Michael A. Hauser and Wolfgang Hörmann and Robert M. Kunst and Jörg Lenneis

A Note on Generation, Estimation and Prediction of Stationary Processes

Paper

Original Citation:

Hauser, Michael A. and Hörmann, Wolfgang and Kunst, Robert M. and Lenneis, Jörg  
(1994)  
A Note on Generation, Estimation and Prediction of Stationary Processes.  
This version is available at: [https://epub.wu.ac.at/142/](https://epub.wu.ac.at/142/)  
Available in ePubWU: November 2005

ePubWU, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.
A Note on Generation, Estimation and Prediction of Stationary Processes

Michael A. Hauser, Wolfgang Hörman, Robert M. Kunst, Jörg Leneis

Department of Applied Statistics and Data Processing
Wirtschaftsuniversität Wien

Preprint Series
Preprint 9
June 1994

http://statmath.wu-wien.ac.at/
A Note on Generation, Estimation and Prediction of Stationary Processes

Michael A. Hauser¹, Wolfgang Hörmann¹, Robert M. Kunst² and Jörg Lenneis¹
¹University of Economics and Business Administration,
Department of Statistics, Vienna
²Institute for Advanced Studies, Vienna.

is published in:

COMPSTAT 1994, Proceedings in Computational Statistics,
Physica-Verlag, Heidelberg
pp. 323-328

Unfortunately there is an error in the printed version in equation (9).
The following

The variance of the linear predictor is, however,

\[ V[Y_t \mid Y_T] = L_{tt} H_{tt} L_{tt}^T \] (9)

with

\[ H_{T+t} = \begin{bmatrix} H_T & 0 \\ 0 & H_r \end{bmatrix}. \]

should be read as

The variance of the linear predictor is, however,

\[ V[Y_t \mid Y_T] = L_{tt} H_{t\mid Y_T} L_{tt}^T \] (9)

with

\[ H_{T+t} = \begin{bmatrix} H_T & 0 \\ 0 & H_r \end{bmatrix} \] and \[ H_{t\mid Y_T} = E[H_{t\mid Y_T}]. \]

The enclosed copy is corrected for that error.
A Note on Generation, Estimation and Prediction of Stationary Processes

Michael A. Hauser¹, Wolfgang Hörmann¹, Robert M. Kunst² and Jörg Lenneis¹
¹University of Economics and Business Administration, Department of Statistics, Vienna
²Institute for Advanced Studies, Vienna.

Keywords. Cholesky decomposition, Toeplitz matrices, fractional integration

1 Introduction

Some recently discussed stationary processes like fractionally integrated processes cannot be described by low order autoregressive or moving average (ARMA) models rendering the common algorithms for generation estimation and prediction partly very misleading [cf. Hosking (1981,1984), Sowell (1992), Ray (1993)]. We offer an unified approach based on the Cholesky decomposition of the covariance matrix which makes these problems exactly solvable in an efficient way.

Our starting point are stationary processes with a Wold representation of the form
\[ y_t - \mu = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \tag{1} \]
where \( \epsilon_t \) is uncorrelated noise with mean zero. The \( \psi_i \) are quadratic summable and the (unconditional) variance of the noise, \( \sigma^2 \), is greater than zero. We assume for simplicity of the presentation that \( \mu = 0 \). \( Y_T \) denotes the vector \( (y_1, \ldots, y_T)' \) and \( E_T = (\epsilon_1, \ldots, \epsilon_T)' \). The covariance matrix of \( Y_T, \Sigma_T \), is positive definite, symmetric and Toeplitz, and thus persymmetric. It may be factorized according to the Cholesky decomposition.
\[ \Sigma_T = L_T L_T'. \tag{2} \]

\( L_T \) is a lower triangular matrix.

One possibility for the generation of a sample of length \( T \) of a given process which possesses exactly the same covariance structure is to use the relation
\[ Y_T = L_T E_T. \tag{3} \]

Under the assumption of normal distributed noise estimation may be performed by maximizing the Gaussian likelihood
\[ f(Y_T; \mu, \Sigma_T) = (2\pi)^{-T/2} |\Sigma_T|^{-1/2} \exp[-(Y_T - \mu)'\Sigma_T^{-1}(Y_T - \mu)/2]. \tag{4} \]
For ARMA models there exist computationally simpler presentations of the likelihood. For fractionally integrated models, however, this is the only known exact form [Li and McLeod (1986) or Sowell (1992)].

