The Uniform Validity of Impulse Response Inference in Autoregressions

Atsushi Inoue[*]
Vanderbilt University

Lutz Kilian[†]
Federal Reserve Bank of Dallas
CEPR

First draft: September 10, 2018. This version: November 5, 2019.

Abstract: Existing proofs of the asymptotic validity of conventional methods of impulse response inference based on higher-order autoregressions are pointwise only. In this paper, we establish the uniform asymptotic validity of conventional asymptotic and bootstrap inference about individual impulse responses and vectors of impulse responses when the horizon is fixed with respect to the sample size. For inference about vectors of impulse responses based on Wald test statistics to be uniformly valid, lag-augmented autoregressions are required, whereas inference about individual impulse responses is uniformly valid under weak conditions even without lag augmentation. We introduce a new rank condition that ensures the uniform validity of inference on impulse responses and show that this condition holds under weak conditions. Simulations show that the highest finite-sample accuracy is achieved when bootstrapping the lag-augmented autoregression using the bias adjustments of Kilian (1999). The conventional bootstrap percentile interval for impulse responses based on this approach remains accurate even at long horizons. We provide a formal asymptotic justification for this result.

JEL code: C22, C52

Key words: Impulse response, autoregression, lag augmentation, asymptotic normality, bootstrap, uniform inference.

[*]Vanderbilt University, Department of Economics, Nashville, TN 37235-1819. E-mail: atsushi.inoue@vanderbilt.edu.

[†]Federal Reserve Bank of Dallas, Research Department, 2200 N. Pearl St., Dallas, TX 75201, USA. E-mail: lkilian2019@gmail.com (corresponding author).

1 Introduction

Impulse response analysis based on autoregressions plays a central role in quantitative economics (see Kilian and Lütkepohl 2017). Many researchers have cautioned against relying on pre-tests to diagnose and remove apparent unit roots in autoregressive processes (e.g., Elliott 1998; Pesavento and Rossi 2007; Gospodinov, Herrera and Pesavento 2011). As result, autoregressions are often estimated based on highly persistent data. A long-standing question has been how to assess the uncertainty about response estimates when the dominant autoregressive root may be close to unity. The asymptotic validity of conventional methods of asymptotic and bootstrap inference for individual impulse responses in stationary processes has been established in Lütkepohl (1990) and Gonçalves and Kilian (2004). Extensions to possibly integrated autoregressive processes are provided in Inoue and Kilian (2002a, 2003), building on Park and Phillips (1989) and Sims, Stock and Watson (1990). Kilian and Lütkepohl (2017) note that the assumptions underlying the analysis of higher-order autoregressions in Inoue and Kilian (2002a) may be relaxed further by fitting lag-augmented autoregressions, as proposed by Dolado and Lütkepohl (1996) and Toda and Yamamoto (1995).[1] All these asymptotic justifications, however, rely on pointwise convergence results. It is unclear whether they are valid uniformly across the parameter space.

In many econometric applications the distinction between pointwise and uniform validity, as discussed in Giraitis and Phillips (2006), Mikusheva (2007a), Andrews and Guggenberger (2009), and Kasy (2019), among others, is of no practical importance. This distinction matters, however, when the distribution of the statistic of interest changes with the value of the population parameter to be estimated, as is the case in the AR(1) model when the autoregressive root approaches unity. The concern is that for a $1-\alpha$ confidence interval C to be asymptotically valid we need to show that

$$\lim_{T\to\infty} \inf_{\rho} P_\rho(\rho \in C) \geq 1-\alpha,$$

where ρ denotes the AR(1) slope parameter and the infimum is taken over the parameter space of ρ. This condition means that there exists a sample size that guarantees the coverage accuracy of the interval for any parameter value ρ. In contrast, under the pointwise approximation, the actual coverage accuracy is not known and may become arbitrarily low, since the true value of ρ is not known.[2]

[1] Related work also includes Kurozumi and Yamamoto (2000) and Bauer and Maynard (2012).

[2] For example, Mikusheva (2007a) demonstrates that the conventional asymptotic normal approximation for the slope parameter in the AR(1) model with near unit roots is pointwise correct, but not uniformly correct, which helps

Concern over the reliability of conventional methods of inference when applied to highly persistent autoregressive processes has subsequently motivated the development of nonstandard asymptotic approximations based on local-to-unity processes. For example, Stock (1991) proposes constructing confidence intervals for the dominant autoregressive root by inverting unit root tests. Phillips (2014), however, proves that inference about the AR(1) slope parameter based on Stock's (1991) confidence interval, while asymptotically valid when the root is local to unity, has zero coverage asymptotically when the root is far enough from unity.[3]

The lack of a uniform asymptotic approximation across the parameter space has undermined the profession's confidence in the accuracy of standard confidence intervals in applied work and has created interest in confidence intervals that remain asymptotically valid whether the AR(1) slope parameter is unity, close to unity or far from unity. For example, under weak conditions, the grid bootstrap of Hansen (1999) can be shown to provide a uniformly asymptotically valid approximation to the distribution of the AR(1) slope parameter under both stationary and local-to-unity asymptotics (see Mikusheva 2007a).

While the AR(1) process has been studied extensively in the literature, there has been much less work on the problem of uniform inference in higher-order autoregressions, which are the workhorse model in applied work. Allowing for additional lags turns out to change the properties of the estimator of the autoregressive model dramatically. Our analysis shows that the lack of uniform validity of the conventional Gaussian asymptotic approximation does not extend to inference on individual slope parameters in higher-order autoregressive models. In the latter case, asymptotic normality holds uniformly across the parameter space. This result has important implications for inference on smooth functions of autoregressive slope coefficients such as impulse responses in autoregressions.

Our contribution to this literature is fivefold. First, we show that conventional asymptotic and bootstrap confidence intervals for individual impulse responses remain uniformly asymptotically valid, as long as the horizon of the impulse response remains fixed with respect to the sample size, generalizing the pointwise asymptotic results in Park and Phillips (1989), Sims, Stock and Watson (1990) and Inoue and Kilian (2002a). Our analysis covers both higher-order autoregressions and lag-augmented autoregressions. We provide a suitable rank condition that ensures that inference

explain the poor coverage accuracy of conventional confidence intervals in this model, as ρ approaches unity (see Nankervis and Savin 1988; Hansen 1999; Kilian 1999).

[3] Phillips' conclusion is consistent with simulation evidence in Hansen (1999), which illustrates the comparatively poor coverage accuracy of Stock's method in the stationary region.

on impulse responses is uniformly valid.[4]

Second, we establish the uniform asymptotic validity of Gaussian inference on vectors of autoregressive slope parameters and vectors of impulse responses. The joint asymptotic normality of the estimator of autoregressive slope parameters has been postulated as a high-level assumption in a range of studies including Guerron-Quintana, Inoue and Kilian (2017), Montiel Olea, Stock and Watson (2018), and Gafarov, Meier and Montiel Olea (2018). Our analysis establishes the unifom joint asymptotic normality of the lag-augmented estimator of the autoregressive slope parameters under conditions not requiring the process to be stationary. We furthermore establish, under the same conditions, the uniform joint asymptotic normality of the impulse response estimator postulated by Granziera, Moon and Schorfheide (2018).

The latter result is central for the construction of joint impulse response confidence intervals based on Wald test statistics. Joint inference on impulse response functions has become increasingly recognized as essential for practitioners interested in understanding the true extent of the uncertainty about estimates of impulse response functions (e.g., Jordà 2009; Lütkepohl; Staszewska-Bystrova and Winker 2015a,b, 2018; Inoue and Kilian 2016; Kilian and Lütkepohl 2017; Bruder and Wolf 2018; Montiel Olea and Plagborg-Møller 2019). Our analysis shows that the use of lag-augmented autoregressions is required for inference about impulse response functions based on Wald test statistics to be uniformly asymptotically valid.

Third, a simulation study involving univariate autoregressions with varying degrees of persistence confirms that the conventional asymptotic approximation based on fixed impulse response horizons remains accurate even uniformly, as long as the horizon is reasonably small relative to the sample size. We find that delta method impulse response confidence intervals based on lag-augmented autoregressions are considerably more accurate in small samples than confidence intervals based on the original autoregression. We then investigate the extent to which further improvements in coverage accuracy may be achieved by bootstrapping the lag-augmented autoregression. The reason that delta method confidence intervals tend to be less accurate than suitably constructed bootstrap confidence intervals is that, even for stationary processes, the finite-sample distribution of impulse responses is far from Gaussian. The longer the horizon, the worse the normal approximation becomes. One potential remedy is the use of the Hall percentile interval, which allows the distribution of the impulse response estimator to be non-Gaussian. We find,

[4] In related work, Mikusheva (2012) established the uniform validity of one-dimensional impulse response inference for a specific form of the grid bootstrap of Hansen (1999) applied to autoregressions. Her uniformity results, however, do not apply to the conventional delta method and bootstrap confidence intervals considered in our analysis.

however, that the use of the Hall percentile interval cannot be recommended because its coverage accuray at all but the shortest horizons tends to be too low. An alternative is the bias-adjusted bootstrap method of Kilian (1999), which was designed to improve the small-sample accuracy of impulse response confidence intervals in stationary autoregressions. This method yields consistently high uniform coverage accuracy when applied to lag-augmented autoregressions. For example, for $T = 240$, the uniform coverage rates for nominal 90% confidence intervals range from 87% to 89% and for $T = 480$ from 89% to 90%, for horizons between 1 and 12 periods. In contrast, without lag augmentation, the same type of bootstrap confidence interval is much less accurate, consistent with earlier simulation evidence in the literature. Our results suggest that persistent autoregressions in applied work should be routinely lag-augmented, if the coverage accuracy of the impulse response confidence intervals is the primary objective.

Fourth, we quantify the increase in average interval length caused by lag augmentation. We show by simulation and analytically that this increase is negligible at short horizons. At long horizons, the cost of maintaining accurate coverage by lag-augmenting the autoregression may be a substantial increase in interval length, when the data generating process is highly persistent. A local power analysis, however, shows that after adjusting for size, impulse-response inference based on the lag-augmented autoregression need not involve a loss in power in finite samples and may even enhance power in some cases.

Finally, we establish the asymptotic validity within the local-to-unity framework of the Efron (1979) percentile interval for impulse responses at long horizons based on the lag-augmented autoregression. Although impulse response inference is not uniformly valid at long horizons, these results help explain the excellent coverage accuracy of this approach when applied to data from persistent autoregressive processes at horizons as long as 60 periods. This result is in stark contrast to earlier theoretical and bootstrap simulation results based on autoregressions that were not lag augmented (see Phillips 1998; Kilian and Chang 2000). Our simulation evidence suggests that there is little need for nonstandard interval estimators based on long-horizon asymptotics in many applications of impulse response analysis. Our results also provide a formal justification for conducting long-horizon inference based on autoregressions in levels rather than in differences. Moreover, we formally show that other bootstrap confidence intervals for impulse responses based on lag-augmented autoregressions such as Hall's (1992) percentile interval or, for that matter, the delta method are not asymptotically valid at long horizons, even when lag-augmenting the autoregression. Likewise, equal-tailed and symmetric percentile-t intervals are not asymptotically valid.

This is a rare example of a situation in which Efron's percentile interval is asymptotically valid for impulse response inference, but other intervals are not.

The remainder of the paper is organized as follows. In section 2, we establish notation and state our assumptions about the data generating process and the estimated model. Section 3 contains the derivation of the uniform validity of the conventional asymptotic Gaussian approximation. We consider inference on individual impulse responses as well as vectors of impulse responses. In section 4, we establish the uniform asymptotic validity of inference based on the recursive-design bootstrap for autoregressions. In section 5, we examine the practical relevance of our asymptotic analysis in finite samples. In section 6, we provide long-horizon asymptotics based on the local-to-unity framework for impulse responses estimated from lag-augmented autoregressions. In section 7, we discuss the robustness of our conclusions to lag order misspecification. We also examine an alternative approach to lag augmentation based on moving average lags. Section 8 contains the concluding remarks. Details of the proofs can be found in Appendix A and B.

2 Preliminaries

2.1 Lag-Augmented Autoregressions

Lag-augmented autoregressions were first introducted by Dolado and Lütkepohl (1996) and Toda and Yamamoto (1995) and have been utilized in a range of applications (see, e.g., Kurozumi and Yamamoto 2000; Bauer and Maynard 2012; Kilian and Lütkepohl 2017). Unlike earlier studies, we are concerned with the use of lag augmentation for impulse response analysis. Let y_t denote a possibly integrated scalar time series variable. Consider fitting an autoregressive model of order p:

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + u_t,$$

where u_t is white noise, the deterministic terms have been suppressed for expository purposes, and $t = 1, ..., T$. Impulse responses may be constructed as a nonlinear transformation of the autoregressive coefficients. Impulse response inference based on consistent estimates of the p autoregressive coefficients in this model works well when the underlying process is known to be covariance stationary, but not necessarily when it is $I(1)$ or near $I(1)$ (see, e.g., Sims, Stock and Watson 1990). If the data generating process (DGP) is possibly (near) $I(1)$, an alternative is to fit the lag-augmented autoregressive model

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \phi_{p+1} y_{t-p-1} + u_t,$$

and to base impulse response inference on estimates of the first p autoregressive slope coefficients only . This is possible because ϕ_{p+1} is known to be zero in population and the estimates of $\phi_1, ..., \phi_p$ are consistent (see, e.g., Dolado and Lütkepohl 1996; Toda and Yamamoto 1995).[5] Our analysis examines the uniform asymptotic validity of impulse response inference based on the original and the lag-augmented autoregression.

