Estimation and Inference for Impulse Response Weights From Strongly Persistent Processes

Richard T. Baillie*
Departments of Economics and Finance, Michigan State University
School of Economics and Finance, Queen Mary, University of London

George Kapetanios
School of Economics and Finance, Queen Mary, University of London

May 23, 2012

Abstract

This paper is concerned with the estimation and construction of confidence intervals for Impulse Response Weights (IRWs) from strongly persistent time series. A non parametric, time domain estimator based on an autoregressive (AR) approximation is shown to have good theoretical and small sample properties for the estimation of $IRWs$. An alternative procedure of using a semi-parametric Local Whittle (LW) estimator of the long memory parameter and then obtaining estimates of the short run parameters and IRWs is also considered. The second part of the paper investigates the most appropriate methods for estimating the variability and the construction of confidence intervals for the estimated $IRWs$. Particular attention is given to a generic semi-parametric sieve bootstrap based on an autoregressive approximation of the unknown data generating mechanism. The validity of bootstrap inference on the IRWs, based on the autoregressive approximation, is proven under mild assumptions. The findings in this paper indicate that a good strategy for analyzing $IRWs$ is to estimate by semi-parametric AR approximations, and to use the sieve bootstrap for estimating confidence intervals. Simulation evidence indicates this approach appears to be a very good strategy for processes with either short or long memory. An empirical example concerning the persistence of real exchange rate series is included.

Key Words: Persistence, Impulse Responses, Autoregressive Approximation, Confidence Intervals.

JEL Codes: C22, C12.

1 Introduction

This paper is concerned with issues relating to the estimation and construction of confidence intervals for Impulse Response Weights (IRWs) from strongly persistent time series. This is an important class of time series processes which includes those with relatively slowly decaying

*Helpful comments have been received from Tim Bollerslev, Luidas Giraitis, Hira Koul, Peter Phillips, Peter Schmidt and Andrew Tremayne. Any remaining errors are our own.

hyperbolic autocorrelations; see Granger and Joyeux (1980), Granger (1980) and Hosking (1981). These models have proved very relevant for representing the behavior of many economic and financial time series.

The paper first considers estimation of the $IRWs$ and focuses on a non parametric time domain estimator based on an autoregressive (AR) approximation. This estimator is shown to have good theoretical and small sample properties. This study also considers the use of the Local Whittle (LW) estimator of the long memory parameter and analyzes the effects of using the LW semi parametric estimator (SPE) of the long memory parameter to obtain estimates of short run parameters and IRWs.

The second part of the paper investigates the most appropriate methods for estimating the variability and the construction of confidence intervals for the estimated $IRWs$. There has in general been a long standing concern in the literature over this issue. For example, Sims (1986) has considered this problem for weakly dependent processes, such as stationary Vector Autoregressions (VAR); while Wright (2000) has considered $IRWs$ from near unit root processes. A major finding of existing work is that confidence intervals based on asymptotic approximations can provide a poor guide to the true finite sample confidence intervals, and one alternative which is pursued in this paper is to use bootstrapping methods.

We extend the previous literature on IRW confidence intervals to the empirically important and relevant class of strongly dependent processes. We consider a generic semi-parametric sieve bootstrap based on an autoregressive approximation of the unknown data generating mechanism. Under mild assumptions we show the validity of bootstrap inference on IRWs based on the AR approximation. Our results indicate that the sieve bootstrap has a number of benefits. Hence, the findings in this paper indicate that a good strategy for analyzing $IRWs$ is to estimate them by semi-parametric AR approximations, and to use the sieve bootstrap for estimating confidence intervals. Furthermore this approach appears to be a very good strategy for processes with either short or long memory.

The emphasis in this paper is on providing a thorough analysis of IRW analysis in univariate time series with strong persistence. We also give empirical examples of the univariate approaches with an investigation of the persistence of real exchange rate series for a number of countries.

The rest of this paper is structured as follows; section 2 outlines the assumptions and basic set up of the models and presents the estimation methods and some basic results. Then, section 3 describes the bootstrap procedures and their theoretical properties; while sections 4 and 5 describe the various Monte Carlo results. Section 6 presents the empirical results. There is also a conclusions section 7, followed by a set of appendices.

2 The Theoretical Foundations

2.1 Model and Assumptions

This paper considers univariate stochastic processes of the form

$$y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, t = 1, ..., T \tag{1}$$

where ϵ_t is an unobserved error term with finite variance σ^2, and ψ_j is a sequence of constants. It is assumed throughout the paper, that $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, so that y_t is a second order stationary process whose spectral density is given by $f_y(\omega) = \frac{\sigma^2}{2\pi}\psi(e^{i\omega})\psi(e^{-i\omega})$, where $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$. The following preliminary assumptions are made concerning the error term and the Wold decomposition coefficients or IRWs, given by ψ_j:

Assumption 1 *is in two parts: (i) ϵ_t is an ergodic martingale difference sequence, so that $E(\epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, ...) = 0$, $E(\epsilon_t^2|\epsilon_{t-1}, \epsilon_{t-2}, ...) = \sigma^2$ and $E(\epsilon_t^3|\epsilon_{t-1}, \epsilon_{t-2}, ...) = \mu_3$ where μ_3 is a finite constant; and also (ii) $E(\epsilon_t^4) < \infty$.*

Assumption 2 *$\psi(z) = \tilde{\psi}(z)/(1-z)^d$, where $\tilde{\psi}(z) = \sum_{j=0}^{\infty} \tilde{\psi}_j z^j$, $\sum_{j=0}^{\infty} |\tilde{\psi}_j| < \infty$ and $d < 0.5$. Also $\psi(z)^{-1} = \sum_{j=0}^{\infty} \kappa_j z_j$ exists.*

Hence the class of above processes is very wide and includes all linear processes considered in the existing literature, and encompasses long memory processes including the leading case of $ARFIMA(p, d, q)$, where $\tilde{\psi}(z) = \phi(z)^{-1}\varphi_1(z)$ and $\phi(z) = \sum_{j=0}^{p} \phi_j z^j$ and $\varphi(z) = \sum_{j=0}^{q} \varphi_j z^j$ and d is the long memory parameter. For the purposes of analyzing both parametric and semi parametric bootstrapped inference on $IRWs$, it is necessary to introduce a parametric representation associated with the above setup, that is more general and encompasses $ARFIMA$ processes. The ψ_j are then allowed to be functions of a finite s dimensional parameter vector, θ, which is defined in a compact subset of $\mathbb{R}^s$, denoted by Θ, and has a nonempty interior. These functions are denoted by $\psi_{j,\theta}$ and the notation $\psi_{j,\theta}$ specifically indicates that subsequent analysis is parametric. The notation ψ_j is used for both the general discussion and also for the semi-parametric setting. The following identifiability assumption is required for the parametric setting.

Assumption 3 *(i) If $\psi_j = \psi_{j,\theta}$ then there exists a unique value of θ, denoted θ_0 such that $y_t = \sum_{j=0}^{\infty} \psi_{j,\theta_0} \epsilon_{t-j}$. Furthermore, $\psi_{\theta_0}(z) \neq \psi_\theta(z)$ for any z and for any θ different to θ_0, where $\psi_\theta(z) = \sum_{j=0}^{\infty} \psi_{j,\theta} z^j$. (ii) $\tilde{\psi}_{j,\theta}$ are twice continuously differentiable, with respect to θ, where $\psi_\theta(z) = \tilde{\psi}_\theta(z)/(1-z)^d$*

2.2 Parametric Estimation of Impulse Response Weights

The purpose of our analysis is to estimate ψ_j for $j = 1, ..., h$, for some finite horizon h, and to conduct inference on the estimated ψ_j, with particular attention to the issue of construction of confidence intervals for the estimated $IRWs$. A standard method is to derive the asymptotic approximation of the distribution of the estimators of ψ_j. The most commonly used approach is to use the parametric estimator given by $\psi_{j,\hat{\theta}}$, where $\hat{\theta}^W$ is the MLE of θ. This paper focuses on the Quasi Maximum Likelihood Estimator, ($QMLE$), which has been previously analyzed in a very general context by Hosoya (1997), and who has elegantly characterized their properties in the frequency domain. In particular, $\hat{\theta}^W$ is defined by $S_T(\hat{\theta}^W) = 0$ where $S_T(\hat{\theta}^W) = (S_{T1}(\hat{\theta}^W), ..., S_{Ts}(\hat{\theta}^W))'$, $S_{Tj}(\theta) = H_j(\theta) + \int_{-\pi}^{\pi} tr\left(h_j(\omega, \theta) I\left(\omega, \theta\right)\right) d\omega$, $H_j(\theta) = \frac{\partial\left(\int_{-\pi}^{\pi} \log\det f_y(\omega,\theta)\right)}{\partial\theta_j}$, $h_j(\theta) = \frac{\partial f_y^{-1}(\omega)}{\partial\theta_j}$, $j = 1, ..., s$, $I\left(\omega, \theta\right)$ is the periodogram for $y_1, ..., y_T$ and $f_y(\omega, \theta)$ is the spectral density, which given the parametric setting, is a function of θ as well as ω. As shown by Robinson (2006), the $QMLE$ is also asymptotically equivalent to an estimator obtained by minimizing the conditional sum of squares,

$$\hat{\theta} = \text{argmax}_{\theta\in\Theta} \sum_{t-1}^{T} \epsilon_t^2(\theta), \qquad \epsilon_t(\theta) = \sum_{j=0}^{t-1} \kappa_{j,\theta} y_{t-j}. \tag{2}$$

However, strictly speaking these estimators have been shown to be equivalent under slightly more restrictive assumptions than those made in Hosoya (1997). For the sake of precision it is desirable to obtain separate results for both estimators. These results relate to the asymptotic distributions of $\psi_{j,\hat{\theta}}$ and $\psi_{j,\hat{\theta}^W}$ and are given in Theorems 1 and 2 of Appendix A. These results provide an operational way, sometimes referred to as the "delta method" for constructing asymptotically valid standard errors for $\psi_{j,\hat{\theta}}$ and $\psi_{j,\hat{\theta}^W}$. However, it is well known that especially in small samples, this method can deliver poor quality approximations for estimated $IRWs$;for example see Kilian (1998a). This is essentially the motivation for the extensive use of the bootstrap for IRWs from such processes and hence is a main focus of this paper.

2.3 Semi Parametric Estimation of Impulse Response Weights

An alternative to the above parametric analysis for estimating and conducting inference on IRWs is to use semi-parametric methods. We consider two alternative approaches in this section.

2.3.1 Inversion of Autoregressive Approximations for Estimation of Impulse Response Weights

This paper suggests an entirely different approach which has a clear semi parametric interpretation. The approach is based on implicitly ignoring the presence of strong dependency in the series and to simply estimate a high order $AR(p_T)$ model. To make more concrete, it should be noted that the $ARFIMA(p, d, q)$ model can be represented by the infinite autoregressive expansion of the form

$$y_t = \sum_{j=1}^{\infty} \kappa_j y_{t-j} + v_t \tag{3}$$

A possible method is to directly estimate by OLS the truncated autoregressive, $AR(p_T)$, expansion

$$y_t = \sum_{j=1}^{p_T} \kappa_j^{(p_T)} y_{t-j} + v_t^{(p_T)} \tag{4}$$

where the order p_T, is obtained by some information criterion. This approach has been theoretically analyzed by Poskitt (2007). The least squares estimates of $\kappa_j^{(p_T)}$ obtained by fitting an $AR(p_T)$ model to the data, are denoted by by $\hat{\kappa}_j^{(p_T)}$. Theorem 5 of Poskitt (2007) states that $\sum_{j=1}^{p_T} \left|\hat{\kappa}_j^{(p_T)} - \kappa_j^{(p_T)}\right|^2 = o_p(1)$ for all p_T such that $p_T \to \infty$ and $p_T = o(T^{\alpha})$ for all $\alpha > 0$. For example, an acceptable sequence for p_T is $(lnT)^{\alpha}$ for some $\alpha > 1$. Further, by the extension of Baxter's inequality proven in Theorem 4.1 of Inoue and Kasahara (2006) it follows that

$$\sum_{j=1}^{p_T} \left|\kappa_j^{(p_T)} - \kappa_j\right| = o(1), \tag{5}$$

as long as $p_T \to \infty$. Then, overall,

$$\sum_{j=1}^{p_T} \left|\hat{\kappa}_j^{(p_T)} - \kappa_j\right|^2 = o_p(1) \tag{6}$$

which implies that the $IRWs$ can be consistently estimated by fitting an approximating AR model to the time series realization. In particular, the $IRWs$ are given by $\hat{\psi}(z) = \sum_{j=1}^{\infty} \hat{\psi}_j z^j = \hat{\kappa}^{-1}(z)$, where $\hat{\kappa}(z) = \hat{\kappa}^{(p_T)}(z) = \sum_{j=1}^{p_T} \hat{\kappa}_j^{(p_T)} z^j$. For subsequent analysis it is convenient to also define $\psi^{(p_T)}(z) = \left(\kappa_j^{(p_T)}(z)\right)^{-1}$. A critical issue is the choice of p_T, and it has been shown by Poskitt (2007), via his Theorem 9, that selecting p_T by information criteria such as the AIC or BIC is asymptotically efficient in the sense of Shibata (1980). In the Monte Carlo study in this paper the value of p_T is fixed at $(\ln T)^2$, which is a valid approximation for finite order $ARFIMA$ processes and even for infinite AR representations. In terms of inference for the $IRWs$,a data dependent method for selecting p_T is also considered in the Monte Carlo study. Consequently, the AR approximation has an interpretation of being a semi-parametric model.

