Research Report
Statistical Research Unit
Göteborg University
Sweden

A comparison of conditioned versus unconditioned forecasts of the VAR(1) process

Thomas Holgersson
Fredrik Lindström

Research Report 2003:5
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Statistical Research Unit	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se/stat
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg			
Sweden			

A comparison of conditioned versus unconditioned forecasts of the VAR(1) process.

By

H. E. T. Holgersson **and** **F. Lindström**

Department of Statistics
Göteborg University
SE-405 30 Göteborg
Sweden

Department of Statistics
Göteborg University
SE-405 30 Göteborg
Sweden

Abstract: The properties of a forecast usually depend upon whether the forecast is conditioned on the final period observation or not. In the case of unconditioned forecasts it is well known that the point predictions are unbiased. If on the other hand the forecast is conditional, then the forecast may be biased. Existing analytical results in literature are insufficient for describing the properties of the conditioned forecast properly, particularly in multivariate models. This paper examines some finite sample properties of conditioned forecasts of the VAR(1) process by means of Monte Carlo experiments. We use a number of parameter settings for the VAR(1) process to demonstrate that the forecast bias of the conditioned forecast may be considerable. Hence, unless the analyst has a clear idea of whether the conditioned or unconditioned forecast is relevant for the time series being analysed, statistical inferences may be seriously erratic.

Keywords: Forecast bias, conditioning, VAR(1) process, Monte Carlo.
JEL Classification: C 32.

I. Introduction

The topic of time series analysis has been an area of intense and active research during the last decades. Developments have been proceeding at a fast pace. Much of this research has strived to develop techniques capable of analysing more and more complicated processes. However, the theoretical basis on which inferences are formed in time series analyses seem to have been less cared less for. Consequently, erratic conclusions may be drawn in the most simple time series models, even when all assumptions imposed hold.

One of the main objectives of time series analysis is that of forecasting. In fact, it is often the *sole* objective of the analysis of a time series. It is therefore important that forecasts are formed on a sound inferential basis and that they provide, in some sense, good forecasts. A wide variety of different forecasting procedures are available through the literature and it is important to realize that no single method is universally optimal. Rather, the analyst must choose the procedure which is most appropriate for a given set of conditions. The properties of a forecast usually depends upon if the forecast is conditioned on the final period observation or not. In the case of unconditioned forecasts it is well known that the point predictions are unbiased, see e.g. Dufour (1985), Malinvaud (1970). If on the other hand the forecast is conditional, then the forecast bias may be substantial (Phillips (1979)). The question then arises whether one should consider the forecast to be conditioned or unconditioned on the last observed period. Quoting Phillips (1979), "*it is the case of conditional forecast that is of greatest interest since, in practice, we do forecast with given final values... . But in the evaluation of the success of the forecasting procedure, on average we might be interested in looking at the unconditional distribution.*" In other words, the question of whether the forecast is conditioned or not should depend upon whether the forecast is performed using a single sample or consecutive samples from the population being analysed. If one decides to consider the forecast to be conditioned, one should take into account that the forecast may well be biased. Further on, this bias will carry over to uncertainty measures such as confidence intervals or hypothesis tests of the forecast. In addition, these effects are expected to be accentuated in high dimensional data.

When it comes to analytical results, the literature on forecasting is incomplete. In a general multivariate time series, one can at the most expect to find analytical results for the first order bias approximation of the parameter estimates (Nicholls and Pope (1988)). (An exception though is the case when the process have started at a known time point, rather than in the infinite past as is the most common case, see e.g. Abadir and Larsson (1996). To describe analytically the properties of a conditioned forecast as compared to its unconditioned counterpart will then be very complicated.

The purpose of this study is to demonstrate that point and interval forecasts can be quite different depending on whether conditional or unconditional forecasts are used. We stress that, unless the analyst has a clear idea of whether the conditioned or unconditioned forecast is relevant for the time series being analysed, statistical inferences may be seriously erratic.

The paper is arranged as follows: In the next section, we present the model of concern in this paper along with some basic assumptions of the data generating process. Section III discusses some relevant measures associated with Vector Auto Regressive process of order one (i.e. the VAR(1) process) and presents expressions showing the difference between the conditioned and unconditioned forecasts. In Section IV we present the Monte Carlo experiment and discuss some scalar-valued measures that are relevant for describing the performance of the conditioned vs unconditioned VAR(1) forecasts. The results of the simulations are presented in Section V. Finally, in Section VI we supply a summary of the main results of the paper.

II. Model specification and assumptions

The VAR(p) model consists of a system of several equations where each of the different dependent variables is dependent on the earlier outcomes and the other endogenous variables. The series is modelled by including p lags of each dependent variable in all or some of the different equations. This makes it possible to catch the dynamics of the system but it also affects the predictions of future values. Since the influential work of Sims (1980) VAR modelling has been used in a broad variety of fields. For example; Edlund and Karlsson (1993) used seasonal dummies to forecast the level of Swedish unemployment, Stergiou and Christou (1996) forecasted the catches of the fisheries and found the predator prey dynamics between the anchovy and sardine. Peiris and McNicol (1996) modelled the daily weather, while Baccala and Sameshima (2001) modelled the neural structure in the brain. These studies are examples of the broad area of applications of the VAR process.

