The Local Projection Residual Bootstrap for AR(1) Models*

Amilcar Velez
Department of Economics
Northwestern University
amilcare@u.northwestern.edu

October 19, 2023

Abstract

This paper proposes a local projection residual bootstrap method to construct confidence intervals for impulse response coefficients of AR(1) models. Our bootstrap method is based on the local projection (LP) approach and a residual bootstrap procedure. We present theoretical results for our bootstrap method and proposed confidence intervals. First, we prove the uniform consistency of the LP-residual bootstrap over a large class of AR(1) models that allow for a unit root. Then, we prove the asymptotic validity of our confidence intervals over the same class of AR(1) models. Finally, we show that the LP-residual bootstrap provides asymptotic refinements for confidence intervals on a restricted class of AR(1) models relative to those required for the uniform consistency of our bootstrap.

KEYWORDS: Bootstrap, Local Projection, Uniform Inference, Asymptotic Refinements.

*I am deeply grateful to Ivan Canay for his guidance and support and to Federico Bugni and Joel Horowitz for extensive discussions that have helped shape the paper. I am also thankful to Federico Crippa, Bruno Fava, Danil Fedchenko, Diego Huerta, Eleftheria Kelekidou, Pepe Montiel-Olea, Filip Obradovic, Mikkel Plagborg-Møller, Sebastian Sardon, and Keli Xu for valuable comments and suggestions.

1 Introduction

This paper contributes to a growing literature on confidence interval construction for impulse response coefficients based on the local projection (LP) approach (Jorda (2005)). In this literature, the LP approach estimates an impulse response coefficient as one of the slope coefficients in a linear regression of a future outcome on (current or lag-augmented) covariates (Ramey (2016); Nakamura and Steinsson (2018); Montiel Olea and Plagborg-Møller (2021)). Recent theoretical results exist for the asymptotic validity of the confidence intervals constructed around the LP estimator over a large class of vector autoregressive models (Xu (2023)). However, the theory of the bootstrap version of the confidence intervals is unknown, even for the AR(1) model.

In this paper we propose an LP-residual bootstrap method to construct confidence intervals for impulse response coefficients of AR(1) models. Our bootstrap method is based on the LP approach and a residual bootstrap procedure. Our bootstrap confidence intervals are centered at the LP estimator and use heteroskedasticity-consistent (HC) standard errors. The main difference with available confidence intervals is the computation of the critical value. Section 3 presents the details.

We rely on the asymptotic distribution theory initially developed in Montiel Olea and Plagborg-Møller (2021) and generalized in Xu (2023). In their framework, the data are generated from a VAR model that belongs to a large class of VAR models, which allows for multiple unit roots and conditional heteroskedasticity of unknown form. For a given horizon h and a sample size n, there is a root $R_n(h)$ (a real-valued function depending on the data and an impulse response coefficient) that is based on the LP approach. They show that the root $R_n(h)$ is asymptotically distributed as a standard normal distribution. As a result, they construct a confidence interval $C_n(h, 1-\alpha)$ for an impulse response coefficient using the root $R_n(h)$ and a normal critical value based on the asymptotic distribution. Moreover, their confidence intervals have asymptotic coverage equal to the nominal level $1-\alpha$ uniformly over the parameter space (coefficients of the VAR model) and intermediate horizons (e.g., $h \leq h_n = o(n)$). See Section 2.1 for further details and discussion.

We propose the LP-residual bootstrap method to approximate the distribution of the root $R_n(h)$ as an alternative to the asymptotic distribution. We use our approximation to compute bootstrap-based critical values that can be used in the confidence interval construction. In particular, we construct a confidence interval $C_n^*(h, 1-\alpha)$ for an impulse response coefficient using the root $R_n(h)$ and a bootstrap critical value.

Our first result proves the uniform consistency of the LP-residual bootstrap over a large class of AR(1) models that allow for a unit root. Moreover, we show that the distribution of the root $R_n(h)$ can be approximated by its bootstrap version uniformly over the parameter space (e.g., $\rho \in [-1, 1]$) and intermediate horizons (e.g., $h \leq h_n = o(n)$). Our result applies to a large class of AR(1) models that allow for a unit root, conditional heteroskedasticity of unknown form as in Gonçalves and Kilian (2004), which includes ARCH and GARCH shocks, and sequence of shocks that satisfy the martingale difference assumption. To prove this result, we extend the existing asymptotic distribution theory for a particular sequence of AR(1) models (Theorem B.1). Moreover, we prove that a high-level assumption (Assumption 3 in Montiel Olea and Plagborg-Møller (2022) and Assumption 4 in Xu (2023)) necessary for the theoretical properties of $C_n(h, 1-\alpha)$ can be verified if the shocks of the AR(1) model are i.i.d (Proposition B.1).

Our first result implies that the LP-residual bootstrap method provides asymptotically valid confidence intervals over a large class of AR(1) models that allow for a unit root. Specifically, our confidence interval $C_n^*(h, 1-\alpha)$ has an asymptotic coverage equal to the nominal level $1-\alpha$ uniformly over $\rho \in [-1, 1]$ and intermediate horizons.

Our second set of results shows that the LP-residual bootstrap provides asymptotic refinements to the confidence intervals on a more restricted class of AR(1) models (e.g., $|\rho| < 1$, i.i.d. shocks, and the existence of positive continuous density), i.e., the rate of convergence of the error in coverage probability (ECP) of the confidence interval $C_n^*(h, 1-\alpha)$ is faster than that of $C_n(h, 1-\alpha)$. Specifically, we show that the rates of convergence of the ECP of the confidence intervals $C_n^*(h, 1-\alpha)$ and $C_n(h, 1-\alpha)$ are $o(n^{-1})$ and $O(n^{-1})$, respectively, under Assumption 5.1 in Section 5.2. We use the rate of convergence of the ECP as a measure to compare the confidence intervals since a faster rate of convergence of the ECP is associated with a coverage probability closer to the nominal level $1-\alpha$ for a sufficiently large sample size n, which indicates a more precise inference method. An informal discussion to derive the mentioned rates of convergence appears in Section 5.1, while the formal results are in Section 5.2. To show these results, we follow the literature and prove the existence of Edgeworth expansions for the distribution of the root $R_n(h)$ and its bootstrap version for a fixed h and $|\rho| < 1$.

In the growing literature on LP inference, various bootstrap methods to construct confidence intervals for the impulse response coefficients have been considered and recommended based on simulation studies. Montiel Olea and Plagborg-Møller (2021) use a wild bootstrap procedure to generate new samples and compute critical values, but the theoretical results for their bootstrap method are unknown. Kilian and Kim (2011) presents a simulation study including a block-bootstrap method to construct confidence intervals based on the LP approach, but the theory of their block-bootstrap method is unknown. Recently, Lusompa (2023) proposes a block wild bootstrap method for confidence interval construction that is point-wise valid for a class of stationary data-generating processes; however, his bootstrap method is not applicable for an AR(1) model with a unit root. In contrast, we present a bootstrap method based on the LP approach with theoretical results. Our bootstrap method guarantees that our confidence intervals provide uniform inference over a large class of AR(1) models that allow for a unit root.

More broadly, we contribute to the literature on confidence interval construction for impulse response coefficients. For short horizons (fixed h), the problem of confidence interval construction has been studied by Andrews (1993), Hansen (1999), Inoue and Kilian (2002), Jorda (2005), Mikusheva (2007, 2015), among others. For long horizons ($h_n = (1-b)n$ for some $b \in (0,1)$), the problem of confidence interval construction was revised by Gospodinov (2004), Pesavento and Rossi (2006), and Mikusheva (2012) since the methods for short horizons produced invalid confidence intervals when the data-generating process allows for unit roots. Recently, the problem of confidence interval construction for intermediate horizons ($h_n = o(n)$) was addressed in Montiel Olea and Plagborg-Møller (2021) and Xu (2023), which was a case not covered in the literature. In this paper, we propose bootstrap confidence intervals that are asymptotically valid at short and intermediate horizons.

Moreover, for short horizons (fixed h), our confidence intervals are a simpler alternative to the available but computationally intensive method known as the grid bootstrap; see Hansen (1999) and Mikusheva (2007, 2012). The grid bootstrap computes bootstrap critical values over a grid of points since it relies on test inversion, which makes the procedure computationally intensive. In contrast, our confidence interval only requires computing one bootstrap critical value. See Remark 4.2 for further discussion.

The remainder of the paper is organized as follows. We first describe in Section 2 the setup and previous results. We then present our bootstrap confidence interval and the LP-residual bootstrap in Section 3. The theoretical properties of the LP-residual bootstrap appear in Section 4 (uniform consistency) and Section 5 (asymptotic refinements). In Section 6, we present a simulation study to investigate the numerical performance of the LP-residual bootstrap. Finally, Section 7 presents concluding remarks. All the proofs are presented in Appendixes A and B, and Supplemental Appendix C. Additional simulation results appear in Supplemental Appendix D.

2 Setup and Previous Results on Local Projection

Consider an AR(1) model,

$$y_t = \rho y_{t-1} + u_t, \quad y_0 = 0, \quad \rho \in [-1, 1] \ . \tag{1}$$

Denote the impulse response coefficient at horizon $h \in \mathbf{N}$ by

$$\beta(\rho, h) \equiv \rho^h \ . \tag{2}$$

Let $\hat{\beta}_n(h)$ be the slope coefficient of y_t in the linear regression of y_{t+h} on y_t and y_{t-1},

$$y_{t+h} = \hat{\beta}_n(h) y_t + \hat{\gamma}_n(h) y_{t-1} + \hat{\xi}_t(h), \quad t = 1, \ldots, n-h \ , \tag{3}$$

where $(\hat{\beta}_n(h),\ \hat{\gamma}_n(h))$ and $\{\hat{\xi}_t(h) : 1 \leq t \leq n-h\}$ are the coefficient vector and residuals of the linear regression (3), respectively. Equation (3) is called the *lag-augmented* LP regression and was developed in Montiel Olea and Plagborg-Møller (2021), where they give conditions under which the coefficient $\hat{\beta}_n(h)$ consistently estimates $\beta(\rho, h)$.

Let $\hat{s}_n(h)$ be the heteroskedasticity-consistent (HC) standard error of $\hat{\beta}_n(h)$ in the lag-augmented LP regression (3), which can be computed as follows

$$\hat{s}_n(h) \equiv \left(\sum_{t=1}^{n-h} \hat{u}_t(h)^2 \right)^{-1/2} \left(\sum_{t=1}^{n-h} \hat{\xi}_t(h)^2 \hat{u}_t(h)^2 \right)^{1/2} \left(\sum_{t=1}^{n-h} \hat{u}_t(h)^2 \right)^{-1/2} , \tag{4}$$

where $\hat{u}_t(h) \equiv y_t - \hat{\rho}_n(h) y_{t-1}$ and

$$\hat{\rho}_n(h) \equiv \left(\sum_{t=1}^{n-h} y_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^{n-h} y_t y_{t-1} \right) \ . \tag{5}$$

For a given $h \in \mathbf{N}$, we consider the following real-valued root for the parameter $\beta(\rho, h)$:

$$R_n(h) \equiv \frac{\hat{\beta}_n(h) - \beta(\rho, h)}{\hat{s}_n(h)} \ , \tag{6}$$

where $\beta(\rho, h)$ is as in (2), $\hat{\beta}_n(h)$ is computed as in (3), and $\hat{s}_n(h)$ is as in (4). We denote the distribution of the root $R_n(h)$ by

$$J_n(x, h, P, \rho) \equiv P_\rho\left(R_n(h) \leq x\right) \ , \tag{7}$$

where $x \in \mathbf{R}$, $h \in \mathbf{N}$, P is the distribution of the shocks $\{u_t : t \geq 1\}$, $\rho \in \mathbf{R}$, and P_ρ denote the probability distribution of the sequence $\{y_t : t \geq 1\}$, which is defined jointly by the distribution P and the parameter ρ in (1).

Let $c_n(h, 1-\alpha)$ be the $1-\alpha$ quantile of $|R_n(h)|$ under the distribution P_ρ,

$$c_n(h, 1-\alpha) \equiv \inf\left\{u \in \mathbf{R} : P_\rho\left(|R_n(h)| \leq u\right) \geq 1-\alpha\right\} \ . \tag{8}$$

Ideally, we would use the root $R_n(h)$ and the critical value $c_n(h, 1-\alpha)$ to construct confidence sets for $\beta(\rho, h)$ with a coverage probability of at least 1-α: collecting all the parameters $\beta(\rho, h)$ such that $|R_n(h)| \leq C_n(h, 1-\alpha)$, which is equivalent to defining the next confidence interval

$$\tilde{C}_n(h, 1-\alpha) \equiv \left[\hat{\beta}_n(h) - c_n(h, 1-\alpha)\ \hat{s}_n(h),\ \hat{\beta}_n(h) + c_n(h, 1-\alpha)\ \hat{s}_n(h)\right] \ .$$

The critical value $c_n(h, 1-\alpha)$ is unknown since the distribution of the root is unknown in general. As a result, the confidence interval $\tilde{C}_n(h, 1-\alpha)$ is infeasible. For this reason, it is common to approximate the distribution of the root $R_n(h)$ relying on asymptotic distribution theory or bootstrap methods to approximate the infeasible $c_n(h, 1-\alpha)$.

2.1 Previous Results

The asymptotic distribution theory developed in Montiel Olea and Plagborg-Møller (2021) and Xu (2023) imply that the distribution $J_n(x, h, P, \rho)$ converges to the standard normal distribution $\Phi(x)$ whenever certain assumptions on the distribution of the shocks P hold. Moreover, this convergence is uniform over the values of $\rho \in [-1, 1]$ and intermediate horizons, that is

$$\sup_{|\rho|\leq 1} \ \sup_{h \leq h_n} \ \sup_{x \in \mathbf{R}} |J_n(x, h, P, \rho) - \Phi(x)| \to 0 \quad \text{as} \quad n \to \infty \ , \tag{9}$$

where $h_n \leq n$ and $h_n = o\left(n\right)$. Assumptions 4.1 and 4.2 in Section 4 are sufficient conditions on the distribution P to obtain (9) due to Theorem 2 in Xu (2023).

The confidence interval for $\beta(\rho, h)$ based on asymptotic distribution theory is defined as

$$C_n(h, 1-\alpha) \equiv \left[\hat{\beta}_n(h) - z_{1-\alpha/2}\ \hat{s}_n(h),\ \hat{\beta}_n(h) + z_{1-\alpha/2}\ \hat{s}_n(h)\right] \ , \tag{10}$$

where $z_{1-\alpha/2} \equiv \Phi^{-1}(1-\alpha/2)$ is the $1-\alpha/2$ quantile of the standard normal distribution. The result in (9) implies that confidence interval $C_n(h, 1-\alpha)$ is uniformly asymptotically valid in the sense that its asymptotic coverage probability is equal to the nominal level $1-\alpha$ uniformly over ρ and h,

$$\sup_{|\rho|\leq 1}\ \sup_{h\leq h_n} |P_\rho\left(\beta(\rho, h) \in C_n(h, 1-\alpha)\right) - (1-\alpha)| \to 0 \quad \text{as} \quad n \to \infty \ ,$$

where $h_n \leq n$ and $h_n = o\left(n\right)$. Three features of $C_n(h, 1-\alpha)$ deserve further discussion. First, it is simpler to compute than available alternatives in the sense that it does not require any tuning parameter. It is common to use heteroskedasticity- and autocorrelation-robust (HAR) standard errors for inference whenever we have dependent data. The major complication of HAR standard errors is the choice of the (truncation) tuning parameter; see Lazarus et al. (2018). In contrast, the HC standard errors $\hat{s}_n(h)$ defined in (4) are simple to compute and sufficient for inference under certain conditions on the distribution P; see Remark 2.1 for further explanation. Second, the uniform asymptotic validity of the confidence interval $C_n(h, 1-\alpha)$ avoids pre-testing procedures about the nature of the data-generating process ($|\rho| < 1$ vs $\rho = 1$) that can distort inference; see Mikusheva (2007). In particular, inference using $C_n(h, 1-\alpha)$ holds regardless of the value of $\rho \in [-1, 1]$. Third, the confidence interval $C_n(h, 1-\alpha)$ has theoretical guarantees at intermediate horizons (e.g., $h = h_n \leq o\left(n\right)$). This is an important feature for inference on impulse response coefficients at intermediate horizons. Other methods to construct confidence intervals that work at short horizons (h fixed) may have problems at long and intermediate horizons; see Gospodinov (2004), Pesavento and Rossi (2006), Mikusheva (2012), and Montiel Olea and Plagborg-Møller (2021) for additional discussion.

Remark 2.1. *The HC standard errors $\hat{s}_n(h)$ defined in (4) are sufficient for the construction of valid confidence intervals under certain conditions on the distribution P. In particular, as it was pointed out by Xu (2023), it is sufficient and necessary that the* scores $\{\xi_t(\rho, h)u_t : 1 \leq t \leq n-h\}$ *be serially uncorrelated, where $\xi_t(\rho, h) \equiv \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell}$. To explain the sufficiency of this condition, we use the derivations presented on page 1811 in Montiel Olea and Plagborg-Møller (2021) that implies that the root $R_n(h)$ defined in (6) can be written as* *follows*

$$\frac{\left((n-h)^{-1/2}\sum_{t=1}^{n-h}\xi_t(\rho,h)u_t\right)}{E\left[\xi_t(\rho,h)^2u_t^2\right]^{1/2}}\times\frac{\left[(n-h)^{-1}\sum_{t=1}^{n-h}\hat{\xi}_t(h)^2\hat{u}_t(h)^2\right]^{-1/2}}{E\left[\xi_t(\rho,h)^2u_t^2\right]^{-1/2}}+\varepsilon_n(\rho,h)\ ,$$

where $\varepsilon_n(\rho,h)$ is a remainder error term. We derive three implications under the Assumptions 4.1 and 4.2 presented in Section 4 that guarantee the scores are serially uncorrelated. First, the term between parentheses converges to a normal distribution with variance correctly scaled by the denominator. Second, the term between brackets converges in probability to its denominator due to serially uncorrelated scores. Third, the remainder error term $\varepsilon_n(\rho,h)$ converges in probability to zero. Importantly, Xu (2023) proposed alternative standard errors for the construction of confidence intervals under serially correlated scores.

3 The LP-Residual Bootstrap

This paper proposes an LP-residual bootstrap for confidence interval construction. Our confidence interval for the impulse response coefficient $\beta(\rho,h)$ is defined as

$$C_n^*(h,1-\alpha)\equiv\left[\hat{\beta}_n(h)-c_n^*(h,1-\alpha)\ \hat{s}_n(h),\ \hat{\beta}_n(h)+c_n^*(h,1-\alpha)\ \hat{s}_n(h)\right]\ , \tag{11}$$

where $\hat{\beta}_n(h)$ is an estimator for $\beta(\rho,h)$ defined in (3), $\hat{s}_n(h)$ is its heteroskedasticity-consistent (HC) standard error defined in (4), and $c_n^*(h,1-\alpha)$ is a bootstrap critical value defined in (15).

3.1 Bootstrap Critical Value

Let $Y^{(n)}\equiv\{y_t:1\leq t\leq n\}$ be data generated by (1). Let $c_n^*(h,1-\alpha)$ be the bootstrap critical value involving the following steps:

Step 1: Estimate ρ in the AR(1) model defined in (1) with the data $Y^{(n)}$ using linear regression, denoted by

$$\hat{\rho}_n\equiv\left(\sum_{t=1}^{n}y_{t-1}^2\right)^{-1}\left(\sum_{t=1}^{n}y_{t-1}y_t\right)\ , \tag{12}$$

and compute the centered residuals

$$\{\tilde{u}_t \equiv \hat{u}_t - n^{-1}\sum_{t=1}^{n}\hat{u}_t : 1 \le t \le n\} \ , \tag{13}$$

where $\hat{u}_t \equiv y_t - \hat{\rho}_n y_{t-1}$.

Step 2: Generate a new sample of size n using (1), (12), and (13). Define the sample as

$$y^*_{b,t} = \hat{\rho}_n y^*_{b,t-1} + u^*_{b,n} \ , \quad y^*_{b,0} = 0 \ , \ t = 1, \dots, n \ ,$$

where $\{u^*_{b,n} : 1 \le t \le n\}$ is a random sample from the empirical distribution of the centered residuals defined in (13). The new sample $\{y^*_{b,n} : 1 \le t \le n\}$ is called the bootstrap sample.

Step 3: Compute $\hat{\beta}^*_{b,n}(h)$ and $\hat{s}^*_{b,n}(h)$ as in (3) and (4) using the lag-augmented LP regression and the bootstrap sample $\{y^*_{b,t} : 1 \le t \le n\}$. Define the bootstrap version of the root

$$R^*_{b,n}(h) = \frac{\hat{\beta}^*_{b,n}(h) - \beta(\hat{\rho}_n, h)}{\hat{s}^*_{b,n}(h)} \ , \tag{14}$$

where $\beta(\rho, h)$ and $\hat{\rho}_n$ are as in (2) and (12), respectively.

Step 4: Define the bootstrap critical value as the $1-\alpha$ quantile of $|R^*_{b,n}(h)|$ conditional on the data $Y^{(n)}$, denoted by

$$c^*_n(h, 1-\alpha) \equiv \inf\left\{u \in \mathbf{R} : P_\rho\left(|R^*_{b,n}(h)| \le u \mid Y^{(n)}\right) \ge 1-\alpha\right\} \ . \tag{15}$$

We named this procedure the LP-residual bootstrap due to steps 2 and 3. Step 2 generates bootstrap samples based on the estimated model and a residual bootstrap procedure. Step 3 computes the bootstrap version of the root based on the lag-augmented LP regression. To our knowledge, this bootstrap procedure is new; see Remark 3.2 and 5.1 for other bootstrap procedures involving roots based on LP estimators.

We use the bootstrap critical value $c^*_n(h, 1-\alpha)$ in the construction of the confidence interval defined in (11). The explicit formula in (15) has two implications. First, the bootstrap critical value $c^*_n(h, 1-\alpha)$ depends on the data, the sample size n, and the horizon h. Second, we can compute $c^*_n(h, 1-\alpha)$ with perfect accuracy whenever we use the exact empirical distribution of the centered residuals defined in (13). However, the computation of an exact distribution can be computationally demanding; therefore, it is common to approximate it using Monte Carlo procedures as we describe in Remark 3.1, which has a theoretical justification due to Glivenko–Cantelli's theorem.

Remark 3.1. *It is a common practice to approximate the bootstrap critical value* $c_n^*(h, 1-\alpha)$ *using a Monte Carlo procedure (Horowitz (2001, 2019)). We generate* B *bootstrap samples of size* T*, where each* b*-th bootstrap sample* $\{y_{b,t}^* : 1 \leq t \leq n\}$ *is generated as in step 2. We then obtain* $\{|R_{b,n}^*(h)| : 1 \leq b \leq B\}$*, where each* $R_{b,n}^*(h)$ *is computed as in step 3. Finally, we approximate the bootstrap critical value* $c_n^*(h, 1-\alpha)$ *by the* $1-\alpha$ *quantile of* $\{|R_{b,n}^*(h)| : 1 \leq b \leq B\}$*, denoted by*

$$c_{b,n}^*(h, 1-\alpha) \equiv \inf \left\{ u \in \boldsymbol{R} : \frac{1}{B} \sum_{b=1}^{B} I\left\{ |R_{b,n}^*(h)| \leq u \right\} \geq 1-\alpha \right\} .$$

The accuracy of the approximation improves as the number of bootstrap samples B *increases. We use* $B = 1000$ *in our simulation study presented in Section 6.*

Remark 3.2. *Another bootstrap procedure to approximate the infeasible critical value* $c_n(h, 1-\alpha)$ *is presented in Section 5 in Montiel Olea and Plagborg-Møller (2021), which they recommend for practical use. They use the wild bootstrap procedure described in Gonçalves and Kilian (2004). For this reason, we name their procedure the LP-wild bootstrap. The only difference with respect to the LP-residual bootstrap is in Step 2. The LP-wild bootstrap defines the shocks as follows:* $u_{b,t}^* = \tilde{u}_t z_{b,t}$ *for all* $t = 1, \ldots, n$*, where* $\{\tilde{u}_t : 1 \leq t \leq n\}$ *are the centered residuals defined in (13) and* $\{z_{b,t} : 1 \leq t \leq n\}$ *is an i.i.d. sequence of standard normal random variables independent of the data* $Y^{(n)}$*. To our knowledge, the theoretical properties of the LP-wild bootstrap are unknown. We include the LP-wild bootstrap in our simulation study presented in Section 6.*

4 Uniform Consistency

We show the uniform consistency of the LP-residual bootstrap (Theorem 4.1) and that our proposed bootstrap confidence interval $C_n^*(h, 1-\alpha)$ defined in (11) is uniformly asymptotically valid (Theorem 4.2). In what follows, we first present and discuss the assumptions, and we then establish the results.

The following assumption imposes restrictions on the distributions of the shocks P. These assumptions are based on the general framework developed by Xu (2023) that generalized the work of Montiel Olea and Plagborg-Møller (2021).

Assumption 4.1.

i) $\{u_t : 1 \leq t \leq n\}$ *is covariance-stationary and satisfies* $E[u_n \mid \{u_s\}_{s<t}] = 0$ *almost surely.*

ii) $E[u_t^2 u_{t-s} u_{t-r}] = 0$ *for all* $s \neq r$*, for all* $t, r, s \geq 1$.

iii) $\{u_t : 1 \leq t \leq n\}$ *is strong mixing with mixing numbers* $\{\alpha(j) : j \geq 1\}$*. There exists* $\zeta > 2$*,* $\epsilon > 1$*, and* $C_\alpha < \infty$*, such that* $\alpha(j) \leq C_\alpha j^{-2\zeta\epsilon/(\zeta-2)}$*, for all* j.

iv) For ζ *defined in (iii),* $E[u_t^{8\zeta}] \leq C_8 < \infty$*, and* $E[u_t^2 \mid \{u_s\}_{s<t}] \geq C_\sigma$ *almost surely.*

Part (i) of Assumption 4.1 assumes the shocks are a martingale difference sequence. This assumption allows for uncorrelated dependent shocks and implies that the shock u_t is uncorrelated with y_{t-1}. Part (ii) in Assumption 4.1 includes a large class of conditional heteroskedastic autoregressive models (e.g., ARCH and GARCH shocks), and it has been common in the literature; for instance, Gonçalves and Kilian (2004) use a similar assumption (Assumption A') to prove the asymptotic consistency of the wild bootstrap for autoregressive processes. Moreover, this assumption implies that the process $\{\xi_t(\rho, h)u_t : 1 \leq t \leq n-h\}$ is serially uncorrelated, where $\xi_t(\rho, h) \equiv \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell}$, which is important for the use of HC standard errors as we discussed in Remark 2.1. Part (iii) and (iv) of Assumption 4.1 are mild regularity conditions on the distribution of the shocks P to establish uniform bounds of approximation errors, which can be relaxed if stronger assumptions are imposed over the serial dependence of the shocks; see Assumption B.1 in Appendix B.

The next assumption is a high-level assumption and imposes additional restrictions on the distributions of the shocks P.

Assumption 4.2.

$$\lim_{M\to\infty} \lim_{n\to\infty} \inf_{|\rho|\leq 1} P_\rho \left(g(\rho, n)^{-2} \; n^{-1} \sum_{t=1}^{n} y_{t-1}^2 \geq 1/M \right) = 1 \; ,$$

where $g(\rho, k) = \left(\sum_{\ell=0}^{k-1} \rho^{2\ell} \right)^{1/2}$.

This assumption implies that the estimator $\hat{\rho}_n(h)$ defined in (5) is well-behaved in the sense that its denominator after scaled by the factor $g(\rho, n-h)$ converges to a strictly positive limit. As a result, we can replace the residual $\hat{u}_t(h) \equiv y_t - \hat{\rho}_n(h) y_{t-1}$ by the shock u_t, which implies the second and third implication discussed in Remark 2.1. We show in Proposition B.1 that Assumption 4.2 can be verified if the shocks are i.i.d. and satisfied mild regularity conditions (Assumption B.1).

Assumptions 4.1 and 4.2 guarantee that the distribution $J_n(\cdot, h, P, \rho)$ defined in (7) can be approximated by the standard normal distribution $\Phi(\cdot)$ uniformly on $\rho \in [-1, 1]$ and h as in (9). Let $\hat{P}_n$ be the empirical distribution of the centered residuals defined in (13) and let $\hat{\rho}_n$ be the estimator of ρ defined in (12). Using this notation $J_n(\cdot, h, \hat{P}_n, \hat{\rho}_n)$ is the distribution of the bootstrap root $R^*_{b,n}(h)$ defined in (14) conditional on the data $Y^{(n)}$. The next theorem shows that the distribution $J_n(\cdot, h, P, \rho)$ can be approximated by the bootstrap distribution $J_n(\cdot, h, \hat{P}_n, \hat{\rho}_n)$ uniformly on $\rho \in [-1, 1]$ and intermediate horizons (e.g., $h \le h_n = o(n)$), i.e., the LP-residual bootstrap is uniformly consistent.

Theorem 4.1. *Suppose Assumptions 4.1 and 4.2 hold. Then, for any* $\epsilon > 0$ *and for any sequence* $h_n \le n$ *such that* $h_n = o(n)$, *we have*

$$\sup_{|\rho| \le 1} P_\rho \left(\sup_{h \le h_n} \sup_{x \in \mathbf{R}} |J_n(x, h, P, \rho) - J_n(x, h, \hat{P}_n, \hat{\rho}_n)| > \epsilon \right) \to 0 \quad as \quad n \to \infty \ , \tag{16}$$

where $J_n(x, h, \cdot, \cdot)$ *is as in* (7), $\hat{P}_n$ *is the empirical distribution of the centered residuals defined in* (13), *and* $\hat{\rho}_n$ *is as in* (12).

