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

Elena Pesavento — Emory University

Barbara Rossi — Duke University

June 2006

Abstract.

This paper provides 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 the methods proposed in Kilian (1999), 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 both Lutz Kilian and Nikolay Gospodinov for sharing their codes, and for providing extraordinary help in implementing them, and an anonymous referee for useful comments. 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 are widely used tools for evaluating the effect of shocks on macroeconomic time series. Confidence intervals for impulse response functions 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 specified in levels or in differences. The objective of this paper is to provide a comprehensive comparison of alternative methods available in the literature. We provide practical recommendations to empirical macroeconomists on how to conduct inference on impulse responses in the presence of highly persistent data.

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 (1999), 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 impulse response 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 to or slightly less than one. In the context of impulse response functions, in particular, there have been some advances in the attempt to solve the problem of constructing confidence intervals with coverage rates that are close to or bounded by the nominal rate even when variables are highly persistent. For example, Andrews and Chen (1994) propose a method to estimate the parameters of an AR process from which median unbiased estimates for the impulse response functions 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 impulse response functions' distribution. Wright (2000) proposes to construct confidence bands based on Bonferroni bounds, while Gospodinov (2004) relies on the inversion of a likelihood ratio test. Both Wright (2000) and Gospodinov (2004) focus on univariate impulse response functions. Finally, Pesavento and Rossi (2005) derive analytic approximations to multivariate confidence intervals for impulse response functions 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 impulse response functions in the presence of highly persistent processes. Although robust methods for inference for impulse response functions 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 different approaches, so we can provide some guidance to practitioners.

This paper aims at providing such guidance. We focus on impulse response functions in univariate models. Although it is common to think of impulse responses in

[1] Some examples are Murray and Papell (2002), Rossi (2005), Lopez, Murray and Papell (2005) for half-life deviations from Purchasing Power Parity, and Wright (2000) for impulse response of aggregrate output.

the context of multivariate models, there are relevant empirical applications in univariate models. Typical examples are the evaluation of the persistence of the effects of shocks to aggregate output (Diebold and Rudebusch, 1989, and Campbell and Mankiw, 1987) or to the real exchange rate (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 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 ε_t is a mean zero independent and identically distributed (i.i.d.) random variable with variance equal to σ^2, $t = 1, 2, ...T$, where T is the sample size.[3] For sake of simplicity of exposition, we ignore deterministic terms which are irrelevant in

[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.

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

$$(1 - \rho L)\, y_t = \theta(L)\, \varepsilon_t, \qquad (2)$$

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 possibly equal to one. Define the impulse response function at horizon h as the effect of a shock on y_{t+h} of size σ in ε_t : $IRF_h = \frac{\partial y_{t+h}}{\partial \varepsilon_t}$.

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 approximation may provide confidence intervals for IRF_h with poor coverage properties. Several methods have been proposed 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 impulse response function 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 when a deterministic trend is present, and the method is not designed for the case in which ρ is exactly equal to one.

As an alternative method for correcting small sample bias, Andrews and Chen (1994) propose the use of an approximately median unbiased estimate of the coefficient on the lagged level variable in the ADF regression. The additional coefficients on the lagged differences of the variable are estimated along with the coefficient on the lagged level variable by an iterative procedure. The confidence interval for the impulse response is obtained by simulating the impulse response functions based on the bias-corrected coefficients, and then taking the $(\alpha/2)\,th$ and $(1-\alpha/2)\,th$ quantiles of that distribution as the end points.[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 improved small-sample approximations. Pesavento and Rossi (2005) propose a method that relies on a local to unity approximation to the asymptotic distribution of the impulse response function. The method is implemented by recognizing that, under the premise that ρ, (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 impulse response function is a fixed fraction of the sample size, so that $\frac{h}{T} \rightarrow \delta$ as $T \rightarrow \infty$, IRF_h 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). After inverting the ADF test for the null of $\rho = 1$, a confidence interval for

