Inference for Impulse Response Functions From Multivariate Strongly Persistent Processes

Richard T. Baillie
Departments of Economics and Finance, 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

February 1, 2013

Abstract

This paper considers a multivariate system of strongly persistent time series and investigates the most appropriate way for estimating the Impulse Response Function (IRF) and their associated confidence intervals. The paper extends the univariate analysis recently provided by Baillie and Kapetanios (2013), and uses a non parametric, time domain estimator, based on a vector autoregressive (VAR) approximation. This VAR is shown to have good theoretical and small sample properties for the estimation of the IRF. The paper also advocates a generic sieve VAR bootstrap for estimating confidence intervals for the estimated IRF. This is shown to be a valid method for conducting inference on the IRF, and is proven under mild assumptions. The theoretical and Monte Carlo findings in this paper indicate that a good strategy for analyzing IRF is to estimate by semi parametric VAR approximations, and to use the sieve VAR bootstrap for estimating confidence intervals. One of the great attractions of the methodology is that it is simple to apply and avoids specification and estimation issues for multivariate time series models. Two empirical examples on predator/prey series and realized volatility are also included.

Key Words: Persistence, Impulse Response Function, Autoregressive Approximation, Sieve VAR, Confidence Intervals, Realized Volatility, Mink-Muskrat series.

JEL Codes: C22, C12.

1 Introduction

Impulse Response Functions (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 IRF from a stationary dynamic simultaneous equation model. This approach was further developed by Schmidt (1974) to derive the asymptotic distribution of predictions from such models. This work led to the derivation of asymptotic distributions of estimated $IRFs$ from Vector Autoregressive (VAR) models in articles by Baillie (1987) and Lutkepohl (1988, 1989). While these theoretical results are quite elegant they essentially rely on a linearized Taylor series expansion of the estimated IRF around the true parameter estimates. Some papers such as Sims (1986), Kilian and Chang (2000) have questioned the applicability of the results to obtaining meaningful confidence intervals for the estimated IRF. The variability and confidence intervals are known to be poorly estimated through the "delta method" and asymptotic distribution theory when the data generating process has very persistent autocorrelation, or has near unit root behavior. The delta method is further discussed in Section 3.

This paper focuses on the problem of inference for estimating the IRF and their confidence intervals in a multivariate setting where the individual processes are generated by strongly persistent as well as weakly persistent processes. Although this does not cover the unit root case, it nevertheless covers many practical problems in financial economics and also empirical macroeconomics. The necessary technical machinery builds on the methodology for the univariate case developed by Baillie and Kapetanios (2013), which uses a non 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 IRF. 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 IRF.

This paper develops the methodology for the multivariate situation and shows that a valid method for conducting inference on the IRF for very persistent processes can be based on estimating an approximating VAR. The validity of this approach is proven under quite mild assumptions. The findings in this paper also indicate that a good strategy for analyzing IRF is to estimate by a semi parametric VAR and to use the sieve bootstrap for estimating confidence intervals. Simulation evidence indicates this approach appears to be a very good strategy for both short, or long memory processes. One of the great attractions of the methodology is that it is relatively simple to apply and that also avoids specification and estimation issues that can make multivariate time series modeling difficult to apply in practice. Hence for the purpose of purely estimating $IRFs$ and their associated confidence intervals, the various difficulties related with the identification and specification of multivariate $ARMA$ models alluded to by Tiao and Tsay (1989), Kapetanios, Pagan and Scott (2007) and Poskitt (2011), do not arise.

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 IRF and the problems with the theory in the presence of persistent 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 IRFs and the Sieve VAR for the derivation of confidence intervals of the estimated $IRFs$. Section 6 provides the empirical applications of the new methodology developed in this paper to (i) the well known bivariate mink and muskrat time series data, which has been the subject of many previous investigations by time series analysts, and (ii) daily realized volatility in currency markets. Section 7 summarizes the conclusions.

2 Basic Theory

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

$$\boldsymbol{y}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\epsilon}_{t-j},$$

where $\boldsymbol{\epsilon}_t$ is an unobserved, vector white noise process, such that $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$, $E\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{/}\right) = \boldsymbol{\Omega}$ which is an m dimensional, positive semi definite, covariance matrix and $E\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_s^{/}\right) = \mathbf{0}$ for $t \neq s$. The sequence of IRF or Wold Decomposition matrices are defined such that $\boldsymbol{\Psi}_0 = \mathbf{I}$, and that $\boldsymbol{\Psi}_j$ is a sequence of $m \times m$ matrices of constants. On defining $\boldsymbol{\Psi}(\mathbf{L}) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j L^j$ then the spectral density function is $\mathbf{f}_y(\omega) = \frac{1}{2\pi} \boldsymbol{\Psi}(e^{i\omega}) \boldsymbol{\Omega} \boldsymbol{\Psi}(e^{-i\omega})$. It is also assumed that

$$\sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\Omega} \boldsymbol{\Psi}_j^{/} < \infty.$$

Let $\|.\|$ denote the Euclidean matrix norm. For subsequent analysis, the following assumptions are invoked:

Assumption 1 *is in two parts: (i)* $\boldsymbol{\epsilon}_t$ *is an* m *dimensional ergodic martingale difference sequence, so that* $E(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, ...) = \mathbf{0}$*, and* $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{/} | \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, ...) = \boldsymbol{\Omega}$ *and the third and fourth moments are matrices of finite constants.*

Assumption 2 $\boldsymbol{\Psi}(L) = \mathbf{D}(L)^{-1} \widetilde{\boldsymbol{\Psi}}(L)$*, 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* $i = 1, 2, ..m$. $\widetilde{\boldsymbol{\Psi}}(\mathbf{L}) = \sum_{j=0}^{\infty} \widetilde{\boldsymbol{\Psi}}_j L^j$ *and* $\sum_{j=0}^{\infty} \left\| \widetilde{\boldsymbol{\Psi}}_j \right\| < \infty$ *. Furthermore,* $\boldsymbol{\Phi}(z) \equiv \boldsymbol{\Psi}(L)^{-1} = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_j z^j$ *exists for all* $|z| \leq 1$ *.*

Hence, the above class of processes is very wide and includes all vector linear time series processes considered in the existing literature, and encompasses long memory processes. The leading univariate case is the basic $ARFIMA$ model of Granger (1980) and Granger and Joyeux (1980). The multivariate version of the $ARFIMA(p,d,q)$ model is the $MVARFIMA(p,d,q)$ which is given by

$$\mathbf{\Pi(L)D(L)y}_t = \mathbf{\Theta}(L)\varepsilon_t$$

where $\widetilde{\mathbf{\Psi}}(\mathbf{L}) = \mathbf{\Pi}(\mathbf{L})^{-1}\mathbf{\Theta}(L)$, $\mathbf{\Pi}(L) = \sum_{j=0}^{p}\mathbf{\Pi}_j L^j$, and $\mathbf{\Theta}(L) = \sum_{j=0}^{q}\mathbf{\Theta}_j L^j$. The population quantities of the multivariate version of this model are to be found in Sowell (1986) and Chung (2002). Note that the standard assumption is that the process is weakly stationary with exponentially decaying $IRFs$ with $\mathbf{\Psi}_j = O\left(\mathbf{C}^j\right)$ 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 in the $IRFs$, so that the $\|\mathbf{\Psi}_j\| = O\left(j^{d-1}\right)$ for $0 \leq d < 0.5$ as in Chung (2002). This implies that $\left\|\sum_{j=0}^{\infty}\mathbf{\Psi}_j\mathbf{\Psi}_j'\right\| < \infty$.

In many practical applications of $VARs$ and of IRF analysis, there can be an issue of identification and the desire to obtain the $IRFs$ is response to a standardized shock with variance equal to the identity matrix rather than $\mathbf{\Omega}$. Hence an investigator may wish to provide estimates of $\left\{\mathbf{\Psi}_j\mathbf{\Omega}^{1/2}\right\}_{j=1}^{h}$ rather than $\{\mathbf{\Psi}_j\}_{j=1}^{h}$. Since $\mathbf{\Omega}^{1/2}$ is not unique, then for a given $\mathbf{\Omega}$, it is necessary to provide further identifying assumptions; e.g. Bernanke, Boivin and Eliasz (2005). See chapter 4 of Canova (2007) for a discussion of this literature. This paper abstracts from this issue, which can be handled in any of the ways discussed in the relevant literature, for example see Baillie (1987) and Lutkepohl (1988) for the $VARMA$ case. Hence this paper focuses on providing estimates of $\{\mathbf{\Psi}_j\}_{j=1}^{h}$.

3 Asymptotic Distributions of Estimated IRF

One standard parametric approach has been to specify a stationary and invertible $MVARFIMA$ model such that $\Pi(\mathbf{L})$ and $\mathbf{\Theta}(L)$, and therefore, $\mathbf{\Psi}(\mathbf{L})$ are parametrized so that they are functions of a vector of parameters given by θ, where θ contains d_i, $i = 1, 2, ..m$, and to estimate θ by either approximate or full MLE so that

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

Note that $\boldsymbol{\theta}_0$ denotes the true value of $\boldsymbol{\theta}$, and the symbol $\xrightarrow{L}$ denotes convergence in distribution. The theory for full and approximate MLE of multivariate $ARFIMA$ models is described by Sowell (1986, 1992a) and Chung (2002)[1]. For examples of some applications of the methodology, see Diebold, Husted and Rush (1991) for the univariate case and see Sowell (1992b), Baillie and Chung (2002) and Jensen (2009) for the multivariate case. In order to obtain confidence intervals for the estimated vectorized IRF matrix at lag j, i.e. $vec(\mathbf{\Psi}_{j,\hat{\theta}})$, where the subscript $\hat{\theta}$ denotes

[1]These reults have been explored in more detail for the univariate m=1 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.

dependence of $\boldsymbol{\Psi}_j$ on $\hat{\theta}$, a traditional method has been to use the delta method, which is based on a linearized Taylor Series expansion around the true vectorized IRF at lag j, $vec(\boldsymbol{\Psi}_{j,\theta_0})$. On further assuming that ϵ_t is an i.i.d. sequence, that $\sum_{j=1}^{\infty} \sup_\theta |\boldsymbol{\Psi}_{j,\theta}| < \infty$ and that $\mathbf{V}$, defined above is nonsingular, then for all $j = 1, ..., h$

$$\sqrt{T}\left(vec(\boldsymbol{\Psi}_{j,\hat{\theta}}) - vec(\boldsymbol{\Psi}_{j,\theta_0})\right) \xrightarrow{L} N(0, \mathbf{D}_j'\mathbf{V}^{-1}\mathbf{D}_j) \tag{2}$$

where $\mathbf{D}_j = \left.\frac{\partial vec(\boldsymbol{\Psi} j,\theta)}{\partial \theta}\right|_{\theta=\theta_0}$. For stationary and invertible $VARMA$ models, there are parametric expressions available for the matrix $\mathbf{D}_j$. However, no corresponding results are yet available for the $MVARFIMA$ model.