2.2 Notation and Assumptions

As in Sims, Stock and Watson (1990) and Inoue and Kilian (2002a), consider a scalar autoregressive process of known order $p > 1$:

$$y_t = d_t^\dagger + y_t^\dagger,$$
$$y_t^\dagger = \phi_1 y_{t-1}^\dagger + \phi_2 y_{t-2}^\dagger + \cdots + \phi_p y_{t-p}^\dagger + u_t,$$

where $d_t^\dagger$ is a deterministic function of time, u_t is iid with zero mean and variance σ^2, and $\Delta y_0^\dagger = \cdots = \Delta y_{1-p}^\dagger = 0$. For expository purposes, we focus on linear time trends of the form $d_t^\dagger = \delta_0^\dagger + \delta_1^\dagger (t/T)$. This process has an augmented Dickey-Fuller representation:

$$\begin{aligned} \Delta y_t &= \delta_0 + \delta_1 \frac{t}{T} + \pi y_{t-1} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_{p-1} \Delta y_{t-p+1} + u_t \\ &= \beta' x_t + u_t, \end{aligned} \quad (1)$$

where $\delta_0 = \phi(1)\delta_0^\dagger + \delta_1^\dagger(\phi_1 + 2\phi_2 + \cdots + p\phi_p)$, $\delta_1 = \phi(1)\delta_1^\dagger$, $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, $\pi = -\phi(1) = \sum_{j=1}^p \phi_j - 1$, $\gamma_j = -(\phi_{j+1} + \cdots + \phi_p)$, $u_t \overset{iid}{\sim} (0, \sigma^2)$, $\beta = [\delta_0 \ \delta_1 \ \pi \ \gamma_1 \ \cdots \ \gamma_{p-1}]'$, and $x_t = [1 \ t/T \ \ y_{t-1} \ \Delta y_{t-1} \ \cdots \ \Delta y_{t-p+1}]'$. Equation (1) may be used to represent either the original autoregressive process or a lag-augmented autoregressive process. When $\phi_p \neq 0$, equation (1) represents the original AR(p) process. Alternatively, when $\phi_p = 0$, the AR(p) process underlying equation (1) without loss of generality may be reinterpreted as an autoregression of order $p - 1$ that has been augmented by one lag.

The autoregressive lag-order polynomial may equivalently be expressed as $\phi(L) = \Pi_{i=1}^p (1 - \rho_i L)$,

[5] In the case of possibly near $I(2)$ or $I(2)$ variables, the autoregressive model would be augmented by two autoregressive lags. In this paper, we restrict attention to variables that are possibly $I(0)$, $I(1)$ or near $I(1)$.

where $|\rho_1| \leq |\rho_2| \leq ... \leq |\rho_{p-1}| \leq |\rho_p|$ are the p autoregressive roots. Let $\omega(\rho_p) = \sigma^2/(\phi^\dagger(1))^2$ where $\phi^\dagger(L) = \phi(L)/(1-\rho_p L)$ and ρ_p is the largest root. Let $\theta = [\beta', \sigma^2]'$. Let the parameter space $\Theta \subset \Re^{d_\theta}$ denote the set of θ where $d_\theta = p+3$. Finally, let $J_{c_T}(r)$ denote an Ornstein-Uhlenbeck process such that $J_{c_T}(r) = \int_0^r e^{c_T(r-s)} dW(s)$, where $W(\cdot)$ is a standard Brownian motion defined on $[0,1]$ and $c_T = T\log(\max(|\rho_p|, \epsilon))$ for some $\epsilon \in (0,1)$.

The model is estimated by least squares, yielding

$$
\begin{aligned}
\widehat{\beta}_T &= \left(\sum_{t=p+1}^{T} x_t x_t'\right)^{-1} \sum_{t=p+1}^{T} x_t y_t, \\
\widehat{u}_t &= y_t - \widehat{\beta}_T' x_t, \\
\widehat{\sigma}_T^2 &= \frac{1}{T-p} \sum_{t=p+1}^{T} \widehat{u}_t^2, \\
\widehat{\sigma}_{4,T} &= \frac{1}{T-p} \sum_{t=1}^{T} (\widehat{u}_t^2 - \widehat{\sigma}_T^2)^2, \\
\widehat{\Sigma}_T &= \begin{bmatrix} \widehat{\sigma}_T^2 \otimes (\sum_{t=p+1}^{T} x_t x_t')^{-1} & 0_{(2+p)\times 1} \\ 0_{1\times(2+p)} & T^{-1}\widehat{\sigma}_{4,T} \end{bmatrix}.
\end{aligned}
$$

Assumption A. The data generating process satisfies:

(i) There are constants $\overline{\rho}$, $\underline{\rho}$, κ and K that do not depend on the data generating process, where $\overline{\rho} \in (0,1)$, $\underline{\rho} \in (0,1)$, and $0 < \kappa < K < \infty$, such that

$$\Theta = \{\theta : \kappa \leq \sigma^2 \leq K,\ -K \leq \delta_0^\dagger \leq K,\ -K \leq \delta_1^\dagger \leq K,\ |\rho_{p-1}| \leq \overline{\rho},\ \text{and either } |\rho_p| \leq \overline{\rho} \text{ or } \underline{\rho} \leq \rho_p \leq 1\}.$$

(ii) $\{u_t\}_{t=1}^T$ is a sequence of iid random variables with $E(u_t) = 0$ and

$$E\left\{\begin{bmatrix} u_t \\ u_t^2 - \sigma^2 \end{bmatrix}\begin{bmatrix} u_t \\ u_t^2 - \sigma^2 \end{bmatrix}'\right\} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma_4 \end{bmatrix},$$

where $\kappa \leq \sigma_4 \leq K$ and $\kappa \leq E(u_t^6) \leq K$.

Remarks.

1. Assumption (i) implies that the roots of $|\phi(z)| = 0$ are either all outside the unit circle in modulus or that $\phi(z) = 0$ has at most one unit root and all the other roots are outside the

unit circle in modulus. We rule out the possibility that the data are generated by an I(2) process or a near-I(2) process or that the process is explosive. This assumption is standard in the literature (e.g., Mikusheva 2012). Assumption (i) also rules out complex near unit roots and roots near -1.

2. When the model is augmented with one lag, the population coefficient on that lag is known to be zero. Although the augmented lag parameter is estimated, the uniform coverage rate is defined as the limit of the infimum of the coverage probabilities with respect to the other parameters, with the augmented lag parameter fixed at zero. This coverage rate is greater than or equal to the uniform coverage rate when the infimum is taken with respect to *all* parameters. Thus, without loss of generality, we focus on the latter.

3. We abstract from the complications introduced by conditional heteroskedasticity in the error term (see Gonçalves and Kilian 2004; Andrews and Guggenberger 2009).

4. As in Sims, Stock, and Watson (1990) and Mikusheva (2012), we deliberately abstract from the lag order selection problem. As discussed in Kilian and Lütkepohl (2017), conditioning on estimates of the lag order invalidates standard asymptotic inference on the autoregressive parameters.

5. In the theoretical analysis, we restrict attention to autoregressions of lag order $p > 1$, building on the insights in Sims, Stock and Watson (1990). This facilitates the comparison across methods of inference. Although standard impulse response inference based on the original autoregressive model breaks down when $p = 1$, lag-augmented inference remains asymptotically valid in that case, as illustrated in section 7.

3 Asymptotic Results for the Delta Method

Let $\widehat{\theta}_T$ denote the least-squares estimator of $\theta = [\beta',\ \sigma^2]' \in \Theta$ and let

$$H(c_T) \;=\; \Upsilon^{-1}(c_T) \begin{bmatrix} 1 & \frac{1}{2} & \omega(\rho_p)\int_0^1 J_{c_T}(r)dr & 0 \\ \frac{1}{2} & \frac{1}{3} & \omega(\rho_p)\int_0^1 rJ_{c_T}(r)dr & 0 \\ \omega(\rho_p)\int_0^1 J_{c_T}(r)dr & \omega(\rho_p)\int_0^1 rJ_{c_T}(r)dr & \omega(\rho_p)^2\int_0^1 J_{c_T}(r)^2dr & 0 \\ 0 & 0 & 0 & M \end{bmatrix} \Upsilon^{-1}(c_T), \tag{2}$$

$$D(c_T) \;=\; \Upsilon^{-1}(c_T) \begin{bmatrix} N_1 \\ N_2 \\ \sigma\omega(\rho_p)\int_0^1 J_{c_T}(r)dW(r) \\ N_3 \end{bmatrix}, \tag{3}$$

where $\Upsilon(c_T) = \text{diag}(1, 1, \frac{1}{\sqrt{-2c_T}}, I_{p+1}, 1)$, $c_T = T\log(\max(|\rho_p|, \epsilon))$ for some $\epsilon \in (0,1)$, ρ_p is the largest root of $\phi(z) = 0$, $T^{-\frac{1}{2}}\sum_{t=1}^{[\cdot T]} u_t \overset{d}{\to} W(\cdot)$, $T^{-\frac{1}{2}}\sum_{t=1}^{T}[1, t/T, \Delta y_{t-1}, \cdots, \Delta y_{t-p}]'u_t \overset{d}{\to} [N_1\ N_2\ N_3]'$ and $T^{-\frac{1}{2}}\sum_{t=1}^{T} u_t^2 \overset{d}{\to} N_4$ with $[N_1, N_2]' = [W(1), \int_0^1 rdW(r)]'$. $[N_1, N_2]'$, N_3 and N_4 are independent normal random vectors with zero means and covariance matrices given by

$$\sigma^2\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix},\ \sigma^2 M \;\equiv\; \sigma^2 E\left(\begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{bmatrix}\begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{bmatrix}'\right) \text{ and } \sigma_4, \tag{4}$$

respectively, and N is the standard normal random vector that is the limit of $\eta(\theta, T)$, as ρ_p approaches one from below.

Proposition 1: Suppose that Θ satisfies Assumption A. Then

$$\lim_{T\to\infty}\sup_{\theta\in\Theta}\sup_{x\in\Re^{d_\theta}}\left|P(\widehat{\Sigma}_T^{-\frac{1}{2}}(\widehat{\theta}_T - \theta) \le x) - P(\eta(\theta, T) \le x)\right| \;=\; 0, \tag{5}$$

where $\eta(\theta, T)$ is

$$\begin{bmatrix} H(c_T)^{-\frac{1}{2}}D(c_T) \\ N_4 \end{bmatrix}, \tag{6}$$

if $|\rho_p| < 1$ and N if $\rho_p = 1$.

Express the $d_\psi \times 1$ vector of impulse responses ψ as a function of θ:

$$\psi \;=\; f(\theta), \tag{7}$$

where $f : X \rightarrow \Re^{d_\psi}$ and $\Theta \subset X \subset \Re^{p+3}$. Our goal is to provide methods for uniform inference on impulse responses over Θ.

Assumption B. Suppose that $f : X \;\rightarrow \Re^{d_\psi}$ is continuously differentiable, that ψ does not depend on $\delta_0^\dagger$ and $\delta_1^\dagger$, and that the rank of

$$Df(\theta)diag(I_2, \sqrt{-2\log(\max(|\rho_p|, \epsilon))}, I_p) \tag{8}$$

is d_ψ for all $\theta \in \Theta$ where $Df(\theta) = \partial f(\theta)/\partial\theta'$, $\epsilon \in (0, 1)$ and ρ_p is the largest autoregressive root.

Remarks.

1. A violation of Assumption B would occur, for example, if the autoregressice parameters were zero in population and the impulse response horizon $h > p$ (see Benkwitz, Lütkepohl and Neumann 2000). We abstract from this well-known problem, as is standard in the literature, since we are concerned with inference about impulse responses estimated from persistent time series processes.

2. Likewise, assumption B may fail for some exact unit root processes. Specifically, standard delta method and bootstrap inference fails when the first-order linear approximation to ψ is proportionate to ρ. In that case, the limiting variance of $\sqrt{T}(\widehat{\psi}_T - \psi)$ is zero, as discussed in Kilian and Lütkepohl (2017), and the rank condition fails. Lag augmenting the autoregression helps rule out this singularity in the asymptotic variance of $f(\theta)$.

3. To appreciate the usefulness of the rank condition in Assumption B, consider the AR(2) process

$$\Delta y_t \;=\; \pi y_{t-1} + \gamma_1 \Delta y_{t-1} + u_t,$$

where $\pi = \phi_1 + \phi_2 - 1$ and $\gamma_1 = -\phi_2$, and the first two impulse responses are ϕ_1 and $\phi_1^2 + \phi_2$.

Then

$$Df(\theta) = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 2(\pi+\gamma_1+1) & 2(\pi+\gamma_1+1)-1 & 0 \end{bmatrix},$$

where the columns of zeros arise because there is no deterministic component, and the impulse responses do not depend on the error variance. Note that the conventional rank condition for the delta method is always satisfied in this example. Matrix (8) can be written as

$$T^{\frac{1}{2}} Df(\theta)\Upsilon_T^{-1}(c) = \begin{bmatrix} 0 & 0 & \sqrt{-2\log(\max(|\rho_2|,\epsilon))} & 1 & 0 \\ 0 & 0 & 2\sqrt{-2\log(\max(|\rho_2|,\epsilon))}(\pi+\gamma_1+1) & 2(\pi+\gamma_1+1)-1 & 0 \end{bmatrix},$$

where $\Upsilon_T(c_T) = \text{diag}(T^{\frac{1}{2}}, T^{\frac{1}{2}}, \frac{T^{\frac{1}{2}}}{\sqrt{-2c_T}}, T^{\frac{1}{2}} I_2)$.

This expression shows that, if one is interested in inference on the first impulse response, the rank condition is always satisfied. As $\phi_1, \phi_2 \to 1/2$, however, the second row approaches a vector of zeros, so inference about the second impulse response is not possible. Even though the conventional rank condition is satisfied, we can conduct uniform inference only on one of the parameters of interest.

