data

$$\sigma_\varepsilon^2 = \mathbb{E}\left[\varepsilon_t^2\right], \quad \sigma_y^2 = \mathbb{E}\left[y_t^2\right], \quad \Sigma_{c,j} = \mathbb{E}\left[c_{t-j}c_t'\right]. \tag{A.6}$$

A.1 LP when means are estimated

In the LP model when means are estimated, the moment conditions for the least-squares estimator are

$$\mathbb{E}\begin{bmatrix} y_{t+h} - \alpha_h - x_t'\beta_h \\ x_t\left(y_{t+h} - \alpha_h - x_t'\beta_h\right) \end{bmatrix} = 0. \tag{A.7}$$

Because, in the notation of Bao and Ullah (2007), $\overline{H}_2 = \mathbf{0}$, to calculate $B_{h,LP}$ we only need to calculate the second element of $\mathbb{E}\left[QH_1Q\psi_{h,T-h}\right]$, which is given by

$$\mathbb{E}\left[QH_1Q\psi_{h,T-h}\right]_2 = -\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\mathbb{E}\left[\phi_1(t,s) + \phi_2(t,s) + \phi_3(t,s)\right], \tag{A.8}$$

where

$$\phi_1(t,s) = \frac{\varepsilon_t}{\sigma_\varepsilon^2}\left(y_{s+h} - \alpha_h - x_s'\beta_h\right), \tag{A.9}$$

$$\phi_2(t,s) = \left(\frac{\varepsilon_t}{\sigma_\varepsilon^2}\right)^2\varepsilon_s\left(y_{s+h} - \alpha_h - x_s'\beta_h\right), \tag{A.10}$$

$$\phi_3(t,s) = \frac{\varepsilon_t}{\sigma_\varepsilon^2}c_{t-1}'\Sigma_{c,0}^{-1}c_{s-1}\left(y_{s+h} - \alpha_h - x_s'\beta_h\right). \tag{A.11}$$

Consider first $\mathbb{E}[\phi_1(t,s)]$. When $t \leq s$ this expectation equals zero by Assumption 2. When $t > s+h$ this expectation is zero by Assumption 1. Finally, when $s < t \leq s+h$, then direct calculation yields $\mathbb{E}[\phi_1(t,s)] = \theta_{s+h-t}$. Next consider $\mathbb{E}[\phi_2(t,s)]$. If $t = s$, then under Assumption 2, this expectation is zero. If $t \neq s$, then the expectation is again zero owing to Assumptions 1 and 2. Finally, consider $\mathbb{E}[\phi_3(t,s)]$. When $t > s+h$, or $t \leq s$ $\mathbb{E}\left[\phi_3(t,s)\right] = 0$ by Assumptions 1 and 2. When $s < t \leq s+h$

$$\begin{aligned}
\mathbb{E}\left[\phi_3(t,s)\right] =& \mathbb{E}\left[\frac{\varepsilon_t}{\sigma_\varepsilon^2}c_{t-1}'\Sigma_{c,0}^{-1}c_{s-1}\left(u_{h-(t-s),t} + x_t'\beta_{h-(t-s)} - x_s'\beta_h\right)\right] && \text{(A.12)}\\
=& \mathbb{E}\left[\frac{\varepsilon_t}{\sigma_\varepsilon^2}c_{t-1}'\Sigma_{c,0}^{-1}c_{s-1}\left(u_{t,h-(t-s)} + x_t'\beta_{h-(t-s)}\right)\right] && \text{(A.13)}\\
=& \mathbb{E}\left[\frac{\varepsilon_t}{\sigma_\varepsilon^2}c_{t-1}'\Sigma_{c,0}^{-1}c_{s-1}\left(x_t'\beta_{h-(t-s)}\right)\right] && \text{(A.14)}\\
=& \mathbb{E}\left[\frac{1}{\sigma_\varepsilon^2}\mathbb{E}_{t-1}\left[\varepsilon_t\left(x_t'\beta_{h-(t-s)}\right)\right] tr\left\{c_{t-1}'\Sigma_{c,0}^{-1}c_{s-1}\right\}\right] && \text{(A.15)}\\
=& \theta_{h-(t-s)}\mathrm{tr}\left\{\Sigma_{c,0}^{-1}\Sigma_{c,t-s}\right\}. && \text{(A.16)}
\end{aligned}$$

Note that equation (A.12) follows from the definition of $u_{h,t}$, equation (A.13) follows from Assumption 1, equation (A.14) follows from Assumption 2, equation (A.15) follows from the law of iterated expectations, and equation (A.16) follows from Assumption 1. Combining these resulting gives expression for $B_{h,LP}$ given in equation (3).

It is instructive to understand the effect of estimating α_h. To do so, we assume the mean is known—so that the intercepts of the regressions α_h are known—and compare the bias of the LP estimator to our benchmark specification. The moment conditions for the LP are

$$\mathbb{E}\left[x_t\left(y_{t+h} - x_t'\beta_h\right)\right] = 0 \tag{A.17}$$

We are interested in the first element of β_h. Similar calculation to the case when α_h is estimated yields

$$\mathbb{E}[\widehat{\theta}_{h,LS} - \theta_h] = \mathbb{E}\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}[\phi_2(t,s) + \phi_3(t,s)] + O(T^{-3/2}), \tag{A.18}$$

where ϕ_2 and ϕ_3 are define in A.10 and A.11. It is immediate that the contribution to $B_{h,LP}$ from estimating the mean is given by

$$-\frac{1}{T-h}\sum_{j=1}^{T-h}\theta_{h-j}. \tag{A.19}$$

A.2 LP with no controls

It is not necessary to include controls in equation (1) to have a consistent estimator of θ_h, and a sub-set fo the related LP literature does not include contorls. In this sub-section, we derive the approximate bias when controls are not included.

To do this, we maintain Assumption 1, but drop Assumption 2, and set c_{t-1} to the empty vector. We assume that the following moment conditions hold

$$\mathbb{E}\begin{bmatrix} y_{t+h} - \alpha_h - \theta_h\varepsilon_t \\ \varepsilon_t(y_{t+h} - \alpha_h - \theta_h\varepsilon_t) \end{bmatrix} = 0. \tag{A.20}$$

To get an expression for the approximate bias, the arguments in A.1 go through under the obvious modifications. The resulting approximate bias is

$$-\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\mathbb{E}\left[\phi_1(t,s) + \phi_2(t,s)\right], \tag{A.21}$$

where $\phi_3(t,s)$ does not appear because no controls are included.

Consider first $\mathbb{E}[\phi_2(t,s)]$. When $t > s+h$ this expectation is zero by Assumption 1. When $t = s$, this expectation is zero by the moment condition. When $s < t \leq s+h$ or $t < s$, direct calculation yields $\mathbb{E}[\phi_1(t,s)] = \theta_{s+h-t}$. Next consider $\mathbb{E}[\phi_1(t,s)]$, which is zero by the moment condition. We can now state the approximate bias for the case without controls

Analytical Result 4 (Bias in LPs without controls.)**.** *Under Assumption 1 and under the assumption that equation A.20 holds, the approximate bias of the least-squares estimator of* θ_h

$$-\frac{1}{T-h}\sum_{j=1}^{T-h-1}\left(1-\frac{j}{T-h}\right)(\theta_{h+j}+\theta_{h-j}). \tag{A.22}$$

A few comments are in order. First, the approximate bias is a function of the impulse response at all horizons up to $T-1$. Relative to the case with controls, this means that the impulse response at many more horizons, including those beyond the horizon of interest, affect the approximate bias. Second, it is not feasible to estimate all of the terms that enter the expression for the approximate bias with a finite sample. Any attempt to bias correct would need to truncate j. Third, it is not the case that the approximate bias without controls is necessarily smaller or larger than the approximate bias with controls. Instead, the magnitude of the bias depends on the particular regression models.

It is instructive to understand the effect of estimating α_h. To do this, we assume that the means of the data (and thus α_h) are known. For the LS estimator, we assume that the following moment condition holds

$$\mathbb{E}\left[\varepsilon_t(y_{t+h}-\theta_h\varepsilon_t)\right]=0. \tag{A.23}$$

The approximate bias is

$$-\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\mathbb{E}\left[\phi_2(t,s)\right], \tag{A.24}$$

where ϕ_2 is defined in equation (A.10). Note that $\phi_1(t,s)$ does not appear in this expression because the means are known, and $\phi_3(t,s)$ does not appear because no controls are included. It is immediate that from the arguments in the previous sub-section that when controls are not included, all of the approximate bias is due to estimating the mean.

A.3 LP with panel data

We use the notation defined in section 4. The moment conditions for the estimator are

$$0=\mathbb{E}\left[\frac{1}{I}\ \textstyle\sum_{i=1}^{I}\ x_{i,t}\left(y_{i,t+h}-x_{i,t}'\beta_h\right)\ \right]. \tag{A.25}$$

Applying the results of Bao and Ullah (2007), the approximate bias of $\widehat{\theta}_{h,LS,I}$ is given by

$$-\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\mathbb{E}\left[\Phi_1(t,s)+\Phi_2(t,s)+\Phi_3(t,s)\right], \tag{A.26}$$

where

$$\Phi_1(t,s)=\mathbb{E}\left\{\frac{1}{I}\sum_{i=1}^{I}\frac{\varepsilon_{i,t}}{\sigma^2_{\varepsilon,I}}\left(y_{i,s+h}-x'_{i,s}\beta_h\right)\right\} \tag{A.27}$$

$$\Phi_2(t,s)=\mathbb{E}\left\{\frac{1}{I}\sum_{i=1}^{I}\frac{\varepsilon^2_{i,t}}{\left(\sigma^2_{\varepsilon,I}\right)^2}\frac{1}{I}\sum_{k=1}^{I}\varepsilon_{k,s}\left(y_{k,s+h}-x'_{k,s}\beta_h\right)\right\} \tag{A.28}$$

$$\Phi_3(t,s)=\mathbb{E}\left\{\frac{1}{I}\sum_{i=1}^{I}\frac{\varepsilon_{i,t}}{\sigma^2_{\varepsilon,I}}\frac{1}{I}\sum_{k=1}^{I}\tilde{c}'_{i,t-1}\Sigma^{-1}_{c,0,I}\tilde{c}_{k,s-1}\left(y_{k,s+h}-x'_{k,s}\beta_h\right)\right\} \tag{A.29}$$