Another important point is concerned with the Theorem 10 of Poskitt (2007), which shows that the asymptotic distribution of $\hat{\kappa}_j^{(p_T)}$ is nonstandard and non Gaussian, which is clearly quite different to the theory relating to weakly dependent processes as described by Lewis and Reinsel (1985). Hence inference based on estimated IRWs obtained from the AR approximations will be problematic. Again, this is another strong motivation to base inference in this semi- parametric setting on the bootstrap. This is the approach adopted later in this paper. Finally, we note that the use of the bootstrap will not only provide valid inference but can be used to correct the bias in the estimates of the IRWs that we observe in finite samples in our Monte Carlo study. This bias correction will be discussed in Section 5.

2.3.2 Semi Parametric Local Whittle Estimation of Impulse Response Weights

An alternative, intuitively interesting approach which does not seem to have been previously implemented in the literature, is to estimate the long memory parameter from the data using a semi parametric estimator, such as the local Whittle (LW) estimator and to then fit a parametric model to a fractionally filtered, or fractionally differenced series. This approach is referred to as the LW two step estimator ($LWTSE$) of the IRWs. The LW estimator of d, is denoted by $\hat{d}_{LW}$ and is obtained by minimizing the objective function $\ln\left[\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}I(\omega_j)\right]-\frac{2d}{m}\sum_{j=1}^{m}\ln(\omega_j)$, with respect to d, where $I(\omega_j)$ is the periodogram given by $I(\omega_j)=\frac{1}{2\pi T}\left|\sum_{j=1}^{T}y_te^{i\omega_jt}\right|^2$, and m is the bandwidth. For the LW estimator of d, it is known that, for linear processes, $m^{1/2}\left(\hat{d}_{LW}-d_0\right)\rightarrow N\{0,1/4\}$ where d_0 denotes the true value of d. It is important to note that $m\leq T^{4/5}$, and m is generally chosen in the range of $T^{1/2}\leq m\leq T^{4/5}$. In the usual case of ignorance of the short run dynamics, the bandwidth is generally selected in an ad hoc way and a popular choice is $m=T^{0.5}$. For a discussion of this issue, see also Henry (2001).

A further method proposed by Andrews and Sun (2004) is the Local Polynomial Whittle, or LPW method which approximates the logarithm of the spectral density of the short memory component by a polynomial. This leads to an estimator of d which has a reduced asymptotic bias, but higher variance. All the simulations involving $\hat{d}_{LPW}$, in this paper, use the first order approximation as in Nielsen and Frederiksen (2004).

In terms of the estimation of IRWs, if the parameter d is known, then the observed y_t series can be fractionally filtered to obtain $u_t=y_t-\sum_{l=1}^{t-p}\pi_l(d)y_{t-l}$ where $(1-L)^dy_t=y_t-\sum_{l=1}^{\infty}\pi_l(d)y_{t-l}$, and $\pi_l(d)$ are the coefficients of the infinite AR representation of y_t in terms of u_t, so that $\pi_l(d)=\Gamma(l-d)\Gamma(-d)^{-1}\Gamma(l+1)$. In practice, d is unknown and can be replaced by the LW estimate, $\hat{d}_{LW}$. Then, the feasible fractionally filtered series based on observable quantities is $\hat{u}_t=y_t-\sum_{l=1}^{t-p}\hat{\pi}_l(\hat{d}_{LW})y_{t-l}$, where $\hat{\pi}_l(\hat{d}_{LW})=\Gamma(l-\hat{d}_{LW})\Gamma(-\hat{d}_{LW})^{-1}\Gamma(l+1)$. For concreteness, this paper focuses on the estimation of the widely used univariate $ARFIMA(p,d,q)$ process. Extensions to models with more complicated short run dynamics are quite manageable. The complete parameter vector is denoted by $\theta=(d,\beta)'$, where the $(p+q)$ $ARMA$ parameters are in the vector $\beta=(\phi_1,..\phi_p,\vartheta_1,...,\vartheta_q)'$. The true parameter values are denoted as $\beta_0(d_0)$, and the LW two step estimator ($LWTSE$) of β, based on the feasible fractionally filtered series are $\hat{\beta}_{LWTSE}(\hat{d}_{LW})$. Then the $ARMA(p,q)$ parameters of the original $ARFIMA(p,d,q)$ process are estimated by minimizing the conditional sum of squares, CSS, conditional on $\hat{d}_{LW}$. The following result provides consistency and a rate of convergence for the two step estimator of the $ARFIMA(p,d,q)$ model.

Theorem 1 *Let y_t be given by an $ARFIMA(p, d, q)$ process, where $\phi(L)$ and $\theta(L)$ are AR and MA polynomials in the lag operator of orders p and q respectively, with all their roots lying outside the unit circle. Let the disturbance ϵ_t be i.i.d.$(0, \sigma^2)$, with $E(\epsilon_t^4) < \infty$. Then,* $\hat{\beta}_{LWTSE}(\hat{d}_{LW}) - \beta_0(d_0) = O_p(m^{-1/2})$.

The above theorem is proven in Appendix C; and it appears that the only previous work investigating the issue of using a semi parametric estimator of d in a two stage analysis is by Wright (1995). Once the parameters of the $ARFIMA$ model have been obtained, it is then straightforward to estimate the IRWs. While the $LWTSE$ approach is semi parametric in the sense that d is estimated semi parametrically, the second step is fully parametric and there does not seem to be any previous literature on how this parametric assumption can be relaxed.

3 Bootstrap Inference

The motivation for using the bootstrap seems very compelling given existing evidence on the poor quality of asymptotic approximations for constructing confidence intervals for $IRWs$ in small samples for weakly dependent processes; see for example Kilian (1998a) and Kilian (1998b). Furthermore, it is clear that unlike semi-parametric autoregressive approximations for weakly dependent processes, such approximations for long memory and the alternative $IRWs$ estimator based on $LWTSE$ are not easily amenable to asymptotic inference, since the relevant distributions are either non Gaussian or unknown. Hence the bootstrap appears to be an attractive alternative approach .

There has been a rapidly increasing literature on the application of the bootstrap to long memory processes; for example, see Poskitt (2008). Andrews and Lieberman (2006) provide results both on the validity of the bootstrap and its ability to provide higher order corrections compared to asymptotic approximations. However, this work assumes Gaussianity and Andrews and Lieberman (2006) conjecture that higher order corrections will not be valid for such processes. The results in this paper prove the validity of the parametric bootstrap for non Gaussian processes for both the parametric estimators introduced in the previous section. This material uses the foundations provided by Hosoya (1997), who has established the validity of MLE for non Gaussian long memory processes. The main contribution of this section of our paper is to provide justification for a semi parametric bootstrap, which can be used for inference on estimated IRWs in either the context of a parametric, or a semi parametric model. The work of Poskitt (2007) is important for these derivations.

It is now convenient to consider the parametric bootstrap for the model given by (1) where $\psi_j = \psi_{j,\theta}$. From assumption 2, it is known that y_t has an infinite AR approximation, which is given by $y_t = \sum_{j=1}^{\infty} \kappa_{j,\theta} y_{t-j} + \epsilon_t$. After estimating θ using one of the methods discussed in the previous section, the residuals can be obtained as $\hat{\epsilon}_t = y_t - \sum_{j=1}^{t-1} \kappa_{j,\hat{\theta}} y_{t-j}$. For the parametric bootstrap, these residuals are then re-centered and re-sampled with replacement, to obtain a vector of bootstrap error terms denoted by $(\epsilon_1^*, ..., \epsilon_T^*)'$. These bootstrap errors can then be used together with the estimated AR coefficients to give the bootstrap sample $(y_1^*, ..., y_T^*)'$. It is important to note that initial conditions are required, and that these are usually set to the estimated unconditional mean of the data. The bootstrap sample can then be used to estimate either by MLE or by the minimization of the conditional sum of squares; and hence obtain bootstrapped estimates $\hat{\theta}^{W*}$ and $\hat{\theta}^*$ respectively. These estimates can then be used to obtain the corresponding bootstrapped estimates of the $IRWs$, denoted by $\psi^*_{j,\hat{\theta}^W}$ and $\psi^*_{j,\hat{\theta}}$. This procedure is replicated B times to generate estimates of the $IRWs$ and their empirical distribution as $B \to \infty$, which can be used for inference on the estimated $IRWs$. On denoting P_y as the probability law of a random vector y and $\mathfrak{m}(y_1, y_2)$ as the Mallows metric between P_{y_1} and P_{y_2}, it is then possible to derive the following theorems concerning the validity of this form of parametric bootstrap for both MLE and the minimization of CSS.

Theorem 2 *Let Assumptions 1-3 and 4-6, of Appendix B, hold. Then, for all* $j = 1, ..., h$

$$\mathfrak{m}\left(\sqrt{T}\left(\psi_{j,\hat{\theta}^W} - \psi_{j,\theta_0}\right), \sqrt{T}\left(\psi_{j,\hat{\theta}^{W*}} - \psi_{j,\hat{\theta}^W}\right)\right) = o_p(1) \tag{7}$$

Theorem 3 *Under assumptions 1(i) and 2-3; and further assuming that (i)* ϵ_t *is an i.i.d. sequence, (ii)* $\sum_{j=1}^{\infty} \sup_\theta |\tilde{\psi}_{j,\theta}| < \infty$, *(iii)* $\tilde{\psi}_{j,\theta}$ *are twice continuously differentiable, with respect to* θ, *and (iv)* Ω, *defined in (14) of Appendix A, is nonsingular. Then, for all* $j = 1, ..., h$

$$\mathfrak{m}\left(\sqrt{T}\left(\psi_{j,\hat{\theta}} - \psi_{j,\theta_0}\right), \sqrt{T}\left(\psi_{j,\hat{\theta}^*} - \psi_{j,\hat{\theta}}\right)\right) = o_p(1) \tag{8}$$

Both Theorems are proven in Appendix D. It is now appropriate to discuss a semi parametric sieve type bootstrap, which can be implemented using the following strategy:

1. Estimate an $AR(p_T)$ model on y_t and obtain the estimated coefficients, $\hat{\kappa}_j^{(p_T)}$, $j = 1, ..., p_T$ and the residuals, $\hat{\epsilon}_t = y_t - \sum_{j=1}^{min(p_T, t-1)} \hat{\kappa}_{j,\hat{\theta}} y_{t-j}$.
2. Invert $\hat{\kappa}^{(p_T)}(z) = \sum_{j=1}^{p_T} \hat{\kappa}_j^{(p_T)} z^j$ to obtain estimates of the IRWs given by $\hat{\psi}_j^{(p_T)}$, $j = 1, ..., h$.
3. Re-center $(\hat{\epsilon}_1, ..., \hat{\epsilon}_T)'$
4. Re-sample with replacement from this vector, to obtain the bootstrap sample of error terms given by $(\epsilon_1^*, ..., \epsilon_T^*)'$.
5. Use the above quantities together with $\hat{\kappa}_j^{(p_T)}$, $j = 1, ..., p_T$, to generate the bootstrap sample $(y_1^*, ..., y_T^*)'$.
6. Estimate an $AR(p_T)$ to $(y_1^*, ..., y_T^*)'$ to obtain the bootstrap estimated autoregressive coefficients given $\hat{\kappa}_j^{*,(p_T)}$, $j = 1, ..., p_T$;
7. Invert $\hat{\kappa}^{*,(p_T)}(z) = \sum_{j=1}^{p_T} \hat{\kappa}_j^{*,(p_T)} z^j$ to obtain bootstrap estimates of the impulse responses given by $\hat{\psi}_j^{*,(p_T)}$, $j = 1, ..., h$.

