Department of Economics
Seminar Series

Richard Baillie
Michigan State University

"Estimation and Inference for Impulse Response Weights From Strongly Persistent Processes"

Friday,
November 12, 2010
3:30 p.m.
210 Middlebush

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

June 14, 2010

Abstract

This paper considers the problem of estimating impulse response weights (IRWs) from processes that may be strongly dependent and the related issue of constructing confidence intervals for the estimated IRWs. We compare several approaches including $QMLE$, and a two step estimator that uses a semi parametric estimate of the long memory parameter in the first step, and also an estimator from fitting an autoregressive approximation. A main focus of the paper concerns the most appropriate method for constructing confidence intervals for the IRWs. We show that the parametric bootstrap is valid under very weak conditions, including non Gaussianity, for making inference on IRW from possibly strongly dependent processes. We also propose, and justify theoretically, a semi-parametric sieve bootstrap based on autoregressive approximations that can be used for IRWs obtained by autoregressive approximations. We find that estimates of IRW based on autoregressive approximations and also confidence intervals of IRWs based on the sieve bootstrap generally have very desirable properties and are shown to perform well in a detailed simulation study. Finally, we apply the methods we develop to an extensive and detailed empirical application on inflation and real exchange rate data.

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

JEL Codes: C22, C12.

1 Introduction

Many economic and financial time series fall into the category of being strongly persistent. One particular class of processes are those with relatively slowly decaying hyperbolic autocorrelations; see Granger and Joyeux (1980), Granger (1980) and Hosking (1981). These models have proved extremely relevant for the conditional mean of many macroeconomic time series and also

*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.

for describing volatility processes and spreads in finance. This paper is concerned with issues that arise in the practical investigation of a time series with these characteristics. In general, apart from estimating the degree of persistence, i.e. the order of fractional integration of a time series, a researcher will typically be concerned with estimating the implied Impulse Response Weights ($IRWs$) for a variety of lags. This paper focuses on the performance of various estimators of IRWs and also derives some new bootstrap related methodology for deriving appropriate corresponding confidence intervals. We consider a variety of estimators including both parametric and semi parametric for both the problems of estimating the mean and dispersion of the estimated IRWs. Since the literature contains virtually nothing on the semi parametric estimation of IRW for possibly strongly dependent processes, we therefore consider some parametric alternatives. We introduce a new semi parametric estimator of IRW based on high order autoregressive approximations and find that our new estimator has good theoretical and small sample properties. We also consider in detail the use of Local Whittle (LW) estimators of the long memory parameter and analyze the effects of using the LW semi parametric estimator (SPE) of the long memory parameter to obtain estimates of short run parameters and $IRWs$.

While point estimates of IRWs are straightforwardly obtained from an estimated dynamic model, there has been a general long standing concern over the most appropriate method for the construction of confidence intervals for the estimated IRWs. For example, Sims (1986) and Phillips (1998) have considered this issue for weakly dependent processes, such as a stationary Vector Autoregression (VAR). Wright (2000) has also 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. One alternative is to use bootstrap methods which have been advocated by Kilian (1998a) and Kilian (1998b).

This paper extends the previous literature to the empirically important and relevant class of strongly dependent processes; the most popular one being the $ARFIMA$ model. We consider the parametric bootstrap for a general class of parametric models for strongly dependent processes, which extends the previous results for weakly dependent autoregressive processes. We show that the parametric bootstrap is valid in this case under quite weak assumptions, and very importantly manages to avoid the wide spread assumption of Gaussianity. We then 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 IRW inference analysis based on the AR approximation and the validity of bootstrap inference on the resulting IRWs. Our new methods and theoretical results for the estimation and construction of confidence intervals for IRWs are accompanied by an extensive Monte Carlo study which investigates the performance of regular bootstrap, and parametric versus semi parametric bootstrap inference. The results indicate that the sieve bootstrap has a number of benefits. Finally, we provide an empirical example of the approaches with an investigation of the persistence of some inflation and real exchange rate series for a large number of countries. The analysis is based on the popular concept of half life for summarizing the estimated IRWs. 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 5 and 4 describe the various Monte Carlo results, while 6 presents the empirical results. There is also a conclusions section 7, followed by an appendix which contains the various proofs.

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 $f_x(\omega) = \frac{\sigma^2}{2\pi}\psi(e^{i\omega})\psi(e^{-i\omega})$, 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 processes, delineated by (1) 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 IRW, it is necessary to introduce a parametric representation associated with the above setup that is more general but encompasses $ARFIMA$ processes. The ψ_j parameters 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 $\psi_{j,\theta}$ 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 $x_t = \sum_{j=0}^{\infty} \psi_{j,\theta_0} \epsilon_{t-j}$. 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 carry out inference on the estimated ψ_j, with special attention to the issue of construction of confidence intervals for the estimated ψ_j. One standard method of inference is to derive the asymptotic approximation of the distribution of estimators of ψ_j. The most commonly used approach is to use the parametric estimator given by $\psi_{j,\hat{\theta}}$, where $\hat{\theta}$ is the MLE of θ. In this paper we focus 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. As shown by Robinson (2006), it 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 clarity we provide separate results for both estimators, where the frequency domain $QMLE$ of Hosoya (1997) is denoted by $\hat{\theta}^W$. The first results of this paper relate to the asymptotic distributions of $\psi_{j,\hat{\theta}}$ and $\psi_{j,\hat{\theta}^W}$; and for this theorem, the following set of technical regularity conditions are required for proofs of the asymptotic normality of $\hat{\theta}^W$. These conditions are given in Appendix A, along with the additional assumptions 3 through 6.

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

$$\sqrt{T} \left(\psi_{j,\hat{\theta}^W} - \psi_{j,\theta_0} \right) \xrightarrow{p} N(0, D_j' W^{-1} U W^{-1} D_j) \tag{3}$$

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 (16) and (17) of Appendix A, and θ_0 denotes the true value of θ.

Theorem 2 *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 (18) of Appendix A, is nonsingular, then for all $j = 1, ..., h$*

$$\sqrt{T}\left(\psi_{j,\hat{\theta}} - \psi_{j,\theta_0}\right) \xrightarrow{p} N(0, D_j'\Omega^{-1}D_j) \tag{4}$$

where D_j is defined in Theorem 1 and $\Omega = WU^{-1}W$.

The proofs of these theorems are given in Appendix B and are often referred to as the "delta method" for obtaining the asymptotic distributions of the estimated IRWs. The above results provide an operational way to construct asymptotically valid standard errors for $\psi_{j,\hat{\theta}}$ and $\psi_{j,\hat{\theta}W}$. It should be noted that the practical calculation of the matrix of partial derivatives, D_j can be straightforwardly achieved by numerical methods as their closed form solution are not easily accessible. This is in contrast to the case for weakly dependent processes, as opposed to strong dependent processes as above. For example, the limiting distribution of IRWs from stationary and invertible vector $ARMA$ models allows the corresponding D_j matrices to be straightforwardly obtained parametrically; see Baillie (1987) for vector $ARMA$ and Lutkepohl (1988) and Lutkepohl (1989) for VARs. It is also known from existing work on such weakly dependent processes, that asymptotic approximations for estimated IRWs are problematic for conducting small sample inference; for example see Kilian (1998a). The existence of these problems is essentially the motivation for the extensive use of the bootstrap for IRWs in 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. An 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. We refer to this approach 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

$$\ln\left[\frac{1}{m}\sum_{j=1}^{m} \omega_j^{2d} I(\omega_j)\right] - \frac{2d}{m}\sum_{j=1}^{m} \ln(\omega_j) \tag{5}$$

with respect to d, where $I(\omega_j)$ is the periodogram given by $I(\omega_j) = \frac{1}{2\pi T}\left|\sum_{j=1}^{T} y_t e^{i\omega_j t}\right|^2$, and m is the bandwidth. For the LW estimator of d, it is known that, for linear processes, $m^{1/2}\left(\hat{d}_{LW} - d_0\right) \to N\{0, 1/4\}$ 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 $m = T^{0.5}$. However, when there is substantial persistence in the short run dynamics, the value of m should potentially be reduced so that more weight is placed on ordinates of the periodogram associated with the low frequency components. On denoting the spectral density function (s.d.f.) of y_t as $f(\omega)$, and the s.d.f. of u_t is $f^*(\omega)$, then $f(\omega) = |1 - \exp(i\omega)|^{-2d} f^*(\omega)$. The s.d.f. of y_t can be approximated as $\omega \to 0+$, by $f(\omega) = \mathcal{L}(1/\omega)\omega^{-2d}\{1+c\omega^{\beta}+o(\omega^{\beta})\}$ where $\mathcal{L}(1/\omega)$ is a slowly varying function with $0 < c < \infty$, usually chosen as one, and $\beta \in (0, 2]$. Henry (2001) has found the optimal bandwidth m^*_{LW} for the LW estimator to be $m^*_{LW} = \left(\frac{3}{4\pi}\right)^{4/5} \left|\tau^* + \frac{d}{12}\right|^{-2/5} T^{4/5}$ where $\tau^* = \left[\dfrac{f^{*''}(0)}{2f^*(0)}\right]_{\omega=0}$ and τ^* has the interpretation of representing the degree of smoothness of the spectral density of the short memory component u_t, as the frequency approaches zero.