The above traditional methodology is known to deliver poor results in the presence of very persistent processes, such as long memory processes and also near unit root processes; see Wright (2000) and Kilian (1998a, 1999). Hence this approach which has been shown to have poor properties in persistent univariate processes, is unlikely to possess desirable properties for persistent multivariate processes.

4 Sieve VAR Bootstrap

Given the results in Baillie and Kapetanios (2013) for the persistent univariate 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 setting. Hence, this paper considers approximating a multivariate persistent 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

$$\boldsymbol{y}_t = \sum_{j=1}^{p_T} \boldsymbol{\Phi}_j^{(p_T)} \boldsymbol{y}_{t-j} + \tilde{\boldsymbol{v}}_t.$$

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

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

for all sequences $\{p_T\}$ such that $p_T \to \infty$ and $p_T = o(T^\alpha)$ for all $\alpha > 0$. An acceptable sequence for p_T is $p_T = [\ln(T)]^\alpha$ where $\alpha > 1$; [.] denotes the integer part. Also, from an extension of Baxter's inequality proven by Inoue and Kasahara (2006)

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

as long as $p_T \to \infty$. The selection of p_T can be done through a data dependent method such as an information criterion (e.g., AIC, or BIC) or by some deterministic rule such as $p_T = [\ln(T)]^2$. Then, IRF analysis proceeds by inverting the polynomial defined by $\widehat{\mathbf{\Phi}}_j^{(p_T)}$, $j = 1, ..., p_T$ to give $\widehat{\mathbf{\Psi}}_j^{(p_T)}$. Inference on $\widehat{\mathbf{\Psi}}_j^{(p_T)}$ can be carried out by means of the bootstrap. A sieve bootstrap, that is a direct extension of the univariate sieve bootstrap of Baillie and Kapetanios (2013), can be implemented using the following algorithm:

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

2. Invert $\widehat{\mathbf{\Phi}}_j^{(p_T)}(z) = \sum_{j=1}^{p_T} \widehat{\mathbf{\Phi}}_j^{(p_T)} z^j$ to obtain estimates of the IRF given by $\widehat{\mathbf{\Psi}}_j^{(p_T)}$, $j = 1, ..., h$.

3. Re-center $(\hat{\boldsymbol{\epsilon}}_1, ..., \hat{\boldsymbol{\epsilon}}_T)'$.

4. Re-sample with replacement from this vector to obtain the bootstrap sample of error terms given by $(\boldsymbol{\epsilon}_1^*, ..., \boldsymbol{\epsilon}_T^*)'$.

5. Use the above quantities together with $\widehat{\mathbf{\Phi}}_j^{(p_T)}$, $j = 1, ..., p_T$, to generate the bootstrap sample $(\boldsymbol{y}_1^*, ..., \boldsymbol{y}_T^*)'$.

6. Estimate a $VAR(p_T)$ for $(\boldsymbol{y}_1^*, ..., \boldsymbol{y}_T^*)'$ to obtain the bootstrap estimated autoregressive coefficients given $\hat{\kappa}_j^{*,(p_T)}$, $j = 1, ..., p_T$.

7. Invert $\widehat{\mathbf{\Phi}}^{*,(p_T)}(z) = \sum_{j=1j}^{p_T} \widehat{\mathbf{\Phi}}^{*,(p_T)} z^j$ to obtain bootstrap estimates of the impulse responses given by $\widehat{\mathbf{\Phi}}_j^{*,(p_T)}$, $j = 1, ..., h$.

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

A further requirement for the valid application of the bootstrap requires an extension of Theorem 1 in Poskitt (2007) to the autocovariances of multivariate processes. In particular

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

where $\mathbf{\Gamma}(\tau) = E\left(\boldsymbol{y}_t \boldsymbol{y}_{t+\tau}'\right)$ and $\mathbf{C}_T(\tau) = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}_t \boldsymbol{y}_{t+\tau}'$. The proof of Theorem 1 of Poskitt (2007) proceeds in the multivariate case exactly as in the univariate case given the result that

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

which can be obtained from a straightforward extension of Theorems 3 and 4 of Hosking (1981). In fact, this multivariate version of the result is provided by Corollary 1 of Chung (2002). This essentially fulfills the requirements for extending the univariate results in Theorem 3.3 of Baillie and Kapetanios (2013) to the multivariate framework of this paper and establishes the validity of the sieve bootstrap given in the above algorithm.

5 Simulation Design

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 that is used 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 can be expressed as

$$\mathbf{D}(\mathbf{L})\mathbf{y}_t = \mathbf{u}_t$$

$$\mathbf{u}_t = \mathbf{\Phi}\mathbf{u}_{t-1} + \varepsilon_t.$$

where $\mathbf{D}(\mathbf{L})$ is an m dimensional square diagonal matrix and is constrained for convenience to have the same long memory parameter across equations so that $\mathbf{D}(\mathbf{L}) = (\mathbf{1}-\mathbf{L})^d \mathbf{I}_m,$ 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. The following three designs were used with the short memory component being weakly persistent, moderately persistent and strongly, or extremely persistent respectively.

Design 1 has the $VAR(1)$ matrix

$$\mathbf{\Phi} = \begin{pmatrix} 1/3 & -1/6 \\ -1/3 & 1/2 \end{pmatrix}$$

which has an associated determinental lag operator polynomial given by $|\mathbf{I} - \mathbf{\Phi}L| = 1-(5/6)\,L+ (1/9)\,L$ which has roots of 6 and 1.5; so that the process is stationary and has eigenvalues of 0.17 and 0.667. This process is clearly stationary with low persistence.

Design 2 has the following $VAR(1)$ coefficient matrix of

$$\mathbf{\Phi} = \begin{pmatrix} 0.7 & 0.4 \\ 0.6 & -0.3 \end{pmatrix}$$

and $|\mathbf{I} - \mathbf{\Phi}L| = 1 - 0.40L - 0.45L^2$ which has roots of 1.11 and -2.00; so that the process is stationary and has eigenvalues of 0.9 and -0.5; and therefore is moderately persistent.

Design 3 has the following $VAR(1)$ coefficient matrix

$$\mathbf{\Phi} = \begin{pmatrix} 0.70 & 0.5 \\ 0.6 & -0.25 \end{pmatrix}$$

and $|\mathbf{I} - \mathbf{\Phi}L| = 1 - 0.45L - 0.475L^2$ which has roots of 1.053 and -2.00; so that the process is stationary and has eigenvalues of 0.95 and -0.5; which implies quite extreme persistence.

For each design, $\mathbf{\Omega}$ was specified to be diagonal, so that the innovation processes are contemporaneously uncorrelated; and hence

$$y_{i,t} = (1-L)^{-d} u_{it}$$

for $i = 1, 2$ and with the same value of d assumed across the two series. The designs were chosen for $d = \{0.2, 0.4\}$ and for sample sizes of $T = \{250, 500, 1000\}$. The number of bootstraps was 599 and $2,000$ Monte Carlo replications were carried out; and the Kilian (1998a) correction was applied using $2,000$ bootstrap replications. The results for the coverage rates for 90% confidence level are reported in Figures 1 through 12. Each figure is divided into four parts. Starting with the top left corner and continuing in a clockwise direction, the first panel depicts the coverage rates for the IRF 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 terminology VAR denotes the sieve $VAR(p_T)$ bootstrap, where $p_T = [\ln(T)]^2$, and $VAR - K$ denotes the sieve $VAR(p_T)$ bootstrap,$p_T = [\ln(T)]^2$, using the Kilian (1998a) correction and $AIC - K$ denotes the sieve VAR bootstrap, where P is chosen by the AIC, using the Kilian (1998a) correction. The maximum order for the AIC is set as $p_T = [\ln(T)]^2$.

In reporting the results the designs for $d = 0.2$ have been excluded; partly for reasons of conserving space, and also since it transpired that they did not produce the degree of persistence to be a real challenge to the methodology proposed in this paper. Hence, all the reported results in Figures 1 through 12 relate only to the case of $d = 0.4$.

Starting with Figure 1, it can be seen that for Design 1, and with $T = 250$, both VAR and $VAR - K$ are close to the nominal 90% most of the time, especially after the first ten lags. As expected the higher sample size of $T = 500$ for the same Design 1 shown in Figure 2 provides improved coverage rates for both the VAR and $VAR - K$. In fact the nominal rate is reached after only the first three lags. This result is further strengthened for the sample of $T = 1,000$ observations in Figure 3. It is worth noting that there are minor differences between VAR and $VAR - K$ where VAR seems to be slightly better in short horizons and $VAR - K$ has slightly better performance for long lags; e.g. see the Figure 3 response of y_2 to y_2.

