Local Projection Inference is Simpler and More Robust Than You Think*

PRELIMINARY AND INCOMPLETE

José Luis Montiel Olea
Columbia University

Mikkel Plagborg-Møller
Princeton University

March 17, 2020

Abstract: A popular method for conducting inference on impulse responses in applied macroeconomics is to compute confidence intervals by local projections, i.e., direct linear regressions of future outcomes on current covariates. This paper proves that local projection inference robustly handles two issues that commonly arise in applications: highly persistent data and the estimation of impulse responses at long horizons. We consider local projections that control for lags of the data. We show that lag-augmented local projections with normal critical values are asymptotically valid uniformly over i) both stationary and non-stationary data, and also over ii) a wide range of impulse response horizons. Moreover, and contrary to conventional wisdom, we show that lag augmentation obviates the need to correct the standard errors for serial correlation in the regression residuals. Hence, local projection inference is arguably both simpler than previously thought and more robust than standard autoregressive impulse response inference, whose validity is known to depend sensitively on the persistence of the data and on the length of the horizon.

Keywords: impulse response, local projection, long horizon, uniform inference.

*Email: `jm4474@columbia.edu`, `mikkelpm@princeton.edu`. We are grateful for comments from Jushan Bai, Marco Del Negro, Domenico Giannone, Michal Kolesár, Simon Lee, Konrad Menzel, Ulrich Müller, Serena Ng, Christian Wolf, and seminar participants at the Atlanta Fed, Columbia, NYU, Princeton, and the 2019 Greater New York Area Econometrics Colloquium. Montiel Olea would like to thank Giovanni Topa for excellent research assistance. Plagborg-Møller acknowledges that this material is based upon work supported by the NSF under Grant #1851665. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

1 Introduction

Impulse response functions are key objects of interest in empirical macroeconomic analysis. It is increasingly popular to estimate these parameters using the method of *local projections* (Jordà, 2005): simple linear regressions of a future outcome on current covariates (Ramey, 2016; Angrist et al., 2018; Nakamura and Steinsson, 2018; Stock and Watson, 2018). Since local projection estimators are regression coefficients, they have a simple and intuitive interpretation. Moreover, inference can be carried out using textbook standard error formulae, adjusting for serial correlation in the (multi-step forecast) regression residuals.

Despite its popularity, there exist no theoretical results justifying the use of local projection *inference* over autoregressive procedures. From an identification and estimation standpoint, Plagborg-Møller and Wolf (2019) argue that neither local projections nor Vector Autoregressions (VARs) dominate the other in terms of mean squared error, and in population the two methods are equivalent. However, from an inference perspective, the only available guidance on the relative performance of local projections comes in the form of a small number of simulation studies, which by necessity cannot cover the entire range of empirically relevant data generating processes.

In this paper we show that—in addition to its intuitive appeal—local projection inference is robust to two common features of macroeconomic applications: highly persistent data and the estimation of impulse responses at long horizons. Key to our result is that we consider *lag-augmented* local projections, which use lags of the data as controls. Formally, we prove that standard confidence intervals based on such lag-augmented local projections have correct asymptotic coverage *uniformly* over the persistence in the data generating process and over a wide range of horizons.[1] This means that confidence intervals remain valid even if the data exhibits unit roots, and even at horizons h that are allowed to grow with the sample size T, e.g., $h = h_T \propto \sqrt{T}$. In fact, when persistence is not an issue, and the data is known to be stationary, local projection inference is also valid at *long* horizons; i.e., horizons that are a non-negligible fraction of the sample size ($h_T \propto T$).

Lag-augmenting local projections not only robustifies inference, it also simplifies the computation of standard errors by obviating the adjustment for serial correlation in the residuals. It is common practice in the local projections literature to compute Heteroskedasticity and Autocorrelation Consistent/Robust (HAC/HAR) standard errors (Jordà, 2005;

[1]We focus on marginal inference on individual impulse responses, not *simultaneous* inference on a vector of several response horizons (see references in Montiel Olea and Plagborg-Møller, 2019).

Ramey, 2016; Kilian and Lütkepohl, 2017; Stock and Watson, 2018). Instead, we prove that the usual Eicker-Huber-White heteroskedasticity-robust standard errors suffice for *lag-augmented* local projections. The reason is that, although the regression residuals are serially correlated, the *regression scores* (the product of the residuals and residualized regressor of interest) are serially uncorrelated under weak assumptions. This finding further simplifies local projection inference, as it side-steps the delicate choice of HAR procedure and associated difficult-to-interpret tuning parameters (e.g., Lazarus et al., 2018).

The robustness properties of lag-augmented local projection stand in stark contrast to the well-known fragility of standard autoregressive inference procedures. Textbook autoregressive inference methods are invalid in the case of near-unit roots or long horizons. We show that lag-augmented local projection inference can overcome these issues, at the cost of being less efficient than textbook autoregressive inference methods when the data is stationary and interest centers on short horizons. Thus, the robustness afforded by our recommended inference procedure is not a free lunch. Another apparent drawback of lag-augmented local projection inference is that it does fail when the data has near-unit roots *and* interest centers on very long horizons $h = h_T \propto T$ (although it is valid under any increasing horizon sequence satisfying $h_T/T \to 0$). Existing VAR-based methods that achieve correct coverage in this case are either highly computationally demanding or result in impractically wide confidence intervals. We provide a detailed comparison with alternative inference procedures in Section 3 below.

Our results rely on assumptions that are similar to those used in the literature on autoregressive inference. In particular, we assume that the true model is a VAR(p) with possibly heteroskedastic innovations and known lag length. The key assumption that we require on the innovations is that they are conditionally mean independent of both past and *future* innovations. This strengthening of the usual martingale difference assumption is crucial to avoid HAC inference, but we show that the assumption is satisfied for a large class of heteroskedastic innovation processes. It is sometimes argued that an advantage of local projections is that this procedure is "robust to misspecification" of the VAR model, but Plagborg-Møller and Wolf (2019) argue that this view is misguided. Hence, in this paper we do not consider the consequences of misspecification.

To illustrate our theoretical results, we present a small-scale simulation study suggesting that lag-augmented local projection confidence intervals achieve a favorable trade-off between coverage and length. Since local projection estimation is subject to small-sample biases just like VAR estimation (Herbst and Johannsen, 2020), we consider a simple and computationally convenient bootstrap implementation of local projection. The simulations suggest that textbook autoregressive procedures have more severe under-coverage problems than local projection inference, especially at moderate and long horizons. Autoregressive confidence intervals can be meaningfully *shorter* than lag-augmented local projection intervals in *relative* terms, but in *absolute* terms the difference in length is surprisingly modest. Our simulations also indicate that lag-augmented local projections with heteroskedasticity-robust standard errors have better coverage/length properties than more standard *non-augmented* local projections with off-the-shelf HAR standard errors. Finally, although the lag-augmented autoregressive bootstrap procedure of Inoue and Kilian (2020) achieves good coverage, it yields prohibitively wide confidence intervals at longer horizons when the data is persistent.

Related literature. It is known that textbook autoregressive (AR) inference methods are neither robust to the persistence of the data nor the length of the impulse response horizon (Phillips, 1998; Benkwitz et al., 2000; Pesavento and Rossi, 2007; Mikusheva, 2012; Inoue and Kilian, 2020). We discuss these well-known issues in detail in Section 3. The practical consequence is that applied researchers cannot rely on AR-based confidence intervals with normal critical values to deliver correct coverage without further analysis. Standard bootstrap methods do not rectify all these issues.

Though we appear to be the first to prove the *uniform* validity of lag-augmented local projection (LP) inference, our paper is inspired by the literature that uses lag augmentation to robustify *autoregressive* inference against the presence of unit roots (Toda and Yamamoto, 1995; Dolado and Lütkepohl, 1996; Inoue and Kilian, 2020). Mikusheva (2007, 2012) and Inoue and Kilian (2020) derive the uniform coverage properties of various autoregressive inference procedures, but they do not consider local projections. The *pointwise* econometric properties of local projection procedures have been discussed by Jordà (2005), Kilian and Lütkepohl (2017), and Stock and Watson (2018), among others. Kilian and Kim (2011) and Brugnolini (2018) present simulation studies comparing AR inference and local projection inference, reaching conflicting conclusions. Brugnolini (2018) finds that the lag length in the local projection matters, which is consistent with our theoretical results.

Several papers have proposed AR-based methods for impulse response inference at *long* horizons $h = h_T \propto T$ (Wright, 2000; Gospodinov, 2004; Pesavento and Rossi, 2007; Mikusheva, 2012; Inoue and Kilian, 2020). With the exception of Mikusheva (2012), this literature has exclusively focused on near-unit root processes as opposed to devising uniformly valid procedures. The Hansen (1999) grid bootstrap analyzed by Mikusheva (2012) is asymptotically valid at short and long horizons. However, it is not valid at intermediate horizons (e.g., $h_T \propto \sqrt{T}$), unlike the LP procedure we analyze. Mikusheva argues, though, that the grid bootstrap is *close* to being valid at intermediate horizons, although it is much more computationally demanding than our recommended procedure, especially in VAR models with several parameters. Inoue and Kilian (2020) show that a version of the Efron bootstrap confidence interval, when applied to lag-augmented AR estimators, is valid at long horizons. We show in Section 3 that, in the context of the AR(1) model, this procedure delivers impractically wide confidence intervals (essentially, the entire positive part of the parameter space) at moderately long horizons when the data is persistent, unlike lag-augmented LP.

Though the theoretical results in this paper appear to be novel, Dufour et al. (2006, Section 5) and Breitung and Brüggemann (2019) have discussed some of the main ideas presented herein. First, both these papers state that lag augmentation in local projections avoids unit root asymptotics, but neither paper considers inference at long horizons or derives uniform inference properties. Second, Breitung and Brüggemann (2019) further argue that HAC inference in local projections can be avoided if the true model is a VAR(p), although it is not clear from their discussion what are the assumptions needed for this to be true. Neither of these papers provide results concerning the efficiency of lag-augmented LP inference relative to other lag-augmented or non-augmented inference procedures, as we do in Section 3.

Local projections are closely related to multi-step forecasts. Richardson and Stock (1989) and Valkanov (2003) develop a non-standard limit distribution theory for long-horizon forecasts. Chevillon (2017) proves a robustness property of direct multi-step inference that involves non-normal asymptotics due to the lack of lag augmentation. Phillips and Lee (2013) develop a novel multi-step inference approach with uniformly normal asymptotics, based on regressions with artificially constructed instruments. Unlike this new method, the local projection procedure we consider does not rely on tuning parameters.

Outline. Section 2 provides a non-technical overview of our results in the context of a simple AR(1) model, including an illustrative simulation study. Section 3 provides an in-depth comparison of lag-augmented LP with other inference procedures. Section 4 states the formal uniformity result for the AR(1) case. Section 5 discusses the general case of a VAR(p) model. Section 6 describes a simple bootstrap implementation of lag-augmented local projection that we recommend for practical use. Section 7 concludes. Further simulation results and theory are relegated to Appendix A, while proofs and auxiliary lemmas can be found in Appendix B.

2 Overview of the Results

This section provides an overview of our results and assumptions in the context of a simple univariate AR(1) model. The discussion here merely intends to illustrate our main points. We show in Section 5 that our recommended procedure can be easily implemented in empirically relevant VAR(p) models.

2.1 Lag-Augmented Local Projection

Model. Consider the AR(1) model for the data $\{y_t\}$:

$$y_t = \rho y_{t-1} + u_t, \quad t = 1, 2, \dots, T, \quad y_0 = 0. \tag{1}$$

The parameter of interest is a nonlinear transformation of ρ, namely the impulse response coefficient at horizon $h \in \mathbb{N}$. We denote this parameter by $\beta(\rho, h) \equiv \rho^h$. Our main assumption in the univariate model is:

Assumption 1. $\{u_t\}$ *is strictly stationary and satisfies* $E(u_t \mid \{u_s\}_{s \neq t}) = 0$ *almost surely.*

The assumption requires the residuals to be mean independent relative to past and future information. This is a slight strengthening of the usual martingale difference assumption on u_t. Assumption 1 is trivially satisfied if $\{u_t\}$ is i.i.d., but it also allows for stochastic volatility and GARCH-type innovation processes.[2] The assumption is in principle testable, but we leave this possibility for future research.

Local projections with and without lag augmentation. We consider the local projection (LP) approach of Jordà (2005) for conducting inference about the impulse response $\beta(\rho, h)$. A common motivation for this approach is that the AR(1) model (1) implies

$$y_{t+h} = \beta(\rho, h) y_t + \xi_t(\rho, h), \tag{2}$$

[2] For example, consider processes $u_t = \tau_t \varepsilon_t$, where ε_t is i.i.d. with $E(\varepsilon_t) = 0$, and for which one of the following two sets of conditions hold: (a) $\{\tau_t\}$ and $\{\varepsilon_t\}$ are independent processes; or (b) τ_t is a function of lagged values of ε_t^2, and the distribution of ε_t is symmetric.

where the regression residual (or *multi-step forecast error*),

$$\xi_t(\rho, h) \equiv \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell},$$

is generally serially correlated, even if the innovation u_t is i.i.d.

The most straight-forward LP impulse response estimator simply regresses y_{t+h} on y_t, as suggested by equation (2), but the validity of this approach is sensitive to the persistence of the data. Specifically, this standard approach leads to a non-normal limiting distribution for the impulse response estimator when $\rho \approx 1$, since the regressor y_t exhibits near-unit-root behavior in this case. Hence, inference based on normal critical values will not be valid uniformly over all values of $\rho \in [-1, 1]$ even for fixed forecast horizons h. If ρ is safely within the stationary region, then the LP estimator is asymptotically normal, but inference generally requires the use of Heteroskedasticity and Autocorrelation Robust (HAR) standard errors to account for serial correlation in the residual $\xi_t(\rho, h)$.

To robustify and simplify inference, we will instead consider a *lag-augmented* local projection, which uses y_{t-1} as an additional control variable. In the autoregressive literature, "lag augmentation" refers to the practice of using more lags for estimation than suggested by the true autoregressive model. Define the covariate vector $x_t \equiv (y_t, y_{t-1})'$. Given any horizon $h \in \mathbb{N}$, the lag-augmented LP estimator $\hat{\beta}(h)$ of $\beta(\rho, h)$ is given by the coefficient on y_t in a regression of y_{t+h} on y_t and y_{t-1}:

$$\begin{pmatrix} \hat{\beta}(h) \\ \hat{\gamma}(h) \end{pmatrix} \equiv \left(\sum_{t=1}^{T-h} x_t x_t' \right)^{-1} \sum_{t=1}^{T-h} x_t y_{t+h}. \tag{3}$$

Here $\hat{\beta}(h)$ is the impulse response estimator of interest, while $\hat{\gamma}(h)$ is a nuisance coefficient.

The purpose of the lag augmentation is to make the effective regressor of interest stationary even when the data y_t has a unit root. Note that equations (1)–(2) imply

$$y_{t+h} = \beta(\rho, h) u_t + \beta(\rho, h+1) y_{t-1} + \xi_t(\rho, h). \tag{4}$$

If u_t were observed, the above equation suggests regressing y_{t+h} on u_t, while controlling for y_{t-1}. Intuitively, this will lead to an asymptotically normal estimator of $\beta(\rho, h)$, since the regressor of interest u_t is stationary by Assumption 1, and we control for the term that involves the possibly non-stationary regressor y_{t-1}. Fortunately, due to the linear relationship $y_t = \rho y_{t-1} + u_t$, the coefficient $\hat{\beta}(h)$ on y_t in the feasible lag-augmented regression (3) on (y_t, y_{t-1}) precisely equals the coefficient on u_t in the desired regression on (u_t, y_{t-1}). This argument for why lag-augmented LP can be expected to have a uniformly normal limit distribution even when $\rho \approx 1$ is completely analogous to the reasoning for using lag augmentation in AR inference (Sims et al., 1990; Toda and Yamamoto, 1995; Dolado and Lütkepohl, 1996; Inoue and Kilian, 2020). In the case of LP inference, lag augmentation has the additional benefit of simplifying the computation of standard errors, as we now discuss.

Standard errors. We now define the standard errors for the lag-augmented LP estimator. We will show that, contrary to conventional wisdom (e.g., Jordà, 2005, p. 166; Ramey, 2016, p. 84), HAR standard errors are *not* needed to conduct inference on *lag-augmented* LP, despite the fact that the regression residual $\xi_t(\rho, h)$ is serially correlated. Instead, it suffices to use the usual heteroskedasticity-robust Eicker-Huber-White standard error of $\hat{\beta}(h)$:[3]

$$\hat{s}(h) \equiv \frac{(\sum_{t=1}^{T-h} \hat{\xi}_t(h)^2 \hat{u}_t(h)^2)^{1/2}}{\sum_{t=1}^{T-h} \hat{u}_t(h)^2}, \tag{5}$$

where we define the lag-augmented LP residuals

$$\hat{\xi}_t(h) \equiv y_{t+h} - \hat{\beta}(h) y_t - \hat{\gamma}(h) y_{t-1}, \quad t = 1, 2, \ldots, T-h, \tag{6}$$

and the residualized regressor of interest

$$\hat{u}_t(h) \equiv y_t - \hat{\rho}(h) y_{t-1}, \quad t = 1, 2, \ldots, T-h,$$

$$\hat{\rho}(h) \equiv \frac{\sum_{t=1}^{T-h} y_t y_{t-1}}{\sum_{t=1}^{T-h} y_{t-1}^2}.$$

As mentioned in the introduction, the fact that we may avoid HAR inference simplifies the implementation of LP inference, as there is no need to choose amongst alternative HAR procedures or specify tuning parameters such as bandwidths (Lazarus et al., 2018).

Why is it not necessary to adjust for serial correlation in the residuals? Since lag-augmented LP controls for y_{t-1}, equation (4) suggests that the estimator $\hat{\beta}(h)$ is asymptotically equivalent with the coefficient in a linear regression of the (population) residualized

[3] This is computed by the `regress, robust` command in Stata, for example. The usual homoskedastic standard error formula suffices if u_t is assumed to be i.i.d.

outcome $y_{t+h} - \beta(\rho, h+1)y_{t-1}$ on the (population) residualized regressor $u_t = y_t - \rho y_{t-1}$:

$$\hat{\beta}(h) \approx \frac{\sum_{t=1}^{T-h}\{y_{t+h} - \beta(\rho, h+1)y_{t-1}\}u_t}{\sum_{t=1}^{T-h} u_t^2}$$
$$= \beta(\rho, h) + \frac{\sum_{t=1}^{T-h} \xi_t(\rho, h)u_t}{\sum_{t=1}^{T-h} u_t^2}.$$

The second term in the decomposition above determines the sampling distribution of the lag-augmented local projection. Although the multi-step regression residual $\xi_t(\rho, h)$ is serially correlated on its own, the *regression score* $\xi_t(\rho, h)u_t$ is serially uncorrelated under Assumption 1.[4] For any $s < t$,

$$\begin{aligned} E[\xi_t(\rho, h)u_t\xi_s(\rho, h)u_s] &= E[E(\xi_t(\rho, h)u_t\xi_s(\rho, h)u_s \mid u_{s+1}, u_{s+2}, \dots)] \\ &= E[\xi_t(\rho, h)u_t\xi_s(\rho, h) \underbrace{E(u_s \mid u_{s+1}, u_{s+2}, \dots)}_{=0}]. \end{aligned} \tag{7}$$

Thus, the heteroskedasticity-robust (but not autocorrelation-robust) standard error $\hat{s}(h)$ suffices for doing inference on $\hat{\beta}(h)$.[5] Notice that this result crucially relies on (i) lag-augmenting the local projections and (ii) the strengthening in Assumption 1 of the usual martingale difference assumption on $\{u_t\}$ (as remarked above, the strengthening still allows for conditional heteroskedasticity and other plausible features of economic shocks).[6]

Lag-augmented Local Projection Inference. Define the nominal $100(1-\alpha)\%$ lag-augmented LP confidence interval for the impulse response at horizon h based on the standard error $\hat{s}(h)$:

$$\hat{C}(h, \alpha) \equiv \left[\hat{\beta}(h) - z_{1-\alpha/2}\, \hat{s}(h)\, ,\, \hat{\beta}(h) + z_{1-\alpha/2}\, \hat{s}(h)\right],$$

where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution.

Our main result shows that the lag-augmented LP confidence interval above is valid regardless of the persistence of the data, i.e., whether or not the data has a unit root.

[4] Breitung and Brüggemann (2019) make this same observation, but they appear to claim that it is sufficient to assume that $\{u_t\}$ is white noise, which is incorrect.

[5] Stock and Watson (2018, p. 152) mention a similar conclusion for the distinct case of LP with an instrumental variable, under some conditions on the instrument.

