SERIE RESEARCH MEMORANDA

ON THE INTEGRATION OF MULTI-STEP PREDICTION
AND MODEL SELECTION FOR STATIONARY TIME SERIES

J.T.C. Kool
A.H.Q.M. Merkies

Researchmemorandum 1986-22 July 1986

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
A M S T E R D A M

On the Integration of Multi-Step Prediction and Model Selection for Stationary Time Series

J.T.C. Kool and A.H.Q.M. Merkies *)

Abstract

An important aim of modelling economic time series is the generation of predictions. However a model, which is suitable for describing the data within a sample, may be suboptimal for the generation of out sample forecasts, especially for increasing time horizons.

The objective of this paper is to analyze a model selection procedure, which is also based on the predictive performance. In first instance the time horizon of prediction is formulated and then the appropriate model is selected, so different time horizons of prediction may lead to different models. After a suitable model has been chosen, we also examine the method of estimation. The properties of the various estimators and the model selection procedure are investigated by simulation.

*) Department of Economics and Department of Econometrics and Actuarial Science, Free University, P.O. Box 7161, 1007 MC Amsterdam, The Netherlands.

I It is quite common in time series analysis to separate model identification and prediction, due to the implicit assumption that prediction is straightforward, once a model is identified. Consequently the time horizon of the predictions usually only appears in the second stage of this procedures; it is ignored in the identification. We will show that if time series are modelled with some ARIMA-process this approach may be not optimal. In practical applications of model building two sources of error occur. First misspecification can hardly be avoided. Even in the case where economic data follow an ARIMA process exactly it is difficult to dectect the true process in moderate sample sizes (Sneek, (1984)). Secondly the assumed model is only an approximation of the data generating process.

Findley (1984) shows that a model which is optimal for one-step ahead prediction in terms of prediction error variance may be suboptimal when the goal is multi-step prediction or spectral estimation. Gersch and Kitagawa (1983) consider model selection and prediction simultaneously. By using monthly data they conclude that a model which is appropriate for one-step ahead prediction differs from the model which is used for the generation of twelve-step ahead predictions. For non-stationary time series the optimal forecast procedure is dependent on the time horizon of the predictions and the prefilter which is used to make the time series stationary (Meese and Geweke, (1984)).

In this paper we consider model identification and prediction for stationary time series simultaneously and we concentrate on the implications of misspecification. In section II asymptotic results are derived, while in section III expressions for small samples are given. The method of estimation and order determination are discussed in section IV. Some simulation results are presented in section V and in the last section some conclusions are drawn.

II Consider the infinite time series $\{X_t, t=0, \pm 1, \pm 2, \ldots\}$ and assume that $\{X_t\}$ can be written as

$$X_t = \psi_0 \xi_t + \psi_1 \xi_{t-1} + \psi_2 \xi_{t-2} + \cdots ,$$

where $\psi_0 = 1$ and $\{\xi_t, t = 0, \pm1, \pm2,\ldots\}$ is a sequence of Gaussian innovations with the properties

$$E[\xi_t] = 0 \ , \quad E[\xi_t^2] = \sigma^2 \ , \ E[\xi_t \, \xi_{t-\tau}] = 0 \qquad \tau \neq 0 \ .$$

We define the function $\psi(z)$ as

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j \, z^j$$

and we require that $|\psi(z)|$ is bounded away from zero for $|z| \leq 1$ and

$$\sum_{j=0}^{\infty} |\psi_j| < M < \infty \ .$$

The spectral density function of $\{X_t\}$ is

$$h(\lambda) = \frac{\sigma^2}{2\pi} |\psi(z)|^2 \quad ,$$

where $z = e^{i\lambda}$, $-\pi \leq \lambda \leq \pi$.

In the class of all predictors of X_{t+1} only linear predictors are considered. The one-step linear predictor $\hat{X}_{t+1}$ is

$$\hat{X}_{t+1} = \sum_{j=1}^{\infty} \alpha_j^{(1)} \, X_{t+1-j}$$

and the expected mean square error of this predictor is

$$\sigma_\eta^2(1) = E[(X_{t+1} - \hat{X}_{t+1})^2] = \frac{\sigma^2}{2\pi i} \oint |\alpha^{(1)}(z)|^2 \, |\psi(z)|^2 \, \frac{1}{z} \, dz$$

where $\alpha^{(1)}(z) = 1 - \sum_{j=1}^{\infty} \alpha_j^{(1)} \, z^j$

and $\oint$ indicates integration around the unit circle.

The minimum of $\sigma_\eta^2(1)$ equals σ^2 and is attained if

$$|\alpha^{(1)}(z)|^2 = |\psi(z)|^{-2} \qquad \text{(Bloomfield, (1972))}.$$

Long term predictions can be calculated recursively for any linear prediction filter $\alpha^{(1)}(z)$ by

$$\hat{X}_{t+k} = \sum_{j=1}^{k-1} \alpha_j^{(1)} \, \hat{X}_{t+k-j} + \sum_{j=k}^{\infty} \alpha_j^{(1)} \, X_{t+k-j}$$

with an expected mean square error of prediction of

$$\sigma_\eta^2(k) = E[(X_{t+k} - \hat{X}_{t+k})^2] = \frac{\sigma^2}{2\pi i} \oint |\alpha^{(1)}(z)|^2 \, |\psi(z)|^2 \, |\sum_{j=0}^{k-1} c_j z^j|^2 \frac{1}{z} \, dz$$

where

$$c(z) = \{\alpha^{(1)}(z)\}^{-1} .$$

In the prediction procedure described above k-step ahead predictions ($k > 1$) can only be obtained if we first compute $\hat{X}_{t+1}$, then $\hat{X}_{t+2}$ etc.

An alternative approach is to form k-step ahead predictions directly by

$$\hat{X}_{t+k} = \sum_{j=1}^{\infty} \alpha_j^{(k)} X_{t+1-j} .$$

By defining

$$\alpha^{(k)}(z) = 1 - \sum_{j=1}^{\infty} \alpha_j^{(k)} z^{k-1+j}$$

it follows that

$$\sigma_\xi^2(k) = E[(X_{t+k} - \hat{X}_{t+k})^2] = \frac{\sigma^2}{2\pi i} \oint |\alpha^{(k)}(z)||\psi(z)|^2 \frac{1}{z} \, dz .$$

The greatest lower bound of both $\sigma_\eta^2(k)$ and $\sigma_\xi^2(k)$ is

$$\tau_k = \sigma^2 \sum_{j=0}^{k-1} \psi_j^2 .$$

When the true model is selected, and the true parameters are found, this implies that

$$\alpha^{(1)}(z) = \psi^{-1}(z)$$

and $\sigma_\eta^2(k)$ reduces to τ_k, while minimization of $\sigma_\xi^2(k)$ with respect to $\{\alpha_j^{(k)}, j \geq 1\}$ would not lead to a reduction of the forecast error variance and the model is optimal in this sense.

