Confidence Intervals for Univariate Impulse Responses With a Near Unit Root

Jonathan H. WRIGHT
International Finance Division, Board of Governors of the Federal Reserve System, Washington, DC 20551
(*jonathan.h.wright@frb.gov*)

This article proposes a method for constructing confidence intervals for the impulse response function of a univariate time series with a near unit root. These confidence intervals control coverage, whereas the existing techniques can all have coverage far below the nominal level. I apply the proposed method to several measures of U.S. aggregate output.

KEY WORDS: Bootstrap; Confidence intervals; Impulse response; Unit roots.

A major topic in empirical macroeconomics is the analysis of impulse response functions. These give the effect of a shock on future values of the time series being considered. Usually, impulse response analysis is considered in the context of a vector autoregression. But there are also univariate applications of impulse response analysis. Researchers are interested in the persistence of shocks to series such as aggregate output (e.g., Diebold and Rudebusch 1989; Koop 1996). A large econometric literature has considered inference concerning aspects of the mean reversion of univariate time series. The many possible references include Cochrane (1988), Stock (1991), Koop, Osiewalski, and Steel (1994), and Andrews and Chen (1994). The spectral density of the differences of a time series at frequency 0 is often used as a persistence measure. But this is just a transformation of the impulse response at an infinite lead time. The impulse response function gives a complete representation of the conditional mean reversion properties of a time series.

Point estimates of an impulse response are, however, of little use without associated confidence intervals. In the empirical literature, confidence intervals for the impulse response function (of a univariate or multivariate system) are usually constructed either by the delta method or by the bootstrap (Runkle 1987; Lütkepohl 1990). These work reasonably well for a time series that is clearly stationary. But the time series that are encountered in empirical macroeconomics appear to be nonstationary or nearly nonstationary. For nonstationary or nearly nonstationary time series, the delta method and the bootstrap may yield confidence intervals with poor coverage properties. Recently, Bayesian and bias-adjusted bootstrap methods have been proposed (Kilian 1998; Sims and Zha 1999). These generally work considerably better than the delta method and the simple bootstrap in the presence of large roots. But the Bayesian and bias-adjusted bootstrap confidence intervals can nevertheless have coverage that is well below the nominal level.

In this article, I propose an algorithm for constructing confidence intervals for the impulse response function of a univariate time series with an autoregressive root that is near to unity. The large root may be an exact unit root, but it does not have to be an exact unit root. This root is modeled as being local to unity. Instead of being a fixed parameter, it converges to 1 as the sample size increases. Modeling a root as being local to unity is a statistical device that gives an excellent approximation to the finite-sample properties of estimators and test statistics when a root is close to unity and the sample size is small; see Stock (1991, 1995) and the references therein. It does not mean that the researcher in fact believes that the largest autoregressive root of an economic time series depends on the observed sample size. The confidence intervals for the impulse response function formed by the algorithm proposed in this article are asymptotically conservative. That is, they will always have coverage that is asymptotically greater than or equal to its nominal level in the limit as the sample size goes to infinity; I know of no other way of controlling asymptotic coverage uniformly in the parameter space. I also report Monte Carlo evidence to show that they successfully control coverage in finite samples.

The layout of the remainder of this article is as follows. Section 1 briefly reviews the existing methods of constructing confidence intervals for a univariate impulse response function. Section 2 describes the proposed new method. Section 3 contains the Monte Carlo simulations. An application to U.S. aggregate output is presented in Section 4, and Section 5 concludes.

1. THE EXISTING METHODS

Consider the scalar autoregression

$$a(L)z_t = \varepsilon_t, \tag{1.1}$$

where ε_t is iid $(0, \sigma^2)$, ε_t has $2 + \delta$ finite moments for some $\delta > 0$, and $a(L)$ is a lag polynomial of order q. The observed time series is y_t, and the model specifies that $y_t = z_t$ (no deterministics), $y_t = \mu_1 + z_t$ (model with constant), or $y_t = \mu_1 + \mu_2 t + z_t$ (model with constant and trend). I define the impulse response at lead time l as the effect of a shock of size σ in ε_t on y_{t+l} and denote this as $h_l(\theta)$, where θ is the vector of parameters in $a(L)$ and σ^2. Defining the impulse response as the effect of a one-standard-deviation shock means that it is scale-independent and is the univariate special case of the orthogonalized impulse responses considered by Sims (1980).