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

the impulse response function can 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), this method relies on the largest root being close to one and on the lead time being large relative to the sample size. We should 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 to be equal to some pre-specified value. As in Pesavento and Rossi (2005), the local to unity assumption together with the assumption that the lead time of the impulse response function is a fixed fraction of the sample size, allows him to derive an asymptotic distribution of the LR test in models with one persistent root. Confidence bands for the impulse response at each horizon can then be constructed by inverting the non-rejection 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 unrestricted 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 $LR_T^{\pm} = sgn\left[\psi_h(\hat{\rho}) - \psi_h(\widetilde{\rho})\right]\sqrt{LR_T}$ 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 ρ is close to one. Even though the asymptotic distribution of the LR test is derived under the assumption that $\frac{h}{T} \to \delta$ as $T \to \infty$, Gospodinov (2004) shows that the coverage of his one-sided LR_T test is good even for short horizons. Our results show that, for the two-sided test $\left(LR_T^{\pm}\right)$, the results may be sensitive to the horizon of the response.

Wright (2000) also relies on a local to unity approximation of the largest root and constructs a $(1-\alpha)$ confidence interval for c (denoted by C_α) by inverting the non-rejection region of the ADF test. For each point c_i in C_α, a $(1-\alpha)$ confidence interval for the responses 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 $\left[IRF_L\left(c_i\right), IRF_U\left(c_i\right)\right]$. By the Bonferroni inequality, the confidence interval constructed as $\left[\inf_{c_i \in C_\alpha} IRF_L\left(c_i\right), \sup_{c_i \in C_\alpha} IRF_U\left(c_i\right)\right]$ has coverage of at least $2\,(1-\alpha)-1$. By the nature of the Bonferroni inequality, Wright's (2000) method controls coverage in the sense that the coverage rate will never be less than the nominal one, but, as shown in Wright (2000), this method can be quite conservative. A common feature of Pesavento and Rossi's (2005), Gospodinov's (2004), and Wright's (2000) approaches is that they perform well only when ρ is close to one, so that y_t is well approximated by a local to unity process.

In the next section, we compare the performance of these 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 impulse response functions, and showed that the grid bootstrap is inferior in most reasonable situations. For this reason we do not include Hansen's method 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 traditional methods in a simple Monte Carlo experiment. The "traditional" methods include the standard bootstrap of Runkle (1987), and the asymptotic normal approximation of Lütkepohl (1991) in which 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% impulse response 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 $\epsilon_t \sim iidN\left(0,1\right)$, λ_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 (see Ivanov and Kilian, 2005) by assuming that p is known. All the methods fit an autoregression with either an intercept (Tables 1-3) or an intercept and a linear time trend (Table 4). 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. Kilian's (1998a) bias corrected bootstrap is implemented by using asymptotic closed form solutions for the bias estimates – see Kilian (1999) and Inoue and Kilian (2002b). Both Runkle (1987) and Kilian (1998a) are implemented with 1,000 bootstrap replications. 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 impulse responses.

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 impulse response function 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. Our framework is rich enough to show the separate effects of increasing the persistence (ρ) and of increasing the additional serial correlation (λ_2). Tables 1-3 report the percentages of samples in which the true value of the impulse response function lays above ("up") and below ("low") the estimated confidence intervals for each method for various values of h when a constant is included in the estimation. We also report the value of δ (the ratio between the impulse response function horizon and the 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 probability of the confidence interval is 90%.

The results show that standard methods (levels, first differences, and pre-test-based methods) do not work well in general in the presence of a persistent root (see Table 1). The impulse response 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 impulse response function is median unbiased if 50% of the probability mass of the impulse response distribution (pointwise in the horizon) lies above its estimate, and that a two-sided, $(1-\alpha)$ nominal confidence interval for the impulse response function is median unbiased if the rejection frequencies on each tail are equal to $\alpha/2$. Therefore, the impulse response confidence bands based on estimating an AR in levels are not median unbiased. This reflects the small-sample bias in the parameters estimated from the regression in levels and the non linearity of the impulse response estimator in the slope parameters. In fact, the estimate of this parameter will be typically downward biased in small samples, therefore over-rejecting in the region close to non-stationarity (i.e., the upper side of the confidence interval).[5] Estimating an AR in first differences gives the correct coverage only when $\rho = 1$. As ρ moves away from unity, confidence intervals computed from the VAR in first differences start to behave poorly, with coverage rates that approach zero as the horizon increases. Similarly, pre-test based intervals have poor coverage properties too, unless the root is exactly unity. Since the root is in general unknown to the researcher, this fact limits the practical usefulness of such approaches. In fact, for large enough values of ρ (though less than one), pre-tests have low power to reject the hypothesis of a unit root, and select an AR in first differences most of the times. As expected, different pre-tests imply very 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. Unreported results show that as ρ 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 impulse response functions, 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