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 the $\hat{d}_{LPW}$ 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 $(1-L)^d y_t = y_t - \sum_{l=1}^{\infty} \pi_l(d) y_{t-l}$, 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

$$\hat{u}_t = y_t - \sum_{l=1}^{t-p} \hat{\pi}_l(\hat{d}_{LW}) y_{t-l} \tag{6}$$

where $\hat{\pi}_l(\hat{d}_{LW}) = \Gamma(l-\hat{d}_{LW})\Gamma(-\hat{d}_{LW})^{-1}\Gamma(l+1)$. These results for long memory processes with a variety of short run dynamics, such as non linear autoregressions considered by For concreteness, this paper focuses on the estimation of the widely used univariate $ARFIMA(p, d, q)$ process; although extensions to models with more complicated short run dynamics are quite manageable. The complete parameter vector is denoted by $\vartheta = (d, \beta)'$, where the $(p+q)$ $ARMA$ parameters are in the vector $\beta = (\phi_1, ..\phi_{\mathbf{p}}, \theta_1, ..., \theta_q)'$. 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 in equation (2) 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 3 *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 Theorem is proven in Appendix B; 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 straightforward to obtain estimates of 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.

2.4 Inversion of Autoregressive Approximations for Estimation of Impulse Response Weights

We now suggest an entirely different alternative approach which has a clearer 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. Again, to make concrete, we note that the $ARFIMA(p, d, q)$ model can be represented by an infinite autoregressive expansion of the form

$$y_t = \sum_{j=1}^{\infty} \kappa_j y_{t-j} + v_t \tag{7}$$

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

$$y_t = \sum_{j=1}^{p_T} \kappa_j^{(p_T)} y_{t-j} + \tilde{v}_t \tag{8}$$

where the order p_T, is obtained by an information criterion. This approach has recently been theoretically analyzed by Poskitt (2005). We denote the least squares estimates of $\kappa_j^{(p_T)}$ obtained by fitting an $AR(p_T)$ model to the data, by $\hat{\kappa}_j^{(p_T)}$. Theorem 5.1 of Poskitt (2005) 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

$$\sum_{j=1}^{P} \left| \kappa_j^{(p_T)} - \kappa_j \right| = o(1), \tag{9}$$

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

$$\sum_{j=1}^{P} \left| \hat{\kappa}_j^{(p_T)} - \kappa_j \right|^2 = o_p(1) \tag{10}$$

which implies that the IRWs can be consistently estimated by fitting an approximating autoregressive 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 $\hat{\kappa}(z) = \sum_{j=1}^{p_T} \hat{\kappa}_j^{(p_T)} z^j$. On the important issue of choosing p_T, it has been shown by Poskitt (2005), via his Theorem 5.3, 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 $(lnT)^2$, which is a valid approximation for finite order $ARFIMA$ processes and even for infinite AR representations. Consequently, the AR approximation has an interpretation of being a semi parametric model.

It should be noted that Theorem 6.1 of Poskitt (2005) shows that the asymptotic distribution of $\hat{\kappa}_j^{(p_T)}$ is nonstandard and non Gaussian, which is clearly quite different for 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, the obvious way to carry out inference in this semi parametric setting is to use the bootstrap as discussed in the next section.

3 Bootstrap Inference

This section considers the adequacy of the asymptotic approximations and the two semi parametric estimators considered earlier. For the IRWs obtained via $LWTSE$ we note that Theorem 3 gives consistency and rates of convergence for the parameters of the $ARFIMA$ model and thereby of the estimated IRWs. However, the theorem does not give an asymptotic distribution; and to provide such a result is non trivial. Hence, the application of the bootstrap appears the obvious approach for carrying out inference on $IRWs$ obtained using $LWTSE$.

It is likely that the small sample performance of asymptotic approximations may be poor given the existing Monte Carlo results for weakly dependent processes. This issue is further explored in the Monte Carlo study of this paper. Furthermore, it is clear that unlike semiparametric autoregressive approximations for weakly dependent processes, such approximations for long memory and the alternative IRW estimator based on $LWTSE$ are not easily amenable to asymptotic inference since the relevant distributions are either non Gaussian or unknown. The above considerations clearly provide strong motivation for a bootstrap approach.

There has been a rapidly increasing literature on the application of the bootstrap to long memory processes; for example see Lazarova (2005) and Poskitt (2006). 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 other contribution of this section of our paper is to provide justification for a semi parametric bootstrap, which can be used for inference on estimated IRW in either the context of a parametric, or a semi parametric model. The work of Poskitt (2005) 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 we assume has the parametric form,

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

On estimating the θ parameters from 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 either the estimated AR or MA 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, to obtain bootstrap estimates $\hat{\theta}^{W*}$ and $\hat{\theta}^*$ respectively. Hence, these estimates can be used to obtain the corresponding bootstrapped estimates of the IRWs $\psi^*_{j,\hat{\theta}^W}$ and $\psi^*_{j,\hat{\theta}}$. This procedure can then be 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 now denoting P_y as the probability law of a random vector y and $d(P_{y1}, P_{y2})$ as the Mallows metric between P_{y1} and P_{y2}, 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 4 *Let Assumptions 1-3 and 4-6 hold. Then, for all* $j = 1, ..., h$

$$d(P_{\sqrt{T}(\psi_{j,\hat{\theta}^W} - \psi_{j,\theta_0})}, P_{\sqrt{T}(\psi_{j,\hat{\theta}^{W*}} - \psi_{j,\hat{\theta}^W})}) = o_p(1) \tag{11}$$

Theorem 5 *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 (18) of Appendix A, is nonsingular. Then, for all* $j = 1, ..., h$

$$d(P_{\sqrt{T}(\psi_{j,\hat{\theta}} - \psi_{j,\theta_0})}, P_{\sqrt{T}(\psi_{j,\hat{\theta}^*} - \psi_{j,\hat{\theta}})}) = o_p(1) \tag{12}$$

Both Theorems are proven in Appendix B. It is now appropriate to discuss a semi parametric sieve type bootstrap, which can be implemented from 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, $\hat{\epsilon}_t = y_t - \sum_{j=1}^{min(p_T,t-1)} \kappa_{j,\hat{\theta}} y_{t-j}$.

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 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 B.

Theorem 6 *Let Assumptions 1-2 hold. Let* $p_T = o\left((\ln T)^a\right)$ *for some* $a > 0$. *Then, for all* $j = 1, ..., h$,

$$d(P_{\hat{\psi}_j^{(p_T)}}, P_{\hat{\psi}_j^{*(p_T)}}) = o_p(1) \tag{13}$$

This Theorem is proved in Appendix B, and does not follow directly from the work of Poskitt (2006), since the statistic being bootstrapped is a function of a statistic that grows with the sample size, rather than being fixed. The validity of an alternative sieve bootstrap whereby the data are generated as above but the statistic being bootstrapped is the parameter vector θ, which can then be used to bootstrap IRWs, follows immediately from Theorem 4.1 and the discussion of Assumption 4 of Poskitt (2006). 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.

4 Monte Carlo Analysis of Estimated IRWs

This section reports the results of a Monte Carlo study of the previously described methods for 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

$$\psi(L) = \theta(L)(1 - L)^{-d}\phi(L)^{-1} \tag{14}$$

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.