More interesting is the case where the underlying model is misspecified, so

$$\alpha^{(1)}(z) \neq \psi^{-1}(z) .$$

By rewriting

$$|\alpha^{(1)}(z)|^2 |\sum_{j=0}^{k-1} c_j z^j|^2 = |\alpha^{(1)}(z) \sum_{j=0}^{k-1} c_j z^j|^2 = |1 - \sum_{j=k}^{\infty} b_j z^j|^2$$

it follows

$$\sigma_\eta^2(k) = \frac{\sigma^2}{2\pi i} \oint |1 - \sum_{j=k}^{\infty} b_j z^j|^2 \, |\psi(z)|^2 \, \frac{1}{z} \, dz$$

which is equivalent with the expressions obtained for $\sigma_\xi^2(k)$. By minimizing $\sigma_\xi^2(k)$ with respect to $\{\alpha_j^{(k)}, j \geq 1\}$ it follows that the prediction error variance of k-step ahead forecast can be reduced by choosing an other model.

Below we consider prediction filters, which contains only a finite number of parameters. The one-step ahead predictor filter $\{\alpha^{(1)}(z)\}$ can be written as

$$\alpha^{(1)}(z) = 1 - \alpha_1^{(1)} z - \ldots - \alpha_p^{(1)} z^p$$

and for the k-step predictor the filter becomes

$$\alpha^{(k)}(z) = 1 - \alpha_1^{(k)} z^k - \ldots - \alpha_p^{(k)} z^{k+p-1} \quad .$$

We introduce the notation AR(p,k) , which specifies a prediction scheme with $p+k-1$ lags and p parameters. If $X_t = AR(p,k)$ this implies

$$X_t = \alpha_1^{(k)} X_{t-k} + \ldots + \alpha_p^{(k)} X_{t-k-p+1} + \eta_t \quad .$$

The prediction filter, which minimizes the error variance can be obtained by minimizing

$$\sigma_\eta^2(p,k) = \frac{\sigma^2}{2\pi i} \oint |\alpha^{(k)}(z)|^2 \, |\psi(z)|^2 \, \frac{1}{z} \, dz =$$

$$= \frac{\sigma^2}{2\pi i} \oint |\alpha^{(k)}(z)|^2 \, \Big| \sum_{j=-\infty}^{\infty} r_j z^j \Big|^2 \frac{1}{z} \, dz$$

where $r_j = E[X_t X_{t-j}]$.

Minimizing $\sigma_\eta^2(p,k)$ with respect to $\{\alpha_1^{(k)} \ldots \alpha_p^{(k)}\}$ is equivalent with solving the Yule-Walker equations, so

$$\begin{bmatrix} \hat{\alpha}_1^{(k)} \\ \hat{\alpha}_2^{(k)} \\ \vdots \\ \hat{\alpha}_p^{(k)} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \ldots & \rho_{p-1} \\ \rho_1 & 1 & \ldots & \rho_{p-2} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \ldots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_k \\ \rho_{k+1} \\ \vdots \\ \rho_{p+k-1} \end{bmatrix} \qquad (2.1)$$

and $\hat{\sigma}_\eta^2(p,k) = \sigma^2 r_0 \{1 - \sum_{j=1}^{p} \hat{\alpha}_j^{(k)} \rho_{k-1+j}\}$

where $\rho_j = r_j / r_0$.

If the 1-step filter is used for the computation of $\hat{X}_{t+k}$ then

$$\text{var}\{X_{t+k} - \hat{X}_{t+k}\} \geq \hat{\sigma}_\eta^2(k)$$

where $\hat{\sigma}_\eta^2(k)$ is the minimal forecast error variance of the k-step filter. By using a model selection procedure which is also based on considerations of the predictive performance it is also possible to reduce the error variance when the 1-step filter is misspecified. The results obtained above are valid in the asymptotic case or when the parameters are known. In the next section results are derived when the parameters have to be estimated from a finite sample.

III The prediction error variances obtained in the preceding section are based on the assumption that the parameters are known. If the parameters have to be estimated from a finite number of observations the prediction error variance shall increase. The bias for misspecified models will be determined along the lines given by Davies and Newbold (1980). In first instance the error variance according to the 1-step filter will be computed,

$$f_{T+k} = \frac{\sigma^2}{2\pi i} \oint |\alpha^{(1)}(z)|^2 |\psi(z)|^2 \left| \sum_{j=0}^{k-1} c_j z^j \right|^2 \frac{1}{z} dz + \tilde{\chi}_k(T) \qquad (3.1)$$

where $\tilde{\chi}_k(T)$ is the bias due to estimation and T is the number of observations. For the derivation of $\tilde{\chi}_k(T)$ we need the results given by Yamamoto (1976) for exact AR(p) processes. If

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \xi_t$$

where ξ_t is white noise, then it is shown by him that

$$\tilde{\chi}_k(T) = \text{tr}\{M_k' V_{\hat{\alpha}} M_k \Sigma\}$$

where $V_{\hat{\alpha}}$ is the asymptotic variance covariance matrix of the maximum likelihood estimates of $(\alpha_1 \ldots \alpha_p)$, M_k is the matrix

$$M_k = \sum_{j=0}^{k-1} c_j \, A^{k-1-j}$$

where A is a $(p*p)$ matrix with $a_{1j} = \text{plim}\,(\hat{\alpha}_j)$ $(j=1\ldots p)$, $a_{j+1j}=1$ $(j=1\ldots p-1)$ and zero otherwise and Σ is also a $(p*p)$ matrix with $\Sigma = E[XX']$, where $X = (X_1 \ldots X_p)'$.

For the derivation of the asymptotic variance covariance matrix of the parameters of the misspecified model we introduce the vectors

$$\alpha^{(k)} = (\alpha_1^{(k)} \ldots \alpha_p^{(k)})' \,, \quad \rho^{(k)} = (\rho_k \ldots \rho_{k+p-1})' \,.$$

Now $\hat{\alpha}^{(k)} = \hat{B}^{-1} \hat{\rho}^{(k)} = g(\hat{\rho}^{(k)})$

where $\hat{B}_{ij} = \hat{\rho}_{|i-j|}$ and g is a function $R^{p+k-1} \rightarrow R^p$.

By expanding $\hat{\alpha}^{(k)}$ in a Taylor-expansion around $\rho^{(k)}$,

$$\hat{\alpha}^{(k)} = g\,(\rho^{(k)}) + \left.\frac{\partial g\,(\hat{\rho}^{(k)})}{\partial \hat{\rho}^{(k)}}\right|_{\hat{\rho}^{(k)} = \rho^{(k)}} (\hat{\rho}^{(k)} - \rho^{(k)})$$

it follows

$$V_{\hat{\alpha}}(k) = Q_k \, W_k \, Q_k'$$

where W_k is a $(p+k-1)*(p+k-1)$ matrix with

$$w_{ij} = E\,[(\hat{\rho}_i - \rho_i)(\hat{\rho}_j - \rho_j)]$$

and Q_k is a $p*(p+k-1)$ matrix with

$$q_{ij} = \frac{\partial g_i}{\partial \rho_j} \,.$$

The elements of W are given by Anderson (1971, p.489) explicitly

$$w_{ij} = \sum_{k=-\infty}^{\infty} \{\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} - 2\rho_i\rho_k\rho_{k+j} + 2\rho_i\rho_j\rho_k^2\} \,.$$