[6] The nuisance coefficient $\hat{\gamma}(h)$ is not interesting *per se*, but note that inference on this coefficient would generally require HAR standard errors, and its limit distribution is in fact non-standard when $\rho \approx 1$.

Crucially, the result does not break down at moderately long horizons h. We give the formal statement of the result in Section 4 and for now just discuss heuristics. Consider any upper bound $\bar{h}_T$ on the horizon which satisfies $\bar{h}_T/T \to 0$. Then Proposition 1 below shows that, for any $a > 0$,

$$\inf_{\rho \in [-1+a,1]} \inf_{1 \le h \le \bar{h}_T} P_\rho \left(\beta(\rho, h) \in \hat{C}(h, \alpha) \right) \to 1 - \alpha \quad \text{as } T \to \infty, \tag{8}$$

where P_ρ denotes the distribution of the data $\{y_t\}$ under the AR(1) model (1) with parameter ρ. In words, the result states that, for sufficiently large sample sizes, LP inference is valid even under the *worst-case* choices of parameter $\rho \in [-1+a, 1]$ and horizon $h \in [1, \bar{h}_T]$. As is well known, such *uniform* validity is a much stronger result than *pointwise* validity for fixed ρ and h. In fact, if we restrict attention to only the stationary region $\rho \in [-1+a, 1-a]$, then the statement (8) is true with the upper bound $\bar{h}_T = (1-a)T$ on the horizon. That is, if we know the time series is not close to a unit root, then local projection inference is valid even at long horizons h that are non-negligible fractions of the sample size T. The reason why our result does not cover the case $\rho \approx -1$ is purely technical and we could easily extend our proof to this case.

2.2 Illustrative Simulation Study

We now present a small simulation study to show that lag-augmented LP achieves a favorable trade-off between robustness and efficiency relative to other procedures. For clarity, we continue to assume the simple AR(1) model (1) with known lag length. Our baseline design considers homoskedastic innovations $u_t \overset{i.i.d.}{\sim} N(0,1)$. In Appendix A.1 we present results for ARCH innovations and for general VAR(p) models.

We stress that, although we use the AR(1) model for illustration here, the central goal of this paper is to develop a procedure that is feasible even in realistic VAR(p) models, as demonstrated in Section 5. Thus, for now we avoid computationally demanding procedures, such as the AR grid bootstrap, which are difficult to apply in applied settings. We provide an extensive theoretical comparison of various inference procedures in Section 3.

Table 1 displays the coverage and median length of impulse response confidence intervals at various horizons. We consider several versions of AR inference and LP inference, either implemented using the bootstrap or using delta method standard errors. "LP" denotes local projection and "AR" autoregressive inference. "LA" denotes lag augmentation. The subscript "b" denotes bootstrap confidence intervals constructed from a wild recursive bootstrap design, as described in Section 6 (for LP we use the percentile-t confidence interval). Columns without the "b" subscript use delta method standard errors. For LA-LP, we always use Eicker-Huber-White standard errors as discussed in Section 2.1, whereas non-augmented LP always uses HAR standard errors.[7] The column "AR-LA" is the Efron bootstrap confidence interval for *lag-augmented* AR estimates developed by Inoue and Kilian (2020) and discussed further in Section 3.[8] The sample size is $T = 240$. We consider data generating processes (DGPs) $\rho \in \{0, .5, .95, 1\}$ and horizons h up to 60 periods (25% of the sample, which is not unusual in applied work). The nominal confidence level is 90%. We use 5,000 Monte Carlo repetitions, with 2,000 bootstrap draws per repetition.

Consistent with our theoretical results, the bootstrap version of lag-augmented local projection (column 1) achieves coverage close to nominal level in almost all cases, whereas the competing procedures either under-cover or return impractically wide confidence intervals. In contrast, non-augmented LP (columns 3 and 4) exhibits larger coverage distortions in almost all cases. As is well known, textbook AR delta method confidence intervals (column 6) severely under-cover when $\rho > 0$ and the horizon is even moderately large.

It is only in the cases where both $\rho = 1$ and $h \geq 36$ that lag-augmented local projection exhibits serious coverage distortions, again consistent with our theory. However, even in these cases, the coverage distortions are similar to or less pronounced than those for non-augmented LP and for delta method AR inference.

Although the Inoue and Kilian (2020) lag-augmented AR bootstrap confidence interval (column 5) achieves correct coverage for $\rho > 0$ at all horizons, this interval is extremely wide in the problematic cases where ρ is close to 1 and the horizon h is intermediate or long. We explain this fact theoretically in Section 3. Confidence intervals with median width greater than 1 would appear to be of little practical use, since the true impulse response parameter is bounded above by 1 in the AR(1) model.[9] Note also that the Inoue and Kilian (2020) interval severely under-covers when $\rho = 0$ at all even (but not odd) horizons h, as explained theoretically in Section 3.

Although outperformed by bootstrap procedures, the lag-augmented local projection

[7] As an off-the-shelf, state-of-the-art HAR procedure, we choose the Equally Weighted Cosine (EWC) estimator with degrees of freedom as recommended by Lazarus et al. (2018, equations 4 and 10). The degrees of freedom depend on the effective sample size $T - h$ and thus differ across horizons h.

[8] We use the Pope (1990) bias-corrected AR estimates to generate the bootstrap samples, as recommended by Inoue and Kilian (2020).

[9] In the AR(1) model, we could intersect all confidence intervals with the interval $[-1, 1]$. In this case, the median length of the Inoue and Kilian (2020) confidence interval is close to 1, cf. Appendix A.2.3.

Table 1: Monte Carlo results: homoskedastic innovations

	Coverage						Median length					
h	LP-LA$_b$	LP-LA	LP$_b$	LP	AR-LA$_b$	AR	LP-LA$_b$	LP-LA	LP$_b$	LP	AR-LA$_b$	AR
						$\rho = 0.00$						
1	0.904	0.895	0.914	0.891	0.894	0.895	0.218	0.211	0.234	0.217	0.211	0.210
6	0.909	0.900	0.909	0.894	0.000	1.000	0.219	0.214	0.233	0.221	0.000	0.000
12	0.903	0.893	0.898	0.893	0.000	1.000	0.221	0.216	0.230	0.226	0.000	0.000
36	0.903	0.896	0.907	0.902	0.000	1.000	0.235	0.229	0.244	0.239	0.000	0.000
60	0.894	0.884	0.895	0.891	0.000	0.978	0.251	0.244	0.261	0.256	0.000	0.000
						$\rho = 0.50$						
1	0.910	0.896	0.915	0.886	0.899	0.897	0.218	0.211	0.202	0.185	0.211	0.184
6	0.908	0.897	0.912	0.880	0.899	0.829	0.252	0.245	0.293	0.267	0.045	0.031
12	0.911	0.903	0.917	0.900	0.899	0.758	0.255	0.248	0.290	0.278	0.002	0.001
36	0.903	0.891	0.900	0.881	0.899	0.628	0.270	0.262	0.310	0.297	0.000	0.000
60	0.909	0.898	0.907	0.887	0.899	0.581	0.291	0.279	0.335	0.319	0.000	0.000
						$\rho = 0.95$						
1	0.912	0.894	0.835	0.827	0.902	0.848	0.220	0.212	0.077	0.073	0.212	0.076
6	0.904	0.839	0.851	0.791	0.902	0.808	0.524	0.452	0.395	0.344	1.013	0.318
12	0.893	0.814	0.853	0.751	0.902	0.765	0.679	0.550	0.642	0.518	1.751	0.429
36	0.876	0.808	0.873	0.667	0.902	0.655	0.726	0.624	0.860	0.616	6.614	0.270
60	0.902	0.835	0.900	0.693	0.902	0.585	0.730	0.649	0.949	0.644	23.317	0.093
						$\rho = 1.00$						
1	0.901	0.877	0.830	0.551	0.882	0.532	0.219	0.211	0.041	0.040	0.210	0.040
6	0.880	0.779	0.838	0.499	0.882	0.494	0.566	0.500	0.243	0.222	1.208	0.217
12	0.854	0.680	0.835	0.416	0.882	0.454	0.826	0.675	0.481	0.388	2.546	0.381
36	0.746	0.436	0.765	0.199	0.882	0.348	1.359	0.955	1.196	0.595	21.403	0.671
60	0.654	0.291	0.722	0.150	0.882	0.288	1.477	0.995	1.695	0.645	165.133	0.734

Coverage probability and median length of nominal 90% confidence intervals at different horizons. AR(1) model with $\rho \in \{0, .5, .95, 1\}$, $T = 240$, i.i.d. standard normal innovations. 5,000 Monte Carlo repetitions.

delta method interval (column 2) performs well among the group of delta method procedures. Its coverage distortions are much less severe than textbook AR delta method inference (column 4) and non-augmented LP inference with HAR standard errors (column 6). Recall that the lag-augmented LP confidence interval is at least as easy to compute as these other delta method confidence intervals. The reason why the bootstrap improves on the coverage properties of the delta method procedures is related to the well-known finite-sample bias of AR and LP estimators (Herbst and Johannsen, 2020). As as side note, our bootstrap implementation of *non-augmented* LP appears to be quite effective at correcting the most severe coverage distortions of the delta method procedure.

Table 1 illustrates the fact that the robustness of lag-augmented local projection inference entails an efficiency loss relative to AR inference when ρ is well below 1, although this loss is not large in absolute terms. In *percentage* terms, local projection confidence intervals are much wider than AR-based confidence intervals when $\rho \ll 1$ and the horizon h is intermediate or long, since AR procedures mechanically impose that the impulse response function tends to 0 geometrically fast with the horizon. Yet, in *absolute* terms, the median length of the LP confidence intervals is not so large as to be a major impediment to applied research. The relative efficiency of lag-augmented LP vs. non-augmented LP cannot be ranked and depends on the DGP and on the horizon. When ρ is close to 1, lag-augmented LP intervals are sometimes (much) narrower than lag-augmented AR intervals. We analytically characterize the various efficiency trade-offs in Appendix A.2.2.

3 Comparison With Other Inference Procedures

The simulations and theoretical results in this paper suggest that lag-augmented local projection is the only inference method that (i) achieves uniformly valid coverage over the DGP and over a wide range of horizons, (ii) has reasonable average length in problematic parts of the parameter space, and (iii) is computationally straight-forward to implement in realistic settings. In this section we discuss in more detail the coverage and length properties of alternative confidence interval procedures for impulse responses. We review the well-known drawbacks of textbook AR inference, provide new results on the relative length of lag-augmented LP vs. non-augmented LP and lag-augmented AR, and discuss the computational challenges of the AR grid bootstrap. We refer the reader back to the small-scale simulation study in Section 2.2 for illustrations of the following arguments.

Textbook autoregressive inference. The uniformity result (8) for lag-augmented LP stands in stark contrast to textbook AR inference on impulse responses, which suffers from several well-known issues. First, for the standard OLS AR estimator, the usual asymptotic normal limiting theory is invalid when the derivative of the impulse response parameter with respect to the AR coefficients has a singular Jacobian matrix. In the AR(1) case, this occurs in the white noise case $\rho = 0$ (Benkwitz et al., 2000). Second, as with non-augmented LP, textbook AR inference suffers from non-uniformity issues when the data is nearly non-stationary (Phillips, 1998, Remark 2.5). Third, pre-testing for the presence of a unit root does not yield uniformly valid inference and can lead to poor finite sample performance (e.g., Mikusheva, 2007, p. 1412). Fourth, plug-in AR inference with normal critical values must necessarily break down at medium-long horizons $h = h_T \propto T^{1/2}$ and at long horizons $h_T \propto T$. We show this analytically in Appendix A.2.1, which builds on insights in Mikusheva (2012). Wright (2000) and Pesavento and Rossi (2006, 2007) construct confidence intervals for persistent processes at long horizons $h = h_T \propto T$ by inverting the non-standard AR limit distribution, but these tailored procedures do not work uniformly over the parameter space or over the horizon.

The severe under-coverage of the delta method AR inference method is starkly illustrated in Section 2.2 (see Column 6 of Table 1). As discussed in detail by Inoue and Kilian (2020), standard bootstrap approaches to AR inference do not solve all the uniformity issues.

We must emphasize, however, that if we restrict attention to stationary processes and short-horizon impulse responses, the standard OLS AR impulse response estimator is more efficient than lag-augmented LP. Hence, there is a trade-off between efficiency in benign settings and robustness to persistence and longer horizons, as is also clear in the simulation results in Section 2.2. We expand upon the efficiency properties of the standard AR estimator in Appendix A.2.2.

Lag-augmented AR inference. The above-mentioned non-uniformity of the textbook AR inference method in the case of near-non-stationary data can be remedied by lag augmentation (Inoue and Kilian, 2020). In the case of an AR(1) model, the lag-augmented AR estimator $\hat{\beta}_{\text{ARLA}}(h)$ is given by $\hat{\rho}_1^h$, where $(\hat{\rho}_1, \hat{\rho}_2)$ are the OLS coefficients from a regression of y_t on (y_{t-1}, y_{t-2}) (i.e., we estimate an AR(2) model). The intuition why this guarantees a normal limiting distribution even in the unit root case is the same as in Section 2.1. Lag-augmented AR and lag-augmented LP coincide at horizon $h = 1$, but not at longer horizons. Lag augmentation involves a loss of efficiency: The lag-augmented AR estimator is strictly less efficient than the non-augmented AR estimator except when the true process is white noise (see Appendix A.2.2). Note that lag augmentation by itself does not solve the above-mentioned issues that occur when the Jacobian of the impulse response transformation is singular, or when doing inference at medium-long or long horizons.[10]

The bootstrap confidence interval for lag-augmented AR proposed by Inoue and Kilian (2020) has valid coverage even at long horizons. Specifically, Inoue and Kilian (2020) show that the *Efron* bootstrap confidence interval—applied to resursive AR bootstrap samples of $\hat{\beta}_{\text{ARLA}}(h)$—has valid coverage even at long horizons $h = h_T \propto T$, as long as the largest autoregressive root is bounded away from 0.[11]

Unfortunately, we show in Appendix A.2.3 that the expected length of the lag-augmented AR estimator is prohibitively large when the data is persistent and the horizon is long. Precisely, in the case of an AR(1) model, $\hat{\beta}_{\text{ARLA}}(h) = \hat{\rho}_1^h$ is *inconsistent* for sequences of DGPs $\rho = \rho_T$ and horizons $h = h_T$ such that $h_T \propto T^\eta, \eta \in [1/2, 1]$, and $h_T(1 - \rho_T) \to a \in [0, \infty)$. The reason is that the lag-augmented coefficient estimator $\hat{\rho}_1$ converges at rate $T^{-1/2}$ even in the unit root case, implying that the estimation error in $\hat{\rho}_1$ is not negligible when raising the estimator to a power of $h = h_T$. We show that this implies that the Efron bootstrap confidence interval does not shrink to zero for such sequences ρ_T and h_T. In fact, when $\eta > 1/2$, the width of the confidence interval for the h_T impulse response is almost equal to the entire positive part of the parameter space $[0, 1]$ with probability equal to the nominal level. This contrasts with the lag-augmented LP confidence interval, which is consistent (i.e., its length shrinks to 0 in probability) for any sequence $\rho_T \in [-1, 1]$ and any sequence h_T such that $h_T/T \to 0$. The large width of the Inoue and Kilian (2020) interval is illustrated in the simulations in Section 2.2 (see the second-to-last column in Table 1).

Interestingly, *if we restrict attention to stationary processes and short horizons, the relative efficiency of lag-augmented AR and lag-augmented LP inference is ambiguous.* We show this analytically in Appendix A.2.2 in the context of a homoskedastic AR(1) model with a fixed horizon h of interest. Lag-augmented AR is more efficient than lag-augmented LP when ρ is small or when the horizon h is large, and vice versa. For any horizon h, there exists

[10]The AR(1) simulations in Section 2.2 show that the coverage of the Inoue and Kilian (2020) confidence interval is 0 at all *even* horizons when $\rho = 0$. This is because the true impulse response is 0, but the bootstrap samples of $\hat{\rho}_1^h$ are all strictly positive. Their procedure achieves uniformly correct coverage at *odd* horizons.

[11]For intuition, consider the AR(1) case. The Efron bootstrap preserves monotonic transformations, and the bootstrap transformation $\beta(\rho, h) = \rho^h$ is monotonic (if we restrict attention to $\rho \in (0, 1]$ or $\rho \in [-1, 0)$). Hence, the Efron confidence interval is valid for ρ^h if it is valid for ρ itself. In more general VAR(p) models, the same argument can be applied at long horizons, since here only the largest autoregressive root matters for impulse responses (if the roots are well-separated).

some cut-off value for $\rho \in (0,1)$, above which lag-augmented LP is more efficient. Intuitively, the nonlinear impulse response transformation $\rho \mapsto \rho^h$ is highly sensitive to values of ρ near 1 whenever h is large, which compounds the effects of estimation error in $\hat{\rho}$, whereas LP is a purely linear procedure.

AR GRID BOOTSTRAP AND PROJECTION. The grid bootstrap of Hansen (1999) represents a computationally intensive approach to doing valid inference at fixed and long horizons, regardless of persistence, but it is invalid at intermediate horizons, as shown by Mikusheva (2012). The grid bootstrap is based on test inversion, so it requires running an autoregressive bootstrap on each point in a fine grid of potential values for the impulse response parameter of interest. It also requires estimating a constrained OLS estimator that imposes the hypothesized null on the impulse response at each point in the grid. Recall that lag-augmented LP inference is computationally simple and valid at any horizon $h = h_T$ satisfying $h_T/T \to 0$. However, in the case of unit roots and long horizons $h_T \propto T$, lag-augmented LP inference with normal critical values is not valid, while the grid bootstrap is valid (Mikusheva, 2012).

Another computationally intensive approach is to form a uniformly valid confidence set for the AR parameters and then map it into a confidence interval for impulse responses by projection. Although doable in an AR(1) or AR(2) model, this approach would appear to be computationally infeasible and possibly highly conservative in realistic VAR(p) settings, unlike lag-augmented LP (see Section 5).

OTHER LOCAL PROJECTION APPROACHES. Non-augmented LP is not robust to non-stationarity, as already discussed in Section 2.1. *If the data is stationary and the horizon h is fixed, the relative efficiency of non-augmented LP and lag-augmented LP is generally ambiguous*, as shown in Appendix A.2.2 in the case of a homoskedastic AR(1) model. The reason is that, although non-augmented LP uses a regressor (y_t) that has higher variance than in the lag-augmented case (where the effective regressor is u_t, as discussed in Section 2.1), the asymptotic variance of the non-augmented LP estimator is affected by the serial correlation of the multi-step forecast error. Thus, lag-augmented LP is relatively more efficient the smaller is ρ and the larger is h.

In some empirical settings, the researcher may directly observe the autoregressive innovation, or some component of the innovation, for example by constructing narrative measures of economic shocks (Ramey, 2016). For concreteness, consider the AR(1) model (1) and assume we observe the innovation u_t. In this case, it is common in empirical practice to simply regress y_{t+h} on u_t, without controls. Although this strategy provides consistent impulse response estimates when the data is stationary, it is inefficient relative to lag-augmented LP, since the latter approach additionally controls for the variable y_{t-1}, which would otherwise show up in the error term in the representation (4). Thus, lag augmentation is desirable on robustness and efficiency grounds even if some shocks are directly observed.

Summary. Existing and new theoretical results confirm the main message of our simulations in Section 2.2: Lag-augmented LP is the only procedure that is computationally feasible in realistic problems and can be shown to have valid coverage under a wide range of DGPs and horizon lengths, without achieving such valid coverage by returning a confidence interval that is impractically wide. This robustness does come at the cost of a loss of efficiency relative to non-robust AR methods. However, the efficiency loss is large in *relative* terms only in stationary, short-horizon cases, where lag-augmented LP confidence intervals do well in *absolute* terms, as illustrated in Section 2.2. Based on these results, we believe that it is only in the case of highly persistent data and very long horizons $h = h_T \propto T$ that the use of alternative robust procedures should be considered, such as the computationally demanding AR grid bootstrap.

4 Technical Result: AR(1) Case

We now formally state our main result for the AR(1) model introduced in Section 2 and discuss further technical assumptions and details of the proof technique.

Additional assumptions. We require some further standard regularity assumptions on the moments of the innovation process.

Assumption 2.

i) $E(u_t^8) < \infty$, *and there exists* $\delta > 0$ *such that* $E(u_t^2 \mid \{u_s\}_{s<t}) \geq \delta$ *almost surely.*

ii) $\{u_t^2\}$ *has absolutely summable cumulants up to 4th order.*

Condition (i) is a standard requirement for consistent estimation of regression standard errors with possibly heteroskedastic residuals. The assumption provides a lower bound on $\sigma^2 \equiv E(u_t^2) \geq \delta$. Condition (ii) is a standard "weak dependence" restriction on the squared innovation process $\{u_t^2\}$, cf. Brillinger (2001, Chapter 2.6).

