



WP 15-46

Richard T. Baillie

Michigan State University, USA
Queen Mary University of London, UK
The Rimini Centre for Economic Analysis, Italy

George Kapetanios

Queen Mary University of London, UK

Fotis Papailias

Queen's University Belfast, UK
quantf research

INFERENCE FOR IMPULSE RESPONSE COEFFICIENTS FROM MULTIVARIATE FRACTIONALLY INTEGRATED PROCESSES

Copyright belongs to the author. Small sections of the text, not exceeding three paragraphs, can be used provided proper acknowledgement is given.

The Rimini Centre for Economic Analysis (RCEA) was established in March 2007. RCEA is a private, nonprofit organization dedicated to independent research in Applied and Theoretical Economics and related fields. RCEA organizes seminars and workshops, sponsors a general interest journal *The Review of Economic Analysis*, and organizes a biennial conference: *The Rimini Conference in Economics and Finance (RCEF)*. The RCEA has a Canadian branch: *The Rimini Centre for Economic Analysis in Canada (RCEA-Canada)*. Scientific work contributed by the RCEA Scholars is published in the RCEA Working Papers and Professional Report series.

The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Rimini Centre for Economic Analysis.

Inference for Impulse Response Coefficients From Multivariate Fractionally Integrated Processes

Richard T. Baillie

Department of Economics, Michigan State University, USA
School of Economics and Finance, Queen Mary, University of London, UK
Rimini Center for Economic Analysis, Italy

George Kapetanios

School of Economics and Finance, Queen Mary, University of London, UK

Fotis Papailias

Queen's University Management School, Queens University Belfast, UK
quantf research, www.quantf.com

April 28, 2015

Abstract

This paper considers a multivariate system of fractionally integrated time series and investigates the most appropriate way for estimating Impulse Response (*IR*) coefficients and their associated confidence intervals. The paper extends the univariate analysis recently provided by Baillie and Kapetanios (2013), and uses a semi parametric, time domain estimator, based on a vector autoregression (*VAR*) approximation. Results are also derived for the orthogonalized estimated *IRs* which are generally more practically relevant. Simulation evidence strongly indicates the desirability of applying the Kilian small sample bias correction, which is found to improve the coverage accuracy of confidence intervals for *IRs*. The most appropriate order of the *VAR* turns out to be relevant for the lag length of the *IR* being estimated.

1 Introduction

The analysis of individual Impulse Responses (*IRs*), and the complete Impulse Response Function (*IRF*), have long been recognized as an important device for interpreting a time series model, or dynamic econometric model. Sims (1980) wrote a seminal article on the practical importance and interpretation of these methods in a Vector Autoregression (*VAR*) context. However, the variability and derivation of confidence intervals for this approach was first derived for weakly stationary, or stable processes by Schmidt (1973, 1977). In particular, Schmidt

developed techniques for matrix differentiation that allowed the derivation of the asymptotic distribution of estimated IRs from a covariance stationary dynamic simultaneous equation model. This work led to the derivation of asymptotic distributions of estimated IRs from covariance stationary Vector Autoregressive (VAR) models in articles by Baillie (1987) and Lütkepohl (1988, 1989). While these theoretical results are quite elegant, they rely on a linearized Taylor series expansion of the estimated IRs around the true parameter values. These articles, and indeed the current study, deal with pointwise inference on IR coefficients, as opposed to the analysis of a vector of IRs and the IRF function. Some analysis of the latter problem for covariance stationary $VARs$ is to be found in Inoue and Kilian (2013).

This paper focuses on the problem of inference for estimating IRs and their confidence intervals for a multivariate fractionally integrated process. These models have been used in several areas of financial economics and empirical macroeconomics; see for example, Sowell (1992) and Jensen (2009). Some of the necessary technical machinery builds on the methodology for the univariate case developed by Baillie and Kapetanios (2013), which uses a semi parametric, time domain estimator, based on an autoregressive (AR) approximation and finds that it has good theoretical and small sample properties for the estimation of the IRs . Baillie and Kapetanios (2013) also recommend using a generic semi parametric sieve bootstrap, based on an autoregressive approximation for the construction of confidence intervals for the estimated IRs . In this context, is important to note the preceding work of Inoue and Kilian (2002a), Inoue and Kilian (2002b) and Goncalves and Kilian (2007) who focus on impulse response analysis for short memory processes.

This paper develops the methodology for the multivariate situation and shows that a valid method for conducting inference on IRs for fractionally integrated processes can be based on estimating an approximating VAR . As a result, it is important to note that our work allows for data generating processes that have not been covered previously in the literature. The validity of this approach is proven under quite mild assumptions. The findings in this paper also indicate that a good strategy for analyzing IRs is to estimate a semi parametric VAR , and to then use a sieve bootstrap for estimating confidence intervals. Simulation evidence indicates this approach appears to be a very good strategy for both short and long memory processes. For the purpose of estimating IRs and their associated confidence intervals, the various difficulties related with the identification and specification of multivariate $ARMA$ models alluded to by Tsay (1989), Kapetanios, Pagan and Scott (2007) and Poskitt (2011), do not arise. It should also be noted that some forms of nonlinearity can be mistaken for long memory, or very persistent processes. The sieve VAR approach advocated in this paper could be used in such a situation and has the advantage of not relying on the parametric specification of a complicated “structural” nonlinear model.

One of the major findings of this paper is that the semi parametric sieve bootstrap, based on a VAR approximation, appears to have a remarkably good coverage rate for the construction of confidence intervals of estimated IRs for a wide range of data generating processes, including

both weakly stationary and fractionally integrated processes.

The plan of the rest of this paper is as follows; Section 2 reviews some basic theory and assumptions, while Section 3 discusses the standard approach of deriving the asymptotic distribution for the estimated IRs and the problems with the theory in the presence of fractionally integrated processes. Section 4 then derives the basic ideas behind the Sieve VAR Bootstrap and the theory behind its validity. Section 5 describes some detailed simulation evidence on the performance of the VAR for estimation of IRs and the Sieve VAR for the derivation of confidence intervals of the estimated IRs . Section 6 provides an empirical application of our new methodology to the long run Fisher effect; and Section 7 provides some conclusions.

2 Basic Theory

This paper considers the vector time series process where \mathbf{y}_t is defined as an m dimensional multivariate stochastic process of the form

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \Psi_j \epsilon_{t-j},$$

where ϵ_t is an unobserved process such that $E(\epsilon_t) = \mathbf{0}$, $E(\epsilon_t \epsilon_t') = \Omega$ which is an m dimensional, positive semi definite, covariance matrix and $E(\epsilon_t \epsilon_s') = \mathbf{0}$ for $t \neq s$. The sequence of IRs or Wold Decomposition matrices are defined such that $\Psi_0 = \mathbf{I}$, and Ψ_j is a sequence of $m \times m$ matrices of constants. On defining $\Psi(\mathbf{L}) = \sum_{j=0}^{\infty} \Psi_j L^j$, the square summability condition

$$\sum_{j=0}^{\infty} \Psi_j \Omega \Psi_j' < \infty$$

is assumed to be satisfied. For subsequent analysis, the following assumptions are invoked:

Assumption 1 ϵ_t is an m dimensional ergodic martingale difference sequence, so that $E(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots) = \mathbf{0}$, and $E(\epsilon_t \epsilon_t' | \epsilon_{t-1}, \epsilon_{t-2}, \dots) = \Omega$ and its third and fourth moments matrices are finite constants.

Assumption 2 $\Psi(L) = \mathbf{D}(L)^{-1} \tilde{\Psi}(L)$, where $\mathbf{D}(L)$ is a diagonal matrix with typical diagonal element given by $(1 - L)^{d_i}$ and d_i is the long memory parameter for the i -th equation and $-0.5 < d_i < 0.5$ for $i = 1, 2, \dots, m$. Also, $\tilde{\Psi}(\mathbf{L}) = \sum_{j=0}^{\infty} \tilde{\Psi}_j L^j$ and $\sum_{j=0}^{\infty} j^{1/2} \|\tilde{\Psi}_j\| < \infty$, where $\|\cdot\|$ denotes the Euclidean matrix norm. Furthermore, $\Phi(z) \equiv \Psi(z)^{-1} = \sum_{j=0}^{\infty} \Phi_j z^j$ exists for all $|z| \leq 1$.

For subsequent use of the bootstrap, Assumption 1 has to be replaced by the condition that ϵ_t is i.i.d. The above framework is essentially restricted to one of linearity with corresponding constraints on the error process. The above class of processes is quite wide and includes many processes considered in the literature. Clearly, the leading univariate case is the basic $ARFIMA$

model of Granger (1980), Granger and Joyeux (1980) and Hosking (1981). The multivariate version of this model is the $MVARFIMA(p, d, q)$ which is given by

$$\mathbf{\Pi}(L)\mathbf{D}(L)(\mathbf{y}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t$$

where $\tilde{\boldsymbol{\Psi}}(L) = \mathbf{\Pi}(L)^{-1}\boldsymbol{\Theta}(L)$, $\mathbf{\Pi}(L) = \sum_{j=0}^p \boldsymbol{\Pi}_j L^j$, $\boldsymbol{\Theta}(L) = \sum_{j=0}^q \boldsymbol{\Theta}_j L^j$ and $\boldsymbol{\mu}$ is a vector of constants. Details of the population aspects of this multivariate model are to be found in Sowell (1986, 1992) and Chung (2002). Note that the standard assumption, in empirical work, is that the process is weakly stationary with exponentially decaying IRs with $\boldsymbol{\Psi}_j = O(\mathbf{C}^j)$ where \mathbf{C} is a matrix of constants with eigenvalues that are bounded from above by one in absolute value. Rather than invoking this standard assumption, this paper considers the case where there is strong persistence due to fractional integration in the IRs , so that the $\|\boldsymbol{\Psi}_j\| = O(j^{d-1})$ for $0 \leq d < 0.5$ as in Chung (2002). This implies that $\left\| \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\Psi}_j' \right\| < \infty$.

In many practical applications of $VARs$ and of IR analysis, there can be an issue of identification and the desire to obtain the orthogonalized IRs , which are standardized in the sense that the covariance matrix of the innovations are equal to the identity matrix rather than $\boldsymbol{\Omega}$. Hence an investigator may wish to provide estimates of $\{\boldsymbol{\Psi}_j \boldsymbol{\Omega}^{-1/2}\}_{j=1}^h$ rather than $\{\boldsymbol{\Psi}_j\}_{j=1}^h$. Since $\boldsymbol{\Omega}^{1/2}$ is not unique, then for a given $\boldsymbol{\Omega}$, it is necessary to provide further identifying assumptions; for example see Inoue and Kilian (2013) and Chapter 4 of Canova (2007) for a discussion. Details of inference from orthogonalized IRs are provided in Sections 4 and 5 of this paper.

3 Asymptotic Distribution of Estimated IRs

One standard parametric approach has been to specify a stationary and invertible $MVARFIMA$ model such that $\mathbf{\Pi}(L)$ and $\boldsymbol{\Theta}(L)$, and therefore, $\boldsymbol{\Psi}(L)$ are parameterized so that they are functions of a vector of parameters given by $\boldsymbol{\theta}$. In particular, for the $MVARFIMA$ model with $\mathbf{D}(L)\mathbf{y}_t = \mathbf{u}_t$ and

$$\mathbf{\Pi}(L)\mathbf{u}_t = \boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t$$

the system can be expressed as the Markovian, or companion form representation of $\mathbf{U}_t = \mathbf{C}\mathbf{U}_{t-1} + \mathbf{v}_t$ where $\mathbf{U}_t' = [\mathbf{y}_t, \mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-r+1}, \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \dots, \boldsymbol{\varepsilon}_{t-r+1}]$. The lag order $r = \max(p, q)$, while $\mathbf{v}_t' = [\boldsymbol{\varepsilon}_t, 0, \boldsymbol{\varepsilon}_t, 0]$, and with null matrices of the appropriate dimension. Then,

$$\mathbf{C} = \begin{bmatrix} \boldsymbol{\Pi}_1 \boldsymbol{\Pi}_2 \dots \boldsymbol{\Pi}_{r-1} & \boldsymbol{\Pi}_r & \boldsymbol{\Theta}_1 \boldsymbol{\Theta}_2 \dots \boldsymbol{\Theta}_{r-1} & \boldsymbol{\Theta}_r \\ & \mathbf{I} & & \\ & 0 & 0 & \\ & 0 & 0 & \mathbf{I} & 0 \end{bmatrix}$$

Then $\boldsymbol{\theta}' = [d_1, \dots, d_m, \text{vec}(\mathbf{N}'\mathbf{C}), \text{vech}(\boldsymbol{\Omega})]$ and $\mathbf{N}' = [\mathbf{I}, \mathbf{0}]$, which is of dimension m by $2mr$. Then on estimation of the structural parameters $\boldsymbol{\theta}$ by either approximate or full MLE ,

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{V}). \tag{1}$$

Note that $\boldsymbol{\theta}_0$ denotes the true value of $\boldsymbol{\theta}$, and the symbol \xrightarrow{L} denotes convergence in distribution¹. The theory for full and approximate *MLE* of multivariate *ARFIMA* models is described by Sowell (1986, 1992) and Chung (2002)². Some applications of the methodology are given in Diebold, Husted and Rush (1991) for the univariate case, while articles by Sowell (1992), Baillie and Chung (2002) and Jensen (2009) describe applications in the multivariate case.

In order to obtain confidence intervals for the estimated vectorized *IR* matrix at lag j , i.e. $\text{vec}(\boldsymbol{\Psi}_{j,\hat{\boldsymbol{\theta}}})$, where the subscript $\hat{\boldsymbol{\theta}}$ denotes dependence of $\boldsymbol{\Psi}_j$ on $\hat{\boldsymbol{\theta}}$, a traditional method has been to use the delta method, which is based on a linearized Taylor Series expansion around the true vectorized *IRs* at lag j , $\text{vec}(\boldsymbol{\Psi}_{j,\boldsymbol{\theta}_0})$.

Assuming that $\sup_{\boldsymbol{\theta}} \sum_{j=1}^{\infty} |\boldsymbol{\Psi}_{j,\boldsymbol{\theta}}| < \infty$ then for all $j = 1, \dots, h$

$$\sqrt{T} \left(\text{vec}(\boldsymbol{\Psi}_{j,\hat{\boldsymbol{\theta}}}) - \text{vec}(\boldsymbol{\Psi}_{j,\boldsymbol{\theta}_0}) \right) \xrightarrow{L} N(\mathbf{0}, \mathbf{D}'_j \mathbf{V} \mathbf{D}_j) \quad (2)$$

where $\mathbf{D}_j = \frac{\partial \text{vec}(\boldsymbol{\Psi}_{j,\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. For stationary and invertible *VARMA* models, there are parametric expressions available for the matrix \mathbf{D}_j ; see Baillie (1987). However, no corresponding results are yet available for the *MVARFIMA* model.

