Econometrics Journal (2004), volume **7**, pp. 505–527.

Asymptotic confidence intervals for impulse responses of near-integrated processes

NIKOLAY GOSPODINOV

Department of Economics, Concordia University and CIREQ, Quebec, Canada
E-mail: gospodin@alcor.concordia.ca

Received: March 2004

Summary Many economic time series are characterized by high persistence which typically requires nonstandard limit theory for inference. This paper proposes a new method for constructing confidence intervals for impulse response functions and half-lives of nearly non-stationary processes. It is based on inverting the acceptance region of the likelihood ratio statistic under a sequence of null hypotheses of possible values for the impulse response or the half-life. This paper shows the consistency of the restricted estimator of the localizing constant which ensures the validity of the asymptotic inference. The proposed method is used to study the persistence of shocks to real exchange rates.

Keywords: *Near-integrated AR process*, *Impulse responses*, *Local-to-unity asymptotics*, *Likelihood ratio test*, *Non-linear restrictions*.

1. INTRODUCTION

Many economic time series are characterized by high persistence which typically requires nonstandard limit theory for inference. For example, real exchange rates, interest rates, unemployment rate, realized and implied volatility are strongly dependent series but imposing an exact unit root may seem too restrictive and in many cases is theoretically unjustified. Some difficulties arise when the interest lies in constructing confidence bands for the impulse response functions and some measures of persistence such as the half-life in unrestricted (near-) non-stationary dynamic models (see Phillips 1998). Recent work on conducting valid inference for impulse responses of highly persistent series includes Hansen (1999), Kilian (1998), Kilian and Zha (2002), Phillips (1998), Rossi (2004), Sims and Zha (1999) and Wright (1997, 2002), among others.

In this paper, we propose a new method for asymptotic inference on impulse response functions and impulse response-based measures of persistence. The approach we adopt takes explicitly into account the presence of large roots in the autoregressive dynamics of the process. The method computes the likelihood ratio (LR) statistic for a sequence of null hypotheses that restrict the values of the impulse responses. Since the impulse responses are non-linear functions of the model parameters, the likelihood function is maximized subject to non-linear constraints. The limiting behaviour of the least squares estimator in linear models subject to non-linear restrictions has been studied by Nagaraj and Fuller (1991). Moon and Schorfheide (2002) established the consistency and derived the limiting distribution of the minimum distance estimator in linear non-stationary time series models subject to non-linear restrictions. This paper builds on their results to derive the limiting distribution of the LR statistic in nearly non-stationary AR models and invert its acceptance region to obtain interval estimators of the statistics of interest.

Our results, however, differ from the previous work in one important dimension. Since our primary interest lies in impulse response analysis, we parameterize the leading time of the impulse response function as a function of the sample size. Hence, the order of the polynomial constraint imposed by the null hypothesis is not a fixed constant but increases linearly with the sample size. We argue that this parameterization is a by-product of our local-to-unity framework and delivers some interesting new results. We show that the restricted estimator of the non-stationary component converges faster than the unrestricted estimator which helps us identify and consistently estimate the localizing constant. The restricted estimate of the localizing constant is then used to evaluate the limiting distribution of the LR statistic.

Our approach to estimation and inference is similar to those proposed by Miller and Newbold (1995), Wright (1997), Bekaert and Hodrick (2001) and Hansen (2000) but is developed to cover the near-non-stationary region of the parameter space which is typically associated with the presence of nuisance parameters (in particular, the local-to-unity parameter) that are not consistently estimable. Moon and Phillips (2000) use panel data to show that $\sqrt{N}$-consistent estimation (where N is the size of the cross section) is possible. Phillips *et al.* (2000) develop a new model of near integration that allows the identification of the localizing constant from blocks of time series data. Valkanov (1998) imposes multivariate restrictions implied by the theory of term structure of interest rates to obtain a consistent estimate of the localizing constant. We establish the consistency of the localizing constant in a different context using univariate time series framework.

Recently, Wright (2000) proposed an algorithm for constructing conservative confidence intervals for univariate impulse responses that controls the asymptotic coverage uniformly in the parameter space. He also showed that this method dominates significantly most of the existing methods in the literature. Unlike Wright's method, the approach proposed in this paper is not conservative and provides asymptotically exact inference. Whereas some other procedures (e.g. Hansen 1999; Rossi 2004) also possess good coverage properties at long horizons, the inversion of the LR statistic has some appealing features. For instance, the LR statistic is criterion function-based and does not require variance estimation for studentization. Also, the inversion of the LR statistic appears to be shifting the confidence intervals away from the non-stationary boundary much more often compared to methods based on inverting the OLS estimator of the largest root. It is also interesting to note that the method controls the coverage at all forecast horizons, not only at long horizons.

The rest of the paper is organized as follows. Section 2 defines the problem of interest and derives the main theoretical results. It also discusses the construction of confidence intervals by test inversion. In Section 3, the coverage properties of the proposed method are assessed by simulation and compared to some existing methods in the literature. Section 4 quantifies the persistence of four major real exchange rates by constructing interval estimators for impulse responses and half-lives. Section 5 summarizes the results.

2. MODEL AND MAIN RESULTS

Suppose that the univariate process of interest is generated by

$$y_t^d = \gamma_1 y_{t-1}^d + \gamma_2 y_{t-2}^d + \cdots + \gamma_p y_{t-p}^d + e_t, \tag{1}$$

where $y_t^d = y_t - \mu$ ($y_t^d = y_t - \mu - \beta t$) denotes the demeaned (detrended) process.[1] By repeated substitution, y_t^d can be expressed (up to initial conditions) as a function of past innovations

$$y_t^d = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots \tag{2}$$

in which the coefficients θ_l are non-linear functions of the AR parameters γ and can be interpreted as impulse responses to a one-standard-deviation shock in e at time $t - l$.

The AR(p) model in (1) can also be rewritten in an augmented Dickey-Fuller (ADF) form as

$$y_t^d = \alpha y_{t-1}^d + \Psi(L) \triangle y_t^d + e_t, \tag{3}$$

where $\alpha = \gamma_1 + \gamma_2 + \cdots + \gamma_p$, $\Psi(L) = \sum_{j=1}^{p-1} \psi_j L^j$ with $\psi_j = -\sum_{i=j+1}^{p} \gamma_i$ for $j = 1, \ldots, p-1$, is a lag polynomial describing the short-run dynamics of the process with no roots near or on the unit circle. The parameter α is the sum of the AR coefficients and can be interpreted as a measure of persistence of the series (Andrews and Chen 1994).

In this paper, we consider highly persistent AR processes with roots near or exactly on the unit circle. If the parameter α is close to the non-stationary boundary, it proves useful to reparameterize it as local-to-unity $\alpha_T = \exp(c/T) \approx 1 + c/T$ for a fixed constant $c < 0$ (near-integrated process), $c = 0$ (unit root process) or $c > 0$ (explosive process) (see Chan and Wei 1987; Phillips 1987; among others). This parameterization removes the discontinuity of the distribution theory in the neighbourhood of one and provides an excellent approximation to the finite-sample distribution of α in this region of the parameter space.

An alternative representation is in terms of the largest autoregressive root and has the form

$$\begin{aligned} y_t &= \mu^* + \beta^* t + y_t^d \\ b(L)(1 - \phi L) y_t^d &= \xi_t, \end{aligned} \tag{4}$$

where ϕ denotes the largest root of the AR polynomial and $b(L) = \sum_{j=0}^{p-1} b_j L^j$ with $b_0 = 1$ is a lag polynomial whose roots are assumed fixed and strictly less than one in absolute value. If the largest AR root is cast in the local-to-unity framework $\phi_T = 1 + c^*/T$, the coefficient on y_{t-1}^d in the ADF representation of the process is $\alpha^* = 1 + c^* b(1)/T$ (Stock 1991). This gives the relationship between the local-to-unity parameters in (3) and (4). In deriving the results in the paper, we focus on model (3) but similar arguments can be used to obtain the same results with model (4).

Often the interest of the econometrician lies in testing hypotheses about the impulse response at lead time l, $\frac{\partial y_{t+l}}{\partial e_t} = \theta_l$. Alternatively, one might want to measure how long it takes for a non-explosive series to complete $100\omega\%$ of the adjustment to the initial shock, $\Lambda_\omega = \sup_{l \in L} |\frac{\partial y_{t+l}}{\partial e_t}| \geq 1 - \omega$ for some fixed $\omega \in (0, 1]$ (Kilian and Zha 2002; Ng and Perron 2002). In both of these cases, the restrictions imposed by the null hypothesis are polynomials of order l in the parameters of the model. In the AR(1) model, the structure of the adjustment is monotonic and $l = \ln(1 - \omega)/\ln(\alpha)$ or $l = [\delta T]$, where $\delta = \ln(1 - \omega)/c$ is a fixed positive constant if $c < 0$ and [.] denotes the greatest lesser integer function. For higher-order AR models with a root near unity, Rossi (2004) showed that $l = [\delta T] + o(1)$. Therefore, the order of the polynomial constraint increases linearly with the sample size as the process approaches the unit root boundary.

[1] In this paper, the process is demeaned (detrended) by OLS. The trend component does not directly affect the shape of the impulse response function but more efficient detrending methods may increase the power of the tests (Elliott *et al.* 1996) and reduce the width of the confidence intervals for the parameter of interest. We have not fully explored alternative detrending methods in this version of the paper.

2.1. Estimation and properties of the constrained estimator

Let $\rho = (\alpha, \psi_1, \ldots, \psi_{p-1})' = (\alpha, \psi')' \in \Xi \subset \mathbb{R}^p$. The quasi log likelihood of (3), conditional on the initial conditions, is given by

$$l_T(\rho) = -\frac{T}{2}\ln(\sigma^2) - \frac{1}{2}\sum_{t=1}^{T}\frac{e_t^2}{\sigma^2} \tag{5}$$

and the maximum likelihood estimator of ρ is $\widehat{\rho} = \arg\max_{\rho\in\Xi} l_T(\rho)$.

Suppose that we are interested in testing the hypothesis that the impulse response at horizon l, θ_l, is equal to a particular value $\theta_{0,l}$, $H_0 : \theta_l = \theta_{0,l}$ versus $H_1 : \theta_l \neq \theta_{0,l}$. Without loss of generality, we can rewrite the restriction under the null as $h(\rho) = 0$, where $h \equiv \theta_l - \theta_{0,l} : \mathbb{R}^p \to \mathbb{R}$ is a polynomial of degree l. Let $\widetilde{\rho} = \arg\max_{h(\rho)=0} l_T(\rho)$ denote the restricted maximum likelihood estimator[2] and $LR_T = T\ \ln(SSR_0/SSR)$ be the likelihood ratio statistic[3] of the hypothesis $H_0 : h(\rho) = 0$, where SSR_0 and SSR are the sum of squares of the restricted and the estimated residuals, respectively. The main assumptions are stated below.

