UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 11-34
Statistics and Econometrics Series 026
October 2011

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

BOOTSTRAP FORECAST OF MULTIVARIATE VAR MODELS WITHOUT USING THE BACKWARD REPRESENTATION

Lorenzo Pascual♦, Esther Ruiz♠ and DiegoFresoli♣

Abstract

In this paper, we show how to simplify the construction of bootstrap prediction densities in multivariate VAR models by avoiding the backward representation. Bootstrap prediction densities are attractive because they incorporate the parameter uncertainty and do not rely on any particular assumption about the error distribution. What is more, the construction of densities for more than one-step-ahead is possible even in situations when these densities are unknown asymptotically. The main advantage of the new procedure is that it is computationally simple without loosing the good performance of bootstrap procedures. Furthermore, by avoiding a backward representation, its asymptotic validity can be proved without relying on the assumption of Gaussian errors as needed by alternative procedures. Finally, the new procedure proposed in this paper can be implemented to obtain prediction densities in models without a backward representation as, for example, models with MA components or GARCH disturbances. By comparing the finite sample performance of the proposed procedure with those of alternatives, we show that nothing is lost when using it. Finally, we implement the procedure to obtain prediction regions for US quarterly future inflation, unemployment and GDP growth.

Keywords: Non-Gaussian VAR models, Prediction cubes, Prediction density, Prediction regions, Prediction ellipsoids, Resampling methods.

♦ EDP-Energías de Portugal, S.A., Unidade de Negócio de Gestao da Energía, Director Adjunto.
♠ Corresponding author: Dpt. Estadística and Instituto Flores de Lemus, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe, Spain. Tel: 34 916249851, Fax: 34 916249849, e-mail: ortega@est-econ.uc3m.es.
♣Dpt. de Estadística, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), Spain, e-mail: dfresoli@est-econ.uc3m.es.
Acknowledgements. The last two authors are grateful for financial support from project ECO2009-08100 of the Spanish Government. We are also grateful to Gloria González-Rivera for her useful comments. The usual disclaims apply.

Bootstrap Forecast of Multivariate VAR Models without using the backward representation

Lorenzo Pascual* Esther Ruiz† Diego Fresoli‡§

October 2011

Abstract

In this paper, we show how to simplify the construction of bootstrap prediction densities in multivariate VAR models by avoiding the backward representation. Bootstrap prediction densities are attractive because they incorporate the parameter uncertainty and do not rely on any particular assumption about the error distribution. What is more, the construction of densities for more than one-step-ahead is possible even in situations when these densities are unknown asymptotically. The main advantage of the new procedure is that it is computationally simple without loosing the good performance of bootstrap procedures. Furthermore, by avoiding a backward representation, its asymptotic validity can be proved without relying on the assumption of Gaussian errors as needed by alternative procedures. Finally, the new procedure proposed in this paper can be implemented to obtain prediction densities in models without a backward representation as, for example, models with MA components or GARCH disturbances. By comparing the finite sample performance of the proposed procedure with those of alternatives, we show that nothing is lost when using it. Finally, we implement the procedure to obtain prediction regions for US quarterly future inflation, unemployment and GDP growth.

KEYWORDS: Non-Gaussian VAR models, Prediction cubes, Prediction density, Prediction regions, Prediction ellipsoids, Resampling methods.

*EDP-Energías de Portugal, S.A., Unidade de Negócio de Gestão da Energía, Director Adjunto.

†Corresponding author: Dpt. Estadística and Instituto Flores de Lemus, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe, Spain. Tel: 34 916249851, Fax: 34 916249849, e-mail: ortega@est-econ.uc3m.es.

‡Dpt. de Estadística, Universidad Carlos III de Madrid, C./ Madrid 126, 28903 Getafe (Madrid), Spain, e-mail dfresoli@est-econ.uc3m.es.

§**Acknowledgements.** The last two authors are grateful for financial support from project ECO2009-08100 of the Spanish Government. We are also grateful to Gloria González-Rivera for her useful comments. The usual disclaims apply.

1 Introduction

Bootstrap procedures are known to be useful when forecasting time series because they allow the construction of prediction densities without imposing particular assumptions on the error distribution and simultaneously incorporating the parameter uncertainty. Note that when the errors are non-Gaussian the prediction densities are usually unknown when the prediction horizon is larger than one-step-ahead. However, the bootstrap can be implemented in these cases to obtain the corresponding prediction densities. They are also attractive because their computational simplicity and wide applicability. However, these advantages are limited by the use of the backward representation that many authors advocate after the seminal paper of Thombs and Schuchany (1990). In particular, Kim (1999) extends the procedure of Thombs and Schuchany (1990) to stationary VAR(p) models. Later, Kim (2001, 2004) considered bias-corrected prediction regions by employing a bootstrap-after-bootstrap approach. On the other hand, Grigoletto (2005) proposes two further alternative procedures based on Kim (1999) that take into account not only the uncertainty due to parameter estimation but also the uncertainty attributable to model specification. In any case, the bootstrap procedures conceived by Kim (1999, 2001, 2004) and Grigoletto (2005) use the backward representation to generate the bootstrap samples used to obtain replicates of the estimated parameters. Using the backward representation has three main drawbacks. First, the resulting procedure is computationally complicate and time consuming. Second, given that the backward residuals are not independent it is necessary to use the relationship between the backward and forward representation of the model in order to resample from independent residuals; see Kim (1997, 1998) for this relationship which can be rather complicate for high order models. Consequently, Kim (1999) resamples directly from the dependent backward residuals. However, the asymptotic validity of the bootstrap resampling can only be proved by imposing i.i.d. and, as a result, it requires assuming Gaussian errors. Finally, these bootstrap alternatives can only be applied to models with a backward representation which excludes their implementation in, for example, multivariate models with Moving Average (MA) components or with GARCH disturbances.

In an univariate framework, Pascual *et al.* (2004a) show that the backward representation can be avoided without loosing the good properties of the bootstrap prediction densities. When dealing with multivariate systems it is even more important avoiding the backward representation due to its larger complexity; see Kim (1997, 1998). In this paper, we propose an extension of the bootstrap procedure proposed by Pascual *et al.* (2004a) for univariate ARIMA models to obtain

joint prediction densities for multivariate VAR(p) models avoiding the backward representation and, consequently, overcoming its limitations. We prove the asymptotic validity of the proposed procedure without relying on particular assumptions about the prediction error distribution. We focus on the construction of multivariate prediction densities from which it is possible to obtain marginal prediction intervals for each of the variables in the system and joint prediction regions for two or more variables within the system.[1] Monte Carlo experiments are carried out to study the finite sample performance of the marginal prediction intervals obtained by the new bootstrap procedure and compare it those of alternative procedures available in the literature. We also compare their corresponding elliptical and Bonferroni regions. We show that, although the bootstrap procedure proposed in this paper is computationally simpler, its finite sample properties are similar to those of previous more complicate bootstrap approaches and clearly better than those of the standard and asymptotic prediction densities. We also show that when the errors are non-Gaussian, the bootstrap elliptical regions are inappropriate with the Bonferroni regions having better properties. The procedures are illustrated with an empirical application which consists of predicting future inflation, unemployment and growth rates in the US.

The rest of the paper is organized as follows. Section 2 describes the asymptotic and bootstrap prediction intervals and regions previously available in the literature. In Section 3, we propose a new bootstrap procedure, derive its asymptotic distribution and analyze its performance in finite samples. We compare the new bootstrap densities and corresponding prediction intervals and regions with the standard, asymptotic and alternative bootstrap procedures. Section 4 illustrates the results with an empirical application. Finally, section 5 concludes the paper with suggestions for further research.

2 Asymptotic and bootstrap prediction intervals and regions for VAR models

In this section, we describe the construction of prediction regions in stationary VAR models based on assuming known parameters and Gaussian errors. We also describe how the parameter uncertainty can be incorporated by using asymptotic and bootstrap approximations of the finite sample distribution of the parameter estimator. The bootstrap procedures can also be implemented to deal with non-Gaussian errors.

[1] Previous paper focus on the Bonferroni regions and not in the bootstrap densities themselves.

2.1 Asymptotic prediction intervals and regions

Consider the following multivariate VAR(p) model

$$\Phi(L)Y_t = \mu + a_t, \tag{1}$$

where Y_t is the $Nx1$ vector of observations at time t, μ is a $Nx1$ vector of constants, $\Phi(L) = I_N - \Phi_1 L - ... - \Phi_p L^p$ with L being the lag operator and I_N the NxN identity matrix. The NxN parameter matrices, Φ_i, $i = 1, ..., p$, satisfy the stationarity restriction. Finally, a_t is a sequence of $Nx1$ independent white noise vectors with nonsingular contemporaneous covariance matrix given by Σ_a.

It is well known that if a_t is an independent vector white noise sequence then the point predictor of Y_{T+h} that minimizes the Mean Square Error (MSE) is its conditional mean which depends on the model parameters. In practice, these parameters are unknown and the predictor of Y_{T+h} is obtained with the parameters substituted by consistent estimates as follows

$$\widehat{Y}_{T+h|T} = \widehat{\mu} + \widehat{\Phi}_1 \widehat{Y}_{T+h-1|T} + ... + \widehat{\Phi}_p \widehat{Y}_{T+h-p|T} \tag{2}$$

where $\widehat{Y}_{T+j|T} = Y_{T+j}$, $j = 0, -1, ...$ Furthermore, the MSE of $\widehat{Y}_{T+h|T}$ is usually estimated as follows

$$\widehat{\Sigma}_{\widehat{Y}}(h) = \sum_{j=0}^{h-1} \widehat{\Psi}_j \widehat{\Sigma}_a \widehat{\Psi}_j', \tag{3}$$

where $\widehat{\Psi}_j$ are the estimated matrices of the MA representation of Y_t and $\widehat{\Sigma}_a = \frac{\widehat{a}\widehat{a}'}{T-Np-1}$ where $\widehat{a} = (\widehat{a}_1, ..., \widehat{a}_T)$ with

$$\widehat{a}_t = Y_t - \widehat{\mu} - \widehat{\Phi}_1 Y_{t-1} - ... - \widehat{\Phi}_p Y_{t-p}. \tag{4}$$

If a_t is further assumed to be Gaussian, then the marginal prediction density of the nth variable in the system is also Gaussian and the standard practice is to construct the (1-α)100% prediction interval for the nth variable in the system as follows

$$GI_{T+h} = \left\{ y_{n,T+h} | y_{n,T+h|T} \in \left[\widehat{y}_{n,T+h|T} \pm z_{\alpha/2} \widehat{\sigma}_{n,h} \right] \right\}, \tag{5}$$

where $\widehat{y}_{n,T+h|T}$ is the nth component of $\widehat{Y}_{T+h|T}$, $\widehat{\sigma}_{n,h}$ is the square root of the nth diagonal element of $\widehat{\Sigma}_{\widehat{Y}}(h)$ and z_α is the α-quantile of the standard Gaussian distribution. The Gaussianity of the forecast errors can also be used to obtain the following (1-α)100% joint ellipsoid for all the

variables in the system

$$GE_{T+h} = \left\{ Y_{T+h} | \left[Y_{T+h} - \widehat{Y}_{T+h|T} \right]' \widehat{\Sigma}_{\widehat{Y}}(h)^{-1} \left[Y_{T+h} - \widehat{Y}_{T+h|T} \right] < \chi^2_\alpha(N) \right\}, \tag{6}$$

where $\chi^2_\alpha(N)$ is the α-quantile of the χ^2 distribution with N degrees of freedom.[2] Constructing the ellipsoids in (6) can be quite demanding when N is larger than two or three. Consequently, Lütkepolh (1991) proposes using the Bonferroni method to construct the following prediction cubes with coverage at least (1-α)100%

$$GC_{T+h} = \left\{ Y_{T+h} | Y_{T+h} \in \cup_{n=1}^{N} \left[\widehat{y}_{n,T+h|T} \pm z_\tau \widehat{\sigma}_{n,h} \right] \right\}, \tag{7}$$

where where $\tau = 0.5(\alpha/N)$. However, note that the prediction intervals and regions above have two main drawbacks. First, they are constructed using the MSE in (3) which does not incorporate the parameter uncertainty. As a consequence, if the sample size is small the uncertainty associated with $\widehat{Y}_{T+h|T}$ is underestimated and the corresponding intervals an regions will have lower coverages than the nominal. The second problem, is related to the Gaussianity assumption. When this assumption does not hold, the quadratic form in (6) is not adequate as well as the width of the intervals in (5) and (7). Even more, when the prediction errors are not Gaussian, the shape of their densities for h >2 is in general unknown.

As an illustration, consider the following VAR(2) bivariate model

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} -0.9 & 0 \\ 0.4 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} -0.5 & -0.7 \\ 0.8 & -0.1 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \tag{8}$$

where $a_t = (a_{1,t},\ a_{2,t})'$ is an independent white noise vector with contemporaneous covariance matrix given by $vec\Sigma_a = (1, 0.8, 0.8, 1)'$ where vec denotes the column stacking operator. The distribution of a_t is a $\chi^2(4)$. Panel (a) of Figures 1 and 2 display the joint one-step-ahead and eight-steps-ahead densities of $y_{1,t}$ and $y_{2,t}$ respectively, which have clear asymmetries, although more pronounced in the former. After generating a time series of size $T = 100$, the VAR(2) parameters are estimated by Least Squares (LS). Panel (b) of Figures 1 and 2 plot kernel estimates of the joint density obtained as usual after assuming that the prediction errors are jointly Gaussian

[2]The same argument can be applied when the interest lies on only a subset of components of Y_t. For example, if the focus is on the first J components of Y_t, the prediction ellipsoid is given by $\{Y_{T+h} | \left[Y_{T+h} - \widehat{Y}_{T+h|T} \right]' \left[C' \left(C \widehat{\Sigma}_{\widehat{Y}}(h) C' \right)^{-1} C \right] \left[Y_{T+h} - \widehat{Y}_{T+h|T} \right] < \chi^2_\alpha(J)\}$, where $C = [I_J\ 0]$.

with zero mean and convariance matrix given by (3).[3] Comparing panels (a) and (b), it is obvious that the Gaussian approach fails to capture the asymmetry of the error distribution. It is usual in practice to construct prediction regions for $y_{1,t}$ and $y_{2,t}$. From the joint Gaussian density plotted in panel (b) of Figure 1, it is possible to obtain the corresponding 95% one-step-ahead ellipsoids and Bonferroni regions. They are shown in Figure 3 together with a realization of Y_{T+1}. We can observe that the shape of both regions is not appropriate to construct a satisfactory prediction region for Y_{T+1}. Finally, as we may also being interested in forecasting only one variable in the system, Figure 4 displays the true marginal one-step-ahead density of $y_{1,t}$ together with the Gaussian approximation. Once more, it is clear that the Gaussian approach fails to capture the skewness of the prediction density.