Finally, we impose a high-level contiguity condition, which simplifies the proof of consistency of the standard errors $\hat{s}(h)$ in the case $\rho \approx 1$.

Assumption 3. *For any sequence* $\{\rho_T\}$ *of real numbers in* $[-1, 1]$ *such that* $\lim_{T\to\infty} T(1 - \rho_T) < \infty$, *the sequence of probability measures* $\{P_{\rho_T}\}$ *is contiguous to the measure* P_1.

The role of this assumption is explained further below. When the innovations u_t are i.i.d., the above high-level condition is satisfied under a mild smoothness condition on the density of u_t, see Jansson (2008). The condition is also known to allow for certain types of conditional heteroskedasticity, see Jeganathan (1995, Section 4).

Main result. Our main result is that the lag-augmented local projection estimator $\hat{\beta}(h)$ is a uniformly asymptotically normal estimator of the true impulse response $\beta(\rho, h) \equiv \rho^h$, and the Eicker-Huber-White standard error $\hat{s}(h)$ is also uniformly valid. The uniformity extends over both data generating processes (i.e., ρ) and over horizons h, up to some limit. Recall that P_ρ denotes the probability measure of the data $\{y_t\}$ under the AR(1) model (1) with parameter ρ. Note that we keep the data generating process for $\{u_t\}$ fixed in our analysis. Let $\Phi(\cdot)$ denote the standard normal cumulative distribution function.

Proposition 1. *Let Assumptions 1 and 2 hold. Let* $a \in (0, 1)$.

i) For all $x \in \mathbb{R}$,

$$\sup_{\rho\in[-1+a,1-a]} \sup_{1\leq h\leq(1-a)T} \left| P_\rho \left(\frac{\hat{\beta}(h) - \beta(\rho, h)}{\hat{s}(h)} \leq x \right) - \Phi(x) \right| \to 0.$$

ii) Let additionally Assumption 3 hold. Consider any sequence $\{\bar{h}_T\}$ *of nonnegative integers such that* $\bar{h}_T < T$ *for all* T *and* $\bar{h}_T/T \to 0$. *Then for all* $x \in \mathbb{R}$,

$$\sup_{\rho\in[-1+a,1]} \sup_{1\leq h\leq\bar{h}_T} \left| P_\rho \left(\frac{\hat{\beta}(h) - \beta(\rho, h)}{\hat{s}(h)} \leq x \right) - \Phi(x) \right| \to 0.$$

Proof. See Appendix B.2 below. □

Part (i) restricts attention to the stationary region, allowing horizons h up to some fraction of the sample size T. Part (ii) considers also local-to-unity and unit root processes, but requires the horizon h to be a negligible fraction of the sample size asymptotically (but allowing for medium-long horizons, as discussed in Section 2.1). The uniform coverage result (8) mentioned in Section 2.1 follows immediately from Proposition 1 by choosing $x = \pm z_{1-\alpha/2}$.

Remarks.

1. We show in Lemmas 4 and 5 below that $\hat{s}(h) \approx (T-h)^{-1/2} v(\rho, h)$, where $v(\rho, h)^2 \leq \min\{\frac{1}{1-\rho^2}, h\}$. Thus, the standard error of $\hat{\beta}(h)$ is of order $(h/(T-h))^{1/2}$ when $|\rho| = 1$ (or ρ is local-to-unity) and of order $(T-h)^{-1/2}$ when $|\rho|$ is bounded away from 1.

2. There are two main technical difficulties in establishing the uniform validity of local projection inference.

   a) The variance of the regression residual $\xi_t(\rho, h)$ is increasing in the horizon h and also depends on ρ. Thus, the simplest laws of large numbers and central limit theorems for stationary processes do not apply. We instead apply a central limit theorem for martingale difference sequences and derive uniform bounds on moments of relevant variables. The central limit theorem is delicate, since the regression scores $\xi_t(\rho, h)u_t$ are not a martingale difference sequence with respect to the natural filtration generated by past u_t's. However, it is possible to "reverse time" in a way that makes the scores a martingale difference sequence with respect to an alternative filtration, see the proof of the auxiliary Lemma 1.

   b) To handle both unit roots, stationary processes, and everything in between, we must consider various kinds of sequences of drifting parameters $\rho = \rho_T$, following the general logic of Andrews et al. (2019). This is primarily an issue when showing consistency of the standard error $\hat{s}(h)$, which requires deriving the convergence rates of the various estimators along drifting parameter sequences. To that end, near-stationary cases are handled via explicit calculation of moment bounds, while near-unit-root cases are handled by appealing to the contiguity imposed by Assumption 3 so that we may restrict attention to the well-studied unit root case $\rho = 1$. This is the only part of the proof that uses Assumption 3 (specifically, the proof of the auxiliary Lemma 9 below).

3. We conjecture that Proposition 1 can be extended to cover the entire range $\rho \in [-1, 1]$ under an additional assumption akin to Assumption 3 applied to the case $\rho = -1$. We refrain from further complicating the assumptions since economic data is rarely highly anti-persistent.

4. The proposition does not cover the case $h \propto T$ when $\rho \approx 1$. Simulation evidence and analytical calculations strongly suggest that the asymptotic normality in Proposition 1 does *not* go through when $\rho = 1$ and $h = \kappa T$ for $\kappa \in (0, 1)$. Indeed, in this case the

sample variance of the regression scores $\xi_t(\rho, h)u_t$ appears to not converge in probability to a constant, thus violating the conclusion of the key auxiliary Lemma 2 below. The behavior of plug-in AR impulse response estimates is also non-standard when $\rho \approx 1$ and $h \propto T$, see Phillips (1998), Pesavento and Rossi (2006, 2007), and Mikusheva (2012).

5. If the innovations u_t were observed, an alternative estimator would regress y_{t+h} onto u_t and y_{t-1}. As discussed in Section 2.1, this estimator is numerically equivalent with $\hat{\beta}(h)$, so the uniformity result carries over.

6. It is not important that the process $\{y_t\}$ is started at $y_0 = 0$ in (1). The proof in Appendix B easily extends to the case where y_0 is random with bounded fourth moment and independent of $\{u_t\}$.

5 General VAR(p) Case

We now generalize the inference procedure and the theoretical uniformity result to cover a VAR(p) model with an intercept. In this case, the lag-augmented LP procedure controls for an intercept and p lags of all the time series that enter into the VAR model. Then all the intuition from the AR(1) model in Section 2.1 goes through. For clarity, we focus on reduced-form impulse responses and leave *structural* inference for future work.

Throughout we assume that the lag length p is known, as in Mikusheva (2012) and Inoue and Kilian (2020). This of course covers the case where the lag length is *over*-stated, since we allow the VAR coefficients to equal zero after some lag.

[Work in progress.]

6 Bootstrap Implementation

In this section we describe the bootstrap implementation of lag-augmented local projection that we recommend for practical use. Although we do not prove that the bootstrap delivers asymptotic refinements relative to the non-bootstrap procedure discussed in Sections 2 and 5, we find in simulations that the bootstrap is effective at correcting small-sample coverage distortions. These distortions arise primarily due to the small-sample bias of local projection, which Herbst and Johannsen (2020) show is analogous to the well-known bias of the AR OLS estimator.

Although other bootstrap procedures are also valid, we have found in simulations that a recursive autoregressive bootstrap design delivers the most accurate results.[12] Our baseline algorithm is based on the wild bootstrap, which allows for heteroskedastic AR innovations (Gonçalves and Kilian, 2004) as in our theoretical results. Again guided by simulation evidence, we construct the bootstrap confidence interval using the equal-tailed percentile-t method, which has a built-in bias correction (Kilian and Lütkepohl, 2017, Chapter 12.2.6).

The bootstrap procedure for computing a $1-\alpha$ confidence interval proceeds as follows, assuming a VAR(p) model:

1. Estimate the impulse response of interest by lag-augmented local projection as in Sections 2 and 5. Denote the estimate and its standard error by $\hat{\beta}(h)$ and $\hat{s}(h)$, respectively.

2. Estimate the VAR(p) model by OLS (without lag augmentation). Compute the corresponding VAR residuals $\hat{u}_t$. Bias-adjust the VAR coefficients using the formula in Pope (1990).

3. Compute the impulse response of interest implied by the VAR model estimated in step 2. Denote this impulse response by $\hat{\beta}_{\text{VAR}}(h)$.

4. For each bootstrap iteration $b = 1, \dots, B$:

   i) Generate bootstrap residuals $\hat{u}_t^* \equiv U_t \hat{u}_t$, $t = 1, \dots, T$, where $U_t \overset{i.i.d.}{\sim} N(0,1)$ are computer-generated random variables that are independent of the data.

   ii) Draw a block of p initial observations $(y_1^*, \dots, y_p^*)$ uniformly at random from the $T-p+1$ blocks of p observations in the original data.

   iii) Generate bootstrap data y_t^*, $t = p+1, \dots, T$, by iterating on the bias-corrected VAR(p) model estimated in step 2, using the innovations $\hat{u}_t^*$.

   iv) Apply the lag-augmented LP estimator to the bootstrap data $\{y_t^*\}$. Denote the impulse response estimate and its standard error by $\hat{\beta}(h)^*$ and $\hat{s}(h)^*$, respectively.

   v) Store $\hat{T}_b^* \equiv (\hat{\beta}(h)^* - \hat{\beta}_{\text{VAR}}(h))/\hat{s}(h)^*$.[13]

[12]It is also valid to use a fixed-design pairs bootstrap, as in any other regression model with serially uncorrelated scores. This bootstrap procedure is the one carried out by Stata using the `bootstrap` command with standard settings. We find in simulations that this design yields accurate coverage except when the data is highly persistent.

[13]It is critical that the bootstrap t-statistic $\hat{T}_b^*$ is centered at the VAR-implied impulse response $\hat{\beta}_{\text{VAR}}(h)$ rather than the LP-estimated impulse response $\hat{\beta}(h)$. This is because the former estimate is the pseudo-true parameter in the recursive bootstrap DGP, and the latter estimate differs from the former by an amount that is not asymptotically negligible.

5. Compute the $\alpha/2$ and $1-\alpha/2$ quantiles of the B draws of $\hat{T}_b^*$, $b = 1, \dots, B$. Denote these by $\hat{Q}_{\alpha/2}$ and $\hat{Q}_{1-\alpha/2}$, respectively.

6. Return the percentile-t confidence interval $[\hat{\beta}(h) - \hat{s}(h)\hat{Q}_{1-\alpha/2}, \hat{\beta}(h) - \hat{s}(h)\hat{Q}_{\alpha/2}]$.[14]

The bias correction in step 2 is optional from the perspective of asymptotic validity, but it helps a bit in small samples. We hope to establish the uniform validity of our suggested bootstrap procedure in future work.

7 Conclusion and Directions for Future Research

The simplicity and statistical robustness of *lag-augmented* local projection inference makes it an attractive option relative to existing inference procedures. We recommend that applied researchers conduct inference based on lag-augmented local projections with heteroskedasticity-robust (Eicker-Huber-White) standard errors. This procedure can be implemented using any regression software and has desirable theoretical properties relative to textbook delta method autoregressive inference and to non-augmented local projection methods. In particular, we showed that confidence intervals based on lag-augmented local projections that use robust standard errors with standard normal critical values are uniformly valid over the persistence in the data and for a wide range of horizons. We also recommended a simple bootstrap implementation in Section 6, which seems to achieve even better finite-sample performance.

In our opinion, there are only two cases in which the lag-augmented local projection inference method is inferior to competitors: i) If the data is known to be at most moderately persistent and interest centers on very short impulse response horizons, in which case textbook AR inference is valid and efficient. ii) When the data has near-unit roots and interest centers on horizons that are a substantial fraction of the sample size, in which case the computationally demanding AR grid bootstrap may be deployed (Hansen, 1999; Mikusheva, 2012). In all other cases, lag-augmented local projection inference appears to achieve a competitive trade-off between robustness and efficiency.

Methodologically, this paper established novel statistical arguments supporting the popular practice of conducting local projection inference on impulse responses. We stress that

[14]It is not valid to use the Efron bootstrap confidence interval based on the bootstrap quantiles of $\hat{\beta}(h)^*$. This is because the bootstrap samples are asymptotically centered around $\hat{\beta}_{\text{VAR}}(h)$, not $\hat{\beta}(h)$. If we used a fixed-design pairs bootstrap design instead of the recursive design, the usual Efron bootstrap confidence interval would be valid, but simulations suggest that this procedure is less accurate in small samples.

conventional AR-based procedures will deliver smaller standard errors than local projections in many cases, but this comes at the cost of fragile coverage properties, especially at longer horizons.

Our results rely on certain simplifying assumptions that may be possible to relax in future work. First, it would appear straight-forward to generalize our results on reduced-form impulse response inference to *structural* inference using the results of Plagborg-Møller and Wolf (2019). Second, whereas we adopt a frequentist perspective in this paper, it remains an open question whether local projection inference is relevant from a Bayesian perspective.

A Appendix

A.1 Simulation Study: Further Results

A.1.1 AR(1) Model With ARCH(1) Innovations

Consider the AR(1) model (1) with innovations u_t that follow an ARCH(1) process

$$u_t = \tau_t \varepsilon_t, \quad \tau_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2, \quad \varepsilon_t \overset{i.i.d.}{\sim} N(0,1). \tag{9}$$

These innovations satisfy Assumption 1. In our simulations, we set $\alpha_1 = .7$ and $\alpha_0 = (1 - \alpha_1)$.[15] Table 2 presents the results, which are qualitatively similar to the i.i.d. case discussed in Section 2.2.

A.1.2 VAR(p) Models

[Work in progress.]

A.2 Comparison With Other Procedures: Further Results

In this subsection we provide details on the under-coverage of delta method AR inference and on the relative efficiency of lag-augmented LP versus other procedures. Throughout, we focus on the tractable case of an AR(1) model as in Section 2.

A.2.1 Undercoverage of Delta Method AR Inference

Here we compute the under-coverage of delta method AR inference (both lag-augmented and non-augmented) for horizons $h \propto T^{1/2}$. This discussion is merely intended to flesh out the insights of Mikusheva (2012, Section 4.3).

Consider a $\sqrt{T}$-consistent asymptotically normal estimator $\hat{\rho}$ of the AR(1) parameter ρ:

$$T^{1/2}(\hat{\rho} - \rho) \xrightarrow{d} N(0, \omega^2).$$

For example, for plain AR(1) inference we would have $\omega^2 = 1 - \rho^2$, and for lag-augmented AR inference we would have $\omega^2 = 1$ (see Appendix A.2.2 below). Suppose we have available a consistent estimator $\hat{\omega} \xrightarrow{p} \omega > 0$. Let $\hat{\beta}_{\text{AR}}(h) \equiv \hat{\rho}^h$ be the plug-in AR estimator of the

[15] This value of α_0 ensures $\mathbb{E}[\tau_t^2] = 1$.

Table 2: Monte Carlo results: ARCH innovations

	Coverage						Median length					
h	LP-LA$_b$	LP-LA	LP$_b$	LP	AR-LA$_b$	AR	LP-LA$_b$	LP-LA	LP$_b$	LP	AR-LA$_b$	AR
						$\rho = 0.00$						
1	0.896	0.863	0.913	0.803	0.835	0.870	0.384	0.355	0.408	0.314	0.334	0.357
6	0.908	0.900	0.906	0.860	0.000	1.000	0.211	0.208	0.219	0.195	0.000	0.000
12	0.907	0.899	0.908	0.891	0.000	1.000	0.209	0.205	0.209	0.201	0.000	0.000
36	0.899	0.890	0.898	0.887	0.000	1.000	0.221	0.216	0.222	0.215	0.000	0.000
60	0.899	0.892	0.904	0.889	0.000	0.988	0.238	0.231	0.238	0.230	0.000	0.000
						$\rho = 0.50$						
1	0.887	0.856	0.907	0.797	0.833	0.870	0.385	0.355	0.332	0.257	0.335	0.292
6	0.905	0.887	0.906	0.842	0.833	0.758	0.245	0.238	0.273	0.234	0.085	0.046
12	0.902	0.892	0.904	0.883	0.833	0.678	0.239	0.232	0.264	0.249	0.007	0.001
36	0.898	0.887	0.896	0.874	0.833	0.554	0.255	0.246	0.279	0.265	0.000	0.000
60	0.908	0.889	0.901	0.881	0.833	0.520	0.274	0.262	0.301	0.284	0.000	0.000
						$\rho = 0.95$						
1	0.894	0.855	0.818	0.807	0.830	0.848	0.396	0.362	0.084	0.079	0.338	0.087
6	0.895	0.811	0.835	0.760	0.830	0.797	0.627	0.518	0.383	0.327	1.744	0.354
12	0.872	0.770	0.836	0.733	0.830	0.750	0.724	0.561	0.603	0.488	3.915	0.468
36	0.867	0.791	0.856	0.664	0.830	0.632	0.716	0.587	0.806	0.593	64.157	0.291
60	0.892	0.821	0.887	0.679	0.830	0.571	0.708	0.616	0.880	0.616	1028.366	0.095
						$\rho = 1.00$						
1	0.894	0.860	0.839	0.584	0.838	0.566	0.385	0.356	0.040	0.041	0.331	0.041
6	0.879	0.770	0.858	0.536	0.838	0.521	0.689	0.590	0.240	0.228	2.089	0.224
12	0.853	0.666	0.842	0.457	0.838	0.472	0.914	0.720	0.469	0.393	5.683	0.391
36	0.740	0.426	0.756	0.213	0.838	0.363	1.386	0.946	1.175	0.604	194.037	0.676
60	0.653	0.288	0.711	0.169	0.838	0.304	1.496	0.981	1.670	0.642	6503.801	0.733

Coverage probability and median length of nominal 90% confidence intervals at different horizons. AR(1) model with $\rho \in \{0, .5, .95, 1\}$, $T = 240$, innovations as in equation (9). 5,000 Monte Carlo repetitions.

impulse response $\beta(\rho, h) \equiv \rho^h$ at horizon h. The usual delta method standard error for $\hat{\beta}_{\mathrm{AR}}(h)$ equals

$$\widehat{\mathrm{se}}_{\mathrm{AR}}(h) \equiv T^{-1/2} h |\hat{\rho}|^{h-1} \hat{\omega}.$$

Consider specifically the horizon $h = h_T = \kappa T^{1/2}$, where $\kappa \in (0, \infty)$ is a fixed constant. Let $\Phi(\cdot)$ denote the standard normal CDF, and let Z denote a standard normal random variable. For any $\rho \neq 0$, the coverage probability of the usual $1-\alpha$ confidence interval equals

$$\begin{aligned}
& P_\rho\left(\frac{|\hat{\beta}_{\mathrm{AR}}(h_T) - \beta(\rho, h_T)|}{\widehat{\mathrm{se}}_{\mathrm{AR}}(h_T)} \le \Phi^{-1}(1-\alpha/2)\right) \\
&= P_\rho\left(\left|\frac{T^{1/2}(\hat{\rho}^{h_T} - \rho^{h_T})}{h_T \hat{\rho}^{h_T - 1}\hat{\omega}}\right| \le \Phi^{-1}(1-\alpha/2)\right) \\
&= P_\rho\left(\left|\frac{\hat{\rho}}{\hat{\omega}\kappa}\left(1 - \left(\frac{\rho}{\hat{\rho}}\right)^{\kappa T^{1/2}}\right)\right| \le \Phi^{-1}(1-\alpha/2)\right) \\
&= P_\rho\left(\left|\frac{\hat{\rho}}{\hat{\omega}\kappa}\left(1 - e^{-\kappa T^{1/2}(\log\hat{\rho} - \log\rho)}\right)\right| \le \Phi^{-1}(1-\alpha/2)\right) \\
&\to P\left(\left|\frac{\rho}{\omega\kappa}\left(1 - e^{-\kappa\rho^{-1}\omega Z}\right)\right| \le \Phi^{-1}(1-\alpha/2)\right) \\
&= P\left(\left|\frac{\rho}{\omega\kappa}\left(1 - e^{\kappa|\rho|^{-1}\omega Z}\right)\right| \le \Phi^{-1}(1-\alpha/2)\right). \qquad (10)
\end{aligned}$$

The last two steps use the delta method and symmetry of the normal distribution:

$$T^{1/2}(\log\hat{\rho} - \log\rho) \xrightarrow{d} \frac{1}{\rho} N(0, \omega^2) \stackrel{d}{=} \frac{\omega}{|\rho|} Z.$$

