Confidence Intervals For Impulse Response Weights From Strongly Dependent Processes

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

George Kapetanios
Department of Economics, Queen Mary University of London

March 11, 2010

Abstract

This paper considers the problem of estimating impulse response (IR)s from processes that are possibly strongly dependent and the related issue of constructing confidence intervals for the estimated IRs. We show that the parametric bootstrap is valid under very weak conditions, including non Gaussianity for making inference on IR from strongly dependent processes. Further, we propose, and justify theoretically, a semi-parametric sieve bootstrap based on autoregressive approximations. We find that the sieve bootstrap generally has very desirable properties and is shown to perform extremely well in a detailed simulation study.

Key Words: Strong Dependence, Impulse Response Analysis, Bootstrap, Autoregressive approximations.

JEL Codes: C22, C12.

1 Introduction

Following the influential article of Sims (1980) and the development of structural macro models, Impulse Response (IR) analysis has come to be regarded as an essential part of the analysis of dynamic models used in macroeconomics and finance. The analysis is most commonly employed on the conditional mean in macroeconomic models and generally on quantities reflecting the second moments such as realized volatility, absolute returns, or conditional variances in many finance applications. The use of IR analysis has now extended to providing information on the validation of dynamic stochastic general equilibrium ($DSGE$) models; e.g. see Kapetanios, Pagan and Scott (2007). IR analysis is also widely used in the determination of the degree of persistence in macroeconomic time series; see Campbell (1987). The latter example has received considerable attention both in applied macroeconomics and time series econometrics, with the development of summary statistics, such as half life measures arising out of IRs.

While point estimates of IRs are relatively unambiguous from an estimated dynamic model, there is a recurrent issue concerning the most appropriate method for the construction of appropriate confidence intervals around the estimated IRs; see Sims (1986) and Phillips (1998). Most work in the area has focused on weakly dependent processes, where the workhorse model has been the autoregressive model and the Vector Autoregression (VAR). A major finding of existing work is that confidence intervals based on asymptotic approximations provide a poor guide to true finite-sample confidence intervals. As a result, bootstrap methods have been repeatedly advocated for the construction of meaningful confidence intervals. Leading examples of such work is Kilian (1998a), Kilian (1998b), Kilian (1999) and Chang and Kilian (2000).

As noted above, all previous work on evaluating alternative methods for constructing confidence intervals for estimated IRs has focused on weakly dependent processes. However, a lot of recent interest has focused on the estimation and interpretation of models involving strongly dependent processes. For example, Baillie, Chung and Tieslau (1996) report estimates of IRs for the conditional mean of inflation which are derived from a parametric long memory model. Also, a particularly active area of research has concerned the dynamic properties of various measures of volatility. For example, Ray and Tsay (2000) consider squared equity market returns, while Andersen et al (2001) and Andersen et al (2003) have found that realized volatility on currency and equity markets tend to be well described as being close to fractional white noise, which is a pure long memory process, and also a strongly dependent process. These papers also report IR analysis from the estimated strongly dependent process from the realized volatility series. Hence, it is of great practical relevance to assess the most appropriate methods for forming confidence intervals for the estimated IRs from such data, which are also typically highly non Gaussian. A further issue investigated in this paper concerns the extent to which asymptotic approximations to finite sample confidence intervals by the delta method, are also valid for strongly dependent processes. Moreover, it is of interest to investigate what type of bootstrap can provide useful approximations to finite sample confidence intervals in these cases, both theoretically, through Monte Carlo simulations and empirically.

This paper addresses some of the outstanding issues mentioned above and begins by considering parametric models for strongly dependent processes, where the most popular example is 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 weak assumptions; and very importantly manages to avoid assuming Gaussianity, which is a very common assumption in this literature. We then consider a generic semi-parametric sieve bootstrap based on an Autoregressive (AR) approximation of the unknown data generating mechanism, which underlies a wide class of strongly dependent processes. The semi-parametric nature of the approximation and the suggested use of the bootstrap links our work to the influential study of Gallant, Rossi and Tauchen (1993), who were among the first to consider both nonparametric approximations and the bootstrap in IR analysis. Under mild assumptions we show the validity of IR analysis based on this AR approximation and the validity of bootstrap inference on the resulting IRs. The theoretical work in this paper on the semi-parametric bootstrap provides a complement to the work of Baillie and Kapetanios (2009), who appear to be the first authors to suggest the use of AR approximations as a unifying framework for constructing IR for both short and long memory processes. This paper complements and extends Baillie and Kapetanios (2009) by providing some practically important theoretical results for inference, on estimated IRs.

Our theoretical results are accompanied with an extensive Monte Carlo study which examines the relative benefits of asymptotic versus bootstrap inference and parametric versus semi-parametric bootstrap inference. Our results suggest that the sieve bootstrap has a number of benefits. Finally, we provide an empirical example of the approach 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 IRs. The rest of this paper is structured as follows; section 2 outlines the assumptions and basic set up of the models and presents some basic results. Then, section 3 describes the bootstrap procedures and their theoretical properties; while sections 4 and 5 present the Monte Carlo study and the empirical results respectively. There is also a conclusions section 6, followed by an appendix which contains the various proofs.