In this paper we focus on first order models without intercept term. This model will be written as

$$Y_t = \mathbf{A}Y_{t-1} + U_t \qquad \text{t} = 0,\ 1,\ 2, \ldots, \tag{2.1}$$

where $Y_t = \left(Y_{1t},\ \mathrm{Y}_{2t} \ldots Y_{Kt}\right)'$ is a random $(K \times 1)$ vector, $\mathbf{A}$ is a $(K \times K)$ coefficient matrix and $U_t = \left(U_{1t},\ \mathrm{U}_{2t} \ldots U_{Kt}\right)'$ is a $(K \times 1)$ white noise vector. K denotes the number of equations in the system and p denotes the number of lags in the equation. The covariance of the error terms will be written as $E\left(U_t U_t'\right) = \Sigma_u$.

The process above can be assumed to have started in an infinite past, i.e. in $t = -\infty$. Further, it will be assumed that the eigenvalues of $\mathbf{A}$ have an absolute value less than one to ensure stationarity of the process. The first order moment of the process is $E\left(Y_t\right) = \mathbf{0}$, since the intercept is assumed to equal zero. The auto covariance over a h-step lag will be defined as $\mathbf{\Gamma}_y(h) := E\left(Y_t Y_{t+h}'\right)$. Moreover, we will assume that the noise vector is Gaussian, i.e. $U_t \overset{iid}{\sim} N_p(\mathbf{0}, \mathbf{\Sigma}_{\mathrm{U}})$. We shall also discriminate between random variables and their realised fixed counterparts by denoting the former by capital letters in order to avoid possible confusions. Hence the realisation of $\{Y_t, \mathrm{t} = 0, 1, 2, \ldots,\}$ will be written as $\{y_t, \mathrm{t} = 0, 1, 2, \ldots,\}$.

The possibly oversimplified model (2.1) will provide us with a tool to examine some relevant properties of estimated VAR(1) processes under ideal conditions. The results of our simulations of Section IV will of course be altered for more complicated processes, such as VARMA processes. However, the main difference between conditioned and unconditioned forecasts discussed in Section III will carry over to any data driven forecast. Hence we expect our simulation results (Section IV) to be most favourable, in the sense that more complicated models with for example heteroscedasticity or misspecifications, will perform even worse as compared to our settings.

The next section presents the most popular methods of point and interval estimation of the VAR(1) process as presented in the standard time series literature e.g. Lutkepohl (1993). In particular, these are the methods used in most statistical packages capable of handling multivariate time series, such as SAS or EViews. Therefore the results of this paper should be valuable to the applied forecaster.

III. Point- and interval predictions.

In this section we discuss some details concerning important properties regarding point and interval forecasts and some measures associated to them. When predicting the future this can be done conditional or unconditional on the last component of the time series (i.e. Y_T, used to form the forecast). Even though the realised forecasts are numerically identical for the conditioned and unconditioned forecasts, their statistical properties will usually differ. In other words, the two different ways of looking at Y_T represent different inferential points of view and may give different properties to the forecasts produced.

In Section I, we discussed the importance of distinguishing between conditional and unconditional forecasts. In this section we present some simple expressions that illustrate the difference between forecasts with respect to conditioned/unconditioned forecasts.

When predicting future outcomes of a time series, it is quite common to use point forecasts solely, but they do not say anything about the uncertainty of the forecast. The uncertainty can be included by constructing a confidence interval for the true unknown outcome (i.e. a forecast interval). For a thorough discussion on the topic of calculating interval forecast see Chatfield (1993). It seems that very little attention has been given to the actual properties of intervals based on unknown estimated parameters, which is the most usual situation. This is particularly apparent for conditioned forecasts where, to our knowledge, no results are available. It will therefore be of great relevance to examine the properties of interval estimates as well as point predictions.

When evaluating the forecasts of a time series, this may be done in several ways. Since we are interested in point- as well as interval forecasts we have chosen to evaluate the properties of the following measures. Firstly, we wish to quantify the precision of the point forecast. A convenient measure for this is the frequently applied Mean Square Error of Prediction (MSEP). The expected value of the forecast error is also of concern, since this measure reflects possible forecast bias. Hence, these two measures will be included in the study. Secondly, the property of the interval estimates needs to be quantified.

Interval predictions can be performed as intervals for each marginal process. These are usually obtained by some multiple inference procedure, such as the Bonferroni correction Lutkepohl (1993). In this case, the coverage rate is a relevant measure. In this section we will describe the different measures discussed above along with some of their important properties.

Before we look at the forecast expressions, and in order to avoid confusion, it is important to distinguish between fixed and stochastic variables. Suppose that we wish to forecast Y_{T+h}, where the integer h is the lead time (i.e. the forecasting horizon). The point forecast of Y_{T+h} made conditional on data up to time T for h steps ahead will be denoted by $\hat{Y}_T(h)$ when regarded as a random variable and by $\hat{y}_T(h)$ when it is a particular value determined by the observed data.