Note that the final expression (10) for the asymptotic coverage probability is a function only of α and the quantity $\zeta \equiv \frac{|\rho|}{\omega\kappa} > 0$. Denote the asymptotic coverage probability by $\mathrm{AsyCov}(\alpha, \zeta)$. We can further simplify this as follows:

$$\begin{aligned}
\mathrm{AsyCov}(\alpha, \zeta) &\equiv P\left(\zeta\left|1 - e^{Z/\zeta}\right| \le \Phi^{-1}(1-\alpha/2)\right) \\
&= P\left(-\frac{\Phi^{-1}(1-\alpha/2)}{\zeta} \le 1 - e^{Z/\zeta} \le \frac{\Phi^{-1}(1-\alpha/2)}{\zeta}\right) \\
&= P\left(1 - \frac{\Phi^{-1}(1-\alpha/2)}{\zeta} \le e^{Z/\zeta} \le 1 + \frac{\Phi^{-1}(1-\alpha/2)}{\zeta}\right) \\
&= P\left(\zeta\log\left(1 - \frac{\Phi^{-1}(1-\alpha/2)}{\zeta}\right) \le Z \le \zeta\log\left(1 + \frac{\Phi^{-1}(1-\alpha/2)}{\zeta}\right)\right)
\end{aligned}$$

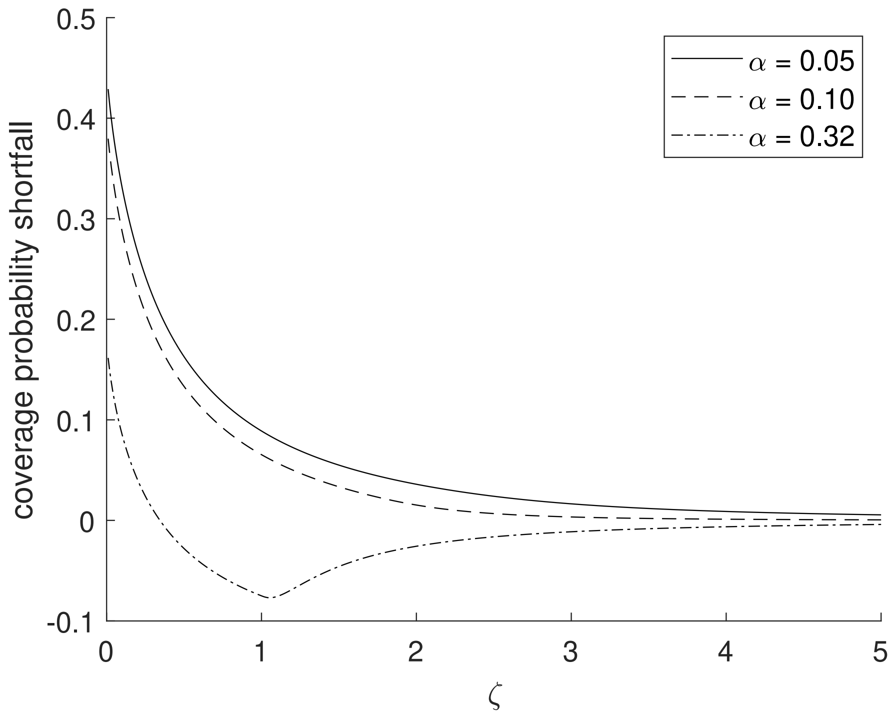

Figure 1: Asymptotic coverage shortfall $(1-\alpha)-\text{AsyCov}(\alpha,\zeta)$ of conventional AR confidence interval at horizon $h_T = \kappa T^{1/2}$. Horizontal axis: $\zeta = \frac{|\rho|}{\omega\kappa}$. The three curves correspond to $\alpha = 0.05, 0.10, 0.32$.

$$= \Phi\left(\zeta \log\left(1 + \frac{\Phi^{-1}(1-\alpha/2)}{\zeta}\right)\right) - \Phi\left(\zeta \log\left(1 - \frac{\Phi^{-1}(1-\alpha/2)}{\zeta}\right)\right).$$

Here we use the convention that $\log(x) = -\infty$ if $x \leq 0$. Observe that

$$\lim_{\zeta\downarrow 0} \text{AsyCov}(\alpha,\zeta) = \Phi\left(\lim_{\zeta\downarrow 0} \zeta \log\left(1 + \frac{\Phi^{-1}(1-\alpha/2)}{\zeta}\right)\right) = \Phi(0) = 1/2,$$

$$\lim_{\zeta\to\infty} \text{AsyCov}(\alpha,\zeta) = \lim_{\zeta\to\infty} \left\{\Phi\left(\zeta\frac{\Phi^{-1}(1-\alpha/2)}{\zeta}\right) + \Phi\left(-\zeta\frac{\Phi^{-1}(1-\alpha/2)}{\zeta}\right)\right\} = 1-\alpha.$$

Thus, for any significance level $\alpha < 1/2$, the AR confidence interval under-covers whenever ζ is sufficiently small (i.e., when $|\rho|/\omega$ is small or $\kappa = h_T/T^{1/2}$ is large). The coverage is close to the nominal level when ζ is large (i.e., $|\rho|/\omega$ is not too small and κ is small).

Figure 1 plots the coverage shortfall $(1-\alpha) - \text{AsyCov}(\alpha,\zeta)$ as a function of $\zeta = \frac{|\rho|}{\omega\kappa}$, for three conventional significance levels α. For $\alpha = 0.05$ and 0.10, the AR confidence interval under-covers asymptotically regardless of the value of ζ. For $\alpha = 0.32$, the conventional interval is conservative at intermediate values of ζ (but liberal for small ζ, of course).

A.2.2 Relative Efficiency of Lag-Augmented LP

Here we compare the efficiency of lag-augmented LP relative to (i) non-augmented AR, (ii) lag-augmented AR, and (iii) non-augmented LP.

For analytical tractability, we restrict attention to a stationary, homoskedastic AR(1) model and to a fixed impulse response horizon h. Specifically, we here assume the AR(1) model (1) with $\rho \in (-1, 1)$ and where the innovations u_t are assumed to be i.i.d. with variance σ^2. This provides useful intuition, even though the main purpose of this paper is to develop methods that work in empirically realistic settings with several variables/lags, high persistence, and longer horizons.

Comparison with non-augmented AR. In a stationary and homoskedastic AR(1) model, the non-augmented AR estimator is the most asymptotically efficient estimator among all regular estimators that are consistent also under heteroskedasticity. This follows from standard semiparametric efficiency arguments, since the non-augmented AR estimator simply plugs the semiparametrically efficient OLS estimator of ρ into the smooth impulse response transformation ρ^h. In particular, non-augmented AR is weakly more efficient than (i) lag-augmented AR, (ii) non-augmented LP, and (iii) lag-augmented LP. As we have discussed in Section 3, however, standard non-augmented AR inference methods perform poorly in situations outside of the benign stationary, short-horizon case.

To gain intuition about the efficiency loss associated with lag augmentation, consider the first horizon $h = 1$. At this horizon, the lag-augmented LP and lag-augmented AR estimators coincide. These estimators regress y_{t+1} on y_t, while controlling for y_{t-1}. As discussed in Section 2.1, this is the same as regressing y_{t+1} directly on the innovation u_t, while controlling for y_{t-1} (which is uncorrelated with u_t). In contrast, the non-augmented AR estimator just regresses y_{t+1} on y_t without controls. Note that (i) the regressor y_t has a higher variance than the regressor u_t, and (ii) the residual in both the augmented and non-augmented regressions equals u_{t+1}. Thus, the usual homoskedastic OLS asymptotic variance formula implies that the non-augmented AR estimator is more efficient than the lag-augmented AR/LP estimator.

Comparison with lag-augmented AR. The relative efficiency of the lag-augmented AR and lag-augmented LP impulse response estimators is ambiguous. In the homoskedastic

AR(1) model, the proof of Proposition 1 implies that the asymptotic variance of the lag-augmented LP estimator $\hat{\beta}(h)$ is

$$\text{AsyVar}_\rho(\hat{\beta}(h)) = \frac{E[u_t^2 \xi_t(\rho, h)^2]}{[E(u_t^2)]^2} = \frac{\sigma^2 E[\xi_t(\rho, h)^2]}{\sigma^4} = \frac{\sigma^2 \sum_{\ell=0}^{h-1} \rho^{2\ell} \sigma^2}{\sigma^4} = \sum_{\ell=0}^{h-1} \rho^{2\ell}. \tag{11}$$

We want to compare this to the asymptotic variance of the plug-in AR estimator $\hat{\beta}_{\text{ARLA}}(h) \equiv \hat{\rho}_{\text{LA}}^h$, where $\hat{\rho}_{\text{LA}}$ is the AR(1) coefficient estimate obtained from a regression with *two* lags (Inoue and Kilian, 2020). Note that $\hat{\rho}_{\text{LA}} = \hat{\beta}(1)$ by definition. By the delta method, the asymptotic variance of $\hat{\beta}_{\text{ARLA}}(h)$ is given by

$$\text{AsyVar}_\rho(\hat{\beta}_{\text{ARLA}}(h)) = (h\rho^{h-1})^2 \times \text{AsyVar}_\rho(\hat{\rho}_{\text{LA}}) = (h\rho^{h-1})^2 \times \text{AsyVar}_\rho(\hat{\beta}(1)) = (h\rho^{h-1})^2.$$

To rank the LP and ARLA estimators in terms of asymptotic variance, note that

$$\text{AsyVar}_\rho(\hat{\beta}(h)) \le \text{AsyVar}_\rho(\hat{\beta}_{\text{ARLA}}(h)) \iff \sum_{\ell=0}^{h-1} \rho^{2(\ell-h+1)} \le h^2 \iff \sum_{m=0}^{h-1} \rho^{-2m} \le h^2.$$

Consider the inequality on the far right of the above display. For $h \ge 2$, the left-hand side is monotonically decreasing from ∞ to h as $|\rho|$ goes from 0 to 1. Hence, there exists an indifference function $\underline{\rho}\colon \mathbb{N} \to (0, 1)$ such that

$$\text{AsyVar}_\rho(\hat{\beta}(h)) \le \text{AsyVar}_\rho(\hat{\beta}_{\text{ARLA}}(h)) \iff |\rho| \ge \underline{\rho}(h).$$

Figure 2 plots the indifference curve between lag-augmented LP standard errors and lag-augmented AR standard errors. Lag-augmented LP is more efficient than lag-augmented AR in the north-west region of the plot, i.e., whenever $|\rho|$ is large enough and the horizon h is small enough. We draw the following conclusions. For $|\rho| \le 0.57$, lag-augmented AR is preferred to LP at all horizons $h \ge 2$ (recall that the two methods coincide at $h = 1$). For $|\rho| = 0.8$, LP is preferred to lag-augmented AR inference at horizons $h \le 8$; for $|\rho| = 0.9$, LP is preferred at horizons $h \le 22$. Of course, when the horizon h is large relative to the sample size, the delta method approximation for lag-augmented AR inference breaks down, as shown in Appendix A.2.1.

Comparison with non-augmented LP. The non-augmented LP estimator $\hat{\beta}_{\text{LPNA}}(h)$ is obtained from a regression of y_{t+h} on y_t without controls. As is clear from the representation

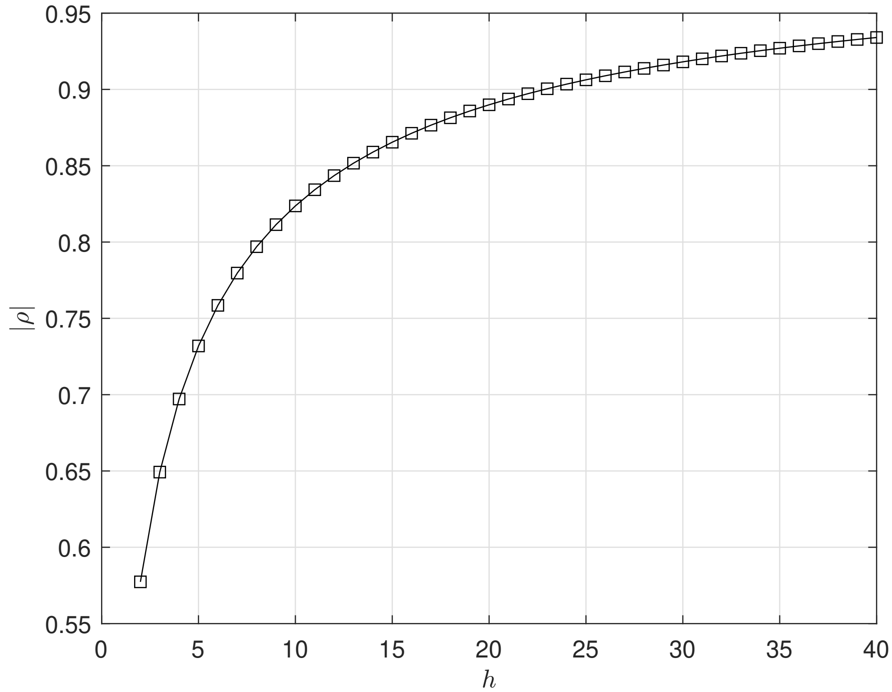

Figure 2: Indifference function $\underline{\rho}(h)$ for $h = 2, 3, \ldots, 40$. Above the curve, lag-augmented LP yields smaller asymptotic variance than lag-augmented AR. Below the curve, lag-augmented AR is preferable to lag-augmented LP.

(2), the asymptotic variance of this estimator is given by

$$\begin{aligned}
\text{AsyVar}_\rho(\hat{\beta}_{\text{LPNA}}(h)) &= \frac{\sum_{\ell=-\infty}^{\infty} E[y_t \xi_t(\rho, h) y_{t-\ell} \xi_{t-\ell}(\rho, h)]}{[E(y_t^2)]^2} \\
&= \frac{\sum_{\ell=-h+1}^{h-1} E[y_t \xi_t(\rho, h) y_{t-\ell} \xi_{t-\ell}(\rho, h)]}{[E(y_t^2)]^2} \\
&= \frac{\sum_{\ell=-h+1}^{h-1} \rho^{|\ell|} E[y_{t-|\ell|}^2] E[\xi_t(\rho, h) \xi_{t-|\ell|}(\rho, h)]}{[E(y_t^2)]^2} \\
&= \frac{\sum_{\ell=-h+1}^{h-1} \rho^{|\ell|} \sum_{m=1}^{h-|\ell|} \rho^{2(h-m)-|\ell|} \sigma^2}{E(y_t^2)} \\
&= \frac{\sum_{\ell=-h+1}^{h-1} \sum_{m=|\ell|}^{h-1} \rho^{2m}}{E(y_t^2)/\sigma^2} \\
&= (1 - \rho^2) \sum_{\ell=-h+1}^{h-1} \sum_{m=|\ell|}^{h-1} \rho^{2m}
\end{aligned}$$

$$= \sum_{\ell=-h+1}^{h-1} (\rho^{2|\ell|} - \rho^{2h})$$

$$= \sum_{\ell=0}^{h-1} \rho^{2\ell} + \sum_{\ell=1}^{h-1} \rho^{2\ell} - (2h-1)\rho^{2h}.$$

Thus, using (11), we find that

$$\text{AsyVar}_\rho(\hat{\beta}(h)) \leq \text{AsyVar}_\rho(\hat{\beta}_{\text{LPNA}}(h)) \iff \sum_{\ell=1}^{h-1} \rho^{2\ell} \geq (2h-1)\rho^{2h} \iff \sum_{\ell=1}^{h-1} \rho^{-2\ell} \geq (2h-1).$$

The last equivalence assumes $\rho \neq 0$, since lag-augmented and non-augmented LP are clearly equally efficient when $\rho = 0$. For $h = 1$, the last inequality above is never satisfied. This is because at this horizon lag-augmented and non-augmented LP reduce to lag-augmented and non-augmented AR, respectively, and the latter is more efficient, as discussed previously. For $h \geq 2$, the left-hand side of the last inequality above decreases monotonically from ∞ to $h-1$ as $|\rho|$ goes from 0 to 1. Thus, there exists an indifference function $\overline{\rho}\colon \mathbb{N} \to (0,1)$ such that

$$\text{AsyVar}_\rho(\hat{\beta}(h)) \leq \text{AsyVar}_\rho(\hat{\beta}_{\text{LPNA}}(h)) \iff |\rho| \leq \overline{\rho}(h).$$

Figure 3 plots this indifference curve between lag-augmented LP and non-augmented LP. The former is more efficient than the latter in the south-east region of the plot, i.e., whenever $|\rho|$ is sufficiently small and the horizon h is large enough. For any $h \geq 2$, lag-augmentation leads to smaller standard errors for all $|\rho| \leq 0.57$; for any $h \geq 5$, lag-augmentation leads to smaller standard errors for all $|\rho| \leq 0.85$. In addition to these efficiency comparisons in stationary cases, recall that non-augmented LP inference is less robust to near-unit roots than lag-augmented LP, as discussed in Section 2.1.

A.2.3 Length of Lag-Augmented AR Bootstrap Confidence Interval

Let $Y^T \equiv (y_1, \dots, y_T)$ denote a sample of size T generated by the AR(1) model (1). Let P_ρ denote the distribution of the data when the autoregressive parameter equals ρ. Let $\hat{\rho}$ denote the lag-augmented autoregressive estimator of the parameter ρ based on the data Y^T (i.e., the first coefficient in an AR(2) regression). Let $\hat{\rho}^*$ be the corresponding lag-augmented autoregressive estimator based on a bootstrap sample. We use $\mathbb{P}^*(\cdot \mid Y^T)$ to denote the distribution of the bootstrap samples conditional on the data.

By the results in Inoue and Kilian (2020) we will assume that (i) $\hat{\rho}$ is uniformly consistent for $\rho \in [-1,1]$ and (ii) the law of $\sqrt{T}(\hat{\rho}^* - \hat{\rho}) \mid Y^T$ converges to $\mathcal{N}(0, \omega^2)$ (in probability).

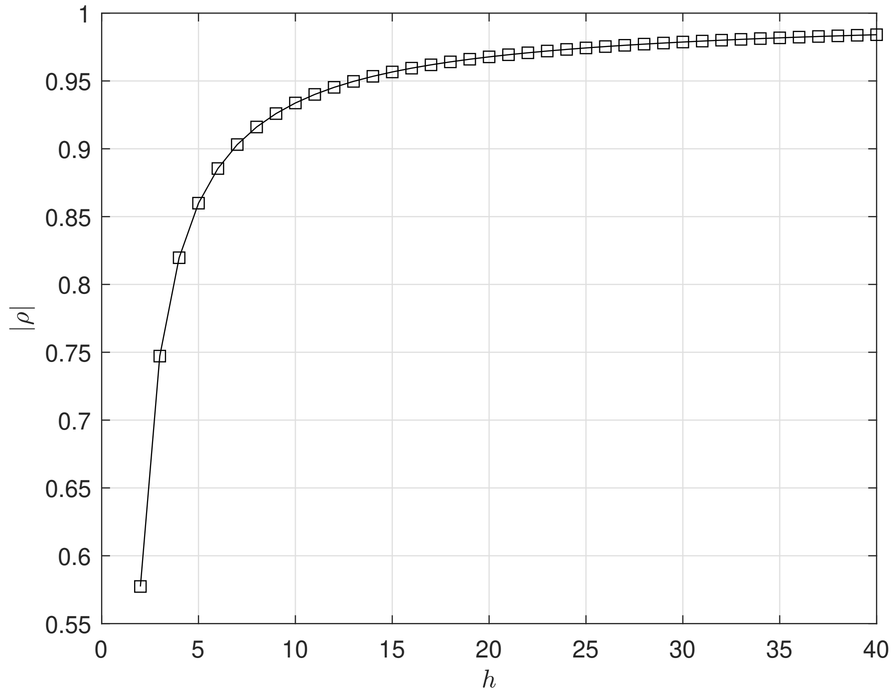

Figure 3: Indifference function $\overline{\rho}(h)$ for $h = 2, 3, \ldots, 40$. Below the curve, lag-augmented LP yields smaller asymptotic variance than non-augmented LP. Above the curve, non-augmented LP is preferable to lag-augmented LP.

We consider a sequence of autoregressive parameters $\{\rho_T\}$ approaching unity as $T \to \infty$, and a sequence of horizons $\{h_T\}$ that increases with the sample size. The restrictions on these sequences are as follows:

$$h_T(1 - \rho_T) \to a \in [0, \infty), \tag{12}$$

$$h_T \propto T^\eta, \quad \eta \in [1/2, 1]. \tag{13}$$

For example, these assumptions cover the cases of (i) local-to-unity DGPs $\rho_T = 1 - a/T$, $a > 0$, at long horizons $h_T \propto T$, and (ii) not-particularly-local-to-unity DGPs $\rho_T = 1 - a/\sqrt{T}$, $a > 0$, at medium-long horizons $h_T \propto \sqrt{T}$.

