Bootstrap prediction regions for multivariate autoregressive processes

Article *in* Statistical Methods and Applications · November 2005

DOI: 10.1007/s10260-005-0113-y · Source: DBLP

1 author:

Matteo Grigoletto
University of Padova
25 PUBLICATIONS **187** CITATIONS

SEE PROFILE

Bootstrap Prediction Regions for Multivariate Autoregressive Processes

Regioni di previsione bootstrap *per processi autoregressivi multidimensionali*

Matteo Grigoletto
Dipartimento di Scienze Statistiche
Università degli Studi di Padova
matteo.grigoletto@unipd.it

Riassunto: L'obiettivo del presente lavoro è studiare il comportamento di una nuova procedura per la determinazione di regioni di previsione per processi autoregressivi multidimensionali. Le regioni di previsione, basate sulla tecnica *bootstrap*, non fanno affidamento su alcuna assunzione distributiva per i disturbi ed inoltre tengono conto della variabilità derivante dalla necessità di stimare i parametri e l'ordine del processo sottostante alle osservazioni. Verrà mostrato che l'approccio proposto permette, rispetto a quelli già presenti in letteratura, di costruire migliori stime della distribuzione previsiva, nel senso che la variabilità degli stimatori dei suoi quantili viene ridotta.

Keywords: vector autoregressive (VAR) processes, prediction, bootstrap.

1. Introduction

The need to study the relationships among several time series may arise in many disciplines. The forecasting perspective is particularly important in Economics and Finance; in these fields multivariate time series are given, e.g., by several stock indexes, the returns of different securities or the interest rates for loans made on various time horizons. The book by Lütkepohl (1993) contains relevant examples.

Among the models that may be used to represent the process underlying a multiple time series, the vector autoregressive (VAR) models have become standard instruments since their introduction in a seminal paper by Sims (1980).

In the present paper we will assess a method, based on the bootstrap technique, that may be applied in order to compute prediction regions for VAR processes. In particular, we will show how the results presented in the study by Kim (1999) can be improved (in the sense that better estimators of the forecasting distribution are obtained) by taking into account a particular property of the prediction error that Kim fails to consider. This property has been previously exploited, in the univariate case, by Grigoletto (1998).

The procedure presented here has the goal of taking into account not only the variability originating from the innovations, but also that implicit in model identification and parameter estimation (Kim, 1999, does not consider model identification). Besides, no parametric assumptions are made about the innovations distribution. Methods having the same objectives were presented, in the univariate case, by Masarotto (1990) and Thombs and Schucany (1990). These procedures all rely on the bootstrap.

Prediction regions may also be based on the asymptotic method presented in section 3.5 of Lütkepohl (1993). This method, that in the following will be used as a term of comparison, considers the variability due to parameter estimation (but not that caused by model identification). As we will see, for small sample sizes the asymptotic procedure can have an unsatisfactory performance. Also, the asymptotic approach is based on the

sometimes unrealistic assumption of Gaussian innovations.

In section 2 we will outline the method used to build the prediction regions. Section 3 contains some simulation results.

2. Framework and Methods

We assume that the data, represented by a k-dimensional time series $\{y_t, t = 1, \ldots, n\}$, are generated according to a VAR(p) model of the form

$$y_t = \nu + A_1\, y_{t-1} + A_2\, y_{t-2} + \ldots + A_p\, y_{t-p} + u_t\ , \tag{1}$$

where $\{u_t\}$ is a zero mean independent white noise process with nonsingular, time invariant covariance matrix Σ_u, ν is a $(k \times 1)$ intercept term and the A_i, $i = 1, \ldots, p$, are $(k \times k)$ coefficient matrices. Because of how it is defined, process (1) can be applied in a natural way for forecasting purposes. This process is stable if $\det(I_k - A_1\, z - \ldots - A_p z^p) \neq 0,\ \text{for}\, |z| \leq 1$, where I_k is the $(k \times k)$ identity matrix. In this case the process admits a Wold moving average representation $y_t = \mu + u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \ldots$, which in turn implies stationarity. The results in this paper will be presented for stable VAR processes, but are trivially extended to situations in which nonstationarity can be removed by differencing.

Our objective is to find a prediction region for y_{n+h}, $h \geq 1$, on the basis of the observed time series $\{y_t, t = 1, \ldots, n\}$. It will be useful to first show how prediction regions may be based on asymptotic considerations. In fact, the asymptotic results rely on a prediction error decomposition that is also central to our bootstrap procedure.

When model parameters are assumed known, the optimal (in the sense of minimizing the mean squared prediction error) predictor for forecast horizon h at forecast origin t is defined recursively by $y_t(h) = \nu + A_1\, y_t(h-1) + \ldots + A_p\, y_t(h-p)$, where $y_t(j) = y_{t+j}$ for $j \leq 0$. By replacing the unknown intercept and coefficient matrices with their estimators $\hat{\nu}$ and $\{\hat{A}_i\}$ (notice, however, that the model order p is still assumed to be known), we get a forecast $\hat{y}_t(h) = \hat{\nu} + \hat{A}_1\, \hat{y}_t(h-1) + \ldots + \hat{A}_p\, \hat{y}_t(h-p)$, where $\hat{y}_t(j) = y_{t+j}$ for $j \leq 0$. In the results discussed below, estimation is based on the least squares method.