Theorem 4.1 shows that the LP-residual bootstrap is uniformly consistent, i.e., the bootstrap distribution $J_n(\cdot, h, \hat{P}_n, \hat{\rho}_n)$ approximates the distribution $J_n(\cdot, h, P, \rho)$ uniformly over the parameter space ($\rho \in [-1, 1]$) and intermediate horizons ($h \le h_n$). Two features of this uniform approximation result deserve further discussion. First, uniform consistency of bootstrap methods over the parameter spaces of autoregressive models is not just a technical detail but a crucial property to guarantee reliable inference methods; see Mikusheva (2007). Otherwise, it is possible to obtain for any sample size n a parameter ρ_n such that the distance between the distributions $J_n(\cdot, h, \hat{P}_n, \hat{\rho}_n)$ and $J_n(\cdot, h, P, \rho)$ is far from zero. Second, the uniform approximation over the horizons is necessary for inference purposes at intermediate horizons. Other valid methods for a fixed h do not necessarily work for h growing with the sample size.

The proof of Theorem 4.1 is presented in Appendix A.1. It has two main ideas. First, we show that the approximation result presented in (9) can be extended for a sequence of AR(1) models,

$$\sup_{P \in \mathbf{P}_{n,0}} \sup_{h \leq h_n} \sup_{|\rho| \leq 1} \sup_{x \in \mathbf{R}} |J_n(x, h, P, \rho) - \Phi(x)| \to 0 \quad \text{as} \quad n \to \infty ,$$

where $\mathbf{P}_{n,0}$ denotes the set of all distributions that satisfy Assumption B.1 in Appendix B.2, h_n is as in Theorem 4.1, $J_n(\cdot, h, P, \rho)$ is as in (7) and $\Phi(\cdot)$ is the standard normal distribution. Assumption B.1 imposes stronger restrictions on the dependence of the shocks (i.i.d. triangular array) and some mild regularity conditions. The formal result is presented in Appendix B.2 as Theorem B.1. Second, we show that Assumptions 4.1 and 4.2 imply the existence of a sequence of events E_n with probability approaching 1 such that the empirical distributions $\hat{P}_n$ conditional on the event E_n verify Assumption B.1. In other words, we show that $\hat{P}_n \in \mathbf{P}_{n,0}$ holds with a probability approaching 1. The construction of the events E_n relies on Lemma B.1 in Appendix B.1. We use the previous two ideas to approximate the distribution $J_n(\cdot, h, \hat{P}_n, \hat{\rho}_n)$ by the standard normal distribution $\Phi(\cdot)$ conditional on the event E_n. Finally, we conclude that the distributions $J_n(\cdot, h, \hat{P}_n, \hat{\rho}_n)$ and $J_n(\cdot, h, P, \rho)$ are asymptotically close since both have the same asymptotic limit.

The next result shows that the confidence interval $C_n^*(h, 1-\alpha)$ defined in (11) is uniformly asymptotically valid in the sense that its asymptotic coverage probability is equal to $1 - \alpha$ uniformly over ρ and h.

Theorem 4.2. *Suppose Assumptions 4.1 and 4.2 hold. Then, for any sequence $h_n \leq n$ such that $h_n = o(n)$, we have*

$$\sup_{|\rho| \leq 1} \sup_{h \leq h_n} |P_\rho (\beta(\rho, h) \in C_n^*(h, 1-\alpha)) - (1-\alpha)| \to 0 \quad \textit{as} \quad n \to \infty , \tag{17}$$

where $\beta(\rho, h)$ and $C_n^*(h, 1-\alpha)$ are as in (2) and (11), respectively.

Theorem 4.2 provides the theoretical justification to conduct inference on the impulse response coefficient $\beta(\rho, h)$ using our bootstrap confidence interval $C_n^*(h, 1-\alpha)$. Note that the only difference with respect to the confidence interval $C_n(h, 1-\alpha)$ defined in (10) is the critical value, which was equal to $z_{1-\alpha/2}$. The critical value $z_{1-\alpha/2}$ was the same for different sample sizes n and horizons h. Instead, we now use a critical value $c_n^*(h, 1-\alpha)$ that depends on the data, the sample size, and the horizon. We evaluate the difference in coverage probability between the confidence intervals $C_n(h, 1-\alpha)$ and $C_n^*(h, 1-\alpha)$ using simulations in Section 6. The simulation results provide evidence that the coverage probability of our proposed confidence interval $C^*_n(h, 1-\alpha)$ is closer to $1-\alpha$ than that of $C_n(h, 1-\alpha)$.

The proof of Theorem 4.2 is presented in Appendix A.2. It only relies on the uniform consistency of the bootstrap procedure. We next sketch the main arguments of the proof. We first note that (17) is equivalent to

$$\sup_{|\rho|\leq 1} \sup_{h\leq h_n} |P_\rho\left(|R_n(h)| \leq c^*_n(h, 1-\alpha)\right) - (1-\alpha)| \to 0 \quad \text{as} \quad n\to\infty \ .$$

We then use that the bootstrap critical value $c^*_n(h, 1-\alpha)$ is included in $[z_{1-\alpha/2-\epsilon},\ z_{1-\alpha/2+\epsilon}]$ with a probability approaching 1 for arbitrary $\epsilon > 0$; see Lemma B.3 in Appendix B.1. This result is possible because the root $R_n(h)$ is asymptotically normal and the LP-residual bootstrap is uniformly consistent. Third, we can conclude using algebra manipulation and the asymptotic normality of the root $R_n(h)$ that

$$\limsup_{n\to\infty} \sup_{|\rho|\leq 1} \sup_{h\leq h_n} |P_\rho\left(|R_n(h)| \leq c^*_n(h, 1-\alpha)\right) - (1-\alpha)| \leq 2\epsilon \ ,$$

which implies (17) since $\epsilon > 0$ was arbitrary.

Remark 4.1. *We can use the LP-residual bootstrap to construct equal-tailed percentile-t confidence intervals denoted by $C^*_{per-t,n}(h, 1-\alpha)$. That is*

$$C^*_{per-t,n}(h, 1-\alpha) \equiv \left[\hat{\beta}_n(h) - q^*_n(h, 1-\alpha/2)\ \hat{s}_n(h),\ \hat{\beta}_n(h) - q^*_n(h, \alpha/2)\ \hat{s}_n(h)\right] \ , \qquad (18)$$

where $\hat{\beta}_n(h)$ is as in (3), $\hat{s}_n(h)$ is as in (4), and $q^*_n(h, \alpha_0)$ is the α_0-quantile of the bootstrap root $R^*_{b,n}(h)$ defined in (14). Three features of $C^*_{per-t,n}(h, 1-\alpha)$ deserve further discussion. First, the bootstrap quantiles $q^*_n(h, \alpha_0)$ can be approximated using Monte Carlo procedures in a similar way as we discussed in Remark 3.1. Second, the confidence interval $C^*_{per-t,n}(h, 1-\alpha)$ can be asymmetric around $\hat{\beta}_n(h)$ by construction, which is not the case of $C^*_n(h, 1-\alpha)$ that is a symmetric one. Third, $C^*_{per-t,n}(h, 1-\alpha)$ is uniformly asymptotically valid,

$$\sup_{|\rho|\leq 1} \sup_{h\leq h_n} \left|P_\rho\left(\beta(\rho, h) \in C^*_{per-t,n}(h, 1-\alpha)\right) - (1-\alpha)\right| \to 0 \quad \text{as} \quad n\to\infty \ ,$$

where $h_n \leq n$ and $h_n = o(n)$. The proof of this claim follows directly by Theorem 4.1, Lemma B.3, and the proof of Theorem 4.2. We include $C^*_{per-t,n}(h, 1-\alpha)$ in our simulation study in Section 6.

Remark 4.2. *For short horizons (fixed h), the confidence interval $C_n^*(h, 1-\alpha)$ based on the LP-residual bootstrap is a simpler alternative to the available but computationally intensive method known as the grid bootstrap (Hansen (1999); Mikusheva (2012)). The grid bootstrap is a method to construct confidence intervals for the parameter $\beta(\rho, h)$ defined in (2) based on test inversion. Mikusheva (2007, 2012) shows that the grid bootstrap provides confidence intervals that are uniformly asymptotically valid in the sense that its asymptotic coverage probability is equal to $1-\alpha$ uniformly on $\rho \in [-1, 1]$; therefore, the grid bootstrap is an alternative method to our confidence intervals for short horizons. However, the grid bootstrap needs to compute bootstrap critical values over a grid of points since it relies on test inversion, which makes its procedure computationally intensive. In contrast, our proposal has the same theoretical properties and requires computing only one bootstrap critical value. Furthermore, our confidence interval $C_n^*(h, 1-\alpha)$ has theoretical guarantees on intermediate horizons (e.g., $h = o(n)$).*

Remark 4.3. *If we restrict our analysis to data-generating processes with weak dependence (e.g., $|\rho| \leq 1-a$ for some $a \in (0, 1)$) and consider stronger assumptions in the distribution of the shocks $\{u_t : 1 \leq t \leq n\}$, then both claims in (16) and (17) can hold for long horizons (e.g., $h_n \leq (1-b)n$ for some $b \in (0, 1)$). In other words, the confidence interval $C_n^*(h, 1-\alpha)$ has theoretical guarantees for long horizons under certain conditions. Assumptions 1-2 in Montiel Olea and Plagborg-Møller (2021) are sufficient to guarantee this claim; a formal proof can be derived following the same strategy presented in Appendix A to prove Theorem 4.1 and 4.2.*

Remark 4.4. *For strictly stationary data, the results in Theorems 4.1 and 4.2 can be extended to vector autoregressive (VAR) models considered in Montiel Olea and Plagborg-Møller (2021) that satisfy their Assumptions 1 and 2. A proof of these extensions may be done using the finite sample inequalities presented in their online appendix and following the approach we presented in Appendixes A and B. We leave the details of a formal proof to future research. For non-stationary data, it is an open question whether the LP-residual bootstrap is consistent for VAR models. Our approach relies on verifying Assumption 4.2 for an appropriate sequence of AR(1) models; therefore, an analogous approach may require a similar step for VAR models, which is out of the scope of this paper.*

5 Asymptotic Refinements

We first discuss informally why the LP-residual bootstrap method provides asymptotic refinements to the confidence intervals in Section 5.1, i.e., the rate of convergence of the error in coverage probability (ECP) of the confidence interval $C_n^*(h, 1-\alpha)$ defined in (11) is faster than that of $C_n(h, 1-\alpha)$ defined in (10). We then present stronger conditions on the data-generating process to provide formal asymptotic refinement results in Section 5.2.

5.1 Why a Bootstrap Method?

The explanation below is not new; see Hall and Horowitz (1996), Horowitz (2001, 2019), and Lahiri (2003). It has the purpose of introducing the main elements and challenges that arise to obtain asymptotic refinements. We first state that the root $R_n(h)$ is *asymptotically pivotal* due to the result in (9), i.e., the distribution of the root $J_n(\cdot, h, P, \rho)$ converges to a limit distribution $\Phi(\cdot)$ that does not depend on the distribution of the shocks P or the parameter ρ. It is often the case that for asymptotically pivotal roots there exist polynomials $q_j(x, h, P, \rho)$ in $x \in \mathbf{R}$ with coefficients that depend on the moments of P and ρ such that (i) $q_j(x, h, P, \rho) = (-1)^{j+1} q_j(-x, h, P, \rho)$ for $j = 1, 2$ and (ii) we can approximate the distribution of the root $R_n(h)$,

$$J_n(x, h, P, \rho) = \Phi(x) + \sum_{j=1}^{2} n^{-j/2} q_j(x, h, P, \rho)\phi(x) + o\left(n^{-1}\right) \ , \tag{19}$$

and the distribution of the bootstrap root $R_{b,n}^*(h)$,

$$J_n(x, h, \hat{P}_n, \hat{\rho}_n) = \Phi(x) + \sum_{j=1}^{2} n^{-j/2} q_j(x, h, \hat{P}_n, \hat{\rho}_n)\phi(x) + o_p\left(n^{-1}\right) \ , \tag{20}$$

where $J_n(x, h, \cdot, \cdot)$ is as in (7), $\hat{P}_n$ is the empirical distribution of the centered residuals defined in (13), and $\hat{\rho}_n$ is the estimator of ρ defined in (12). The approximations in (19) and (20) are known as *Edgeworth expansions* and are commonly used to show that the bootstrap methods provide improvements over the asymptotic distribution theory (e.g., faster rates of convergence); see Hall (1992) for a textbook reference for the case of i.i.d. data. To our knowledge, there are no available theoretical results about valid Edgeworth expansions for the root we use in this paper. In that sense, the approximations in (19)-(20) and the discussion presented below are informal.

We can use the Edgeworth expansions defined in (19)-(20) to compute the rate of convergence of the ECP of the confidence intervals $C_n(h, 1-\alpha)$ and $C_n^*(h, 1-\alpha)$. First, the exact coverage level $P_\rho\left(\beta(\rho, h) \in C_n(h, 1-\alpha)\right)$ is equal to $P_\rho\left(|R_n(h)| \leq z_{1-\alpha/2}\right)$ due to the definitions of $C_n(h, 1-\alpha)$ and $R_n(h)$ in (10) and (6), respectively. Second, note that (19) and the properties of $q_j(\cdot, h, P, \rho)$ implies that for any $x > 0$, we have

$$P_\rho\left(|R_n(h)| \leq x\right) = 2\Phi(x) - 1 + n^{-1} 2 q_2(x, h, P, \rho)\phi(x) + o\left(n^{-1}\right) \ . \tag{21}$$

We conclude taking $x = z_{1-\alpha/2}$ that the rate of convergence of the ECP of the confidence interval $C_n(h, 1-\alpha)$ is $O\left(n^{-1}\right)$. Similarly, we use (20) and the properties of $q_j(\cdot, h, \hat{P}_n, \hat{\rho}_n)$ to obtain

$$\begin{aligned} P_\rho\left(|R_{b,n}^*(h)| \leq x \mid Y^{(n)}\right) &= 2\Phi(x) - 1 + n^{-1} 2 q_2(x, h, \hat{P}_n, \hat{\rho}_n)\phi(x) + o_p\left(n^{-1}\right) \\ &= 2\Phi(x) - 1 + n^{-1} 2 q_2(x, h, P, \rho)\phi(x) + o_p\left(n^{-1}\right) \ , \end{aligned} \tag{22}$$

where the last equality uses that $q_2(x, h, \hat{P}_n, \hat{\rho}_n) = q_2(x, h, P, \rho) + o_p(1)$. Now, we can conclude that $c_n^*(h, 1-\alpha) = c_n(h, 1-\alpha) + o_p\left(n^{-1}\right)$ taking $x = c_n(h, 1-\alpha)$ in (21) and $x = c_n^*(h, 1-\alpha)$ in (22). Since the exact coverage level of $C_n^*(h, 1-\alpha)$ can be written as $P_\rho\left(|R_n(h)| \leq c_n^*(h, 1-\alpha)\right)$ and $c_n^*(h, 1-\alpha) = c_n(h, 1-\alpha) + o_p\left(n^{-1}\right)$, it follows that

$$P_\rho\left(|R_n(h)| \leq c_n^*(h, 1-\alpha)\right) = P_\rho\left(|R_n(h)| \leq c_n(h, 1-\alpha)\right) + o\left(n^{-1}\right) = 1 - \alpha + o\left(n^{-1}\right) \ .$$

We conclude that the rate of convergence of the ECP of the confidence interval $C_n^*(h, 1-\alpha)$ is $o\left(n^{-1}\right)$, which is faster than the rate of convergence of the ECP of $C_n(h, 1-\alpha)$ that is $O\left(n^{-1}\right)$. The informal explanation presented above implies that the LP-residual bootstrap provides asymptotic refinements when (i) there exist valid Edgeworth expansions for the distribution of the root and its bootstrap version as in (19)-(20) and (ii) the polynomials defined in the Edgeworth expansions satisfy $q_j(x, h, P, \rho) = (-1)^{j+1} q_j(-x, h, P, \rho)$ for $j = 1, 2$ and $q_2(x, h, \hat{P}_n, \hat{\rho}_n) = q_2(x, h, P, \rho) + o_p(1)$. We present in Section 5.2 conditions under which the previous informal discussion can be formalized.

Remark 5.1. *The bootstrap methods proposed in Hall and Horowitz (1996) and Andrews (2002) can be adapted for the construction of confidence intervals for the impulse response* $\beta(h, \rho)$ *defined in* (2)*. Four points based on their framework and results deserve further discussion. First, their bootstrap method consists of the nonoverlapping block bootstrap scheme*

(Carlstein (1986)) and overlapping block bootstrap (Kunsch (1989)). Second, they show that their bootstrap methods provide asymptotic refinements to the critical values of t-tests based on generalized method of moments (GMM) estimators $\hat{\theta}_T$ and weakly dependent data $\{Z_t : 1 \leq t \leq n\}$. One of their main conditions is that the series of moment functions $\{g(Z_t, \theta) : t \geq 1\}$ are uncorrelated beyond some finite lags, i.e. for some $\kappa > 0$ we have $E[g(Z_t, \theta)g(Z_s, \theta)'] = 0$ for any $t, s \geq 1$ such that $|t - s| > \kappa$. Third, the LP estimator $\hat{\beta}_n(h)$ defined in (3) can be presented as a GMM estimator using the following dependent data $\{Z_t = (y_{t-1}, y_t, y_{t+h}) : 1 \leq t \leq n\}$ and moment function: $g(y_{t+h}, x_t, \theta) = (y_{t+h} - \theta x_t)x_t$, where $x_t = (y_t, y_{t-1})'$. Then, we can invoke their results and use their bootstrap methods but only for the case of $|\rho| < 1$ and under additional assumptions. Note that their main condition can be verified with $\kappa = h$. Fourth, we can construct confidence intervals for $\beta(\rho, h)$ based on their asymptotic distribution theory. Note that the constructions of these confidence intervals do not require a tuning parameter.

5.2 Formal results

This section presents conditions under which the LP-residual bootstrap provides asymptotic refinements to the confidence interval. Under these conditions, we compute the rate of convergence of the ECP for $C_n(h, 1-\alpha)$ and $C_n^*(h, 1-\alpha)$ in Theorems 5.1 and 5.2, respectively. As a result, we conclude the coverage probability of $C_n^*(h, 1-\alpha)$ is closer to $1-\alpha$ than that of $C_n(h, 1-\alpha)$ for a sufficiently large sample size.

The following assumption imposes stronger conditions on the distribution of the shocks P than the ones presented in Assumption 4.1. We use this assumption to formalize the informal explanation about asymptotic refinements presented in Section 5.1.

Assumption 5.1.

i) $\{u_t : 1 \leq t \leq n\}$ *is a sequence of i.i.d. random variables with* $E[u_t] = 0$.

ii) u_t *has a positive continuous density.*

iii) $E[e^{xu_t}] \leq e^{x^2 c_u^2}$ *for all* $|x| \leq 1/c_u$ *and* $E[u_t^2] \geq C_\sigma$ *for some constants* $c_u, C_\sigma > 0$.

Part (i) of Assumption 5.1 imposes stronger conditions over the serial dependence of the shocks. This assumption is common for theoretical analysis of the asymptotic refinement of the bootstrap method in autoregressive models. An incomplete list of previous research that uses this assumption includes Bose (1988), Park (2003, 2006), and Mikusheva (2015). Parts (ii) and (iii) of Assumption 5.1 are sufficient technical conditions on the distribution of the shocks P to establish the existence of the Edgeworth expansions presented in (19)-(20). Part (ii) implies that the distribution $J_n(\cdot, h, P, \rho)$ defined in (7) is continuous and guarantees that a data-dependent version of the Cramér condition holds, which is a common condition to guarantee the existence of Edgeworth expansions; see Remark 5.4 for further discussion. Part (iii) implies that any sufficiently large number of moments exist and are uniformly bounded by a function of the constant c_u, which is important to guarantee the Edgeworth expansion for the bootstrap distribution $J_n(\cdot, h, \hat{P}_n, \hat{\rho}_n)$. Although this condition is strong, it is not atypical in the literature of the asymptotic refinement of the bootstrap method with dependent data; for instance, Hall and Horowitz (1996) and Inoue and Shintani (2006) assume the existence of 33rd and 36th moments, respectively, while Andrews (2002) assumes that all the moment exists.

We rely on Assumption 5.1, the approach and results presented in Bhattacharya and Ghosh (1978) and Bhattacharya (1987), and the general framework developed by Götze and Hipp (1983) to prove the existence of Edgeworth expansions with dependent data. The framework of Götze and Hipp (1983) requires weakly dependent data and verifying stronger regularity conditions than the ones needed in the case of i.i.d. data; see Hall (1992) and Lahiri (2003) for textbook references. Therefore, we restrict our analysis to data-generating processes with weak dependence (e.g., $|\rho| \leq 1 - a$ for some $a \in (0, 1)$) in a similar way to previous research on asymptotic refinements involving dependent data that includes Bose (1988), Hall and Horowitz (1996), Lahiri (1996), Andrews (2002, 2004), and Inoue and Shintani (2006). It is an open question whether there exist Edgeworth expansion as in (19)-(20) for the case $\rho = 1$.

Theorem 5.1. *Suppose Assumption 5.1 hold. Fix a given $h \in \boldsymbol{N}$ and $a \in (0, 1)$. Then, for any $\rho \in [-1 + a, 1 - a]$, we have*

$$|P_\rho\left(\beta(\rho, h) \in C_n(h, 1 - \alpha)\right) - (1 - \alpha)| = O\left(n^{-1}\right) \ , \tag{23}$$

where $\beta(\rho, h)$ is as in (2) and $C_n(h, 1 - \alpha)$ is as in (10).

Theorem 5.1 provides in (23) the rate of convergence of the ECP of the confidence interval $C_n(h, 1 - \alpha)$. The rate of convergence is the same as the one derived in our informal explanation in Section 5.1. Similar convergence rates were obtained for the ECP of symmetrical confidence intervals in the i.i.d. data case; see Hall (1992) and Horowitz (2001, 2019).

The proof of Theorem 5.1 is presented in Appendix A.3. It uses two main ideas developed previously in the literature. First, we approximate the distribution $J_n(\cdot, h, P, \rho)$ by another distribution $\tilde{J}_n(\cdot, h, P, \rho)$ up to an error of size $O\left(n^{-1-\epsilon}\right)$ for a fixed $\epsilon \in (0, 1/2)$; similar approach has been used in Hall and Horowitz (1996) and Andrews (2002, 2004). Second, we use that the distribution $\tilde{J}_n(\cdot, h, P, \rho)$ admits an Edgeworth expansion up to an error of size $O\left(n^{-3/2}\right)$ based on the results of Bhattacharya and Ghosh (1978) and Götze and Hipp (1983, 1994); see Theorem B.2 in Appendix B.3. These two ideas guarantee the existence of the Edgeworth expansion presented in (19). We then conclude the proof by standard derivations similar to the one derived in our informal explanation presented in Section 5.1.

The next theorem shows that the LP-residual bootstrap provides asymptotic refinements to the confidence intervals. In other words, the rate of convergence of the ECP of our bootstrap confidence interval defined in (11) for $\beta(\rho, h)$ is faster than $O\left(n^{-1}\right)$.

Theorem 5.2. *Suppose Assumption 5.1 hold. Fix a given $h \in \boldsymbol{N}$ and $a \in (0, 1)$. Then, for any $\rho \in [-1 + a, 1 - a]$ and $\epsilon \in (0, 1/2)$, we have*

$$|P_\rho\left(\beta(\rho, h) \in C_n^*(h, 1-\alpha)\right) - (1-\alpha)| = O\left(n^{-1-\epsilon}\right) \ , \tag{24}$$

where $\beta(\rho, h)$ is as in (2) and $C_n^*(h, 1-\alpha)$ is as in (11).

Theorem 5.2 provides in (24) the rate of convergence of the ECP of the confidence interval $C_n^*(h, 1-\alpha)$. The rate of convergence is similar to the one derived in our informal explanation in Section 5.1, but it is slower than the convergence rates obtained for the ECP of symmetrical confidence intervals using bootstrap methods in the i.i.d. data case; see Hall (1992) and Horowitz (2001, 2019).

The proof of Theorem 5.2 is presented in Appendix A.4. It relies on two claims: the existence of the Edgeworth expansion for the distribution $J_n(\cdot, h, P, \rho)$ and the existence of constants C_1 and C_2 such that $P_\rho(|\Delta_n| > C_1 n^{-1-\epsilon}) \leq C_2 n^{-1-\epsilon}$, where $\Delta_n = c_n^*(h, 1-\alpha) - c_n(h, 1-\alpha)$, and $c_n(h, 1-\alpha)$ and $c_n^*(h, 1-\alpha)$ are defined in (8) and (15), respectively. We next sketch the proof based on those two claims. We can derive

$$\begin{aligned}
P_\rho\left(\beta(\rho, h) \in C_n^*(h, 1-\alpha)\right) &= P_\rho\left(|R_n(h)| \leq c_n^*(h, 1-\alpha)\right) \\
&= P_\rho\left(|R_n(h)| \leq C_n(h, 1-\alpha) + \Delta_n, |\Delta_n| \leq n^{-1-\epsilon}\right) + O\left(n^{-1-\epsilon}\right) \\
&= 1 - \alpha + O\left(n^{-1-\epsilon}\right) \ ,
\end{aligned}$$

where the last equality follows from the existence of the Edgeworth expansion for the distribution $J_n(\cdot, h, P, \rho)$ (our first claim), which implies

$$P_\rho \left(|R_n(h)| \le C_n(h, 1-\alpha) + O\left(n^{-1-\epsilon}\right)\right) = 1 - \alpha + O\left(n^{-1-\epsilon}\right) \ .$$

Note that the first claim follows from Theorem 5.1. To prove our second claim, we first show that there is an event E_n such that (i) $J_n(\cdot, h, \hat{P}_n, \hat{\rho}_n)$ has an Edgeworth expansion as in (20) conditional on E_n and (ii) the probability of the complement of E_n is equal to $O\left(n^{-1-\epsilon}\right)$ for a fixed $\epsilon \in (0, 1/2)$; see Lemma B.5 in Appendix B.1. We then follow standard arguments in the literature to prove this claim.

Remark 5.2. *As we mentioned in Remark 5.1, we can use the bootstrap methods presented in Hall and Horowitz (1996) and Andrews (2002, 2004) to construct confidence intervals for $\beta(\rho, h)$ since the LP estimator $\hat{\beta}_n(h)$ defined in (3) can be presented as a GMM estimator. Their results provide rates of convergence of the ECP of these confidence intervals that are qualitatively similar to the one found in Theorem 5.2.*

Remark 5.3. *The rate of convergence of the ECP of $C^*_{per-t,n}(h, 1-\alpha)$ is $O(n^{-1})$. We presented and discussed the equal-tailed percentile-t confidence interval $C^*_{per-t,n}(h, 1-\alpha)$ in Remark 4.1. To compute its rate of convergence, we can use the existence of the Edgeworth expansions presented in (19)-(20) and Theorem 5.2 in Hall (1992). Note that the rate of convergence is similar to the one obtained in (23) for the ECP of the confidence interval $C_n(h, 1-\alpha)$; therefore, the LP-residual bootstrap does not provide asymptotic refinement for equal-tailed percentile-t confidence intervals. Similar conclusions were obtained for the case of i.i.d. data; see Hall (1992) and Horowitz (2001, 2019).*

Remark 5.4. *We use part (ii) of Assumption 5.1 to verify that a dependent-data version of the Cramer condition required in Götze and Hipp (1983) holds, which is an important condition for the existence of the Edgeworth expansion in the dependent-data case. However, verifying that condition is quite difficult in general, as pointed out by Hall and Horowitz (1996) and Götze and Hipp (1994), among others. Therefore, we proceed in two steps based on the results by Götze and Hipp (1994) that propose simple and verifiable conditions to guarantee the conditions required by Götze and Hipp (1983), including the dependent-data version of the Cramer condition. We first approximate the distribution $J_n(\cdot, h, P, \rho)$ by a distribution $\tilde{J}_n(\cdot, h, P, \rho)$. We then use part (ii) of Assumption 5.1 to verify the conditions required in Theorem 1.2 of Götze and Hipp (1994), which guarantee the existence of Edgeworth expansion for the distribution $\tilde{J}_n(\cdot, h, P, \rho)$.*

Remark 5.5. *For strictly stationary data-generating processes, the results in Theorems 5.1 and 5.2 can be extended to the family of vector autoregressive (VAR) models that satisfy similar assumptions to the ones presented in Assumption 5.1, which are stronger than Assumptions 1 and 2 in Montiel Olea and Plagborg-Møller (2021). These extensions can be shown by verifying the conditions required in Götze and Hipp (1994). We leave the details of a formal proof for the VAR models for future research.*

6 Simulation Study

We examine the finite sample performance of $C_n^*(h, 1-\alpha)$ defined in (11) using different data-generating processes. We consider a sample size $n = 95$, which is the median sample size based on 71 papers that have utilized the LP approach; see Herbst and Johannsen (2021). Additionally, we examine other confidence intervals presented in the paper.

6.1 Monte-Carlo Design

We use four designs for the distribution of the shocks $\{u_t : 1 \le t \le n\}$ and two values for the parameter $\rho \in \{0.95, 1\}$ in our Monte-Carlo simulation. The shocks are defined according to the GARCH(1,1) model:

$$u_t = \tau_t v_t, \quad \tau_t^2 = \omega_0 + \omega_1 u_{t-1}^2 + \omega_2 \tau_{t-1}^2, \quad v_t \text{ are } i.i.d. \,,$$

where the distribution of v_t and the parameter vector $(\omega_0, \omega_1, \omega_2)$ are specified as follows:

Design 1: $v_t \sim N(0,1)$, $\omega_0 = 1$, and $\omega_1 = \omega_2 = 0$.

Design 2: $v_t \sim N(0,1)$, $\omega_0 = 0.05$, $\omega_1 = 0.3$, and $\omega_2 = 0.65$.