Consider first the problem of incorporating the parameter uncertainty. As pointed out above, the MSE in (3) underestimates the true prediction uncertainty and, consequently, they can be inappropriate in small samples sizes. Granted that a good estimator is used, the importance of taking into account the parameter uncertainty could be small in systems consisting in few variables; see Riise and Tjostheim (1984). But, in empirical applications we often found VAR(p) models fitted to large systems; see, for example, Simkins (1995) for a VAR(6) model for a system of 5 macroeconomic variables, Waggoner and Zha (1999) who fit a VAR(13) model to a system of 6 macroeconomic variables, Chow and Choy (2006) who fit a VAR(5) model to a system of 5 variables related with the global electronic system, Gómez and Guerrero (2006) for a VAR(3) model fitted to a system of 6 macroeconomic variables and Chevillon (2009) for a VAR(2) model for a system of 4 macroeconomic variables just to mention a few empirical applications. Additionally, as these examples show, when dealing with real systems of time series, their adequate representation often requires a rather large order p. If the number of variables in the system and/or the number of lags of the model are relatively large, the estimation precision in finite samples could be rather low and predictions based on VAR(p) models with estimated parameters may suffer severely from the uncertainty in the parameter estimation. In these cases, it is important to construct prediction intervals and regions that take into account this uncertainty; see, for instance, Schmidt (1977), Lütkepolh (1991), West (1996), West and McCracken (1998), Sims and Zha (1998, 1999) for existing evidence on the importance of taking into account parameter uncertainty in unconditional forecasts and Waggoner and Zha (1999) for the same result in conditional forecasts.

[3]The smoothed density is obtained by applying a Gaussian kernel density estimator with a diagonal bandwidth matrix with elements given by the Gaussian "rule of thumb".

To incorporate the parameter uncertainty, Lütkepolh (1991) suggests approximating the sample distribution of the estimator by its asymptotic distribution density.[4] In this case, the MSE of $\widehat{Y}_{T+h|T}$ that incorporates the parameter uncertainty can be approximated by

$$\widehat{\Sigma}^{l}_{\widehat{Y}}(h) = \widehat{\Sigma}_{\widehat{Y}}(h) + \frac{1}{T}\widehat{\Omega}(h), \tag{9}$$

where

$$\widehat{\Omega}(h) = \sum_{i=0}^{h-1}\sum_{j=0}^{h-1} tr\left[(\widehat{\mathbf{B}}')^{h-1-i}\widehat{\Upsilon}^{-1}\widehat{\mathbf{B}}^{h-1-j}\widehat{\Upsilon}\right]\widehat{\Psi}_i\widehat{\Sigma}_a\widehat{\Psi}'_j, \tag{10}$$

with $\widehat{\Upsilon} = \frac{ZZ'}{T}$ and $\widehat{\mathbf{B}}$ is the following $(Np+1)x(Np+1)$ matrix

$$\widehat{\mathbf{B}} = \begin{bmatrix} 1 & 0 & 0 & ... & 0 & 0 \\ \widehat{\mu} & \widehat{\Phi}_1 & \widehat{\Phi}_2 & ... & \widehat{\Phi}_{p-1} & \widehat{\Phi}_p \\ 0 & I_N & 0 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & I_N & 0 \end{bmatrix}.$$

In order to asses the effect of the parameter uncertainty on the MSE given by (9), consider the case of one-step-ahead predictions, i.e. when $h = 1$. In this situation, $\widehat{\Omega}(1) = (Np+1)\widehat{\Sigma}_a$, and $\widehat{\Sigma}^{l}_{\widehat{Y}}(1)$ can be approximated by

$$\widehat{\Sigma}^{l}_{\widehat{Y}}(1) = \frac{T+Np+1}{T}\widehat{\Sigma}_a.$$

This expression shows that the contribution of the parameter uncertainty to the one-step-ahead MSE matrix depends on the dimension of the system, N, the VAR order, p, and the sample size, T. As long as N and/or p, or both, are big enough compared to the sample size T, the effect of parameter uncertainty can be substantial. Obviously, as the sample size gets larger then $\lim_{T\to\infty}\frac{T+Np+1}{T} = 1$ and the parameter uncertainty contribution to the MSE in (9) vanishes.

Once the MSE is computed as in (9), the corresponding prediction intervals, ellipsoids and cubes are constructed using the Gaussianity assumption as follows

$$AI_{T+h} = \left\{y_{n,T+h} | y_{n,T+h} \in \left[\widehat{y}_{n,T+h|T} \pm z_{\alpha/2}\widehat{\sigma}^{l}_{n,h}\right]\right\}, \tag{11}$$

[4] Alternatively, some authors propose using Bayesian methods which could be rather complicated from a computational point of view; see, for example, Simkins (1995) and Waggoner and Zha (1999) who need the Gaussianity assumption to derive the likelihood and posterior distribution.

$$AE_{T+h} = \left\{ Y_{T+h} | \left[Y_{T+h} - \widehat{Y}_{T+h|T} \right] \widehat{\Sigma}^{l}_{\widehat{Y}}(h)^{-1} \left[Y_{T+h} - \widehat{Y}_{T+h|T} \right] < \chi^2_{\alpha}(N) \right\}, \tag{12}$$

$$AC_{T+h} = \left\{ Y_{T+h} | Y_{T+h} \in \cup_{n=1}^{N} \left[\widehat{y}_{n,T+h|T} \pm z_{\tau} \widehat{\sigma}^{l}_{n,h} \right] \right\}, \tag{13}$$

where $\widehat{\sigma}^{l}_{n,h}$ is the square root of the nth diagonal element of $\widehat{\Sigma}^{l}_{\widehat{Y}}(h)$.

Panel (c) of Figure 1 which plots the density of $y_{1,T+1}$ and $y_{2,T+1}$ for the same example as above constructed assuming that the forecast error are Gaussian with zero mean and covariance matrix given by (9), shows that this density is not very different from that plotted in panel (b) and, obviously, it is not able to capture the asymmetries of the error distribution. Similarly, the joint density of $y_{1,T+8}$ and $y_{2,T+8}$ in panel (c) of Figure 2 does not look different from that in panel (b). The similarity between the standard and the asymptotic densities is even more clear in Figure 3 that plots the elliptical and Bonferroni regions constructed using (12) and (13), respectively. As we can observe they are slightly larger than the standard but still located very close to them. They cannot cope with the lack of symmetry of the prediction error distribution. This similarity could be expected as we are estimating 8 parameters with $T = 100$. Similar comments deserve Figure 4, where we can observe that the asymptotic marginal density for the first component of the system $y_{1,T+1}$ only differs from the standard density in the variability, which is slightly larger in the former.

Note that the asymptotic approximation of the distribution of the LS estimator can be inadequate in small samples depending on the number of parameters to be estimated and the true distribution of the innovations.

2.2 Bootstrap procedures for prediction intervals and regions

To overcome the limitations of the Gaussian densities described before, Kim (1999, 2001, 2004) and Grigoletto (2005) propose using bootstrap procedures which incorporate the parameter uncertainty even when the sample size is small and do not rely on the Gaussianity assumption.

In order to take into account the conditionality of VAR forecasts on past observations, Kim (1999) proposes to obtain bootstrap replicates of the series based on the following backward recursion

$$Y^*_t = \widehat{\omega} + \widehat{\Lambda}_1 Y^*_{t+1} + ... + \widehat{\Lambda}_p Y^*_{t+p} + \widehat{v}^*_t \tag{14}$$

where $Y^*_{T-i} = Y_{T-i}$ for $i = 0, 1, ..., p-1$ are p starting values which coincide with the last values of the original series, $\widehat{\omega}$, $\widehat{\Lambda}_1, ..., \widehat{\Lambda}_p$, are LS estimates of the parameters of the backward representation, and $\widehat{v}^*_t$ are obtained by resampling from the empirical distribution function of the

centered and rescaled backward residuals. Then, bootstrap LS estimates of the parameters of the forward representation are obtained by estimating the VAR(p) model in (1) using $\{Y_1^*, ..., Y_T^*\}$. Denote these estimates by $\hat{B}^* = (\widehat{\mu}^*, \widehat{\Phi}_1^*, ..., \widehat{\Phi}_p^*)$. The bootstrap forecast for period $T+h$ is then given by

$$\widehat{Y}_{T+h|T}^* = \widehat{\mu}^* + \widehat{\Phi}_1^* \widehat{Y}_{T+h-1|T}^* + ... + \widehat{\Phi}_p^* \widehat{Y}_{T+h-p|T}^* + \widehat{a}_{T+h}^* \tag{15}$$

where $\widehat{a}_{T+h}^*$ are random draws from the empirical distribution function of centered and rescaled forward residuals. Having obtained R bootstrap replicates of $\widehat{Y}_{T+h|T}^*$, Kim (2001) defines the bootstrap (1-α)100% prediction interval for the nth variable in the system as follows

$$KI_{T+h} = \left\{ y_{n,T+h} | y_{n,T+h} \in \left[q_K^* \left(\frac{\alpha}{2} \right), q_K^* \left(1 - \frac{\alpha}{2} \right) \right] \right\} \tag{16}$$

where $q_K^*(\gamma)$ is the empirical γ-quantile of the bootstrap distribution of the nth component of $\widehat{Y}_{T+h|T}^*$ approximated by $G_{n,K}^*(x) = \#(\widehat{y}_{n,T+h|T}^* < x)/R$. Similarly, Kim (1999) proposes to construct bootstrap prediction ellipsoids with probability conte[illegible]$1-\alpha)100\%$ are given by

$$KE_{T+h} = \left\{ Y_{T+h} | \left[Y_{T+h} - \overline{\widehat{Y}_{T+h|T}^*} \right]' S_{\hat{Y}^*}^K(h)^{-1} \left[Y_{T+h} - \overline{\widehat{Y}_{T+h|T}^*} \right] < Q_K^* \right\} \tag{17}$$

where $\overline{\widehat{Y}_{T+h|T}^*}$ is the sample mean of the R bootstrap replicates $\widehat{Y}_{T+h|T}^{*(r)}$ and $S_{\hat{Y}^*}^K(h)$ is the corresponding sample covariance.[5] The quantity Q_K^* in (17) is the $(1-\alpha)100\%$ percentile of the bootstrap distribution of the following quadratic form

$$\left[\widehat{Y}_{T+h|T}^* - \overline{\widehat{Y}_{T+h|T}^*} \right]' S_{\hat{Y}^*}^K(h)^{-1} \left[\widehat{Y}_{T+h|T}^* - \overline{\widehat{Y}_{T+h|T}^*} \right]. \tag{18}$$

Furthermore, Kim (1999) proposes using the Bonferroni approximation to obtain prediction cubes with nominal coverage of at least (1-α)100% which are given by

$$KC_{T+h} = \left\{ y_{n,T+h} | y_{n,T+h} \in \cup_{n=1}^N \left[q_K^* \left(\frac{\tau}{2} \right), q_K^* \left(1 - \frac{\tau}{2} \right) \right] \right\} \tag{19}$$

where $\tau = \alpha/N$.[6]

[5] Kim (1999) does not explicitly show how $S_{\hat{Y}^*}^K(h)$ should be defined. Alternatively, one can obtain $S_{\hat{Y}^*}(h)$ by substituting the parameters in (9) by their bootstrap estimates and computing the average through all bootstrap replicates. By calculating it with the sample covariance or by substituting the bootstrap parameters in the corresponding expressions we get similar results.

[6] Actually, what Kim (1999) defines as KC is slightly different from (19) as he uses the percentile and percentile-t methods of Hall (1992). Here we prefer to use the Bonferroni prediction regions in (19) because they are better suited to deal with potential asymmetries of the error distribution; see Hall (1992).

The bootstrap procedure just described is illustrated by considering again the same time series of size $T = 100$ simulated by model (8). Panel (d) of Figures 1 and 2 plot the joint bootstrap density of $y_{1,T+1}$ and $y_{2,T+1}$ and $y_{1,T+8}$ and $y_{2,T+8}$ respectively, based on $R = 2999$ bootstrap replicates. When comparing these densities with their Gaussian counterparts, it is clear that the bootstrap can reproduce the asymmetry and is much closer to the true density plotted in panel (a) of the same figures. Figure 3 plots the corresponding ellipsoids and cubes defined in (17) and (19). First of all, note that the bootstrap ellipsoid is only slightly larger than the two Gaussian ellipsoids and it is not adequate to represent the shape of the realization of $y_{1,T+1}$ and $y_{2,T+1}$ plotted in Figure 2. This is due to the fact the ellipsoid in (17) is still based on a Gaussian assumption and only differs from the ellipsoids described in the previous subsection in the way the MSE is computed. However, Figure 3 clearly illustrates that when dealing with non-Gaussian prediction errors, the prediction regions constructed from the bootstrap joint densities cannot be based on ellipsoids. An alternative is to use High Density Regions (HDR) proposed by Hyndman (1996) which have also been plotted in Figure 3. These regions are based on kernel estimates of the joint densities as those plotted in Figure 1. Although the shape of these regions seem to be more adequate, they are unfeasible when the dimension of the system is large as, in this case, there are not satisfactory kernel estimators of the bootstrap densities. On the other hand, when the prediction regions are constructed using the Bonferroni approximation, Figure 3 shows that the bootstrap cube is located towards the northeast so it is more adequate than the ellipsoid. This is in fact reflecting that the quantiles of the marginal densities used in (19) can cope with the asymmetries while the ellipsoids use a quadratic form based on the wrong Gaussianity assumption. Finally, Figure 4 displays the marginal one-step-ahead kernel density of $y_{1,T+1}$. We observe that it is clearly closer to the true density than its Gaussian counterparts.

Kim (1999) justifies the use of his bootstrap procedure in small samples by suggesting that the asymptotic results of Thombs and Schuchany (1990) can be extended to a multivariate framework. However, the asymptotic validity of the bootstrap procedures based on the backward representation relies on the assumption of Gaussian innovations; see Kim (2001). This is due to the fact that the serial independence of the backward errors is needed and this can only be guarantee under the assumption of Gaussian disturbances. Note that alternatively one could use the relationship between the forward and backward residuals and resampling from the former to obtain the latter. However, obtaining the backward representation can be very complicated in VAR(p) models with

large order; see Kim (1997, 1998) for the expression of the backward representation.[7]

3 A new bootstrap procedure

In this section, we propose an extension of the bootstrap procedure proposed by Pascual *et al.* (2004a) for univariate ARIMA models, to obtain the joint prediction density of Y_{T+h} in VAR(p) models. The new bootstrap procedure avoids the backward representation by resampling without fixing the observations of the available sample when incorporating the parameter uncertainty. It is important to note that although the predictions are conditional on the available time series, the sample distribution of the parameter estimator is defined as the distribution through different replicates; see Harvey (1989) and Lütkepohl (1991) who argue that the distribution of the predictions based on estimated parameters is obtained as if the sample used for prediction is different from the sample used for estimation. Therefore, using the backward representation is not necessary theoretically and only adds complexity into the bootstrap procedure without any advantage; see Pascual *et al.* (2004a) for the same arguments in univariate ARIMA models and Rodríguez and Ruiz (2009) for univariate state space models. By avoiding the backward representation, the new procedure is simpler from a computational point of view, can be implemented in models without such representation and its asymptotic validity can be established without assuming Gaussian errors. In this section, we describe the proposed procedure and prove its asymptotic validity to estimate the joint density of future values of Y_{T+h}. We also carry out Monte Carlo experiments to analyse the performance of prediction intervals constructed from the corresponding marginal bootstrap densities and ellipsoids and cubes constructed from the joint bootstrap density.

