One-dimensional inference in autoregressive models with the potential presence of a unit root.

By Anna Mikusheva [1]

Abstract

This paper examines the problem of testing and confidence set construction for one-dimensional functions of the coefficients in AR(p) models with potentially persistent time series. The primary example concerns inference on impulse responses. A new asymptotic framework is suggested, and some new theoretical properties of known procedures are demonstrated. I show that the LR and $\mathrm{LR}^{\pm}$ statistics for a linear hypothesis in an AR(p) can be uniformly approximated by a weighted average of local-to-unity and normal distributions. The corresponding weights depend on the weight placed on the largest root in the null hypothesis. The suggested approximation is uniform over the set of all linear hypotheses. The same family of distributions approximates the LR and $LR^{\pm}$ statistics for tests about impulse responses, and the approximation is uniform over the horizon of the impulse response. I establish the size properties of tests about impulse responses proposed by Inoue and Kilian (2002) and Gospodinov (2004), and theoretically explain some of the empirical findings of Pesavento and Rossi (2007). An adaptation of the grid bootstrap for IRFs is suggested and its properties are examined.

`Key words: impulse response, grid bootstrap, uniform inferences`

1 Introduction

Impulse response function (IRF) estimates and confidence sets are the most common way of reporting results for AR/VAR estimation, and also for describing dynamic interactions between variables. IRF confidence sets also serve as means for assessing the fit between theoretical macro models and the data as in, for example, Gali (1999) and Christiano, Eichenbaum and Vigfusson (2004). Estimation of impulse responses is also a first step in the estimation of DSGE models via the IRF matching method introduced by Rotemberg

[1]Department of Economics, M.I.T., 50 Memorial Drive, Building E52, Cambridge, MA, 02142. Email: amikushe@mit.edu. I would like to thank three anonymous referee, co-editor, Nikolay Gospodinov, Pierre Perron, Barbara Rossi, Jim Stock and participants of the Econometric Seminar at Harvard and Boston University, the Winter Econometric Society Meeting 2008, and the CIREQ conference in Montreal 2009 for comments. I am grateful to Denis Chetverikov, JB Doyle and Ruitian Lang for research assistance. Financial support from the Castle-Krob Career Development Chair is gratefully acknowledged.

and Woodford (1997) and Christiano, Eichenbaum and Evans (1999).

The main problem discussed in this paper is how to make inferences on IRFs in AR(p) models when the process is highly persistent. The question of how much a practitioner should be concerned about the potential presence of a unit root is a controversial one. At one extreme, Inoue and Kilian (2002) showed that the usual bootstrap works asymptotically for impulse responses at fixed horizons in an AR(p), regardless of whether there is a unit root. At the same time, it was shown in simulations by Kilian (1998), Kilian and Chang (2000), Pesavento and Rossi (2007), among others, that the bootstrap possesses very poor size properties, and the normal approximation is inaccurate for impulse responses at very long horizons if a near-unit root is present. This was also demonstrated theoretically by Phillips (1998) and Gospodinov (2004).

The reason for such a difference of opinion is the use of different asymptotic assumptions. There are two distinct strands of literature, which are referred below as the "short" and "long" horizon approaches. In the "short" horizon approach, one assumes that the horizon of the impulse response being estimated is fixed as the sample size increases. Under this assumption, one indeed arrives at the same conclusion as Inoue and Kilian (2002), in particular, the t-statistic for the impulse response converges to a normal distribution regardless whether the unit root is present. However, if the horizon is modeled as growing proportionally to the sample size, then the asymptotic distribution of the LR statistic will be non-standard if a weak unit root is present. The assumption of a growing horizon may be a reasonable one for the estimation of an impulse responses on long horizons. Methods suggested by Gospodinov (2004) and Pesavento and Rossi (2006) employ this asymptotic setting. The confidence sets based on the classical methods (delta-method, bootstrap) are significantly different from those produced by methods that are consistent under linearly-growing horizon asymptotics. The incompatibility of these two different approaches requires practitioners to somewhat artificially choose a set of asymptotic assumptions, making inferences subjective. The main goal of this paper is to provide a method that will work uniformly well for both short and long horizons independent of the presence of a unit root.

In this paper, I introduce and examine the properties of the grid bootstrap, a procedure initially suggested by Hansen (1999) to do inferences on the sum of AR coefficients in an AR(p). In an AR(p) with a root near one the distributions of the usual test statistics depend on a so-called local-to-unity parameter, which cannot be consistently estimated by OLS. The idea underlying the grid bootstrap is that the null hypothesis contains information that can be used for estimating the local-to-unity parameter. The grid bootstrap uses the restricted estimates of unknown parameters to construct critical values. The grid bootstrap is so named because it constructs confidence sets by grid testing, that is, one tests all possible values of the impulse response (on a fine grid) and composes the confidence set from the accepted values. For each value tested, I use the same test statistic but different critical values. For each null hypothesis, the critical values depend on the restricted estimates of the parameters. Mikusheva (2007a) shows that the grid bootstrap produces uniformly correct confidence sets for the sum of autoregressive parameters in an AR(p), while the classical bootstrap or equi-tailed subsampling fails to control size uniformly.

The paper contains two main results. First, I establish an approximation of the finite sample distributions of the LR and $LR^{\pm}$ test statistics for making inferences about IRFs by a weighted sum of local-to-unity and normal distributions. The approximation is uniform over the parameter space, and it works uniformly well for all horizons. The weights in the approximation are determined by how important the persistence parameter is for the null hypothesis. For short horizons, the approximating distribution is very close to normal, while for very long horizons it approaches the local-to-unity limit. Thus the new approximation includes as special cases the classical asymptotic setting with a fixed horizon as well as the non-standard asymptotic setting of Gospodinov (2004), and it allows for a comparison of the asymptotic size obtained by the different methods. The generalization of the results to heteroscedastic and multivariate models is discussed.

The second result presented in this paper is that the grid bootstrap controls size well for the two extremes- a fixed horizon as well as a linearly-growing horizon. Surprisingly enough, the grid bootstrap fails to fully control size on medium horizons, but the distortions are relatively small. The intuition for the grid bootstrap performance is as follows. An impulse response at a very long horizon heavily depends on the persistence of the process, as characterized by the largest root, and therefore a hypothesis about such an impulse response contains a lot of information about the degree of persistence. As a result, the restricted estimate of the local-to-unity parameter is consistent if the horizon is growing at a faster rate than $\sqrt{T}$. This allows the grid bootstrap to use asymptotically correct critical values and to effectively control size for long horizons. The impulse responses at short horizons do not contain as much information about the largest root, but the distribution of the test statistic is close to normal and does not depend on persistence in any significant way. As a result, the grid bootstrap controls size on short horizons as well. Some distortions occur in the middle, but they are smaller than the distortions produced by the usual bootstrap since the grid bootstrap uses a better estimate. The grid bootstrap is a working method to produce confidence intervals for the impulse responses that behaves favorably when compared with currently used methods.

Along the way I obtain three results that are interesting on their own. First, I generalize the results obtained in Mikusheva (2007a) to an AR(p). I establish a uniform asymptotic approximation for some basic OLS statistics for an AR(p). The difficulties of normalizing and separating statistics with standard and non-standard limits are discussed. Related results describing the behavior of linear and quadratic forms for the stationary region only are obtained in Giraitis and Phillips (2009).

Second, establishing the uniform approximation for LR and $LR^{\pm}$ statistics for IRFs allows me to discuss theoretical properties of known procedures for testing hypotheses about IRFs on different horizons. I provide a theoretical justification to many of the findings Pesavento and Rossi (2007) obtained in their simulation study.

Third, I show that the LR and Wald statistics may have asymptotically different distributions. I prove that the LR statistic for an IRF has the same asymptotic approximation as the LR statistic for the linearized version of the hypothesis, and this approximation is uniform over the horizon of the IRF as well as over the parameter space. At the same time, the Wald statistic is very sensitive to the curvature of the problem, and it is not uniformly well approximated by the above-described family. The curiosity found in this paper is one of the first examples where LR and Wald statistics are asymptotically different.

The remainder of the paper is organized in the following way. In section 2, the uniform approximation for the basic OLS-type statistics in an AR(p) is established. Section 3 establishes the uniform approximation for a set of linear hypotheses, and it introduces the grid bootstrap procedure. The issues related to the non-linearity of IRFs are discussed in section 4, which also provides a generalization of the results to multi-dimensional and heteroskedastic settings. Section 5 obtains and discusses the asymptotic coverage properties of different methods of constructing confidence sets. The results of a simulation study appear in section 6. The proofs of the main results are collected in the Appendix. Since the proofs are long and technically involved, some proofs, in particular, those for the results of Section 4.2, extensions to heteroskedasticity as well as additional discussions of the Wald statistic for non-linear functions are placed in the Supplementary Appendix, which can be found on the author's web-site.[2]

2 The Asymptotic Approximation for AR(p)

2.1 Autoregressive Models

Consider a sample of size T from an AR(p) process written in the Augmented Dickey-Fuller (ADF) form

$$y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \alpha_j \Delta y_{t-j} + e_t. \tag{1}$$

Assume that the process is initialized as $y_0 = y_{-1} = ... = y_{-p+1} = 0$.

Assumption A. Let $\{e_t\}_{t=1}^{\infty}$ be i.i.d. error terms with a zero mean Ee_t=0, a unit variance $Ee_t^2 = 1$ and a finite fourth moment $Ee_t^4 < C < \infty$.

I restrict the parameter space of coefficients $(\rho, \alpha = (\alpha_1, ..., \alpha_{p-1}))$ in such a way that there is at most one root close to a unit root. In doing so I abstract from multiple unit roots as well as harmonic unit roots. For each (ρ, α), one can find $\lambda_i = \lambda_i(\rho, \alpha)$ such that

$$1 - \rho L - \sum_{j=1}^{p-1} \alpha_j L^j (1 - L) = (1 - \lambda_1 L)...(1 - \lambda_p L),$$

where the (complex) roots are ordered in ascending order $|\lambda_1| \le |\lambda_2| \le ... \le |\lambda_p|$. I assume that $(\rho, \alpha) \in \mathfrak{R}_\delta$, and

$$\mathfrak{R}_\delta = \{(\rho, \alpha) : |\lambda_{p-1}(\rho, \alpha)| < \delta, \text{ and if } \lambda_p(\rho, \alpha) \in \mathbb{R}, \text{ then } -\delta \le \lambda_p < 1\}.$$

Consider a family of finite-sample distributions of some statistic ξ_1: $F^{(1)}_{\rho,\alpha,T}(x) = P_{\rho,\alpha,T}\{\xi_1 < x\}$. I approximate this family by a family of limiting distributions indexed by a parameter c: $F^{(2)}_c(x) = P_c\{\xi_2(c) < x\}$, where $c = c(T, \rho, \alpha)$ is related to the true parameter value (ρ, α) and to the sample size T in a known way. The approximation is said to be uniform over $(\rho, \alpha) \in \mathfrak{R}_\delta$ if

$$\lim_{T\to\infty} \sup_{(\rho,\alpha)\in\mathfrak{R}_\delta} \sup_x \left| F^{(1)}_{\rho,\alpha,T}(x) - F^{(2)}_{c(\rho,\alpha,T)}(x) \right| = 0.$$

[2]http://econ-www.mit.edu/files/6612

This corresponds to the notion of uniformity given in Mikusheva (2007a). For simplicity, when no ambiguity arises, I write $\xi_1 \cong \xi_2(c)$. The uniform approximation is a much stronger requirement than the point-wise approximation

$$\lim_{T\to\infty} \sup_x \left| F^{(1)}_{\rho,\alpha,T}(x) - F^{(2)}_{c(\rho,\alpha,T)}(x) \right| = 0 \text{ for each } (\rho,\alpha) \in \mathfrak{R}_\delta.$$

The former requires that for any $\varepsilon > 0$ there exists a sample size that guarantees that the accuracy of the approximation at *any* parameter value (ρ, α) is no worse than ε. In contrast, the point-wise approximation allows the speed of convergence to differ depending on the value of (ρ, α). Since the true value of (ρ, α) is usually unknown, the accuracy of the point-wise procedure is unknown, and the finite sample size of the procedure may be much larger than the declared. The distinction between uniform and point-wise approximations is important for many econometric problems such as inference for persistent time series, weak instruments, and the parameter-on-the-boundary problem. A nice treatment of this distinction with additional examples and explanations is given in Andrews and Guggenberger (2009).

The importance of uniform inferences for an AR(1) with near unit roots was demonstrated in Mikusheva (2007a). For any value of the AR parameter different from the unit root, the Central Limit Theorem and Law of Large Numbers hold for $\frac{1}{\sqrt{T}}\sum_{t=1}^T y_{t-1}e_t$ and $\frac{1}{T}\sum_{t=1}^T y_t^2$ respectively. However, the convergence in those laws becomes slower as the AR parameter approaches a unit root. This leads to the observation that the classical OLS t-statistic inferences are point-wise correct but not uniformly correct. This explains the extremely poor coverage of classical OLS procedures for very persistent processes.

2.2 Uniform Approximation for AR(1)

Here I briefly describe the results for an AR(1) established in Mikusheva (2007a) because they provide useful intuition in an AR(p) case.

Assume that $\overline{y}_t = \rho\overline{y}_{t-1} + e_t$ with $\overline{y}_0 = 0$, and the error term satisfies Assumption A. Then the following asymptotic approximations hold uniformly over $\rho \in [-1+\delta, 1)$ for some positive $\delta > 0$:

$$\sqrt{\frac{1-\rho^2}{T}} \sum_{t=1}^T \overline{y}_{t-1}e_t \cong \frac{1}{\sqrt{g(c)}} \int_0^1 J_c(t)dW(t); \tag{2}$$

$$\frac{1-\rho^2}{T} \sum_{t=1}^T \overline{y}_{t-1}^2 \cong \frac{1}{g(c)} \int_0^1 J_c^2(t)dt, \tag{3}$$

where $J_c(s) = \int_0^s e^{c(t-s)}dW(t)$ is an Ornstein-Uhlenbeck process, $g(c) = E\int_0^1 J_c^2(t)dt = \frac{e^{2c}-1-2c}{4c^2}$, and $c = c(\rho, T) = T\log(\rho)$.

Several observations are worth making. First, the expressions on the right side of equations (2) and (3) are the point-wise limits of the expressions staying on the left in the so-called local-to-unity asymptotic approach, where the parameter ρ is modeled as approaching a unit root as the sample size T increases: $\rho_T = 1 + c/T$. The local-to-unity asymptotics were suggested in a sequence of papers by Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987), and Stock (1991).

Second, in the classical asymptotic setting, that is, $|\rho| < 1$ is fixed and $T \to \infty$, the first statistic has normalization proportional to $\frac{1}{\sqrt{T}}$, while the second has $\frac{1}{T}$. If the parameter ρ is modeled as in the local-to-unity approach $\rho_T = 1 + c/T$, then the normalization in (2) and (3) becomes much stronger: for the first statistic, it is proportional to $\frac{1}{T}$, while for the second it is proportional to $\frac{1}{T^2}$.

Third, if $|\rho| < 1$ is fixed and $T \to \infty$ then $c(\rho, T) \to -\infty$. Phillips (1987) showed that the right side of (2) weakly converges to a standard normal distribution, and the right side of (3) converges in probability to one as $c \to -\infty$. That is, even though the approximating family resembles the local-to-unity limit distribution, it nests the classical Central Limit Theorem and the Law of Large Numbers approximations for strictly stationary processes.