The implicit noise vector may be obtained by

$$E_T = L_T^{-1}Y_T.$$  \hspace{1cm} (5)

The linear prediction for one step to \( \tau \) steps ahead may simply be performed by extending the above equation to \( T + \tau \) and replacing the future noises by their expectation which is zero. This is

$$Y_{T+\tau} = \begin{bmatrix} Y_T \\ Y_\tau \end{bmatrix} = L_{T+\tau}E_{T+\tau} = \begin{bmatrix} L_T & 0 \\ L_{T\tau} & L_{\tau\tau} \end{bmatrix} \begin{bmatrix} E_T \\ E_\tau \end{bmatrix}$$

$$E[Y_{T+\tau} Y_T] = \begin{bmatrix} L_{T\tau} & L_{\tau\tau} \\ 0 & 0 \end{bmatrix} E_T = L_{\tau\tau}E_T.$$  \hspace{1cm} (6)

The variance of the linear forecast \( Y_T \) given \( Y_T, E_T \) respectively, is given by means of the covariance matrix \( \Sigma_{T+\tau} \), with \( \Sigma_{T+\tau} = L_{T+\tau}H_{T+\tau} \).

$$V[Y_{T+\tau} Y_T] = \text{E}[\{Y_T - E[Y_T Y_T]\}(Y_T - E[Y_T Y_T])'] = L_{\tau\tau}L_{\tau\tau}'. \hspace{1cm} (7)$$

If the innovations are conditional heteroscedastic and Gaussian - i.e. \( \alpha_t \) are uncorrelated and normal with non-constant variances, which depend on the past - the process likelihood is given by (4) by replacing the covariance matrix \( \Sigma_T \) by a process dependent covariance matrix [see Hauser and Kunst (1993)]

$$\Sigma_T = L_TH_TL_T'$$  \hspace{1cm} (8)

where \( H_T \) is diagonal and contains the conditional variances of the normalized \( \alpha_t \). In case of homoscedasticity the \( H_T \) matrix reduces to \( I_T \).

Generation and linear prediction is analogous to the homoscedastic case once the heteroscedastic innovations are given. The variance of the linear predictor is, however,

$$V[Y_{T+\tau} Y_T] = L_{\tau\tau}H_{\tau\tau}L_{\tau\tau}'$$  \hspace{1cm} (9)

with \( H_{T+\tau} = \begin{bmatrix} H_T & 0 \\ 0 & H_\tau \end{bmatrix} \) and \( H_{\tau\tau} = E[H_T Y_T] \).

The numerical problems addressed above can be summarized as follows: Generation and prediction require the calculation of the Cholesky factor, the inverse of the Cholesky factor, and the repeated multiplication of the Cholesky factor with an arbitrary vector. Estimation, i.e. the inversion of the covariance matrix, may be implemented by factorizing \( \Sigma_T^{-1} \) in a \( MDM' \), \( M \) a lower triangular matrix with ones in the diagonal, \( D \) a diagonal matrix, via the Levinson algorithm. The determinant of the covariance matrix is then equal \( |D| \). The calculation of the variance of the predictor may be obtained
by calculating only the lower right $\tau \times \tau$ part of the Cholesky matrix. How the necessary operations can be performed in an efficient way is discussed below.

2 The multiplication of the Cholesky factor with an arbitrary vector

Notation and some properties of Toeplitz matrices:

$$
\Sigma_{T+1} = \begin{bmatrix} \Sigma_T & \Sigma_T^2 \\ \Sigma_T & \Sigma_{11} \end{bmatrix},
L_{T+1} = \begin{bmatrix} L_T & 0 \\ L_{1T} & L_{11} \end{bmatrix},
R_{T+1} = \begin{bmatrix} R_T & E_T \\ (E_T^T)^T & 1 \end{bmatrix},
$$

$E$ is a square matrix with ones in the secondary diagonal and zeros else. It holds that $EE = I$, $E^{-1} = E$. $R_T$ is the correlation matrix, $\Sigma_T = \sigma_T^2 R_T$. It is symmetric and Toeplitz, so that $E R_T E = R_T$ and $E R_T = R_T E$ holds. $R_T^{-1}$ is also symmetric and persymmetric.