3.1 Description of the bootstrap procedure

The new procedure proposed in this paper to obtain the bootstrap prediction density of Y_{T+h} is similar to that proposed by Kim (1999) but avoiding the backward representation in (14). The algorithm to obtain the bootstrap replicates of Y_{T+h} is the following.

Step 1. Estimate by LS the parameters of model (1) and obtain the corresponding vector of

[7] For the simpler expression of the backward representation in which the lag values of the variables in (1) are substituted by forward values, Tong and Zhang (2005) and Chan *et al.* (2006) show that a necessary condition for the VAR(p) model to have this backward representation is that the covariance matrices $\Upsilon(h) = E\left[(Y_t - E(Y_t))(Y_{t-h} - E(Y_t))'\right]$ are symmetric for all h. This is a very strong restriction not likely to be satisfied in real data systems.

residuals defined as in (4). Center and scale the residuals by using the factor $[(T-p)/(T-2p)]^{0.5}$ recommended by Stine (1987). Denote by $\widehat{F}_a$ the empirical distribution function of the centered and rescaled residuals, $\widehat{F}_a(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(\widehat{a}_t < x)$ where $\mathbf{1}(\cdot)$ is an indicator variable which takes value 1 if the argument is true.

Step 2. From a set of p initial values, say $\mathbf{Y}_0^* = \{Y_{-p+1}^*, \ldots, Y_0^*\}$, construct a bootstrap series $\{Y_1^*, \ldots, Y_T^*\}$ as follows

$$Y_t^* = \widehat{\mu} + \widehat{\Phi}_1 Y_{t-1}^* + \cdots + \widehat{\Phi}_p Y_{t-p}^* + \widehat{a}_t^*, \ t = 1, \ldots, T, \tag{20}$$

where $\widehat{a}_t^*$ are independent draws from $\widehat{F}_a$. Obtain $\widehat{B}^* = (\widehat{\mu}^*, \widehat{\Phi}_1^*, ..., \widehat{\Phi}_p^*)$, a bootstrap replicate of the LS estimates by fitting model (1) to the bootstrap replicate $\{Y_1^*, ..., Y_T^*\}$.

Step 3. Forecast using model (1) with the parameters substituted by their bootstrap estimates and fixing the last p observations of the original series, as follows

$$\widehat{Y}_{T+h|T}^* = \widehat{\mu}^* + \widehat{\Phi}_1^* \widehat{Y}_{T+h-1|T}^* + \cdots + \widehat{\Phi}_p^* \widehat{Y}_{T+h-p|T}^* + \widehat{a}_{T+h}^*, \tag{21}$$

with $\widehat{a}_{T+h}^*$ being a random draw from $\widehat{F}_a$ and $\widehat{Y}_{T+h}^* = Y_{T+h}$, $h \leq 0$.

Step 4. Repeat steps 1 to 3 R times.

Using this procedure, we obtain R bootstrap replicates of Y_{T+h}, denoted by $\{\widehat{Y}_{T+h|T}^{*(1)}, ..., \widehat{Y}_{T+h|T}^{*(R)}\}$ and their corresponding bootstrap distribution which can be used to delimit prediction intervals, ellipsoids and cubes with appropriate probability content just as before. For instance, if we are interested in the nth component of Y_{T+h}, we can approximate the bootstrap density of the future value by using $\{\widehat{y}_{n,T+h}^{*(1)}, \widehat{y}_{n,T+h}^{*(2)}, ..., \widehat{y}_{n,T+h}^{*(R)}\}$, so that a (1-$\alpha$)100% bootstrap prediction interval for the nth variable is given by

$$BI_{T+h} = \{y_{n,T+k} | y_{n,T+k} \in [q_B^*(\tau), q_B^*(1-\tau)]\} \tag{22}$$

where $q_B^*(\tau) = G_{n,B}^{*-1}$ is the τth percentile of $G_{n,B}^*(x) = \#(\widehat{y}_{n,T+k}^{*(b)} \leq x)/R$. Using similar arguments, we can obtain the following prediction ellipsoids and cubes

$$BE_{T+h} = \left\{ Y_{T+h} | \left[Y_{T+h} - \overline{\widehat{Y}_{T+h}^*} \right]' S_{\widehat{Y}^*}^B(h)^{-1} \left[Y_{T+h} - \overline{\widehat{Y}_{T+h}^*} \right] < Q_B^* \right\} \tag{23}$$

$$BC_{T+h} = \left\{ y_{n,T+h} | y_{n,T+h} \in \cup_{n=1}^{N} \left[q_B^*\left(\frac{\tau}{2}\right), q_B^*\left(1-\frac{\tau}{2}\right) \right] \right\} \tag{24}$$

respectively, where $\tau = \alpha/N$, $\overline{\widehat{Y}^*_{T+h|T}}$ is the mean of $\{\widehat{Y}^{*(1)}_{T+h|T}, \widehat{Y}^{*(2)}_{T+h}, ..., \widehat{Y}^{*(R)}_{T+h}\}$ and Q^*_B is obtained as in (18).

To illustrate the implementation of the new bootstrap procedure proposed in this paper we consider again the simulated bivariate time series previously described. Panel (e) of Figure 1 displays the kernel estimate of the bootstrap joint density of $y_{1,T+1}$ and $y_{2,T+1}$. It is clear that it is more adequate to approximate to the true density than Gaussian procedures and similar to the bootstrap procedure proposed by Kim (1999). This is also evident from panel (e) of Figure 2, which looks more alike to that of panel (a) than those plotted in the other panels. Nothing seems to be lost by not using the backward representation. Figure 3 plots the bootstrap cube and elliptical regions obtained from the new bootstrap density. Although the bootstrap densities are very different from the densities obtained by the standard and asymptotic approximations, we cannot observe a big difference in the location between the new bootstrap ellipsoid and those ellipsoids obtained with the alternative procedures. This is due to the fact that first two moments estimates involved in the definition of the ellipsoids do not differ significantly for all the procedures; i.e., all estimate similar centers and dispersions of the future value. For instance, the center of the ellipsoid in Figure 3 is $(-2.21, -0.27)$ for Gaussian alternatives and $(-2.21, -0.26)$ and $(-2.19, -0.26)$ for Kim's and the new bootstrap approaches, respectively, while the one-step-ahead MSE estimates are given by

$$\widehat{\Sigma}_{\widehat{Y}}(1) = \begin{bmatrix} 0.98 & 0.79 \\ 0.79 & 0.98 \end{bmatrix}, \qquad \widehat{\Sigma}^l_{\widehat{Y}}(1) = \begin{bmatrix} 1.03 & 0.83 \\ 0.83 & 1.00 \end{bmatrix},$$

$$S^K_{\widehat{Y}^*}(h) = \begin{bmatrix} 1.04 & 0.84 \\ 0.84 & 1.03 \end{bmatrix} \text{ and } S^B_{\widehat{Y}^*}(h) = \begin{bmatrix} 1.09 & 0.90 \\ 0.90 & 1.08 \end{bmatrix}.$$

However, we see different locations between the bootstrap cubes and the Gaussian cubes, reflecting the ability of the former to adapt to the asymmetry. Finally, Figure 4 displays the one-step-ahead kernel estimate of the density of $y_{1,T+1}$ obtained by the new bootstrap which like Kim's is much closer to the true density than the Gaussian alternatives. After all, this example suggests that both bootstrap densities resemble better the true density than the traditional procedures. Nevertheless, our bootstrap procedure is much simpler than Kim (1999) and its asymptotic validity can be proven without assuming Gaussianity as it is shown in the next subsection.

3.2 Asymptotic validity

Consider again the stationary VAR(p) model in (1) and assume that there exists p presample values Y_{-p+1}, Y_{-p+2},..., Y_0. The error is given by

$$a_t(B) = Y_t - \Phi_1 Y_{t-1} - ... - \Phi_p Y_{t-p}, t = 1, ..., T \tag{25}$$

where we make explicit that it depends on the unknown parameters contained in B. Denote by F the distribution of a_t. Assume that we estimate B by a $\sqrt{T}$-consistent estimator denoted by $\hat{B}$. The estimated residual is then given by

$$a_t(\widehat{B}) = Y_t - \widehat{\Phi}_1 Y_{t-1} - ... - \widehat{\Phi}_p Y_{t-p}, t = 1, ..., T, \tag{26}$$

which after being centered they have $F(\widehat{B})$ as distribution function.[8] The asymptotic validity of the bootstrap depends to a large extend on the approximation given by $F(\hat{B})$ to F as the sample size T increases. In particular, Theorem 2.4 of Paparoditis (1996) states that $d_2(F(\hat{B}), F(B)) \to 0$ in probability as $T \to 0$, where d_2 is a Mallow's metric.[9]

In this paper we consider the LS estimator $\hat{b} = vec(\widehat{B})$, where vec is the column stacking operator. The second part of the asymptotic validity relies on the approximation of $\| \widehat{b} - b \|$ given by $\| \widehat{b^*} - \widehat{b} \|$. Let's define the sequence $\{\| I(p) \|^2\}$ of pN^2x1 vectors of constant such that $0 < K_1 \leq \| I(p) \|^2 \leq K_2 < \infty$ and $s_T = \sqrt{T} I(p)'(\widehat{b} - b)$. Theorem 3 of Lewis and Reinsel (1985, p.399) states $s_T = \sqrt{T} I(p)'(\widehat{b} - b) \xrightarrow{d} N(0, I(p)'(\Upsilon^{-1} \otimes \Sigma_a) I(p))$ in probability where $\Upsilon = P \lim \left(\frac{ZZ'}{T} \right)$ The bootstrap counterpart of s_T is given by $s_T^* = \sqrt{T} I(p)'(\widehat{b^*} - \widehat{b})$. Denote the laws of s_T and s_T^* by ℓ and ℓ^*, respectively. Theorem 3.2 of Paparoditis (1996, p.284) establishes that $d_2(\ell, \ell^*) \to 0$ ℓ^* in probability as $T \to \infty$, where ℓ^* is conditioned on a given sample Y. As the convergence in Mallow's metric implies the convergence of the corresponding random variables then $s_T^* \xrightarrow{d} N(0, I(p)'(\Upsilon^{-1} \otimes \Sigma_a) I(p))$. As far as s_T and s_T^* converge weakly to the same Gaussian distribution in probability, the bootstrap validity is established. Specifically, Theorem 3.3 of Pararoditis (1996, p.285) sets the asymptotic validity of the Yule Walker estimator of B for the bootstrap procedure, but it still remains valid for any estimator which satisfies

[8] The asymptotic validity is established for centered residuals, but according to Stine (1987) it remains valid if they are also rescaled.

[9] Assume that X and Y are with distribution G_X and G_Y random variables in R^N. The Mallows distance between G_X and G_Y is define as $d_p(F, G) = inf\{E(X - Y)^p\}^{p/2}$, where the minimum is taken over all joint probability distribution F for (X, Y) such that the marginal distribution of X and Y are G_X and G_Y, respectively. In what follows we set $p = 2$.

$\| \widehat{B} - B \| = O_p(p^{1/2}/T^{1/2})$. Consequently, it is valid for the LS estimator considered in this paper.[10]

Finally, the asymptotic validity of the bootstrap approximation of the distribution of a future value focuses on $Y^*_{T+h|T}$, which is given by

$$\widehat{Y}^*_{T+h|T} = \widehat{\Phi}^*_0 + \widehat{\Phi}^*_1 \widehat{Y}^*_{T+h-1|T} + ... + \widehat{\Phi}^*_p \widehat{Y}^*_{T+h-p|T} + \widehat{a}^*_{T+h|T} \tag{27}$$

Theorem. Let $\{Y_t, t = -p+1, ..., 1, 2, ...T\}$ be a realization of a stationary VAR(p) process $\{Y_t\}$ with $E(a_t) = 0$ and $E|a_{it}a_{jt}a_{lt}a_{rt}| < \infty$ for $1 \leq i, j, l, r \leq N$, $\hat{B}$ the LS estimator of B and $\widehat{Y}^*_{T+h|T}$ obtained by following the steps 1 to 4 in the previous subsection. Then, $\widehat{Y}^*_{T+h|T}$ conditioned on $\{Y_t, t = -p+1, ..., 1, 2, ...T\}$ converges weakly in probability to Y_{T+h} as $T \to \infty$.

Proof. Following the arguments in Pascual *et al.* (2004a), let's first consider the one-step-ahead bootstrap future value given by

$$\widehat{Y}^*_{T+1|T} = \widehat{\Phi}^*_0 + \widehat{\Phi}^*_1 Y_T + ... + \widehat{\Phi}^*_p Y_{T-p+1} + \widehat{a}^*_{T+1} \tag{28}$$

For $h = 2$ we have

$$\widehat{Y}^*_{T+2|T} = \widehat{\Phi}^*_0 + \widehat{\Phi}^*_1 \widehat{Y}^*_{T+1|T} + ... + \widehat{\Phi}^*_p Y_{T-p+2} + \widehat{a}^*_{T+2} \tag{29}$$

and replacing the $\widehat{Y}^*_{T+1|T}$ in (28) it follows that

$$\widehat{Y}^*_{T+2|T} = \widehat{\Phi}^*_0(1 + \widehat{\Phi}^*_1) + N_1(\widehat{B}^*)Y_T + ... + N_p(\widehat{B}^*)Y_{T-p+1} + M_1(\widehat{B}^*)\widehat{a}^*_{T+1} + \widehat{a}^*_{T+2} \tag{30}$$

which expresses the bootstrap future values as a function of the given realization $\{Y_{T-p+1}, ..., Y_T\}$, the independent random draws $\hat{a}^*_{T+h}$ and continuous functions of the bootstrap parameter estimates $\widehat{B}^*$. Proceeding in this way we obtain the following expression

$$\begin{aligned} Y^*_{T+h|T} = {} & N_0(\widehat{B}^*) + N_1(\widehat{B}^*)Y_T + ... + N_p(\widehat{B}^*)Y_{T-p+1} \\ & + M_1(\widehat{B}^*)\widehat{a}^*_{T+1} + M_2(\widehat{B}^*)\widehat{a}^*_{T+2} + ... + \widehat{a}^*_{T+h} \end{aligned} \tag{31}$$

where N_0 maps $\widehat{B}^*$ to $Nx1$, N_j, $j = 1, ..., p$, and M_i, $i = 1, ..., h-1$, maps $\widehat{B}^*$ to NxN matrices. The functions N_js and M_is are continuous functions and different for each forecast horizon.