Finally, the variables are normalized in such a way that $\frac{1}{g(c)}\int_0^1 J_c^2(t)dt$ is bounded in probability from above and separated in probability from zero, while $\frac{1}{\sqrt{g(c)}}\left|\int_0^1 J_c(t)dW(t)\right|$ is bounded in probability uniformly over $-\infty \leq c < 0$.

2.3 Main result on the uniform approximation for AR(p)

If one wishes to establish some analogs to equations (2) and (3) for an AR(p), one faces several difficulties. The regressor $X_t = (y_{t-1}, \Delta y_{t-1}, ..., \Delta y_{t-p+1})'$ in the auto-regression (1) has components $Z_t = (\Delta y_{t-1}, ..., \Delta y_{t-p+1})'$ which are strictly stationary and obey the classical Central Limit Theorem and the Law of Large Numbers, as well as a component y_{t-1} that may be stationary or persistent depending on how close the largest root is to the unit circle. As in the case of an AR(1), these components may require different normalizations. When ρ is close to 1, statistics involving y_{t-1} will be asymptotically independent from those containing Δy_{t-j}. However, when ρ is far from 1 and the classical normal asymptotics apply, there is a non-zero correlation between these two sets of statistics. To deal with this issue, I rotate X_t in such a way that the resulting statistics will be asymptotically uncorrelated in the stationary region as well.

Let $\Sigma(\rho, \alpha) = \lim_{t\to\infty} E_{\rho,\alpha} X_t X_t'$ be the limit of the variance matrix of X_t.[3] Assume also that F is a lower-triangular matrix such that $F\Sigma(\rho, \alpha)F' = I_p$.

Theorem 1 *Let y_t be an AR(p) process defined by equation (1) with error terms satisfying Assumption A. Then the following approximations hold uniformly over $(\rho, \alpha) \in \mathfrak{R}_\delta$:*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F X_t e_t \cong \begin{pmatrix} \frac{1}{\sqrt{g(c)}} \int_0^1 J_c(s) dW(s) \\ \xi \end{pmatrix}, \tag{4}$$

$$\frac{1}{T} \sum_{t=1}^{T} F X_t X_t' F' \cong \begin{pmatrix} \frac{1}{g(c)} \int_0^1 J_c^2(s) ds & \mathbf{0}' \\ \mathbf{0} & I_{p-1} \end{pmatrix}, \tag{5}$$

where $c = T\log(|\lambda_p|)$, $\xi \sim N(0, I_{p-1})$ is independent of W, $\mathbf{0}$ is a $(p-1) \times 1$ vector of zeros, and I_{p-1} is an identity matrix of size $(p-1) \times (p-1)$.

In an AR(p), the persistence of the process can be characterized by the largest root λ_p. Local-to-unity asymptotics assumes $\lambda_p = \exp\{c/T\}$. However, the concept of the largest root is well defined only for very persistent processes. A strictly stationary AR(p) process may have complex roots. I resolve this issue by defining $c = T\log(|\lambda_p|)$.

According to Lemma 2 in the Appendix, the first component of the vector FX_t is $\frac{\sqrt{1-\rho^2}}{C(\rho,\alpha)} y_{t-1}$, where $C(\rho, \alpha)$ is bounded and separated from the zero. Theorem 1 shows that $\frac{\sqrt{1-\rho^2}}{C(\rho,\alpha)\sqrt{T}} \sum_{t=1}^{T} y_{t-1} e_t$ and $\frac{1-\rho^2}{C^2(\rho,\alpha)T} \sum_{t=1}^{T} y_{t-1}^2$ are approximated by the same family of distributions as in the AR(1) case. The normalization term $\sqrt{1-\rho^2}$ allows for stronger normalization in local-to-unity asymptotics. Function $C(\rho, \alpha)$ captures the long-run variance of the quasi-differenced series $y_t - \lambda_p y_{t-1}$ within the local-to-unity framework. The last $p-1$ components of FX_t (denote them $\widetilde{x}_t$) are strictly stationary in the sense that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widetilde{x}_t e_t$ satisfies the Central Limit Theorem and $\frac{1}{T} \sum_{t=1}^{T} \widetilde{x}_t \widetilde{x}_t'$ obeys the Law of Large Numbers. The transformation F rotates X_t in such a way that the standard and non-standard components are asymptotically independent.

[3]It is equal to the covariance matrix of an analog of X_t for a strictly stationary process as defined in equation (1).

3 Inference for linear functions

3.1 Approximation of the test statistic

The main goal of the paper is to perform a test about or to construct a confidence set for a function of the parameters $\gamma = f(\rho, \alpha, T)$. These are dual problems, since a confidence set is an acceptance set for a sequence of the hypotheses $H_0 : \gamma = \gamma_0$. Let $(\widehat{\rho}, \widehat{\alpha}')'$ be the unrestricted OLS estimates and

$$(\widetilde{\rho}(\gamma_0), \widetilde{\alpha}(\gamma_0)')' = \arg \max_{f(\rho,\alpha)=\gamma_0} \max_{\sigma} l_T(\rho, \alpha, \sigma)$$

be the restricted estimates of coefficients given the null $H_0 : f(\rho, \alpha) = \gamma_0$. Here

$$l_T(\rho, \alpha, \sigma) = -\frac{T}{2} \ln(\sigma^2) - \frac{1}{2} \sum_{t=1}^{T} \frac{(y_t - \rho y_{t-1} - \alpha' Z_t)^2}{\sigma^2}$$

is the quasi-log-likelihood function. The LR statistic is $LR(\gamma_0) = T \ln(SSR_0/SSR)$. It is easy to verify that $LR = T \log(1 + \frac{\sum \tilde{e}_t^2 - \sum \hat{e}_t^2}{\sum \hat{e}_t^2}) = \frac{\sum \tilde{e}_t^2 - \sum \hat{e}_t^2}{\hat{\sigma}^2} + o_p(1)$ uniformly. For simplicity, I will use LR to denote the last expression, namely $\frac{\sum \tilde{e}_t^2 - \sum \hat{e}_t^2}{\hat{\sigma}^2}$. I also consider a signed version of the LR statistic $LR^{\pm}(\gamma_0) = sign(f(\hat{\rho}, \hat{\alpha}) - \gamma_0) \cdot \sqrt{LR(\gamma_0)}$. The signed statistic can be employed to create equi-tailed confidence sets. It also can be used to create a median-unbiased estimator as was mentioned in Gospodinov (2004).

This section considers only linear hypotheses: $f(\rho, \alpha) = A_1 \rho + A_2' \alpha$, where $A = (A_1, A_2')'$ is a $p \times 1$ vector. Since in the linear case the LR and the Wald statistics coincide and $LR = t^2$, it is enough to study the t-statistic in this case:

$$t = \frac{A' \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{T} X_t e_t}{\widehat{\sigma} \sqrt{A' \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} A}} = \frac{(FA)' \left(\frac{1}{T} \sum_{t=1}^{T} F X_t X_t' F' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F X_t e_t}{\widehat{\sigma} \sqrt{(FA)' \left(\frac{1}{T} \sum_{t=1}^{T} F X_t X_t' F' \right)^{-1} FA}}. \quad (6)$$

Consider a family of distributions indexed by two parameters, c and u:

$$t(c, u) = \frac{t^c + u \sqrt{\frac{\int_0^1 J_c^2(s) ds}{g(c)}} N(0, 1)}{\sqrt{1 + u^2 \frac{\int_0^1 J_c^2(s) ds}{g(c)}}}, \quad (7)$$

where $t^c = \frac{\int_0^1 J_c(t) dW(t)}{\sqrt{\int_0^1 J_c^2(t) dt}}$ is the limit of the t-statistic for ρ in the local-to-unity asymptotics, and $N(0, 1)$ is a standard normal variable independent of Brownian motion $W(t)$. Let

$$u = \frac{\sqrt{f_2' f_2}}{|f_1|} = \sqrt{\frac{A' F' F A - (i_1' F A)^2}{(i_1' F A)^2}}, \quad (8)$$

where $f_1 = i_1' FA$ is the first component of the vector FA, and f_2 is the vector consisting of the last $p-1$ components.

Theorem 2 *Consider a sample $\{y_t\}$ from an AR(p) process as defined in equation (1) with error terms satisfying assumption A. Assume that the linear hypothesis about coefficients $H_0 : (\rho, \alpha')A = \gamma_0$ is tested, and $t(A)$ is the OLS t-statistic for testing this hypothesis. For each value of A, ρ and α let u be defined as in equation (8), and $c = T\log(|\lambda_p|)$. Then the finite sample distribution of t-statistic is asymptotically approximated by the distribution of the random variable $t(c,u)$ defined in equation (7) uniformly with respect to parameters of the process ρ, α and the hypothesis tested:*

$$\lim_{T\to\infty} \sup_{A\in\mathbb{R}^p} \sup_{(\rho,\alpha)\in\Re_\delta\cap H_0} \left|P_{\rho,\alpha}\left\{t(A) < x\right\} - P\left\{t(c,u) < x\right\}\right| = 0.$$

An important peculiarity of the above result is that the suggested approximation is uniform with respect to the linear *null hypothesis* tested along with parameters ρ and α. The intuition of the result is as follows. Under the local-to-unity assumption the OLS estimate $\widehat{\rho}$ is T-consistent and has a non-standard and non-pivotal asymptotic distribution, while the OLS estimate $\widehat{\alpha}$ is $\sqrt{T}$-consistent and asymptotically normal. The parameter u is the measure of sensitivity of the testing problem to a potential presence of the unit root. The value of u is equal to the ratio of the stochastic uncertainty introduced into the estimation of $\gamma = A_1\rho + A_2'\alpha$ by the normally behaving coefficients α to the uncertainty infused by the potentially non-standardly behaving coefficient ρ.

Assume for simplicity that one has an AR(2) process, and thus α is a scalar. If one uses the local-to-unity modeling assumptions, $\rho_T = 1+c/T$, then u will be asymptotically proportional to the ratio of weights $\frac{A_2\sqrt{T}}{A_1}$ put on the coefficients by the hypothesis. So, if there is a sequence of hypotheses for which $\frac{A_2\sqrt{T}}{A_1} \to 0$, then no weight is put on α asymptotically ($u \to 0$), and I obtain the local-to-unity limit $t \Rightarrow t^c$. If, on the contrary, $\frac{A_2\sqrt{T}}{A_1} \to \infty$, then ($u \to \infty$), that is, the short-term dynamics dominate the stochastic component, and $t \Rightarrow N(0,1)$. In general, the finite sample distribution of the t-statistic is approximated by a random mixture of the two extremes.

The uniformity of the approximation over the set of linear hypotheses will be important since impulse responses at different horizons place different asymptotic weights on α and ρ. Let f_k denote the IRF at horizon k. The ratio of derivatives $\frac{\partial f_k}{\partial \rho} / \frac{\partial f_k}{\partial \alpha}$ changes approximately as $const \cdot k$ when k increases (i.e. the dependence of the IRF on ρ increases with the horizon). Consider a linear hypothesis that places weights on ρ and α that are proportional to the corresponding partial derivatives. When one models the horizon k as fixed, asymptotic normality coming from the estimation of α stochastically dominates the non-standard stochastic component introduced by the super-consistent estimation of ρ. As a result, the asymptotic distribution of the test statistic for a linearized hypothesis about f_k will be the same as if ρ were known. That is, this asymptotic distribution does not depend on any nuisance parameters, and the usual bootstrap will work (Inoue and Kilian (2002)). However, if $k_T = [\delta T]$, the horizon proportionally increases with the sample size as in Gospodinov (2004) and Pesavento and Rossi (2006), then the ratio of derivatives $\frac{\partial f_{k_T}}{\partial \rho} / \frac{\partial f_{k_T}}{\partial \alpha}$ increases at speed T. As a result, the influence of the non-standard distribution of ρ on the test statistic stochastically dominates the normal component, and the non-standard asymptotics should be used.

When the horizon k_T is changing between the two described extremes, the limit distribution of the t-statistic is a mixture of the classical (normal) term and the non-standard local-to-unity limit distribution t^c. The weights with which two distributions enter the mixture depend on the parameter u, which is proportional to $\frac{k_T}{\sqrt{T}}$. Section 4 shows that the test statistic for the IFR behaves in the same manner as its linearized version, and the analog of Theorem 2 holds for IRF uniformly over different horizons.

3.2 Accuracy of the restricted estimator

For the grid bootstrap discussed in the sections below, I will need the restricted estimates of (ρ, α); that is, the estimates calculated under the restriction that the null hypothesis be satisfied. In this section, I show that the parameter u introduced in the previous section is closely related to the accuracy of the restricted estimate of c.

Lemma 1 *Let us have an AR(p) process y_t defined in (1) with error terms satisfying Assumption A. Assume that the null hypothesis $H_0 : (\rho, \alpha')A = \gamma_0$ holds. Then uniformly over the parameter space $\mathfrak{R}_\delta$ and over the space of all linear hypotheses $A \in \mathbb{R}^p$, the following representation holds:*

$$\sqrt{T} F^{-1} \begin{pmatrix} \widetilde{\rho} - \rho \\ \widetilde{\alpha} - \alpha \end{pmatrix} \cong \begin{pmatrix} 0 \\ \xi \end{pmatrix} + u\eta(u, c), \tag{9}$$

where $\xi \sim N(0, I_{p-1})$, and the vector of random variables $\eta(u, c)$ is uniformly bounded in probability over c and $0 \le u \le U$ for any fixed, positive U. That is, for any $U > 0$ and *any $\varepsilon > 0$ there exists a constant $C < \infty$ such that*

$$\sup_{-\infty \le c \le 0} \sup_{0 \le u < U} P\{|\eta(u,c)| > C\} < \varepsilon.$$

To interpret Lemma 1, consider the local-to-unity asymptotics. That is, assume that the largest root, $\lambda_p = 1 + c/T$, converges to a unit root while all other roots $|\lambda_1| \le ... \le |\lambda_{p-1}| < \delta$ remain fixed. In the described setting the OLS estimate $\widehat{\rho}$ is T-consistent, that is, $T(\widehat{\rho} - \rho)$ has a non-degenerate limit distribution depending on c, while $\hat{\alpha}$ is $\sqrt{T}$-asymptotically normal. As a result, $\widehat{c} = -T\frac{1-\widehat{\rho}}{1-\sum_{j=1}^{p-1}\widehat{\alpha}_j}$, a natural OLS estimate of c, will not be consistent. Rather it converges to a stochastic limit. Given that the approximating family $t(c,u)$ obtained in the previous section depends on c in a non-trivial way, the inability to estimate c may pose a problem for making inferences. Lemma 1 shows that the restricted estimate may be better if u is small. If I consider a sequence of testing problems indexed by the sample size T such that $u_T \to 0$, then I can estimate c consistently, and u_T will be related to the speed of convergence.

Indeed, under the local-to-unity assumptions matrix F^{-1} has asymptotically a block diagonal form. It follows from Lemma 2 that the upper left element of F^{-1} is asymptotically behaving like $\sqrt{\frac{T}{-2c}}$. That is, the normalization $\sqrt{T}F^{-1}$ in statement (9) treats ρ and α differently, it pre-multiplies $(\widetilde{\alpha} - \alpha)$ by $\sqrt{T}$ while it pre-multiplies $(\widetilde{\rho} - \rho)$ by T. The main message of Lemma 1 is that, unlike the OLS estimate, $T(\widetilde{\rho} - \rho)$ converges to 0 if $u_T \to 0$. That is, in such a case the parameter c is consistently estimable, and u_T characterizes the accuracy of the estimator. Therefore, $\widetilde{c} = -T\frac{1-\widetilde{\rho}}{1-\sum\widetilde{\alpha}_j}$ is a consistent estimate of c if $u_T \to 0$.