Analogous analysis can be conducted for orthogonalized *IRs* which are based on the quantities $\hat{\boldsymbol{\Psi}}_j \hat{\boldsymbol{\Omega}}^{-1/2}$, where $\hat{\boldsymbol{\Omega}} = T^{-1} \sum_t \hat{\boldsymbol{\epsilon}}_t \hat{\boldsymbol{\epsilon}}_t'$, and $\hat{\boldsymbol{\Omega}}^{1/2}$ denotes some suitably defined square root of $\hat{\boldsymbol{\Omega}}$. Then,

$$\sqrt{T} \left(\text{vec}(\boldsymbol{\Psi}_{j,\hat{\boldsymbol{\theta}}} \hat{\boldsymbol{\Omega}}_j^{-1/2}) - \text{vec}(\boldsymbol{\Psi}_{j,\boldsymbol{\theta}_0} \boldsymbol{\Omega}_j^{-1/2}) \right) \xrightarrow{L} N(\mathbf{0}, \mathbf{G}'_j \mathbf{V} \mathbf{G}_j) \quad (3)$$

where $\mathbf{G}_j = \frac{\partial \text{vec}(\boldsymbol{\Psi}_{j,\boldsymbol{\theta}} \boldsymbol{\Omega}_j^{-1/2})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. It is important to note that high persistence can be a problem with the above traditional methodology, or “delta method”. In particular, issues arise in the presence of very persistent processes, such as long memory processes and also near unit root processes; see Wright (2000), Kilian (1998, 1999) and Pesavento and Rossi (2007). This motivates the need for investigation of alternative approaches, as in Baillie and Kapetanios (2013) for univariate fractionally integrated processes.

4 Sieve VAR Bootstrap

Given the results in Baillie and Kapetanios (2013) for the univariate fractionally integrated process, the estimation and inference approach taken in this paper is to extend the sieve *AR* approximations with bootstrapped confidence intervals in the univariate case to a multivariate

¹For ease of exposition, $\boldsymbol{\mu}$ is assumed to either be the null vector, or known; so that the *MVARFIMA* model is expressed in deviation form. The *MLE* of $\boldsymbol{\mu}$ converges at a slower rate than $T^{1/2}$ and hence the full *MLE* of the system requires different normalizations for the estimated $\boldsymbol{\mu}$ parameters. The properties of the other parameter estimates remain unchanged.

²These results have been explored in more detail for the univariate case by Baillie and Kapetanios (2013), who use results of Hosoya (1997) which do not necessarily require the innovations to be i.i.d.. The extension to vector long memory processes has not been attempted.

setting. Hence, this paper considers approximating a multivariate fractionally integrated process with a $VAR(\infty)$ representation which exists under Assumption 2 by means of a $VAR(p_T)$ model where the lag order, p_T , is allowed to tend to infinity with the sample size. In particular

$$\mathbf{y}_t = \sum_{j=1}^{p_T} \Phi_j^{(p_T)} \mathbf{y}_{t-j} + \mathbf{v}_t$$

where $\sum_{j=1}^{p_T} \Phi_j^{(p_T)} \mathbf{y}_{t-j}$ is the linear projection of \mathbf{y}_t on $\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p_T}$. The *OLS* estimates of $\Phi_j^{(p_T)}$ are obtained by fitting a $VAR(p_T)$ model to the data and are denoted by $\widehat{\Phi}_j^{(p_T)}$. The estimated *IRs* are then obtained by inverting the truncated *VAR*. It follows from a straightforward extension of Theorem 5 of Poskitt (2007) that

$$\sum_{j=1}^{p_T} \left\| \widehat{\Phi}_j^{(p_T)} - \Phi_j^{(p_T)} \right\|^2 = o_p(1)$$

for all sequences $\{p_T\}$ such that $p_T \rightarrow \infty$ and $p_T = o(T^\alpha)$ for all $\alpha > 0$. Full details of the proof of this result are provided in Appendix 8.1 in this paper. An acceptable sequence for p_T is $p_T = \lfloor \ln(T)^\alpha \rfloor$ where $\alpha > 1$ and $\lfloor \cdot \rfloor$ denotes the integer part. Also, from an extension of Baxter's inequality proven by Inoue and Kasahara (2006),

$$\sum_{j=1}^{p_T} \left\| \Phi_j^{(p_T)} - \Phi_j \right\| = o(1)$$

as long as $p_T \rightarrow \infty$. The selection of p_T can be through a data dependent method such as an information criterion like the *AIC*,³ or by some deterministic rule such as $p_T = \lfloor \ln(T)^a \rfloor$, where $a > 0$. Then, the *IR* analysis proceeds by inverting the autoregressive polynomial with coefficient matrices $\widehat{\Phi}_j^{(p_T)}$, $j = 1, \dots, p_T$ to give estimated *IRs* of the form $\widehat{\Psi}_j^{(p_T)}$. Inference on $\widehat{\Psi}_j^{(p_T)}$ can be carried out by means of the bootstrap. The sieve bootstrap, as usually applied in the literature; e.g. Inoue and Kilian (2002b), and as analyzed in the univariate long memory case by Baillie and Kapetanios (2013), can also be implemented in this study. For ease of reference, the relevant algorithm is given below:

1. Estimate a $VAR(p_T)$ model on \mathbf{y}_t and obtain the estimated coefficients, $\widehat{\Phi}_j^{(p_T)}$, $j = 1, \dots, p_T$ and the residuals, $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \sum_{j=1}^{\min(p_T, t-1)} \widehat{\Phi}_j^{(p_T)} \mathbf{y}_{t-j}$.

³The use of information criterion necessitates the specification of lower and upper bounds for the possible value of p_T . Usually for the short memory case the lower bound $p_{\min} = p_{\min, T}$ is specified to diverge with T while the upper bound is of the form $p_{\max, T} = T^\alpha$ and $\alpha < 1/2$. Of course $p_{\min, T} < p_{\max, T}$. For the long memory case a different upper bound is needed. This is because consistency of the parameter estimates cannot be obtained for all $d < 1/2$ when $p_{\max, T} = T^\alpha$ for any $\alpha > 0$, and a lower rate is needed. Therefore, it is usually assumed that $p_{\max, T} = \lfloor \ln(T)^\alpha \rfloor$, where $\alpha > 0$. More details may be found in Poskitt (2007).

2. Invert the autoregressive polynomial with coefficient matrices $\widehat{\Phi}_j^{(p_T)}$, $j = 1, \dots, p_T$ to obtain the $\widehat{\Psi}_j^{(p_T)}$, for $j = 1, \dots, h$. In some applications it is desirable to orthogonalize the *IRs*, which are defined as $\widehat{\Psi}_j^{(p_T)} \widehat{\Omega}^{-1/2}$, where $\widehat{\Omega} = \frac{1}{T} \sum_t \widehat{\varepsilon}_t \widehat{\varepsilon}_t'$, and $\widehat{\Omega}^{1/2}$ denotes some suitably defined square root of $\widehat{\Omega}$.
3. Re-center $(\widehat{\varepsilon}_1, \dots, \widehat{\varepsilon}_T)'$.
4. Re-sample with replacement from this vector to obtain the bootstrap sample of error terms given by $(\varepsilon_1^*, \dots, \varepsilon_T^*)'$.
5. Use the above error terms together with $\widehat{\Phi}_j^{(p_T)}$, $j = 1, \dots, p_T$, to generate the bootstrap sample $(\mathbf{y}_1^*, \dots, \mathbf{y}_T^*)'$ with initial values set to zeros.
6. Estimate a $VAR(p_T)$ for $(\mathbf{y}_1^*, \dots, \mathbf{y}_T^*)'$ to obtain the bootstrap estimated autoregressive coefficients given $\widehat{\Phi}_j^{*,(p_T)}$ for $j = 1, \dots, p_T$.
7. Invert $\widehat{\Phi}^{*,(p_T)}(z) = \sum_{j=1}^{p_T} \widehat{\Phi}_j^{*,(p_T)} z^j$ to obtain bootstrap estimates of the impulse responses given by $\widehat{\Psi}_j^{*,(p_T)}$, $j = 1, \dots, h$. Similarly, bootstrap estimates of orthogonalized impulse responses are derived from $\widehat{\Psi}_j^{*,(p_T)} \widehat{\Omega}^{-1/2}$.
8. Repeat the above algorithm B times and then use the resulting estimates of the *IR* to construct an empirical distribution of the *IR*. We follow Kilian (1998) in the construction of the percentile intervals; see Kilian (1998) for more details.

The properties of the sieve estimator discussed in this section, as well as those of the sieve bootstrap, are summarized in the following Theorem, which together with Corollary 1 of Inoue and Kilian (2002a) imply the validity of the sieve bootstrap for *IR* analysis. The proof of the Theorem is in the appendix.

Theorem 1 (*Proof in Appendix 8.1*). *Let Assumptions 1 and 2 hold and set $p_T = o((\ln T)^a)$ for some $a > 0$. Then,*

$$\sum_{j=1}^{p_T} \left\| \widehat{\Phi}_j^{(p_T)} - \Phi_j \right\|^2 = o_p(1). \quad (4)$$

Further, for all $j = 1, \dots$, then

$$d(P_{\widehat{\Psi}_j^{(p_T)}}, P_{\widehat{\Psi}_j^{*,(p_T)}}) = o_p(1) \quad (5)$$

where $P_{\mathbf{Y}}$ denotes the probability law of a random matrix \mathbf{Y} , $d(P_{\mathbf{Y}_1}, P_{\mathbf{Y}_2})$ is the Mallows metric between $P_{\mathbf{Y}_1}$ and $P_{\mathbf{Y}_2}$ and

$$d(P_{\widehat{\Omega}^*}, P_{\widehat{\Omega}}) = o_p(1). \quad (6)$$

5 Simulations

This section investigates the properties of the sieve *VAR* bootstrap procedure compared with various other schemes. In order to have a meaningful comparison, the data generating process is a *MVARFIMA*(1, d , 0) which can have substantial persistence given the choice of the long memory parameter, d and with $m = 2$ dimensions. The data generating process is given by

$$\begin{aligned} \mathbf{D}(L)\mathbf{y}_t &= \mathbf{u}_t \\ \mathbf{u}_t &= \mathbf{\Phi}\mathbf{u}_{t-1} + \boldsymbol{\varepsilon}_t, \\ (\mathbf{I} - \mathbf{\Phi}L) &= \begin{pmatrix} 1 - \beta L & 0 \\ -0.5L & -0.5L \end{pmatrix} \end{aligned}$$

and

$$\boldsymbol{\varepsilon}_t \stackrel{iid}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix} \right),$$

where $\mathbf{D}(L)$ is an m dimensional square diagonal matrix and is constrained for convenience in the simulation study to have the same long memory parameter across equations so that $\mathbf{D}(L) = (1 - L)^d \mathbf{I}_m$, where \mathbf{I}_m is an m dimensional identity matrix. A critical feature of the different simulation designs concerns the persistence of the short memory *VAR*(1) components. This paper adopts the same design as used by Kilian (1998), where $(\mathbf{I} - \mathbf{\Phi}L)$ is constructed so that β determines the degree of short memory persistence, since the eigenvalues of $\mathbf{\Phi}$ are 0.5 and β , where $\beta \in \{0.5, 0.9\}$. In all the experiments d was chosen as 0.4⁴. The sample size is $T \in \{250, 500\}$ and the number of Monte Carlo replications is 500. For each realization, 1,000 bootstrap replications to obtain the first order mean bias “Kilian corrected”, (“bootstrap-after-bootstrap”) version of the parameter estimates are used, as in Kilian (1998); and 2,000 bootstrap replications are used for investigating the properties of the estimated *IRs*. Note that the Kilian bias correction is obtained using a double bootstrap as discussed on page 220 of Kilian (1998). Finally, and as previously noted, it can be desirable to use either orthogonalized or non orthogonalized *IRs*. The orthogonalized *IRs* are derived from imposing a recursive causal structure, from the first to the last variable, and is consistent with the standard Cholesky decomposition method.

It must be noted that the Kilian bias correction is derived under the assumption of weak stationarity. However, as we see from the simulation experiments, the correction provides adequate results in long memory data as well.

Given the various theoretical methods and issues discussed earlier, three different *VAR* estimation approaches were implemented in the simulations. The application of the *VARs* essentially vary in the way the maximum lag length is selected. The methods are:

⁴Simulation work was also completed for the case of $d = 0.2$, but the results do not add significant extra information and are not reported for reasons of conserving space. These results are available from the authors on request.

(i) *VAR* represents the case where the sieve $VAR(p_T)$ bootstrap has been applied with $p_T = \lfloor c \ln(T)^2 \rfloor$, and with the constant c allowed to take the values 0.25, 0.50, 0.75 and 1.00. For example, for the sample sizes of $T = 250$ and 500 and for $c = 1$, the lag length, p_T is 30 and 39 respectively. However, the results for the $c = 0.75$ case are suppressed from this version of the paper since they turned out to be extremely similar to those for $c = 1$; but are available from the authors on request. Similarly, the results for $c = 0.5$ are only presented for one design.

(ii) *VAR-K* denotes the sieve $VAR(p_T)$ bootstrap, $p_T = \lfloor c \ln(T)^2 \rfloor$, using the Kilian (1998) correction for bias. The same values for c are used.

(iii) *AIC-K* involves the sieve *VAR* bootstrap, where p is chosen by the *AIC*, and also the bias correction method of Kilian (1998) is implemented. The maximum order for the *AIC* is set as $p_T = \lfloor \ln(T)^2 + 20 \rfloor$.

All the results for the coverage rates are calculated at the 90% level. The first set of results presented in Figures 1 through 4 report results for the design with the less persistent short memory case of $\beta = 0.5$, for the sample size of $T = 250$. Each figure is divided into four panels. Starting with the top left corner and continuing in a clockwise direction, the first panel depicts the coverage rates for the *IRs* of y_1 to y_1 , then y_1 to y_2 , then y_2 to y_2 and finally y_2 to y_1 . The first two figures present both orthogonalized and non-orthogonalized *IRs*. As expected, the latter perform better since they do not require estimates of the Ω matrix. This pattern is consistent for all other cases considered in the simulation experiment. Since non-orthogonalized *IRs* are far less relevant for applied work, it was decided not to report these results after the first two figures for each *DGP*. This increases the clarity of the graphical exposition in the remaining figures.