[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.

provide an accurate approximation to the small sample distribution of the impulse response functions, and it will result in confidence intervals with coverage well below the nominal level. This result might be surprising given that we know from Inoue and Kilian (2002a) that the standard bootstrap approximation should work well for slope parameters in AR(2) processes. However, here we are interested in inference on non-linear functions of slope parameters, for which much larger sample sizes are required for accurate bootstrap inference at a given horizon. Also, the confidence intervals are highly asymmetric, rejecting only on one side. When in addition to a constant, the model includes also a trend, Runkle's (1987) bootstrap performance is even worse (see Inoue and Kilian, 2002a). In contrast, Andrews and Chen's (1994) method performs well in the close to stationary case with no serial correlation, but quickly worsens as soon as the largest root approaches unity. The quality of the approximation, while quite erratic, seems to worsen as the degree of additional serial correlation introduced by λ_2 increases. 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 as long as the largest root less than or equal to 0.95. For ρ around 0.97, the coverage is still approximately correct as long as there is not much additional serial correlation. However, its performance worsens as the persistence increases further and, in fact, its validity so far has only been established for weakly stationary processes; the tail probabilities of its bias-corrected confidence intervals are also highly asymmetric, and, thus, it might be unappealing if the objective is to obtain median unbiased confidence intervals. To summarize, we found that the overall performance of Andrews and Chen's (1994) and Runkle's (1987) methods is not satisfactory, while the use of the Kilian's (1998a, 1999) bias correction provides good overall coverage in models for all but the most persistent processes.

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 "ρ"). 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. Alternatively, Gospodinov's two-sided test, labeled "*Gospodinov (II)*", can be used if the goal is to design a symmetric confidence interval. 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.[6]

Unreported results show that in the presence of an exact unit root the performance of the VAR in levels and the bootstrap methods further worsen, as we would expect, whereas the VAR in differences and the pre-tests have the correct size. Wright's (2000) method is less conservative and Pesavento and Rossi (2005) and Gospodinov (2004) behave similarly to the $\rho = 0.99$ case.

When a time trend is also included in the estimation, traditional methods with or without bias correction are known to perform poorly in small samples (Kilian, 1998a, Hansen, 1999, and Inoue and Kilian, 2002a) so we do not report their rejections rates for this case. Table 4 shows that methods that rely on local to unity approximations are able to control size even in the presence of a deterministic trend. Wright (2000) has coverage still bounded by the nominal rate and it is one-sided, but now slightly less conservative. Gospodinov's (2004) one-sided test has coverage close to 90% at all horizons as in the case for an intercept, but the two-sided test is now conservative and asymmetric especially at short horizons, when there are no short run dynamics and when the largest root is very close to one. Finally, the coverage of Pesavento and Rossi (2005) is similar to the case with only an intercept.