Figures 4, 5, and 6 present the results from simulation Design 2 using $d = 0.4$. In this case the use of the $AIC - K$ results in superior estimates of the IRF in all three sample sizes. The quality of the estimates only deteriorates after the first thirty five lags. Particularly, in Figure 6, where the sample size of $T = 1,000$ is illustrated, it is obvious that $AIC - K$ is closer to the nominal level in all cases. However, we see that in the response of y_1 to y_1 $VAR - K$ crosses the coverage rate of $AIC - K$ from below and approaches the nominal rate after the first twenty seven lags. However, in the other responses that same fact occurs after the first twenty lags. Furthermore, analyzing the response of y_2 to y_2 it is clear that $AIC - K$ decreases for longer lags and the coverage rate line of $VAR - K$ crosses $AIC - K$ after the first ten lags.

Finally, the results from the last and very persistent Design 3 are illustrated in Figures 7 through 9 for the case of $d = 0.4$. The extreme persistence of the short memory $VAR(1)$ in this design is clearly a more challenging situation for the Sieve bootstrapping methodology presented in this paper. It can be seen that the $AIC - K$ provides coverage rates closest to the nominal value and the choice of $p_T = [\ln(T)]^2$ seems to not lead in any significant results.

Another important fact that should be noted in all the above designs is the effect of the Kilian (1998a) correction. Focusing particularly in the small sample size of $T = 250$ where the correction is more meaningful, it can be seen that for Design 1, the bias correction does not add any significant value. However, on moving to the more persistent systems in Designs 2 and 3, the differences between VAR and $VAR-K$ are obvious, with $VAR-K$ providing by far better coverage rates compared to VAR.

While some of the coverage rates are poor for the small sample size of $T = 250$, the methodology nevertheless works extremely well for medium and large samples. In general the $VAR-AIC-K$ is seen to provide superior estimates and coverage rates.

6 Empirical Applications

6.1 Example of Mink and Muskrat

A number of previous studies have considered the bivariate process of the annual numbers of mink and muskrat trapped by the Hudson Bay Company in northern Canada. The series are considered to represent a proxy for the population numbers of these animals since the trapping activity is considered to be constant over time. The series have fascinated time series analysts for several reasons. First, the series display strong components of periodicity between nine and eleven years which have been represented in articles by Bulmer (1974) and Chen and Wallis (1978) by high order AR processes with complex conjugate roots. Secondly, the series are an example of predator prey relationships with mink being a predator of muskrat. This feature has been confirmed by the presence of a negative estimated coefficient in the off diagonal element of the MA coefficient matrix in estimated $VARMA(2,1)$ models by Chen and Wallis (1978) and in canonical analysis by Terasvirta (1985). A third feature of the series is the uncertainty over whether both series are weakly stationarity. This has led Bulmer (1974) to impose a unit root in the muskrat series, but not the mink series; and also led Chen and Wallis (1978) to detrend the series before estimating $VARMA$ models. The borderline decision on whether or not to difference the series has been dealt with by Terasvirta (1985) through estimating canonical $VARMA$ models; while Tong (1980) and Lim and Tong (1983) have estimated threshold AR models.

In this study the bivariate VAR models were estimated over the period 1842 through to 1957, which was a period which appeared to be stationary and also to possess very persistent autocorrelation. Given the lack of any clear theory for the ordering of the two types of innovations on mink and muskrat, the bivariate $VARs$ were estimated for both possible orderings and the sieve VARs were inverted to derive estimated IRFs and the bootstrapped sieve $VARs$ used to find associated confidence intervals. The four panels in Figure 10 show the estimated $IRFs$ and their bootstrapped confidence intervals. The estimated $IRFs$ show that both mink and muskrat respond positively to their own past innovations. The lower left hand panel indicates how recent past innovations in mink promote a negative response in current muskrat. Furthermore, the top right hand panel indicates that the innovations on past muskrats lead to positive responses in current levels of mink. Both of these off diagonal IRF responses are consistent with the basic predator prey theory of how increases in mink lead to subsequent reductions in the muskrat population; and also that increases in the muskrat population leads to subsequent increases in the mink population. Figure 11 reports corresponding analysis with the ordering of the mink and muskrat series being reversed. The results are extremely similar with the general pattern in the estimated $IRFs$ very similar to the previous figure; only with the patterns in the off diagonals being reversed as expected.

6.2 Example of Realized Volatility

Anderson, Bollerslev, Diebold and Labys (2003) consider three Realized Volatility (RV) series composed of daily observations from 1 December 1986 through 30 June 1999 for the RV of the US dollar, German Deutschmark and Japanese Yen. They conduct univariate analysis of the three RV series and estimate $ARFIMA(0, d, 0)$ models and also use a multivariate two step filtering method which uses semi parametric estimates of d from each series and then filters out the long memory component. One interesting aspect of their study is that they found no evidence of common stochastic persistence (trends) in the data in the sense that a linear combination of the three series has virtually the same order of fractional integration as the component series. Figures 12 and 13 report the estimated $IRFs$ and their bootstrapped confidence intervals for the Yen-$ and DM-$ realized volatility series. As expected the $IRFs$ for both realized volatility series, indicate significant long memory effects with past innovations having long and substantial effects on the current level of both own series' realized volatility. An interesting feature is that past innovations of the DM-$ realized volatility has persistent effects on current Yen-$ realized volatility; yet there was no significant effects from past innovations of Yen-$ realized volatility on current DM-$ realized volatility. Figure 13 presents the reverse causal orderings with very similar conclusions.

7 Conclusion

This paper considers a multivariate system of strongly persistent time series and investigates the most appropriate method for estimating the Impulse Response Function (IRF) and its 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. This approximation is shown to have good theoretical and small sample properties for the estimation of the IRF. The paper also advocates a generic sieve VAR bootstrap for estimating confidence intervals for the estimated IRF. This is shown to be a valid method for conducting inference on the IRF, and is proven under mild assumptions. The theoretical and Monte Carlo findings in this paper indicate that a good strategy for analyzing IRF is to estimate using semi-parametric VAR approximations, and to use the sieve VAR bootstrap for estimating confidence intervals. One of the great attractions of the methodology is that it is simple to apply and avoids specification and estimation issues for multivariate time series models.

References

Anderson T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71:579-625.

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

Baillie, R. T., Chung, S.-K. (2002). Modeling and forecasting from trend stationary long memory models with applications to climatology. *International Journal of Forecasting*, 18:215-226.

Baillie, R. T., Kapetanios, G. (2008). Nonlinear models for strongly dependent processes and financial applications. *Journal of Econometrics*, 147:60-71.

Baillie, R. T., Kapetanios, G. (2013). Estimation and inference for impulse response functions form univariate strongly persistent processes. *Econometrics Journal*, forthcoming.

Benanke, B., Boivin, J., Eliasz, P. (2005). Measuring the effects of monetary policy: a factor augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 50:387-422.

Bulmer, M.G. (1974). A statistical analysis of the 10 year cycle in Canada. *Journal of Animal Ecology,* 43:701-718.

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

Chen, W.-Y. T., Wallis, K. F. (1978). Multiple time series modeling: another look at the mink-muskrat interaction. *Journal of the Royal Statistical Association, C ,* 27:168-175.

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

Diebold, F.X., Husted, S., Rush, M. (1991). Real exchange rates under the gold standard. *Journal of Political Economy*, 99:1252-1271.

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

Granger, C. W. J., Joyeux, R. (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.

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

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

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., Scott, A. (2007). Making a match: combining theory and evidence in policy orientated macroeconomic modelling. *Journal of Econometrics*, 136:565-594.

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

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

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.

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

Lim, K.S., Tong, H. (1983). A statistical approach to difference - delay equation modeling in ecology: two case studies. *Journal of Time Series Analysis,* 4:239-267.

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

Lutkepohl, H. (1989). A note on the asymptotic distribution of impulse response functions of estimated VAR models with orthogonal residuals. *Journal of Econometrics*, 42:71–376.

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

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

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. (1974). The asymptotic distribution of forecasts in the dynamic simulation of an econometric model. *Econometrica,* 42:303-309.

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.

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

SOWELL, F. (1986). Fractionally integrated vector time series. *PhD dissertation, Duke University.*

SOWELL, F. (1992a). Maximum likelihood estimation of stationary univariate fractionally-integrated time-series models. *Journal of Econometrics,* 53:165-188.

SOWELL, F. (1992b). Modeling long-run behavior with the fractional ARIMA model. *Journal of Monetary Economics,* 29:277-302.

TERASVIRTA, T. (1985). Mink and muskrat interaction: a structural analysis. *Journal of Time Series Analysis,* 6:171-180.

TIAO, G. C., TSAY, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society, B,* 51:157-213.

TONG, H. (1980). Threshold autoregression, limit cycles, and cyclical data. *Journal of the Royal Statistical Society, B* 42:245-292.