Design 3: $v_t \sim t_4/\sqrt{2}$, $\omega_0 = 1$, and $\omega_1 = \omega_2 = 0$.

Design 4: $v_t | B_t = j \sim N(m_j, \sigma_j^2)$, where $B_n \in \{0,1\}$, $B_n = 1$ with probability $p = 0.25$, $m_0 = 2/\sigma_2$, $m_1 = -6/\sigma_2$, $\sigma_0 = 0.5/\sigma_2$, $\sigma_1 = 2/\sigma_2$, and $\sigma_2^2 = p(m_1^2 + \sigma_1) + (1-p)(m_0^2 + \sigma_0)$, $\omega_0 = 0.05$, $\omega_1 = 0.3$, and $\omega_2 = 0.65$.

For each design for the distribution of the shock and values of ρ, we consider eight different confidence intervals. All our confidence intervals use the HC standard errors $\hat{s}_n(h)$ defined in (4). Additionally, we consider alternative HC standard errors $\hat{s}_{j,n}(h)$ defined as

$$\hat{s}_{j,n}(h) \equiv \left(\sum_{t=1}^{n-h} \hat{u}_t(h)^2\right)^{-1/2} \left(\sum_{t=1}^{n-h} \hat{\xi}_{j,t}(h)^2 \hat{u}_t(h)^2\right)^{1/2} \left(\sum_{t=1}^{n-h} \hat{u}_t(h)^2\right)^{-1/2},$$

for $j = 2, 3$, where $\hat{\xi}_{2,t}(h)^2 = \hat{\xi}_t(h)^2/(1 - \mathbb{P}_{h,tt})$ and $\hat{\xi}_{3,t}(h)^2 = \hat{\xi}_t(h)^2/(1 - \mathbb{P}_{h,tt})^2$. We use the projection matrix $\mathbb{P}_h = \mathbb{X}_h(\mathbb{X}_h'\mathbb{X}_h)^{-1}\mathbb{X}_h'$, where $\mathbb{X}_h$ is a matrix with row elements equal to $(\hat{u}_t(h),\ y_{t-1})$ for $t = 1, \ldots, n-h$. The confidence intervals that we use are listed below

1. **RB:** confidence interval as in (11) based on the LP-residual bootstrap.
2. **RB**$_{per-t}$**:** equal-tail percentile-t confidence interval as in (18). It is based on the LP-residual bootstrap and discussed in Remark 4.1.
3. **RB**$_{hc3}$**:** confidence interval as in (11) but using $\hat{s}_{3,n}(h)$ and $c^*_{3,n}(h, 1-\alpha)$ instead of $\hat{s}_n(h)$ and $c^*_n(h, 1-\alpha)$, where $c^*_{3,n}(h, 1-\alpha)$ is computed as in Section 3.1 but using $\hat{s}^*_{3,n}(h)$ instead of $\hat{s}^*_n(h)$.
4. **WB:** confidence interval as in (11) but using $c^{wb,*}_n(h, 1-\alpha)$ instead of $c^*_n(h, 1-\alpha)$, where $c^{wb,*}_n(h, 1-\alpha)$ is based on the LP-wild bootstrap; see Remark 3.2.
5. **WB**$_{per-t}$**:** equal-tail percentile-t confidence interval as in (18) but using $q^{wb,*}_n(h, \alpha_0)$ instead of $q^*_n(h, \alpha_0)$, where $q^{wb,*}_n(h, \alpha_0)$ is based on the LP-wild bootstrap discussed in Remark 3.2.
6. **AA:** standard confidence interval as in (10).
7. **AA**$_{hc2}$**:** standard confidence interval as in (10) but using $\hat{s}_{2,n}(h)$ instead of $\hat{s}_n(h)$.
8. **AA**$_{hc3}$**:** standard confidence interval as in (10) but using $\hat{s}_{3,n}(h)$ instead of $\hat{s}_n(h)$.

6.2 Discussion and Results

In all the designs, the shocks have zero mean and variance one. Designs 1-2 verify Assumption 4.1 presented in Section 4. Design 1 also verifies Assumption 4.2 due to Proposition B.1 in Appendix B.2. Assumption 4.2 can be tedious to verify in general since it involves computing a probability for all the parameters ρ in the parameter space and taking their infimum. In contrast, designs 3-4 do not verify all the parts of Assumption 4.1. Design 3 considers shocks

ρ	h	RB	RB_{per-t}	RB_{hc3}	WB	WB_{per-t}	AA	AA_{hc2}	AA_{hc3}
Design 1: Gaussian i.i.d. shocks									
0.95	1	90.04	89.60	90.08	90.38	90.32	88.26	89.12	89.60
	6	89.36	88.98	89.38	90.46	90.22	85.00	85.58	86.44
	12	88.12	86.96	88.08	89.60	88.28	83.78	84.44	85.34
	18	87.96	86.08	87.88	89.46	88.08	84.44	85.16	85.86
1.00	1	90.20	89.80	90.30	90.48	90.34	88.30	88.90	89.66
	6	89.80	89.44	89.80	90.68	90.22	83.54	84.42	85.28
	12	87.92	87.60	87.90	88.78	89.02	80.32	81.30	81.94
	18	86.22	84.76	86.22	87.02	86.36	78.34	79.16	79.98
Design 2: Gaussian GARCH shocks									
0.95	1	88.86	89.00	89.40	90.18	90.02	86.84	88.10	89.16
	6	87.94	88.00	88.26	90.12	90.74	83.64	84.52	85.60
	12	87.08	85.72	87.28	88.72	88.18	82.96	83.90	84.88
	18	86.36	84.36	86.40	87.98	86.94	82.76	83.44	84.38
1.00	1	88.64	88.82	89.14	89.96	89.94	86.72	87.84	88.90
	6	88.96	88.52	89.08	90.76	90.96	82.34	83.76	84.52
	12	86.64	86.08	86.60	88.56	88.68	79.14	80.46	81.32
	18	84.90	83.74	84.78	86.56	86.52	76.64	77.74	78.70

Table 1: Coverage probability (in %) of confidence intervals for $\beta(\rho, h)$ with a nominal level of 90% and $n = 95$. 5,000 simulations and 1,000 bootstrap iterations.

without a fourth moment, i.e., it does not verify part (iv) of Assumption 4.1, which was a regularity condition. Design 4 considers a distribution of the shocks (GARCH errors with asymmetric v and nonzero skewness) that lie outside the class of conditional heteroskedastic processes that we consider in this paper, i.e., it does not verify part (ii) of Assumption 4.1. As we discussed in Remark 2.1, this assumption was a sufficient one for the validity of the HC standard errors $\hat{s}_n(h)$ in the construction of confidence intervals.

Tables 1 and 2 report the coverage probabilities (in %) of our simulations. Columns are labeled as the confidence intervals we specified in Section 6.1. For all the designs on the distribution of the shock and values of ρ, we use 5000 simulations to generate data with a sample size $n = 95$ based on the AR(1) model (1). In each simulation, we compute the eight confidence intervals described above for horizons $h \in \{1, 6, 12, 18\}$. The confidence intervals have a nominal level equal to $1 - \alpha = 90\%$. The bootstrap critical values are computed using $B = 1000$ as described in Remark 3.1. We summarize our findings from the simulations below.

Four features of Table 1 deserve discussion. First, it shows that our recommended confidence interval **RB** has a coverage probability closer to 90% than the confidence intervals

ρ	h	RB	RB_{per-t}	RB_{hc3}	WB	WB_{per-t}	AA	AA_{hc2}	AA_{hc3}
Design 3: t-student i.i.d. shocks									
0.95	1	90.00	90.08	90.36	90.52	90.32	88.04	89.24	90.26
	6	89.08	88.48	89.28	89.76	89.64	84.04	85.40	86.66
	12	87.74	86.18	87.90	88.46	87.42	82.78	84.24	85.46
	18	88.08	85.38	88.26	89.12	87.52	83.36	84.80	86.20
1.00	1	89.96	89.88	90.16	90.36	89.98	87.74	88.82	90.16
	6	89.78	88.60	89.84	90.52	89.84	82.88	84.54	85.78
	12	87.56	86.82	87.64	88.40	88.22	79.04	80.30	81.56
	18	85.64	84.40	86.00	86.80	86.24	77.50	78.84	80.22
Design 4: mix-gaussian GARCH shocks)									
0.95	1	89.00	89.86	89.32	88.80	89.60	86.38	87.20	87.88
	6	87.90	90.62	88.14	89.12	92.04	84.30	85.30	86.18
	12	84.14	86.64	84.00	85.58	87.98	80.70	81.52	82.32
	18	83.48	84.70	83.66	85.32	86.88	80.46	81.40	82.56
1.00	1	88.84	90.24	89.04	88.98	89.70	86.60	87.24	88.00
	6	88.24	91.26	88.50	89.62	92.66	82.78	83.82	84.64
	12	84.96	88.54	85.08	86.74	89.86	77.40	78.32	79.50
	18	82.30	84.62	82.34	83.90	86.30	74.18	75.28	76.14

Table 2: Coverage probability (in %) of confidence intervals for $\beta(\rho, h)$ with a nominal level of 90% and $n = 95$. 5,000 simulations and 1,000 bootstraps iterations.

AA, $\mathbf{AA}_{hc2}$, and $\mathbf{AA}_{hc3}$ for all the designs 1-3, values of ρ, and horizons h, with some few exceptions. The lowest coverage probability of **RB**, **AA**, $\mathbf{AA}_{hc2}$, and $\mathbf{AA}_{hc3}$ are 85%, 77%, 78%, and 79%, respectively, and occur when $\rho = 1$ and horizon $h = 18$. Second, **RB** and $\mathbf{RB}_{hc3}$ have better performance than $\mathbf{RB}_{per-t}$, especially when $\rho = 1$ and the horizon is a significant fraction of the sample size ($h \in \{12, 18\}$). Third, **WB** and $\mathbf{WB}_{per-t}$ have larger coverage probability than **RB** for all the designs 1-3, values of ρ, and horizons h, with some few exceptions. The larger coverage of **WB** and $\mathbf{WB}_{per-t}$ is associated with a larger median length of their confidence intervals, as we reported in Table D.1 in Supplemental Appendix D. Fourth, $\mathbf{AA}_{hc3}$ presents a coverage probability closer to 90% and larger than **AA** and $\mathbf{AA}_{hc2}$ for all the designs 1-3, values of ρ, and horizons h. This finding suggests that using $\hat{s}_{3,n}(h)$ instead of $\hat{s}_n(h)$ can improve the coverage probability of the confidence interval; however, confidence intervals based on bootstrap methods (e.g., **RB** and $\mathbf{WB}_{per-t}$) report coverage probability closer to 90%.

Table 2 presents results for designs 3-4. Our findings for design 3 are qualitatively similar to Table 1, which was discussed above. This suggests that failing part (iv) of Assumption 4.1 (a regularity condition) does not have a major effect on the coverage probability of the confidence intervals that we considered. In contrast, design 4 shows that some of our qualitative findings can change if we fail to verify part (ii) of Assumption 4.1. This result is consistent with existing theory since this assumption was a sufficient condition for the validity of confidence intervals that use HC standard errors $\hat{s}_n(h)$; see Remark 2.1. In particular, $\mathbf{RB}_{per-t}$ has a coverage probability closer to 90% and larger than **RB** and $\mathbf{RB}_{hc3}$. The small sample size ($n = 95$) does not explain the findings for design 5. We obtain similar results for a sample size $n = 240$ in Table D.2 in Supplemental Appendix D.

7 Concluding Remarks

This paper contributes to a growing literature on confidence interval construction for impulse response coefficients based on the local projection approach. Specifically, we propose the LP-residual bootstrap method to construct confidence intervals for the impulse response coefficients of AR(1) models at intermediate horizons. We prove two theoretical properties of this method: uniform consistency and asymptotic refinements. For a large class of AR(1) models that allow for a unit root, we show that the proposed confidence interval $C_n^*(h, 1-\alpha)$ defined in (11) has an asymptotic coverage probability equal to its nominal level $1-\alpha$ uniformly over the parameter space (e.g., $\rho \in [-1, 1]$) and intermediate horizons. For a restricted class of AR(1) models, we demonstrate that the rate of convergence of the error in coverage probability of $C_n^*(h, 1-\alpha)$ is faster than that of $C_n(h, 1-\alpha)$, where $C_n(h, 1-\alpha)$ is a confidence interval for the impulse response coefficients at intermediate horizons based on the existing asymptotic distribution theory in Section 2.1. Our finding implies that the coverage probability of $C_n^*(h, 1-\alpha)$ can be closer to $1-\alpha$ than that of $C_n(h, 1-\alpha)$ for sufficiently large sample sizes.

This paper considered the AR(1) model as the first step in understanding the theoretical properties of the LP-residual bootstrap. Three possible directions exist for future research. First, the uniform consistency of the LP-residual bootstrap method is an open question for the general vector auto-regressive (VAR) model. Second, the asymptotic refinement property of this method is unknown for the unit-root model ($\rho = 1$) or general VAR models. Third, future work is needed to prove the uniform consistency of the LP-wild bootstrap discussed in Remark 3.2.

A Proofs of Result in Main Text

A.1 Proof of Theorem 4.1

We prove a stronger result:

$$\sup_{|\rho|\le 1} P_\rho\left(\sup_{h\le h_n}\ \sup_{|\tilde\rho|\le 1}\ \sup_{x\in\mathbf{R}} |J_n(x,h,P,\tilde\rho) - J_n(x,h,\hat P_n,\hat\rho_n)| > \epsilon\right) \to 0 \quad \text{as} \quad n\to\infty ,$$

which is sufficient to conclude (16). The proof has three steps.

Step 1: Let $E_{n,1} = \{g(\rho,n)\ n^{1/2}\ |\hat\rho_n - \rho| > M\}$, $E_{n,2} = \{|n^{-1}\sum_{t=1}^n \tilde u_t^2 - \sigma^2| > \sigma^2/2\}$, and $E_{n,3} = \{n^{-1}\sum_{t=1}^n \tilde u_t^4 > \tilde K_4\}$ be events, where M and $\tilde K_4$ are constant defined next. Fix $\eta > 0$. We use Lemma B.1 to guarantee the existence of M, $\tilde K_4$, and $N_0 = N_0(\eta)$ such that $P_\rho(E_{n,j}) < \eta/3$ for $j = 1,2,3$, $n \ge N_0$ and $\rho \in [-1,1]$. Define $E_n = E_{n,1}^c \cap E_{n,2}^c \cap E_{n,3}^c$. By construction $P_\rho(E_n) > 1-\eta$ for $n \ge N_0$ and for any $\rho\in[-1,1]$.

Step 2: Conditional on the event E_n, we have $|\hat\rho_n - \rho| \le Mn^{-1/2}/g(\rho,n)$ for $n\ge N_0$ and for any $\rho\in[-1,1]$. Therefore, conditional on the event E_n, we can use Lemma B.2 to conclude the existence of $\tilde M$ and $N_1 \ge N_0$ such that $|\hat\rho_n| \le 1 + \tilde M/n$ for all $n \ge N_1$. Note also that conditional on the event E_n, we have that distribution $\hat P_n$ of the centered residuals defined in (13) verifies Assumption B.1 taking $K_4 = M$, $\underline{\sigma} = \sigma^2/2$, and $\overline{\sigma} = 3\sigma^2/2$, i.e., $\hat P_n \in \mathbf{P}_{n,0}$, where $\mathbf{P}_{n,0}$ is defined in Appendix B.2.

Step 3: We use Theorem B.1 taking $M = \tilde M$. This implies that for any $\epsilon > 0$, there exists $N_2 = N_2(\epsilon,\eta) \ge N_1$ such that $\sup_{x\in\mathbf{R}}\ |J_n(x,h,P_n,\rho) - \Phi(x)| < \epsilon/2$, for any $n \ge N_2$, $|\rho| \le 1 + \tilde M/n$, $h \le h_n \le n$ and $h_n = o(n)$, and $P_n \in \mathbf{P}_{n,0}$. Conditional on E_n, we have $\hat P_n \in \mathbf{P}_{n,0}$ due to Step 2, then

$$\sup_{h\le h_n}\ \sup_{x\in\mathbf{R}} \left|J_n(x,h,\hat P_n,\hat\rho_n) - \Phi(x)\right| < \epsilon/2 , \tag{A.1}$$

for any $n \ge N_2$, $h_n \le n$ and $h_n = o(n)$. By (9) there exists $N_3 \ge N_2$ such that

$$\sup_{h\le h_n}\ \sup_{\tilde\rho\in[-1,1]}\ \sup_{x\in\mathbf{R}} |J_n(x,h,P,\tilde\rho) - \Phi(x)| < \epsilon/2 ,$$

for any $n \ge N_3$, $h_n \le n$, and $h_n = o(n)$. Therefore, conditional on the event E_n and using triangular inequality, we conclude that

$$\sup_{h \leq h_n} \sup_{\tilde{\rho} \in [-1,1]} \sup_{x \in \mathbf{R}} \left| J_n(x,h,P,\tilde{\rho}) - J_n(x,h,\hat{P}_n,\hat{\rho}_n) \right| < \epsilon \ ,$$

for any $n \geq N_3$, $h_n \leq n$, and $h_n = o(n)$. Since $P_\rho(E_n) \geq 1-\eta$ for any $\rho \in [-1,1]$, the previous conclusion is equivalent to

$$\sup_{\rho \in [-1,1]} P\left(\sup_{h \leq h_n} \sup_{\tilde{\rho} \in [-1,1]} \sup_{x \in \mathbf{R}} \left| J_n(x,h,P,\tilde{\rho}) - J_n(x,h,\hat{P}_n,\hat{\rho}_n) \right| < \epsilon \right) \geq 1-\eta \ ,$$

for any $n \geq N_3$, $h_n \leq n$ and $h_n = o(n)$, which concludes the proof of the theorem.

A.2 Proof of Theorem 4.2

By Lemma B.3, for any $\epsilon > 0$, there exists $N_0 = N_0(\epsilon)$ such that

$$P_\rho \left(z_{1-\alpha/2-\epsilon/2} \leq c_n^*(h, 1-\alpha) \leq z_{1-\alpha/2+\epsilon/2} \right) \geq 1-\epsilon \ , \tag{A.2}$$

for any $n \geq N_0$, $\rho \in [-1,1]$ and any $h \leq h_n \leq n$ and $h_n = o(n)$. Assumptions 4.1 and 4.2 guarantee (9); therefore, there exist $N_1 \geq N_0$ such that

$$P_\rho \left(|R_n(h)| \leq z_{1-\alpha/2+\epsilon/2} \right) \leq 1-\alpha+2\epsilon \quad \text{and} \quad P_\rho \left(|R_n(h)| \leq z_{1-\alpha/2-\epsilon/2} \right) \geq 1-\alpha-2\epsilon \ , \tag{A.3}$$

for any $n \geq N_1$, $\rho \in [-1,1]$ and any $h \leq h_n \leq n$ and $h_n = o(n)$. Consider the derivation

$$\begin{aligned} P_\rho \left(\beta(\rho,h) \in C_n^*(h,1-\alpha) \right) &= P_\rho \left(|R_n(h)| \leq c_n^*(h,1-\alpha) \right) \\ &= P_\rho \left(|R_n(h)| \leq c_n^*(h,1-\alpha), c_n^*(h,1-\alpha) > z_{1-\alpha/2+\epsilon/2} \right) \\ &\quad + P_\rho \left(|R_n(h)| \leq c_n^*(h,1-\alpha), c_n^*(h,1-\alpha) \leq z_{1-\alpha/2+\epsilon/2} \right) \\ &\leq P_\rho \left(c_n^*(h,1-\alpha) > z_{1-\alpha/2+\epsilon/2} \right) + P_\rho \left(|R_n(h)| \leq z_{1-\alpha/2+\epsilon/2} \right) \\ &\leq \epsilon + 1 - \alpha + 2\epsilon \ , \end{aligned}$$

where the last inequality follows by (A.2) and (A.3). Similarly, we obtain the inequality

$$P_\rho \left(|R_n(h)| \leq z_{1-\alpha/2-\epsilon/2} \right) \leq P_\rho \left(\beta(\rho,h) \in C_n^*(h,1-\alpha) \right) + P_\rho \left(c_n^*(h,1-\alpha) < z_{1-\alpha/2-\epsilon/2} \right) \ ,$$

which implies that $P_\rho\left(\beta(\rho,h)\in C_n^*(h,1-\alpha)\right)\geq 1-\alpha-2\epsilon-\epsilon$. We conclude that for any $n\geq N_1$, $\rho\in[-1,1]$ and any $h\leq h_n\leq n$ and $h_n=o\left(n\right)$, we have

$$\left|P_\rho\left(\beta(\rho,h)\in C_n^*(h,1-\alpha)\right)-(1-\alpha)\right|\leq 3\epsilon\ ,$$

which completes the proof of Theorem 4.2.

A.3 Proof of Theorem 5.1

We first show that $J_n(x,h,P,\rho)$ admits a valid Edgeworth expansion, that is

$$\sup_{x\in\mathbf{R}}\left|J_n(x,h,P,\rho)-\left(\Phi(x)+\sum_{j=1}^{2}n^{-j/2}q_j(x,h,P,\rho)\phi(x)\right)\right|=O\left(n^{-1-\epsilon}\right)\qquad\text{(A.4)}$$

for some $\epsilon\in(0,1/2)$, where $q_j(x,h,P,\rho)$ are polynomials on x with coefficients that are continuous functions of the moments of P (up to order 12) and ρ. Furthermore, we have $q_1(x,h,P,\rho)=q_1(-x,h,P,\rho)$ and $q_2(x,h,P,\rho)=-q_2(-x,h,P,\rho)$.

To show (A.4), we first use Lemma B.4 to approximate $J_n(x,h,P,\rho)$ by $\tilde{J}_n(x,h,P,\rho)$,

$$\sup_{x\in\mathbf{R}}\left|J_n(x,h,P,\rho)-\tilde{J}_n(x,h,P,\rho)\right|=D_n+O\left(n^{-1-\epsilon}\right)\ ,$$

for some $\epsilon\in(0,1/2)$, where

$$D_n=\sup_{x\in\mathbf{R}}\left|\tilde{J}_n(x+n^{-1-\epsilon},h,P,\rho)-\tilde{J}_n(x-n^{-1-\epsilon},h,P,\rho)\right|\ .$$

Due to Theorem B.2, we can conclude $D_n=O\left(n^{-1-\epsilon}\right)$. We then use Theorem B.2 to approximate $\tilde{J}_n(x,h,P,\rho)$ by a valid Edgeworth expansion,

$$\sup_{x\in\mathbf{R}}\left|\tilde{J}_n(x,h,P,\rho)-\left(\Phi(x)+\sum_{j=1}^{2}n^{-j/2}q_j(x,h,P,\rho)\phi(x)\right)\right|=O\left(n^{-3/2}\right)\ .$$

Note that we can use Theorem B.2 since Assumption 5.1 implies Assumption B.2 and the distribution $\tilde{J}_n(x,h,P,\rho)$ that we obtain from Lemma B.4 satisfy the required conditions. We conclude (A.4) by triangular inequality. The polynomials q_j that appear in (A.4) are the polynomials in the Edgeworth expansion of $\tilde{J}_n(x,h,P,\rho)$.

Now, we show that $P_\rho(|R_n(h)| \le x)$ also admits an asymptotic approximation, that is

$$\sup_{x\in\mathbf{R}} \left| P_\rho(|R_n(h)| \le x) - \left(2\Phi(x) - 1 + 2n^{-1}q_2(x,h,P,\rho)\phi(x)\right)\right| = O\left(n^{-1-\epsilon}\right) , \tag{A.5}$$

where $q_2(x,h,P,\rho)$ and $\epsilon \in (0,1/2)$ are defined in (A.4). Note that (23) follows from (A.5) since we can write (23) as follows

$$\left|P_\rho\left(|R_n(h)| \le z_{1-\alpha/2}\right) - (1-\alpha)\right| = O\left(n^{-1}\right) ,$$

and the previous expression is what we obtain taking $x = z_{1-\alpha/2}$ in (A.5), where we used that $1-\alpha = 2\Phi(z_{1-\alpha/2}) - 1$ holds by definition of $z_{1-\alpha/2}$.

To show (A.5), we first write

$$P_\rho(|R_n(h)| \le x) = J_n(x,h,P,\rho) - J_n(-x,h,P,\rho) + r_n(x) ,$$

where $r_n(x) = P_\rho(R_n(h) = -x)$. We then use (A.4) to approximate $J_n(\cdot,h,P,\rho)$ and the properties of the polynomials $q_j(\cdot,h,P,\rho)$ to obtain the following approximation

$$\sup_{x\in\mathbf{R}} \left| P_\rho(|R_n(h)| \le x) - \left(2\Phi(x) - 1 + 2n^{-1}q_2(x,h,P,\rho)\phi(x) + r_n(x)\right)\right| = O\left(n^{-1-\epsilon}\right) .$$

Finally, $\sup_{x\in\mathbf{R}} r_n(x) = O\left(n^{-1-\epsilon}\right)$ since $r_n(x) \le P_\rho\left(R_n(h) \in (-x - n^{-1-\epsilon}, -x]\right)$ and (A.4) holds. We use this in the previous expression to complete the proof of (A.5).

A.4 Proof of Theorem 5.2

The proof has two parts. In the first part we assume that $P(|\Delta_n| > C_1 n^{-1-\epsilon}) \le C_2 n^{-1-\epsilon}$ for some constants C_1 and C_2, where $\Delta_n = c_n^*(h,1-\alpha) - c_n(h,1-\alpha)$. We use this assumption to prove the theorem. In the second part, we prove the assumption of the first part.

Part 1: By (11), we have $P_\rho\left(\beta(\rho,h) \in C_n^*(h,1-\alpha)\right) = P_\rho\left(|R_n(h)| \le c_n^*(h,1-\alpha)\right)$. We can write this term as the sum of $P_\rho\left(|R_n(h)| \le c_n(h,1-\alpha) + \Delta_n, |\Delta_n| \le C_1 n^{-1-\epsilon}\right)$ and $P_\rho\left(|R_n(h)| \le c_n(h,1-\alpha) + \Delta_n, |\Delta_n| > C_1 n^{-1-\epsilon}\right)$. Using our assumption, we conclude $P_\rho\left(\beta(\rho,h) \in C_n^*(h,1\right.$ is equal to

$$P_\rho\left(|R_n(h)| \le c_n(h,1-\alpha) + \Delta_n, |\Delta_n| \le C_1 n^{-1-\epsilon}\right) + O\left(n^{-1-\epsilon}\right) .$$

By (A.5) in the proof of Theorem 5.1, we have

$$P_\rho\left(|R_n(h)| \le x + zn^{-1-\epsilon}\right) = P_\rho\left(|R_n(h)| \le x\right) + O\left(n^{-1-\epsilon}\right)$$

for $z = -C_1, C_1$ and any $x \in \mathbf{R}$. Since

$$P_\rho\left(|R_n(h)| \le x + \Delta_n, |\Delta_n| \le C_1 n^{-1-\epsilon}\right) \le P_\rho\left(|R_n(h)| \le x + C_1 n^{-1-\epsilon}\right)$$

and

$$P_\rho\left(|R_n(h)| \le x + \Delta_n, |\Delta_n| \le C_1 n^{-1-\epsilon}\right) \ge P_\rho\left(|R_n(h)| \le x - C_1 n^{-1-\epsilon}\right) + O\left(n^{-1-\epsilon}\right) ,$$

we conclude $P_\rho\left(|R_n(h)| \le x + \Delta_n, |\Delta_n| \le n^{-1-\epsilon}\right) = P_\rho\left(|R_n(h)| \le x\right) + O\left(n^{-1-\epsilon}\right)$. Taking $x = c_n(h, 1-\alpha)$ and using that $P_\rho\left(|R_n(h)| \le c_n(h, 1-\alpha)\right) = 1 - \alpha$ (due to part 2 in Assumption 5.1), we conclude $P_\rho\left(\beta(\rho, h) \in C_n^*(h, 1-\alpha)\right) = 1 - \alpha + O\left(n^{-1-\epsilon}\right)$.