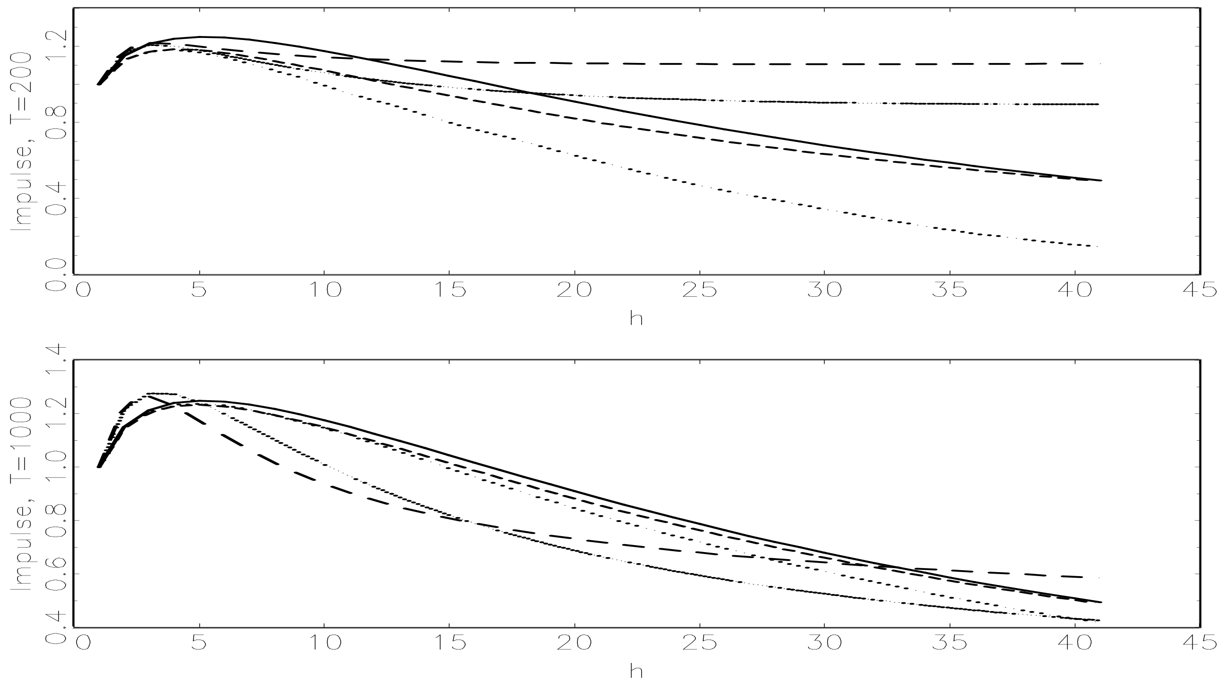
Figures 1 through 4 report results for different $IRWs$ for horizons $k = 1, 2, ...40$ for $ARFIMA(1, d, 0)$ models; and for designs of $d = 0.2, 0.4, 0.6, 0.8$ and $\phi = 0.95$. Although the previous theoretical analysis focused on stationary processes, a considerable amount of applied econometric work has found estimates of d in the range of $0.5 < d < 1$, which implies a non stationary process with finite cumulative IRWs. Hence it seems important to extend the Monte Carlo analysis to consider some mildly nonstationary long memory processes.

The IRW are estimated from AR approximations, MLE, $LWTSE$ for all the cases, and also from using LPW, rather than LW, for the $LWTSE$ estimation, in the stationary cases. The estimated $IRWs$ from using the LW and LPW methods are constructed using a bandwidth of $m = \; T^{0.5}$. [1] For models with $d = 0.2, 0.4$ and quite persistent short memory, Figures 1 and 2 indicate 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 $d = 0.6$, or $d = 0.8$ when there is persistent autocorrelation of $\phi = 0.95$. In this case the $AR(p_T)$ approximation does 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 designs of ($d = 0.6$, $\phi = 0.95$) and ($d = 0.8$, $\phi = 0.95$) in figures 3 and 4 respectively, 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 raises important issues as to whether it is worth an investigator being concerned with the presence of long memory if the investigator's main interest is to only to assess the impact of shocks or innovations on a series. With this in mind, noting that estimates of IRWs based on $LWTSE$ do not perform as well as one might possibly have hoped, and due to space constraints, we decide not to consider them for the Monte Carlo study of confidence interval estimators carried out in

[1] 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 suppressed since they are broadly similar to reported results in this paper.

Figure 1: Impulse Responses: d=0.2, ar=0.95



Notes: Solid Line (—–): True IRF; Long Dashed Line (– – –): Two-Step LW; Dotted Line (. . .): AR Approximation; Short Dashed Line (- - -): MLE; Dense Dotted Line (...): Two-Step LPW.

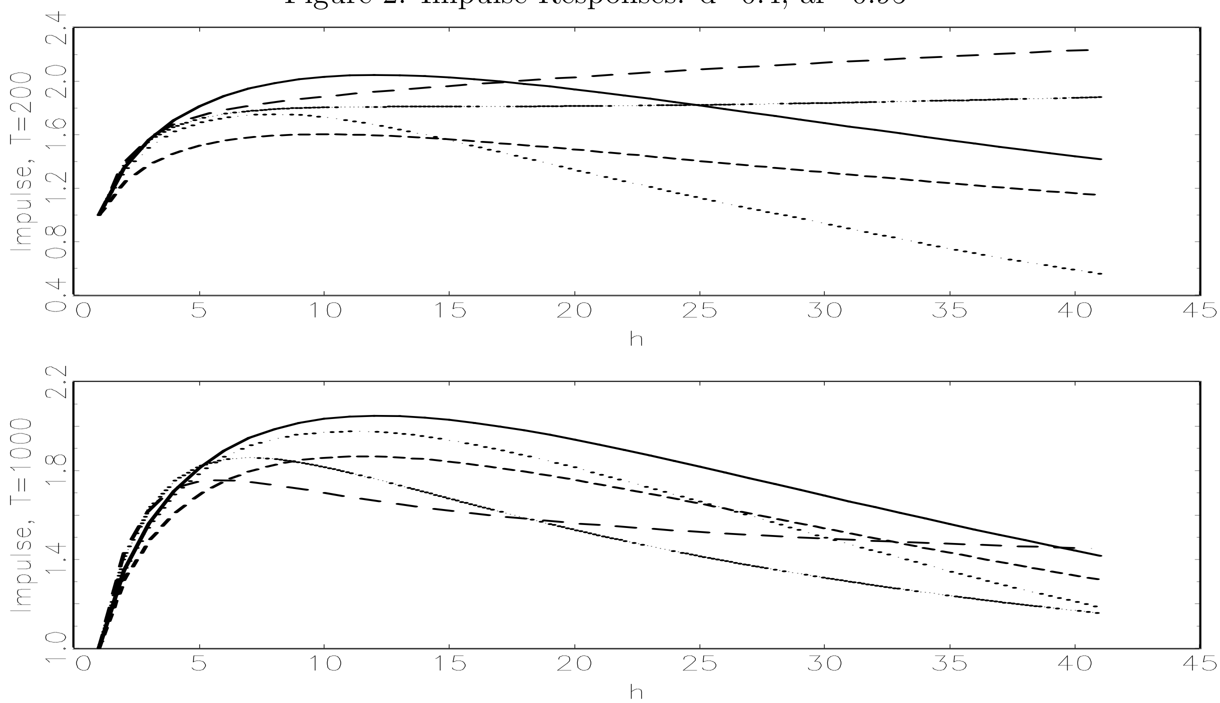
the next section.

5 Monte Carlo Analysis for Confidence Intervals for Estimated IRWs

This subsection investigates the small sample properties of the methods analyzed in the previous sections for constructing confidence intervals for IRWs. The focus is on simple parametric models as the data generating process 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 semiparametric method used to construct confidence intervals. However, our results reported below, still give quite clear indications as to the superiority of the various methods.