The above figures contain interesting information on the coverage rates when estimating *IRs* at lags $h = 2, 4, 8, 12$ and $h = 16$. In particular there is the substantial increase in coverage provided from using the Kilian bias correction method compared to the regular sieve bootstrap *VAR*. Another interesting feature is the extra degree of difficulty introduced by estimating the orthogonalized, rather than non orthogonalized *IRs* and the relative importance of the Kilian bias correction in both cases.⁵

It can generally be seen that the use of $p_T = \lfloor c \ln(T)^2 \rfloor$, is extremely sensitive to the choice of c . In Figure 1 with $c = 0.25$, the coverage is quite good for low order lags Ψ_h for $h < 10$, but very poor for higher order lags of the *IRs* for $h \geq 10$. For higher values of c , beginning with $c = 0.50$ in Figure 2, the performance is seen to be radically improved for the higher order lags beyond $h > 10$, but conversely very poor for the low order lags of $h < 10$. This pattern is almost identical to the case where $c = 1.00$ in Figure 3. As previously mentioned the situation for $c = 0.75$ is extremely similar to the $c = 1.00$ case and is therefore not reported. Figure 4, summarizes results for the same design, only with the lag length determined by the *AIC*. These coverage rates are extremely poor and appear to be due to excessively high lag values for the *VAR* being chosen by the *AIC* method. The general poor performance of the *AIC* turned out

⁵Full details of these results can be obtained from two tables that are available on request from the authors.

to be typical for all the designs being considered and hence are not reported in the remaining figures in order to conserve space.

Figures 5 and 6 present corresponding information to Figures 1 through 4, only for the larger sample size of $T = 500$ and with $c = 0.25$ and $c = 1.00$ respectively. Figures 7 through 8 are analogous to the Figures 5 through 6 and differ only in terms of the short memory parameter β being fixed at the higher level of persistence of $\beta = 0.9$ and with a sample size of $T = 250$. Finally, Figures 9 and 10 are for the cases of higher level of short memory persistence with $\beta = 0.9$ and with the higher sample size of $T = 500$ and for $c = 0.25$ and $c = 1.00$ respectively. The conclusions are generally the same as for the initial tables 1 through 4.

The following broad conclusions can be reached on the basis of the simulation results:

(a) The orthogonalized estimated *IRs* perform considerably worse than the corresponding non-orthogonalized *IRs*. The intuition behind this is quite clear since the latter also use the estimate of the covariance matrix of the error term. This is a very important conclusion, since it demonstrates that the extension from univariate to multivariate *IRs* is not a trivial one in practice and hence many of the results obtained by Baillie and Kapetanios (2013) for the univariate case need to be re-examined for the multivariate case.

(b) The bias correction method is extremely useful and very substantially improves the performance of both the estimated orthogonalized and the non-orthogonalized *IRs* estimators.

(c) Superior estimates of the *IRs* at short horizons can be obtained from using *VARs* with fewer lags. Whereas the opposite seems to be the case for long horizons of the *IRs* being better estimated with high order *VARs*. The corollary is that a sensible approach is to make the lag order of the *VAR* depend on the horizon of the *IRs* one desires to estimate. A reasonable horizon cut-off point appears to be around 10 periods ahead. Given these results, such a mixed *VAR* approach was used for the *IR* analysis and the results are given in Figures 11 and 12 for the sample size of 250. Figure 11 reports results for using two different order *VARs*; with the lag order chosen on the basis for $c = 0.25$ when estimating *IRs* up to lag 10 and then from using a *VAR* with lags chosen on the basis of $c = 0.5$ for estimating *IRs* for lags 11 through 25. Figure 12 presents results for when the *VAR* has a lag order equal $h + 1$ for estimating *IRs* up to lag $h = 1, \dots, 10$; and on the basis of $c = 0.5$ is chosen for longer horizons. These both choices appear to be feasible ways of conducting *IR* analysis and achieve good results.

(d) The persistence of the short memory process, as defined by β , does not seem to affect the performance of any of the methods, or any of the above three conclusions.

(e) As expected, the performance of all the methods improves with increasing sample size.

(f) *AIC* seems to choose too high a value for p , the lag length of the *VAR*, for small sample size; and too low a value of p for larger sample sizes. Consequently it does well at long horizons in smaller samples and vice versa for longer samples. An interesting feature of *AIC* is that it performs much better on the non-orthogonalized *IRs*, which are of rather less economic interest than the orthogonalized version. In general it does not appear to be a very good method to use in this context.

6 Application to the Long Run Fisher equation

An interesting application of the above methodology concerns the relationship between inflation and nominal interest rates and the long run Fisher equation. In a recent article, Jensen (2009), has considered the reasons that previous literature has failed to find support for the long run Fisher effect; and attributes that to the inappropriate assumption that inflation may be a unit root process. Jensen assumes that both inflation and nominal interest rates are described by a fractionally integrated, mean reverting, bivariate *MVARFIMA* model. Jensen (2009) argues that since there is no evidence that inflation is a unit root process and since permanent changes to inflation do not occur, then the issue on whether permanent changes to inflation affect the long run nominal interest rate will be uninformative, as to the truth or otherwise of the long run Fisher hypothesis.

The following analysis uses the same data set as Jensen (2009) with the *US* and *UK* variables being analysed using quarterly series published in the International Financial Series database. The inflation rate is the usual definition of the differenced logarithm of each country's consumer price index; and the nominal interest rate is the annualized rate of return of the relevant Treasury bill. The data is from 1957–*Q1* through 2004–*Q4* for the *US* and 1957–*Q1* through 2004–*Q3* for the *UK*. The *VAR* maximum lag length for the *AIC* was set as $p_T = \lfloor 1.00 \ln(T)^2 \rfloor$. On adopting the standard monetary policy rule assumption that inflation is causally prior to nominal interest rates, the ordering of variables and the Cholesky decomposition is clear. For example see Bernanke et al. (2005) for discussion of this modeling approach. Following the conclusions of the above simulation study, the lag length of the *VAR* was chosen on the basis of a lag order equal $h + 1$ for estimating *IRs* up to lag $h = 1, \dots, 10$; and subsequently on the basis of $c = 0.5$ for longer horizons. As expected the *IRs* for own shocks; i.e. the effect of inflationary shocks on current levels of inflation, and the analogous effect of interest rate shocks on current interest rates, are both positive and significant. Figures 13 and 14 indicate that interestingly, the *UK* has a larger *IR* at short horizons for a *CPI* shock compared with an interest rate shock. Also, past *CPI* innovations have a positive effect on interest rates. However, this effect is relatively small, which possibly indicates that before the Volcker regime authorities did not really try seriously to combat inflation.

Furthermore, past interest rate shocks appear to positively effect inflation. This is a common finding in the literature and has been termed as the 'price puzzle'. According to Castelnuovo and Surico (2010), during periods of passive monetary policy, such as prior to the Volcker regime, a price puzzle in the *US* is to be expected. More specifically, these authors find that a positive response of prices to a monetary policy shock is historically limited to the sub-samples that are typically associated with a weak interest rate response to inflation.

7 Conclusion

This paper considers a fractionally integrated system and investigates the most appropriate method for estimating Impulse Responses and their associated confidence intervals. The paper has extended the univariate analysis recently provided by Baillie and Kapetanios (2013), and has used a semi parametric, time domain estimator, based on a vector autoregressive (*VAR*) approximation. We find that a deterministic rule of choice of the *VAR* lag length based on $c \ln(T)^2$ seems to perform reasonably well. Simulation evidence strongly indicates the desirability for applying the Kilian (1998) small sample bias correction method, which improves both the estimated orthogonalized and non-orthogonalized *IRs*.

One interesting finding is that superior estimates of the *IRs* for short horizons can be obtained from using *VARs* with fewer lags, while *IRs* at long horizons are better estimated with high order *VARs*. This method appears quite attractive and is used in an empirical example of the inflation and interest rate relationship.

One of the major findings of this paper is that the semi parametric sieve bootstrap, based on a *VAR* approximation, appears to have a remarkably good coverage rate for the construction of confidence intervals of estimated *IRs* for a wide range of data generating processes, including both weakly stationary and fractionally integrated processes.

8 Appendix

8.1 Proof of Theorem 1

This Appendix provides a proof of Theorem 1 in the paper. For the validity of equation (4) in our Theorem 1, it is necessary to show that Theorem 1 of Poskitt (2007) extends to the autocovariances of a multivariate fractionally integrated processes. In particular, it is required to prove that

$$\max_{0 \leq \tau \leq p_T} \|\mathbf{C}_T(\tau) - \mathbf{\Gamma}(\tau)\| = O\left(\left(\frac{\log T}{T}\right)^{1/2-d}\right) \quad (7)$$

where $\mathbf{\Gamma}(\tau) = E(\mathbf{y}_t \mathbf{y}'_{t+\tau})$ and $\mathbf{C}_T(\tau) = \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t \mathbf{y}'_{t+\tau}$. However, the proof of Theorem 1 of Poskitt (2007) proceeds in the multivariate case exactly as in the univariate case once it is established that

$$E(\|\mathbf{C}_T(\tau) - \mathbf{\Gamma}(\tau)\|^2) = O(T^{2d-1}). \quad (8)$$

It is necessary to prove this result, which for the univariate case is provided by the Theorem 3 of Hosking (1996). The multivariate version of this result can be obtained by considering the arguments in Chung (2002). In particular, the proof of Theorem 2 and the expression on pp.74 of Chung (2002) involves the conditional covariances $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t | \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \dots)$. It should be noted that if we assume that $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t | \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \dots) = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t)$ and then replacing $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t | \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \dots)$

with $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$ in (A.10) of Chung (2002), it then follows that (A.10) provides the variance of each element of $T^{0.5-d}(\mathbf{C}_T(\tau) - \boldsymbol{\Gamma}(\tau))$.

The proof of Theorem 2 of Chung (2002) then proceeds by showing that (A.10) is bounded, under the assumption that $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$ is bounded, proving (8). Using similar arguments to those in the proof of Theorems 4 and 5 of Poskitt (2007), (8) implies that

$$\sum_{j=1}^{p_T} \left\| \widehat{\boldsymbol{\Phi}}_j^{(p_T)} - \boldsymbol{\Phi}_j^{(p_T)} \right\| = o(1) \quad (9)$$

It is further required to show that

$$\sum_{j=1}^{p_T} \left\| \boldsymbol{\Phi}_j^{(p_T)} - \boldsymbol{\Phi}_j \right\| = o(1) \quad (10)$$

as long as $p_T \rightarrow \infty$. However, Theorem 4.1 of Inoue and Kasahara (2006) can be extended to this case, with some tedious but relatively straightforward algebra, which proves equation (4).

The next step is to prove equation (5), which is done by extending Theorem 6 of Baillie and Kapetanios (2013) to multivariate processes. This extension is carried out in the next subsection.

Finally, for a proof of equation (6), it is sufficient to show the results of Theorem 1 of Inoue and Kilian (2002a) and Inoue and Kilian (2002a) for $\boldsymbol{\Omega}$. First, it is necessary to show that the second block of (A.4) in Inoue and Kilian (2002a) converges to the normal distribution, which is sufficient for proving the above result. This is facilitated by obtaining versions of (A.1) through (A.3) in Inoue and Kilian (2002a). It is important to note that deriving (A.4) requires only that (A.1) through (A.3) hold with the right hand side being replaced by $o_p(1)$. This is first done by reviewing the results of Paparoditis (1996). It is then necessary to prove that

$$\|\mathbf{C}_T(\tau) - \boldsymbol{\Gamma}(\tau)\| = o_p(1) \quad (11)$$

$$\sum_{j=1}^{p_T} \left\| \widehat{\boldsymbol{\Phi}}_j^{(p_T)} - \boldsymbol{\Phi}_j \right\| = o_p(1) \quad (12)$$

$$\sum_{j=1}^{p_T} \left\| \widehat{\boldsymbol{\Psi}}_j^{(p_T)} - \boldsymbol{\Psi}_j \right\| = o_p(1) \quad (13)$$

$$\|\boldsymbol{\Gamma}^*(\tau) - \boldsymbol{\Gamma}(\tau)\| = o_p(1) \quad (14)$$

$$\|\boldsymbol{\Gamma}^{*-1}(\tau) - \boldsymbol{\Gamma}^{-1}(\tau)\| = o_p(1) \quad (15)$$

where $\mathbf{\Gamma}^*(\tau)$ is the bootstrap analog of $\mathbf{\Gamma}(\tau)$. Equations (11) and (12) have been shown above; while Equation (12) together with Theorem 2.2 of Paparoditis (1996) gives Equation (13). Similarly, Theorem 2.5 of Paparoditis (1996) and Equations (11) through (13) can be used to prove Equations (14) and (15). On using Equations (11) through (15) gives (A.4) of Inoue and Kilian (2002a). The rest of the steps of the proof of Theorem 1 of Inoue and Kilian (2002a) follow straightforwardly. In particular, Inoue and Kilian (2002a) uses Theorem 24.3 of Davidson (1994) to prove that

$$(T - k)^{1/2} \text{vech} \left(\widehat{\mathbf{\Omega}}^* - \mathbf{\Omega}^* \right) \sim N(0, \mathbf{V})$$

where $\mathbf{\Omega}^*$ and $\widehat{\mathbf{\Omega}}^*$ are the bootstrap analogs of $\mathbf{\Omega}$ and $\widehat{\mathbf{\Omega}}$ and \mathbf{V} is a positive definite covariance matrix. Then Equation (A.6) of Inoue and Kilian (2002a) which shows that $\widehat{\mathbf{\Omega}}^* - \mathbf{\Omega}^*$ is made up of terms that follow a martingale difference, is self contained and does not change between the short and long memory cases. The rest of the steps using Theorems 2.3 and 2.4 of Paparoditis (1996) again follow from Equation (12). Only the rate in Theorem 2.3 of Paparoditis (1996) changes but this is inconsequential as only convergence in probability is used in the relevant steps of the proof of Theorem 1 of Inoue and Kilian (2002a) and not any rates (see second half of page 325 and first half of page 326 of Inoue and Kilian (2002a)).