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

[6] 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 necessarily at short horizons. For an extension that improves the performance of this method at short horizons, see Pesavento and Rossi (2005).

there is evidence of high persistence and serial correlation beyond the largest root, 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 confidence intervals will not be median unbiased. If the researcher wants a method that is likely to produce impulse response 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. If the researcher is only interested in one-sided response bands and is satisfied with median biased confidence intervals, 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 impulse response bands, Wright's (2000) and Pesavento and Rossi (2005) are the best options, with the following trade-offs. Wright (2000) is a good choice if the researcher is interested in confidence intervals at both short and long horizons and is satisfied with a conservative confidence interval. Pesavento and Rossi (2005) provides median unbiased confidence intervals 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.[7] Pesavento and Rossi's (2005) "robust" method

[7] 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

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 provides guidance to 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 confidence intervals. Wright's (2000) method is a good choice if the researcher is interested in confidence intervals at both short and long horizons and is satisfied with a conservative confidence interval. Pesavento and Rossi's (2005) method tends to provide median unbiased confidence intervals with accurate coverage no matter whether the root is unity or close to unity (even mildly explosive), but is

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).

advised in general only for horizons that are bigger than ten percent of the sample size. If one is content with one-sided impulse response intervals, the method proposed by Gospodinov (2004) 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 (1987), "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 (2002a), "Bootstrapping Autoregressive Processes with Possible Unit Roots", *Econometrica* 70(1), 377-391.

Inoue, Atsushi, and Lutz Kilian (2002b), "Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in VAR(∞) Models", *International Economic Review* 43(2), 309-332.

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 (1998a), "Small-sample Confidence Intervals for Impulse Response Functions", *Review of Economics and Statistics* 80(2), 218-230.

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

Kilian, Lutz (1999), "Finite-Sample Properties of Percentile and Percentile-t Bootstrap Confidence Intervals for Impulse Responses", *Review of Economics and Statistics* 81(4), 652-660.

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.

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

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	
ρ	λ_2	h	δ	low	up	low	up	low	up	low	up
0.99	0	5	0.05	0.00	0.37	0.09	0.03	0.05	0.07	0.05	0.06
	0	10	0.1	0.00	0.41	0.21	0.01	0.13	0.06	0.12	0.06
	0	20	0.2	0.00	0.41	0.62	0.00	0.47	0.06	0.46	0.06
	0	30	0.3	0.00	0.40	0.94	0.00	0.85	0.06	0.82	0.06
	0.4	5	0.05	0.01	0.30	0.05	0.05	0.03	0.07	0.03	0.07
	0.4	10	0.1	0.00	0.40	0.09	0.03	0.06	0.07	0.05	0.08
	0.4	20	0.2	0.00	0.42	0.28	0.00	0.20	0.05	0.18	0.07
	0.4	30	0.3	0.00	0.42	0.58	0.00	0.44	0.05	0.42	0.07
	0.8	5	0.05	0.05	0.12	0.10	0.04	0.06	0.03	0.06	0.02
	0.8	10	0.1	0.01	0.32	0.04	0.08	0.03	0.07	0.02	0.06
	0.8	20	0.2	0.00	0.44	0.06	0.06	0.04	0.08	0.03	0.08
	0.8	30	0.3	0.00	0.46	0.12	0.03	0.08	0.07	0.07	0.07
0.97	0	5	0.05	0.00	0.33	0.29	0.00	0.17	0.07	0.16	0.07
	0	10	0.1	0.00	0.31	0.89	0.00	0.76	0.08	0.72	0.07
	0	20	0.2	0.00	0.30	1.00	0.00	0.92	0.08	0.87	0.07
	0	30	0.3	0.00	0.29	1.00	0.00	0.92	0.08	0.87	0.07
	0.4	5	0.05	0.01	0.28	0.12	0.03	0.07	0.07	0.06	0.06
	0.4	10	0.1	0.00	0.33	0.44	0.00	0.31	0.08	0.30	0.07
	0.4	20	0.2	0.00	0.31	0.99	0.00	0.90	0.08	0.84	0.07
	0.4	30	0.3	0.00	0.30	1.00	0.00	0.92	0.08	0.85	0.07
	0.8	5	0.05	0.05	0.12	0.13	0.03	0.09	0.03	0.08	0.03
	0.8	10	0.1	0.01	0.28	0.09	0.04	0.06	0.07	0.05	0.06
	0.8	20	0.2	0.00	0.35	0.32	0.01	0.24	0.07	0.20	0.07
	0.8	30	0.3	0.00	0.34	0.79	0.00	0.68	0.07	0.61	0.08
0.95	0	5	0.05	0.00	0.29	0.60	0.00	0.44	0.09	0.40	0.07
	0	10	0.1	0.00	0.26	1.00	0.00	0.88	0.10	0.79	0.06
	0	20	0.2	0.00	0.24	1.00	0.00	0.88	0.10	0.79	0.06
	0	30	0.3	0.00	0.23	1.00	0.00	0.88	0.10	0.79	0.06
	0.4	5	0.05	0.01	0.27	0.20	0.01	0.12	0.07	0.10	0.06
	0.4	10	0.1	0.00	0.28	0.87	0.00	0.71	0.09	0.64	0.07
	0.4	20	0.2	0.00	0.25	1.00	0.00	0.88	0.10	0.80	0.07
	0.4	30	0.3	0.00	0.24	1.00	0.00	0.88	0.10	0.80	0.07
	0.8	5	0.05	0.04	0.12	0.15	0.02	0.08	0.03	0.07	0.02
	0.8	10	0.1	0.01	0.26	0.16	0.02	0.09	0.07	0.07	0.05
	0.8	20	0.2	0.00	0.30	0.75	0.00	0.60	0.09	0.53	0.07
	0.8	30	0.3	0.00	0.27	1.00	0.00	0.89	0.09	0.81	0.07