The results reported here focus on various $ARFIMA(1, d, 0)$ models. Previous work by Baillie and Kapetanios (2006), Baillie and Kapetanios (2008) and Nielsen and Frederiksen (2004) has suggested that the sole 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 components can be mistaken

Figure 2: Impulse Responses: d=0.4, ar=0.95



Notes: Solid Line (—–): True IRF; Long Dashed Line (– – –): Two-Step LW; Dotted Line (. . .): AR Approximation; Short Dashed Line (- - -): MLE; Dense Dotted Line (...): Two-Step LPW.

for long memory. Hence we consider a parsimonious short memory $AR(1)$ structure, which gives an overall $ARFIMA(1, d, 0)$ model. Thus simple model can be very informative for more general models.

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 $\epsilon_t \sim NID(0, 1)$. The following four different methods were used to construct confidence intervals for the estimated IRWs:

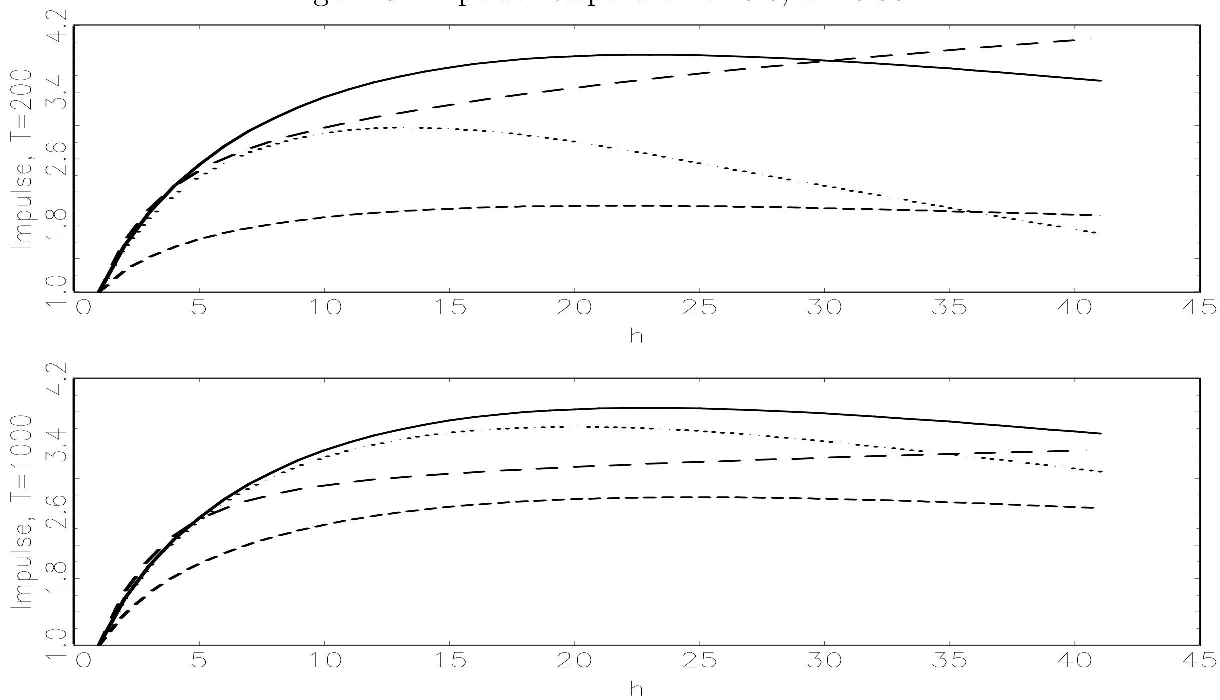
1. The asymptotic approximation to the limiting distribution of the IRWs discussed in Theorem 2.

2. The parametric bootstrap applied to the parametric model, i.e., the method analyzed in Theorem 4.

3. The sieve bootstrap analyzed in Theorem 5, using a lag order equal to $(\ln T)^2$.

4. The sieve bootstrap applied to the parameter vector θ, as discussed immediately below Theorem 5.

Figure 3: Impulse Responses: d=0.6, ar=0.95



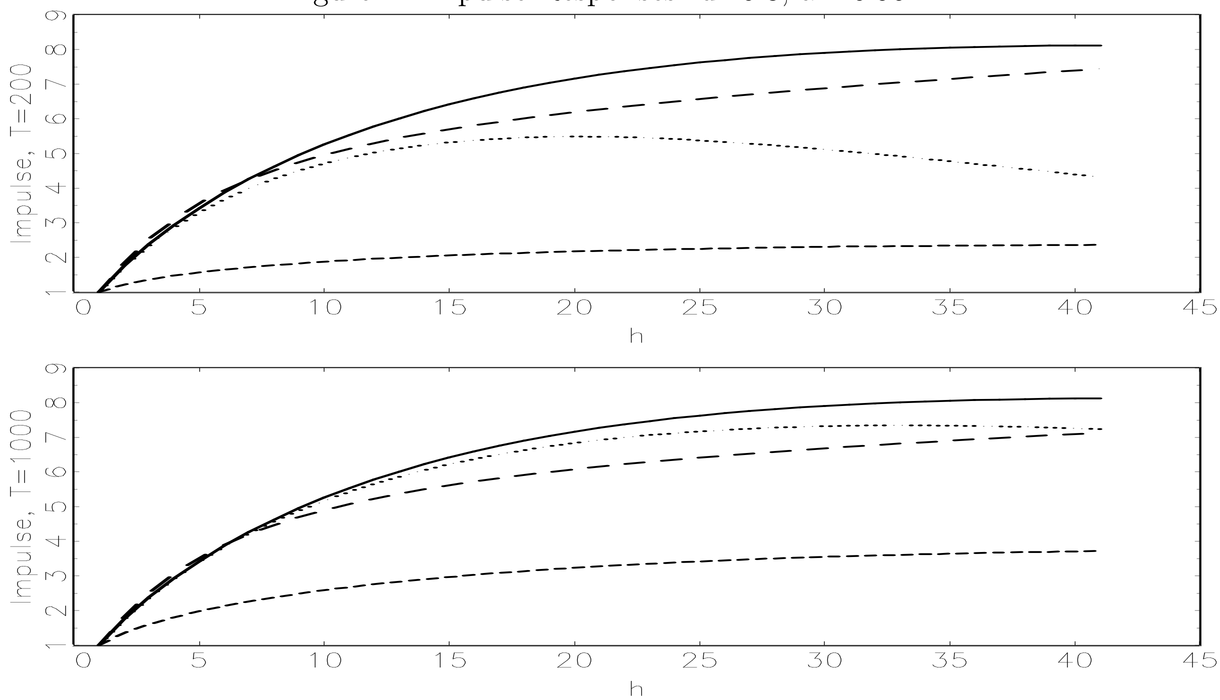
Notes: Solid Line (—–): True IRF; Long Dashed Line (– – –): Two-Step LW; Dotted Line (. . .): AR Approximation; Short Dashed Line (- - -): MLE.

Figures 5 through 8 report the results of 1,000 replications and averages of the resulting 95% and 5% confidence intervals, together with estimates of the true confidence intervals as obtained directly from 1,000 Monte Carlo replications. Note that 399 bootstrap replications are used for each Monte Carlo replication.

There are several interesting aspects of the results of the Monte Carlo analysis. First, the confidence intervals of the IRWs based on parametric estimation. are approximately one half the width of the corresponding confidence intervals of the IRWs from semi parametric estimation. Second, the width of the true confidence intervals and the accuracy with which they are estimated, considerably improves with sample size. Third, the confidence intervals generally become both wider and more imprecisely estimated as the persistence of the AR parameter increases, with the long memory parameter remaining constant. This effect becomes amplified as the short memory persistence is kept constant at a high level but the long memory parameter increases.

In general, all the methods perform very well for the low persistence case of $\phi = 0.5$ and $d = 0.2$. For $T = 400$ and $T = 1000$ the true and average estimated quantiles are indistinguishable. For $T = 200$, the performance is only very slightly worse, across all methods. For the middle case of $\phi = 0.95$ and $d = 0.2$, all methods deteriorate to an extent, and particularly, and

Figure 4: Impulse Responses: d=0.8, ar=0.95



Notes: Solid Line (—–): True IRF; Long Dashed Line (– – –): Two-Step LW; Dotted Line (. . .): AR Approximation; Short Dashed Line (- - -): MLE.

perhaps surprisingly, the parametric bootstrap deteriorates more than the rest. It is, by a small margin, the worst method for this case. The other three methods are very comparable with the semi parametric method performing slightly worse for small samples. Finally, all the methods perform substantially worse with highly persistent processes. Focusing on the larger sample sizes, it can be noted that the worse method appears to be the sieve bootstrap on the parametric model, followed closely by the asymptotic approximation and the parametric bootstrap. Perhaps most surprisingly, the best performer by a small margin is the sieve bootstrap for the semi-parametrically estimated IRWs. Note that the width of the true confidence intervals for this method is comparable to the parametric methods. This finding is surprising since the semi parametric method is slightly better than the parametric methods. This is not the case for the smallest sample size and even then, the methods are eminently comparable in their performance.