8. Repeat the above algorithm B times and then use the resulting estimates of the IRWs to construct an empirical distribution of the IRWs.

The following theorem justifies the above bootstrap approach, and is proven in Appendix E.

Theorem 4 *Let Assumptions 1-2 hold. Let $p_T = o\left((\ln T)^a\right)$ for some $a > 0$. Then, for all $j = 1, ..., h$,*

$$\mathfrak{m}\left(r_T\left(\hat{\psi}_j^{(p_T)} - \psi_j^{(p_T)}\right), r_T\left(\hat{\psi}_j^{*,(p_T)} - \hat{\psi}_j^{(p_T)}\right)\right) = o_p(1) \tag{9}$$

where $r_T = p_T^{-3/2}\left(\frac{T}{\ln(T)}\right)^{1/2-d}$.

This theorem does not follow directly from the work of Poskitt (2008), since the statistic being bootstrapped is a function of a statistic that grows with the sample size, rather than being fixed. It is worth briefly commenting on this bootstrap. It can be classified as an 'other percentile' bootstrap in the taxonomy of Hall (1992). Further, the statistics on which it is based do not have the desirable pivotalness property that can also lead to higher order asymptotic refinements. However, in this respect we note the important contribution of Kilian (1999) who notes that studentising IRWs, to induce asymptotic pivotalness, can be counterproductive, and lead to worse finite sample performances. An extension of this bootstrap approach that compensates for the small sample bias involved in the autoregressive estimation in Step 1 of the above algorithm, along the lines of Kilian (1998a), can be straightforwardly envisaged. We consider this extension in our Monte Carlo study.

An alternative sieve bootstrap is obtained by generating the data as above, but with the parameter vector θ, being bootstrapped and used to generate the $IRWs$. The validity of this bootstrap follows immediately from Theorem 2 and the discussion of Assumption 4 of Poskitt (2008). This argument also clearly applies to the IRWs obtained via $LWTSE$. The only difference here is that d is estimated semi parametrically rather than parametrically within an $ARFIMA$ model. A final point worth mentioning here is that this method need not be restricted to univariate models. In many applied situations, it is often desirable to consider vector processes. Appendix F extends the above estimation and inferential methodology to IRWs from $VAR(p_T)$ models. These results show that there is no difficulty in extending the methodology to the vector case.

4 Monte Carlo Analysis of Estimated IRWs

This section reports the results of the previously described Monte Carlo study of the estimation of the IRWs. Given the $ARFIMA(p, d, q)$ process in equation (3) the implied $IRWs$, denoted by ψ_k for $k = 1, , 2, ...$ are generated from $\psi(L) = \vartheta(L)(1-L)^{-d}\phi(L)^{-1}$, where $\psi(L) = \sum_{k=1}^{\infty}\psi_k L^k$.

The estimated $IRWs$ are obtained by replacing the true theoretical parameters with their corresponding estimates. For large lag k, these Wold decomposition coefficients decay at the approximate rate of $\psi_k \sim c_1 k^{d-1}$. However, the presence of a relatively persistent $AR(1)$ component process can considerably alter the appearance of the $IRWs$ for short to moderate impulse response horizons.

Figures 1 through 2 report some of the results for different $IRWs$ for horizons $k = 1, 2, ...40$ for $ARFIMA(1, d, 0)$ models; and for designs of $d = 0.4, 0.8$ and $\phi = 0.95$.[1] Although the previous theoretical analysis has focused on stationary processes, a considerable amount of applied econometric work has found estimates of d in the range of $0.5 < d < 1$, which implies a non stationary process, but with finite cumulative IRWs. Hence, it seems very important to extend the Monte Carlo analysis to consider some mildly non stationary long memory processes.

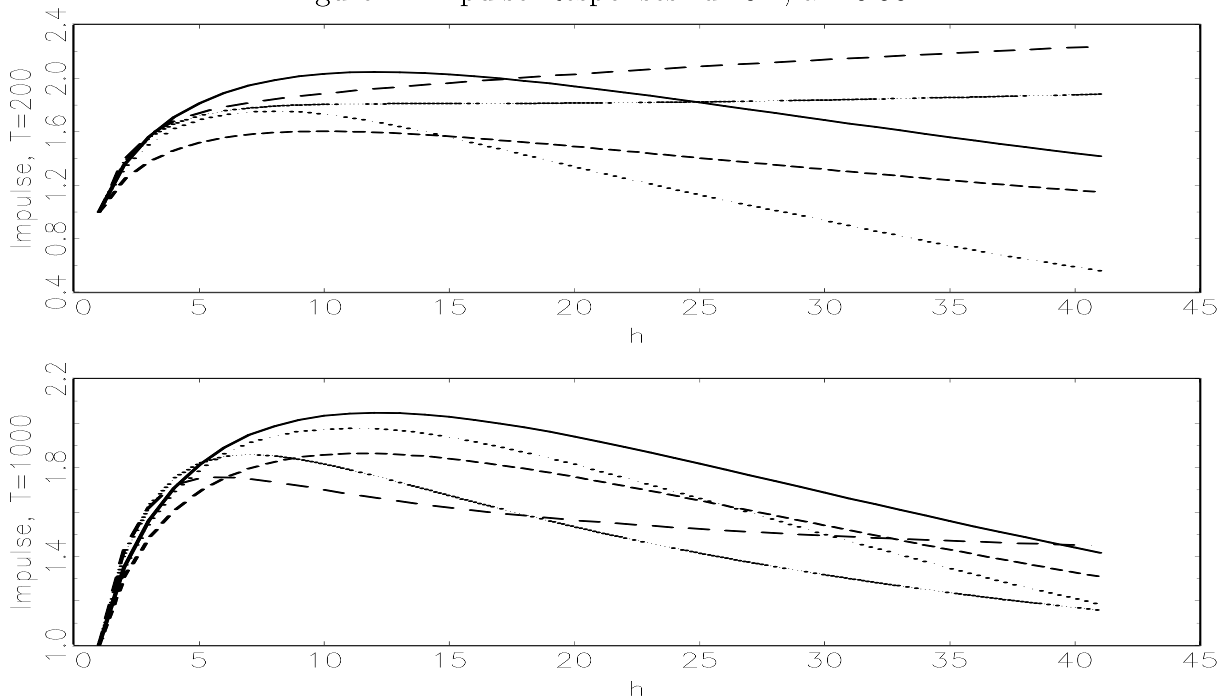
The $IRWs$ are all estimated from the three different methods of (i) AR approximations, (ii) MLE and (iii) $LWTSE$. The $LWTSE$ method is based on using the LPW, rather than LW, for the initial estimation of d in the stationary cases. The estimated $IRWs$ from using the LW and LPW methods are constructed using a bandwidth of $m = T^{0.5}$.[2] For a model with $d = 0.4$ and quite persistent short memory, Figure 1 indicates that $IRWs$ estimated from the $LWTSE$ approach perform poorly in comparison with corresponding estimates from MLE. The $IRWs$ estimated from MLE with d in the stationary region dominate alternative methods; however MLE estimated $IRWs$ are poor for $d = 0.8$ when there is persistent autocorrelation of $\phi = 0.95$. In this case the $AR(p_T)$ approximation performs surprisingly well and is the preferred method.

For the large sample size of $T = 1,000$ and for designs of ($d = 0.6$, $\phi = 0.5$) and ($d = 0.8$, $\phi = 0.5$),which are not reported to save space, the MLE performs extremely well, with the high order AR approximation generally being slightly superior to the $LWTSE$. For the design of ($d = 0.8$, $\phi = 0.95$) in figure 2, the high order AR approximation performs outstandingly well, with the MLE a poor third compared with the $LWTSE$. Hence, there seems some evidence that MLE works well for non stationary long memory processes provided that there is only moderate degree of persistence in the short run dynamics. However, when a non stationary long memory process has a very persistent short run component, the high order AR approximation method is extraordinarily accurate compared with MLE and the $LWTSE$. The excellent performance of the high order $AR(p_T)$ method strongly suggests that it should be the main analytic tool if an investigator is principally interested in assessing the impact of shocks or innovations on a series. This recommendation is also reinforced by the fact that in practice an investigator will be unaware of whether the data generating process is I(0), or stationary long memory, or non-stationary long memory. Also, the results suggest that the application of the LW should probably be reserved for only obtaining an estimate of the long memory parameter. Hence the

[1] Results were also obtained for the cases of $d = 0.2$ and $d = 0.6$. They are qualitatively very similar to the results presented and are omitted for reasons of conserving space. They are available from the authors on request.

[2] Results for the cases of $\phi = 0.5$ and $\phi = 0.8$ are omitted for reasons of conserving space, and are available from the authors on request. Similarly results based on optimal bandwidth given knowledge of the true data generating process are also omitted since they are broadly similar to reported results in this paper.

Figure 1: Impulse Responses: d=0.4, ar=0.95



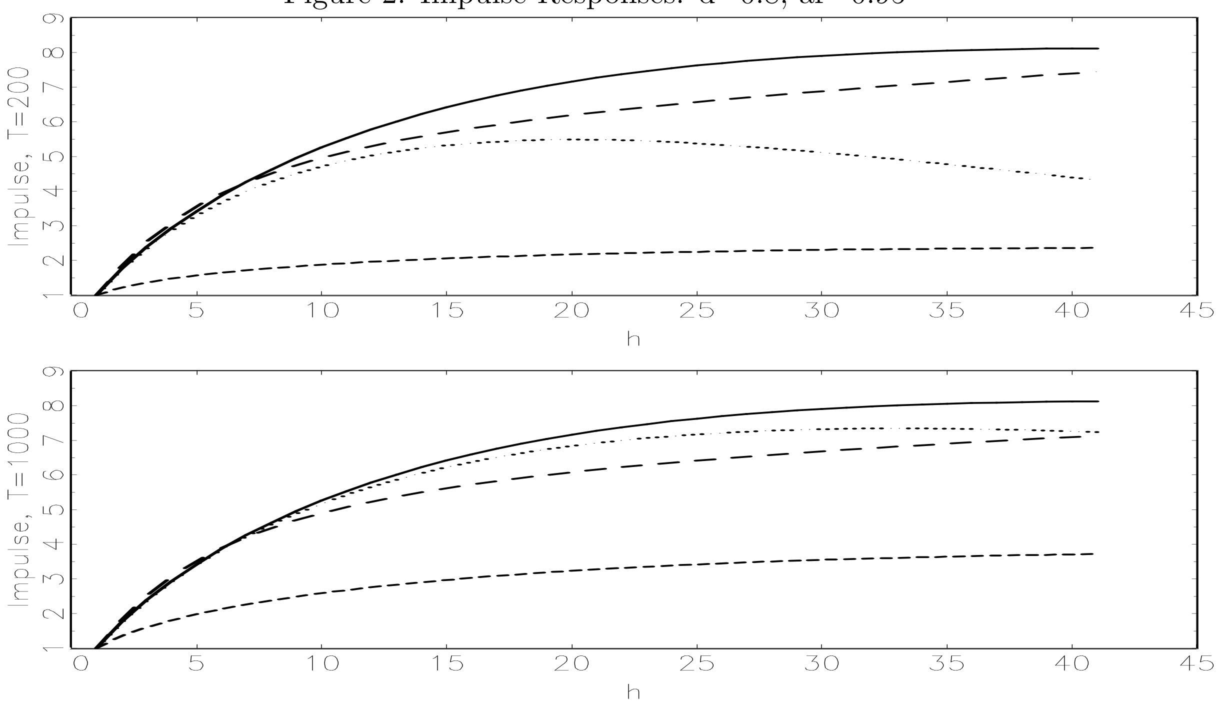
Key: Solid Line (—–) represents the true IRW; Long Dashed Line (– – –) represents the Two-Step LW; Dotted Line (. . .) represents the AR Approximation; Short Dashed Line (- - -) represents the MLE; Dense Dotted Line (...) represents the Two-Step LPW.

estimates of $IRWs$ based on $LWTSE$ are omitted for space constraints and are not considered for the Monte Carlo study of confidence interval estimators reported in the next section.