Following Lutkepohl (1993), the linear minimum MSE point predictor of the VAR(1) process Y_{t+h} is $Y_t(h) = \mathbf{A}^h Y_t = \mathbf{A} Y_t(h-1)$. Hence the forecast error of the h-step optimal forecast is

$$Y_t(h) - Y_{t+h} = \sum_{i=0}^{h-1} \mathbf{A}^i U_{t+h-i}, \tag{3.1}$$

with its corresponding mean square error covariance matrix

$$\Sigma_Y(h) := MSE\left[Y_t(h)\right] = E\left(\sum_{i=0}^{h-1} \mathbf{A}^i U_{t+h+i}\right)\left(\sum_{i=0}^{h-1} \mathbf{A}^i U_{t+h+i}\right)' = \sum_{i=0}^{h-1} \mathbf{A}^i \Sigma_U \left(\mathbf{A}^i\right)' \tag{3.2}$$

If $U_t \sim N(0, \Sigma_U)$, we have $\{Y_{t+h} - Y_t(h)\} \sim N(0, \Sigma_Y(h))$ and it follows that

$$\frac{Y_{k,t+h} - Y_{k,t}(h)}{\sigma_k(h)} \sim N(0,1), \tag{3.3}$$

where $Y_{k,t+h}$ is the k-th component of Y_{t+h} and $\sigma_k(h)$ is the square root of the k-th diagonal element of $\Sigma_Y(h)$. The expression (3.3) is useful when inferences are to be formed (e.g. hypothesis tests or confidence intervals) on one specific marginal process or if simultaneous inferences of the full process are to be formed. Now, the expressions above are not feasible (operational) since the autoregressive parameters are unknown, and have to be estimated. Following Lutkepohl (1993) the OLS estimator of the autoregressive parameter $\mathbf{A}$ of (2.2) is given by

$$\hat{\mathbf{A}} = \mathbf{YZ}'(\mathbf{ZZ}')^{-1} = \mathbf{A} + \mathbf{UZ}'(\mathbf{ZZ}')^{-1} \tag{3.4}$$

where $\mathbf{Z}$ is the matrix of 1:st order lag of $\mathbf{Y}$. Furthermore,

$$\sqrt{T} Vec\left(\hat{\mathbf{A}} - \mathbf{A}\right) \xrightarrow{\ell} N\left[\mathbf{0},\ \mathbf{\Gamma}^{-1} \otimes \mathbf{\Sigma_u}\right] \tag{3.5}$$

where $\mathbf{\Gamma} = plim(\mathbf{ZZ}'/T)$, Vec is the vec operator stacking the columns of a matrix into one elongated column vector, $\otimes$ is the Kronecker product and ℓ denotes convergence in law. It is well known that (3.4) is biased towards zero, though the bias vanishes asymptotically according to (3.5). Further analytical results of $\hat{\mathbf{A}} - \mathbf{A}$, are given by e.g. White (1961), for univariate $\mathbf{A}$ and Nicholls and Pope (1988) for the general multivariate case. As we will soon see, the impact of the bias of $\hat{\mathbf{A}}$ is directly connected to the issue of conditioned versus unconditioned forecasts. Replacing the unknown parameter $\mathbf{A}$ with its OLS estimate, the one-step forecast becomes $\hat{Y}_t(1) = \hat{\mathbf{A}} Y_t$ with its corresponding forecast error given by

$$\hat{Y}_t(1) - Y_{t+1} = \hat{\mathbf{A}} Y_t - \mathbf{A} Y_t - U_{t+1} = \left(\hat{\mathbf{A}} - \mathbf{A}\right) Y_t - U_{t+1} \tag{3.6}$$

The above forecast error is known to have zero mean when U_t is symmetric, see e.g Malinvaud (1970), Dufour (1985). These results refer to the *unconditional* forecast. However, the expectation of the error of the forecast *conditional* on $Y_t = y_t$ is usually not zero. For example, for a univariate process, the conditioned forecast error is given by

$$E\left[\hat{Y}_t(1) - Y_{t+1} \middle| Y_t = y_t\right] = -2(A/T) y_t + \mathrm{O}\left(T^{-2}\right) \tag{3.7}$$

(Phillips (1979)). Hence, in general, we have for the multivariate process

$$E\left[\hat{Y}_t(1)-Y_{t+1}\middle|Y_t=y_t\right]=E\left\{\left(\hat{\mathbf{A}}-\mathbf{A}\right)Y_t\middle|Y_t=y_t\right\}\neq\mathbf{0} \tag{3.8}$$

The bias of (3.8) may be reduced either by increasing the sample size, or by replacing $\hat{\mathbf{A}}$ with a bias reduced estimate, if such can be found. However the Y_t component of the forecast error remains outside our control since it is given from the data.