In contrast, when fitting an AR(3) model to data generated by this AR(2) DGP, the rank of the matrix (8),

$$\begin{bmatrix} 0 & 0 & \sqrt{-2\log(\max(|\rho_3|,\epsilon))} & 1 & 0 & 0 \\ 0 & 0 & 2\sqrt{-2\log(\max(|\rho_3|,\epsilon))}(\pi+\gamma_1+1) & 2(\pi+\gamma_1+1)-1 & 1 & 0 \end{bmatrix},$$

is always 2. Thus, lag augmentation allows inference about the second impulse response as well as joint inference about both of the impulse responses.[6]

4. This example may be generalized. It can be shown that the rank condition in Assumption B is always satisfied for the first p impulse responses for all θ in the parameter space specified in Assumption A and for all $p = 1, 2, ...$, when the autoregression is augmented by one lag. Suppose that y_t follows an $AR(p)$ process. The companion matrix for the first p coefficients

[6] The concept of lag augmentation to overcome an asymptotic rank reduction in the Jacobian matrix of the impulse responses is conceptually similar to adding moments to overcome the asymptotic rank reduction in the moment Jacobian matrix that occurs in first-order underidentified or weakly identified GMM models (see Lee and Liao 2018).

of the lag-augmented model is given by

$$F = \begin{bmatrix} \pi+\gamma_1+1 & \gamma_2-\gamma_1 & \gamma_3-\gamma_2 & \cdots & \gamma_p-\gamma_{p-1} & \gamma_{p+1}-\gamma_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \tag{9}$$

Then the h-step-ahead impulse response is given by the (1,1) element of F^h. The responses are given by

$$\begin{aligned} &\pi+\gamma_1+1, \\ &(\pi+\gamma_1+1)^2+\gamma_2-\gamma_1, \\ &(\pi+\gamma_1+1)^3+2(\pi+\gamma_1+1)(\gamma_2-\gamma_1)+(\gamma_3-\gamma_2), \\ &\vdots \\ &(\pi+\gamma_1+1)^p+p(\gamma_{p+1}-\gamma_p)^{p-1}(\gamma_2-\gamma_1)+\cdots+\gamma_{p+1}-\gamma_p, \end{aligned}$$

for $h = 1, 2, 3, ..., p$, respectively. We are concerned about uniform joint inference about the first p impulse responses. Our claim is that the submatrix M_p obtained from eliminating the first two columns and the last column of the $p \times (p+3)$ matrix in Assumption B has rank p. Note that the first $p-1$ impulse responses are identical to those from an $AR(p)$ model with $\gamma_p = 0$. Thus, the $(p-1) \times (p-1)$ upper-left submatrix of M_p matches the corresponding $(p-1) \times (p-1)$ submatrix for the $AR(p)$ model that is obtained from lag-augmenting an $AR(p-1)$ model. We prove this claim by mathematical induction. When $p = 1$ (i.e., the DGP is an AR(1) process and an AR(2) model is fitted), the 1×2 matrix in Assumption B always has rank 1 satisfying Assumption B. Suppose that the rank condition is satisfied for $p = k$, i.e., the $k \times k$ submatrix has rank k. Denote that matrix by M_k. Then the $(k+1) \times (k+1)$ submatrix for the $AR(k+1)$ model can be written as

$$\begin{bmatrix} M_k & 0_{k\times 1} \\ 0_{1\times k} & 0 \end{bmatrix} + \begin{bmatrix} 0_{k\times k} & & 0_{k\times 1} \\ 0_{1\times(k-1)} & -1 & 1 \end{bmatrix}. \tag{10}$$

Because M_k has rank k, this matrix has rank $k+1$ anywhere in the parameter space specified in Assumption A. Thus, the claim holds for $p = k+1$. Since this result holds for the first p impulse responses jointly, it holds also for individual elements in this vector.

Proposition 2. Under Assumptions A and B,

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} \sup_{x\in\Re^{d_\psi}} \left| P\left((Df(\widehat{\theta}_T)\widehat{\Sigma}_T Df(\widehat{\theta}_T)')^{-\frac{1}{2}}(f(\widehat{\theta}_T) - f(\theta)) \leq x \right) - \Phi(x) \right| = 0, \quad (11)$$

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} \sup_{x\in\Re} \left| P\left((f(\widehat{\theta}_T) - f(\theta))'(Df(\widehat{\theta}_T)\widehat{\Sigma}_T Df(\widehat{\theta}_T)')^{-1}(f(\widehat{\theta}_T) - f(\theta)) \leq x \right) - F_{\chi^2_{d_\psi}}(x) \right| = 0, \quad (12)$$

where $F_{\chi^2_{d_\psi}}(\cdot)$ is the cdf of the chi-square distribution with d_ψ degrees of freedom.

It follows from Lemma 2 of Kasy (2019) that confidence sets constructed from quantiles of the standard normal and chi-square distributions have confidence level $1-\alpha$ uniformly on Θ.

4 Asymptotic Results for Bootstrap Inference

Bootstrap approximations of the asymptotic distribution of impulse response estimators may be generated by standard recursive residual-based bootstrap algorithms for autoregressions (see Kilian and Lütkepohl 2017). Let $\widehat{\theta}^*_T$ denote the bootstrap estimator of $\widehat{\theta}_T$, constructed by bootstrapping the original or the lag-augmented autoregressive model. Similarly, let P^* and $\widehat{\Sigma}^*_T$ denote the bootstrap analogue of P and $\widehat{\Sigma}_T$, respectively.

Proposition 3. Under Assumption A,

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} \sup_{x\in\Re^{d_\theta}} |P^*(\widehat{\Sigma}^{*-\frac{1}{2}}_T(\widehat{\theta}^*_T - \widehat{\theta}_T) \leq x) - P^*(\eta^*(\widehat{\theta}_T, T) \leq x)| = 0, \quad (13)$$

almost surely conditional on the data, where $c^*_T = c_T + \int_0^1 J^\tau_{c_T}(r)dW(r)/\int_0^1 (J^\tau_{c_T}(x))^2 dr$, $J^\tau_{c_T}(r) = J_{c_T}(r) - \int_0^1 (4-6s)J_{c_T}(s)ds - r\int_0^1 (12s-6)J_{c_T}(s)ds$, and $\eta^*(\cdot,\cdot)$ is $\eta(\cdot,\cdot)$ with c_T replaced by c^*_T.

Proposition 4. Under Assumptions A and B,

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} \sup_{x\in\Re^{d_\Psi}} |P^*((Df(\widehat{\theta}^*_T)\widehat{\Sigma}^*_T Df(\widehat{\theta}^*_T)')^{-\frac{1}{2}}(f(\widehat{\theta}^*_T) - f(\widehat{\theta}_T)) \leq x) - \Phi(x)| = 0, \quad (14)$$

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} \sup_{x\in\Re} |P^*(f(\widehat{\theta}^*_T) - f(\widehat{\theta}_T))(Df(\widehat{\theta}^*_T)\widehat{\Sigma}^*_T Df(\widehat{\theta}^*_T)')^{-1}(f(\widehat{\theta}^*_T) - f(\widehat{\theta}_T)) \leq x)) - F_{\chi^2_{d_\psi}}(x)| = 0, \quad (15)$$

almost surely conditional on the data.

To summarize, Propositions 1 through 4, extend the pointwise asymptotic validity results in Park and Phillips (1989), Sims, Stock and Watson (1990), and Inoue and Kilian (2002a) by establishing the uniform validity of asymptotic and bootstrap inference about individual slope parameters and impulse responses based on higher-order autoregressions. They also establish the corresponding results for asymptotic and bootstrap inference based on lag-augmented autoregressions. Finally, they establish the uniform validity of asymptotic and bootstrap inference based on lag-augmented autoregressions about vectors of impulse responses.

5 Simulation Evidence

In this section, we demonstrate that our asymptotic analysis helps understand the finite-sample accuracy of delta method and bootstrap confidence intervals for impulse responses. The bootstrap algorithms are reviewed in detail in Appendix D. For expository purposes, we generate 5,000 samples of $\{y_t\}_{t=1}^T$ from the data generating process $y_t = \rho y_{t-1} + u_t$, $u_t \overset{iid}{\sim} N(0,1)$, where $\rho \in \{0.2, 0.5, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 1\}$, $T \in \{80, 120, 240, 480, 600\}$ and $y_0 = 0$. Of particular interest are roots exceeding 0.95 because for smaller roots conventional bootstrap approximations are known to work well (see Kilian 1999).[7] For each sample of length T, we fit an AR(p) model, $p \in \{2, 4, 6\}$, with intercept and construct the implied responses to a unit shock at horizons $h \in \{1, ..., 12\}$. Lag-augmented autoregressions include an additional lag, but the impulse responses are based on the estimates of the first p slope coefficients only. We do not include a linear time trend in the fitted model because the inclusion of deterministic time trends is rare in applied work.[8]

Since the results are not sensitive to the lag order p, the tables shown in this section concentrate on the case of $p = 4$. Our analysis focuses on confidence intervals for individual impulse responses. The nominal confidence level is 90%. In constructing the uniform coverage rates as the infimum of

[7] Alternative specifications of the data generating process with standardized t_4 or standardized χ^2_3 errors, as in Kilian (1998a), yield results very similar to the baseline specification with $N(0, 1)$ errors and, hence, are not shown to conserve space. These specific distributions were chosen because their moments resemble those of residual distributions often encountered in applied work (see Kilian 1998a). Although the standardized t_4 distribution does not satisfy our sufficient condition A(iii), the simulation results are robust to this violation.

[8] Likewise, we do not consider autoregressions excluding an intercept. Normal asymptotic appoximations tend to work better when the regression model does not include an intercept because the exclusion of deterministic regressors reduces the small-sample bias of the least-squares estimator. This regression specification is hardly ever used in applied work, however.

the coverage rates for a given impulse response across $\rho \in \{0.2, 0.5, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 1\}$ for given T, one has to account for the bias caused by data mining across ρ. The reason is that the coverage rates in the simulation study (like any estimate of a proportion) are subject to estimation error. They have an approximate Gaussian distribution. Thus, even if the estimate of the coverage rate were centered on 0.90 for each ρ, there would be sampling variation in the simulated coverage rates. It can be shown by that under the null hypothesis that the coverage is truly 0.90, the lowest coverage rate across all ρ would be 0.8934, which biases downward our estimate of the uniform coverage rate. This means that we need to adjust upward the infimum across ρ obtained in the simulation by 0.0066 to control for data mining. This adjustment is independent of the sample size because it only reflects the Monte Carlo simulation error. Details of the rationale of this adjustment can be found in Appendix C.[9]

5.1 Coverage Accuracy at Short Horizons

The delta method intervals are based on closed-form solutions for the impulse-response standard error, as discussed in Lütkepohl (1990). Table 1 shows that the uniform coverage rates of the delta method interval converge to 0.90, as $T \to \infty$, as predicted by asymptotic theory, whether inference is based on the AR(4) model or the lag-augmented AR(5) model. There is strong evidence that delta method intervals based on the lag-augmented AR(5) model are considerably more accurate in small samples than delta method intervals based on the AR(4) model. For example, for $T = 480$, the uniform coverage accuracy at horizon 12 is 86% for the lag-augmented model compared with only 63% for the original model. These differences are not predicted by our asymptotic analysis in section 3. For much larger T, as expected, there is nothing to choose between these approaches.

Not surprisingly, the coverage accuracy is excellent at very short horizons, but deteriorates as h increases. This finding mirrors the conclusions of Kilian and Chang (2000) and Phillips (1998) that the conventional asymptotic approximation remains accurate, as long as the horizon is small relative to the sample size. Only when the horizon of the impulse response is allowed to grow with the sample size, conventional asymptotic approximations for impulse response estimators become asymptotically invalid in a local-to-unity setting. Table 1 illustrates that even for horizons as large as $h = 12$, for moderately large samples, the conventional asymptotic Gaussian approximation

[9] An alternative approach would have been to view ρ as local to unity and to report results for the implied ρ, given T and a grid of Pitman drifts. Since our asynptotic results do not hinge on this particular asymptotic thought experiment, it is more natural to focus on the grid of possible ρ values in the simulation study.

remains reasonably accurate. For example, for $T = 480$, coverage rates for the lag-augmented model range from 90% at short horizons to 86% at horizon 12.

An important question is whether the accuracy of impulse response inference may be improved by bootstrapping the impulse responses. In Table 2, we examine the coverage accuracy of several commonly used bootstrap confidence intervals.[10] The distribution of impulse response estimators is known to be non-normal in small samples (see Kilian 1999). The first two panels in Table 2 show results based on the Hall percentile interval with endpoints $\left[2\widehat{\psi}_T - \widehat{\psi}^*_{0.95,T}, 2\widehat{\psi}_T - \widehat{\psi}^*_{0.05,T}\right]$, where $\widehat{\psi}_T$ is the impulse response estimator based on the original sample and $\widehat{\psi}^*_{\gamma,T}$ denotes the critical point defined by the γ quantile of the distribution of the bootstrap estimator of the impulse response. This interval accounts for small-sample bias in the impulse response estimator and does not require normality to hold (see Hall 1992). The bootstrap estimators are generated based on the standard recursive design-bootstrap discussed in Appendix D. All results are based on 1,000 bootstrap replications. Table 2 shows that bootstrap confidence intervals may greatly improve the accuracy of inference based on the non-augmented AR(4) model. For example, the uniform coverage rates for $T = 480$ range from 90% at horizon 1 to 85% at horizon 12, which is much higher than for the delta method in Table 1 at longer horizons. When using the lag-augmented model, however, the Hall interval is not systematically more accurate than the delta method.

An alternative approach that has been shown to work well in bootstrapping stationary autoregressions is the bias-adjusted bootstrap of Kilian (1999), which replaces the least-squares estimates of the slope parameters by first-order mean bias-adjusted estimates when implementing the bootstrap (see Pope 1990). Impulse response intervals are based on the standard percentile interval of Efron (1979) with the interval endpoints defined as $\left[\widehat{\psi}^*_{0.05,T}, \widehat{\psi}^*_{0.95,T}\right]$. Table 2 shows that this method greatly improves the uniform coverage accuracy of the bootstrap confidence intervals in small samples, whether the model is lag-augmented or not, but by far the most accurate coverage rates are obtained based on the lag-augmented model. For $T = 80$, the coverage rates are between 80% and 87%, depending on the horizon. For $T = 120$, the coverage accuracy improves to between 83% and 88%. For $T = 240$, they are at least 87%, for $T = 480$ at least 88%, and for $T = 600$ at least 89%. This evidence suggests that the conventional asymptotic approximation remains accurate at longer horizons than previously thought possible. Performance deteriorates, when the autoregression is not lag-augmented, whether inference is based on the delta method or

[10] A general introduction to bootstrap methods for autoregressions and details of the construction of each of these intervals can be found in Kilian and Lütkepohl (2017).

the bias-adjusted bootstrap.

5.2 The Effect of Lag-Augmentation on Average Interval Length

Our coverage results suggest that accurate coverage for impulse response confidence intervals requires highly persistent autoregressions in applied work to be lag-augmented. This improvement in accuracy, however, tends to come at the cost of an increase in the average length of the interval. Since lag augmentation is necessary to control coverage accuracy in finite samples, it is difficult in general to compare the average length of intervals based on the original and on the lag-augmented model.

One way of assessing the cost of lag augmentation while controlling for coverage is to compare the average length of the Efron interval based on the bias adjusted autoregressive model augmented with two autoregressive lags to the same type of interval based on the bias-adjusted autoregressive model augmented by only one autoregressive lag. Since the coverage accuray of these intervals can be shown to be virtually the same, we can directly compare the average interval length. As Table A1 in the online appendix shows, the inclusion of the extra augmented lag has negligible effects on average interval length. For $T = 80$, the average interval length increases by between 1.4% at horizon 1 and 2.2% at horizon 12. For $T = 600$, this increase shrinks to between 0.2% at horizon 1 and 0.7% at horizon 12.