Consider $\Phi_1(t,s)$. Under Assumptions 3 and 4, $\Phi_1(t,s)=0$ when $t\le s$ or $t>s+h$. If $s<t\le s+h$, $\Phi_1(t,s)=\theta_{s+h-t}$. Consider $\Phi_2(t,s)$. Under Assumptions 3 and 4, $\Phi_2(t,s)=0$ for all t and s. Consider $\Phi_3(t,s)$. Under Assumptions 3 and 4, $\Phi_3(t,s)=0$ if $t\le s$ or $t>s+h$. When $s<t\le s+h$

$$\begin{aligned}
\Phi_3(t,s)&=\mathbb{E}\left\{\frac{1}{I}\sum_{i=1}^{I}\frac{\varepsilon_{i,t}}{\sigma^2_{\varepsilon,I}}\frac{1}{I}\sum_{k=1}^{I}\tilde{c}'_{i,t-1}\Sigma^{-1}_{c,0,I}\tilde{c}_{k,s-1}\left(u_{k,h-(t-s),t}+x'_{k,t}\beta_{h-(t-s)}\right)\right\} && \text{(A.30)}\\
&=\mathbb{E}\left\{\frac{1}{I}\sum_{i=1}^{I}\frac{\varepsilon_{i,t}}{\sigma^2_{\varepsilon,I}}\frac{1}{I}\sum_{k=1}^{I}\tilde{c}'_{i,t-1}\Sigma^{-1}_{c,0,I}\tilde{c}_{k,s-1}x'_{k,t}\beta_{h-(t-s)}\right\} && \text{(A.31)}\\
&=\mathbb{E}\left\{\frac{1}{I}\sum_{i=1}^{I}\frac{1}{\sigma^2_{\varepsilon,I}}\frac{1}{I}\sum_{k=1}^{I}\mathbb{E}_{t-1}\left[\varepsilon_{i,t}x'_{k,t}\beta_{h-(t-s)}\right]\tilde{c}'_{i,t-1}\Sigma^{-1}_{c,0,I}\tilde{c}_{k,s-1}\right\} && \text{(A.32)}\\
&=\frac{1}{I}\sum_{i=1}^{I}\frac{1}{I}\sum_{k=1}^{I}\mathrm{tr}\left\{\Sigma^{-1}_{c,0,I}\mathbb{E}\left[\tilde{c}_{k,s-1}\tilde{c}'_{i,t-1}\right]\right\}\frac{\sigma_{\varepsilon,i,k}}{\sigma^2_{\varepsilon,I}}\theta_{s+h-t}. && \text{(A.33)}
\end{aligned}$$

Note that A.30 follows from the defnition of $u_{k,h-(t-s),t}$ and Assumption 3, A.31 follows from Assumption 4, A.32 follows from the law of iterated expectations, and A.33 follows from Assumptions 3 and 4. Combining these observations delivers teh results reported in Analytic Result 3.

It is instructive to understand the effect of estimating $\alpha_{i,h}$. In this case, the moment conditions are

$$0=\mathbb{E}\left[\frac{1}{I}\ \textstyle\sum_{i=1}^{I}\ \ [c'_{i,t},c'_t]'\left(y_{i,t+h}-x'_{i,t}\beta_h\right)\ \ \right], \tag{A.34}$$

where the first I elements of β_h are known. Algebra similar to the previous sub-section yields that the approximate bias is given by

$$-\frac{1}{(T-h)^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\mathbb{E}\left[\Phi_2(t,s)+\Phi_3(t,s)\right], \tag{A.35}$$

where $\Phi_1(t,s)$ and $\Phi_2(t,s)$ are as defined in equations (A.28) and (A.29). It is immediate that the contribution to $B_{h,LP,I}$ from estimating the mean is given by

$$-\frac{1}{T-h}\sum_{j=1}^{T-h}\theta_{h-j}. \tag{A.36}$$

B Analytic comparison between LP and parametric model

Here, we provide details of the derivation of the bias in the AR estimator $\widehat{\theta}_0\widehat{\rho}^h$ discussed in Section 3.2.3 and under the data generating process given in equation 5. To derive the bias, we work with the following moment conditions

$$\mathbb{E}\begin{bmatrix} y_t-\alpha-\varepsilon_t\theta_0-\rho y_{t-1} \\ \varepsilon_t\left(y_t-\alpha-\varepsilon_t\theta_0-\rho y_{t-1}\right) \\ y_{t-1}\left(y_t-\alpha-\varepsilon_t\theta_0-\rho y_{t-1}\right) \\ \theta_h-\theta_0\rho^h \end{bmatrix}=\mathbf{0}. \tag{B.1}$$

Note that Assumptions 1 and 2 are satisfied. For notational ease, and without loss of generality we assume that $\mathbb{E}[y_t]=0$, meaning $\alpha_h=0$.

With appropriately defined vectors and matrices analogous to those discussed in Appendix A that also are in the notation of Bao and Ullah (2007), $\mathbb{E}\left[Q\left(H_1-\overline{H}_1\right)Q\psi_{h,T-1}\right]_4$ is given by

$$\left(\frac{1}{T-1}\right)^2\sum_{t=2}^{T}\sum_{s=2}^{T}\mathbb{E}\left[-\left(\frac{\rho^h}{\sigma_\varepsilon^2}\varepsilon_t+\frac{\theta_0 h\rho^{h-1}}{\sigma_y^2}y_{t-1}\right)\left(1+\frac{\varepsilon_t\varepsilon_s}{\sigma_\varepsilon^2}+\frac{y_{t-1}y_{s-1}}{\sigma_y^2}\right)\nu_s\right]. \tag{B.2}$$

It can be shown that this quantity can be expressed as

$$\mathbb{E}\left[Q\left(H_1-\overline{H}_1\right)Q\psi_{h,T-1}\right]_4=-\frac{1}{T-1}\theta_0 h\rho^{h-1}\gamma(1+3\rho)+O(T^{-3/2}), \tag{B.3}$$

where $\gamma=\frac{\sigma_\nu^2}{\sigma_\nu^2+\theta_0^2\sigma_\varepsilon^2}$.

Additionally,

$$\begin{aligned}\left[\frac{1}{2}Q\overline{H}_2\left[a_{-1/2}\otimes a_{-1/2}\right]\right]_4 =&-\frac{1}{2}h\rho^{h-1}\left(\frac{1}{T-1}\right)^2\sum_{t=2}^{T}\sum_{s=2}^{T}\frac{\varepsilon_s}{\sigma_\varepsilon^2}\nu_s\frac{y_{t-1}}{\sigma_y^2}\nu_t \\ &-\frac{1}{2}h\rho^{h-1}\left(\frac{1}{T-1}\right)^2\sum_{t=2}^{T}\sum_{s=2}^{T}\frac{\varepsilon_t}{\sigma_\varepsilon^2}\nu_t\frac{y_{s-1}}{\sigma_y^2}\nu_s \\ &-\frac{1}{2}\theta_0 h\left(h-1\right)\rho^{h-2}\left(\frac{1}{T-1}\right)^2\sum_{t=2}^{T}\sum_{s=2}^{T}\frac{y_{s-1}}{\sigma_y^2}\nu_s\frac{y_{t-1}}{\sigma_y^2}\nu_t \qquad \text{(B.4)} \\ =&-\frac{1}{2}\theta_0 h\left(h-1\right)\rho^{h-2}\frac{1}{T-1}\gamma\left(1-\rho^2\right). \qquad \text{(B.5)}\end{aligned}$$

Then

$$B_{h,AR} = -\frac{1}{T-1}\gamma\theta_0 h\rho^{h-1}\left[(1+3\rho) - \frac{1}{2}(h-1)\left(1-\rho^2\right)\rho^{-1}\right], \tag{B.6}$$

which is the expression given in Section 3.2.3.

C Additional analysis in the context of simple data generating processes

In section 3.2 we analzed $B_{h,LP}$ in the context of an AR(1) data generating process given by equation (5). In this Appendix, we analyze the quality of the approximation offered by $B_{h,LP}$ when means are known, different parameterizations of equation (5) and an AR(2) model that generates hump-shaped impulse response functions.

C.1 The quality of the approximation $B_{h,LP}$ in an AR(1) when the means are known

Here, we analyze the quality of the approximation $B_{h,LP}$ in the context of the AR(1) data generating process in equation (5) when the means are known. Uner this assumption, figure C.1 shows analogous results to those reported in figure 2.

As in the case when α_h is estimated, for $\rho = 0.9$ and $\rho = 0.95$, $B_{h,LP}$ is a good approximation to the exact finite-sample bias in $\widehat{\theta}_{h,LS}$ for all h shown. The quality of the approximation improves somewhat as T increases. Clearly, when $\rho = 0.99$, $B_{h,LP}$ is not as good of an approximation as it is for smaller values of ρ. We conclude that when the means are known and for empirically relevant sample sizes, $B_{h,LP}$ offers a reasonable approximation to the bias in LPs, though the quality of the approximation is somewhat worse for smaller values of T and larger values of ρ.

C.2 Alternative parameterizations of the AR(1) data generating process

In the main text, we assumed $\sigma_\varepsilon = \sigma_\nu = 1$. Here, we consider different settings.

First, consider the case where $\sigma_\varepsilon = 10\sigma_\nu = 1$. In this case, the structural shock ε explains almost all of the variation in y_t. Using this parameterization, figure C.2 shows analogous results to those reported in figure 2. Note that the values of σ_ε and σ_ν do not appear in $B_{h,LP}$. So, the line representing $B_{h,LP}$ are identical in figures 2 and C.2. Additionally, it is apparent that the Monte Carlo means of $\widehat{\theta}_{LS}$ are also very similar when comparing figures figures 2 and C.2.

Figure C.3 shows analogous results to those reported in Figure C.3. While the average bias correction (left panels) is little changed by the change in σ_ν, the perfmance of $\widehat{\theta}_{h,BCC}$ improves relative to $\widehat{\theta}_{h,LS}$ as measured by RMSE.