Part 2: Fix $\epsilon \in (0, 1/2)$. Define $E_{n,1} = \{|\hat{\rho}_n| \le 1 - a/2\}$, $E_{n,2} = \{n^{-1}\sum_{t=1}^n \tilde{u}_t^2 \ge \tilde{C}_\sigma\}$, $E_{n,3} = \{n^{-1}\sum_{t=1}^n \tilde{u}_t^{4k} \le M\}$, and $E_{n,4} = \{\max_{1\le r\le 12} |n^{-1}\sum_{t=1}^n \tilde{u}_t^r - E[u_t^r]| \le n^{-\epsilon}\}$, where $\tilde{C}_\sigma$ and M are as in Lemma B.5. Define $E_n = E_{n,1} \cap E_{n,2} \cap E_{n,3} \cap E_{n,4}$. By Lemma B.5 and Assumption 5.1, it follows that $P(E_n^c) \le C_2 n^{-1-\epsilon}$ for some constant $C_2 = C_2(a, h, k, C_\sigma, \epsilon, c_u)$. Note that conditional on the event E_n, we can use Lemma B.4 for the distribution of the bootstrap root $R_n^*(h)$. That is

$$\sup_{x\in\mathbf{R}} |J_n(x, h, \hat{P}_n, \hat{\rho}_n) - \tilde{J}_n(x, h, \hat{P}_n, \hat{\rho}_n)| \le D_n + n^{-1-\epsilon} C\left(n^{-1}\sum_{t=1}^n |\tilde{u}_t|^k + \tilde{u}_t^{2k} + \tilde{u}_t^{4k}\right) ,$$

for some constant C, where

$$D_n = \sup_{x\in\mathbf{R}} \left|\tilde{J}_n(x + n^{-1-\epsilon}, h, \hat{P}_n, \hat{\rho}_n) - \tilde{J}_n(x - n^{-1-\epsilon}, h, \hat{P}_n, \hat{\rho}_n)\right| .$$

By Theorem B.3, there is an Edgeworth expansion for $\tilde{J}_n(x, h, \hat{P}_n, \hat{\rho}_n)$ conditional on E_n. This implies $D_n \le Cn^{-1-\epsilon}$ conditional on E_n, for some constant C. Similarly, conditional on E_n, $n^{-1}\sum_{t=1}^n \left(|\tilde{u}_i|^k + \tilde{u}_t^{2k} + \tilde{u}_t^{4k}\right) \le C$, for some constant C that depends on M. We conclude that, conditional on E_n, $J_n(x, h, \hat{P}_n, \hat{\rho}_n)$ has the following Edgeworth expansion,

$$\sup_{x\in\mathbf{R}} \left|J_n(x, h, \hat{P}_n, \hat{\rho}_n) - \left(\Phi(x) + \sum_{j=1}^2 n^{-j/2} q_j(x, h, \hat{P}_n, \hat{\rho}_n)\phi(x)\right)\right| \le Cn^{-1-\epsilon} .$$

The properties of $q_j(x, h, \hat{P}_n, \hat{\rho}_n)$ from Theorem B.3 and arguments from the proof of Theorem 5.1 imply

$$\sup_{x \in \mathbf{R}} \left| P_\rho \left(|R_n^*(h)| \le x \mid Y^{(n)} \right) - \Big(2\Phi(x) - 1 + 2n^{-1} q_2(x, h, \hat{P}_n, \hat{\rho}_n)\phi(x) \Big) \right| \le C n^{-1-\epsilon} \ .$$

Recall that the coefficients of $q_2(x, h, \hat{P}_n, \hat{\rho}_n)$ are polynomial of the moments of $\hat{P}_n$ (up-to order 12) and $\hat{\rho}_n$. Conditional on E_n, we known the moments of $\hat{P}_n$ are close to the moments of P: $|n^{-1}\sum_{t=1}^n \tilde{u}_t^r - E[u_t^r]| \le n^{-\epsilon}$ for $r = 1, \dots, 12$. Therefore, conditional on E_n, we have

$$\sup_{x \in \mathbf{R}} \left| P_\rho \left(|R_n^*(h)| \le x \mid Y^{(n)} \right) - \left(2\Phi(x) - 1 + 2n^{-1} q_2(x, h, P, \rho)\phi(x) \right) \right| \le C n^{-1-\epsilon} \ ,$$

for some constant C. By (A.5) in the proof of Theorem 5.1, the previous inequality, and the definition of $c_n^*(h, 1-\alpha)$ and $c_n(h, 1-\alpha)$ as quantiles, we conclude that

$$|c_n^*(h, 1-\alpha) - c_n(h, 1-\alpha)| \le C_1 n^{-1-\epsilon}$$

for some constant C_1. This completes the proof of our assumption in part 1.

B Auxiliary Results

B.1 Lemmas

Lemma B.1. *Suppose Assumptions 4.1 and 4.2 hold. Then, for any fixed $\eta > 0$, there exist constants $M > 0$, $\tilde{K}_4 > 0$, and $N_0 = N_0(\eta)$ such that*

1. $P_\rho \left(g(\rho, n)\ n^{1/2}\ |\hat{\rho}_n - \rho| > M \right) < \eta$,
2. $P_\rho \left(|n^{-1}\sum_{t=1}^n \tilde{u}_t^2 - \sigma^2| > \sigma^2/2 \right) < \eta$,
3. $P_\rho \left(n^{-1}\sum_{t=1}^n \tilde{u}_t^4 > \tilde{K}_4 \right) < \eta$,

for $n \ge N_0$ and $\rho \in [-1, 1]$, where $g(\rho, k) = \left(\sum_{\ell=0}^{k-1} \rho^{2\ell} \right)^{1/2}$, $\hat{\rho}_n$ is as in (12), and $\{\tilde{u}_t : 1 \le t \le n\}$ are centered residuals as in (13).

Proof. See Supplemental Appendix C.1. ■

Lemma B.2. *For any fixed $M > 0$. Suppose that for any $\rho \in [-1,1]$ we have*

$$|\hat{\rho}_n - \rho| \leq \frac{M}{n^{1/2} g(\rho, n)} \ ,$$

where $g(\rho,k) = \left(\sum_{\ell=0}^{k-1} \ \rho^{2\ell}\right)^{1/2}$. Then, there exist constants $\tilde{M} = \tilde{M}(M) > 0$ and $N_0 = N_0(M) > 0$ such that $|\hat{\rho}_n| \leq 1 + \tilde{M}/n$ for all $n \geq N_0$.

Proof. See Supplemental Appendix C.2. ∎

Lemma B.3. *Suppose Assumptions 4.1 and 4.2 hold. Fix $\epsilon > 0$. Then, for any $\alpha \in (0,1)$ and for any sequence $h_n \leq n$ such that $h_n = o(n)$, we have*

1. $\lim_{n\to\infty} \ \sup_{h \leq h_n} \ \sup_{\rho\in[-1,1]} P_\rho \left(z_{1-\alpha/2-3\epsilon/2} \leq c_n^*(h, 1-\alpha) \leq z_{1-\alpha/2+3\epsilon/2}\right) = 1$,
2. $\lim_{n\to\infty} \ \sup_{h \leq h_n} \ \sup_{\rho\in[-1,1]} P_\rho \left(z_{\alpha_0-\epsilon/2} \leq q_n^*(h, \alpha_0) \leq z_{\alpha_0+\epsilon/2}\right) = 1$,

where z_{α_0} is the α_0-quantiles of the standard normal distribution, $c_n^*(h, 1-\alpha)$ is as in (15), and $q_n^*(h,\alpha_0)$ is the α_0-quantile of $R_{b,n}^*(h)$ defined in (14).

Proof. See Supplemental Appendix C.3. ∎

Lemma B.4. *Suppose Assumption 5.1 holds. For any fixed $h \in \mathbf{N}$ and $a \in (0,1)$. Then, for any $\rho \in [-1+a, 1-a]$ and $\epsilon \in (0, 1/2)$, there exist constant $C = C(a,h,k,C_\sigma) > 0$ and a real-valued function*

$$\mathcal{T}(\cdot; \sigma^2, \psi_4^4, \rho) : \mathbf{R}^8 \to \mathbf{R} \ ,$$

such that

1. $\mathcal{T}(\mathbf{0}; \sigma^2, \psi_4^4, \rho) = 0$,
2. $\mathcal{T}(x; \sigma^2, \psi_4^4, \rho)$ *is a polynomial of degree 3 in $x \in \mathbf{R}^8$ with coefficients depending continuously differentiable on σ^2, ψ_4^4, and ρ,*
3. $\sup_{x\in\mathbf{R}} |J_T(x,h,P,\rho) - \tilde{J}_n(x,h,P,\rho)| \leq D_n + n^{-1-\epsilon} C \left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right)$,

where $\sigma^2 = E_P[u_1^2]$, $\psi_4^4 = E_P[u_1^4]$, $k \geq 8(1+\epsilon)/(1-2\epsilon)$,

$$\tilde{J}_n(x,h,P,\rho) \equiv P_\rho \left((n-h)^{1/2} \mathcal{T}\left(\frac{1}{n-h} \sum_{t=1}^{n-h} X_t; \sigma^2, \psi_4^4, \rho \right) \leq x \right) \ ,$$

and

$$D_n = \sup_{x\in\mathbf{R}} \left| \tilde{J}_n(x+n^{-1-\epsilon}, h, P, \rho) - \tilde{J}_n(x-n^{-1-\epsilon}, h, P, \rho) \right| \ .$$

The sequence $\{X_t : 1 \le t \le n-h\}$ *is defined in* (A.9)*. Furthermore, the asymptotic variance of* $(n-h)^{1/2}\mathcal{T}((n-h)^{-1}\sum_{t=1}^{n-h} X_t; \sigma^2, \psi_4^4, \rho)$ *is equal to one.*

Proof. See Supplemental Appendix C.4. ∎

Lemma B.5. *Suppose Assumptions 5.1 holds. For any fixed* $h \in \mathbf{N}$ *and* $a \in (0,1)$*. Then, for any* $|\rho| \le 1-a$ *and* $\epsilon \in (0, 1/2)$*, there exist* $C = C(a, k, h, C_\sigma, \epsilon, c_u)$*,* $\tilde{C}_\sigma$*, and* M *such that*

1. $P\left(|\hat{\rho}_n| > 1 - a/2\right) \le Cn^{-1-\epsilon}$
2. $P\left(|n^{-1}\sum_{t=1}^{n}\tilde{u}_t^r - E[u_t^r]| > n^{-\epsilon}\right) \le Cn^{-1-\epsilon}$
3. $P\left(n^{-1}\sum_{t=1}^{n}\tilde{u}_t^2 < \tilde{C}_\sigma\right) \le Cn^{-1-\epsilon}$
4. $P\left(n^{-1}\sum_{t=1}^{n}\tilde{u}_t^{4k} > M\right) \le Cn^{-1-\epsilon}$

for fixed $r \ge 1$*,* $k \ge 8(1+\epsilon)/(1-2\epsilon)$*, where* $\hat{\rho}_n$ *and the centered residuals* $\{\tilde{u}_t : 1 \le t \le n\}$ *are defined in* (12) *and* (13)*, respectively.*

Proof. See Supplemental Appendix C.5. ∎

B.2 Uniform Consistency

For any fixed $M > 0$, consider the sequence of models:

$$y_{n,t} = \rho_n y_{n,t-1} + u_{n,t} \ , \ y_{n,0} = 0 \ , \quad \text{and} \quad \rho_n \in [-1 - M/n, 1 + M/n] \ ,$$

where $\{u_{n,t} : 1 \le t \le n\}$ is a sequence of shocks with probability distribution denoted by P_n. We use P_n and E_n to compute respectively probabilities and expected values of the sequence $\{(y_{n,t}, u_{n,t}) : 1 \le t \le n\}$. This appendix presents results for a sequence of AR(1) models.

We extend the notation introduced in Section 2 for the sequence of models. For fixed any $h < n$, the coefficients in the linear regression of $y_{n,t+h}$ on $(y_{n,t}, y_{n,t-1})$ are defined by

$$\begin{pmatrix} \hat{\beta}_n(h) \\ \hat{\gamma}_n(h) \end{pmatrix} = \left(\sum_{t=1}^{n-h} x_{n,t}x_{n,t}'\right)^{-1} \left(\sum_{t=1}^{n-h} x_{n,t}y_{n,t+h}\right) \ , \tag{A.6}$$

where $x_{n,t} \equiv (y_{n,t}, y_{n,t-1})'$. And the HC standard error $\hat{s}_n(h)$ is defined by

$$\hat{s}_n(h) \equiv \left(\sum_{t=1}^{n-h} \hat{u}_{n,t}(h)^2 \right)^{-1/2} \left(\sum_{t=1}^{n-h} \hat{\xi}_{n,t}(h)^2 \hat{u}_{n,t}(h)^2 \right)^{1/2} \left(\sum_{t=1}^{n-h} \hat{u}_{n,t}(h)^2 \right)^{-1/2} ,$$

where $\hat{\xi}_{n,t}(h) = y_{n,t+h} - \hat{\beta}_n(h) y_{n,t} - \hat{\gamma}_n(h) y_{n,t-1}$, $\hat{u}_{n,t}(h) = y_{n,t} - \hat{\rho}_n(h) y_{n,t-1}$, and $\hat{\rho}_n(h)$ is defined as

$$\hat{\rho}_n(h) \equiv \left(\sum_{t=1}^{n-h} y_{n,t-1}^2 \right)^{-1} \left(\sum_{t=1}^{n-h} y_{n,t} y_{n,t-1} \right) . \tag{A.7}$$

For any fixed positive constants $K_4 > 0$ and $\overline{\sigma} \geq \underline{\sigma} > 0$, we consider the next assumption that imposes restrictions on the distribution of the shocks P_n.

Assumption B.1.

i) $\{u_{n,t} : 1 \leq t \leq n\}$ *is row-wise i.i.d. triangular array with mean zero and variance* σ_n^2.

ii) $E_n[u_{n,t}^4] < K_4$ *and* $\sigma_n^2 \in [\underline{\sigma}, \overline{\sigma}]$.

We denote by $\mathbf{P}_{n,0}$ the set of all distributions P_n that verify Assumption B.1. Theorem B.1 below extends the results presented in Xu (2023) and Montiel Olea and Plagborg-Møller (2021) for a large family of models under stronger assumptions over the serial dependence of the shocks. We adapt their proof and simplify some steps based on our stronger assumptions. For instance, we assume only bounded 4th moments, while they assume bounded at least 8th bounded moments. One remarkable difference is that we do not need to assume a high-level assumption such as Assumption 4.2 since this can be verified using Assumption B.1; we present the claim of this result in the next proposition.

Proposition B.1. *Suppose Assumption B.1 holds. Then, we have*

$$\lim_{K \to \infty} \lim_{n \to \infty} \inf_{P_n \in \mathbf{P}_{n,0}} \inf_{|\rho_n| \leq 1+M/n} P \left(g(\rho, n)^{-2} \ n^{-1} \sum_{t=1}^{n} y_{n,t-1}^2 \geq 1/K \right) = 1 ,$$

where $g(\rho, k) = \left(\sum_{\ell=0}^{k-1} \ \rho^{2\ell} \right)^{1/2}$.

Theorem B.1. *Suppose Assumption B.1 holds. Then, for any sequence* $h_n \leq n$ *such that* $h_n = o(n)$*, we have*

$$\sup_{h \leq h_n} \sup_{P_n \in \mathbf{P}_{n,0}} \sup_{|\rho| \leq 1+M/n} \sup_{x \in \mathbf{R}} |J_n(x, h, P_n, \rho) - \Phi(x)| \to 0 , \quad as \quad n \to \infty,$$

where $J_n(\cdot, h, P_n, \rho)$ is as in (7) and $\Phi(x)$ is the cdf of the standard normal distribution.

Proof. See Supplemental Appendix C.6. ■

B.3 Asymptotic Refinements

Consider the sequence $\{z_t : 1 \le t \le n\}$ defined as

$$z_t = \rho z_{t-1} + u_t \ , \quad \text{and} \quad z_0 = \sum_{\ell=0}^{\infty} \rho^\ell u_{-\ell} \ ,$$

where $\{u_{-\ell} : \ell \ge 0\}$ is an i.i.d. sequence with the same distribution as u_1. This appendix presents asymptotic expansion results for distributions of real value functions based on sample averages of the sequence $\{X_t = F(z_{t-1}, z_t, z_{t+h}) : 1 \le t \le n-h\}$, where F is a function that we define below. Our approach in this section relies on the framework and results presented in Götze and Hipp (1994) and Bhattacharya and Ghosh (1978).

Let $F(\cdot\ ; \sigma^2, V, \rho) : \mathbf{R}^3 \to \mathbf{R}^8$ be a function defined at (x, y, z) equal to

$$\Big((z-\rho^h y)(y-\rho x),\ (y-\rho x)^2 - \sigma^2,\ ((z-\rho^h y)(y-\rho x))^2 - V,\ (z-\rho^h y)(y-\rho x)^3, \\ (y-\rho x)x,\ (z-\rho^h y)x,\ (z-\rho^h y)^2(y-\rho x)x,\ (z-\rho^h y)(y-\rho x)^2 x\Big) \ , \tag{A.8}$$

where $\sigma^2 = \sigma^2(P) = E_P[u_1^2]$, $V = V(\rho, h, P) = E_P[\xi_1^2 u_1^2]$, $\xi_1 = \xi_1(\rho, h) \equiv \sum_{\ell=1}^{h} \rho^{h-\ell} u_{1+\ell}$, and P is the distribution of the shocks that verified Assumption B.2 that we define below. Using that $u_t = z_t - \rho z_{t-1}$, $\xi_t = z_{t+h} - \rho^h z_t$, and the definition of F in (A.8), we can write the sequence of random vectors $\{X_t = F(z_{t-1}, z_t, z_{t+h}; \sigma^2, V, \rho)) : 1 \le t \le n-h\}$ as follows

$$X_t = (\xi_t u_t, u_t^2 - \sigma^2, (\xi_t u_t)^2 - V, \xi_t u_t^3, u_t z_{t-1}, \xi_t z_{t-1}, \xi_t^2 u_t z_{t-1}, \xi_t u_t^2 z_{t-1}) \ . \tag{A.9}$$

We assume in this section that $h \in \mathbf{N}$ is fixed and $|\rho| < 1$. Moreover, for any fixed positive constants $C_{18} > 0$ and $C_\sigma > 0$, we consider the next assumption that imposes restrictions on the distribution of the shocks P.

Assumption B.2.

i) $\{u_t : 1 \le t \le n\}$ is independent and identically distributed with $E[u_t] = 0$.

ii) u_t has a positive continuous density.

iii) $E[u_t^{18}] \leq C_{18} < \infty$ and $E[u_t^2] \geq C_\sigma$.

Assumption B.2 implies that the sequence $\{z_t : 1 \leq t \leq n\}$ is strictly stationary. By construction, $E[X_t] = \mathbf{0} \in \mathbf{R}^8$. Define

$$\Sigma = \lim_{n\to\infty} Cov\left((n-h)^{-1/2}\sum_{t=1}^{n-h} X_t\right) . \tag{A.10}$$

The asymptotic covariate matrix Σ is non-singular due to Lemma 2.1 in Götze and Hipp (1994), Assumption B.2, and how we defined the sequence $\{X_t : 1 \leq t \leq n-h\}$. Let $\mathcal{T} : \mathbf{R}^8 \to \mathbf{R}$ be a polynomial with coefficients depending on ρ, $E_P[u_1^2]$, and $E_P[u_1^4]$ such that $\mathcal{T}(\mathbf{0}) = 0$. Define

$$\tilde{J}_n(x,h,P,\rho) \equiv P_\rho\left(\frac{(n-h)^{1/2}}{\tilde{\sigma}}\mathcal{T}\left(\frac{1}{n-h}\sum_{t=1}^{n-h} X_t\right) \leq x\right) , \tag{A.11}$$

where $\tilde{\sigma}^2$ is the asymptotic variance of $(n-h)^{1/2}\mathcal{T}((n-h)^{-1}\sum_{t=1}^{n-h} X_t)$. The next theorem shows that the distribution $\tilde{J}_n(\cdot,h,P,\rho)$ admits a valid Edgeworth expansion.

Theorem B.2. *Suppose Assumption B.2 holds. Fix a given $h \in \mathbf{N}$ and $a \in (0,1)$. Then, for any $\rho \in [-1+a, 1-a]$, we have*

$$\sup_{x\in\mathbf{R}}\left|\tilde{J}_n(x,h,P,\rho) - \left(\Phi(x) + \sum_{j=1}^{2} n^{-j/2} q_j(x,h,P,\rho)\phi(x)\right)\right| = O\left(n^{-3/2}\right) ,$$

where $\tilde{J}_n(x,h,P,\rho)$ is as in (A.11), $\Phi(x)$ and $\phi(x)$ are the cdf and pdf of the standard normal distribution, and $q_1(x,h,P,\rho)$ and $q_2(x,h,P,\rho)$ are polynomials on x with coefficients that are continuous function of moments of P (up to order 12) and ρ. Furthermore, we have $q_1(x,h,P,\rho) = q_1(-x,h,P,\rho)$ and $q_2(x,h,P,\rho) = -q_2(-x,h,P,\rho)$.

The proof of Theorem B.2 is presented in Supplemental Appendix C.7. It relies on Götze and Hipp (1983, 1994) to guarantee the existence of Edgeworth expansion for sample averages and in the results of Bhattacharya and Ghosh (1978) to complete the proof.

For the empirical distribution $\hat{P}_n$ defined in (13) and the estimator $\hat{\rho}_n$ defined in (12), we consider the bootstrap sequence $\{z^*_{b,t} : 1 \leq t \leq n\}$ defined as

$$z^*_{b,t} = \hat{\rho}_n z^*_{b,t-1} + u^*_{b,t} \ , \quad \text{and} \quad z^*_{b,0} = \sum_{\ell=0}^{\infty} \hat{\rho}_n^{\ell} u^*_{b,-\ell} \ ,$$

where $\{u^*_{b,j} : j \leq n\}$ is an i.i.d. sequence draw from the distribution $\hat{P}_n$. Define the sequence of random vectors $\{X^*_{b,t} = F(z^*_{b,t-1}, z^*_{b,t}, z^*_{b,t+h}; \hat{\sigma}^2_n, \hat{V}_n, \hat{\rho}_n) : 1 \leq t \leq n-h\}$, where $F(\cdot)$ is as in (A.8) and $\hat{\sigma}^2_n, \hat{V}_n, \hat{\rho}_n$ are the defined using $\hat{P}_n$ and $\hat{\rho}_n$.

Theorem B.3. *Suppose Assumption 5.1 holds. Fix a given $h \in \mathbf{N}$ and $a \in (0,1)$. Then, for any $\rho \in [-1+a, 1-a]$ and $\epsilon \in (0, 1/2)$, there exist constants C_1 and C_2 such that*

$$P\left(\sup_{x \in \mathbf{R}} \left| \tilde{J}_n(x, h, \hat{P}_n, \hat{\rho}_n) - \left(\Phi(x) + \sum_{j=1}^{2} n^{-j/2} q_j(x, h, \hat{P}_n, \hat{\rho}_n) \phi(x) \right) \right| > C_1 n^{-3/2} \right) \ ,$$

is lower than $C_2 n^{-1-\epsilon}$, where $\tilde{J}_n(x, h, \cdot, \cdot)$ is as in (A.11) and $X^*_{b,t}$ is replacing X_i, $\Phi(x)$ and $\phi(x)$ are the cdf and pdf of the standard normal distribution, and $q_1(x, h, \hat{P}_n, \hat{\rho}_n)$ and $q_2(x, h, \hat{P}_n, \hat{\rho}_n)$ are polynomials on x with coefficients that are continuous function of moments of $\hat{P}_n$ (up to order 12) and $\hat{\rho}_n$. Furthermore, we have $q_1(x, h, \hat{P}_n, \hat{\rho}_n) = q_1(-x, h, \hat{P}_n, \hat{\rho}_n)$ and $q_2(x, h, \hat{P}_n, \hat{\rho}_n) = -q_2(-x, h, \hat{P}_n, \hat{\rho}_n)$.

The proof of Theorem B.3 is presented in Supplemental Appendix C.8. It relies on Götze and Hipp (1983, 1994), Bhattacharya and Ghosh (1978), and Lemma B.5.

B.4 Proof of Proposition B.1

Proof of Proposition B.1. We use the general subsequence approach of Andrews et al. (2020) to show that the uniform result in the proposition holds. We prove that for any sequence $\{\rho_n : n \geq 1\}$ such that $|\rho_n| \leq 1 + M/n$ and any sequence $\{\sigma^2_n : n \geq 1\} \subset [\underline{\sigma}, \overline{\sigma}]$, there exists subsequences $\{\rho_{n_k} : k \geq 1\}$ and $\{\sigma^2_{n_k} : k \geq 1\}$ such that

$$\lim_{K \to \infty} \ \lim_{k \to \infty} \ P\left(\ g(\rho_{n_k}, n_k)^{-2} n_k^{-1} \sum_{i=1}^{n_k} y^2_{n_k, i-1} \geq 1/K \ \right) = 1 \ . \tag{A.12}$$

We consider two cases to prove (A.12). The first case is $n_k(1 - |\rho_{n_k}|) \to \infty$ for some subsequence $\{n_k : k \geq 1\}$, which considers the subsequence of ρ_n that stay on the stationary region or go to the boundary at slower rates. The second case is $n_k(1-|\rho_{n_k}|) \to c \in [-M, +\infty)$ for some subsequence $\{n_k : k \geq 1\}$, which considers the subsequence of ρ_n that goes to the boundary (local-unit-model) or are on it (unit-root model). For both cases, we assume $\sigma^2_{n_k} \to \sigma^2_0$ since any sequence $\{\sigma^2_n : n \geq 1\} \subset [\underline{\sigma},\, \overline{\sigma}]$ has always a convergent subsequence. To avoid complicated sub-index notation, we present the algebra derivation using the original sequence.

Case 1: Suppose $n(1-|\rho_n|) \to \infty$ and $\sigma^2_n \to \sigma^2_0$ as $n \to \infty$. This condition implies that there exists N_0 such that $|\rho_n| \leq 1$ for all $n \geq N_0$, otherwise there is a subsequence $|\rho_{n_k}|$ in $(1, 1+M/n_k]$ but this cannot occur since $n_k(1-|\rho_{n_k}|) \in [-M, 0]$. As a result, we have $g(\rho_n, n)^2 = \sum_{\ell=0}^{n-1} \ \rho_n^{2\ell} \leq (1-\rho_n^2)^{-1}$ that implies

$$P\left(\ g(\rho_n, n)^{-2} \ n^{-1} \sum_{t=1}^{n} y^2_{n,t-1} \geq 1/K \ \right) \geq P\left(\ \frac{(1-\rho_n^2)}{n} \sum_{t=1}^{n} y^2_{n,t-1} \ \geq 1/K \ \right) \ .$$

Therefore, to verify (A.12) is sufficient to prove that

$$\lim_{K\to\infty} \ \lim_{n\to\infty} \ P\left(\ \frac{(1-\rho_n^2)}{n} \sum_{t=1}^{n} y^2_{n,t-1} \ \geq 1/K \ \right) = 1 \ ,$$

which follows if we prove

$$\frac{(1-\rho_n^2)}{n} \sum_{t=1}^{n} y^2_{n,t-1} \xrightarrow{p} \sigma^2_0 \ . \tag{A.13}$$

We prove (A.13) in two steps.

Step 1: Using Assumption B.1 and $y_{n,t-1} = \sum_{\ell=1}^{i-1} \rho_n^{i-1-\ell} u_{n,\ell}$, we derive the following:

$$\begin{aligned} E\left[\frac{(1-\rho_n^2)}{n} \sum_{t=1}^{n} y^2_{n,t-1} \right] &= \frac{(1-\rho_n^2)}{n} \sum_{t=1}^{n} \sum_{\ell=1}^{n-1} E[u^2_{n,\ell}] \ \rho_n^{2(t-1-\ell)} \ I\{1 \leq \ell \leq i-1\} \\ &= \frac{\sigma^2_n (n-1)}{n} - \frac{\sigma^2_n}{n} \sum_{\ell=1}^{n-1} \rho_n^{2(n-\ell)} \end{aligned}$$

We conclude the right-hand side of the previous display converges to σ^2_0 since $\sigma^2_n \to \sigma^2_0$, $n^{-1} \sum_{\ell=1}^{n} \rho_n^{2(n-\ell)} \leq \{n(1-|\rho_n|)(1+|\rho_n|)\}^{-1}$, and $n(1-|\rho_n|) \to \infty$.

Step 2: We use $E[y_{n,t-1}^2] = g(\rho_n, t-1)^2\sigma_n^2$ to derive the following decomposition

$$\sum_{i=1}^{n} y_{n,t-1}^2 - E[y_{n,t-1}^2] = \sum_{\ell=1}^{n-1} \left(u_{n,\ell}^2 - \sigma_n^2\right) b_{n,\ell} + 2\sum_{\ell=1}^{n-1} u_{n,\ell}\, d_{n,\ell} \ ,$$

where $b_{n,\ell} = \sum_{i=1+\ell}^{n} \rho_n^{2(i-1-\ell)}$ and $d_{n,\ell} = b_{n,\ell} \sum_{\ell_2=1}^{\ell-1} u_{n,\ell_2}\rho_n^{\ell-\ell_2}$. Note $d_{n,\ell}$ is measurable with respect to the σ-algebra defined by $\{u_{n,k} : 1 \le k \le \ell - 1\}$.

The decomposition above, Loeve's inequality (see Theorem 9.28 in Davidson (1994)), and Assumption B.1 imply that the variance of $(1-\rho_n^2)n^{-1}\sum_{t=1}^{n} y_{n,t-1}^2$ is lower than

$$\frac{2(1-\rho_n^2)^2}{n^2}\left(\sum_{\ell=1}^{n-1} E\left[\left(u_{n,\ell}^2 - \sigma_n^2\right)^2\right] b_{n,\ell}^2 + 4\sum_{\ell=1}^{n-1} E\left[u_{n,\ell}^2\right] E\left[d_{n,\ell}^2\right]\right) \ .$$

Since $b_{n,\ell}^2 \le (1-\rho_n^2)^{-2}$ and $E[d_{n,\ell}^2] = b_{n,\ell}^2\sigma_n^2 \sum_{\ell_2=1}^{\ell-1} \rho_n^{2(\ell-\ell_2)} \le \sigma_n^2(1-\rho_n^2)^{-3}$ for all $\ell = 1, \dots, n-1$, and $E[(u_{n,\ell}^2 - \sigma_n^2)^2] \le E[u_{n,\ell}^4] \le K_4$ by Assumption B.1, the previous display is lower than

$$\frac{2(n-1)K_4}{n^2} + \frac{8(n-1)\sigma_n^4}{n^2(1-\rho_n^2)} \ ,$$

which goes to 0 since $\sigma_n^4 = E[u_{n,\ell}^2]^2 \le E[u_{n,\ell}^4] \le K_4$ and $n(1-\rho_n^2) = n(1-|\rho_n|)(1+|\rho_n|) \to \infty$ as $n \to \infty$. This proves that the variance of the left-hand side on (A.13) goes to zero, which prove (A.13) due to step 1.