Now, if Y_t is distributed symmetrically and centred around zero, it will obviously have its maximum density at zero. Hence, if one regularly performs forecasts of a system, such as daily air temperature using the latest 24 hourly observations, then one will typically be interested in the average performance of the conditioned forecast. The expectation of such a forecast is

$$E\left[E\left[\hat{Y}_t(1)-Y_{t+1}\middle|Y_t=y_t\right]\right]=E\left[\hat{Y}_t(1)-Y_{t+1}\right]=\mathbf{0} \tag{3.9}$$

where the first identity is given by the law of iterated expectations. The last identity of (3.9) is given by Dufour (1985). In other words, in the light of (3.8-9), it is crucial that the forecaster has a clear idea of whether the success of the forecast should be evaluated with given final period values (i.e. 3.8) or if it is more relevant to evaluate the performance on average (i.e. 3.9). However, the approach of most relevance will vary from case to case. For most situations the conditioned forecast might be the more relevant while in some situations the average performance (i.e. the unconditional forecast) may be the relevant approach. An informative example of the possible implications of conditioning on what is likely to happen (in our case the event that $Y_T=E[Y_T]$) instead of conditioning on the realisation of Y_T that actually happened (in our case the event $Y_T=y_T$) are given by Cox and Hinkley (1974) p. 38.

When it comes to the analytical properties of the point estimator (3.4) the literature is rather sparse. The one dimensional case (i.e the AR(1) process) has been examined by White (1961) who supplies asymptotic bias expressions for the first two powers of the autoregressive parameter up to order $O\left(T^{-5/2}\right)$, while Vinod and Shenton (1996) supply exact moments. However, in the latter case, numerical integrations are needed which may be considered as a high cost of effort for obtaining a precision superior to $O\left(T^{-5/2}\right)$. When it comes to the multidimensional case, the only result (to our knowledge) which provides bias expressions of (3.4) for the case of a stochastic initial value, is that of Nicholls and Pope (1988). These authors supply an asymptotic bias expression up to $O\left(T^{-3/2}\right)$. Even though their derivation is a nice piece of applied mathematics, their bias expression does not allow for bias adjustment of the autoregressive parameter, since it is too complex. In fact, their paper suggests that it is doubtful if it is at all possible to obtain bias reducing constants of a multi dimensional VAR process. Hence, in view of the bias of **A** and (3.8), it seems as forecasting a multivariate VAR(1) process in finite samples is a major problem.

IV. The simulation experiment

The exact properties of the estimated point and interval forecasts of the VAR process are unknown. In addition, the forecast properties will be even more complicated when the forecast is made conditional on the last observed data. There exist very few analytical results in the literature on this somewhat complicated issue and we are left to evaluate the properties by means of Monte Carlo simulations. In this section the Monte Carlo experiment is presented along with a discussion on some of the matters involved in the simulation experiment.

Criteria for judging the quality of forecasts. We previously discussed various measures of concern as regarding the properties of conditioned and unconditioned forecasts. In particular, we argued that the precision of the point predictor and the precision of the interval predictor are of interest. The interval predictor may be considered as a measure of the uncertainty of the point predictor. Its precision, or quality, is usually measured by its covering rate, i.e. the frequency of covering the true unknown future outcome being predicted. Hence we will include the forecast error and the coverage rate of the conditioned/unconditioned forecasts in our simulation experiment. The coverage rate is defined as the frequency of times (in repeated sampling) that the multi dimensional cube contains the true future realisation being forecasted.

Further, we previously discussed the relevance of the MSEP measure. This measure is quite informative as it combines the squared bias of the forecast with the forecast variance. Since the MSEP is a measure in $\mathbb{R}^k$, it will be difficult to evaluate. Therefore we will map it into $\mathbb{R}$ by using the mean value of the diagonal elements of the MSEP matrix as a scalar valued average measure of multivariate MSEP, referred to as TMSEP.

The prediction interval of the VAR process is usually constructed via the marginal statistics $\left\{Y_{k,t+h} - \hat{Y}_{k,t}(h)\right\} / \hat{\sigma}_k(h)$ which have an asymptotic $N(0,1)$ distribution (Lutkepohl (1993)). The k marginal statistics can be joined in a global statistic by using the Bonferroni correction. The procedure may be described as follows: let ψ be the

$(1-\alpha)\cdot 100\%$ percentile from the $N(0,1)$ distribution. Then the joint percentile for the k dimensional $N_k(\mathbf{0},\mathbf{I})$ distribution may be found by taking the $(1-\alpha/k)\cdot 100\%$ percentile, $\tilde{\psi}$, say, from the $N(0,1)$ distribution. By constructing marginal confidence intervals using $\tilde{\psi}$ rather than ψ, i.e. $\hat{Y}_{k,t}(h)\pm\hat{\sigma}_k(h)\tilde{\psi}$, we get (asymptotically) a simultaneous prediction interval with a covering rate of at least $(1-\alpha)\cdot 100\%$. We will refer to this measure in our experiment as the covering rate (CR).