Next, consider the case where $10\sigma_\varepsilon = \sigma_\nu = 1$. In this case, the structural shock ε explains almost none of the variation in y_t. Using this parameterization, figure C.4 shows analogous results to those reported in figure 2. Note that the values of σ_ε and σ_ν do not appear in $B_{h,LP}$. So, the line representing $B_{h,LP}$ are identical in figures 2 and C.4. Additionally, the Monte Carlo means of $\widehat{\theta}_{LS}$ are also similar when comparing figures figures 2 and C.4.

C.3 An AR(2) example

Here, we analyze the performance of the approximation offered by $B_{h,LP}$ using a simple AR(2) model. We specify the data generating process so that

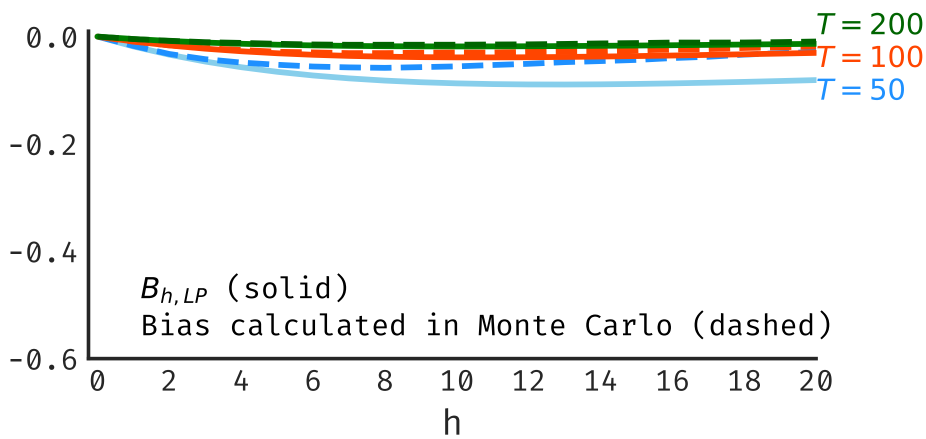
$$y_t = (\rho + \psi)y_{t-1} - \psi\rho y_{t-2} + \theta_0\varepsilon_t + \nu_t \tag{C.1}$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $\nu_t \sim N(0, \sigma_\nu^2)$, and each is independent of each other and over time. We set $\psi = 0.4$ and $\theta_0 = 1$. For the values of ρ that we consider, this data generating process delivers hump-shaped impulse response functions. This particular formulation of the AR(2) is useful because ρ plays a similar role to the AR(1) in that it determines the persistence of y_t. We consider different values for ρ within each figure. The LP is specified so that $x_t = [\varepsilon_t, y_{t-1}, y_{t-2}]'$.

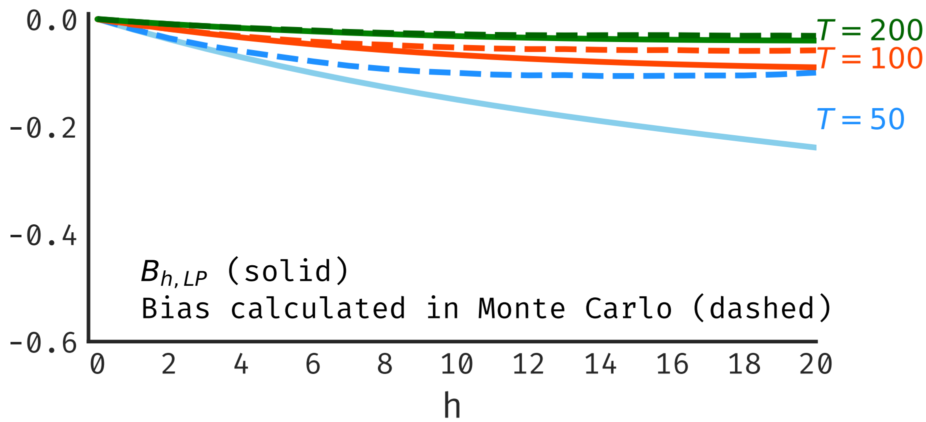
As was the case for the AR(1), For $\rho = 0.9$ and $\rho = 0.95$, $B_{h,LP}$ is a good approximation to the exact finite-sample bias in $\widehat{\theta}_{h,LS}$ for all h shown. The quality of the approximation improves somewhat as T increases. Clearly, when $\rho = 0.99$, $B_{h,LP}$ is not as good of an approximation as it is for smaller values of ρ. Nevertheless, even with $\rho = 0.99$ and $T = 50$ $B_{h,LP}$ captures salient features of the finite-sample bias, including that it is growing in magnitude with h over the values of h shown. We conclude that the quality of the approximation is somewhat worse for smaller values of T and larger values of ρ.

Figure C.1: $B_{h,LP}$ performs well in empirically-relevant samples when y_t is an AR(1) and the means are known.

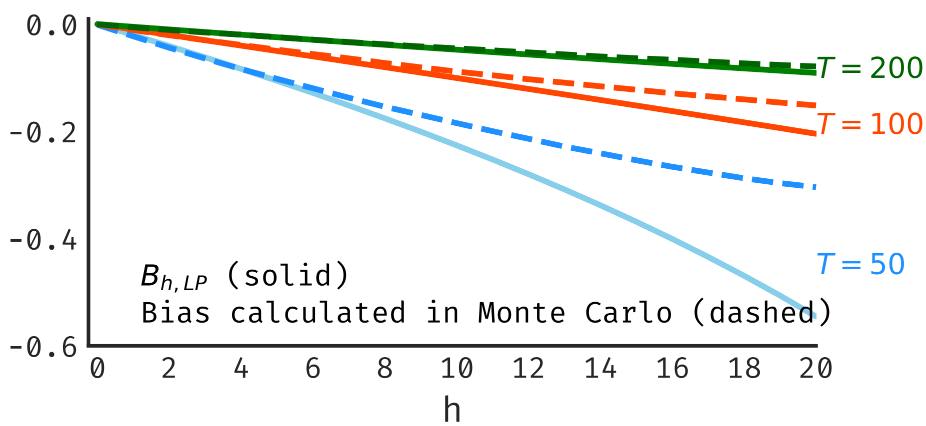
(a) $\rho = 0.90$



(b) $\rho = 0.95$



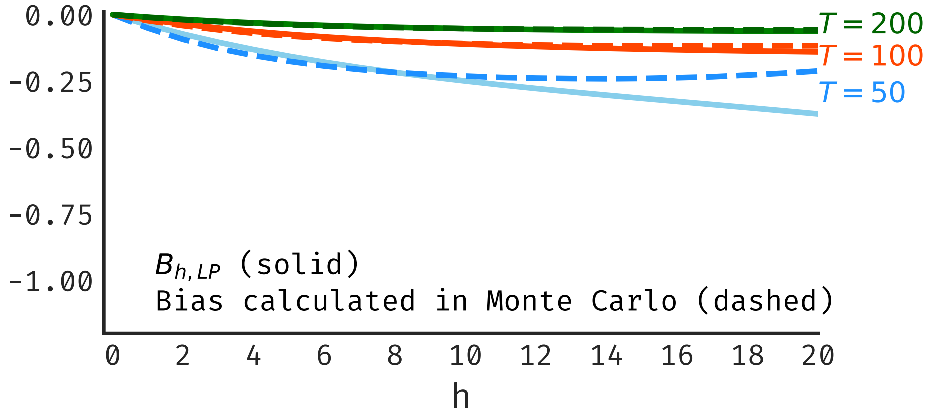
(c) $\rho = 0.99$



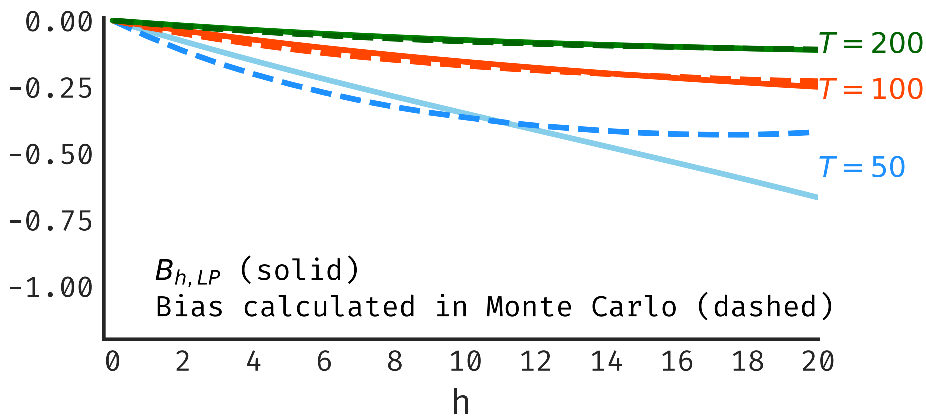
Note: This figure is analogous to Figure 2, but the estimation is done under the assumption that the means of the data are known. The sub-figures show the value of $B_{h,LP}$ and the Monte Carlo means of $\widehat{\theta}_{LS}$ estimated on data simulated from equation (5) under the assumption that α_h is known. We use 1,000,000 Monte Carlo simulations. We set $\sigma_\varepsilon = \sigma_\nu = 1$ and $\theta_0 = 1$.

Figure C.2: $B_{h,LP}$ performs well in empirically-relevant samples when y_t is an AR(1) and σ_ν is small.