Case 2: Suppose $n(1-|\rho_n|) \to c \in [-M, +\infty)$ and $\sigma_n^2 \to \sigma_0^2$ as $n \to \infty$. We first observe that $g(\rho_n, n)^2 \le n\exp(2M)$ due to $|\rho_n| \le 1 + M/n$ and the following derivation:

$$g(\rho_n, n)^2 = \sum_{\ell=0}^{n-1} \rho_n^{2\ell} \le n(1+M/n)^{2n} = n\exp\left(2n\log\left(1+M/n\right)\right) \le n\exp\left(2M\right) \ ,$$

where we used that $\log(1+x) \le x$ for all $x > -1$. By the previous observation

$$P\left(\ g(\rho_n, n)^{-2}n^{-1}\sum_{t=1}^{n} y_{n,t-1}^2 \ge 1/K\ \right) \ge P\left(\ \frac{\exp(-2M)}{n^2}\sum_{t=1}^{n} y_{n,t-1}^2\ \ \ge 1/K\ \right) \ ,$$

where $\exp(-2M)$ is constant that does not change as $n \to \infty$ and $K \to \infty$. Therefore, it is sufficient to prove that $\lim_{K\to\infty}\ \lim_{n\to\infty}\ P\left(\ \frac{1}{n^2}\sum_{t=1}^{n} y_{n,t-1}^2\ \ \ge 1/K\ \right) = 1$ to verify (A.12), which follows if we prove

$$\frac{1}{n^2}\sum_{t=1}^{n} y_{n,t-1}^2 \xrightarrow{d} \sigma_0^2 \int_0^1 J_{-c}(r)^2 dr \ , \tag{A.14}$$

where $J_{-c}(r) = \int_0^r e^{-(r-s)c} dW(s)$ and $W(s)$ is a standard Brownian motion.

To prove (A.14), we rely on the results and techniques presented in Phillips (1987). Specifically, we adapt his Lemma 1 part (c) for the sequence of models and the drifting parameter that we consider in this paper. We proceed in two steps. First, we construct a triangular array $\{\tilde{y}_{n,t} : 1 \leq t \leq n; n \geq 1\}$ that verify (A.14). Then, we prove that the constructed sequence approximates the original triangular array $\{y_{n,t} : 1 \leq t \leq n\}$.

Step 1: Define $\tilde{u}_{n,t} = u_{n,t} I\{\rho_n \geq 0\} + (-1)^t u_{n,t} I\{\rho_n < 0\}$ for all $t = 1, \ldots, n$. Note that the sequence $\{\tilde{u}_{n,t} : 1 \leq t \leq n\}$ defines a martingale difference array with the same variance $E[\tilde{u}_{n,t}^2] = \sigma_n^2$ and satisfies that $E[\tilde{u}_{n,t}^2] \in [\underline{\sigma}, \overline{\sigma}]$, and $E[\tilde{u}_{n,t}^4] \leq K_4$. Using this notation, we construct the following triangular array:

$$\tilde{y}_{n,t} = e^{-c/n} \tilde{y}_{n,t-1} + \tilde{u}_{n,t} \ , \quad \tilde{y}_{n,0} = 0 \ ,$$

where $c = \lim_{n\to\infty} n(1 - |\rho_n|)$. Denote the sequence of partial sums by $S_{n,j} = \sum_{t=1}^{j} u_{n,t}$ for any $j = 1, \ldots, n$ and $S_{n,0} = 0$. Let us define the following random process

$$X_n(r) = \frac{1}{\sqrt{n}}\frac{1}{\sigma_n} S_{n,[nr]} = \frac{1}{\sqrt{n}}\frac{1}{\sigma_0} S_{n,j-1} \ \text{ if } (j-1)/n \leq r < j/n \ ,$$

and $X_n(1) = \frac{1}{\sqrt{n}}\frac{1}{\sigma_n} S_{n,n}$. By a functional central limit theorem for martingale difference arrays (see Theorem 27.14 in Davidson (1994)), we claim that $\{X_n(r) : r \in [0,1]\}$ converges to the standard Brownian motion process $\{W(r) : r \in [0,1]\}$. To use this result, we prove

$$\text{(a)} \quad \sum_{t=1}^{n} \frac{\tilde{u}_{n,t}^2}{n\sigma_n^2} \xrightarrow{p} 1 \ , \quad \text{(b)} \quad \max_{t \leq n} \left| \frac{\tilde{u}_{n,t}}{\sqrt{n}\sigma_n} \right| \xrightarrow{p} 0 \ , \quad \text{(c)} \quad \lim_n \sum_{t=1}^{[nr]} \frac{E[\tilde{u}_{n,t}^2]}{n\sigma_n^2} = r \ .$$

We can verify condition (a) using $\tilde{u}_{n,t}^2 = u_{n,t}^2$, Chebyshev's inequality and Assumption B.1:

$$P\left(\left| \sum_{t=1}^{n} \frac{\tilde{u}_{n,t}^2}{n\sigma_n^2} - 1 \right| > \epsilon \right) \leq \frac{n^{-2}}{\epsilon^2} \sum_{i=1}^{n} E[u_{n,t}^4] \leq \frac{K_4}{\epsilon^2 n} \to 0 \quad \text{as } n \to \infty \ ,$$

for any $\epsilon > 0$. To verify condition (b) holds is sufficient to show that

$$nE\left[\frac{\tilde{u}_{n,t}^2}{n\sigma_n^2} I\left\{\left|\frac{\tilde{u}_{n,t}}{\sqrt{n}\sigma_n}\right| > c\right\}\right] \to 0 \ ,$$

for any $c > 0$, where $I\{\cdot\}$ is the indicator function. If the previous display holds, then condition (b) follows by theorem 23.16 in Davidson (1994). To verify the previous condition, note that

$$nE\left[\frac{\tilde{u}_{n,t}^2}{n\sigma_n^2} I\left\{\left|\frac{\tilde{u}_{n,t}}{\sqrt{n}\sigma_n}\right| > c\right\}\right] \leq nE\left[\frac{\tilde{u}_{n,t}^4}{n^2\sigma_n^4 c^2} I\left\{\left|\frac{\tilde{u}_{n,t}}{\sqrt{n}\sigma_n}\right| > c\right\}\right] \leq \frac{E[\tilde{u}_{n,t}^4]}{n\sigma_n^4 c^2} \ \leq \ \frac{K_4}{n\underline{\sigma}^2 c^2} \ ,$$

were the last inequality uses $\tilde{u}_{n,t}^4 = u_{n,t}^4$ and Assumption B.1. Finally, condition (c) holds since $E[\tilde{u}_{n,t}^2] = \sigma_n^2$.

Using the functional central limit theorem, the continuous mapping theorem, and $\sigma_n \to \sigma_0$, we can repeat the arguments presented in the proof of Lemma 1 in Phillips (1987) to conclude that $n^{-2}\sum_{t=1}^n \tilde{y}_{n,t-1}^2 \xrightarrow{d} \sigma_0^2 \int_0^1 J_{-c}(r)^2 dr$.

Step 2: Define $a_n = |\rho_n| e^{c/n}$. We know $y_{n,t} = \sum_{\ell=1}^i \rho_n^{i-\ell} u_{n,t}$ and $\tilde{y}_{n,t} = \sum_{\ell=1}^i e^{-c(i-\ell)/n} \tilde{u}_{n,t}$; therefore, $\tilde{y}_{n,t} = y_{n,t} - R_{n,t}$ if $\rho_n \geq 0$, and $\tilde{y}_{n,t} = (-1)^t y_{n,t} - R_{n,t}$ if $\rho_n < 0$, where $R_{n,t} = \sum_{\ell=1}^t \left(a_n^{t-\ell} - 1\right) e^{-c(t-\ell)/n} \tilde{u}_{n,\ell}$. Therefore, we conclude that $y_{n,t}^2 = \tilde{y}_{n,t}^2 + 2\tilde{y}_{n,t}R_{n,t} + R_{n,t}^2$. This implies that

$$\left|\frac{1}{n^2}\sum_{t=1}^n y_{n,t-1}^2 - \frac{1}{n^2}\sum_{t=1}^n \tilde{y}_{n,t-1}^2\right| \leq \left|\frac{2}{n^2}\sum_{t=1}^n \tilde{y}_{n,t-1} R_{n,t}\right| + \left|\frac{1}{n^2}\sum_{t=1}^n R_{n,t}^2\right| \ .$$

By Cauchy–Schwartz's inequality, the right-hand side of the previous expression is lower or equal than

$$2\left(\frac{1}{n^2}\sum_{i=1}^n \tilde{y}_{n,t-1}^2\right)^{1/2} \left(\frac{1}{n^2}\sum_{i=1}^n R_{n,t}^2\right)^{1/2} + \left|\frac{1}{n^2}\sum_{i=1}^n R_{n,t}^2\right| \ . \tag{A.15}$$

By the result at the end of Step 1, we have $n^{-2}\sum_{i=1}^n \tilde{y}_{n,t-1}^2$ is $O_p(1)$. Therefore, it is sufficient to show $n^{-2}\sum_{i=1}^n R_{n,t}^2 \xrightarrow{p} 0$ to conclude that (A.15) converges to zero in probability.

To verify the claim, we first observe that $a_n^j e^{-(i-\ell)c/n} = |\rho_n|^j e^{-(i-\ell-j)c/n} \leq |\rho_n|^j$ for all

$j = 0, \ldots, i - \ell - 1$, which implies that

$$|R_{n,t}| = \left| \sum_{\ell=1}^{t} (a_n - 1)(1 + a_n + ... + a_n^{i-\ell-1}) e^{-(t-\ell)c/n} \tilde{u}_{n,\ell} \right|$$
$$\leq |a_n - 1| \sum_{\ell=1}^{t} (1 + |\rho_n| + ... + |\rho_n|^{t-\ell-1}) |\tilde{u}_{n,\ell}| \ .$$

Using the previous inequality and $|\rho_n|^j \leq (1 + M/n)^j \leq (1 + M/n)^n \leq e^M$, we obtain

$$|R_{n,t}| \leq e^M |a_n - 1| \sum_{\ell=1}^{t} (t - \ell) |\tilde{u}_{n,\ell}| \leq e^M |n(a_n - 1)| \sum_{\ell=1}^{n} |u_{n,\ell}| \ ,$$

for all $t = 1, \ldots, n$, where we used that $|\tilde{u}_{n,\ell}| = |u_{n,\ell}|$ in the last inequality. Then, we derive

$$n^{-2} \sum_{t=1}^{n} R_{n,t}^2 \leq e^{2M} |n(a_n - 1)|^2 \left(n^{-1} \sum_{\ell=1}^{n} |u_{n,\ell}| \right)^2 \ .$$

By Markovs's inequality and Assumption B.1, we obtain that $n^{-1} \sum_{\ell=1}^{n} |u_{n,\ell}|$ is $O_p(1)$. Analyzing $a_n - 1 = e^{c/n} \left(|\rho_n| - e^{-c/n} \right)$, we can conclude that $n(a_n - 1) = o(1)$, which implies that the right-hand side of the previous display converges to zero in probability. As a result, we conclude that (A.15) converges to zero in probability, which implies that $n^{-2} \sum_{i=1}^{n} y_{n,t-1}^2 - n^{-2} \sum_{i=1}^{n} \tilde{y}_{n,t-1}^2 = o_p(1)$, and by the result at the end of Step 1, we conclude (A.14). ■

References

Andrews, D. W. (1993): "Exactly median-unbiased estimation of first order autoregressive/unit root models," *Econometrica: Journal of the Econometric Society*, 139–165.

——— (2002): "Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators," *Econometrica*, 70, 119–162.

——— (2004): "The block–block bootstrap: improved asymptotic refinements," *Econometrica*, 72, 673–700.

Andrews, D. W., X. Cheng, and P. Guggenberger (2020): "Generic results for establishing the asymptotic size of confidence sets and tests," *Journal of Econometrics*, 218, 496–531.

Bhattacharya, R. (1987): "Some aspects of Edgeworth expansions in statistics and probability," *New perspectives in theoretical and applied statistics*, 157, 171.

Bhattacharya, R. N. and J. K. Ghosh (1978): "On the validity of the formal Edgeworth expansion," *Ann. Statist*, 6, 434–451.

Bose, A. (1988): "Edgeworth correction by bootstrap in autoregressions," *The Annals of statistics*, 1709–1722.

Carlstein, E. (1986): "The use of subseries values for estimating the variance of a general statistic from a stationary sequence," *The annals of statistics*, 1171–1179.

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

Dharmadhikari, S. W., V. Fabian, and K. Jogdeo (1968): "Bounds on the Moments of Martingales," *The Annals of Mathematical Statistics*, 39, 1719 – 1723.

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," *The Econometrics Journal*, 7, 505–527.

Götze, F. and C. Hipp (1983): " Asymptotic expansions for sums of weakly dependent random vectors," *Z. Wahrscheinlichkeitstheorie verw Gebiete*, 64, 211–239.

——— (1994): "Asymptotic distribution of statistics in time series," *The Annals of Statistics*, 2062–2088.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*, Berlin: Springer, 1 ed.

Hall, P. and J. L. Horowitz (1996): "Bootstrap critical values for tests based on generalized-method-of-moments estimators," *Econometrica: Journal of the Econometric Society*, 891–916.

Hansen, B. E. (1999): "The grid bootstrap and the autoregressive model," *Review of Economics and Statistics*, 81, 594–607.

Herbst, E. and B. K. Johannsen (2021): "Bias in local projections," *FEDS Working Paper No. 2020-010R1.*

Horowitz, J. L. (2001): "The Bootstrap," *Handbook of Econometrics*, 5, 3159–3228.

——— (2019): "Bootstrap Methods in Econometrics," *Annual Review of Economics*, 11, 193–224.

Inoue, A. and L. Kilian (2002): "Bootstrapping autoregressive processes with possible unit roots," *Econometrica*, 70, 377–391.

Inoue, A. and M. Shintani (2006): "Bootstrapping GMM estimators for time series," *Journal of Econometrics*, 133, 531–555.

Jorda, B. (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?" *The Review of Economics and Statistics*, 93, 1460–1466.

Kunsch, H. R. (1989): "The jackknife and the bootstrap for general stationary observations," *The annals of Statistics*, 1217–1241.

Lahiri, S. (2003): *Resampling methods for dependent data*, Springer Science & Business Media.

Lahiri, S. N. (1996): "On Edgeworth Expansion and Moving Block Bootstrap for StudentizedM-Estimators in Multiple Linear Regression Models," *Journal of Multivariate Analysis*, 56, 42–59.

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.

Lusompa, A. (2023): "Local projections, autocorrelation, and efficiency," *Quantitative Economics, Forthcoming.*

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.

——— (2015): "Second Order Expansion of the t-statistic in AR (1) Models," *Econometric Theory*, 31, 426–448.

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

——— (2022): "Corrigendum: Local Projection Inference is Simpler and More Robust Than You Think," *available at https://scholar.princeton.edu/sites/default/files/lp_inference_corrigendum.pdf*.

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

Park, J. Y. (2003): "Bootstrap unit root tests," *Econometrica*, 71, 1845–1895.

——— (2006): "A bootstrap theory for weakly integrated processes," *Journal of Econometrics*, 133, 639–672.

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.

Phillips, P. C. (1987): "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74, 535–547.

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

White, H. (2000): *Asymptotic theory for econometricians*, Academic press.

Xu, K.-L. (2023): "Local Projection Based Inference under General Conditions," Tech. rep., No. 2023-001, Center for Applied Economics and Policy Research, Department of Economics, Indiana University Bloomington.

Appendix C: Proof of Auxiliary Results

C.1 Proof of the Lemma B.1

Proof. **Item 1:** Consider the following derivation:

$$g(\rho, n)\ n^{1/2}\ \ (\hat{\rho}_n - \rho) = \left(\frac{g(\rho, n)^{-2} \sum_{t=1}^{n} y_{t-1}^2}{n} \right)^{-1} \left(\frac{\sum_{t=1}^{n} u_t y_{t-1}}{g(\rho, n) n^{1/2}} \right) ,$$

where the first term is uniformly $O_p(1)$ due to Assumption 4.2. The second term is also uniformly $O_p(1)$ due to the following derivation:

$$E\left[\left(\frac{\sum_{t=1}^{n} u_t y_{t-1}}{g(\rho, n) n^{1/2}} \right)^2 \right] = \frac{1}{g(\rho, n)^2 n} \sum_{t=1}^{n} E[u_t^2 y_{t-1}^2] \leq E[u_t^4]^{1/2} \max_{1 \leq i \leq n} \left(\frac{E[y_{t-1}^4]}{g(\rho, n)^4} \right)^{1/2} \leq C_8^{1/(4\zeta)} C_{y4}^{1/2} ,$$

where the first inequality follows by Cauchy's and algebra manipulation and the second inequality follows by Assumption 4.1 and part(i) of Lemma MOMT-Y in Xu (2023). The constant C_{y4} depends on the distribution of the sequence $\{u_i : i \geq 1\}$ but does not depend on ρ. Therefore, we conclude $g(\rho, n)\ n^{1/2}\ \ (\hat{\rho}_n - \rho)$ is uniformly $O_p(1)$ for any $\rho \in [-1, 1]$, which conclude the proof of the lemma.

Item 2: Recall that $\tilde{u}_t = \hat{u}_t - n^{-1} \sum_{t=1}^{n} \hat{u}_t$, where $\hat{u}_t = y_t - \hat{\rho}_n y_{t-1}$ and $\hat{\rho}_n$ is defined in (12). By Bonferroni's inequality, it is sufficient to prove that there exists $N_0 = N_0(\eta)$ such that

$$P\left(\left| n^{-1} \sum_{t=1}^{n} \hat{u}_t^2 - \sigma^2 \right| > \sigma^2/4 \right) < \eta/2 \tag{C.1}$$

and

$$P\left(\left(n^{-1} \sum_{t=1}^{n} \hat{u}_t \right)^2 > \sigma^2/4 \right) < \eta/2 \tag{C.2}$$

for any $n \geq N_0$ and any $\rho \in [-1, 1]$. Lemma SIG in Xu (2023) adapted for the case of the AR(1) model implies (C.1). To prove (C.2), we derive the following inequality

$$\begin{aligned} \left(n^{-1} \sum_{t=1}^{n} \hat{u}_t \right)^2 &= \left(n^{-1} \sum_{t=1}^{n} u_t + (\rho - \hat{\rho}_n) n^{-1} \sum_{t=1}^{n} y_{t-1} \right)^2 \\ &\leq 2 \left(n^{-1} \sum_{t=1}^{n} u_t \right)^2 + 2 g(\rho, n)^2 (\hat{\rho}_n - \rho)^2 \left(n^{-1} \sum_{t=1}^{n} \frac{y_{t-1}^2}{g(\rho, n)^2} \right) , \end{aligned}$$

where we used Loeve's inequality (see Theorem 9.28 in Davidson (1994)) in the inequality above. Note that the first term is uniformly $o_p(1)$ due to the law of large numbers for α-mixing sequences (see Corollary 3.48 in White (2000)) and Assumption 4.1. Since $g(\rho,n)^2(\hat{\rho}_n-\rho)^2$ is uniformly $o_p(1)$ due to Part 1, it is sufficient to prove that $n^{-1}\sum_{t=1}^{n}\frac{y_{t-1}^2}{g(\rho,n)^2}$ is uniformly $O_p(1)$. The last claim follows by the next inequality

$$E\left[n^{-1}\sum_{t=1}^{n}\frac{y_{t-1}^2}{g(\rho,n)^2}\right]\leq n^{-1}\sum_{t=1}^{n}\left(E\left[\frac{y_{t-1}^4}{g(\rho,n)^4}\right]\right)^{1/2}\leq C_{y4}^{1/2} ,$$

where the last inequality follows by part(i) of Lemma MOMT-Y in Xu (2023). The constant C_{y4} depends on the distribution of the sequence $\{u_i : i\geq 1\}$ but does not depend on ρ. Therefore, $n^{-1}\sum_{t=1}^{n}\frac{y_{t-1}^2}{g(\rho,n)^2}$ is uniformly $O_p(1)$, which concludes the proof of the lemma.

Item 3: Recall that $\tilde{u}_t=\hat{u}_t-n^{-1}\sum_{t=1}^{n}\hat{u}_t$, where $\hat{u}_t=y_t-\hat{\rho}_n y_{t-1}$ and $\hat{\rho}_n$ is defined in (12). By Loeve's inequality (see Theorem 9.28 in Davidson (1994)), we obtain

$$(\hat{u}_t-n^{-1}\sum_{t=1}^{n}\hat{u}_t)^4=((\hat{u}_t-u_t)+n^{-1}\sum_{t=1}^{n}\hat{u}_t+u_i)^4\leq 3^3\left((\hat{u}_t-u_t)^4+\left(n^{-1}\sum_{t=1}^{n}\hat{u}_t\right)^4+u_t^4\right) .$$

Therefore, it is sufficient to prove that there exists $N_0=N_0(\eta)$ and $\tilde{K}_4$ such that

$$P\left(n^{-1}\sum_{t=1}^{n}(\hat{u}_t-u_t)^4>\tilde{K}_4/81\right)<\eta/3 , \tag{C.3}$$

$$P\left(\left(n^{-1}\sum_{t=1}^{n}\hat{u}_t\right)^4>\tilde{K}_4/81\right)<\eta/3 , \tag{C.4}$$

$$P\left(n^{-1}\sum_{t=1}^{n}u_t^4>\tilde{K}_4/81\right)<\eta/3 . \tag{C.5}$$

To prove (C.3), we use $\hat{u}_t-u_t=(\rho-\hat{\rho}_n)y_{t-1}$, the following equality

$$n^{-1}\sum_{t=1}^{n}(\hat{u}_t-u_t)^4=(\hat{\rho}_n-\rho)^4\ g(\rho,n)^4\ n^{-1}\sum_{t=1}^{n}\frac{y_{t-1}^4}{g(\rho,n)^4} ,$$

Markov inequality, and part (i) of Lemma MOMT-Y in Xu (2023). To verify (C.4), we use (C.2) from the proof of Part 2. Finally, Markov's inequality and Assumption 4.1 implies (C.5). ■

C.2 Proof of the Lemma B.2

Proof. We first prove that there exists $\tilde{M} = \tilde{M}(M) > 0$ and $N_0 = N_0(M) > 0$ such that

$$\rho + \frac{M}{n^{1/2} g(\rho, n)} \leq 1 + \frac{\tilde{M}}{n} \ , \tag{C.6}$$

for all $n \geq N_0$. Note that (C.6) is sufficient to conclude that $\hat{\rho}_n \leq 1 + \tilde{M}/n$ since $\hat{\rho}_n \leq \rho + Mn^{-1/2}/g(\rho_n, n)$.

Let us prove (C.6) by contradiction. That is: suppose that there exist sequences ρ_k, $\tilde{M}_k \to \infty$, and $n_k \to \infty$ such that $\rho_k + M/(n_k^{1/2} g(\rho_k, n_k)) > 1 + \tilde{M}_k/n_k$, for all k. The previous expression is equivalent to

$$M > n_k^{1/2} g(\rho_k, n_k)(1 - \rho_k) + \tilde{M}_k \frac{g(\rho_k, n_k)}{n_k^{1/2}} \ . \tag{C.7}$$

Define $a_k = n_k(1 - |\rho_{n_k}|)$. Consider the derivation to get a lower bound for $g(\rho_k, n_k)$:

$$\begin{aligned} g(\rho_k, n_k)^2 &= 1 + (1 - a_k/n_k)^2 + \ldots + (1 - a_k/n_k)^{2(n_k - 1)} \\ &= \frac{\{1 - (1 - a_k/n_k)^{n_k}\}\{1 + (1 - a_k/n_k)^{n_k}\}}{a_k/n_k\{2 - a_k/n_k\}} \\ &\geq \frac{n_k}{a_k} \times \frac{1 - e^{-a_k}}{2} \ , \end{aligned}$$

where the last inequality use that $(1 - a_k/n_k)^{n_k} = \exp(n_k \log(1 - a_k/n_k)) \leq \exp(-a_k)$. Without loss of generality, suppose that $a_k \to a \in [0, +\infty]$; otherwise, we can use a subsequence. We now consider two cases. For the first case, suppose $a_k \to +\infty$. This implies that

$$n_k^{1/2} g(\rho_k, n_k)(1 - \rho_k) \geq \left(\frac{1 - e^{-a_k}}{2}\right)^{1/2} a_k^{1/2} \to \infty \ ,$$

which contradicts (C.7). For the second case, suppose $a_k \to a$. This implies that

$$\frac{\tilde{M}_k}{\sqrt{n_k}} g(\rho_k, n_k) \geq \left(\frac{1 - e^{-a_k}}{2a_k}\right)^{1/2} \tilde{M}_k \to \infty \ ,$$

which contradicts (C.7). Therefore, there exists $N_0 = N_0(M)$ and $\tilde{M} = \tilde{M}(M)$ such that (C.6) holds for $n \geq N_0$. We can adapt the proof to conclude that $\hat{\rho}_n > -1 - \tilde{M}/n$ for all $n \geq N_0$. ∎

C.3 Proof of the Lemma B.3

Proof. We prove only item 1 since the proof of item 2 is analogous. The proof of item 1 has three steps. First, we can write $P_\rho\left(|R^*_{n,b}(h)| \leq x \mid Y^{(n)}\right) - (2\Phi(x) - 1) = I_1 + I_2 + I_3$, where $I_1 = J_n(x, h, \hat{P}_n, \hat{\rho}_n) - \Phi(x)$, $I_2 = \Phi(-x) - J_n(-x, h, \hat{P}_n, \hat{\rho}_n)$, and

$$I_3 = P_\rho(R^*_{n,b}(h) = -x \mid Y^{(n)}) \leq P_\rho(R^*_{n,b}(h) \in (-x - \epsilon/2, -x + \epsilon/2] \mid Y^{(n)}) \ .$$

Second, conditional on the event E_n defined in the proof of Theorem 4.1, the inequality (A.1) in the proof of Theorem 4.1 implies that $|I_1| < \epsilon/2$ and $|I_2| < \epsilon/2$ for any $n \geq N_2 = N_2(\epsilon, \eta)$ and any $h \leq h_n$ such that $h_n \leq n$ and $h_n = o(n)$. Also, the inequality (A.1) and algebra manipulation implies $|I_3| < 2\epsilon$. Therefore, we conclude

$$\sup_{h \leq h_n} \sup_{x \in \mathbf{R}} \left| P_\rho\left(|R^*_{n,b}(h)| \leq x \mid Y^{(n)}\right) - (2\Phi(x) - 1)\right| < 3\epsilon \ , \tag{C.8}$$

for any $n \geq N_2$ and any $h_n \leq n$ such that $h_n = o(n)$. Third, taking $x = z_{1-\alpha/2-3\epsilon/2}$ in (C.8), it follows that $\left|P_\rho\left(|R^*_{n,b}(h)| \leq z_{1-\alpha/2-3\epsilon/2} \mid Y^{(n)}\right) - (1 - \alpha - 3\epsilon)\right| < 3\epsilon$, which implies

$$P_\rho\left(|R^*_{n,b}(h)| \leq z_{1-\alpha/2-3\epsilon/2} \mid Y^{(n)}\right) \leq 1 - \alpha \ .$$

By definition of $c^*_n(h, 1-\alpha)$ as in (15), it follows that $c^*_n(h, 1-\alpha) \geq z_{1-\alpha/2-3\epsilon/2}$ holds conditional on the event E_n. We similarly obtain that $c^*_n(h, 1-\alpha) \leq z_{1-\alpha/2+3\epsilon/2}$ holds conditional on E_n. ■

C.4 Proof of the Lemma B.4

Proof. By Lemma C.12, there exists a random variable $\tilde{R}_n(h)$ such that

$$P_\rho\left(\left|R_n(h) - \tilde{R}_n(h)\right| > n^{-1-\epsilon}\right) \leq C n^{-1-\epsilon}\left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right) \ ,$$

where

$$\tilde{R}_n(h) \equiv (n-h)^{1/2} \mathcal{T}\left(\frac{1}{n-h}\sum_{t=1}^{n-h} X_t; \sigma^2, \psi_4^4, \rho\right)$$

and the sequence $\{X_t : 1 \leq t \leq n-h\}$ is defined in (A.12). Due to Lemma C.12, we know $\mathcal{T}$ is a polynomial. Define $\tilde{J}_n(x, h, P, \rho) = P_\rho\left(\tilde{R}_n(h) \leq x\right)$. Using Bonferroni's inequality, we conclude

$$|P_\rho\left(R_n(h)\leq x\right)-P_\rho\left(\tilde{R}_n(h)\leq x\right)|\leq D_n+P_\rho\left(\left|R_n(h)-\tilde{R}_n(h)\right|>n^{-1-\epsilon}\right)\ .$$

Therefore, $\sup_{x\in\mathbf{R}}|J_n(x,h,P,\rho)-\tilde{J}_n(x,h,P,\rho)|\leq D_n+Cn^{-1-\epsilon}\left(E[|u_t|^k]+E[u_t^{2k}]+E[u_t^{4k}]\right)$, which completes the proof of the Lemma. Note that the constant C is defined in Lemma C.12 and only depends on a, h, k, and C_σ. ■

C.5 Proof of the Lemma B.5

Proof. For item 1, for any fixed $\epsilon>0$, there exist $N_0=N_0(\epsilon)$ such that the next inclusion

$$\{|\hat{\rho}_n|>1-a/2\}\subseteq\{n^{1/2}|\hat{\rho}_n-\rho|>n^{1/2-\epsilon}\}\cup\{|\rho|>1-a\}\ ,$$

holds for any $n\geq N_0$. Since $|\rho|\leq 1-a$, we conclude $P\left(|\hat{\rho}_n|>1-a/2\right)\leq P\left(n^{1/2}|\hat{\rho}_n-\rho|>n^{1/2-\epsilon}\right)\leq Cn^{-1-\epsilon}\left(E[|u_t|^{k_1}]+E[u_t^{2k_1}]\right)$, for $k_1\geq 2(1+\epsilon)/(1-\epsilon)$, where the last inequality follows from Lemma C.15. This proves item 1.