Details on the simulations. The distributional properties of the unconditioned forecast are straight forward and impose no particular difficulties regarding the simulations. The conditioned forecasts on the other hand need more attention. There is no technique to generate a stochastic process in such a way that the last observation attains a certain pre-defined value. Hence we are forced to generate a large number of series and save only those where the last observation lies in close proximity to the pre-defined value of the last observation. Lets say we have sufficient computer power at hand to save only those replications which are within an interval of $y_t \pm \Delta$. For example, if we are interested in the forecast properties conditional on $Y_t = 4$ and chose $\Delta = 0.1$, we would then use only the realised processes with a last observation within $3.9-4.1$. One may wonder how this choice of Δ will affect the results of the simulations as compared to the extremes $\Delta = \lim_{\Delta\to 0}$ and $\Delta = \lim_{\Delta\to\infty}$ respectively. The latter case, i.e. $\Delta = \lim_{\Delta\to\infty}$, corresponds to the whole real line and is hence equal to the unconditional case. In other words, the choice $\Delta = 0.1$ is slightly closer to the unconditional case as compared to $\Delta = \lim_{\Delta\to 0}$. Hence our results are conservative in the sense that the simulated biases are underestimated, i.e. the true biases are larger as compared to the simulated.

Choice of parameters in the simulation. The bias in the point estimate is a function of the true parameter values of the process, i.e. $\mathbf{A}$ of (2.2). To choose the elements of $\mathbf{A}$ one by one becomes quite tedious when the dimension of the process increases. When there are three equations present there will be 9 parameters to determine, and this will be messy unless some systematic technique is applied. The properties of the **A** matrix is a function of its eigenvalues in the sense that $\mathbf{A} = \mathbf{P \Lambda P^{-1}}$, or equivalently, $\mathbf{\Lambda} = \mathbf{PAP^{-1}}$ where $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues of **A**, and **P** its corresponding eigenvectors, see Lutkepohl (1993). If any of the eigenvalues have modulus one, then the process will be non-stationary which is a non-standard situation and will therefore not be considered in this paper. Further, if the eigenvalues are zero the VAR process reduces to a white noise vector, which is irrelevant in this paper. Hence, we will use a setting that has high but stationary autocorrelation. This is done by putting the largest eigenvalue of **A**, λ_{max}, equal to 0.8 in all models.

The VAR process is assumed to have started in an infinite past. In order to mimic this behaviour we have used 50 values to start up the process. These first 50 values have then been removed from the analysed data. Since the autocovariance of the VAR(1) process may be expressed as $\Gamma(h) = A^h \Sigma_Y = \left(P \Lambda^h P^{-1}\right) \Sigma_Y$, and because the largest eigenvalue in this study is 0.8, it follows that the largest autocovariance for 50 lags is smaller than $1.5 \cdot 10^{-5}$. Hence 50 start-up values should suffice for all parameterisations of **A** considered here. The factors on which the simulations depend are displayed in Table i. and Table ii. below.

Table i. Factors used in the experiment

FACTOR	**SYMBOL**	**VALUE**
Number of MC Replications	R	10 000
Number of observations	T	15, 22, 32, 46, 66.
Number of equations	K	1, 2, 3.
Number of forecast periods	h	1, 2.
Autoregressive parameter	**A**	$\lambda_{\max} = 0.8$ (the highest eigenvalue)
Covariance matrix	Σ_U	$5 \cdot \mathbf{I}$
Last observed values	Y_t	**Fixed** in the conditional forecasts $y_t = 0, 2, 4$. **Stochastic** $E(Y_t) = 0$ in the unconditional forecast

Table ii. Measures used in the simulation study to describe the properties of the forecast.

MEASSURE	DEFINITION
Coverage probability	Percentage of the replicate where none of the predictions is outside the interval.
Forecast bias	$E\left[y_{t+h}-y_t(h)\right]$
TMSEP*	$TMSEP = trace\left[E\left(Y_{t+h}-Y_t(h)\right)\right]^2/K = \Sigma_y(h)/K$

*The TMSEP for the h-step forecast, i.e. $trace\{\Sigma_y(h)\}/K$, may be calculated by $trace\{\Sigma_y(1)\}/K = trace\{\Sigma_u\}/K$ and $trace\{\Sigma_y(2)\}/K = trace\{\Sigma_u + A\Sigma_u A'\}/K$, Lutkepohl (1993).