The standard point estimate of $h_l(\theta)$ is simply $h_l(\hat{\theta})$, where $\hat{\theta}$ is the ordinary least squares (OLS) estimator of θ. This article is concerned with forming confidence intervals for $h_l(\theta)$. As in most of this literature (e.g., Runkle 1987; Lütkepohl 1990; Kilian 1998), I construct pointwise confidence intervals—that is, confidence intervals for any single value of l. When the roots of $a(L)$ are strictly outside the unit circle, $T^{1/2}(\hat{\theta} - \theta) \Rightarrow \mathrm{N}(0, \Omega)$ for some nonsingular covariance matrix Ω. Consequently, by the delta method,

$$T^{1/2}(h_l(\hat{\theta}) - h_l(\theta)) \Rightarrow \mathrm{N}\left(0, \frac{dh_l(\theta)'}{d\theta}\Omega\frac{dh_l(\theta)}{d\theta}\right),$$

giving a confidence interval for $h_l(\theta)$ (Runkle 1987; Lütkepohl 1990).

A bootstrap can also be used to construct a confidence interval for the impulse response function. This involves calculating the residuals in (1.1). The researcher then draws from these with replacement, constructing artificial datasets, and then estimates the impulse responses for each of these bootstrap datasets. To construct each bootstrap sample, q initial conditions are required. I follow Kilian (1998) in randomly choosing a block of q observations from the underlying time series for this purpose. If the lower and upper $(1-\alpha)/2$ quantiles of the bootstrap distribution of $h_l(\theta)$ are t_l and t_u, respectively, then the $100\alpha\%$ bootstrap confidence interval is simply $[t_l, t_u]$. There are other ways of constructing a bootstrap confidence interval that were discussed and compared in the context of impulse response analysis by Kilian (1999). I follow Runkle (1987) and Kilian (1998, 1999) in using this method.

The bootstrap and delta methods work reasonably well in finite samples if the time series has no roots close to the unit circle. But otherwise they may have coverage that is far below the nominal level. This is shown in the simulations in this article and (in a multivariate setting) by many others, including Griffiths and Lütkepohl (1993), Fachin and Bravetti (1996), and Kilian (1998). In fact, the delta method and the bootstrap are not necessarily even asymptotically valid. In the case of a univariate random walk, the delta method is based on a premise of asymptotic normality of the autoregressive coefficient estimate, which has in fact a highly nonnormal asymptotic distribution. Basawa, Mallik, McCormick, Reeves, and Taylor (1991) proved the asymptotic failure of the bootstrap in the random-walk model (if the unit root is not imposed in estimation). In higher-order autoregressions with a unit root, the delta method is asymptotically valid if the lead time of the impulse response is treated as fixed, but not if it is nested as being proportional to the sample size (Phillips 1998).

Kilian (1998) proposed a bias-adjusted bootstrap algorithm for forming confidence intervals. This uses a bootstrap to estimate the bias of the autoregressive parameters and adjusts the OLS estimate for this bias before running the standard bootstrap. A correction is made to the bias adjustment to prevent it from inducing an explosive root in the autoregressive parameters. This bias adjustment does not resolve the asymptotic failure of the bootstrap in the simple random-walk model. It can still give confidence intervals with coverage well below the nominal level (see Kilian and the simulations following). But the Monte Carlo results of Kilian indicate that it often gives much better coverage properties than the delta method or the ordinary bootstrap.

Sims and Zha (1999) proposed forming equal-tailed Bayesian confidence intervals for the impulse response function, based on Monte Carlo integration. A $100\alpha\%$ interval is formed by simulating the posterior of $h_l(\theta)$ and then simply taking the upper and lower $(1-\alpha)/2$ quantiles of this distribution. This can be interpreted as a classical confidence interval. It does not consistently achieve coverage equal to the nominal level, but it can also represent a major improvement on the delta method or the ordinary bootstrap (see Kilian 1998, Sims and Zha 1999, and the simulations following).

This article is concerned with a new method for forming confidence intervals in a time series with a single large root. It is a two-step procedure, recognizing that inference about the near unit root is the source of the problems with existing methods of inference for impulse responses. The first step involves forming a confidence interval for the largest autoregressive root of the time series by an algorithm due to Stock (1991). This would yield a confidence interval for the impulse response of an AR(1) (autoregression of order 1). But, for a higher-order autoregression there are additional nuisance parameters. These are allowed for in the second step of the proposed method.