5 Monte Carlo Analysis for Confidence Intervals for Estimated IRWs

This section investigates the small sample properties of some of the methods analyzed in the previous sections for constructing confidence intervals for $IRWs$. The focus is on data generating process which have simple parametric models and it is assumed that the parametric methods for constructing the confidence intervals use the correct specification of the process. This is of course, disadvantageous to the semi-parametric method used to construct confidence intervals. However, our results reported below, still give quite clear indications as to the superiority of the various methods. It was decided to focus on various $ARFIMA(1, d, 0)$ models as the benchmark. Previous work by Baillie and Kapetanios (2007), Baillie and Kapetanios (2008) and Nielsen and Frederiksen (2004) has suggested that the most important reason for problematic inference in small samples for a variety of long memory models, hinges on the presence of persistent short memory components. This is intuitively very reasonable since such persistent stationary

Figure 2: Impulse Responses: d=0.8, ar=0.95



Key: Solid Line (——) represents the true IRW; Long Dashed Line (– – –) represents the Two-Step LW; Dotted Line (. . .) represents the AR Approximation; Short Dashed Line (- - -) represents the MLE.

components can be mistaken for long memory. Hence this study considers a parsimonious short memory $AR(1)$ structure, which gives an overall $ARFIMA(1,d,0)$ model.

For the Monte Carlo experiment, realizations of $ARFIMA(1,d,0)$ processes were generated for three different sample sizes of $T = 200$, $T = 400$ and $T = 1,000$; and for three simulation designs of the AR coefficient, ϕ, and long memory parameter, d. The designs were $(\phi, d) = (0.50, 0.2), (0.95, 0.2), (0.95, 0.4)$; with $\epsilon_t \sim NID(0,1)$. This is our baseline Monte Carlo setup. However, in practice an investigator would have no knowledge as to whether or not a series has long memory. Hence this study considers the performance of the various approaches in the presence of stationary, but very persistent processes, where the data generating process is $AR(1)$ with coefficients of $\phi = 0.9, 0.95, 0.98, 0.99$. The following four different approaches were used to construct confidence intervals (CI) for the estimated IRWs:

- *Approach 1:* CI obtained by using the asymptotic normal approximation to the finite sample distribution of the IRWs assuming the correct parametric $ARFIMA(1,d,0)$ model. This method is theoretically justified by standard theoretical results on the consistency and asymptotic normality of the parameter estimates of the $ARFIMA(1,d,0)$ model. 4.

- *Approach 2:* CI obtained from bootstrap IRW obtained by fitting $AR(p_T)$ models to samples generated by the sieve bootstrap where $p_T = (\ln T)^2$. This method is theoretically justified in Theorem 4.

- *Approach 3:* CI obtained from bootstrap IRW obtained by fitting $AR(p_T)$ models to samples generated by the sieve bootstrap where $p_T = (\ln T)^2$. The AR parameters are bias corrected using the method that is presented and theoretically justified in Kilian (1998a). The algorithm used for the bias correction is given in Section B of Kilian (1998a). 4.

- *Approach 4:* CI obtained from bootstrap IRW obtained by fitting $AR(p_T)$ models to samples generated by the sieve bootstrap where p_T is obtained by using the Akaike Information Criterion (AIC). The AR parameters are bias corrected using the method that is presented and theoretically justified in Kilian (1998a). The algorithm used for the bias correction is given in Section B of Kilian (1998a).

It is clear that many alternative methods could be used and we spend some time discussing some preliminary results that motivate the selection of methods whose performance is presented in detail. Approach 1 is a useful benchmark and its computational tractability suggests that it should be considered whenever a parametric assumption is waranteed. However, it is clear that one should avoid making parametric assumptions on the form of the short component of the series, if possible. We have considered the question of misspecification by generating data from an $ARFIMA(1, d, 1)$ and fitting an $ARFIMA(1, d, 0)$ model. We find that the coverage rates of the asymptotic approximation are very far away from their nominal level as expected. This is simply a confirmation that parametric assumptions can be extremely problematic. Nevertheless, it is useful to compare our semiparametric methods to a parametric benchmark to gauge the efficiency loss of the semiparametric approach when the model is correct. Moving on to potential bootstrap schemes that have been discussed in Section 3 we note the following. We can consider two bootstrap variants that have a parametric component. The first is fully parametric both generating the data from an $ARFIMA$ model and estimating such a model on both original and bootstrap data. This suffers from two problems. The first is the obvious one relating to the parametric assumption that is discussed above. The second relates to the interesting fact, that was observed in preliminary Monte Carlo work, that parametric estimation of $ARFIMA$ models produces parameter estimates which when used to generate bootstrap samples result in coverage rates that again considerably deviate from their nominal level. So this parametric scheme is both conditional on using the correct model and even then does not seem to work well. The second bootstrap which is partly parametric is the same as the above parametric bootstrap but uses a sieve to generate the bootstrap samples. We find this has much better coverage rates that are usually very close to their nominal values when the model is correct. But, under misspecification, the coverage rates again deviate considerably from their nominal level. For this reason, and because of the fact that its hybrid parametric/semiparametric nature is not that intuitive, we do not consider it in further detail.[3] As a result we focus on two

[3] It is worth discussing briefly a related question concerning these parametric bootstraps. This question relates

semiparametric bootstrap schemes: Approach 2 simply uses a sieve to generate the data and then fits autoregressive models with a lag order that increases as a deterministic function of sample size. This seems to work very well but it is clear a persistent short memory component results in performance deterioration as one would expect. Therefore, it is productive to consider schemes that alleviate this effect. The one proposed by Kilian (1998a) is the most popular of such schemes. We consider it and find that it performs extremely well. There are other corrections that could be considered. Of those two stand out. The first is the method of Wright (2000) which assumes a near unit root process. Initial experimentation suggests that while it works reasonably well it is considerably inferior to the Kilian (1998a) correction. This and the fact that its near unit root motivation bears little relation to the current long memory framework leads us to disregard this method. Finally, the approach of Sims and Zha (1999) is another approach that is worth considering in this respect. However, that alternative has a clear Bayesian motivation and the paper that proposed it does not clarify a number of important aspects needed for its automatic application. As a result we view it as less attractive operationally, compared to the work of Kilian (1998a). Finally, we consider Approach 4 as a way to investigate how a data dependent lag selection approach for the best performing method compares to the deterministic lag selection rule we use as a baseline case.

Figure 3 reports the coverage rates for the above methods using 2,000 replications with 599 bootstrap replications being used for each Monte Carlo replication and a nominal level of 90%. This set of results covers the case when the data generating process has long memory. The benchmark Approach 1 is found to generally perform quite well. This is reasonable since it is the case when the correct specification of the model is known and used. Hence this is a benchmark for the other approaches. In general, Approach 2 works very well but has a deteriorating performance when persistence rises either by increasing d or ϕ. Moving on to the bias corrected Approach 3 we see that it works extremely well for all sample sizes and all levels of persistence. It outperforms both Approaches 1 and 2. In the case of Approach 1, this is extremely interesting as Approach 3 is semiparametric yet performs better than the parametric method even though the correct model is assumed to be used. Allowing for a data dependent lag selection method, as Approach 4 does, results in performance deterioration It is clear that the Akaike information criterion chooses too few lags in smaller samples. As a result we recommend the use of the deterministic lag section rule. Figure 4 reports the average width of the confidence intervals for the above Monte Carlo experiment. It is clear that the conclusions from this figure correspond closely with those reached by analysing the results in Figure 3.

For the AR experiments, for which results are presented in Figure 5, the approaches work better in general. As expected Approach 3 works extremely well here too for all samples and all levels of persistence. This is not surprising given this approach was designed for such data

to whether studentising IRWs produces better performing parametric bootstrap schemes or not. As we noted earlier, Kilian (1999) clearly illustrates that studentising is counterproductive for some short memory processes and our preliminary Monte Carlo study confirms that this is also the case for the long memory processes we consider.

generating processes.

Overall, Approach 3 seems to be a robust and very useful method for constructing CI for IRW. It seems that, for relatively large sample sizes, lag selection is not that helpful and the use of a large lag order provides robustness to model misspecification without large costs in terms of performance due to over parametrisation.

Finally, we note that a very valuable byproduct of the above analysis, related to the bias correction of Kilian (1998a) is that in addition to improved confidence bands we have a bootstrap-based bias estimate for the autoregressive coefficients of the autoregressive approximation that is used in Section 2.3.1 to estimate IRWs. So we can actually get an improved point estimate of these IRWs. In a final short Monte Carlo study we evaluate this IRW estimation approach and replicate Figure 1 where we now compare the AR approximation with and without bias correction. We see that the bias corrected approach performs much better. As a result we have a unified robust and very effective estimation and inference approach for IRWs for long memory processes.

6 Empirical Application

The previous findings in this paper have indicated that a good strategy for analyzing $IRWs$ is to estimate them by semi-parametric AR approximations, and to use the sieve bootstrap for estimating confidence intervals. Furthermore this approach appears to be a very good strategy for processes with either persistence or short memory. This section provides an illustration of this approach to the analysis of two reasonably large macroeconomic quarterly data sets comprising real exchange rates.

The real exchange rate (RER) data is from 10 countries: UK, Switzerland, South Africa, Norway, New Zealand, Mexico, South Korea, Japan, Canada and Australia. Note that Euro zone countries are excluded from the RER data due to the introduction of the Euro in January 1998, and the possibility of structural breaks occurring around January 1998. The data span the period of 1957Q1 through 2009Q1; and all the data are obtained from the IMF (International Financial Statistics (IFS)). The bilateral real exchange rate q is constructed as the *i-th* currency at time t as $q_{i,t} = s_{i,t} + p_{j,t} - p_{i,t}$, where $s_{i,t}$ is the corresponding nominal exchange rate (*i-th* currency units per one unit of the *j-th* currency), $p_{j,t}$ the price level (CPI) in the *j-th* country, and $p_{i,t}$ the price level of the *i-th* country. That is, a rise in $q_{i,t}$ implies a real appreciation of the *j-th* country's currency against the *i-th* country's currency.

The IRW analysis was conducted by implementing our Approach 3. Hence an AR approximation with a lag order of $(\ln T)^2$ was used to estimate the IRWs and then a sieve bootstrap was used to correct the bias of the estimated IRW and construct 90% confidence intervals. The half lives were measured for each of the impulse responses. For the purposes of this paper, the half life is defined as $h = i,$ for which $\psi_i = \psi_0/2$ where linear interpolation is used to define ψ_i for non-integer i. Note that the usual closed form solution for h, given by $h = \frac{\ln(1/2)}{ln(\hat{\rho})}$, where ρ

Table 1: Half-Life Estimates, 5% quantiles and 95% quantiles for CPI inflation and Real Exchange Rates

Real Exchange Rates			
Country	Half Life	5% quantile	95% quantile
UK	13.464	6.533	17.506
Switzerland	26.309	10.787	111.644
South Africa	20.129	8.643	71.743
Norway	25.262	11.139	85.452
New Zealand	12.379	7.854	15.517
Mexico	9.771	7.134	26.716
South Korea	84.417	9.520	120.512
Japan	17.767	11.087	120.480
Canada	24.315	12.917	59.949
Australia	16.540	9.114	26.137

denotes the AR coefficient of an $AR(1)$ model, is only valid for $AR(1)$ models. There is no closed form solution for general $AR(p)$ models. All the empirical work uses 599 bootstrap replications.

Plots of the IRWs for the real exchange rate series are presented in Figure 7. It should be noted that the corresponding IRWs are relatively smooth and suggest that the real exchange rate series are very persistent processes. In some cases, e.g. New Zealand and UK, there is a smooth oscillatory pattern reminiscent of $AR(2)$ structures with complex roots. The increased persistence is reflected in the half life measures which range from 9.7 for Mexico to 84 for South Korea. We note the problem of non-monotonicity of some of the IRWs associated with the UK and New Zealand, where initial IRWs fall below 0.5 but are above 0.5 at longer horizons.