Percentage of times the true IRF lays above or below the confidence interval. 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	
ρ	λ_2	h	δ	low	up	low	up	low	up
0.99	0	5	0.05	0.00	0.62	0.17	0.09	0.00	0.19
	0	10	0.1	0.00	0.84	0.15	0.24	0.00	0.17
	0	20	0.2	0.00	0.91	0.09	0.06	0.00	0.14
	0	30	0.3	0.00	0.91	0.07	0.06	0.00	0.13
	0.4	5	0.05	0.00	0.44	0.22	0.06	0.00	0.19
	0.4	10	0.1	0.00	0.77	0.23	0.08	0.00	0.21
	0.4	20	0.2	0.00	0.91	0.22	0.11	0.00	0.17
	0.4	30	0.3	0.00	0.92	0.20	0.02	0.00	0.14
	0.8	5	0.05	0.00	0.37	0.23	0.19	0.02	0.09
	0.8	10	0.1	0.00	0.58	0.22	0.15	0.00	0.22
	0.8	20	0.2	0.00	0.84	0.24	0.07	0.00	0.27
	0.8	30	0.3	0.00	0.92	0.26	0.06	0.00	0.24
0.97	0	5	0.05	0.00	0.54	0.10	0.09	0.00	0.14
	0	10	0.1	0.00	0.69	0.06	0.14	0.00	0.12
	0	20	0.2	0.00	0.71	0.02	0.08	0.00	0.10
	0	30	0.3	0.00	0.71	0.01	0.08	0.00	0.10
	0.4	5	0.05	0.00	0.39	0.17	0.04	0.00	0.14
	0.4	10	0.1	0.00	0.67	0.17	0.04	0.00	0.13
	0.4	20	0.2	0.00	0.74	0.14	0.03	0.00	0.10
	0.4	30	0.3	0.00	0.75	0.12	0.02	0.00	0.10
	0.8	5	0.05	0.00	0.31	0.17	0.15	0.03	0.08
	0.8	10	0.1	0.00	0.51	0.17	0.10	0.00	0.17
	0.8	20	0.2	0.00	0.75	0.21	0.02	0.00	0.16
	0.8	30	0.3	0.00	0.81	0.26	0.01	0.00	0.13
0.95	0	5	0.05	0.00	0.47	0.05	0.10	0.00	0.13
	0	10	0.1	0.00	0.56	0.02	0.13	0.00	0.11
	0	20	0.2	0.00	0.57	0.00	0.08	0.00	0.10
	0	30	0.3	0.00	0.57	0.00	0.09	0.00	0.10
	0.4	5	0.05	0.00	0.35	0.14	0.04	0.00	0.12
	0.4	10	0.1	0.00	0.57	0.14	0.03	0.00	0.10
	0.4	20	0.2	0.00	0.60	0.09	0.02	0.00	0.09
	0.4	30	0.3	0.00	0.59	0.07	0.02	0.00	0.09
	0.8	5	0.05	0.00	0.28	0.14	0.10	0.03	0.07
	0.8	10	0.1	0.00	0.47	0.15	0.06	0.00	0.14
	0.8	20	0.2	0.00	0.66	0.22	0.01	0.00	0.11
	0.8	30	0.3	0.00	0.68	0.31	0.00	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	
ρ	λ_2	h	δ	low	up	low	up	low	up	low	up