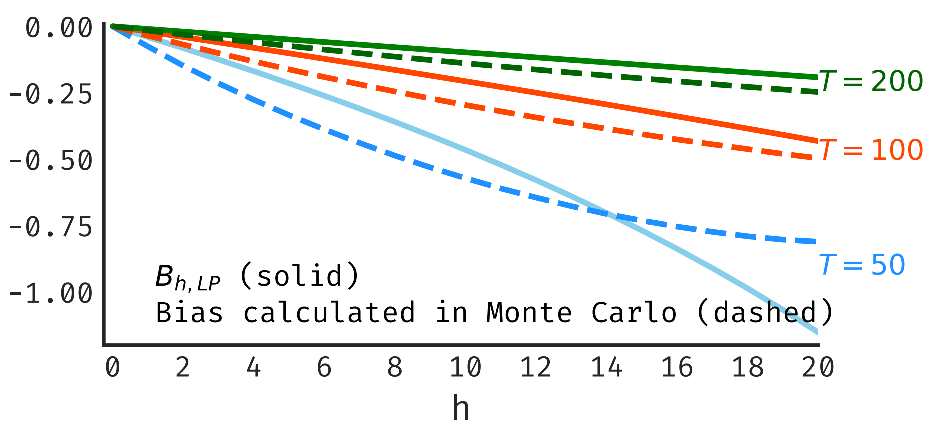
(a) $\rho = 0.90$



(b) $\rho = 0.95$



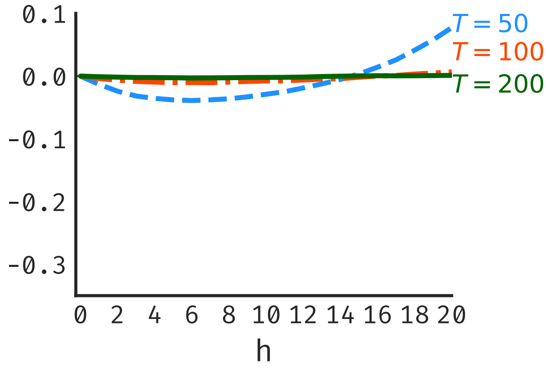
(c) $\rho = 0.99$



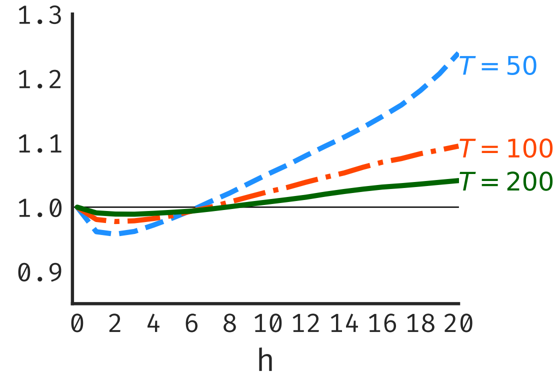
Note: This figure is analogous to Figure 2, but with $\sigma_\varepsilon = 10\sigma_\nu = 1$. The sug-figures show the value of $B_{h,LP}$ and the Monte Carlo means of $\widehat{\theta}_{LS}$ estimated on data simulated from equation (5). We use 100,000 Monte Carlo simulations.

Figure C.3: Performance of $B_{h,BCC}$ in the AR(1) example when σ_ν is small.

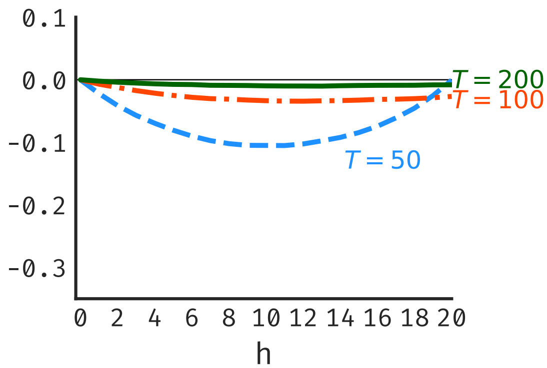
(a) $\rho = 0.90$, $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$



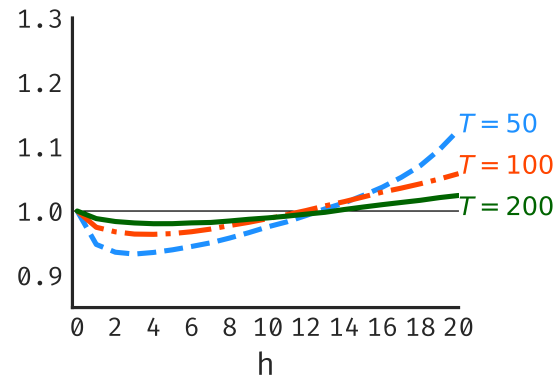
(b) $\rho = 0.90$, RMSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



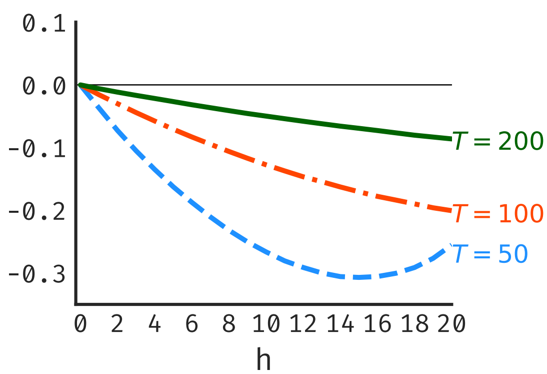
(c) $\rho = 0.95$, $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$



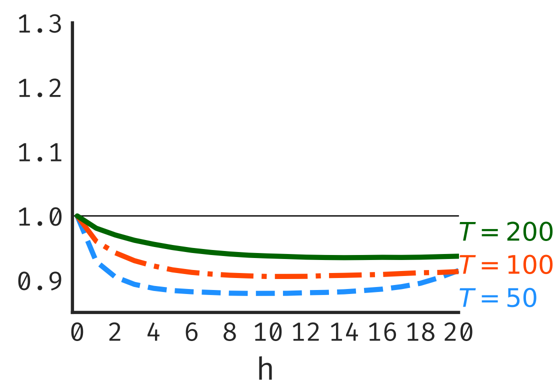
(d) $\rho = 0.95$, RMSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



(e) $\rho = 0.99$, $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$



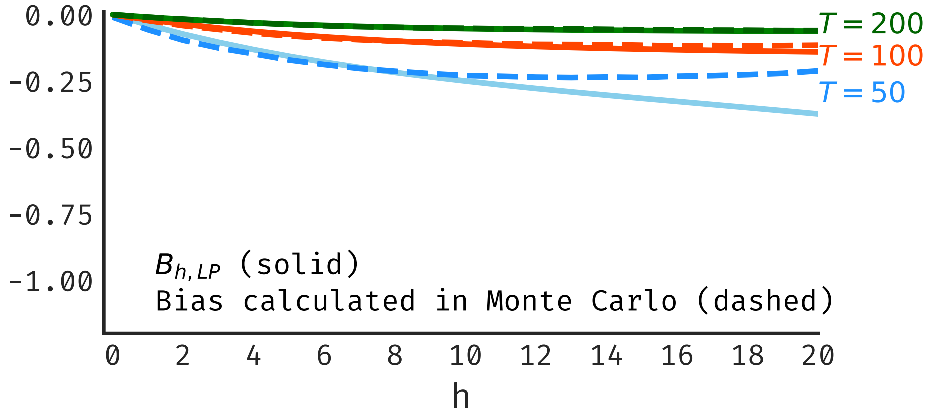
(f) $\rho = 0.99$, RMSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



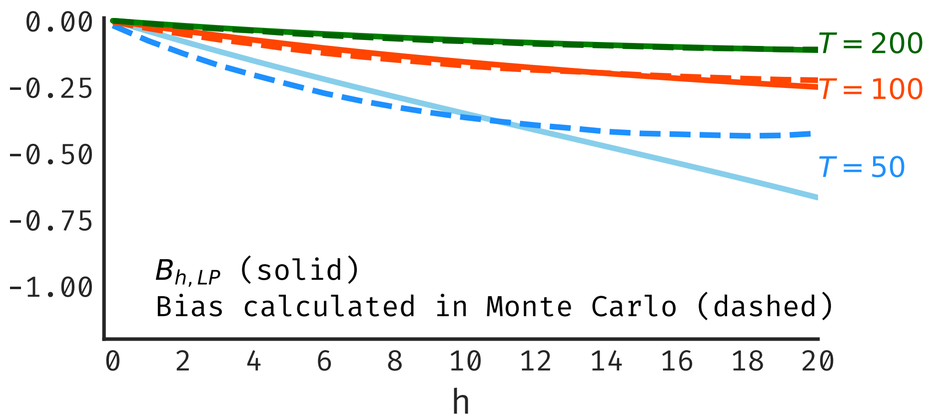
Note: This figure is analogous to Figure 3, but here we set $\sigma_\varepsilon = 10\sigma_\nu = 1$ and $\theta_0 = 1$. The sub-figures on the left show the Monte Carlo means of $\widehat{\theta}_{h,BCC}$ estimated on data simulated from equation (5). The sub-figures on the right show the Monte Carlo value of the RMSE of $\widehat{\theta}_{h,BCC}$ relative to the RMSE of $\widehat{\theta}_{h,LS}$. A number larger than unity indicates that the RMSE of $\widehat{\theta}_{h,BCC}$ is larger than $\widehat{\theta}_{h,LS}$. We use 100,000 Monte Carlo simulations.

Figure C.4: $B_{h,LP}$ performs well in empirically-relevant samples when y_t is an AR(1) and σ_ε is small.