By inserting $V_{\hat{\alpha}}(1)$, $\tilde{\chi}_{K(T)}$ becomes:

$$\tilde{\chi}_K(T) = \text{tr}\,\{M_k' \, Q_k \, W_k \, Q_k' \, M_k \, \Sigma\} \,.$$

For the k-step filter the error variance is

$$g_{T+k} = \frac{\sigma^2}{2\pi i} \oint |\alpha^{(k)}(z)|^2 |\psi(z)|^2 \frac{1}{z} dz + \hat{\chi}_k(T) \qquad (3.2)$$

where $\hat{\chi}_k(T)$ is the bias due to estimation. For the k-step model M_k is equal to the identity matrix, so $\chi_k(T)$ reduces to

$$\hat{\chi}_k(T) = \text{tr}\{Q_k W_k Q_k' \Sigma\}$$

The derivations given above are based on the assumption that the series which is used for estimation is independent of the series for which the predictions are made, but the two series have the same stochastic structure. For practical work this is a rather unrealistic assumption, but there is some Monte-Carlo evidence that the bias due to the dependence between the observations used for the estimation and the predictions is small (Reinsel, 1980).

IV Having observed a finite realization of the series the parameters can be estimated by solving the equations (2.1). A related problem is the determination of the order of the process. A well known selection procedure and frequently used is the one suggested by Akaike (1974). The order of an AR-process can be determined by minimizing

$$AIC(p) = \ln \hat{\sigma}^2 + \frac{2p}{T} \qquad (p = 1,\ldots,M) \qquad (4.1)$$

where $\hat{\sigma}^2$ is an estimate of the residual variance and M a prespecified upper limit. Shibata (1976) shows that if the true process is finite AR, AIC tends to overestimate the true order, so this procedure is asymptotically not consistent. Although the AIC criterion is originally designed for Y-W estimates several other estimates which are asymptotically equivalent with these estimates have been used for the estimation of σ^2.

By developing order selection criteria a lot of attention has been paid to the second term of the RHS of (4.1) and not to the estimation of the residual variance. However there is some evidence that the method of estimation can influence the order determination (Paulsen and Tsjøtheim (1985), Beamish and Priestley (1981)). Especially when roots of the AR-

polynomial are closely to the unit circle and the process is strongly autocorrelated the Y-W estimate of the residual variance has a severe bias upwards.

The processes considered by us are AR(∞) processes, so the optimal order is determined by minimizing

$$AIC\ (p,k) = \ln \hat{\sigma}_\eta^2\ (p,k) + \frac{2p}{T} \qquad (p=1,\ldots,M)$$

If the process is not degenerated to a finite AR process, this is selection procedure is asymptotically efficient (Shibata (1980)). By applying this rule some care has to be taken. First the parameters are estimated only by using a finite number of observations and secondly the estimation of the error variance can influence the order determination.

In first instance we give some details of the evaluation of $\hat{\sigma}_\eta^2(p,k)$, $(p=1 \ldots M,\ k=1 \ldots K)$ which can be calculated without estimating the parameters $(\alpha_1^{(k)} \ldots \alpha_p^{(k)})$. For $k = 1$, first $\hat{\rho}_1 \ldots \hat{\rho}_{M+1}$ have to be estimated and the matrix R has been computed, where $R_{ij} = \rho_{|i-j|}$ $(i,j = 1 \ldots M+1)$. By making the decomposition

$$R = LDL' \ ,$$

where L is lower triagonal. $l_{ii} = 1$ $(i=1 \ldots M+1)$ and D is diagonal it follows

$$\hat{\sigma}_\eta^2\ (p,1) = \hat{f}_0\ d_{p+1}$$

where $\hat{f}_0$ is the estimated variance of the process (Pagano (1972)).

For $k \geq 2$ the matrix R(p,k) is composed

$$R(p,k) = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} & 0 & \cdots & 0 & \rho_{k+p-1} \\ \rho_1 & 1 & & \rho_{p-2} & \cdot & & \cdot & \cdot \\ \vdots & \vdots & & \vdots & \cdot & & \cdot & \cdot \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 & 0 & \cdots & 0 & \rho_k \\ 0 & \cdots & \cdots & 0 & & & & 0 \\ \vdots & & & \vdots & & I_{k-1} & & \vdots \\ 0 & \cdots & \cdots & 0 & & & & 0 \\ \rho_{k+p-1} & \cdots & & \rho_k & 0 & \cdots & 0 & 1 \end{bmatrix}$$

By decomposing

$$R(p,k) = L(p,k)\ D(p,k)L'(p,k)$$

at the same way as R , then

$$\hat{\sigma}_{\eta}^{2}\ (p,k) = \hat{r}_{0}\ d_{p+k}(p,k) \qquad \text{(McClave (1975))}.$$

For the decomposition of $R(p,k)$ the results of R can be used and only the last row of $L(p,k)$ has to be calculated, while $\hat{\sigma}_{\eta}^{2}(p-1,k+1)$ can be directly obtained from the decomposition of $R(p,k)$ and R , so all residual variances can very efficiently be determined. After the optimal model for each k has been obtained the parameters can be estimated by:

$$\hat{\alpha}^{(k)} = L_{p}'^{-1}\ D_{p}^{-1}\ L_{p}^{-1}\ \hat{\rho}^{(k)}$$

where L_{p}^{-1} and D_{p}^{-1} are the $p \times p$ submatrices of L^{-1} and D^{-1} respectively.

It is well known that for small samples Y-W estimates are biased. An alternative is reestimation by the method of maximum likelihood. Suppose the true distribution of a random variable x can be written as $g(x)$ and the class of possible distributions is $f(x,\theta)$. The quasi-maximum likelihood estimator (QMLE) of θ is the estimator $\bar{\theta}$ obtained by solving

$$\max_{\theta \in \Theta}\ L_{T}\ (x,\theta)$$

where

$$L_{T}\ (x,\theta) = \frac{1}{T} \sum_{t=1}^{T} \log\ f(x_{t}\ ,\ \theta) \quad .$$

If $f(x,\theta)$ contains the true distribution, $f(x,\theta_{0}) = g(x)$ for some $\theta_{0} \in \Theta$, then $\bar{\theta}$ is a consistent estimate of θ_{0}. If $g(x)$ does not belong to the set $f(x,\theta)$, then $\bar{\theta}$ is an estimate of $\hat{\theta}$, which minimizes the Kullback-Leibler Information Criterion

$$I\ (g,\ f,\ \theta) = \int g(x) \log g(x)\ dx - \int f(x,\theta) \log g(x)\ dx$$

and $I(g,\ f,\ \theta) \geq 0$, while $I(g,\ f,\ \theta) = 0$ if and only if $g(x) = f(x,\theta)$ almost everywhere (White (1982)).