We now derive an expression for the quantiles of the bootstrap distribution of the impulse response estimates. For any $c \in \mathbb{R}$,

$$\mathbb{P}^*((\hat{\rho}^*)^{h_T} \leq c \mid Y^T) \;\; = \;\; \mathbb{P}^*((\hat{\rho}^*)^{h_T} \leq c \text{ and } \hat{\rho}^* \geq 0 \mid Y_T) + o_{P_{\rho_T}}(1),$$

$$
\begin{aligned}
&= \mathbb{P}^*(\hat{\rho}^* \leq c^{1/h_T} \mid Y^T) + o_{P_{\rho_T}}(1), \\
&\quad \text{(since } \mathbb{P}^*(\hat{\rho}^* \geq 0 \mid Y^T) \approx 1 \text{ when } (\rho_T, h_T) \text{ satisfy (12)–(13))}, \\
&= \mathbb{P}^*(\sqrt{T}(\hat{\rho}^* - \hat{\rho}) \leq \sqrt{T}(c^{1/h_T} - \hat{\rho}) \mid Y^T) + o_{P_{\rho_T}}(1).
\end{aligned}
$$

The equation above implies that the $1-\alpha$ bootstrap quantile of $(\hat{\rho}^*)^{h_T}$, which we denote $c^*_{1-\alpha}$, can be written as

$$
c^*_{1-\alpha} = \Big(\hat{\rho} + (\omega z_{1-\alpha}/\sqrt{T})\Big)^{h_T} + o_{P_{\rho_T}}(1), \tag{14}
$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. Thus, the Inoue and Kilian (2020) nominal $1-\alpha$ Efron bootstrap confidence interval is

$$
\left[(\hat{\rho})^{h_T}\left(1 + \frac{\omega z_{\alpha/2}}{\sqrt{T}\hat{\rho}}\right)^{h_T} + o_{P_{\rho_T}}(1)\ ,\ (\hat{\rho})^{h_T}\left(1 + \frac{\omega z_{1-\alpha/2}}{\sqrt{T}\hat{\rho}}\right)^{h_T} + o_{P_{\rho_T}}(1)\right]. \tag{15}
$$

Its length is given by

$$
(\hat{\rho})^{h_T}\left[\left(1 + \frac{\omega z_{1-\alpha/2}}{\sqrt{T}\hat{\rho}}\right)^{h_T} - \left(1 + \frac{\omega z_{\alpha/2}}{\sqrt{T}\hat{\rho}}\right)^{h_T}\right] + o_{P_{\rho_T}}(1). \tag{16}
$$

We now analyze the shape of this confidence interval and its length under two different kinds of sequences h_T.

Case 1: $h_T \propto T^\eta, \eta \in (1/2, 1]$. We analyze the lower and upper ends points—denoted $\hat{\ell}$ and $\hat{u}$ respectively—of the interval (16). A Taylor expansion of the logarithm function around 1 shows that

$$
\begin{aligned}
\hat{\ell} &= \exp\left(\frac{h_T}{\sqrt{T}}\left(\sqrt{T}\log(\hat{\rho}) + \frac{\omega z_{\alpha/2}}{\hat{\rho}}\right) - \frac{h_T}{2T}\left(\frac{\omega z_{\alpha/2}}{\hat{\rho}}\right)^2 + O_{P_{\rho_T}}(1/\sqrt{T})\right), \\
\hat{u} &= \exp\left(\frac{h_T}{\sqrt{T}}\left(\sqrt{T}\log(\hat{\rho}) + \frac{\omega z_{1-\alpha/2}}{\hat{\rho}}\right) - \frac{h_T}{2T}\left(\frac{\omega z_{1-\alpha/2}}{\hat{\rho}}\right)^2 + O_{P_{\rho_T}}(1/\sqrt{T})\right).
\end{aligned}
$$

Since

$$
\sqrt{T}(\log(\hat{\rho}) - \log(\hat{\rho})) \xrightarrow{d} Z \equiv \mathcal{N}(0, \omega^2),
$$

and by equation (12),

$$
h_T \log(\hat{\rho}) = h_T \log((\hat{\rho} - 1) + 1) \to -a,
$$

then for any $\epsilon > 0$ and T large enough

$$P_{\rho_T}(\hat{\ell} < \epsilon \text{ and } \hat{u} > 1 + \epsilon) \quad = \quad P(Z + \omega z_{\alpha/2} < 0 \text{ and } Z + \omega z_{1-\alpha/2} > 0) + o(1).$$

Consequently,

$$P_{\rho_T}(\, [\epsilon, 1+\epsilon] \subseteq [\hat{\ell}, \hat{u}] \,) = P_{\rho_T}(\hat{\ell} \leq \epsilon \text{ and } \hat{u} \geq 1 + \epsilon) \to 1 - \alpha.$$

This means that the lag-augmented autoregressive bootstrap of Inoue and Kilian (2020) has correct coverage, but it does so at the expense of reporting—with probability $1 - \alpha$—confidence intervals that are almost as large as the $[0, 1]$ interval (which is half of the parameter space for the impulse response of interest). In contrast, as long as $\eta < 1$, the lag-augmented LP confidence interval has valid coverage and length that tends to zero in probability asymptotically.

Case 2: $h_T = \kappa\sqrt{T}, \kappa \in (0, 1]$. An analogous argument to one made above shows that

$$P_{\rho_T}\left((\rho_T)^{h_T} \in [\hat{\ell}, \hat{u}]\right) \to 1 - \alpha.$$

If we restrict the upper and lower bounds $\hat{u}$ and $\hat{\ell}$ to be smaller than one (i.e., intersect the confidence interval with the parameter space $[-1, 1]$), then

$$(\hat{u} - \hat{\ell}) \quad \overset{d}{\to} \quad \min\left\{\exp\left(c(Z + \omega z_{1-\alpha/2} - a)\right), 1\right\} - \min\left\{\exp\left(c(Z + \omega z_{\alpha/2} - a)\right), 1\right\},$$

where a is given in equation (12). This means that the lag-augmented autoregressive bootstrap of Inoue and Kilian (2020) has correct coverage, but its length remains random in large samples. Again, in this case, the length of the lag-augmented LP confidence interval shrinks to zero in probability asymptotically.

B Proofs and Auxiliary Lemmas

We prove Proposition 1 in Appendix B.2, which in turn relies on several auxiliary lemmas stated in Appendix B.3.

B.1 Notation for proofs

We first define additional notation. Define, for any $\rho \in [-1, 1]$ and $h \in \mathbb{N}$,

$$v(\rho, h) \equiv \frac{\{E[\xi_t(\rho, h)^2 u_t^2]\}^{1/2}}{\sigma^2} \tag{17}$$

where we remind the reader that u_t (with mean zero and variance σ^2) is the innovation in the AR(1) model in equation (1) and

$$\xi_t(\rho, h) \equiv \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell}$$

is the multi-step forecast error in equation (2). As we mentioned in the main body of the paper, the regression *score*

$$\xi_t(\rho, h) u_t,$$

will play a crucial role in deriving our asymptotic results.

We have defined the LP estimator as the coefficient on y_t in the regression of y_{t+h} on $x_t = (y_t, y_{t-1})'$. By the Frisch-Waugh theorem, we can also obtain the coefficient of interest by regressing y_{t+h} on the residuals from the regression of y_t on y_{t-1}:

$$\hat{\beta}(h) \equiv \frac{\sum_{t=1}^{T-h} y_{t+h} \hat{u}_t(h)}{\sum_{t=1}^{T-h} \hat{u}_t(h)^2}, \tag{18}$$

where we recall the definitions

$$\hat{u}_t(h) \equiv y_t - \hat{\rho}(h) y_{t-1}, \quad \hat{\rho}(h) \equiv \frac{\sum_{t=1}^{T-h} y_t y_{t-1}}{\sum_{t=1}^{T-h} y_{t-1}^2}.$$

Recall also from (4) that

$$y_{t+h} = \beta(\rho, h) u_t + \underbrace{\beta(\rho, h+1)}_{\equiv \eta(\rho, h)} y_{t-1} + \xi_t(\rho, h).$$

If we define the lag-augmented LP residuals $\hat{\xi}_t(h) \equiv y_{t+h} - \hat{\beta}(h)y_t - \hat{\gamma}(h)y_{t-1}$ as in equation (6), then

$$\begin{aligned}
\hat{\xi}_t(h) - \xi_t(\rho,h) &= (y_{t+h} - \hat{\beta}(h)y_t - \hat{\gamma}(h)y_{t-1}) - (y_{t+h} - \beta(\rho,h)u_t - \eta(\rho,h)y_{t-1}) \\
&= -\hat{\beta}(h)\underbrace{(y_t - \rho y_{t-1})}_{=u_t} - (\underbrace{\rho\hat{\beta}(h) + \hat{\gamma}(h)}_{\equiv\hat{\eta}(\rho,h)})y_{t-1} + \beta(\rho,h)u_t + \eta(\rho,h)y_{t-1} \\
&= [\beta(\rho,h) - \hat{\beta}(h)]u_t + [\eta(\rho,h) - \hat{\eta}(\rho,h)]y_{t-1}. \qquad (19)
\end{aligned}$$

Finally, geometric series of the form $\sum_{\ell=0}^{h-1}\rho^{2\ell}$ will show up repeatedly in the proofs. Observe that, for all $\rho \in [-1,1]$ and $h \in \mathbb{N}$,

$$1 \leq \sum_{\ell=0}^{h-1}\rho^{2\ell} \leq \min\left\{\frac{1}{1-\rho^2}, h\right\} \leq \min\left\{\frac{1}{1-|\rho|}, h\right\} \equiv g(\rho,h)^2.$$

Here we slightly abuse notation by defining $\frac{1}{1-1} = \infty$.

In the proofs below we simplify notation by omitting the subscript ρ (which indexes the data generating process) from expectations, variances, covariances, and so on.

B.2 Proof of Proposition 1

The autoregressive model in equation (1) implies the following linear regression model for the outcome variable y_{t+h}:

$$y_{t+h} = \beta(\rho,h)u_t + \underbrace{\beta(\rho,h+1)}_{\equiv\eta(\rho,h)}y_{t-1} + \xi_t(\rho,h),$$

cf. equation (4). Using the definition of the lag-augmented local projection estimator which is given in equation (18) above, we have

$$\begin{aligned}
\hat{\beta}(h) &= \frac{\sum_{t=1}^{T-h} y_{t+h}\hat{u}_t(h)}{\sum_{t=1}^{T-h}\hat{u}_t(h)^2} \\
&= \frac{\sum_{t=1}^{T-h}[\beta(\rho,h)u_t + \eta(\rho,h)y_{t-1} + \xi_t(\rho,h)]\hat{u}_t(h)}{\sum_{t=1}^{T-h}\hat{u}_t(h)^2} \\
&\quad \text{(by equation (4))} \\
&= \frac{\sum_{t=1}^{T-h}[\beta(\rho,h)u_t + \xi_t(\rho,h)]\hat{u}_t(h)}{\sum_{t=1}^{T-h}\hat{u}_t(h)^2} \\
&\quad \text{(because } \textstyle\sum_{t=1}^{T-h} y_{t-1}\hat{u}_t(h) = 0 \text{ by definition of } \hat{u}_t(h))
\end{aligned}$$

$$= \beta(\rho, h) + \frac{\sum_{t=1}^{T-h}[\beta(\rho, h)(u_t - \hat{u}_t(h)) + \xi_t(\rho, h)]\hat{u}_t(h)}{\sum_{t=1}^{T-h} \hat{u}_t(h)^2}$$

$$= \beta(\rho, h) + \frac{\sum_{t=1}^{T-h} \xi_t(\rho, h)\hat{u}_t(h)}{\sum_{t=1}^{T-h} \hat{u}_t(h)^2},$$

where the last equality uses $u_t - \hat{u}_t(h) = (\hat{\rho}(h) - \rho)y_{t-1}$ and again $\sum_{t=1}^{T-h} y_{t-1}\hat{u}_t(h) = 0$ by definition of $\hat{u}_t(h)$. Consequently

$$\frac{\hat{\beta}(h) - \beta(\rho, h)}{\hat{s}(h)} = \frac{\sum_{t=1}^{T-h} \xi_t(\rho, h)\hat{u}_t(h)}{(\sum_{t=1}^{T-h} \hat{\xi}_t(h)^2\hat{u}_t(h)^2)^{1/2}}$$

$$= \left(\frac{\sum_{t=1}^{T-h} \xi_t(\rho, h)u_t}{\sigma^2(T-h)^{1/2}v(\rho, h)} + \frac{\sum_{t=1}^{T-h} \xi_t(\rho, h)[\hat{u}_t(h) - u_t]}{\sigma^2(T-h)^{1/2}v(\rho, h)}\right)$$

$$\times \frac{\sigma^2(T-h)^{1/2}v(\rho, h)}{(\sum_{t=1}^{T-h} \hat{\xi}_t(h)^2\hat{u}_t(h)^2)^{1/2}}.$$

Using the general subsequence approach of Andrews et al. (2019), both statements (i) and (ii) of the proposition follow if we can show the following: For any sequence $\{\rho_T\}$ of real numbers satisfying $\rho_T \in [-1 + a, 1]$ for all T, and for any sequence $\{h_T\}$ of nonnegative integers satisfying $h_T \leq (1-a)T$ for all T and $g(\rho_T, h_T)^2/T \to 0$, there exists a subsequence $\{k_T\}$ of $\{T\}$ such that:

i) $\frac{\sum_{t=1}^{k_T - h_{k_T}} \xi_t(\rho_{k_T}, h_{k_T})u_t}{\sigma^2(k_T - h_{k_T})^{1/2}v(\rho_{k_T}, h_{k_T})} \underset{P_{\rho_{k_T}}}{\overset{d}{\to}} N(0, 1).$

ii) $\frac{\sum_{t=1}^{k_T - h_{k_T}} \hat{\xi}_t(h_{k_T})^2\hat{u}_t(h_{k_T})^2}{\sigma^4(k_T - h_{k_T})v(\rho_{k_T}, h_{k_T})^2} \underset{P_{\rho_{k_T}}}{\overset{p}{\to}} 1.$

iii) $\frac{\sum_{t=1}^{k_T - h_{k_T}} \xi_t(\rho_{k_T}, h_{k_T})[\hat{u}_t(h_{k_T}) - u_t]}{(k_T - h_{k_T})^{1/2}v(\rho_{k_T}, h_{k_T})} \underset{P_{\rho_{k_T}}}{\overset{p}{\to}} 0.$

Observe that $g(\rho_T, h_T)^2/T \to 0$ is implied by either (i) $|\rho_T| \leq 1-a$ for all T, or (ii) $h_T/T \to 0$.

Result (i) is a central limit theorem for the (standardized) process $\xi_t(\rho, h)u_t$ (which we refer to as *regression score*). We establish this result in Lemma 1 below, relying on intermediate results obtained in Lemma 2 (consistency of the sample variance of $\xi_t(\rho, h)u_t$), Lemma 3 (bounds on the fourth moments of $\xi_t(\rho, h)u_t$ and $\xi_t(\rho, h)$), and Lemma 4 (lower and upper bounds on $E[\xi_t(\rho, h)^2u_t^2]$).

Result (ii) follows from Lemma 5 below (consistency of the sample variance of $\hat{\xi}_t(p, h)\hat{u}_t$), which shows that the variance of $\xi_t(\rho, h)u_t$ can be estimated by the numerator of the Eicker-Huber-White standard errors. The lemma relies on the intermediate results in Lemma 6 (negligibility of estimation error in $\hat{\xi}_t(h)$), Lemma 7 (negligibility of estimation error in $\hat{u}_t(h)$), Lemma 8 (bound on $E[y_t^4]$), Lemma 9 (uniform convergence rates of OLS estimators), Lemma 10 (uniform convergence rate of the OLS numerators), and Lemma 11 (convergence rate of the OLS denominators in the stationary case).

Result (iii) allows us to ignore the term

$$\frac{\left|\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T)[\hat{u}_t(h_T) - u_t]\right|}{(T-h_T)^{1/2} v(\rho_T, h_T)},$$

To establish this result, apply Cauchy-Schwarz to get the upper bound

$$\left(\frac{\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T)^4}{(T-h_T) v(\rho_T, h_T)^4}\right)^{1/4} \times \left(\frac{\sum_{t=1}^{T-h_T} [\hat{u}_t(h_T) - u_t]^4}{(T-h_T) v(\rho_T, h_T)^4}\right)^{1/4}.$$

The first fraction above is uniformly bounded in probability by a simple application of Markov's inequality and the bounds for the fourth moments derived in Lemma 3 below. The second fraction above tends to zero in probability along some subsequence of $\{T\}$ by Lemma 7 below. This gives the desired result. □

B.3 Auxiliary lemmas

We now state and prove a series of auxiliary lemmas that are used in the proof of in Proposition 1.

Lemma 1 (Central limit theorem for $\xi_t(\rho, h)u_t$). *Let Assumptions 1 and 2 hold. Let $\{\rho_T\}$ be a sequence of real numbers satisfying $\rho_T \in [-1, 1]$ for all T, and let $\{h_T\}$ be a sequence of nonnegative integers satisfying $T - h_T \to \infty$ and $g(\rho_T, h_T)^2/T \to 0$. Then*

$$\frac{\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T) u_t}{\sigma^2 (T-h_T)^{1/2} v(\rho_T, h_T)} \underset{P_{\rho_T}}{\overset{d}{\to}} N(0, 1).$$

Proof. The definition of the multi-step forecast error implies

$$\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T) u_t = \sum_{t=1}^{T-h_T} (\rho_T^{h_T-1} u_{t+1} + \rho_T^{h_T-2} u_{t+2} + \ldots + u_{t+h_T}) u_t. \tag{20}$$

Note that the summands above do not form a martingale difference sequence with respect to a conventionally defined filtration of the form $\sigma(u_{t+h_T}, u_{t+h_T-1}, u_{t+h_T-2}, \ldots)$, even if $\{u_t\}$ is i.i.d. Instead, we will define a process that "reverses time". For any T and any time period

$1 \leq t \leq T - h_T$, define the triangular array and filtration

$$\chi_{T,t} = \frac{\xi_{T-h_T+1-t}(\rho_T, h_T) u_{T-h_T+1-t}}{\sigma^2 (T-h)^{1/2} v(\rho_T, h_T),},$$
$$\mathcal{F}_{T,t} = \sigma(u_{T-h_T+1-t}, u_{T-h_T+2-t}, \ldots).$$

We say that we have reversed time because $\chi_{T,1}$ corresponds to the (scaled) last term that appears in the summation (20); the term $\chi_{T,2}$ to the second-to-last term, and so on. By reversing time we have achieved three things. First, the sequence of σ-algebras is a *filtration*:

$$\mathcal{F}_{T,1} \subseteq \mathcal{F}_{T,2} \subseteq \ldots \subseteq \mathcal{F}_{T,T-h_T}.$$

Second, the process $\{\chi_{T,t}\}$ is adapted to the filtration $\{\mathcal{F}_{T,t}\}$, as $\chi_{T,t}$ is measurable with respect to $\mathcal{F}_{T,t}$ for all t. Third, the pair $\{\chi_{T,t}, \mathcal{F}_{T,t}\}$ form a martingale difference array:

$$\begin{aligned} E[\chi_{T,t} \mid \mathcal{F}_{T,t-1}] &\propto E[(\rho_T^{h_T-1} u_{T-h_T+2-t} \ldots + u_{T+1-t}) u_{T-h_T+1-t} \mid u_{T-h_T+2-t}, u_{T-h_T+3-t}, \ldots] \\ &= (\rho_T^{h_T-1} u_{T-h_T+2-t} \ldots + u_{T+1-t}) E[u_{T-h_T+1-t} \mid u_{T-h_T+2-t}, u_{T-h_T+3-t}, \ldots] \\ &= 0, \end{aligned}$$

where the last equality follows from Assumption 1.