V. Results.

The graphs and tables in the following section contain results from the simulations of one, two and three dimensional VAR(1) models. The simulations were performed using all assumptions from Section II. This means that all our simulation results correspond to the ideal situation and can hence be viewed as the optimal performance of the model. The series were simulated with varying factors according to Table i. above. We have also supplied the asymptotic (i.e. when $T \to \infty$) measures so that the estimated finite sample measures may be compared to this. Clearly then, the closer the estimated measures are to its asymptotic limits, the better the forecast property. The results are shown in the graphs and tables below. In Table iii. below, we present the estimated confidence interval covering rate for conditioned and unconditioned forecasts, with forecast horizon one step ahead. In comparison between the unconditioned forecast and the forecast conditioned on $Y_t = 0$, it is striking that the two covering rates are nearly equal. The covering rates are slightly underestimated, particularly in the three-dimensional process, though they limit the nominal covering rates asymptotically. Moving on to table vi., where the same parameter settings are used as in Table v., though with the forecast horizon two steps ahead, no further news is provided. The covering rates are underestimated, but limit their nominal covering rate asymptotically. Further on, in Table v. we examine the TMSEP of the conditioned and unconditioned forecasts. The forecasts conditioned on $y_t = 0$ are uniformly closer to its limiting value over the whole range of sample sizes and dimensions, as compared to the unconditioned case. However, when $y_t = 4$, the conditioned forecasts converge slower than the unconditioned forecasts, in terms of TMSEP. Moving on to Table iv., the properties of the two-step predictions show similar, but accentuated, behaviour as that of the one-step predictions. Also, in Tables iii.-vi, the biases increase with the value of y_t. However, this is not the case with the forecast error. Following Tables vii. and viii., the forecast bias of the conditioned forecast when $y_t = 2$ is larger than the bias of the forecast conditioned on $y_t = 4$, when $K = 2$. In fact, in this case, the forecast is heavily biased upwards when $y_t = 2$ but slightly downwards when $y_t = 4$. When $K = 3$, the bias of the forecast conditioned on $y_t = 2$ is close to zero, while the bias of the forecast conditioned on $y_t = 4$ is highly negative (note though, that the behaviour of the forecast conditioned on $y_t = 0$ is more or less identical to that of the unconditioned forecast). This seemingly odd performance needs further investigation.

In Figures i. and ii. below we present simulations regarding forecast error as a function of y_t. Figure i. visualises the forecast error for the two dimensional processes two step ahead. Obviously, the forecast error is not monotonically increasing with y_t. The forecast error reaches a maximum when the forecasts are conditioned on values close to $y_t = 2$. This explains the seemingly odd results of Tables vii-viii, where the bias of the forecast conditioned on $y_t = 2$ is high whereas it is zero at $y_t = 4$. When it comes to Figure ii, the forecast properties are altered as compared to the two-dimensional case in Figure 1. The forecast bias at $y_t = 2$ is slightly larger than 0, though it is highly negative at $y_t = 4$.

Table i. One step (h=1) covering rate for conditioned and unconditioned forecast.

	One dimension (k=1)				Two dimensions (k=2)				*Three dimensions (k=3)*			
NOBS	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	*Yt=4*
15	0,94	0,91	0,92	0,93	0,88	0,90	0,89	0,89	0,84	0,91	0,89	*0,85*
22	0,95	0,93	0,93	0,93	0,90	0,91	0,91	0,91	0,90	0,92	0,91	*0,89*
32	0,95	0,93	0,94	0,94	0,92	0,93	0,93	0,92	0,93	0,94	0,93	*0,92*
46	0,95	0,94	0,94	0,94	0,94	0,94	0,94	0,93	0,95	0,94	0,94	*0,92*
66	0,95	0,94	0,95	0,94	0,94	0,94	0,93	0,94	0,95	0,94	0,95	*0,94*
∞	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*

Table ii. Two step (h=2) covering rate for conditioned and unconditioned forecast.

	One dimension (k=1)				Two dimensions (k=2)				*Three dimensions (k=3)*			
NOBS	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	*Yt=4*
15	0,93	0,89	0,90	0,91	0,87	0,88	0,87	0,89	0,85	0,90	0,90	*0,86*
22	0,93	0,91	0,92	0,92	0,90	0,90	0,90	0,91	0,90	0,92	0,91	*0,91*
32	0,94	0,92	0,92	0,94	0,91	0,92	0,92	0,92	0,93	0,92	0,92	*0,93*
46	0,94	0,93	0,94	0,94	0,93	0,93	0,93	0,92	0,95	0,93	0,93	*0,94*
66	0,94	0,94	0,94	0,94	0,93	0,93	0,93	0,94	0,95	0,94	0,94	*0,94*
∞	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*	*0,95*

Table iii. One step (h=1) Trace Mean Square Error of Prediction.

	One dimension (k=1)				Two dimensions (k=2)				*Three dimensions (k=3)*			
NOBS	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	*Yt=4*
15	5,50	5,12	5,25	5,56	5,87	4,90	5,37	6,01	6,64	5,02	5,95	*7,89*
22	5,20	5,02	5,16	5,42	5,70	5,18	5,33	5,55	5,90	5,09	5,54	*6,71*
32	5,25	5,13	5,03	5,25	5,41	4,97	5,06	5,59	5,61	5,05	5,42	*5,92*
46	5,04	4,99	5,05	5,08	5,15	5,01	4,98	5,37	5,43	5,04	5,15	*5,64*
66	5,07	5,09	4,90	5,14	5,23	5,06	5,11	5,10	5,28	5,11	5,08	*5,31*
∞	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*

Table iv. Two step (h=2) Trace Mean Square Error of Prediction.