The QMLE is a handsome tool for estimating the prediction filters

$$\{\alpha_1^{(k)} \ \ldots \ \alpha_p^{(k)}\} \ .$$

For the 1-step filter

$$x_t = \alpha_1^{(1)} x_{t-1} + \ldots + \alpha_p^{(1)} x_{t-p} + \xi_t \qquad (4.2)$$

where ξ_t is assumed to be white noise with $E[\xi_t^2] = \sigma_\xi^2$. The likelihood function is:

$$L_0 \propto -\frac{T}{2}\log\sigma_\xi^2 - \tfrac{1}{2}\log|\Omega(p)| - \frac{1}{2\sigma_\xi^2}\{\sum_{i=1}^{p}\sum_{j=1}^{p} w^{ij}(p)x_i x_j + \sum_{i=p+1}^{T}(x_i - \ldots \alpha_p^{(1)} x_{i-p})^2\}$$

where $\Omega_{ij}(p) = r_{|i-j|}$, $w^{ij}(p)$ are the elements of $\Omega^{-1}(p)$ and $r_0 \ldots . r_{p-1}$ are the theoretical autocovariances of the process.
For the multi-step filters there are several possibilities for the formulation of the likelihood. A direct approach is that the underlying distribution can be approximated by

$$x_t = \alpha_1^{(k)} x_{t-k} + \ldots + \alpha_p^{(k)} x_{t-k-p+1} + \eta_t \qquad (4.3)$$

where η_t is assumed to be white noise and $E[\eta_t^2] = \sigma_\eta^2$. The likelihood function of the fitted distribution can be written as:

$$L_1 \propto -\frac{T}{2}\log\sigma_\eta^2 - \tfrac{1}{2}\log|\Omega(p+k)| - \frac{1}{2\sigma_\eta^2}\{\sum_{i=1}^{p+k-1}\sum_{j=1}^{p+k-1} w^{ij}(p+k)x_i x_j + \sum_{i=p+k}^{T}(x_i - \ldots \alpha_p^{(k)} x_{i-p-k+1})^2\}$$

However, there is a remarkable difference between L_0 and L_1 . In evaluating L_0 the assumed underlying process is (4.2), whereas for L_1 it is assumed that (4.3) is the true distribution. A predictive likelihood for $k \geq 2$, while (4.2) is the assumed distribution, can be obtained by the likelihood function

$$L_2 \propto -\tfrac{1}{2}\sum_{i=k+1}^{T}\log F_i - \tfrac{1}{2}\sum_{i=k+1}^{T}\frac{v_i^2}{F_i}$$

where $v_i = x_i - x_{i\,|i-k}$ and $F_i = \text{var}\,[v_i]$ (Gersch and Kitagawa (1983)).

The parameters of interest $\{\alpha_1^{(k)}, \ \ldots, \ \alpha_p^{(k)}\}$ can be obtained in two stages. In first instance maximize L_2 with respect to the unknown

parameters and in the second stage calculate $\{\hat{\alpha}_1^{(k)}, \ldots, \hat{\alpha}_p^{(k)}\}$ recursively from the estimates $\{\hat{\alpha}_1^{(1)}, \ldots, \hat{\alpha}_p^{(1)}\}$. For practical applications of L_2, v_i and F_i $(i=k+1, \ldots, T)$ can be determined by the Kalman filter, but differentiation of L_2 can be cumbersome.
By maximizing L_1 and L_2 respectively the parameters of interest may not converge to the same value, even if T tends to infinity, because by assuming different distributions the stationary conditions can change. An example can make this point clear. Suppose the true distribution is AR(2), and the assumed distribution for the 1-step filter is also AR(2),

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \xi_t \quad . \tag{4.4}$$

For this process the stationary conditions are given by the region K_1, where

$$\phi_1 + \phi_2 < 1 \,, \quad \phi_2 - \phi_1 < 1 \,, \quad |\phi_2| < 1 \tag{4.5}$$

(see figure 1). For the sake of simplicity we evaluate the multi-step filter only for $k = 2$. If the assumed distribution is:

$$x_t = \beta_2 x_{t-2} + \beta_3 x_{t-3} + \eta_t \tag{4.6}$$

then this process is stationary if

$$\beta_2 + \beta_3 < 1 \,, \quad \beta_2 - \beta_3 < 1 \,, \quad \beta_3^2 - \beta_2 < 1 \,, \quad |\beta_2| < 1 \tag{4.7}$$

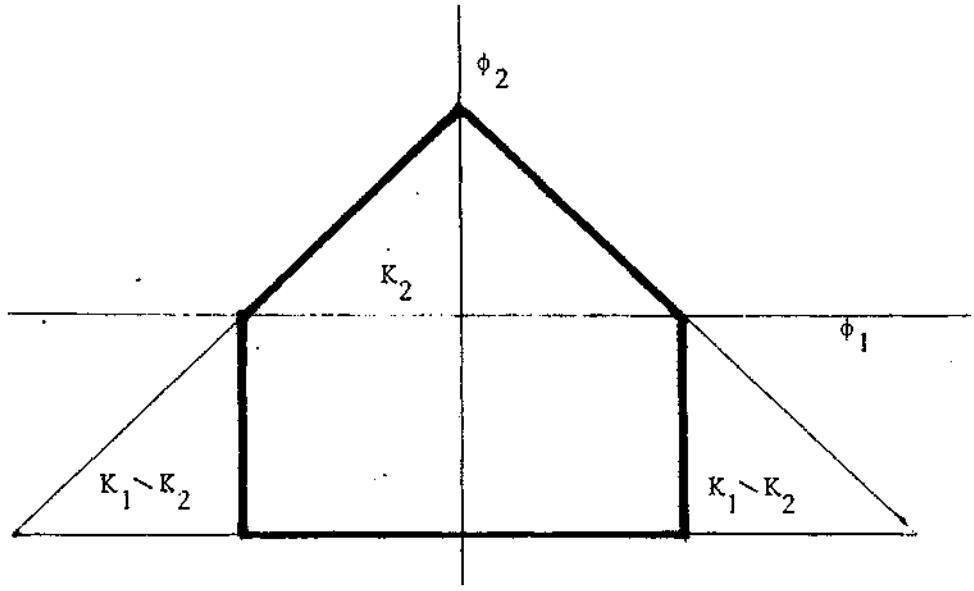

Fig. 1. Stationary regions for the 1-step and 2-step filter of the process $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \xi_t$.

In terms of (ϕ_1,ϕ_2) , (4.6) is:

$$x_t = (\phi_1^2 + \phi_2)x_{t-2} + \phi_1 \phi_2 x_{t-3} + \eta_t \quad .$$

By rewriting

$$1 - (\phi_1^2 + \phi_2)z^2 - \phi_1 \phi_2 z^3 = (1 + \phi_1 z)(1 - \phi_1 z - \phi_2 z^2)$$

it follows that the region K_2 implied by the stationary conditions (4.7) is given by (4.5) and the additional constraint $|\phi_1| < 1$. However $K_2 \subset K_1$, and $K_1 \setminus K_2$ is not of the measure zero, so if $(\phi_1,\phi_2) \in K_1 \setminus K_2$ and when the direct approach is done by using L_1 , the parameters β_2 and β_3 do not converge to the true values, although both likelihood functions are dominated by the residual sum of squares.

In section II optimal prediction filters are obtained by solving the Y-W equations (2.1). For $k = 1$ the roots of the prediction filter lay outside the unit circle (Pagano (1972)). For multi-step predictions the roots of the optimal filter $\{\alpha^{(k)}(z)\}$ do not have necessarily this property, so the application of quasi-likelihood via L_1 can lead to a filter, which is not optimal in terms of prediction error variance.