The intuition of this result is as follows. If $u_T \to 0$, the null hypothesis of interest puts disproportionately large weight on the "persistence coefficient" ρ, and thus, it is highly informative about ρ. The restricted estimator uses this information. Think about the limit case, imagine that the null is $H_0 : \rho = \rho_0$ a hypothesis about ρ only. It will produce $u = 0$. In this case the restricted estimate is equal to the true value of ρ: $\widetilde{\rho} = \rho_0$.

It also worth noticing that one is better at estimating c in exactly the situations one needs it the most. According to Theorem 2 the smaller u corresponds to the approximation that puts greater weight on the local-to-unity distribution t^c.

3.3 Grid bootstrap and why it works

As will be shown formally in the next section, the $\mathrm{LR}^{\pm}$ statistic for an impulse response can be approximated by a family of distributions $t(u,c)$ indexed by two parameters - the local-to-unity parameter c and the relative importance of normal and local-to-unity component u. And the approximating family of distributions consists of weighted sums of normal and local-to-unity limits. The grid bootstrap procedure, inspired by Hansen (1999), is an attempt to base inferences on the both terms.

The idea of the grid bootstrap was initially introduced by Hansen (1999) for the sum of AR coefficients' confidence set construction. It has been shown that the grid bootstrap produces uniform inference on ρ (Mikusheva (2007a)). Here I modify Hansen's (1999) idea for the IRF.

The key problem here is that the local-to-unity parameter c needed to produce valid critical values cannot be consistently estimated by OLS. The idea behind the grid bootstrap is to use the information contained in the null hypothesis whenever one is tested. The restricted estimate under the null tends to be better, and in fact, sometimes will even be consistent.

The grid bootstrap procedure is as follows. To test the hypothesis $H_0 : f(\rho,\alpha) = \gamma_0$ based on a sample $Y_T = (y_1, ..., y_T)$ one should:

(1) calculate the test statistic $W(Y_T)$ (in our case $W(Y_T)$ is either LR or $LR^{\pm}$) and the restricted estimates $(\widetilde{\rho}, \widetilde{\alpha})$ from the sample Y_T;

(2) simulate samples $Y^*_{b,T} = \{y^*_{b,1}, ..., y^*_{b,T}\}, b = 1, ..., B$ from an AR(p) with coefficients $(\widetilde{\rho}, \widetilde{\alpha})$ and errors drawn randomly with replacement from the OLS residuals or simulated from the standard normal distribution, that is,

$$y^*_{b,t} = \widetilde{\rho} y^*_{b,t-1} + \sum_{j=1}^{p-1} \widetilde{\alpha}_j \Delta y^*_{t-j} + \varepsilon^*_{b,t},$$

where $\varepsilon^*_{b,t}$ are i.i.d. The number of simulations B handles the accuracy of quantile simulation and should be large.

(3) for each simulated sample, $Y^*_{b,T}$, calculate the test statistic $w_b = W(Y^*_{b,T})$;

(4) sort w_b in ascending order: $w_{(1)} \leq ... \leq w_{(B)}$. Perform the test by comparing the test statistic $W(Y_T)$ with the quantiles of the simulated distribution of $W(Y^*_T)$. If

one uses the LR statistic, then the test accepts whenever $W(Y_T) \leq w_{(\lfloor B(1-\alpha)\rfloor)}$ (here $\lfloor x \rfloor$ stays for the whole part of x). When LR$^{\pm}$ is used, the test accepts whenever $w_{(\lfloor B\alpha/2\rfloor)} \leq W(Y_T) \leq w_{(\lfloor B(1-\alpha/2)\rfloor)}$.

As an asymptotically equivalent way of realizing grid bootstrap,[4] one may substitute steps (3) and (4) with the following step

(3') calculate the implied values of $\tilde{c}$ and $\tilde{u}$ using formula (8) and $\widetilde{c} = -T\frac{1-\widetilde{\rho}}{1-\sum\widetilde{\alpha}_j}$. Then find quantiles of the distribution $t(\tilde{c}, \tilde{u})$ for the LR$^{\pm}$ statistic or of distribution $t(\tilde{c}, \tilde{u})^2$ for the LR statistic.

The needed quantiles of $t(c, u)$ and $t(c, u)^2$ may be tabulated and stored for a fine grid of values of c and u to improve the speed of the algorithm. The algorithm with step (3') is less computationally intensive, but it may have less favorable finite-sample properties, since the bootstrap tends to provide a second-order improvement in the classical setting, which may be important for non-linear functions.

If one wants to construct a confidence set for the parameter $\gamma = f(\rho, \alpha)$ via the grid bootstrap, then one has to test all possible hypotheses $H_0 : f(\rho, \alpha) = \gamma_0$ as described above. The confidence set consists of all γ_0 for which the corresponding hypothesis is accepted. Notice that for different γ_0 one must use different critical values since the corresponding restricted estimate $(\widetilde{\rho}(\gamma_0), \widetilde{\alpha}(\gamma_0))$ changes. In practice this is accomplished on a fine grid of γ_0. In our extensive simulation experiments we have never observed a disjoint confidence set for an IRF. Even though I do not formally prove that the confidence set is an interval, if one agrees to use it as an assumption, then the task is to find the two ends of the confidence interval. This task can be solved with very high accuracy at a relatively low computational cost if one uses the bisection method. The OLS estimator always belongs to the confidence set, as LR statistic is equal to zero, and the corresponding hypothesis is accepted. If one wants to find the left end of the interval, one should move to the left from the OLS estimator with a step of length one until the hypothesis is rejected. In such a way one finds the interval of unit length containing the left end of the confidence set. Then one has to divide the resulting interval into halves, test the middle point and determine to which half the end point belongs. The division should be repeated several times until the desired accuracy is achieved.

[4] I am grateful to an anonymous referee for pointing this out.

The grid bootstrap approximates the finite-sample distribution of the LR statistic by $t^2(\widetilde{c}, \widetilde{u})$ and the distribution of $LR^{\pm}$ by $t(\widetilde{c}, \widetilde{u})$. Unlike the classical bootstrap, the grid bootstrap uses the restricted estimates $(\widetilde{\rho}, \widetilde{\alpha})$ to simulate the critical values, while the classical bootstrap uses $(\hat{\rho}_{OLS}, \hat{\alpha}_{OLS})$. OLS produces an always-inconsistent estimate of c with a large bias in a stationary direction, while the restricted estimate can be consistent in some situations, in particular when $u_T \to 0$.

Lemma 1 shows that the accuracy of the restricted estimate is directly connected with u. The impulse response at long horizons (small u) heavily depends on the persistence ρ, and thus, the null hypothesis about it contains a lot of information about ρ. This allows one to estimate c well, and the grid bootstrap critical values are close to correct. A hypothesis about a short horizon impulse response does not contain as much information about the persistence, but a good estimate of c is not as needed either, since both the sample and simulated statistics are asymptotically close to χ^2 (normal) anyway. In principle, problems could arise for medium horizons. Results from the simulation of the asymptotic size of the grid bootstrap reported in Section 6 show that the size distortions are relatively small.

4 Inferences about Impulse responses

In the previous sections, I circumvented non-linearity issues by assuming that the parameter of interest is a linear function of the coefficients. One of the implications of this assumption is that the Wald statistic is equal to the LR, and the t-statistic is equal to LR$^{\pm}$. However, the IRF is a very non-linear function of the coefficients, and the degree of non-linearity increases with the horizon. This section examines uniform asymptotic approximations of the LR and Wald statistics for IRFs. Our goal is to obtain an asymptotic approximation that is uniform over the parameter space, $\mathfrak{R}_\delta$, as well as over all possible horizons of the impulse response.

4.1 LR statistic for IRF

Let $f_k(\rho, \alpha)$ be the impulse response of an AR(p) process with parameters (ρ, α) at the horizon k; that is, $f_k(\rho, \alpha) = \frac{\partial y_{t+k}}{\partial e_t}$. Let $\mathcal{L}$ be the set of all impulse response functions: $\mathcal{L} = \{f_k(\rho, \alpha), k = 1, 2, ...\}$. Let $LR(f, \gamma_0)$ be the LR statistic and $LR^{\pm}(f, \gamma_0)$ be the LR$^{\pm}$

statistic for testing the null $H_0 : f(\rho, \alpha) = \gamma_0$. The theorem below shows that the LR test for IRFs is asymptotically uniformly approximated by the same family of distributions as LR statistics for the linearized hypothesis.

Theorem 3 *Assume $\{y_t\}$ is an AR(p) process defined as in (1) with error terms satisfying Assumption A. Assume the null hypothesis $H_0 : f(\rho, \alpha) = \gamma_0$ holds. Let $A = \frac{\partial f}{\partial(\rho,\alpha')}(\rho, \alpha)$ and u be calculated as in (8). Then*

$$\lim_{T\to\infty} \sup_{f\in\mathcal{L}} \sup_{(\rho,\alpha)\in\Re_\delta\cap H_0} \left| P_{\rho,\alpha}\left\{LR(f,\gamma_0) < x\right\} - P\left\{t(c,u)^2 < x\right\}\right| = 0,$$

$$\lim_{T\to\infty} \sup_{f\in\mathcal{L}} \sup_{(\rho,\alpha)\in\Re_\delta\cap H_0} \left| P_{\rho,\alpha}\left\{LR^{\pm}(f,\gamma_0) < x\right\} - P\left\{t(c,u) < x\right\}\right| = 0.$$

The non-linearity of the IRF does not matter asymptotically if the LR statistic is used. The same asymptotic approximation holds as for the linearized versions of IRFs. The most important feature of the above stated approximation is that the approximation holds uniformly over the horizon as well as over the parameter space; that is, it works well for short, long and medium horizons, no matter how one defines them.

The LR statistic has one obvious drawback; it is not robust to heteroskedasticity. One way to robustify the inferences to conditional heteroskedasticity is to consider the GMM-based distance metric statistic[5]

$$DM_T = Q_T(\tilde{\theta}) - Q_T(\hat{\theta}),$$

where $\theta = (\rho, \alpha')'$, $Q_T(\theta) = e(\theta)'X\Omega_T^{-1}X'e(\theta)$, $X = (X_{p+1}, ..., X_T)'$ is a $(T - p) \times p$ regressor matrix, $\Omega_T = \frac{1}{T}\sum_{t=p+1}^{T} X_t X_t' e_t^2(\hat{\theta})$, $e(\theta) = Y - X\theta$, $\hat{\theta}$ is the OLS estimate, and $\tilde{\theta}$ is the restricted estimate of θ. Many of the results pertaining to IRF inferences can be directly generalized to conditionally heteroskedastic processes. The interested reader should refer to the Supplementary Appendix for more details.

4.2 IRFs in a VAR with a potential unit root

In this section, some results of the paper are generalized to VAR systems in which at most one near unit root is present.

Consider a k-dimensional VAR(p) process

$$y_t = B_1 y_{t-1} + ... + B_p y_{t-p} + e_t. \quad (10)$$

[5] In the context of testing IRFs, the Distance Metric statistic was mentioned in Gospodinov (2004).

Imagine for simplicity that one knows the co-integration (near co-integration) relation and can locate the problematic root. That is, assume that the first component $y_{1,t}$ has a local-to-unity root, while all other components $y_{-1,t} = (y_{2,t}, ..., y_{k,t})'$ are strictly stationary. Formally, assume that the VAR lag polynomial $B(L) = I_k - B_1 L - ... - B_p L^p$ can be factorized in the following way $B(L) = (I_k - diag(\lambda, 0, ..., 0)L)\tilde{B}(L)$.

Assumption VAR1

(i) All roots of the characteristic polynomial $\tilde{B}$ lie strictly inside and are bounded away from the unit circle. In particular, the process x_t given by $\tilde{B}(L)x_t = e_t$ can be written as an MA (∞) process $x_t = \Theta(L)e_t = \sum_{j=0}^{\infty} \Theta_j e_{t-j}$ with MA coefficients satisfying the following condition: $\sum_{j=0}^{\infty} j\|\Theta_j\| < \infty$, where $\|\Theta_j\| = \sqrt{trace(\Theta_j \Theta_j')}$.

(ii) $y_t = \Lambda y_{t-1} + x_t, y_0 = 0$, where $\Lambda = diag(\lambda, ..., 0)$; that is, $y_{1,t} = \lambda y_{1,t-1} + x_{1,t}; y_{-1,t} = x_{-1,t}$. The problematic root λ is local-to-unity, $\lambda = \lambda_T = 1 - c/T$.

(iii) e_t is a martingale-difference sequence with respect to the sigma-algebra $\mathcal{F}_t$, with $E\left(e_t e_t' | \mathcal{F}_{t-1}\right) = \Omega$ and four finite moments.

The assumption above is a direct generalization of the local-to-unity asymptotic embedding to a multivariate setting. If Assumption VAR1 holds, the OLS estimator of the regression in (10) has non-standard asymptotic behavior due to some linear combination of coefficients being estimated super-consistently. A survey of local-to-unity multivariate models can be found in Phillips (1988).

Let us interest ourselves in testing a hypothesis about the coefficients $H_0 : f(B_1, ..., B_p) = 0$. The following statistic is a generalization of the LR statistic to the multi-dimensional case:

$$LR = T \cdot trace(\hat{\Omega}^{-1}(\tilde{\Omega} - \hat{\Omega})) \tag{11}$$

with $\Omega(B) = \frac{1}{T}\sum_{t=1}^{T}(B(L)y_t)(B(L)y_t)'$, $\hat{\Omega} = \Omega(\hat{B}), \tilde{\Omega} = \Omega(\tilde{B})$, where $\hat{B}$ is the OLS estimator of coefficients in regression (10), while

$$\tilde{B} = \arg \min_{B=(B_1,...,B_p):f(B)=0} Ttrace(\hat{\Omega}^{-1}(\hat{\Omega} - \Omega(B)))$$

is the restricted estimator.

Consider a hypothesis about the impulse response of the nearly-non-stationary series $y_{1,t}$ to the j^{th} shock at the horizon h, denote it $\theta_h = \frac{\partial y_{1,t+h}}{\partial e_{j,t}}$. Assume that the horizon

$h = [q\sqrt{T}]$ is increasing in proportion to $\sqrt{T}$. This embedding implies that u_T converges to a constant in the AR(p) case and delivers the mixture of local-to-unity and normal distributions as the limit distribution of the $\text{LR}^{\pm}$ statistic. Lemma 4 in the Supplementary Appendix points out that the linearized hypothesis about such an impulse response puts $\sqrt{T}-$increasing weight on the coefficients estimated super-consistently relative to the weights put on the asymptotically normal coefficients on the stationary regressors. Let $\tilde{A} = \frac{\partial \theta_h}{\partial B}$. Let the hypothesis $H_0 : \tilde{A}'B = \gamma_0$ be the linearized version of the hypothesis $H_0 : \theta_h = \gamma_0$.

Theorem 4 *Let y_t be a $k \times 1$ VAR(p) process satisfying Assumption VAR1. Assume that the linearized version of the hypothesis $H_0 : \theta_h = \frac{\partial y_{1,t+h}}{\partial e_{j,t}} = \gamma_0$ at the horizon $h_T = q\sqrt{T}$ is tested using the statistic defined in equation (11). Then $LR \Rightarrow (t(u,c))^2$ as $T \to \infty$ for some u.*

Theorem 4 states that in the multivariate VAR model with at most one local-to unity root the asymptotic behavior of the LR test statistic for the impulse response at horizon proportional to $\sqrt{T}$ is of the same nature as the corresponding statistic in the univariate AR(p).