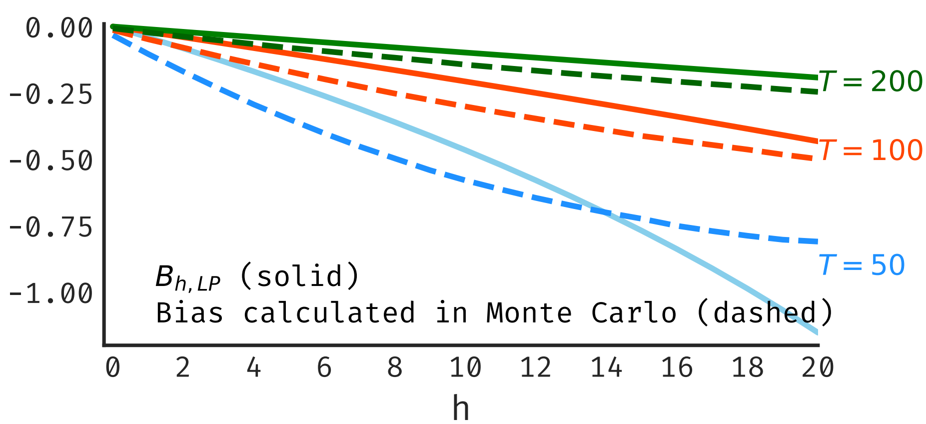
(a) $\rho = 0.90$



(b) $\rho = 0.95$

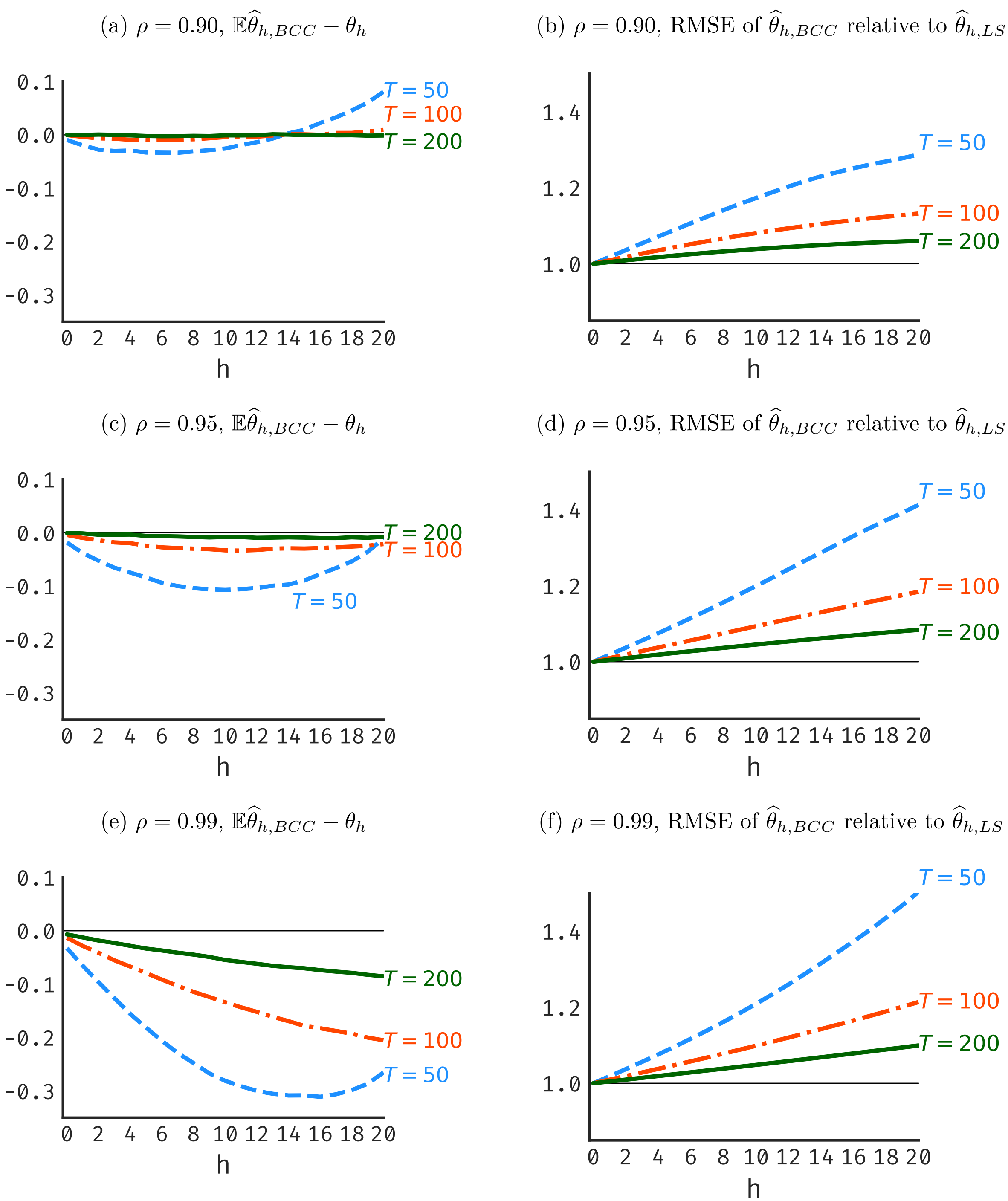


(c) $\rho = 0.99$



Note: This figure is analogous to Figure 2, but with $10\sigma_\varepsilon = \sigma_\nu = 1$. The sug-figures show the value of $B_{h,LP}$ and the Monte Carlo means of $\widehat{\theta}_{LS}$ estimated on data simulated from equation (5). We use 100,000 Monte Carlo simulations.

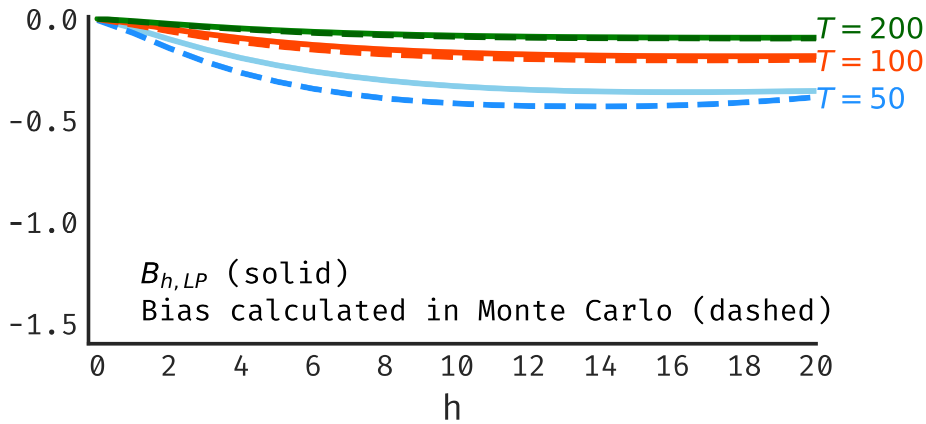
Figure C.5: Performance of $B_{h,BCC}$ in the AR(1) example when σ_ε is small.



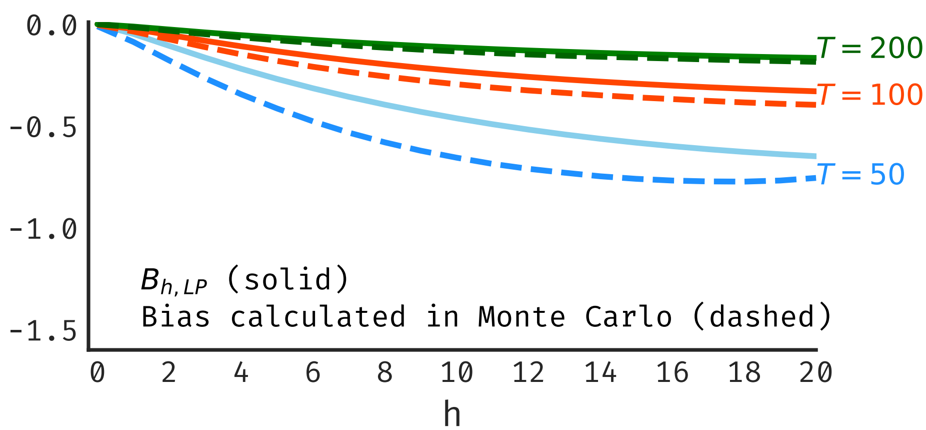
Note: This figure is analogous to Figure 3, but here we set $10\sigma_\varepsilon = \sigma_\nu = 1$ and $\theta_0 = 1$. The sub-figures on the left show the Monte Carlo means of $\widehat{\theta}_{h,BCC}$ estimated on data simulated from equation (5). The sub-figures on the right show the Monte Carlo value of the RMSE of $\widehat{\theta}_{h,BCC}$ relative to the RMSE of $\widehat{\theta}_{h,LS}$. A number larger than unity indicates that the RMSE of $\widehat{\theta}_{h,BCC}$ is larger than $\widehat{\theta}_{h,LS}$. We use 100,000 Monte Carlo simulations.

Figure C.6: $B_{h,LP}$ performs well in empirically-relevant samples when y_t is an AR(2).

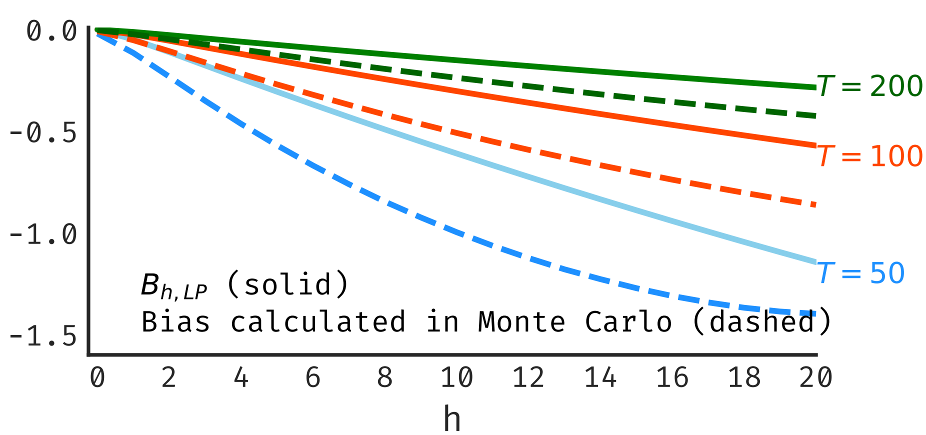
(a) $\rho = 0.90$



(b) $\rho = 0.95$



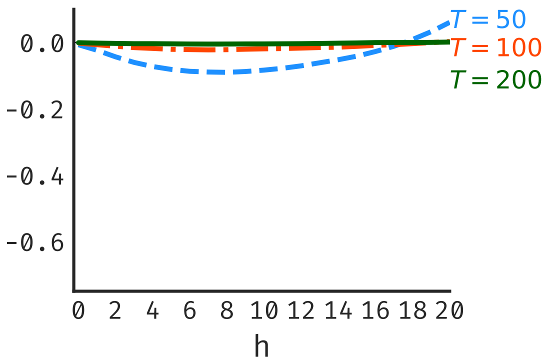
(c) $\rho = 0.99$



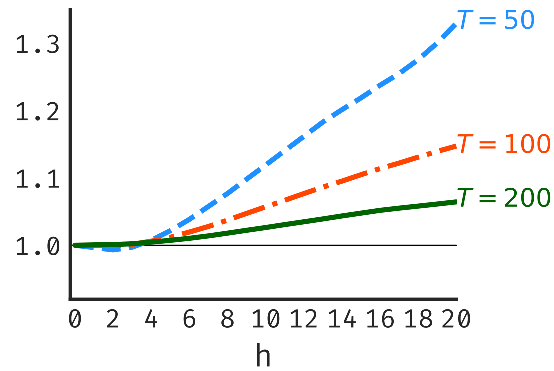
Note: the sub-figures show the value of $B_{h,LP}$ and the Monte Carlo means of $\widehat{\theta}_{LS}$ estimated on data simulated from equation (C.1). We use 100,000 Monte Carlo simulations. We set $\sigma_\varepsilon = \sigma_\nu = 1$ and $\theta_0 = 1$.