Note that $N_0(\widehat{B}^*)$ and $N_j(\widehat{B}^*)Y_{T-j+1}$ are just continuous functions of $\widehat{B}^*$ and Y_{T-j+1} is a

[10]The VAR process may have an infinite order. In particular, the requirement for the validity of the bootstrap estimator is that $p/T \to 0$ as $T \to 0$ which is fulfilled in the case of finite order VAR process.

given value, so that both quantities converge weakly in probability to $N_0(B)$ and $N_j(B)Y_{T-j+1}$, respectively. Furthermore, $M_i(\widehat{B}^*)$ converges weakly in probability to $M_i(B)$ and $\widehat{a}^*_{T+j}$ converge weakly to a^*_{T+j} in probability; therefore, by applying the bootstrap version of Slutsky's Theorem we obtain that $M_i(\widehat{B}^*)\widehat{a}^*_{T+j} \xrightarrow{d} M_i(B)\widehat{a}_{T+j}$. Finally, observe that $\widehat{a}^*_{T+j}$ are independent and thus all terms in (31) converge weakly in probability. Consequently, $Y^*_{T+h|T} \xrightarrow{d} Y_{T+h}$ as T goes to infinity, in probability.

■

3.3 Small sample properties

In this subsection, we carry out Monte Carlo experiments to analyse the finite sample properties of the marginal intervals, ellipsoids and cubes constructed by using the new bootstrap procedure proposed in this paper to obtain bootstrap densities. They are compared with those of the Gaussian approximations and the bootstrap procedure based on the backward representation.

The DGP considered all through the experiments is the following bivariate VAR(1) model

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = + \begin{bmatrix} -0.5 & 0 \\ 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \tag{32}$$

where $a_t = (a_t^{(1)}, a_t^{(2)})'$ is an independent white noise vector with contemporaneous covariance matrix given by $vec\Sigma_a = (1, 0.8, 0.8, 1)'$. This model have been previously considered by Kim (2001)[11]. We examine three alternative distributions for a_t, namely, Gaussian, Student-5 and $\chi^2(4)$. The last two distributions are proposed to capture fat-tailed and asymmetric distributions. In order to have the covariance matrix described above, the Student-5 and $\chi^2(4)$ have been centered and rescaled. The number of Monte Carlo replicates is 1000 and the sample sizes are $T = 25$ and 100. We generate $F = 3000$ future values of of the process Y_{T+h} using the assumed distribution with mean and MSE given by (2) and (3), respectively, which approximate the density of the future value conditional on $\{Y_-p+1, ..., Y_0, ..., Y_T\}$. The parameters are estimated by LS. Then, the 90%, 95% and 99% marginal prediction intervals for forecast horizons, h, from 1 to 8 are computed for each of the variables in the system by using i) the Gaussian prediction interval without incorporating the parameter uncertainty (GI) in equation (5), ii) the asymptotic (AI) in equation (11), iii) the KI in (16) and, iv) the new bootstrap procedure proposed in

[11]The model is stationary with roots given by (-2,2). Results for other alternative VAR(1) and VAR(2) models are similar to those reported in this paper. They are available from the authors upon request.

this paper as in equation (22). The same is done for the prediction regions, whether they are based on the elliptical assumption such as GE in (6), AE in (12), KE in (17) and BE in (23), or the Bonferroni approximations GC in (7), AC in (13), KC in (19) and BC in (24). The number of bootstrap replicates is $R = 4999$. The number of future values inside the interval and regions is then counted to obtained the empirical coverage of each procedure. We also compute the volume which is given by i) the length of the individual prediction intervals, ii) $V = [\pi^{0.5N}/\Gamma(1+0.5N)][\chi^2_{1-\alpha}(N)]^{0.5N}\{det[\Sigma_y(h)^{-1}]\}^{-0.5}$, where $\Gamma(\cdot)$ is the gamma function, for elliptical regions, and (iii) the product of the lengths of the intervals jointly making the Bonferroni approximation. Finally, for individual prediction intervals we calculate the coverage on the left and right sides of the empirical density for the future value.

To compare the alternative prediction densities considered, we rely on an approximation of the Mallows distance based on the following factorization of a joint density

$$g(y_{1,T+h}, y_{2,T+h}|Y_T) = g_{12}(y_{1,T+h}, |y_{2,T+h}, Y_T) * g_2(y_{2,T+h}|Y_T). \tag{33}$$

In our setting, we consider the conditional densities of $y_{1,T+h}$ given three different values of $y_{2,T+h}$, which are the 25th, 50th and 75th quantiles of the real density of $y_{2,T+h}$; these values are called by q_1, q_2 and q_3, respectively. To approximate it given a set of realization that characterized a distribution $\{\widehat{Y}^{(1)}_{T+h}, ..., \widehat{Y}^{(R)}_{T+h}\}$, we proceed by considering the realizations on the neighborhood of these quantiles of $\widehat{y}_{2,T+h}$. For instance, the $g_{12}(y_{1,T+h}, |q_j, Y_T)$ was approximated by the set of realization satisfying

$$C_j = \{Y_{T+h}|\widehat{Y}_{T+h} \in [\widehat{y}_{1,T+h}, q_j]' \pm [0, e]').$$

Call $y^C_{1,T+h}$ the R_{12} observations of $\widehat{Y}_{T+h}$ that fulfill that condition.[12] Then, the conditional distribution given q_j is $G_{1|2}(x) = \#(\widehat{y}^C_{1,T+k} \leq x)/R_{12}$. The marginal density is obtained just by considering all the realization in the dimension $y_{2,T+h}$, just as before $G_2(x) = \#(\widehat{y}^{(r)}_{2,T+k} \leq x)/R$. Once we have the conditional and marginal distributions, we compute the Mallows distance by computing the difference among 100 quantiles of the distributions involved. Giving a mass $\frac{1}{100}$ to each quantile, then the Mallows distance between, for instance, the real distribution G_2 (or G_{12} in the case of the conditional distribution) and its bootstrap counterpart G^*_2 (or G^*_{12}) can be

[12]Different values of e were proved. Although here we considerer $e = 0.1$, those values within the limits $(0.025, 0.125)$ yield practically the same results. Obviously, for the choice of e scale consideration matters.

approximated as follows

$$d_2(G_2, G_2^*) \approx \left[\frac{1}{100} \sum_{i=1}^{100} \parallel y_{2,T+h}^{(q)} - y_{2,T+h}^{*(q)} \parallel^2 \right]^{\frac{1}{2}}$$

where $y_{1,T+h}^{(q)}$ and $y_{1,T+h}^{*(q)}$ are the qth quantile of the real and bootstrap marginal density of the first component. We incorporate this distance calculation exercise within the Monte Carlo simulation framework described above.

We describe first the results obtained when the objective is the prediction of the marginal distribution of the first variable in the system, $y_{1,T+h}$. Table 1 reports the Monte Carlo means and standard deviations of the coverages and lengths of its one-step-ahead and eight-steps-ahead prediction intervals when the nominal coverage is 95%. Table 1 also reports the average of the coverages left out on the right and left of the intervals. When looking at the coverages, Table 1 shows that, regardless of the error distribution, when the sample size is small, $T = 25$, all intervals have coverages smaller than the nominal; Lütkepohl (1991) and Kim (1999) also observe undercoverages. In general, the coverage of the GI is the smallest with the AI being slightly closer to the nominal than the bootstrap when the prediction horizon is 1. However, when predicting eight-steps-ahead into the future, the bootstrap intervals are superior to the GI and AI. On the other hand, when $T = 100$, the coverages of all intervals are similar irrespective of the distribution and horizon. When looking at the results reported on average lengths, we can observe that they are similar for all procedures and distributions and slightly larger when the horizon is larger. Finally, Table 1 also reports the average coverage on the left and right of the prediction intervals which show that GI and AI are not able to cope with the asymmetry of the $\chi^2(4)$ error distribution. These intervals left most observations on just one of their sides. Note that the gains of bootstrap intervals are clearer for longer horizons and when the errors are asymmetric. It is important to remark that although the bootstrap procedure proposed in this paper is much simpler than that proposed by Kim (1999) by avoiding the use of the backward representation, the performance of both prediction intervals is comparable. Therefore, these experiments illustrate that there is not price to pay for not using the backward representation.

Consider now that the objective is to obtain joint prediction regions that contain the two variables in the system with a given nominal coverage. First, we focus on the performance of the prediction cubes constructed by using the Bonferroni's approximation. Table 2 reports the corresponding empirical coverages and volumes when the nominal coverage is 90%. The results show that, regardless of the error distribution, when $T = 25$, the AC has empirical coverage

closer to the nominal than any other procedure. However, when $h = 8$ the bootstrap regions have better properties. On the other hand, when $T = 100$ all procedures tend to overestimate the nominal coverage with the bootstrap having coverages closer to the nominal. With respect to volumes, both bootstrap procedures provide regions that are generally larger than the regions obtained by the standard methodology, as they incorporate the parameter uncertainty due to the estimation process. Furthermore, Table 2 shows that the volume associated with non-normal errors are, in general, larger than those corresponding to Gaussian innovations, a feature that is also highlighted by Kim (1999). Note that, as before, there are not differences between the sample performance of the bootstrap procedure propose by Kim (1999) and the simpler one proposed in this paper.

Regarding the performance of the ellipsoidal prediction regions, Table 3 displays similar information to that in Table 2 for a VAR(1) model in (8) for 90% nominal coverage. The main fact to note is that the empirical coverages in Table 2 are below the corresponding coverages in Table 3. This is not surprising since the Bonferroni's regions provide a cubical approximation with larger volume than ellipsoids, so that it is expected at least the same coverages for the former; see Figure 3. The other features of Table 3 are similar to those commented before for the Bonferroni's regions.

Finally, Table 4 contains mean distances obtained for conditional and marginal densities. Regarding to the marginal density of $y_{2,T+h}$, we observe that when the errors are non-Gaussian and the sample size is medium ($T = 100$), bootstrap densities perform similarly and better than the alternatives based on Gaussian assumptions, regardless of the forecast horizon. On the other hand, when we consider the conditional densities note that, as expected, in the presence of a Gaussian error and sample size is $T = 25$, the standard and asymptotic prediction densities are closer to the real than bootstrap densities no matter the forecast horizon. But when the sample size increases to $T = 100$ and the forecast horizon gets longer $h = 8$, the differences between the bootstrap densities and those based on a Gaussian assumptions tend to vanish. In case of the error distribution with heavy tails such as the Student-5, the Gaussian and asymptotic procedures do provide good approximations of the real conditional prediction densities, as bootstrap alternatives do, when the sample is either small or medium size and the forecast horizon is $h = 1$. Furthermore, as the forecast horizon gets longer, bootstrap prediction densities are closer to the real than their alternatives. Finally, when the error is $\chi^2(4)$ and the sample size is $T = 25$, bootstrap procedures perform better for the q_1 and q_2, regardless of the forecast horizon. When $T = 25$ and $h = 8$, the Gaussian approximation overcome bootstrap densities probably because the need to raise

the number of replication when q_3 to capture the large asymmetry of that part of the error distribution; see Figure 1 and Figure 3. On the other hand, when $T = 100$ bootstrap densities perform clearly better than any other alternative.

After all, the simulations carried out show that the new bootstrap procedure performs better than traditional methods and does not worse than the procedure based on the backward representation with the advantage that it is easier to implement.

4 Empirical application

In this section, we implement the proposed bootstrap procedure to construct prediction intervals, ellipsoids and cubes for quarterly US inflation (π_t), unemployment rate (u_t) and GDP growth (g_t) observed quarterly from 1954Q3 to 2010Q4.[13] The inflation rates are computed by $\pi_t = log(IPI_t/IPI_{t-1}) * 100$ where IPI is the US Implicit Price Deflactor. The unemployment is measured by the civilian unemployment rate[14] and, finally, the GDP growth is given by $g_t = log(GDP_t/GDP_{t-1}) * 100$ where GDP is the US Real Gross Domestic Product. The whole sample period has been split into an estimation sample from 1954Q3 to 2008Q4 ($T = 218$) and an out-of-sample period from 2009Q1 to 2010Q4. Table 5, which reports some descriptive statistics for the estimation period, shows that for the null of a unit root in the inflation cannot be rejected when using the Dickey-Fuller test; see for instance Stock and Watson (2007) for a discussion of the existence of a unit root in US quarterly IPI inflation. On the other hand, unemployment is stationary at 5% and GDP growth at 1%. Hence a reduce VAR model in which the first difference of inflation ($\triangle\pi$) and the current values of unemployment rate and GDP growth rates depend on their lagged values seems to be appropriate for the sample period 1954Q3-2008Q4. On the other hand, the first four columns of Table 5 report the mean, standard deviation, skewness and kurtosis. When considering all the series together the measures of skewness and kurtosis are based on Mardia (1970). It can be observed that all the series are characterized by significant skewness and kurtosis, suggesting non-Gaussian distributions. This is corroborated by the Gaussianity test of Doornik and Hansen (2008) which rejects the null hypothesis of Gaussian series both individually and jointly. To choose the lag-order p, we use several order selection measures, such as Akaike, Schwarz, Hannan-Quin and the final prediction error criterions. All of them suggest a VAR(3), which is the model finally fitted.

[13] The data was obtained from the Federal Reserve Bank of St. Louis, webpage: www.stlouisfed.org.

[14] As the unemployment data is monthly, we chose the value of the last month of each quarter to measure its quarterly.

The descriptive statistics for the centered and scaled residuals, $\widehat{a}_t$, are displayed in Table 6. All the estimated residuals have significant excess of kurtosis and Gaussianity is rejected. Overall, this fact suggests that the standard approach to forecasting in the context of VAR models may be misleading when working with these US quarterly series, and therefore the implementation of bootstrap prediction intervals, ellipsoids and cubes is advisable. The estimated VAR(3) model is used to construct the multi-step out-of-sample prediction densities up to horizon 8. Figure 5 plots one-step-ahead and eight-steps-ahead bootstrap marginal prediction densities for each of the variables in the system together with the corresponding Gaussian densities. It can be observed a positive skewness in case of unemployment, just in line with the descriptive statistic of the prediction error. A similar analysis can be done for the multivariate prediction. Figure 6 and 7 plot bivariate one-step and eight-steps-ahead Gaussian and bootstrap densities for the variables considered two by two, respectively. The skewness and the kurtosis of the series individually are also clearly manifested in the shape of the kernel estimates of the bivariate densities. When $h = 1$, this is more evident for first difference of inflation-unemployment and unemployment-GDP growh densities, which are affected by the skewness of unemployment and the kurtosis of inflation rate and unemployment.