Let us now consider the following prediction error decomposition:

$$y_{t+h} - \hat{y}_t(h) \;\; = \;\; [y_{t+h} - y_t(h)] + [y_t(h) - \hat{y}_t(h)] = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} + [y_t(h) - \hat{y}_t(h)]\ , \tag{2}$$

where matrices $\{\Phi_i, i \geq 1\}$ have been defined implicitly in the Wold representation of the autoregressive process, and $\Phi_0 = I_k$. If the process $\{y_t\}$ is Gaussian, we have $\sqrt{n}\,[y_t(h) - \hat{y}_t(h)] \xrightarrow{d} \mathcal{N}(0, \Omega(h))$, for an appropriate matrix $\Omega(h)$. Setting $\Sigma_y(h) = \sum_{i=0}^{h-1} \Phi_i\, \Sigma_u\, \Phi_i'$ and $\Sigma_{\hat{y}}(h) = \Sigma_y(h) + \Omega(h)/n$, we then have $[y_{t+h} - \hat{y}_t(h)]'\, \Sigma_{\hat{y}}^{-1}(h)\, [y_{t+h} - \hat{y}_t(h)] \xrightarrow{d} \chi^2_k$, under the normality assumption. This result also holds when replacing $\Sigma_{\hat{y}}(h)$ by a consistent estimator $\hat{\Sigma}_{\hat{y}}(h)$ (here the least squares estimator will be adopted) and we are therefore able to define the prediction region of level $1 - \alpha$ that, following Kim (1999), will be called *asymptotic prediction ellipsoid* (APE):

$$\text{APE}_{1-\alpha} = \left\{ y_{t+h} \,:\, [y_{t+h} - \hat{y}_t(h)]'\, \hat{\Sigma}_{\hat{y}}^{-1}(h)\, [y_{t+h} - \hat{y}_t(h)] \leq \chi^2_{k;1-\alpha} \right\}\ , \tag{3}$$

where $\chi^2_{k;1-\alpha}$ is the $1 - \alpha$ quantile of the χ^2_k distribution, with $0 < \alpha < 1$.

When k is greater than two or three, constructing a prediction ellipsoid can require an appreciable computational effort. It may then be more practical to base a joint prediction region on the Bonferroni's method (see section 2.2.3 in Lütkepohl, 1993). The prediction region in this case is an *asymptotic prediction cube* (APC) of level at least $1-\alpha$ defined by

$$\text{APC}_{1-\alpha} = \{y_{n+h} \,:\, -z_\tau \leq [y_{j,t+h} - \hat{y}_{j,t}(h)]/\hat{\sigma}_j(h) \leq z_\tau;\; j = 1, \ldots, k\} \;, \qquad (4)$$

where $y_{j,t+h}$ and $\hat{y}_{j,t}(h)$, $j = 1, \ldots, k$, are the j-th coordinates of vectors y_{t+h} and $\hat{y}_t(h)$, respectively, $\hat{\sigma}_j(h)$ is the square root of the j-th diagonal element of $\hat{\Sigma}_{\hat{y}}(h)$, $\tau = \alpha/(2\,k)$ and z_τ is such that $P(Z > z_\tau) = \tau$ when $Z \sim \mathcal{N}(0, 1)$.

The small sample properties of the asymptotic procedure described above are unknown. Besides, the asymptotic method assumes the process is Gaussian and it does not consider the variability due to model identification. These problems can be overcome by resorting to the bootstrap technique.

Our approach to the construction of multidimensional bootstrap prediction regions for autoregressive models is an extension of the method by Kim (1999), who adapted to the multivariate case the results by Thombs and Schucany (1990) and Masarotto (1990).

We should first remark that, following Thombs and Schucany (1990), the bootstrap pseudo-series are generated conditionally on the last p values of the observed series. This is because we want to study the future with respect to the past we have observed. However, for a VAR(p) model, conditioning on the past amounts to fixing the last p observed values. In order to generate conditional pseudo-series we need to consider the backward VAR model $y_t = \delta + H_1\, y_{t+1} + \ldots + H_p\, y_{t+p} + v_t$, where the various terms have a meaning analogous to those in the corresponding forward model (1).

Once estimates $(\hat{\delta}, \hat{H}_1, \ldots, \hat{H}_p)$ for $(\delta, H_1, \ldots, H_p)$ have been obtained (here, in particular, the least squares estimates will be adopted), the pseudo-time series $\{y_t^*, t = 1, \ldots, n\}$ can be generated recursively according to $y_t^* = \hat{\delta} + \hat{H}_1\, y_{t+1}^* + \ldots + \hat{H}_p\, y_{t+p}^* + v_t^*$, where the distribution of v_t^* is the empirical distribution of the backward least squares residuals $\{\hat{v}_t\}$. The values $y_{n-p+1}^*, \ldots, y_n^*$ needed to start the recursion are set equal to the last p observed values $y_{n-p+1}, \ldots, y_n$.