Assumption 1. *Assume that $\alpha_T = 1 + c/T$ for some fixed c, $y^d_{-p+1}, \ldots, y^d_0$ are fixed and e_t is a martingale difference sequence with $Ee_t^2 = \sigma^2$ and $\sup_t Ee_t^4 < \infty$.*

Assumption 2. *The parameter space Ξ is compact and there is a unique ρ_0 in the interior of Ξ that maximizes $l(\rho)$ subject to $h(\rho_0) = 0$.*

Assumption 3. *The function $h(\rho)$ is a polynomial constraint that is continuous and differentiable in a region about ρ_0 and the order of the polynomial is assumed to be $l = [\delta T]$ for some fixed $\delta > 0$.*

Assumption 1 parameterizes the sum of all AR parameters as local-to-unity and specifies the error term as a homoskedastic martingale difference sequence. Assumptions 2 and 3 provide the identification condition and restrict the class of non-linear constraints on the parameters. Assumption 3 parameterizes the lead time of the impulse response as a function of the sample size. The parameterization $l = [\delta T]$ for some fixed $\delta > 0$ has been used previously in the literature by Stock (1996), Phillips (1998), Rossi (2004) and Gospodinov (2002), among others, for impulse response analysis and long-horizon forecasting of nearly non-stationary processes. This parameterization allows the asymptotic approximation to preserve the parameter estimation uncertainty present in the finite sample distribution. By contrast, in the fixed forecast horizon case,

[2]The constrained optimization problem is performed using the CO library in Gauss. Although the degree of the constraint (the forecasting horizon of the impulse response) can be large, the numerical implementation using the existing optimization algorithms is not difficult. Since the likelihood function is quadratic, the computation of the analytical gradient and Hessian is straightforward. We can also provide analytical derivatives for the impulse responses as in Lütkepohl (1990). This reduces the computation of the constrained estimator to a standard optimization problem.

[3]An asymptotically equivalent alternative to the LR test is the GMM-based distance metric (DM) statistic $DM_T = Q_T(\widetilde{\rho}) - Q_T(\widehat{\rho})$, where $Q_T(\rho) = e(\rho)'X\Omega_T^{-1}X'e$, X is the regressor matrix in (3), Ω_T is a consistent estimate of $TE(x_t x_t' e_t^2)$ and $\widetilde{\rho} = \arg\min_{h(\rho)=0} Q_T(\rho)$. This test, by construction, is robust to conditional heteroskedasticity although the interpretation of the impulse responses in the case of heteroskedasticity is not obvious.

the parameter estimation uncertainty vanishes asymptotically as T goes to ∞. Parameterizing the degree of the polynomial constraint as $O(T)$ may seem as an unusual assumption but it arises naturally in inference for impulse responses and half-lives in local-to-unity framework.

The following theorem establishes the rate of convergence of the restricted estimators of the stationary and non-stationary components.

Theorem 1. *If Assumptions* 1–3 *are satisfied, then as* $T \to \infty$,

(i) $\Pi_{2,T}(\widetilde{\rho} - \rho_0) = \left(O_p(1), \ldots, O_p(1)\right)'$ *and* $\sqrt{T}(\widetilde{c} - c) = O_p(1)$, *where* $\Pi_{2,T} = diag(T^{3/2}, \sqrt{T}, \ldots, \sqrt{T})$
(ii) $T(\widehat{\alpha} - \widetilde{\alpha}) = O_p(1)$ *and* $\Pi_{3,T}(\widehat{\psi} - \widetilde{\psi}) = o_p(1)$, *where* $\Pi_{3,T} = diag(\sqrt{T}, \ldots, \sqrt{T})$
(iii) $\sqrt{T}\,(h(\widetilde{\rho}) - h(\rho_0)) = O_p(1)$.

Proof. See Appendix.

Theorem 1 shows that the restricted estimate of α is converging at a faster rate ($T^{3/2}$-consistency) than the (T-consistent) unrestricted estimator. As a result, the localizing constant c can be consistently estimated under the imposed restrictions. Also, the restricted estimation provides a consistent estimate of the impulse response function. It is interesting to note that the impulse response function is written as a function of a non-stationary component and stationary components. These components have different rates of convergence and since the faster converging non-stationary component determines the shape of the impulse response as $T \to \infty$, the stationary components $\sqrt{T}(\widehat{\psi} - \widetilde{\psi}) = o_p(1)$ are not affected asymptotically by the imposition of the constraint.

The consistency of the restricted estimator of the localizing constant deserves some further remarks.[4] The result appears to be driven by the parameterization of the lead time of the impulse response as a function of the sample size. It is worth pointing out that this parameterization is not arbitrary and is a natural by-product of the local-to-unity specification for the parameter α. This parameterization forces the parameter on the (near-) non-stationary component to satisfy a highly non-linear polynomial constraint whose degree increases with the sample size which in turn accelerates the rate of convergence of the estimator. It is precisely the parameterization $l = [\delta T]$ in Assumption 3 that makes our results differ from those obtained in Nagaraj and Fuller (1991). If the degree of the polynomial constraint is fixed, the rates of convergence of the restricted and the unrestricted estimators are the same as shown in Nagaraj and Fuller (1991).

To clarify the notation and illustrate the main result in Theorem 1, consider the AR(2) model

$$y_t^d = \alpha y_{t-1}^d + \psi \bigtriangleup y_{t-1}^d + e_t. \tag{6}$$

We can rewrite this model as

$$\begin{pmatrix} y_t^d \\ \bigtriangleup y_t^d \end{pmatrix} = \begin{pmatrix} \alpha & \psi \\ \alpha - 1 & \psi \end{pmatrix} \begin{pmatrix} y_{t-1}^d \\ \bigtriangleup y_{t-1}^d \end{pmatrix} + \begin{pmatrix} e_t \\ e_t \end{pmatrix}$$

[4] Since the local-to-unity parameterization is an artificial statistical device, attaching any structural interpretation to this result will be inappropriate. Nevertheless, this result helps us identify the localizing constant under the null that can be later used to evaluate the limiting distribution of the LR statistic.

or

$$Y_t = AY_{t-1} + E_t.$$

Then, $Y_{t+l} = A^{l+1}Y_{t-1} + A^l E_t + A^{l-1}E_{t+1} + \cdots + E_{t+l}$ and the impulse response at lead time l is $\theta_l \equiv \frac{\partial y^d_{t+l}}{\partial e_t} = (1,0)'A^l(1,1)$. Furthermore, the impulse response can be rewritten in polynomial form as $\theta_l = \sum_{i=0}^{l/2}(-1)^i \pi_i(l)\psi^i(\alpha+\psi)^{l-2i}$, for even $l \geq 2$, and $\theta_l = \sum_{i=0}^{(l-1)/2}(-1)^i \pi_i(l)\psi^i(\alpha+\psi)^{l-2i}$, for odd $l \geq 3$, where $\pi_0(0) = 0$, $\pi_0(l) = 1$ and $\pi_i(l) = \pi_i(l-1) + \pi_{i-1}(l-2)$. For example, $\theta_1 = \alpha+\psi$, $\theta_2 = (\alpha+\psi)^2 - \psi$, $\theta_3 = (\alpha+\psi)^3 - 2\psi(\alpha+\psi)$, $\theta_4 = (\alpha+\psi)^4 - 3\psi(\alpha+\psi)^2 + \psi^2$, $\theta_5 = (\alpha+\psi)^5 - 4\psi(\alpha+\psi)^3 + 3\psi^2(\alpha+\psi)$ etc. At point $\theta_{0,l}$ in the parameter space, we can obtain the restricted ML estimator of ρ from the constrained optimization problem $\max l_T(\rho)$ subject to $\theta_l - \theta_{0,l} = 0$. For the half-life, the constraint is $\theta_l - 0.5 = 0$.

Now let $\alpha = 1 + c/T$, where the dependence of α on the sample size T is omitted for notational simplicity. Since $\alpha - 1 = c/T = o_p(1)$,

$$\begin{aligned}
(1\ 0)\begin{pmatrix} \alpha & \psi \\ o_p(1) & \psi \end{pmatrix}^l \begin{pmatrix} 1 \\ 1 \end{pmatrix} &= (1\ 0)\begin{pmatrix} \alpha^l + o_p(1) & \alpha^{l-1}\psi + \alpha^{l-2}\psi^2 + \cdots + \psi^l + o_p(1) \\ o_p(1) & \psi^l + o_p(1) \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
&= \alpha^l + \alpha^{l-1}\psi + \alpha^{l-2}\psi^2 + \cdots + \psi^l + o_p(1) \\
&= \alpha^l\left(1 + \frac{\psi}{\alpha} + \cdots + \left(\frac{\psi}{\alpha}\right)^l\right) + o_p(1).
\end{aligned}$$

Further, if $l = [\delta T]$ and as $T \to \infty$, $\alpha^l = \exp(c\delta)$, $1 + \frac{\psi}{\alpha} + \cdots + (\frac{\psi}{\alpha})^l = \frac{\alpha}{\alpha-\psi} = \frac{1}{1-\psi} + o_p(1)$ and $\theta_l \to \frac{\exp(c\delta)}{1-\psi}$ (Rossi 2004).

We want to test $H_0 : h(\rho) \equiv \theta_l - \theta_{0,l} = 0$, where $\theta_{0,l}$ is a fixed constant and $\rho = (\alpha\ ,\ \psi)$ or equivalently $\rho = (c, \psi)$. Then, $\widetilde{c} = \frac{\ln(\theta_{0,l}[1-\widetilde{\psi}])}{\delta} = g(\widetilde{\psi})$, i.e. the restricted estimate $\widetilde{c}$ is a non-linear function of $\widetilde{\psi}$. Taking a first-order Taylor series expansion of $\widetilde{c}$ about c, we have

$$\sqrt{T}(\widetilde{c} - c) = \frac{\partial g(\psi)}{\partial \psi}\sqrt{T}(\widetilde{\psi} - \psi) + o_p(1),$$

i.e. the estimator of $\widetilde{c}$ has to share the rate of convergence of the estimator $\widetilde{\psi}$.

2.2. Limiting Distribution of the LR Test and Confidence Interval Construction

In this section, we derive the limiting distribution of the LR statistic and discuss the construction of confidence intervals for impulse responses and half-lives.

Theorem 2. *Consider model* (1)–(3) *and let* $\rho = \alpha$ *for* $p = 1$ *and* $\rho = (\alpha\ ,\ \psi_1, \ldots, \psi_{p-1})'$ *for* $p > 1$. *Under the null* $H_0 : h(\rho) = 0$ *and* Assumptions 1–3,

$$LR_T \Rightarrow \frac{\left[\int_0^1 J_c^\tau(s)\,dW(s)\right]^2}{\int_0^1 J_c^\tau(s)^2\,ds},$$

where $J_c^\tau(r) = J_c(r)$ for $y_t^d = y_t$, $J_c^\tau(r) = J_c(r) - \int_0^1 J_c(s)ds$ for $y_t^d = y_t - \mu$, $J_c^\tau(r) = J_c(r) - \int_0^1(4-6s)J_c(s)\,ds - r\int_0^1(12s-6)J_c(s)\,ds$ for $y_t^d = y_t - \mu - \beta t$, $J_c(r) = \int_0^r \exp[(r -$

$s)c]\, dW(s)$ is a homogeneous Ornstein-Uhlenbeck process generated by the stochastic differential equation $dJ_c(r) = cJ_c(r) + dW(r)$ with $J_c(0) = 0$, $W(r)$ is the standard Brownian motion defined on [0,1] *and $\Rightarrow$ denotes weak convergence.*

Proof. See Appendix.

