EUI WORKING PAPERS
ECO No. 2006/19

ECO

# Impulse Response Confidence Intervals for Persistent Data: What Have We Learned?

ELENA PESAVENTO AND BARBARA ROSSI

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

**EUROPEAN UNIVERSITY INSTITUTE**
**DEPARTMENT OF ECONOMICS**

# *Impulse Response Confidence Intervals for Persistent Data: What Have We Learned?*

**ELENA PESAVENTO AND BARBARA ROSSI**

EUI Working Paper **ECO** No. 2006/19)

ISSN 1725-6704



Printed in Italy
European University Institute
Badia Fiesolana
I – 50016 San Domenico di Fiesole (FI)
Italy

http://www.iue.it/
http://cadmus.iue.it/dspace/index.jsp

# Impulse Response Confidence Intervals for Persistent Data: What Have We Learned?

Elena Pesavento — Emory University

Barbara Rossi — Duke University

March 2006

Abstract.

This paper is a comprehensive comparison of existing methods for constructing confidence bands for univariate impulse response functions in the presence of high persistence. Monte Carlo results show that Kilian (1998a), Wright (2000), Gospodinov (2004) and Pesavento and Rossi (2005) have favorable coverage properties, although they differ in terms of robustness at various horizons, median unbiasedness, and reliability in the possible presence of a unit or mildly explosive root. On the other hand, methods like Runkle's (1987) bootstrap, Andrews and Chen (1994), and regressions in levels or first differences (even when based on pre-tests) may not have accurate coverage properties. The paper makes recommendations as to the appropriateness of each method in empirical work.

*Keywords: Local to unity asymptotics, persistence, impulse response functions.*

JEL Classification: C1, C2.

Acknowledgements: We are extremely grateful to Lutz Kilian for suggesting many constructive ways to improve both the contents and the exposition of the paper, and for accurate and detailed comments on an earlier draft. We also thank Nikolay Gospodinov for sharing his codes, and for providing extraordinary help in implementing them. Part of this research was done while the first author was a Jean Monnet Fellow in the Economics Department of the European University Institute.

Corresponding author: Barbara Rossi, Department of Economics, Duke University, Durham NC27708, USA. E-mail: brossi@econ.duke.edu. Phone: (919) 660 1801. Fax: (919) 684 8974.

## 1. Introduction

Impulse response functions (IRFs) are the most commonly used tool to evaluate the effect of shocks on macroeconomic time series. Confidence intervals (CIs) for IRFs are commonly based on Lütkepohl's (1990) asymptotic normal approximation or bootstrap approximations to that distribution (see Runkle (1987) and Kilian (1998a, 1999)). The properties of these traditional approximations, however, may crucially depend on whether the series are assumed to be stationary or integrated. The objective of this paper is to provide a comprehensive comparison of methods available in the literature, and to give constructive recommendations to empirical macroeconomists who need to estimate IRFs and construct their confidence bands in the presence of a root close to unity.

Until recently, many researchers dealing with the possibility of a root close to unity chose to specify autoregressions either in levels or first differences. However, even when standard methods of inference are justified asymptotically, in practice confidence bands may have poor coverage properties in small samples in the presence of highly persistent variables, as shown by Kilian and Chang (2000), Ashley and Verbrugge (2001), Rossi (2005), and Pesavento and Rossi (2005). This literature shows that there are cases when none of the traditional methods applied to regressions in levels or first differences are a good choice. Unit root pre-tests do not solve the problem, as the actual coverage of IRF bands obtained after a pre-test can be quite different from the nominal one (see Cavanagh et al. (1995) and Elliott (1998) for a

discussion of the theoretical reasons behind the poor performance of unit root pre-tests). We will show in Section 3 that even unit root pre-tests with high power will not result in inference with the correct rejection probabilities.

It is well known that most macroeconomic variables have roots that are close to unity, so it is unsettling that traditional methods may fail exactly in the situations typical of most macroeconomic applications. Given the deficiencies of traditional methods, the recent literature has moved in the direction of devising methods that are robust to the presence of variables with roots equal or slightly less than one. In the context of the IRF, in particular, there have been some advances in the attempt to solve the problem of constructing confidence intervals with coverage rates that are close or bounded by the nominal rate even when variables are highly persistent. Andrews and Chen (1994) propose a median unbiased method to estimate the parameters of an AR process from which median unbiased estimates for the IRFs can be computed (see also Andrews, 1993). Alternative methods based on bootstrap approximations are recommended by Kilian (1998a) and Hansen (1999). Whereas the former attempts to extend the range of statistical models for which the bootstrap works, the latter proposes a grid bootstrap method for local to unity processes. More recently, Wright (2000), Gospodinov (2004) and Pesavento and Rossi (2005) have suggested the use of local-to-unity devices to obtain better approximations to the IRFs' distribution. Wright (2000) proposes to construct confidence bands based on Bonferroni bounds, while Gospodinov (2004) relies on the inversion of a likelihood

ratio test in which the constrained estimate exploits a null hypothesis on the value of the IRF at some horizon of interest. Both Wright (2000) and Gospodinov (2004) focus on univariate IRFs. Finally, Pesavento and Rossi (2005) derive analytic approximations to multivariate IRFs CI by using local-to-unity approximations at long horizons.

Given the diversity of recently developed methods, the empirical macroeconomist is left with a variety of choices. But what are their relative strengths and their weaknesses? Which method should be chosen by the researcher facing a specific problem? This paper provides an answer to these questions by comparing existing methods for constructing confidence intervals for univariate IRFs in the presence of highly persistent processes. Although the currently proposed robust methods have been used in empirical applications[1] and some comparisons have been proposed by the original authors, to our knowledge, none has provided a systematic comparison. While the current literature agrees on the need to use robust methods for inference, it is important that we understand the relative performance of the different approaches, so we can provide guidance to practitioners. This paper aims at providing such a guide.