2. THE PROPOSED METHOD

Consider the univariate autoregression given by Equation (1.1), where $a(L)$ has a single large root. It can be factorized as $a(L) = b(L)(1 - \rho L)$, where $b(L)$ is a lag polynomial of order $q-1$ with no roots on or near the unit circle and ρ is the largest autoregressive root of the process, which is sufficiently close to 1 that it is modeled as local to unity, $\rho = 1 + c/T$. As discussed in the introduction, this nesting is simply a statistical device that is helpful for approximating the finite-sample distribution of estimators and test statistics in the presence of large autoregressive roots. Although it is conceptually possible to extend the proposed algorithm to multivariate models with matrices of roots local to unity, its implementation in that context is extremely cumbersome and is omitted from this article.

The parameter c is not consistently estimable, but Stock (1991) showed how to construct a $100\alpha\%$ confidence interval for the scalar c, or equivalently for ρ. The idea is to derive the asymptotic distribution of the augmented Dickey–Fuller test for different values of c and to form the confidence interval for c as the inverse of the acceptance region of the test. Denote this confidence interval as C_α. I also impose the restriction that $c \leq 0$, ruling out explosive roots (though not ruling out exact unit roots). Although this restriction is not necessary as a matter of theory, it greatly reduces the width of the confidence intervals for the impulse response function constructed by the method proposed in this article. In finite samples, when ρ is close to unity, the confidence interval C_α has coverage that is very close to the nominal level (Stock 1991). Exact confidence intervals for ρ are available in the special case of a Gaussian AR(1). The advantage of Stock's algorithm is that it applies more generally.

Partition the vector of parameters θ as $(\rho, \gamma')'$. For each c in C_α, a $100\alpha\%$ confidence interval for $h_l(\theta)$ may be formed by the delta method treating c as fixed, with γ estimated by fitting a stationary autoregression to $(1-\rho L)y_t$. Let this confidence interval be $[h_l^A(c), h_l^B(c)]$. By the Bonferroni inequality,

$$H_l = [\inf_{c \varepsilon C_\alpha} h_l^A(c), \sup_{c \varepsilon C_\alpha} h_l^B(c)]$$

is a confidence interval for $h_l(\theta)$ with coverage of at least $100[1-2(1-\alpha)]\%$, asymptotically. For example, if $\alpha = 95\%$, then the asymptotic coverage of H_l is at least 90%. Neither $h_l^A(c)$ nor $h_l^B(c)$ is necessarily monotone in c, so H_l must in practice be computed by taking the minimum and maximum of $h_l^A(c)$ and $h_l^B(c)$ over a discrete grid of values of c in C_α. In this article, I use a 100-point, equally spaced grid.

This is the algorithm that I propose for forming confidence intervals for univariate impulse responses. It can indeed be used to form a confidence interval for any nonlinear function of the parameters of a univariate autoregression, but impulse responses are a leading case. Nesting ρ as being local to unity, this method is asymptotically conservative and can be expected to yield finite-sample coverage that is typically greater than the nominal level. This is the sense in which I say that it controls coverage. Conservative confidence intervals have been proposed in other contexts by many authors (e.g., Campbell and Dufour 1997). For the coverage to be too high is not a problem in itself (it means that there are "too few" Type I errors), but the confidence intervals can be expected to be relatively wide.

A conventional rule for evaluating alternative tests is that the test with the highest power is chosen among the set of tests with size no greater than a prespecified probability of making a Type I error. The equivalent rule for evaluating alternative confidence intervals is that the confidence interval with the shortest average width is chosen among all those confidence intervals that control coverage. This classical rule specifies lexicographic preferences: A confidence interval with coverage at or above the nominal level must be preferred to any confidence interval with coverage below the nominal level. All of the existing algorithms for forming confidence intervals for impulse responses have coverage that can be far below the nominal level. So, if the new method succeeds in controlling coverage, then it must be preferred to all these existing algorithms in the simple model in which it is applicable.