Hence the semi-parametric method for estimating IRWs and conducting bootstrap inference appears to be the superior method and is largely robust to the underlying data generating process. This semi parametric method also has comparable performance to parametric methods; and also importantly does not require knowledge of the presence or not of long memory. Hence exactly the same approach can be used to estimate IRWs and to conduct inference if the process was weakly dependent. The only caveat is that the lag order has to be of the order

($\ln T$), which is essential for when the process has long memory. Another point of interest is that the asymptotic approximation is not substantially worse in any respect to the bootstrap approaches. This is in contrast to the small sample results in the existing literature for short memory processes.

6 Empirical Application

6.1 Data and Setup

This section provides an illustration of the preceding ideas and theory to the investigation of the persistence properties of two reasonably large macroeconomic quarterly data sets comprising CPI inflation and real exchange rates. The empirical work builds on our finding that semiparametrically estimated IRWs based on AR approximations, combined with inference obtained using the sieve bootstrap, appears to be a very good strategy for IRW analysis for persistent processes, and processes with short memory.

The CPI inflation data comprises of 26 countries: UK, US, Switzerland, Sweden, Spain, South Africa, Portugal, Norway, New Zealand, Netherlands, Mexico, Malta, Luxembourg, South Korea, Japan, Italy, Greece, Germany, France, Finland, Denmark, Cyprus, Canada, Belgium, Austria and Australia. 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 is 1957Q1 to 2009Q1; and all 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

$$q_{i,t} = s_{i,t} + p_{j,t} - p_{i,t} \tag{15}$$

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.

We use an AR approximation with a lag order of $(\ln T)^2$ to construct IRWs and then carry out a sieve bootstrap as discussed in the previous sections to construct 95% confidence intervals. Furthermore, we calculate half life measures for each of the impulse responses. For the purposes of this paper, we define half life 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 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. We use 399 bootstrap replications.

6.2 Empirical Results

The IRW results are reported in Figure 9 for the CPI inflation data and in Figure 10 for the real exchange rates. Half life measures and their sieve bootstrap confidence intervals are reported in Table 1.

The estimated IRWs for the inflation series are plotted in Figure 9, and as expected have a jagged appearance since they are derived from a high order AR model. However, it is clear from these plots that the inflation series are not very persistent, since the IRWs of most of the series appear to have a clear monotonically declining trend. Only for New Zealand and Spain do the IRWs have a hump shape. The estimated half lives and their confidence intervals are given in Table 1. The half lives are quite low, ranging from 1.5 to about 3 for the majority of cases. Exceptions include Mexico and Italy whose half lives exceed four years.

However, one interesting issue 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 9 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. Chortareas and Kapetanios (2004) have provided a solution to this problem by suggesting alternative measures of half life, whose consideration is beyond the scope of the current paper. It is sufficient for the purposes of this paper to simply note that the standard measure of half life may misrepresent the persistence of CPI inflation.

Plots of the IRWs for the real exchange rate series are presented in Figure 10. We note that the corresponding IRW are much smoother than for the inflation series and suggest that the real exchange rate series are much more 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 8 for Mexico to 34 for South Korea. Again there is the problem of non-monotonicity of some of the IRWs associated with the UK and New Zealand, where initial IRW falls below 0.5 but is above 0.5 at longer horizons. Another interesting feature 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 and significantly more so than CPI inflation.

7 Conclusions

This paper has considered both the issue of estimating impulse response weights (IRWs) from processes that may be strongly dependent and the related issue of constructing confidence intervals for the estimated IRWs. In terms of estimation, we compared the $QMLE$, a two step estimator that uses a semi parametric estimate of the long memory parameter in the first step known as $LWTSE$ and also an estimator from fitting an autoregressive approximation. The simulation evidence finds that the high order AR approximation method works very well for both stationary and non-stationary processes irrespective of the persistence of the short memory component of the data. In general, we conclude that this method is preferable compared with MLE and the $LWTSE$, when its semiparametric nature is taken into account.

The other main focus of the paper concerns the most appropriate method for constructing confidence intervals for the IRWs. We show that the parametric bootstrap is valid under very weak conditions, including non Gaussianity, for making inference on IRW from possibly strongly dependent processes. We also propose, and justify theoretically, a semi-parametric sieve bootstrap based on autoregressive approximations that can be used both for IRWs obtained by autoregressive approximations and more generally. We find that the confidence intervals of IRWs based on the sieve bootstrap generally have very desirable properties and are shown to perform well in a detailed simulation study.

Finally, we consider the methods we develop in an extensive and detailed empirical application on inflation and real exchange rate data.

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 (2006): “Testing for neglected nonlinearity in long memory models,” *Journal of Business and Economic Statistics*, 25, 447–461.

Baillie, R. T., and G. Kapetanios (2008): “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.

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

Chortareas, G., and G. Kapetanios (2004): "How puzzling is the PPP puzzle? An alternative half-life measure of convergence to PPP, " *Queen Mary University of London Working Paper No. 522.*

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.

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.

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.

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

Killian, 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.

Lazarova, S. (2005): "Testing for structural change in regression with long memory processes," *Journal of Econometrics*, 129, 329–372.

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. (19885): "Asymptotic distribution of the moving average coefficients of an estimated vector autoregressive process," *Econometric Theory*, 4, 77–85.

LUTKEPOHL, H. (19885): "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.

PHILLIPS, P.C.B. (1998): "Impulse response and forecast error variance asymptotics in non stationary VARs," *Journal of Econometrics*, 83, 21–56.

POSKITT, D. (2005): "Autoregressive approximation in nonstandard situations: The non-invertible and fractionally integrated case," *Annals of the Institute of Statistical Mathematics,* 59, 697–725.

POSKITT, D. (2006): "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. (2003): "Confidence Intervals for Half-Life Deviations from Purchasing Power Parity," *Journal of Business and Economic Statistics,* 23, 432–442.

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.

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.

Appendix A

This appendix sets out a set of technical regularity conditions that are required for the validity of the results of Hosoya (1997). 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

$$H_j(\theta) = \frac{\partial\left(\int_{-\pi}^{\pi} \log\det f_x(\omega, \theta)\right)}{\partial\theta_j}, j = 1, ..., s,$$

$$h_j(\theta) = \frac{\partial f_x^{-1}(\omega)}{\partial\theta_j}, j = 1, ..., s,$$

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

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

$$W_{ij} = \frac{\partial R_i(\theta)}{\partial\theta_j}, \quad i, j = 1, ..., s, \tag{16}$$

and

$$U_{ij} = 4\pi \int_{-\pi}^{\pi} h_i(\omega, \theta) h_j(\omega, \theta) f_x^2(\omega, \theta) d\omega + \tag{17}$$
$$2\pi \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} h_i(\omega_1, \theta) h_j(\omega_2, \theta) \psi_\theta(e^{i\omega_1})\psi_\theta(e^{-i\omega_1})\psi_\theta(e^{i\omega_2})\psi_\theta(e^{-i\omega_2}) d\omega_1 d\omega_2$$

respectively. Finally, let

$$\Omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \varpi(\omega)\varpi(\omega)' d\omega \tag{18}$$

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.*

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

Assumption 5 *(i) $f_x(\omega)$ 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*

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

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

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

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.

8 Appendix B

Proof of Theorems 1 and 2

Under the assumptions of the Theorems, the results for Theorems 1 and 2 follow immediately from Theorem 2.2 of Hosoya (1997) and Theorem 2 of Robinson (2006), respectively, and the application of the delta method.