0.99	0	5	0.05	0.01	0.02	0.00	0.10	0.07	0.05	0.06	0.06
		10	0.1	0.01	0.02	0.00	0.10	0.07	0.05	0.05	0.05
		20	0.2	0.02	0.03	0.00	0.10	0.07	0.06	0.06	0.05
		30	0.3	0.04	0.03	0.00	0.10	0.07	0.08	0.06	0.05
	0.4	5	0.05	0.01	0.01	0.00	0.11	0.16	0.05	0.06	0.15
		10	0.1	0.01	0.02	0.00	0.09	0.10	0.05	0.05	0.07
		20	0.2	0.01	0.02	0.00	0.09	0.07	0.05	0.06	0.06
		30	0.3	0.02	0.02	0.00	0.09	0.07	0.05	0.06	0.06
	0.8	5	0.05	0.01	0.01	0.00	0.07	0.29	0.03	0.31	0.09
		10	0.1	0.01	0.01	0.00	0.09	0.21	0.05	0.06	0.18
		20	0.2	0.01	0.01	0.00	0.08	0.12	0.05	0.03	0.10
		30	0.3	0.01	0.01	0.00	0.08	0.09	0.05	0.04	0.07
0.97	0	5	0.05	0.01	0.01	0.00	0.09	0.08	0.05	0.05	0.06
		10	0.1	0.02	0.03	0.00	0.09	0.08	0.05	0.06	0.06
		20	0.2	0.05	0.04	0.00	0.09	0.08	0.05	0.06	0.06
		30	0.3	0.05	0.05	0.00	0.09	0.08	0.06	0.06	0.06
	0.4	5	0.05	0.01	0.01	0.00	0.09	0.14	0.05	0.05	0.11
		10	0.1	0.01	0.01	0.00	0.07	0.11	0.04	0.04	0.06
		20	0.2	0.03	0.02	0.00	0.08	0.10	0.04	0.05	0.06
		30	0.3	0.05	0.03	0.00	0.08	0.10	0.04	0.06	0.06
	0.8	5	0.05	0.01	0.02	0.00	0.06	0.30	0.03	0.25	0.08
		10	0.1	0.01	0.01	0.00	0.07	0.22	0.04	0.05	0.13
		20	0.2	0.01	0.01	0.00	0.07	0.15	0.03	0.04	0.06
		30	0.3	0.03	0.00	0.00	0.06	0.12	0.03	0.05	0.06
0.95	0	5	0.05	0.01	0.01	0.00	010	0.07	0.05	0.04	0.05
		10	0.1	0.03	0.03	0.00	0.10	0.07	0.05	0.05	0.06
		20	0.2	0.05	0.04	0.00	0.10	0.07	0.05	0.05	0.06
		30	0.3	0.06	0.04	0.00	0.09	0.07	0.06	0.06	0.06
	0.4	5	0.05	0.01	0.01	0.00	0.08	0.12	0.04	0.03	0.09
		10	0.1	0.02	0.01	0.00	0.08	0.12	0.03	0.04	0.06
		20	0.2	0.04	0.02	0.00	0.08	0.10	0.04	0.05	0.06
		30	0.3	0.06	0.03	0.00	0.08	0.13	0.02	0.06	0.05
	0.8	5	0.05	0.00	0.02	0.00	0.06	0.27	0.03	0.18	0.08
		10	0.1	0.00	0.01	0.00	0.06	0.20	0.04	0.03	0.12
		20	0.2	0.02	0.01	0.00	0.06	0.16	0.03	0.04	0.07
		30	0.3	0.05	0.00	0.00	0.05	0.16	0.03	0.06	0.07