We focus on IRFs in univariate models. Although it is common to think of IRFs in the context of multivariate models, there are relevant empirical applications in

[1] Some examples are Murray and Papell (2002), Rossi (2004), Lopez, Murray and Papell (2003) for half-life deviations from Purchasing Power Parity, and Wright (2000) for impulse response of aggregrate output.

univariate models. Typical examples are the evaluation of the persistence of shocks on aggregate output (Diebold and Rudebusch, 1989, and Campbell and Mankiw, 1987) or of the effect of shocks on real exchange rates as a measure of the deviations of the nominal exchange rates from their PPP level (see Murray and Papell,2002; Kilian and Zha, 2002; Busetti et al., 2005, and references therein). Furthermore, some of the methods that we compare are available only for univariate models.[2]

The reminder of this article is organized as follow. Section 2 briefly reviews each of the existing methods and their implied assumptions. Section 3 compares the coverage probabilities of the recently proposed methods with traditional approaches, including autoregressions in levels or first differences and autoregressions based on unit root pre-tests. Section 4 concludes.

## 2. Review of Robust Methods

Consider the standard scalar autoregression

$$a(L)\, y_t = \varepsilon_t \tag{1}$$

where $\varepsilon_t$ is a mean zero independent and identically distributed (i.i.d.) with variance equal to $\sigma^2$, $t = 1, 2, ...T$, where $T$ is the total sample size.[3] For sake of simplicity of exposition, let's ignore deterministic terms which are irrelevant in the construction

[2] A comparison of some of the methods that can be generalized to the multivariate case is presented in Pesavento and Rossi (2005).

[3] The i.i.d. assumption is stronger than needed, but convenient for expository purposes.

of the impulse response function of $y_t$. If we isolate the largest root, the process in (1) can be equivalently written as

$$(1 - \rho L)\, y_t = \theta\,(L)\, \varepsilon_t, \tag{2}$$

where $\theta\,(L) = 1 + \theta_1 L + \theta_2 L^2 + ...$ and the roots of $\theta\,(L)$ are all outside the unit circle, so there is no more than one root equal to one. Define the impulse response function at horizon $h$ as the effect of a shock of size $\sigma$ in $\varepsilon_t$ on $y_{t+h}$: $IRF_h = \frac{\partial y_{t+h}}{\partial \varepsilon_t}$.

When the largest root in the process (2) is one or close to one, traditional approximations like the delta method (whether based on analytical solutions or simulations) or the standard bootstrap approximations may provide confidence intervals for $IRF_h$ with poor coverage properties. Several methods have been proposed in the current literature to deal with this problem.

To improve on the traditional bootstrap approach, Kilian (1998a) proposes a bias-adjusted bootstrap method for constructing confidence intervals for persistent but stationary autoregressions. Kilian (1998b) recognizes the need to account for the small-sample bias and skewness of the small-sample distribution of the IRF estimator. The bias-adjusted bootstrap is based on bias corrected estimates of the autoregressive parameters. Kilian (1998a,b) shows that the bias-adjusted bootstrap provides a significant improvement in coverage accuracy over the standard bootstrap and the delta method. At the same time, the bias adjustment may not work well in the case in which a deterministic trend is present, and the method is not designed for the case

in which $\rho$ is exactly equal to one.

As an alternative method for correcting small sample bias, Andrews and Chen (1994) propose the use of an approximated median unbiased estimate of the coefficient on the lagged variable in level in the $ADF$ regression to simulate IRFs quantiles. The CI for the IRF is obtained by simulating the IRFs based on the corrected coefficients, and then taking the $(\alpha/2)\,th$ and $(1-\alpha/2)\,th$ quantiles of that distribution as the end points (cfr. Murray and Papell (2001)).[4]

Pesavento and Rossi (2005), Wright (2000) and Gospodinov (2004) rely on local to unity approximations of the largest root of the process to obtain approximations that perform better in small samples. Pesavento and Rossi (2005) propose a method that relies on a local to unity approximation to the asymptotic distribution of the IRF. The method is implemented by recognizing that, under the premise that $\rho$, (the largest root of $y_t$) is close to one, it can be modeled as local to unity, so that $\rho = 1+c/T$. Assuming that the lead time of the IRF is a fixed fraction of the sample size, so that $\frac{h}{T} \underset{T\to\infty}{\to} \delta$, the IRF can be approximated by $e^{c\delta}\theta(1)$. Although $c$, the local to unity parameter, cannot be consistently estimated, methods for constructing valid confidence intervals for $c$ are available (e.g., Stock, 1991) by simply inverting the ADF test for the null of $\rho = 1$. A confidence interval for the IRF can then be constructed by using the confidence interval for $c$, say $[c_L, c_u]$, and a consistent estimate of $\theta(1)$, as $\left[e^{c_L\delta}\hat{\theta}(1), e^{c_U\delta}\hat{\theta}(1)\right]$. As stated in Pesavento and Rossi (2005),

[4] We are grateful to C. Murray for providing the codes to implement the Andrews and Chen's method.

this method relies on the largest root being close to one and on the lead time being large relatively to the sample size. We should then not expect this method to work well at short horizons or when the process in (1) is strictly stationary.

Gospodinov (2004) notes that the localizing constant $c$ in autoregressive models can be consistently estimated under a sequence of null hypotheses that restrict the value of the impulse response at each horizon. As in Pesavento and Rossi (2005), the local to unity assumption together with the assumption that the lead time of the IRF is a fixed fraction of the sample size, allows him to derive an asymptotic distribution intended to approximate better the behavior of the LR test in models with one persistent root. Confidence bands for the IRF at each horizon can then be constructed by inverting the acceptance region of the LR test, $LR_T = T \ln (SSR_0/SSR)$, where $SSR_0$ and $SSR$ are the sum of squares of the restricted and the estimated residuals. By construction, inverting $LR_T$ will result in one sided confidence intervals. To construct two-sided confidence intervals, Gospodinov (2004) suggests inverting the statistics $LR_T^{\pm} = sgn [\psi_h (\hat{\rho}) - \psi_h (\widetilde{\rho})] \sqrt{LR_T}$ where $sgn (\cdot)$ denotes the sign of the expression in the brackets, $\psi (L) = a (L)^{-1} = (1 - \rho L)^{-1} \theta (L)$, and $\hat{\rho}$ and $\widetilde{\rho}$ are the unrestricted and the restricted estimates respectively. As in Pesavento and Rossi (2005), the key assumption for the validity of this approach is that $\rho$ is close to one. Even though the asymptotic distribution of the LR test is derived under the assumption that $\frac{h}{T} \underset{T \to \infty}{\to} \delta$, Gospodinov (2004) shows that the coverage of his one sided $LR_T$ test is good even for short horizons. Table 2 shows that, for the two-sided

test $\left(LR_T^{\pm}\right)$, the results may be sensitive to the horizon of the IRF.