Finally, we construct the prediction intervals and regions. Figure 8 plots the multi-step point predictions together with the observed series and the 95% new bootstrap and Gaussian prediction intervals for inflation, unemployment and growth up to horizon 8. Note that as expected the bootstrap prediction intervals are usually larger than those obtained by the standard approach, a fact that is related to the shape of out-of-sample prediction densities. On the other hand, the one-step-ahead density for the unemployment has high kurtosis, positive skewness which is manifested in wider intervals stemming mainly from upper bound. For inflation rate (panel a) and GDP grwoth (panel c), the observed series lie within the 95% prediction bands for both GI and BI. However, for the unemployment series we see that standard method fails to capture the first two out-of-sample values; though the BI captures all the unemployment rates in 2009Q1 and 2010Q4.

Similarly, we construct one-step-ahead and eight-steps-ahead prediction ellipsoids and cubes which are plotted in Figure 9 for the first difference of inflation-unemployment, unemployment-GDP growth and first difference of inflation-GDP growth for the Gaussian and the new bootstrap procedures. The point predictors and the observed out-of-sample values for the series are also plotted in Figure 9. Observe that the later values lie outside the GE and GC for the first difference of inflation-unemployment regions, but it is still within the border of the BE and BC. In panel

(b) of Figure 9 we see that something similar occurs for the unemployment-GDP, in which case the observed value lies outside SC but inside the rest of regions; i.e., GE, BC and BE. At last, the out-of-sample value lies inside the cubes and ellipsoids of all procedures for the inflation-GDP growth. Overall, the shape of the regions constructed by each procedure depends on the estimated prediction densities. For instance, in panel (c) of Figure 9 we see that the prediction cube for the unemployment-GDP growth is larger mainly from the unemployment side, which is in part caused by the positive skewness of the unemployment density.

At last, using the out-of-sample observation, we calculate the coverages and volumes of the prediction ellipsoids and cubes obtained using each of the procedures described in this paper. Our strategy is to run a rolling-window estimation for the VAR(3) model starting with data from 1954Q3 to 1966Q4. The sample size is $T = 50$. Then we construct one-step-ahead elliptical and cubical regions for all the procedures. We count how many real observations belong to the prediction regions constructed and compute their volumes. Table 7 displays the results. With respect to coverage, there are only slight differences between Kim's and the new bootstrap procedures. However, our method provides smaller regions than the former. Note also that our bootstrap attains better coverage than those approaches based on Gaussian assumptions. After all, the results reported in Table 7 are in line with those obtained with simulated data in the sense that there are no large differences between using the bootstrap procedure proposed in this paper or that proposed by Kim (1999), with the advantage that the our algorithm is simpler from a computational point of view.

5 Conclusions

In this paper we extend the bootstrap procedure proposed by Pascual *et al.* (2004a) to construct prediction regions in multivariate VAR(p) models. The main attractive of the new bootstrap procedure when compared with alternative bootstrap procedures previously proposed in the literature, is that it does not require the backward representation. As a result, it is possible to prove its asymptotic validity without assuming Gaussian errors. Furthermore, the new procedure can be implemented in multivariate models without a backward representation while its computational burden is reduced. We show that the procedure works properly in incorporating the parameter uncertainty and is robust in presence of non-Gaussian errors.

Given its simplicity, the bootstrap procedure proposed in this paper can be easily extended to models with MA components and to cointegrated and non-cointegrated non-stationary systems.

Also, as proposed by Pascual *et al.* (2004b) in univariate systems, our proposed procedure can also be implemented to obtain prediction intervals for the original observations when a VAR model is fitted to transformed observations; see Ariño and Franses (2000) and Bardsen and Lütkepohl (2011).

Finally, it is often of interest to predict future values of one of the variables in the system conditional on particular values of other variables of the system; see, for example, Waggoner and Zha (1999) for a macroeconomic example. The conditional densities can be easily obtained in the context of the bootstrap algorithm proposed in this paper by keeping the bootstrap replicates $\widehat{Y}^{*(b)}_{T+h|T}$ that satisfy the conditions. Another interesting application of the proposed procedure is the construction of confidence intervals for impulse-response functions; see, for example, Killian (1998) for a bootstrap procedure based on the backward representation and Fachin and Bravetti (1996) who proposed a bootstrap alternative which is not. Related with response-impulse functions is the construction of prediction paths; see Staszewska-Bystova (2010) for bootstrap prediction bands based on the backward representation.

Further effort should be directed to the construction of prediction regions. In this sense, it is worth noting that the prediction ellipsoids are only appropriate when the distribution of the future values of the variables in the system is approximately multivariate Gaussian. When the distribution of Y_{T+h} departs from Gaussianity, the quality of such approximation deteriorates. The Bonferroni's approximation to the ellipses does provide a better solution capturing the asymmetry of the distribution. However, the shape of the cube could not be appropriate in some cases. Consequently, it would be interesting to give regions that depart either from the elliptical or rectangular shapes. On the other hand, using HDR as proposed by Hyndman (1996) and applied to GARCH models by Eklund (2005) and more recently by Terärvirta and Zhao (2011), could not be adequate when the number of components in the system is large.

References

[1] Ariño, M.A. and P.H. Franses (2000), Forecasting the levels of vector autoregressions log-transformed time series, *International Journal of Forecasting,* 16, 111-116.

[2] Bardsen, G. and H. Lütkepolh (2011), Forecasting levels of log variables in vector autoregressions, *International Journal of Forecasting,* 27(4), 1108-1115.

[3] Chan, K.-S., L.-H. Ho and H. Tong (2006), A note on time-reversibility of multivariate linear processes, *Biometrika,* 93(1), 221-227.

[4] Chevillon, G. (2009), Multi-step forecasting in emerging economies: An investigation of the South African GDP, *International Journal of Forecasting,* 25, 602-628.

[5] Chow, H.K. and K.M. Choy (2006), Forecasting the global electronic cycle with leading indicators: A Bayesian VAR approach, *International Journal of Forecasting,* 22, 301-315.

[6] Clements, M.P. and J. Smith (2002), Evaluating multivariate forecast densities: a comparison of two approaches, *International Journal of Forecasting,* 18, 397-407.

[7] Diebold, F.X., J. Hahn and A.S. Tay (1999), Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange, *The Review of Economics and Statistics,* 81(4), 661-673.

[8] Doornik, J.A. and Hansen, H. (2008), An Omnibus Test for Univariate and Multivariate Normality, *Oxford Bulletin of Economics and Statistics,* 70, 927-939.

[9] Eklund, B. (2005), Estimating confidence regions over bounded domains, *Journal Computational Statistics & Data Analysis,* 49(2), 349-360.

[10] Fachin, S. and L. Bravetti (1996), Asymptotic Normal and Boostrap Inference in Structural VAR Analysis, *Journal of Forecasting,* 15, 329-341.

[11] Gómez, N. and V. Guerrero (2006), Restricted forecasting with VAR models: An analysis of a test for joint compatibility between restrictions and forecasts, *International Journal of Forecasting,* 22, 751-770.

[12] Grigoletto, M. (2005), Bootstrap prediction regions for multivariate autoregressive processes, *Statistical Methods and Applications,* 14, 179-207.

[13] Hall, P. (1992), *The bootstrap and Edgeworth expansions*, Springer-Verlag, New York.

[14] Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

[15] Hyndman, R.J. (1996), Computing and graphing highest density regions, *The American Statistician,* 50(2), 120-126.

[16] Kilian, L. (1998). Confidence intervals for impulse responses under departures from normality, *Econometric Reviews*, 17, 1-29.

[17] Kim, J.H. (1997), Relationship between the forward and backward representation of the stationary VAR model, Problem 97.5.2, *Econometric Theory,* 13, 899-990.

[18] Kim, J.H. (1998), Relationship between the forward and backward representation of the stationary VAR model, Solution 97.5.2, *Econometric Theory,* 14, 691-693.

[19] Kim, J.H. (1999), Asymptotic and bootstrap prediction regions for Vector Autoregression, *International Journal of Forecasting,* 15, 393-403.

[20] Kim, J.H. (2001), Bootstrap after bootstrap prediction intervals for autoregressive models, *Journal of Business & Economic Statistics,* 19(1), 117-128.

[21] Kim, J.H. (2004), Bias-corrected bootstrap prediction regions for Vector Autoregression, *Journal of Forecasting,* 23, 141-154.

[22] Lewis R. and Reinsel, G.C. (1985), Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis,* 16, 393-411.

[23] Lütkepolh, H. (1991), *Introduction to Multiple Time Series Analysis,* 2nd ed., Springer-Verlag, Berlin.

[24] Lütkepolh, H. (2006), Forecasting with VARMA models, in Elliot, G., C.W.J. Granger and A. Timmerman (eds.), *Handbook of Economic Forecasting,* Vol. 1, 287-325.

[25] Mardia, K.V. (1970), Measures of multivariate skewness and kurtosis with applications, *Biometrika,* 57(3), 519-530.

[26] Paparoditis, E. (1996). Bootstrapping autoregressivce and moving average parameter estimates of infinite order vector autoregressive processes, *Journal of Multivariate Analysis,* 57, 277-296.

[27] Pascual, L., J. Romo, and E. Ruiz (2004a). Bootstrap predictive inference for ARIMA processes, *Journal of Time Series Analysis,* 25, 449-465.

[28] Pascual, L., J. Romo, and E. Ruiz (2004b). Bootstrap prediction intervals for power-transformed time series, *International Journal of Forecasting,* 21(2), 219-235.

[29] Riise, T. and D. Tjostheim (1984), Theory and practice of multivariate ARMA forecasting, *Journal of Forecasting,* 3, 309-317.

[30] Rodríguez A. and Ruiz E. (2009), Bootstrap prediction intervals in statespace models, *Journal of Time Series Analysis,* 30(2), 167-178.

[31] Runkle (1987), Vector autoregressions and reality, *Journal of Business & Economic Statistics,* 5(4), 437-442.

[32] Schmidt, P. (1977), Some small sample evidence on the distribution of dynamic simulation forecasts, *Econometrica,* 45(4), 997-1005.

[33] Simkins, S. (1995), Forecasting with vector autoregressive (VAR) models subject to business cycle restrictions, *International Journal of Forecasting,* 11, 569-583.

[34] Sims, C.A. and T. Zha (1998), Error bands for impulse responses, *Econometrica,* 67(5), 1113-1155.

[35] Sims, C.A. and T. Zha (1999), Bayesian methods for dynamic multivariate models, *International Economis Review,* 39, 949-968.

[36] Staszewska-Bystova, A. (2010), Bootstrap prediction bands for forecast paths from vector autoregression models, *Journal of Forecasting,* doi:10.1002 /for.1205.

[37] Stine, J.H. (1987), Estimating properties of autoregressive forecasts, *Journal of Economic Perspectives,* 15(4), 101-115.

[38] Stock J.H. and M.W. Watson (2001), Vector autoregressions, *Journal of Economic Perspective,* 15(4), 101-115.

[39] Stock J.H. and M.W. Watson (2007), Why Has U.S. Inflation Become Harder to Forecast?, *Journal of Money, Credit and Banking,* 39, 13-33.

[40] Teräsvirta, T. and Zhao, Z. (2011). Stylized facts of return series, robust estimates, and three popular models of volatility, *Applied Financial Economics,* 21(1),67-94.

[41] Thombs L. A. and W.R. Schucany (1990), Bootstrap prediction intervals for autoregression, *Journal of the American Statistical Association,* 85, 486-92.

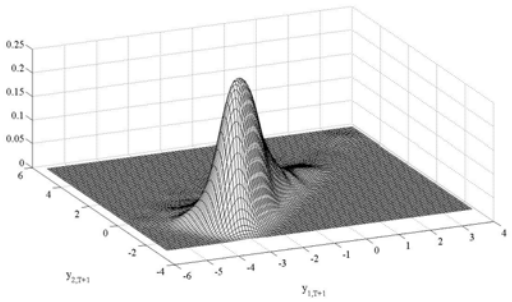
[42] Tong, H. and Z. Zhang (2005), On time-reversibility of multivariate linear processes, *Statistica Sinica,* 15, 495-504.

[43] Tay, A.S. and K.F. Wallis (2000). Density forecasting: a survey, *Journal of Forecasting,* 19, 235-254.

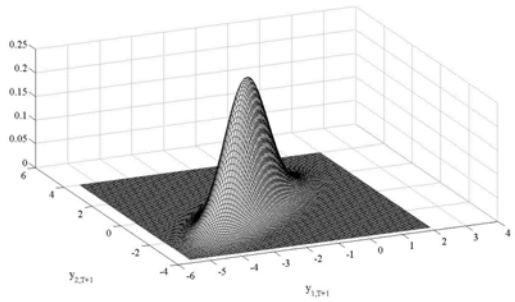
[44] Waggoner, D.F. and T. Zha (1999), Conditional forecasts in dynamic multivariate models, *The Review of Economics and Statistics,* 81(4), 639-651.

[45] West, K.D. (1996), Asymptotic inference about predictive ability, *Econometrica,* 64(5), 1067-1084.

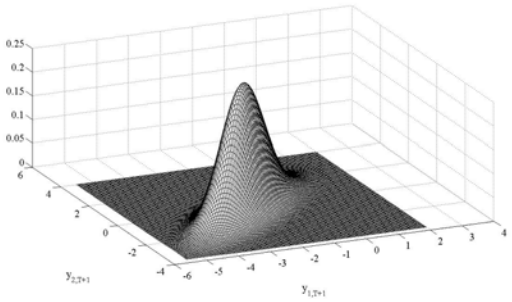
[46] West, K.D. and M.W. McCracken (1998), Regression based test of predictive ability, *International Economic Review,* 39, 817-840.