For item 2, we use the definition of $\tilde{u}_t$ in (13), $\hat{u}_t=y_t-\hat{\rho}_n y_{t-1}$, where $\hat{\rho}_n$ is as in (12), and the model (1) to obtain $n^{-1}\sum_{t=1}^n\tilde{u}_t^r=n^{-1}\sum_{t=1}^n\left(\hat{u}_t-\bar{\hat{u}}\right)^r=n^{-1}\sum_{t=1}^n\left(u_t+(\rho-\hat{\rho}_n)y_{t-1}-\bar{\hat{u}}\right)^r$, where $\bar{\hat{u}}=n^{-1}\sum_{t=1}^n u_t+(\rho-\hat{\rho}_n)n^{-1}\sum_{t=1}^n y_{t-1}$. Using the multinomial formula and the previous expression, we have that $n^{-1}\sum_{t=1}^n\tilde{u}_t^r$ is equal to

$$n^{-1}\sum_{t=1}^n\sum_{r_1+r_2+r_3=r}\binom{r}{r_1,r_2,r_3}u_t^{r_1}\left((\rho-\hat{\rho}_n)y_{t-1}\right)^{r_2}\bar{\hat{u}}^{r_3}=I_1+I_2+E[u_t^r]\ ,$$

where

$$\begin{aligned}
I_1&=\sum_{r_1+r_2+r_3=r}\binom{r}{r_1,r_2,r_3}(\rho-\hat{\rho}_n)^{r_2}\ \bar{\hat{u}}^{r_3}\left\{n^{-1}\sum_{t=1}^n u_t^{r_1}y_{t-1}^{r_2}-m_{r_1,r_2}\right\}\\
I_2&=\sum_{r_1+r_2+r_3=r}\binom{r}{r_1,r_2,r_3}(\rho-\hat{\rho}_n)^{r_2}\ \bar{\hat{u}}^{r_3}m_{r_1,r_2}-E[u_t^r]\\
m_{r_1,r_2}&=E\left[u_t^{r_1}\left(\sum_{j=1}^{i-1}\rho^{i-1-j}u_j\right)^{r_2}\right]\ .
\end{aligned}$$

Note that Lemmas C.8, C.14, and C.15 imply that $P(|\rho-\hat{\rho}_n|>n^{-\epsilon})\leq Cn^{-1-\epsilon}$, $P(|\bar{\hat{u}}|>n^{-\epsilon})\leq Cn^{-1-\epsilon}$, and $P(|n^{-1}\sum_{t=1}^n u_t^{r_1}y_{t-1}^{r_2}-m_{r_1,r_2}|>n^{-\epsilon})\leq Cn^{-1-\epsilon}$ for some constant C.

Therefore, Lemmas C.8 implies $P(|I_j| > n^{-\epsilon}) \leq Cn^{-1-\epsilon}$ for $j = 1, 2$. This implies that, for a fixed r, we have $P(|n^{-1}\sum_{t=1}^{n} \tilde{u}_t^r - E[u_t^r]| > n^{-\epsilon}) \leq Cn^{-1-\epsilon}$.

For item 3, we note that item 2 implies $P(|n^{-1}\sum_{t=1}^{n} \tilde{u}_t^2 - E[u_t^2]| > E[u_t^2]/2) \leq Cn^{-1-\epsilon}$. Therefore, we conclude item 3 by taking $\tilde{C}_\sigma = E[u_t^2]/2$. For item 4, we note that item 2 implies $P(|n^{-1}\sum_{t=1}^{n} \tilde{u}_t^{4k} - E[u_t^{4k}]| > 1) \leq Cn^{-1-\epsilon}$. Then, we conclude item 4 by taking $M = E[u_t^{4k}] + 1$. ■

C.6 Proof of Theorem B.1

Additional Notation: Define $\xi_{n,t}(\rho_n, h_n) = y_{n,t+h} - \beta(\rho_n, h_n) y_{n,t}$ and recall $\beta(\rho, h) \equiv \rho^h$. Algebra shows

$$\xi_{n,t}(\rho_n, h_n) = \sum_{\ell=1}^{h_n} \rho_n^{h_n-\ell} u_{n,t+\ell} \quad . \tag{C.9}$$

Proof. The derivations presented on page 1811 in Montiel Olea and Plagborg-Møller (2021) implies

$$R_n(h) = \frac{\sum_{t=1}^{n-h} \xi_{n,t}(h)\hat{u}_{n,t}(h)}{\left(\sum_{t=1}^{n-h} \hat{\xi}_{n,t}(h)^2 \hat{u}_{n,t}(h)^2\right)^{1/2}} ,$$

which is equal to

$$\left(\frac{\sum_{t=1}^{n-h} \xi_{n,t}(h) u_{n,t}}{(n-h)^{1/2} V(\rho, h)^{1/2}} + \frac{\sum_{t=1}^{n-h} \xi_{n,t}(\rho, h)(\hat{u}_{n,t}(h) - u_{n,t})}{(n-h)^{1/2} V(\rho, h)^{1/2}} \right) \times \left(\frac{(n-h) V(\rho, h)}{\sum_{t=1}^{n-h} \hat{\xi}_{n,t}(h)^2 \hat{u}_{n,t}(h)^2} \right)^{1/2} ,$$

where $V(\rho, h) = E[\xi_{n,t}(h)^2 u_{n,t}^2]$. We then follow their approach and prove that under Assumption B.1: for any sequences $\{\rho_n : n \geq 1\} \subset [-1, 1]$, $\{\sigma_n^2 : n \geq 1\} \subset [\underline{\sigma}, \overline{\sigma}]$, and $\{h_n : n \geq 1\}$ satisfying $h_n = o(n)$ and $h_n \leq n$, we have

$$\begin{aligned}
(i) \quad & \frac{\sum_{t=1}^{n-h_n} \xi_{n,t}(h_n) u_{n,t}}{(n-h_n)^{1/2} V(\rho_n, h_n)^{1/2}} \xrightarrow{d} N(0, 1) , \\
(ii) \quad & \frac{\sum_{t=1}^{n-h} \xi_{n,t}(\rho, h)(\hat{u}_{n,t}(h) - u_{n,t})}{(n-h)^{1/2} V(\rho, h)^{1/2}} \xrightarrow{p} 0 , \\
(iii) \quad & \frac{\sum_{t=1}^{n-h} \hat{\xi}_{n,t}(h)^2 \hat{u}_{n,t}(h)^2}{(n-h) V(\rho, h)} \xrightarrow{p} 1
\end{aligned}$$

Finally, Lemmas C.4, C.5, and C.7 imply (i), (ii), and (iii), respectively. ■

Lemma C.1. *Suppose Assumption B.1 holds. Then, for any* (ρ_n, σ_n, h_n) *such that* $|\rho_n| \leq 1 + M/n$, $\sigma_n^2 \in [\underline{\sigma}, \overline{\sigma}]$, *and* $h_n \leq n$, *we have*

$$E\left[\xi_{n,t}(\rho_n, h_n)^4\right] \leq 4g(\rho_n, h_n)^4 K_4 \ ,$$

where $g(\rho, k) = \left(\sum_{\ell=0}^{k-1} \ \rho^{2\ell}\right)^{1/2}$ *and* $\xi_{n,t}(\rho_n, h_n)$ *is in* (C.9).

Proof. It follows from the proof of Lemma A.7 in Montiel Olea and Plagborg-Møller (2021). ∎

Lemma C.2. *Suppose Assumption B.1 holds. Then, for any* (ρ_n, σ_n, h_n) *such that* $|\rho_n| \leq 1 + M/n$, $\sigma_n^2 \in [\underline{\sigma}, \overline{\sigma}]$, *and* $h_n \leq n$, *we have*

$$E\left[\left(\frac{\sum_{t=1}^{n-h_n} \xi_{n,t}(\rho_n, h_n) y_{n,t-1}}{(n-h_n) g(\rho_n, n-h_n) g(\rho_n, h_n) \sigma_n^2}\right)^2\right] \leq \frac{n}{n-h_n} \times \frac{h_n}{n-h_n} \times \frac{\sqrt{4K_4}}{\underline{\sigma}} \ .$$

where $g(\rho, k) = \left(\sum_{\ell=0}^{k-1} \ \rho^{2\ell}\right)^{1/2}$ *and* $\xi_{n,t}(\rho_n, h_n)$ *is in* (C.9).

Proof. The definition of $\xi_{n,t}(\rho_n, h_n)$ in (C.9) implies $\sum_{t=1}^{n-h_n} \xi_{n,t}(\rho_n, h_n) y_{n,t-1} = \sum_{j=1}^{n} u_{n,j} b_{n,j}$, where $b_{n,j} = \sum_{t=j-h_n}^{j-1} \rho_n^{t+h_n-j} \ y_{n,t-1} I\{1 \leq t \leq n-h\}$. Note that $b_{n,j}$ is measurable with respect to the σ-algebra defined by $\{u_{n,k} : 1 \leq k \leq j-2\}$. Using Assumption B.1, we obtain

$$E\left[\left(\sum_{j=1}^{n} u_{n,j} b_{n,j}\right)^2\right] = \sum_{j=1}^{n} E[u_{n,j}^2 b_{n,j}^2] = \sigma_n^2 \sum_{j=1}^{n} E[b_{n,j}^2] \ .$$

Therefore, the derivation above implies $E\left[\left(\sum_{t=1}^{n-h_n} \xi_{n,t}(\rho_n, h_n) y_{n,t-1}\right)^2\right] = \sigma_n^2 \sum_{j=1}^{n} E[b_{n,j}^2]$. We claim that

$$E[b_{n,j}^2] \leq h_n g(\rho_n, h_n)^2 g(\rho_n, n-h_n)^2 \sqrt{4K_4} \ , \tag{C.10}$$

for any $j = 1, \ldots, n$. The previous claim and Assumption B.1 imply

$$E\left[\left(\frac{\sum_{t=1}^{n-h_n} \xi_{n,t}(\rho_n, h_n) y_{n,t-1}}{(n-h_n) g(\rho_n, n-h_n) g(\rho_n, h_n) \sigma_n^2}\right)^2\right] \leq \frac{n h_n}{(n-h_n)^2 \sigma_n^2} \times \sqrt{4K_4} \leq \frac{n h_n}{(n-h_n)^2 \underline{\sigma}} \times \sqrt{4K_4} \ .$$

To verify (C.10), we consider three cases. The first case is $j \leq h_n$, in which we derive that

$$E[b_j^2] = E[(\sum_{t=1}^{j-1} \rho^{t+h-j} y_{n,t-1})^2] \leq (j-1)E[\sum_{t=1}^{j-1} \rho^{2(t+h-j)} y_{n,t-1}^2]$$
$$\leq h_n \left(\sum_{t=1}^{j-1} \rho^{2(t+h-j)} \right) g(\rho_n, n-h_n)^2 \sqrt{4K_4}$$
$$\leq h_n g(\rho_n, h_n)^2 g(\rho_n, n-h_n)^2 \sqrt{4K_4} \ ,$$

where we use Loeve's inequality (see Theorem 9.28 in Davidson (1994)) in the first inequality above. In the second inequality, we use $E[y_{n,t-1}^2] \leq E[y_{n,t-1}^4]^{1/2}$, $y_{n,t-1} = \xi_{n,0}(\rho_n, t-1)$, Lemma C.1, and $g(\rho_n, t-1)^2 \leq g(\rho_n, n-h_n)^2$ for all $t = 1, \dots, n-h_n$. Note that we also use $j \leq h_n$, which also implies the last inequality above. The second case is $h_n+1 \leq j \leq n-h_n+1$. We follow the same approach as before and conclude $E[b_j^2] \leq h_n g(\rho_n, h_n)^2 g(\rho_n, n-h_n)^2 \sqrt{4K_4}$. In the final case, we have $j \geq n-h_n+2$. As before, we obtain $E[b_j^2] \leq h_n g(\rho_n, h_n)^2 g(\rho_n, n-h_n)^2 \sqrt{4K_4}$. ■

Lemma C.3. *Suppose Assumption B.1 holds. Then, for any (ρ_n, σ_n, h_n) such that $|\rho_n| \leq 1 + M/n$, $\sigma_n^2 \in [\underline{\sigma}, \overline{\sigma}]$, and $h_n \leq n$, we have*

$$E\left[\left(\frac{\sum_{t=1}^{n-h_n} u_{n,t} y_{n,t-1}}{(n-h_n)^{1/2} g(\rho_n, n-h_n)} \right)^2 \right] \leq 2K_4 \ ,$$

where $g(\rho, k) = \left(\sum_{\ell=0}^{k-1} \rho^{2\ell} \right)^{1/2}$.

Proof. It follows from the proof of Lemma E.8 in Montiel Olea and Plagborg-Møller (2021). ■

Lemma C.4. *Suppose Assumptions B.1 hold. Then, for any sequences $\{\rho_n : n \geq 1\}$ such that $|\rho_n| \leq 1 + M/n$, $\{\sigma_n^2 : n \geq 1\} \subset [\underline{\sigma}, \overline{\sigma}]$, and $\{h_n : n \geq 1\}$ satisfying $h_n = o(n)$ and $h_n \leq n$, we have*

$$\frac{\sum_{t=1}^{n-h_n} \xi_{n,t}(\rho_n, h_n) u_{n,t}}{(n-h_n)^{1/2} g(\rho_n, h_n) \sigma_n^2} \xrightarrow{d} N(0,1) \ , \tag{C.11}$$

where $\xi_{n,t}(\rho_n, h_n)$ is in (C.9) and $g(\rho, h)^2 = \sum_{\ell=1}^{h} \rho^{2h}$.

Proof. We adapt the proof of Lemma A1 in Montiel Olea and Plagborg-Møller (2021). We start by writing the term on the left-hand side term in (C.11) as follows

$$\sum_{t=1}^{n-h_n} \chi_{n,t} \ ,$$

where

$$\chi_{n,t} = \frac{\xi_{n,n-h_n+1-t}(\rho_n, h_n) u_{n,n-h_n+1-t}}{(n-h_n)^{1/2} g(\rho_n, h_n)\sigma_n^2} \ ,$$

for $i = 1, \ldots, n - h_n$. Define the σ-algebra $\mathcal{F}_{n,t} = \sigma\left(u_{n-h_n+j-t} : j \geq 1\right)$. Note that for any $t = 1, \ldots, n-h_n$, $\chi_{n,t}$ is measurable with respect to $\mathcal{F}_{n,t}$. Therefore, the sequence $\{\chi_{n,t} : 1 \leq t \leq n-h_n\}$ is adapted to the filtration $\{\mathcal{F}_{n,t} : 1 \leq t \leq n-h_n\}$. Moreover, $\xi_{n,n-h_n+1-t}(\rho_n, h_n)$ is measurable with respect to $\mathcal{F}_{n,t-1}$ since it is a function of $\{u_{n,n-h_n+j-(t-1)} : 1 \leq j \leq h_n\}$. This implies that $E[\chi_{n,t}|\mathcal{F}_{n,t-1}] = 0$ since

$$E[\chi_{n,t}|\mathcal{F}_{n,t-1}] = \frac{\xi_{n,n-h_n+1-t}(\rho_n, h_n)}{(n-h_n)^{1/2} g(\rho_n, h_n)\sigma_n^2} E[u_{n,n-h_n+1-t}|\mathcal{F}_{n,t-1}]$$

and by Assumption B.1 we conclude $E[u_{n,n-h_n+1-t}|\mathcal{F}_{n,t-1}] = E[u_{n,n-h_n+1-t}] = 0$.

The derivation presented above proves that the sequence $\{\chi_{n,t} : 1 \leq t \leq n - h_n\}$ is a martingale difference array with respect to the filtration $\{\mathcal{F}_{n,t} : 1 \leq t \leq n - h_n\}$. The result in (C.11) then follows by a martingale central limit theorem (Theorem 24.3 in Davidson (1994)), which requires

$$(i) \quad \sum_{t=1}^{n-h_n} E[\chi_{n,t}^2] = 1 \ , \quad (ii) \quad \sum_{t=1}^{n-h_n} \chi_{n,t}^2 \xrightarrow{p} 1 \ , \quad (iii) \quad \max_{1 \leq t \leq n-h_n} |\chi_{n,t}| \xrightarrow{p} 0 \ .$$

The condition (i) follows by using that $E[\xi_{n,t}(\rho_n, h_n)^2 u_{n,t}^2] = g(\rho_n, h_n)^2 \sigma_n^4$. To prove the condition (ii) is sufficient to show

$$\text{Var}\left(\sum_{t=1}^{n-h_n} \chi_{n,t}^2\right) \to 0 \ . \tag{C.12}$$

To prove (C.12), we first recall that

$$\sum_{t=1}^{n-h_n} \chi_{n,t}^2 = \frac{1}{(n-h_n) g(\rho_n, h_n)^2 \sigma_n^4} \times \sum_{t=1}^{n-h_n} \xi_{n,t}(\rho_n, h_n)^2 u_{n,t}^2 \ ,$$

where the second term of the right-hand side of the previous display can be decomposed into the sum of its expected value and another three zero mean terms:

$$(n-h_n)g(\rho_n,h_n)^2\sigma_n^4+\sum_{j=1}^{n}(u_{n,j}^2-\sigma_n^2)b_{n,j}+\sum_{j=1}^{n}u_{n,j}d_{n,j}+\sum_{j=1}^{n}(u_{n,j}^2-\sigma_n^2)g(\rho_n,h_n)^2\sigma_n^2\ ,$$

where $b_{n,j}=\sum_{t=j-h_n}^{j-1}\rho_n^{2(h_n+t-j)}u_{n,t}^2 I\{1\leq t\leq n-h_n\}$ and

$$d_{n,j}=\sum_{t=j-h_n}^{j-1}\sum_{\ell_2=1}^{j-t-1}u_{n,t+\ell_2}\rho_n^{2h-j+t-\ell_2}u_{n,t}^2 I\{1\leq t\leq n-h_n\}\ .$$

Note that $b_{n,j}$ and $d_{n,j}$ are measurable with respect to the $\sigma-$algebra $\sigma(u_{n,k}:1\leq k\leq j-1\}$. By Assumption B.1 and Loeve's inequality (Theorem 9.28 in Davidson (1994)), we conclude

$$E[b_{n,j}^2]\leq h_n\sum_{t=j-h_n}^{j-1}E[\rho_n^{4(h_n+t-j)}u_{n,t}^4]\leq h_n g(\rho_n,h_n)^4 K_4$$

and

$$E[d_{n,j}^2]\leq h_n E\left[\left(\sum_{\ell_2=1}^{j-i-1}u_{n,i+\ell_2}\rho_n^{2h-j+i-\ell_2}u_{n,t}^2\right)^2\right]\leq h_n g(\rho_n,h_n)^4\sigma_n^2 K_4$$

for all $j=1,\ldots,n$.

We use the decomposition presented above, Assumption B.1, and Loeve's inequality (see Theorem 9.28 in Davidson (1994)) imply that the left-hand side of (C.12) is lower or equal than

$$\frac{3\sum_{j=1}^{n}E[(u_{n,j}^2-\sigma_n^2)^2 b_{n,j}^2]+3\sum_{j=1}^{n}E[u_{n,j}^2 d_{n,j}^2]+3g(\rho_n,h_n)^4\sigma_n^4\sum_{j=1}^{n}E[(u_{n,j}^2-\sigma_n^2)^2]}{(n-h_n)^2 g(\rho_n,h_n)^4\sigma_n^8}\ .$$

By Assumption B.1 and the upper bounds that we found for $E[b_{n,j}^2]$ and $E[d_{n,j}^2]$, the previous expression is lower or equal than

$$\frac{3nh_nK_4^2}{(n-h_n)^2\underline{\sigma}^4}+\frac{3nh_nK_4}{(n-h_n)^2\underline{\sigma}^2}+\frac{3nK_4}{(n-h_n)^2\underline{\sigma}^2}\ .$$

The previous expression is $o(1)$ since $h_n=o(n)$ as $n\to\infty$. This implies that (C.12) holds.

Finally, to verify that condition (iii) holds is sufficient to show

$$(n-h_n)E\left[\chi_{n,t}^2 I\{|\chi_{n,t}^2|>c\}\right]\to 0\ , \tag{C.13}$$

for any $c>0$, where $I\{\cdot\}$ is the indicator function. If the condition in (C.16) holds, then condition (iii) follows by Theorem 23.16 in Davidson (1994). To verify (C.13), note that

$$\begin{aligned}
(n-h_n)E\left[\chi_{n,t}^2 I\{|\chi_{n,t}^2|>c\}\right] &\le\ (n-h_n)E\left[\frac{\chi_{n,t}^4}{c^2} I\{|\chi_{n,t}^2|>c\}\right]\\
&\le\ (n-h_n)\frac{E\left[\chi_{n,t}^4\right]}{c^2}\\
&=\ \frac{E\left[\xi_{n,n-h_n+1-t}(\rho_n,h_n)^4\right]E\left[u_{n,n-h_n+1-t}^4\right]}{(n-h_n)\sigma_n^8 g(\rho_n,h_n)^4c^2}\\
&\le\ \frac{4K_4E\left[u_{n,n-h_n+1-t}^4\right]}{(n-h_n)\sigma_n^8c^2}\ ,
\end{aligned}$$

where the equality above uses Assumption B.1, and the last inequality follows by Lemma C.1. By Assumption B.1 we obtain $(n-h_n)E\left[\chi_{n,t}^2 I\{|\chi_{n,t}^2|>c\}\right]\le 4K_4^2/((n-h_n)\underline{\sigma}^4c^2)$, which is sufficient to conclude (C.13). ∎

Lemma C.5. *Suppose Assumption B.1 holds. Then, for any sequences $\{\rho_n : n\ge 1\}$ such that $|\rho_n|\le 1+M/n$, $\{\sigma_n^2 : n\ge 1\}\subset[\underline{\sigma},\ \overline{\sigma}]$, and $\{h_n : n\ge 1\}$ satisfying $h_n=o(n)$ and $h_n\le n$, we have*

$$\frac{\sum_{t=1}^{n-h}\xi_{n,t}(\rho_n,h_n)(\hat{u}_{n,t}(h_n)-u_{n,t})}{(n-h_n)^{1/2}g(\rho_n,h_n)\sigma_n^2}\xrightarrow{p} 0\ , \tag{C.14}$$

where $\hat{u}_{n,t}(h_n)=y_{n,t}-\hat{\rho}_n(h_n)y_{n,t-1}$, $\hat{\rho}_n(h_n)$ is defined in (A.7), and $g(\rho,h)^2=\sum_{\ell=1}^h\rho^{2h}$.

Proof. A proof can be adapted from the proof of Lemma A.4 and Lemma E.8 in Montiel Olea and Plagborg-Møller (2021). Importantly, their Assumption 3 (relevant for the proof) holds due to Proposition B.1 in Appendix B. ∎

Lemma C.6. *Suppose Assumption B.1 holds. Then, for any sequences $\{\rho_n : n\ge 1\}$ such that $|\rho_n|\le 1+M/n$, $\{\sigma_n^2 : n\ge 1\}\subset[\underline{\sigma},\ \overline{\sigma}]$, and $\{h_n : n\ge 1\}$ satisfying $h_n=o(n)$ and*

$h_n \leq n$, we have

$$(i) \quad \frac{\hat{\beta}_n(h_n) - \beta(\rho_n, h_n)}{g(\rho_n, h_n)} \xrightarrow{p} 0 \ ,$$
$$(ii) \quad \frac{g(\rho_n, n-h_n)\left(\hat{\eta}(\rho_n, h_n) - \eta(\rho_n, h_n)\right)}{g(\rho_n, h_n)} \xrightarrow{p} 0 \ ,$$
$$(iii) \quad (n-h_n)^{1/2} \times g(\rho_n, n-h_n) \times (\hat{\rho}_n(h_n) - \rho_n) = O_p(1) \ ,$$

where $\hat{\eta}(\rho_n, h_n) = \rho_n \hat{\beta}_n(h_n) + \hat{\gamma}_n(h_n)$, $\eta(\rho_n, h_n) = \rho_n \beta(\rho_n, h_n) = \rho_n^{h_n+1}$, $\hat{\beta}_n(h_n)$ and $\hat{\gamma}_n(h_n)$ are defined in (A.6), $\hat{\rho}_n(h_n)$ is in (A.7), and $g(\rho, h)^2 = \sum_{\ell=1}^{h} \rho^{2h}$.

Proof. A proof can be adapted from the proof of Lemma A.3 in Montiel Olea and Plagborg-Møller (2021). Importantly, their Assumption 3 (relevant for the proof) holds due to Proposition B.1 in Appendix B. ■

Lemma C.7. *Suppose Assumption B.1 holds. Then, for any sequences $\{\rho_n : n \geq 1\}$ such that $|\rho_n| \leq 1 + M/n$, $\{\sigma_n^2 : n \geq 1\} \subset [\underline{\sigma}, \overline{\sigma}]$, and $\{h_n : n \geq 1\}$ satisfying $h_n = o(n)$ and $h_n \leq n$, we have*

$$\frac{\sum_{t=1}^{n-h_n} \hat{\xi}_{n,t}(h_n)^2 \hat{u}_{n,t}(h_n)^2}{(n-h_n) g(\rho_n, h_n)^2 \sigma_n^4} \xrightarrow{p} 1 \ ,$$

where $\hat{\xi}_{n,t}(h_n) = y_{n,t+h_n} - \hat{\beta}_n(h_n) y_{n,t} - \hat{\gamma}_n(h_n) y_{n,t-1}$, $\hat{u}_{n,t}(h_n) = y_{n,t} - \hat{\rho}_n(h_n) y_{n,t-1}$, $\hat{\beta}_n(h_n)$ and $\hat{\gamma}_n(h_n)$ are defined in (A.6), $\hat{\rho}_n(h_n)$ is defined in (A.7), and $g(\rho, h)^2 = \sum_{\ell=1}^{h} \rho^{2h}$.

Proof. We adapt the proof of Lemma A.2 in Montiel Olea and Plagborg-Møller (2021) presented in their Supplemental Appendix E.2. They claim that is sufficient to prove

$$\frac{\sum_{t=1}^{n-h_n} \hat{\xi}_{n,t}(h_n)^2 \hat{u}_{n,t}(h_n)^2}{(n-h_n) g(\rho_n, h_n)^2 \sigma_n^4} - \frac{\sum_{t=1}^{n-h_n} \xi_{n,t}(\rho_n, h_n)^2 u_{n,t}^2}{(n-h_n) g(\rho_n, h_n)^2 \sigma_n^4} \xrightarrow{p} 0 \ , \tag{C.15}$$

since they then can conclude using their Lemma A6, which implies

$$\frac{\sum_{t=1}^{n-h_n} \xi_{n,t}(\rho_n, h_n)^2 u_{n,t}^2}{(n-h_n) g(\rho_n, h_n)^2 \sigma_n^4} \xrightarrow{p} 1 \ .$$

We avoid using their Lemma A6 since its proof requires that the shocks have a finite 8th moment. Instead, we observe that (C.12) presented in the proof of Lemma C.4 implies the previous claim.

To verify (C.15), Montiel Olea and Plagborg-Møller prove that is sufficient to show that

$$\frac{\sum_{t=1}^{n-h_n}\left[\hat{\xi}_{n,t}(h_n)^2\hat{u}_{n,t}(h_n)^2 - \xi_{n,t}(\rho_n,h_n)^2u_{n,t}^2\right]^2}{(n-h_n)g(\rho_n,h_n)^2\sigma_n^4} \tag{C.16}$$

converges in probability to zero. To prove that, they derive the following upper bound for (C.16):

$$3[(\hat{R}_1)^{1/2}\times(\hat{R}_2)^{1/2}] + 3[(\hat{R}_3)^{1/2}\times(\hat{R}_4)^{1/2}] + 3[(\hat{R}_1)^{1/2}\times(\hat{R}_3)^{1/2}] ,$$

where

$$\hat{R}_1 = \frac{\sum_{t=1}^{n-h_n}\left[\xi_{n,t}(\rho_n,h_n) - \hat{\xi}_{n,t}(h_n)\right]^4}{(n-h_n)g(\rho_n,h_n)^4\sigma_n^8} , \qquad \hat{R}_2 = \frac{\sum_{t=1}^{n-h_n}u_{n,t}^4}{n-h_n}$$

$$\hat{R}_3 = \frac{\sum_{t=1}^{n-h_n}[\hat{u}_{n,t}(h_n)-u_{n,t}]^4}{n-h_n} , \qquad \hat{R}_4 = \frac{\sum_{t=1}^{n-h_n}\xi_{n,t}(\rho_n,h_n)^4}{(n-h_n)g(\rho_n,h_n)^4\sigma_n^8} .$$

In what follows, we use Assumption B.1 to prove that (i) $\hat{R}_1$ and $\hat{R}_3$ are $o_p(1)$ and (ii) $\hat{R}_2$ and $\hat{R}_4$ are $O_p(1)$, which are sufficient to conclude that (C.16) converges to zero in probability.

To verify $\hat{R}_1$ is $o_p(1)$, let us first observe that

$$\xi_{n,t}(\rho_n,h_n) - \hat{\xi}_{n,t}(h_n) = [\hat{\beta}_n(h_n) - \beta(\rho_n,h_n)]u_{n,t} + [\hat{\eta}_n(\rho_n,h_n) - \eta(\rho_n,h_n)]y_{n,t-1} ,$$

where $\hat{\eta}(\rho_n,h_n) = \rho_n\hat{\beta}_n(h_n) + \hat{\gamma}_n(h_n)$ and $\eta(\rho_n,h_n) = \rho_n\beta(\rho_n,h_n)$. Then, using Loeve's inequality (see Theorem 9.28 in Davidson (1994)), we obtain

$$\begin{aligned}\hat{R}_1 \le\, & 8\left(\frac{[\hat{\beta}_n(h_n)-\beta(\rho_n,h_n)]}{g(\rho_n,h_n)}\right)^4\left(\frac{\sum_{t=1}^{n-h_n}u_{n,t}^4}{(n-h_n)\sigma_n^8}\right)\\ &+ 8\left(\frac{g(\rho_n,n-h_n)[\hat{\eta}_n(\rho_n,h_n)-\eta(\rho_n,h_n)]}{g(\rho_n,h_n)}\right)^4\left(\frac{\sum_{t=1}^{n-h_n}y_{n,t-1}^4}{(n-h_n)g(\rho_n,n-h_n)^4\sigma_n^8}\right) .\end{aligned}$$

Note that the first term on the right-hand side in the previous expression goes to zero in probability due to part (i) in Lemma C.6, Markov's inequality, and Assumption B.1. The second term on the right-hand side in the previous expression goes to zero in probability due to part (ii) in Lemma C.6, Markov's inequality, and using that

$$E[y_{n,t-1}^4] = E[\xi_{n,0}(\rho_n,t-1)^4] \le g(\rho_n,n-h_n)^4 4K_4 , \tag{C.17}$$

where the inequality holds due to Lemma C.1 and $g(\rho_n, t-1)^4 \leq g(\rho_n, n-h_n)^4$ for all $i \leq n - h_n$. This completes the proof of $\hat{R}_1$ is $o_p(1)$.