Additional insights may be gained by comparing the average length of the Efron percentile interval at short horizons, because at these horizons both the intervals based on the bias-adjusted autoregression augmented by one lag and the intervals based on the bias-adjusted original autoregression are about equally accurate (see Table 2). Table A2 in the online appendix shows the percentage increase in the average interval length at these horizons, computed as the average of the percentage increases in average interval length obtained for each ρ. We find that the increase in average interval length caused by the lag augmentation tends to be negligible. Even for $T = 80$ and $T = 120$ the average interval length increases by at most 1.5% and 1%, respectively, when a fifth autoregressive lag is added. For $T = 240$, that increase drops to 0.5% and for larger samples sizes the increase further reduces to 0.25% and 0.15%.

Our findings for short-horizon impulse responses are consistent with simulation evidence in Dolado and Lütkepohl (1996) that the power loss of Granger causality tests based on the lag-augmented model is quite modest when the autoregressive lag order is reasonably large. This does not mean that the benefits of lag augmentation come without cost, however. One important difference from the analysis in Dolado and Lütkepohl (1996) is that the distribution of the long-run impulse response depends on the sum of the autoregressive coefficients, when the dominant autoregressive root is modeled as local to unity. This sum is more accurately measured in the original autoregression, where $\widehat{\phi}_1 + \widehat{\phi}_2 + \cdots + \widehat{\phi}_p$ converges at rate T to a nonstandard distribution, than in the lag-augmented autoregression, where $\widehat{\phi}_1 + \widehat{\phi}_2 + \cdots + \widehat{\phi}_p$ is $\sqrt{T}$-consistent and asymptotically Gaussian. This reasoning suggests that impulse response inference based on lag-augmented models at longer horizons may involve a loss in power compared to inference based on the original autoregression, when the data generating process is highly persistent.

In Table 3, we investigate this conjecture by comparing the size-adjusted local power of impulse response inference based on the lag-augmented autoregression to that based on the original autoregression. We focus on the delta method, given the computational challenges in evaluating the power of the bootstrap procedures. Our analysis appeals to the duality of inference based on delta method confidence intervals and inference based on two-sided t-tests. All results shown are exact finite-sample results. The local power differences are expressed in percentage points with positive entries indicating a loss in size-adjusted power from lag-augmenting the autoregression and negative entries indicating a gain in size-adjusted power. Table 3 shows that there are situations when lag augmentation causes the power of the test to decline substantially at longer horizons, but there are also situations when the power remains unchanged or even increases. Thus, once the size is controlled for, finite-sample inference based on the lag-augmented model does not necessarily involve a loss in power, although it may involve a substantial loss in power in some cases.

5.3 Coverage Accuracy at Longer Horizons

An important question is how quickly the accuracy of our asymptotic approximation deteriorates with the impulse response horizon. Table 4 shows that the coverage accuracy of the Efron percentile interval based on the bias-adjusted lag-augmented autoregression is preserved even at much longer horizons. For example, for $T = 240$ the uniform coverage accuracy of the bootstrap confidence interval for the lag-augmented model is at least 88% at every impulse response horizon from 12 to 60. For $T = 480$ the lowest uniform coverage rate at these horizons is 89% and for $T = 600$ it is 90%.[11]

[11] Our analysis in Tables 1 through 4 focuses on nominal 90% intervals. Confidence levels of 90% or smaller are conventional for element-wise impulse response inference. In constructing joint confidence intervals based on

These results suggest that for many applications of impulse response analysis there is no need to rely on nonstandard interval estimators based on long-horizon asymptotics for impulse responses, as long as we apply the bias-adjusted bootstrap method to the lag-augmented autoregression. The superior accuracy of this method at longer horizons is not explained by our fixed-horizon asymptotics in sections 3 and 4, however. In the next section, we formally establish the pointwise asymptotic validity of this method (and this method alone) under the assumption that the impulse response horizon increases linearly with the sample size, which helps explain its greater robustness to the impulse response horizon.

6 Impulse Responses at Long Horizons When the DGP is Local to Unity

In this section, we show that Efron's percentile interval based on the lag-augmented autoregressive model (employed with or without bias adjustments of the autoregressice slope parameters) is asymptotically valid for long-horizon impulse response inference when there is a near unit root. We first consider the AR(1) model for illustration:

$$y_t \;=\; \phi_{1,T} y_{t-1} + u_t, \tag{16}$$

where $\phi_{1,T} = e^{c/T}$ for some constant $c \leq 0$, $y_0 = 0$, and $u_t \overset{iid}{\sim} (0, \sigma^2)$. As is well known, the estimator $\widehat{\phi}_{1,T}$ in this model has a nonstandard distribution, as does the $[\lambda T]$-step-ahead impulse response $\widehat{\phi}_{1,T}^{[\lambda T]}$. In contrast, in the lag-augmented model

$$y_t \;=\; \phi_{1,T} y_{t-1} + \phi_2 y_{t-2} + u_t, \tag{17}$$

which may equivalently be expressed as

$$\Delta y_t \;=\; \pi_T y_{t-1} + \gamma_1 \Delta y_{t-2} + u_t, \tag{18}$$

the Bonferroni method, it is customary to rely on much higher element-wise confidence levels (see Lütkepohl et al. 2015, 2018). It is therefore useful to note that, for $T = 240$ or larger, nominal 99% impulse response confidence intervals based on our preferred method have effective uniform coverage rates of between 98% and 99% at all horizons considered.

$\sqrt{T}(\widehat{\phi}_{1,T}-1) \;=\; \sqrt{T}(\widehat{\pi}_T+\widehat{\gamma}_{1,T}) = \sqrt{T}\widehat{\gamma}_{1,T}+o_p(1)$ is asymptotically normally distributed.

Consider the t-statistic for testing the null hypothesis that the $[\lambda T]$-step-ahead impulse response is 1:

$$\frac{\widehat{\phi}_{1,T}^{[\lambda T]}-1}{[\lambda T]\widehat{\phi}_{1,T}^{[\lambda T]-1}\widehat{ASE}(\widehat{\phi}_{1,T})}, \tag{19}$$

where $\widehat{ASE}(\widehat{\phi}_{1,T})$ is the estimate of the asymptotic standard error of $\widehat{\phi}_{1,T}$. Rewrite the t-statistic as

$$\left(1-\frac{1}{\widehat{\phi}_{1,T}^{[\lambda T]}}\right)\frac{\widehat{\phi}_{1,T}}{[\lambda T]\widehat{ASE}(\widehat{\phi}_{1,T})}. \tag{20}$$

It follows from

$$\frac{1}{\widehat{\phi}_{1,T}} \;=\; \frac{1}{1+\frac{1}{\sqrt{T}}\sqrt{T}(\widehat{\phi}_{1,T}-1)}, \tag{21}$$

that

$$\begin{aligned}\frac{1}{\lambda T^{\frac{1}{2}}}\left(1-\frac{1}{\widehat{\phi}_{1,T}^{[\lambda T]}}\right) \;&=\; -\frac{1}{\lambda T^{\frac{1}{2}}}\left(1+\frac{1}{\sqrt{T}}\sqrt{T}(\widehat{\phi}_{1,T}-1)\right)^{-[\lambda T]}+o_p(1)\\ &=\; -\frac{1}{\lambda T^{\frac{1}{2}}}(e^{-z})^{\lambda T^{\frac{1}{2}}}+o_p(1),\end{aligned} \tag{22}$$

where z is a zero-mean normal random variable such that $\sqrt{T}(\widehat{\gamma}_{1,T}-\gamma_1) \;\overset{d}{\to}\; z$. Combining (20) and (22), the t statistic can be approximated by

$$-\frac{1}{\lambda T^{\frac{1}{2}}}e^{-\lambda T^{\frac{1}{2}}z}\frac{\widehat{\phi}_{1,T}}{T^{\frac{1}{2}}\widehat{ASE}(\widehat{\phi}_{1,T})}. \tag{23}$$

Note that $\frac{\sqrt{T}\widehat{\phi}_{1,T}}{\widehat{ASE}(\widehat{\phi}_{1,T})}$ converges in probability to a constant. Thus,

$$z_T \;=\; T^{-\frac{1}{2}}\log(\text{-}t) \tag{24}$$

converges in distribution to a zero-mean normal random variable. As a result, z_T is positive with probability approaching 1/2 and is negative with probability approaching 1/2. As $T \;\to\; \infty$, the t-statistic,

$$t \;=\; -\exp(T^{\frac{1}{2}}z_T), \tag{25}$$

converges to zero with probability 1/2 and diverges to minus infinity with probability 1/2 as T

goes to infinity. Thus, the delta method interval fails. Noting that $T^{\frac{1}{2}}\widehat{ASE}(\widehat{\phi}_{1,T})$ converges in probability to a constant, the same argument applies to the unstudentized statistic, $\sqrt{T}(\widehat{\phi}_{1,T}^{[\lambda T]}-1)$ and $\sqrt{T}(\widehat{\phi}_{1,T}^{*[\lambda T]}-\widehat{\phi}_{1,T}^{[\lambda T]})$, invalidating Hall's percentile interval.

In contrast, impulse response inference based on Efron's percentile interval remains asymptotically valid at long horizons, because, as the horizon lengthens, the impulse response can be expressed as a monotonic function of the asymptotically normal estimator $\widehat{\phi}_{1,T}$. Unlike other confidence intervals, the Efron percentile interval is transformation-respecting. In other words, the interval for a given monotonic transformation of the original parameter may be obtained by applying the same transformation to the interval endpoints for the original parameter (see Efron 1979). The implications of this point for long-horizon impulse response inference are formalized in the following proposition.

Assumption C. The data generating process satisfies:

(i) $|\rho_j| < 1$ for $j = 1, ..., p-1$ and $\rho_p = e^{c/T}$ for some constant $c \leq 0$.

(ii) $\{u_t\}_{t=1}^{T}$ is a sequence of iid random variables with $E(u_t) = 0$ and $E(u_t^2) < \infty$.

Proposition 5: Suppose that the DGP is an AR(p) process, as described in section 2.2, and that Assumption C is satisfied. Then the Efron (1979) percentile interval based on the lag-augmented autoregressive model allows asymptotically valid inference about the impulse responses at long horizons.

Remarks.

1. Proposition 5 establishes the asymptotic validity of impulse response inference based on Efron's percentile interval within the local-to-unity framework. It does not establish its asymptotic validity in the stationary region. Thus, long-horizon inference is not uniformly valid in the parameter space. It is valid only for roots close to unity. Further simulations (not shown to conserve space) suggest that for reasonably large samples our asymptotic approximation is excellent for roots of 0.8 or larger. An immediate implication is that impulse response inference is likely to be more reliable at long horizons when persistent autoregressive processes are expressed in levels rather than in differences.

2. Proposition 5 may be generalized to vector autoregressive processes, as long as there is only one large root, as is commonly assumed in related studies (see, e.g., Pesavento and Rossi 2006; Mikusheva 2012). For a potential alternative approach that allows for multiple large roots see Phillips and Lee (2016) and the references therein.

3. In related work, Mikusheva (2012) proposes a generalization of the grid bootstrap of Hansen (1999) for autoregressions that allows inference on individual impulse responses that, like our approach, is uniformly asymptotically valid in the parameter space. The advantage of Mikusheva's asymptotic approximation is that it nests as special cases the conventional normal approximation for short-horizon impulse response estimators and the nonstandard asymptotic approximation for long-horizon impulse response estimators, as proposed by Phillips (1998), Wright (2000), Gospodinov (2004) and Pesavento and Rossi (2006). The disadvantage of Mikusheva's procedure is that its computational cost tends to be prohibitive for all but the simplest autoregressive processes.[12] Our approach provides a computationally less costly alternative to Mikusheva's grid bootstrap in many applied settings for short as well as long horizons. For example, at horizons up to 12 periods, even for $T = 240$, the infimum of the impulse response coverage rates based on our conventional bootstrap asymptotics for lag-augmented autoregressive models ranges from 87% to 89%. For $T = 480$, the coverage rates reach 89% to 90%, depending on the horizon. The latter coverage rates are at least as accurate as the grid bootstrap coverage rates for $T = 500$ reported in Mikusheva (2012), which range from 87% to 92% at similar horizons.

4. Unlike Efron's percentile interval, Hall's percentile interval and percentile-t intervals are not transformation respecting and hence are invalid at long horizons under Assumption C.

It is useful to illustrate these results by simulation. Table 5 restricts the impulse response horizon to a fraction $\lambda \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ of the sample size T. All methods of inference are implemented exactly as in section 5. We focus on impulse response inference based on the lag-augmented autoregression, since impulse response inference based on the original autoregression fails at long horizons, regardless of how the interval is constructed. This fact follows from Phillips (1998) and may easily be verified by simulation. The more interesting question is how well our asymptotic approximation works for the lag-augmented model. As predicted by our theoretical

[12] The largest process considered by Mikusheva (2012) in her simulation analysis is an AR(2) process. Her method does not appear to have been applied to autoregressions with more lags.

analysis, conventional delta method inference breaks down at long horizons, even when working with the lag-augmented model. Even for $T = 600$, the coverage rates of the nominal 90% delta method interval remain between 49% and 68%, depending on λ. Similarly, the coverage rates of the Hall percentile interval range from 43% to 52% for $T = 600$, illustrating the failure of this method. Likewise, equal-tailed and symmetric percentile-t impulse response intervals show no tendency to converge to their nominal probability content (results not shown to conserve space).

In contrast, even without bias adjustments, the coverage rates of Efron's percentile interval range from 85% for $T = 240$ to 89% for $T = 600$. Applying bias adjustments for the slope parameters, further increases the finite-sample accuracy. The coverage accuracy increases to about 89% for $T = 240$ and 90% for $T = 600$, regardless of λ. We conclude that the Efron percentile interval based on the bias-adjusted lag-augmented autoregression which performed best in section 5 is also the preferred approach in the current setting. It should be noted that the type of local power analysis we conducted in section 5.2 is infeasible in the context of long-horizon inference when one of the roots is local to unity because inference based on the non-augmented model is asymptotically invalid.

7 Extensions to ARMA Processes

The practice of augmenting the AR(p) model by one autoregressive lag, resulting in an AR($p + 1$) model, of which only the first p slope coefficients are used for inference, presumes that p is known. In section 2, p was defined as the true lag order, assuming that this lag order is finite. It should be noted that all our results would go through if p exceeded the true lag order. The lag order p could, alternatively, be defined as the approximating lag order for an AR(∞) process (e.g., Inoue and Kilian 2002b). While a theoretical analysis of this setting is beyond the scope of this paper, in this section we present some preliminary simulation evidence based on ARMA(1,1)-DGPs with MA roots of +0.25 and -0.25, respectively. Table 6 illustrates that the uniform coverage of our preferred method based on the lag augmented AR(6) model is nearly perfect in this case at all horizons from 1 to 60. The coverage rates range from 0.88 to 0.91 for $T = 600$, matching the results for purely autoregressive processes. Only slightly less accurate results hold for lag augmented AR(4) and AR(8) models, suggesting that the autoregressive approximation continues to work well for reasonably large lag orders.