Wright (2000) also relies on a local to unity approximation of the largest root and constructs a $(1-\alpha)$ confidence interval for $c$ (let's denote it by $C_\alpha$) by inverting the acceptance region of the ADF test. For each point $c_i$ in $C_\alpha$, a $(1-\alpha)$ confidence interval for the IRF can be computed by the delta method treating $c$ as fixed and running an autoregression on $(1-\rho_i L)\, y_t$ where $\rho_i = 1 + \frac{c_i}{T}$. Let this confidence interval be $[IRF_L(c_i), IRF_U(c_i)]$. By the Bonferroni inequality, the confidence interval constructed as $\left[\inf_{c_i \in C_\alpha} IRF_L(c_i), \sup_{c_i \in C_\alpha} IRF_U(c_i)\right]$ has coverage that is at least equal to $2(1-\alpha)-1$. By the nature of the Bonferroni inequality, Wright's (2000) method controls coverage in the sense that the the coverage will never be less than the nominal one, but, as we will see in our simulations and as it is shown in Wright (2000), it can be quite conservative. As Pesavento and Rossi's (2005) and Gospodinov's (2004), Wright's (2000) approach performs well only when $\rho$ is close to one, so that $y_t$ can be well approximated by a local to unity process.

In the next section, we compare the performance of these recent robust methods and compare them to the more traditional bootstrap and asymptotic normal approximation. An alternative method not considered here is Hansen's (1999) grid bootstrap. Gospodinov (2004) and Rossi (2005) already analyzed Hansen's (1999) small sample properties for confidence intervals for half-lives and IRFs, and show that the grid bootstrap is inferior in most reasonable situations. For this reason we do not include Hansen's methods here.

## 3. Monte Carlo Evidence on Coverage Accuracy

The goal of this section is to compare the empirical coverage probabilities of the methods described in the previous section with the traditional methods in a simple Monte Carlo experiment. The "traditional" methods include the standard bootstrap of Runkle (1987), and the asymptotically normal approximation of Lütkepohl (1991) when the practitioner decides to run the regression in levels or in first differences. We also consider the common procedure of deciding between levels and first differences based on the outcome of a unit root pre-test, where the pre-test can either be the standard ADF test or the more powerful ADF-GLS test. In this case, if the nominal coverage rate is 0.90, we do a first stage pre-test of size 5%, and then construct a 95% IRF confidence band in a second stage. If the two stages of the pre-test were independent, the final coverage should be roughly 0.90. Due to the lack of independence of the two steps, the total coverage is not 0.90, but in practice nobody really corrects for that.

The Monte Carlo design is as follows. Let the DGP be: $\Pi_{j=1}^{p}\left(1-\lambda_j L\right) y_t = \epsilon_t$, where $\epsilon_t \sim iidN\left(0,1\right)$, $\lambda_j$ are the possible roots of the process, and $\lambda_1 \equiv \rho = 1+c/T$ is the largest root. We abstract from small sample problems associated with the choice of the lag length (cfr. Ivanov and Kilian, 2005) by assuming that $p$ is known. All the methods fit an autoregression or and ADF regression (depending on the method) with an intercept. The nominal coverage is 0.90, $T = 100$, and the number of Monte Carlo replications is $5,000$ with the exception of Gospodinov's (2004) method, which is

computationally intensive. In the latter case, we only did $1,000$ replications. Runkle (1987) is implemented with 2,000 bootstrap replications, whereas Kilian's (1998a) bias corrected bootstrap is implemented with additional $2,000$ bootstrap replications and the "bootstrap after bootstrap" short-cut. For the levels, first difference and pre-test-based methods, we compute the coverage by simulating the normal asymptotic distribution of the parameters $5,000$ times and by taking the $5^{th}$ and $95^{th}$ percentile of the simulated IRFs.

We consider a variety of representative AR(2) processes, where $\rho = (0.99, 0.97, 0.95)$ and $\lambda_2 = (0, 0.4, 0.8)$, and the horizons of the IRF are $(5, 10, 20, 30)$. We chose to keep the Monte Carlo design simple enough to shed light on when each method's performance worsens, while at the same time rich enough to highlight the most important insights regarding the performance of the various methods. In fact, the simplicity coming from considering only two separate roots (one that captures the effects of persistence local to a unit root, and one that captures any additional stationary serial correlation) allows us to understand their respective roles without the complications coming from additional serial correlation, but, at the same time, is rich enough to understand the separate effects of increasing the persistence ($\rho$) and of increasing the additional serial correlation ($\lambda_2$). Tables 1-3 report the percentages of samples in which the true value of the IRF lays above ("up") and below ("low") the estimated CI's for each method for various values of $h$. We also report the value of $\delta$, the ratio between the IRF horizon and the sample size because our analysis focuses on small

samples, so the horizon per se is less important than its ratio to the available sample size. For example, in a sample of 100 monthly observations, a horizon of 12 months would corresponds to $\delta = 0.12$. The nominal value of each one-sided rejection rate in the Tables is 0.05, so that the total nominal coverage of the CI is 90%.