One interesting issue related to this oscillatory behaviour concerns the previous definition of half life, which is not fully robust. In particular, when the IRWs oscillate, rather than monotonically decline, it is possible that the IRW will fall below half their original value, only to rise again before falling back. This oscillation may in fact be repeated and in this case the definition breaks down. One definition that has been used is to define the half life as either the smallest i for which $\psi_i = 1/2\psi_0$; see for example Rossi (2005), or alternatively the largest such i; see for example Ng (2003). This study follows Rossi (2005) and uses the smallest i. Examination of the IRWs in Figure 5 suggests that in a number of cases, including the US, Switzerland and Spain, the oscillatory nature of the IRW implies that the reported half life may be misleading. It is sufficient for the purposes of this illustrative empirical work to note that the standard measure of half life may misrepresent the persistence of CPI inflation.

Another interesting feature of the analysis is that the IRW exceed unity at horizons of about 2 to 10 quarters for a majority of countries, which indicates quite extreme persistence. Overall, it seems that the new methodology proposed in this paper provide a reliable and robust method for carrying out IRW analysis. The empirical findings confirm that real exchange rates are very persistent.

7 Conclusions

This paper has considered the estimation and construction of confidence intervals for Impulse Response Weights (IRWs) from strongly persistent time series, which include fractionally integrated processes with slowly decaying hyperbolic autocorrelations. One of the main contributions of the paper in terms of estimation of the $IRWs$ is to consider a non parametric time domain estimator based on an autoregressive (AR) approximation. This estimator is shown to have surprisingly good theoretical and small sample properties. The paper has also examined the application of a procedure where the Local Whittle (LW) estimator is initially used to estimate the long memory parameter and to then subsequently estimate the short memory parameters and hence to estimate the $IRWs$. In general, Monte Carlo results indicate that this method does not work as well as the AR approximation for the estimation of the $IRWs$, especially when the AR approximation uses a bias correction.

The second part of the paper investigates the most appropriate methods for estimating the variability and the construction of confidence intervals for the estimated $IRWs$. As previously discussed there has been a long standing concern in the literature over this issue. As with weakly dependent processes confidence intervals based on the "delta method", and asymptotic approximations can prove very unreliable. This paper has considered a generic semi-parametric sieve bootstrap based on an autoregressive approximation of the unknown data generating mechanism. Under mild assumptions, we show the validity of IRW inference analysis based on the AR approximation and the validity of bootstrap inference on the resulting IRWs.

The results in the paper indicate that the sieve bootstrap has a number of advantages and that a good strategy for analyzing $IRWs$ is to estimate them by semi-parametric AR approximations, and to use the sieve bootstrap for estimating confidence intervals. Furthermore this approach appears to be a very good strategy for processes with either short or long memory. The objective in this paper has been to provide a detailed analysis of the IRW analysis in univariate time series with strong persistence. The application to real exchange rate series indicates that the prescribed methodology is reasonably easy to implement in practice and gives intuitively reasonable results. The extension of the methodology to the multivariate case with the use of high order VAR approximations is possible. However, while such an extension is straightforward in principle, many practical issues, beyond the scope of the present paper, would need to be adequately addressed in future research.

References

Andrews, D.W.K. and Y. Sun (2004): Adaptive local polynomial Whittle estimation of long range dependence, *Econometrica*, 72, 569-614.

Andrews, D. W. K., and O. Lieberman (2006): "Higher-order improvements of the parametric bootstrap for long-memory Gaussian processes," *Journal of Econometrics*, 133, 673–

702.

Baillie, R. T. (1987): "Inference in dynamic models containing surprise variables," *Journal of Econometrics*, 35, 101–117.

Baillie, R. T., and G. Kapetanios (2007): "Testing for neglected nonlinearity in long memory models," *Journal of Business and Economic Statistics*, 25, 447–461.

Baillie, R. T., and G. Kapetanios (2007): "Nonlinear models for strongly dependent processes and financial applications, " *Journal of Econometrics*, 147, 60-71.

Bickel, P. J., and D. A. Freeman (1981): "Some asymptotic theory for the bootstrap," *Annals of Statistics*, 9, 1196–1217.

Canova, F. (2007): *Methods for Applied Macroeconomic Research.* Princeton University Press.

Chang, P.-L. and Killian, L. (2000): "How accurate are confidence intervals for impulse responses in large VAR models?," *Economics Letters*, 69, 299-307.

Chung, C.-F. (2002): "Sample means, sample autocovariances and linear regression of stationary multivariate long memory processes," *Econometric Theory*, 18, 51–78.

Granger, C. W. J. (1980): "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227–238.

Granger, C. W. J., and R. Joyeux (1980): "An introduction to long memory time series models and fractional differencing," *Journal of Time Series Analysis*, 1, 15–39.

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

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

Henry, M. (2001): "Robust automatic bandwidth for long memory," *Journal of Time Series Analysis*, 22, 292-316.

Hosking, J. R. M. (1981): "Fractional differencing," *Biometrika*, 65, 165–176.

Hosking, J. R. M. (1981): "Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series," *Journal of Econometrics*, 73, 261–283.

Hosoya, Y. (1997): "A limit theory for long range dependence and statistical inference on related models," *Annals of Statistics*, 25, 105–137.

Inoue, A. and Y. Kasahara (2006): "Explicit representation of finite predictor coefficients and its applications," *Annals of Statistics*, 34, 973–993.

Kilian, L. (1998a): "Small-sample confidence intervals for impulse response functions," *Review of Economics and Statistics*, 80, 218–230.

Kilian, L. (1998b): "Confidence intervals for impulse responses under departures from normality," *Econometric Reviews*, 17, 1–29.

Killian, L. (1999): "Finite sample properties of percentile and percentile t bootstrap confidence intervals for impulse response," *Review of Economics and Statistics*, 81, 652–660.

Lewis, S. and G. C. Reinsel (1985): "Prediction on multivariate autoregressive time series by autoregressive model fitting," *Journal of Multivariate Analysis*, 16, 393–411.

Lutkepohl, H. (1988): "Asymptotic distribution of the moving average coefficients of an estimated vector autoregressive process," *Econometric Theory*, 4, 77–85.

Lutkepohl, H. (1989): "A note on the asymptotic distribution of impulse response functions of estimated VAR models with orthogonal residuals," *Journal of Econometrics*, 42, 371–376.

Ng, S. (2003): "Can sticky prices account for the variations and persistence in real exchange rates?," *Journal of International Money and Finance*, 22(1), 65–85.

Nielsen, M. O. and P. H. Frederiksen (2004): "Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration" *Econometric Reviews*, 24, 405-443.

Poskitt, D. (2007): "Autoregressive approximation in nonstandard situations: The non-invertible and fractionally integrated case," *Annals of the Institute of Statistical Mathematics,* 59, 697–725.

Poskitt, D. (2008): "Properties of the sieve bootstrap for fractionally integrated and non-invertible processes," *Journal of Time Series Analysis,* 29, 224–250.

Robinson, P. M. (2006): "Conditional - sum - of - squares estimation of models for stationary time series with long memory," *London School of Economics working paper.*

Rossi, B. (2005): "Confidence Intervals for Half-Life Deviations from Purchasing Power Parity," *Journal of Business and Economic Statistics,* 23, 432–442.

Rossi, B. and E. Pesevanto (2006): "Small Sample Confidence Bands for Multivariate Impulse Response Functions," *Journal of Applied Econometrics,* 21, 1135–1155.

Shibata, R. (1980): "Asymptotically efficient selection of the order of the model for estimating parameters of a linear process," *Annals of Statistics*, 8, 147–164.

Sims, C.A. (1986): "Are forecasting models usable for policy analysis?," *Quarterly Review of the Federal Reserve Bank of Minneapolis,* Winter, 2–16.

Sims, C.A. and T. Zha (1999): "Error Bands for Impulse Responses," *Econometrica,* 67(5), 1113–1156.

Wright, J H (1995): "Stochastic orders of magnitude associated with two-stage estimators of fractional ARMA systems," *Journal of Time Series Analysis*, 16, 119-125.

Wright, J H (2000): "Confidence intervals for univariate impulse responses with a near unit root," *Journal of Business and Economic Statistics*, 20, 368-373.

Appendices

Appendix A

In this Appendix we present the distributional results referred to in Section 2.2. Assumptions 4-6 are presented in Appendix B.

Theorem 5 *Under the assumptions 1-3 and 4-6, and for all $j = 1, ..., h$, where h is the maximum lag of the IRW weights being considered,*

$$\sqrt{T}\left(\psi_{j,\hat{\theta}^W} - \psi_{j,\theta_0}\right) \xrightarrow{p} N(0, D_j' W^{-1} U W^{-1} D_j) \tag{10}$$

where $D_j = \left.\frac{\partial \psi_{j,\theta}}{\partial \theta}\right|_{\theta=\theta_0}$, the (i,j)-th elements of W and U are defined in (12) and (13) of Appendix A, and θ_0 denotes the true value of θ.

Theorem 6 *Under the assumptions 1(ii) and 2, 3, and further assuming that ϵ_t is an i.i.d. sequence, that $\sum_{j=1}^{\infty} \sup_\theta |\tilde{\psi}_{j,\theta}| < \infty$ and that Ω, defined in (14) of Appendix A, is nonsingular, then for all $j = 1, ..., h$*

$$\sqrt{T}\left(\psi_{j,\hat{\theta}} - \psi_{j,\theta_0}\right) \xrightarrow{p} N(0, D_j' \Omega^{-1} D_j) \tag{11}$$

where D_j is defined in Theorem 5.

Under the assumptions of the Theorems, the results for Theorems 5 and 6 follow immediately from Theorem 2.2 of Hosoya (1997) and Theorem 2 of Robinson (2006), respectively, and the application of the delta method.

Appendix B

This Appendix sets out a set of technical regularity conditions that are required for the validity of the results of Hosoya (1997) and Theorem 5. It is necessary to define the following terms; in particular $Q^\epsilon(\omega_1, \omega_2, \omega_3)$ denotes the fourth order spectral density of ϵ_t, and is

$$Q^\epsilon(\omega_1, \omega_2, \omega_3) = \frac{1}{8\pi^3} \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} \sum_{t_3=-\infty}^{\infty} \exp\left(-i\left(\omega_1 t_1 + \omega_2 t_2 + \omega_3 t_3\right)\right) \tilde{Q}^\epsilon(t_1, t_2, t_3)$$

where $\tilde{Q}^\epsilon(t_1, t_2, t_3)$ is the joint fourth-order cumulant of ϵ_t, ϵ_{t+t_1}, ϵ_{t+t_2} and ϵ_{t+t_3}. Let

$$R_j(\theta) = H_j(\theta) + \int_{-\pi}^{\pi} h_j(\omega, \theta) f_y(\omega, \theta) d\omega,$$

Let W and U be matrices whose $ij-$th element is given by

$$W_{ij} = \frac{\partial R_i(\theta)}{\partial \theta_j}, \quad i, j = 1, ..., s, \tag{12}$$

and

$$U_{ij} = 4\pi \int_{-\pi}^{\pi} h_i(\omega, \theta) h_j(\omega, \theta) f_y^2(\omega, \theta) d\omega + \tag{13}$$
$$2\pi \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h_i(\omega_1, \theta) h_j(\omega_2, \theta) \psi_\theta(e^{i\omega_1}) \psi_\theta(e^{-i\omega_1}) \psi_\theta(e^{i\omega_2}) \psi_\theta(e^{-i\omega_2}) d\omega_1 d\omega_2$$

respectively. Finally, let

$$\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varpi(\omega) \varpi(\omega)' d\omega \tag{14}$$

where

$$\varpi(\omega) = \left[\log \left|1 - e^{i\omega}\right|^2 - 2\frac{\partial}{\partial \omega} \log \left|\psi_{\theta_0}(e^{i\omega})\right| \right].$$

The relevant technical regularity conditions are:

Assumption 4 *$Q^\epsilon(\omega_1, \omega_2, \omega_3)$ is γ-Lipschitz, uniformly in ω_1, ω_2 and ω_3, i.e.*

$$|Q^\epsilon(\omega_1 + \varepsilon_1, \omega_2 + \varepsilon_2, \omega_3 + \varepsilon_3) - Q^\epsilon(\omega_1, \omega_2, \omega_3)| < \left\{ \max_i |\varepsilon_i| \right\}^\gamma.$$

Assumption 5 *(i) $f_y(\omega)$ is bounded away from zero (ii) $\int_{-\pi}^{\pi} \psi(e^{i\omega})^{2u} d\omega < \infty$, for some u such that $1 < u \leq 2$. (iii) There exists $c > 1/2$, such that*

$$\sup_{|\lambda| < \varepsilon} \left(\int_{-\pi}^{\pi} \left| f_y^{-1}(\omega) \left(f_y(\omega) - f_y(\omega - \lambda) \right) \right|^u d\omega \right)^{1/u} = O(\varepsilon^c)$$

for some u such that $1 < u \leq 2$. (iv) For any $\varepsilon > 0$ and θ, there exists $a > 0$, and functions $\tilde{h}_j(\omega)$ and $\bar{h}_j(\omega)$, such that, if $|\theta_1 - \theta| < a$, $\tilde{h}_j(\omega) \leq h_j(\omega, \theta_1) \leq \bar{h}_j(\omega)$ and

$$\left(\int_{-\pi}^{\pi} \left| f_y(\omega) \left(\bar{h}_j(\omega) - \tilde{h}_j(\omega) \right) \right|^v d\omega \right)^{1/v} < \varepsilon,$$

for $v = (u-1)/u$ and $1 < u \leq 2$.