It may seem that the conventional lag augmentation approach could be further improved upon by augmenting the original autoregression by one moving average lag instead of one autoregressive lag. This is not the case. First, the resulting ARMA(p,1) model would not be more robust against lag order misspecification because – after discarding the augmented MA coefficient – the implied AR(p) model is only as good as the maintained AR(p) specification. Second, applied users for good reason tend to avoid the estimation of higher-dimensional ARMA models because ARMA estimates tend to be unreliable in small samples. Third, and most importantly, it can be shown that this MA lag augmentation procedure is asymptotically invalid. Consider the ARMA(1,1) model:

$$y_t = \phi y_{t-1} + u_t - \theta u_{t-1}$$

where $u_t \overset{iid}{\sim} (0,1)$. Closed-form results are facilitated by focusing on the IV estimator. When this process is stationary, ϕ and θ satisfy $\gamma_1 = \phi\gamma_0 - \theta$ and $\gamma_2 = \phi\gamma_1$, where $\gamma_j \equiv E(y_t y_{t-j})$ for $j = 0, 1, 2$. The implied IV estimator of ϕ and θ is $\widehat{\phi} = \widehat{\gamma}_2/\widehat{\gamma}_1$ and $\widehat{\theta} = (\widehat{\gamma}_2/\widehat{\gamma}_1)\widehat{\gamma}_0 - \widehat{\gamma}_1$, where $\widehat{\gamma}_j = (1/T)\sum_{t=j+1}^{T} y_t y_{t-j}$ for $j = 0, 1, 2$. When $\phi = 1$, it can be shown that $\widehat{\phi}$ and $\widehat{\theta}$ remain consistent for ϕ and θ. Hall (1989) shows that $\widehat{\phi}$ converges weakly to a Dickey-Fuller distribution. Because the impulse responses, $\widehat{\phi}^h$, are not asymptotically normally distributed, conventional methods of inference based on the asymptotic normal approximation fail.

To illustrate this point, we conduct a small-scale Monte Carlo experiment. The data are generated from the Gaussian random walk process:

$$y_t = y_{t-1} + u_t,$$

where $y_0 = 0$, $u_t \overset{iid}{\sim} N(0,1)$ and $t = 1, 2, ..., T$ with $T = 600$. AR(1), AR(2) and ARMA(1,1) models are fitted and the first twelve impulse responses are calculated using the estimate of the AR(1) coefficient parameter of these models. Table A3 reports the coverage rates for nominal 90% impulse response confidence intervals based on the delta method. As expected, the delta method fails for the AR(1) model, as does the delta method for the AR(1) model augmented by one MA lag. The effective coverage rates are near 50%. In contrast, the delta method produces coverage rates between 85% and 89% in the AR(1) model augmented by one autoregressive lag. Even more accurate results would be obtained by suitable bootstrap methods.

8 Concluding Remarks

Although impulse response inference has played an important role in macroeconometrics since the 1980s, all existing proofs of the asymptotic validity of conventional delta method and bootstrap confidence intervals are based on pointwise Gaussian asymptotic approximations. We established the uniform asymptotic validity of conventional asymptotic and bootstrap inference about individual impulse responses and vectors of impulse responses when the horizon is fixed with respect to the sample size. We showed that for inference about vectors of impulse responses based on Wald test statistics to be uniformly valid in the parameter space, autoregressions must be lag augmented. Inference about individual impulse responses, in contrast, under weak conditions is uniformly valid even without lag augmentation, provided the model includes more than one lag.

We documented by simulation that the conventional asymptotic approximation works well in moderately large samples, as long as the impulse response horizon remains reasonably small relative to the sample size. The highest small-sample accuracy is achieved when bootstrapping the lag-augmented autoregressive model using the bias-adjusted bootstrap method of Kilian (1999). We provided formal asymptotic arguments why this method of inference retains its accuracy even at very long impulse response horizons, when other methods do not. Although the latter argument does not hold uniformly across the parameter space, it does hold in a local-to-unity setting. Our results suggest that highly persistent autoregressions in applied work should be routinely lag-augmented when conducting impulse response analysis, if coverage accuracy is the primary objective.

Appendix A: Proof of Propositions 1, 2, 3, 4 and 5

To prove Proposition 1, we follow the steps taken in Mikusheva (2007a,b). First we show that the estimation uncertainty about the asymptotic covariance matrix is asymptotically negligible (Lemma B1). Next, we show that the distribution of the least-squares estimator can be uniformly approximated by that based on Gaussian autoregressive processes (Lemma B2). Third, we show that the latter can be uniformly approximated by the local-to-unity asymptotic distribution (Lemma B3). Proposition 2 follows from Proposition 1 and the rank condition in Assumption B.

As in Mikusheva (2007a, 2012), we split the parameter space into two overlapping parts:

$$\begin{aligned} \mathcal{A}_T &= \{\theta \in \Theta : |1-\rho_p| < T^{1-\alpha}\}, \\ \mathcal{B}_T &= \{\theta \in \Theta : |1-\rho_p| > T^{1-\alpha}\}, \end{aligned}$$

for some $0 < \alpha < 1$.

Proof of Proposition 1. First, it follows from Lemma B1 that

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} \sup_{x\in\Re^{d_\theta}} \left| P(\widehat{\Sigma}_T^{-\frac{1}{2}}(\widehat{\theta}_T - \theta) \le x) - P(\Sigma_T^{-\frac{1}{2}}(\widehat{\theta}_T - \theta) \le x) \right| = 0, \tag{A.1}$$

where

$$\widehat{\Sigma}_T = \begin{bmatrix} \widehat{\sigma}_T^2(\sum_{t=p+1}^T x_t x_t')^{-1} & 0_{(p+2)\times 1} \\ 0_{1\times(p+2)} & \widehat{\sigma}_{4,T} \end{bmatrix}, \quad \Sigma_T = \begin{bmatrix} \sigma^2(\sum_{t=p+1}^T x_t x_t')^{-1} & 0_{(p+2)\times 1} \\ 0_{1\times(p+2)} & \sigma_4 \end{bmatrix}. \tag{A.2}$$

Next, it follows from Lemma B2 that the distribution of $\Sigma_T^{-\frac{1}{2}}(\widehat{\theta}_T - \theta)$ based on $\{y_t\}_{t=1}^T$ can be uniformly approximated by that based on $\{\bar{y}_t\}_{t=1}^T$ where

$$\begin{aligned} \Delta\bar{y}_t &= \delta_0 + \delta_1\left(\frac{t}{T}\right) + \pi\bar{y}_{t-1} + \gamma_1\Delta\bar{y}_{t-1} + \cdots + \gamma_{p-1}\Delta\bar{y}_{t-p+1} + \bar{u}_t \\ &= \beta\bar{x}_t + \bar{u}_t, \end{aligned} \tag{A.3}$$

and $\bar{u}_t \overset{iid}{\sim} N(0,\sigma^2)$. It follows from Lemma 5 of Mikusheva (2007a) and Lemma B3 that

$$\begin{aligned}
&\Upsilon_T^{-1}(c_T)\sum_{t=p+1}^{T}\bar{x}_t\bar{x}_t'\Upsilon_T^{-1}(c_T)-\Upsilon^{-1}(c_T)\begin{bmatrix} 1 & \frac{1}{2} & \omega(\rho_p)\int_0^1 J_{c_T}(r)dr & 0\\ \frac{1}{2} & \frac{1}{3} & \omega(\rho_p)\int_0^1 rJ_{c_T}(r)dr & 0\\ \omega(\rho_p)\int_0^1 J_{c_T}(r)dr & \omega(\rho_p)\int_0^1 rJ_{c_T}(r)dr & \omega(\rho_p)^2\int_0^1 J_{c_T}(r)^2dr & 0\\ 0 & 0 & 0 & M\end{bmatrix}\Upsilon^{-1}(c_T)\\
= \;& o_p(1) \qquad\qquad (A.4)
\end{aligned}$$

and

$$\Upsilon_T^{-1}(c_T)\sum_{t=p+1}^{T}\bar{x}_t\bar{u}_t-\Upsilon^{-1}(c_T)\begin{bmatrix} N_1\\ N_2\\ \sigma\omega(\rho_p)\int_0^1 J_{c_T}(r)dW(r)\\ N_3\end{bmatrix} = o_p(1), \qquad (A.5)$$

uniformly over $\mathcal{A}_T$, where $c_T = T\log(\max(|\rho_p|,\epsilon))$ and $\Upsilon_T(c_T)=T^{\frac{1}{2}}\Upsilon(c_T)$.

Third, it follows from Lemma 12(a) and (b) of Mikusheva (2007b), (A.4) and (A.5) that

$$\begin{aligned}
T^{\frac{1}{2}}(\widehat{\sigma}_T^2-\sigma^2) &= \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}(\widehat{u}_t^2-\sigma^2)\\
&= \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}((u_t-(\widehat{\beta}_T-\beta)'x_t)^2-\sigma^2)\\
&= \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}(u_t^2-\sigma^2-2(\widehat{\beta}_T-\beta)'x_tu_t+(\widehat{\beta}_T-\beta)'x_tx_t'(\widehat{\beta}_T-\beta))\\
&= \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}(u_t^2-\sigma^2)-2T^{-\frac{1}{2}}(\widehat{\beta}_T-\beta)'\Upsilon_T(c_T)\Upsilon_T^{-1}(c_T)\sum_{t=p+1}^{T}x_tu_t\\
&\quad +T^{-\frac{1}{2}}(\widehat{\beta}_T-\beta)'\Upsilon_T(c)\Upsilon_T^{-1}(c_T)\sum_{t=p+1}^{T}x_tx_t'\Upsilon_T^{-1}(c_T)\Upsilon_T(c_T)(\widehat{\beta}_T-\beta)\\
&= \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}(u_t^2-\sigma^2)+O_p(T^{-\frac{1}{2}})\\
&\overset{d}{\to} N_4, \qquad\qquad (A.6)
\end{aligned}$$

uniformly on Θ.

It follows from (A.4), (A.5) and (A.6) that

$$\lim_{T\to\infty}\sup_{\theta\in\mathcal{A}_T}\sup_{x\in\Re^{d_\theta}}\left|P(\Sigma_T^{-\frac{1}{2}}(\widehat{\theta}_T-\theta)\le x)-P(\eta(\theta,T)\le x)\right| = 0. \qquad (A.7)$$

Using

$$\sqrt{-2c_T}\int_0^1 J_{c_T}(r)dW(r) \quad \overset{d}{\to} \quad N(0,1), \tag{A.8}$$

$$(-2c_T)\int_0^1 J_{c_T}^2(r)dx \quad \overset{p}{\to} \quad 1, \tag{A.9}$$

as $c_T \to -\infty$ (Phillips, 1987), it follows from Lemma 12(a) and (b) of Mikusheva (2007b) and (A.6) that

$$\lim_{T\to\infty}\sup_{\theta\in\mathcal{B}_T}\sup_{x\in\Re^{d_\theta}}\left|P(\Sigma_T^{-\frac{1}{2}}(\widehat{\theta}_T-\theta)\le x)-P(\eta(\theta,T)\le x)\right| \;=\; 0. \tag{A.10}$$

Therefore, Proposition 1 follows from (A.7) and (A.10). ■

Proof of Proposition 2. Because $f(\cdot)$ is continuously differentiable,

$$\begin{aligned} T^{\frac{1}{2}}(f(\widehat{\theta}_T)-f(\theta)) &= T^{\frac{1}{2}}Df(\bar{\theta}_T)(\widehat{\theta}_T-\theta) \\ &= T^{\frac{1}{2}}Df(\bar{\theta}_T)\Sigma_T^{\frac{1}{2}}\Sigma_T^{-\frac{1}{2}}(\widehat{\theta}_T-\theta) \\ &= T^{\frac{1}{2}}Df(\bar{\theta}_T)\Upsilon_T^{-1}(c_T)(\Upsilon_T(c_T)\Sigma_T\Upsilon_T(c_T))^{\frac{1}{2}} \\ &\quad \times\Sigma_T^{-\frac{1}{2}}(\widehat{\theta}_T-\theta), \end{aligned} \tag{A.11}$$

where $\bar{\theta}_T$ is a point between $\widehat{\theta}_T$ and θ.

Note that $T^{\frac{1}{2}}Df(\theta)\Upsilon_T(c)$ equals (8), that the first two columns consist of zeros by Assumption B, and that in the nonstationary region $\mathcal{A}_T$, the elements of the third column converge to zero. Thus, it follows from (A.4), (A.5) and (A.11) that

$$(Df(\widehat{\theta}_T)\widehat{\Sigma}_T Df(\widehat{\theta}_T)')^{-\frac{1}{2}}T^{\frac{1}{2}}(f(\widehat{\theta}_T)-f(\theta)) \tag{A.12}$$

converges in distribution to the standard normal random vector uniformly on $\mathcal{A}_T$. In the stationary region $\mathcal{B}_T$, it follows from Lemma 12(a) and (b) of Mikusheva (2007b) and (A.6) that (A.12) converges in distribution to the standard normal random vector uniformly on $\mathcal{B}_T$. Thus, (11) follows. (12) follows from the second remark about Theorem 1 in Kasy (2019). ■

Proof of Proposition 3. Because we assume that the variance is uniformly bounded away from zero and uniformly bounded from above in Assumption A(i), the arguments in the proof of Lemma 6 of Mikusheva (2007a) carry through after scaling the residual in her proof by its standard deviation.