Theorem 2 shows that the limiting theory for the LR test is dominated by the near non-stationary component and is not affected by the presence of stationary components. The shape of the asymptotic distribution is determined by the shape of the limiting representation of the estimator of α.[5]

The following remarks may prove useful for understanding how the procedure works. When $p = 1$, the restriction $h(\rho) = 0$ has the form $(1 + c/T)^l - \theta_{0,l} = 0$, where $\theta_{0,l}$ and T are given, and uniquely determines the value of c (the $O_p(1)$ term for c in part (i) of Theorem 1 is degenerate) that can be used to invert the acceptance region of the LR statistic. In this case, there is a one-to-one mapping between c and θ_l and the limits of the confidence interval for the impulse response are monotonic transformations of the limits of the confidence interval for c obtained by inverting the LR test of $H_0 : c = c_0$ versus $H_1 : c \neq c_0$ using the asymptotic representation in Theorem 2. One could also follow Elliott and Stock (2001) and construct more accurate confidence intervals for c by inverting a sequence of most powerful tests of $H_0 : c = c_0$ and use them to obtain confidence intervals for the impulse response.

When $p > 1$, there are extra parameters that describe the short-run dynamics of the process and the localizing constant cannot be exactly pinned down by the restriction imposed by the null and needs to be estimated. Since in this case the relationship between c and the parameter of interest (impulse response) is not one-to-one and not necessarily monotonic, the methods of Stock (1991) and Elliott and Stock (2001), among others, can no longer be applied directly and we need to resort to Bonferroni inference (Wright 2000). In this paper we adopt an alternative approach and estimate simultaneously the non-stationary and stationary components of the process under the null that the impulse response takes on a particular value. Even though the transitory component of the process affects the shape of the impulse response at short horizons, its effect that causes non-monotonicity in the IRF vanishes at long horizons which helps us identify the value of c under the null hypothesis (more precisely, Theorem 1 shows that, under the specified assumptions, the restricted estimator of c is consistent). Then, the restricted estimate of c can be used for inverting the distribution of the LR test by reading off the corresponding critical value for the LR test from look-up tables constructed from Theorem 2.

Sims *et al.* (1990) and Inoue and Kilian (2002) showed that the standard asymptotics and standard bootstrap, respectively, are valid in large samples for (linear combinations of) the slope parameters γ_i in the near-integrated AR(p) model (1) with $p > 1$. Although this certainly is a very useful result, there are two drawbacks. First, it introduces a discontinuity in the large sample inference for AR(1) and AR(p) models for $p > 1$. Second, it does not guarantee uniformly good coverage rates of the confidence intervals for different combinations of parameter values for any given finite sample. By contrast, our approach explicitly parameterizes the persistence parameter as a function of the sample size and is expected to deliver better small-sample performance than the standard asymptotics (or bootstrap) in the presence of large AR roots.

[5] As pointed out by Elliott and Stock (2001), when the largest AR root is near one, a uniformly most powerful of the null $h(\rho) = 0$ against the alternative $h(\rho) \neq 0$ does not exist.

We should also note that even though the limiting result in Theorem 2 appears standard (for $c = 0$, it is the square of the asymptotic distribution of the ADF unit root test), it is not a trivial extension of the results in Phillips and Durlauf (1986) and Sims *et al.* (1990) to non-linear constraints. For $p > 1$ and $l = [\delta T]$, the properties of the restricted estimator of the non-stationary component were unknown. The result in Theorem 2 is driven by the $T^{3/2}$-consistency of the constrained estimator of α, $T(\widehat{\alpha} - \widetilde{\alpha}) = T(\widehat{\alpha} - \alpha_0) + o_p(1)$ and $\sqrt{T}(\widehat{\psi} - \widetilde{\psi}) = o_p(1)$.

Given the limiting representation in Theorem 2, the $100\eta\%$ confidence set for the l-period ahead of impulse response is given by the set of values of θ_l satisfying $LR_T \leq q_\eta(c)$ or equivalently $C_\eta(\theta) = \{\theta_l \in \Theta : \theta \in A(\theta_l)\}$, where Θ is the parameter space of θ, $q_\eta(c)$ is the ηth quantile of the asymptotic distribution in Theorem 2 and $A(\theta_l)$ is the acceptance region of the test LR_T.[6] The endpoints of the confidence set are the infimum and the supremum over $C_\eta(\theta)$, respectively.[7] The 100η % confidence interval for the half-life is $C_\eta(l) = \{l \in L : LR_T \leq q_\eta(c)\}$, where $\widetilde{\rho} = \arg\max l_T(\rho)$ subject to $\theta_l - 0.5 = 0$. Similarly, we can construct confidence intervals for this persistence measure for any $0 < \omega \leq 1$.

3. MONTE CARLO STUDY

To evaluate the small sample properties of the proposed method for constructing asymptotic confidence intervals for the impulse responses, we conducted a small Monte Carlo study. The simulation study includes only classical methods that have been shown to be asymptotically valid in the presence of a root near or on the unit circle. Therefore, we do not consider explicitly the bias-adjusted bootstrap of Kilian (1998), the standard delta method and the Bayesian confidence intervals of Sims and Zha (1999) as well as methods that are not transformation respecting such as the Hall's percentile and percentile-*t* methods (for a definition of a transformation respecting method, see Hall, 1992). One asymptotically valid method with good coverage properties which is also omitted from the simulation study is the conservative method proposed by Wright (2000). But because the design of our experiment is similar to Wright's (2000) Monte Carlo simulation study which includes all the above-mentioned methods, their coverage rates can be read off from Wright's (2000) figures and compared to our results.

The data for the first set of results are generated from a Gaussian AR model $(1 - bL)(1 - \phi L)y_t = e_t$ where $e_t \sim i.i.d.N(0, 1)$ with parameters $\phi = 0.9, 0.97, 0.99, 1$ or 1.005 and $b = -0.6$, $-0.3, 0, 0.3$ or 0.6. The sample size is $T = 100$ and the number of Monte Carlo replications is 5,000. Given the invariance of the limiting distribution to the values of the deterministic components, we set them, without loss of generality, to zero in the data-generating process. We consider both the demeaned and the detrended case, where the process is demeaned and detrended by OLS. The asymptotic confidence intervals are constructed by inverting the LR statistic using

[6] In practice, one can compute the LR statistic on a finite discretized grid of equally spaced points in the relevant range of Θ or using the bisection algorithm.

[7] Another useful test statistic can be constructed as

$$LR_T^{\pm} = sgn\,[\theta_l(\widehat{\rho}) - \theta_l(\widetilde{\rho})]\,\sqrt{LR_T},$$

where $sgn(.)$ denotes the sign of the expression in the brackets and $\widehat{\rho}$ and $\widetilde{\rho}$ are the unrestricted and restricted estimates, respectively. The limiting distribution of $LR_T^{\pm}$ is given by the square root of the expressions in Theorem 2. This test statistic can then be used for constructing two-sided confidence intervals and median unbiased estimates.

the limiting result[8] in Theorem 2 (*new ASY*) and the standard χ^2 asymptotics for stationary processes (χ^2 *ASY*). For comparison, we also include two popular bootstrap methods: the grid percentile (Hansen, 1999) and Efron's percentile methods. The number of bootstrap repetitions for the conventional percentile bootstrap (*CBOOT*) is set to 999. The grid bootstrap (*GBOOT*) approximates the distribution at 20 grid points with 699 repetitions at each point.

The empirical coverage rates of 95% confidence intervals of the impulse responses at lead times 1 to 24 ($b = -0.6, 0, 0.6.$) are plotted in Figures 1 and 2 for the demeaned and the detrended case, respectively. The method based on the inversion of the asymptotic acceptance region of the LR test has excellent coverage properties. A closer inspection of the figures reveals only slight overcoverage (1–2 percentage points) for some parameter configurations at lead times 1 and 2. Table 1 presents numerical values for the effective coverage rates of the method along with median lengths of the confidence intervals for the impulse response at lead time 12 ($b = -0.3, 0, 0.3$). Since our method does not exclude explosive roots, in some cases, the constructed confidence intervals may appear relatively wide.

It is interesting to note that even though the asymptotic distribution of the LR test is derived under the assumption that $l = O(T)$, the coverage rates of the method at very short horizons are also very good. A possible explanation for this result is that the long-horizon assumption $l = O(T)$ is used only for deriving the critical value but not in the construction of the test statistic. Approximating the constraint using the assumption $l = O(T)$, as we did at the end of Section 2.1, imposes monotonicity on the impulse response that may lead to nontrivial size distortions of the test at short horizons.

The grid bootstrap confidence intervals also perform well for lead times greater than 10 and stable AR roots but they tend to undercover at short horizons and explosive roots. As the magnitudes of the largest (ϕ) and stable (b) roots increase, the performance of the grid bootstrap deteriorates. One reason for this could be the imprecise estimation of the transitory component. Another feature of the grid bootstrap intervals is that they typically lie to the right of the *new ASY* intervals as the LR-based asymptotic confidence intervals are shifted more towards the stationary region.

As expected, the coverage rates of the conventional bootstrap and the χ^2 asymptotics deteriorate as the largest root approaches the unit root boundary and the number of estimated deterministic components increases. The extremely bad performance of the conventional bootstrap for impulse response inference of strongly persistent time series has already been reported elsewhere (for instance, Wright 2000) with the striking result of 0% coverage for many parameter configurations. (The coverage rates for ($\phi = 0.99, b = 0$), ($\phi = 1, b = 0$) and ($\phi = 1.005, b = 0$) in the detrended case are not visible on the graph because they lie on top of the horizontal axis.) Furthermore, the simulation study shows that the conventional bootstrap provides a very poor approximation even for parameter values that are away from the non-stationary boundary. This is in agreement with the theoretical results that the Efron's percentile bootstrap does not achieve the first-order asymptotic accuracy of the limit theory in non-stationary AR models.

To assess the properties of the method for higher-order models, we also generated data from an AR(4) model $(1 - b_1 L)(1 - b_2 L)(1 - b_3 L)(1 - \phi\, L)y_t = e_t$ with parameters $\phi = 0.9, 0.97, 0.99, 1$ or 1.005 and $(b_1, b_2, b_3) = (-0.4, 0.6, 0.2)$. The coverage rates are presented in Figure 3. The proposed method is characterized again by very good coverage properties although in some of the cases the effective coverage rates lie slightly above the nominal level.

[8]Critical values for the LR statistic are obtained by simulating the limiting representation in Theorem 2 with sample size 5,000 and 50,000 replications. A table of critical values is downloadable from the author's website.