To prove that $\hat{R}_3$ is $o_p(1)$, note that we can write:

$$\hat{R}_1 = (g(\rho_n, h_n)[\hat{\rho}_n(h_n) - \rho])^4 \frac{\sum_{t=1}^{n-h_n} y_{t-1}^4}{(n-h_n)g(\rho_n, h_n)^4}$$

since $\hat{u}_{n,t}(h_n) - u_{n,t} = (\hat{\rho}_n(h_n) - \rho)y_{t-1}$. Note that the right-hand side in the previous expression goes to zero in probability due to part (iii) in Lemma C.6, Markov's inequality, and using (C.17). This completes the proof of $\hat{R}_3$ is $o_p(1)$. Finally, Markov's inequality and Assumption B.1 implies that $\hat{R}_2$ is $O_p(1)$. While Markov's inequality and Lemma C.1 implies $\hat{R}_4$ is $O_p(1)$. ■

C.7 Proof of Theorem B.2

Proof. The proof of this theorem has two steps.

Step 1: The sample average $(n-h)^{-1/2}\sum_{t=1}^{n-h} X_t$ has a valid Edgeworth expansion up to an error $o\left(n^{-3/2}\right)$ due to the results in Götze and Hipp (1983). Assumption B.2 and the definition of X_t in (A.8) guarantees that we can use Theorem 1.2 in Götze and Hipp (1994), and this in turn implies that we can use the results in Götze and Hipp (1983) (Theorem 2.8 and Remark 2.12). We obtain an approximation error of $o\left(n^{-3/2}\right)$ since $E[|X_t|^6] < +\infty$, which holds due to Assumption B.2.(iii).

Step 2: The proof of Theorem 2 in Bhattacharya and Ghosh (1978) and the Edgeworth expansion for the sample average $(n-h)^{-1/2}\sum_{t=1}^{n-h} X_t$ guarantee the existence of Edgeworth expansion for the distribution $\tilde{J}_n$ defined in (A.11). Furthermore, the function $q_j(x, h, P, \rho)$ for $j = 1, 2, 3$ is a polynomial in x with coefficients that are polynomials of the moments of X_t (up to order $j+2$) since the sequence X_t is strictly stationary ($|\rho| < 1$). In particular, the coefficients of the polynomial $q_j(x, h, P, \rho)$ for $j = 1, 2$ are polynomials of moments of P (up to order 12) and ρ since the moments of X_t can be computed using the moments of u_t and ρ. Moreover, $q_j(x, h, P, \rho) = (-1)^j q_j(-x, h, P, \rho)$ since the sequence X_t is strictly stationary.
■

C.8 Proof of Theorem B.3

Proof. The proof has two steps.

Step 1: Define the events $E_{n,1} = \{|\hat{\rho}_n| \leq 1 - a/2\}$, $E_{n,2} = \{n^{-1} \sum_{t=1}^{n} \tilde{u}_t^2 \geq \tilde{C}_\sigma\}$, and $E_{n,3} = \{n^{-1} \sum_{t=1}^{n} \tilde{u}_t^{4k} < M\}$, where $\tilde{C}_\sigma$ and M are as in Lemma B.5. Define $E_n = E_{n,1} \cap E_{n,2} \cap E_{n,3}$. By Lemma B.5 and Assumption 5.1 it follows that $P(E_n^c) \leq C_2 n^{-1-\epsilon}$ for some constant C_2 that depends on the moments of u_t. Since $k > 8$, it follows that, conditional on E_n, the empirical distribution $\hat{P}_n$ verifies part (i) and (iii) of Assumption B.2. It is important to mention that Götze and Hipp (1994) use part (ii) of Assumption B.2 to guarantee the dependent-data version of the Cramer condition that appears in Götze and Hipp (1983); see Lemma 2.3 in Götze and Hipp (1994).

Step 2: Condition (iii) in Lemma 2.3 in Götze and Hipp (1994) holds for the the bootstrap sequence $X_{b,t}^*$ since it holds for the original sequence X_t, otherwise the function F in (A.8) verifies equation (8) in Götze and Hipp (1994). Therefore, the dependent-data version of the Cramer condition holds for the bootstrap sequence $X_{b,t}^*$. The results in Götze and Hipp (1994) implied that Edgeworth expansion exists for the sample average. Then, conditional on the event E_n we can repeat the arguments presented in the proof of Theorem B.2. ■

Lemma C.8. *Let $\{W_{n,j} : 1 \leq j \leq r\}$ be a sequence of random variables. Suppose that there exist constants c_j and C such that*

$$P(n^{1/2}|W_{n,j}| > c_j\delta) \leq C\delta^{-k} ,$$

for $j = 1, .., r$ and some $k \in \mathbf{N}$. Then, for any $r \geq 2$ and $\delta < n^{1/2}$, we have

1. $P(n^{1/2}|\sum_{j=1}^{r} W_{n,j}| > (\sum_{j=1}^{r} c_j)\delta) \leq rC\delta^{-k}$.
2. $P(n^{r/2}|\prod_{j=1}^{r} W_{n,j}| > (\prod_{j=1}^{r} c_j)\delta^r) \leq rC\delta^{-k}$.
3. $P(n^{1/2}|W_{n,1} + \prod_{j=2}^{r} W_{n,j}| > (c_1 + \prod_{j=2}^{r} c_j)\delta) \leq 2C\delta^{-k}$.
4. *If $c_1 n^{-1/2}\delta < 1$. Then, $P(|W_{n,1}| > 1 - b) \leq C\delta^{-k}$ for any $b \in (0, 1 - c_1 n^{-1/2}\delta)$.*
5. *If $c_1 n^{-1/2}\delta < 1$. Then, for any $b \in (0, 1 - c_1 n^{-1/2}\delta)$, we have*

$$P\left(n^{3/2}|(1 + W_{n,1})^{-1/2} - (1 - \frac{1}{2}W_{n,1} + \frac{3}{8}W_{n,1}^2)| > \frac{5}{16}b^{-7/2}c_1^3\delta^3\right) \leq 4C\delta^{-k}$$

and

$$P\left(n^{2/2}|(1+W_{n,1})^{-1}-(1-W_{n,1})|>\frac{2}{1+(1-b)^3}\delta^2\right)\leq 3C\delta^{-k} .$$

Proof. Bonferroni's inequality and $\{|W_{n,1}|>1-b\}\subseteq\{n^{1/2}|W_{n,1}|>c_1\delta\}$ for $b\leq 1-c_1n^{-1/2}\delta$ imply the proof of items 1–4. To prove the first part of item 5, we use Bonferroni's inequality to conclude that the left-hand side in item 5 is lower or equal to the sum of $P(|W_{n,1}|>1-b)$ and

$$P\left(n^{3/2}|(1+W_{n,1})^{-1/2}-(1-\frac{1}{2}W_{n,1}+\frac{3}{8}W_{n,1}^2)|>\frac{5}{16}b^{-7/2}c_1^3\delta^3,|W_{n,1}|\leq 1-b\right) .$$

Item 4 implies that the former term is bounded by $C\delta^{-k}$, while the latter term is lower or equal to

$$P\left(n^{3/2}\frac{5}{16}b^{-7/2}|W_{n,1}|^3>\frac{5}{16}b^{-7/2}c_1^3\delta^3,|W_{n,1}|\leq 1-b\right) ,$$

where the left-hand side term inside the previous probability used the Taylor Polynomial error and $|W_{n,1}|\leq 1-b$. By item 2, the above probability is lower or equal to $3C\delta^{-k}$. Finally, adding the upper and lower bounds concludes the first part of item 5. The second part is analogous. ■

Lemma C.9. *Let $\{Z_t : 1\leq it\leq n\}$ be a martingale difference sequence. Then, for any $k\geq 2$, we have*

$$E\left[\left|n^{-1/2}\sum_{t=1}^{n}Z_t\right|^k\right]\leq d_k\beta_{n,k} ,$$

where $\beta_{n,k}=n^{-1}\sum_{t=1}^{n}E[|Z_t|^k]$ and $d_k=(8(k-1)\max\{1,2^{k-3}\})^k$.

Proof. See Dharmadhikari et al. (1968), where this lemma is the main theorem. ■

Lemma C.10. *Suppose Assumption 5.1 holds. For any fixed $h,k\in\mathbf{N}$ and $a\in(0,1)$. Define*

$$f_n=\frac{\sum_{t=1}^{n-h}\xi_t(\rho,h)\hat{u}_t(h)}{n-h}$$

where $\xi_t(\rho,h)=\sum_{\ell=1}^{h}\rho^{h-\ell}u_{t+\ell}$, $\hat{u}_t(h)=y_t-\hat{\rho}_n(h)y_{t-1}$, and $\hat{\rho}_n(h)$ is as in (5). Then, for any $\rho\in[-1+a,1-a]$, there exists a constant $C=C(h,k,a,C_\sigma)$ such that we can write $f_n=f_{1,n}+f_{2,n}+f_{3,n}+f_{4,n}$, where $P\left((n-h)^{j/2}|f_{j,n}|\geq\delta^j\right)\leq C\delta^{-k}\left(E[|u_t|^k]+E[u_t^{2k}]+E[u_t^{4k}]\right)$ for any $\delta<(n-h)^{1/2}$ and $j\in\{1,2,3,4\}$.

Proof. In what follows we use $\xi_t = \xi_t(\rho, h)$. Using the definition of $\hat{\rho}_n(h)$, we obtain

$$f_n = \frac{\sum_{t=1}^{n-h} \xi_t u_t}{n-h} - \left(\frac{\sum_{t=1}^{n-h} u_t y_{t-1}}{n-h}\right) \left(\frac{\sum_{t=1}^{n-h} \xi_t y_{t-1}}{n-h}\right) \left(\frac{\sum_{t=1}^{n-h} y_{t-1}^2}{n-h}\right)^{-1} .$$

Let us define the components of f_n as follows:

$$\begin{aligned}
f_{1,n} &= \left(\frac{\sum_{t=1}^{n-h} \xi_t u_t}{n-h}\right) \\
f_{2,n} &= -\frac{(1-\rho^2)}{\sigma^2} \left(\frac{\sum_{t=1}^{n-h} u_t y_{t-1}}{n-h}\right) \left(\frac{\sum_{t=1}^{n-h} \xi_t y_{t-1}}{n-h}\right) \\
f_{3,n} &= \frac{(1-\rho^2)}{\sigma^4} \left(\frac{\sum_{t=1}^{n-h} u_t y_{t-1}}{n-h}\right) \left(\frac{\sum_{t=1}^{n-h} \xi_t y_{t-1}}{n-h}\right) \left(2\rho \frac{\sum_{t=1}^{n-h} u_t y_{t-1}}{n-h} + \frac{\sum_{t=1}^{n-h} u_t^2 - \sigma^2}{n-h}\right) \\
f_{4,n} &= f_n - (f_{1,n} + f_{2,n} + f_{3,n}) \ .
\end{aligned}$$

Note that by construction $f_n = \sum_{j=1}^4 f_{j,n}$. Lemma C.14 guarantees that each sample average in $f_{1,n}$, $f_{2,n}$, and $f_{3,n}$ verify the conditions to use Lemma C.8, which imply that $P\left((n-h)^{j/2} |f_{j,n}| \geq \delta^j\right) \leq C\delta^{-k} E[u_t^{2k}]$ for $j = 1, 2, 3$, where the constant C includes C_σ and a.

To prove $P\left((n-h)^{4/2} |f_{j,n}| \geq \delta^4\right) \leq C\delta^{-k} \left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right)$, we proceed in two steps.

Step 1: Define

$$W_n = \frac{\sum_{t=1}^{n-h} (1-\rho^2)\sigma^{-2} y_{t-1}^2}{n-h} - 1 \ .$$

We can use (1) and algebra to derive the following identity

$$W_n = -\frac{\sigma^{-2} y_{n-h}^2}{n-h} + 2\rho\sigma^{-2} \frac{\sum_{t=1}^{n-h} u_t y_{t-1}}{n-h} + \sigma^{-2} \frac{\sum_{t=1}^{n-h} u_t^2 - \sigma^2}{n-h} \ .$$

This implies that

$$f_{3,n} = \frac{(1-\rho^2)}{\sigma^2} \left(\frac{\sum_{t=1}^{n-h} u_t y_{t-1}}{n-h}\right) \left(\frac{\sum_{t=1}^{n-h} \xi_t y_{t-1}}{n-h}\right) \left(W_n + \frac{\sigma^{-2} y_{n-h}^2}{n-h}\right) \ .$$

Therefore, $f_{4,n} = f_n - f_{1,n} - f_{2,n} - f_{3,n}$ is equal to

$$-\frac{(1-\rho^2)}{\sigma^2}\left(\frac{\sum_{t=1}^{n-h} u_t y_{t-1}}{n-h}\right)\left(\frac{\sum_{t=1}^{n-h} \xi_t y_{t-1}}{n-h}\right)\left((1+W_n)^{-1} - (1-W_n) + \frac{\sigma^{-2} y_{n-h}^2}{n-h}\right) .$$

Step 2: Due to Lemma C.8, it is sufficient to show that

$$P\left((n-h)^{2/2}|(1+W_n)^{-1} - (1-W_n)| > \delta^2\right) \leq C\delta^{-k}\left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right) \quad \text{(C.18)}$$

and

$$P\left((n-h)^{2/2}\left|\frac{\sigma^{-2} y_{n-h}^2}{n-h}\right| > \delta^2\right) \leq C\delta^{-k}\left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right) . \quad \text{(C.19)}$$

Note that (C.18) follows by part 5 in Lemma C.8 since $P(n^{1/2}|W_n| > \delta) \leq C\delta^{-k}E[u_t^{2k}]$ due to Lemma C.14 and $\delta < n^{1/2}$. Finally, (C.19) follows by Markov's inequality and Lemma C.13. As we mentioned before, the constant C includes the constants C's that appear in Lemmas C.13 and C.14 that only depends on a, h, k, and C_σ. ■

Lemma C.11. *Suppose Assumption 5.1 holds. For any fixed $h, k \in \mathbf{N}$ and $a \in (0,1)$. Define*

$$g_n = \frac{\sum_{t=1}^{n-h} \hat{\xi}_t(h)^2 \hat{u}_t(h)^2}{V(n-h)} - 1$$

where $V = \sigma^2 \sum_{\ell=1}^{h} \rho^{2(h-\ell)}$, $\hat{\xi}_t(h)$ is as in (3), $\hat{u}_t(h) = y_t - \hat{\rho}_n(h) y_{t-1}$, and $\hat{\rho}_n(h)$ is as in (5). Then, for any $\rho \in [-1+a, 1-a]$, there exists a constant $C = C(h,k,a,C_\sigma)$ such that we can write $g_n = g_{1,n} + g_{2,n} + g_{3,n}$, where $P\left((n-h)^{j/2}|g_{j,n}| \geq \delta^j\right) \leq C\delta^{-k}\left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right)$ for any $\delta < (n-h)^{1/2}$ and $j \in \{1,2,3\}$.

Proof. In what follows we use $\xi_t = \xi_t(\rho, h) = \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell}$. As we did for the case of f_n in Lemma C.10, we utilize the linear regression formulas to define the components of g_n as functions of the sample average of functions of ξ_t, u_t, and y_{t-1}:

$$
\begin{aligned}
Vg_{1,n} &= \frac{\sum_{t=1}^{n-h}(\xi_t^2 u_t^2 - V)}{n-h} \\
Vg_{2,n} &= \frac{\psi_4^4}{\sigma^4}\left(\frac{\sum_{t=1}^{n-h}\xi_t u_t}{n-h}\right)^2 - \frac{2}{\sigma^2}\left(\frac{\sum_{t=1}^{n-h}\xi_t u_t}{n-h}\right)\left(\frac{\sum_{t=1}^{n-h}\xi_t u_t^3}{n-h}\right) \\
&\quad + (1-\rho^2)g(\rho,h)^2\left(\frac{\sum_{t=1}^{n-h}u_t y_{t-1}}{n-h}\right)^2 - \frac{2(1-\rho^2)}{\sigma^2}\left(\frac{\sum_{t=1}^{n-h}u_t y_{t-1}}{n-h}\right)\left(\frac{\sum_{t=1}^{n-h}\xi_t^2 u_t y_{t-1}}{n-h}\right) \\
&\quad + (1-\rho^2)\left(\frac{\sum_{t=1}^{n-h}\xi_t y_{t-1}}{n-h}\right)^2 - \frac{2(1-\rho^2)}{\sigma^2}\left(\frac{\sum_{t=1}^{n-h}\xi_t y_{t-1}}{n-h}\right)\left(\frac{\sum_{t=1}^{n-h}\xi_t u_t^2 y_{t-1}}{n-h}\right) \\
Vg_{3,n} &= Vg_n - V\left(g_{1,n} + g_{2,n}\right) \ ,
\end{aligned}
$$

where $\psi_4^4 = E[u_t^4]$ and $g(\rho,h)^2 = \sum_{\ell=1}^h \rho^{2(h-\ell)}$. By construction $g_n = \sum_{j=1}^3 g_{j,n}$. Lemma C.14 guarantees that each sample average in $g_{1,n}$ and $g_{2,n}$ verify the conditions to use Lemma C.8, which imply $P\left((n-h)^{j/2}|g_{j,n}| \geq \delta^j\right) \leq C\delta^{-k}\left(E[u_t^{4k}]\right)$ for $\delta < (n-h)^{1/2}$ and $j \in \{1,2\}$, , where the constant C includes C_σ and a. In what follows we prove $P\left((n-h)^{3/2}|g_{3,n}| \geq \delta^3\right) \leq C\delta^{-k}\left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right)$.

First, we write $Vg_{3,n} = R_{g,1} + R_{g,2}$, where $R_{g,1}$ and $R_{g,2}$ are specified below. We will prove that $P((n-h)^{3/2}|V^{-1}R_{g,j}| > \delta^3) \leq C\delta^{-k}E[u_t^{4k}]$ for $j = 1, 2$. To compute $g_{3,n}$ we use equation (3) and the following equality

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

where $\hat{\eta}_n(\rho,h) = \rho\hat{\beta}_n(h) + \hat{\gamma}_n(h)$. We also use that $\hat{u}_t(h) = y_t - \hat{\rho}_n(h)y_{t-1} = u_t - (\hat{\rho}_n(h) - \rho)y_{t-1}$. In what follows, we use $\hat{\beta} = \hat{\beta}_n(h)$, $\hat{\rho} = \hat{\rho}_n(h)$ and $\hat{\eta} = \hat{\eta}_n(\rho,h)$, and $\sum = \sum_{t=1}^{n-h}$ to simplified the heavy notation. We obtain $R_{g,2}$ is equal to

$$
\begin{aligned}
&2\left[(\eta-\hat{\eta}) + \frac{(1-\rho^2)}{\sigma^2}\left(\frac{\sum\xi_t y_{t-1}}{n-h}\right)\right]\left(\frac{\sum\xi_t u_t^2 y_{t-1}}{n-h}\right) + \frac{\sigma^4}{1-\rho^2}\left\{(\eta-\hat{\eta})^2 - \frac{(1-\rho^2)^2}{\sigma^4}\left(\frac{\sum\xi_t y_{t-1}}{n-h}\right)^2\right\} \\
&+ 2\left[(\rho-\hat{\rho}) + \frac{(1-\rho^2)}{\sigma^2}\left(\frac{\sum u_t y_{t-1}}{n-h}\right)\right]\left(\frac{\sum\xi_t^2 u_i y_{t-1}}{n-h}\right) + \psi_4^4\left\{(\beta-\hat{\beta})^2 - \frac{1}{\sigma^4}\left(\frac{\sum\xi_t u_t}{n}\right)^2\right\} \\
&+ 2\left[(\beta-\hat{\beta}) + \frac{1}{\sigma^2}\left(\frac{\sum\xi_t u_t}{n-h}\right)\right]\left(\frac{\sum\xi_t u_t^3}{n-h}\right) + \frac{\sigma^4 g_2^2}{1-\rho^2}\left\{(\rho-\hat{\rho})^2 - \frac{(1-\rho^2)^2}{\sigma^4}\left(\frac{\sum u_t y_{t-1}}{n-h}\right)^2\right\} \ ,
\end{aligned}
$$

and we obtain that $R_{g,1}$ is equal to

$$
\begin{aligned}
&(\beta-\hat{\beta})^2\left(\frac{\sum u_t^4}{n-h}-\psi_4^4\right)+(\rho-\hat{\rho})^2\left(\frac{\sum \xi_t^2 y_{t-1}^2}{n-h}-\frac{\sigma^4 g(\rho,h)^2}{1-\rho^2}\right)+(\eta-\hat{\eta})^2\left(\frac{\sum u_t^2 y_{t-1}^2}{n-h}-\frac{\sigma^4}{1-\rho^2}\right)\\
&2(\beta-\hat{\beta})^2(\rho-\hat{\rho})\left(\frac{\sum u_t^3 y_{t-1}}{n-h}\right)+2(\eta-\hat{\eta})^2(\rho-\hat{\rho})\left(\frac{\sum u_t y_{t-1}^3}{n-h}\right)+2(\beta-\hat{\beta})(\rho-\hat{\rho})^2\left(\frac{\sum \xi_t u_t y_{t-1}^2}{n-h}\right)\\
&+4(\beta-\hat{\beta})(\rho-\hat{\rho})\left(\frac{\sum \xi_t u_t^2 y_{t-1}}{n-h}\right)+2(\eta-\hat{\eta})(\rho-\hat{\rho})^2\left(\frac{\sum \xi_t y_{t-1}^3}{n-h}\right)+4(\eta-\hat{\eta})(\rho-\hat{\rho})\left(\frac{\sum \xi_t u_t y_{t-1}^2}{n-h}\right)\\
&+2(\beta-\hat{\beta})(\eta-\hat{\eta})\left(\frac{\sum u_t^3 y_{t-1}}{n-h}\right)+2(\beta-\hat{\beta})(\eta-\hat{\eta})(\rho-\hat{\rho})^2\left(\frac{\sum u_t y_{t-1}^3}{n-h}\right)\\
&+(\beta-\hat{\beta})^2(\rho-\hat{\rho})^2\left(\frac{\sum u_t^2 y_{t-1}^2}{n-h}-\frac{\sigma^4}{1-\rho^2}\right)+4(\beta-\hat{\beta})(\eta-\hat{\eta})(\rho-\hat{\rho})\left(\frac{\sum u_t^2 y_{t-1}^2}{n-h}-\frac{\sigma^4}{1-\rho^2}\right)\\
&+(\eta-\hat{\eta})^2(\rho-\hat{\rho})^2\left(\frac{\sum y_{t-1}^4}{n-h}-\frac{E[u_t^4]}{1-\rho^4}-6\frac{\rho^2\sigma^4}{(1-\rho^2)(1-\rho^4)}\right)+(\eta-\hat{\eta})^2(\rho-\hat{\rho})^2\frac{E[u_t^4]}{1-\rho^4}\\
&+\frac{\sigma^4}{1-\rho^2}(\beta-\hat{\beta})^2(\rho-\hat{\rho})^2+4(\beta-\hat{\beta})(\eta-\hat{\eta})(\rho-\hat{\rho})\frac{\sigma^4}{1-\rho^2}+(\eta-\hat{\eta})^2(\rho-\hat{\rho})^2\frac{6\rho^2\sigma^4}{(1-\rho^2)(1-\rho^4)}
\end{aligned}
$$

Note that $P((n-h)^{3/2}|V^{-1}R_{g,1}|>\delta^3)\leq C\delta^{-k}E[u_t^{4k}]$ follows by Lemmas C.8, C.14, and C.15, since each term between parenthesis in the definition of $R_{g,1}$ appears in Lemma C.14 or in items 5-8 of Lemma C.15. Similarly, $P((n-h)^{3/2}|V^{-1}R_{g,2}|>\delta^3)\leq C\delta^{-k}E[u_t^{4k}]$ follows by Lemmas C.8, C.14, and C.15, since each term between parenthesis in the definition of $R_{g,2}$ appears in Lemma C.14, the terms in brackets appears in items 1-4 of Lemma C.15, and the terms between curly brackets can be written as the product of terms like parenthesis and brackets terms. ■

Lemma C.12. *Suppose Assumption 5.1 holds. For any fixed $h\in\mathbf{N}$ and $a\in(0,1)$. Then, for any $\rho\in[-1+a,1-a]$ and $\epsilon\in(0,1/2)$, there exist a constant $C=C(a,h,k)>0$, where $k\geq 8(1+\epsilon)/(1-2\epsilon)$, and a real-valued function $\mathcal{T}(\cdot\,;\sigma^2,\psi_4^4,\rho):\mathbf{R}^8\to\mathbf{R}$, such that*

1. *$\mathcal{T}(\mathbf{0};\sigma^2,\psi_4^4,\rho)=0$,*
2. *$\mathcal{T}(x;\sigma^2,\psi_4^4,\rho)$ is a polynomial of degree 3 in $x\in\mathbf{R}^8$ with coefficients depending continuously differentiable on σ^2, ψ_4^4, and ρ,*
3. *$P_\rho\left(\left|R_n(h)-\tilde{R}_n(h)\right|>n^{-1-\epsilon}\right)\leq Cn^{-1-\epsilon}\left(E[|u_t|^k]+E[u_t^{2k}]+E[u_t^{4k}]\right)$,*

where $\sigma^2=E_P[u_1^2]$, $\psi_4^4=E_P[u_1^4]$,

$$\tilde{R}_n(h)\equiv(n-h)^{1/2}\mathcal{T}\left(\frac{1}{n-h}\sum_{t=1}^{n-h}X_t;\sigma^2,\psi_4^4,\rho\right)$$

and the sequence $\{X_t : 1 \le t \le n-h\}$ is defined in (A.12). Furthermore, the asymptotic variance of $\tilde{R}_n(h)$ equals one.

Proof. The proof has two main parts. We first use Lemmas C.10 and C.11 to approximate $R_n(h)$ using functions based on ξ_t, u_t, and y_{t-1}. We then replace y_{t-1} by z_{t-1}. We specifically define the polynomial $\mathcal{T}$.

Part 1: The derivations presented on page 1811 in Montiel Olea and Plagborg-Møller (2021) implies

$$R_n(h) = \frac{\sum_{t=1}^{n-h} \xi_t(h)\hat{u}_t(h)}{\left(\sum_{t=1}^{n-h} \hat{\xi}_t(h)^2 \hat{u}_t(h)^2\right)^{1/2}} ,$$

where $\xi_t(\rho, h) = \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell}$, $\hat{u}_t(h) = y_t - \hat{\rho}_n(h) y_{t-1}$, $\hat{\xi}_t(h) = y_{t+h} - \left(\hat{\beta}_n(h) y_t + \hat{\gamma}_n(h) y_{t-1}\right)$, and the coefficients $(\hat{\beta}_n(h), \hat{\gamma}_n(h))$ is as in (3) and $\hat{\rho}_n(h)$ is defined in (5). Define

$$f_n = \frac{\sum_{t=1}^{n-h} \xi_t(\rho, h)\hat{u}_t(h)}{n-h} \quad \text{and} \quad g_n = \frac{\sum_{t=1}^{n-h} \hat{\xi}_t(h)^2 \hat{u}_t(h)^2}{V(n-h)} - 1 ,$$

where $V = E[\xi_t(\rho, h)^2 u_t^2] = \sigma^4 \sum_{\ell=1}^{h} \rho^{2(h-\ell)}$. It follows that $R_n(h) = (n-h)^{1/2} V^{-1/2} f_n \left(1 + g_n\right)^{-1/2}$. Lemmas C.8, C.10 and C.11 imply $P((n-h)^{1/2} |V^{-1/2} f_n| > \delta) \le C\delta^{-k} \left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right)$ and $P((n-h)^{1/2} |g_n| > \delta) \le C\delta^{-k} \left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right)$.

Step 1: Define $\tilde{R}_{g,n} = (n-h)^{1/2} f_n V^{-1/2} \left(1 - \frac{1}{2} g_n + \frac{3}{8} g_n^2\right)$. Due to Lemma C.8, we have

$$P\left((n-h)^{3/2} \left|R_n(h) - \tilde{R}_{g,n}\right| > \delta^4\right) \le C\delta^{-k} \left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right)$$

since $P\left(n^{3/2} \left|(1+g_n)^{-1/2} - \left(1 - \frac{1}{2} g_n + \frac{3}{8} g_n^2\right)\right| > \delta^3\right) \le C\delta^{-k} \left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right)$.

Step 2: Define $\tilde{R}_{f,n} = (n-h)^{1/2} V^{-1/2} \left(f_{1,n} + f_{2,n} + f_{3,n}\right) \left(1 - \frac{1}{2} g_n + \frac{3}{8} g_n^2\right)$, where $f_n = \sum_{j=1}^{4} f_{j,n}$ as in Lemma C.10. We conclude

$$\begin{aligned} P\left((n-h)^{3/2} \left|\tilde{R}_{g,n} - \tilde{R}_{f,n}\right| > \delta^4\right) &= P\left((n-h)^{4/2} \left|V^{-1/2} f_{4,n} \left(1 - \frac{1}{2} g_n + \frac{3}{8} g_n^2\right)\right| \ge \delta^4\right) \\ &\le C\delta^{-k} \left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right) , \end{aligned}$$

where the last inequality follows by Lemmas C.8 and C.10.