2 Theoretical Setup

Consider a 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 x_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 about the error term and 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.*

The class of processes, delineated by (1) is very wide and includes all linear processes considered in the existing literature, and encompasses the leading case of $ARFIMA(p, d, q)$ processes where $\tilde{\psi}(z) = \phi(z)^{-1}\varphi_1(z)$, where $\phi(z) = \sum_{j=0}^{p} \phi_j z^j$ and $\varphi(z) = \sum_{j=0}^{q} \varphi_j z^j$. For the purposes of analyzing both parametric and semi parametric bootstrap inference on IR, we now introduce a parametric representation associated with the above setup that is more general but encompasses $ARFIMA$ processes. The ψ_j parameters are 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 the analysis is in a parametric setting. The notation ψ_j is used for both the general discussion and also for the semi-parametric setting. The following indentifiability 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 $\psi_\theta(z) = \tilde{\psi}_\theta(z)/(1-z)^d$*

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. The standard method 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 a general Quasi Maximum Likelihood Estimator, ($QMLE$), which has been previously analyzed in a very general context by Hosoya (1997). Although the properties of this $QMLE$ can be most elegantly characterized in the frequency domain, it is also asymptotically equivalent to an estimator obtained by minimizing the conditional sum of squares,

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

See Theorem 2 of Robinson (2006). 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 we denote the frequency domain based estimator of Hosoya (1997) by $\hat{\theta}^W$.

The first results of the paper relate to the asymptotic distributions of $\psi_{j,\hat{\theta}}$ and $\psi_{j,\hat{\theta}^W}$. For this theorem, we require a set of technical regularity conditions, which are needed to justify the asymptotic normality of $\hat{\theta}^W$. These conditions are given in Appendix A.

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 (9) and (10) 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 (11) 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.

The proofs of these theorems are given in Appendix B. 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 of the limiting distribution of IRs from stationary and invertible vector $ARMA$ models where the equivalent D_j matrices can be easily 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 IRs 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 IRs in such processes and hence is the focus in this paper.

A similar approach for carrying out asymptotic inference on a semi parametric estimate of ψ_j could be based on the use of autoregressive approximations. In particular, an estimate of the ψ_j, denoted by $\hat{\psi}_j$, can be obtained by simply fitting an $AR(p_T)$ model to the y_t series. It is necessary for the p_T to grow at an appropriate rate, which for long memory process is of the order $(\ln T)^a$, for some $a > 0$. Then the estimated coefficients from the $AR(p_T)$ estimation are $\hat{\psi}(z) = \sum_{j=1}^{\infty} \hat{\psi}_j z^j = \hat{\kappa}^{-1}(z)$, where $\hat{\kappa}(z) = \sum_{j=1}^{p_T} \hat{\kappa}_j^{(p_T)} z^j$ and $\hat{\kappa}_j^{(p_T)}$ are the estimated coefficients from the $AR(p_T)$ estimation. This approach has been advocated by Poskitt (2005), who showed the consistency of $\hat{\kappa}_j^{(p_T)}$ and derived their rates of convergence. The same approach has been advocated for IR estimation by Baillie and Kapetanios (2009), who showed that it performed exceptionally well in small sample for both stationary and non-stationary long memory processes. Baillie and Kapetanios (2009) also discussed some results for this approach in the case of non-stationary long memory processes. As discussed in Theorem 6.1 of Poskitt (2005), the asymptotic distribution of $\hat{\kappa}_j^{(p_T)}$ is nonstandard and non Gaussian. As a result, inference based on asymptotic considerations is problematic. One way to carry out inference in this semi parametric setting is to use the bootstrap as discussed in the next section. It is worth noting that the above non Gaussian result contrasts with existing work on the asymptotic distributions of autoregressive coefficient estimates or autocorrelations, for weakly dependent processes, which have been proven to be asymptotically normal, see, e.g., Lewis and Reinsel (1985).

3 Bootstrap Inference

The previous section provided theory for the asymptotic distribution of IRs for parametric long memory models. However, it is likely that their small sample performance 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 semi-parametric autoregressive approximations for weakly dependent processes, such approximations for long memory are not easily amenable to asymptotic inference since the relevant distributions are non Gaussian. 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; see Lazarova (2005), Hidalgo (2003), Andrews and Lieberman (2006) and Poskitt (2006). Some of this work is not necessarily applicable to our general framework. For example, the work of Hidalgo (2003) relates to regression models involving long memory processes. Such a framework is encompassed by our approach, or relatively straightforward extensions of it such as assuming that y_t is a vector rather than a scalar process.

The work of Andrews and Lieberman (2006) and Poskitt (2006) are closest to our approach for both the parametric and semi parametric methods respectively. Andrews and Lieberman (2006) provides 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. Andrews and Lieberman (2006) conjecture that both forms of the parametric bootstrap are valid for non-Gaussian processes and that higher order corrections will not be valid for such processes. However, any formal analysis of these issues is not currently available and one contribution of this paper is to examine the validity of the parametric bootstrap for non Gaussian processes for both the parametric estimators introduced in the previous section. This is relatively straightforward given the existing work of Andrews and Lieberman (2006), and also that of Hosoya (1997), which establishes the validity of the 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 IR in either the context of a parametric, or a semi parametric model. The work of Poskitt (2005) is instrumental for these derivations.