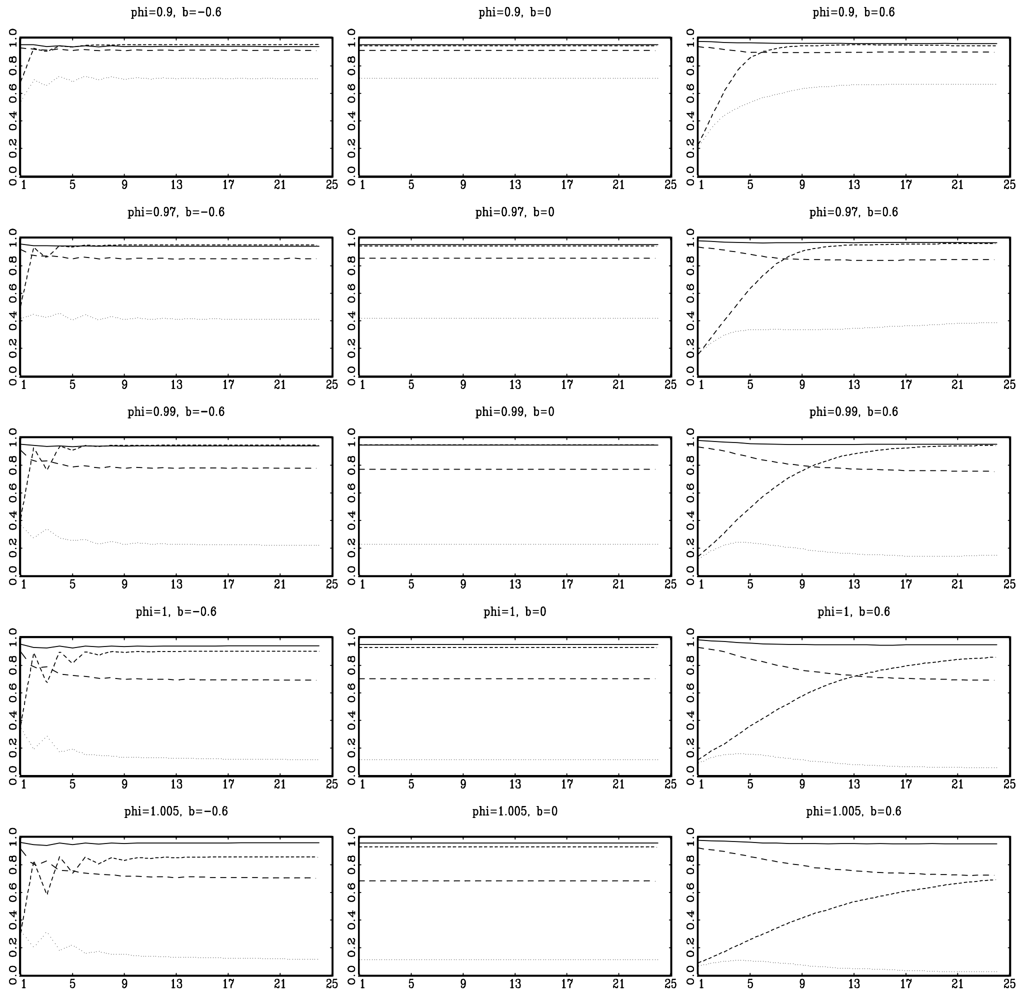

Figure 1. Coverage rates of 95% confidence intervals from the demeaned AR model $(1 - bL)(1 - \phi L)y_t = e_t$, where $e_t \sim i.i.d.N(0, 1)$. Solid line: LR test inversion with asymptotic approximation from Theorem 2; Dashed line: LR test inversion with χ^2 asymptotic approximation. Dotted line: Efron's percentile method; Short dashes: grid percentile method.

4. EMPIRICAL ANALYSIS OF PERSISTENCE IN REAL EXCHANGE RATES

The simultaneous presence of strong persistence and high short-run volatility in the real exchange rates constitutes one of the major puzzles in economics. Economic theory predicts a long-run mean reversion of the real exchange rates to their purchasing power parity (PPP). The speed of adjustment to the purchasing power parity is typically unspecified and needs to be determined empirically. Most of the empirical studies, however, find the mean reversion of real exchange rates to be very slow which appears to be inconsistent with many traditional macroeconomic models. The 'consensus estimates for the half-life of three to five years' reported by Rogoff (1996) have

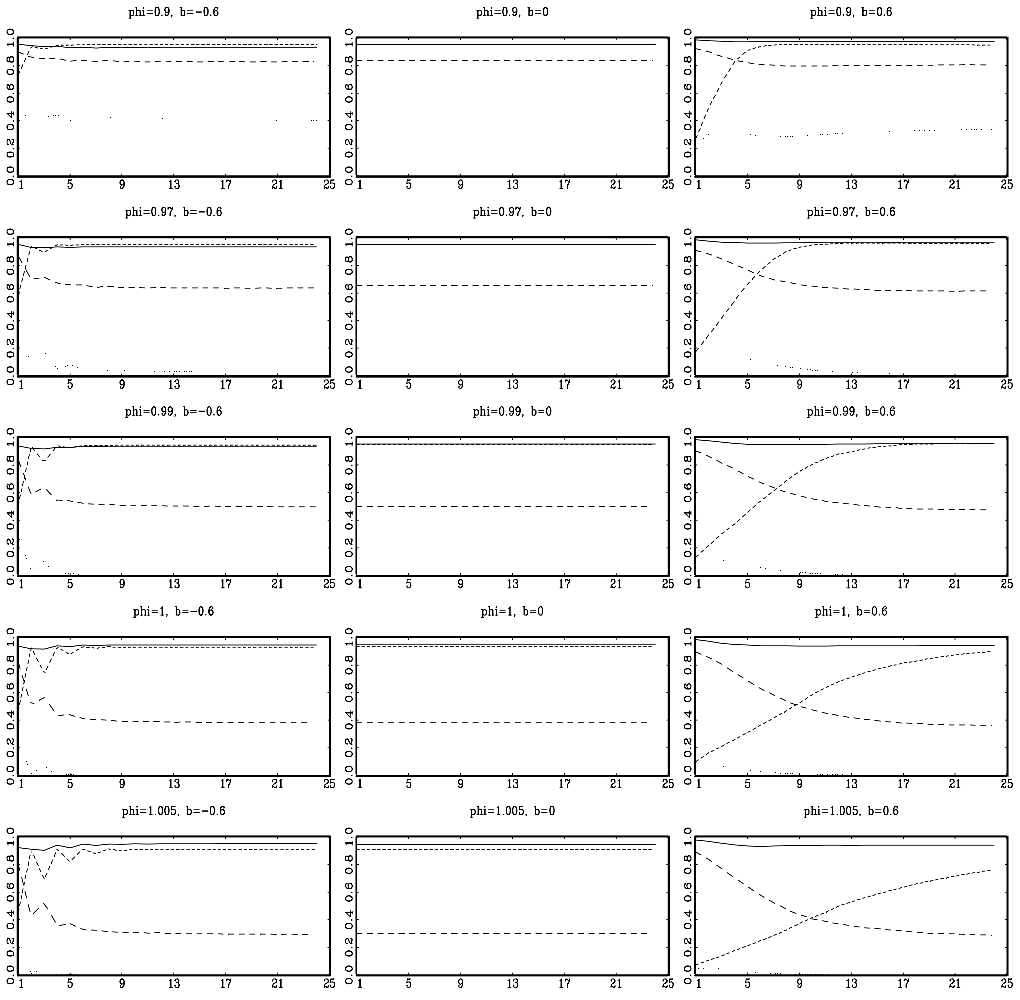

Figure 2. Coverage rates of 95% confidence intervals from the detrended AR model $(1 - bL)(1 - \phi L)y_t = e_t$, where $e_t \sim i.i.d.N(0, 1)$. Solid line: LR test inversion with asymptotic approximation from Theorem 2; Dashed line: LR test inversion with χ^2 asymptotic approximation; Dotted line: Efron's percentile method; Short dashes: grid percentile method.

been challenged recently by Murray and Papell (2002) and Kilian and Zha (2002). These studies report that the interval estimators of several univariate measures of persistence provide little support of the hypothesis of 3–5 years half-lives of purchasing power parity deviations.[9] Some potential pitfalls for the empirical work on the PPP puzzle have been pointed out by Taylor (2001).

[9]Note, however, that the 'consensus estimate' for the half-life of 3–5 years implies that in an AR(1) model the corresponding intervals for the AR parameter are [0.944, 0.966] for quarterly data and [0.981, 0.989] for monthly data. Given the short samples in the floating exchange period these intervals appear to be extremely tight at the commonly used significance levels and should be interpreted and compared with caution.

Table 1. Effective coverage rates and medium lengths of 95% confidence intervals for impulse response at lead time $l = 12$, constructed by the proposed (*newASY*) method.

	Demeaned		Detrended	
	Coverage (%)	Length	Coverage (%)	Length
$\phi = 0.9, b = -0.3$	94.54	0.876	94.26	1.005
$\phi = 0.9, b = 0$	95.60	1.180	95.36	1.388
$\phi = 0.9, b = 0.3$	95.78	1.661	95.92	2.084
$\phi = 0.97, b = -0.3$	94.48	0.969	94.74	1.088
$\phi = 0.97, b = 0$	94.92	1.262	94.68	1.464
$\phi = 0.97, b = 0.3$	95.50	2.052	95.30	2.527
$\phi = 0.99, b = -0.3$	94.12	0.963	95.36	1.094
$\phi = 0.99, b = 0$	95.16	1.232	94.66	1.459
$\phi = 0.99, b = 0.3$	94.84	2.080	95.14	2.620
$\phi = 1, b = -0.3$	94.50	0.944	94.86	1.095
$\phi = 1, b = 0$	94.62	1.202	95.26	1.463
$\phi = 1, b = 0.3$	95.14	2.100	94.30	2.626
$\phi = 1.005, b = -0.3$	95.62	0.913	95.28	1.099
$\phi = 1.005, b = 0$	95.16	1.129	95.22	1.462
$\phi = 1.005, b = 0.3$	95.00	2.143	94.58	2.623

This section presents empirical findings from interval estimators of the impulse responses for the British Pound, German Mark, Swiss Franc and French Franc real exchange rates. The nominal exchange rates against the US dollar and the consumer price level data are extracted from the IFS database of IMF and cover the period 1973:Q2–1998:Q4 at quarterly frequency and 1973:M4–1998:M12 at monthly frequency. The real exchange rates are constructed as $y_t = \ln(S_t P_t^*/P_t)$, where S_t is the nominal exchange rate (domestic currency per 1 unit of foreign currency), P_t and P_t^* are the domestic and foreign price levels. The quarterly frequency is our preferred frequency for computing persistence measures but we also include the monthly data to check the robustness of the findings.