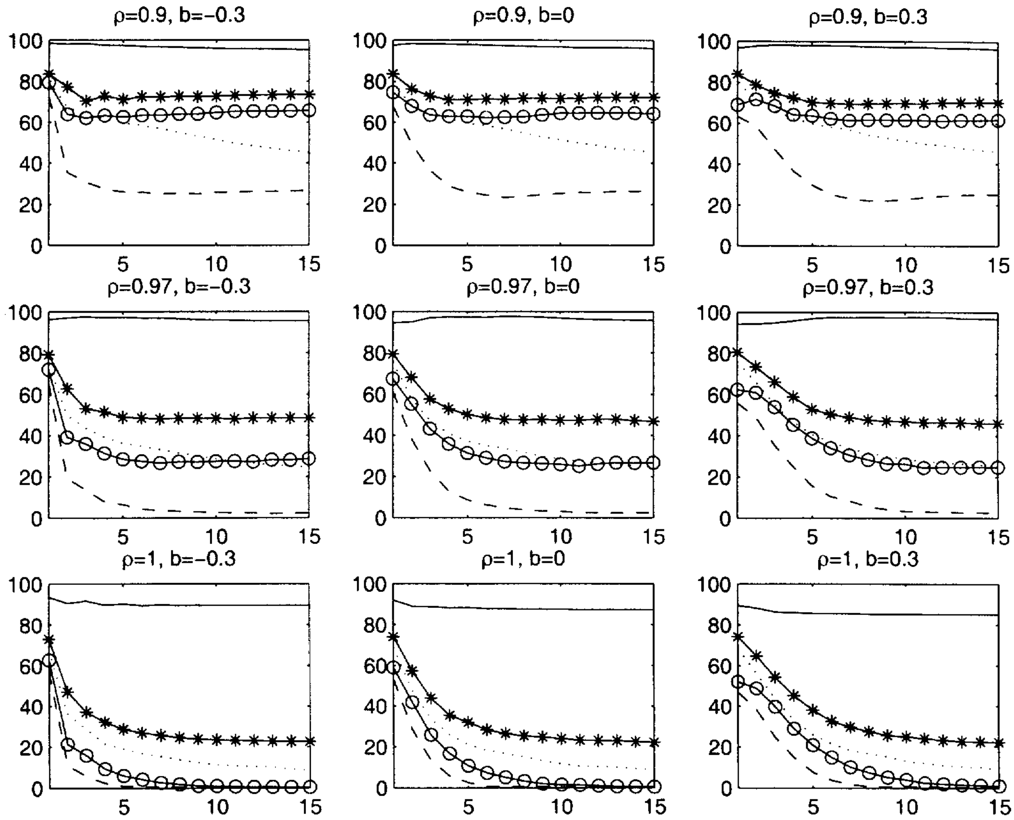

Figure 1. Percentage Coverage of Confidence Intervals: Solid Line, New Method; Dots, Delta Method; Dashes, Bootstrap; Solid Line With Circles, Bias-Adjusted Bootstrap; Solid Line With Stars, Bayesian Method. The lead time of the impulse response function is plotted on the horizontal axis. The design of the experiment is described in the text.

The width and coverage of the alternative confidence intervals will, in practice, both be of interest to researchers. Section 3 reports a Monte Carlo simulation comparing their width and coverage.

3. MONTE CARLO EVIDENCE

In this section, I report a simulation comparing the performance of the delta method, the bootstrap, the bias-adjusted bootstrap, the Bayesian method, and the proposed new method in a simple univariate autoregression. The model considered is (1.1) with a constant and trend $y_t = \mu_1 + \mu_2 t + z_t$, where $(1 - bL)(1 - \rho L)z_t = \varepsilon_t$ and ε_t is iid $\mathrm{N}(0, \sigma^2)$ with $\sigma^2 = 1$. The parameters μ_1 and μ_2 are both set to 0, but this is not imposed in estimation. The parameter ρ is 1, .97, or .9, and b is $-.3, 0$, or .3. The sample size is 100, and 1,000 replications were conducted for each experiment. The lag order of the autoregression is treated as known. In each application of the bootstrap, throughout this article, 1,000 bootstrap draws are used. The coverage rates and average width of the confidence intervals for the impulse response at lead times 1–15 are shown in Figures 1 and 2.

In this experiment, the new method has actual coverage that is nearly always above the 90% nominal level. In no case does it fall much below 90%. All the other methods have coverage that can be well below the nominal level, especially at the longer lead times and especially if $\rho = 1$. The bias-adjusted bootstrap represents an improvement on the ordinary bootstrap but still has coverage that is much too low. The least squares estimate of ρ is below the true value of ρ with probability close to 1. This causes the delta method and the bootstrap to have poor coverage. The bias-adjusted bootstrap and the Bayesian method may help, but they are not complete solutions to the problem.