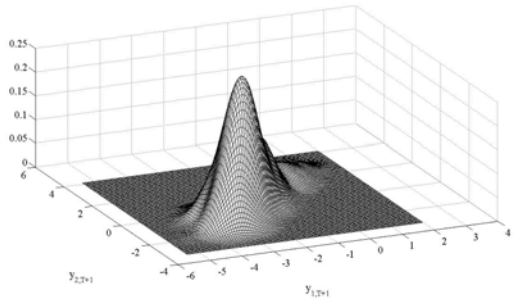
(a) Empirical



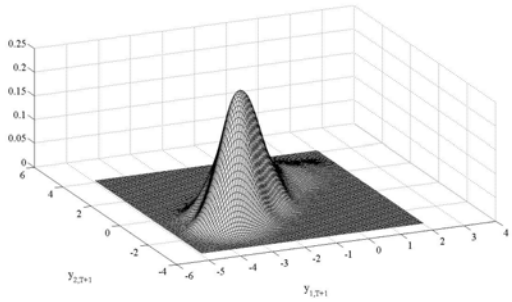
(b) Gaussian



(c) Asymptotic Gaussian

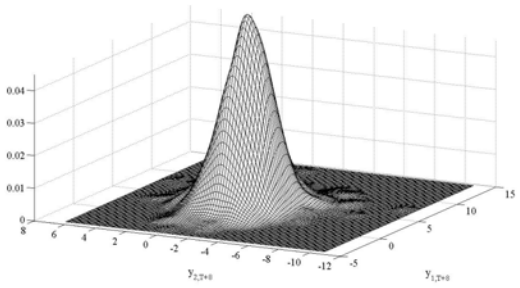


(d) Kim's bootstrap

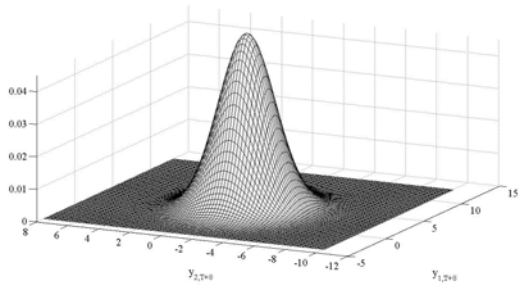


(e) New bootstrap

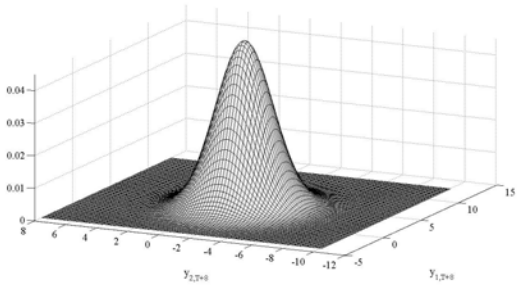
Figure 1: Kernel estimates of joint densities of one-step predictions for a bivariate series generated by a VAR(2) with $T = 100$ and $\chi^2(4)$ errors.



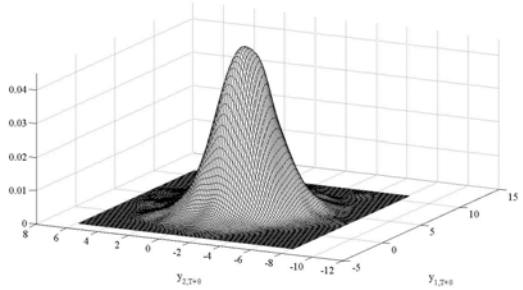
(a) Empirical



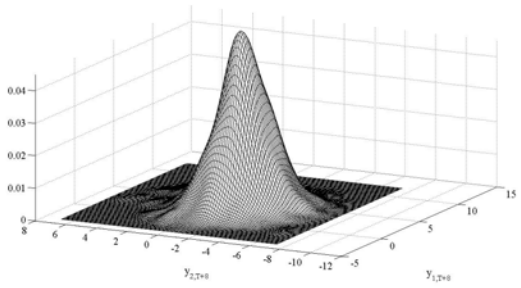
(b) Gaussian



(c) Asymptotic Gaussian



(d) Kim's bootstrap



(e) New bootstrap

Figure 2: Kernel estimates of joint densities of eight-steps predictions for a bivariate series generated by a VAR(2) with $T = 100$ and $\chi^2(4)$ errors.

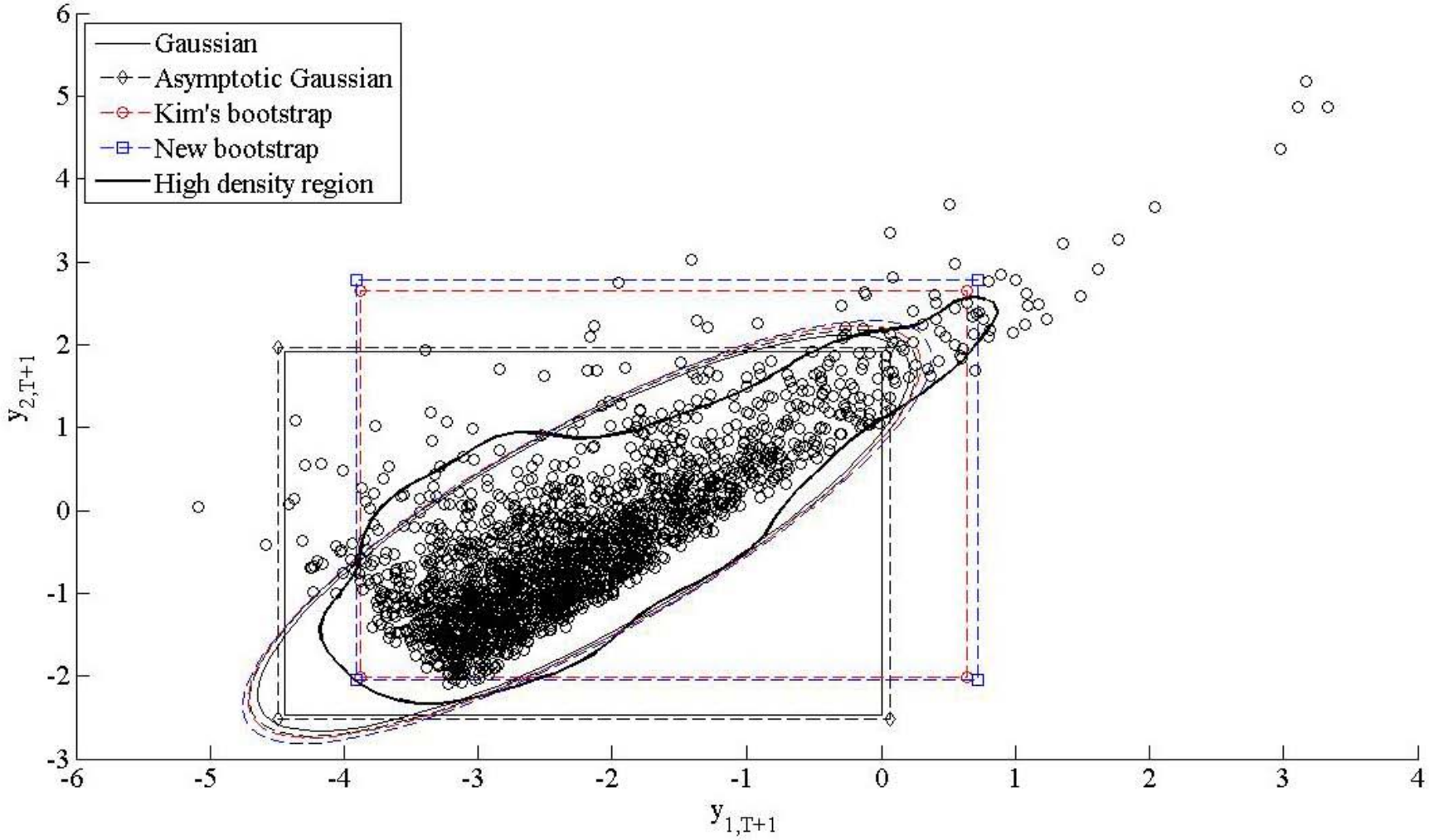


Figure 3: 95% Bonferroni cubes and elliptical regions for one-step-ahead prediction for a bivariate VAR(2) model with $T = 100$ and $\chi^2(4)$ errors.

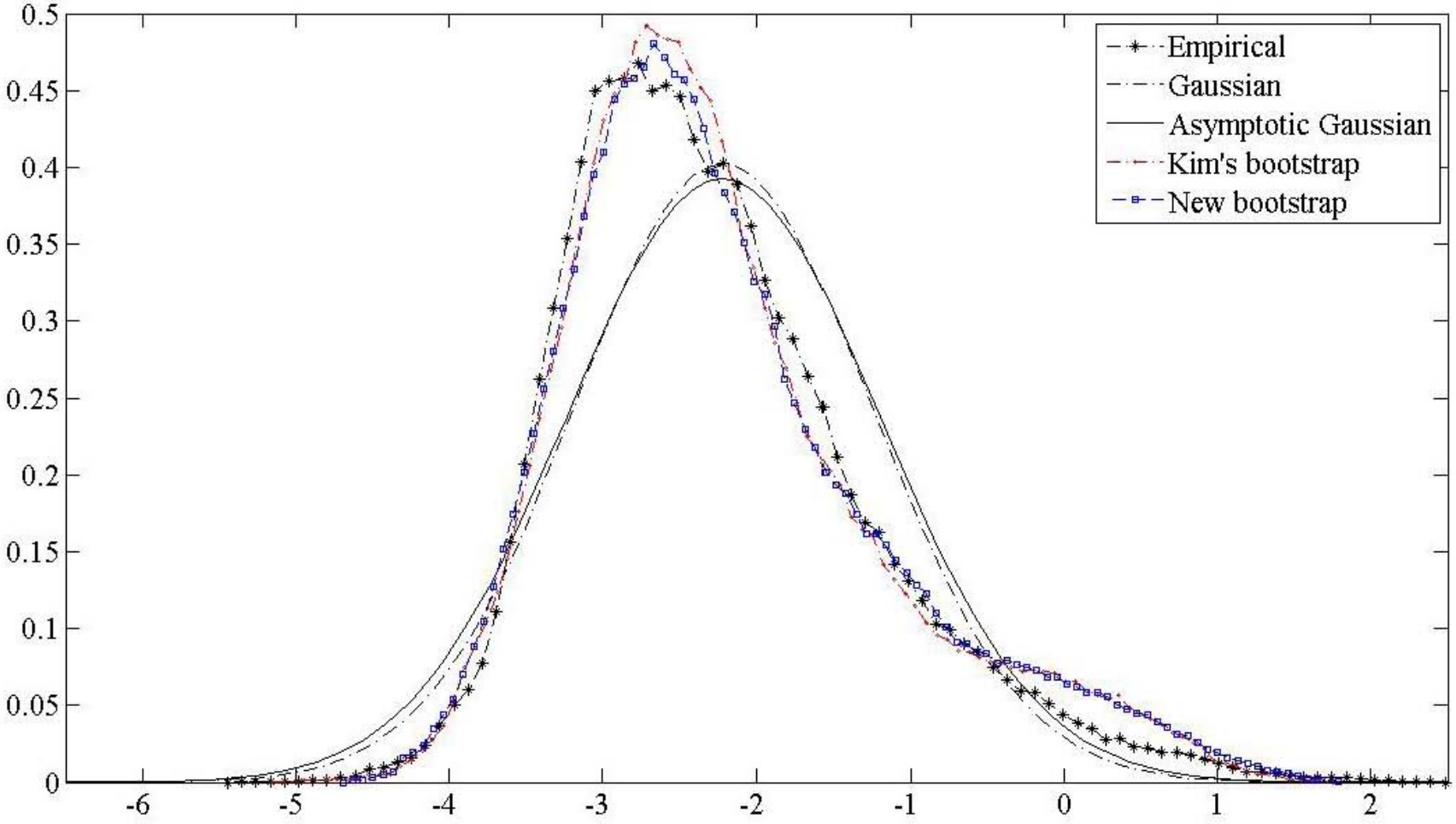


Figure 4: Estimated kernel densities for one-step-ahead predictions of the first variable generated by model Model 1 with $T = 100$ and $\chi^2(4)$ errors.

Table 1: Monte Carlo means and standard deviations (in parenthesis) of coverages of one-step-ahead and eight-steps-ahead prediction intervals for the first variable of a bivariate VAR(1) model with Gaussian, Student-5 and $\chi^2(4)$ errors, constructed using the Gaussian (GI), asymptotic (AI) and bootstrap procedures (KI and BI). Nominal coverage 95%. Samples sizes $T = 25$ and $T = 100$.

		One-step-ahead			Eight-steps-ahead		
		Coverage	Length	Left/Right	Coverage	Length	Left/Right
Gaussian	T=25						
	GI	91.89 (5.91)	3.83 (0.51)	4.04 /4.07	94.08 (4.15)	4.59 (0.72)	2.95 /2.97
	AI	93.40 (5.37)	4.05 (0.54)	3.28 /3.32	94.33 (4.02)	4.64 (0.73)	2.83 /2.85
	KI	92.41 (5.55)	4.03 (0.62)	3.75 /3.84	95.11 (3.92)	4.89 (0.87)	2.43 /2.46
	BI	92.40 (5.64)	4.03 (0.62)	3.75 /3.85	95.23 (3.91)	4.93 (0.85)	2.38 /2.39
	T=100						
	GI	94.48 (1.57)	3.90 (0.18)	2.73 /2.80	94.71 (1.66)	4.51 (0.30)	2.65 /2.64
	AI	94.83 (1.52)	3.96 (0.18)	2.56 /2.62	94.77 (1.65)	4.52 (0.30)	2.62 /2.61
	KI	94.41 (2.01)	3.95 (0.29)	2.80 /2.79	95.05 (1.83)	4.61 (0.36)	2.48 /2.47
	BI	94.37 (1.99)	3.95 (0.28)	2.80 /2.84	95.08 (1.82)	4.61 (0.36)	2.44 /2.48
Student-5	T=25						
	GI	91.28 (7.40)	3.80 (0.68)	4.47 /4.25	93.56 (4.50)	4.58 (0.92)	3.25 /3.19
	AI	92.54 (6.84)	4.02 (0.72)	3.82 /3.63	93.78 (4.38)	4.62 (0.93)	3.13 /3.09
	KI	92.41 (6.24)	4.25 (1.00)	3.92 /3.67	94.74 (4.10)	5.08 (1.29)	2.66 /2.60
	BI	92.28 (6.30)	4.23 (0.98)	3.99 /3.73	94.83 (4.15)	5.10 (1.22)	2.60 /2.57
	T=100						
	GI	94.30 (1.66)	3.89 (0.26)	2.84 /2.86	94.51 (1.75)	4.51 (0.38)	2.73 /2.76
	AI	94.57 (1.60)	3.95 (0.27)	2.70 /2.73	94.56 (1.74)	4.52 (0.38)	2.71 /2.74
	KI	94.54 (2.04)	4.06 (0.44)	2.80 /2.65	95.11 (1.82)	4.72 (0.47)	2.48 /2.41
	BI	94.48 (2.13)	4.05 (0.44)	2.85 /2.68	95.08 (1.86)	4.71 (0.47)	2.48 /2.44
$\chi^2(4)$	T=25						
	GI	92.47 (4.95)	3.79 (0.70)	1.38 /6.15	93.42 (4.56)	4.55 (0.90)	1.94 /4.64
	AI	93.55 (4.42)	4.01 (0.74)	0.99 /5.45	93.62 (4.46)	4.60 (0.91)	1.86 /4.52
	KI	93.10 (5.27)	4.13 (0.96)	2.99 /3.91	94.78 (4.34)	5.04 (1.18)	2.06 /3.16
	BI	93.07 (5.28)	4.11 (0.94)	2.95 /3.99	94.87 (4.36)	5.08 (1.20)	1.97 /3.17
	T=100						
	GI	94.69 (1.59)	3.91 (0.26)	0.52 /4.79	94.52 (1.69)	4.53 (0.37)	1.46 /4.02
	AI	94.90 (1.54)	3.97 (0.26)	0.47 /4.63	94.57 (1.68)	4.55 (0.38)	1.44 /3.99
	KI	94.68 (2.13)	3.98 (0.39)	2.55 /2.77	95.09 (1.97)	4.74 (0.48)	2.35 /2.56
	BI	94.69 (2.05)	3.98 (0.38)	2.54 /2.77	95.07 (2.00)	4.74 (0.49)	2.33 /2.59

Table 2: Monte Carlo means and standard deviations (in parenthesis) of coverages of one-step-ahead and eight-steps-ahead Bonferroni prediction regions for components of a bivariate VAR(1) model with Gaussian, Student-5 and $\chi^2(4)$ errors, constructed using the Gaussian (GC), asymptotic (AC) and bootstrap procedures (KC and BC). Nominal coverage 90%. Samples sizes $T = 25$ and $T = 100$.