8.2 Multivariate Version of Theorem 6 of Baillie and Kapetanios (2013)

This section proves that the sieve bootstrap is valid for *IR* analysis based on the estimation of an $VAR(p_T)$ model. We use the results of Poskitt (2007) and Poskitt (2008). Let $\widehat{\mathbf{\Phi}}^{(p_T)}$ denote the $mp_T \times m$ matrix of parameter estimates of the coefficients of an $AR(p_T)$ model fitted to the original sample. Let $\widehat{\mathbf{\Phi}}^{*,(p_T)}$ denote the same estimates obtained from a bootstrap sample constructed using the sieve bootstrap. Let $\mathbf{X}_t^{(p_T)} = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p_T})'$, $\mathbf{X}^{(p_T)} = (\mathbf{X}_{p_T+1}^{(p_T)}, \dots, \mathbf{X}_T^{(p_T)})'$, $\mathbf{x} = (\mathbf{y}_{p_T+1}, \dots, \mathbf{y}_T)'$. Starred variables represent bootstrap versions of non-starred variables. C denotes a generic constant. Then,

$$\widehat{\mathbf{\Phi}}^{(p_T)} = \left(\mathbf{X}^{(p_T)'} \mathbf{X}^{(p_T)} \right)^{-1} \mathbf{X}^{(p_T)'} \mathbf{x}$$

and

$$\widehat{\mathbf{\Phi}}^{*,(p_T)} = \left(\mathbf{X}^{*,(p_T)'} \mathbf{X}^{*,(p_T)} \right)^{-1} \mathbf{X}^{*,(p_T)'} \mathbf{x}^*$$

Let $\{A\}_{ij}$ denote the i, j -th element of a matrix \mathbf{A} . We first need to determine the rate at which $\widehat{\mathbf{\Phi}}_j^{(p_T)}$ converges to $\mathbf{\Phi}_j^{(p_T)}$ and $\widehat{\mathbf{\Psi}}_j^{(p_T)}$ converges to $\mathbf{\Psi}_j^{(p_T)}$. By the extension of Theorem 5 of Poskitt (2007) to the multivariate setting, we present in the next section, we have that

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

Further,

$$\widehat{\Psi}_j^{(p_T)} - \Psi_j^{(p_T)} = O_p \left(\left\| \Phi^{(p_T)} \right\|^2 \left\| \widehat{\Phi}^{(p_T)} - \Phi^{(p_T)} \right\| \right) = O_p \left(p_T^{3/2} \left(\frac{\ln(T)}{T} \right)^{1/2-d} \right) = O_p(r_T)$$

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

$$\mathbf{m} \left(q_T \left(\lambda' \left(\widehat{\Phi}^{(p_T)} - \Phi^{(p_T)} \right) \right), q_T \left(\lambda' \left(\widehat{\Phi}^{*,(p_T)} - \widehat{\Phi}^{(p_T)} \right) \right) \right) \rightarrow 0$$

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

$$\begin{aligned} \mathbf{m} \left(q_T \left(\lambda' \left(\widehat{\Phi}^{(p_T)} - \Phi^{(p_T)} \right) \right), q_T \left(\lambda' \left(\widehat{\Phi}^{*,(p_T)} - \widehat{\Phi}^{(p_T)} \right) \right) \right)^2 &\leq E \left[E^* \left(\left\| q_T \left(\widehat{\Phi}^{*,(p_T)} - \widehat{\Phi}^{(p_T)} \right) \right\|^2 \right) \right] \leq \\ E \left[E^* \left(\left\| \left(\mathbf{X}^{*,(p_T)'} \mathbf{X}^{*,(p_T)} \right)^{-1} - \left(\mathbf{X}^{(p_T)'} \mathbf{X}^{(p_T)} \right)^{-1} \right\|^2 \right) \right] &E \left[E^* \left(\left\| q_T \left(\mathbf{X}^{*,(p_T)'} \mathbf{v}^{*,(p_T)} - \mathbf{X}^{(p_T)'} \mathbf{v}^{(p_T)} \right) \right\|^2 \right) \right] \end{aligned}$$

Examining each of the two terms above we have

$$\begin{aligned} E \left[E^* \left(\left\| \left(\mathbf{X}^{*,(p_T)'} \mathbf{X}^{*,(p_T)} \right)^{-1} - \left(\mathbf{X}^{(p_T)'} \mathbf{X}^{(p_T)} \right)^{-1} \right\|^2 \right) \right] &\leq \\ C p_T^4 E \left[E^* \left(\left\| \mathbf{X}^{*,(p_T)'} \mathbf{X}^{*,(p_T)} - \mathbf{X}^{(p_T)'} \mathbf{X}^{(p_T)} \right\|^2 \right) \right] &\leq \\ C p_T^6 \sup_{1 \leq i, j \leq m p_T} E \left[E^* \left(\left\| \left\{ \mathbf{X}^{*,(p_T)'} \mathbf{X}^{*,(p_T)} \right\}_{ij} - \left\{ \mathbf{X}^{(p_T)'} \mathbf{X}^{(p_T)} \right\}_{ij} \right\|^2 \right) \right] \end{aligned}$$

But

$$\begin{aligned} \sup_{1 \leq i, j \leq m p_T} E \left[E^* \left(\left\| \left\{ \mathbf{X}^{*,(p_T)'} \mathbf{X}^{*,(p_T)} \right\}_{ij} - \left\{ \mathbf{X}^{(p_T)'} \mathbf{X}^{(p_T)} \right\}_{ij} \right\|^2 \right) \right] &\leq \\ C p_T^2 E \left[E^* \left(\left\| \left\{ \mathbf{X}^{*,(p_T)'} \mathbf{X}^{*,(p_T)} \right\}_{11} - \left\{ \mathbf{X}^{(p_T)'} \mathbf{X}^{(p_T)} \right\}_{11} \right\|^2 \right) \right] \end{aligned}$$

Further,

$$\begin{aligned} E \left[E^* \left(\left\| \mathbf{X}^{*,(p_T)'} \mathbf{v}^{*,(p_T)} - \mathbf{X}^{(p_T)'} \mathbf{v}^{(p_T)} \right\|^2 \right) \right] &\leq C p_T \sup_{1 \leq i \leq m p_T} E \left[E^* \left(\left\| \left\{ \mathbf{X}^{*,(p_T)'} \mathbf{v}^{*,(p_T)} \right\}_i - \left\{ \mathbf{X}^{(p_T)'} \mathbf{v}^{(p_T)} \right\}_i \right\|^2 \right) \right] \\ C p_T^2 E \left[E^* \left(\left\| \left\{ \mathbf{X}^{*,(p_T)'} \mathbf{v}^{*,(p_T)} \right\}_1 - \left\{ \mathbf{X}^{(p_T)'} \mathbf{v}^{(p_T)} \right\}_1 \right\|^2 \right) \right] \end{aligned}$$

But by the proof of Theorem 2 of Poskitt (2008) we have that

$$E \left[E^* \left(\left\| \left\{ \mathbf{X}^{*,(p_T)'} \mathbf{v}^{*,(p_T)} \right\}_1 - \left\{ \mathbf{X}^{(p_T)'} \mathbf{v}^{(p_T)} \right\}_1 \right\|^2 \right) \right] = O \left(p_T^{5/2} \left(\frac{\log T}{T} \right)^{1-2d} \right)$$

and

$$E \left[E^* \left(\left\| \left\{ \mathbf{X}^{*,(p_T)'} \mathbf{v}^{*,(p_T)} \right\}_1 - \left\{ \mathbf{X}^{(p_T)'} \mathbf{v}^{(p_T)} \right\}_1 \right\|^2 \right) \right] = O \left(p_T^{5/2} \left(\frac{\log T}{T} \right)^{1-2d} \right)$$

Hence,

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

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

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

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

8.3 Proof of the Multivariate Version of Theorem 5 of Poskitt (2007)

The proof follows exactly the same steps as Theorem 5 of Poskitt (2007) once we replace the univariate versions of the Yule-Walker and OLS normal equations with their multivariate counterparts and show the multivariate version of Theorems 1 and 2 of Poskitt (2007). The next subsection proves the multivariate version of these Theorems.

8.4 Proof of the Multivariate Version of Theorems 1 and 2 of Poskitt (2007)

8.4.1 Proof of Theorem 1

Assume that $\frac{1}{4} < d < \frac{1}{2}$. By (8), $E(\|\mathbf{C}_T(\tau) - \Gamma(\tau)\|^2) = O(T^{-(1-2d)})$ and from Chebychev's inequality

$$\Pr \left(\|\mathbf{C}_T(\tau) - \Gamma(\tau)\| > \delta \left(\frac{\log T}{T} \right)^{\frac{1}{2}-d} \right) \leq \frac{\mathbf{C}}{\delta^2} \frac{1}{(T \log T)^{1-2d}}.$$

Now set $\Delta_\tau(T) = \|\mathbf{C}_T(\tau) - \Gamma(\tau)\| (\log T/T)^{d-\frac{1}{2}}$. Then for $T' = N^{4/(1-2d)}$

$$\sum_{N'=1}^{\infty} \Pr \left(\max_{|\tau| \leq p_{T'}} |\Delta_\tau(T')| > \delta \right) \leq \frac{\mathbf{C}}{\delta^2} \sum_{T=1}^{\infty} \left(\frac{1-2d}{4 \log T} \right)^{3(1-2d)/2} \frac{1}{T^2} < \infty$$

and by the Borel-Cantelli lemma $\Delta_\tau(T') \rightarrow 0$ a.s. uniformly in τ , $|\tau| \leq p_{T'}$.

Now, let $N^2 = T'$. Then for all T such that $N^2 < T < (N+1)^2$

$$\max_{|\tau| \leq p} |\Delta_\tau(T)| \leq \max_{|\tau| \leq p} \left(\frac{T}{\log T} \frac{\log(N+1)^2}{(N+1)^2} \right)^{\frac{1}{2}(1-2d)} \left(1 + \frac{1}{N} \right)^2 |\Delta_\tau((N+1)^2)| + \quad (16)$$

$$\max_{|\tau| \leq p} \left(\frac{T}{\log T} \right)^{\frac{1}{2}(1-2d)} \frac{1}{T} \sum_{t=T+1}^{(N+1)^2} \|\mathbf{y}_t \mathbf{y}'_{t-\tau} - \Gamma(\tau)\|. \quad (17)$$

But

$$\frac{T}{\log T} \frac{\log(N+1)^2}{(N+1)^2} \leq \frac{\log(N+1)}{\log N} \rightarrow 1 \text{ as } N \rightarrow \infty$$

and by what has already be shown follows that the first term on the right hand of (16) converges to zero a.s. .

Moreover using Chebychev's inequality once more we can bound

$$\Pr \left(\max_{|\tau| \leq p_T} \sum_{t=T+1}^{(N+1)^2} \|\mathbf{y}_t \mathbf{y}'_{t-\tau} - \mathbf{\Gamma}(\tau)\| \geq \delta (\log T)^{\frac{1}{2}(1-2d)} T^{\frac{1}{2}(1+2d)} \right) \quad (18)$$

by

$$\left(\frac{(N+1)^2}{\log(N+1)^2} \right)^{\frac{1}{2}(1-2d)} \cdot \frac{\mathbf{C}}{\delta^2 (\log T)^{(1-2d)} T^{(1+2d)}} \cdot (2N+1)^{4d}$$

where the first factor accounts for the maximum being taken over $p_T < p_{(N+1)^2}$ terms and the last factor arises because the sum contains $(N+1)^2 - T < (2N+1)$ summands. Thus we can deduce that the probability in (18) is less than

$$\frac{(2N+1)^{4d} (N+1)^{1-2d}}{(2 \log N)^{3(1-2d)/2} N^{2+4d}} \leq \frac{18}{N^{1+2d}}$$

and hence, via the Borel-Cantelli lemma, that the second term of (16) converges to zero with probability one since the series $\{N^{-(1+2d)}\}$ is convergent.

A similar proof using the method of subsequences can be employed to establish the result for the remaining cases, $d = \frac{1}{4}$, when $E [\|\mathbf{C}_T(\tau) - \mathbf{\Gamma}(\tau)\|^2] = O(\log T/T)$, and $d \in (-\frac{1}{2}, \frac{1}{4})$, when $E [\|\mathbf{C}_T(\tau) - \mathbf{\Gamma}(\tau)\|^2] = O(T^{-1})$.

8.4.2 Proof of Theorem 2

The following relationship exists between the elements of the sequence $c_T(\tau)$ and the $c_T(j, k)$ for $j - k = \tau = 0, \pm 1, \dots, \pm p_T$:

$$T \{\mathbf{C}_T(\tau) - \mathbf{C}_T(j, k)\} = \sum_{s=B_T(j, k)}^T \mathbf{y}_{s-|\tau|} \mathbf{y}'_s = \mathbf{D}_T(j, k, \tau) \quad (19)$$

where $B_T(j, k) = T + 1 - \min\{j, k\}$. Note that $\mathbf{D}_T(j, k, \tau)$ contains $\min\{j, k\}$, or at most, p_T summands. Now, since $\boldsymbol{\varepsilon}(t)$ has finite fourth moment, $E [\|\mathbf{y}_t \mathbf{y}'_t\|^2] < \infty$.

Thus

$$\Pr \left(\max_{|\tau| \leq p_T} \|\mathbf{D}_T(j, k, t)\| \geq \delta T \right) < p_T \frac{\mathbf{C} p_T^2}{\delta^2 T^2} \leq \frac{\mathbf{C}}{(\log T)^{\frac{3}{2}} T^{\frac{1}{2}}} \quad (20)$$

where the final inequality follows since $0 < d < \frac{1}{2}$ $p_T = o\left\{(T/\log T)^{\frac{1}{2}}(\log T/T)^d\right\}$ and $(\log T/T)^d < 1$ and $p_T = o\left\{(T/\log T)^{\frac{1}{2}}\right\}$ for $-\frac{1}{2} < d \leq 0$. Along the subsequence $T' = N'^4$ it follows that $\lim_{N' \rightarrow \infty} \max_{|\tau| \leq p_{T'}} T' - 1 \|\mathbf{D}_{T'}(j, k, \tau)\| = 0$ a.s. because $\{N'^{-2}\}$ is a convergent series. Furthermore, lettings $N^2 = T'$, then for all T between N^2 and $(N+1)^2$ we can bound $\|N^{-2}\mathbf{D}_{N^2}(j, k, \tau) - T^{-1}\mathbf{D}_T(j, k, \tau)\|$ by

$$\left\| \frac{(T - N^2)\mathbf{D}_{N^2}(j, k, \tau)}{TN^2} \right\| + \left\| \frac{\sum_{s=B_{N^2}(j,k)}^{B_T(j,k)} \mathbf{y}_{s-|\tau|} \mathbf{y}'_s - \sum_{s=N^2+1}^T \mathbf{y}_{s-|\tau|} \mathbf{y}'_s}{T} \right\|. \quad (21)$$

The first term in (21) converges to zero uniformly in j, k and τ by what has already been proved since $(T - N^2)/TN^2 \leq (2N + 1)/N^4$ and the second term converges similarly via an application of Chebychev's inequality and the Borrel-Cantelli lemma.

To show the latter, consider the case $d \in (-\frac{1}{2}, \frac{1}{4})$ for example. By (8) the variance of the numerator can be bounded by $C(T - N^2) \leq C(2N + 1)$ uniformly in $j, k = 0, 1, \dots, p_T$ so

$$\Pr \left(\max_{|\tau| \leq p_T} \left\| \sum_{s=B_{N^2}(j,k)}^{B_T(j,k)} \mathbf{y}_{s-|\tau|} \mathbf{y}'_s - \sum_{s=N^2+1}^T \mathbf{y}_{s-|\tau|} \mathbf{y}'_s \right\| \geq \delta T \right) < p_T \cdot \frac{C(2N + 1)}{\delta^2 T^2}$$

and $p_T(2N + 1)/T^2 \leq 2(N + 1)(2N + 1)/N^4 \log(N + 1) < 6/N^2$.