The results show that standard methods (levels, first differences, and pre-test-based methods) don't work well in general in the presence of a persistent root (see Table 1). The IRF confidence bands based on estimating an AR in levels slightly over reject and are highly asymmetric: almost all of the rejections happen on the upper tail. We say that an IRF is median unbiased if 50% of the distribution (pointwise in the horizon) lies above its estimate, and that a two-sided, $(1-\alpha)$ nominal CI for the IRF is median unbiased is the rejection frequencies on each tail are equal to $\alpha/2$. Therefore, the IRF confidence bands based on estimating an AR in levels are not median unbiased. This reflects the bias in the parameter $\rho$ estimated from a regression in levels. In fact, the estimate of this parameter will be downward biased in small samples, therefore over-rejecting in the region close to non-stationarity (i.e. the upper side of the CI).[5] Estimating an AR in first differences gives the correct coverage only when $\rho = 1$ (for brevity not reported). As $\rho$ moves away from unity, confidence intervals computed from the VAR in first differences start to behave poorly, with coverage that approaches zero as the horizon increases. Similarly, pre-tests show

[5] The rejection frequencies are based on a Monte Carlo approximation to the distribution of the IRF bands, and not on an analytic application of the delta method. The latter would perform much worse in practice, and was therefore not included.

bad coverage properties too, unless the root is exactly unity. However, the root is in general unknown to the researcher, thus limiting the practical usefulness of such approaches. In fact, for large enough (though less than one) values of $\rho$, pre-tests have low power to reject the hypothesis of a unit root, and select an AR in first difference most of the times. As expected, different pre-tests have significantly different coverage properties. The better coverage rate of the ADF-GLS test relative to the ADF test reflects the higher power of ADF-GLS test against alternatives that are close to one. Therefore, if a researcher has to choose between using an ADF pre-test and using an ADF-GLS pre-test, he should use the latter. As $\rho$ moves further away from unity, the pre-tests are able to reject the hypothesis of a unit root more often, and their coverage improves.

Methods designed to solve the problems of standard methods typically rely on better approximations to the small sample distribution of the IRFs, either by using the bootstrap or by using iterative methods. Results for such methods are reported in Table 2. Table 2 shows that Runkle's bootstrap does not provide an accurate approximation to the small sample distribution of the IRFs, and it will result in confidence intervals with coverage below the nominal level. This result might be surprising given that we know from Inoue and Kilian (2002) that the standard bootstrap approximation should work well for estimating AR(2) processes. However, here we are really interested in the small sample performance, and the bias is introduced by the inaccurate estimate of the largest root which plays a key role at making infer-

ence on IRF at horizons that are large relative to the sample size.[6] Also, the CIs are highly asymmetric, rejecting only on one side. Finally, as shown in Inoue and Kilian (2002), when the data generating process does include a constant or a time trend, Runkle (1987) bootstrap's performance is even worse. Andrews and Chen's (1994) method performs well in the close to stationary case and at short horizons, but quickly worsens as soon as the largest root approaches unity and the amount of serial correlation ($\lambda_2$) becomes non-negligible. The quality of the approximation, while quite erratic, seems to worsen as the horizon and the degree of additional serial correlation introduced by $\theta(L)$ increase. This is not surprising: Andrews and Chen (1994) also reported similar results (see their Table 2, last DGP). Furthermore, as Murray and Papell (2001) point out, this method is computationally very intensive. Kilian's (1998a) improved bootstrap method, on the other hand, works very well in terms of overall coverage. However, the tail probabilities of its bias-corrected CI are also highly asymmetric, and, thus, it might be unappealing if the objective is to obtain median unbiased CI's. Furthermore, it is well known that the performance of the bootstrap worsens if the DGP includes deterministic components (especially time trends), and its validity so far has only been established for weakly stationary processes. To summarize, we found that both Andrews and Chen (1994) and Runkle (1987) generally have coverage closer to the nominal level the smaller the largest root

[6] The argument is really similar to the small sample properties of inference on AR(2) processes based on asymptotic approximations, as in Sims et al. (1990): both AR(2) coefficients may have standard normal asymptotic distributions, but at long horizons the largest root drives the asymptotic distribution, which becomes non-standard, as explained in Phillips (1998) and Rossi (2005).

and the degree of additional serial correlation are.

Methods that rely on the local to unity approximation seem to perform better than Andrews and Chen (1994) and Runkle's bootstrap (see Table 3). Wright's (2000) method is conservative, but effectively controls coverage. Also, in unreported results, we found that its performance is poor in the presence of an explosive root (since it imposes a unit upper bound on "$\rho$"). Gospodinov's (2004) one-sided test, labeled "*Gospodinov (I)*" in Table 3, has coverage that is close to the nominal level for various horizons, even short ones. However, by construction the test will reject only on one side and never reject on the other. Gospodinov's two sided test, labeled "*Gospodinov (II)*", can be used if the goal is to design a symmetric CI. Unfortunately, the inversion of $LR_T^{\pm}$ can result in a coverage below 90%, at times as low as 70%, especially for short horizons and when the second root of the process is large. Finally, Pesavento and Rossi's (2005) method (see the last columns of Table 3) has coverage close to nominal at medium to long horizons (that is, $\delta \geq 0.10$) and it is median unbiased, with similar rejection probabilities on both sides. However, by construction, its coverage is not accurate at short horizons, as discussed in Section 2.[7]

To summarize, when the largest root of an AR process is close to unity, levels, first differences, pre-tests and Runkle's bootstrap should in general be avoided. If there is evidence of high persistence and serial correlation beyond the largest root,

[7]Since in this paper we were interested mainly in comparing the performances of the various methods at long horizons, we implemented Pesavento and Rossi's (2005) long horizon method, which is designed to work well at long horizons but not necessarilty at short horizons. For an extension that improves the performance of this method at short horizons, see Pesavento and Rossi (2005).