	One dimension (k=1)				Two dimensions (k=2)				*Three dimensions (k=3)*			
NOBS	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	*Yt=4*
15	9,18	8,31	8,58	9,22	8,24	6,88	7,49	8,45	7,29	5,94	6,58	*9,42*
22	8,96	8,21	8,46	8,91	7,90	7,01	7,54	7,86	6,71	6,11	6,58	*7,73*
32	8,75	8,39	8,34	8,38	7,61	6,95	7,26	7,66	6,45	6,05	6,39	*6,94*
46	8,64	8,25	8,15	8,41	7,27	6,94	7,22	7,62	6,30	6,07	6,27	*6,96*
66	8,29	8,24	8,13	8,53	7,41	6,99	7,23	7,22	6,24	6,11	6,10	*6,49*
∞	*8,20*	*8,20*	*8,20*	*8,20*	*7,00*	*7,00*	*7,00*	*7,00*	*6,04*	*6,04*	*6,04*	*6,04*

Table v. One step (h=1) forecast error.

	One dimension (k=1)				Two dimensions (k=2)				*Three dimensions (k=3)*			
NOBS	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	*Yt=4*
15	0,03	0,02	0,30	0,24	0,03	-0,01	0,25	-0,02	0,00	-0,01	0,16	*-0,46*
22	0,07	-0,00	0,18	0,18	-0,02	0,02	0,22	0,02	0,01	0,01	0,13	*-0,27*
32	-0,03	-0,01	0,14	0,18	-0,02	-0,00	0,11	0,05	-0,01	0,01	0,09	*-0,20*
46	-0,01	0,02	0,09	0,12	-0,02	0,01	0,09	0,06	0,01	-0,01	0,08	*-0,12*
66	0,01	0,01	0,06	0,08	0,01	-0,00	0,05	0,02	0,01	0,02	0,05	*-0,10*
∞	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*

Table vi. Two step (h=2) forecast error.

	One dimension (k=1)				Two dimensions (k=2)				*Three dimensions (k=3)*			
NOBS	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	**Yt=4**	**Unc.**	**Yt=0**	**Yt=2**	*Yt=4*
15	0,01	-0,03	0,35	0,25	0,03	0,03	0,30	-0,13	0,01	-0,02	0,09	*-0,73*
22	0,09	0,02	0,23	0,19	-0,01	0,00	0,28	-0,03	0,00	-0,01	0,10	*-0,45*
32	-0,02	-0,01	0,15	0,18	-0,03	-0,01	0,11	0,03	0,00	0,01	0,05	*-0,29*
46	0,00	0,04	0,18	0,17	0,00	0,01	0,10	0,05	0,00	-0,02	0,06	*-0,16*
66	0,00	-0,02	0,06	0,09	0,03	0,01	0,00	0,02	0,01	0,01	0,01	*-0,12*
∞	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*

Figure i. Bias of the forecast conditioned on Y_t in the two dimensional process, h = 2, N = 22.

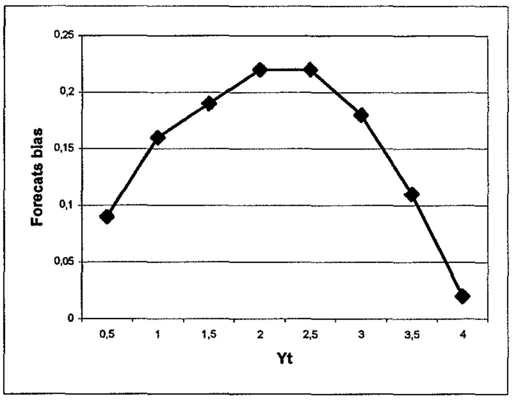
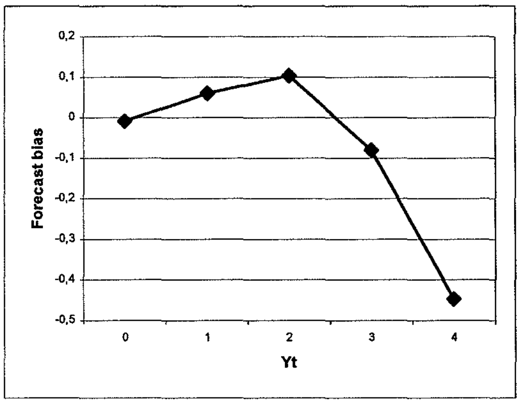


Figure ii. Bias of the forecast conditioned on Y_t in the three dimensional process, h = 2, N = 22.



To sum up, our graphs and tables show some interesting findings regarding conditioned versus unconditioned forecasts. In particular, the conditioned forecast might possess extremely bad performances, in terms of coverage rate of the intervals and in terms of forecasts errors. However, there does not seem to exist a simple linear relationship between the performance of the conditioned forecast and the value of the final observation being conditioned on, nor on the dimension of the process. In some situations the bias is increasing with the number of dimensions of the process, but for some values of the last observation, the bias decreases when the number of dimensions grows. Further, the unconditioned forecast is inferior to the conditioned forecast for some combinations of the last observation y_t and the dimension of the process, *h*, whereas it is superior for other combinations of y_t and *h*. Hence, it is not possible to conclude that the conditioned forecast is better or worse as compared to the unconditioned forecast. Consequently, since the performance of the two ways of regarding the forecast (conditioned or unconditioned) may differ considerably, it is important that the forecaster has a clear idea whether the forecast being performed should be considered conditioned or unconditioned on the final observation. In particular, we have demonstrated that the forecast bias of the conditioned forecast may be substantial.