Proof of Theorem 3

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 $\pi_j(d) = dX_j^1 + d^2 X_j^2$, 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

$$T^{-1}\sum_{t=1}^{T-k} \hat{u}_t \hat{u}_{t+k} = T^{-1}\sum_{t=1}^{T-k} u_t u_{t+k} + O_p(m^{-1/2}) \tag{19}$$

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 $L(\beta)$. The form of the likelihood may be found in, e.g., (5.6.3) of Hamilton (1994). It is then clear that the likelihood function is differentiable and as long as (19) 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 $L\left(\hat{\beta}(\hat{d}_{LW})\right) = L\left(\beta_0(d_0)\right) + \frac{\partial L}{\partial \beta}\Big|_{\beta=\bar{\beta}}\left(\hat{\beta}(\hat{d}_{LW}) - \beta_0(d_0)\right)$. Hence, the result of the Theorem holds for $ARFIMA(p,d,q)$ models completing the proof.

Proof of Theorems 4 and 5

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 4, i.e. for the conditional sum of squares (CSS) estimator of θ. The proof of Theorem 3 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

$$d(P_\epsilon, P_{\epsilon^*}) = 0$$

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

$$d(P_{\epsilon^*}, P_{\hat{\epsilon}^*}) \to 0 \tag{20}$$

Then, by Lemma 8.5 of Bickel and Freeman (1981), using Ψ as a relevant function, it follows from (20) that

$$d(P_y, P_{\hat{y}^*}) \to 0$$

Then, it immediately follows that

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

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

Proof of Theorem 6

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 (2005). 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, we have that

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

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 wish to show that

$$d\left(P_{\hat{\kappa}^{*,(p_T)}}, P_{\hat{\kappa}^{(p_T)}}\right) \to 0$$

where $d(P_{y1}, P_{y2})$ is the Mallows metric between P_{y1} and P_{y2}. It follows that

$$d\left(P_{\hat{\kappa}^{*,(p_T)}}, P_{\hat{\kappa}(p_T)}\right) \leq E\left[E^*\left(\left\|\hat{\kappa}^{*,(p_T)} - \hat{\kappa}(p_T)\right\|^2\right)\right] \leq$$

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

Looking at each of the two terms above we have

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

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

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

But

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

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

Further,

$$E\left[E^*\left(\left\|X^{*,(p_T)\prime}x^* - X^{(p_T)\prime}x\right\|^2\right)\right] \leq p_T \sup_{1\leq i\leq p_T} E\left[E^*\left(\left\|\left\{X^{*,(p_T)\prime}x^*\right\}_{i1} - \left\{X^{(p_T)\prime}x\right\}_{i1}\right\|^2\right)\right] \leq$$

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

But by the proof of Theorem 4.2 of Poskitt (2005) 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

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

So, overall

$$d\left(P_{\hat{\kappa}^{*,(p_T)}}, P_{\hat{\kappa}(p_T)}\right) = O\left(p_T^{21/2}\left(\frac{\log T}{T}\right)^{1-2d}\right)$$

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

$$d\left(P_{\hat{\kappa}^{*,(p_T)}}, P_{\hat{\kappa}(p_T)}\right) = O\left(\frac{\log T^{\frac{21a}{2}+1-2d}}{T^{1-2d}}\right) = o(1)$$

proving that $\hat{\kappa}^{*,(p_T)}$ has the same probability law as $\hat{\kappa}^{(p_T)}$. The result for the impulse responses follows by noting that they are continuous functions of $\hat{\kappa}^{(p_T)}$.

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

CPI Inflation			
Country	Half Life	5% quantile	95% quantile
UK	2.994	1.855	3.384
US	2.364	2.111	2.560
Switzerland	1.768	1.632	1.965
Sweden	1.651	1.538	1.766
Spain	1.651	1.548	1.768
South Africa	1.837	1.666	2.062
Portugal	1.702	1.586	1.867
Norway	1.623	1.527	1.730
New Zealand	1.927	1.733	2.744
Netherlands	1.715	1.588	1.885
Mexico	5.836	2.799	7.589
Malta	1.647	1.539	1.773
Luxemburg	1.744	1.617	1.868
South Korea	2.029	1.827	2.295
Japan	1.710	1.598	1.851
Italy	4.613	1.888	5.439
Greece	1.610	1.527	1.697
Germany	1.738	1.611	1.913
France	2.244	1.808	2.956
Finland	1.977	1.765	2.533
Denmark	1.582	1.501	1.673
Cyprus	1.529	1.462	1.601
Canada	1.947	1.733	2.303
Belgium	1.966	1.754	2.979
Austria	1.584	1.501	1.678
Australia	1.729	1.600	1.862
Real Exchange Rates			
Country	Half Life	5% quantile	95% quantile
UK	13.061	6.620	16.163
Switzerland	24.579	10.693	28.209
South Africa	21.575	6.964	26.238
Norway	23.114	6.970	25.438
New Zealand	10.997	7.413	13.444
Mexico	8.853	6.306	10.925
South Korea	34.692	8.327	40.500
Japan	16.613	9.441	35.228
Canada	24.164	12.797	30.083
Australia	15.776	8.274	21.956

Figure 5: Monte Carlo Results: Standard Errors for Impulse Responses: Asymptotic Approximation

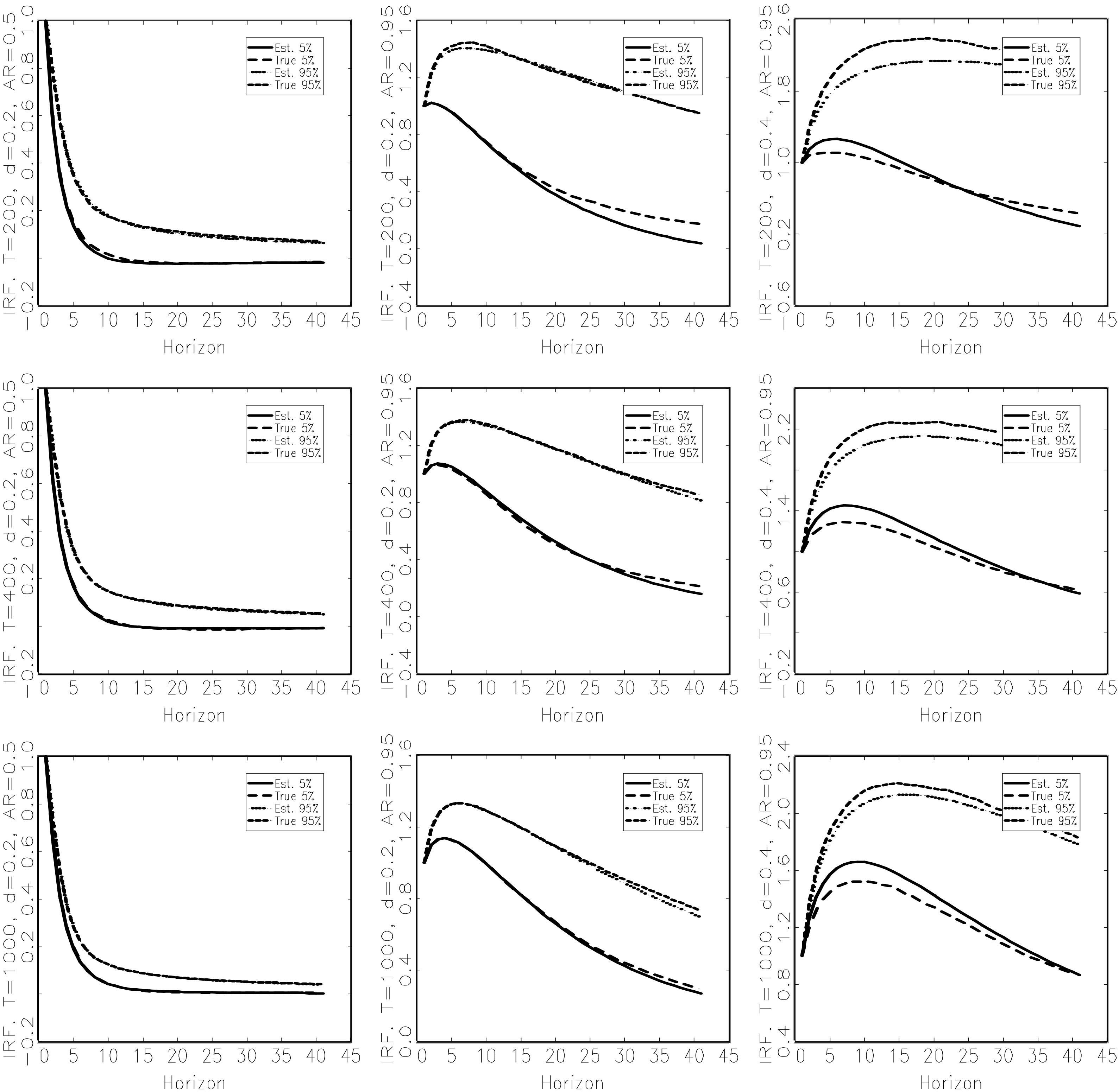


Figure 6: Monte Carlo Results: Standard Errors for Impulse Responses: Parametric Bootstrap on Parametric Model