Andrews and Chen (1994) might perform poorly. The researcher instead may rely on Kilian (1998a), Gospodinov (2004), Wright (2000) or Pesavento and Rossi (2005). Kilian's (1998a) method is a good choice if the researcher knows the process does not have an exact unit root, nor an explosive one, and if the researcher cares about total coverage, although the CI's will not be median unbiased. If the researcher wants to be completely agnostic about the largest root and use a method that is likely to produce IRF bands with coverage close to the nominal one whether the root is one or close to one, then Wright (2000), Gospodinov's (2004) method (I) or Pesavento and Rossi (2005) should be used as follows. If the researcher is only interested in one-sided IRF bands and is satisfied with median biased CIs, then Gospodinov's (2004) one-sided method could be used, which works well at both short and long horizons. In the more common situation in which the researcher is interested in two-sided IRF bands, Wright's (2000) and Pesavento and Rossi (2005) are the two available options, with the following trade-offs. Wright's (2000) is a good choice if the researcher is interested in CIs at both short and long horizons and is satisfied with a conservative CI. Pesavento and Rossi (2005) provides median unbiased CIs no matter whether the root is unity or close to unity (even mildly explosive),[8] but is advised in general only for horizons that are bigger than ten percent of the sample size.[9]

[8] In unreported results, we also compared the length of the confidence bands of the different methods we considered above. To make a fair comparison, since some methods impose an upper bound of unity on the largest root, we impose the upper bound on all methods. We found that there are not huge differences between Wright (2001), Kilian (1998a) and Pesavento and Rossi (2004).

[9] Pesavento and Rossi's (2005) "robust" method could be used instead if the researcher is interested in short horizons. However, that method is conservative. See Pesavento and Rossi (2005) for more details.

## 4. Conclusions

This paper is a guide for empiricists who face the problem of computing confidence bands for univariate impulse response functions when variables are highly persistent. We confirm previous results and show that traditional methods (asymptotic normal approximations and the standard Runkle's (1987) bootstrap) may be highly unreliable. Even inference based on unit root pre-tests with high power will not result in inference with the correct rejection probabilities. We compare a number of methods that have been recently developed to provide more robust approximations in the presence of variables with roots equal or slightly less than one. The Monte Carlo results show that, among the preferred methods, Kilian (1998a) is a good choice for a researcher who knows that the process does not have a unit root, nor an explosive one, and cares only about total coverage, as it may not deliver median unbiased CIs. Wright's (2000) is a good choice if the researcher is interested in CIs at both short and long horizons and is satisfied with a conservative CI. Pesavento and Rossi's (2005) method tends to provide median unbiased CIs with accurate coverage no matter whether the root is unity or close to unity (even mildly explosive), but is advised in general only for horizons that are bigger than ten percent of the sample size. Gospodinov's (2004) method (I) can be used to construct one sided confidence intervals, but it performs well at both short and long horizons.

## 5. References

Andrews, Donald (1993), "Exactly Median_Unbiased Estimation of First Order Autoregressive/Unit Root Models", *Econometrica* 61(1), 139-165.

Andrews, Donald, and Hong-Yuan Chen (1994), "Approximately Median-Unbiased Estimation of Autoregressive Models", *Journal of Business and Economic Statistics* 12(2), 187-204.

Ashley, Richard, and Randal Verbrugge (1999), "To Difference or Not to Difference: a Monte Carlo Investigation of Inference in Vector Autoregressive Models", VPI Economics Department Working Paper E99-15.

Busetti, Fabio, Silvia Fabiani, and Andrew Harvey (2005), "Convergence of Prices and Rates of Inflation", *mimeo*, Bank of Italy.

Campbell, John Y., and Gregory N. Mankiw, "Are Output Fluctuations Transitory?", *Quarterly Journal of Economics* 102(4), 857-880.

Cavanagh, Christopher L., Graham Elliott, and James H. Stock (1995), "Inference in Models with nearly Integrated Regressors", *Econometric Theory* 11(5), 1131-47.

Diebold, Frank and Glenn D. Rudebusch (1989), "Long Memory and Persistence in Aggregate Output", *Journal of Monetary Economics* 24, 189-209.

Elliott, Graham (1998), "On the Robustness of Cointegration Methods When Regressors Almost Have Unit Roots", *Econometrica* 66(1), 149-158.

Gospodinov, Nikolay (2004), "Asymptotic Confidence Intervals for Impulse Responses of Near-integrated Processes", *Econometrics Journal* 7(2), 505-527.

Hansen, Bruce (1999), "Bootstrapping the Autoregressive Model", *Review of Economics and Statistics* 81(4), 594-607.

Inoue, Atsushi, and Lutz Kilian (2002), "Bootstrapping Autoregressive Processes with Possible Unit Roots", *Econometrica* 70(1), 377-391.

Ivanov, Ventzislav, and Lutz Kilian (2005), "A Practitioner's Guide to Lag-Order Selection for VAR Impulse Response Analysis", *Studies in Nonlinear Dynamics and Econometrics* 9(1).

Kilian, Lutz (1998(a)), "Small-sample Confidence Intervals for Impulse Response Functions", *Review of Economics and Statistics* 80, 218-230.

Kilian, Lutz (1998(b)), "Confidence Intervals for Impulse Responses under Departures from Normality", *Econometric Reviews* 17, 1-29.

Kilian, Lutz, and Paolo-Li Chang (2000), "How accurate are Confidence Intervals for Impulse Responses in Large VAR Models?", *Economics Letters* 69, 299-307.

Kilian, Lutz, and Tao Zha (2002), "Quantifying the Uncertainty about the Half-life of Deviations from PPP", *Journal of Applied Econometrics* 17(2).

Lopez, Claude, Christian Murray and David Papell (2005), "State of the Art Unit Root Tests and Purchasing Power Parity,", *Journal of Money, Credit and Banking* 37, 361-369.

Lütkepohl, Helmut (1990), "Asymptotic Distribution of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models", *Review of Economics and Statistics* 72(1), 116-125.

Murray, Chris, and David Papell (2002), "The Purchasing Power Parity Persistence Paradigm", *Journal of International Economics* 56, 1-19.