V The properties of the prediction filters for some series are investigated by simulation. The following time series are considered

AR(1) $\quad x_t = 0.9\, x_{t-1} + \xi_t$

MA(1) $\quad x_t = \xi_t - 0.9\, \xi_{t-1}$

ARMA(1,2) $\quad x_t = 0.8\, x_t + \xi_t + 0.4\, \xi_{t-1} + 0.8\, \xi_{t-2}$.

The AR(1) process has been chosen to analyse the predictors when the class of models which are considered consists also the true model. The sample sizes $T = 50$ and $T = 100$ are investigated, because in applied work this sample sizes often occured. For the three processes the optimal values of p, f_{T+k}, g_{T+k} and $\sigma_\eta^2(k)$ are given in the tables 1 till 3. The variance of the innovation sequence ξ_t is set at 1.0 . Predictions up to a time horizon of 8 are investigated.

The tables should be read als follows: k is the time horizon of prediction, p is the order of the optimal AR(p,k) filter, f_{T+k} is the prediction error variance when the optimal AR(p,1) model has been used for prediction, g_{t+k} is the error variance of the optimal AR(p,k) filter and $\sigma_\eta^2(k)$ is the corresponding asymptotic prediction error variance and is equal to the variance of the assumed distribution. For the ARMA(1,2) process the minimum absolute value of the roots are also given, because all roots of this filter may not lay outside the unit circle. The optimal orders are obtained by evaluating the expressions (3.1) and (3.2) for $p=1...M$ and then select p for which the error variance is minimal. For the selected p, $\sigma_\eta^2(k)$ is obtained by letting T go to infinity.

For the AR(1) model holds asymptotically $f_{T+k} = g_{T+k}$, but for small samples $f_{T+k} < g_{T+k}$, which is due to the fact that f_{T+k} is based on the estimation of ρ_1, whereas for g_{T+k} the estimate of ρ_k is used, so efficiency is lost. For the MA(1) model $g_{T+k} < f_{T+k}$ for $k \geq 2$ and the multi-step prediction filters reduces to a random walk, so g_{T+k} is equal to the variance of the process. In all cases g_{T+k} is less than f_{T+k} for the ARMA(1,2) model, although for $k = 2$ the difference is negligible and for $T = 100$ the difference is smaller than for $T = 50$. For the simulation random variables were generated by making use of the IMSL routine GGNPM, and time series of 407 observations were computed. The first 100 realizations were not used and the following T observations were used for estimation. To make the estimates independent from the predictions, the observations 300 till 407 were used to evaluate the forecast error variances. The number of replications for the AR(1) and MA(1) process were 500, and for the ARMA(1,2) 250 samples were generated. For each process the same innovation sequence was used.

In first instance the order is determined by the Y-W estimates of $\sigma_\eta^2(p,k)$, which don't require the estimation of the remaining parameters. The forecast error variances in small samples are influenced by the method of estimation (Ansley and Newbold (1980, 1981)), so three estimation methods have been applied. First the parameters are obtained by solving the Y-W equations (2.1). Secondly the likelihood L_1 is maximized, although for some prediction filters of the ARMA(1,2) process the estimates cannot

converge to the theoretical values, so the error variances must be biased upwards. The likelihood functions L_1 and L_2 are dominated by the residual sum of squares and we have reestimated the parameters by least squares,

$$\hat{\alpha}^{(k)} = B^{-1} q$$

and

$$\hat{\sigma}_\eta^2(p,k) = \frac{1}{T-p-k+1} \sum_{t=p+k}^{T} (x_t - \hat{\alpha}_1^{(k)} x_{t-k} - \ldots - \hat{\alpha}_p^{(k)} x_{t-k-p+1})^2$$

where

$$B_{ij} = \sum_{t=p+k}^{T} x_{t-k+1-i}\, x_{t-k+1-j}\,, \quad q_i = \sum_{t=p+k}^{T} x_t\, x_{t-k+1-i} \qquad (i,j=1\ldots p).$$

For the analysis of the predictive performance and the order determination the following statistics are computed:

1) The estimated asymptotic forecast error variance

$$\hat{\sigma}_\eta^2(k) = \frac{1}{NR} \sum_{j=1}^{NR} \hat{\sigma}_\eta^2(\hat{p},k)(j)$$

where NR is the number of replications, j is the replication index and $\hat{p}$ is the estimated order of replication j.

2) The mean square error of the variance relative to the theoretical value

$$\text{MSE}\,(\hat{\sigma}_\eta^2(k)) = \{\frac{1}{NR} \sum_{j=1}^{NR} \{\hat{\sigma}_\eta^2(\hat{p},k)(j) - \sigma_\eta^2(k)\}^2\}^{\frac{1}{2}} .$$

3) The bias of the predictions

$$\text{Bias}\,(f_{T+k}) = \frac{1}{NR} \sum_{j=1}^{NR} \{\frac{1}{100} \sum_{i=1}^{100} x_{T+k+i}\,(j) - x_{T+k+i|T+i}\,(j)\}$$

4) The prediction error variance

$$\hat{f}_{T+k} = \frac{1}{NR} \sum_{j=1}^{NR} \left\{\frac{1}{100} \sum_{i=1}^{100} \{(x_{T+k+i}(j) - x_{T+k+i|T+i}(j)) - \overline{(x_{T+k+i}(j) - x_{T+k+i|T+i}(j))}\}^2\right.$$

where the bar implies the estimated mean within a replication.

5) The mean square error of this variance relative to the theoretical value

$$\text{MSE}\,(\hat{f}_{T+k}) = \left(\frac{1}{NR} \sum_{j=1}^{NR} (\hat{f}_{T+k}\,(j) - f_{T+k})^2\right)^{\frac{1}{2}}$$

6) The mean of the estimated orders

$$\bar{p} = \frac{1}{NR} \sum_{j=1}^{NR} \hat{p}(j)$$

7) The mode of the order distribution,

$$\bar{p}_{mod}$$

The statistics 3, 4 and 5 are also computed for g_{T+k} . For each time horizon hundred predictions are computed. The results of the simulation are summarized in the tables 1 till 3. We first discuss briefly the order determination. For the AR(1) process the maximum number of parameters was set at 10 and for the other two processes the upper limit was 15. For $k = 1$ the order distribution of the AR(1) process coincide with the asymptotic distribution given by Shibata (1976). For increasing k the distribution becomes more fat-tailed. Especially for $T = 50$ the distribution tends towards zero, which explains the decreasing mean when k increases. For $k = 1$ the order of the optimal filter of the MA(1) process is slightly underestimated and for increasing k the distribution is dominated by $p = 0$. For the ARMA(1,2) process and small k the order is strongly underestimated, which may be caused by the very erratic behaviour of the sequence $\sigma_\eta^2(p,k)$ (see table 4).

p	1	2	3	4	5	6	7	8	9	10	11	12
k=1,T=50	1.647	1.589	1.391	1.274	1.286	1.230	1.260	1.226	1.243	1.253	1.259	1.277
k=1,T=100	1.640	1.547	1.342	1.228	1.224	1.600	1.175	1.141	1.146	1.145	1.143	1.151
k=2,T=50	3.865	3.249	3.291	2.997	3.059	3.009	3.012	3.032	3.036	3.092	3.094	3.147
k=2,T=100	3.823	3.142	3.146	2.858	2.884	2.811	2.793	2.793	2.774	2.802	2.789	2.815