Naturally, the new method produces confidence intervals that are wider than the existing methods, especially at long lead times. The average width of confidence intervals for impulse responses at long lead times, generated by the new method, is close to $\sigma/(1-b)$. But this makes sense because in all the models considered in the Monte Carlo experiment it is very hard to tell if the time series has a unit root or not. Unit-root tests do not have that much power. If the series *does* have a unit root, then the impulse response at an infinite lead time is $\sigma/(1-b)$. If it *does not* have a unit root, then the impulse response at an infinite lead time is 0. If the researcher cannot tell whether the series has a unit root or not, then the confidence interval at long lead times should include both these possibilities and the confidence interval should have a width of at least $\sigma/(1-b)$. The new method of forming confidence intervals controls coverage but at the price that the confidence intervals are wider. In the light of the poor coverage of all the existing methods in these simulations and the preceding intuition for what the width of a properly constructed confidence interval should be, the width of the confidence intervals generated by the new method seems to be an appropriate statement about our uncertainty concerning the long-run effect of shocks.

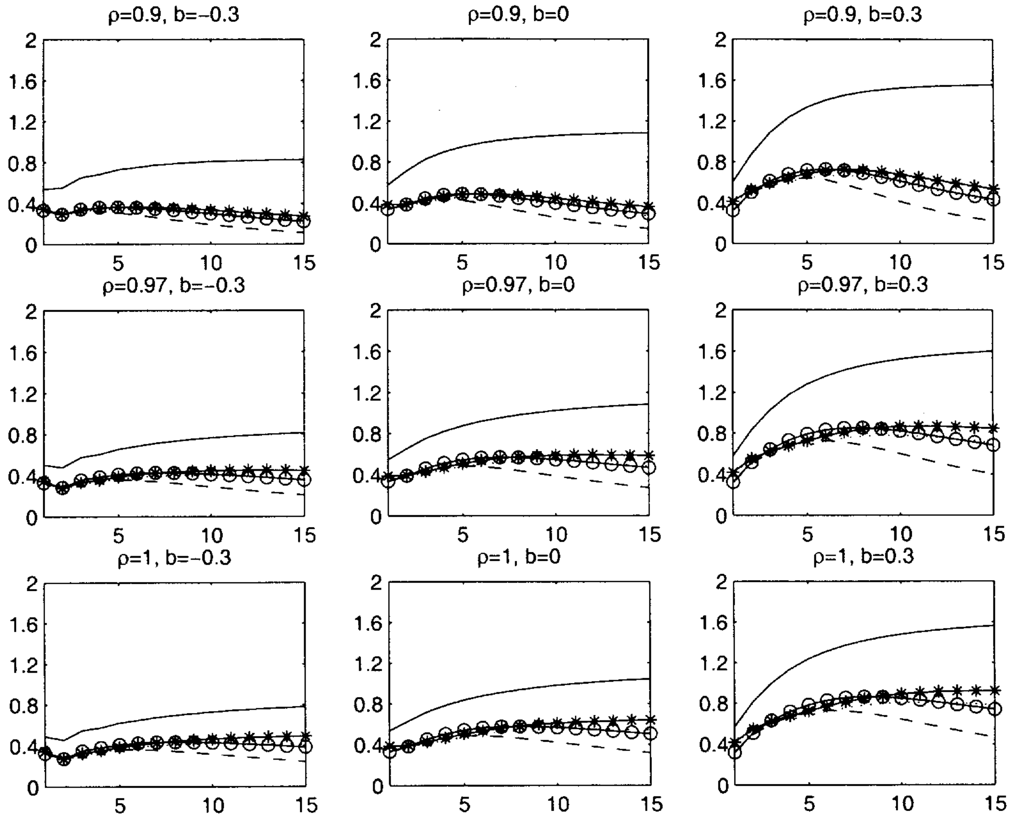

Figure 2. Average Width of Confidence Intervals: Solid Line, New Method; Dots, Delta Method; Dashes, Bootstrap; Solid Line With Circles, Bias-Adjusted Bootstrap; Solid Line With Stars, Bayesian Method. The lead time of the impulse response function is plotted on the horizontal axis. The design of the experiment is described in the text.