		One-step-ahead		Eight-steps-ahead	
		Coverage	Volume	Coverage	Volume
Gaussian	T=25				
	GC	87.67 (7.23)	14.77 (3.51)	85.98 (7.47)	24.59 (6.29)
	AC	89.85 (6.63)	16.55 (3.93)	87.26 (7.16)	26.00 (6.88)
	KC	88.27 (6.80)	16.45 (4.61)	88.27 (6.77)	28.19 (19.55)
	BC	88.26 (6.90)	16.46 (4.57)	88.29 (6.90)	28.04 (7.58)
	T=100				
	GC	91.41 (1.96)	15.21 (1.24)	89.27 (2.38)	24.66 (2.38)
	AC	91.93 (1.90)	15.66 (1.28)	89.60 (2.35)	25.03 (2.44)
	KC	91.22 (2.42)	15.62 (1.87)	90.04 (2.58)	25.86 (2.90)
	BC	91.17 (2.37)	15.57 (1.83)	90.06 (2.58)	25.88 (2.94)
Student-5	T=25				
	GC	87.60 (8.93)	14.76 (5.03)	86.12 (7.99)	24.48 (8.67)
	AC	89.39 (8.24)	16.53 (5.63)	87.19 (7.66)	25.88 (9.31)
	KC	89.08 (7.57)	18.86 (8.95)	88.24 (7.02)	30.27 (27.18)
	BC	88.84 (7.72)	18.69 (8.68)	88.06 (7.41)	29.62 (12.43)
	T=100				
	GC	92.12 (2.14)	15.24 (1.94)	89.92 (2.71)	24.64 (3.35)
	AC	92.50 (2.06)	15.69 (2.00)	90.17 (2.68)	25.00 (3.42)
	KC	92.35 (2.61)	16.71 (3.32)	90.95 (2.85)	27.10 (4.52)
	BC	92.27 (2.67)	16.62 (3.20)	90.92 (2.88)	27.12 (4.53)
$\chi^2(4)$	T=25				
	GC	89.45 (7.26)	14.71 (5.15)	87.28 (7.83)	24.37 (8.89)
	AC	91.01 (6.46)	16.48 (5.77)	88.18 (7.51)	25.74 (9.52)
	KC	89.26 (7.63)	17.61 (8.16)	89.09 (7.11)	28.90 (12.22)
	BC	89.25 (7.62)	17.43 (7.83)	89.15 (7.22)	29.07 (12.09)
	T=100				
	GC	93.18 (1.93)	15.32 (1.86)	90.96 (2.46)	24.78 (3.25)
	AC	93.44 (1.87)	15.78 (1.92)	91.17 (2.42)	25.15 (3.32)
	KC	91.51 (3.49)	15.60 (2.84)	90.78 (2.72)	26.49 (4.08)
	BC	91.49 (3.45)	15.62 (2.76)	90.78 (2.72)	26.45 (4.12)

Table 3: Monte Carlo means and standard deviations (in parenthesis) of coverages of one-step-ahead and eight-steps-ahead elliptical prediction regions for components of a bivariate VAR(1) model with Gaussian, Student-5 and $\chi^2(4)$ errors, constructed using the Gaussian (GE), asymptotic (AE) and bootstrap procedures (KE and BE). Nominal coverage 90%. Sample sizes $T = 25$ and $T = 100$.

		One-step-ahead		Eight-steps-ahead	
		Coverage	Volume	Coverage	Volume
Gaussian	T=25				
	GE	82.86 (8.10)	8.02 (1.55)	85.17 (7.92)	22.34 (5.70)
	AE	85.90 (7.48)	8.98 (1.74)	86.38 (7.59)	23.46 (6.07)
	KE	85.30 (7.27)	8.92 (1.95)	88.15 (7.12)	25.39 (8.77)
	BE	85.11 (7.46)	8.86 (1.86)	88.15 (7.30)	25.55 (6.82)
	T=100				
	GE	88.76 (2.16)	8.53 (0.57)	88.64 (2.53)	22.45 (2.16)
	AE	89.47 (2.09)	8.78 (0.58)	88.96 (2.51)	22.75 (2.21)
	KE	89.07 (2.57)	8.70 (0.63)	89.75 (2.79)	23.61 (2.54)
	BE	89.07 (2.58)	8.71 (0.65)	89.73 (2.84)	23.61 (2.59)
Student-5	T=25				
	GE	82.28 (10.31)	7.83 (2.17)	85.05 (8.30)	22.20 (7.79)
	AE	84.71 (9.62)	8.77 (2.43)	86.03 (7.98)	23.30 (8.23)
	KE	85.69 (7.99)	8.87 (2.86)	88.27 (7.21)	25.67 (12.89)
	BE	85.28 (8.41)	8.71 (2.54)	88.10 (7.63)	25.56 (9.86)
	T=100				
	GE	89.13 (2.34)	8.52 (0.86)	88.97 (2.83)	22.46 (3.02)
	AE	89.62 (2.27)	8.78 (0.89)	89.20 (2.79)	22.76 (3.06)
	KE	89.66 (2.68)	8.72 (0.97)	89.98 (3.15)	23.58 (3.35)
	BE	89.53 (2.74)	8.69 (0.95)	89.94 (3.18)	23.62 (3.45)
$\chi^2(4)$	T=25				
	GE	84.78 (8.33)	7.84 (2.13)	86.65 (8.05)	22.13 (7.99)
	AE	86.77 (7.59)	8.78 (2.38)	87.44 (7.73)	23.22 (8.45)
	KE	86.77 (7.12)	8.81 (2.82)	88.85 (7.30)	25.30 (9.96)
	BE	86.55 (7.22)	8.66 (2.52)	88.84 (7.48)	25.33 (9.49)
	T=100				
	GE	90.21 (1.98)	8.57 (0.82)	90.28 (2.55)	22.57 (2.93)
	AE	90.57 (1.91)	8.83 (0.85)	90.44 (2.52)	22.87 (2.97)
	KE	89.87 (2.35)	8.77 (0.91)	90.24 (2.82)	23.69 (3.28)
	BE	89.80 (2.43)	8.74 (0.90)	90.26 (2.83)	23.69 (3.33)

Table 4: Monte Carlo means of Mallows distances between of one-step-ahead ($h = 1$) and eight-steps-ahead ($h = 8$) marginal and conditional prediction densities and corresponding true densities obtained for a bivariate VAR(1) model with Gaussian, Student-5 and $\chi^2(4)$ errors, constructed using the Gaussian (GD), asymptotic (AD) and bootstrap procedures (KD and BD). Sample sizes $T = 25$ and $T = 100$.

		Marginal		Condition 1 (q_1)		Condition 2 (q_2)		Condition 3 (q_3)	
		h=1	h=8	h=1	h=8	h=1	h=8	h=1	h=8
Gaussian	T=25								
	GD	0.329	0.468	0.226	0.271	0.210	0.251	0.224	0.274
	AD	0.334	0.473	0.224	0.267	0.210	0.252	0.222	0.277
	KD	0.378	0.466	0.308	0.352	0.282	0.327	0.304	0.349
	BD	0.385	0.495	0.308	0.353	0.281	0.332	0.300	0.352
	T=100								
	GD	0.145	0.188	0.113	0.193	0.109	0.173	0.115	0.194
	AD	0.145	0.188	0.115	0.191	0.107	0.171	0.116	0.193
	KD	0.195	0.197	0.208	0.217	0.196	0.198	0.217	0.223
	BD	0.196	0.200	0.208	0.220	0.195	0.200	0.216	0.220
Student-5	T=25								
	GD	1.126	1.253	0.243	0.335	0.237	0.364	0.235	0.334
	AD	1.118	1.245	0.251	0.338	0.250	0.368	0.245	0.336
	KD	1.190	1.408	0.277	0.279	0.249	0.256	0.273	0.280
	BD	1.187	1.382	0.277	0.279	0.247	0.258	0.272	0.282
	T=100								
	GD	0.994	1.143	0.212	0.320	0.210	0.354	0.202	0.317
	AD	0.992	1.139	0.216	0.320	0.215	0.354	0.207	0.319
	KD	0.984	1.041	0.237	0.236	0.198	0.211	0.221	0.231
	BD	0.983	1.040	0.238	0.235	0.197	0.214	0.220	0.234
$\chi^2(4)$	T=25								
	GD	0.583	0.687	0.299	0.466	0.284	0.316	0.363	0.351
	AD	0.579	0.686	0.320	0.471	0.291	0.320	0.358	0.349
	KD	0.561	0.843	0.263	0.308	0.287	0.307	0.402	0.415
	BD	0.558	0.783	0.256	0.316	0.282	0.316	0.397	0.417
	T=100								
	GD	0.468	0.482	0.222	0.391	0.196	0.220	0.248	0.261
	AD	0.467	0.480	0.230	0.391	0.199	0.218	0.246	0.265
	KD	0.365	0.289	0.146	0.153	0.181	0.173	0.274	0.252
	BD	0.364	0.296	0.146	0.152	0.182	0.174	0.275	0.252

Table 5: Descriptive statistics of quarterly US inflation, unemployment and GDP growth observed from 1954Q3 to 2008Q4.

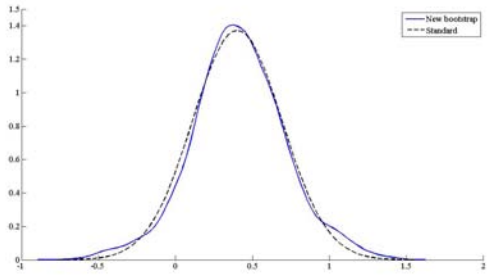
Series	Mean	Sd	Skewness	Kurtosis	Normality test	A. D-F[1]
Inflation	0.87	0.59	1.21*	4.27*	93.65*	-2.35
Unemployment	5.77	1.42	0.80*	3.77**	38.59*	-3.16**
GDP growth	0.80	0.93	-0.33**	4.29*	13.81*	-10.25*
First difference of inflation ($\triangle\pi$)	-0.00	0.34	-0.24	5.01*	27.07*	-12.53*
Joint (with π)[1]			2.69*	19.00*	125.89*	
Joint (with $\triangle\pi$)[1]			1.63*	8.12*	77.20*	

(*) and (**) significant at 1% and 5%, respectively.
[1] The Augmented Dickey-Fuller test critical value is -2.87 at 5%.

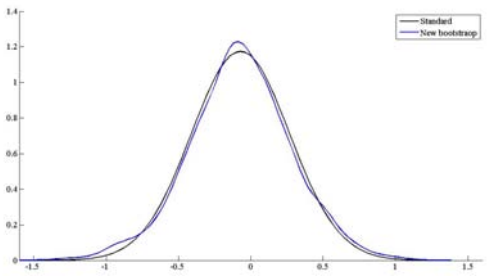
Table 6: Diagnostics of the VAR(3) residuals corresponding to the US first difference of inflation, unemployment and GDP growth series.

Series	Mean	Sd	Skewness	Kurtosis	Normality test
$\hat{a}_{\triangle\pi}$	-0.00	0.28	-0.04	3.71**	5.89**
$\hat{a}_u$	-0.00	0.30	0.99*	5.00*	29.24*
$\hat{a}_g$	0.00	0.82	0.01	4.13*	12.12*
Joint			1.70*	20.22*	35.31*

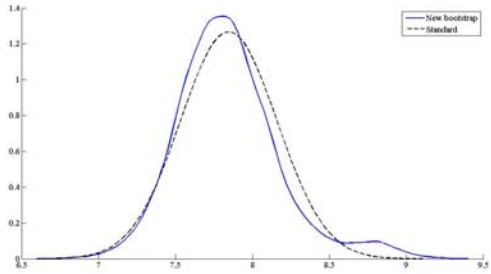
(*) and (**) significant at 1% and 5%, respectively.



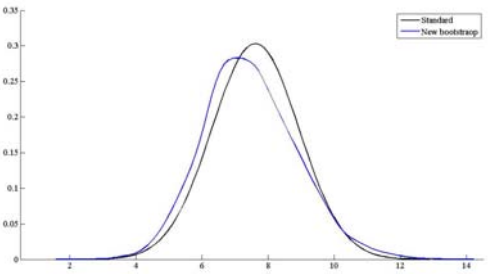
(a) First difference of inflation, $h = 1$



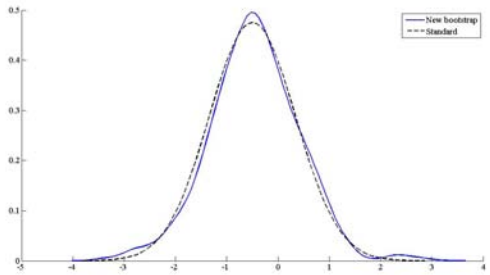
(b) First difference of inflation, $h = 8$



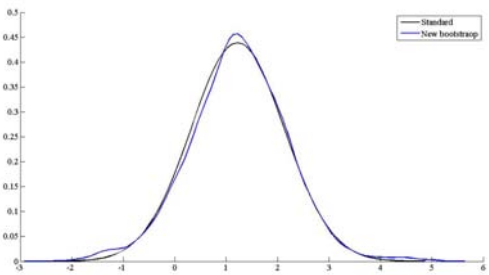
(c) Unemployment, $h = 1$



(d) Unemployment, $h = 8$

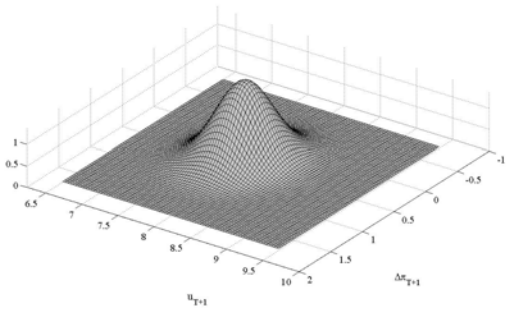


(e) GDP growth, $h = 1$

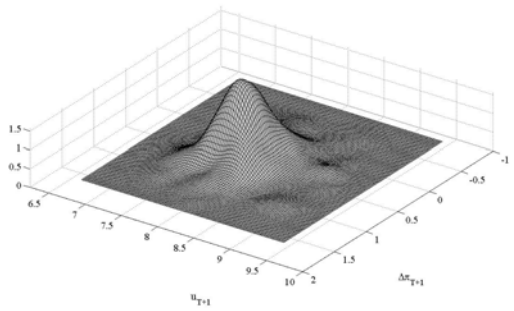


(f) GDP growth, $h = 8$

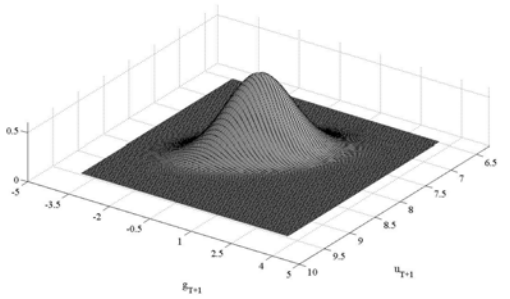
Figure 5: Kernel estimates of the densities of the one-step-ahead and four-steps-ahead predictions for the first difference of (a-b) inflation, (c-d) unemployment and (e-f) GDP growth.