Step 3: Define $\tilde{R}_{fg,n} = (n-h)^{1/2} V^{-1/2} \left(f_{1,n} + f_{2,n} + f_{3,n} - \frac{1}{2} f_{1,n} g_n - \frac{1}{2} f_{2,n} g_n + \frac{3}{8} f_{1,n} g_n^2\right)$. Lemmas C.8, C.10, and C.11 imply $P\left((n-h)^{3/2} \left|\tilde{R}_{f,n} - \tilde{R}_{fg,n}\right| > \delta^4\right) \le C\delta^{-k} \left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right)$

Step 4: Define $\tilde{R}_{y,n}(h) = (n-h)^{1/2}V^{-1/2}\left(\sum_{j=1}^{3} f_{j,n} - \frac{1}{2}f_{1,n}g_{1,n} - \frac{1}{2}f_{1,n}g_{2,n} - \frac{1}{2}f_{2,n}g_{1,n} + \frac{3}{8}f_{1,n}g_{1,n}^2\right)$, where $g_n = \sum_{j=1}^{3} g_{j,n}$ as in Lemma C.11. We use Lemmas C.8, C.10, and C.11 to conclude

$$P\left((n-h)^{3/2}\left|\tilde{R}_{fg,n} - \tilde{R}_{y,n}(h)\right| > \delta^4\right) \leq C\delta^{-k}\left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right) \ .$$

Step 5: By Bonferroni's: $P\left((n-h)^{3/2}\left|R_n(h) - \tilde{R}_{y,n}(h)\right| > \delta^4\right) \leq C\delta^{-k}\left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right)$.

Part 2: We consider $V^{1/2}\mathcal{T}(x;\sigma^2,\psi_4^4,\rho)$ is equal to the following polynomial

$$\left(f_1(x) + f_2(x) + f_3(x) - \frac{1}{2}\left(f_1(x)g_1(x) + f_1(x)g_2(x) + f_2(x)g_1(x)\right) + \frac{3}{8}f_1(x)g_1(x)^2\right) \ , \tag{C.20}$$

where $f_1(x) = x_1,\ f_2(x) = -\sigma^{-2}(1-\rho^2)x_5x_6,\ f_3(x) = \sigma^{-4}(1-\rho^2)x_5x_6(2\rho x_5 + x_2),\ g_1(x) = V^{-1}x_3$, and

$$g_2(x) = V^{-1}\left(\frac{\psi_4^4}{\sigma^4}x_1^2 - \frac{2}{\sigma^2}x_1x_4 + (1-\rho^2)\left(g(\rho,h)^2x_5^2 - \frac{2}{\sigma^2}x_5x_7 + x_6^2 - \frac{2}{\sigma^2}x_6x_8\right)\right)$$

Note that $\tilde{R}_{y,n}(h) = V^{1/2}\mathcal{T}\left((n-h)^{-1}\sum_{t=1}^{n-h}\tilde{X}_i;\sigma^2,\psi_4^4,\rho\right)$, where

$$\tilde{X}_t = (\xi_tu_t, u_t^2 - \sigma^2, (\xi_tu_i)^2 - V, \xi_tu_t^3, u_ty_{t-1}, \xi_ty_{t-1}, \xi_t^2u_ty_{t-1}, \xi_tu_t^2y_{t-1}) \ .$$

Since $z_t = y_t + \rho^t z_0$, it follows that

$$P\left((n-h)^{1/2}\left|\frac{\sum_{t=1}^{n-h} f_ty_{t-1}}{n-h} - \frac{\sum_{t=1}^{n-h} f_tz_{t-1}}{n-h}\right| > \delta\right) \leq C\delta^{-k}\left(E[u_t^{2k}] + E[u_t^{4k}]\right)$$

for $f_t = u_t, \xi_t, \xi_t^2u_i, \xi_tu_t^2$. Then, Lemma C.8 and step 5 in part 1 implies

$$P\left((n-h)^{3/2}\left|R_n(h) - \tilde{R}_n(h)\right| > \delta^4\right) \leq C\delta^{-k}\left(E[|u_t|^k] + E[u_t^{2k}] + E[u_t^{4k}]\right) \ ,$$

where $\tilde{R}_n(h) = V^{1/2}\mathcal{T}\left((n-h)^{-1}\sum_{t=1}^{n-h}X_t;\sigma^2,\psi_4^4,\rho\right)$ and the sequence $\{X_t : 1 \leq t \leq n-h\}$ is defined in (C.9). As we mentioned before, the constant C includes the constants C's that appear in Lemmas C.10, C.11, C.13 and C.14 that only depends on a, h, k, and C_σ. Finally, we take $\delta = n^{(1/2-\epsilon)/4}$ ■

Lemma C.13. *Suppose Assumption 5.1 holds. For fixed $a \in (0,1)$ and $k \geq 1$. Then, for* *any* $|\rho| \leq 1-a$*, there exists a constant* $C = C(a,k) > 0$ *such that*

$$E[|y_n|^k] \leq C\, E[|u_n|^k] \ , \ \forall n \geq 1 \ .$$

and $P(n^{-1/2}|y_n^s| > \delta) \leq C\, \delta^{-k} E[|u_n|^{sk}]$ *,* $\forall n \geq 1$ *.*

Proof. The proof goes by induction. For $k=1$, we have $|y_n| \leq |\rho||y_{n-1}| + |u_n|$, which implies that

$$E[|y_n|] \leq E[|u_n|] \left(1 + |\rho| + ... + |\rho|^{n-1}\right) \leq E[|u_n|]\ a^{-1} \ .$$

Therefore, the constant $C = a^{-1}$. We can also derive $y_n^2 = \rho^2 y_{n-1}^2 + 2\rho y_{n-1} u_n + u_n^2$, which implies $E[y_n^2] \leq \rho^2 E[y_{n-1}^2] + 2|\rho| E[|y_{n-1} u_n|] + E[u_n^2]$. Using that $E[|y_{n-1}u_n|] = E[|y_{n-1}|]E[|u_n|] \leq a^{-1}E[u_n^2]$, we conclude $E[y_n^2] \leq \rho^2 E[y_{n-1}^2] + (2|\rho|a^{-1}+1)E[u_n^2]$, which implies $E[y_n^2] \leq (2|\rho|a^{-1}+1)E[u_n^2]\left(1+\rho^2+...+\rho^{2(n-1)}\right) \leq C_1(a)E[u_n^2]$, where $C_1(a) = (2(1-a)/a+1)/a$. In this case, the constant $C = C_1(a)$.

Now, let us use $C_1(a)$ to construct $C_2(a)$ and so on. Suppose we already compute $C_k(a)$. Now, let us compute $C_{k+1}(a)$. We have

$$y_n^{2k+1} = \rho^{2k+1} y_{n-1}^{2k+1} + \sum_{\ell=1}^{2k} C_\ell^{2k+1} \rho^{2k+1-\ell} y_{n-1}^{2k+1-\ell} u_n^\ell + u_n^{2k+1}$$

By triangular inequality and similar arguments as before, we obtain

$$E[y_n^{2k+1}] \leq |\rho|^{2k+1} E[|y_{n-1}|^{2k+1}] + \sum_{\ell=1}^{2k} C_\ell^{2k+1} |\rho|^{2k+1-\ell} E[|y_{n-1}|^{2k+1-\ell}] E[|u_n^\ell|] + E[|u_n|^{2k+1}] \ ,$$

and by the inductive hypothesis, we know $E[|y_{n-1}|^{2k+1-\ell}]E[|u_n^\ell|] \leq C_{k-\ell/2} E[|u_n|^{2k+1}]$, where we used that $E[|X|^a]E[|X|^b] \leq E[|X|^{a+b}]$. Thus, the inductive hypothesis implies that

$$E[y_n^{2k+1}] \leq |\rho|^{2k+1} E[|y_{n-1}|^{2k+1}] + E[|u_n|^{2k+1}] \left(\sum_{\ell=1}^{2k} C_\ell^{2k+1} |\rho|^{2k+1-\ell} C_{k-\ell/2} + 1 \right) \ ,$$

in a similar way as in the initial case, we conclude. Note that the final constant C only involves a and k. The other case is analogous. ■

Lemma C.14. *Suppose Assumption 5.1 holds. For a given* $h, k \in \mathbf{N}$ *and* $a \in (0,1)$*. Then, for any* $|\rho| \leq 1-a$ *and* $h \in \mathbf{N}$*, there exist a constant* $C = C(a,k,r,s,h,C_\sigma) > 0$ *such that*

1. $P(\ (n-h)^{1/2}|(n-h)^{-1}\sum_{t=1}^{n-h} u_t^r y_{t-1}^s - m_{r,s}| > \delta) \leq C\delta^{-k} E\left[|u_t|^{(r+s)k}\right]$
2. $P(\ (n-h)^{1/2}|(n-h)^{-1}\sum_{t=1}^{n-h} \xi_t u_t^r y_{t-1}^s| > \delta) \leq C\delta^{-k} E\left[|u_t|^{(1+r+s)k}\right]$
3. $P(\ (n-h)^{1/2}|(n-h)^{-1}\sum_{t=1}^{n-h} \xi_t^2 u_t^2 - V| > 3\delta) \leq C\delta^{-k} E[u_t^{4k}]$
4. $P(\ (n-h)^{1/2}|(n-h)^{-1}\sum_{t=1}^{n-h} \xi_t^2 u_t y_{t-1}| > 3\delta) \leq C\delta^{-k} E[u_t^{4k}]$
5. $P(\ (n-h)^{1/2}|(n-h)^{-1}\sum_{t=1}^{n-h} \xi_t^2 y_{t-1}^2 - \sigma^4 g(\rho,h)^2(1-\rho^2)^{-1}| > 5\delta) \leq C\delta^{-k} E[u_t^{4k}]$

for any $\delta > 0$ *and any* $n > h$*, where* $\xi_t = \xi_t(\rho, h) = \sum_{\ell=1}^h \rho^{h-\ell} u_{t+\ell}$*,* $g(\rho,h) = \left(\sum_{\ell=1}^h \rho^{2(h-\ell)}\right)^{1/2}$*, and* $m_{r,s} = E\left[u_t^r \left(\sum_{j\geq 1} \rho^{j-1} u_{t-j}\right)^s\right]$.

Proof. Define the filtration $\mathcal{F}_j = \sigma(u_t : t \leq j)$. In what follows, we use Markov's inequality, Lemmas C.8, C.9, and C.13. The constant C will replace other constants and will only depend on a, k, r, s, h, C_σ, and the constants that appears in Lemmas C.9 and C.13.

Item 1: We prove this item by induction on s. First, consider $s = 0$ and $r \geq 0$. Note that $\{(u_t^r - E[u_t^r], \mathcal{F}_t) : 1 \leq t \leq n-h\}$ define a martingale difference sequence. Therefore, Markov's inequality and Lemma C.9 imply $P(\ (n-h)^{1/2}|(n-h)^{-1}\sum_{t=1}^{n-h} u_t^r - m_{r,0}| > \delta) \leq C\delta^{-k} E\left[|u_t|^{rk}\right]$, since $m_{r,0} = E[u_t^r]$. Let us suppose that item 1 holds for any (r, s) such that $r \geq 0$ and $s \leq s_0$ (this is a strong inductive hypothesis). Next, let us prove item 1 for (r, s_0+1). We write

$$(n-h)^{-1}\sum_{t=1}^{n-h} u_t^r y_{t-1}^{s_0+1} - m_{r,s_0+1} = I_1 + I_2\ ,$$

where $I_1 = (n-h)^{-1}\sum_{t=1}^{n-h}(u_t^r - m_{r,0})\, y_{t-1}^{s_0+1}$ and $I_2 = (n-h)^{-1}\sum_{t=1}^{n-h} m_{r,0}\left(y_{t-1}^{s_0+1} - m_{0,s_0+1}\right)$. Note that $\{((u_t^r - m_{r,0})\, y_{t-1}^{s_0+1}, \mathcal{F}_t) : 1 \leq t \leq n-h\}$ define a martingale difference sequence; therefore, we conclude that $P((n-h)^{1/2}|I_1| > \delta) \leq C\delta^{-k} E[|u_t|^{k(r+s_0+1)}]$ using Markov's inequality and Lemmas C.9 and C.13. Now, let us write

$$y_t^{s_0+1} = (\rho y_{t-1} + u_t)^{s_0+1} = \rho^{s_0+1} y_{t-1}^{s_0+1} + \sum_{j=1}^{s_0+1} \binom{s_0+1}{j} \rho^{s_0+1-j} u_t^j y_{t-1}^{s_0+1-j}\ ,$$

which implies the following identity

$$(1-\rho^{s_0+1})(n-h)^{-1}\sum_{t=1}^{n-h} y_{t-1}^{s_0+1} = \frac{-y_t^{s_0+1}}{n-h} + \sum_{j=1}^{s_0+1}\binom{s_0+1}{j}\rho^{s_0+1-j}\left((n-h)^{-1}\sum_{t=1}^{n-h} u_t^j y_{t-1}^{s_0+1-j}\right)\ .$$

In a similar way, using $z_t = \sum_{j\geq 1} \rho^{j-1} u_{t-j}$ instead of y_t, we can derive the following identity

$$(1-\rho^{s_0+1}) m_{0,s_0+1} = \sum_{j=1}^{s_0+1} \binom{s_0+1}{j} \rho^{s_0+1-j} m_{j,s_0+1-j} \ .$$

Using that $|m_{r,0}|^k = |E[u_t^r]|^k \leq E[|u_t|^{rk}]$, the previous two identities, the inductive hypothesis to $(n-h)^{-1}\sum_{t=1}^{n-h} u_t^j y_{t-1}^{s_0+1-j} - m_{j,s_0+1-j}$ for $j = 1, \ldots, s_0+1$, and Lemmas C.13 and C.8, we conclude that $P((n-h)^{1/2}|I_2| > \delta) \leq C\delta^{-k} E[|u_t|^{k(r+s_0+1)}]$, which completes the proof due to Lemma C.8.

Item 2: Consider the following derivation for a sequence of random variable $f_t \in \mathcal{F}_t$:

$$\sum_{t=1}^{n-h} \xi_t f_i = \sum_{t=1}^{n-h}\sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell} \ f_t = \sum_{t=1}^{n-h}\sum_{j=t+1}^{t+h} \rho^{t+h-j} u_j \ f_t = \sum_{j=1}^{n} u_j b_{n,j} \ ,$$

where $b_{n,j} = \sum_{t=j-h}^{j-1} \rho^{t+h-j} \ f_t \ I\{1 \leq t \leq n-h\}$. Note that $\{u_j b_{n,j} \in \mathcal{F}_j : 1 \leq j \leq n-h\}$ defines a martingale difference sequence and

$$E[|u_j b_{n,j}|^k] \leq h^{k-1} \sum_{t=j-h}^{j-1} |\rho|^{(t+h-j)k} \ E[|u_j|^k] E[|f_t|^k] I\{1 \leq t \leq n-h\} \leq C_{k,h,a} E[|u_t|^k] \max_{1\leq t\leq n-h} E[|f_t|^k] \ ,$$

where $C_{k,h,a} = h^{k-1}\sum_{\ell=1}^{h}(1-a)^{(h-\ell)k}$. Finally, we take $f_t = u_t^r y_{t-1}^s$ and use $E[|u_t^r y_{t-1}^s|^k] \leq CE[|u_t|^{rk}]E[|u_t|^{sk}]$ due to Lemma C.13. We conclude by Jensen that $E[|u_t|^k]E[|f_t|^k] \leq E[|u_t|^{(1+s+r)k}]$.

Item 3–4, we write each expression as the sum of its expected value and three martingale difference sequences. For item 3, we did a decomposition to prove (C.12) in the proof of Lemma C.4.

Item 5: We proceed in a similar (but not exactly) way as in the proof of items 3 and 4. Let us write $\sum_{t=1}^{n-h} \xi_t^2 y_{t-1}^2 - \sigma^4 g(\rho,h)^2 (1-\rho^2)^{-1}$ as the sum of three terms:

$$\sum_{j=1}^{n} (u_j^2 - \sigma^2) b_{n,j} + \sum_{j=1}^{n} u_j d_{n,j} + g(\rho,h)^2 \sigma^4 (1-\rho^2)^{-1} \sum_{j=1}^{n-h} (\sigma^{-2}(1-\rho^2) y_{t-1}^2 - 1) \ ,$$

where $b_{n,j} = \sum_{t=j-h}^{j-1} \rho^{2(h+t-j)} y_{t-1}^2 I\{1 \leq t \leq n-h\}$ and

$$d_{n,j} = \sum_{t=j-h}^{j-1} \sum_{\ell_2=1}^{j-t-1} u_{t+\ell_2} \rho^{2h-j+t-\ell_2} y_{t-1}^2 I\{1 \leq t \leq n-h\} \ .$$

Note that as before that $(u_j^2 - \sigma^2) b_{n,j}$ and $u_j d_{n,j}$ define a martingale difference sequence with respect to $\mathcal{F}_{j-1}$, and we can proceed as before. What is new is the third term, which can be controlled using item 1 and using that $m_{0,2} = \sigma^2 (1-\rho^2)^{-1}$. In particular, we obtain that

$$P(\ (n-h)^{1/2} |(n-h)^{-1} \sum_{t=1}^{n-h} \xi_t^2 y_{t-1}^2 - \sigma^4 g(\rho,h)^2 (1-\rho^2)^{-1}| > 5\delta) \leq I_1 + I_2 + I_3 \ ,$$

by Bonferroni's inequality, is lower or equal to

$$I_1 = P(\ |(n-h)^{-1/2} \sum_{j=1}^{n} (u_j^2 - \sigma^2) b_{n,j}| > \delta) \leq \delta^{-k} C_{k,h,a,1} d_k E[u_t^{2k}]^2$$

$$I_2 = P(\ |(n-h)^{-1/2} \sum_{j=1}^{n} u_j d_{n,j}| > \delta) \leq \delta^{-k} C_{k,h,a,2} d_k E[|u_t|^k]^2 E[u_t^{2k}]^2$$

$$I_3 = P(\ |(n-h)^{-1/2} \frac{g(\rho,h)^2 \sigma^4}{1-\rho^2} \sum_{j=1}^{n-h} (\sigma^{-2}(1-\rho^2) y_{t-1}^2 - 1)| > \delta) \leq \delta^{-k} d_k C_{k,h,a,3} E[u_t^{2k}]^2 \ ,$$

where we compute the first and second constant as we did in the proof of items 3 and 4, $C_{k,h,a,1} = 2^k \tilde{C}_{2k} h^{k-1} \sum_{\ell=1}^{h} (1-a)^{2(h-\ell)k}$,$C_{k,h,a,2} = h^{2(k-1)} \tilde{C}_{2k} \sum_{\ell_1=1}^{h} \sum_{\ell_2=1}^{h} (1-a)^{(2h-\ell_1-\ell_2)k}$, while the third one follows by item 1 and using that $m_{0,2} = \sigma^2(1-\rho^2)^{-1}$, $C_{k,h,a,2} = 2^k (\tilde{C}_k + 2) \left(\sum_{\ell=1}^{h} (1-a)^{2(h-\ell)} \right)^k$. We conclude $I_1 + I_2 + I_3 \leq C\delta^{-k} E[u_t^{4k}]$, where the constant C absorb all the previous constants. ■

Lemma C.15. *Suppose Assumption 5.1 holds. For fixed $h, k \in \mathbf{N}$ and $a \in (0,1)$. Then, for any $|\rho| \leq 1-a$, there exist a constant $C = C(a,k,h,C_\sigma) > 0$ such that*

1. $P\left((n-h)^{2/2} \left| \hat{\rho}_n(h) - \rho - (1-\rho^2)\sigma^{-2} \frac{\sum_{t=1}^{n-h} u_t y_{t-1}}{n-h} \right| > \delta^2 \right) \leq C\delta^{-k} \left(E[|u_t|^k] + E[u_t^{2k}] \right)$
2. $P\left((n-h)^{2/2} \left| \hat{\beta}_n(h) - \beta(\rho,h) - \sigma^{-2} \frac{\sum_{t=1}^{n-h} \xi_t(\rho,h) u_t}{n-h} \right| > \delta^2 \right) \leq C\delta^{-k} \left(E[|u_t|^k] + E[u_t^{2k}] \right)$
3. $P\left((n-h)^{2/2} \left| \hat{\gamma}_n(h) - \frac{(1-\rho^2) \sum_{t=1}^{n-h} \xi_t(\rho,h) y_{t-1}}{\sigma^2 (n-h)} + \frac{\rho \sum_{t=1}^{n-h} \xi_t(\rho,h) u_t}{\sigma^2(n-h)} \right| > \delta^2 \right) \leq C\delta^{-k} \left(E[|u_t|^k] + E[u_t^{2k}] \right)$

4. $P\left((n-h)^{2/2}\left|\hat{\eta}_n(\rho,h)-\eta(\rho,h)-\frac{(1-\rho^2)\sum_{t=1}^{n-h}\xi_t y_{t-1}}{\sigma^2(n-h)}\right|>\delta^2\right)\leq C\delta^{-k}\left(E[|u_t|^k]+E[u_t^{2k}]\right)$

5. $P(n^{1/2}|\hat{\rho}_n-\rho|>\delta)\leq C\delta^{-k}\left(E[|u_t|^k]+E[u_t^{2k}]\right)$

6. $P(n^{1/2}|\hat{\beta}_n(h)-\beta(\rho,h)|>\delta)\leq C\delta^{-k}\left(E[|u_t|^k]+E[u_t^{2k}]\right)$

7. $P(n^{1/2}|\hat{\gamma}_n(h)|>\delta)\leq C\delta^{-k}\left(E[|u_t|^k]+E[u_t^{2k}]\right)$

8. $P(n^{1/2}|\hat{\eta}_n-\eta|>\delta)\leq C\delta^{-k}\left(E[|u_t|^k]+E[u_t^{2k}]\right)$

for any $\delta<n^{1/2}$, where $\hat{\rho}_n(h)$ is as in (5), $(\hat{\beta}_n(h),\hat{\gamma}_n(h))$ is as in (3), and $\xi_t(\rho,h)=\sum_{\ell=1}^{h}\rho^{h-\ell}u_{t+\ell}$. $\hat{\eta}_n(\rho,h)=\rho\hat{\beta}_n(h)+\hat{\gamma}_n(h)$, $\eta(\rho,h)=\rho\beta(\rho,h)$.

Proof. To prove item 1, we first use the definition of $\hat{\rho}_n(h)$,

$$\hat{\rho}_n(h)-\rho=\frac{(n-h)^{-1}\sum_{t=1}^{n-h}u_t y_{t-1}}{(n-h)^{-1}\sum_{t=1}^{n-h}y_{t-1}^2}=\frac{(1-\rho^2)\sum u_t y_{t-1}}{\sigma^2(n-h)}(1+W_n)^{-1}\ ,$$

where $W_n=(1-\rho^2)\sigma^{-2}(n-h)^{-1}\sum_{t=1}^{n-h}y_{t-1}^2-1$. Using this notation, we have

$$\hat{\rho}_n(h)-\rho-\frac{(1-\rho^2)\sum u_t y_{t-1}}{\sigma^2(n-h)}=\frac{(1-\rho^2)\sum u_t y_{t-1}}{\sigma^2(n-h)}\left((1+W_n)^{-1}-1\right)\ .$$

Since $P(n^{1/2}|(n-h)^{-1/2}\sum_{t=1}^{n-h}u_t y_{t-1}|>\delta)\leq C\delta^{-k}E[u_t^{2k}]$ holds by Lemma C.14, it is sufficient to show that $P(n^{1/2}|(1+W_n)^{-1}-1|>\delta)\leq C\delta^{-k}E[u_t^{2k}]$ due to Lemma C.8. To prove the last inequality we use $P(n^{1/2}|W_n|>\delta)\leq C\delta^{-k}E[u_t^{2k}]$ (which holds by Lemma C.14) and part 5 in Lemma C.8.

The proof of items 2–3 follows from the same arguments as before. Finally, the proof of item 4 follows by the results of items 2 and 3, the definition of $\hat{\eta}_n(\rho,h)$ and $\eta(\rho,h)$, and Bonferroni's inequality. Items 5-8 are implied by items 1-4, Bonferroni's inequality, and Lemma C.14. ∎

Appendix D: Additional Tables

This appendix presents the additional results of the simulations.

ρ	h	RB	RB_{per-t}	RB_{hc3}	WB	WB_{per-t}	AA	AA_{hc2}	AA_{hc3}
Design 1: Gaussian i.i.d. shocks									
0.95	1	0.35	0.35	0.35	0.35	0.35	0.33	0.34	0.35
	6	0.83	0.81	0.83	0.86	0.84	0.71	0.73	0.74
	12	1.07	1.03	1.07	1.12	1.09	0.89	0.91	0.93
	18	1.15	1.11	1.15	1.21	1.17	0.98	1.00	1.03
1.00	1	0.35	0.35	0.35	0.35	0.35	0.33	0.34	0.35
	6	0.97	0.93	0.97	1.00	0.96	0.80	0.82	0.84
	12	1.51	1.41	1.51	1.57	1.48	1.12	1.15	1.17
	18	2.01	1.83	2.01	2.09	1.92	1.36	1.39	1.42
Design 2: Gaussian GARCH shocks									
0.95	1	0.44	0.43	0.44	0.46	0.45	0.41	0.43	0.44
	6	0.93	0.91	0.94	1.00	0.98	0.80	0.82	0.84
	12	1.10	1.06	1.11	1.19	1.15	0.91	0.94	0.97
	18	1.13	1.09	1.13	1.22	1.18	0.95	0.98	1.01
1.00	1	0.44	0.43	0.44	0.45	0.45	0.41	0.42	0.44
	6	1.10	1.06	1.11	1.17	1.13	0.91	0.93	0.96
	12	1.60	1.50	1.61	1.73	1.63	1.18	1.22	1.25
	18	2.04	1.86	2.05	2.21	2.04	1.37	1.41	1.45
Design 3: t-student i.i.d. shocks									
0.95	1	0.33	0.33	0.34	0.33	0.33	0.31	0.32	0.33
	6	0.81	0.79	0.82	0.84	0.82	0.68	0.71	0.73
	12	1.05	1.02	1.06	1.10	1.07	0.86	0.89	0.93
	18	1.14	1.10	1.15	1.19	1.16	0.94	0.98	1.02
1.00	1	0.33	0.33	0.34	0.34	0.33	0.31	0.32	0.33
	6	0.94	0.91	0.95	0.97	0.94	0.77	0.79	0.82
	12	1.49	1.39	1.50	1.54	1.45	1.07	1.11	1.16
	18	1.96	1.79	1.98	2.03	1.87	1.30	1.35	1.41
Design 4: mix-gaussian GARCH shocks									
0.95	1	0.46	0.45	0.46	0.46	0.45	0.42	0.43	0.44
	6	0.89	0.87	0.90	0.96	0.94	0.77	0.79	0.82
	12	1.01	0.98	1.02	1.11	1.07	0.86	0.88	0.91
	18	1.02	1.00	1.03	1.12	1.08	0.89	0.91	0.94
1.00	1	0.46	0.45	0.46	0.46	0.46	0.42	0.43	0.44
	6	1.06	1.01	1.07	1.12	1.09	0.87	0.90	0.92
	12	1.50	1.40	1.51	1.62	1.53	1.11	1.14	1.18
	18	1.88	1.71	1.89	2.06	1.90	1.28	1.31	1.35

Table D.1: Median length of confidence intervals for $\beta(\rho, h)$ with a nominal level of 90% and $n = 95$. 5,000 simulations and 1,000 bootstrap iterations.

ρ	h	RB	RB_{per-t}	RB_{hc3}	WB	WB_{per-t}	AA	AA_{hc2}	AA_{hc3}
Design 1: Gaussian i.i.d. shocks									
0.95	18	89.44	87.52	89.46	89.66	88.74	87.20	87.40	87.68
	40	90.30	87.92	90.28	91.06	88.98	88.74	89.08	89.48
1.00	18	88.62	88.78	88.56	89.14	89.58	82.16	82.70	83.02
	40	86.56	84.80	86.52	86.64	85.78	78.56	78.88	79.16
Design 2: Gaussian GARCH shocks									
0.95	18	86.64	86.44	86.70	88.58	88.20	84.28	84.62	85.22
	40	89.18	86.84	89.24	90.36	87.90	87.46	87.88	88.32
1.00	18	87.10	87.50	87.22	89.26	89.72	80.82	81.28	81.94
	40	84.10	82.56	84.10	85.46	85.16	76.54	76.88	77.30
Design 3: t-student i.i.d. shocks									
0.95	18	89.06	88.04	89.06	89.82	88.78	86.04	86.68	87.38
	40	89.84	87.26	89.94	90.66	88.70	88.04	88.72	89.46
1.00	18	89.36	88.96	89.54	89.98	89.42	82.36	83.00	83.64
	40	85.96	84.90	85.82	86.52	86.24	77.90	78.54	79.28
Design 4: mix-gaussian GARCH shocks									
0.95	18	83.00	86.32	83.10	84.50	87.50	81.22	81.54	82.00
	40	87.20	86.18	87.24	88.64	87.56	86.00	86.50	87.02
1.00	18	84.06	88.68	84.22	86.04	90.42	76.98	77.34	77.80
	40	81.24	84.06	81.20	82.76	85.90	73.44	73.98	74.36

Table D.2: Coverage probability (in %) of confidence intervals for $\beta(\rho, h)$ with a nominal level of 90% and $n = 240$. 5,000 simulations and 1,000 bootstrap iterations.