Assumption 6 *Given $\varepsilon > 0$, there exists integer $m(\varepsilon)$, a partition $U^{(1)}(r), ..., U^{(m(\varepsilon))}(r)$ of the ball in Θ with centre θ_0 and radius r and square integrable functions $\tilde{h}_j^i(\omega)$ and $\bar{h}_j^i(\omega)$ such that for all sufficiently small r and for all j, $\tilde{h}_j^l(\omega) \leq h_j(\omega, \theta) \leq \bar{h}_j^l(\omega)$ if $\theta \in U^{(l)}(r)$. Also,*

$$\left(\int_{-\pi}^{\pi} \left| \psi_\theta(e^{i\omega}) \psi_\theta(e^{-i\omega}) \left(\bar{h}_j^l(\omega) - h_j(\omega, \theta_0) \right) \right|^v d\omega \right)^{1/v} \leq \varepsilon r$$

and

$$\left(\int_{-\pi}^{\pi}\left|\psi_\theta(e^{i\omega})\psi_\theta(e^{-i\omega})\left(\tilde{h}_j^l(\omega)-h_j(\omega,\theta_0)\right)\right|^v d\omega\right)^{1/v}\leq\varepsilon r,$$

for all l*, where* $v=(u-1)/u$ *and* $1<u\leq 2$*. Further, Condition B of Hosoya (1997), holds for the pairs* $\left\{\tilde{h}_j^l,\psi\right\}$, $\left\{\bar{h}_j^l,\psi\right\}$ *and* $\{h_j(.,\theta_0),\psi\}$*, for all* l,j*.*

There are several connections between these technical regularity conditions, the assumptions made in the body of the text and the assumptions needed for Theorem 2.2 of Hosoya (1997). Assumption 3(ii) and 5(i) is sufficient for differentiability of the spectral density function, its logarithm, its inverse and Assumptions C(iv) and D(ii) of Hosoya (1997), as required for Theorem 2.2 of Hosoya (1997). The identifiability conditions of Assumption 3(i) imply Assumptions C(iii) and D(iv) of Hosoya (1997). Assumption 4, the ergodicity and martingale difference assumption of Assumption 1 imply Assumption A of Hosoya (1997). Finally, Assumption 6 implies Assumption D (iii) and the second part of Assumption D(iv) of Hosoya (1997), needed for the bracketing function approach taken in that paper.

Appendix C

This Appendix provides the proof of Theorem 1.

Since all the roots of the polynomials in the lag operator $\phi(L)$ and $\theta(L)$ lie outside the unit circle, it follows that $\sum_{k=0}^{\infty}\pi_k^2<\infty$ and hence that

$$\sum_{k=1}^{t-1}\pi_k y_{t-k}=O_p(1).$$

The Local Whittle estimator $\hat{d}_{LW}$ will generate the fractionally filtered series

$$\hat{u}_t=(1-L)^{\hat{d}_{LW}}y_t=y_t-\sum_{l=1}^{t-p}\hat{\pi}_l(\hat{d}_{LW})y_{t-l},$$

where

$$\hat{\pi}_l(\hat{d}_{LW})=\Gamma(l-\hat{d}_{LW})\Gamma(-\hat{d}_{LW})^{-1}\Gamma(l+1).$$

Since $\hat{u}_t=(1-L)^{\hat{d}_{LW}}y_t$, then

$$(\hat{u}_t-u_t)=\sum_{j=1}^{\infty}\pi_j(\hat{d}_{LW}-d_0)u_{t-j}.$$

Since

$$(\hat{d}_{LW}-d_0)=O_p(m^{-1/2})$$

and $u_t = (1-L)^d y_t$, then following the same approach as Wright (1995),

$$T^{-1}\sum_{j=1}^{\infty}(\hat{u}_t - u_t)^2 = T^{-1}\sum_{t=1}^{T}\left(\sum_{j=1}^{t-1}\pi_j(\hat{d}_{LW} - d_0)u_{t-j}\right)^2.$$

Then, using the mean value theorem we have that $\pi_j(d) = dX_j^1 + d^2X_j^2$, where X_j^1 denotes the first derivative and X_j^2 the second derivative of $\pi_j(.)$. Then,

$$\sum_{k=1}^{t-1}\pi_k u_{t-k} = d\sum_{k=1}^{t-1}X_j^1 u_{t-j} + d^2\sum_{k=1}^{t-1}X_j^2 u_{t-j},$$

and following the same arguments as in Wright (1995),

$$(\hat{d}_{LW} - d_0)\sum_{k=1}^{t-1}X_j^1 u_{t-j} = O_p(m^{-1/2}),$$

and

$$T^{-1}(\hat{d}_{LW} - d_0)\sum_{k=1}^{t-1}X_j^2 u_{t-j} = O_p(m^{-1/2}),$$

and hence

$$T^{-1}\sum_{t=1}^{T-k}\hat{u}_t\hat{u}_{t+k} = T^{-1}\sum_{t=1}^{T-k}u_t u_{t+k} + O_p(m^{-1/2}) \tag{15}$$

This suffices to prove the result for an $ARFIMA(p,d,0)$ model. For the general case of an $ARFIMA(p,d,q)$ model we have that for the second step $ARMA$ estimation, the conditional MLE needs to be numerically maximized. Let us denote the likelihood function by $L(\beta;d)$. The form of the likelihood may be found in, e.g., (5.6.3) of Hamilton (1994) and is given by

$$L(\beta(d)) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\left(\sigma^2\right) - \sum_{t=1}^{T}\frac{\epsilon_t^2\left(\beta(d)\right)}{\sigma^2}$$

where

$$\epsilon_t\left(\beta(d)\right) = u_t - \sum_{j=1}^{p}\phi_j u_{t-j} - \sum_{j=1}^{q}\vartheta_j\epsilon_{t-j}\left(\beta(d)\right)$$

It is clear that the likelihood function is differentiable. It is further clear that, once initial conditions for $\epsilon_q\left(\beta(d)\right), ..., \epsilon_1\left(\beta(d)\right)$ are set, $L(\beta(d))$ is a function of autocovariances of u_t. Then, as long as (15) holds we have that

$$L\left(\hat{\beta}_{LWTSE}(\hat{d}_{LW})\right) - L\left(\beta_0(d_0)\right) = O_p(m^{-1/2}).$$

But, by an application of the mean value theorem we have that

$$L\left(\hat{\beta}(\hat{d}_{LW})\right) = L\left(\beta_0(d_0)\right) + \left.\frac{\partial L}{\partial\beta}\right|_{\beta=\bar{\beta}}\left(\hat{\beta}(\hat{d}_{LW}) - \beta_0(d_0)\right).$$

Hence, the result of the Theorem holds for $ARFIMA(p,d,q)$ models completing the proof.

Appendix D

We wish to prove that the parametric bootstrap for the parameter estimates, of parametric long memory models is valid. We will focus on the proof of Theorem 2, i.e. for the conditional sum of squares (CSS) estimator of θ. The proof of Theorem 1 is very similar and is not reported. We do not assume Gaussianity of the data unlike most of the literature including Andrews and Lieberman (2006). Let $\sim^d$ denote asymptotic equivalence in weak law possibly in different probability spaces. Formally, we wish to show that

$$\sqrt{T}(\hat{\theta}^* - \hat{\theta}) \sim^d \sqrt{T}(\hat{\theta} - \theta_0).$$

The assumed model is of the form

$$y_t = \sum_{i=0}^{\infty} \psi_{i,\theta_0} \epsilon_{t-i}$$

which by Assumption 2 is invertible, so that

$$y_t = \sum_{i=1}^{\infty} \kappa_{i,\theta_0} y_{t-i} + \epsilon_t$$

Without loss of generality, we set

$$\kappa_{i,\theta_0} = \tilde{\kappa}_{i,\theta_0} i^{-d(\theta_0)-1}$$

such that $\sup_i \tilde{\kappa}_{i,\theta_0} < \infty$ and $0 < d(\theta_0) < 1/2$. This implies that, for some $\tilde{\psi}_{i,\theta_0}$, such that $\sup_i \tilde{\psi}_{i,\theta_0} < \infty$,

$$\psi_{i,\theta_0} = \tilde{\psi}_{i,\theta_0} i^{d(\theta_0)-1}$$

The parametric bootstrap we investigate is based on constructing bootstrap samples by either

$$\hat{y}_t^* = \sum_{i=0}^{\infty} \psi_{i,\hat{\theta}}(\hat{\theta}) \hat{\epsilon}_{t-i}^*$$

or

$$\hat{y}_t^* = \sum_{i=1}^{t-1} \kappa_{i,\hat{\theta}} y_{t-i} + \hat{\epsilon}_t^*$$

where $\hat{\epsilon}_t^*$ is an i.i.d. re-sample with replacement of $\hat{\epsilon}_t$, where $\hat{\epsilon}_t$ is the residual resulting from the estimation giving $\hat{\theta}$. The CSS estimator of θ is given by

$$\hat{\theta} = \arg\min_{\theta \in \Theta} s_T(\theta)$$

where

$$s_T(\theta) = \sum_{t=1}^{T} \epsilon_t(\theta)^2$$

and

$$\epsilon_t(\theta) = y_t - \sum_{i=1}^{t-1} \kappa_{i,\theta} y_{t-i}$$

Theorem 2 of Robinson (2006) shows that $\sqrt{T}(\hat{\theta}-\theta_0)$ has a normal probability law. We introduce the following notation:

$$y_t^* = \sum_{i=0}^{t-1} \psi_{i,\theta_0} \epsilon_{t-i}^*$$

where ϵ_t^* is an i.i.d. resample with replacement of ϵ_t. Define

$$\hat{\theta}^* = \arg\min_{\theta\in\Theta} \hat{s}_T^*(\theta), \quad \hat{s}_T^*(\theta) = \sum_{t=1}^{T} \hat{\epsilon}_t^*(\theta)^2, \quad \hat{\epsilon}_t^*(\theta) = \hat{y}_t^* - \sum_{i=1}^{t-1} \kappa_{i,\theta} \hat{y}_{t-i}^*$$

and

$$\theta^* = \arg\min_{\theta\in\Theta} s_T^*(\theta), \quad s_T^*(\theta) = \sum_{t=1}^{T} \epsilon_t^*(\theta)^2, \quad \epsilon_t^*(\theta) = y_t^* - \sum_{i=1}^{t-1} \kappa_{i,\theta} y_{t-i}^*$$

Let $\epsilon = (\epsilon_1, ..., \epsilon_T)'$, $\epsilon^* = (\epsilon_1^*, ..., \epsilon_T^*)$, $\hat{\epsilon}^* = (\hat{\epsilon}_1^*, ..., \hat{\epsilon}_T^*)$, $y = (y_1, ..., y_T)'$, $y^* = (y_1^*, ..., y_T^*)$ and $\hat{y}^* = (\hat{y}_1^*, ..., \hat{y}_T^*)$. Recall that P_y denotes the probability law of a random vector x and $d(P_{y_1}, P_{y_2})$ the Mallows metric between P_{y_1} and P_{y_2}. Finally, define a continuous function $\Psi(\epsilon;\theta)$ to describe the mapping from ϵ to y. Then, we have

$$\mathfrak{m}(\epsilon, \epsilon^*) \to 0$$

But the fact that (3.7)-(3.9) of Robinson (2006) are $o_p(T^{-1/2})$, is sufficient for,

$$\mathfrak{m}(\hat{\epsilon}, \hat{\epsilon}^*) \to 0 \tag{16}$$

Then, by Lemma 8.5 of Bickel and Freeman (1981), using Ψ as a relevant function, it follows from (16) that

$$\mathfrak{m}(y, \hat{y}) \to 0$$

Then, it immediately follows that

$$\sqrt{T}(\hat{\theta}^* - \hat{\theta}) \sim^d \sqrt{T}(\hat{\theta} - \theta_0).$$

and so $\sqrt{T}(\hat{\theta}^* - \hat{\theta})$ has an asymptotic Normal distribution. The result follows immediately by noting that the IRWs are, by assumption, a continuous function of the model parameters.