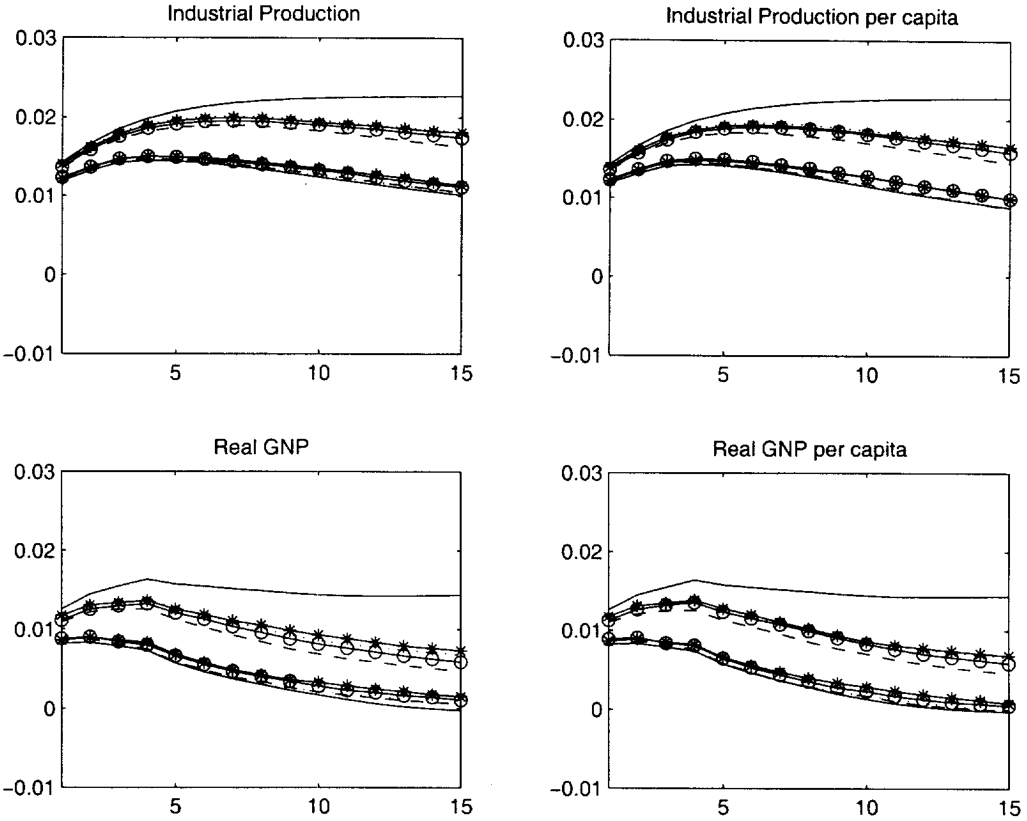

Figure 3. Confidence Intervals for Impulse Response of Log Output: Solid Line, New Method; Dots, Delta Method; Dashes, Bootstrap; Solid Line With Circles, Bias-Adjusted Bootstrap; Solid Line With Stars, Bayesian Method. The lead time of the impulse response function is plotted on the horizontal axis.

Qualitatively similar results are obtained using a sample of size 50 and in a model in which there is no trend term (either in the fitted model or the data-generating process). In the latter case, the Bayesian method and bias-adjusted bootstrap do much better in terms of coverage, but the coverage can still fall well short of the required nominal level. These and some other simulation results are not shown but are available from the author on request.

4. AN APPLICATION TO U.S. AGGREGATE OUTPUT

The mean reversion of macroeconomic time series, especially aggregate output, has received a lot of attention in the applied econometric literature over the past two decades. In this section, I present an application of the methodology proposed in this article to the construction of confidence intervals for the impulse response function for several measures of U.S. aggregate output. The univariate impulse response analysis of U.S. aggregate output has been considered in several articles including that of Diebold and Rudebusch (1989).

I measure aggregate output by monthly industrial production (covering the period 1947:1 to 1996:7) and by quarterly real gross national product (GNP) (covering the period 1959:3 to 1996:2). In addition, I consider both these series in per capita terms. The data were all obtained from CITIBASE—mnemonics IP, GNPQ, and POP, for industrial production, real GNP, and population, respectively.

I then formed a 90% confidence interval for the effect of a one-standard-deviation innovation on the value of log output after l periods. I used the new method and each of the existing methods—the delta method, the Bayesian method, the bootstrap, and the bias-adjusted bootstrap. In each case, the time series was modeled as being an AR(p) with a deterministic time trend, with p being selected so as to minimize the Akaike information criterion, as suggested by Kilian (1997).