4.3 Wald statistic for the IRF

It is well known that different statistics behave differently for very curved parameters. The t-statistic is not invariant to a monotonic transformation of the null-hypothesis or a re-parametrization of the model. Gregory and Veall (1985) showed that the results of the Wald test can change dramatically under a monotonic transformation of the null hypothesis. The LR test, however, is invariant to such transformations.[6] Hansen (2006) shows that in a classical OLS setting that the GMM-distance statistic (of which the LR statistic is a special case) is second-order better approximated by a χ^2-distribution when compared to a Wald statistic.

Below I show that the approximation as stated in Theorem 3 for the LR statistic does not hold for the Wald statistic. For simplicity, the problem is illustrated in an

[6]Due to the invariance of the LR statistic to monotonic transformations, the LR statistic for a test about an IRF in an AR(1) model is equal to the LR statistic for a test about the coefficient ρ only. The validity of the grid bootstrap for such a test in an AR(1) model is proven in Mikusheva (2007a).

AR(1) model. A treatment of the more general AR(p) case is left to the Supplementary Appendix.

Consider an AR(1) process in a classical setting; that is, $y_t = \rho y_{t-1} + e_t$, with $0 < \rho < 1$ being fixed, and the sample size growing to infinity. Consider an IRF $f_k(\rho) = \rho^k$ where the horizon is modeled as changing with the sample size $k_T = [\sqrt{T}]$. In this case $\frac{df_k(\rho)}{d\rho} = k\rho^{k-1}$, and I can write the t-statistic (Wald statistic equals a t-statistic squared).

$$t = \frac{\widehat{\rho}^{k_T} - \rho^{k_T}}{k_T \widehat{\rho}^{k_T - 1}} \sqrt{\frac{T}{1-\widehat{\rho}^2}} = \frac{\widehat{\rho}^{\sqrt{T}} - \rho^{\sqrt{T}}}{\widehat{\rho}^{\sqrt{T}-1}} \sqrt{\frac{1}{1-\widehat{\rho}^2}}.$$

Since $|\rho| < 1$ is fixed, $\widehat{\rho} \to^p \rho$ and $\sqrt{T}(\widehat{\rho} - \rho) \Rightarrow N(0, (1-\rho^2)^{-1})$. I can re-write the t-statistic as follows:

$$t = \left(1 - \left(1 + \frac{1}{\sqrt{T}} \frac{\sqrt{T}(\widehat{\rho} - \rho)}{\widehat{\rho}}\right)^{\sqrt{T}}\right) \frac{\rho}{\sqrt{1-\rho^2}}.$$

Since $(1 + \frac{x}{\sqrt{T}})^{\sqrt{T}} \to e^x$ as $T \to \infty$ uniformly over $|x| < C$ for any positive constant C, one obtains that

$$t \Rightarrow \left(\exp\left(\sqrt{\frac{1-\rho^2}{\rho^2}}\xi\right) - 1\right) \frac{\rho}{\sqrt{1-\rho^2}},$$

where $\xi \sim N(0,1)$. The limit distribution is not normal, as one might have expected. The limit distribution of the t-statistic does not belong to the family of distributions $t(u,c)$. This implies that the approximating family of distributions for the t-statistic (or Wald statistic) for a highly non-linear hypothesis may differ from that for a linear hypothesis. This section demonstrated that handling the t-statistic (Wald statistic) for a highly non-linear null hypothesis is a more delicate task than dealing with LR-type statistics. I leave finding approximations for Wald statistics for future research.

5 The Asymptotic Size of Different Methods

Section 4 establishes a uniform approximation for the LR, and $\text{LR}^{\pm}$ statistics for testing hypotheses about impulse responses. The suggested asymptotic approximation holds uniformly over the horizon of the impulse response. In this section, I discuss the *asymptotic* size properties of some known methods for impulse response inference.

Let $Y = \{y_1, ..., y_T\}$ be a sample from an AR(p) process described in (1) with error terms satisfying Assumption A. The hypothesis $H_0 : f(\rho, \alpha) = \gamma_0$ is tested, where $f \in \mathcal{A}$.

In our case, $\mathcal{A}$ is the set of impulse responses. Let $\psi(Y, f) \in \{0, 1\}$ be a test of H_0 (if $\psi(Y) = 1$ the test will reject the null). Then the finite sample size of ϕ is $E_{\rho,\alpha}\psi(Y, f)$. The uniform approximation obtained before allows one to characterize for many known tests how well the test maintains size asymptotically. The characterization is done through the function $\phi(c, u)$ such that:

$$\lim_{T\to\infty} \sup_{f\in\mathcal{A}} \sup_{(\rho,\alpha)\in\mathfrak{R}_\delta\cap H_0} |E_{\rho,\alpha}\psi(Y, f) - \phi(c, u)| = 0.$$

A test that maintains its size uniformly well for all parameter values and all functions from $\mathcal{A}$ would have $\phi(c, u) \leq \phi_0$ for all $c \leq 0$ and $u \geq 0$, where ϕ_0 is the declared size. It should be emphasized that $\phi(c, u)$ characterizes the *asymptotic* size of a procedure. The finite sample size properties are discussed in Section 6.

Table I provides the values of the parameter u for tests of the impulse response at different horizons in the AR(2) model. One can observe that u becomes very close to 0, i.e. the majority of the weight is put on the local-to-unity term, when the horizon increases, however, u is large at shorter horizons, especially when there is a positive second root. Due to the fact that the impulse response is a non-linear function of coefficients, the value of u depends on both roots. As Table I shows, u is an increasing function of both roots in the AR(2) case.

		$h=2$	$h=3$	$h=4$	$h=5$	$h=10$	$h=15$
$\lambda = 0.99$	$\mu = -0.6$	1.2	1.7	0.2	0.9	0.1	0.1
	$\mu = -0.3$	1.3	1.6	1.0	0.8	0.4	0.2
	$\mu = 0$	3.4	2.2	1.6	1.3	0.6	0.3
	$\mu = 0.3$	5.9	3.9	2.8	2.2	0.9	0.5
	$\mu = 0.6$	9.6	7.1	5.4	4.3	1.8	1.0
$\lambda = 0.98$	$\mu = -0.6$	1.0	1.2	0.3	0.5	0.0	0.1
	$\mu = -0.3$	0.8	1.0	0.6	0.5	0.1	0.0
	$\mu = 0$	2.4	1.5	1.1	0.8	0.3	0.1
	$\mu = 0.3$	4.1	2.7	1.9	1.5	0.5	0.3
	$\mu = 0.6$	6.8	4.9	3.8	2.9	1.2	0.6

Table I: Value of u for a hypothesis about the impulse response at horizon h for an AR(2) process with roots λ and μ. A simplified explicit formula for u in the AR(2) model can be found in the Supplementary Appendix.

Let $q^t_{\phi_0,i}(c, u)$ be the ϕ_0-equi-quantiles of statistics $t(c, u)$; that is, $P\{t(c, u) < q^t_{\phi_0,1}(c, u)\} = \phi_0/2$ and $P\{t(c, u) > q^t_{\phi_0,2}(c, u)\} = \phi_0/2$. Also let $q^W_{\phi_0}(c, u)$ be the $1 - \phi_0$ quantile of

$t^2(c,u)$, that is, $P\{t^2(c,u) > q^W_{\phi_0}(c,u)\} = \phi_0$.

5.1 Classical bootstrap

The classical bootstrap is the following procedure. One estimates the coefficients $(\widehat{\rho}, \widehat{\alpha})$ by OLS and calculates the test statistic (signed or squared). Then one simulates an AR(p) model with coefficients $(\widehat{\rho}, \widehat{\alpha})$, calculates the test statistic for the simulated model, and computes its quantiles.

As a direct corollary of Theorem 3 the asymptotic size for the bootstrap of the LR is

$$\phi(c,u) = P\left\{t^2(c,u) > q^W_{\phi_0}\left(\widehat{u}, \widehat{c}\right)\right\};$$

and for signed statistics is

$$\phi(c,u) = 1 - P\left\{q^t_{\phi_0,1}(\widehat{u}, \widehat{c}) < t(c,u) < q^t_{\phi_0,2}(\widehat{u}, \widehat{c})\right\}.$$

Here $\widehat{c} = c + \frac{\int_0^1 J_c(t)dW(t)}{\int_0^1 J_c^2(t)dt}$ is a random variable defined on the same probability space as $t(c,u)$, where $W(t)$ and $J_c(t)$ are the same as in the definition of $t(c,u)$ (see (7)), and $\widehat{u} = u\sqrt{\frac{\widehat{c}}{c}}$.

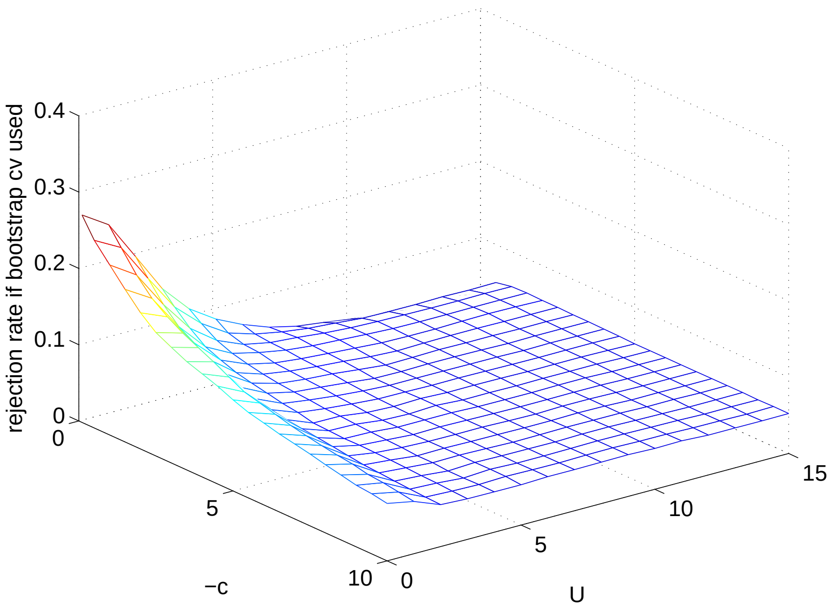

Figure 1: Rejection rate of the classical bootstrap LR test. The nominal significance level is 5%. A linear trend is assumed. Based on 5000 simulations.

It is interesting to note that even though the bootstrap was initially justified for impulse responses by proving that the t-statistic converges to a normal distribution, it

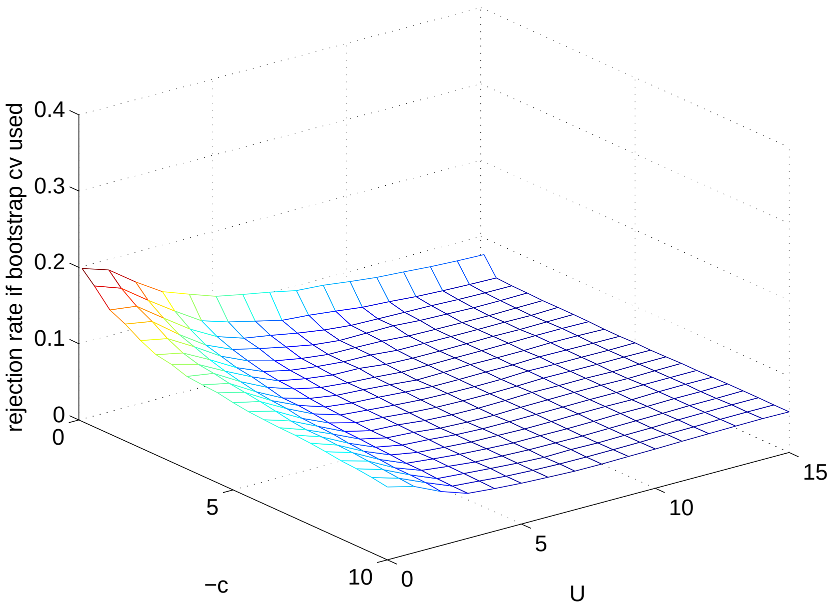

Figure 2: Rejection rate of the classical bootstrap $LR^{\pm}$ test. The nominal significance level is 5%. A linear trend is assumed. Based on 5000 simulations from an asymptotic distribution.

eventually uses approximation with both terms through simulations. It employs the OLS estimates of c and u, rather than the true values (which are not known). It is well-known that the OLS estimate of c is not consistent. In fact, $\hat{c}$ is asymptotically highly biased to the left, thus the estimated model looks more stationary than it actually is. The OLS estimate $\hat{u}$ tends to somewhat overestimate u, and thus, it puts more weight on the stationary component than it should. Both facts lead to size distortions, which become unimportant if either $c \to -\infty$ (very stationary series) or $u \to \infty$ (hypothesis is about stationary coefficients only). The function $\phi(c, u)$ for the bootstrap applied to signed and squared statistics is depicted in Figures 1 and 2. It has very significant distortions.

5.2 Gospodinov's method

Here I consider a method for constructing a confidence set that was suggested by Gospodinov (2004). For the null hypothesis $H_0 : f(\rho, \alpha) = \gamma_0$, one calculates the LR test statistic (signed or unsigned) and the restricted estimates $(\widetilde{\rho}, \widetilde{\alpha})$. One also calculates the implied value of the local-to-unity parameter $\widetilde{c}$ corresponding to the restricted estimates of the coefficients. The test compares the value of the statistic with the quantiles of the local-to-unity limit t^c (or $(t^c)^2$) evaluated at $\widetilde{c}$. Gospodinov's method uses only $t^c = t(c, u = 0)$

ignoring the stationary part of $t(c, u)$.

Gospodinov (2004) based his method on asymptotic approximations obtained under the assumption that the length of the horizon is proportional to the sample size, $k_T = [\delta T]$. This corresponds in our notation to $u_T \to 0$, and it leads to using quantiles of $t^c = t(c, u = 0)$. Gospodinov (2004) checks the robustness of the method in a Monte-Carlo study, and he finds that his LR test becomes conservative on short horizons. He also notices that the rejections of the LR tests are mainly one-sided and suggests using the signed statistic LR$^{\pm}$.

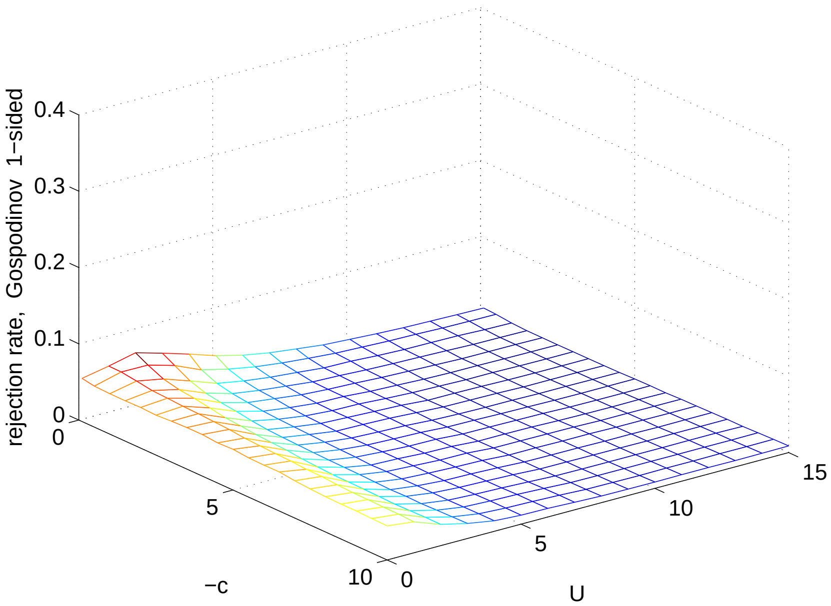

Figure 3: Rejection rate of Gospodinov's LR test. The nominal significance level is 5%. A linear trend is assumed. Based on 5000 simulations.