The results are obtained from estimating a demeaned AR (2) model for each of the four real exchange rates with quarterly data and a demeaned AR(4) model for monthly data. The AR(2) model for quarterly data is selected by BIC but the diagnostic checks reveal that a small amount of serial correlation is still present in the residuals.[10] Also, for monthly data the ARCH-LM test detects conditional heteroskedasticity in the residuals for some of the real exchange rates. While our method is robust to the presence of some heterogeneity in the data, strong persistence in the conditional variance may lead to some distortions in the inference procedures although these distortions are expected to be small (Kim and Schmidt 1993; Seo 1999).

Table 2 reports the median unbiased estimates along with 68% and 90% confidence intervals for the localizing constant c of the series. These are constructed by inverting the acceptance region

[10]We also estimated AR(4) models for quarterly data that are favoured by the AIC criterion but the results are very similar to the AR(2) model and are omitted to preserve space.

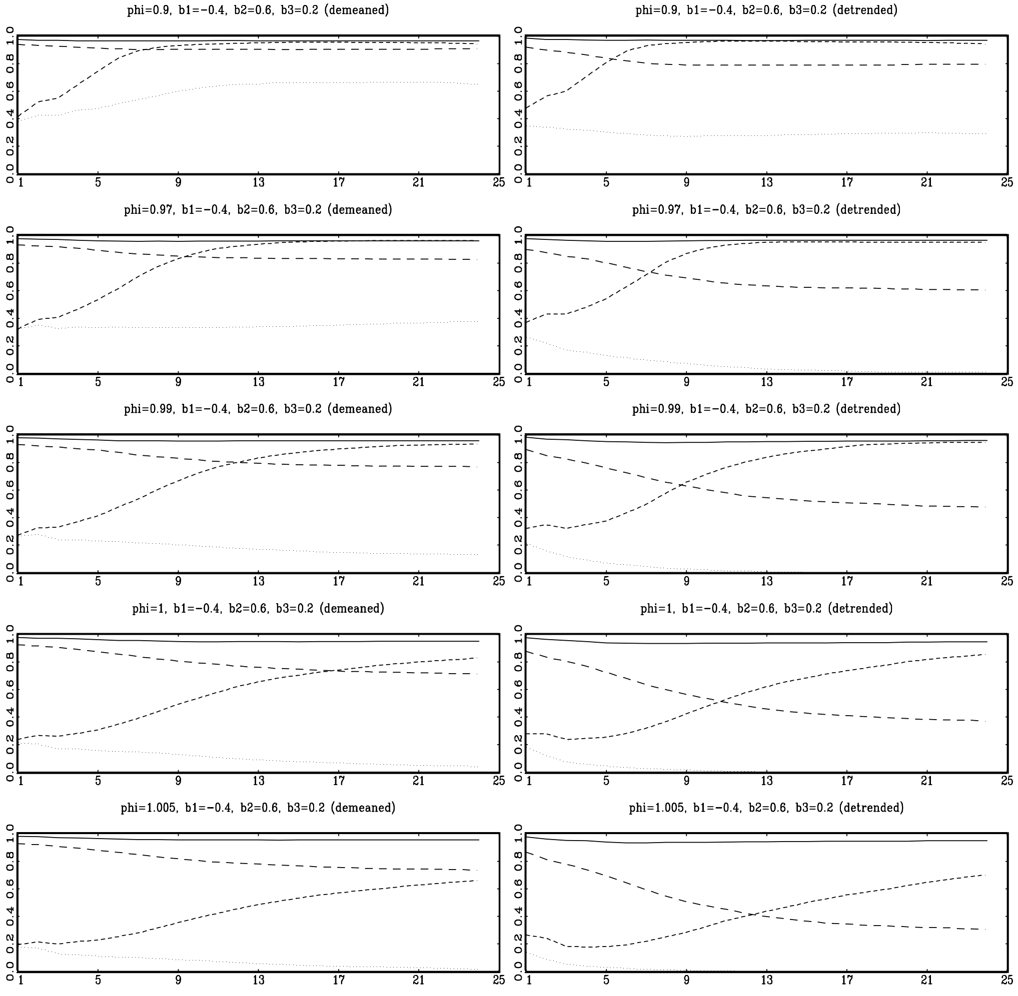

Figure 3. Coverage rates of 95% confidence intervals from the demeaned and detrended AR model $(1-b_1L)(1-b_2L)(1-b_3L)(1-\phi L)y_t = e_t$, where $e_t \sim i.i.d.N(0,\ 1)$. Solid line: LR test inversion with asymptotic approximation from Theorem 2; Dashed line: LR test inversion with χ^2 asymptotic approximation; Dotted line: Efron's percentile method; Short dashes: Grid percentile method.

of the DF-GLS test for $H_0 : c = c_0$ versus $H_1 : c \neq c_0$ (Elliott *et al.* 1996) whose finite-sample distribution is approximated by the grid bootstrap (Hansen 1999). The median unbiased estimates of c range from -8.3 to -5.9 for quarterly data and -7.2 and -3.5 for monthly data which is indicative of strong persistence in the series. Only the 90% confidence intervals for the Swiss Franc includes 0 (unit root case) although all of the 90% upper limits are very close to the non-stationary boundary. This supports our conjecture that the real exchange rates are very persistent processes but the presence of exact unit roots does not seem likely. Also, the confidence intervals appear to be rather tight which confirms the advantages of using more powerful tests for interval estimation (Elliott and Stock 2001).

Table 2. Point and interval estimates of the localizing constant c.

		Parameter c		
Real exchange rate	Frequency	MU estimate	68% CI	90% CI
British Pound	Quarterly	−8.34	[−12.45, −4.37]	[−15.11, −1.37]
	Monthly	−7.16	[−11.17, −3.00]	[−13.59, −0.10]
German Mark	Quarterly	−5.88	[−9.35, −2.12]	[−11.65, 0.83]
	Monthly	−6.51	[−10.17, −2.90]	[−12.61, −0.52]
Swiss Franc	Quarterly	−6.89	[−10.65, −3.03]	[−13.06, 0.26]
	Monthly	−3.46	[−6.26, −0.46]	[−7.89, 2.00]
French Franc	Quarterly	−6.55	[−9.94, −2.66]	[−12.44, −0.09]
	Monthly	−6.85	[−10.41, −3.11]	[−12.98, −0.55]

Note: The median unbiased (MU) estimates and the confidence intervals (CI) for c are constructed by the grid bootstrap method proposed by Hansen (1999) using the efficiently demeaned DF-GLS statistic of $H_0 : c = c_0$.

Figures 4 and 5 plot the 68% error bands of the impulse responses obtained by the likelihood-based test inversion method for quarterly and monthly data, respectively. The upper limits of the 68% interval estimators of the impulse response functions suggest that 25% of the adjustment to the shock is completed within 4 (5) years for all four real exchange rate using quarterly (monthly) data. At the specified confidence level, none of the real exchange rates appear to contain a unit root and the estimated speed of adjustment can be justified by some existing macroeconomic models with persistent real shocks. From the information provided by the impulse responses, we can construct a sequence of interval estimators that traces out the adjustment of the series at 1, 2, etc. periods ahead.

The results for the 68%, 80%, 90% and 95% confidence intervals of the half-life $\Lambda_{.5}$ are given in Tables 3 and 4. The confidence intervals for quarterly (monthly) data indicate that the half-life falls in the range 1.3–8 (1.3–9) years with 68% probability. At the same time, the 90% and 95% confidence intervals reveal that the uncertainty associated with the speed of mean reversion of the real exchange rates is substantial. We also constructed median unbiased estimates of the half-life using the $LR_T^{\pm}$ statistic. For monthly data, these estimates are between 3 and 4.2 years which lends some support to Rogoff's 'consensus estimate' if the latter is considered a range of likely point estimates for the different currencies.

In an innovative paper, Kilian and Zha (2002) obtain the posterior probability distribution of the half-life under the 'consensus prior' based on a survey of economists in the field of international macroeconomics. One interesting result that arises from Tables 3 and 4 is that the estimated probability distribution of the half-life appears to be very similar to the 'consensus prior' reported in Kilian and Zha (2002) for the post Bretton-Woods period. Our empirical analysis seems to provide formal support for the consensus beliefs of the economists expressed in the survey conducted by Kilian and Zha (2002). This interpretation could also reconcile the similarity of our classical confidence intervals and the Bayesian intervals based on the posterior of the half-life in Kilian and Zha (2002). Needless to say, all these comparisons must be interpreted with extreme caution due to the conceptual differences of the two approaches.

This paper adopts a univariate linear framework which implicitly imposes the restriction that the adjustment to the purchasing power parity occurs continuously with a constant speed. A natural extension of our approach is to explore the possibility that the mean reversion is non-linear due to the presence of transaction costs or noise trading and arbitrage (see Taylor 2001; Kilian and Taylor 2003; among others). In this case, the linear testing procedures might have difficulties detecting

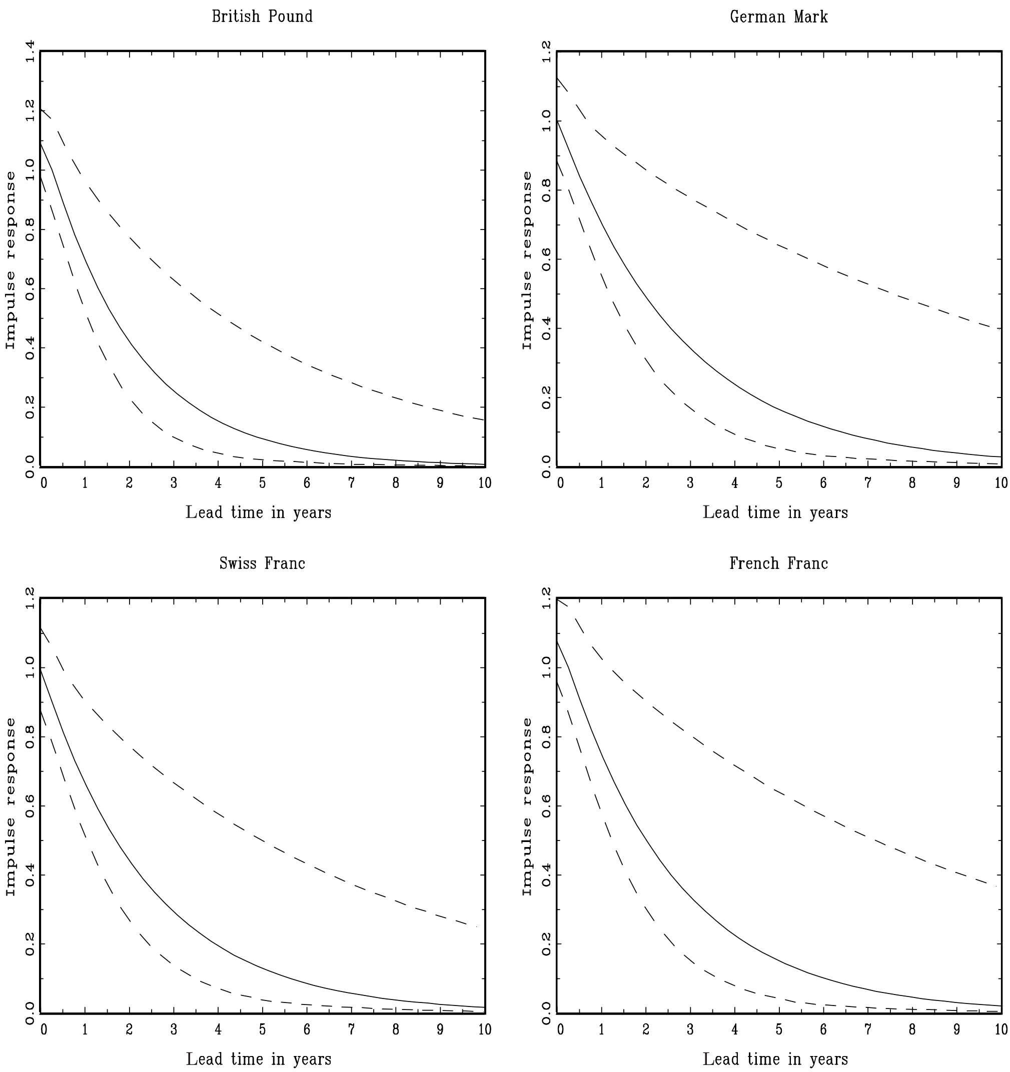

Figure 4. The OLS estimate (solid line) and 68% confidence intervals for the impulse response function from an estimated AR(2) model for quarterly real exchange rates.