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

Figures

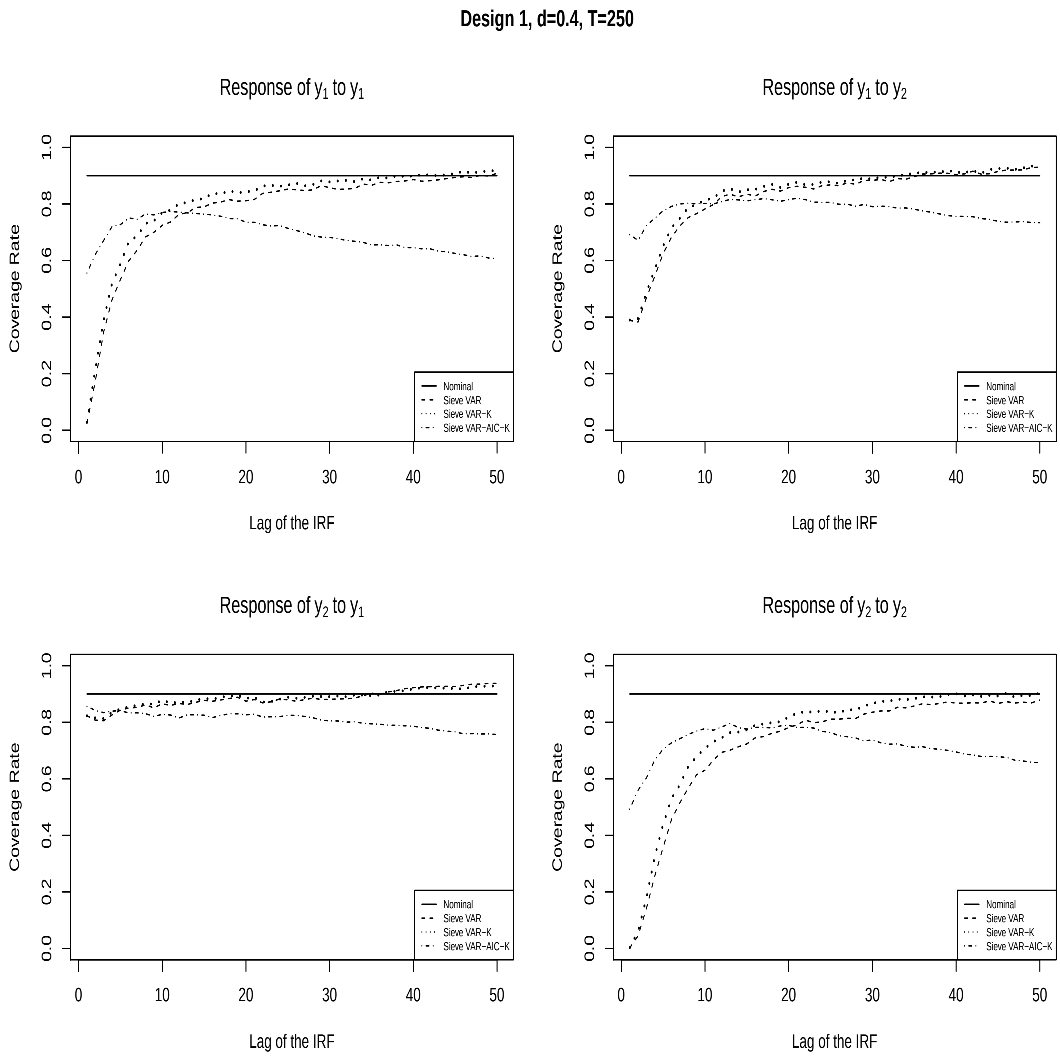

Figure 1: Design 1, $d = 0.4$, $T = 250$, Monte Carlo averages of IRWs and coverage rates.

Notes. Black solid line represents the nominal coverage rate. **Sieve VAR** denotes the sieve VAR bootstrap with $p = ln(T)^2$, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction, **Sieve VAR-AIC-K** denotes the sieve VAR bootstrap with p chosen using the Akaike's criterion and Kilian (1998a) correction.

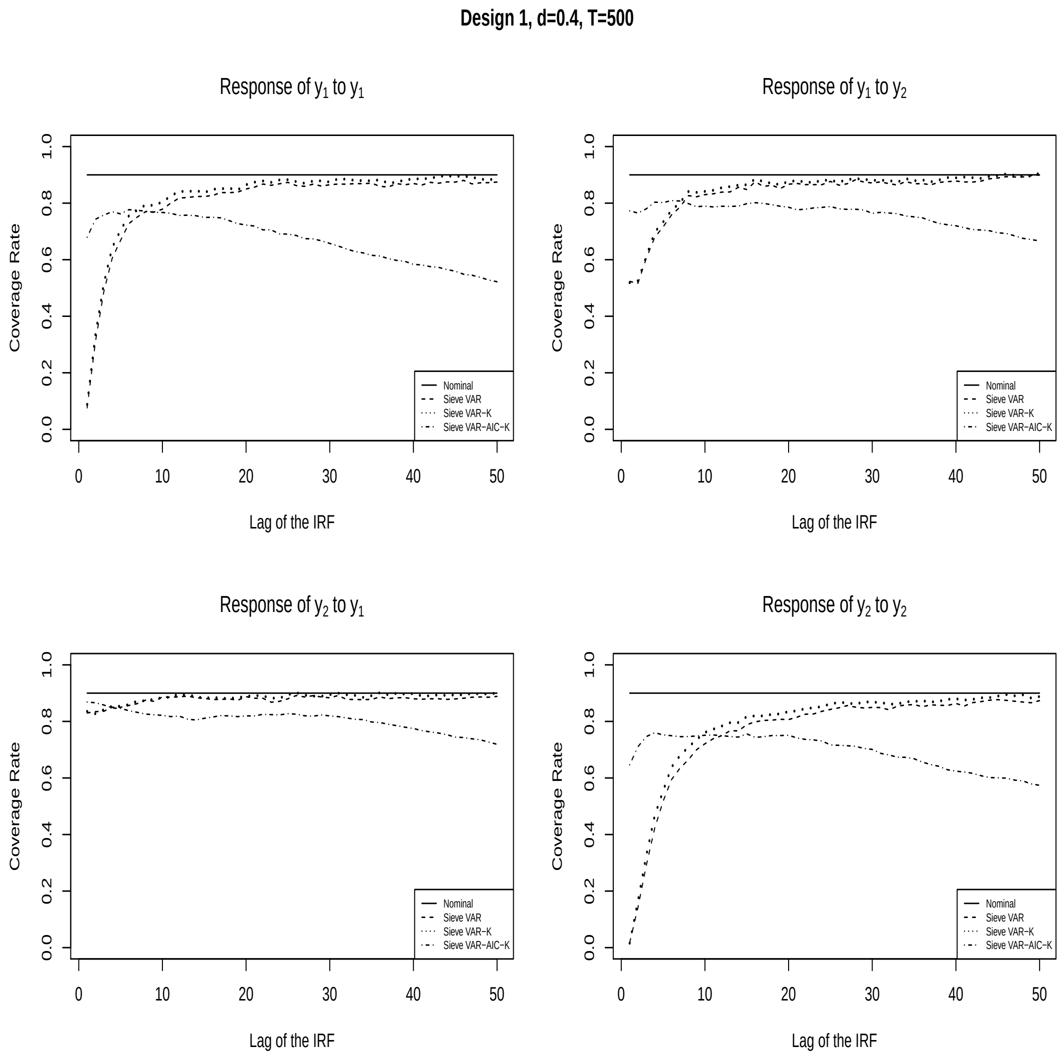

Figure 2: Design 1, $d = 0.4$, $T = 500$, Monte Carlo averages of IRWs and coverage rates.

Notes. Black solid line represents the nominal coverage rate. **Sieve VAR** denotes the sieve VAR bootstrap with $p = ln(T)^2$, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction, **Sieve VAR-AIC-K** denotes the sieve VAR bootstrap with p chosen using the Akaike's criterion and Kilian (1998a) correction.

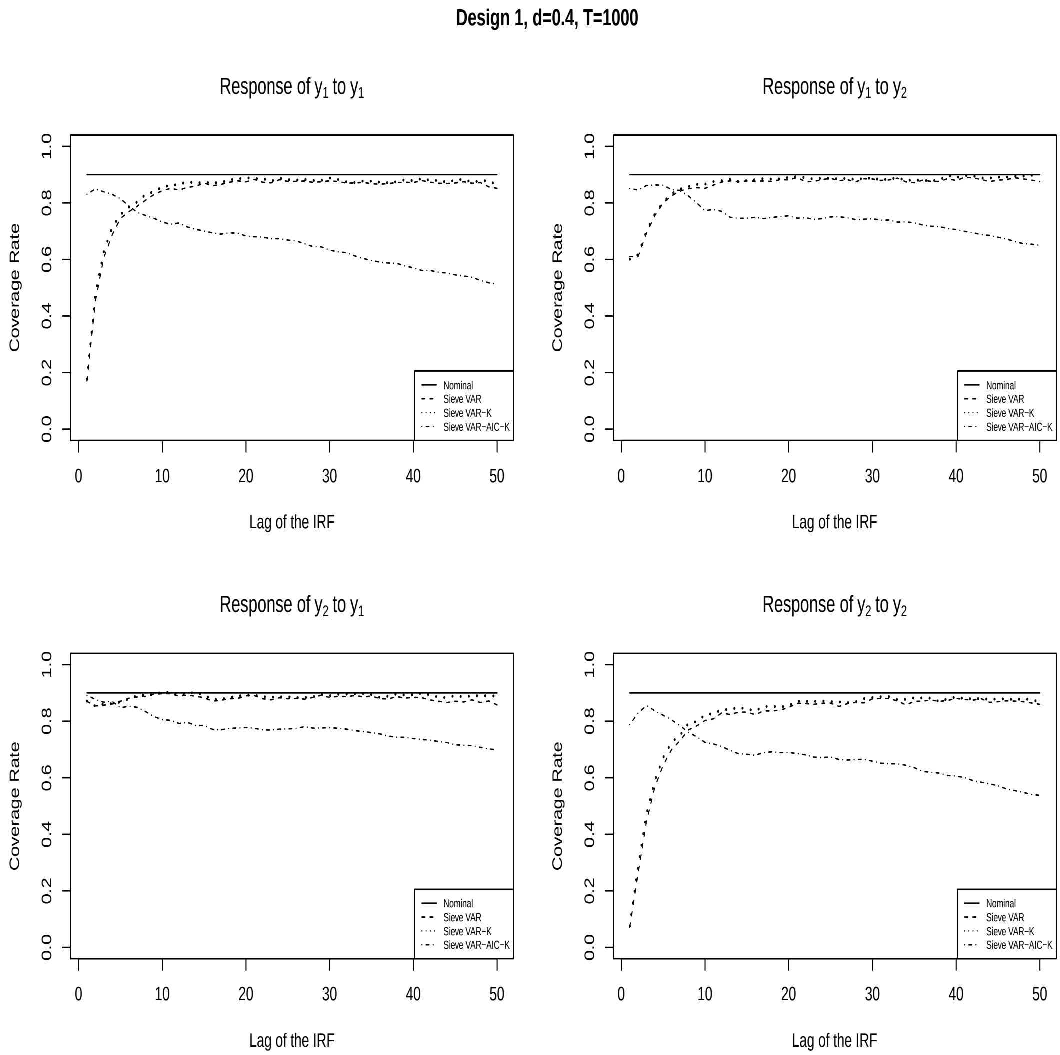

Figure 3: Design 1, $d = 0.4$, $T = 1000$, Monte Carlo averages of IRWs and coverage rates.