Thus, we can apply the martingale central limit theorem in Davidson (1994, Thm. 24.3) to show that

$$\sum_{t=1}^{T-h_T} \chi_{T,t} \xrightarrow{d} N(0,1),$$

which is the statement of the lemma. We now verify the conditions of this theorem. First, by definition of $v(\rho, h)$,

$$\sum_{t=1}^{T-h_T} E[\chi_{T,t}^2] = 1.$$

Second, in Lemma 2 below we show (by means of Chebyshev's inequality)

$$\sum_{t=1}^{T-h_T} \chi_{T,t}^2 = \frac{\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T)^2 u_t^2}{\sigma^4 (T-h_T) v(\rho, h_T)^2} \xrightarrow{p} 1.$$

Finally, we argue that $\max_{1 \leq t \leq T-h_T} |\chi_{T,t}(\rho_T, h_T)| \xrightarrow{p} 0$. By Davidson (1994, Thm. 23.16), it is sufficient to prove that, for arbitrary $c > 0$, we have

$$(T-h_T)E\left[\chi_{T,t}^2 \mathbb{1}(|\chi_{T,t}|>c)\right] \to 0.$$

Indeed,

$$\begin{aligned}
&(T-h_T)E\left[\chi_{T,t}^2 \mathbb{1}(|\chi_{T,t}|>c)\right] \\
&\le (T-h_T)E\left[\chi_{T,t}^2 \mathbb{1}(|\chi_{T,t}|>c) \times \frac{\chi_{T,t}^2}{c^2}\right] \\
&= (T-h_T)\frac{E[\chi_{T,t}^4]}{c^2} \\
&= \frac{1}{(T-h_T)\sigma^8 c^2} E\left[\left|v(\rho_T,h_T)^{-1}\xi_{T-h_T+1-t}(\rho_T,h_T)u_{T-h_T+1-t}\right|^4\right] \\
&\le \frac{6E(u_t^8)}{(T-h_T)\sigma^4\delta^2 c^2},
\end{aligned}$$

where the last inequality uses Lemma 3 below (recall that δ is the constant in Assumption 2(i)). The right-hand side tends to zero as $T \to \infty$, as required. □

Lemma 2 (Consistency of the sample variance of $\xi_t(\rho,h)u_t$)**.** *Let the conditions of Lemma 1 hold. Then*

$$\frac{\sum_{t=1}^{T-h_T}\xi_t(\rho_T,h_T)^2 u_t^2}{\sigma^4 (T-h_T) v(\rho,h_T)^2} \xrightarrow{p} 1.$$

Proof. We would like to show $\hat{\varsigma} \xrightarrow{p} 1$, where

$$\hat{\varsigma} \equiv \frac{1}{T-h_T}\sum_{t=1}^{T-h_T}\frac{\xi_t(\rho_T,h_T)^2 u_t^2}{v(\rho_T,h_T)^2\sigma^4}.$$

Note that the summands could be serially correlated under our assumptions. We establish the desired convergence in probability by showing that the variance of $\hat{\varsigma}$ tends to 0 (since its mean is 1). Observe that

$$\begin{aligned}
\mathrm{Var}(\hat{\varsigma}) &= \frac{1}{(T-h_T)^2 v(\rho_T,h_T)^4\sigma^8}\sum_{t=1}^{T-h_T}\sum_{s=1}^{T-h_T}\mathrm{Cov}\left(\xi_t(\rho_T,h_T)^2u_t^2, \xi_s(\rho_T,h_T)^2u_s^2\right) \\
&= \frac{1}{(T-h_T)v(\rho_T,h_T)^4\sigma^8}\sum_{|m|<T-h_T}\left(1-\frac{|m|}{T-h_T}\right)\mathrm{Cov}\left(\xi_0(\rho_T,h_T)^2u_0^2, \xi_m(\rho_T,h_T)^2u_m^2\right) \\
&\le \frac{2}{(T-h_T)v(\rho_T,h_T)^4\sigma^8}\sum_{m=0}^{T-h_T}|\Gamma_T(m)|, \qquad (21)
\end{aligned}$$

where we define

$$\Gamma_T(m) \equiv \mathrm{Cov}\left(\xi_0(\rho_T, h_T)^2 u_0^2, \xi_m(\rho_T, h_T)^2 u_m^2\right), \quad m = 0, 1, 2, \ldots$$

By expanding the squares $\xi_0(\rho, h)^2$ and $\xi_m(\rho, h)^2$, we obtain

$$\Gamma_T(m) = \sum_{\ell_1=1}^{h_T} \sum_{\ell_2=1}^{h_T} \sum_{\ell_3=1}^{h_T} \sum_{\ell_4=1}^{h_T} \rho_T^{4h_T-\ell_1-\ell_2-\ell_3-\ell_4} \mathrm{Cov}\left(u_{\ell_1} u_{\ell_2} u_0^2, u_{m+\ell_3} u_{m+\ell_4} u_m^2\right).$$

Consider any summand on the right-hand side above defined by indices $(\ell_1, \ell_2, \ell_3, \ell_4)$. If $\ell_1 = \ell_2$, then Assumption 1 implies that the covariance in the summand equals zero whenever $\ell_3 \neq \ell_4$, since in this case at most one of the subscripts $m + \ell_3$ or $m + \ell_4$ can equal ℓ_1 $(= \ell_2)$. Thus, if $\ell_1 = \ell_2$, then the summand can only be nonzero when $\ell_3 = \ell_4$. If instead $\ell_1 \neq \ell_2$, then Assumption 1 implies that the summand can only be nonzero when $\{\ell_1, \ell_2\} = \{m+\ell_3, m+\ell_4\}$, which in turn requires that $m < h_T$. Putting these facts together, we obtain

$$\begin{aligned}
|\Gamma_T(m)| &\leq \sum_{\ell_1=1}^{h_T} \sum_{\ell_3=1}^{h_T} |\rho_T|^{4h_T-2\ell_1-2\ell_3} \left|\mathrm{Cov}\left(u_{m+\ell_1}^2 u_m^2, u_{\ell_3}^2 u_0^2\right)\right| \\
&\quad + \mathbb{1}(m < h_T) 2 \sum_{\ell_1=1}^{h_T} \sum_{\ell_2 \neq \ell_1} |\rho_T|^{4h_T-\ell_1-\ell_2-(\ell_1-m)-(\ell_2-m)} \left|\mathrm{Cov}\left(u_{\ell_1} u_{\ell_2} u_m^2, u_{\ell_1} u_{\ell_2} u_0^2\right)\right| \\
&= \sum_{\ell_1=1}^{h_T} \sum_{\ell_3=1}^{h_T} |\rho_T|^{4h_T-2(\ell_1+\ell_3)} \left|\mathrm{Cov}\left(u_{m+\ell_1}^2 u_m^2, u_{\ell_3}^2 u_0^2\right)\right| && (22) \\
&\quad + \mathbb{1}(m < h_T) 2 \rho_T^{2m} \sum_{\ell_1=1}^{h_T} \sum_{\ell_2 \neq \ell_1} |\rho_T|^{4h_T-2(\ell_1+\ell_2)} \left|E\left(u_{\ell_1}^2 u_{\ell_2}^2 u_m^2 u_0^2\right)\right|, && (23)
\end{aligned}$$

where the last equality also uses Assumption 1. Let $\tilde{\Gamma}_{1,T}(m)$ and $\tilde{\Gamma}_{2,T}(m)$ denote expressions (22) and (23), respectively. We will now bound $\sum_{m=0}^{T-h_T} \tilde{\Gamma}_{1,T}(m)$ and $\sum_{m=0}^{T-h_T} \tilde{\Gamma}_{2,T}(m)$, so that we can ultimately insert these bounds into (21). First, since $|\rho_T| \leq 1$,

$$\begin{aligned}
\sum_{m=0}^{T-h_T} \tilde{\Gamma}_{1,T}(m) &\leq \sum_{m=0}^{T-h_T} \sum_{\ell_1=1}^{h_T} \sum_{\ell_3=1}^{h_T} \rho_T^{2(h_T-\ell_3)} \left|\mathrm{Cov}\left(u_{m+\ell_1}^2 u_m^2, u_{\ell_3}^2 u_0^2\right)\right| \\
&\leq \sum_{b_1=1}^{h_T} \rho_T^{2(h_T-b_1)} \sum_{b_2=-\infty}^{\infty} \sum_{b_3=-\infty}^{\infty} \left|\mathrm{Cov}\left(u_{b_1}^2 u_0^2, u_{b_2}^2 u_{b_3}^2\right)\right|. && (24)
\end{aligned}$$

According to Brillinger (2001, Thm. 2.3.2),

$$\operatorname{Cov}\left(u_0^2 u_{b_1}^2, u_{b_2}^2 u_{b_3}^2\right) = \operatorname{Cov}\left(u_0^2, u_{b_2}^2\right)\operatorname{Cov}\left(u_{b_1}^2, u_{b_3}^2\right) + \operatorname{Cov}\left(u_0^2, u_{b_3}^2\right)\operatorname{Cov}\left(u_{b_1}^2, u_{b_2}^2\right) \\ + \operatorname{Cum}\left(u_0^2, u_{b_1}^2, u_{b_2}^2, u_{b_3}^2\right),$$

where "Cum" denotes the joint fourth-order cumulant. Thus, the expression (24) is bounded above by

$$\begin{aligned} &\sum_{b_1=1}^{h_T} \rho_T^{2(h_T-b_1)} \sum_{b_2=-\infty}^{\infty} \sum_{b_3=-\infty}^{\infty} \left|\operatorname{Cov}\left(u_0^2, u_{b_2}^2\right)\operatorname{Cov}\left(u_{b_1}^2, u_{b_3}^2\right)\right| \\ &+ \sum_{b_1=1}^{h_T} \rho_T^{2(h_T-b_1)} \sum_{b_2=-\infty}^{\infty} \sum_{b_3=-\infty}^{\infty} \left|\operatorname{Cov}\left(u_0^2, u_{b_3}^2\right)\operatorname{Cov}\left(u_{b_1}^2, u_{b_2}^2\right)\right| \\ &+ \sum_{b_1=-\infty}^{\infty} \sum_{b_2=-\infty}^{\infty} \sum_{b_3=-\infty}^{\infty} \left|\operatorname{Cum}\left(u_0^2, u_{b_1}^2, u_{b_2}^2, u_{b_3}^2\right)\right|. \end{aligned}$$

The third term above is finite by Assumption 2(ii). The analysis of the first and second terms above is similar, so we focus on the first term. Note that

$$\begin{aligned} &\sum_{b_1=1}^{h_T} \rho_T^{2(h_T-b_1)} \sum_{b_2=-\infty}^{\infty} \sum_{b_3=-\infty}^{\infty} \left|\operatorname{Cov}\left(u_0^2, u_{b_2}^2\right)\operatorname{Cov}\left(u_{b_1}^2, u_{b_3}^2\right)\right| \\ &= \sum_{b_1=1}^{h_T} \rho_T^{2(h_T-b_1)} \sum_{b_2=-\infty}^{\infty} \left|\operatorname{Cov}\left(u_0^2, u_{b_2}^2\right)\right| \sum_{b_3=-\infty}^{\infty} \left|\operatorname{Cov}\left(u_{b_1}^2, u_{b_3}^2\right)\right| \\ &= \sum_{b_1=1}^{h_T} \rho_T^{2(h_T-b_1)} \sum_{b_2=-\infty}^{\infty} \left|\operatorname{Cov}\left(u_0^2, u_{b_2}^2\right)\right| \sum_{\ell=-\infty}^{\infty} \left|\operatorname{Cov}\left(u_0^2, u_\ell^2\right)\right| \\ &= \left(\sum_{\ell=0}^{h_T-1} \rho_T^{2\ell}\right)\left(\sum_{\ell=-\infty}^{\infty} \left|\operatorname{Cov}\left(u_0^2, u_\ell^2\right)\right|\right)^2, \end{aligned}$$

where the second equality uses stationarity of $\{u_t^2\}$. By Assumption 2(ii), the autocovariances of $\{u_t^2\}$ are absolutely summable, implying that the sum in the final expression above is finite. Thus, we have shown that there exist constants $K_1, K_2 \in (0, \infty)$ (which only depend on the fixed data generating process for $\{u_t\}$) such that

$$\sum_{m=0}^{T-h_T} \tilde{\Gamma}_{1,T}(m) \le K_1 + K_2 \sum_{\ell=0}^{h_T-1} \rho_T^{2\ell}. \tag{25}$$

Next, a straight-forward application of Cauchy-Schwarz gives

$$\begin{aligned}\sum_{m=0}^{T-h_T} \tilde{\Gamma}_{2,T}(m) &= 2\sum_{m=0}^{h_T-1} \rho_T^{2m} \sum_{\ell_1=1}^{h_T} \sum_{\ell_2\neq \ell_1} |\rho_T|^{4h_T-2(\ell_1+\ell_2)} \left| E\left(u_{\ell_1}^2 u_{\ell_2}^2 u_m^2 u_0^2\right)\right| \\ &\leq 2\sum_{m=0}^{h_T-1} \rho_T^{2m} \sum_{\ell_1=1}^{h_T} \sum_{\ell_2=1}^{h_T} |\rho_T|^{4h_T-2(\ell_1+\ell_2)} E(u_t^8) \\ &= 2E(u_t^8)\left(\sum_{\ell=0}^{h_T-1} \rho_T^{2\ell}\right)^3. \end{aligned} \tag{26}$$

Putting together (21), (22), (23), (25), and (26), we get

$$\begin{aligned}\mathrm{Var}(\hat{\varsigma}) &\leq \frac{2}{(T-h_T)v(\rho_T,h_T)^4\sigma^8}\left\{K_1 + K_2\sum_{\ell=0}^{h_T-1}\rho_T^{2\ell} + 2E(u_t^8)\left(\sum_{\ell=0}^{h_T-1}\rho_T^{2\ell}\right)^3\right\} \\ &\leq \frac{2}{(T-h_T)\left(\sum_{\ell=0}^{h_T-1}\rho_T^{2\ell}\right)^2\sigma^4\delta^2}\left\{K_1 + K_2\sum_{\ell=0}^{h_T-1}\rho_T^{2\ell} + 2E(u_t^8)\left(\sum_{\ell=0}^{h_T-1}\rho_T^{2\ell}\right)^3\right\} \\ &\leq \frac{2(K_1+K_2)}{(T-h_T)\sigma^4\delta^2} + \frac{4E(u_t^8)\sum_{\ell=0}^{h_T-1}\rho_T^{2\ell}}{(T-h_T)\sigma^4\delta^2},\end{aligned}$$

where the second inequality uses Lemma 4, and the last inequality uses $\sum_{\ell=0}^{h_T-1}\rho_T^{2\ell} \geq 1$. The final expression above tends to zero as $T\to\infty$, since $(T-h_T)\to\infty$ and

$$\frac{\sum_{\ell=0}^{h_T-1}\rho_T^{2\ell}}{T-h_T} \leq \frac{g(\rho_T,h_T)^2}{T-h_T} \to 0.$$

Thus, $\mathrm{Var}(\hat{\varsigma})\to 0$, as we had set out to prove. □

Lemma 3 (Bounds on the fourth moments of $\xi_t(\rho,h)u_t$ and $\xi_t(\rho,h)$)**.** *Let Assumption 1 and Assumption 2(i) hold. Then*

$$E\left[\left(v(\rho,h)^{-1}\xi_t(\rho,h)u_t\right)^4\right] \leq \frac{6E(u_t^8)\sigma^4}{\delta^2}$$

and

$$E\left[\left(v(\rho,h)^{-1}\xi_t(\rho,h)\right)^4\right] \leq \frac{6E(u_t^4)\sigma^4}{\delta^2}$$

for all $\rho\in[-1,1]$ and $h\in\mathbb{N}$.

Proof. We prove only the first statement of the lemma, as the proof is completely analogous for the second part. Expanding the four-fold product $\xi_t(\rho,h)^4$, we obtain

$$E[\xi_t(\rho,h)^4 u_t^4] = \sum_{\ell_1=1}^{h}\sum_{\ell_2=1}^{h}\sum_{\ell_3=1}^{h}\sum_{\ell_4=1}^{h} \rho^{4h-(\ell_1+\ell_2+\ell_3+\ell_4)} E(u_{t+\ell_1}u_{t+\ell_2}u_{t+\ell_3}u_{t+\ell_4}u_t^4). \tag{27}$$

By Assumption 1, the summands above equal zero if one of the indices ℓ_j is different from the three other indices. Hence, the only possibly nonzero summands are those for which the four indices appear in two pairs, e.g., $\ell_1 = \ell_3$ and $\ell_2 = \ell_4$. The typical nonzero summand can thus be written in the form $\rho^{4h-2(\ell+m)}E(u_{t+\ell}^2 u_{t+m}^2 u_t^4)$ where $\ell, m \in \{1,\dots,h\}$. For given ℓ and m, this specific type of summand is obtained precisely when either (i) $\ell_1 = \ell_2 = \ell$ and $\ell_3 = \ell_4 = m$, or (ii) $\ell_1 = \ell_3 = \ell$ and $\ell_2 = \ell_4 = m$, or (iii) $\ell_1 = \ell_4 = \ell$ and $\ell_2 = \ell_3 = m$, or (iv) $\ell_1 = \ell_2 = m$ and $\ell_3 = \ell_4 = \ell$, or (v) $\ell_1 = \ell_3 = m$ and $\ell_2 = \ell_4 = \ell$, or (vi) $\ell_1 = \ell_4 = m$ and $\ell_2 = \ell_3 = \ell$. That is, there are six summands in (27) of this form. Thus,

$$\begin{aligned}
E[\xi_t(\rho,h)^4 u_t^4] &= 6\sum_{\ell=1}^{h}\sum_{m=1}^{h} \rho^{4h-2(\ell+m)} E(u_{t+\ell}^2 u_{t+m}^2 u_t^4) \\
&\le 6E(u_t^8)\sum_{\ell=1}^{h}\sum_{m=1}^{h} \rho^{4h-2(\ell+m)} \\
&\quad \text{(by applying Cauchy-Schwarz twice)} \\
&= 6E(u_t^8)\sum_{\ell=1}^{h}\sum_{m=1}^{h} \rho^{2(h-\ell)}\rho^{2(h-m)} \\
&= 6E(u_t^8)\left(\sum_{\ell=1}^{h}\rho^{2(h-\ell)}\right)\left(\sum_{m=1}^{h}\rho^{2(h-m)}\right) \\
&= 6E(u_t^8)\left(\sum_{\ell=0}^{h-1}\rho^{2\ell}\right)^2.
\end{aligned}$$

It follows from Lemma 4 that

$$E\left[\left(v(\rho,h)^{-1}\xi_t(\rho,h)u_t\right)^4\right] \le \frac{6E(u_t^8)\sigma^4}{\delta^2},$$

as was to be shown. ∎

Lemma 4 (Bounds on $v(\rho,h)$)**.** *Let Assumption 1 and Assumption 2(i) hold. Then*

$$\frac{\delta}{\sigma^2} \le \frac{v(\rho,h)^2}{\sum_{\ell=0}^{h-1}\rho^{2\ell}} \le \frac{E[u_t^4]}{\sigma^4}$$

for all $\rho \in [-1,1]$ and $h \in \mathbb{N}$.