Pesavento, Elena, and Barbara Rossi (2005), "Small Sample Confidence Intervals for Multivariate IRFs at Long Horizons", *Journal of Applied Econometrics*, forthcoming.

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

Rossi, Barbara (2005), "Confidence Intervals for Half-Life Deviations From PPP", *Journal of Business and Economic Statistics* 23(4), 432-442.

Runkle, David E. (1987), "Vector Autoregression and Reality", *Journal of Business and Economic Statistics* 5, 437-442.

Sims, Christopher, James H. Stock and Mark W. Watson (1990), "Inference in linear time series models with some unit roots", *Econometrica* 58, 113-144.

Wright, Jonathan (2000), "Confidence Intervals for Univariate Impulse Responses with a Near Unit Root", *Journal of Business and Economic Statistics* 18(3), 368-373.

**Table 1: Comparison of coverage of IRF confidence bands: Traditional asymptotic methods (Lütkepohl) and pre-tests**

| | | | | Levels | | Differences | | Pre-test: ADF | | ADF-GLS | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | $\lambda_2$ | $h$ | $\delta$ | low | up | low | up | low | up | low | up |
| 0.99 | 0 | 5 | 0.05 | 0.018 | 0.125 | 0.094 | 0.031 | 0.039 | 0.054 | 0.049 | 0.028 |
| | | 10 | 0.1 | 0.017 | 0.108 | 0.207 | 0.009 | 0.096 | 0.050 | 0.127 | 0.021 |
| | | 20 | 0.2 | 0.018 | 0.101 | 0.624 | 0.000 | 0.339 | 0.048 | 0.477 | 0.018 |
| | | 30 | 0.3 | 0.020 | 0.100 | 0.942 | 0.000 | 0.592 | 0.049 | 0.828 | 0.019 |
| | 0.4 | 5 | 0.05 | 0.029 | 0.119 | 0.068 | 0.046 | 0.032 | 0.051 | 0.038 | 0.035 |
| | | 10 | 0.1 | 0.019 | 0.122 | 0.111 | 0.027 | 0.048 | 0.053 | 0.068 | 0.027 |
| | | 20 | 0.2 | 0.018 | 0.106 | 0.317 | 0.004 | 0.144 | 0.047 | 0.214 | 0.018 |
| | | 30 | 0.3 | 0.021 | 0.101 | 0.608 | 0.001 | 0.327 | 0.045 | 0.470 | 0.016 |
| | 0.8 | 5 | 0.05 | 0.107 | 0.047 | 0.143 | 0.028 | 0.084 | 0.018 | 0.091 | 0.014 |
| | | 10 | 0.1 | 0.026 | 0.115 | 0.058 | 0.054 | 0.024 | 0.054 | 0.033 | 0.033 |
| | | 20 | 0.2 | 0.015 | 0.141 | 0.090 | 0.041 | 0.034 | 0.066 | 0.048 | 0.036 |
| | | 30 | 0.3 | 0.015 | 0.131 | 0.171 | 0.024 | 0.071 | 0.059 | 0.104 | 0.030 |
| 0.97 | 0 | 5 | 0.05 | 0.014 | 0.132 | 0.285 | 0.003 | 0.111 | 0.061 | 0.173 | 0.022 |
| | | 10 | 0.1 | 0.014 | 0.106 | 0.900 | 0.000 | 0.455 | 0.054 | 0.724 | 0.019 |
| | | 20 | 0.2 | 0.019 | 0.099 | 1.000 | 0.000 | 0.547 | 0.052 | 0.862 | 0.018 |
| | | 30 | 0.3 | 0.021 | 0.097 | 1.000 | 0.000 | 0.548 | 0.049 | 0.862 | 0.019 |
| | 0.4 | 5 | 0.05 | 0.022 | 0.129 | 0.123 | 0.018 | 0.045 | 0.057 | 0.069 | 0.026 |
| | | 10 | 0.1 | 0.018 | 0.120 | 0.465 | 0.002 | 0.201 | 0.058 | 0.311 | 0.022 |
| | | 20 | 0.2 | 0.017 | 0.109 | 0.992 | 0.000 | 0.554 | 0.054 | 0.844 | 0.021 |
| | | 30 | 0.3 | 0.022 | 0.104 | 1.000 | 0.000 | 0.563 | 0.054 | 0.856 | 0.021 |
| | 0.8 | 5 | 0.05 | 0.087 | 0.053 | 0.149 | 0.024 | 0.070 | 0.023 | 0.092 | 0.014 |
| | | 10 | 0.1 | 0.025 | 0.126 | 0.107 | 0.034 | 0.033 | 0.053 | 0.057 | 0.028 |
| | | 20 | 0.2 | 0.015 | 0.130 | 0.351 | 0.006 | 0.143 | 0.061 | 0.226 | 0.025 |
| | | 30 | 0.3 | 0.016 | 0.117 | 0.809 | 0.000 | 0.410 | 0.056 | 0.636 | 0.024 |
| 0.95 | 0 | 5 | 0.05 | 0.011 | 0.127 | 0.631 | 0.001 | 0.210 | 0.070 | 0.402 | 0.031 |
| | | 10 | 0.1 | 0.014 | 0.106 | 1.000 | 0.000 | 0.430 | 0.058 | 0.784 | 0.027 |
| | | 20 | 0.2 | 0.018 | 0.094 | 1.000 | 0.000 | 0.432 | 0.051 | 0.785 | 0.023 |
| | | 30 | 0.3 | 0.019 | 0.092 | 1.000 | 0.000 | 0.431 | 0.048 | 0.785 | 0.022 |
| | 0.4 | 5 | 0.05 | 0.016 | 0.148 | 0.210 | 0.010 | 0.056 | 0.073 | 0.115 | 0.033 |
| | | 10 | 0.1 | 0.014 | 0.120 | 0.870 | 0.000 | 0.342 | 0.064 | 0.634 | 0.028 |
| | | 20 | 0.2 | 0.017 | 0.107 | 1.000 | 0.000 | 0.429 | 0.057 | 0.783 | 0.022 |
| | | 30 | 0.3 | 0.020 | 0.102 | 1.000 | 0.000 | 0.429 | 0.053 | 0.783 | 0.020 |
| | 0.8 | 5 | 0.05 | 0.066 | 0.060 | 0.147 | 0.020 | 0.058 | 0.025 | 0.079 | 0.011 |
| | | 10 | 0.1 | 0.020 | 0.122 | 0.163 | 0.018 | 0.047 | 0.058 | 0.082 | 0.023 |
| | | 20 | 0.2 | 0.017 | 0.115 | 0.771 | 0.000 | 0.314 | 0.059 | 0.550 | 0.025 |
| | | 30 | 0.3 | 0.019 | 0.098 | 0.999 | 0.000 | 0.484 | 0.051 | 0.810 | 0.021 |