Figure C.7: Performance of $B_{h,BCC}$ in an AR(2) example.

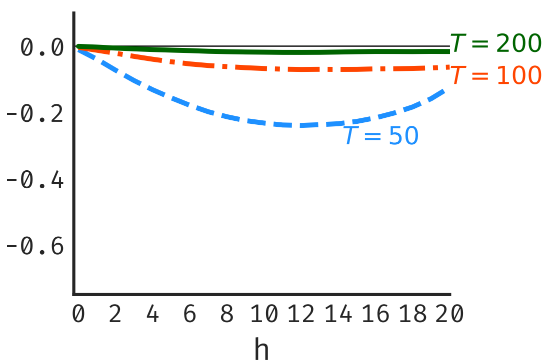
(a) $\rho = 0.90$, $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$



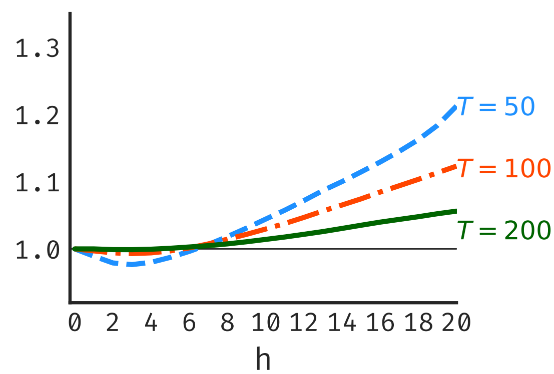
(b) $\rho = 0.90$, RMSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



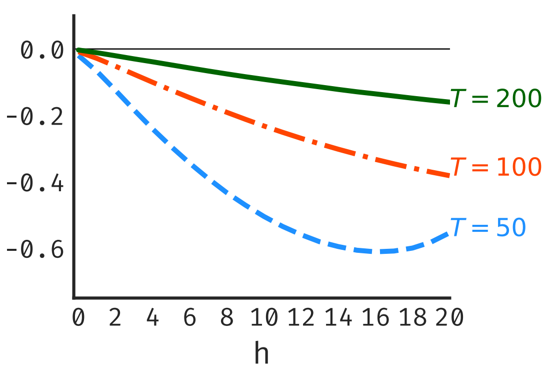
(c) $\rho = 0.95$, $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$



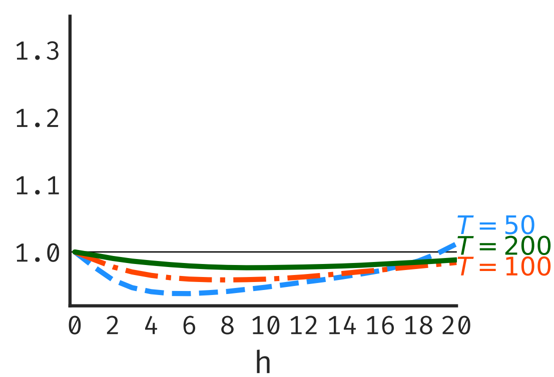
(d) $\rho = 0.95$, RMSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



(e) $\rho = 0.99$, $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$



(f) $\rho = 0.99$, RMSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



Note: the sub-figures on the left show the Monte Carlo means of $\widehat{\theta}_{h,BCC}$ estimated on data simulated from equation (5). The sub-figures on the right show the Monte Carlo value of the RMSE of $\widehat{\theta}_{h,BCC}$ relative to the RMSE of $\widehat{\theta}_{h,LS}$. A number larger than unity indicates that the RMSE of $\widehat{\theta}_{h,BCC}$ is larger than $\widehat{\theta}_{h,LS}$. We use 100,000 Monte Carlo simulations. We set $\sigma_\varepsilon = \sigma_\nu = 1$ and $\theta_0 = 1$.

D Derivations related to standard errors

Here, we provide details of the derivation of the bias in the GMM estimator of the autocovariance of the regression score under the assumption that $\theta_h = 0$ for all h. Our notation mirrors that of Appendix A.

D.1 Derivations related to SEs in LP with no controls

To derive the approximate bias in $\widehat{\gamma}_{h,\ell}$, in an LP with controls, we work with the following moment conditions

$$\mathbb{E}\begin{bmatrix} y_{t+h} - \alpha - x_t'\beta_h \\ \varepsilon_t\left(y_{t+h} - \alpha - x_t'\beta_h\right) \\ c_{t-1}\left(y_{t+h} - \alpha - x_t'\beta_h\right) \\ \varepsilon_t\left(y_{t+h} - \alpha - x_t'\beta_h\right)\varepsilon_{t-\ell}\left(y_{t-\ell+h} - \alpha - x_{t-\ell}'\beta_h\right) - \gamma_{h,\ell} \end{bmatrix} = \mathbf{0}, \tag{D.1}$$

where $\ell > 0$. We maintain Assumptions 1 and 2 and assume that Assumptions A-C in Rilstone et al. (1996) are satisfied. For notational ease, and without loss of generality, we assume that all data have zero mean.

With appropriately defined vectors and matrices analogous to those discussed in Appendix A that also are in the notation of Bao and Ullah (2007), the final element of $\mathbb{E}\left[Q\left(H_1 - \overline{H}_1\right)Q\psi_{h,T-h-\ell}\right]$ is given by

$$\left(\frac{1}{T-h-\ell}\right)^2 \sum_{t=1+\ell}^{T-h}\sum_{s=1+\ell}^{T-h} \mathbb{E}\left[\Upsilon_1(t,s) + \Upsilon_2(t,s) + \Upsilon_3(t,s) + \Upsilon_4(t,s)\right], \tag{D.2}$$

where

$$\Upsilon_1(t,s) = -\left[\varepsilon_{t-\ell}\varepsilon_t u_{h,t} + \varepsilon_t\varepsilon_{t-\ell}u_{h,t-\ell}\right]u_{s,t}, \tag{D.3}$$

$$\Upsilon_2(t,s) = -\left[\varepsilon_t\varepsilon_{t-\ell}^2 u_{h,t} + \varepsilon_t^2\varepsilon_{t-\ell}u_{h,t-\ell}\right]\frac{\varepsilon_s}{\sigma_\varepsilon^2}u_{h,s}, \tag{D.4}$$

$$\Upsilon_3(t,s) = -\left[c_{t-\ell-1}'\varepsilon_{t-\ell}\varepsilon_t u_{h,t} + \varepsilon_t c_{t-1}'\varepsilon_{t-\ell}u_{h,t-\ell}\right]\Sigma_c^{-1}c_{s-1}u_{h,s}, \tag{D.5}$$

$$\Upsilon_4(t,s) = -\varepsilon_s u_{s,t}\varepsilon_{s-\ell}u_{h,s-\ell} + \gamma_{h,u}. \tag{D.6}$$

It can be shown that for all t and s, $\mathbb{E}\Upsilon_1(t,s) = 0$, $\mathbb{E}\Upsilon_3(t,s) = 0$, $\mathbb{E}\Upsilon_4(t,s) = 0$. If $t = s$ or $t = s+\ell$, then

$$\mathbb{E}\Upsilon_2(t,s) = -\sigma_\varepsilon^2\mathbb{E}\left[\left(y_{t+h} - \alpha_h - x_t'\beta_h\right)^2\right]. \tag{D.7}$$

If $t \neq s$ and $t \neq s+\ell$, then $\mathbb{E}\Upsilon_2(t,s)=0$. So, the final element of $\mathbb{E}\left[Q\left(H_1-\overline{H}_1\right)Q\psi_{h,T-h-\ell}\right]$ is given by

$$-2\left(\frac{1}{T-h-\ell}\right)\sigma_\varepsilon^2\mathbb{E}\left[\left(y_{t+h}-\alpha_h-x_t'\beta_h\right)^2\right]+O\left(T^{-3/2}\right). \tag{D.8}$$

Additionally, the final element of $\mathbb{E}\left[\frac{1}{2}Q\overline{H}_2\left[a_{-1/2}\otimes a_{-1/2}\right]\right]$ is given by

$$\begin{aligned}&-\left(\frac{1}{T-h-\ell}\right)^2\sum_{t=1+\ell}^{T-h}\sum_{s=1+\ell}^{T-h}\mathbb{E}\left[\varepsilon_t\left(y_{t+h}-\alpha_h-x_t'\beta_h\right)\varepsilon_s\left(y_{s+h}-\alpha_h-x_s'\beta_h\right)\right]\\&=-\left(\frac{1}{T-h-\ell}\right)\sigma_\varepsilon^2\mathbb{E}\left[\left(y_{s+h}-\alpha_h-x_s'\beta_h\right)^2\right]\end{aligned} \tag{D.9}$$

Noting that $\gamma_{h,0}=\sigma_\varepsilon^2\mathbb{E}\left[(y_{s+h}-\alpha_h-x_s'\beta_h)^2\right]$, the approximate bias of $\widehat{\gamma}_{h,\ell}$ is

$$-\frac{1}{T-h-\ell}\gamma_{h,0}. \tag{D.10}$$

It is useful to parse the effect of estimating α_h. If the means of the data are known (α_h is known), the moment conditions are

$$\mathbb{E}\begin{bmatrix}\varepsilon_t\left(y_{t+h}-\alpha-x_t'\beta_h\right)\\ c_{t-1}\left(y_{t+h}-\alpha-x_t'\beta_h\right)\\ \varepsilon_t\left(y_{t+h}-\alpha-x_t'\beta_h\right)\varepsilon_{t-\ell}\left(y_{t-\ell+h}-\alpha-x_{t-\ell}'\beta_h\right)-\gamma_{h,\ell}\end{bmatrix}=\mathbf{0}, \tag{D.11}$$

where $\ell>0$. It can be shown thet the final element of $\mathbb{E}\left[Q\left(H_1-\overline{H}_1\right)Q\psi_{h,T-h-\ell}\right]$ is given by

$$\left(\frac{1}{T-h-\ell}\right)^2\sum_{t=1+\ell}^{T-h}\sum_{s=1+\ell}^{T-h}\mathbb{E}\left[\Upsilon_2(t,s)+\Upsilon_3(t,s)+\Upsilon_4(t,s)\right], \tag{D.12}$$

and that the final element of $\frac{1}{2}Q\overline{H}_2\left[a_{-1/2}\otimes a_{-1/2}\right]$ is unchanged from the case when controls are included. Because $\mathbb{E}\Upsilon_1(t,s)=0$ for all t and s, the approximate bias of $\widehat{\gamma}_{h,\ell}$ is the same as in the case when α_h is estimated.