Thus, the empirical distribution of the scaled residuals belongs to the $\mathcal{L}_r(K, M, \theta)$ class[13], and the Skorohod representation result in Lemma 12 of Mikusheva (2007a) applies. In other words, for any realization of the disturbance term, there exists $K > 0$, $M > 0$ and θ such that the empirical distribution function of the residual-based bootstrap, $\widehat{F}_T$, belongs to $\mathcal{L}_r(K, M, \theta)$ for all $\theta \in \Theta$. Thus, there is an almost sure approximation of the partial sum process by Brownian motions: For any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\lim_{T\to\infty} \sup_{F_T \in \mathcal{L}_r(K,M,\theta)} P^* \left(\sup_{0\leq s\leq 1} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} u_t^* - \sigma W(s)| > \varepsilon T^{-\delta} \right| \right) = 0. \tag{A.13}$$

Moreover, by Lemma B2 with c_T, u_t, x_t, y_t replaced by $c_T^*, u_t^*, x_t^*, y_t^*$, respectively, the relevant bootstrap sample moments can be approximated by those generated from a Gaussian autoregressive process with $\beta = \widehat{\beta}_T$ almost surely conditional on the data. A bootstrap version of Lemma B1 may be constructed by replacing $\widehat{\sigma}_{4,T}$ and $\sigma_{4,T}$ by $\widehat{\sigma}_{4,T}^*$ and $\widehat{\sigma}_{4,T}$, respectively. Repeating the arguments in the proof of Lemma 5 of Mikusheva (2007a) yields a bootstrap version of Lemma B3 in which c_T is replaced by c_T^* from which we obtain the desired result. ■

Proof of Proposition 4. The proof of Proposition 4 is analogous to that of Proposition 2. ■

Proof of Proposition 5. The AR(p) model is augmented by one lag. Let $\widehat{F}$ and $\widehat{F}^*$ denote the companion matrices for the first p coefficients of the estimated lag-augmented model and its bootstrap analogue,

$$\widehat{F} = \begin{bmatrix} \widehat{\phi}_{1,T} & \widehat{\phi}_{2,T} & \cdots & \widehat{\phi}_{p-1,T} & \widehat{\phi}_{p,T} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \widehat{F}^* = \begin{bmatrix} \widehat{\phi}_{1,T}^* & \widehat{\phi}_{2,T}^* & \cdots & \widehat{\phi}_{p-1,T}^* & \widehat{\phi}_{p,T}^* \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

respectively. The h-step-ahead impulse response estimate and bootstrap estimate are the (1,1)

[13] Mikusheva (2007a) defines this class to be the class of sequences of distributions F_T such that the mean is zero, the variance σ_T^2 satisfies $|\sigma_T^2| \leq MT^{-\theta}$ and the supremum of the rth moment with respect to T is less than K.

elements of $\widehat{F}^h$ and $\widehat{F}^{*h}$, respectively.

F can be written as

$$F = PJP^{-1}, \tag{A.14}$$

where J is the Jordan normal form of F, and P consists of eigenvectors and generalized eigenvectors of P. Thus

$$F^h = PJ^hP^{-1} \tag{A.15}$$

The hth power of the Jordan normal form is given by

$$J^h = \begin{bmatrix} \rho_p^h & 0 & \cdots & 0 \\ 0 & J_{m_2}^h(\rho_{p-1}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{m_q}^h(\rho_1) \end{bmatrix}, \tag{A.16}$$

where m_j is the multiplicity of the jth largest root with $m_1 = 1$ such that $\sum_{j=1}^q m_j = p$, and

$$J_{m_j}^h(\rho_k) = \begin{bmatrix} \rho_k^h & \binom{h}{1}\rho_k^{h-1} & \binom{h}{2}\rho_k^{h-2} & \cdots & \binom{h}{m_j-1}\rho_k^{h-m_j+1} \\ 0 & \rho_k^h & \binom{h}{1}\rho_k^{h-1} & \cdots & \binom{h}{m_j-2}\rho_k^{h-m_j+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \rho_k^h \end{bmatrix} \tag{A.17}$$

for $j = 1, 2, ..., q$. Because there is one and only one local-to-unity root (ρ_p), the (1,1) element of $F^{[\lambda T]}$ can be approximated by

$$P_{11}\rho_p^{[\lambda T]}P^{11} + o(1), \tag{A.18}$$

where P_{11} is the (1,1) element of P and P^{11} is the (1,1) element of P^{-1}. Similarly, the (1,1) element of $\widehat{F}^{[\lambda T]}$ and that of $\widehat{F}^{*[\lambda T]}$ can be approximated by

$$\widehat{P}_{11}\widehat{\rho}_p^{[\lambda T]}\widehat{P}^{11} + o_p(1), \tag{A.19}$$

$$\widehat{P}_{11}^*\widehat{\rho}_p^{*[\lambda T]}\widehat{P}^{11*} + o_p^*(1), \tag{A.20}$$

respectively. Taking the log on both sides

$$\log(\widehat{P}_{11}) + \log(\widehat{P}^{11}) + [\lambda T] \log(\widehat{\rho}_p) + o_p(1), \tag{A.21}$$

$$\log(\widehat{P}^*_{11}) + \log(\widehat{P}^{11*}) + [\lambda T] \log(\widehat{\rho}^*_p) + o^*_p(1). \tag{A.22}$$

Because there is one and only one local-to-unity root (ρ_p), $\widehat{\rho}_p$ and $\widehat{\rho}^*_p$ are continuously differentiable in $(\widehat{\phi}_{1,T}, ..., \widehat{\phi}_{p,T})$ and $(\widehat{\phi}^{*,T}_{1,T}, ..., \widehat{\phi}^*_{p,T})$, respectively. Thus, $\widehat{\rho}_{p,T}$ is asymptotically normally distributed in the lag-augmented model. Because Efron's percentile bootstrap interval is transformation-respecting and because $\widehat{\rho}_{p,T}$ is asymptotically normally distributed, it remains asymptotically valid, when other intervals fail. Since first-order mean bias adjustments of the slope parameters are of order T, this argument remains valid when using Efron's interval in conjunction with bias adjustments (see Kilian 1998b). ■

Appendix B: Proofs of Lemmas B1, B2, B3 and B4

Throughout Appendix B, suppose that Assumptions A and B are satisfied.

The following lemma builds on Lemma 3 of Mikusheva (2007a):

Lemma B1.

$$\widehat{\sigma}_{4,T} = \sigma_4 + o_p(1) \tag{B.1}$$

uniformly on Θ.

Proof of Lemma B1. $\widehat{\sigma}_{4,T}$ can be approximated by

$$\begin{aligned}
\widehat{\sigma}_{4,T} &= \frac{1}{T-p} \sum_{t=p+1}^{T} \widehat{u}_t^4 - \widehat{\sigma}_T^4 \\
&= \frac{1}{T-p} \sum_{t=p+1}^{T} \widehat{u}_t^4 - \sigma^4 + o_p(1) \\
&= \frac{1}{T-p} \sum_{t=p+1}^{T} (u_t - (\widehat{\beta}_T - \beta)' x_t)^4 - \sigma^4 + o_p(1)
\end{aligned}$$

$$= \frac{1}{T-p}\left[\sum_{t=p+1}^{T} u_t^4 - 4((\widehat{\beta}_T-\beta)'x_t)u_t^3 + 6((\widehat{\beta}_T-\beta)'x_t)^2u_t^2 - 4((\widehat{\beta}_T-\beta)'x_t)^3u_t + ((\widehat{\beta}_T-\beta)'x_t)^4\right]$$

$$= \frac{1}{T-p}\left[\sum_{t=p+1}^{T} u_t^4 - 4(\zeta_T'\tilde{x}_t)u_t^3 + 6(\zeta_T'\tilde{x}_t)^2u_t^2 - 4(\zeta_T'\tilde{x}_t)^3u_t + (\zeta_T'x_t)^4\right] \tag{B.2}$$

where the second equality follows from Lemma 3 of Mikusheva (2007a), $\zeta_T = (\sum_{t=p+1}^T x_t x_t')^{\frac{1}{2}}(\widehat{\beta}_T - \beta)$ and $\tilde{x}_t = (\sum_{t=p+1}^T x_t x_t')^{-\frac{1}{2}}x_t$. As shown in the proof of Proposition 1, $\zeta_T = O_p(1)$ uniformly on Θ. Because $\sum_{t=p+1}^T \tilde{x}_t\tilde{x}_t' = I_{p+2}$,

$$\sum_{t=p+1}^{T}\sum_{j=1}^{p+2}\tilde{x}_t^k \;\leq\; 1, \tag{B.3}$$

for $k = 4, 6, 8$. Thus, it follows from the Cauchy-Schwarz inequality that

$$\left|\frac{1}{T-p}\sum_{t=p+1}^{T}(\zeta_T'\tilde{x}_t)u_t^3\right| \leq \left(\frac{1}{T-p}\sum_{t=p+1}^{T}(\zeta_T'\tilde{x}_t\tilde{x}_t'\zeta_T)\right)^{\frac{1}{2}}\left(\frac{1}{T-p}\sum_{t=p+1}^{T}u_t^6\right)^{\frac{1}{2}} = O_p(T^{-\frac{1}{2}}), \tag{B.4}$$

$$\begin{aligned}\left|\frac{1}{T-p}\sum_{t=p+1}^{T}(\zeta_T'\tilde{x}_t)^2u_t^2\right| &\leq \left(\frac{1}{T-p}\sum_{t=p+1}^{T}(\zeta_T'\tilde{x}_t)^4\right)^{\frac{1}{2}}\left(\frac{1}{T-p}\sum_{t=p+1}^{T}u_t^4\right)^{\frac{1}{2}} \\ &\leq C\left(\frac{\|\zeta_T\|^4}{T-p}\sum_{t=p+1}^{T}\sum_{j=1}^{p+2}\tilde{x}_{j,t}^4\right)^{\frac{1}{2}}\left(\frac{1}{T-p}\sum_{t=p+1}^{T}u_t^4\right)^{\frac{1}{2}} \\ &= O_p(T^{-\frac{1}{2}}),\end{aligned} \tag{B.5}$$

$$\left|\frac{1}{T-p}\sum_{t=p+1}^{T}(\zeta_T'\tilde{x}_t)^3u_t\right| = O_p(T^{-\frac{1}{2}}), \tag{B.6}$$

$$\left|\frac{1}{T-p}\sum_{t=p+1}^{T}(\zeta_T'\tilde{x}_t)^4\right| = O_p(T^{-\frac{1}{2}}), \tag{B.7}$$

where the last two results follow from arguments similar to the one used in the second result and the $O_p(T^{-\frac{1}{2}})$ terms are uniformly on Θ. Thus, Lemma B1 follows. ■

The next lemma is a slight extension of Lemma 11 of Mikusheva (2007b) which we present for completeness.

Lemma B2. Suppose that $\bar{y}_t$ follows

$$\begin{aligned}
\Delta\bar{y}_t &= \delta_0 + \delta_1(t/T) + \pi\bar{y}_{t-1} + \gamma_1\Delta\bar{y}_{t-1} + \cdots + \gamma_{p-1}\Delta\bar{y}_{t-p+1} + \bar{u}_t \\
&= \beta\bar{x}_t + \bar{u}_t, \qquad \text{(B.8)}
\end{aligned}$$

where $\bar{u}_t \overset{iid}{\sim} N(0,\sigma^2)$. Then thre exists a completion of the initial probability space and the realization of $\bar{y}_t$ on this probability space such that

$$\sup_{\theta\in\mathcal{A}_T}\sup_{t=1,\ldots,T}\left\|\frac{y_t}{\sqrt{T}} - \frac{\bar{y}_t}{\sqrt{T}}\right\| = o(1) \text{ a.s.,} \qquad \text{(B.9)}$$

$$\sup_{\theta\in\mathcal{A}_T}\sup_{t=1,\ldots,T}\left\|\frac{y_t}{\sqrt{T}}\right\| = O(1) \text{ a.s.,} \qquad \text{(B.10)}$$

$$\sup_{\theta\in\mathcal{A}_T}\left\|\Upsilon_T^{-1}(c)\sum_{t=p+1}^{T} x_t u_t - \Upsilon_T^{-1}(c)\sum_{t=p+1}^{T}\bar{x}_t\bar{u}_t\right\| = o(1), \qquad \text{(B.11)}$$

$$\sup_{\theta\in\mathcal{A}_T}\left\|\Upsilon_T^{-1}(c)\sum_{t=p+1}^{T} x_t x_t'\Upsilon_T^{-1}(c) - \Upsilon_T^{-1}(c)\sum_{t=p+1}^{T}\bar{x}_t\bar{x}_t'\Upsilon_T^{-1}(c)\right\| = o(1), \qquad \text{(B.12)}$$

$$\sup_{\theta\in\mathcal{A}_T}\left\|(\sum_{t=p+1}^{T} x_t x_t')^{-\frac{1}{2}}\sum_{t=p+1}^{T} x_t u_t' - (\sum_{t=p+1}^{T}\bar{x}_t\bar{x}_t')^{-\frac{1}{2}}\sum_{t=p+1}^{T}\bar{x}_t\bar{u}_t'\right\| = o(1). \qquad \text{(B.13)}$$

Proof of Lemma B2.

Because $\kappa \leq \sigma^2 \leq K$, (B.9) and (B.10) follow from Lemma 11(a) and (b), respectively, of Mikusheva (2007b) who normalizes σ^2 to one. Similarly, (B.11) follows from Lemma 11(c), (d), (e), and (f) of Mikusheva (2007b) and (B.12) from her Lemma 11(g), (h), and (i). (B.13) follows from (B.11) and (B.12). ■

The next two results follow from the arguments used in the proof of Lemma 5 of Mikusheva

(2007a).

Lemma B3.

$$
\begin{aligned}
&\lim_{T\to\infty}\sup_{\theta\in\mathcal{A}_T} E\left\|\text{vech}\left(\Upsilon_T(c_T)^{-1}\sum_{t=p+1}^{T}\bar{x}_t\bar{x}_t\Upsilon_T(c_T)^{-1}\right.\right.\\
&\left.\left.-\Upsilon^{-1}(c_T)\begin{bmatrix} 1 & \frac{1}{2} & \omega(\rho_p)\int_0^1 J_{c_T}(r)dr & 0\\ \frac{1}{2} & \frac{1}{3} & \omega(\rho_p)\int_0^1 rJ_{c_T}(r)dr & 0\\ \omega(\rho_p)\int_0^1 J_{c_T}(r)dr & \omega(\rho_p)\int_0^1 rJ_{c_T}(r)dr & \omega(\rho_p)^2\int_0^1 J_{c_T}(r)^2dr & 0\\ 0 & 0 & 0 & M\end{bmatrix}\Upsilon^{-1}(c_T)\right)\right\|^2\\
&= 0, \qquad \text{(B.14)}
\end{aligned}
$$

$$
\lim_{T\to\infty}\sup_{\theta\in\mathcal{A}_T} E\left\|\Upsilon_T(c_T)^{-1}\sum_{t=p+1}^{T}\bar{x}_t\bar{u}_t-\Upsilon^{-1}(c_T)\begin{bmatrix} N_1\\ N_2\\ \sigma\omega(\rho_p)\int_0^1 J_{c_T}(r)dW(r)\\ N_3\end{bmatrix}\right\|^2 = 0. \qquad \text{(B.15)}
$$

The following result is a slight modification of equation (21) of Inoue and Kilian (2002a):

Lemma B4.