and measuring the mean reversion of the process. If the adjustment is effective only if a critical threshold is reached, the real exchange rates would exhibit a random-walk-like behaviour in the regime where the PPP equilibrium forces are weak and a mean-reverting behaviour when the threshold barrier is passed. The presence of high persistence in the real exchange rates, however, complicates the testing and modelling of these non-linearities and the inference procedure needs to be modified appropriately (Caner and Hansen 2000). Finally, the relatively large uncertainty associated with the inference about the mean reversion of the real exchange rates undoubtedly reflects the limitations posed by the univariate model and calls for a more general approach that allows explicitly for underlying economic fundamentals and interdependencies among the different currencies.

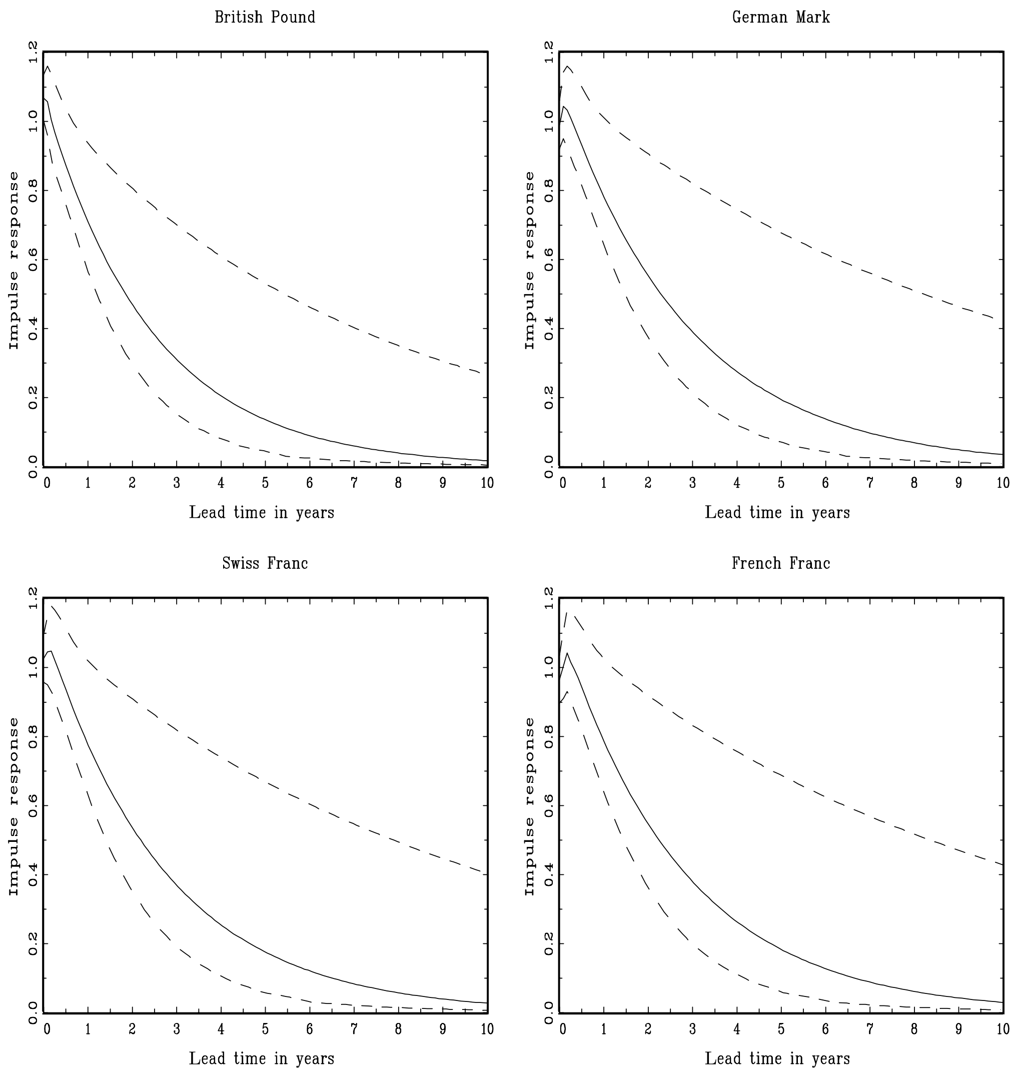

Figure 5. The OLS estimate (solid line) and 68% confidence intervals for the impulse response function from an estimated AR(4) model for monthly real exchange rates.

Table 3. Asymptotic confidence intervals for half-life (in years) from AR(2) model (quarterly data).

	68% CI		80% CI		90% CI		95% CI	
Real exchange rate	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
British Pound	1.30	4.37	1.20	8.37	1.11	>20	1.00	>20
German Mark	1.41	8.04	1.27	>20	1.17	>20	1.07	>20
Swiss Franc	1.27	5.28	1.18	17.04	1.06	>20	0.98	>20
French Franc	1.46	7.60	1.35	>20	1.22	> 20	1.13	>20

Table 4. Asymptotic confidence intervals for half-life (in years) from AR(4) model (monthly data).

	68% CI		80% CI		90% CI		95% CI	
Real exchange rate	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
British Pound	1.44	4.13	1.34	7.26	1.22	>20	1.15	>20
German Mark	1.68	6.76	1.51	>20	1.35	>20	1.23	>20
Swiss Franc	1.39	4.91	1.25	12.45	1.16	>20	1.07	>20
French Franc	1.65	6.02	1.50	19.25	1.37	>20	1.25	>20

5. CONCLUSION

This paper considers the properties of the constrained estimator of the non-stationary component in a near-integrated AR model subject to a polynomial restriction whose order increases linearly with the sample size. This situation is of great practical relevance since it arises often in testing hypothesis on impulse response-based measures of persistence (such as the half-life) for local-to-unity processes. Interestingly, we find that the restricted estimator of the non-stationary component converges faster than the unrestricted estimator which helps us obtain a consistent estimate of the localizing constant. The paper develops a LR test inversion method for constructing confidence intervals which evaluates the LR statistic for a sequence of null hypotheses that restrict the values of the half-life or the impulse responses. Furthermore, we derive the limiting distribution of the LR statistic and show that the validity of the asymptotic inference is ensured by the consistent estimation of the localizing constant. We find that the inversion of the LR appears to be controlling the coverage over a wide range of parameter configurations and across different forecasting horizons. The application of the method to real exchange rates reveals substantial uncertainty about their speed of mean reversion and half-lives.

ACKNOWLEDGEMENTS

I would like to thank the Editor (Pierre Perron), three anonymous referees, Bruce Hansen, Lutz Kilian, Serena Ng, Barbara Rossi and Jonathan Wright for very helpful comments and suggestions. Cristian Troncoso provided excellent research assistance. Financial support from the FRDP (Concordia University), FCAR (Quebec) and SSHRC of Canada is gratefully acknowledged.

REFERENCES

Andrews, D. W. K. and H-Y. Chen (1994). Approximately median-unbiased estimation of autoregressive models. *Journal of Business and Economic Statistics 12*, 187–204.

Basawa, I. V., A. K., Mallik, W. P., McCormick, J. H. Reeves and R. L. Taylor (1991). Bootstrapping unstable first-order autoregressive processes. *Annals of Statistics 19*, 1098–1101.

Bekaert, G. and R. J. Hodrick (2001). Expectations hypotheses tests. *Journal of Finance 56*, 1357–94.

Caner, M. and B. E. Hansen (2001). Threshold autoregressions with a unit root. *Econometrica 69*, 1555–96.

Chan, N. H. and C. Z. Wei (1987). Asymptotic inference for nearly non-stationary AR(1) processes. *Annals of Statistics 15*, 1050–63.

Elliott, G., T. J. Rothenberg and J. H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica 64*, 813–36.

Elliott, G. and J. H. Stock (2001). Confidence intervals for autoregressive coefficients near one. *Journal of Econometrics 103*, 155–81.

Gospodinov, N. (2002). Median unbiased forecasts for highly persistent autoregressive processes. *Journal of Econometrics 111*, 85–101.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hansen, B. E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics and Statistics 81*, 594–607.

Hansen, B. E. (2000). Edgeworth expansions for the Wald and GMM statistics for non-linear restrictions. Unpublished manuscript, University of Wisconsin-Madison.

Inoue, A. and L. Kilian (2002). Bootstrapping autoregressive processes with possible unit roots. *Econometrica 70*, 377–91.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics 80*, 218–30.

Kilian, L. and M. P. Taylor (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics 60*, 85–107.

Kilian, L. and T. Zha (2002). Quantifying the uncertainty about the half-life of deviations from PPP. *Journal of Applied Econometrics 17*, 107–25.

Kim, K. and P. Schmidt (1993). Unit root tests with conditional heteroskedasticity. *Journal of Econometrics 59*, 287–300.

Lütkepohl, H. (1990). Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models. *Review of Economics and Statistics 72*, 116–25.

Miller, J. P. and P. Newbold (1995). Uncertainty about the persistence of economic shocks. *Journal of Business and Economic Statistics 13*, 435–40.

Moon, H. R. and P. C. B. Phillips (2000). Estimation of autoregressive roots near unity using panel data. *Econometric Theory 16*, 927–97.

Moon, H. R. and F. Schorfheide (2002). Minimum distance estimation of non-stationary time series models. *Econometric Theory 18*, 1385–1407.

Murray, C. J. and D. H. Papell (2002). The purchasing power parity persistence paradigm. *Journal of International Economics 56*, 1–19.

Nagaraj, K. N. and W. A. Fuller (1991). Estimation of the parameters of linear time series models subject to non-linear restrictions. *Annals of Statistics 19*, 1143–54.