It is convenient to begin by considering the parametric bootstrap in the context of 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. On further assuming a parametric form for this approximation,

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

once θ is estimated using one 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 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 be used to estimate by MLE or the minimization of the conditional sum of squares, to obtain bootstrap estimates of θ, denoted by $\hat{\theta}^{W*}$ and $\hat{\theta}^*$ respectively. Hence, these estimates can be used to obtain the bootstrapped estimates of the IRs $\psi^*_{j,\hat{\theta}^W}$ or $\psi^*_{j,\hat{\theta}}$. On replicating the above procedure, B times the B estimates of the IR and the empirical distribution of them, as $B \to \infty$, can be used for inference on the IRs. Next, we denote 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}. Then we have the following Theorems on the validity of this form of parametric bootstrap for both MLE and the minimization of CSS.

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

$$d(P_{\sqrt{T}(\psi_{j,\hat{\theta}^W} - \psi_{j,\theta_0})}, P_{\sqrt{T}(\psi_{j,\hat{\theta}^{W*}} - \psi_{j,\hat{\theta}^W})}) = o(1) \tag{5}$$

Theorem 4 *Under the assumptions 1(i) and 2-3; and assuming further that* ϵ_t *is an i.i.d. sequence, that* $\sum_{j=1}^{\infty} \sup_\theta |\tilde{\psi}_{j,\theta}| < \infty$*, that* $\tilde{\psi}_{j,\theta}$ *are twice continuously differentiable, with respect to* θ*, and that* Ω*, defined in (11) of Appendix A, is nonsingular. Then, for all* $j = 1, ..., h$

$$d(P_{\sqrt{T}(\psi_{j,\hat{\theta}} - \psi_{j,\theta_0})}, P_{\sqrt{T}(\psi_{j,\hat{\theta}^*} - \psi_{j,\hat{\theta}})}) = o(1) \tag{6}$$

Both Theorems are proven in Appendix B. It is now appropriate to discuss a semi parametric sieve type bootstrap. Given the results of Poskitt (2006), it is reasonable to expect that a sieve bootstrap can be useful for inference with estimated IRs. We propose the following strategy for the implementation of the bootstrap:

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, $\hat{\epsilon}_t = y_t - \sum_{j=1}^{min(p_T,t-1)} \kappa_{j,\hat{\theta}} y_{t-j}$, from the $AR(p_T)$ estimation.

2. Invert $\hat{\kappa}^{(p_T)}(z) = \sum_{j=1}^{p_T} \hat{\kappa}_j^{(p_T)} z^j$ to obtain estimates of the IRs 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 $(x_1^*, ..., x_T^*)'$ 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 IRs to construct an empirical distribution of the IRs.

The following theorem justifies the above bootstrap approach.

Theorem 5 *Let Assumptions 1-2 hold. Let $p_T = o\left((\ln T)^a\right)$ for some $a > 0$. Then, for all $j = 1, ..., h$,*

$$d(P_{\hat{\psi}_j^{(p_T)}}, P_{\hat{\psi}_j^{*(p_T)}}) = o(1) \tag{7}$$

This Theorem is proved in Appendix B. Note that this theorem does not follow directly from the work of Poskitt (2006), since the statistic being bootstrapped is a function of a statistic that does not have a fixed dimension, but its dimension grows with the sample size. 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 IRs, follows immediately from Theorem 4.1 and the discussion of Assumption 4 of Poskitt (2006). In general the approach we advocate above has close links with the work of Gallant, Rossi and Tauchen (1993), who were among the first, both in suggesting nonparametric approximations for constructing IRs and in using the bootstrap for carrying out statistical inference on them.

4 Monte Carlo Analysis

This section investigates the small sample properties of all the methods analyzed in the previous sections for constructing confidence intervals for IRs. The focus is on simple parametric models as the data generating processes 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 single 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.

The results reported here focus on various $ARFIMA(1,d,0)$ models. Previous work by Baillie and Kapetanios (2006), Baillie and Kapetanios (2008), Baillie and Kapetanios (2009) 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 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 emphasis is on the minimization of the CSS, which is a popular estimator for practical applications and is asymptotically equivalent to MLE under Gaussianity. This estimator essentially neglects the effects of starting values, which are asymptotically negligible for stationary processes.

The following four different methods were used to construct confidence intervals for the estimated IRs:

1. The asymptotic approximation to the limiting distribution of the IRs 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.

Figures 1 through 4 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.

The results of the Monte Carlo are very interesting, and indicate that the confidence intervals of the IRs based on parametric estimation are approximately one half the width of the corresponding confidence intervals of the IRs from semi parametric estimation. The second interesting finding confirms that as expected the width of the true confidence intervals and the accuracy with which they are estimated, considerably improves with sample size. A third general finding is that the confidence intervals get both wider and more imprecisely estimated as the persistence of the AR component 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 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. On 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 IRs. 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.

The above findings of this Monte Carlo indicate that the semi-parametric method for estimating IRs 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 IRs 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.

5 Empirical Application

5.1 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 semi-parametrically estimated IRs based on AR approximations, combined with inference obtained using the sieve bootstrap, appears to be the best strategy for IR analysis for persistent processes, and processes with short memory.

The CPI inflation data comprises 26 countries of: 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. While 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

$$q_{i,t} = s_{i,t} + p_{j,t} - p_{i.t} \tag{8}$$

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 IRs 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.