Percentage of times the true IRF lays above or below the CI. Nominal values are 5% on each side.

T=100

**Table 2: Comparison of coverage of IRF bands: Methods designed to provide better approximations to the small sample distribution of the IRFs**

| | | | | Runkle | | Andrews-Chen | | Kilian | |
|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | $\lambda_2$ | $h$ | $\delta$ | low | up | low | up | low | up |
| 0.99 | 0 | 5 | 0.05 | 0.01 | 0.15 | 0.00 | 0.15 | 0.02 | 0.10 |
| | | 10 | 0.1 | 0.00 | 0.21 | 0.00 | 0.17 | 0.00 | 0.12 |
| | | 20 | 0.2 | 0.00 | 0.22 | 0.00 | 0.16 | 0.00 | 0.13 |
| | | 30 | 0.3 | 0.00 | 0.22 | 0.00 | 0.16 | 0.00 | 0.13 |
| | 0.4 | 5 | 0.05 | 0.01 | 0.13 | 0.01 | 0.13 | 0.03 | 0.06 |
| | | 10 | 0.1 | 0.00 | 0.18 | 0.00 | 0.17 | 0.01 | 0.08 |
| | | 20 | 0.2 | 0.00 | 0.20 | 0.00 | 0.16 | 0.00 | 0.09 |
| | | 30 | 0.3 | 0.00 | 0.19 | 0.00 | 0.15 | 0.00 | 0.10 |
| | 0.8 | 5 | 0.05 | 0.00 | 0.16 | 0.02 | 0.06 | 0.02 | 0.07 |
| | | 10 | 0.1 | 0.00 | 0.20 | 0.00 | 0.14 | 0.02 | 0.07 |
| | | 20 | 0.2 | 0.00 | 0.24 | 0.00 | 0.18 | 0.01 | 0.09 |
| | | 30 | 0.3 | 0.00 | 0.24 | 0.00 | 0.18 | 0.00 | 0.10 |
| 0.97 | 0 | 5 | 0.05 | 0.00 | 0.14 | 0.00 | 0.14 | 0.02 | 0.11 |
| | | 10 | 0.1 | 0.00 | 0.17 | 0.00 | 0.14 | 0.00 | 0.12 |
| | | 20 | 0.2 | 0.00 | 0.17 | 0.00 | 0.13 | 0.00 | 0.12 |
| | | 30 | 0.3 | 0.00 | 0.17 | 0.00 | 0.13 | 0.00 | 0.12 |
| | 0.4 | 5 | 0.05 | 0.01 | 0.14 | 0.01 | 0.13 | 0.03 | 0.08 |
| | | 10 | 0.1 | 0.00 | 0.18 | 0.00 | 0.13 | 0.01 | 0.10 |
| | | 20 | 0.2 | 0.00 | 0.17 | 0.00 | 0.11 | 0.00 | 0.11 |
| | | 30 | 0.3 | 0.00 | 0.16 | 0.00 | 0.10 | 0.00 | 0.11 |
| | 0.8 | 5 | 0.05 | 0.00 | 0.15 | 0.03 | 0.06 | 0.03 | 0.07 |
| | | 10 | 0.1 | 0.00 | 0.19 | 0.01 | 0.12 | 0.02 | 0.08 |
| | | 20 | 0.2 | 0.00 | 0.20 | 0.01 | 0.09 | 0.01 | 0.10 |
| | | 30 | 0.3 | 0.00 | 0.17 | 0.00 | 0.05 | 0.00 | 0.10 |
| 0.95 | 0 | 5 | 0.05 | 0.00 | 0.12 | 0.00 | 0.15 | 0.02 | 0.09 |
| | | 10 | 0.1 | 0.00 | 0.14 | 0.00 | 0.14 | 0.01 | 0.11 |
| | | 20 | 0.2 | 0.00 | 0.14 | 0.00 | 0.12 | 0.01 | 0.11 |
| | | 30 | 0.3 | 0.00 | 0.14 | 0.00 | 0.12 | 0.01 | 0.11 |
| | 0.4 | 5 | 0.05 | 0.01 | 0.13 | 0.01 | 0.12 | 0.03 | 0.08 |
| | | 10 | 0.1 | 0.00 | 0.15 | 0.00 | 0.11 | 0.01 | 0.10 |
| | | 20 | 0.2 | 0.00 | 0.14 | 0.00 | 0.09 | 0.00 | 0.10 |
| | | 30 | 0.3 | 0.00 | 0.13 | 0.00 | 0.09 | 0.00 | 0.10 |
| | 0.8 | 5 | 0.05 | 0.00 | 0.14 | 0.03 | 0.06 | 0.03 | 0.07 |
| | | 10 | 0.1 | 0.00 | 0.17 | 0.01 | 0.10 | 0.03 | 0.08 |
| | | 20 | 0.2 | 0.00 | 0.17 | 0.02 | 0.06 | 0.01 | 0.10 |
| | | 30 | 0.3 | 0.00 | 0.14 | 0.01 | 0.03 | 0.00 | 0.10 |