References

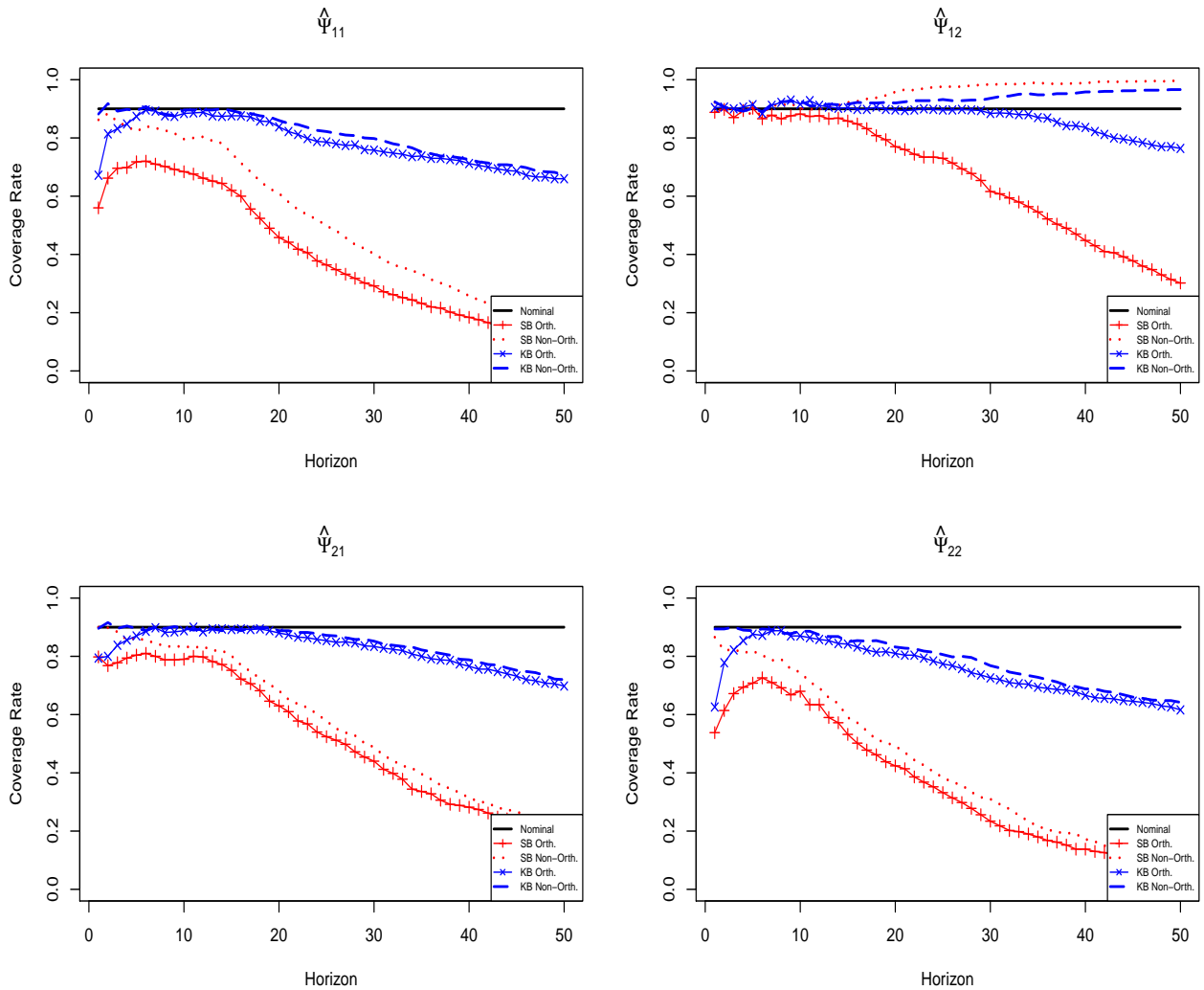
- BAILLIE, R. T. (1987): "Inference in dynamic models containing surprise variables", *Journal of Econometrics*, 35, 101-117.
- BAILLIE, R. T., AND S.-K. CHUNG (2002): "Modeling and forecasting from trend stationary long memory models with applications to climatology", *International Journal of Forecasting*, 18, 215-226.
- BAILLIE, R. T., AND G. KAPETANIOS (2013): "Estimation and inference for impulse response functions form univariate strongly persistent processes", *Econometrics Journal*, 16, 373-399.
- BERNANKE, B., BOIVIN, J. AND P. ELIASZ (2005): "Measuring the Effects of Monetary Policy: A factor augmented vector autoregressive (FAVAR) approach", *Quarterly Journal of Economics*, 50, 387-422.
- CANOVA, F. (2007): *Methods for Applied Macroeconomic Research*. Princeton University Press.
- CASTELNUOVO, E. AND SURICO, P. (2010): "Monetary policy, inflation expectations and the price puzzle", *Economic Journal*, 120, 1262-1283.
- CHUNG, C.-F. (2002): "Sample means, sample autocovariances and linear regression of stationary multivariate long memory processes", *Econometric Theory*, 18, 51-78.

- DAVIDSON, J.E.H. (1994): “Stochastic Limit Theory”, *Oxford University Press*.
- DIEBOLD, F.X., S, HUSTED AND M. RUSH (1991): “Real exchange rates under the gold standard”, *Journal of Political Economy*, 99, 1252-1271.
- GONCALVES, S. AND L. KILIAN (2007): “Asymptotic and bootstrap inference for $AR(\infty)$ processes with conditional heteroskedasticity”, *Econometric Reviews*, 26, 609–641.
- 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.
- HOSKING, J. R. M. (1981): “Fractional differencing”, *Biometrika*, 65, 165–176.
- HOSKING, J. R. M. (1996): “Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long memory time series”, *Journal of Econometrics*, 73, 261-284.
- 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.
- INOUE, A. AND L. KILIAN (2002a): “Bootstrapping autoregressive processes with possible unit roots”, *Econometrica*, 70, 377-391.
- INOUE, A. AND L. KILIAN (2002b): “Bootstrapping smooth functions of slope parameters and innovation variances in $VAR(\infty)$ models”, *International Economic Review*, 43, 309–332.
- INOUE, A. AND L. KILIAN (2013): “Inference on impulse response functions in structural VAR models”, *Journal of Econometrics*, 177, 1-13.
- JENSEN, M.(2009): “The long-run Fisher effect: can it be tested?”, *Journal of Money, Credit and Banking*, 41, 221-231.
- KAPETANIOS, G., PAGAN, A.R. AND SCOTT, A. (2007): “Making a match: combining theory and evidence in policy orientated macroeconomic modelling”, *Journal of Econometrics*, 136, 565-594.
- KILIAN, L. (1998): “Small-sample confidence intervals for impulse response functions”, *Review of Economics and Statistics*, 80, 218–230.
- KILIAN, L. (1999): “Finite sample properties of percentile and percentile t bootstrap confidence intervals for impulse response”, *Review of Economics and Statistics*, 81, 652–660.

- LÜTKEPOHL, H. (1988): “Asymptotic distribution of the moving average coefficients of an estimated vector autoregressive process”, *Econometric Theory*, 4, 77–85.
- LÜTKEPOHL, H. (1989): “A note on the asymptotic distribution of impulse response functions of estimated VAR models with orthogonal residuals”, *Journal of Econometrics*, 42, 371–376.
- PAPARODITIS, E. (1996): “Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes”, *Journal of Multivariate Analysis*, 57, 277–296.
- PESAVENTO, E. AND B. ROSSI (2007): “Impulse response confidence intervals with persistent data: what have we learned?”, *Journal of Economic Dynamics and Control*, 31, 2398–2412.
- POSKITT, D. (2007): “Autoregressive approximation in nonstandard situations: The non-invertible and fractionally integrated case”, *Annals of the Institute of Statistical Mathematics*, 59, 697–725.
- POSKITT, D. (2011): “Vector autoregressive moving average identification for macroeconomic modeling: a new methodology”, *Monash University working paper*.
- SCHMIDT, P. (1973): “The asymptotic distribution of dynamic multipliers”, *Econometrica*, 41, 161-164.
- SCHMIDT, P. (1977): “Some small sample evidence on the distribution of dynamic simulation forecasts”, *Econometrica*, 45, 997-1005.
- SIMS, C.A. (1980): “Macroeconomics and reality”, *Econometrica*, 48, 1-48.
- SOWELL, F.B. (1986): “Fractionally integrated vector time series”. *PhD dissertation, Duke University*.
- SOWELL, F.B.(1992): “Modeling long run behavior with the fractional ARIMA model”, *Journal of Econometrics*, 29, 277-302.
- TSAY, R.(1989): “Identifying multivariate time series models”, *Journal of Time Series Analysis*, 10, 357-372.
- WRIGHT, J. (2000): “Confidence intervals for univariate impulse responses with a near unit root”, *Journal of Business and Economic Statistics*, 20, 368-373.

Simulation Results: Figures

Figure 1: Effective Coverage Rates from Design with $\beta = 0.5$, $T = 250$ and $p_T = \lfloor 0.25 \ln(T)^2 \rfloor = 8$.



Key to all Simulations Figures: The solid line represents the nominal coverage rate at the 90 % level in all cases. The line denoted by "+" represents the coverage rate of the orthogonalized *IRs* from a standard Sieve *VAR* bootstrap (*SB Orth.*). The line denoted by "x" represents the coverage rate of the orthogonalized *IRs* from a Sieve *VAR* bootstrap using the Kilian (1998) bias correction (*KB Orth.*). For Figures 1 and 2, the dotted line represents the coverage rate of the non-orthogonalized *IRs* from a standard Sieve *VAR* bootstrap (*SB Non-Orth.*). The dashed line represents the coverage rate of the non-orthogonalized *IRs* from a Sieve *VAR* bootstrap using the Kilian (1998) bias correction (*KB Non-Orth.*). The variable $\hat{\Psi}_{ij}$ denotes the response of variable *i* to a shock in variable *j*; for $i, j = 1, 2$. The results are from 500 Monte Carlo replications; while the Kilian (1998) bias corrected parameter estimates are obtained from using 1,000 bootstrap resamples. The *IRs* are obtained from using 2,000 bootstrap resamples.

Figure 2: Effective Coverage Rates from Design with $\beta = 0.5$, $T = 250$ and $p_T = \lfloor 0.50 \ln(T)^2 \rfloor = 15$.

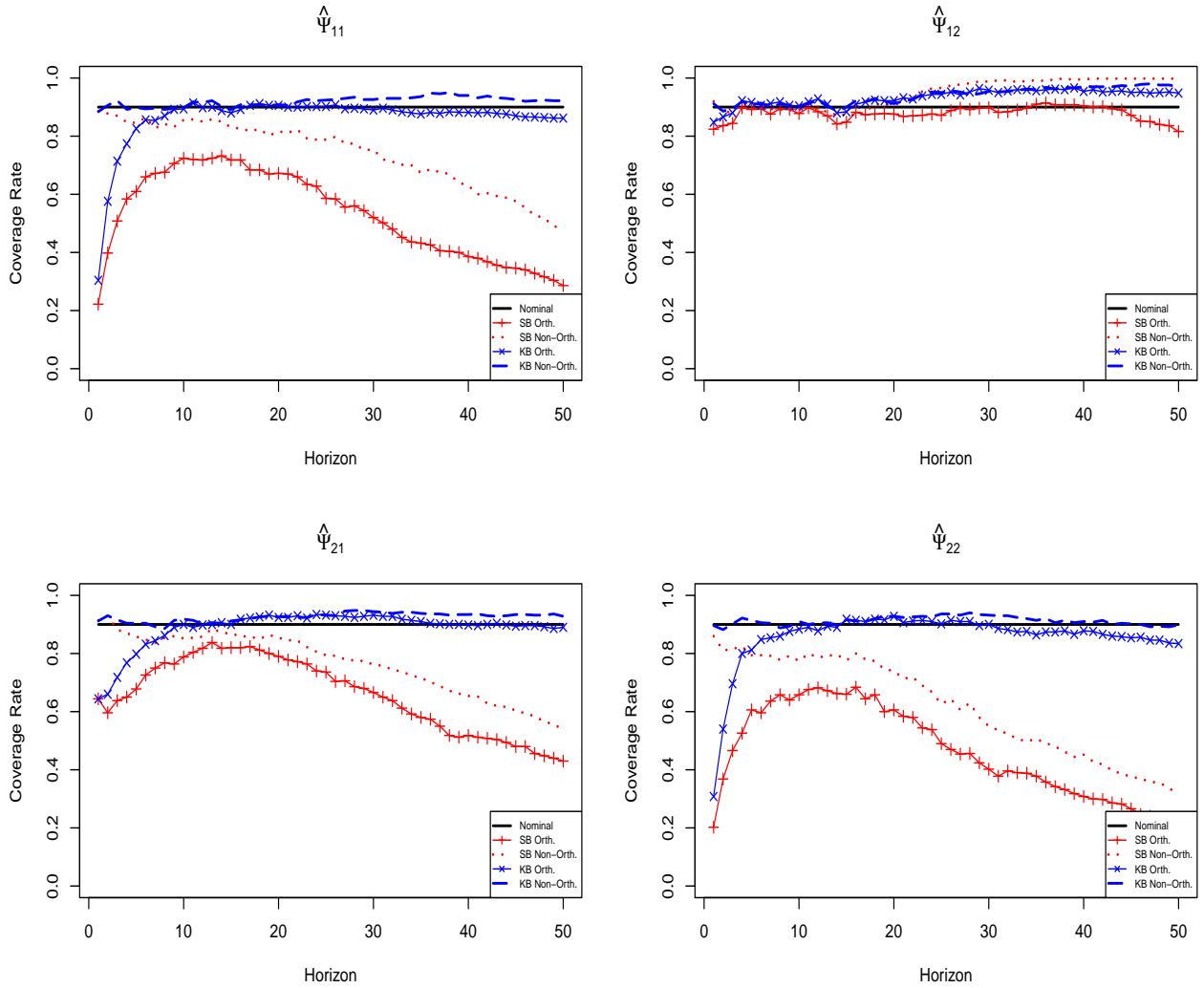


Figure 3: Effective Coverage Rates from Design with $\beta = 0.5$, $T = 250$ and $p_T = \lfloor 1.00 \ln(T)^2 \rfloor = 30$.