5.2 Empirical Results

The IR results are reported in Figure 5 for the CPI inflation data and in Figure 6 for the real exchange rates. Half life measures and their sieve bootstrap confidence intervals are reported in Tables 1 and 2.

The estimated IRs for the inflation series are plotted in Figure 5, 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 IRs of most of the series appear to have a clear monotonically declining trend. Only for New Zealand and Spain do the IRs 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 IRs oscillate, rather than monotonically decline, it is possible that the IR will fall below half its original value (which is unity in our normalization), 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 IRs in Figure 5 suggests that in a number of cases, including the US, Switzerland and Spain, the oscillatory nature of the IR 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 understate the persistence of CPI inflation.

Plots of the IRs for the real exchange rate series are presented in Figure 6. We note that the corresponding IR 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 IRs associated with the UK and New Zealand, where an initial IR falls below 0.5 but is above 0.5 at longer horizons. Another interesting feature is that the IR exceeds 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 IR analysis. The empirical findings confirm that real exchange rates are very persistent and significantly more so than CPI inflation.

6 Conclusions

This paper has considered the issue of the most appropriate methodology for estimating impulse response (IR)s from processes are possibly, but not necessarily, strongly dependent and also the issue of constructing confidence intervals for such estimated IRs. Our results have extended known results to include strongly dependent processes; and we show the parametric bootstrap is valid under very weak assumptions. We consider four main alternative methods for conducting inference in these situations; namely (i) using a standard asymptotic approximation to the limiting distribution of the IRs, (ii) the parametric bootstrap applied to the parametric model, (iii) the sieve bootstrap with a lag order equal to $(\ln T)^2$, and (iv) the sieve bootstrap applied to the estimated model's parameter vector.

The sieve bootstrap based on the AR approximation turns out to have very desirable properties for a wide class of strongly dependent processes such as $ARFIMA$ and performs extremely well in a detailed simulation experiment. The methodology is illustrated with examples of inflation and real exchange rates. The proposed new approach promises to be an important development for estimating and conducting inference on impulse responses for a wide variety of weakly and strongly dependent processes.

References

Andersen, T., Bollerslev, T., Diebold, F. X. and Labys (2001): The distribution of realized exchange rate volatilty, *Journal of the American Statistical Association*, 96, 42-55.

Andersen, T., Bollerslev, T., Diebold, F. X. and Labys (2003): Modelling and forecasting realised volatilty, *Econometrica*, 71, 579-625.

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., C.-F. Chung and M. A. Tieslau (1996): "Analysing inflation by the fractionally integrated ARFIMA-GARCH model," *Journal of Applied Econometrics, 11, 23-40.*

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.

Baillie, R. T., and G. Kapetanios (2009): "Semi parametric estimation of long memory: comparisons and some attractive alternatives, " *Paper presented at conference in honor of Peter Phillip's Sixtieth Birthday at Singapore Management University*,

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

Campbell, J.Y. and N.G. Mankiw (1987): "Are output fluctuations transitory?," *Quarterly Journal of Economics*, 102, 857–880.

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.*

Gallant, A. R., Rossi, P. E. and G. Tauchen (1993): "Nonlinear dynamic structures," *Econometrica*, 61, 871–907.

Hidalgo, J. (2003): "An alternative bootstrap to moving blocks for time series regression models," *Journal of Econometrics*, 117, 369–399.

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

Kapetanios, G., Pagan, A.R., and A. Scott (2007): "Making a match: Combining theory and evidence in policy-oriented macroeconomic modeling," *Journal of Econometrics*, 136, 565–594.

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.

Ray, B. and R. Tsay (2000): "Long-range depenence in daily stock volatilities," *Journal of Business and Economic Statistics, 18, 254–262.*

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.

Sims, C.A. (1980): "Macroeconomics and reality," *Econometrica,* 48, 1–48.

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

Appendix A

This appendix sets out a set of technical regularity conditions needed in the main paper for the results of Hosoya (1997) to hold. We first define a number of terms. Let $Q^{\epsilon}(\omega_1, \omega_2, \omega_3)$ denote the fourth order spectral density of ϵ_t, defined as

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

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

$$W_{ij} = \frac{\partial R_i(\theta)}{\partial \theta_j}, \quad i, j = 1, ..., s, \tag{9}$$

and

$$U_{ij} = 4\pi \int_{-\pi}^{\pi} h_i(\omega, \theta) h_j(\omega, \theta) f_x^2(\omega, \theta) d\omega + \tag{10}$$
$$2\pi \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h_i(\omega_1, \theta) h_j(\omega_2, \theta) \psi_\theta(e^{i\omega_1}) \psi_\theta(e^{-i\omega_1}) \psi_\theta(e^{i\omega_2}) \psi_\theta(e^{-i\omega_2}) d\omega_1 d\omega_2$$

respectively. Finally, let

$$\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varpi(\omega) \varpi(\omega)' d\omega \tag{11}$$

where

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

Next we present the relevant technical regularity conditions:

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 $\tilde{h}^i_j(\omega)$ and $\bar{h}^i_j(\omega)$ such that for all sufficiently small r and for all j, $\tilde{h}^l_j(\omega) \leq h_j(\omega, \theta) \leq \bar{h}^l_j(\omega)$ if $\theta \in U^{(l)}(r)$. Also,*

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

and

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

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

We note a number of 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 funtion approach taken in that paper.