VI. Summary and conclusions.

In this paper we have examined some differences in the performance of the conditioned vs the unconditioned forecast of the VAR(1) process. We have used measures of relevance for evaluation of forecasts, namely TMSEP, FE and CR. These measures have then been used in a Monte Carlo experiment in order to quantify the differences in the performance of the two approaches of forecasting. Our main findings are that unconditioned forecast is unbiased, as expected, but the conditioned forecast may have a substantial bias. Also, the squared measure TMSEP, is shown to be much higher for the conditioned forecast as compared to the unconditioned counterpart when the last observation of the time series is large. In particular, the difference in performance of the two forecasts may be expected to worsen as the number of equations in the process increases. The main source to the poor performance of the conditioned forecast is that of the bias of the autoregressive parameter, which in turn is a small sample problem as it vanishes asymptotically. But on the other hand, we find that the conditioned forecast may perform better than the unconditioned forecast whenever the last observation is small. Hence, if one has to perform forecasts based on small samples, it is extremely important to have a clear idea of whether the forecast should be conditioned on the last data, or if it should be unconditioned. The last situation usually appears when performing successive forecasts on data from the same population, in which case the average performance, i.e. the unconditioned forecast, is the relevant occurrence.

Since the conditioned forecast might be rather different as compared to the unconditioned forecast, when based on a small sample, one may question whether the performance can be improved in some way. Unfortunately the bias of the point estimate of the autoregressive parameter, which is the source of the forecast bias, seems too complicated to be useful as a bias adjustment. Hence the well-known problem that forecasts of VAR processes have rather wide prediction intervals is extended by having a prediction bias. The seemingly trivial problem of forecasting a VAR process is hence subject for further research.

References

Abadir, K. and R. Larsson (1996). "The joint moment generating function of quadratic forms in multivariate autoregressive series." *Econometric theory* 12: 682-704.

Baccala, L. A. and K. Sameshima (2001). "Partial directed coherence: a new concept in neural structure determination." *Biological Cybernetics* 84: 463-474.

Chatfield, C. (1993). "Calculating Interval forecasts." *Journal of Business & Economic Statistics* 11(2): 121-130.

Cox, D. R. and D. V. Hinkley (1974). Theoretical Statistics, Chapman and Hall Ltd.

Dufour, J. M. (1985). "Unbiasedness of predictions from etimated vector autoregressions." *Econometric Theory* 1: 387-402.

Edlund, P. O. and S. Karlsson (1993). "Forecasting the Swedish Unemployment Rate Var Vs Transfer- Function Modeling." *International Journal of Forecasting* 9(1): 61-76.

Lutkepohl, H. (1993). Introduction to Multiple Time Series Analysis, Pringer-Verlag.

Malinvaud, E. (1970). Statistical Methods of Econometrics. Amsterdam, North Holland Publishing Company.

Nicholls, D. F. and A. L. Pope (1988). "Bias in the estimation of multivariate autoregression." *Australian Journal of Statistics*(30A): 296-309.

Peiris, D. R. and J. W. McNicol (1996). "Modeling daily weather with multivariate time series." *Agricultural and forest meteorology* 79: 219-231.

Phillips, P. C. B. (1979). "The sampling distribution of forecasts from a first-order autoregression." *Journal of Econometrics* 9(2): 241-261.

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

Stergiou, K. I. and E. D. Christou (1996). "Modelling and forecasting annual fisheries catches:

Comparison of regression, univariate and multivariate time series methods." *Fisheries Research* 25: 105-138.

White, J. S. (1961). "Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient." *Biometrika* 48(1/2): 85-94.

Vinod, H. D. and L. R. Shenton (1996). "Exact moments for autoregressive and random walk models for a stationary initial value." *Econometric Theory* 12: 481-499.

Research Report

2002:1	Frisén, M. & Sonesson, C.:	Optimal surveillance. Based on exponentially weighted moving averages.
2002:2	Frisén, M.:	Statistical surveillance. Optimality and methods.
2002:3	Jonsson, R. & Persson, A.:	Bayes prediction of binary outcomes based on correlated discrete predictors.
2002:4	Persson, A.:	Prediction of work resumption among men and women with lower back- and neck pain in a Swedish population.
2002:5	Persson, A.:	Prediction of work resumption in theory and practice.
2002:6	Andersson, E., Bock, D. & Frisén, M.:	Detection of turning points in business cycles.
2002:7	Andersson, E., Bock, D. & Frisén, M.:	Some statistical aspects on methods for detection of turning points in business cycles.
2002:8	Andersson, E., Bock, D. & Frisén, M.:	Statistical surveillance of cyclical processes with application to turns in business cycles.
2002:9	Holgersson, T.:	Testing for non-normality in multivariate regression with nonspherical disturbances.
2003:1	Holgersson, T. & Shukur, G.:	Testing for multivariate heteroscedasticity.
2003:2	Holgersson, T.:	Testing for multivariate autocorrelation.
2003:3	Petzold, M. & Sonesson, C.:	Detection of intrauterine growth restriction.
2003:4	Bock, D.:	Similarities and differences between statistical surveillance and certain decision rules in finance.