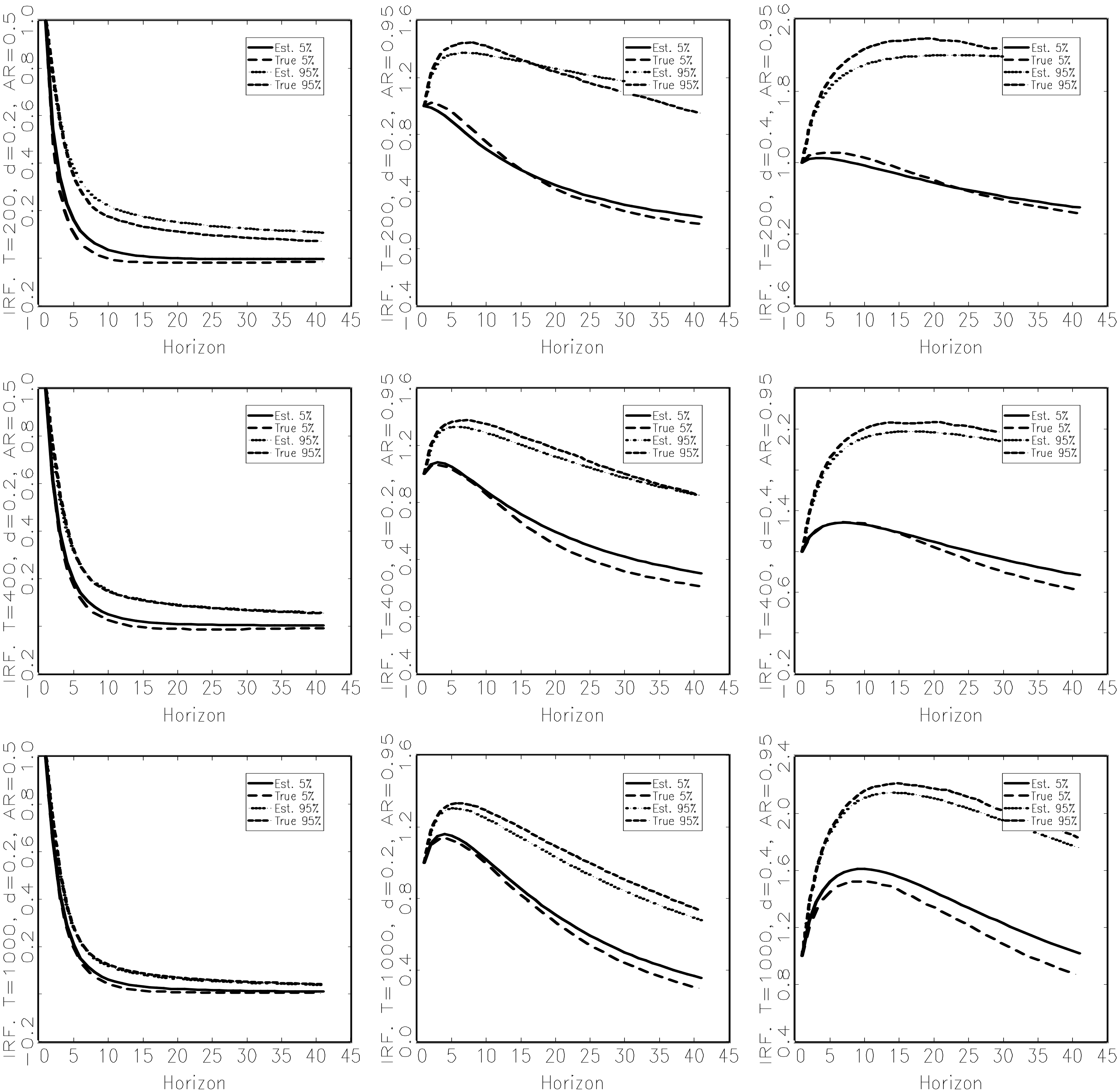


Figure 7: Monte Carlo Results: Standard Errors for Impulse Responses: Sieve Bootstrap for Long AR

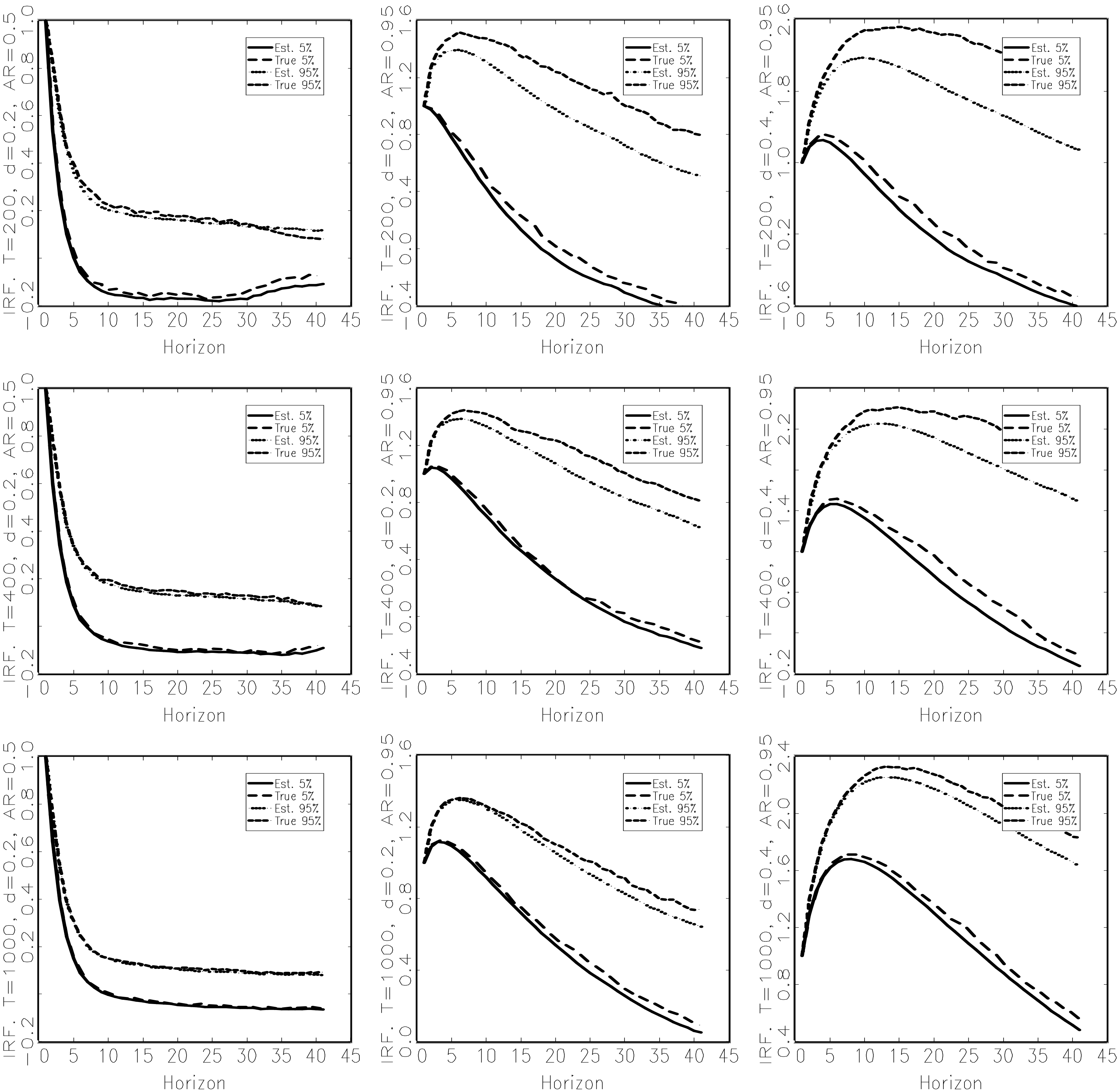


Figure 8: Monte Carlo Results: Standard Errors for Impulse Responses: Sieve Bootstrap on Parametric Model