Ng, S. and P. Perron (2002). PPP may not hold afterall: A further investigation. *Annals of Economics and Finance 3*, 43–64.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika 74*, 535–47.

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. and S. N. Durlauf (1986). Multiple time series regression with integrated processes. *Review of Economic Studies 53*, 473–95.

Phillips, P. C. B., H. R. Moon, and Z. Xiao (2000). How to estimate autoregressive roots near unity. *Econometric Theory 17*, 29–69.

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic Literature 34*, 647–68.

Rossi, B. (2004). Confidence intervals for half-life deviations from purchasing power parity, Unpublished manuscript, Duke University.

Seo, B. (1999). Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics 91*, 113–44.

Sims, C. A., J. H. Stock and M. W. Watson (1990). Inference in linear time series with some unit roots. *Econometrica 58*, 113–44.

Sims, C. A. and T. Zha (1999). Error bands for impulse responses. *Econometrica 67*, 1113–55.
Stock, J. H. (1991). Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics 28*, 435–59.
Stock, J. H. (1996). VAR, error-correction and pretest forecasts at long horizons. *Oxford Bulletin of Economics and Statistics 58*, 685–701.
Taylor, A. M. (2001). Potential pitfalls for the purchasing-power-parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price. *Econometrica 69*, 473–98.
Valkanov, R. (1998). The term structure with highly persistent interest rates, Unpublished manuscript, Princeton University.
Wright, J. H. (1997). Inference for impulse response functions, Unpublished manuscript, Harvard University.
Wright, J. H. (2000). Confidence intervals for univariate impulse responses with a near unit root. *Journal of Business and Economic Statistics 18*, 368–73.

APPENDIX A

Mathematical Proofs

Proof of Theorem 1. **Part(i)** The structure of this proof follows Nagaraj and Fuller (1991). First, we show the consistency of the restricted estimator.

Recall that ρ_0 denotes the true value of the parameter vector, $\widehat{\rho} = \arg\max_{\rho\in\Xi} Q_T(\rho)$ is the unrestricted estimator and $\widetilde{\rho} = \arg\max_{h(\rho)=0} Q_T(\rho)$ is the constrained estimator. From Phillips (1987) and Sims *et al.* (1990), $\Pi_{1,T}(\widehat{\rho} - \rho_0) = O_p(1)$, where $\Pi_{1,T} = diag(T, \sqrt{T}, \ldots, \sqrt{T})$, i.e. the estimator of the non-stationary component is T-consistent and the estimator of the stationary component is $\sqrt{T}$-consistent. Define the matrix $B_T = \Pi_{1,T}^{-1}(\sum_{t=1}^{T} x_{t-1}x_{t-1}')\Pi_{1T}^{-1}$, where $x_{t-1} = (y_{t-1}, \triangle y_{t-1}, \ldots, \triangle y_{t-p+1})'$, and let $b_{\min}$ and $b_{\max}$ be the smallest and largest roots of B_T, respectively. Because B_T is properly standardized, $b_{\min} = O_p(1)$ and $b_{\max} = O_p(1)$. Then,

$$|\Pi_{1,T}(\widetilde{\rho} - \rho_0)|^2 \leq |\Pi_{1,T}(\widetilde{\rho} - \widehat{\rho})|^2 + |\Pi_{1,T}(\widehat{\rho} - \rho_0)|^2$$

and

$$|\Pi_{1,T}(\widetilde{\rho} - \widehat{\rho})|^2 \leq b_{\min}^{-1}(\widetilde{\rho} - \widehat{\rho})'\,\Pi_{1,T}B_T\Pi_{1,T}\,(\widetilde{\rho} - \widehat{\rho}).$$

Note that the objective function of the constrained minimization problem can be expressed as

$$\begin{aligned} Q_T(\rho_0) &= \sum_{t=1}^{T}(y_t - x_{t-1}'\widehat{\rho})^2 + (\widehat{\rho} - \rho_0)'\,\Pi_{1,T}B_T\Pi_{1,T}\,(\widehat{\rho} - \rho_0) \\ Q_T(\widetilde{\rho}) &= \sum_{t=1}^{T}(y_t - x_{t-1}'\widehat{\rho})^2 + (\widehat{\rho} - \widetilde{\rho})'\,\Pi_{1,T}B_T\Pi_{1,T}\,(\widehat{\rho} - \widetilde{\rho}) \end{aligned}$$

since ρ_0 and $\widetilde{\rho}$ satisfy the restriction. Because $\widetilde{\rho}$ minimizes $Q_T(\rho)$ subject to the constraints, $Q_T(\widetilde{\rho}) \leq Q_T(\rho)$ and hence,

$$b_T^{-1}(\widetilde{\rho} - \widehat{\rho})'\,\Pi_{1,T}B_T\Pi_{1,T}\,(\widetilde{\rho} - \widehat{\rho}) \leq b_{\min}^{-1}(\widehat{\rho} - \rho_0)'\,\Pi_{1,T}B_T\Pi_{1,T}\,(\widehat{\rho} - \rho_0).$$

Therefore,

$$|\Pi_{1,T}(\widetilde{\rho} - \rho_0)|^2 \leq 2b_{\min}^{-1}b_{\max}|\Pi_{1,T}(\widehat{\rho} - \rho_0)|^2 \tag{A.1}$$

which implies that the restricted estimator is consistent and its rate of convergence is at least as fast as that of the unrestricted estimator.

Next, we determine the exact rate of convergence of the restricted estimator. For the subsequent results, it will be useful to define the restricted estimator as the value of ρ that minimizes the Lagrangian

$$\frac{1}{2}\sum_{t=1}^{T}(y_t - x'_{t-1}\rho)^2 + \lambda h(\rho), \tag{A.2}$$

where λ is the Lagrange multiplier. By the mean value theorem for $h(\rho)$, the first-order conditions of the constrained minimization problem are given by

$$\begin{bmatrix} \left(\sum_{t=1}^{T} x_{t-1}x'_{t-1}\right)(\widetilde{\rho} - \rho_0) + d(\widetilde{\rho})\lambda \\ d(\rho^*)'(\widetilde{\rho} - \rho_0) \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{T} x_{t-1}e_t \\ 0 \end{bmatrix}, \tag{A.3}$$

where $d(\rho) = \frac{\partial h(\rho)}{\partial \rho'}$ is a $p \times 1$ vector of partial derivatives and ρ^* is a point on the line segment joining $\widetilde{\rho}$ and ρ_0. This system of equations can be rewritten in matrix form as

$$\begin{bmatrix} \left(\sum_{t=1}^{T} x_{t-1}x'_{t-1}\right) & d(\widetilde{\rho}) \\ d(\rho^*)' & 0 \end{bmatrix} \begin{bmatrix} (\widetilde{\rho} - \rho_0) \\ \lambda \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{T} x_{t-1}e_t \\ 0 \end{bmatrix}. \tag{A.4}$$

Let $G_T = [d(\rho_0)'(\sum_{t=1}^{T} x_{t-1}x'_{t-1})^{-1}d(\rho_0)]^{-1}$ and premultiply both sides of (A.4) by $\Pi_{1,T}^{-1}$ and $G_T^{1/2}$. Define $\widetilde{R}_T = \Pi_{1,T}^{-1}\widetilde{d}G_T^{1/2}$ and $R_T^* = \Pi_{1,T}^{-1}d^*G_T^{1/2}$, where $\widetilde{d} = d(\widetilde{\rho})$ and $d^* = d(\rho^*)$. Then, we can rewrite the system of equations in (A.4) as

$$\begin{bmatrix} B_T & \widetilde{R}_T \\ R_T^{*'} & 0 \end{bmatrix} \begin{bmatrix} \Pi_{1,T}(\widetilde{\rho} - \rho_0) \\ G_T^{-1/2}\lambda \end{bmatrix} = \begin{bmatrix} \Pi_{1,T}^{-1}\sum_{t=1}^{T} x_{t-1}e_t \\ 0 \end{bmatrix}.$$

From the consistency of $\widetilde{\rho}$ shown above and the continuity of d, it follows that $\widetilde{R}_T \to R_T$ and $R_T^* \to R_T$, where $R_T = R_T(\rho_0)$. Then,

$$\begin{bmatrix} \Pi_{1,T}(\widetilde{\rho} - \rho_0) \\ G_T^{-1/2}\lambda \end{bmatrix} = \begin{bmatrix} B_T & R_T \\ R'_T & 0 \end{bmatrix}^{-1} \begin{bmatrix} \Pi_{1,T}^{-1}\sum_{t=1}^{T} x_{t-1}e_t \\ 0 \end{bmatrix} + o_p(1) \tag{A.5}$$

provided that $\begin{bmatrix} B_T & R_T \\ R'_T & 0 \end{bmatrix}^{-1} = O_p(1)$. From the formula for the inverse of a partitioned matrix,

$$\begin{bmatrix} B_T & R_T \\ R'_T & 0 \end{bmatrix}^{-1} = \begin{bmatrix} I & -B_T^{-1}R_T \\ 0 & I \end{bmatrix} \begin{bmatrix} B_T^{-1} & 0 \\ 0 & -\left(R'_T B_T^{-1} R_T\right) \end{bmatrix} \begin{bmatrix} I & 0 \\ -R'_T B_T^{-1} & I \end{bmatrix}. \tag{A.6}$$

By construction, $R'_T B_T^{-1} R_T = I$ and $B_T^{-1} = O_p(1)$. Because $R'_T B_T^{-1/2'} B_T^{-1/2} R_T = I$, $R'_T B_T^{-1/2'} = O_p(1)$ and $R'_T B_T^{-1} = R'_T B_T^{-1/2'} B_T^{-1/2} = O_p(1)$ which ensures that the inverse of the above matrix is $O_p(1)$.

Using (A.6), we can rewrite (A.5) as

$$\begin{bmatrix} \Pi_{1,T}\,(\widetilde{\rho}-\rho_0) \\ G_T^{-1/2}\lambda \end{bmatrix} = \begin{bmatrix} \Sigma_T & \Gamma_T \\ \Gamma_T' & \Omega_T \end{bmatrix} \begin{bmatrix} \Pi_{1,T}^{-1}\sum_{t=1}^{T} x_{t-1}e_t \\ 0 \end{bmatrix} + o_p(1), \tag{A.7}$$

where $\Sigma_T = B_T^{-1/2'}(I - B_T^{-1/2}R_T(R_T'B_T^{-1}R_T)^{-1}R_T'B_T^{-1/2'})B_T^{-1/2}$, $\Phi_T = -(R_T'B_T^{-1}R_T)^{-1}$ and $\Gamma_T = -\Phi_T R_T' B_T^{-1}$. Next, we rewrite the first set of equations as

$$\Pi_{2,T}\,(\widetilde{\rho}-\rho_0) = \Psi_T\,\Sigma_T\,\Pi_{1,T}^{-1}\sum_{t=1}^{T} x_{t-1}e_t + o_p(1), \tag{A.8}$$

where $\Pi_{2,T} = diag(T^{3/2}, \sqrt{T}, \ldots, \sqrt{T}) = \Psi_T\Pi_{1,T}$ and $\Psi_T = diag(T^{1/2}, 1, \ldots, 1)$.