Lemma 1: [Brockwell and Davis(1991, p.168)]

The best linear 1-step ahead predictor of $y_{T+1}$ of $y_{T+1}$ in terms of $Y_T$ and its mean squared error are

$$
\hat{y}_{T+1} = \Sigma_{1T} \Sigma_T^{-1} Y_T, \quad v_T = \Sigma_{11} - \Sigma_{1T} \Sigma_T^{-1} \Sigma_{1T}.
$$

(10)

In case of multivariate normal distributed $Y_{T+1}$ this is identical to the moments given by the conditional normal distribution [Johnson(1987, p.50)].

The coefficient in front of $y_T$ may be interpreted as the $T$-th partial autocovariance.

Proposition 1:

The best linear 1-step ahead predictor of $y_{T+1}$ of $y_{T+1}$ in terms of the Cholesky factors and past innovation vector $E_T$ and its mean squared error are

$$
\hat{y}_{T+1} = L_{1T} E_T, \quad v_T = L_{11} L_{11}.
$$

(11)

Proof:

This may be easily seen by using $\Sigma_{T+1} = L_{T+1} L_{T+1}^T$ in the partition representation as given above, multiplying out, and replacing the $\Sigma$-matrices by the corresponding expressions in terms of the $L$-matrices in (10). For $Y_T$ use $Y_T = L_T E_T. \square$

The predictor is given by the multiplication of the last line of the Cholesky matrix by the vector $(E_T, 0)^T$.

For the generation of samples of a process with given true covariance matrix the best linear predictor can be easily used recursively in the following way
starting at \( T = 0 \) with \( v_0 = \sigma_y^2 \) [cf. Hosking(1984, p.1900)]:

\[
y_{T+1} = \Sigma_T \Sigma_T^{-1} Y_T + \sqrt{\tau_T} \epsilon_{T+1},
\]

(12)

where the \( \epsilon_t \) are an \( \text{(possibly heteroscedastic)} \) innovation sequence.

In notation of the Cholesky matrix this amounts to

\[
y_{T+1} = L_T E_T + L_{11} \epsilon_{T+1}.
\]

(13)

This is the multiplication of the last line of \( L_{T+1} \) with \( E_{T+1} \), or more compactly for the whole vector \( Y_{T+1} \), \( Y_{T+1} = L_{T+1} E_{T+1} \).

An efficient algorithm to compute the best linear predictor and its mean squared error is the Durbin-Levinson algorithm [Brockwell and Davis(1991, p.169)]. Thus the Durbin-Levinson algorithm does multiply the Cholesky matrix with the vector \( E_{T+1} \) by requiring \( O(T^2) \) flops and \( O(T) \) storage. More generally, this algorithm performs the multiplication of the Cholesky matrix of a symmetric Toeplitz matrix with any arbitrary vector. This is remarkable, since there is no procedure known for the simply structured Toeplitz matrices to compute the Cholesky matrix with less than \( O(T^3) \) flops and \( O(T^2) \) storage. Below we will give a derivation of an equivalent algorithm based on matrix computations and the use of the Durbin algorithm which solves the Yule-Walker equations [Golub and VanLoan(1989, p.185)].

**Derivation of the algorithm:**

The idea for the algorithm is identical to the first step of the recursion of the Trench algorithm as presented in Golub and VanLoan(1989, p.188). For simplicity we reformulate the problem in correlations instead of covariances, which implies \( \sigma_y^2 = 1 \), \( \Sigma_T = \sigma_y^2 R_T \) respectively. The first two moments of \( y_{T+1} \) as given in (10) simplify to \( (Er') R_T^{-1} Y_T \) and \( 1 - r' R_T^{-1} r \) using the properties of Toeplitz matrices and the matrix \( E \) and the notation given above.

\[
R_T^{-1} = \begin{bmatrix}
R_T & Er \\
(Er)' & 1
\end{bmatrix}^{-1} = \begin{bmatrix}
B & v \\
v' & \gamma
\end{bmatrix}.
\]

This implies that

\[
\begin{bmatrix}
R_T \\
(Er)
\end{bmatrix}^{-1} = \begin{bmatrix}
1 \\
1
\end{bmatrix}.
\]

Solving this system for \( v \) and \( \gamma \) yields \( R_T v = -\gamma Er \) from the first equation. So \( v \) can be expressed via the solution \( y \) of the Yule-Walker equations, \( R_T y = -r \), \( y = -R_T^{-1} r \) and \( v = \gamma Ey \).