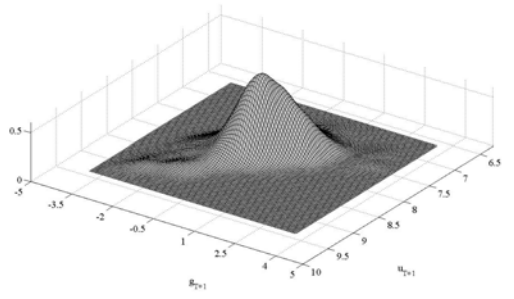
(a) Gaussian △Inflation-Unemployment



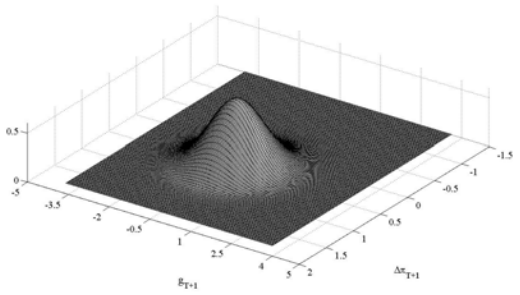
(b) New bootstrap △Inflation-Unemployment



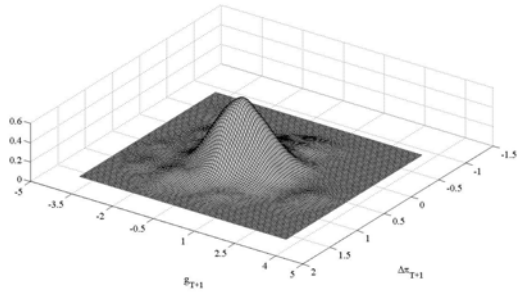
(c) Gaussian Unemployment-GDP growth



(d) New bootstrap Unemployment-GDP growth

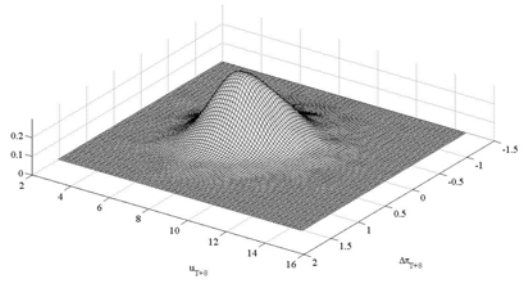


(e) Gaussian △Inflation-GDP growth

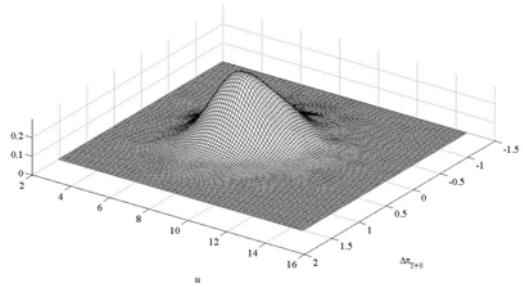


(f) New bootstrap △Inflation-GDP growth

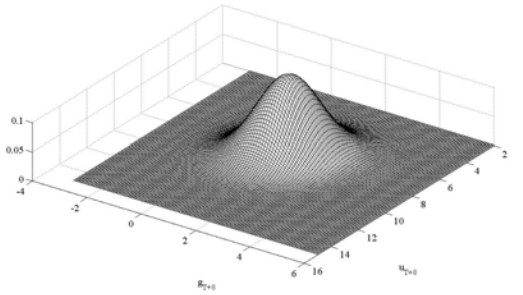
Figure 6: Kernel estimates of the joint densities of the one-step-ahead predictions for a bivariate US series.



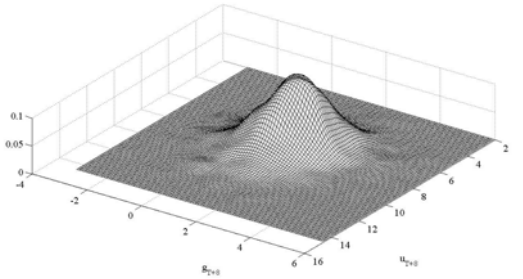
(a) Gaussian △Inflation-Unemployment



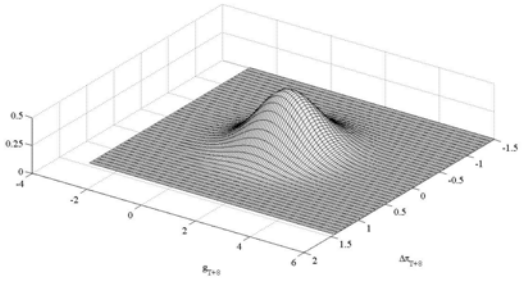
(b) New bootstrap △Inflation-Unemployment



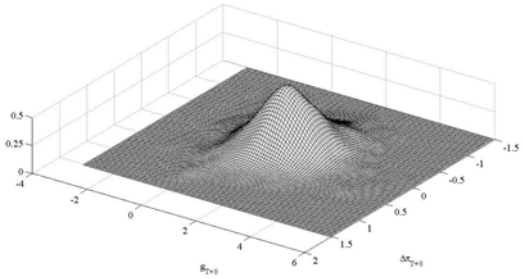
(c) Gaussian Unemployment-GDP growth



(d) New bootstrap Unemployment-GDP growth

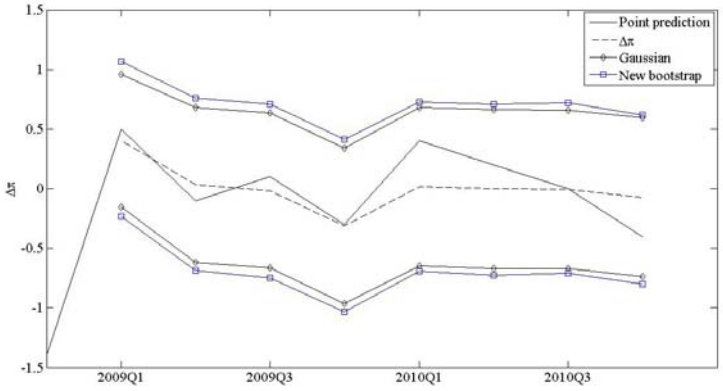


(e) Gaussian △Inflation-GDP growth

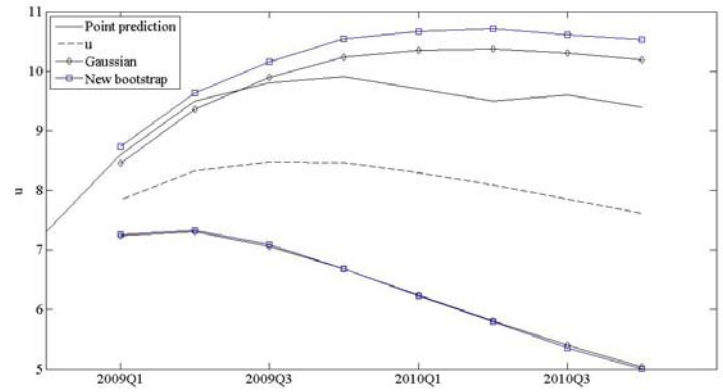


(f) New bootstrap △Inflation-GDP growth

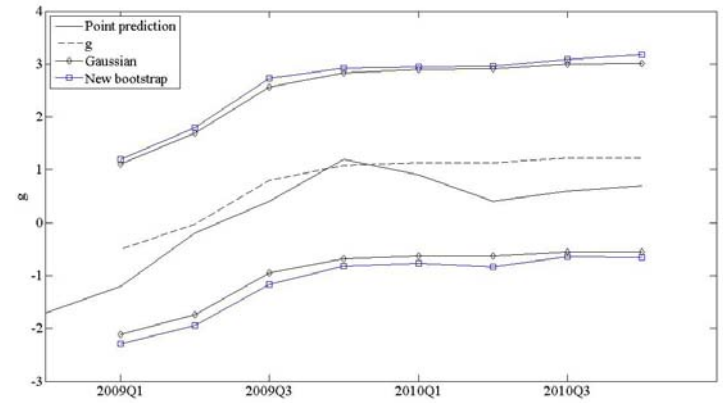
Figure 7: Kernel estimates of the joint densities of the eight-steps-ahead predictions for a bivariate US series.



(a) First difference of inflation

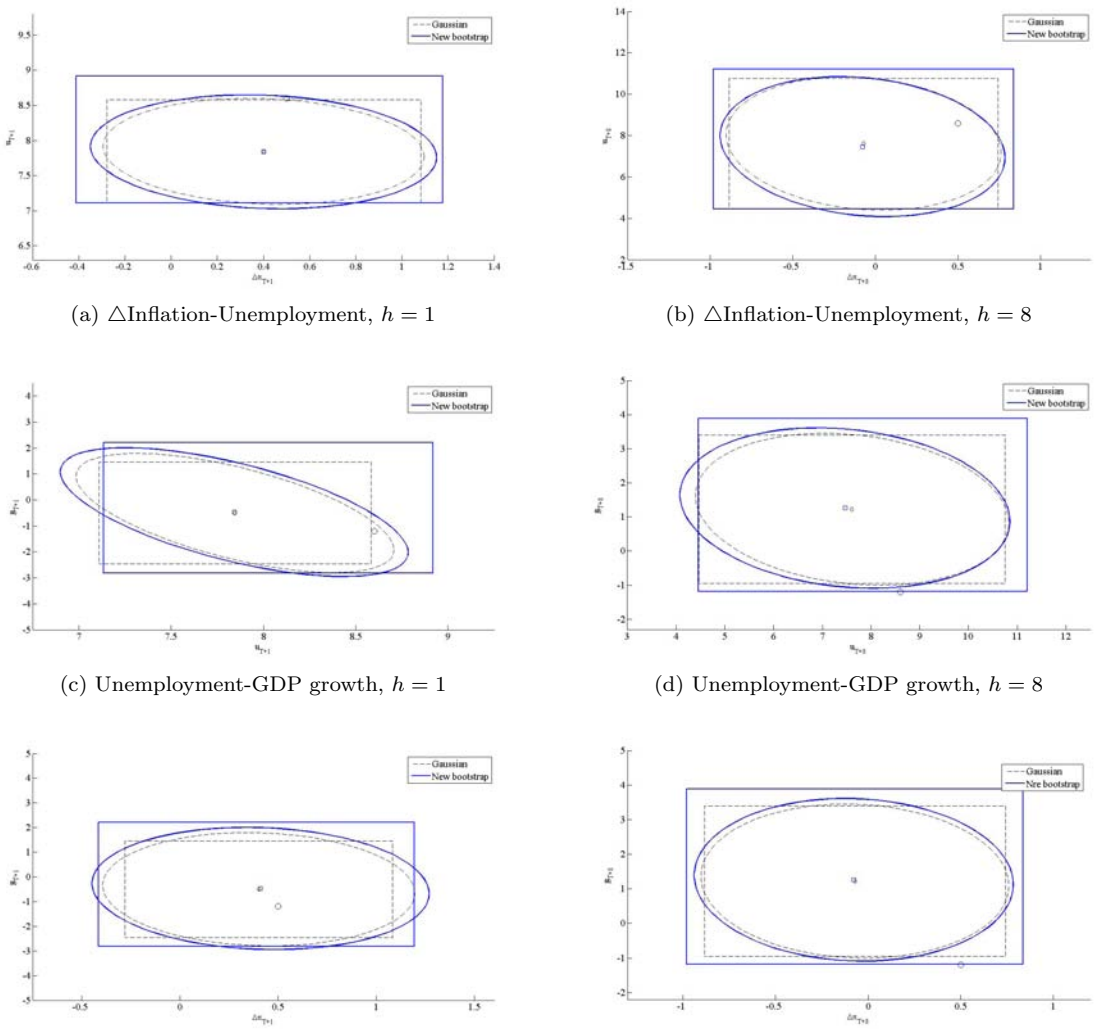


(b) Unemployment



(c) GDP growth

Figure 8: 95% prediction intervals for US quarterly (a) inflation, (b) unemployment and (c) GDP growth from 2009Q1 to 2010Q4.



(a) △Inflation-Unemployment, $h = 1$

(b) △Inflation-Unemployment, $h = 8$

(c) Unemployment-GDP growth, $h = 1$

(d) Unemployment-GDP growth, $h = 8$

(e) △Inflation-GDP growth, $h = 1$

(f) △Inflation-GDP growth, $h = 8$

Figure 9: 95% Bonferroni cubes and elliptical regions for the one-step-ahead and eight-steps-ahead predictions for (a-b) △Inflation-Unemployment, (c-d) Unemployment-GDP growth and (e-f) △Inflation-GDP growth.

Table 7: Mean coverages and volumes of one-step-ahead Bonferroni and elliptical prediction regions constructed using the Gaussian (GC), asymptotic (AC) and bootstrap procedures (KC and BC) for a VAR(3) model for US inflation, unemployment and GDP growth. Nominal coverages 90%, 95% and 99%.

	Bonferroni		Elliptical	
	Coverage	Volume	Coverage	Volume
90%				
Gaussian	71.26	3.75	67.66	2.43
Asymptotic	75.45	4.93	71.26	3.19
Kim's bootstrap	82.63	7.07	77.84	4.45
New bootstrap	83.23	6.78	77.84	4.19
95%				
Gaussian	79.04	5.33	71.86	3.39
Asymptotic	81.44	7.01	79.04	4.46
Kim's bootstrap	86.83	9.70	86.23	6.22
New bootstrap	86.23	9.27	86.83	5.86
99%				
Gaussian	88.02	9.83	85.03	5.94
Asymptotic	89.22	12.92	89.22	7.80
Kim's bootstrap	92.22	15.85	94.01	10.88
New bootstrap	92.22	15.22	92.81	10.25