D.2 Derivations related to SEs in LP with no controls

Here, we consider the approximate bias of $\widehat{\gamma}_{h,\ell}$ in an LP without controls. We maintain Assumption 1. The moment conditions are the same as in the LP with controls, but with $c_{t-1}=\emptyset$. It can be shown thet the final element of $\mathbb{E}\left[Q\left(H_1-\overline{H}_1\right)Q\psi_{h,T-h-\ell}\right]$ is given by

$$\left(\frac{1}{T-h-\ell}\right)^2\sum_{t=1+\ell}^{T-h}\sum_{s=1+\ell}^{T-h}\mathbb{E}\left[\Upsilon_1(t,s)+\Upsilon_2(t,s)+\Upsilon_4(t,s)\right], \tag{D.13}$$

and the final element of $\frac{1}{2}Q\overline{H}_2\left[a_{-1/2}\otimes a_{-1/2}\right]$ is unchanged in form from when controls are included. Because $\theta_h=0$, it is immediate that the approximate bias of $\widehat{\gamma}_{h,\ell}$ is

$$-\frac{1}{T-h-\ell}\sigma_\varepsilon^2\mathbb{E}\left[y_t^2\right]=-\frac{1}{T-h-\ell}\gamma_{h,0}. \tag{D.14}$$

Following similar arguments to those made in the case of controls, it is easily shown that the approximate bias of $\widehat{\gamma}_{h,\ell}$ is unchanged in the case without controls if α_h is known.

E CEE VAR

Christiano et al. (2005) estimate a 9 variable VAR(4) using quarterly U.S. data on real GDP, real consumption, real investment, GDP deflator prices, real wages, labor productivity, federal funds rate, real profits, and the growth rate of M2. All variables except the federal funds rate and M2 growth rate enter the VAR in log levels.

E.1 Data Construction

Following Christiano et al. (2005), our sample begins in 1965Q3 and ends in 1995Q3. Details on the variable construction are given below.

1. **Real GDP.** Take the level of real gross domestic product (FRED mnemonic `GDPC1`). The VAR observable is constructed as:

$$\text{Real GDP}_t=\log\left(GDPC1_t\right).$$

2. **Real Consumption.** Take the level of real consumption (FRED mnemonic `PCECC96`). The VAR observable is constructed as:

$$\text{Real Consumption}_t=\log\left(PCECC96_t\right).$$

3. **GDP Deflator Prices.** Take the level of GDP deflator prices (FRED mnemonic `GDPDEF`). The VAR observable is constructed as:

$$\text{Price Level}_t=\log\left(GDPDEF_t\right)).$$

4. **Real Investent.** Take the level of real gross domestic private investment (FRED mnemonic `GPDIC1`). The VAR observable is constructed as:

$$\text{Real Investment}_t=\log\left(GPDIC1_t\right).$$

5. **Real Wages.** Take hourly compensation for all employmed persons (FRED mnemonic `COMPNFB`), the level of the GDP deflator prices, and the average weekly hours worked (FRED mnemonic `PRS85006023`). The VAR observable is constructed as:

$$\text{Real Wages}_t = \log\left(\frac{COMPNFB_t}{GDPDEF_t \times PRS85006023_t}\right).$$

6. **Labor Productivity.** Take labor productivity (FRED mnemonic `OPHNFB`). The VAR observable is constructed as:

$$\text{Labor Productivity}_t = \log\left(OPHNFB_t\right).$$

7. **Federal Funds Rate.** Take the quarterly average of the monthly federal funds rate (FRED mnemonic `FEDFUNDS`). The VAR observable is constructed as:

$$\text{Federal Funds Rate}_t = FEDFUNDS_t.$$

8. **Real Profits.** Take corporate profits after tax (FRED mnemonic `CP`). The VAR observable is constructed as:

$$\text{Real Profits}_t = \log\left(\frac{CP_t}{GDPDEF_t}\right).$$

9. **The growth rate of M2.** Take the M2 monetary aggregate (FRED mnemonic `M2`). The VAR observable is constructed as:

$$\Delta M2_t = 100 \times \left(\frac{M2_t - M2_{t-1}}{M2_{t-1}}\right).$$

E.2 Identification of the Monetary Policy Shock in the VAR

The structural shocks underpinning the VAR are identified using the Cholesky factorization of the estimated covariance matrix of the one step ahead forecast errors. The observables enter the VAR in the following order: real GDP, real consumption, real investment, GDP deflator prices, real wages, labor productivity, federal funds rate, real profits, and the growth rate of M2. The monetary policy shock is assumed to be the one associated with the federal funds "equation" (i.e., the 6th one). The Cholesky identification scheme can thus be interpreted as a set of timing assumptions. The monetary policy shock cannot contemporaneously affect GDP, consumption, investment, the price level, wages, or labor productivity.

E.3 Local Projections and VARs, Redux

Here we put finite sample issues in LPs in the context a VAR-type estimator. Unless, the estimated VAR coincides with the data generating process, there will be an *asymptotic* bias and variance trade-off between the LP and VAR. In this section, we gauge the extent of this issue in the context of finite sample approximation for our data generating process. We compare the impulse response estimator $\widehat{\theta}_{BCC}$ as well as the estimators constructed by iterating the $h = 0$ impact estimate using either an estimated VAR(1) or estimated VAR(4). We denote these estimates by $\widehat{\theta}_{VAR(1)}$ and $\widehat{\theta}_{VAR(4)}$, respectively. The VAR(4) represents the idealized case where the specification coincides with the data generating process. The VAR(1) stands in as a misspecified alternative that might be attractive to an investigator for it's parsimony.

The left column of figure E.1 displays Monte Carlo estimates of $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid lines), $\mathbb{E}\widehat{\theta}_{h,VAR(1)} - \theta_h$ (dotted lines), and $\mathbb{E}\widehat{\theta}_{h,VAR(4)} - \theta_h$ (dash-dotted lines) for for output, the price level, and the federal funds (rows) for $T = 100$. The VAR(4), the dash-dotted lines, being correctly specified, exhibits the least amount of bias in general. Note that while $\widehat{\theta}_{LS}$ and the VAR(4) estimate are both unbiased asymptotically in this setting, the finite sample bias associated with the (even bias-corrected) LP estimator is substantially larger. While this is not a general statement–our theoretical results confirm that this difference cannot be signed in even a simple AR(1) setting–it is instructive that in empirically realistic settings this difference is sizable, as the LP bias and VAR bias are generated by the statisical considerations. Thus, the common practice of ignoring issues of finite sample bias in VARs maybe not be appropriate for LPs. The VAR(1), being misspecified, exhibits bias coming from both finite sample and asymptotic considerations. For real GDP and the price level, the bias associated with the VAR(1) is substantial, sometimes much worse than even the non-corrected LP (not shown), as their responses display more complex dynamics.

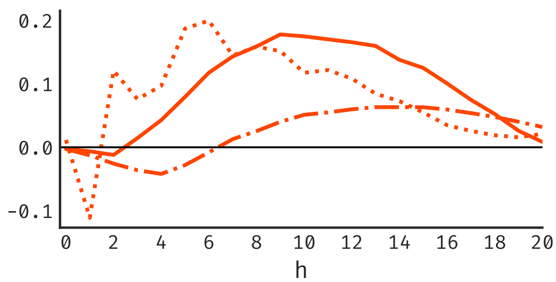
The right column of Figure E.1 displays the MSE ratios of $\hat{\theta}_{BCC}$ and $\widehat{\theta}_{VAR(1)}$ relative to the $\widehat{\theta}_{VAR(4)}$ for $T = 100$. The VAR(1) exhibits lower MSE than $\hat{\theta}_{BCC}$ for essentially all variables and horizons—indeed it is typically preferred to the (correctly specified) VAR(4) estimator owing to finite sample considerations. One way to interpret this result in the context of increasing popularity of LPs is that investigators care much more about bias than variance of estimators–otherwise we would see much use of misspecified VARs.

E.4 Additional Results

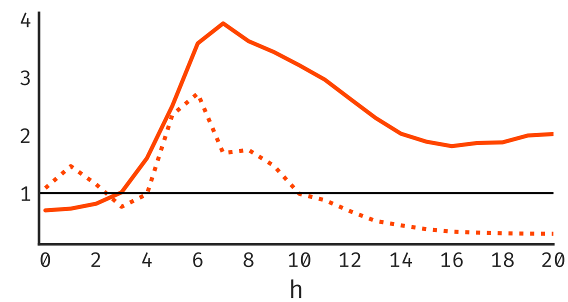
Figure E.1: Bias and MSE of $\widehat{\theta}_{BCC}$ compared to VAR-based estimates under a CEE-type VAR Data Generating Process

Real GDP

(a) $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid), $\mathbb{E}\widehat{\theta}_{h,VAR(1)} - \theta_h$ (dotted) and $\mathbb{E}\widehat{\theta}_{h,VAR(4)} - \theta_h$ (dash-dotted)

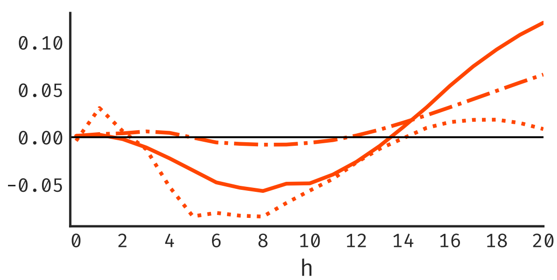


(b) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,VAR(4)}$ (solid) and $\widehat{\theta}_{h,VAR(1)}$ relative to $\widehat{\theta}_{h,VAR(4)}$ (dotted)