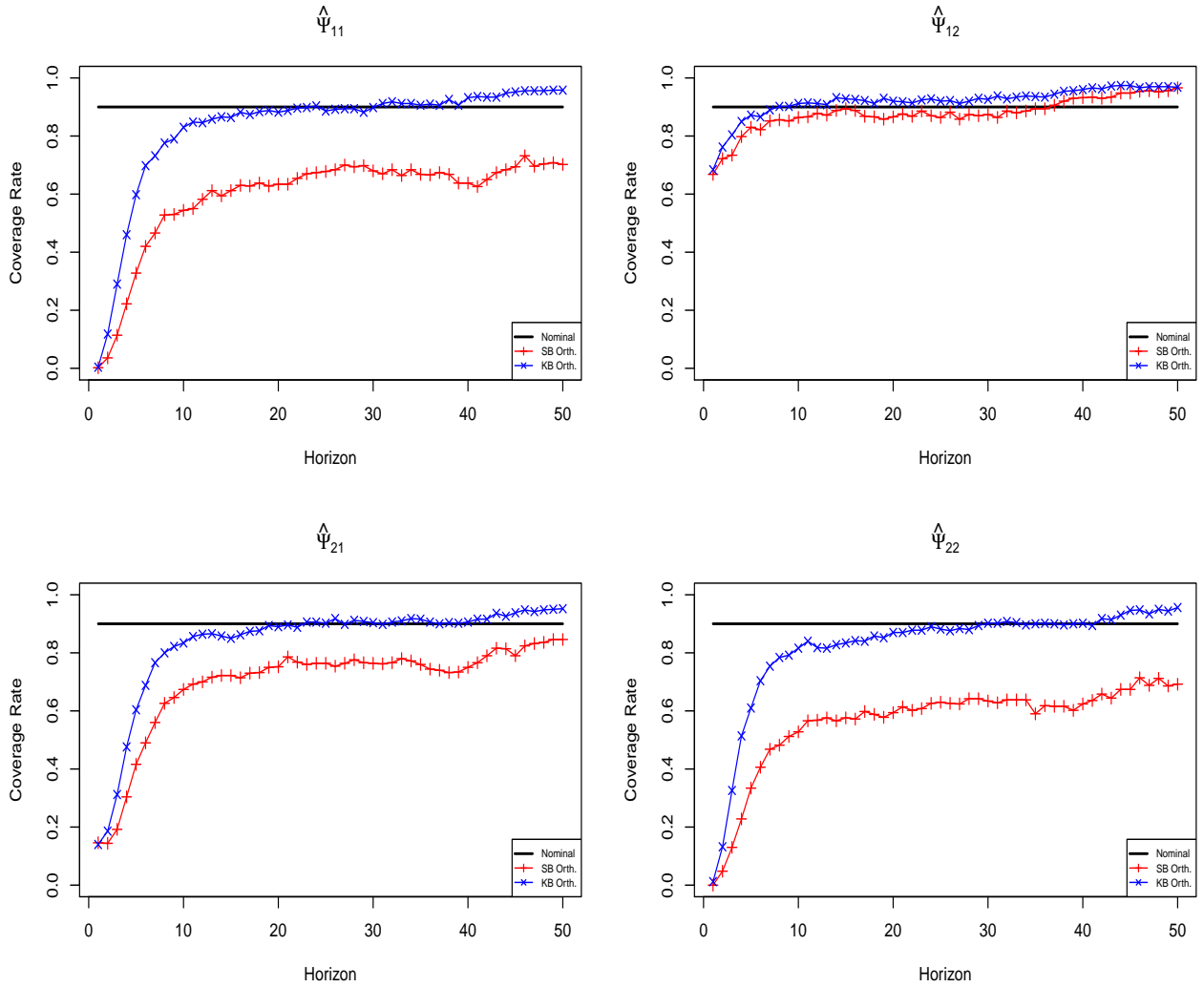


Figure 4: Effective Coverage Rates from Design with $\beta = 0.5$, $T = 250$ and $p_T = AIC$.

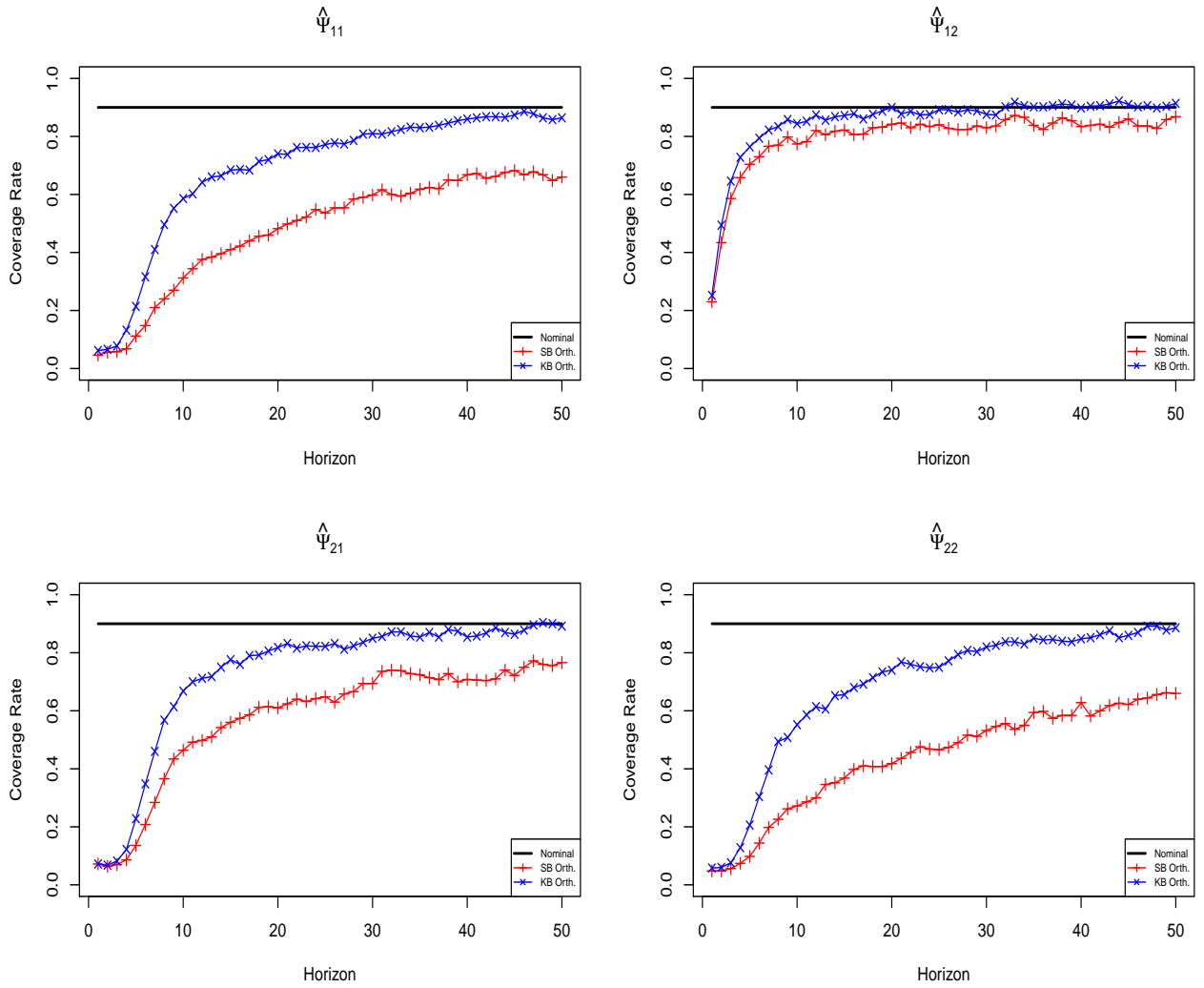


Figure 5: Effective Coverage Rates from Design with $\beta = 0.5$, $T = 500$ and $p_T = \lfloor 0.25 \ln(T)^2 \rfloor = 10$.

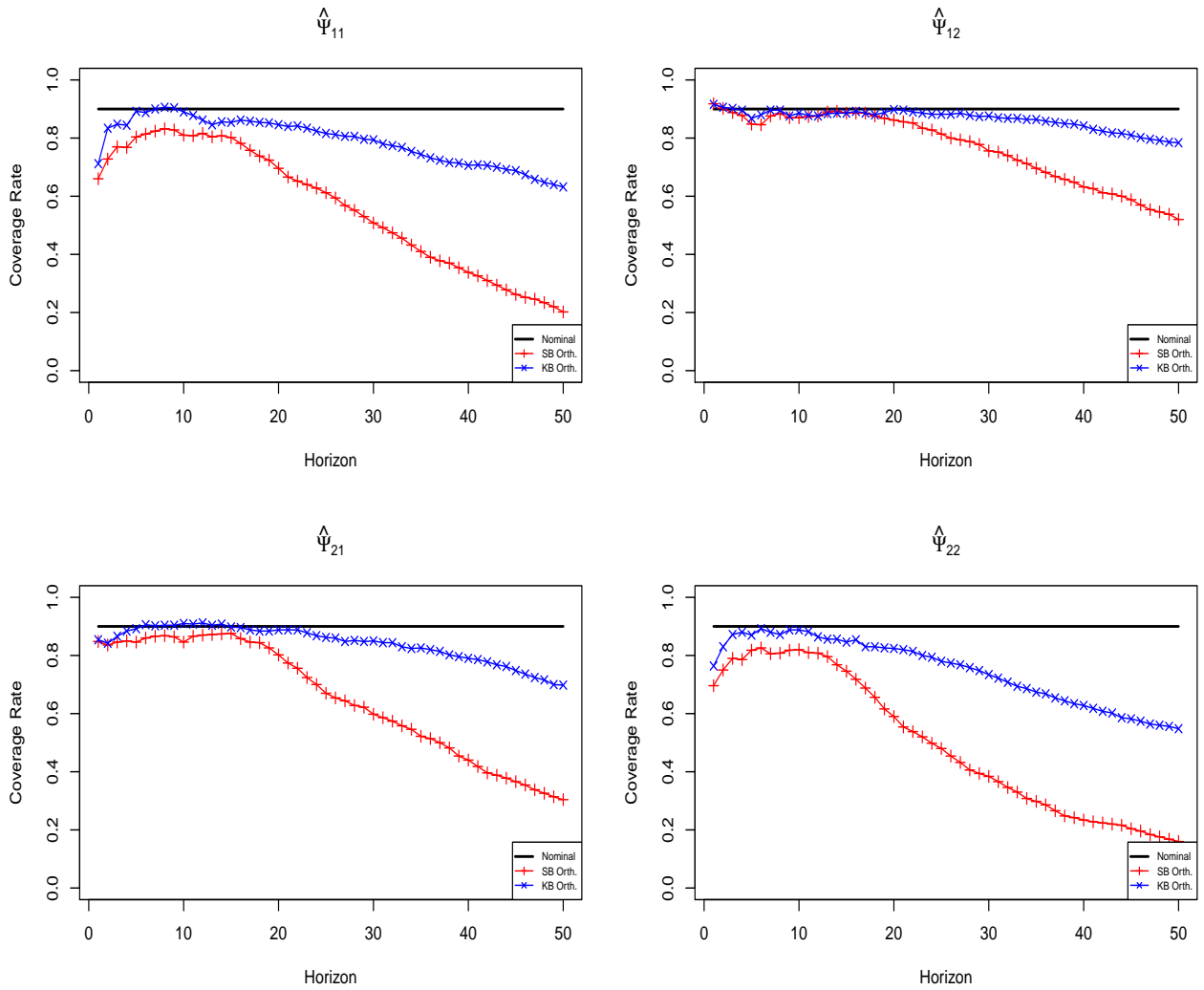


Figure 6: Effective Coverage Rates from Design with $\beta = 0.5$, $T = 500$ and $p_T = \lfloor 1.00 \ln(T)^2 \rfloor = 39$.

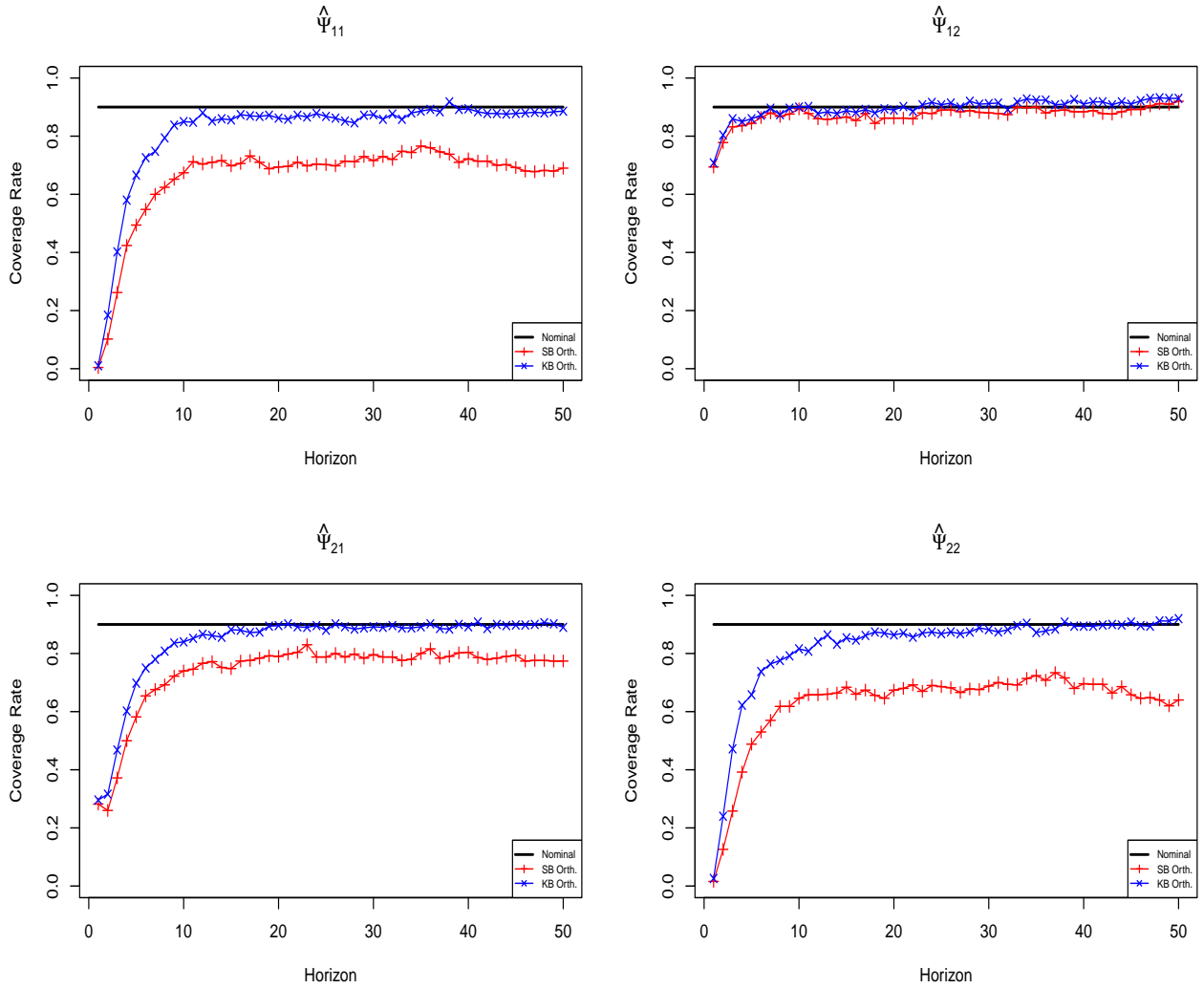


Figure 7: Effective Coverage Rates from Design with $\beta = 0.9$, $T = 250$ and $p_T = \lfloor 0.25 \ln(T)^2 \rfloor = 8$.