Let $P_T = B_T^{-1/2}R_T(R_T'B_T^{-1}R_T)^{-1}R_T'B_T^{-1/2'}$. To simplify the notation, we omit the $o_p(1)$ term and suppress the dependence of the matrices on T. From (A.8),

$$\begin{aligned}
|\Pi_2\,(\widetilde{\rho}-\rho_0)|^2 &= \left(\Pi_1^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right)' B^{-1/2'}\,(I-P)\,B^{-1/2}\Psi\Psi B^{-1/2'}\,(I-P)\,B^{-1/2} \\
&\quad\times \left(\Pi_1^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right) \\
&= tr\left[\left(\Pi_1^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right)' \Psi B^{-1/2'}\,(I-P)\,B^{-1/2}B^{-1/2'}\,(I-P)\,B^{-1/2}\Psi \right. \\
&\quad\left.\times \left(\Pi_1^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right)\right] \\
&= \left(\Pi_1^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right)' \Psi B^{-1/2'}\,(I-P)\,B^{-1/2}B^{-1/2'}\,(I-P)\,B^{-1/2}\Psi \\
&\quad\times \left(\Pi_1^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right) \\
&\le b_{\min}^{-1}\left(\Pi_1^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right)' \Psi B^{-1/2'}\,(I-P)\,B^{-1/2}BB^{-1/2'}\,(I-P)\,B^{-1/2}\Psi \\
&\quad\times \left(\Pi_1^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right) \\
&= b_{\min}^{-1}\left(\Pi_1^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right)' \Psi B^{-1/2'}\,(I-P)\,B^{-1/2}\Psi\left(\Pi_1^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right),
\end{aligned}$$

because $B^{-1/2}BB^{-1/2'} = I$, $(I-P)$ is an idempotent matrix and, by the properties of the trace, $a = tr(a)$ for a scalar a and $tr(AB) = tr(BA)$.

Note that if $l = [\delta T]$ for some fixed $\delta > 0$, then $\alpha^l = (1+\frac{c}{T})^l = (1+\frac{c}{T})^{[\delta T]} \to \exp(c\delta)$ and $d_0 = (O(T), O(1), \ldots, O(1))'$. Given this result, it can be verified that the elements of matrix $P = [p_{ij}]$ are $p_{11} = O_p(1)$, $p_{ii} = O_p(T^{-1})$ for $i > 1$ and $p_{ij} = O_p(T^{-1/2})$, for $i \neq j$. Hence, all elements of $\Psi\ B^{-1/2'}(I - P)\ B^{-1/2}\Psi$ are $O_p(1)$. Since $\Pi_{1,T}^{-1}\sum_{t=1}^{T} x_{t-1}e_t = (O_p(1), \ldots, O_p(1))'$ (Phillips 1987; Sims *et al.* 1990), we obtain that $\Pi_{2,T}(\widetilde{\rho}-\rho_0) = (O_p(1), \ldots, O_p(1))'$.

Finally, from $T^{3/2}(\widetilde{\alpha}-\alpha_0) = O_p(1)$, $\widetilde{\alpha} = 1+\widetilde{c}/T$ and $\alpha_0 = 1 + c/T$, we get $T^{1/2}(\widetilde{c}-c) = O_p(1)$ and $(\widetilde{c}-c) = o_p(1)$.

Part(ii): By the triangle inequality,

$$|T(\widehat{\alpha}-\widetilde{\alpha})| \leq |T(\widehat{\alpha}-\alpha_0)| + |T(\widetilde{\alpha}-\alpha_0)|\,.$$

From Phillips (1987) and part (i) of this theorem, $|T(\widehat{\alpha}-\alpha_0)| = O_p(1)$ and $|T(\widetilde{\alpha}-\alpha_0)| = o_p(1)$. Thus, $|T(\widehat{\alpha}-\widetilde{\alpha})| = O_p(1)$.

Also, $\Pi_{1,T}(\widehat{\rho}-\widetilde{\rho}) = \Pi_{1,T}(\widehat{\rho}-\rho_0) - \Pi_{1,T}(\widetilde{\rho}-\rho_0) = (B_T^{-1}-\Sigma_T)\Pi_{1,T}^{-1}\sum_{t=1}^{T} x_{t-1}e_t + o_p(1)$. Since $\left|\Pi_{1,T}^{-1}\sum_{t=1}^{T} x_{t-1}e_t\right| = O_p(1)$ and $B_T^{-1} - \Sigma_T = P$ with $p_{ii} = O_p(T^{-1})$ for $i > 1$ and $p_{ij} = O_p(T^{-1/2})$, for $i \neq j$ (see the proof of part (i) above), $\sqrt{T}(\widehat{\psi}-\widetilde{\psi}) = o_p(1)$.

Part(iii): Take a first-order Taylor series expansion of $h(\widetilde{\rho})$ around ρ_0. Since $d(\rho_0) = O(T)$, the proper standardization is

$$\sqrt{T}h(\widetilde{\rho}) = \sqrt{T}h(\rho_0) + \left[T^{1/2}\Pi_{2T}^{-1}d(\rho^*)\right]'\left[\Pi_{2T}(\widetilde{\rho}-\rho_0)\right].$$

Because the second term on the right-hand side is $O_p(1)$, then $\sqrt{T}\,(h(\widetilde{\rho})-h(\rho_0)) = O_p(1)$.□

Proof of Theorem 2. If $p = 1$, the proof is straightforward and follows directly from Phillips (1987). For $p > 1$, let $Y_{t-1} = (\triangle y_{t-1}, \ldots, \triangle y_{t-p+1})'$, $x_{t-1} = (y_{t-1}, Y'_{t-1})'$ and $\psi = (\psi_1, \ldots, \psi_{p-1})'$ in the zero-mean case. The residual sum of squares under the null $H_0{:}h(\rho) = 0$ is given by

$$\begin{aligned}
SSR_0 &= \sum_{t=1}^{T}(y_t - x'_{t-1}\widetilde{\rho})^2\\
&= \sum_{t=1}^{T}[\widehat{e}_t + x'_{t-1}(\widehat{\rho}-\widetilde{\rho})]^2\\
&= \sum_{t=1}^{T}[\widehat{e}_t + (\widehat{\alpha}-\widetilde{\alpha})y_{t-1} + Y'_{t-1}(\widehat{\psi}-\widetilde{\psi})]^2\\
&= SSR + (\widehat{\alpha}-\widetilde{\alpha})^2\sum_{t=1}^{T}y_{t-1}^2 + (\widehat{\psi}-\widetilde{\psi})'\sum_{t=1}^{T}Y_{t-1}Y'_{t-1}(\widehat{\psi}-\widetilde{\psi})\\
&\quad + 2(\widehat{\alpha}-\widetilde{\alpha})\sum_{t=1}^{T}\widehat{e}_t y_{t-1} + 2(\widehat{\psi}-\widetilde{\psi})'\sum_{t=1}^{T}\widehat{e}_t Y_{t-1} + 2(\widehat{\alpha}-\widetilde{\alpha})(\widehat{\psi}-\widetilde{\psi})'\sum_{t=1}^{T}y_{t-1}Y_{t-1}
\end{aligned}$$

using that $\widehat{e}_t = \widetilde{e}_t - x'_{t-1}(\widehat{\rho}-\widetilde{\rho})$, where $\widehat{e}_t$ and $\widetilde{e}_t$ are the unrestricted and restricted residuals, respectively.

Since $\sqrt{T}(\widehat{\psi}-\widetilde{\psi}) = o_p(1)$ from part (ii) of Theorem 1, $(\widehat{\psi}-\widetilde{\psi})'\sum_{t=1}^{T}\widehat{e}_t Y_{t-1} = o_p(1)$ and $(\widehat{\psi}-\widetilde{\psi})'\sum_{t=1}^{T}Y_{t-1}Y'_{t-1}(\widehat{\psi}-\widetilde{\psi}) = o_p(1)$. Also from Theorem 1, $T(\widehat{\alpha}-\widetilde{\alpha}) = O_p(1)$ and hence $(\widehat{\alpha}-\widetilde{\alpha})(\widehat{\psi}-\widetilde{\psi})'\sum_{t=1}^{T}y_{t-1}Y_{t-1} = o_p(1)$.

From the $T^{3/2}$-consistency of $\widetilde{\alpha}$, it follows that $T(\widehat{\alpha}-\widetilde{\alpha}) = T(\widehat{\alpha}-\alpha_0) + o_p(1)$. After substituting for $\widehat{e}_t = e_t - x'_{t-1}(\widehat{\rho}-\rho_0)$ in the fourth term and using the last result in the previous paragraph, we obtain $2(\widehat{\alpha}-\widetilde{\alpha})\sum_{t=1}^{T}[e_t - (\widehat{\alpha}-\alpha_0)y_{t-1}]\,y_{t-1} + o_p(1) = 2(\widehat{\alpha}-\alpha_0)\sum_{t=1}^{T}y_{t-1}e_t - 2(\widehat{\alpha}-\alpha_0)\sum_{t=1}^{T}y_{t-1}^2$. Finally, from $\sum_{t=1}^{T}y_{t-1}Y_{t-1} = O_p(T)$, we have that $(\widehat{\alpha}-\widetilde{\alpha})(\widehat{\psi}-\widetilde{\psi})'\sum_{t=1}^{T}y_{t-1}Y_{t-1} = o_p(1)$.

Collecting the results gives

$$SSR_0 - SSR = 2\,[T(\widehat{\alpha}-\alpha_0)]\left[T^{-1}\sum_{t=1}^{T}y_{t-1}e_t\right] - [T(\widehat{\alpha}-\alpha_0)]^2\left[T^{-2}\sum_{t=1}^{T}y_{t-1}^2\right] + o_p(1).$$

From Phillips (1987), $T^{-1}\sum_{t=2}^{T}y_{t-1}e_t \Rightarrow \sigma^2\psi(1)\int_0^1 J_c(s)\,dW(s)$, $T^{-2}\sum_{t=2}^{T}y_{t-1}^2 \Rightarrow \sigma^2\psi(1)^2\int_0^1 J_c(s)^2\,ds$ and $T(\widehat{\alpha}-\alpha_0) \Rightarrow \psi(1)^{-1}(\int_0^1 J_c(s)^2\,ds)^{-1}\int_0^1 J_c(s)\,dW(s)$, where $\psi\ (1) = 1 - \psi_1 - \cdots - \psi_{p-1}$. Then, using $T^{-1}SSR \xrightarrow{p} \sigma^2$ and $LR_T = T(SSR_0 - SSR)/SSR + o_p(1)$, the asymptotic distribution of the LR_T statistic becomes

$$LR_T \Rightarrow \frac{\left[\int_0^1 J_c(s)\, dW(s)\right]^2}{\int_0^1 J_c(s)^2\, ds}$$

which proves the zero-mean case of Theorem 2. The demeaned and detrended cases can be proved analogously. □