7 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 Theorems 3 and 4

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. resample with replacement of $\hat{\epsilon}_t$ and $\hat{\epsilon}_t$ 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,

$$d(P_{\epsilon^*}, P_{\hat{\epsilon}^*}) \to 0 \tag{12}$$

Then, by Lemma 8.5 of Bickel and Freeman (1981), using Ψ as a relevant function, it follows from (12) 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 asymptotically a normal law. The result follows immediately by noting that impulse responses are, by assumption, a continuous function of the model parameters.

Proof of Theorem 5

Next, we wish to prove that the sieve boostrap 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 $X_t^{(p_T)} = (x_{t-1},..., x_{t-p_T})'$, $X^{(p_T)} = (X_{p_T+1}^{(p_T)}, ..., X_T^{(p_T)})'$, $x = (x_{p_T+1}, ..., x_T)'$ 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

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

We have 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

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

Table 2: Half-Life Estimates, 5% quantiles and 95% quantiles for 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 1: Monte Carlo Results: Standard Errors for Impulse Responses: Asymptotic Approximation

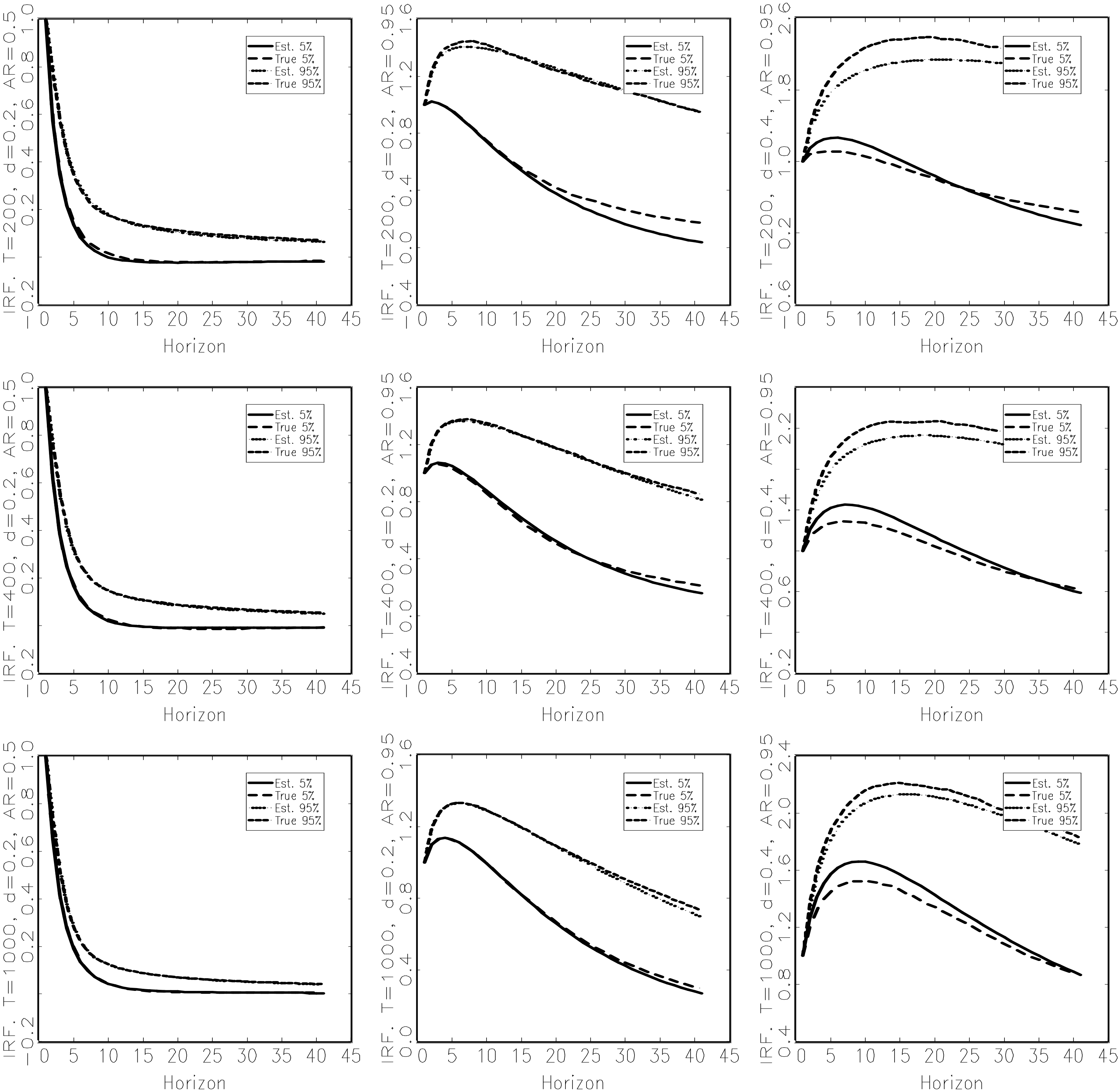

Figure 2: Monte Carlo Results: Standard Errors for Impulse Responses: Parametric Bootstrap on Parametric Model