The results are shown in Figure 3, at lead times from 1 to 15. The results are very similar for all four measures of aggregate output. The Bayesian and the bias-adjusted bootstrap give results that are nearly identical. Both indicate slightly more persistence than is found using the regular bootstrap. Interestingly, all the methods of constructing confidence intervals are in quite close agreement on the *lower* bound of the confidence interval for the impulse response function. But the new method gives an *upper* bound of the confidence interval that is considerably higher than that obtained by any of the other four methods. So the new method gives somewhat wider confidence intervals, which include some considerably more persistent impulse response functions. The new method (which was the only one to consistently achieve coverage over 30% in the simulations of Section 3) gives different conclusions about the mean reversion of output.

5. CONCLUSION

In this article, I have considered the problem of forming a confidence interval for the impulse response function of a univariate time series with a single large root. In a Monte Carlo simulation, it is the only method that yields coverage that is consistently at or above the nominal level. The conventional methods can all have coverage that is far too low. Especially at long lead times, the confidence intervals yielded by the new method are wide. In small samples, unit-root tests have little power against slowly mean-reverting alternatives. If we are not able to tell whether a time series has a unit root or not, then great uncertainty about the long-term effect of a shock seems entirely appropriate. The difficulties with the conventional approaches for forming confidence intervals for impulse responses in the presence of large roots also apply in multivariate vector autoregressions and motivate further research; this article has proposed a simple and practical solution in the univariate case.

ACKNOWLEDGMENTS

I am grateful to Lutz Kilian, Guido Imbens, Jim Stock, Luis Viceira, Mark Watson, and three anonymous referees for helpful comments on earlier versions of this article. All errors are my sole responsibility.

[Received May 1997. Revised September 1999.]

REFERENCES

Andrews, D. W. K., and Chen, H. Y. (1994), "Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business & Economic Statistics*, 12, 187–204.

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H., and Taylor, R. L. (1991), "Bootstrapping Unstable First Order Autoregressive Processes," *The Annals of Statistics*, 19, 1098–1101.

Campbell, B., and Dufour, J. M. (1997), "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter," *International Economic Review*, 38, 151–173.

Cochrane, J. H. (1988), "How Big is the Random Walk in GNP?" *Journal of Political Economy*, 96, 893–920.

Diebold, F. X., and Rudebusch, G. D. (1989), "Long Memory and Persistence in Aggregate Output," *Journal of Monetary Economics*, 24, 189–209.

Fachin, S., and Bravetti, L. (1996), "Asymptotic Normal and Bootstrap Inference in Structural VAR Analysis," *Journal of Forecasting*, 15, 329–341.

Griffiths, W., and Lütkepohl, H. (1993), "Confidence Intervals for Impulse Responses From VAR Models: A Comparison of Asymptotic Theory and Simulation Approaches," unpublished manuscript, Humboldt University, Institute for Statistics and Econometrics.

Kilian, L. (1997), "Impulse Response Analysis in Vector Autoregressions With Unknown Lag Order," unpublished manuscript, University of Michigan, Dept. of Economics.

——— (1998), "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, 80, 218–230.

——— (1999), "Finite-Sample Properties of Percentile and Percentile-*t* Bootstrap Confidence Intervals for Impulse Responses," *Review of Economics and Statistics*, 81, 1–10.

Koop, G. (1996), "Parameter Uncertainty and Impulse Response Analysis," *Journal of Econometrics*, 72, 135–149.

Koop, G., Osiewalski, J., and Steel, M. F. J. (1994), "Posterior Properties of Long Run Impulse Responses," *Journal of Business & Economic Statistics*, 12, 489–492.

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.

Phillips, P. C. B. (1998), "Impulse Response and Forecast Variance Asymptotics in Nonstationary VARs," *Journal of Econometrics*, 83, 21–56.

Runkle, D. E. (1987), "Vector Autoregressions and Reality," *Journal of Business & Economic Statistics*, 5, 437–442.

Sims, C. A. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

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

Stock, J. H. (1991), "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435–459.

——— (1995), "Cointegration, Long-Run Comovements and Long Horizon Forecasting," in *Advances in Economics and Econometrics: Theory and Applications* (Vol. 3), eds. K. F. Wallis and D. M. Kreps, Cambridge, U.K.: Cambridge University Press, pp. 34–60.