As a direct corollary of Theorems 2 and 3 and Lemma 1, one has that for the LR test

$$\phi(c, u) = P\{t^2(c, u) \geq q^W_{\phi_0}(\widetilde{c}, u = 0)\}$$

and for LR$^{\pm}$:

$$\phi(c, u) = 1 - P\{q^t_{\phi_0,1}(\widetilde{c}, u = 0) \leq t(c, u) \leq q^t_{\phi_0,2}(\widetilde{c}, u = 0)\},$$

where $\widetilde{c}$ is defined as the right side of (9).

To understand the size properties of Gospodinov's method several facts should be taken into account: 1) if $u \neq 0$, then $\widetilde{c}$ is biased negatively; 2) the distribution of $(t^c)^2$ first-order stochastically dominates the distribution of $(t^{c_1})^2$ if $-c < -c_1$; 3) the distribution of

$(t^c)^2$ first-order stochastically dominates χ^2_1; 4) distributions t^c and $N(0,1)$ have different locations.

First, let us consider the LR statistic. From 1) and 2) one might expect that there would be slight over-rejection for small non-zero values of u, since for those values of u the persistence is slightly underestimated. For relatively large values of u (that correspond to moderate or short horizons) the true distribution of the LR statistic is a squared mixture of t^c and normal, but the quantiles of the local-to-unity distribution are used, which in light of fact 3) are larger. As a result, Gospodinov's method for the LR statistic is conservative at medium and short horizons. This fact was observed before in simulations, but only now has received any explanation.

Now let me turn to the signed statistic LR$^{\pm}$. In light of fact 4), one should expect that Gospodinov's (LR$^{\pm}$) should over-reject when u is high (short horizons) and c is close to 0 (persistent case). The rejection rate increases with u and at the limit $(c = 0, u = \infty)$ becomes close to 40% (for a 5% declared level). The conservativeness of the method for the LR and over-rejection of the $LR^{\pm}$ are demonstrated on Figures 3 and 4.

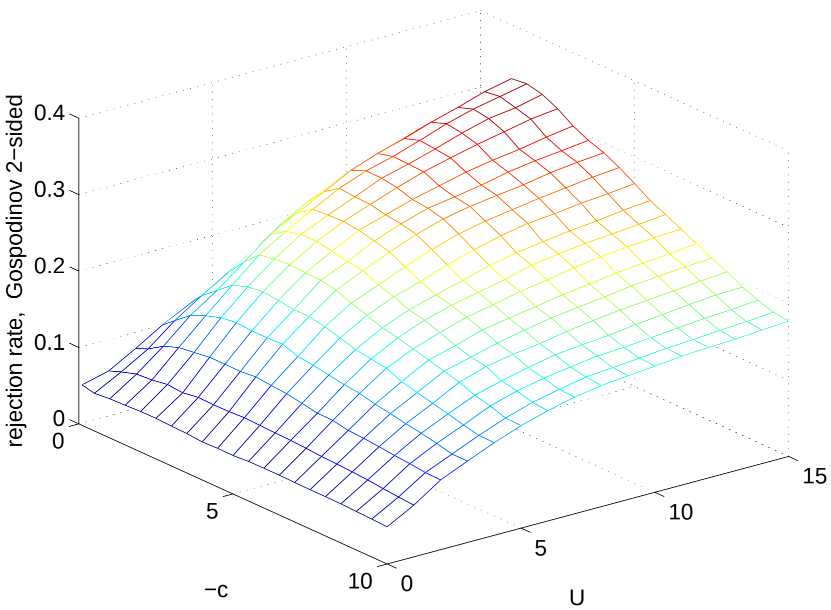

Figure 4: Rejection rate of Gospodinov's LR$^{\pm}$ test. The nominal significance level is 5%. A linear trend is assumed. Based on 5000 simulations.

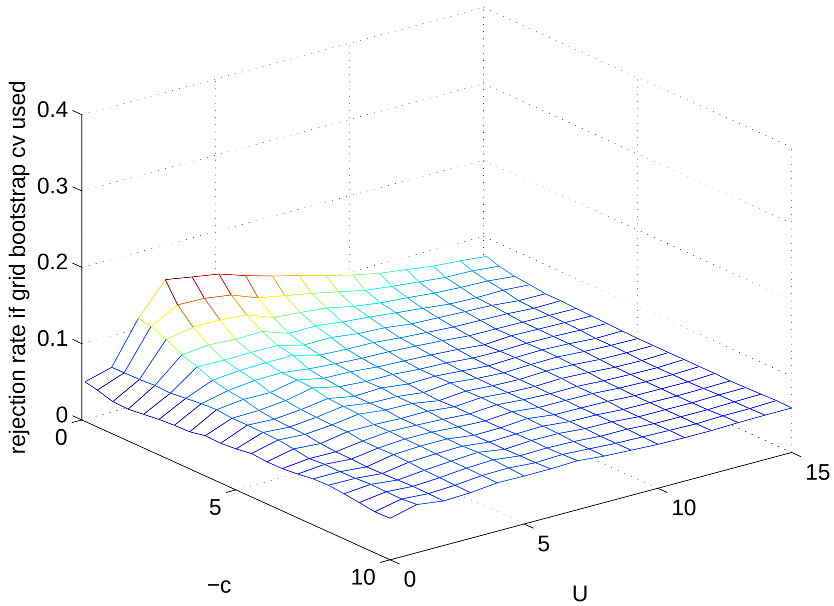

Figure 5: Rejection rate of grid bootstrap LR test. The nominal significance level is 5%. A linear trend is assumed. Based on 5000 simulations.

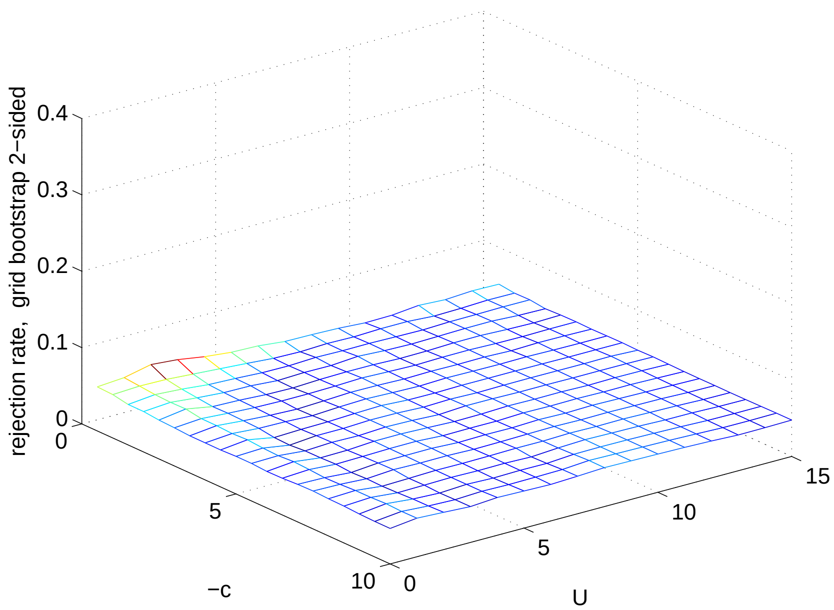

Figure 6: Rejection rate of grid bootstrap $LR^{\pm}$ test. The nominal significance level is 5%. A linear trend is assumed. Based on 5000 simulations.

5.3 Grid bootstrap

As we have seen above, the classical bootstrap approximates the finite sample distribution of the test statistic by a mixture of the $t(c, u)$-type, but it uses a very biased estimate of c ($\hat{c}_{OLS}$). The bias of the estimate is particularly bad when the process is very persistent. The highest distortions of the classical bootstrap happen when the local-to-unity distribution has the highest weight ($u = 0$). Gospodinov's method uses a better estimate of c, the restricted estimate $\widetilde{c}$. The estimate $\widetilde{c}$ is especially good when u is very close to 0 (Lemma 1). The idea of marrying the two methods in order to fix the drawbacks of each is realized in the grid bootstrap. The procedure is described in Section 3.3, and it approximates the unknown finite-sample distribution of the test statistic by $t^2(\widetilde{c}, \widetilde{u})$ for squared and by $t(\widetilde{c}, \widetilde{u})$ for signed statistics.

Maximal actual size	Nominal size		
	0.025	0.05	0.1
1-sided(LR)	0.052	0.099	0.18
2-sided ($LR^{\pm}$)	0.036	0.064	0.12

Table II: Maximal over u and c asymptotic size of grid bootstrap procedure for a given nominal size. Based on 10000 simulations.

The potential usefulness of the grid bootstrap is that if $0 < u < \infty$, then the asymptotic distribution of both the sampled and simulated statistics will have the classical and local-to-unity terms, unlike the delta-method or Gospodinov's method that have only one of the two terms.

Note that if u is close to zero, the estimator $\widetilde{c}$ is extremely precise, and the critical values are close to being correct. If u is large enough, then both the sample and simulated statistics are asymptotically close to χ^2 (normal), and the coverage is close to the declared coverage. In principle, problems could arise for intermediate values of u. For a series of experiments that keeps $u \neq 0$ constant, the estimate of $\tilde{c}$ is not consistent. As a result, the grid bootstrap does not deliver the asymptotically correct size uniformly, if uniformity is required over two directions simultaneously: over the AR parameters and over the hypothesis tested (in this case the horizon of the IRF). As shown in this paper, no existing procedure is uniform over both dimensions. What we show below is that the size distortions of the grid bootstrap are very small and in this sense the grid bootstrap dominates all currently known procedures. Moreover, the grid bootstrap is uniformly asymptotically correct if uniformity is required only with respect to the AR parameters.

Simulation results of the *asymptotic* size of the grid bootstrap appear in Figures 5 and 6. One can see that the grid bootstrap does not to maintain the size uniformly well, but at the same time the size distortions are relatively small. As mentioned above, due to the asymmetry of the limit distribution, the $LR^{\pm}$-statistic based equi-tailed interval has better size properties than the symmetric one. The maximum asymptotic size of the grid bootstrap at different nominal significance levels is reported in Table II.

6 Simulation study.

The comprehensive simulation study of Pesavento and Rossi (2007) compared coverage (size) of classical methods for performing inference on impulse responses such as the delta-method, pretest and bootstrap, as well as methods based on a linearly-growing horizon such as Wright (2000), Gospodinov (2004) and Pesavento and Rossi (2006). Many of the observations of Pesavento and Rossi (2007) are explained and theoretically justified in previous sections.

The main aim of this section is to complete and refine Pesavento and Rossi's (2007) study by exploring the finite-sample properties of the grid bootstrap introduced in this paper. I intend to answer the following two questions: what are the finite-sample size (coverage) properties of the grid bootstrap? If the grid bootstrap is better at controlling size, will it provide any gain in terms of length of the corresponding confidence sets when compared with conservative procedures?

I compare the grid bootstrap with Gospodinov's method. In Table III, I report the coverage for sample size of $T = 500$, while in Table IV the coverage and the average length of confidence sets for a sample size of $T = 250$ are reported. In both cases, the samples are generated from an AR(2). The largest root μ is chosen to produce the same local parameter $c = 5$ in both cases: in particular, $\mu = 0.99$ for Table III and $\mu = 0.98$ for Table IV. The second root takes values inside the unit circle $\lambda = -0.6, -0.3, 0, 0.3, 0.6$. A linear time trend is assumed. I denote as h the horizon of the corresponding impulse response. The nominal level of all procedures is 90%. For calculation of coverage it is enough to merely simulate the acceptance rate for the test of the true value; no grid search is needed. I report coverage probabilities based on 1000 simulations. The average length of confidence sets in Table IV is calculated based on 100 simulations.

AR roots	method	statistics	horizon						
			h=1	h=3	h=5	h=10	h=15	h=20	h=30
$\mu = 0.99$	Grid bootstrap	LR	0.90	0.87	0.87	0.90	0.90	0.90	0.89
$\lambda = -0.6$		$LR^{\pm}$	0.91	0.90	0.89	0.90	0.92	0.91	0.89
	Gospodinov	LR	0.95	0.92	0.89	0.89	0.90	0.90	0.88
		$LR^{\pm}$	0.77	0.82	0.87	0.90	0.91	0.90	0.89
$\mu = 0.99$	Grid bootstrap	LR	0.90	0.89	0.87	0.89	0.90	0.89	0.88
$\lambda = -0.3$		$LR^{\pm}$	0.89	0.91	0.89	0.91	0.90	0.91	0.88
	Gospodinov	LR	0.96	0.91	0.90	0.90	0.90	0.89	0.88
		$LR^{\pm}$	0.73	0.84	0.84	0.86	0.89	0.91	0.88
$\mu = 0.99$	Grid bootstrap	LR	0.89	0.89	0.87	0.88	0.90	0.90	0.90
$\lambda = 0$		$LR^{\pm}$	0.89	0.90	0.90	0.90	0.90	0.90	0.90
	Gospodinov	LR	0.96	0.94	0.92	0.89	0.89	0.90	0.89
		$LR^{\pm}$	0.74	0.80	0.82	0.86	0.87	0.89	0.89
$\mu = 0.99$	Grid bootstrap	LR	0.90	0.89	0.90	0.86	0.90	0.89	0.91
$\lambda = 0.3$		$LR^{\pm}$	0.91	0.90	0.92	0.88	0.90	0.90	0.91
	Gospodinov	LR	0.97	0.95	0.95	0.88	0.90	0.89	0.90
		$LR^{\pm}$	0.72	0.77	0.80	0.80	0.88	0.89	0.90
$\mu = 0.99$	Grid bootstrap	LR	0.89	0.91	0.88	0.88	0.89	0.88	0.90
$\lambda = 0.6$		$LR^{\pm}$	0.89	0.90	0.88	0.91	0.91	0.90	0.91
	Gospodinov	LR	0.96	0.95	0.92	0.89	0.89	0.87	0.89
		$LR^{\pm}$	0.74	0.75	0.79	0.84	0.88	0.87	0.90

Table III: Comparison of the finite-sample coverage for Gospodinov's method and the grid bootstrap (LR and $LR^{\pm}$ statistics). Samples of size $T = 500$ are taken from an AR(2) with roots μ and λ and standard normal errors. A linear trend is assumed. The nominal coverage of all procedures is 90%. Based on 1000 simulations.

According to our theory, the grid bootstrap should control size very well for short and long horizons but display some small distortions on medium horizons. As expected, the grid bootstrap controls size relatively well for both LR and its signed version, $LR^{\pm}$. The results of Gospodinov's method are essentially indistinguishable from those of the grid bootstrap on long horizons for both statistics. The horizon for which the results of the two methods start to look alike depends on the both roots, for Table III it happens around horizons 10 or 15, while for Table IV it starts from horizons 5 or 10. The two procedures produce more close results if the second root is negative. On short horizons Gospodinov's method has size distortions for the $LR^{\pm}$ statistic, while it is somewhat conservative for the LR statistic, as expected from Section 5.2.