Table 4. Theoretical values of the asymptotic forecast variance for some selected values of the ARMA(1,2) process.

	k	1		2		3		4		5		6		7		8	
	T	50	100	50	100	50	100	50	100	50	100	50	100	50	100	50	100
	p	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	$\sigma_\eta^2(k)$	1.000	1.000	1.810	1.810	2.466	2.466	2.998	2.998	3.428	3.428	3.777	3.777	4.059	4.059	4.288	4.288
	f_{T+k}^{η}	1.020	1.010	1.875	1.842	2.584	2.525	3.168	3.083	3.643	3.536	4.028	3.902	4.336	4.198	4.581	4.434
	g_{T+k}			1.879	1.844	2.600	2.533	3.201	3.102	3.714	3.571	4.139	3.958	4.495	4.277	4.793	4.540
	$\tilde{p}$	1.430	1.444	1.494	1.666	1.462	1.774	1.394	1.900	1.364	1.914	1.202	1.880	1.086	1.850	1.072	1.808
	$\tilde{p}_{mod}$	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
YW	$\hat{\sigma}_\eta^2(k)$	1.045	1.029	1.771	1.802	2.287	2.385	2.631	2.799	2.854	3.115	3.018	3.348	3.123	3.519	3.182	3.643
	$MSE(\hat{\sigma}_\eta^2(k))$	.275	.173	.542	.352	.821	.551	1.080	.738	1.320	.924	1.539	1.086	1.716	1.233	1.869	1.365
	bias f	.012	.006	.020	.011	.026	.016	.031	.021	.034	.026	.038	.031	.040	.035	.041	.038
	$\hat{f}_{T+k}$	1.107	1.018	2.048	1.846	2.776	2.516	3.292	3.022	3.659	3.405	3.906	3.690	4.079	3.908	4.199	4.074
	$MSE(\hat{f}_{T+k})$	.330	.174	.762	.412	1.084	.686	1.310	.954	1.499	1.161	1.643	1.337	1.749	1.485	1.835	1.618
	bias g			.022	.014	.029	.024	.037	.032	.046	.042	.047	.049	.044	.057	.050	.062
	$\hat{g}_{T+k}$			2.102	1.860	2.912	2.558	3.490	3.099	3.937	3.521	4.216	3.832	4.413	4.089	4.507	4.280
	$MSE(\hat{g}_{T+k})$			.881	.415	1.283	.639	1.535	.993	1.848	1.275	1.949	1.472	2.116	1.690	2.176	1.884
QML	$\hat{\theta}_\eta^2(k)$	.935	.960	1.667	1.723	2.233	2.333	2.627	2.787	2.894	3.150	3.091	3.411	3.238	3.608	3.334	3.757
	$MSE(\sigma_\eta^2(k))$	.201	.140	.436	.307	.713	.495	.989	.682	1.229	.860	1.459	1.027	1.618	1.182	1.771	1.310
	bias f	-.004	.003	.001	.005	.004	.009	.006	.012	.010	.017	.013	.021	.016	.024	.018	.027
	$\hat{f}_{T+k}$	1.045	1.009	1.902	1.825	2.588	2.486	3.106	2.991	3.501	3.382	3.791	3.678	4.014	3.911	4.185	4.094
	$MSE(\hat{f}_{T+k})$	.231	.163	.509	.366	.757	.588	.985	.814	1.194	1.019	1.370	1.209	1.513	1.377	1.637	1.524
	bias g			.001	.006	.011	.013	.020	.020	.017	.023	.018	.032	.014	.033	.017	.043
	$\hat{g}_{T+k}$			1.946	1.834	2.729	2.512	3.331	3.054	3.820	3.494	4.163	3.827	4.343	4.088	4.420	4.298
	$MSE(\hat{g}_{T+k})$			.698	.371	1.020	.609	1.269	.879	1.606	1.169	1.801	1.365	1.887	1.595	1.839	1.776
LS	$\hat{\sigma}_\eta^2(k)$	.930	.959	1.650	1.719	2.202	2.322	2.585	2.773	2.864	3.141	3.052	3.403	3.220	3.598	3.291	3.746
	$MSE(\hat{\sigma}_\eta^2(k))$	.203	.140	.437	.307	.728	.502	1.021	.696	1.267	.879	1.514	1.051	1.675	1.208	1.849	1.334
	bias f	-.001	.002	-.001	.005	.002	.008	.005	.012	.008	.017	.012	.021	.015	.025	.018	.028
	$\hat{f}_{T+k}$	1.048	1.010	1.911	1.827	2.603	2.489	3.127	2.996	3.528	3.387	3.821	3.685	4.047	3.919	4.221	4.104
	$MSE(\hat{f}_{T+k})$	.232	.163	.516	.366	.778	.589	1.022	.816	1.244	1.020	1.425	1.211	1.570	1.378	1.695	1.525
	bias g			.000	.006	.011	.013	.022	.020	.020	.024	.022	.033	.017	.034	.020	.044
	$\hat{g}_{T+k}$			1.964	1.841	2.754	2.525	3.365	3.068	3.855	3.506	4.195	3.838	4.383	4.105	4.480	4.318
	$MSE(\hat{g}_{T+k})$			.702	.373	1.023	.602	1.282	.873	1.607	1.159	1.792	1.361	1.894	1.593	1.919	1.771

Table 1. Theoretical and simulation results of the AR(1) process.