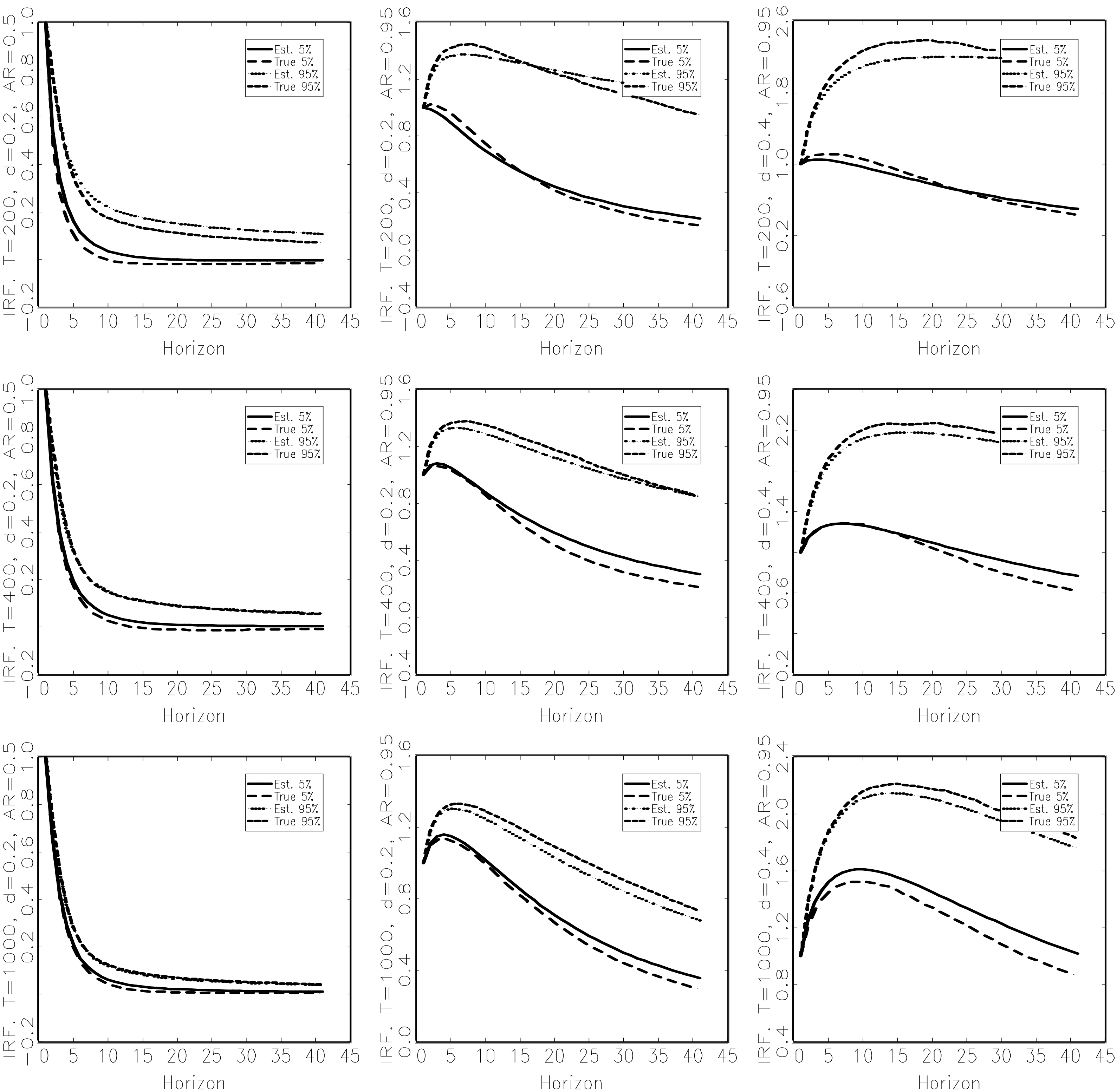

Figure 3: Monte Carlo Results: Standard Errors for Impulse Responses: Sieve Bootstrap for Long AR