Proof. Algebra shows

$$\begin{aligned} v(\rho,h)^2 &= \frac{E[\xi_t(\rho,h)^2 u_t^2]}{\sigma^4} \\ &= \frac{1}{\sigma^4} E\left[(\rho^{h-1}u_{t+1} + \ldots + u_{t+h})^2 u_t^2\right] \\ &= \frac{1}{\sigma^4} E\left[\left(\sum_{\ell=1}^{h}\sum_{m=1}^{h} \rho^{2h-\ell-m} u_{t+\ell}u_{t+m}\right) u_t^2\right]. \end{aligned}$$

Assumption 1 implies that the last expression above equals

$$\frac{1}{\sigma^4}\sum_{\ell=1}^{h} \rho^{2(h-\ell)} E(u_{t+\ell}^2 u_t^2).$$

An application of Cauchy-Schwarz gives the upper bound

$$v(\rho,h)^2 = \frac{1}{\sigma^4}\sum_{\ell=1}^{h} \rho^{2(h-\ell)} E(u_{t+\ell}^2 u_t^2) \leq \frac{E[u_t^4]}{\sigma^4}\sum_{\ell=0}^{h-1} \rho^{2\ell}.$$

By Assumption 2(i),

$$E[u_{t+\ell}^2 u_t^2] = E\left[E(u_{t+\ell}^2 \mid \{u_s\}_{s<t+\ell}) u_t^2\right] \geq \delta E[u_t^2],$$

so we also obtain the lower bound

$$v(\rho,h)^2 = \frac{1}{\sigma^4}\sum_{\ell=1}^{h} \rho^{2(h-\ell)} E(u_{t+\ell}^2 u_t^2) \geq \frac{\delta E(u_t^2)}{\sigma^4}\sum_{\ell=0}^{h-1} \rho^{2\ell}.$$

This concludes the proof. □

Lemma 5 (Consistency of standard errors)**.** *Let Assumptions 1 and 2 hold. Let $a > 0$. Let $\{\rho_T\}$ be a sequence of real numbers satisfying $\rho_T \in [-1+a, 1]$ for all T, and let $\{h_T\}$ be a sequence of nonnegative integers satisfying $h_T \leq (1-a)T$ for all T and $g(\rho_T, h_T)^2/T \to 0$. Moreover, assume that either (i) Assumption 3 holds, or (ii) $|\rho_T| \leq 1-a$ for all T. Then there exists a subsequence $\{k_T\}$ of $\{T\}$ such that*

$$\frac{\sum_{t=1}^{k_T - h_{k_T}} \hat{\xi}_t(h_{k_T})^2 \hat{u}_t(h_{k_T})^2}{\sigma^4 (k_T - h_{k_T}) v(\rho_{k_T}, h_{k_T})^2} \xrightarrow[P_{\rho_{k_T}}]{p} 1.$$

Proof. The result follows from Lemma 2 if we can show that

$$\frac{\sum_{t=1}^{T-h_T} \hat{\xi}_t(h_T)^2 \hat{u}_t(h_T)^2 - \sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T)^2 u_t^2}{(T-h_T)v(\rho_T, h_T)^2} \xrightarrow[P_{\rho_T}]{p} 0,$$

along some subsequence of $\{T\}$. Algebra shows that

$$\begin{aligned}
&\frac{\left|\sum_{t=1}^{T-h_T} \left[\hat{\xi}_t(h_T)^2 \hat{u}_t(h_T)^2 - \xi_t(\rho_T, h_T)^2 u_t^2\right]\right|}{(T-h_T)v(\rho_T, h_T)^2} \\
&\leq \frac{\sum_{t=1}^{T-h_T} \left|\hat{\xi}_t(h_T)^2 \hat{u}_t(h_T)^2 - \xi_t(\rho_T, h_T)^2 u_t^2\right|}{(T-h_T)v(\rho_T, h_T)^2} \\
&= \frac{\sum_{t=1}^{T-h_T} \left|\hat{\xi}_t(h_T)\hat{u}_t(h_T) - \xi_t(\rho_T, h_T)u_t\right| \times \left|\hat{\xi}_t(h_T)\hat{u}_t(h_T) - \xi_t(\rho_T, h_T)u_t + 2\xi_t(\rho_T, h_T)u_t\right|}{(T-h_T)v(\rho_T, h_T)^2} \\
&\text{(as } (a+b)(a-b) = a^2 - b^2) \\
&\leq \left(\frac{\sum_{t=1}^{T-h_T} \left[\hat{\xi}_t(h_T)\hat{u}_t(h_T) - \xi_t(\rho_T, h_T)u_t\right]^2}{(T-h_T)v(\rho_T, h_T)^2}\right)^{1/2} \\
&\quad \times \left(\frac{\sum_{t=1}^{T-h_T} \left[\hat{\xi}_t(h_T)\hat{u}_t(h_T) - \xi_t(\rho_T, h_T)u_t + 2\xi_T(\rho_T, h_T)u_t)\right]^2}{(T-h_T)v(\rho_T, h_T)^2}\right)^{1/2},
\end{aligned}$$

where the last line follows from an application of Cauchy-Schwarz. Consider the expression in the last line above. By Loève's inequality (Davidson, 1994, Thm. 9.28)—which states that for any $r > 0$, $E\left[\left|\sum_{i=1}^m X_i\right|^r\right] \leq c_r \sum_{i=1}^m E[|X_i|^r]$, where $c_r = 1$ when $r \leq 1$ and $c_r = m^{r-1}$ when $r \geq 1$—this expression is bounded above by

$$\left(2\frac{\sum_{t=1}^{T-h_T} \left[\hat{\xi}_t(h_T)\hat{u}_t(h_T) - \xi_t(\rho_T, h_T)u_t\right]^2}{(T-h_T)v(\rho_T, h_T)^2} + 8\frac{\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T)^2 u_t^2}{(T-h_T)v(\rho_T, h_T)^2}\right)^{1/2}.$$

The last fraction above is bounded in probability along any subsequence by Lemma 2. Thus, it is sufficient to show that

$$\frac{\sum_{t=1}^{T-h_T} \left[\hat{\xi}_t(h_T)\hat{u}_t(h_T) - \xi_t(\rho_T, h_T)u_t\right]^2}{(T-h_T)v(\rho_T, h_T)^2}$$

converges in probability to zero along some subsequence.

Rewrite

$$\hat{\xi}_t(h_T)\hat{u}_t(h_T) - \xi_t(\rho_T, h_T)u_t$$

as the sum of the following three terms:

$$(\hat{\xi}_t(h_T) - \xi_t(\rho_T, h_T))u_t + (\hat{u}_t(h_T) - u_T)\xi_t(\rho_T, h_T) + (\hat{\xi}_t(h_T) - \xi_t(\rho_T, h_T))(\hat{u}_t(h_T) - u_T).$$

By another application of Loève's inequality,

$$\begin{aligned}
&\frac{\sum_{t=1}^{T-h_T} \left[\hat{\xi}_t(h_T)\hat{u}_t(h_T) - \xi_t(\rho_T, h_T)u_t\right]^2}{(T-h_T)v(\rho_T, h_T)^2} \\
&\leq 3\frac{\sum_{t=1}^{T-h_T}[\hat{\xi}_t(h_T) - \xi_t(\rho_T, h_T)]^2 u_t^2}{(T-h_T)v(\rho_T, h_T)^2} \\
&\quad + 3\frac{\sum_{t=1}^{T-h_T}[\hat{u}_t(h_T) - u_t]^2 \xi_t(\rho_T, h_T)^2}{(T-h_T)v(\rho_T, h_T)^2} \\
&\quad + 3\frac{\sum_{t=1}^{T-h_T}[\hat{\xi}_t(h_T) - \xi_t(\rho_T, h_T)]^2[\hat{u}_t(h_T) - u_t]^2}{(T-h_T)v(\rho_T, h_T)^2} \\
&\leq 3\left(\frac{\sum_{t=1}^{T-h_T}[\hat{\xi}_t(h_T) - \xi_t(\rho_T, h_T)]^4}{(T-h_T)v(\rho_T, h_T)^4}\right)^{1/2} \times \left(\frac{\sum_{t=1}^{T-h_T} u_t^4}{T-h_T}\right)^{1/2} \\
&\quad + 3\left(\frac{\sum_{t=1}^{T-h_T}[\hat{u}_t(h_T) - u_t]^4}{T-h_T}\right)^{1/2} \times \left(\frac{\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T)^4}{(T-h_T)v(\rho_T, h_T)^4}\right)^{1/2} \\
&\quad + 3\left(\frac{\sum_{t=1}^{T-h_T}[\hat{\xi}_t(h_T) - \xi_t(\rho_T, h_T)]^4}{(T-h_T)v(\rho_T, h_T)^4}\right)^{1/2} \times \left(\frac{\sum_{t=1}^{T-h_T}[\hat{u}_t(h_T) - u_t]^4}{T-h_T}\right)^{1/2} \\
&\text{(by Cauchy-Schwarz)} \\
&\equiv 3\left[(\hat{R}_1)^{1/2} \times (\hat{R}_2)^{1/2}\right] + 3\left[(\hat{R}_3)^{1/2} \times (\hat{R}_4)^{1/2}\right] + 3\left[(\hat{R}_1)^{1/2} \times (\hat{R}_3)^{1/2}\right].
\end{aligned}$$

It follows from Lemmas 6 and 7, respectively, that $\hat{R}_1$ and $\hat{R}_3$ both tend to zero in probability along some subsequence of $\{T\}$. $\hat{R}_2$ is bounded in probability by Assumption 2(i) and a standard application of Markov's inequality. Finally, another standard application of Markov's inequality combined with Lemma 3 implies that $\hat{R}_4$ is also uniformly bounded in probability. Hence, the entire expression above tends to zero in probability along some subsequence. □

Lemma 6 (Negligibility of estimation error in $\hat{\xi}_t(h)$)**.** *Let the conditions of Lemma 5 hold.*

Then there exists a subsequence $\{k_T\}$ of $\{T\}$ such that

$$\frac{\sum_{t=1}^{k_T-h_{k_T}}[\hat{\xi}_t(h_{k_T})-\xi_t(\rho_{k_T},h_{k_T})]^4}{(k_T-h_{k_T})v(\rho_{k_T},h_{k_T})^4} \underset{P_{\rho_{k_T}}}{\overset{p}{\to}} 0.$$

Proof. Recall equation (19):

$$\hat{\xi}_t(h)-\xi_t(\rho,h)=[\beta(\rho,h)-\hat{\beta}(h)]u_t+[\eta(\rho,h)-\hat{\eta}(\rho,h)]y_{t-1}.$$

By Loève's inequality (Davidson, 1994, Thm. 9.28),

$$\begin{aligned}\frac{\sum_{t=1}^{T-h_T}[\hat{\xi}_t(h_T)-\xi_t(\rho_T,h_T)]^4}{(T-h_T)v(\rho_T,h_T)^4} &\leq 8\frac{[\hat{\beta}(h)-\beta(\rho_T,h_T)]^4}{v(\rho_T,h_T)^4}\frac{\sum_{t=1}^{T-h_T}u_t^4}{T-h_T} \\ &\quad + 8\frac{[\hat{\eta}(\rho_T,h_T)-\eta(\rho_T,h_T)]^4}{v(\rho_T,h_T)^4}\frac{\sum_{t=1}^{T-h_T}y_{t-1}^4}{T-h_T}.\end{aligned}$$

The second term on the right-hand side of the above display tends to zero in probability along a subsequence by Lemmas 8 and 9 and Markov's inequality.

By Assumption 2(i) and Markov's inequality, we have $(T-h_T)^{-1}\sum_{t=1}^{T-h_T}u_t^4=O_p(1)$ (along any subsequence). Lemma 9 then implies that the first term on the right-hand side in the above display tends to zero in probability along a subsequence. □

Lemma 7 (Negligibility of estimation error in $\hat{u}_t(h)$)**.** *Let the conditions of Lemma 5 hold. Then there exists a subsequence $\{k_T\}$ of $\{T\}$ such that*

$$\frac{\sum_{t=1}^{k_T-h_{k_T}}[\hat{u}_t(h_{k_T})-u_t]^4}{k_T-h_{k_T}} \underset{P_{\rho_{k_T}}}{\overset{p}{\to}} 0.$$

Proof. Since $\hat{u}_t(h_T)-u_t=[\rho-\hat{\rho}(h_T)]y_{t-1}$, we have

$$\frac{\sum_{t=1}^{T-h_T}[\hat{u}_t(h_T)-u_t]^4}{T-h_T}=[\hat{\rho}(h_T)-\rho]^4\frac{\sum_{t=1}^{T-h_T}y_{t-1}^4}{T-h_T}.$$

The right-hand tends to zero in probability along a subsequence by Lemmas 8 and 9, and Markov's inequality. □

Lemma 8 (Moment bound for y_t^4)**.** *Let Assumption 1 and Assumption 2(i) hold. Then, for* *all* $T \in \mathbb{N}$ *and* $\rho \in [-1, 1]$,

$$\max_{1 \le t \le T} E(y_t^4) \le \frac{6(E(u_t^4))^3}{\sigma^4 \delta^2} \times g(\rho, T)^4$$

under P_ρ.

Proof. We have defined

$$\xi_t(\rho, h) \equiv \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell},$$

Since we have set the initial condition $y_0 = 0$, we have

$$y_t = \sum_{\ell=1}^{t} \rho^{t-\ell} u_\ell = \xi_0(\rho, t).$$

Lemma 3 gives

$$\begin{aligned}\max_{1 \le t \le T} E(y_t^4) &= \max_{1 \le t \le T} E[\xi_0(\rho, t)^4] \\ &\le \frac{6E(u_0^4)\sigma^4}{\delta^2} \times \max_{1 \le t \le T} v(\rho, t)^4.\end{aligned}$$

Lemma 4 showed that

$$v(\rho, t)^2 \le \frac{E[u_t^4]}{\sigma^4} \sum_{\ell=0}^{t-1} \rho^{2\ell} \le \frac{E[u_t^4]}{\sigma^4} g(\rho, t)^2.$$

Thus,

$$\begin{aligned}\max_{1 \le t \le T} E[y_t^4] &\le \frac{6E(u_0^4)\sigma^4}{\delta^2} \times \frac{(E[u_0^4])^2}{\sigma^8} \times \max_{1 \le t \le T} g(\rho, t)^4 \\ &\le \frac{6(E(u_t^4))^3}{\sigma^4 \delta^2} \times g(\rho, T)^4,\end{aligned}$$

as was to be shown. □

Lemma 9 (Convergence rates of estimators)**.** *Let the conditions of Lemma 5 hold. Then there exists a subsequence* $\{k_T\}$ *of* $\{T\}$ *such that the following statements all hold:*

i) $\frac{\hat{\beta}(h_{k_T}) - \beta(\rho_{k_T}, h_{k_T})}{v(\rho_{k_T}, h_{k_T})} \underset{P_{\rho_{k_T}}}{\overset{p}{\to}} 0$.

ii) $g(\rho_{k_T}, k_T - h_{k_T})^2 \times \frac{[\hat{\eta}(\rho_{k_T}, h_{k_T}) - \eta(\rho_{k_T}, h_{k_T})]^2}{v(\rho_{k_T}, h_{k_T})^2} \xrightarrow[P_{\rho_{k_T}}]{p} 0.$

iii) $g(\rho_{k_T}, k_T - h_{k_T})^2 \times [\hat{\rho}(h_{k_T}) - \rho_{k_T}]^2 \xrightarrow[P_{\rho_{k_T}}]{p} 0.$

Proof. We shall only prove the statements (i)–(ii), since statement (iii) can be proved in a very similar manner.

For brevity, denote $g_T \equiv g(\rho_T, T - h_T) = \min\{(1 - |\rho_T|)^{-1/2}, (T - h_T)^{1/2}\}$. Recall the definition $\hat{\eta}(\rho, h) \equiv \rho\hat{\beta}(h) + \hat{\gamma}(h)$ in equation (19). Since $(\hat{\beta}(h), \hat{\eta}(\rho, h))'$ is a linear transformation of the OLS coefficients $(\hat{\beta}(h), \hat{\gamma}(h))'$, the former vector equals the OLS coefficients in a regression of y_{t+h} on (u_t, y_{t-1}), using $u_t = y_t - \rho y_{t-1}$. By the representation

$$y_{t+h} = \beta(\rho, h)u_t + \underbrace{\beta(\rho, h+1)}_{=\rho\beta(\rho,h)\equiv\eta(\rho,h)} y_{t-1} + \xi_t(\rho, h)$$

in equation (4), we can therefore write

$$\begin{pmatrix} \frac{\hat{\beta}(h_T) - \beta(\rho_T, h_T)}{v(\rho_T, h_T)} \\ g_T \frac{\hat{\eta}(\rho_T, h_T) - \eta(\rho_T, h_T)}{v(\rho_T, h_T)} \end{pmatrix} = \begin{pmatrix} \frac{\sum_{t=1}^{T-h_T} u_t^2}{T-h_T} & \frac{\sum_{t=1}^{T-h_T} u_t y_{t-1}}{(T-h_T)g_T} \\ \frac{\sum_{t=1}^{T-h_T} u_t y_{t-1}}{(T-h_T)g_T} & \frac{\sum_{t=1}^{T-h_T} y_{t-1}^2}{(T-h_T)g_T^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T)u_t}{(T-h_T)v(\rho_T, h_T)} \\ \frac{\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T)y_{t-1}}{(T-h_T)v(\rho_T, h_T)g_T} \end{pmatrix}$$

$$\equiv \hat{M}^{-1} \begin{pmatrix} \hat{m}_1 \\ \hat{m}_2 \end{pmatrix}.$$

We must prove that both elements of the vector on the left-hand side tend to zero in probability along some subsequence. Note that $\hat{m}_1$ tends to zero in probability by the central limit theorem in Lemma 1, and so does $\hat{m}_2$ by Lemma 10. Hence, it just remains to show that $\hat{M}^{-1}$ is bounded in probability along some subsequence.

We consider separately the two possible cases $\limsup_T (T - h_T)(1 - \rho_T) < \infty$ and $\limsup_T (T - h_T)(1 - \rho_T) = \infty$. By passing to a subsequence if necessary, we may assume that one of the following two conditions are true about the sequences $\{\rho_T\}$ and $\{h_T\}$ (we suppress the subsequence in our notation for brevity):

A) $(T - h_T)(1 - \rho_T) \to c \in [0, \infty)$.

B) $(T - h_T)(1 - \rho_T) \to \infty$.

Case (A). We must have $\rho_T \to 1$. Thus, $(T - h_T)(1 - |\rho_T|) \to c$ as well, and so $(T - h_T)^{-1/2} g_T \to \tilde{c} \equiv \min\{c^{-1}, 1\} \in (0, \infty)$. By passing to a further subsequence if necessary, we may assume $h_T/T \to \kappa \in [0, 1-a]$. Thus, $T(1-\rho_T) = \frac{T}{T-h_T} \times (T-h_T)(1-\rho_T) \to \frac{1}{1-\kappa} \times c \in [0, \infty)$, so the sequence of probability measures P_{ρ_T} is contiguous to P_1 by Assumption 3. (Note: This is the only place where we use Assumption 3 in any proof.) Thus, according to Pollard (2008, Lemma <4>)—which states that if a sequence of probability measure Q_n is continguous to P_n, then terms that are $O_{P_n}(1)$ are also $O_{Q_n}(1)$—we just need to prove that

$$\begin{pmatrix} \frac{\sum_{t=1}^{T-h_T} u_t^2}{T-h_T} & \frac{\sum_{t=1}^{T-h_T} u_t y_{t-1}}{(T-h_T)^{3/2}\tilde{c}} \\ \frac{\sum_{t=1}^{T-h_T} u_t y_{t-1}}{(T-h_T)^{3/2}\tilde{c}} & \frac{\sum_{t=1}^{T-h_T} y_{t-1}^2}{(T-h_T)^2\tilde{c}} \end{pmatrix}^{-1} = O_p(1)$$

under the probability measure P_1 (i.e., when the parameter ρ is fixed at 1). Note that $\{u_t\}$ satisfies a Functional Central Limit Theorem under Assumption 1 and Assumption 2(i), cf. Davidson (1994, Thm. 27.14). Hence, the convergence in distribution of the above matrix follows from standard unit root asymptotics, e.g., Hamilton (1994, Ch. 17.4).