	k	1		2		3		4		5		6		7		8	
	T	50	100	50	100	50	100	50	100	50	100	50	100	50	100	50	100
	p	5	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$\sigma_\eta^2(k)$	1.075	1.043	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810
	f_{T+k}	1.181	1.116	2.011	1.949	1.975	1.931	1.939	1.913	1.907	1.895	1.871	1.878	1.833	1.860	1.817	1.841
	g_{T+k}			1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810	1.810
	$\bar{p}$	4.038	5.920	.900	1.096	.776	1.024	.690	.938	.576	.844	.554	.880	.528	.874	.456	.796
	$\bar{p}_{mod}$	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
YW	$\hat{\sigma}_\eta^2(k)$	1.000	.991	1.656	1.710	1.665	1.713	1.675	1.718	1.685	1.720	1.685	1.719	1.669	1.721	1.698	1.725
	$MSE(\hat{\sigma}_\eta^2(k))$	.246	.163	.451	.316	.445	.318	.436	.316	.434	.317	.434	.320	.433	.316	.430	.315
	bias f	.002	.002	.001	.001	.002	.002	.000	.000	.001	.001	.001	.001	.000	.000	.000	.000
	$\hat{f}_{T+k}$	1.230	1.138	2.025	1.943	1.976	1.927	1.933	1.907	1.898	1.886	1.872	1.867	1.853	1.852	1.842	1.843
	$MSE(\hat{f}_{T+k})$	.240	.195	.462	.393	.412	.384	.386	.372	.374	.362	.360	.349	.348	.341	.339	.338
	bias g			.001	.001	.002	.002	.000	.000	.001	.001	.000	.000	.000	.001	.001	.000
	$\hat{g}_{T+k}$			1.900	1.856	1.892	1.859	1.883	1.851	1.869	1.848	1.865	1.846	1.862	1.848	1.861	1.848
	$MSE(\hat{g}_{T+k})$			.438	.373	.404	.372	.403	.365	.385	.353	.387	.348	.368	.349	.362	.343
QML	$\hat{\sigma}_\eta^2$	.943	.953	1.645	1.705	1.651	1.707	1.661	1.712	1.669	1.714	1.667	1.711	1.670	1.713	1.679	1.716
	$MSE(\hat{\sigma}_\eta^2(k))$	.259	.175	.462	.321	.455	.324	.449	.322	.444	.323	.446	.327	.447	.324	.444	.322
	bias f	.002	.002	.001	.001	.002	.002	.000	.000	.001	.001	.000	.000	.000	.000	.001	.000
	$\hat{f}_{T+k}$	1.246	1.146	2.072	1.966	2.017	1.948	1.968	1.926	1.928	1.903	1.895	1.881	1.870	1.864	1.856	1.852
	$MSE(\hat{f}_{T+k})$	.246	.199	.491	.406	.438	.396	.404	.381	.391	.370	.376	.355	.361	.347	.348	.343
	bias g			.001	.001	.002	.002	.000	.000	.001	.001	.000	.000	.000	.001	.000	.000
	$\hat{g}_{T+k}$			1.914	1.862	1.906	1.865	1.896	1.857	1.886	1.854	1.882	1.854	1.870	1.858	1.881	1.857
	$MSE(\hat{g}_{T+k})$			.465	.380	.430	.382	.436	.374	.411	.360	.415	.356	.390	.358	.389	.351
LS	$\hat{\sigma}_\eta^2(k)$	.940	.953	1.646	1.708	1.654	1.710	1.662	1.717	1.666	1.717	1.666	1.713	1.673	1.714	1.685	1.717
	$MSE(\hat{\sigma}_\eta^2(k))$	.265	.175	.464	.319	.456	.321	.450	.320	.444	.321	.448	.326	.454	.324	.453	.325
	bias f	.002	.002	.001	.001	.002	.002	.000	.000	.001	.001	.000	.000	.000	.000	.001	.000
	$\hat{f}_{T+k}$	1.255	1.148	2.084	1.970	2.028	1.952	1.978	1.929	1.935	1.905	1.903	1.883	1.878	1.866	1.863	1.854
	$MSE(\hat{f}_{T+k})$	.264	.201	.511	.408	.456	.399	.417	.384	.399	.372	.382	.356	.366	.348	.354	.344
	bias g			.001	.001	.002	.002	.000	.000	.001	.001	.000	.000	.000	.001	.000	.000
	$\hat{g}_{T+k}$			1.919	1.863	1.908	1.867	1.899	1.859	1.890	1.855	1.887	1.854	1.886	1.858	1.885	1.857
	$MSE(\hat{g}_{T+k})$			.473	.381	.430	.384	.424	.374	.421	.360	.425	.355	.401	.358	.394	.350

Table 2. Theoretical and simulation results of the MA(1) process

	k	1		2		3		4		5		6		7		8	
	T	50	100	50	100	50	100	50	100	50	100	50	100	50	100	50	100
	p	8	8	4	9	2	4	1	2	1	2	1	1	0	1	0	0
	$\sigma_\eta^2(k)$	1.055	1.055	2.720	2.513	5.919	5.717	8.069	7.764	9.140	8.945	9.826	9.826	11.044	10.265	11.044	11.044
	f_{T+k}	1.226	1.141	3.004	2.782	6.755	6.205	9.157	8.397	10.665	9.785	11.552	10.635	12.035	11.135	12.302	11.435
	g_{T+k}			2.997	2.513	6.450	5.717	8.570	7.764	9.850	8.945	10.771	9.826	11.044	10.265	11.044	11.044
	$\lvert\lambda\rvert_{min}$	1.126	1.126	.926	.849	.879	.819	1.178	.933	1.192	.993	1.202	1.202		1.209		
	$\bar{p}$	3.748	6.032	3.296	5.508	2.468	3.680	2.064	3.240	1.820	2.744	1.528	2.656	1.428	2.484	1.272	2.312
	$\bar{p}_{mod}$	3	6	2	6	2	2	1	2	1	1	0	1	0	0	0	0
YW	$\hat{\sigma}_\eta^2(k)$	1.471	1.248	3.088	2.789	5.330	5.390	6.556	6.883	7.199	7.769	7.590	8.236	7.765	8.512	7.876	8.668
	$MSE(\hat{\sigma}_\eta^2(k))$	.713	.368	1.190	.724	1.967	1.222	2.947	1.924	3.602	2.420	4.047	2.863	4.861	3.121	4.874	3.590
	bias f	-.001	-.001	-.003	-.003	-.002	-.002	-.006	-.001	-.010	.002	-.013	.006	-.013	.010	-.010	.014
	$\hat{f}_{T+k}$	1.489	1.225	3.532	2.904	6.983	6.111	8.980	8.046	10.069	9.239	10.470	9.902	10.580	10.264	10.572	10.463
	$MSE(\hat{f}_{T+k})$	.572	.266	1.654	.742	3.194	1.613	4.163	2.423	4.447	3.080	4.382	3.570	4.351	3.887	4.378	4.109
	bias g			-.009	0.000	.001	.003	-.004	.014	-.004	.023	0.000	.036	-.004	.041	-.005	.037
	$\hat{g}_{T+k}$			3.534	2.933	7.085	6.154	9.116	8.053	10.250	9.209	10.721	9.917	10.929	10.344	10.843	10.582
	$MSE(\hat{g}_{T+k})$			1.541	.776	3.178	1.664	3.996	2.415	4.527	2.976	4.627	3.385	4.667	3.798	4.280	4.057
QML	$\hat{\sigma}_\eta^2(k)$	1.078	1.001	2.842	2.513	5.587	5.662	6.909	7.214	7.593	8.122	8.044	8.627	8.240	8.901	8.378	9.052
	$MSE(\hat{\sigma}_\eta^2(k))$	.327	.182	1.048	.604	1.958	1.260	2.897	1.903	3.508	2.404	3.892	2.830	4.669	3.066	4.754	3.460
	bias f	-.007	-.001	-.016	-.004	-.024	-.006	-.031	-.007	-.036	-.007	-.035	-.004	-.030	.001	-.022	.007
	$\hat{f}_{T+k}$	1.404	1.204	3.244	2.859	6.698	6.188	8.805	8.206	10.102	9.454	10.713	10.150	10.972	10.539	11.027	10.752
	$MSE(\hat{f}_{T+k})$	.384	.233	.904	.654	1.895	1.501	2.806	2.283	3.474	2.915	3.882	3.374	4.133	3.690	4.304	3.946
	bias g			-.028	-.005	-.028	-.002	-.036	.006	-.033	.017	-.037	.021	-.038	.016	-.034	.012
	$\hat{g}_{T+k}$			3.512	2.956	6.793	6.320	8.813	8.109	10.019	9.267	10.557	9.938	10.939	10.399	11.025	10.636
	$MSE(\hat{g}_{T+k})$			1.089	.790	2.027	1.720	2.856	2.364	3.534	2.943	3.862	3.365	4.275	3.821	4.339	4.100
LS	$\hat{\sigma}_\eta^2(k)$	1.070	1.000	2.494	2.405	5.081	5.262	6.572	6.959	7.372	7.988	7.903	8.547	8.150	8.907	8.275	9.062
	$MSE(\hat{\sigma}_\eta^2(k))$	.326	.183	.808	.515	1.926	1.226	2.938	1.938	3.586	2.465	3.984	2.896	4.830	3.184	4.931	3.551
	bias f	-.008	-.001	-.017	-.004	-.025	-.005	-.031	-.006	-.035	-.005	-.034	-.002	-.029	.003	-.023	.008
	$\hat{f}_{T+k}$	1.409	1.207	3.257	2.863	6.720	6.194	8.843	8.212	10.156	9.463	10.793	10.171	11.073	10.571	11.154	10.797
	$MSE(\hat{f}_{T+k})$	.385	.233	.907	.651	1.905	1.499	2.837	2.286	3.529	2.931	3.980	3.407	4.261	3.734	4.465	3.997
	bias g			-.023	-.004	-.027	-.005	-.032	.002	-.032	.014	-.033	.018	-.036	.011	-.033	.007
	$\hat{g}_{T+k}$			3.234	2.882	6.691	6.111	8.824	8.021	10.129	9.253	10.753	9.969	11.160	10.424	11.244	10.660
	$MSE(\hat{g}_{T+k})$			.844	.661	1.900	1.513	2.815	2.208	3.554	2.843	3.976	3.314	4.455	3.763	4.614	4.022