			Gospodinov's coverage		Grid bootstrap coverage		Gospodinov's length		Grid bootstrap length	
μ	λ	h	LR	$LR^{\pm}$	LR	$LR^{\pm}$	LR	$LR^{\pm}$	LR	$LR^{\pm}$
0.98	-0.6	1	0.95	0.77	0.90	0.90	0.24	0.17	0.20	0.19
		3	0.90	0.84	0.88	0.89	0.25	0.19	0.24	0.20
		5	0.90	0.90	0.90	0.91	0.26	0.20	0.26	0.21
		10	0.88	0.88	0.89	0.89	0.36	0.29	0.36	0.29
		15	0.89	0.88	0.89	0.89	0.42	0.35	0.43	0.35
0.98	-0.3	1	0.96	0.76	0.89	0.90	0.27	0.19	0.22	0.19
		3	0.89	0.88	0.88	0.91	0.26	0.20	0.25	0.21
		5	0.89	0.89	0.89	0.91	0.31	0.25	0.31	0.25
		10	0.90	0.90	0.90	0.91	0.44	0.35	0.44	0.35
		15	0.88	0.88	0.89	0.89	0.54	0.46	0.54	0.47
0.98	0	1	0.95	0.75	0.89	0.89	0.28	0.20	0.22	0.21
		3	0.88	0.83	0.86	0.89	0.33	0.27	0.31	0.27
		5	0.91	0.86	0.91	0.91	0.43	0.36	0.42	0.37
		10	0.91	0.90	0.92	0.91	0.61	0.52	0.62	0.53
		15	0.89	0.90	0.90	0.91	0.70	0.60	0.72	0.61
0.98	0.3	1	0.95	0.74	0.89	0.89	0.27	0.18	0.22	0.19
		3	0.92	0.81	0.89	0.88	0.50	0.41	0.45	0.41
		5	0.89	0.85	0.88	0.91	0.64	0.57	0.60	0.56
		10	0.88	0.89	0.89	0.91	0.87	0.73	0.88	0.73
		15	0.89	0.90	0.91	0.91	1.02	0.89	1.05	0.90
0.98	0.6	1	0.93	0.75	0.88	0.89	0.22	0.15	0.17	0.18
		3	0.94	0.79	0.89	0.91	0.70	0.53	0.59	0.56
		5	0.90	0.80	0.88	0.89	1.02	0.85	0.94	0.87
		10	0.86	0.86	0.88	0.91	1.56	1.49	1.55	1.48
		15	0.90	0.89	0.92	0.92	1.86	1.69	1.91	1.69

Table IV: Comparison of Gospodinov's method and the grid bootstrap in terms of the coverage and length of the corresponding confidence set. The samples of size $T = 100$ are taken from an AR(2) with roots μ and λ and standard normal errors. A linear trend is present. The nominal coverage of all procedures is 90% Number of simulations for coverage is $N = 1000$, and for the length $N = 100$.

Table IV answers our second question comparing the length of the resulting confidence sets. Again I compare the grid bootstrap with Gospodinov's method. One can make several observations. First, using the $LR^{\pm}$ rather than the LR statistic is beneficial in terms of the length of the confidence set. It is due to the asymmetry of the signed statistic $LR^{\pm}$, which translates to the observation made by Gospodinov that almost all rejections for LR are made on one side of the alternative. Second, the length of the grid bootstrap LR confidence set at horizon 1 is around 20% shorter than that of Gospodinov's LR confidence set. The difference in the length between the two methods for the LR statistic essentially disappears after horizon 3. The third observation is that the grid bootstrap $LR^{\pm}$ in most cases produces intervals of the same average length as Gospodinov's $LR^{\pm}$ method, while coverage of the grid bootstrap $LR^{\pm}$ confidence set is much better than that of Gospodinov's $LR^{\pm}$ due to better centering.

7 Appendix.

Lemma 2 *Let F be a lower-triangular matrix as defined in Section 2.3. Then there exist constants $C_1 > 0$ and $C_2 < \infty$ such that for all $(\rho, \alpha) \in \mathfrak{R}_\delta$ the following holds:*

(i) $F_{11} = \frac{\sqrt{1-\rho^2}}{C(\rho,\alpha)}$, where $C_1 < C(\rho,\alpha) < C_2$, and $\lim_{\rho\to 1} C^2(\rho,\alpha) = \frac{1}{(1-\lambda_1)...(1-\lambda_{p-1})}$;

(ii) $F_{1i} = (1-\rho^2)a_{i-1}(\rho,\alpha)$, where $|a_i(\rho,\alpha)| < C_2$.

Proof of Lemma 2. By definition,

$$1 - \rho L + \sum_{j=1}^{p-1} \alpha_j L^j (1-L) = (1-\lambda_1 L)...(1-\lambda_p L)$$

with $|\lambda_1| \leq ... \leq |\lambda_p|$. Notice that $1-\rho = (1-\lambda_1)...(1-\lambda_p)$. I divide the parameter space $\mathfrak{R}_\delta$ into two regions $A_T = \mathfrak{R}_\delta \cap (|\lambda_p| > \delta + \varepsilon)$ and $B_T = \mathfrak{R}_\delta \cap (|\lambda_p| \leq \delta + \varepsilon)$.

Assume that $(\rho, \alpha) \in A_T$. Then λ_p is real and positive (see the definition of $\mathfrak{R}_\delta$). Consider a strictly stationary process u_t with $(1-\lambda_1 L)...(1-\lambda_{p-1}L)u_t = e_t$, and $\widetilde{y}_t = \lambda_p \widetilde{y}_{t-1} + u_t, \widetilde{y}_0 = 0$. The process u_t has absolutely summable auto-covariances $\widetilde{\gamma}_k$: $\sum_{i=0}^{\infty} |\widetilde{\gamma}_j| < C(\delta) < \infty$ (Lemma 8 (b) in Mikusheva (2007b)).Then,

$$\lim_{t\to\infty} Var(\widetilde{y}_t) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \lambda_p^{i+j} E(u_{t-j}u_{t-i}) = 2\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty} \lambda_p^{i+j}\widetilde{\gamma}_{j-i} + \sum_{i=0}^{\infty} \lambda_p^{2i}\widetilde{\gamma}_0 =$$
$$= \frac{1}{1-\lambda_p^2}(\widetilde{\gamma}_0 + 2\sum_{k=1}^{\infty} \lambda_p^k \widetilde{\gamma}_k)$$

It is easy to show that the expression $\widetilde{\gamma}_0 + 2\sum_{k=1}^{\infty} \lambda_p^k \widetilde{\gamma}_k$ is bounded from above and away from zero uniformly over $\mathfrak{R}_\delta$.

As a result

$$F_{11} = \frac{1}{\sqrt{\lim_{t\to\infty} Var(\widetilde{y}_t)}} = \sqrt{1-\lambda_p^2}(\widetilde{\gamma}_0 + 2\sum_{k=1}^{\infty} \lambda_p^k \widetilde{\gamma}_k)^{-1/2} = \frac{\sqrt{1-\rho^2}}{C(\rho,\alpha)}$$

$C(\rho,\alpha)^2 = (1-\lambda_1)...(1-\lambda_{p-1})\frac{1+\rho}{1+\lambda_p}(\widetilde{\gamma}_0 + 2\sum_{k=1}^{\infty}\lambda_p^k\widetilde{\gamma}_k)$. Given that $|\lambda_i| < \delta$ for $i = 1,...,p-1$ and $\lambda_p > \delta+\varepsilon$ it is easy to see that $C(\rho,\alpha)$ is uniformly bounded and separated from zero uniformly over A_T.

Notice that for $\rho \to 1$ one has $\lambda_p \to 1$ and

$$\widetilde{\gamma}_0 + 2\sum_{k=1}^{\infty}\lambda_p^k\widetilde{\gamma}_k \to \widetilde{\gamma}_0 + 2\sum_{k=1}^{\infty}\widetilde{\gamma}_k = \frac{1}{(1-\lambda_1)^2...(1-\lambda_{p-1})^2}$$

The last equality is a formula for the long-run variance for an AR(p-1) process u_t. From the formula for $C(\rho,\alpha)$, one has $C(\rho,\alpha)^2 \to \frac{1}{(1-\lambda_1)...(1-\lambda_{p-1})}$.

Next consider $(\rho,\alpha) \in B_T$. Then $(1-\rho) > (1-(\delta+\varepsilon))^p$. Since $\lim_{t\to\infty} Var(y_t)$ is uniformly bounded and separated from zero uniformly over B_T(Lemma 8 in Mikusheva (2007b)), the first half of statement (i) holds uniformly over B_T.

Part (ii): Again consider $(\rho,\alpha) \in A_T$. The matrix F is obtained as the limit (as $t\to\infty$) in a process of orthogonalization of $(\widetilde{y}_{t-1}, \Delta\widetilde{y}_{t-1}, ..., \Delta\widetilde{y}_{t-p+1})$, and as a result, $F_{1i+1} = \lim \frac{cov(\widetilde{y}_{t-1},\Delta\widetilde{y}_{t-i})}{Var(\widetilde{y}_{t-1})}$, $i = 1,...,p-1$. Notice that $\Delta\widetilde{y}_{t-j} = \widetilde{y}_{t-j} - \widetilde{y}_{t-j-1} = u_{t-j} - (1-\lambda_p)\widetilde{y}_{t-j-1}$:

$$\begin{aligned} F_{1i+1} =& \frac{E(\widetilde{y}_{t-1}u_{t-j})}{Var(\widetilde{y}_t)} - (1-\lambda_p)\frac{cov(\widetilde{y}_{t-1},\widetilde{y}_{t-j-1})}{Var(\widetilde{y}_{t-1})} = \\ =&(1-\lambda_p)\left(\frac{(1+\lambda_p)E(\widetilde{y}_{t-1}u_{t-j})}{\widetilde{\gamma}_0 + 2\sum_{k=1}^{\infty}\lambda_p^k\widetilde{\gamma}_k} - corr(\widetilde{y}_{t-1},\widetilde{y}_{t-j-1})\right) = (1-\rho^2)a_i. \end{aligned}$$

So,

$$a_j = \left(\frac{(1+\lambda_p)E(\widetilde{y}_{t-1}u_{t-j})}{\widetilde{\gamma}_0 + 2\sum_{k=1}^{\infty}\lambda_p^k\widetilde{\gamma}_k} - corr(\widetilde{y}_{t-1},\widetilde{y}_{t-j-1})\right)\frac{1}{(1-\lambda_1)...(1-\lambda_{p-1})}.$$

Since $|corr(\widetilde{y}_{t-1},\widetilde{y}_{t-j-1})| \le 1$,

$$|E(\widetilde{y}_{t-1}u_{t-j})| = |\sum_{i=0}^{\infty}\lambda_p^i\widetilde{\gamma}_{j-i-1}| \le \sum_{i=0}^{\infty}|\widetilde{\gamma}_{j-i-1}| < 2C(\delta).$$

This proves (ii).

Proof of Theorem 1. I divide the space into two regions, as in Mikusheva (2007a), $\mathcal{A}_T = \{(\rho,\alpha) \in \mathfrak{R}_\delta : |1-\rho| < T^{-1+\varepsilon})\}$ for some fixed and positive $\varepsilon > 0$ and $\mathcal{B}_T = \{(\rho,\alpha) \in \mathfrak{R}_\delta : |1-\rho| > T^{-1+\varepsilon})\}$. The former is called the local-to-unity region, and the latter is the stationary region. Lemma 12 in Mikusheva (2007b) states that uniformly over the stationary region $\mathcal{B}_T$ the classical CLT and Law of Large Numbers hold. Namely

$$\lim_{T\to\infty}\sup_{(\rho,\alpha)\in\mathcal{B}_T}\sup_{x\in\mathbb{R}^p}\left|P_{\rho,\alpha}\left\{\frac{1}{\sqrt{T}}\sum_t^{T-j}FX_te_t < x\right\} - P\{N(0,I_p) < x\}\right| = 0,$$

and for any $x > 0$

$$\lim_{T\to\infty} \sup_{(\rho,\alpha)\in\mathcal{B}_T} P_{\rho,\alpha}\left\{\left\|\frac{1}{T}\sum_{t}^{T-j} FX_tX_t'F' - I_p\right\|_2 > x\right\} = 0.$$

According to Phillips (1987)

$$\sqrt{-2c}\int_0^1 J_c(t)dw(t) \Rightarrow N(0,1) \text{ and } (-2c)\int_0^1 J_c^2(t)dt \to^p 1 \text{ as } c \to -\infty.$$

We also have that $g(c)(-2c) \to 1$ as $c \to -\infty$. Finally, notice that for $c(T,\rho,\alpha) = T\log(|\lambda_p|)$, $\inf_{(\rho,\alpha)\in\mathcal{B}_T} c(T,\rho,\alpha) \to -\infty$ as $T \to \infty$. Putting all of the statements above together one can see that (4) and (5) hold uniformly over the stationary region $\mathcal{B}_T$.

Now let us turn to the local-to-unity region $\mathcal{A}_T$. Since $\rho \to 1$ uniformly over $\mathcal{A}_T$ one can guarantee that for some large T the root $|\lambda_p| > \delta$ and thus it is not complex and is positive (see the definition of the parameter space $\mathfrak{R}_\delta$). As a result, one can represent our process as $y_t = \lambda_p y_{t-1} + u_t$ with $(1-\lambda_1 L)...(1-\lambda_{p-1}L)u_t = e_t$. I apply Lemma 3.1 of Phillips (2007) to u_t. It says that there exists on some expanded probability space a realization of all random variables u_t and $S_k = \sum_{t=1}^k u_t$ and a Brownian motion $B(\cdot)$ with variance ω^2 that is equal to the long-run variance of u_t, such that

$$\sup_{0\le k\le T}\left|\frac{S_k}{\sqrt{T}} - B(k/T)\right| = o_p(T^{-1/2+1/4}). \tag{12}$$

In our case $\omega^2 = \frac{1}{(1-\lambda_1)^2...(1-\lambda_{p-1})^2} = \lim_{\rho\to 1} C^4(\rho,\alpha)$.