Let $\delta_{0,T}=\phi_T(1)\delta_{0,T}^{\dagger}+\delta_{1,T}^{\dagger}(\phi_{T,1}+2\phi_{T,2}+\cdots+p\phi_{T,p})$, $\delta_{0,T}^{\dagger}=\delta_0^{\dagger}+\xi_cT^{-1/2}+o(T^{-1/2})$, $\delta_{1,T}=\phi_T(1)\delta_{1,T}^{\dagger}$, $\delta_{1,T}^{\dagger}=\delta_1^{\dagger}+\xi_dT^{-1/2}+o(T^{-1/2})$, $\phi_T(L)=1-\phi_{T,1}L-\cdots-\phi_{T,p}L^p$, $\pi_T=-\phi_T(1)=\sum_{j=1}^{p}\phi_{T,j}-1=\xi_0T^{-1}+o(T^{-1})$, $\gamma_{T,j}=-(\phi_{T,j+1}+\cdots+\phi_{T,p}) = \gamma_j+\xi_jT^{-1/2}+o(T^{-1/2})$, $u_{T,t}\overset{iid}{\sim}(0_{n\times 1},\sigma_T^2)$, and $\sigma_T^2=\sigma^2+\xi_{\sigma^2}T^{-1/2}+o(T^{-1/2})$ for some $[\xi_c\ \xi_d\ \xi_0,\xi_1\ \cdots\ \xi_{p-1}\ \xi_{\sigma^2}]'w$. Define a triangular array

$$
\begin{aligned}
\Delta y_{T,t} &= \delta_{0,T}+\delta_{1,T}\left(\frac{t}{T}\right)+\pi_Ty_{T,t-1}+\gamma_{T,1}\Delta y_{T,t-1}+\cdots+\gamma_{T,p-1}\Delta y_{T,t-p+1}+u_{T,t}\\
&= \beta_T'x_{T,t}+u_{T,t}, \qquad \text{(B.16)}
\end{aligned}
$$

where $\beta_T=[\delta_{0,T}\ \delta_{1,T}\ \pi_T\ \gamma_{T,1}\ \cdots\ \gamma_{T,p}]'$, and $x_t=[1\ t/T\ \ y_{t-1}\ \Delta y_{t-1}\ \cdots\ \Delta y_{t-p+1}]'$. Then

$$
\lim_{T\to\infty}\sup_{\theta\in\Theta}\sup_{x\in\Re^{d_\theta}}\left|P(\Upsilon_T(c_T)(\widehat{\theta}_T-\theta)\le x)-P(\eta_T(\theta,T)\le x)\right| = 0, \qquad \text{(B.17)}
$$

where $c_T = T \log(1 + \rho_{p,T})$.

The proof of Lemma B4 is based on the same arguments already used in the proof of Proposition 1 and is omitted. ■

Appendix C: Data Mining Correction

Let the uniform coverage rate of the $1-\alpha$ level confidence set C_i be

$$\min_{i \in \{1,2,\ldots,d\}} P_{\rho_i}(\psi_i \in C_i), \tag{C.1}$$

where ψ_i is the true parameter value and ρ_i is the ith value of ρ.

Because $P_{\rho_i}(\psi_i \in C_i)$ is not analytically tractable in finite samples, it is approximated by simulation:

$$\widehat{P}_{\rho_i}(\psi_i \in C_i) = \frac{1}{M} \sum_{j=1}^{M} I(\psi_i \in C_i^{(j)}) \tag{C.2}$$

where M is the number of Monte Carlo simulations and $C_i^{(j)}$ is a level $1-\alpha$ confidence set for the jth Monte Carlo iteration.

The problem is that the estimate of (C.1) is downward biased due to "data mining". To estimate this bias, express the Monte Carlo estimate of the coverage rate as:

$$X_i = \frac{1}{M} \sum_{j=1}^{M} d_{ij}, \tag{C.3}$$

where d_{ij} is an iid Bernoulli random variable with parameter $1-\alpha$ for $i = 1, ..., d$. Thus, MX_i is a binomial random variable with parameters M and $1-\alpha$. Then the expectation of

$$Y = \min_{i \in \{1,2,\ldots,d\}} X_i \tag{C.4}$$

minus $(1-\alpha)$ is the data mining bias. Thus, the expectation of Y can be estimated from

$$\frac{1}{M} \sum_{j=1}^{M} \min_{i \in \{1,2,\ldots,d\}} X_i^{(j)} \tag{C.5}$$

When $\alpha = 0.1$, $d = 10$, and $M = 5000$, as in our simulation study, the normal approximation yields 0.894, implying a data mining bias of 0.0066. The same answer is obtained when simulating this bias rather than relying on the normal approximation.

Appendix D: How to Bootstrap Lag-Augmented AR Models

This appendix summarizes the bootstrap algorithms employed in generating the simulation results in Tables 1 through 5. First, consider the standard bootstrap algorithm for autoregressions without bias adjustments. We approximate the data generating process

$$y_t = \delta_0 + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + u_t, \tag{D.1}$$

where u_t is iid white noise, by the bootstrap data generating process

$$y_t^* = \widehat{\delta}_{0,T} + \widehat{\phi}_{1,T} y_{t-1}^* + ... + \widehat{\phi}_{p,T} y_{t-p}^* + u_t^*. \tag{D.2}$$

The bootstrap data set $\{y_t^*\}_{t=1}^T$ is recursively generated from model (D.2), given a randomly chosen block of pre-sample observations and the least-squares estimates $\left\{\widehat{\delta}_{0,T}, \widehat{\phi}_{1,T}, ..., \widehat{\phi}_{p,T}\right\}$ of the parameters in model (D.1) from the observed sample $\{y_t\}_{t=1}^T$. The bootstrap innovation u_t^* is generated by drawing with replacement from the set of least-squares residuals $\{\widehat{u}_t\}_{t=1}^T$. Repeated application of this procedure allows us to generate a large number of bootstrap data sets of length T, each of which is evaluated by fitting an AR(p) model with intercept and calculating the implied impulse response coefficients. The empirical distribution of the bootstrap estimates of a given impulse response may then be used to construct either the Efron (1979) or the Hall (1992) percentile interval.

When using bias adjustments, as proposed in Kilian (1998b, 1999), the data generating process (1) is approximated by the bootstrap data generating process

$$y_t^* = \widehat{\delta}_{0,T} + \widehat{\phi}_{1,T}^{BC} y_{t-1}^* + ... + \widehat{\phi}_{p,T}^{BC} y_{t-p}^* + u_t^*. \tag{D.3}$$

where the least-squares slope parameter estimates have been replaced by first-order mean bias corrected least-squares estimates $\widehat{\phi}_T^{BC} = \left\{\widehat{\phi}_{1,T}^{BC}, ..., \widehat{\phi}_{p,T}^{BC}\right\}$. Let $\phi = \{\phi_1, ..., \phi_p\}$. Then, under some regularity conditions, $E(\widehat{\phi}_T - \phi) = b(\widehat{\phi}_T)/T + O(T^{-3/2})$, so $\widehat{\phi}_T^{BC} = \widehat{\phi}_T - b(\widehat{\phi}_T)/T$, where $b(\widehat{\phi}_T)/T$ may be estimated based on the closed-form solutions in Pope (1990). The motivation for this bias adjustment is that we want the bootstrap data generating process to be centered on parameter values that are not systematically different from their population values. We implement this procedure with all the refinements discussed in Kilian and Lütkepohl (2017). Repeated application of this procedure allows us to generate a large number of bootstrap data sets of length T, each of which is evaluated by fitting an AR(p) model with intercept and calculating the implied impulse response coefficients based on similarly bias-adjusted estimates of the slope parameters. The empirical distribution of the bootstrap estimates of a given impulse response then is used to construct the Efron (1979) percentile interval.

The same algorithms may be applied to the lag augmented model. The only difference is that we approximate

$$y_t = \delta_0 + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \phi_{p+1} y_{t-p-1} + u_t, \tag{D.4}$$

where $\phi_{p+1} = 0$, by

$$y_t^* = \widehat{\delta}_{0,T} + \widehat{\phi}_{1,T} y_{t-1}^* + ... + \widehat{\phi}_{p,T} y_{t-p}^* + \widehat{\phi}_{p+1,T} y_{t-p-1}^* + u_t^*. \tag{D.5}$$

where the slope parameter estimates in equation (D.5) may be adjusted for small-sample estimation bias or not. Each bootstrap data set of length T is evaluated by fitting the lag-augmented autoregressive model (and possibly applying bias adjustments to the bootstrap slope parameter estimates). Of course, as noted in section 2, for the construction of the bootstrap approximation of the impulse response distribution only the first p bootstrap slope parameter estimates are relevant, and $\widehat{\phi}_{p+1,T}^*$ must be discarded. The reason for bootstrapping the AR($p+1$) model (D.4) rather than the AR(p) model is that we are not testing the restriction on the $p+1$st lag. Rather our confidence sets are based on inverting Wald tests for nonlinear restrictions on the first p parameters of the AR($p+1$) model. Even though our Wald tests do not involve ϕ_{p+1}, it is necessary to include the $p+1$st lag to correctly mimic the joint asymptotic distributon of the estimator of the remaining parameters.

Appendix E: Supplementary Appendix

Tables A1 through A3 can be found online in the supplementary appendix at https://doi.org/10.1016/j.jeconom.201

Acknowledgments

The views expressed in the paper are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Dallas or the Federal Reserve System. We thank Xu Cheng, Nikolay Gospodinov, Helmut Lütkepohl, Mikkel Plagborg-Møller, Barbara Rossi, Katsumi Shimotsu, Xiaoqing Zhou, the editor, the associate editor, and two anonymous referees for helpful discussions. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Andrews, D.W.K., Guggenberger, P., 2009. Hybrid and size-corrected subsample methods. Econometrica 77, 721-762.

2. Bauer, D., Maynard, A., 2012. Persistence-robust surplus-lag Granger causality testing. Journal of Econometrics 169, 293-300.

3. Benkwitz, A., Lütkepohl, H., Neumann, M.H., 2000. Problems related to confidence intervals for impulse responses of autoregressive processes. Econometric Reviews 19, 69-103.

4. Bruder, S., Wolf, M., 2018. Balanced bootstrap joint confidence bands for structural impulse response functions. Journal of Time Series Analysis 39, 641-664.

5. Dolado, J., Lütkepohl, H., 1996. Making Wald tests work for cointegrated VAR systems. Econometric Reviews 15, 369-386.

6. Efron, B., 1979. Bootstrap Methods: Another Look at the Jackknife. Annals of Statistics 7, 1-26.

7. Elliott, G., 1998. On the robustness of cointegration methods when regressors almost have unit roots. Econometrica 66, 149-158.

8. Gafarov, B., Meier, M., Montiel Olea, J.L., 2018. Delta method inference for a class of set-identified SVARs. Journal of Econometrics 203, 316-327.

9. Giraitis, L., Phillips, P.C.B., 2006. Uniform limit theory for stationary autoregression. Journal of Time Series Analysis 27, 51-60.

10. Gonçalves, S., Kilian, L., 2004. Bootstrapping autoregressions in the presence of conditional heteroskedasticity of unknown form. Journal of Econometrics 123, 89-120.

11. Gospodinov, N., 2004. Asymptotic confidence intervals for impulse responses of near-integrated processes. Econometrics Journal 7, 505-527.

12. Gospodinov, N., Herrera, A.M., Pesavento, E., 2011. Unit roots, cointegration and pre-testing in VAR models. Advances in Econometrics 31, 81-115.

13. Granziera, E., Moon, H.R., Schorfheide, F., 2018. Inference for VARs identified with sign restrictions. Quantitative Economics 9, 1087-1121.

14. Guerron-Quintana, P., Inoue, A., Kilian, L., 2017. Impulse response matching estimators for DSGE models. Journal of Econometrics 196, 144-155.

15. Hall, A.R., 1989. Testing for a unit root in the presence of moving average errors. Biometrika 76, 49–56.

16. Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York.

17. Hansen, B.E., 1999. The grid bootstrap and the autoregressive model. Review of Economics and Statistics 81, 594-607.

18. Inoue, A., Kilian, L., 2002a. Bootstrapping autoregressive processes with possible unit roots. Econometrica 70, 377-391.

19. Inoue, A., Kilian, L., 2002b. Bootstrapping smooth functions of slope parameters and innovation variances in VAR(∞) models. International Economic Review 43, 309–332.

20. Inoue, A., Kilian, L., 2003. The continuity of the limit distribution in the parameter of interest is not essential for the validity of the bootstrap. Econometric Theory 19, 944-961.

21. Inoue, A., Kilian, L., 2016. Joint confidence sets for structural impulse responses. Journal of Econometrics 192, 421-432.

22. Jordà, Ò., 2009. Simultaneous confidence regions for impulse responses. Review of Economics and Statistics 91, 629-647.

23. Kasy, M., 2019. Uniformity and the delta method. Journal of Econometric Methods 8, 1-19.

24. Kilian, L., 1998a. Confidence intervals for impulse responses under departures from normality. Econometric Reviews 17, 1-29.

25. Kilian, L., 1998b. Small-sample confidence intervals for impulse response functions. Review of Economics and Statistics 80, 218-230.

26. Kilian, L., 1999. Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses. Review of Economics and Statistics 81, 652-660.

27. Kilian, L, Chang, P.L., 2000. How accurate are confidence intervals for impulse responses in large VAR models? Economics Letters 69, 299-307.

28. Kilian, L., Lütkepohl, H.L., 2017. Structural Vector Autoregressive Analysis. Cambridge University Press, New York.

29. Kurozumi, E., Yamamoto, T., 2000, Modified lag augmented vector autoregressions. Econometric Reviews 19, 207-231.

30. Lee, J.H., Liao, Z., 2018. On Standard inference for GMM with local identification failures of known forms. Econometric Theory 34, 790-814.

31. Lütkepohl, H., 1990. Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models. Review of Economics and Statistics 72, 116-125.

32. Lütkepohl, H., Staszewska-Bystrova, A., Winker, P., 2015a. Confidence bands for impulse responses: Bonferroni vs. Wald. Oxford Bulletin of Economics and Statistics 77, 800-821.

33. Lütkepohl, H., Staszewska-Bystrova, A., Winker, P., 2015b. Comparison of methods for constructing joint confidence bands for impulse response functions. International Journal of Forecasting 31, 782-798.

34. Lütkepohl, H., Staszewska-Bystrova, A., Winker, P., 2018. Calculating joint confidence bands for impulse response functions using highest density regions. Empirical Economics 55, 1389-1411.

35. Mikusheva, A., 2007a. Uniform inference in autoregressive models. Econometrica 75, 1411-1452.

36. Mikusheva, A., 2007b. Uniform inference in autoregressive models: Supplementary appendix. https://economics.mit.edu/files/9467.