Appendix E

This Appendix proves that the sieve bootstrap is valid for impulse response analysis based on the estimation of an $AR(p_T)$ model. We use the results of Poskitt (2007) and Poskitt (2008). Let $\hat{\kappa}^{(p_T)}$ denote the $p_T \times 1$ vector of parameter estimates of the coefficients of an $AR(p_T)$ model fitted to the original sample. Let $\hat{\kappa}^{*,(p_T)}$ denote the same estimates obtained from a bootstrap sample constructed using the sieve bootstrap. Let $X_t^{(p_T)} = (x_{t-1},...,x_{t-p_T})'$, $X^{(p_T)} = (X_{p_T+1}^{(p_T)},...,X_T^{(p_T)})'$, $x = (x_{p_T+1},...,x_T)'$. Starred variables represent bootstrap versions of non-starred variables. Then,

$$\hat{\kappa}^{(p_T)} = \left(X^{(p_T)\prime} X^{(p_T)}\right)^{-1} X^{(p_T)\prime} x$$

and

$$\hat{\kappa}^{*,(p_T)} = \left(X^{*,(p_T)\prime} X^{*,(p_T)}\right)^{-1} X^{*,(p_T)\prime} x^*$$

Let $\{A\}_{ij}$ denote the i,j-th element of a matrix A. We first need to determine the rate at which $\hat{\kappa}_j^{(p_T)}$ converges to $\kappa^{(p_T)}$ and $\hat{\psi}_j^{(p_T)}$ converges to $\psi_j^{(p_T)}$. By Theorem 5 of Poskitt (2007), we have that

$$\left\|\hat{\kappa}^{(p_T)} - \kappa^{(p_T)}\right\| = O_p\left(p_T\left(\frac{\ln(T)}{T}\right)^{1/2-d}\right)$$

Further,

$$\hat{\psi}_j^{(p_T)} - \psi_j^{(p_T)} = O_p\left(\left\|\kappa^{(p_T)}\right\|^2\left\|\hat{\kappa}^{(p_T)} - \kappa^{(p_T)}\right\|\right) = O_p\left(p_T^{3/2}\left(\frac{\ln(T)}{T}\right)^{1/2-d}\right) = O_p\left(r_T\right)$$

Define $q_T = \left(\frac{T}{\ln(T)}\right)^{1/2-d}$. The Theorem is proven if we show that

$$\mathfrak{m}\left(q_T\left(\lambda'\left(\hat{\kappa}^{(p_T)} - \kappa^{(p_T)}\right)\right), q_T\left(\lambda'\left(\hat{\kappa}^{*,(p_T)} - \hat{\kappa}^{(p_T)}\right)\right)\right) \to 0$$

whereλ is some, finite dimensional, selector vector and $\mathfrak{m}(y_1, y_2)$ is the Mallows metric between the probability measures, P_{y_1} and P_{y_2}, of two random vectors y_1 and y_2. We have that

$$\mathfrak{m}\left(q_T\left(\lambda'\left(\hat{\kappa}^{(p_T)} - \kappa^{(p_T)}\right)\right), q_T\left(\lambda'\left(\hat{\kappa}^{*,(p_T)} - \hat{\kappa}^{(p_T)}\right)\right)\right)^2 \leq E\left[E^*\left(\left\|q_T\left(\hat{\kappa}^{*,(p_T)} - \hat{\kappa}^{(p_T)}\right)\right\|^2\right)\right] \leq$$

$$E\left[E^*\left(\left\|\left(X^{*,(p_T)\prime}X^{*,(p_T)}\right)^{-1} - \left(X^{p_T\prime}X^{(p_T)}\right)^{-1}\right\|^2\right)\right] E\left[E^*\left(\left\|q_T\left(X^{*,(p_T)\prime}v^{*,(p_T)} - X^{(p_T)\prime}v^{(p_T)}\right)\right\|^2\right)\right]$$

Examining each of the two terms above we have

$$E\left[E^*\left(\left\|\left(X^{*,(p_T)\prime}X^{*,(p_T)}\right)^{-1} - \left(X^{(p_T)\prime}X^{(p_T)}\right)^{-1}\right\|^2\right)\right] \leq$$

$$p_T^4 E\left[E^*\left(\left\|X^{*,(p_T)\prime}X^{*,(p_T)} - X^{(p_T)\prime}X^{(p_T)}\right\|^2\right)\right] \leq$$

$$p_T^6 \sup_{1\le i,j\le p_T} E\left[E^*\left(\left\|\left\{X^{*,(p_T)\prime}X^{*,(p_T)}\right\}_{ij}-\left\{X^{(p_T)\prime}X^{(p_T)}\right\}_{ij}\right\|^2\right)\right]$$

But

$$\sup_{1\le i,j\le p_T} E\left[E^*\left(\left\|\left\{X^{*,(p_T)\prime}X^{*,(p_T)}\right\}_{ij}-\left\{X^{(p_T)\prime}X^{(p_T)}\right\}_{ij}\right\|^2\right)\right]\le$$

$$p_T^2 E\left[E^*\left(\left\|\left\{X^{*,(p_T)\prime}X^{*,(p_T)}\right\}_{11}-\left\{X^{(p_T)\prime}X^{(p_T)}\right\}_{11}\right\|^2\right)\right]$$

Further,

$$E\left[E^*\left(\left\|X^{*,(p_T)\prime}v^{*,(p_T)}-X^{(p_T)\prime}v^{(p_T)}\right\|^2\right)\right]\le p_T \sup_{1\le i\le p_T} E\left[E^*\left(\left\|\left\{X^{*,(p_T)\prime}v^{*,(p_T)}\right\}_{i}-\left\{X^{(p_T)\prime}v^{(p_T)}\right\}_{i}\right\|^2\right)\right]\le$$

$$p_T^2 E\left[E^*\left(\left\|\left\{X^{*,(p_T)\prime}v^{*,(p_T)}\right\}_{1}-\left\{X^{(p_T)\prime}v^{(p_T)}\right\}_{1}\right\|^2\right)\right]$$

But by the proof of Theorem 4.1 of Poskitt (2008) we have that

$$E\left[E^*\left(\left\|\left\{X^{*,(p_T)\prime}X^{*,(p_T)}\right\}_{11}-\left\{X^{(p_T)\prime}X^{(p_T)}\right\}_{11}\right\|^2\right)\right]=O\left(p_T^{5/2}\left(\frac{\log T}{T}\right)^{1-2d}\right)$$

and

$$E\left[E^*\left(\left\|\left\{X^{*,(p_T)\prime}v^{*,(p_T)}\right\}_{1}-\left\{X^{(p_T)\prime}v^{(p_T)}\right\}_{1}\right\|^2\right)\right]=O\left(p_T^{5/2}\left(\frac{\log T}{T}\right)^{1-2d}\right)$$

Hence,

$$\mathfrak{m}\left(q_T\left(\lambda'\left(\hat{\kappa}^{(p_T)}-\kappa^{(p_T)}\right)\right),q_T\left(\lambda'\left(\hat{\kappa}^{*,(p_T)}-\hat{\kappa}^{(p_T)}\right)\right)\right)=O\left(p_T^{21/2}\left(\frac{\log T}{T}\right)^{1-2d}\right)$$

But since $p_T = O\left(\log T^a\right)$, it follows that

$$\mathfrak{m}\left(q_T\left(\lambda'\left(\hat{\kappa}^{(p_T)}-\kappa^{(p_T)}\right)\right),q_T\left(\lambda'\left(\hat{\kappa}^{*,(p_T)}-\hat{\kappa}^{(p_T)}\right)\right)\right)=O\left(\frac{\log T^{\frac{21a}{2}+1-2d}}{T^{1-2d}}\right)=o(1)$$

proving that $q_T\left(\lambda'\left(\hat{\kappa}^{*,(p_T)}-\hat{\kappa}^{(p_T)}\right)\right)$ has the same probability law as $q_T\left(\lambda'\left(\hat{\kappa}^{(p_T)}-\kappa^{(p_T)}\right)\right)$ and, therefore, $r_T\left(\hat{\psi}_j^{*,(p_T)}-\hat{\psi}_j^{(p_T)}\right)$ has the same probability law as $r_T\left(\hat{\psi}_j^{(p_T)}-\psi_j^{(p_T)}\right)$

Figure 3: Monte Carlo Results: Coverage Rates under Long Memory. Key: Solid Line (—–) represents Approach 1; Long Dashed Line (– – –) represents Approach 2; Dotted Line (. . .) represents Approach 3; Short Dashed Line (- - -) represents Approach 4. Horizontal line Represents 90% nominal level.

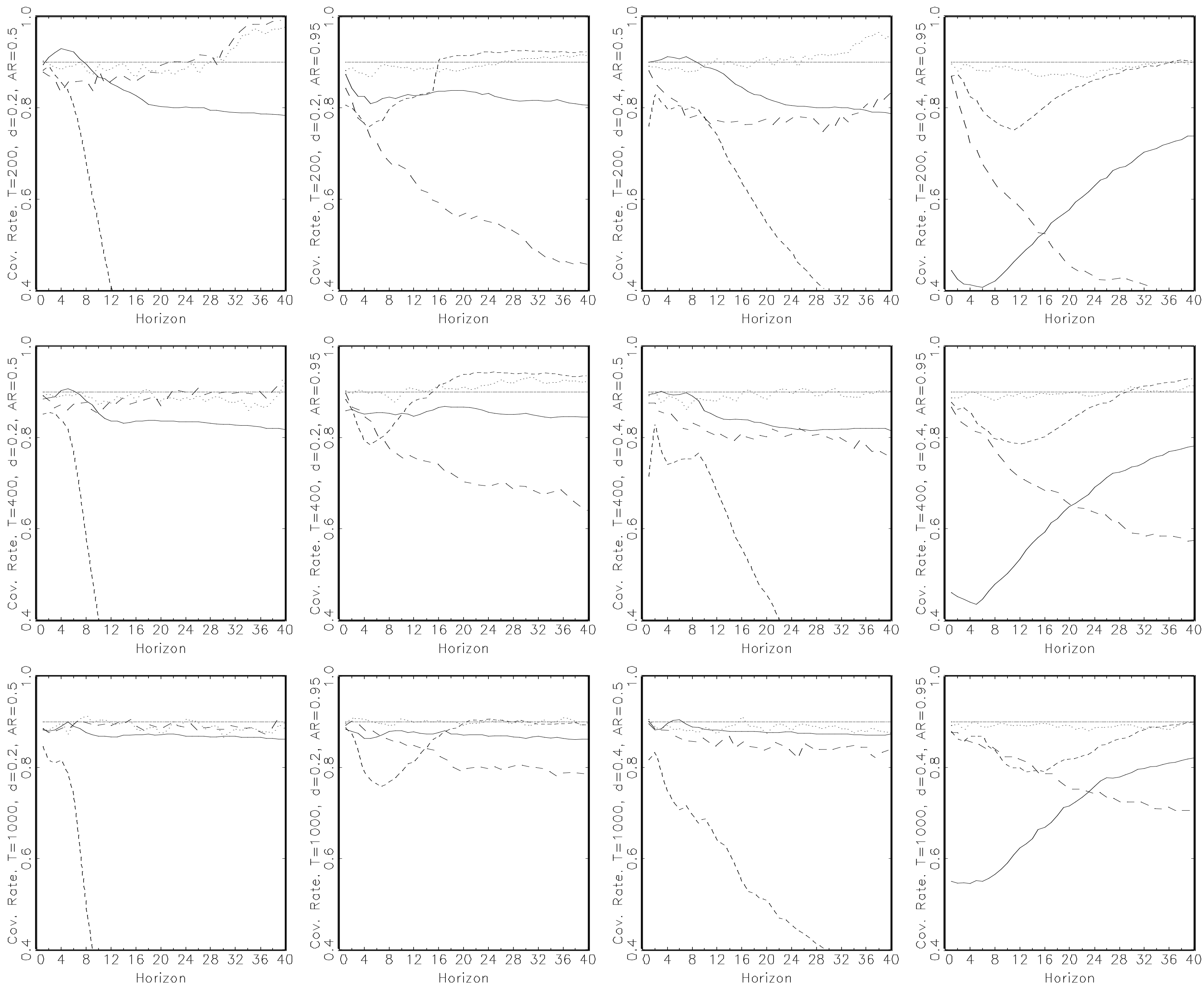


Figure 4: Monte Carlo Results: Interval Width under Long Memory. Key: Solid Line (——) represents Approach 1; Long Dashed Line (– – –) represents Approach 2; Dotted Line (. . .) represents Approach 3; Short Dashed Line (- - -) represents Approach 4.