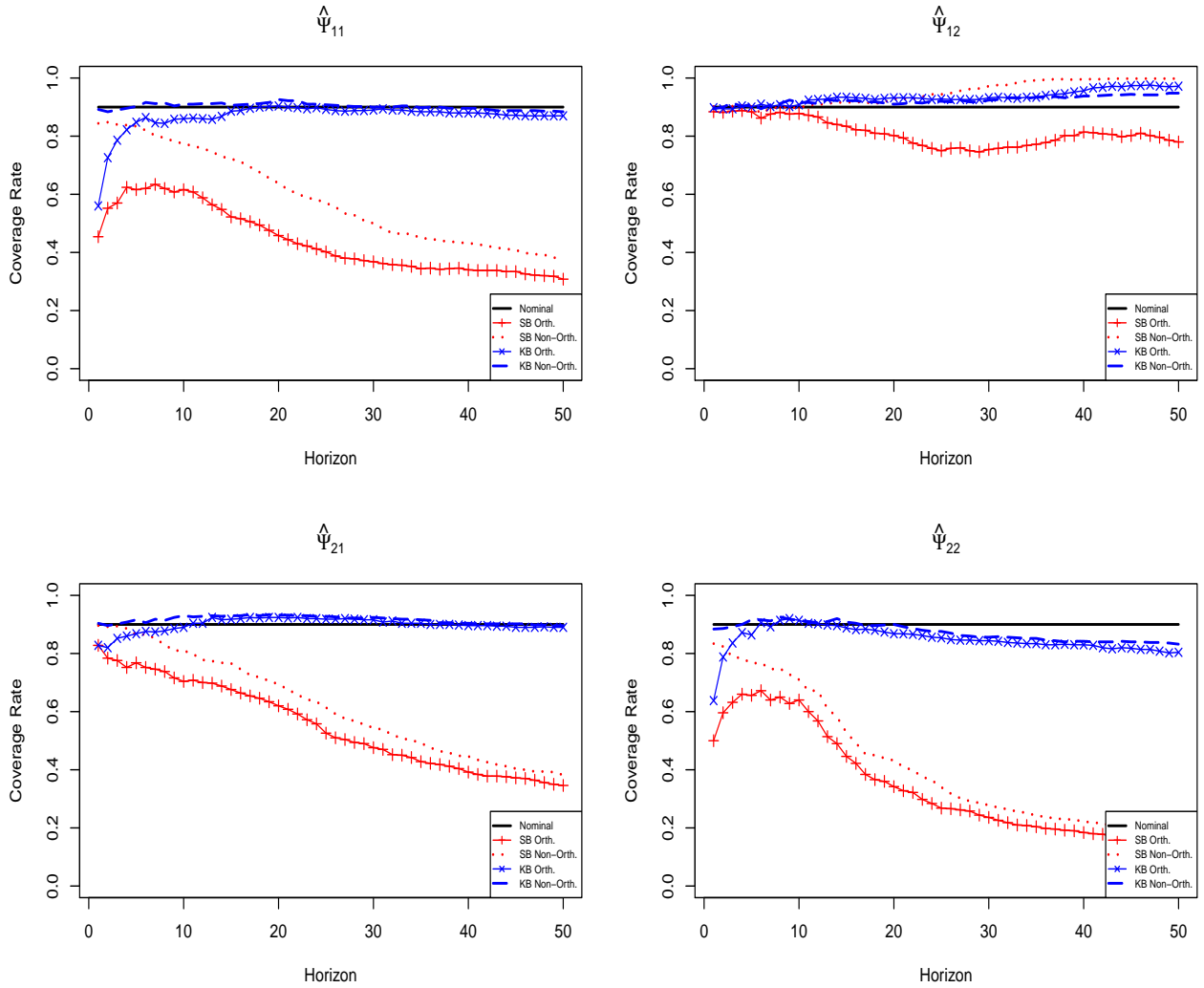


Figure 8: Effective Coverage Rates from Design with $\beta = 0.9$, $T = 250$ and $p_T = \lfloor 1.00 \ln(T)^2 \rfloor = 30$.

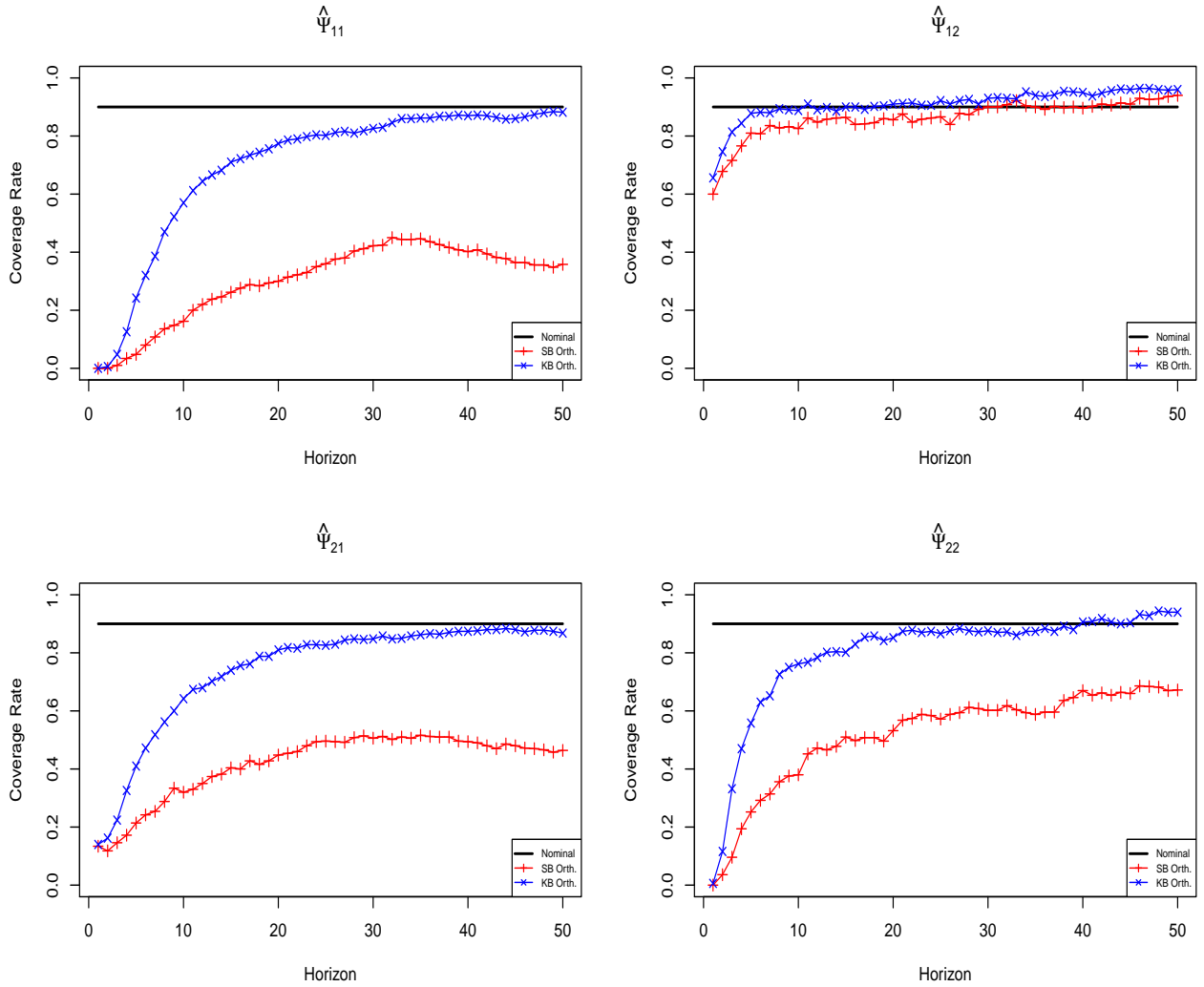


Figure 9: Effective Coverage Rates from Design with $\beta = 0.9$, $T = 500$ and $p_T = \lfloor 0.25 \ln(T)^2 \rfloor = 10$.

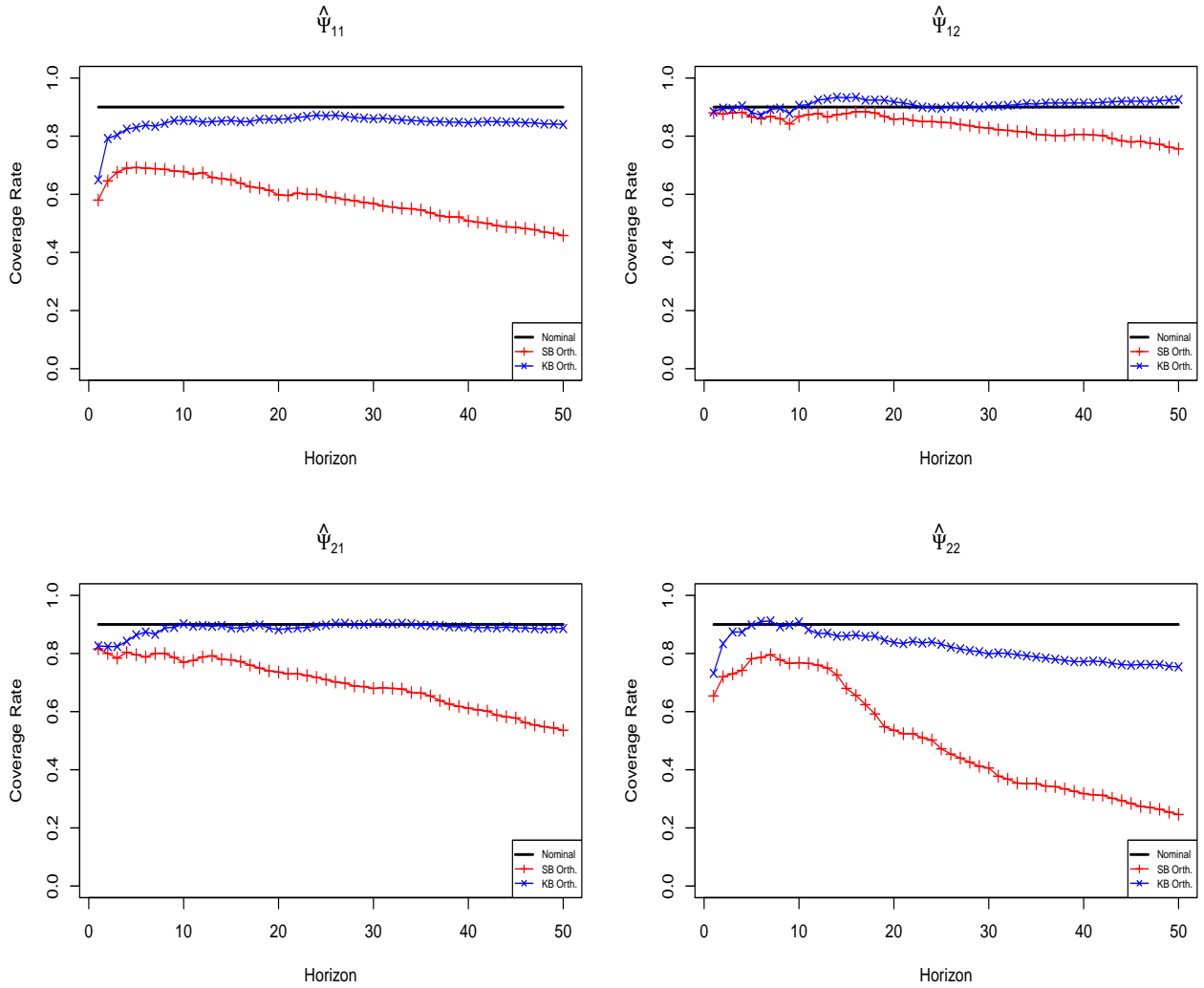


Figure 10: Effective Coverage Rates from Design with $\beta = 0.9$, $T = 500$ and $p_T = \lfloor 1.00 \ln(T)^2 \rfloor = 39$.