From (12) one obtains statements (g) and (h) of Lemma 4 in Mikusheva (2007a) and following the proof of Lemma 5 (d) and (e) in Mikusheva (2007a), one obtains that uniformly over $\mathcal{A}_T$:

$$\frac{1-\lambda_p^2}{T}\sum_{t=1}^T y_{t-1}^2 \cong \omega^2\frac{1}{g(c)}\int_0^1 J_c^2(t)dt,$$

and for any fixed $j \ge 1$

$$\sqrt{\frac{1-\lambda_p^2}{T}}\sum_{t=1}^T y_{t-j}u_t \cong \omega^2\frac{1}{\sqrt{g(c)}}\int_0^1 J_c(t)dw(t). \tag{13}$$

I note that $\frac{(1-\rho^2)\omega^2}{C(\rho,\alpha)^2(1-\lambda_p^2)} \to 1$ uniformly over $\mathcal{A}_T$. For the last statement I observe that $\rho \to 1$ and thus $\lambda_p \to 1$ over $\mathcal{A}_T$, $(1-\rho^2) = (1+\rho)(1-\lambda_1)...(1-\lambda_p)$, and I use statement (i) of Lemma 2 and the formula for ω. I obtain that the following statement holds uniformly over $\mathcal{A}_T$:

$$\frac{1-\rho^2}{TC(\rho,\alpha)^2}\sum_{t=1}^T y_{t-1}^2 \cong \frac{1}{g(c)}\int_0^1 J_c^2(t)dt. \tag{14}$$

Now I prove that the following statement holds uniformly over $\mathcal{A}_T$:

$$\frac{\sqrt{1-\rho^2}}{\sqrt{T}C(\rho,\alpha)}\sum_{t=1}^{T} y_{t-1}e_t \cong \frac{1}{\sqrt{g(c)}}\int_0^1 J_c(t)dw(t). \tag{15}$$

Let us define β_i in the following way: $(1-\lambda_1 L)...(1-\lambda_{p-1}L) = 1-\sum_{j=1}^{p-1}\beta_j L^j$. Then $e_t = u_t - \sum_{j=1}^{p-1}\beta_j u_{t-j}$. We also know that $y_{t-1} = \lambda_p^p y_{t-p-1} + \sum_{j=1}^{p}\lambda_p^{j-1}u_{t-j}$. As a result,

$$\begin{aligned}\sum_{t=p+1}^{T} y_{t-1}e_t =& \lambda_p^p \sum_{t=p+1}^{T} y_{t-p-1}e_t + \sum_{j=1}^{p}\lambda_p^j \sum_{t=p+1}^{T} e_t u_{t-j} = \\ =& \lambda_p^p \sum_{t=p+1}^{T} y_{t-p-1}(u_t - \sum_{j=1}^{p-1}\beta_j u_{t-j}) + \sum_{j=1}^{p}\lambda_p^j \sum_{t=p+1}^{T} e_t u_{t-j}\end{aligned}$$

Now one would notice that $Ee_tu_{t-j} = 0$, for any $j \geq 1$ and the series u_t is stationary while e_t is i.i.d., so the Central Limit Theorem holds. It means that uniformly over $\mathfrak{R}_\delta$ values we have $\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} e_t u_{t-j} = O_p(1)$. Given that $\frac{\sqrt{1-\rho^2}}{C(\rho,\alpha)} \to 0$ uniformly over $\mathcal{A}_T$ one obtains that

$$\frac{\sqrt{1-\rho^2}}{\sqrt{T}C(\rho,\alpha)}\sum_{t=1}^{T} y_{t-1}e_t = \frac{\sqrt{1-\rho^2}}{\sqrt{T}C(\rho,\alpha)}\lambda_p^p \sum_{t=j+1}^{T} y_{t-p-1}(u_t - \sum_{j=1}^{p-1}\beta_j u_{t-j}) + o_p(1). \tag{16}$$

One can notice that $\frac{\sqrt{1-\rho^2}}{C(\rho,\alpha)}\lambda_p^p(1-\sum_{j=1}^{p-1}\beta_j)\frac{\omega^2}{\sqrt{1-\lambda_p^2}} \to 1$ uniformly over $\mathcal{A}_T$. Indeed, $\frac{\sqrt{1-\rho^2}}{\sqrt{1-\lambda_p^2}} \to \sqrt{(1-\lambda_1)...(1-\lambda_{p-1})}$ uniformly over $\mathcal{A}_T$, and

$$1-\sum_{j=1}^{p-1}\beta_j = (1-\lambda_1)...(1-\lambda_{p-1}) = \frac{1}{\omega} = \lim_{\rho\to 1}\frac{1}{C(\rho,\alpha)^2}.$$

Putting the last statement together with (16) and (13) one obtains that (15) holds uniformly over $\mathcal{A}_T$.

Lemma 11 of Mikusheva (2007b) (in particular, statements (f), (h), (i)) implies the following:

$$\lim_{T\to\infty}\sup_{(\rho,\alpha)\in\mathcal{A}_T}\sup_{x\in\mathbb{R}^p}\left|P_{\rho,\alpha}\left\{\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} Z_t e_t < x\right\} - P\{N(0,\Gamma)<x\}\right| = 0, \tag{17}$$

$$\lim_{T\to\infty}\sup_{(\rho,\alpha)\in\mathcal{A}_T} P_{\rho,\alpha}\left\{\left\|\frac{\sqrt{1-\rho^2}}{T}\sum_{t=p+1}^{T} Z_t y_{t-1}\right\|_2 > x\right\} = 0, \tag{18}$$

$$\lim_{T\to\infty}\sup_{(\rho,\alpha)\in\mathcal{A}_T} P_{\rho,\alpha}\left\{\left\|\frac{1}{T}\sum_{t=p+1}^{T} Z_t Z_t' - \Gamma\right\|_2 > x\right\} = 0, \tag{19}$$

and for any $x > 0$. Here $Z_t = (\Delta y_{t-1}, ..., \Delta y_{t-p+1})'$ and $\Gamma = \lim_{t\to\infty} EZ_tZ_t'$. Let $\tilde{x}_t$ be the last $p-1$ components of FX_t. According to statement (ii) of Lemma 2

$$\widetilde{x}_t = (1-\rho^2)a(\rho,\alpha)y_{t-1} + \widetilde{F}Z_t, \tag{20}$$

where $a(\rho,\alpha)$ is a vector with $\sup_{\mathcal{A}_T} \|a_i\|_2 < Const$, and $\widetilde{F}$ is the $p-1\times p-1$ sub-matrix of F.

$$\frac{\sqrt{1-\rho^2}}{T}\sum_t^{T-j}\widetilde{x}_t y_{t-1} = \sqrt{1-\rho^2}a(\rho,\alpha)\frac{1-\rho^2}{T}\sum_t^{T-j}y_{t-1}^2 + \widetilde{F}\frac{\sqrt{1-\rho^2}}{T}\sum_t^{T-j}Z_t y_{t-1}.$$

Statement (14), the fact that the family of distributions $\frac{1}{g(c)}\int_0^1 J_c^2(t)dt$ is uniformly bounded in probability and that $1-\rho^2 \to 0$ uniformly over $\mathcal{A}_T$ implies that

$$\sqrt{1-\rho^2}a(\rho,\alpha)\frac{1-\rho^2}{T}\sum_t^{T-j}y_{t-1}^2 = o_p(1)$$

uniformly over $\mathcal{A}_T$. The last observation together with (18) gives us that the following holds uniformly over $\mathcal{A}_T$: $\frac{\sqrt{1-\rho}}{T}\sum_{t=1}^T y_{t-1}\widetilde{x}_t \to^p 0$.

Similarly, (15), (17) and (20) imply asymptotic normality of the statistics $\frac{1}{\sqrt{T}}\sum_{t=1}^T \widetilde{x}_t e_t$ uniformly over $\mathcal{A}_T$. The observation that the limiting covariance matrix is I_{p-1} follows from the normalization. As a result, one has $\frac{1}{\sqrt{T}}\sum_{t=1}^T \widetilde{x}_t e_t \Rightarrow N(0, I_{p-1})$ uniformly over $\mathcal{A}_T$. Finally, (18), (19), (20) and the definition of normalization lead to $\frac{1}{T}\sum_{t=1}^T \widetilde{x}_t\widetilde{x}_t' \to^p I_{p-1}$, which holds uniformly over $\mathcal{A}_T$. This implies that (4) and (5) hold uniformly over $\mathcal{A}_T$.

Proof of Theorem 2. Let me denote $Q_e = \begin{pmatrix} \frac{1}{\sqrt{g(c)}}\int_0^1 J_c(s)dW(s) \\ \xi \end{pmatrix}$ as the right-hand side, and $Q_{eT} = \frac{1}{\sqrt{T}}\sum_{t=1}^T FX_t e_t$ the left of equation (4), while $Q_T = \frac{1}{T}\sum_{t=1}^T FX_tX_t'F'$ and $Q = \begin{pmatrix} \frac{1}{g(c)}\int_0^1 J_c^2(s)ds & \mathbf{0}' \\ \mathbf{0} & I_{p-1} \end{pmatrix}$, are the left and the right sides of equation (5) correspondingly. I can re-write equation (6) as follows: $t = \frac{(FA)'Q_T^{-1}Q_{eT}}{\hat{\sigma}\sqrt{(FA)'Q_T^{-1}FA}}$.

One needs to prove that one can substitute Q_{eT} and Q_T with their uniform limits Q_e and Q and the uniform approximation would hold. Namely, that $t \cong \frac{(FA)'Q^{-1}Q_e}{\hat{\sigma}\sqrt{(FA)'Q^{-1}FA}}$ uniformly over $\mathfrak{R}_\delta$ and $A \in \mathbb{R}^p$.

According to Theorem 1, $Q_T \cong Q$ and $Q_{eT} \cong Q_e$. In addition, given the uniform boundedness in probability of Q_e and Q and that Q is uniformly separated from zero (see Lemma 10 in Mikusheva (2007a)) one obtains that $Q_T^{-1} \cong Q^{-1}$ and $Q_T^{-1}Q_{eT} \cong Q^{-1}Q_e$.

One also has that $\frac{B'Q^{-1}B}{B'B}$ is bounded in probability and separated from zero uniformly over c and uniformly over all possible vectors $B \in \mathbb{R}^p$. Let me denote $FA = B$ and re-write $t = \frac{B'Q_T^{-1}Q_{eT}}{\widehat{\sigma}\sqrt{B'B}}\sqrt{\frac{B'B}{B'Q_T^{-1}B}}$. By a standard argument one can obtain that uniformly over $\mathfrak{R}_\delta$ and over $B \in \mathbb{R}^p$ it holds that $t \cong \frac{B'Q^{-1}Q_e}{\widehat{\sigma}\sqrt{B'B}}\sqrt{\frac{B'B}{B'Q^{-1}B}}$. It finishes the Proof of Theorem 2.

Proof of Lemma 1. I write down the first-order condition for the restricted maximization problem: $\begin{pmatrix} \sum_{t=1}^T X_tX_t' & A \\ A' & 0 \end{pmatrix}\begin{pmatrix} \widetilde{\rho}-\rho \\ \widetilde{\alpha}-\alpha \\ \mathcal{L} \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^T X_te_t \\ 0 \end{pmatrix}$, where $\mathcal{L}$ is a Lagrange multiplier. As a result,

$$\begin{pmatrix} \widetilde{\rho}-\rho \\ \widetilde{\alpha}-\alpha \end{pmatrix} = \left(\sum_{t=1}^T X_tX_t'\right)^{-1}\sum_{t=1}^T X_te_t - \frac{\left(\sum_{t=1}^T X_tX_t'\right)^{-1}AA'\left(\sum_{t=1}^T X_tX_t'\right)^{-1}\sum_{t=1}^T X_te_t}{A'\left(\sum_{t=1}^T X_tX_t'\right)^{-1}A},$$

and

$$F^{-1}\begin{pmatrix} \widetilde{\rho}-\rho \\ \widetilde{\alpha}-\alpha \end{pmatrix} = \left(\sum_{t=1}^T FX_tX_t'F'\right)^{-1}\sum_{t=1}^T FX_te_t -$$

$$-\frac{\left(\sum_{t=1}^T FX_tX_t'F'\right)^{-1}FAA'F'\left(\sum_{t=1}^T FX_tX_t'F'\right)^{-1}\sum_{t=1}^T FX_te_t}{A'F'\left(\sum_{t=1}^T FX_tX_t'F'\right)^{-1}FA}.$$

Now using the approximation obtained in Theorem 1 I have the following:

$$\sqrt{T}F^{-1}\begin{pmatrix} \widetilde{\rho}-\rho \\ \widetilde{\alpha}-\alpha \end{pmatrix} \cong \begin{pmatrix} \frac{\sqrt{g(c)}\int_0^1 J_c(s)dw(s)}{\int_0^1 J_c^2(s)ds} \\ \xi \end{pmatrix} - \frac{\frac{f_1\sqrt{g(c)}\int_0^1 J_c(s)dw(s)}{\int_0^1 J_c^2(s)ds} + f_2'\xi}{\frac{f_1^2 g(c)}{\int_0^1 J_c^2(s)ds} + f_2'f_2}\begin{pmatrix} \frac{f_1 g(c)}{\int_0^1 J_c^2(s)ds} \\ f_2 \end{pmatrix} =$$

$$= \begin{pmatrix} 0 \\ \xi \end{pmatrix} + u\begin{pmatrix} \frac{u\frac{1}{\sqrt{g(c)}}\int_0^1 J_c(s)dw(s)+\eta}{1+u^2\frac{1}{g(c)}\int_0^1 J_c^2(s)ds} \\ -\frac{\frac{1}{\sqrt{g(c)}}\int_0^1 J_c(s)dw(s)e_f+u\eta e_f}{1+u^2\frac{1}{g(c)}\int_0^1 J_c^2(s)ds} \end{pmatrix},$$

where $\xi \sim N(0, I_{p-1})$ and $\eta = \frac{f_2'\xi}{f_2'f_2} \sim N(0,1)$.

Proof of Theorem 3.

Let us consider a null hypothesis $H_0 : f(\rho,\alpha) = \gamma_0$ for some $f \in \mathcal{L}$. And let $(\widetilde{\rho},\widetilde{\alpha}) = (\widetilde{\rho}(\gamma_0),\widetilde{\alpha}(\gamma_0))$ be the restricted estimate. Let us consider the true value of (ρ,α) (for which the null is satisfied). According to the intermediate value theorem, there exists α^* lying between $\widetilde{\alpha}$ and α, and ρ^* lying between $\widetilde{\rho}$ and ρ, such that:

$$0 = f(\widetilde{\rho},\widetilde{\alpha}) - f(\rho,\alpha) = A_1^*(\widetilde{\rho}-\rho) + A_2^{*'}(\widetilde{\alpha}-\alpha);$$

here $A^* = (A_1^*, A_2^{*\prime})' = \frac{\partial f}{\partial(\rho,\alpha')}(\rho^*, \alpha^*)$. Let also $\widetilde{A} = \frac{\partial f}{\partial(\rho,\alpha')}(\widetilde{\rho}, \widetilde{\alpha})$, and $f(\hat{\rho}, \hat{\alpha}) - \gamma_0 = (\hat{\rho} - \tilde{\rho}, (\hat{\alpha} - \tilde{\alpha})')\hat{A}$. Let us define as before $Q_{eT} = \frac{1}{\sqrt{T}} \sum_{t=1}^T FX_t e_t$ and $Q_T = \frac{1}{T} \sum_{t=1}^T FX_t X_t' F'$, where F is the normalization described in Lemma 2. The proof is contained in the following three lemmas.

Lemma 3 *Uniformly over $\mathfrak{R}_\delta$ and over the set of functions $\mathcal{L}$ one has:*

$$LR \cong \frac{\left((FA^*)' Q_T^{-1} Q_{eT}\right)^2 (F\widetilde{A})' Q_T^{-1} F\widetilde{A}}{\left((FA^*)' Q_T^{-1} F\widetilde{A}\right)^2}; \tag{21}$$

$$LR^\pm \cong \sqrt{LR} \cdot sign\left(\frac{\left((F\hat{A})' Q_T^{-1} F\widetilde{A}\right)(FA^*)' Q_T^{-1} Q_{eT}}{(FA^*)' Q_T^{-1} F\widetilde{A}}\right). \tag{22}$$

Proof of Lemma 3 It is easy to observe that

$$LR = ((\widetilde{\rho} - \widehat{\rho}), (\widetilde{\alpha} - \widehat{\alpha})') J_T \begin{pmatrix} (\widetilde{\rho} - \widehat{\rho}) \\ (\widetilde{\alpha} - \widehat{\alpha}) \end{pmatrix},$$

where $J_T = \sum_{t=1}^T X_t X_t'$.