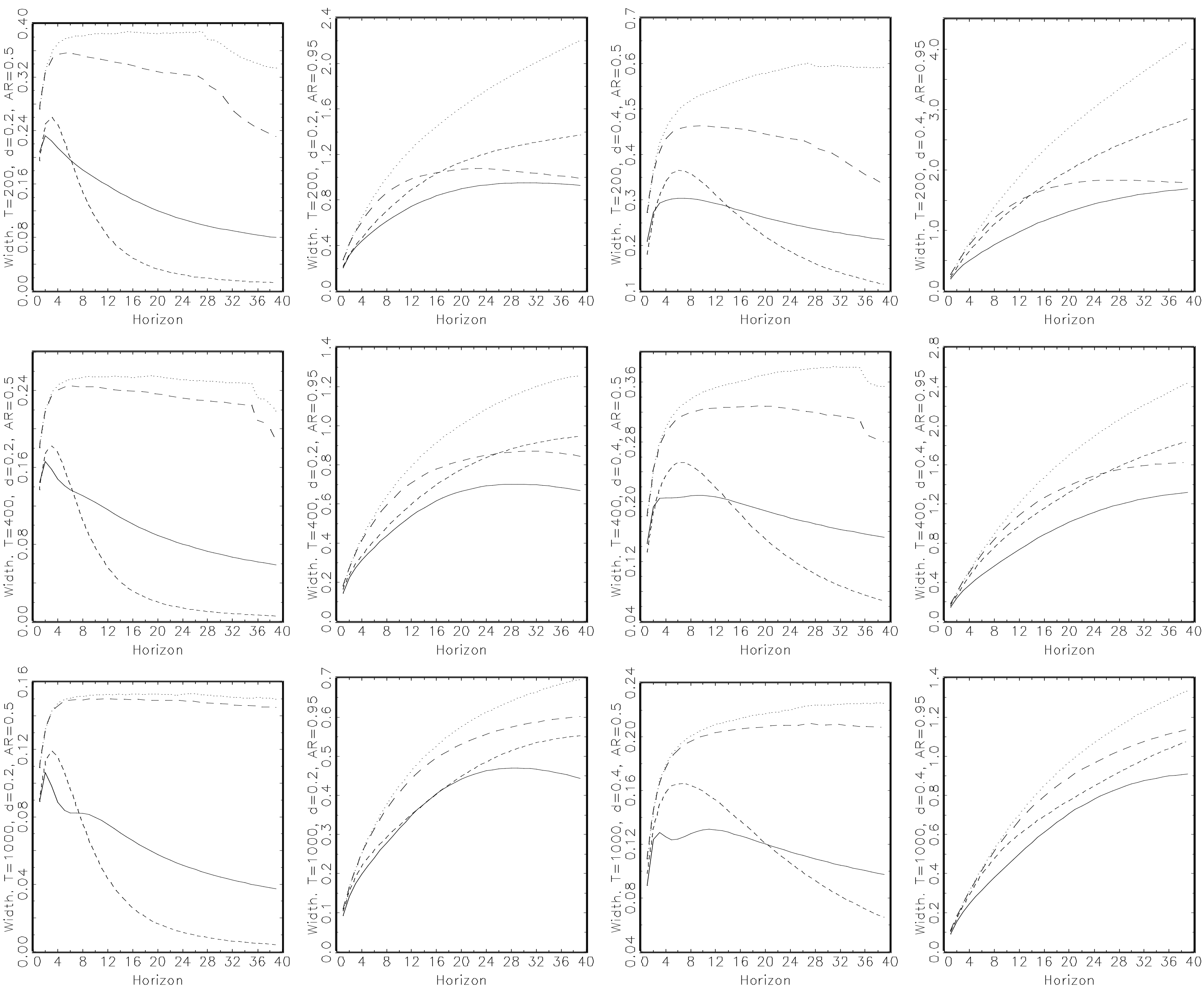


Figure 5: Monte Carlo Results: Coverage Rates under Short Memory. Key: Solid Line (——) represents Approach 1; Long Dashed Line (– – –) represents Approach 2; Dotted Line (. . .) represents Approach 3; Short Dashed Line (- - -) represents Approach 4. Horizontal line Represents 90% nominal level.

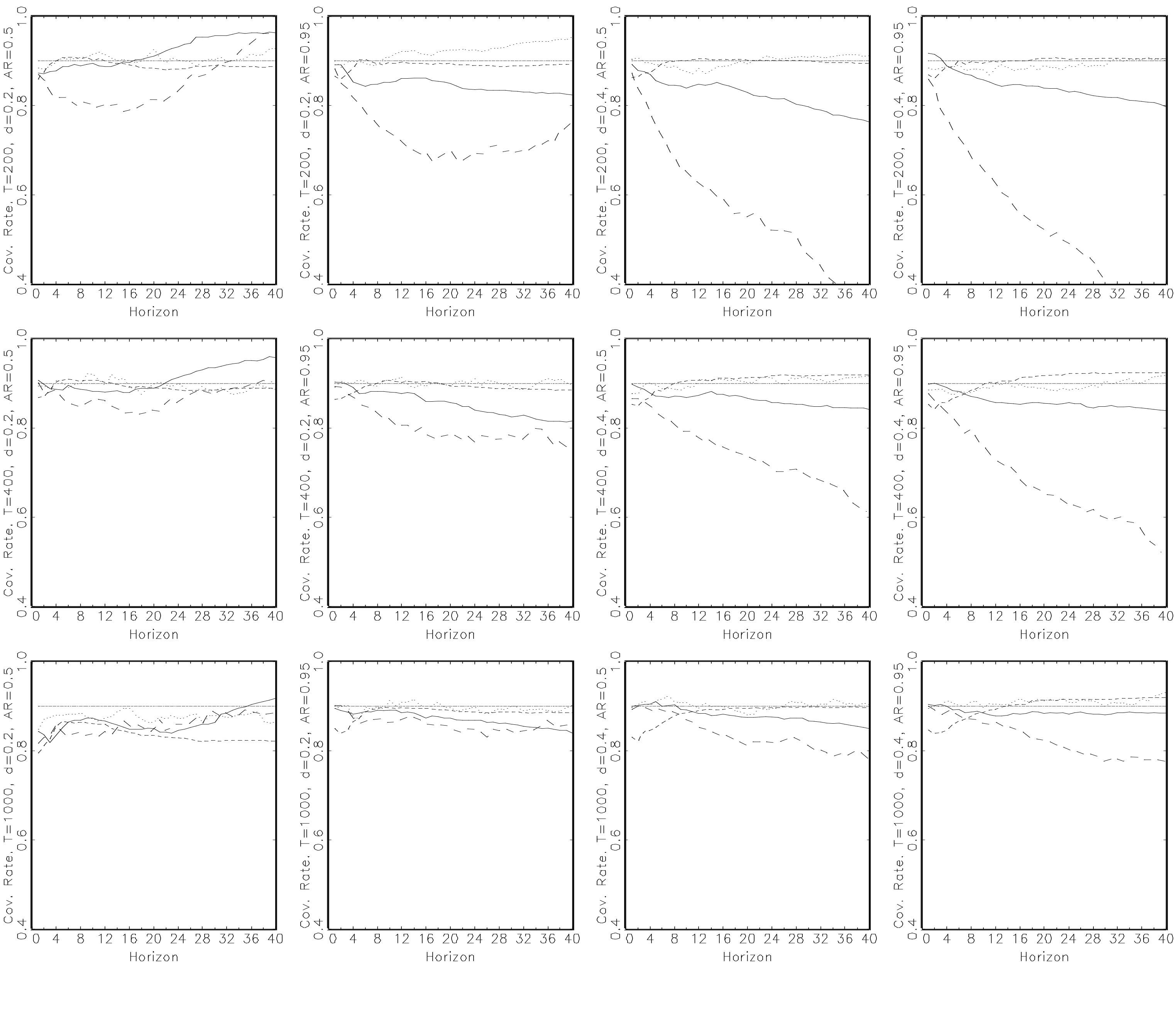
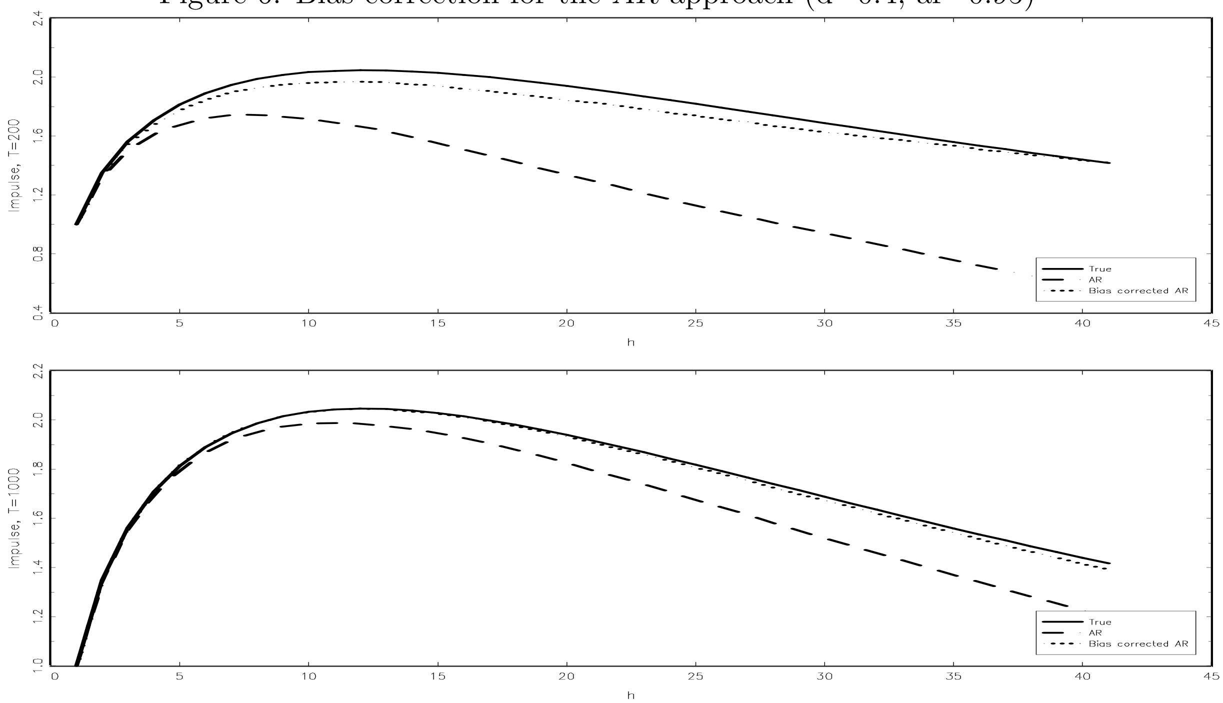


Figure 6: Bias correction for the *AR* approach (d=0.4, ar=0.95)



Key: Solid Line (——) represents the true IRW; Long Dashed Line (– – –) represents the AR Approximation; Dotted Line (. . .) represents the bias corrected AR Approximation.

Figure 7: Empirical Results: Impulse Responses for Real Exchange Rates