Notes: as per Table 1

**Table 3: Comparison of coverage of IRF bands: Methods based on local to unity approximations**

| | | | | Wright | | Gospodinov (I) | | Gospodinov (II) | | Pesavento-Rossi | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | $\lambda_2$ | $h$ | $\delta$ | low | up | low | up | low | up | low | up |
| 0.99 | 0 | 5 | 0.05 | 0.01 | 0.06 | 0.00 | 0.10 | 0.07 | 0.05 | 0.05 | 0.07 |
| | | 10 | 0.1 | 0.00 | 0.07 | 0.00 | 0.10 | 0.07 | 0.05 | 0.05 | 0.06 |
| | | 20 | 0.2 | 0.01 | 0.05 | 0.00 | 0.10 | 0.07 | 0.06 | 0.05 | 0.06 |
| | | 30 | 0.3 | 0.02 | 0.05 | 0.00 | 0.10 | 0.07 | 0.08 | 0.06 | 0.06 |
| | 0.4 | 5 | 0.05 | 0.01 | 0.04 | 0.00 | 0.11 | 0.16 | 0.05 | 0.07 | 0.14 |
| | | 10 | 0.1 | 0.00 | 0.06 | 0.00 | 0.09 | 0.10 | 0.05 | 0.05 | 0.06 |
| | | 20 | 0.2 | 0.00 | 0.05 | 0.00 | 0.09 | 0.07 | 0.05 | 0.05 | 0.05 |
| | | 30 | 0.3 | 0.01 | 0.04 | 0.00 | 0.09 | 0.07 | 0.05 | 0.06 | 0.05 |
| | 0.8 | 5 | 0.05 | 0.03 | 0.01 | 0.00 | 0.07 | 0.29 | 0.03 | 0.31 | 0.09 |
| | | 10 | 0.1 | 0.00 | 0.05 | 0.00 | 0.09 | 0.21 | 0.05 | 0.06 | 0.17 |
| | | 20 | 0.2 | 0.00 | 0.08 | 0.00 | 0.08 | 0.12 | 0.05 | 0.04 | 0.10 |
| | | 30 | 0.3 | 0.00 | 0.08 | 0.00 | 0.08 | 0.09 | 0.05 | 0.05 | 0.07 |
| 0.97 | 0 | 5 | 0.05 | 0.00 | 0.04 | 0.00 | 0.09 | 0.08 | 0.05 | 0.05 | 0.05 |
| | | 10 | 0.1 | 0.01 | 0.04 | 0.00 | 0.09 | 0.08 | 0.05 | 0.06 | 0.05 |
| | | 20 | 0.2 | 0.04 | 0.05 | 0.00 | 0.09 | 0.08 | 0.05 | 0.06 | 0.05 |
| | | 30 | 0.3 | 0.05 | 0.05 | 0.00 | 0.09 | 0.08 | 0.06 | 0.06 | 0.05 |
| | 0.4 | 5 | 0.05 | 0.00 | 0.05 | 0.00 | 0.09 | 0.14 | 0.05 | 0.05 | 0.11 |
| | | 10 | 0.1 | 0.00 | 0.03 | 0.00 | 0.07 | 0.11 | 0.04 | 0.05 | 0.06 |
| | | 20 | 0.2 | 0.02 | 0.04 | 0.00 | 0.08 | 0.10 | 0.04 | 0.06 | 0.06 |
| | | 30 | 0.3 | 0.04 | 0.05 | 0.00 | 0.08 | 0.10 | 0.04 | 0.06 | 0.06 |
| | 0.8 | 5 | 0.05 | 0.02 | 0.01 | 0.00 | 0.06 | 0.30 | 0.03 | 0.25 | 0.08 |
| | | 10 | 0.1 | 0.00 | 0.05 | 0.00 | 0.07 | 0.22 | 0.04 | 0.04 | 0.13 |
| | | 20 | 0.2 | 0.00 | 0.03 | 0.00 | 0.07 | 0.15 | 0.03 | 0.04 | 0.07 |
| | | 30 | 0.3 | 0.00 | 0.02 | 0.00 | 0.06 | 0.12 | 0.03 | 0.05 | 0.07 |
| 0.95 | 0 | 5 | 0.05 | 0.00 | 0.04 | 0.00 | 010 | 0.07 | 0.05 | 0.04 | 0.05 |
| | | 10 | 0.1 | 0.02 | 0.05 | 0.00 | 0.10 | 0.07 | 0.05 | 0.05 | 0.05 |
| | | 20 | 0.2 | 0.04 | 0.05 | 0.00 | 0.10 | 0.07 | 0.05 | 0.05 | 0.06 |
| | | 30 | 0.3 | 0.05 | 0.06 | 0.00 | 0.09 | 0.07 | 0.06 | 0.05 | 0.06 |
| | 0.4 | 5 | 0.05 | 0.00 | 0.04 | 0.00 | 0.08 | 0.12 | 0.04 | 0.03 | 0.09 |
| | | 10 | 0.1 | 0.00 | 0.03 | 0.00 | 0.08 | 0.12 | 0.03 | 0.04 | 0.06 |
| | | 20 | 0.2 | 0.03 | 0.04 | 0.00 | 0.08 | 0.10 | 0.04 | 0.05 | 0.06 |
| | | 30 | 0.3 | 0.05 | 0.05 | 0.00 | 0.08 | 0.13 | 0.02 | 0.06 | 0.06 |
| | 0.8 | 5 | 0.05 | 0.01 | 0.02 | 0.00 | 0.06 | 0.27 | 0.03 | 0.18 | 0.08 |
| | | 10 | 0.1 | 0.00 | 0.04 | 0.00 | 0.06 | 0.20 | 0.04 | 0.03 | 0.11 |
| | | 20 | 0.2 | 0.00 | 0.02 | 0.00 | 0.06 | 0.16 | 0.03 | 0.04 | 0.07 |
| | | 30 | 0.3 | 0.01 | 0.02 | 0.00 | 0.05 | 0.16 | 0.03 | 0.06 | 0.06 |

Notes: as per Table 1.