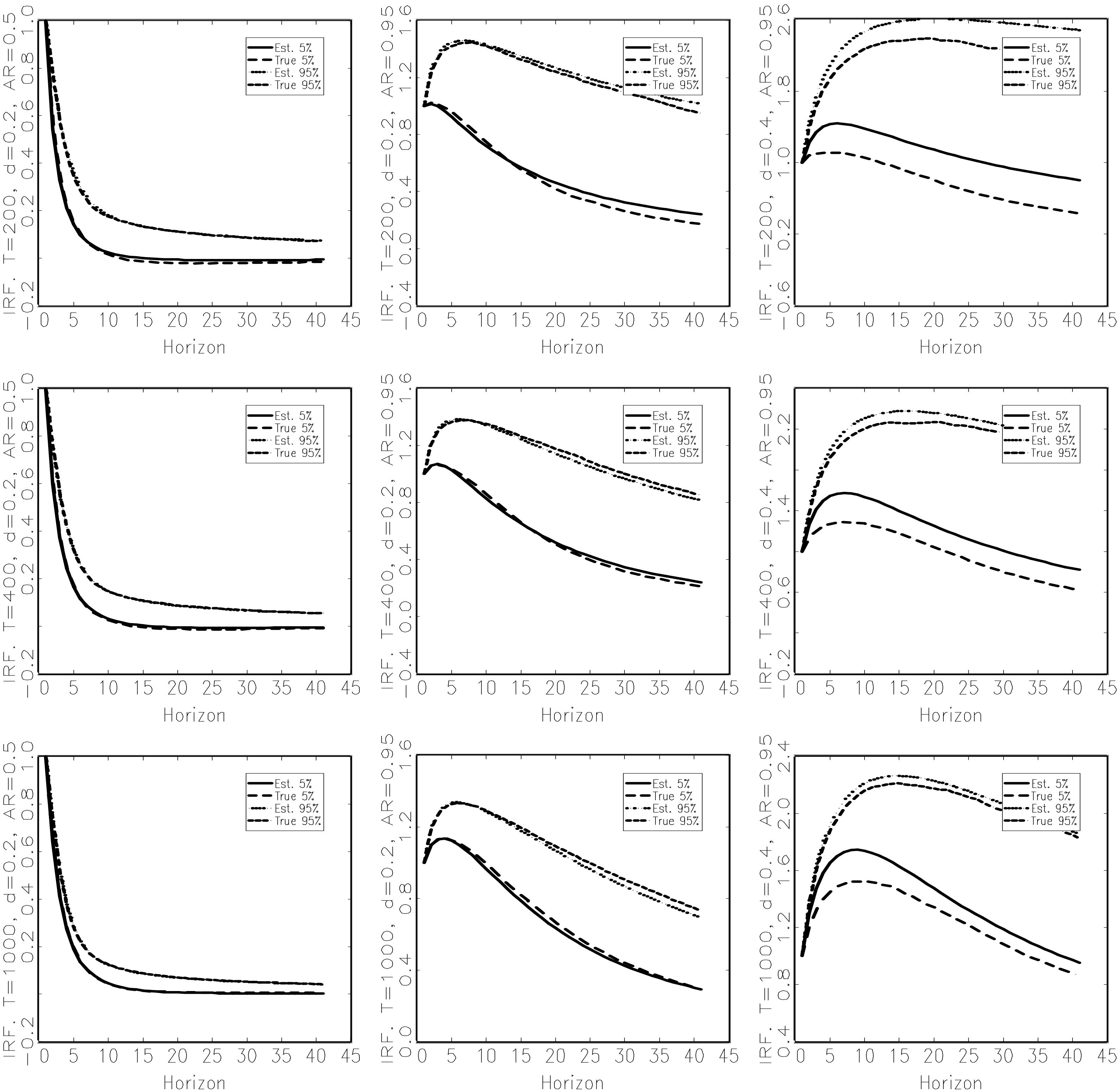


Figure 9: Empirical Results: Impulse Responses for CPI Inflation

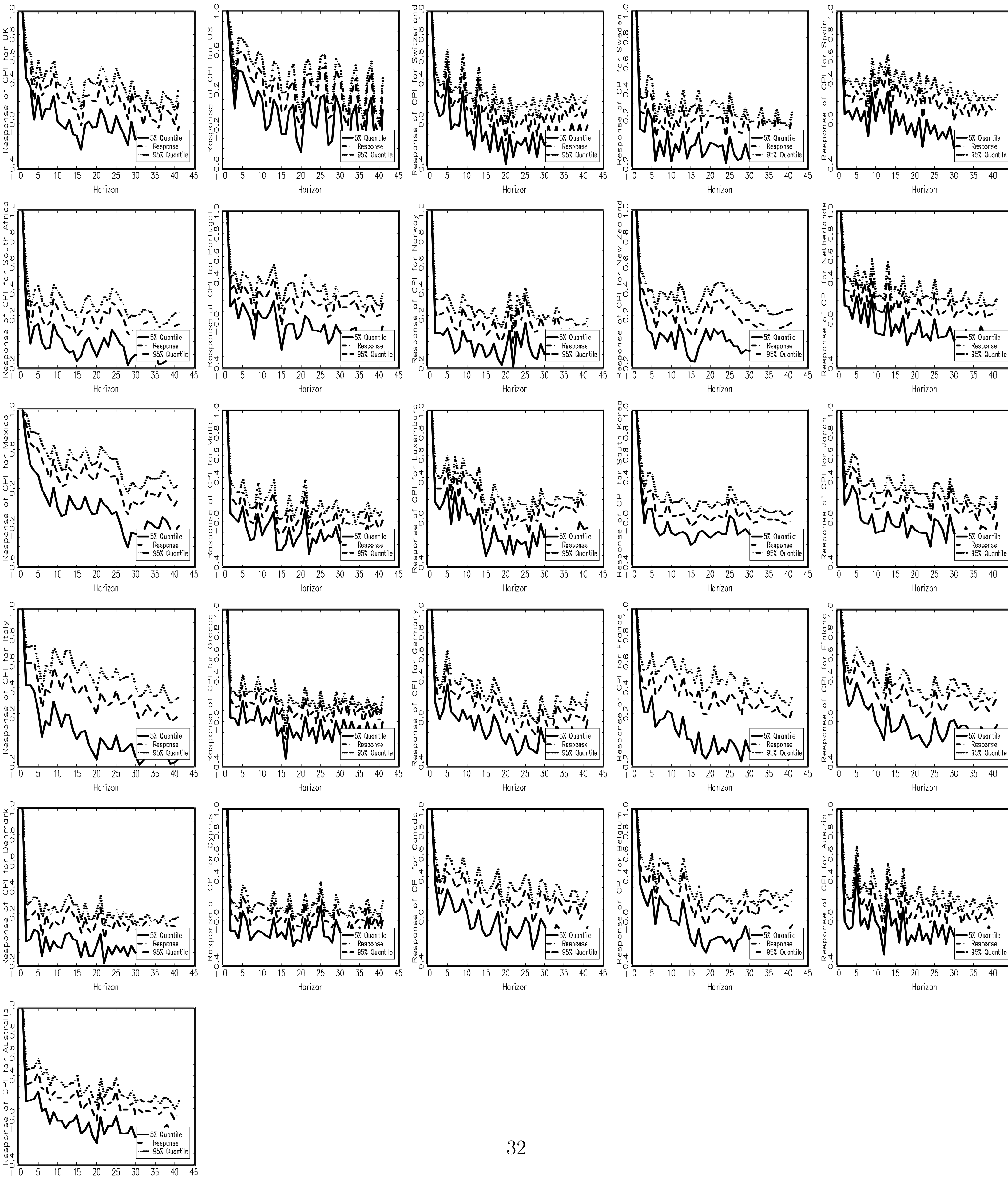


Figure 10: Empirical Results: Impulse Responses for Real Exchange Rates