Notes. Black solid line represents the nominal coverage rate. **Sieve VAR** denotes the sieve VAR bootstrap with $p = ln(T)^2$, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction, **Sieve VAR-AIC-K** denotes the sieve VAR bootstrap with p chosen using the Akaike's criterion and Kilian (1998a) correction.

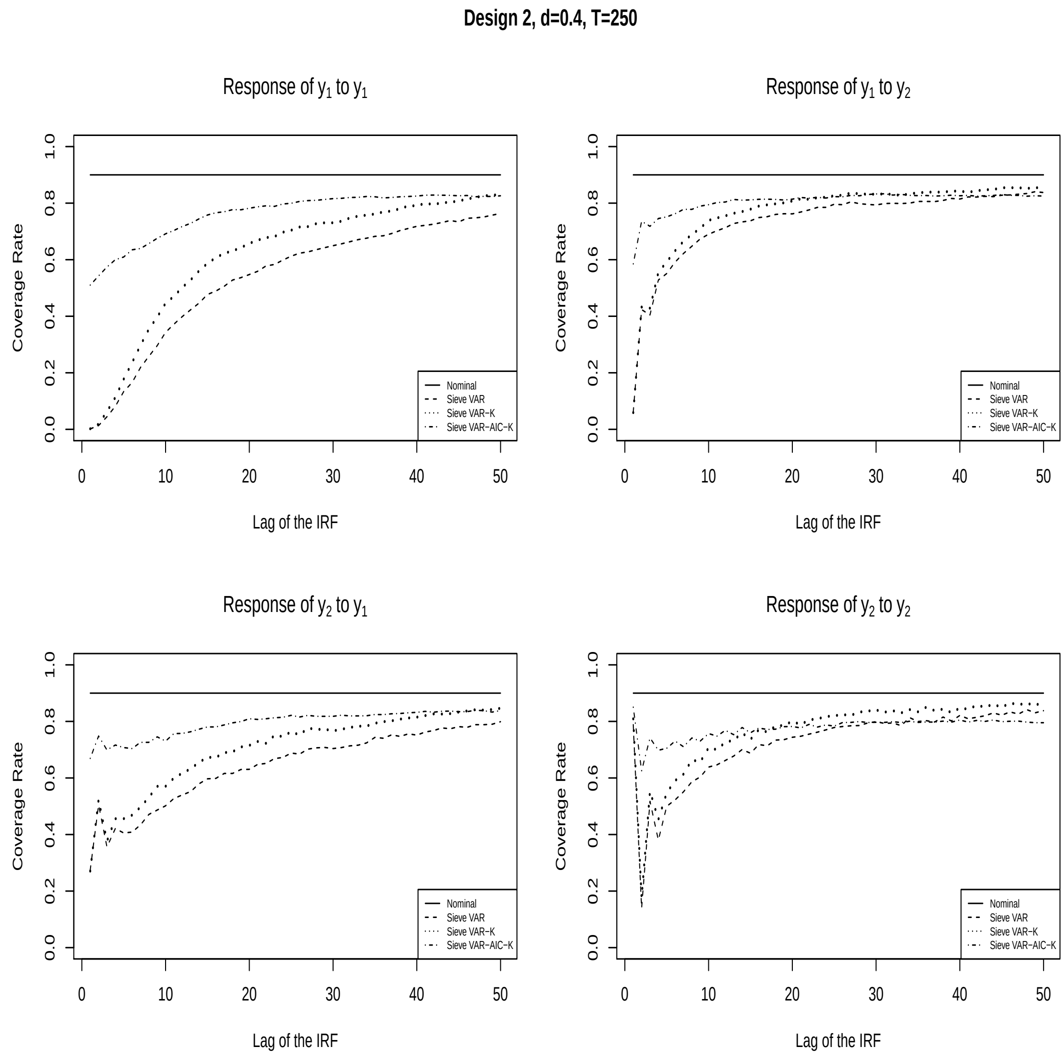

Figure 4: Design 2, $d = 0.4$, $T = 250$, Monte Carlo averages of IRWs and coverage rates.

Notes. Black solid line represents the nominal coverage rate. **Sieve VAR** denotes the sieve VAR bootstrap with $p = ln(T)^2$, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction, **Sieve VAR-AIC-K** denotes the sieve VAR bootstrap with p chosen using the Akaike's criterion and Kilian (1998a) correction.

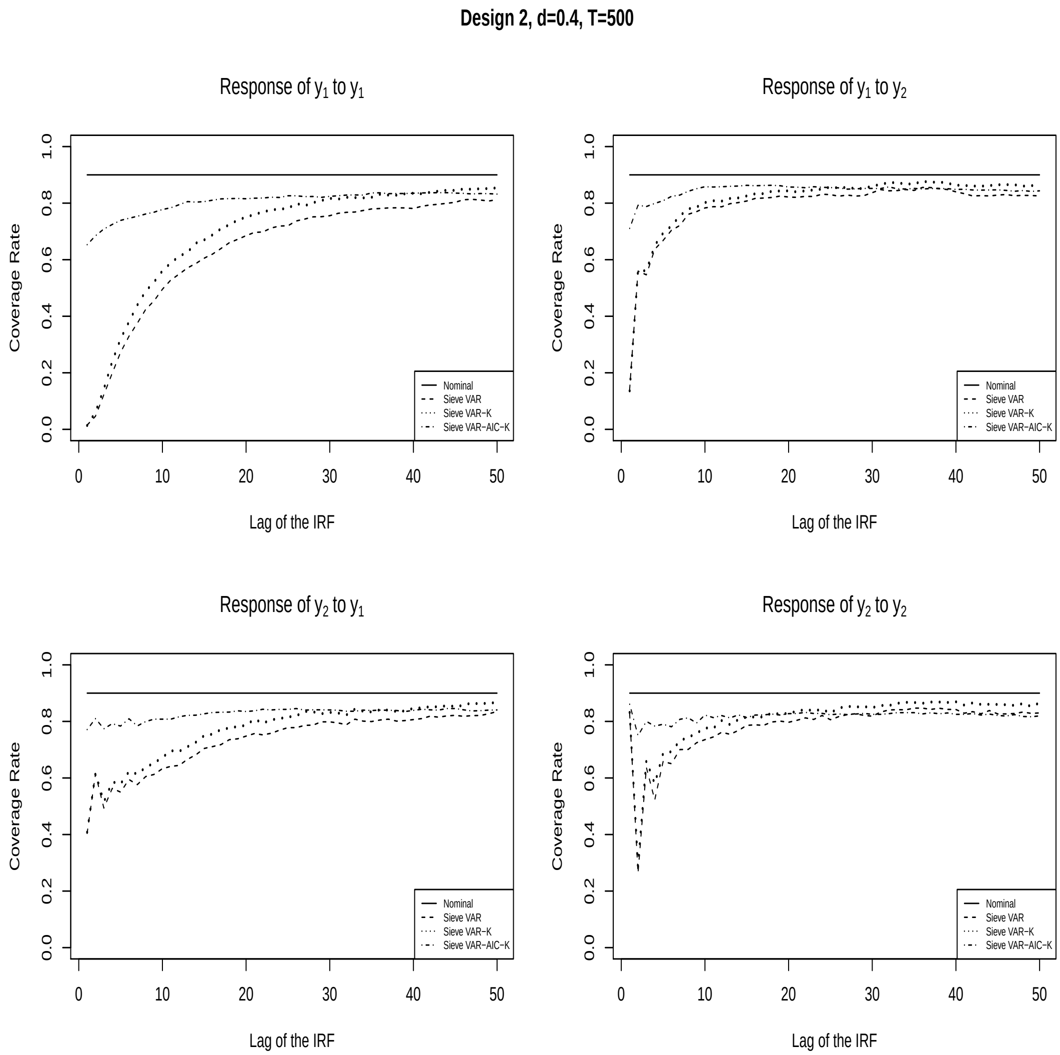

Figure 5: Design 2, $d = 0.4$, $T = 500$, Monte Carlo averages of IRWs and coverage rates.

Notes. Black solid line represents the nominal coverage rate. **Sieve VAR** denotes the sieve VAR bootstrap with $p = ln(T)^2$, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction, **Sieve VAR-AIC-K** denotes the sieve VAR bootstrap with p chosen using the Akaike's criterion and Kilian (1998a) correction.