The restricted estimates $\widetilde{\rho}$ and $\widetilde{\alpha}$ satisfy the following system of equations (the restricted optimization first-order condition):

$$\begin{pmatrix} J_T & \widetilde{A} \\ A^* & 0 \end{pmatrix} \begin{pmatrix} (\widetilde{\rho} - \rho) \\ (\widetilde{\alpha} - \alpha) \\ \lambda \end{pmatrix} = \begin{pmatrix} J_{eT} \\ 0 \end{pmatrix}, \tag{23}$$

here λ is the Lagrange multiplier for the restricted optimization and $J_{eT} = \sum_{t=1}^T X_t e_t$. The last equation in the system is the statement that $f(\widetilde{\rho}, \widetilde{\alpha}) = f(\rho, \alpha)$ and uses the definition of A^* from above. System (23) implies that,

$$((\widetilde{\rho} - \rho), (\widetilde{\alpha} - \alpha)')' = J_T^{-1} J_{eT} - \frac{J_T^{-1} \widetilde{A} A^{*\prime} J_T^{-1} J_{eT}}{A^{*\prime} J_T^{-1} \widetilde{A}}$$

or $((\widetilde{\rho} - \widehat{\rho}), (\widetilde{\alpha} - \widehat{\alpha})')' = -\frac{J_T^{-1} \widetilde{A} A^{*\prime} J_T^{-1} J_{eT}}{A^{*\prime} J_T^{-1} \widetilde{A}}$. This implies that $LR = \frac{\left(A^{*\prime} J_T^{-1} J_{eT}\right)^2 \widetilde{A}' J_T^{-1} \widetilde{A}}{\left(A^{*\prime} J_T^{-1} \widetilde{A}\right)^2}$. Given that $Q_T = \frac{1}{T} F J_T F'$ and $Q_{eT} = \frac{1}{\sqrt{T}} F J_{eT}$ I arrive at (21). Since $LR^\pm = \sqrt{LR} \cdot sign\left(\hat{A}'(\hat{\rho} - \tilde{\rho}, (\hat{\alpha} - \tilde{\alpha})')'\right)$, Lemma 3 is proven.

Lemma 4 *If $\frac{FA^*}{\sqrt{(FA^*)' FA^*}} \to^p \frac{FA}{\sqrt{(FA)' FA}}$, $\frac{F\widetilde{A}}{\sqrt{(F\widetilde{A})' F\widetilde{A}}} \to^p \frac{FA}{\sqrt{(FA)' FA}}$ and $\frac{F\hat{A}}{\sqrt{(F\hat{A})' F\hat{A}}} \to^p \frac{FA}{\sqrt{(FA)' FA}}$ as $T \to \infty$ uniformly over $\mathfrak{R}_\delta$ and uniformly over $f \in \mathcal{L}$, then $LR \cong t(c,u)^2$ and $LR^\pm \cong t(c,u)$ uniformly over $\mathfrak{R}_\delta$ and uniformly over $f \in \mathcal{L}$.*

Proof of Lemma 4. Due to Lemma 3 formulas (21) and (22) hold. According to Theorem 1, $\frac{1}{T}FJ_TF' \cong Q$ and $\frac{1}{\sqrt{T}}FJ_{eT} \cong Q_e$, where Q and Q_e are right sides of equations (5) and (4) correspondingly. It is also known that Q_T is uniformly separated from zero, and both Q_T and Q_{eT} are bounded in probability uniformly over $\mathfrak{R}_\delta$. Given the conditions of the Lemma one can apply a direct generalization of Slutsky's Theorem and the Continuous mapping Theorem to uniform convergence. As a result, $LR \cong \frac{\left((FA)'Q^{-1}Q_e\right)^2}{(FA)'Q^{-1}FA}$, and the right hand side of the last expression equals to $t(u,c)^2$. The proof for $LR^{\pm}$ is similar.

Lemma 5 $\frac{FA^*}{\sqrt{(FA^*)'FA^*}} \to^p \frac{FA}{\sqrt{(FA)'FA}}$, $\frac{F\widetilde{A}}{\sqrt{(F\widetilde{A})'F\widetilde{A}}} \to^p \frac{FA}{\sqrt{(FA)'FA}}$ *and* $\frac{F\hat{A}}{\sqrt{(F\hat{A})'F\hat{A}}} \to^p \frac{FA}{\sqrt{(FA)'FA}}$ *as* $T \to \infty$ *uniformly over* $\mathfrak{R}_\delta$ *and uniformly over* $f \in \mathcal{L}$.

Proof of Lemma 5. I will prove it only for $\widetilde{A}$, the proof for A^* and $\hat{A}$ is analogous. Let $\|A\| = \sqrt{A'A}$ be the Euclidian norm. Let me consider separately a subset of $\mathfrak{R}_\delta$ when $|\lambda_p| < \delta + \varepsilon$ for some small positive $\varepsilon > 0$ and a subset for which $\lambda_p > \delta + \varepsilon$.

Let assume that $|\lambda_p| < \delta + \varepsilon$. Then F is positively definite, uniformly separated from zero and uniformly bounded (Lemma 2). This implies that it is enough to prove: $\frac{\widetilde{A}}{\|\widetilde{A}\|} \to^p \frac{A}{\|A\|}$. Indeed, if $\left\|\frac{\widetilde{A}}{\|\widetilde{A}\|} - \frac{A}{\|A\|}\right\| \to 0$ then under the condition that $\|F\|$ is bounded one has

$$\left|\frac{\|F\widetilde{A}\|}{\|\widetilde{A}\|} - \frac{\|FA\|}{\|A\|}\right| \leq \left\|\frac{F\widetilde{A}}{\|\widetilde{A}\|} - \frac{FA}{\|A\|}\right\| \leq \|F\| \left\|\frac{\widetilde{A}}{\|\widetilde{A}\|} - \frac{A}{\|A\|}\right\| \to 0.$$

If F separated from zero , the last statement also implies that $\frac{\|\widetilde{A}\|}{\|F\widetilde{A}\|} \to \frac{\|A\|}{\|FA\|}$. Since $\frac{FA}{\|FA\|} = \frac{FA}{\|A\|}\frac{\|A\|}{\|FA\|}$, one obtains the needed statement. In fact from the reasoning above, it is enough to prove $\frac{G\widetilde{A}}{\|G\widetilde{A}\|} \to^p \frac{GA}{\|GA\|}$ for some positive definite bounded matrix G.

I consider the derivative with respect to the AR coefficients rather than a derivative with respect to ADF coefficients (ρ, α). Let me denote the AR coefficients $\phi = (\phi_1, ..., \phi_p)$, that is, $y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + e_t$. There is a linear transformation between (ρ, α) and ϕ, this transformation is separated from zero and bounded. Let $\widetilde{\phi}$ is the AR coefficients for $(\widetilde{\rho}, \widetilde{\alpha})$. Let $\theta_k = f_k(\phi_1, ..., \phi_p)$ is k-th impulse response. Assume that $\nabla_k = \frac{\partial \theta_k}{\partial \phi}(\phi)$ is the vector of derivatives.So, it is enough to show that for two set of values ϕ and $\widetilde{\phi}$ I have $\frac{\nabla_k}{\|\nabla_k\|} - \frac{\widetilde{\nabla}_k}{\|\widetilde{\nabla}_k\|} \to 0$ uniformly over k if $\phi - \widetilde{\phi} \to 0$. In fact, it is enough to check that $\frac{\|\nabla_k - \widetilde{\nabla}_k\|}{\|\nabla_k\|} \to 0$.

First, I write down formula for ∇_k. Here I use formula from Lütkepohl (1990) (Proposition 1(i)): $\nabla_{k,j} = \frac{\partial \theta_k}{\partial \phi_j}(\phi) = \sum_{s=1}^{k-j} \theta_s \theta_{k-j-s}$. Let introduce a $k \times k$ matrix L that consists of zeros, except the diagonal just below the main diagonal that has all ones. L is the matrix for the lag operator. Then $\nabla_k = e_k'(\sum_{j=0}^k \phi_j L^j)(\sum_{j=0}^k \phi_j L^j) = e_k'\Phi_k\Phi_k$, where $L^0 = I_k$ and $\Phi_k = \sum_{j=0}^k \phi_j L^j$. Now

$$\frac{\|\nabla_k - \widetilde{\nabla}_k\|}{\|\nabla_k\|} = \frac{\|e_k'(\Phi_k^2 - \widetilde{\Phi}_k^2)\|}{\|e_k'\Phi_k^2\|} \leq \|I_k - \Phi_k^{-2}\widetilde{\Phi}_k^2\| \leq \|\widetilde{\Phi}_k^2\|\|\Phi_k^{-2} - \widetilde{\Phi}_k^{-2}\|$$

here $\|\|$ is the operator norm if applied to a matrix. It follows from Lemma A.5.(i) in Saikkonen and Lütkepohl (2000) that $\|\Phi_k\| \leq \sum_{j=0}^{\infty} \theta_j \leq const(\delta + \varepsilon)$ is bounded uniformly over k. Formula (A.21) in Saikkonen and Lütkepohl (2000) implies that $\|\Phi_k^{-1} - (\widetilde{\Phi}_k)^{-1}\| \leq \|\phi - \widetilde{\phi}\|$. Given that $\widetilde{\phi}$ uniformly converges to ϕ as the sample size increases, I proved Lemma 5 in this case.

Now turn to the case when $\lambda_p > \delta + \varepsilon$. According to formula (2.4.16) in Hamilton(1994) the impulse response function has the form of $\theta_s = \sum_{j=1}^p c_j \lambda_j^s$, where λ_j are roots of the autoregressive polynomial. Then

$$\frac{\partial\theta_k}{\partial\phi_j}(\phi) = \sum_{s=1}^{k-j} \theta_s\theta_{k-j-s} = \sum_{i=1}^p \sum_{l=1}^p c_i c_l \sum_{s=1}^{k-j} \lambda_i^s \lambda_l^{k-j-s} =$$
$$= \sum_{i=1}^p c_i^2 \lambda_i^{k-j}(k-j) + \sum_{i=1}^p \sum_{l=1, l\neq i}^p c_i c_l \frac{\lambda_i^{k-j} - \lambda_l^{k-j}}{\lambda_i - \lambda_l}.$$

I divide the above expression by $k\lambda_p^k$ for a fixed $1 \leq j \leq p$. Then

$$\frac{1}{k\lambda_p^k}\frac{\partial\theta_k}{\partial\phi_j}(\phi) = c_p\frac{k-j}{k} + \frac{1}{\lambda_p^j}\sum_{i=1}^{p-1} c_i^2 \left(\frac{\lambda_i}{\lambda_p}\right)^{k-j}\frac{(k-j)}{k} + \frac{1}{k\lambda_p^j}\sum_{i=1}^p \sum_{l=1, l\neq i}^p c_i c_l \frac{\left(\frac{\lambda_i}{\lambda_p}\right)^{k-j} - \left(\frac{\lambda_l}{\lambda_p}\right)^{k-j}}{\lambda_i - \lambda_l}.$$

Since $(\rho, \alpha) \in \mathfrak{R}_\delta$ and $\lambda_p > \delta + \varepsilon$ I have that $\left|\frac{\lambda_i}{\lambda_p}\right| < 1 - \varepsilon_1$ for some $\varepsilon_1 > 0$ and all $i < p$. Now I notice that function $l(x) = x^k$ is continuous in x uniformly over $|x| < 1 - \varepsilon_1$ and uniformly over all positive k. Since autoregressive roots are continuous functions of the autoregressive coefficients, and c_i are continuous functions of roots, and given that $(\widetilde{\rho}, \widetilde{\alpha}) \to^p (\rho, \alpha)$ uniformly over $\mathfrak{R}_\delta$, I obtain that $\frac{1}{k\widetilde{\lambda}_p^k}\frac{\partial\theta_k}{\partial\phi_j}(\widetilde{\phi}) \to^p \frac{1}{k\lambda_p^k}\frac{\partial\theta_k}{\partial\phi_j}(\phi)$ uniformly over $\mathfrak{R}_\delta$ and *uniformly* over k. I also note that $\frac{1}{k\lambda_p^k}\frac{\partial\theta_k}{\partial\phi_j}(\phi)$ is separated from zero uniformly over $\mathfrak{R}_\delta$ and uniformly over k. This implies that $\frac{\widetilde{\nabla}_k}{\|\widetilde{\nabla}_k\|} = \frac{\widetilde{\nabla}_k}{k\widetilde{\lambda}_p^k}\frac{k\widetilde{\lambda}_p^k}{\|\widetilde{\nabla}_k\|} \to \frac{\nabla_k}{\|\nabla_k\|}$

8 References

Andrews D.W.K. and P. Guggenberger (2009): "Hybrid and Size-corrected Subsample Methods," *Econometrica*, 77(3), 721-762.

Bobkoski M.J. (1983): "Hypothesis Testing in Nonstationary Time Series," Unpublished PhD thesis (Dept. of Statistics, University of Wisconsin, Madison).

Cavanagh, C. (1985): "Roots Local to Unity", Manuscript (Dept. of Economics, Harvard University, Cambridge, MA).

Chan N.H. and C.Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics*, 15(3), 1050-1063.

Christiano L., M. Eichenbaum and C.L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What End?" In *Handbook of Macroeconomics*, vol. 1A, J.B. Taylor and M. Woodford, eds. Amsterdam: Elsevier.

Christiano L., M. Eichenbaum and R. Vigfusson (2004): "What Happens after a Technology Shock?" NBER Working Paper 9819.

Gali, J. (1999): "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 89(1), 249-271.

Giraitis L., and P.C.B. Phillips (2009):"Mean and Autocovariance Function Estimation Near the Boundary of Stationarity," Cowles Foundation working paper 1690.

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

Gregory, A.W. and M.R. Veall (1985): "On Formulating Wald Tests for Non-linear Restrictions," *Econometrica*, 53, 1465-68.

Hamilton, J.D. (1994): "Time Series Analysis", *Princeton University Press*

Hansen, B.E. (1999): "The Grid Bootstrap and the Autoregressive Model," *The Review of Economics and Statistics*, 81(4), 594-607.

Hansen, B.E. (2006): "Edgeworth Expansions for the Wald and GMM Statistics for Nonlinear Restrictions," *Econometric Theory and Practice*, D. Corbae, S. N. Durlauf, and B. E. Hansen (eds.), New York: Cambridge University Press.

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

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

Kilian, L. (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, L. and P.L. Chang (2000):"How Accurate are Confidence Intervals for Impulse

Responses in Large VAR Models?", *Economic Letters*, 69, 299-307.

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

Mikusheva, A. (2007a): "Uniform Inference in Autoregressive Models," *Econometrica*, 75(5), 1411-1452.

Mikusheva, A. (2007b): "Supplement to "Uniform Inferences in Autoregressive Models": Supplementary Appendix,"
http://www.econometricsociety.org/ecta/supmat/6254_proofs.pdf

Pesavento, E. and B. Rossi (2006): "Small Sample Confidence Intervals for Multivariate Impulse Response Functions at Long Horizons," *Journal of Applied Econometrics* 21(8), December 2006, 1135-1155.

Pesavento, E. and B. Rossi (2007): "Impulse Response Confidence Intervals for Persistent Data: What Have We Learned?" *Journal of Economic Dynamics and Control* 31, 2398-2412.

Phillips P.C.B. (1987): "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74(3), 535-547.

Phillips P.C.B. (1998): "Impulse Response and Forecast Error Variance Asymptotics in Non-stationary VARs," *Journal of Econometrics*, 83, 21-56.

Phillips P.C.B. (1988): "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56 (5), 1021-1043.

Phillips P.C.B. (2007)"Unit Root Log Periodogram Regression," *Journal of Econometrics*, 138, 104-124.

Rotemberg, J.J and M. Woodford (1997): "An Optimization-based Econometric Framework for the Evaluation of Monetary Policy," *NBER Macroeconomics Annual 1997.*

Runkle, D.E. (1987): "Vector Autoregressions and Reality," *Federal Reserve Bank of Minneapolis*, Staff Report 107.

Saikkonen and Lütkepohl (2000): "Asymptotic Inference on Nonlinear Functions of the Coefficients of Infinite Order Cointegrated VAR Processes," *Nonlinear econometric modeling in time series*, eds. Barnett et al., *Cambridge Press*

Stock J. (1991): "Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435-459.

Wright, J.H. (2000): "Confidence Intervals for Univariate Impulse Responses with a

Near Unit Root," *Journal of Business and Economic Statistics*, 18, 368-73.