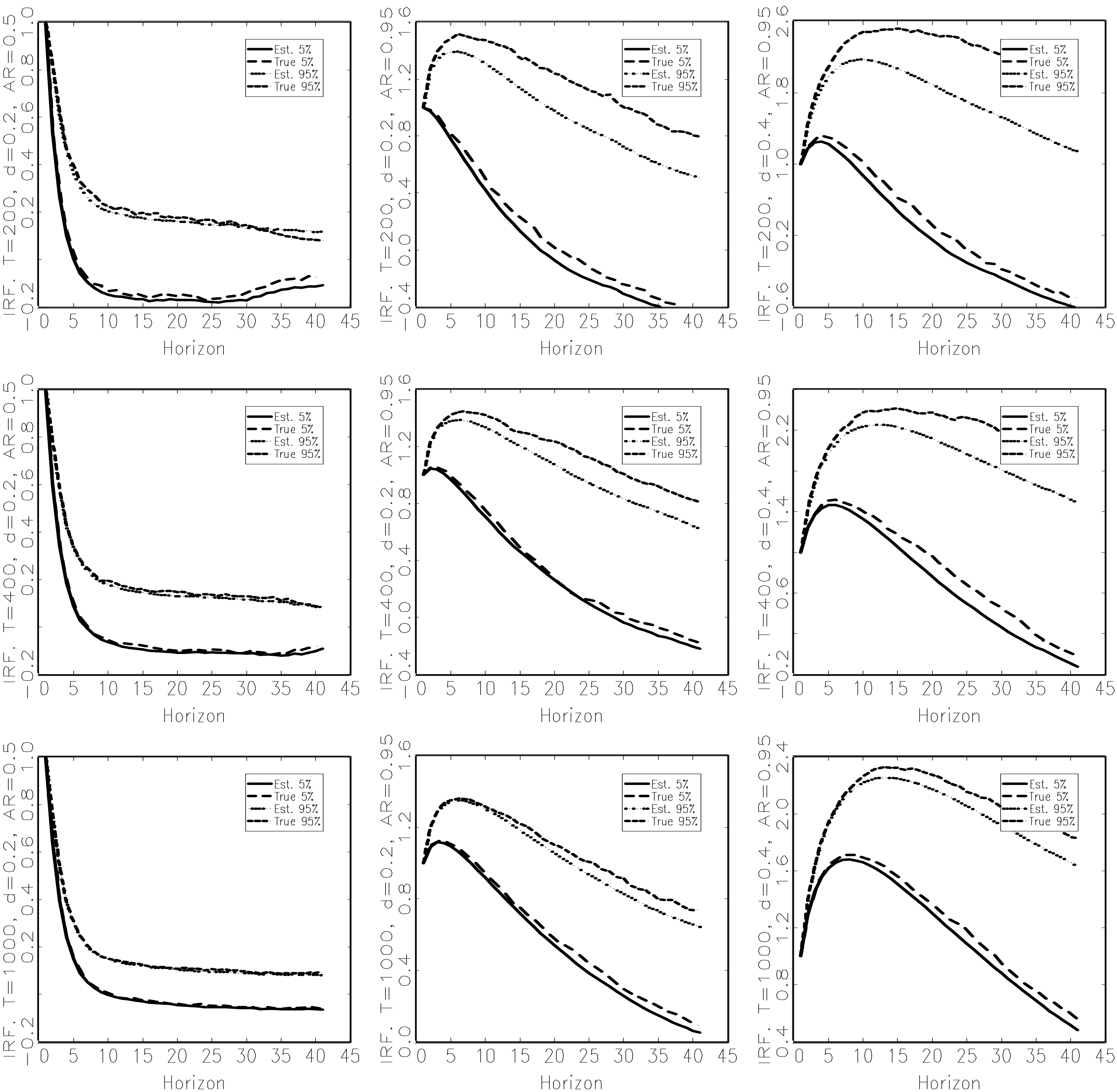

Figure 4: Monte Carlo Results: Standard Errors for Impulse Responses: Sieve Bootstrap on Parametric Model