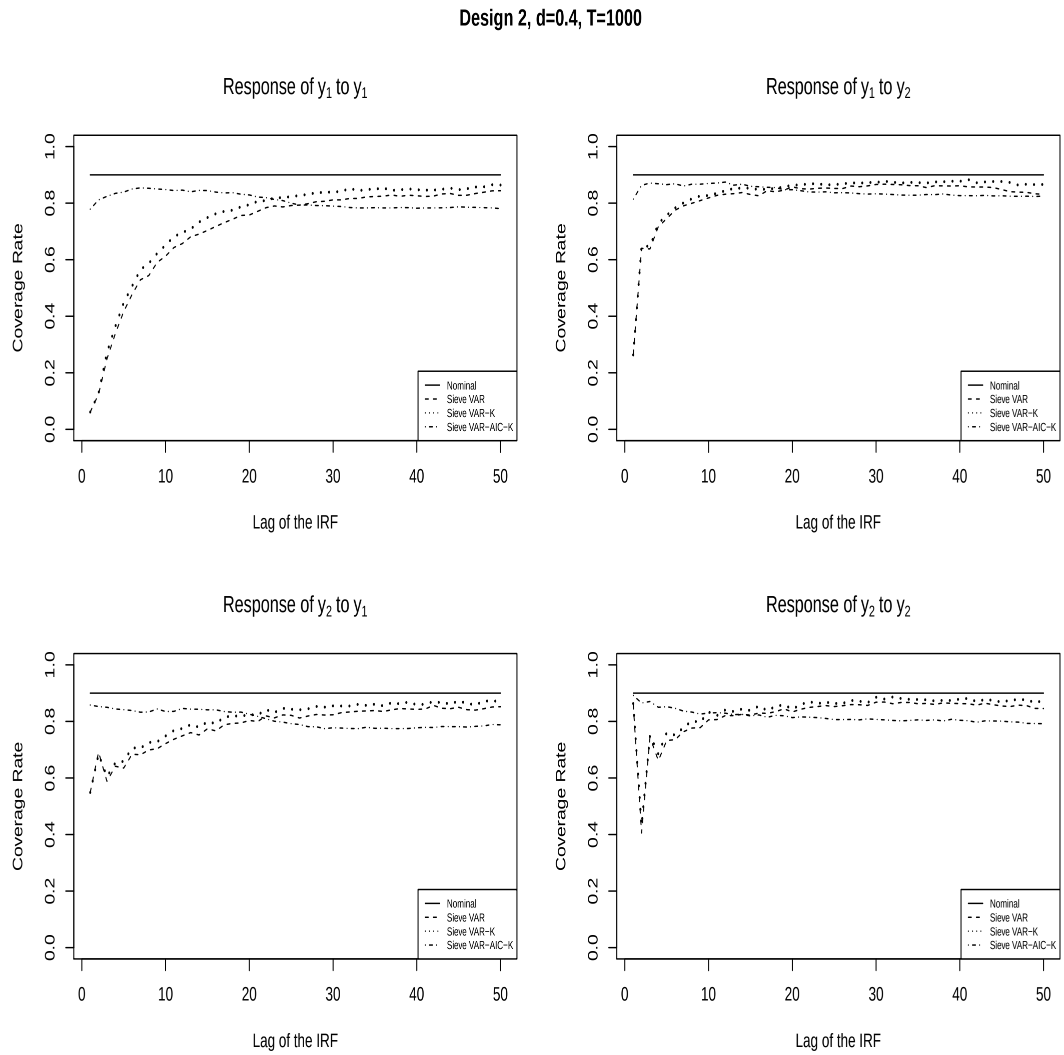

Figure 6: Design 2, $d = 0.4$, $T = 1000$, Monte Carlo averages of IRWs and coverage rates.

Notes. Black solid line represents the nominal coverage rate. **Sieve VAR** denotes the sieve VAR bootstrap with $p = ln(T)^2$, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction, **Sieve VAR-AIC-K** denotes the sieve VAR bootstrap with p chosen using the Akaike's criterion and Kilian (1998a) correction.

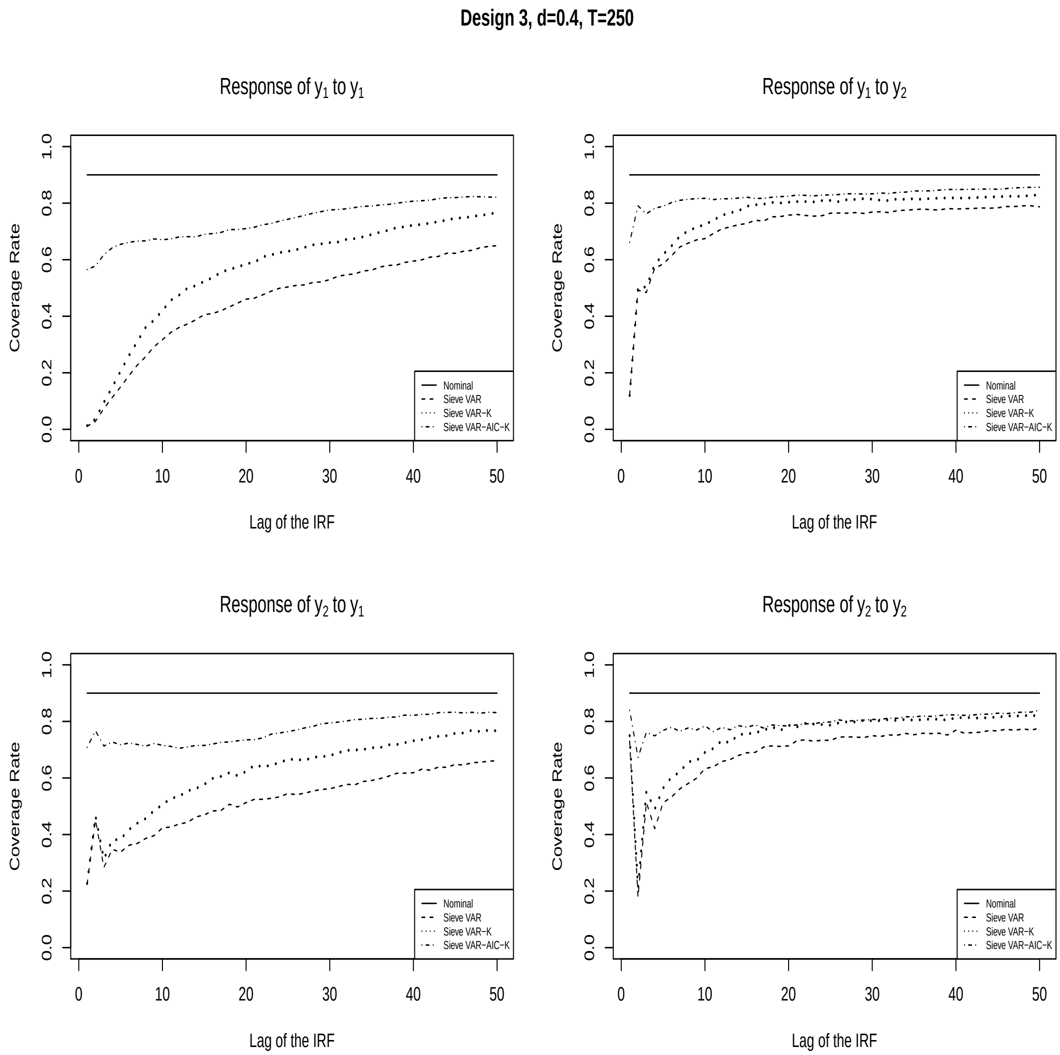

Figure 7: Design 3, $d = 0.4$, $T = 250$, Monte Carlo averages of IRWs and coverage rates.

Notes. Black solid line represents the nominal coverage rate. **Sieve VAR** denotes the sieve VAR bootstrap with $p = ln(T)^2$, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction, **Sieve VAR-AIC-K** denotes the sieve VAR bootstrap with p chosen using the Akaike's criterion and Kilian (1998a) correction.

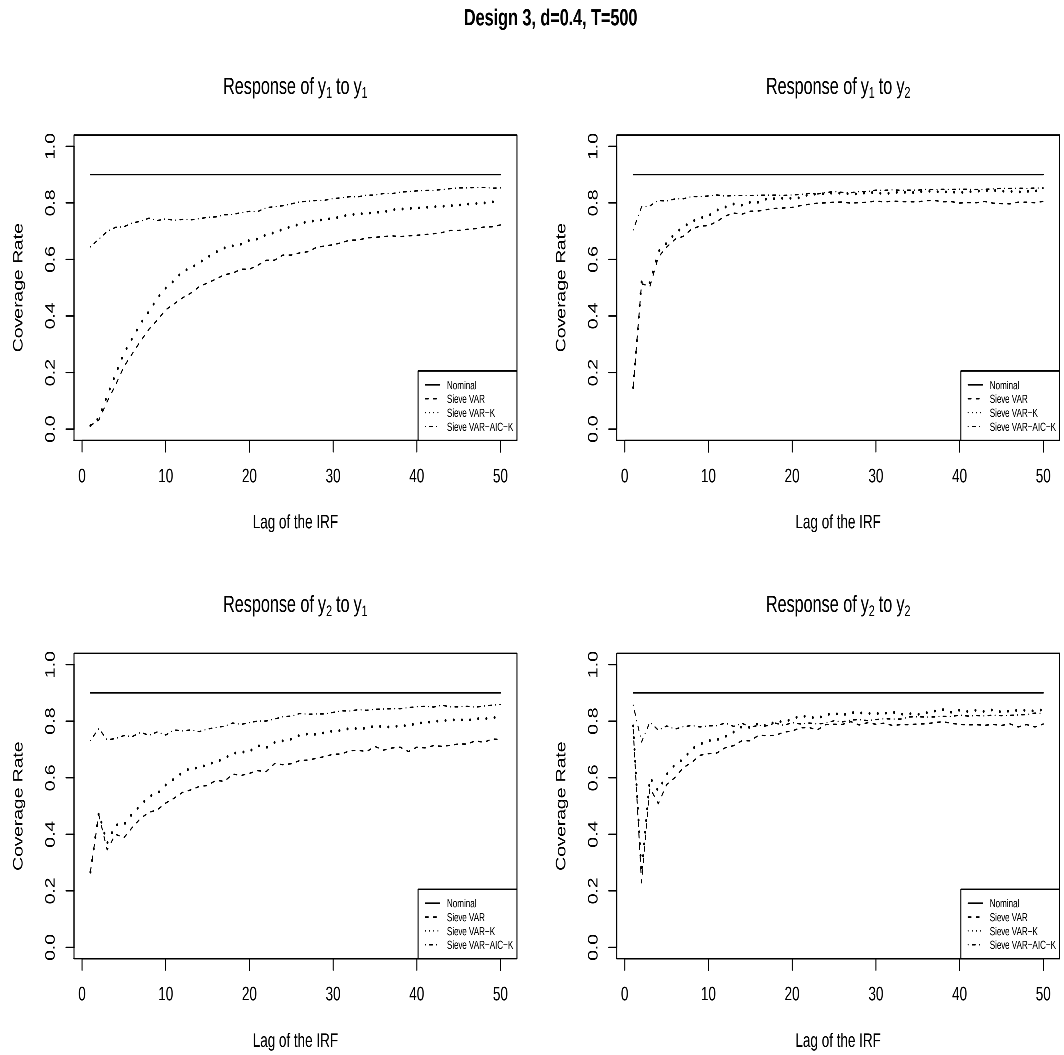

Figure 8: Design 3, $d = 0.4$, $T = 500$, Monte Carlo averages of IRWs and coverage rates.