37. Mikusheva, A., 2012. One-dimensional inference in autoregressive models with the potential presence of a unit root. Econometrica 80, 173-212.

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

39. Montiel Olea, J.L., Stock, J.H., Watson, M.W., 2018. Inference in SVARs identified with an external instrument. manuscript, Columbia University.

40. Nankervis J.C., Savin, N.E., 1988. The Student's t approximation in a stationary first order autoregressive model. Econometrica 56, 119-145.

41. Park, J.Y., Phillips, P.C.B., 1989. Statistical inference in regressions with integrated processes: Part 2. Econometric Theory 5, 95-131.

42. Pesavento, E., Rossi, B., 2006. Small-sample confidence intervals for multivariate impulse response functions at long horizons. Journal of Applied Econometrics 21, 1135-1155.

43. Pesavento, E., Rossi, B., 2007. Impulse response confidence intervals for persistent data: What have we learned? Journal of Economic Dynamics and Control 31, 2398-2412.

44. Phillips, P.C.B., 1987. Toward a unified asymptotic theory for autoregression. Biometrika 74, 535-547.

45. Phillips, P.C.B., 1998. Impulse response and forecast error variance asymptotics in nonstationary VARs. Journal of Econometrics 83, 21-56.

46. Phillips, P.C.B., 2014. On confidence intervals for autoregressive roots and predictive regression. Econometrica 82, 1177-1195.

47. Phillips, P.C.B., Lee, J.H., 2016. Robust econometric inference with mixed integrated and mildly explosive regressors. Journal of Econometrics 192, 433-450.

48. Pope, A.L., 1990. Biases of estimators in multivariate non-Gaussian autoregressions. Journal of Time Series Analysis 11, 249-258.

49. Sims, C.A., Stock, J.H., Watson, M.W., 1990. Inference in linear time series models with some unit roots. Econometrica 58, 113-144.

50. Stock, J.H., 1991. Confidence intervals for the largest autoregressive root in U.S. economic time series. Journal of Monetary Economics 28, 435–460.

51. Toda, H., Yamamoto, T., 1995. Statistical inference in vector autoregressions with possibly integrated processes. Journal of Econometrics 66, 225-250.

52. Wright, J.H., 2000. Confidence intervals for univariate impulse responses. Journal of Business and Economic Statistics 18, 368-373.

Table 1: Uniform coverage rates of nominal 90% impulse response confidence intervals based on the delta method

	Impulse response horizon											
T	1	2	3	4	5	6	7	8	9	10	11	12
AR model												
80	0.865	0.808	0.768	0.665	0.583	0.538	0.497	0.467	0.441	0.419	0.402	0.387
120	0.877	0.840	0.794	0.714	0.646	0.605	0.567	0.538	0.509	0.487	0.469	0.452
240	0.893	0.876	0.853	0.800	0.740	0.702	0.667	0.638	0.611	0.590	0.565	0.545
480	0.898	0.889	0.877	0.842	0.812	0.782	0.755	0.730	0.704	0.682	0.661	0.640
600	0.897	0.896	0.889	0.864	0.834	0.806	0.780	0.755	0.733	0.714	0.693	0.674
Lag-augmented AR model												
80	0.866	0.808	0.766	0.724	0.722	0.734	0.728	0.723	0.713	0.700	0.684	0.666
120	0.876	0.839	0.793	0.771	0.775	0.781	0.779	0.776	0.769	0.757	0.745	0.731
240	0.895	0.878	0.857	0.829	0.830	0.837	0.837	0.836	0.833	0.827	0.818	0.809
480	0.898	0.889	0.878	0.862	0.862	0.869	0.871	0.873	0.871	0.866	0.862	0.854
600	0.898	0.893	0.889	0.877	0.870	0.874	0.876	0.878	0.878	0.873	0.869	0.866

Notes: The data are generated from $y_t = \rho y_{t-1} + u_t$, $u_t \overset{iid}{\sim} N(0,1)$, $\rho \in \{0.2, 0.5, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 1\}$. The original model is an AR(4) with intercept. The lag-augmented model is an AR(5) model with intercept. Inference on the individual impulse responses is based on the delta method based on closed-form solutions for the impulse response standard error in Lütkepohl (1990). The uniform coverage rates are computed as the infimum of the coverage rates across the ρ values, adjusted for data mining bias.

Table 2: Uniform coverage rates of nominal 90% impulse response confidence intervals based on the bootstrap

	Impulse response horizon											
T	1	2	3	4	5	6	7	8	9	10	11	12
AR model: Hall percentile interval												
80	0.888	0.858	0.825	0.802	0.786	0.758	0.709	0.658	0.608	0.557	0.515	0.477
120	0.893	0.871	0.849	0.824	0.820	0.803	0.780	0.753	0.715	0.677	0.643	0.609
240	0.899	0.892	0.876	0.859	0.855	0.856	0.858	0.849	0.822	0.801	0.777	0.752
480	0.900	0.896	0.890	0.882	0.874	0.881	0.882	0.883	0.874	0.864	0.853	0.843
600	0.897	0.902	0.898	0.893	0.888	0.890	0.891	0.886	0.875	0.866	0.857	0.845
Lag-augmented AR model: Hall percentile interval												
80	0.887	0.852	0.828	0.790	0.768	0.738	0.705	0.681	0.645	0.607	0.567	0.524
120	0.889	0.872	0.849	0.825	0.811	0.796	0.778	0.753	0.726	0.693	0.658	0.619
240	0.900	0.894	0.879	0.874	0.863	0.855	0.845	0.831	0.815	0.795	0.768	0.740
480	0.897	0.895	0.890	0.886	0.882	0.881	0.881	0.875	0.867	0.854	0.835	0.813
600	0.898	0.898	0.896	0.895	0.887	0.887	0.885	0.880	0.875	0.862	0.852	0.829
AR model with bias adjustment: Efron percentile interval												
80	0.877	0.846	0.818	0.797	0.755	0.751	0.734	0.726	0.714	0.709	0.700	0.697
120	0.881	0.872	0.834	0.811	0.777	0.776	0.762	0.755	0.741	0.735	0.726	0.721
240	0.896	0.894	0.876	0.863	0.840	0.834	0.821	0.809	0.796	0.787	0.779	0.773
480	0.899	0.898	0.887	0.881	0.866	0.861	0.852	0.847	0.839	0.831	0.822	0.813
600	0.898	0.900	0.898	0.893	0.880	0.875	0.868	0.862	0.853	0.845	0.840	0.835
Lag-augmented AR model with bias adjustment: Efron percentile interval												
80	0.874	0.845	0.812	0.799	0.796	0.813	0.808	0.813	0.819	0.823	0.826	0.828
120	0.879	0.868	0.830	0.829	0.827	0.841	0.836	0.839	0.840	0.844	0.849	0.851
240	0.894	0.893	0.872	0.867	0.868	0.870	0.873	0.873	0.874	0.875	0.877	0.881
480	0.899	0.896	0.886	0.879	0.883	0.890	0.890	0.889	0.891	0.892	0.893	0.893
600	0.896	0.897	0.896	0.894	0.885	0.891	0.892	0.889	0.892	0.893	0.893	0.894

Notes: The data are generated from $y_t = \rho y_{t-1} + u_t$, $u_t \overset{iid}{\sim} N(0,1)$, $\rho \in \{0.2, 0.5, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 1\}$. The original model is an AR(4) with intercept. The lag-augmented model is an AR(5) model with intercept. The bootstrap data are generated using a recursive-design bootstrap method, as discussed in Appendix D. The uniform coverage rates are computed as the infimum of the coverage rates across the ρvalues, adjusted for data mining bias.

Table 3: Percentage loss in average size-adjusted power of the t test based on the lag-augmented autoregression relative to the t-test based on the original autoregression at the 10% significance level

		δ							
h	T	-5.0	-2.0	-1.0	-0.5	0.5	1.0	2.0	5.0
1	120	0.005	0.231	0.080	0.019	0.056	0.098	0.580	0.023
	240	-0.002	0.112	-0.002	-0.009	0.057	0.065	0.229	0.005
	480	0.001	0.066	0.006	0.025	0.082	0.122	0.098	0.003
	600	0.001	0.029	-0.002	-0.018	0.071	0.074	0.063	0.002
4	120	9.092	2.942	0.990	0.268	-0.273	-0.663	-1.447	-1.312
	240	8.931	4.102	1.496	0.629	-0.522	-1.373	-2.095	-0.168
	480	9.820	5.829	2.444	0.852	-0.857	-0.975	-1.135	2.070
	600	10.239	5.465	2.162	0.953	-0.887	-0.941	-0.707	2.616
8	120	17.847	4.629	1.987	1.011	-0.632	-1.266	-1.978	-6.016
	240	25.465	7.483	3.325	1.764	-1.138	-1.785	-3.352	-0.617
	480	31.110	10.525	4.512	1.981	-1.338	-2.505	-2.783	13.153
	600	33.164	11.283	4.852	1.687	-1.539	-2.433	-2.334	17.969
12	120	9.780	2.664	1.343	0.539	-0.531	-0.887	-1.427	-2.509
	240	20.147	5.672	2.407	1.152	-0.788	-1.412	-2.232	-2.267
	480	27.441	8.089	3.369	1.413	-1.107	-1.888	-2.692	2.929
	600	28.782	8.335	3.229	1.510	-0.864	-1.727	-2.708	5.762

Notes: The data are generated from $(1-\rho L)(1-\delta T^{-\frac{1}{2}}L)y_t = u_t$, $u_t \overset{iid}{\sim} N(0,1)$, with $y_1 = y_2 = 0$, $\rho \in \{0.2, 0.5, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 1\}$ and $\delta \in \{-2, -1, -0.5, 0.5, 1, 2\}$. The lag-augmented model is an AR(5) model with intercept. The original model is an AR(4) model with intercept. The null hypothesis that the h-step-ahead impulse response is given by ρ^h is tested based on the two-sided t test at the 10% significance level. The size-adjusted finite-sample critical values are simulated based on 10,000 draws for each ρ and δ. The percentage loss in the average power is based on the average of the percentage losses for each ρ and it is computed based on 10,000 simulation draws. A positive (negative) number indicates the t-test is more powerful when based on the original autoregression (lag-augmented autoregression).

Table 4: Uniform coverage rates of nominal 90% impulse response confidence intervals based on the bootstrap

T	Impulse response horizon								
	12	18	24	30	36	42	48	54	60
	Lag-augmented AR model with bias adjustment: Efron percentile interval								
80	0.831	0.899	0.857	0.906	0.889	0.894	0.905	0.894	0.907
120	0.852	0.889	0.868	0.873	0.886	0.876	0.888	0.885	0.882
240	0.881	0.884	0.886	0.885	0.887	0.888	0.888	0.888	0.888
480	0.893	0.894	0.896	0.897	0.897	0.897	0.897	0.897	0.898
600	0.891	0.894	0.896	0.897	0.897	0.898	0.898	0.898	0.898

Notes: The data are generated from $y_t = \rho y_{t-1} + u_t$, $u_t \overset{iid}{\sim} N(0,1)$, $\rho \in \{0.2, 0.5, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 1\}$. The lag-augmented model is an AR(5) model with intercept. The bootstrap data are generated using a recursive-design bootstrap method, as discussed in Appendix D. The uniform coverage rates are computed as the infimum of the coverage rates across the ρ values, adjusted for data mining bias.

Table 5: Uniform coverage rates of nominal 90% impulse response confidence intervals based on lag-augmented autoregressions

T	$\lambda = h/T$ 0.1	0.3	0.5	0.7	0.9
			Delta method interval		
120	0.729	0.546	0.487	0.454	0.433
240	0.704	0.554	0.477	0.477	0.464
480	0.697	0.568	0.499	0.499	0.488
600	0.682	0.552	0.494	0.494	0.484
			Hall percentile interval		
120	0.621	0.401	0.376	0.363	0.357
240	0.543	0.429	0.407	0.399	0.393
480	0.542	0.458	0.440	0.433	0.429
600	0.524	0.457	0.443	0.435	0.432
			Efron percentile interval		
120	0.745	0.797	0.795	0.805	0.804
240	0.840	0.847	0.850	0.848	0.850
480	0.878	0.881	0.882	0.882	0.882
600	0.887	0.887	0.888	0.888	0.887
			Efron percentile interval after bias adjustment		
120	0.855	0.887	0.888	0.891	0.890
240	0.884	0.887	0.889	0.889	0.891
480	0.896	0.888	0.898	0.888	0.896
600	0.901	0.900	0.898	0.900	0.900

Notes: The data are generated from $y_t = \rho y_{t-1} + u_t$, $u_t \overset{iid}{\sim} N(0,1)$, $\rho \in \{0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 1\}$. The lag-augmented model is an AR(5) model with intercept. The bootstrap data are generated using a recursive-design bootstrap method, as discussed in Appendix D. The uniform coverage rates are computed as the infimum of the coverage rates across the ρ values, adjusted for data mining bias.

Table 6: Uniform coverage rates of nominal 90% impulse response confidence intervals in AR(∞) models based on lag-augmented autoregressions

θ	p	Horizon											
		3	6	9	12	18	24	30	36	42	48	54	60
+0.25	4	0.894	0.882	0.880	0.882	0.885	0.886	0.887	0.887	0.887	0.886	0.887	0.888
	6	0.897	0.881	0.893	0.895	0.899	0.902	0.909	0.901	0.900	0.900	0.906	0.903
	8	0.898	0.899	0.899	0.890	0.883	0.881	0.878	0.877	0.875	0.878	0.880	0.885
-0.25	4	0.900	0.887	0.881	0.884	0.885	0.885	0.885	0.885	0.885	0.885	0.885	0.885
	6	0.897	0.885	0.885	0.885	0.888	0.891	0.894	0.894	0.895	0.894	0.894	0.894
	8	0.896	0.903	0.895	0.891	0.881	0.884	0.880	0.881	0.873	0.882	0.881	0.886

Notes: The data are generated from $y_t = \rho y_{t-1} + u_t + \theta u_{t-1}$, $u_t \overset{iid}{\sim} N(0,1)$, $\rho \in \{0.2, 0.5, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 1\}$, $\theta \in \{-0.25, 0.25\}$. The lag-augmented model is an AR(p+1) model with intercept. The bootstrap data are generated using a recursive-design bootstrap method, as discussed in Appendix D. The results are based on the Efron percentile interval based on the bias-adjusted lag augmented autoregression. The sample size is $T = 600$. The uniform coverage rates are computed as the infimum of the coverage rates across the ρ values, adjusted for data mining bias.