Table 3. Theoretical and simulation results of the ARMA(1,2) process.

We now discuss the predictive performance of the various filters. It must be noted that the estimates are influenced by the order determination, so it might be expected that the variability of the estimates shall increase. More comparable results can be obtained by fixing the order. In almost all cases $\hat{\sigma}_\eta^2(k)$ is biased downwards. Some notable exceptions are Y-W estimates of the AR(1) and the ARMA(1,2) process for small k. For the strongly autocorrelated processes the bias becomes larger for increasing time horizons and the Y-W estimates are more biased than the QML and LS estimates. In general we can see that for the AR(1) and the ARMA(1,2) process f_{T+k}) and g_{T+k} are overestimated for small k, whereas the error variances are underestimated for increasing k. For the correctly specified AR(1) process $\hat{f}_{T+k}$ is less than $\hat{g}_{T+k}$ for all time horizons and for the MA(1) process in nearly all cases the converse holds true, which is in accordance with the theoretical results. For the ARMA(1,2) process however $\hat{f}_{T+k} < \hat{g}_{T+k}$ for the Y-W estimates, $\hat{f}_{T+k} > \hat{g}_{T+k}$ for the LS estimates and the method of QML gives mixed results.

The most important feature is that for all processes the difference $\hat{f}_{T+k} - \hat{g}_{T+k}$ is biased downwards. This implies that although for misspecified models multi-step prediction filters are theoretically better, it is hardly to detect in practice. For the AR(1) process especially the multi-step filter produces more badly results than expected.

A comparison of the different estimation methods leads to the following conclusions. First, there is little difference between QML and LS, except for those filters for which one or more roots lay inside the unit circle. Secondly, QML and LS estimate the parameters $(\alpha_1^{(k)},\ldots,\alpha_p^{(k)},\sigma_\eta^2(p,k))$ more accurate than Y-W, but the differences between the forecast variances are less pronounced. By comparing the influence of the sample size, we see that the bias of the estimate $\hat{\sigma}_\eta^2(k)$ becomes less if T increases. The prediction error variances of the MA(1) process are more accurate for $T = 100$, but for the other two processes there are mixed results.

VI In this paper we have examined the properties of the various prediction filters for stationary time series. By using different models for the various time horizons the forecast error variance can be reduced if the 1-step filter is misspecified, although the simulation results show that the reduction is less than theoretically expected. We suspect that the order determination and the method of estimation have influenced the results, so both issues require further investigation. Especially the evaluation of a likelihood function, which is also based on the predictive accuracy might be useful, although the analytical expressions can become quite complicated. It seems worthwhile to apply the proposed method to real economic data and to compare the results with results obtained by the conventional methods. However in the case of non-stationary series, special attention should be paid to the prefilter, which is used for detrending the series.

JK/AM/mt
860414

References

Akaike, H. (1974): 'A new look at the statistical model identification', IEEE Transactions on Automatic Control, 19, 716-723.

Anderson, T.W. (1971): 'The Statistical Analysis of Time Series', New York, Wiley.

Ansley, C.F., and P. Newbold (1980): 'Finite sample properties of estimators for autoregressive moving average models', Journal of Econometrics, 13, 159-183.

Ansley, C.F., and P. Newbold (1981): 'On the bias in estimates of forecast mean squared error', Journal of the American Statistical Association, 76, 569-578.

Beamish, N., and M.B. Priestley (1981): 'A study of autoregressive and window spectral estimation', Applied Statistics, 30, 41-58.

Bloomfield, P. (1972): 'On the error of prediction of a time series', Biometrika, 59, 501-507.

Davies, N., and P. Newbold (1980): 'Forecasting with misspecified models', Applied Statistics, 29, 87-92.

Findley, D.F. (1984): 'On some ambiguities associated with the fitting of ARMA models to time series', Journal of Time Series Analysis, 5, 213-225.

Gersch, W., and G. Kitagawa (1983): 'The prediction of time series with trends and seasonalities', Journal of Business & Economic Statistics, 1, 253-264.

McClave, J. (1975): 'Subset autoregression', Technometrics, 17, 213-220.

Meese, R., and J. Geweke (1984): 'A comparison of autoregressive univariate forecasting procedures for macroeconomic time series', Journal of Business & Economic Statistics, 2, 191-200.

Pagano, M. (1972): 'An algorithm for fitting autoregressive schemes', Applied Statistics, 21, 274-281.

Paulsen, J., and D. Tjøstheim (1985): 'On the estimation of residual variance and order in autoregressive time series', Journal of the Royal Statistical Society, Series B, 47, 216-228.

Reinsel, G. (1980): 'Asymptotic properties of prediction errors for the multivariate autoregressive model using estimated parameters', Journal of the Royal Statistical Society, Series B, 42, 328-333.

Shibata, R. (1976): 'Selection of the order of an autoregressive model by Akaike's information criterion', Biometrika, 63, 117-126.

Shibata, R. (1980): 'Asymptotically efficient selection of the order of the model for estimating parameters of a linear process', The Annals of Statistics, 8, 147-164.

Sneek, J.M. (1984): 'Modelling Procedures for Univariate Economic Time Series', Amsterdam, Free University Press.

White, H. (1982): 'Maximum likelihood estimation of misspecified models', Econometrica, 50, 1-25.

Yamamoto, T. (1976): 'Asymptotic mean square prediction error for an autoregressive model with estimated coefficients', Applied Statistics, 25, 123-127.

JK/AM/mt
860414