Notes: as per Table 1.

Table 4: Comparison of coverage of IRF bands:
Methods based on local to unity approximations, trend included in estimation

				Wright		Gospodinov (I)		Gospodinov (II)		Pesavento-Rossi	
ρ	λ_2	h	δ	low	up	low	up	low	up	low	up
0.99	0	5	0.05	0.01	0.01	0.00	0.10	0.00	0.05	0.05	0.05
		10	0.1	0.01	0.01	0.00	0.09	0.00	0.05	0.05	0.05
		20	0.2	0.03	0.02	0.00	0.10	0.00	0.05	0.05	0.05
		30	0.3	0.05	0.02	0.00	0.10	0.00	0.11	0.05	0.05
	0.4	5	0.05	0.02	0.01	0.00	0.10	0.02	0.05	0.05	0.11
		10	0.1	0.01	0.01	0.00	0.09	0.00	0.06	0.06	0.05
		20	0.2	0.01	0.01	0.00	0.09	0.00	0.05	0.06	0.05
		30	0.3	0.03	0.01	0.00	0.09	0.00	0.10	0.06	0.05
	0.8	5	0.05	0.00	0.01	0.00	0.07	0.10	0.03	0.17	0.10
		10	0.1	0.00	0.01	0.00	0.09	0.05	0.05	0.04	0.17
		20	0.2	0.00	0.01	0.00	0.09	0.03	0.06	0.05	0.07
		30	0.3	0.01	0.01	0.00	0.09	0.02	0.13	0.06	0.06
0.97	0	5	0.05	0.01	0.01	0.00	0.10	0.00	0.05	0.06	0.05
		10	0.1	0.03	0.02	0.00	0.10	0.01	0.05	0.06	0.06
		20	0.2	0.05	0.03	0.00	0.10	0.01	0.05	0.06	0.06
		30	0.3	0.05	0.03	0.00	0.10	0.00	0.09	0.06	0.06
	0.4	5	0.05	0.01	0.01	0.00	0.09	0.03	0.05	0.05	0.08
		10	0.1	0.02	0.01	0.00	0.08	0.01	0.05	0.05	0.06
		20	0.2	0.04	0.01	0.00	0.09	0.00	0.04	0.06	0.06
		30	0.3	0.05	0.01	0.00	0.09	0.00	0.08	0.06	0.06
	0.8	5	0.05	0.00	0.01	0.00	0.06	0.09	0.03	0.16	0.09
		10	0.1	0.01	0.01	0.00	0.07	0.06	0.04	0.05	0.11
		20	0.2	0.02	0.01	0.00	0.07	0.03	0.04	0.06	0.06
		30	0.3	0.05	0.00	0.00	0.07	0.02	0.08	0.07	0.06
0.95	0	5	0.05	0.01	0.01	0.00	0.10	0.00	0.05	0.05	0.05
		10	0.1	0.04	0.02	0.00	0.10	0.00	0.04	0.06	0.06
		20	0.2	0.06	0.02	0.00	0.10	0.00	0.05	0.06	0.06
		30	0.3	0.06	0.03	0.00	0.10	0.00	0.07	0.06	0.06
	0.4	5	0.05	0.01	0.00	0.00	0.07	0.02	0.04	0.04	0.07
		10	0.1	0.03	0.01	0.00	0.07	0.00	0.04	0.05	0.06
		20	0.2	0.05	0.01	0.00	0.07	0.00	0.04	0.05	0.06
		30	0.3	0.06	0.01	0.00	0.07	0.00	0.04	0.06	0.0
	0.8	5	0.05	0.00	0.01	0.00	0.05	0.07	0.03	0.12	0.08
		10	0.1	0.01	0.01	0.00	0.06	0.04	0.03	0.04	0.08
		20	0.2	0.03	0.00	0.00	0.06	0.02	0.03	0.05	0.07
		30	0.3	0.06	0.00	0.00	0.06	0.01	0.06	0.06	0.06

Notes: as per Table 1.