Case (B). Since $\rho_T \geq -1 + a$ for all T, we must also have $(T - h_T)(1 - |\rho_T|) \to \infty$. Thus, $g_T = (1 - |\rho_T|)^{-1/2}$ for sufficiently large T. It then follows from Lemma 11 that $\hat{M}$ converges in probability to a non-singular matrix along some subsequence. This completes the proof. □

Lemma 10 (OLS numerator)**.** *Let Assumptions 1 and 2 hold. Let $\{\rho_T\}$ be a sequence of real numbers satisfying $\rho_T \in [-1, 1]$ for all T, and let $\{h_T\}$ be a sequence of nonnegative integers satisfying $T - h_T \to \infty$ and $g(\rho_T, h_T)^2/T \to 0$. Then*

$$\frac{\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T) y_{t-1}}{(T-h_T)v(\rho_T, h_T)g(\rho_T, T-h_T)} \overset{p}{\underset{P_{\rho_T}}{\to}} 0.$$

Proof. We will show that

$$E\left[\left(\frac{\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T) y_{t-1}}{(T-h_T)v(\rho_T, h_T)g(\rho_T, T-h_T)}\right)^2\right] \to 0.$$

To that end, observe that if $t \geq s + h_T$, then

$$E[\xi_t(\rho_T, h_T)y_{t-1}\xi_s(\rho_T, h_T)y_{s-1}] = E\left[E(\xi_t(\rho_T, h_T) \mid u_t, u_{t-1}, \dots)y_{t-1}\xi_s(\rho_T, h_T)y_{s-1}\right] = 0,$$

by Assumption 1. By symmetry, the left-hand side above equals 0 also if $s \geq t + h_T$. Thus,

$$E\left[\left(\frac{\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T) y_{t-1}}{(T-h_T)v(\rho_T, h_T)g(\rho_T, T-h_T)}\right)^2\right]$$
$$\leq \sum_{t=1}^{T-h_T} \sum_{s=1}^{T-h_T} \mathbb{1}(|s-t| < h_T) \frac{|E[\xi_t(\rho_T, h_T) y_{t-1} \xi_s(\rho_T, h_T) y_{s-1}]|}{(T-h_T)^2 v(\rho_T, h_T)^2 g(\rho_T, T-h_T)^2}. \tag{28}$$

We now bound the summands on the right-hand side above. Consider first the case $s \in (t - h_T, t]$ (we will handle the case $t \in (s - h_T, s]$ by symmetry). By writing out the definitions of $\xi_t(\rho_T, h_T)$ and $y_{t-1} = \xi_0(\rho_T, t-1)$, we find

$$E[\xi_t(\rho_T, h_T) y_{t-1} \xi_s(\rho_T, h_T) y_{s-1}]$$
$$= \sum_{\ell_1=1}^{h_T} \sum_{\ell_2=1}^{h_T} \sum_{m_1=1}^{t-1} \sum_{m_2=1}^{s-1} \rho_T^{2h_T-\ell_1-\ell_2} \rho_T^{m_1+m_2-2} E(u_{t+\ell_1} u_{t-m_1} u_{s+\ell_2} u_{s-m_2}).$$

Consider any summand above defined by its indices $(\ell_1, \ell_2, m_1, m_2)$. Since $t + \ell_1 > \max\{t - m_1, s - m_2\}$, Assumption 1 implies that the summand can only be nonzero if $s + \ell_2 = t + \ell_1$, which requires $\ell_1 \leq h_T + s - t$. Moreover, when $s + \ell_2 = t + \ell_1$, we also need $t - m_1 = s - m_2$ for the summand to be nonzero, which in turn requires $m_1 \geq t - s + 1$. Thus,

$$|E[\xi_t(\rho_T, h_T) y_{t-1} \xi_s(\rho_T, h_T) y_{s-1}]|$$
$$\leq \sum_{\ell_1=1}^{h_T+s-t} \sum_{m_1=t-s+1}^{t-1} |\rho_T|^{2h_T-\ell_1-(t-s+\ell_1)} |\rho_T|^{m_1+(s-t+m_1)-2} E(u_{t+\ell_1}^2 u_{t-m_1}^2)$$
$$\leq E(u_0^4) \rho_T^{2(t-s)} \sum_{\ell_1=1}^{h_T+s-t} \rho_T^{2(h_T+s-t-\ell_1)} \sum_{m_1=t-s+1}^{t-1} \rho_T^{2[m_1-(t-s+1)]}$$
$$= E(u_0^4) \rho_T^{2(t-s)} \sum_{\ell=0}^{h_T+s-t-1} \rho_T^{2\ell} \sum_{m=0}^{s-2} \rho_T^{2m}$$
$$\leq E(u_0^4) \rho_T^{2(t-s)} \sum_{\ell=0}^{h_T-1} \rho_T^{2\ell} \sum_{m=0}^{T-h_T-1} \rho_T^{2m}$$
$$\leq E(u_0^4) \rho_T^{2(t-s)} \times v(\rho_T, h_T)^2 \frac{\sigma^2}{\delta} \times g(\rho_T, T-h_T)^2,$$

where the last inequality uses Lemma 4 and $\sum_{m=0}^{T-h_T-1} \rho_T^{2m} \leq \min\{(1-|\rho_T|)^{-1}, T-h_T\}$. We have derived the bound in the above display under the assumption $s \in (t - h_T, t]$, but by symmetry, it also applies when $t \in (s - h_T, s]$ if we replace $(t-s)$ with $|t-s|$. Inserting into

(28), we get

$$
\begin{aligned}
&E\left[\left(\frac{\sum_{t=1}^{T-h_T}\xi_t(\rho_T,h_T)y_{t-1}}{(T-h_T)v(\rho_T,h_T)g(\rho_T,T-h_T)}\right)^2\right] \\
&\leq \frac{1}{(T-h_T)^2}\sum_{t=1}^{T-h_T}\sum_{s=1}^{T-h_T}\mathbb{1}(|s-t|<h_T)E(u_0^4)\rho_T^{2|t-s|}\times\frac{\sigma^2}{\delta} \\
&= \frac{\sigma^2 E(u_0^4)}{(T-h_T)^2\delta}\sum_{t=1}^{T-h_T}\sum_{s=1}^{T-h_T}\mathbb{1}(|s-t|<h_T)\rho_T^{2|t-s|} \\
&= \frac{\sigma^2 E(u_0^4)}{(T-h_T)\delta}\sum_{|m|<h_T}\left(1-\frac{|m|}{T-h_T}\right)\rho_T^{2|m|} \\
&\leq \frac{2\sigma^2 E(u_0^4)}{(T-h_T)\delta}\sum_{m=0}^{h_T-1}\rho_T^{2m} \\
&\leq \frac{2\sigma^2 E(u_0^4)}{\delta}\times\frac{g(\rho_T,h_T)^2}{T-h_T} \\
&\to 0,
\end{aligned}
$$

as was to be shown. □

Lemma 11 (OLS denominator)**.** *Let Assumptions 1 and 2 hold. Let there be given a sequence $\{\rho_T\}$ of real numbers satisfying $\rho_T \in [-1,1]$ for all T and a sequence $\{h_T\}$ of nonnegative integers satisfying $T-h_T\to\infty$ and $(T-h_T)(1-|\rho_T|)\to\infty$. Then all the following statements hold:*

i) $\frac{\sum_{t=1}^{T-h_T}u_t^2}{T-h_T}\xrightarrow[P_{\rho_T}]{p}\sigma^2$.

ii) $\frac{\sum_{t=1}^{T-h_T}u_t y_{t-1}}{(T-h_T)(1-|\rho_T|)^{-1/2}}\xrightarrow[P_{\rho_T}]{p}0$.

iii) There exists a subsequence $\{k_T\}$ of $\{T\}$ and a constant $\bar{c}\in[\sigma^2/2,\sigma^2]$ such that

$$
\frac{\sum_{t=1}^{k_T-h_{k_T}}y_{t-1}^2}{(k_T-h_{k_T})(1-|\rho_{k_T}|)^{-1}}\xrightarrow[P_{\rho_{k_T}}]{p}\bar{c}.
$$

Proof. We consider each statement separately.

PART (I). Since $E(u_t^2)=\sigma^2$ by definition, this statement follows from a standard application of Chebyshev's inequality, exploiting the summability of the autocovariances of $\{u_t^2\}$, cf. Assumption 2(ii). See for example Davidson (1994, Thm. 19.2).

Part (ii). Note that $\{u_t y_{t-1}\}_t$ is a martingale difference array with respect to the natural filtration $\tilde{\mathcal{F}}_t = \sigma(u_t, u_{t-1}, \dots)$. Thus, the sequence is serially uncorrelated, implying that

$$E\left[\left(\frac{\sum_{t=1}^{T-h_T} u_t y_{t-1}}{(T-h_T)(1-|\rho_T|)^{-1/2}}\right)^2\right] = \frac{1-|\rho_T|}{(T-h_T)^2}\sum_{t=1}^{T-h_T} E[u_t^2 y_{t-1}^2]$$
$$\leq \frac{1-|\rho_T|}{T-h_T} \times [E(u_t^4)]^{1/2} \times \max_{1\leq t\leq T-h_T}(E[y_{t-1}^4])^{1/2}$$
$$\leq \frac{1-|\rho_T|}{T-h_T} \times \frac{\sqrt{6}E(u_t^4)^2}{\sigma^2\delta} \times g(\rho_T, T-h_T)^2, \tag{29}$$

where the last step uses Lemma 8. Since $g(\rho_T, T-h_T)^2 = \min\{(1-|\rho_T|)^{-1}, T-h_T\}$, it follows from the assumption $(T-h_T)(1-|\rho_T|) \to \infty$ that the expression (29) is uniformly of order $O((T-h_T)^{-1})$. This proves the statement.

Part (iii). By passing to a further subsequence if necessary, we may assume that $\rho_T \to \bar{\rho} \in [-1, 1]$, since the space $[-1, 1]$ is compact (we suppress the subsequence to simplify notation). We remind the reader that since $(T-h_T)(1-|\rho_T|) \to \infty$, we have $\rho_T < 1$ for large enough T. We can write

$$\frac{\sum_{t=1}^{T-h_T} y_{t-1}^2}{(T-h_T)(1-|\rho_T|)^{-1}} = \frac{1}{1+|\rho_T|} \times \frac{\sum_{t=1}^{T-h_T} y_{t-1}^2}{(T-h_T)(1-\rho_T^2)^{-1}},$$

and the first factor on the right-hand side above converges to $1/(1+|\bar{\rho}|) \in [1/2, 1]$ along the subsequence. Hence, we just need to show that

$$\frac{\sum_{t=1}^{T-h_T} y_{t-1}^2}{(T-h_T)(1-\rho_T^2)^{-1}} \xrightarrow[P_{\rho_T}]{p} \sigma^2.$$

To do this, we show that the mean converges and that the variance vanishes asymptotically. Note that it must be the case that $|\rho_T| < 1$ for sufficiently large T. Then

$$E\left[\frac{\sum_{t=1}^{T-h_T} y_{t-1}^2}{(T-h_T)(1-\rho_T^2)^{-1}}\right] = \frac{\sum_{t=1}^{T-h_T} E(y_{t-1}^2)}{(T-h_T)(1-\rho_T^2)^{-1}}$$
$$= \frac{\sum_{t=1}^{T-h_T} E\left((\sum_{\ell=1}^{t-1}\rho_T^{t-1-l}u_\ell)^2\right)}{(T-h_T)(1-\rho_T^2)^{-1}}$$

$$
\begin{aligned}
&= \sigma^2 \frac{\sum_{t=1}^{T-h_T} (1-\rho_T^2) \sum_{\ell=0}^{t-2} \rho_T^{2\ell}}{T-h_T} \\
&= \sigma^2 \frac{\sum_{t=1}^{T-h_T} (1-\rho_T^{2(t-1)})}{T-h_T} \\
&= \sigma^2 \left(1 - \frac{\sum_{t=1}^{T-h_T} \rho_T^{2(t-1)}}{T-h_T}\right) \\
&\to \sigma^2,
\end{aligned}
$$

where the last step uses the fact that ρ_T is eventually smaller than 1 for T large enough and

$$
\frac{\sum_{t=1}^{T-h_T} \rho_T^{2(t-1)}}{T-h_T} \le \frac{\sum_{t=0}^{\infty} \rho_T^{2t}}{T-h_T} = \frac{1}{(T-h_T)(1-\rho_T^2)} \le \frac{1}{(T-h_T)(1-|\rho_T|)} \to 0.
$$

It remains to show that

$$
\mathrm{Var}\left[\frac{\sum_{t=1}^{T-h_T} y_{t-1}^2}{(T-h_T)(1-\rho_T^2)^{-1}}\right] \to 0.
$$

We first derive an upper bound on $|\operatorname{Cov}(y_t^2, y_s^2)|$. Assume first that $t \ge s$. Since $y_t = \sum_{\ell=1}^{t} \rho_T^{t-\ell} u_\ell$, we can write

$$
\operatorname{Cov}(y_t^2, y_s^2) = \sum_{\ell_1=1}^{t} \sum_{\ell_2=1}^{t} \sum_{m_1=1}^{s} \sum_{m_2=1}^{s} \rho_T^{2(t+s)-(\ell_1+\ell_2+m_1+m_2)} \operatorname{Cov}(u_{\ell_1} u_{\ell_2}, u_{m_1} u_{m_2}).
$$

Due to Assumption 1, the only way that an expression of the form $\operatorname{Cov}(u_{\ell_1} u_{\ell_2}, u_{m_1} u_{m_2})$ can be non-zero is if either (a) $\ell_1 = \ell_2$ and $m_1 = m_2$, or (b) $\{\ell_1, \ell_2\} = \{m_1, m_2\}$. Note that case (b) is only possible if $\ell_1, \ell_2 \le s$ (since $s \le t$). By enumerating the various possibilities, we thus obtain, again for $t \ge s$,

$$
\begin{aligned}
\operatorname{Cov}(y_t^2, y_s^2) = &\sum_{\ell_1=1}^{t} \sum_{m_1=1}^{s} \rho_T^{2(t+s-\ell_1-m_1)} \operatorname{Cov}(u_{\ell_1}^2, u_{m_1}^2) \\
&+ 2 \sum_{\ell_1=1}^{s} \sum_{1 \le \ell_2 \le s,\, \ell_2 \ne \ell_1} \rho_T^{2(t+s-\ell_1-\ell_2)} \operatorname{Cov}(u_{\ell_1} u_{\ell_2}, u_{\ell_1} u_{\ell_2}).
\end{aligned}
$$

Note that $|\operatorname{Cov}(u_{\ell_1} u_{\ell_2}, u_{\ell_1} u_{\ell_2})| = \mathrm{Var}(u_{\ell_1} u_{\ell_2}) \le E(u_0^4) < \infty$ by Assumption 2(i). Consequently, by changing summation variables in the above display, we have for $t \ge s$ that

$$
|\operatorname{Cov}(y_t^2, y_s^2)| \le \sum_{\tilde{\ell}_1=0}^{t-1} \sum_{\tilde{m}_1=0}^{s-1} \rho_T^{2(\tilde{\ell}_1+\tilde{m}_2)} |\operatorname{Cov}(u_{t-\tilde{\ell}_1}^2, u_{s-\tilde{m}_1}^2)| + 2E(u_0^4) \rho_T^{2(t-s)} \left(\sum_{\tilde{\ell}_1=0}^{s-1} \rho_T^{2\tilde{\ell}_1}\right)^2
$$

$$\leq \sum_{\tilde{\ell}_1=0}^{t-1} \sum_{\tilde{m}_1=0}^{s-1} \rho_T^{2(\tilde{\ell}_1+\tilde{m}_2)} |\operatorname{Cov}(u_{t-\tilde{\ell}_1}^2, u_{s-\tilde{m}_1}^2)| + 2E(u_0^4)\rho_T^{2(t-s)} \frac{1}{(1-\rho_T^2)^2}.$$

We derived the above bound under the assumption $t \geq s$, but by symmetry, the bound above holds for all t, s as long as we replace $(t-s)$ with $|t-s|$. Note that, by Assumption 2(ii), we have

$$K \equiv \sum_{s=-\infty}^{\infty} |\operatorname{Cov}(u_t^2, u_s^2)| < \infty,$$

and this expression does not depend on t by stationarity. Putting everything together, we find

$$\begin{aligned}
\operatorname{Var}\left[\frac{\sum_{t=1}^{T-h_T} y_{t-1}^2}{(T-h_T)(1-\rho_T^2)^{-1}}\right] &\leq \frac{(1-\rho_T^2)^2}{(T-h_T)^2} \sum_{t=1}^{T-h_T} \sum_{s=1}^{T-h_T} |\operatorname{Cov}(y_{t-1}^2, y_{s-1}^2)| \\
&\leq \frac{(1-\rho_T^2)^2}{(T-h_T)^2} \sum_{t=1}^{T-h_T} \sum_{s=1}^{T-h_T} \sum_{\tilde{\ell}_1=0}^{t-1} \sum_{\tilde{m}_1=0}^{s-1} \rho_T^{2(\tilde{\ell}_1+\tilde{m}_2)} |\operatorname{Cov}(u_{t-\tilde{\ell}_1}^2, u_{s-\tilde{m}_1}^2)| \\
&\quad + \frac{(1-\rho_T^2)^2}{(T-h_T)^2} \sum_{t=1}^{T-h_T} \sum_{s=1}^{T-h_T} 2E(u_0^4)\rho_T^{2|t-s|} \frac{1}{(1-\rho_T^2)^2} \\
&\leq \frac{(1-\rho_T^2)^2}{T-h_T} \sum_{\tilde{\ell}_1=0}^{t-1} \sum_{\tilde{m}_1=0}^{s-1} \rho_T^{2(\tilde{\ell}_1+\tilde{m}_2)} \sum_{s=-\infty}^{\infty} |\operatorname{Cov}(u_{-\tilde{\ell}_1}^2, u_s^2)| \\
&\quad + \frac{2E(u_0^4)}{(T-h_T)^2} \sum_{t=1}^{T-h_T} \sum_{s=1}^{T-h_T} \rho_T^{2|t-s|} \\
&\leq \frac{K(1-\rho_T^2)^2}{T-h_T} \left(\sum_{\tilde{\ell}_1=0}^{t-1} \rho_T^{2\tilde{\ell}_1}\right)^2 + \frac{4E(u_0^4)}{(T-h_T)^2} \sum_{k=0}^{T-h_T-1} (T-h_T-k)\rho_T^{2k} \\
&\leq \frac{K(1-\rho_T^2)^2}{T-h_T} \left(\sum_{\tilde{\ell}_1=0}^{\infty} \rho_T^{2\tilde{\ell}_1}\right)^2 + \frac{4E(u_0^4)}{T-h_T} \sum_{k=0}^{\infty} \rho_T^{2k} \\
&= \frac{K}{T-h_T} + \frac{4E(u_0^4)}{(T-h_T)(1-\rho_T^2)} \\
&\leq \frac{K}{T-h_T} + \frac{4E(u_0^4)}{(T-h_T)(1-|\rho_T|)} \\
&\to 0,
\end{aligned}$$

as needed. Here the last step uses the assumption $(T-h_T)(1-|\rho_T|) \to \infty$. □

References

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2019): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," *Journal of Econometrics*, forthcoming.

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.

Benkwitz, A., M. H. Neumann, and H. Lütkepohl (2000): "Problems related to confidence intervals for impulse responses of autoregressive processes," *Econometric Reviews*, 19, 69–103.

Breitung, J. and R. Brüggemann (2019): "Projection estimators for structural impulse responses," University of Konstanz Department of Economics Working Paper Series 2019-05.

Brillinger, D. R. (2001): *Time Series: Data Analysis and Theory*, Classics in Applied Mathematics, SIAM.

Brugnolini, L. (2018): "About Local Projection Impulse Response Function Reliability," Manuscript, University of Rome "Tor Vergata".

Chevillon, G. (2017): "Robustness of Multistep Forecasts and Predictive Regressions at Intermediate and Long Horizons," ESSEC Working Paper 1710.

Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*, Advanced Texts in Econometrics, Oxford University Press.

Dolado, J. J. and H. Lütkepohl (1996): "Making Wald tests work for cointegrated VAR systems," *Econometric Reviews*, 15, 369–386.

Dufour, J.-M., D. Pelletier, and Éric Renault (2006): "Short run and long run causality in time series: inference," *Journal of Econometrics*, 132, 337–362.

Gonçalves, S. and L. Kilian (2004): "Bootstrapping autoregressions with conditional heteroskedasticity of unknown form," *Journal of Econometrics*, 123, 89–120.

Gospodinov, N. (2004): "Asymptotic confidence intervals for impulse responses of near-integrated processes," *Econometrics Journal*, 7, 505–527.

Hamilton, J. D. (1994): *Time Series Analysis*, Princeton University Press.

Hansen, B. E. (1999): "The Grid Bootstrap and the Autoregressive Model," *Review of Economics and Statistics*, 81, 594–607.

Herbst, E. P. and B. K. Johannsen (2020): "Bias in Local Projections," Board of Governors of the Federal Reserve System Finance and Economics Discussion Series 2020-010.

Inoue, A. and L. Kilian (2020): "The uniform validity of impulse response inference in autoregressions," *Journal of Econometrics*, 215, 450–472.

Jansson, M. (2008): "Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis," *Econometrica*, 76, 1103–1142.

Jeganathan, P. (1995): "Some Aspects of Asymptotic Theory with Applications to Time Series Models," *Econometric Theory*, 11, 818–887.

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

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 H. Lütkepohl (2017): *Structural Vector Autoregressive Analysis*, Cambridge University Press.

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.

Mikusheva, A. (2007): "Uniform Inference in Autoregressive Models," *Econometrica*, 75, 1411–1452.

——— (2012): "One-Dimensional Inference in Autoregressive Models With the Potential Presence of a Unit Root," *Econometrica*, 80, 173–212.

Montiel Olea, J. L. and M. Plagborg-Møller (2019): "Simultaneous confidence bands: Theory, implementation, and an application to SVARs," *Journal of Applied Econometrics*, 34, 1–17.

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

Pesavento, E. and B. Rossi (2006): "Small-sample confidence intervals for multivariate impulse response functions at long horizons," *Journal of Applied Econometrics*, 21, 1135–1155.

——— (2007): "Impulse response confidence intervals for persistent data: What have we learned?" *Journal of Economic Dynamics and Control*, 31, 2398–2412.

Phillips, P. C. and J. H. Lee (2013): "Predictive regression under various degrees of persistence and robust long-horizon regression," *Journal of Econometrics*, 177, 250–264, special issue on "Dynamic Econometric Modeling and Forecasting".

Phillips, P. C. B. (1998): "Impulse response and forecast error variance asymptotics in nonstationary VARs," *Journal of Econometrics*, 83, 21–56.

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

Pollard, D. (2008): "Contiguity," Preprint of *Asymptopia*, Chapter 7.

Pope, A. L. (1990): "Biases of Estimators in Multivariate Non-Gaussian Autoregressions," *Journal of Time Series Analysis*, 11, 249–258.

Ramey, V. A. (2016): "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, chap. 2, 71–162.

Richardson, M. and J. H. Stock (1989): "Drawing inferences from statistics based on multiyear asset returns," *Journal of Financial Economics*, 25, 323–348.

Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113–144.

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

Toda, H. Y. and T. Yamamoto (1995): "Statistical inference in vector autoregressions with possibly integrated processes," *Journal of Econometrics*, 66, 225–250.

Valkanov, R. (2003): "Long-horizon regressions: theoretical results and applications," *Journal of Financial Economics*, 68, 201–232.

Wright, J. H. (2000): "Confidence Intervals for Univariate Impulse Responses With a Near Unit Root," *Journal of Business & Economic Statistics*, 18, 368–373.