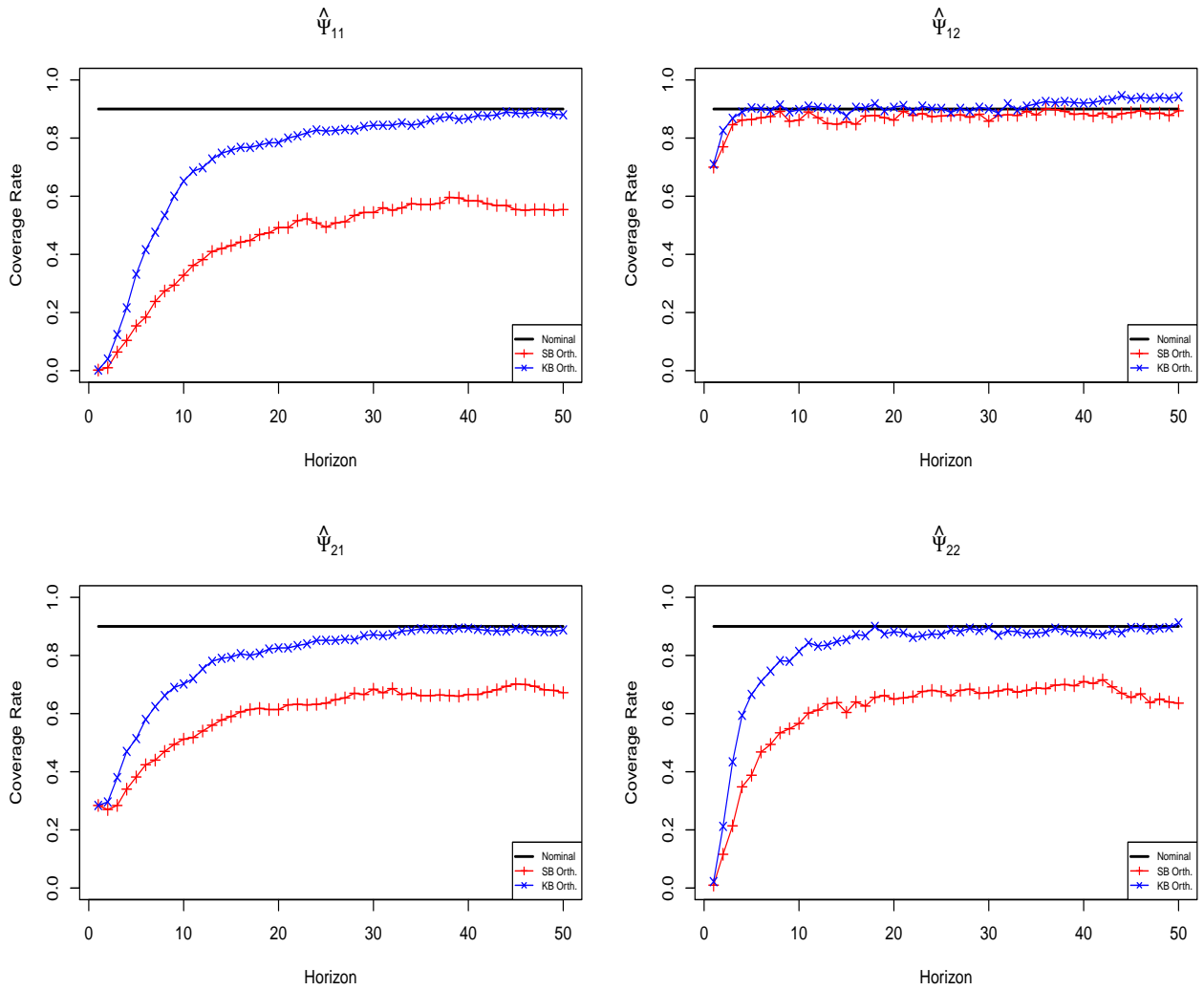


Figure 11: Effective Coverage Rates at 90% confidence intervals from Design with $\beta = 0.5$, $T = 250$; with the Impulse Responses estimated from Hybrid VARs. For lags $h = 1, 2, \dots, 10$ use $p_h = \lfloor 0.25 \ln(T)^2 \rfloor = 8$ and for lags $h = 11, 12, \dots, 25$ use $p_h = \lfloor 0.50 \ln(T)^2 \rfloor = 15$.

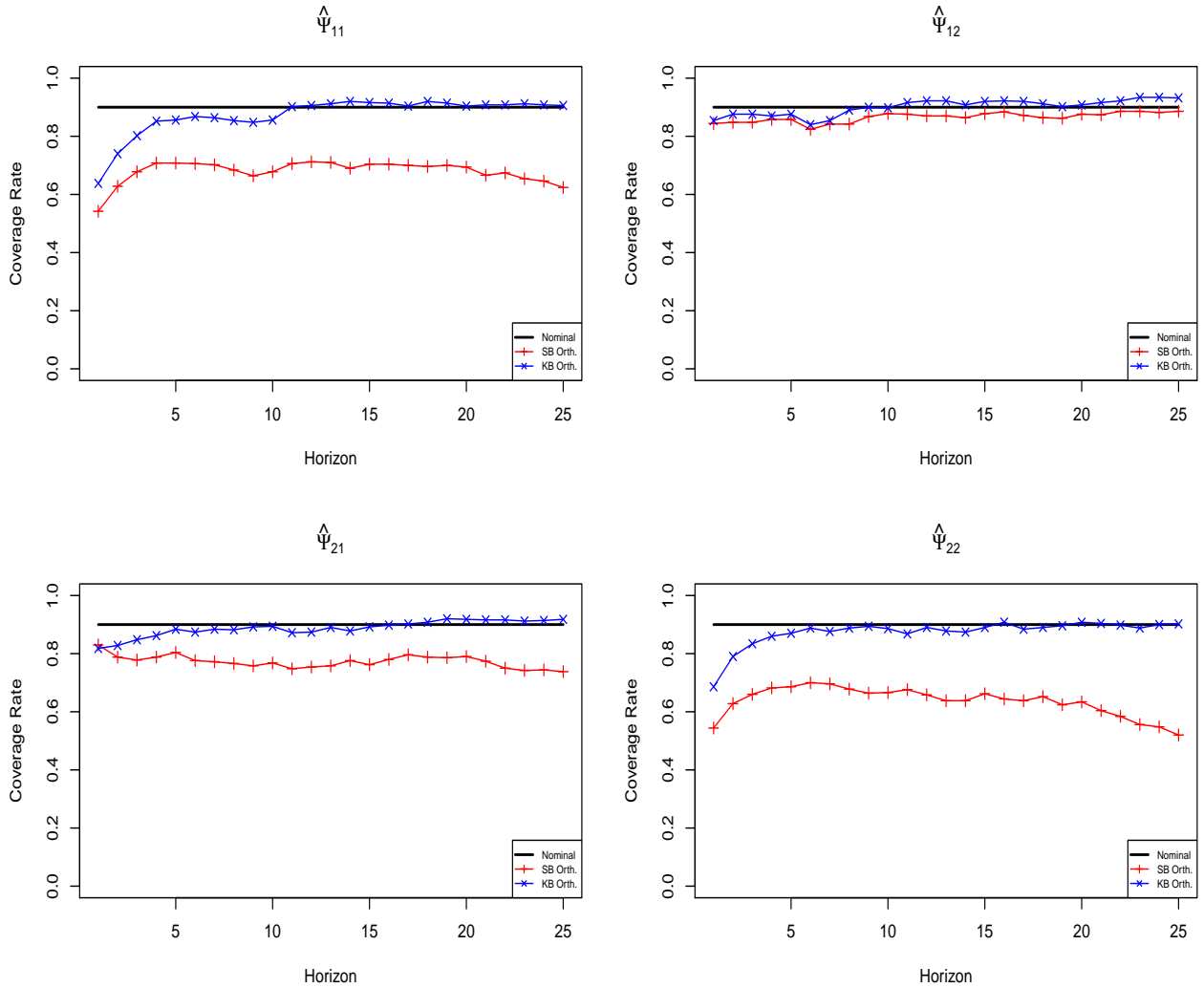
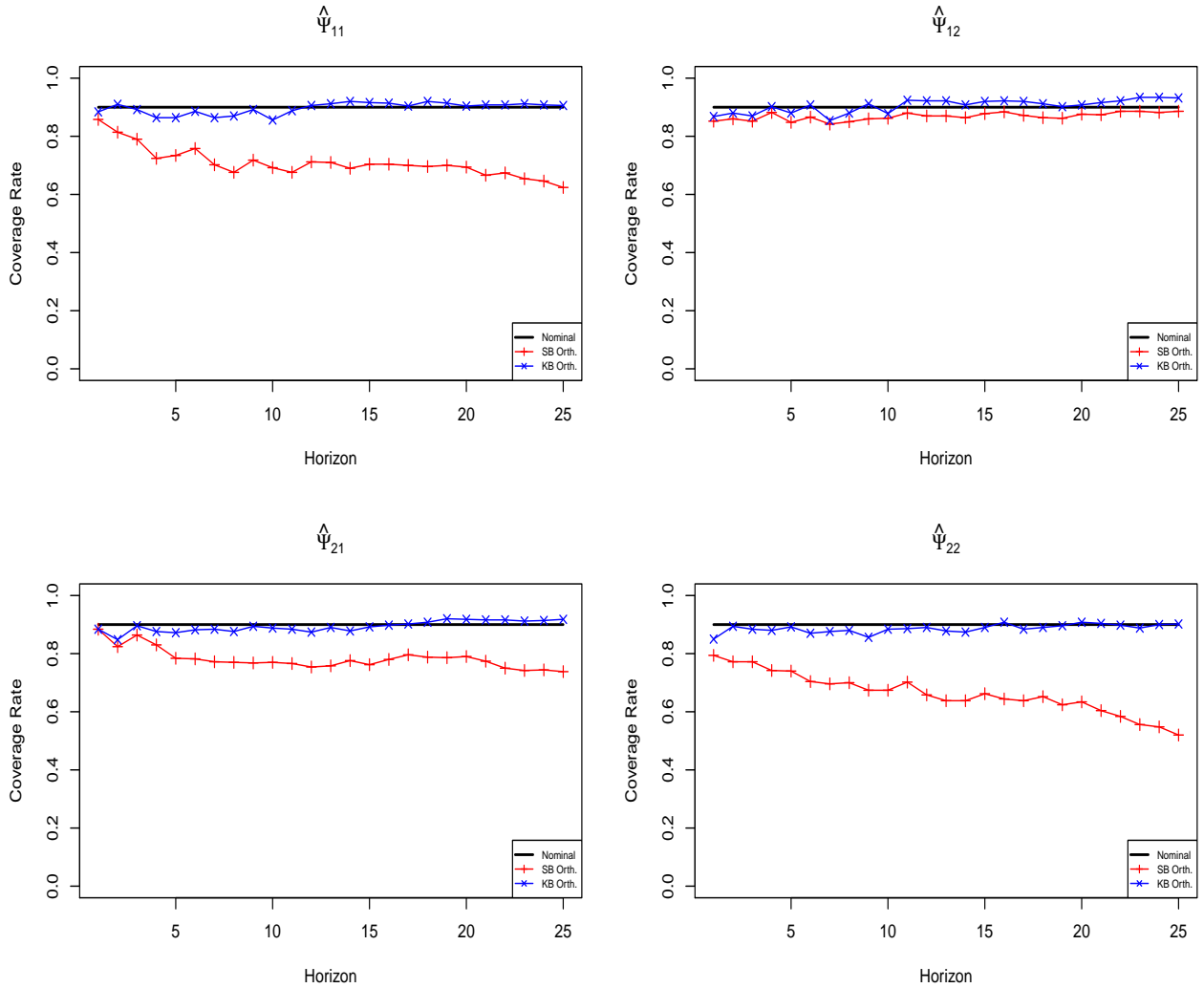
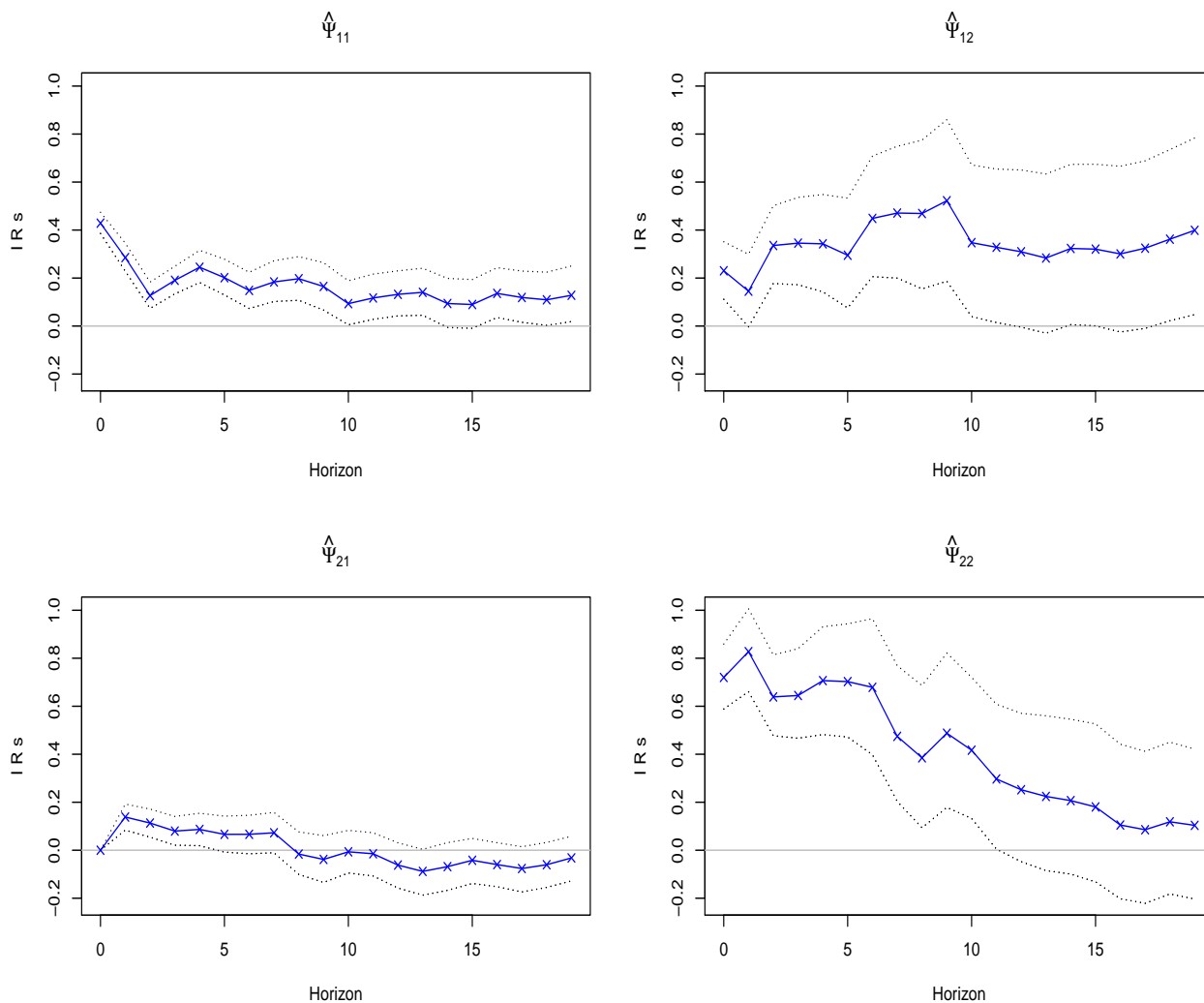


Figure 12: Effective Coverage Rates at 90% confidence intervals from Design with $\beta = 0.5$, $T = 500$; with the Impulse Responses estimated from Hybrid VARs. For lags $h = 1, 2, \dots, 10$ use $p_h = (h + 1)$ and for lags $h = 11, 12, \dots, 25$ use $p_h = \lfloor 0.50 \ln(T)^2 \rfloor = 15$.



Empirical Application: Figures

Figure 13: US



Key to all Figures: The line with "x" represents the orthogonalized *IRs* from a Sieve *VAR* bootstrap using Kilian (1998) bias correction (KB Orth.) along with the 90% confidence bounds (dotted lines). The reported results are produced using different lag orders for short and long horizons: $p_{T,h=\{1,\dots,10\}} = h + 1$, and $p_{T,h=\{11,\dots,20\}} = \lfloor 0.50 \ln(T)^2 \rfloor = 14$. Kilian (1998) corrected parameter estimates are obtained using 1,000 bootstrap resamples. The *IRs* are obtained using 2,000 bootstrap resamples.

Figure 14: UK