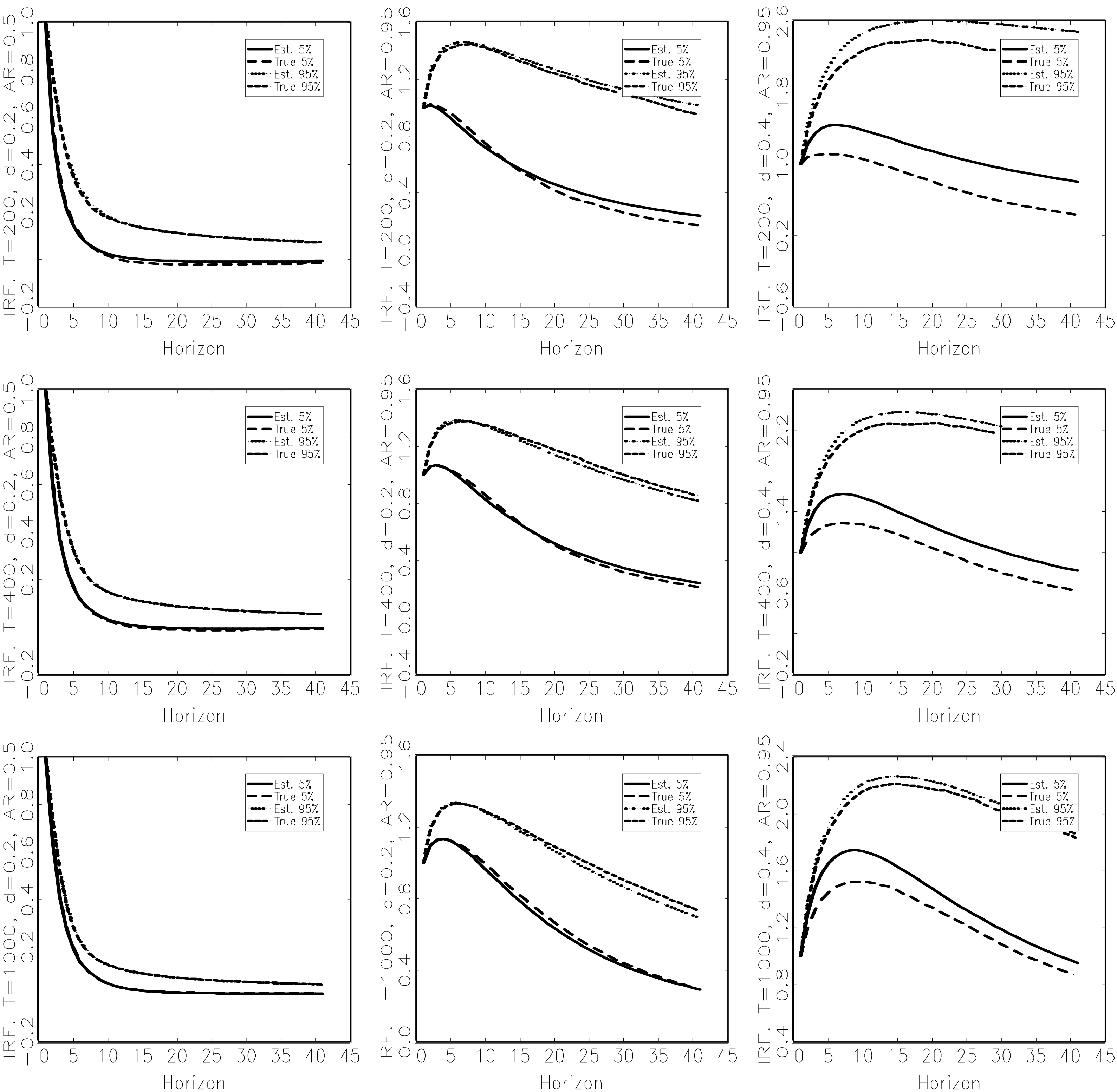

Figure 5: Empirical Results: Impulse Responses for CPI Inflation

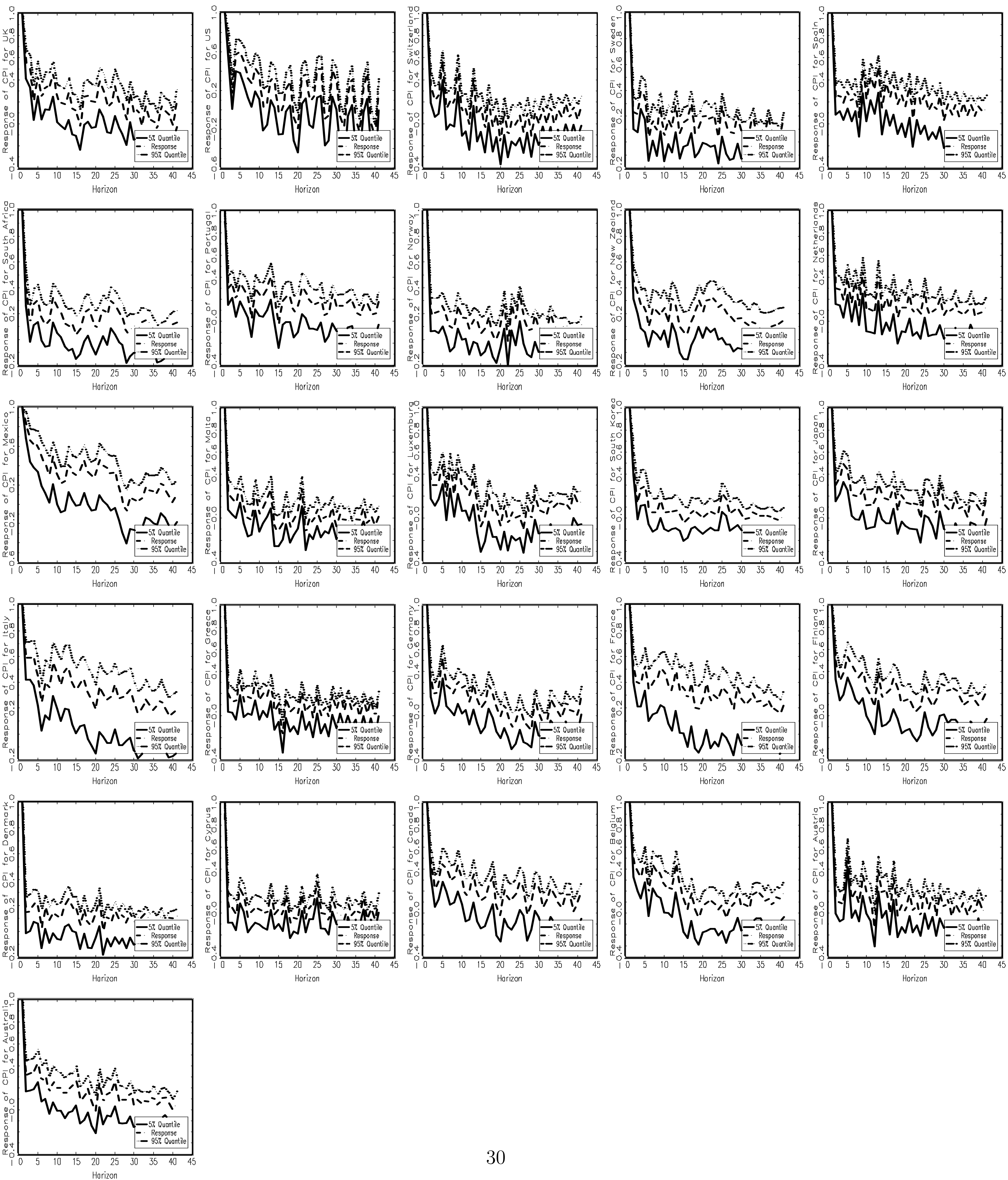

Figure 6: Empirical Results: Impulse Responses for Real Exchange Rates