The bootstrap procedure relies on the fact that, within the bootstrap framework, the prediction errors are known (this, of course, is not true in the real world). In order to compute them, we need to generate pseudo-future values and this is achieved on the basis of the recursion $y_t^* = \hat{\nu} + \hat{A}_1\, y_{t-1}^* + \hat{A}_2\, y_{t-2}^* + \ldots + \hat{A}_p\, y_{t-p}^* + u_t^*$. Here, as in the backward generation, the starting values are the last p observations. Bootstrap innovations $\{u_t^*\}$ are distributed according to the empirical distribution of the forward least squares residuals $\{\hat{u}_t\}$.

The method proceeds by generating B pseudo-series and then computing the bootstrap prediction ellipsoids and bootstrap prediction cubes, that will be based on expressions analogous to (3) and (4), respectively, with all the involved quantities replaced by their bootstrap counterparts. This also means that the quantiles of the chi-squared and standard Gaussian distributions have to be replaced by the relevant bootstrap quantiles.

Now, the results obtained with the procedure just described can be substantially improved (in the sense that better estimates of the predictive distribution are found) by generalizing to the multivariate case the method described in Grigoletto (1998). This method is founded on a prediction error decomposition analogous to that defined by equation (2), applied to the bootstrap setting. Due to space reasons, this procedure cannot be described in greater detail here. However, an example of the improvement it leads to is outlined in the next section.

3. Simulation Results

The results discussed in this section are based on a bi-dimensional VAR(1) model with zero intercept and coefficient matrix defined by $\text{vec}(A_1) = (0.5, -0.6, 0.3, 1.3)'$, where vec denotes the column stacking operator. This matrix is interesting since it im-

plies near-nonstationarity: the characteristic roots have moduli $(1.1, 1.1)$. Innovations were generated from a Gaussian distribution with variance-covariance matrix Σ_u such that $\text{vec}(\Sigma_u) = (1, 0.5, 0.5, 1)'$. The results refer to a sample size $n = 80$ and a prediction horizon $h = 1$. The order selection procedure is based on the Akaike's information criterion, and the maximum order tried is 4.

Estimation of the predictive distribution was based on 999 bootstrap resamplings. The nominal level $1 - \alpha$ considered for the prediction regions is 0.95. In order to compute the observed coverages, means and variances 500 Monte Carlo iterations were performed.

Concerning the acronyms APE, APC, BPE and BPC used in Table 1, PE and PC indicate prediction ellipsoids and cubes, while the leading A and B refer to asymptotic and bootstrap procedures. Subscripts 1 and 2 are used for regions based on the method by Kim and on the procedure proposed here, respectively.

At the present sample size $n = 80$, the observed coverages for the asymptotic prediction regions are below the nominal level (0.926 both for the APE and APC). Bootstrap prediction ellipsoids and cubes yield coverages generally closer to the nominal ones (0.942, 0.944, 0.926 and 0.930 for BPE_1, BPE_2, BPC_1 and BPC_2, respectively).

While from the point of view of observed coverages there are not significant differences between the procedure by Kim and the one considered here, important advantages brought by the use of the proposed method become apparent when considering the variability of the predictive distribution estimated quantiles. Table 1 displays the observed means and variances of the (scalar) quantiles that define the prediction ellipsoids. For the prediction cubes, the observed bi-dimensional means of the lower and upper quantiles (denoted by l.c. and u.c., respectively) are shown, together with the observed determinants of the variance-covariance matrices of the quantile estimators. These determinants give a measure of variability in the sense that they are proportional to the square of the volume of the concentration ellipsoids for the quantile estimators.

As Table 1 shows, the method proposed in the present paper leads always to a (often strong) reduction in the variability of the predictive distribution quantile estimators, yielding therefore more reliable prediction regions.

Table 1: *Average and variability behaviour of the quantiles that define the prediction regions.*

BPE_1		BPE_2		BPC_1			BPC_2	
Mean	Variance	Mean	Variance	Mean		Σ	Mean	Σ
6.894	0.452	6.932	0.353	(l.q.)	$(-2.413, -2.440)'$	0.956	$(-2.383, -2.408)'$	0.729
				(u.q.)	$(2.372, 2.366)'$	0.899	$(2.380, 2.370)'$	0.730

References

Grigoletto M. (1998) Bootstrap prediction intervals for autoregression: some alternatives, *International Journal of Forecasting*, 14, 447–456.

Kim J.H. (1999) Asymptotic and bootstrap prediction regions for vector autoregression, *International Journal of Forecasting*, 15, 393–403.

Lütkepohl H. (1993) *Introduction to Multiple Time Series Analysis*, Springer Verlag, New York.

Masarotto G. (1990) Bootstrap prediction intervals for autoregressions, *International Journal of Forecasting*, 6, 229–239.

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

Thombs L.A. and Schucany W.R. (1990) Bootstrap prediction intervals for autoregression, *Journal of the American Statistical Association*, 85, 486–492.