Notes. Black solid line represents the nominal coverage rate. **Sieve VAR** denotes the sieve VAR bootstrap with $p = ln(T)^2$, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction, **Sieve VAR-AIC-K** denotes the sieve VAR bootstrap with p chosen using the Akaike's criterion and Kilian (1998a) correction.

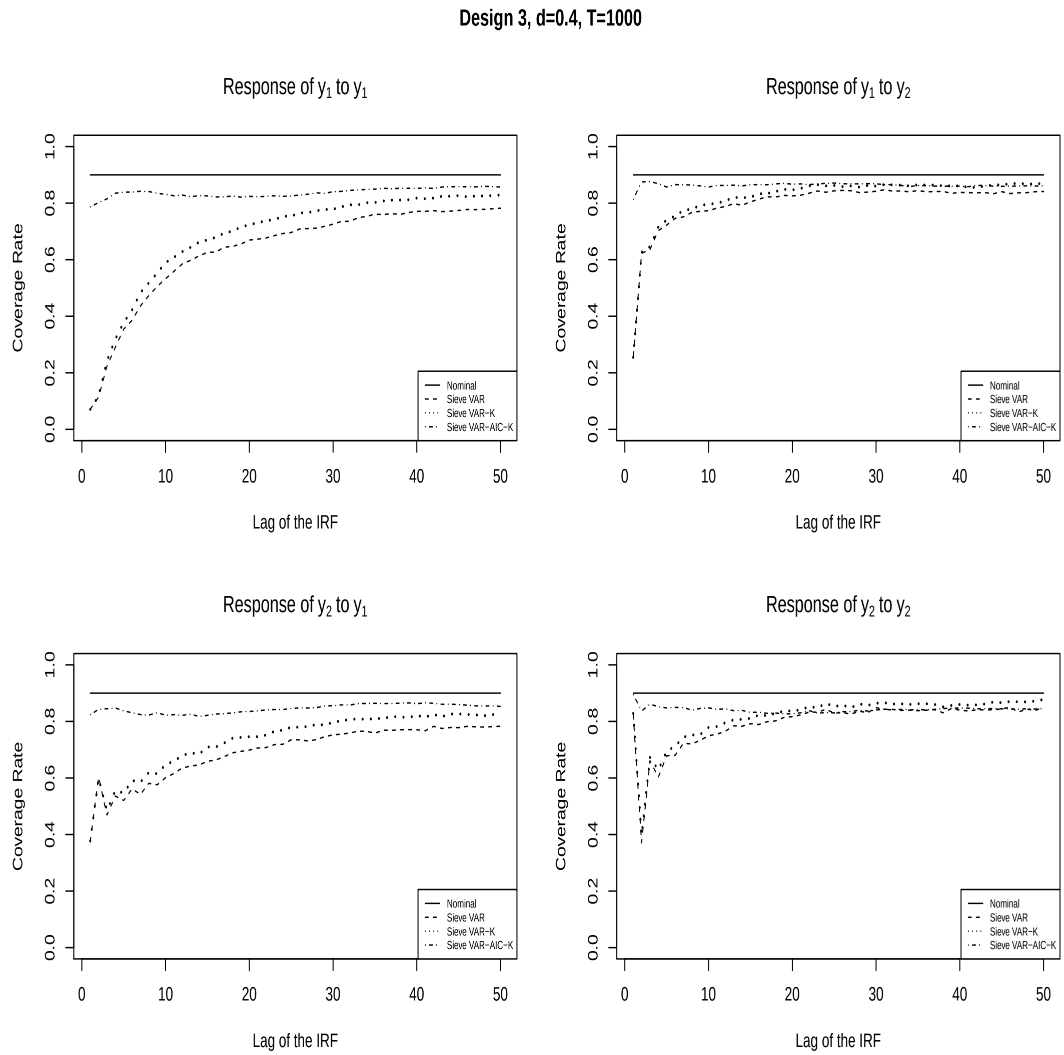

Figure 9: Design 3, $d = 0.4$, $T = 1000$, Monte Carlo averages of IRWs and coverage rates.

Notes. Black solid line represents the nominal coverage rate. **Sieve VAR** denotes the sieve VAR bootstrap with $p = ln(T)^2$, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction, **Sieve VAR-AIC-K** denotes the sieve VAR bootstrap with p chosen using the Akaike's criterion and Kilian (1998a) correction.

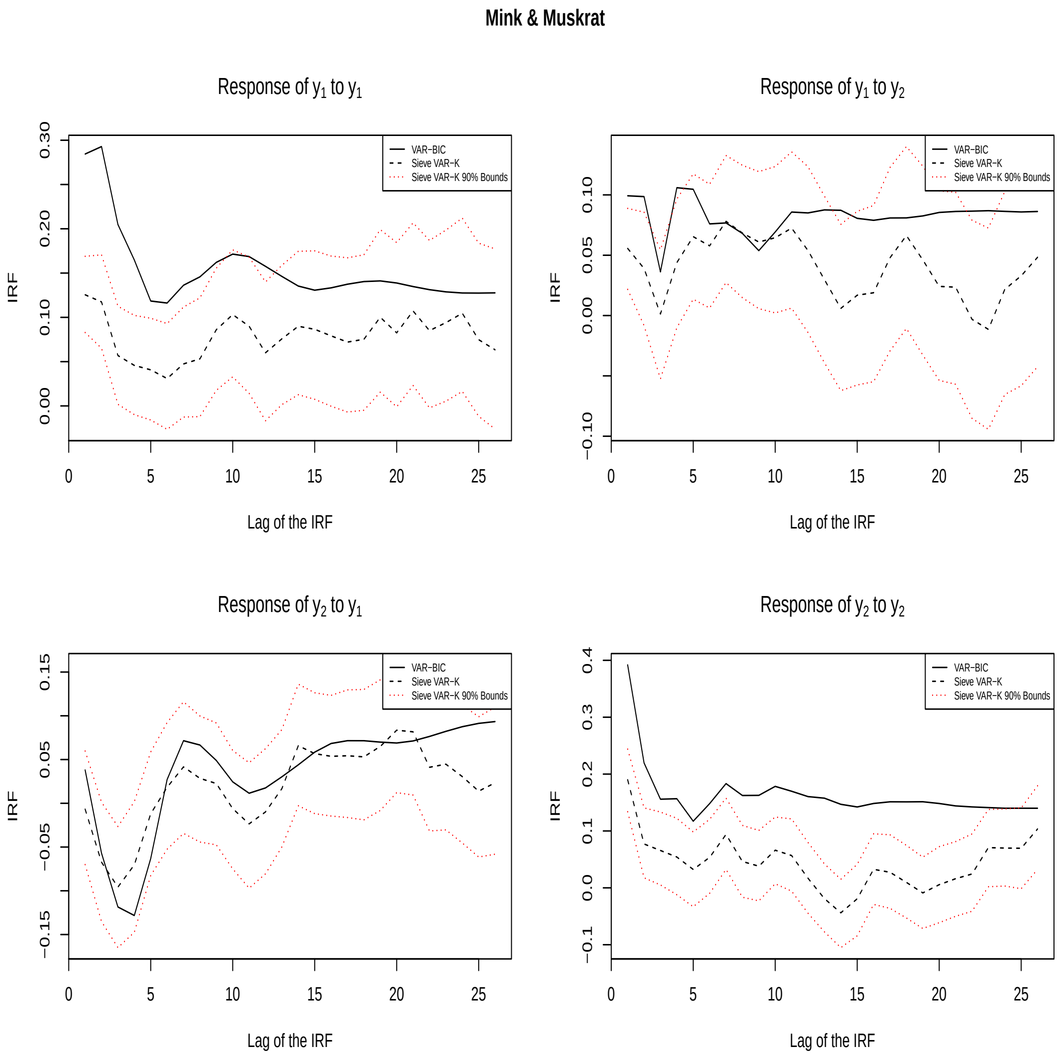

Figure 10: Mink & Muskrat Impulse Response Function.

Notes. **Sieve VAR-BIC** denotes the VAR(p) with p chosen using the Bayesian criterion, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction; the red dotted lines represent its bounds.