By replacing \( r \) in the second equation \( \gamma \) can be expressed as \( \gamma = 1/(1 + r'y) = 1/(1 + r' R_T^{-1} r) \).

The first two moments of \( y_{T+1} \) can be then expressed in terms of \( y \). That is:

\[
(Er') R_T^{-1} Y_T = - (Ey') Y_T \quad \text{and} \quad 1 - r' R_T^{-1} r = 1/\gamma = 1 + r'y.
\]

\( \square \)
The algorithm gives the multiplication of the Cholesky factor with an arbitrary vector. The storage requirements are those of Durbin algorithm which are linear. Its number of flops are $O(T^2)$ which increases by two vector multiplications.

3 The inverse of the Cholesky factor

**Proposition 2:**
The inverse Cholesky matrix is related to the Cholesky matrix of the inverse by transposing with respect to the secondary diagonal.

**Proof:**
\[ \Sigma \text{ is positive definite, symmetric and persymmetric. The Cholesky decompositions of } \Sigma \text{ and its inverse, which is also symmetric and persymmetric, are } \Sigma = AA' \text{ and } \Sigma^{-1} = BB'. \]
The inversion of the first decomposition is $\Sigma^{-1} = (A')^{-1}A^{-1}$. $A, A^{-1}$ and $B$ are lower triangular matrices. So there is a lower triangular decomposition and an upper triangular decomposition of the same matrix.

\[ \Sigma^{-1} = E\Sigma^{-1}E = E(A')^{-1}EA^{-1}E = (E(A')^{-1}E)(EA^{-1}E) = BB'. \]
Since the Cholesky decomposition is well defined $EA^{-1}E = B'$ and, thus, $A^{-1} = EB'E$ follows. \[\square\]

4 Computations

As given above generation of samples of the process $(y_t)$ may be obtained efficiently in linear storage requirements, once the autocorrelation function is given. [For the calculation of the autocovariance function of fractional integrated processes see Sowell(1992).]

If the estimation is performed via the likelihood function given in (4) the Levinson algorithm [see Marple(1987, p.87)] may be used to calculate the Cholesky decomposition of $\Sigma_T^{-1} \Sigma_T^{-1} = MDM'$, and thus also the required determinant. This algorithm is $O(T^2)$ in storage and $O(T^2)$ in flops.

The resulting innovations may be calculated using the Cholesky decomposition of the last iteration of the optimization procedure, Proposition 2 and (5).

The linear 1- to $\tau$-step prediction (forecast) vector given $Y_T$ may be calculated via the (estimated) residual vector and (6) - linear in storage and quadratic in flops - using the (estimated) autocovariance function.

Especially in case of calculating the variance of the linear predictor, (7), Proposition 1 is very helpful since $\tau$ is typically small. Multiplying $L_{T+\tau}$ by a vector with zeros and a 1 in position $(T + j)$ picks out exactly the $(T + j)$-
th column which is the column $j$ in $L_{T\times T}$. Without storing the intermediate results of the multiplication of $L_{T\times T}$ with the first $T$ zeros the number of flops is $O(T^3)$. The storage is linear if the diagonal elements are needed only.

The procedure can be easily generalized for heteroscedastic innovations. The 1's have to be replaced by the square root of the conditional variances.

5 Summary

An efficient algorithm - $O(T)$ in storage and $O(T^2)$ in flops - for multiplying the Cholesky factor by an arbitrary vector is presented. It may be used for generation of linear processes, linear prediction and calculation of the predictor variance.

It is shown that the Cholesky factor of an inverse symmetric Toeplitz matrix is a simple function of the inverse Cholesky factor of the Toeplitz matrix itself. Thus, given the Cholesky factor of the inverse covariance matrix the noise vector may be easily obtained.

We have outlined that for the simulation of stationary processes, for estimation and prediction two different algorithms are sufficient: the Levinson algorithm for calculating the Cholesky decomposition of the inverse covariance matrix and the algorithm giving a multiplication of a vector with the Cholesky matrix of the covariance matrix. Moreover this way is also very efficient.

References