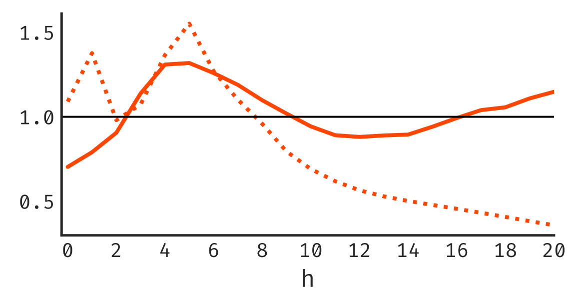


Price Level

(c) $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid), $\mathbb{E}\widehat{\theta}_{h,VAR(1)} - \theta_h$ (dotted) and $\mathbb{E}\widehat{\theta}_{h,VAR(4)} - \theta_h$ (dash-dotted)

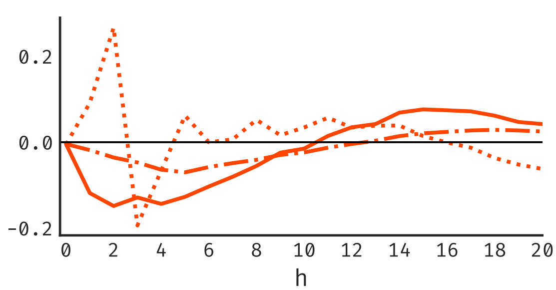


(d) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,VAR(4)}$ (solid) and $\widehat{\theta}_{h,VAR(1)}$ relative to $\widehat{\theta}_{h,VAR(4)}$ (dotted)

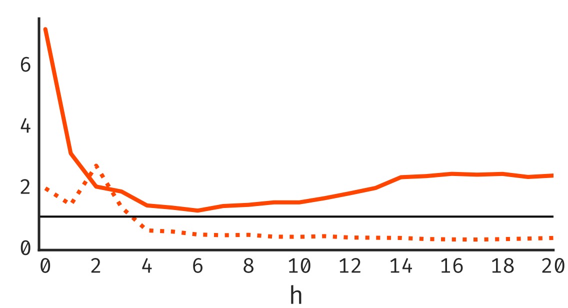


Federal Funds Rate

(e) $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid), $\mathbb{E}\widehat{\theta}_{h,VAR(1)} - \theta_h$ (dotted) and $\mathbb{E}\widehat{\theta}_{h,VAR(4)} - \theta_h$ (dash-dotted)



(f) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,VAR(4)}$ (solid) and $\widehat{\theta}_{h,VAR(1)}$ relative to $\widehat{\theta}_{h,VAR(4)}$ (dotted)

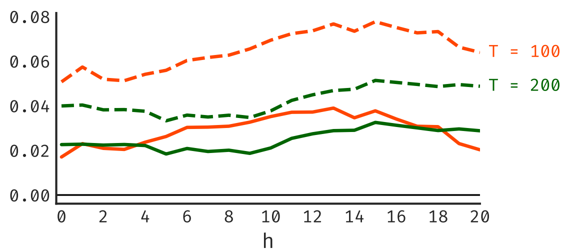


The left column of figure shows the bias of $\mathbb{E}\widehat{\theta}_{h,BCC}$ (solid lines), $\mathbb{E}\widehat{\theta}_{h,VAR(1)}$ (dotted lines) and $\mathbb{E}\widehat{\theta}_{h,VAR(4)}$ (dash-dotted lines) for $T = 100$. The right column shows the ratio of the mean squared error of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,VAR(4)}$ (solid lines) and for $\widehat{\theta}_{h,VAR(1)}$ relative to $\widehat{\theta}_{h,VAR(4)}$ (dotted lines) for $T = 100$.

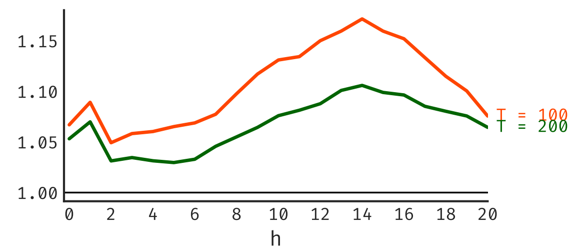
Figure E.2: Bias and MSE under a CEE-type VAR DGP (No controls)

Output

(a) $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid)

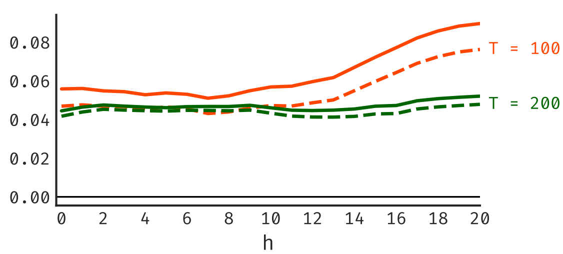


(b) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$

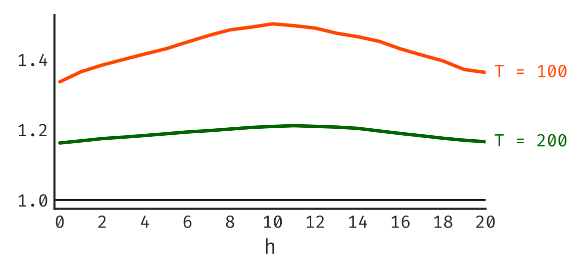


Price Level

(c) $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid)

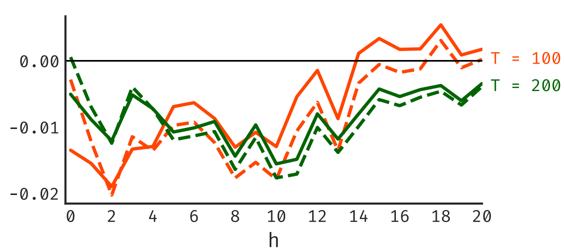


(d) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$

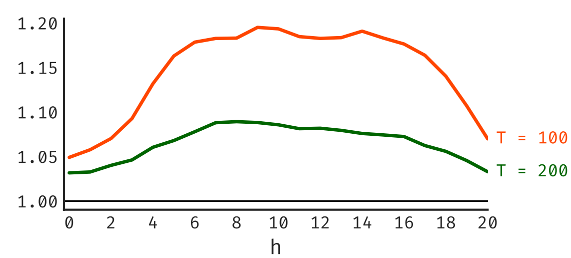


Federal Funds Rate

(e) $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid)



(f) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$

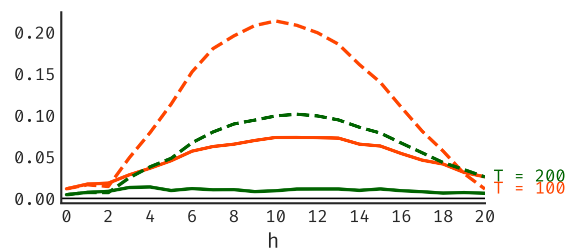


The sub-figures in the left column show the $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed lines) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid lines) for $T = 100$ (red) and $T = 200$ (green). The sub-figures in the right column show the ratio of the MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$ for $T = 100$ (red) and $T = 200$ (green).

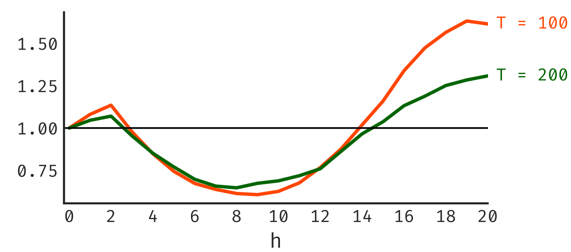
Figure E.3: Bias and MSE under a CEE-type VAR DGP (Partial Controls)

Output

(a) $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid)

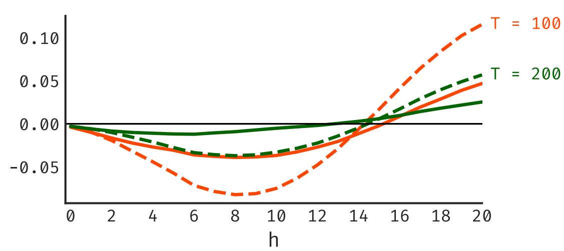


(b) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$

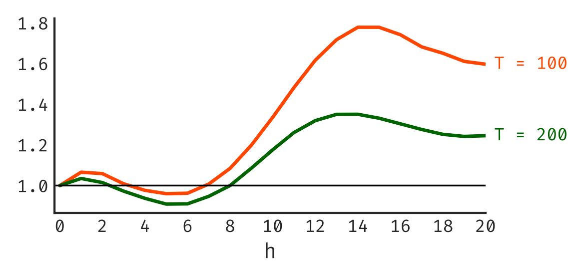


Price Level

(c) $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid)

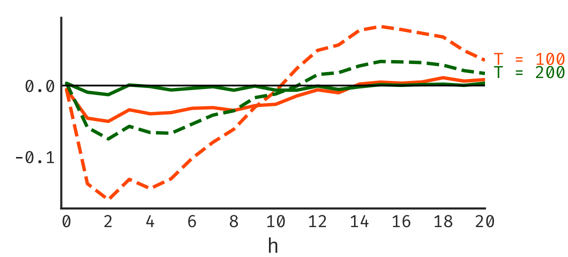


(d) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$

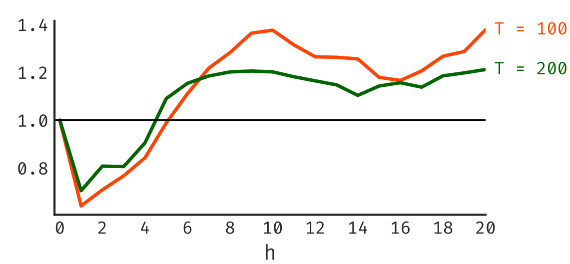


Federal Funds Rate

(e) $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid)



(f) MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$



The sub-figures in the left column show the $\mathbb{E}\widehat{\theta}_{h,LS} - \theta_h$ (dashed lines) and $\mathbb{E}\widehat{\theta}_{h,BCC} - \theta_h$ (solid lines) for $T = 100$ (red) and $T = 200$ (green). The sub-figures in the right column show the ratio of the MSE of $\widehat{\theta}_{h,BCC}$ relative to $\widehat{\theta}_{h,LS}$ for $T = 100$ (red) and $T = 200$ (green).

F 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 (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.