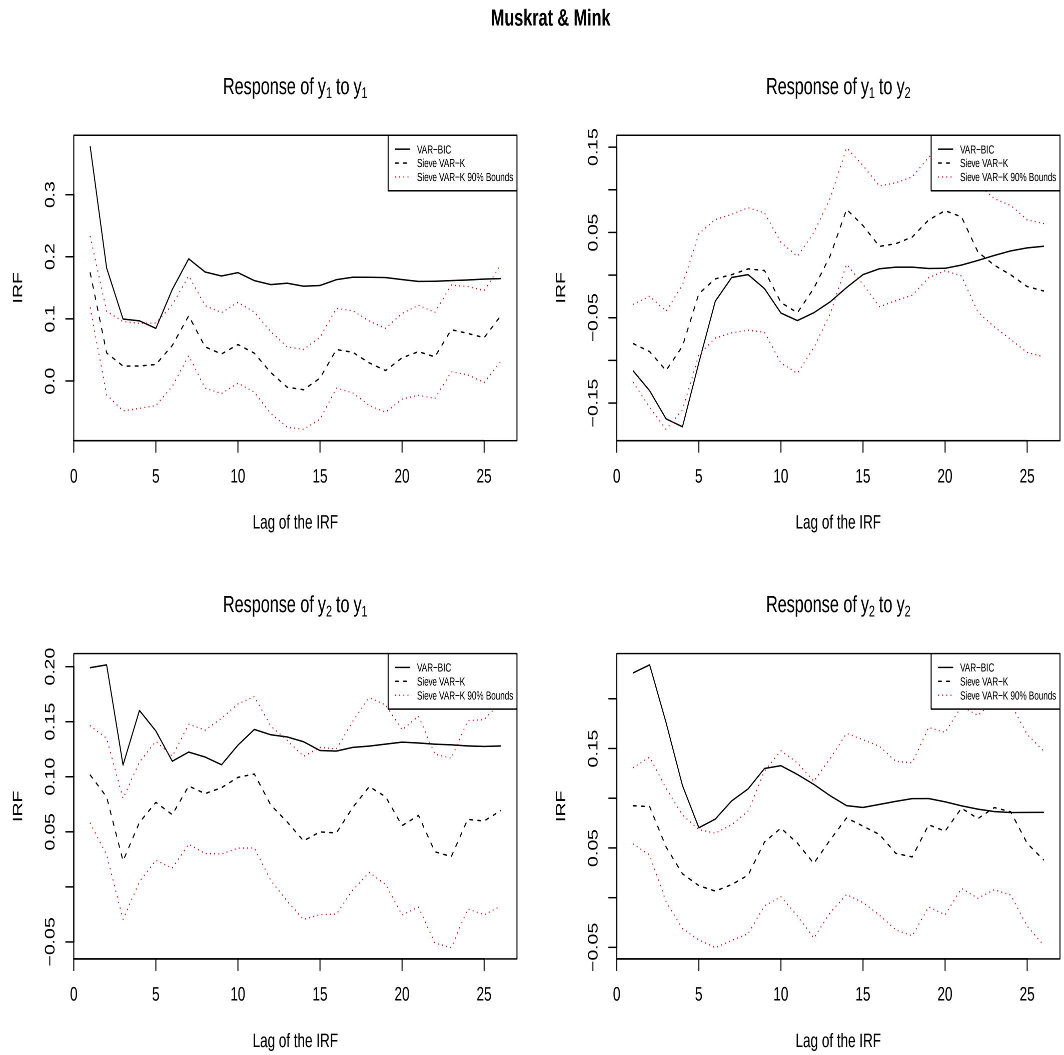

Figure 11: Muskrat & Mink Impulse Response Function.

Notes. **Sieve VAR-BIC** denotes the VAR(p) with p chosen using the Bayesian criterion, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction; the red dotted lines represent its bounds.

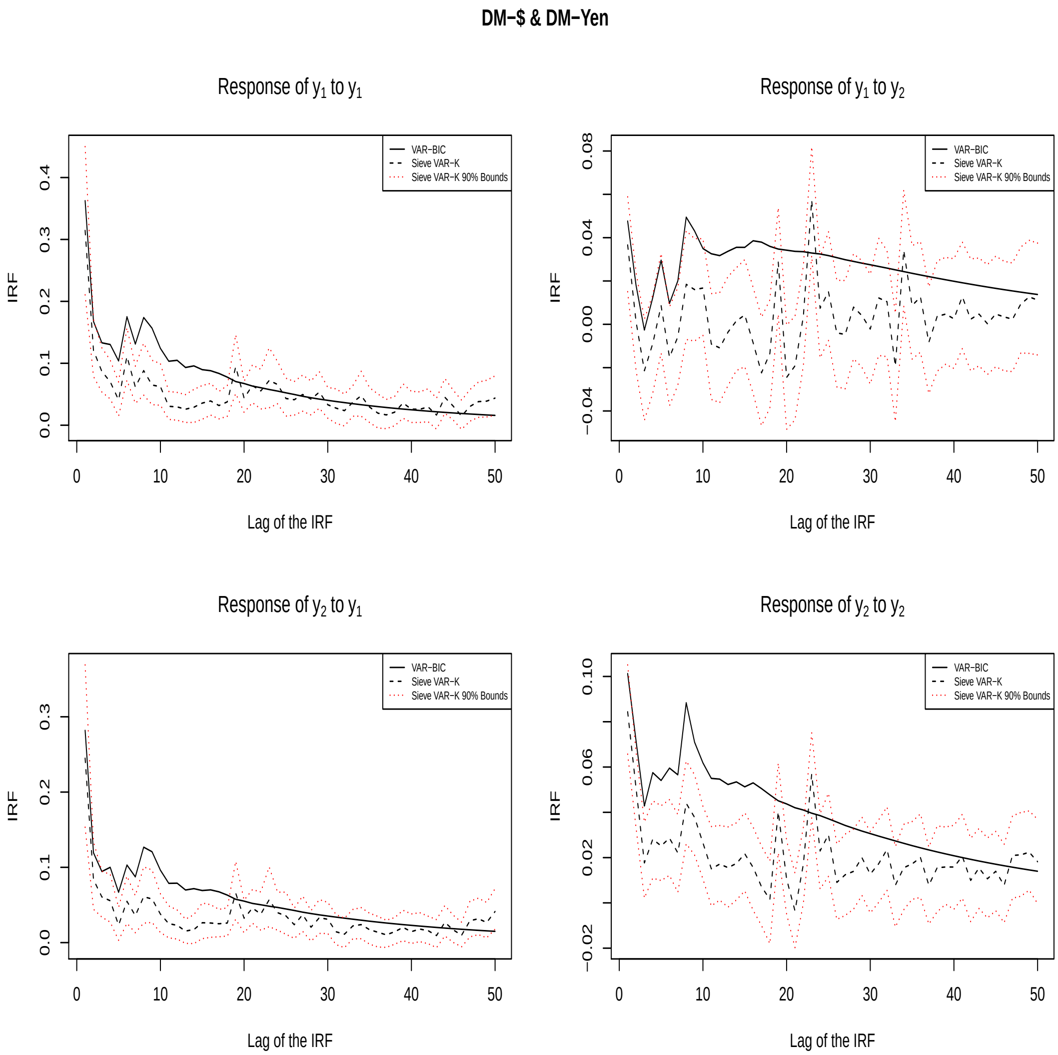

Figure 12: DM-$ & DM-Yen Impulse Response Function.

Notes. **Sieve VAR-BIC** denotes the VAR(p) with p chosen using the Bayesian criterion, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction; the red dotted lines represent its bounds.

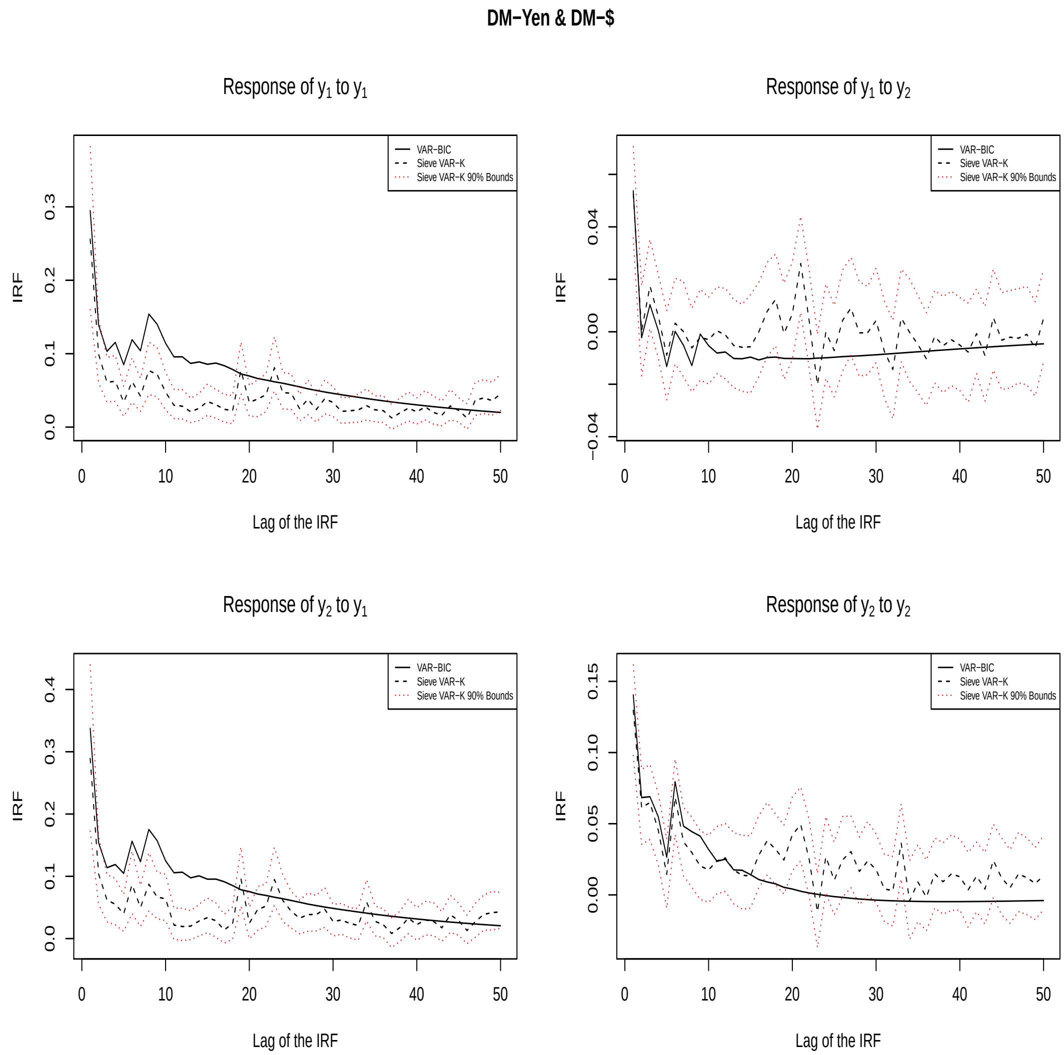

Figure 13: DM-Yen & DM-$ Impulse Response Function.

Notes. **Sieve VAR-BIC** denotes the VAR(p) with p chosen using the Bayesian criterion, **Sieve VAR-K** denotes the sieve VAR bootstrap with $p = ln(T)^2$ and Kilian (1998a) correction; the red dotted lines represent its bounds.