Autoregressive and Rolling Moving Average Processes using the K-matrix with Discrete but Unequal Time Steps

Stephen P. Smith September, 2018

Abstract. The autoregressive and rolling moving average time series models are describe with discrete time steps that may be unequal. The standard time series is described, as well as a two-dimensional spatial process that is separable into two one-dimensional processes. The K-matrix representations for each of these are presented, which can then be subjected to standard matrix handling techniques.

1. Introduction

In this paper, autoregressive and rolling moving average (ARMA) processes that are typical to time-series models (see Hamilton 1994) are investigated with renewed interest so that the models can be fitted to the K-matrix formulation described by Smith (2001). Once in this formulation, standard matrix handling techniques can be utilized to permit estimation, prediction and restricted maximum likelihood where the details are already described by Smith (2001) and by Smith, Nikolic and Smith (2012). A reformulation based on discrete time with unequal time steps was sought, merely to provide an alternative to the continuous time formulations that are equally suitable. Section 2 describes the ARMA model, with its specifications. In Section 3, this model is represented as the K-matrix first as a time series, and then as a two-dimensional spatial processe involving two separable ARMA processes.

2. Formulation Based on Discrete Time Steps

Definitions

$$\mathbf{x}_t = \begin{bmatrix} x_t \\ x_{t-1} \\ x_{t-2} \end{bmatrix};\ \mathbf{B} = \begin{bmatrix} w_2 & w_1 & w \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix};\ \mathbf{w} = \begin{bmatrix} w_2 \\ w_1 \\ w \end{bmatrix}$$

$$\mathbf{z}_t = \begin{bmatrix} z_t \\ 0 \\ 0 \end{bmatrix}$$; where z_t is independent normal $(0,\sigma^2)$

The 3rd order autoregressive model in state-space form is provided by:

(1) $\mathbf{x}_t = \mathbf{B}\, \mathbf{x}_{t-1} + \mathbf{z}_t$

where Var{$\mathbf{x}_t$}=$\mathbf{V}$, a Toeplitz matrix, Cov{$\mathbf{x}_t$,$\mathbf{x}_{t-1}$}=$\mathbf{BV}$

$$Var\begin{Bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \end{Bmatrix} = \begin{bmatrix} \mathbf{V} & \mathbf{VB}^T \\ \mathbf{BV} & \mathbf{V} \end{bmatrix}$$

For n-th order autoregressive models define the following.

$$\mathbf{x}_t = \begin{bmatrix} x_t \\ x_{t-1} \\ \bullet \\ \bullet \\ \bullet \\ x_{n-1} \end{bmatrix}; \mathbf{B} = \begin{bmatrix} \vec{\mathbf{w}}^T & w \\ \mathbf{I}_{n-1\times n-1} & \mathbf{0}_{n-1\times 1} \end{bmatrix}; \mathbf{w} = \begin{bmatrix} \vec{\mathbf{w}}_{n-1\times 1} \\ w \end{bmatrix}; \mathbf{z}_t = \begin{bmatrix} z_t \\ \mathbf{0}_{n-1\times 1} \end{bmatrix}$$

Equation (1) applies as before.

Because $\mathbf{V}$ is Toeplitz the following holds.

$$\mathbf{V} = \begin{bmatrix} \overline{\mathbf{V}} & \mathbf{v} \\ \mathbf{v}^T & v \end{bmatrix} = \begin{bmatrix} v & \mathbf{v}^T \\ \mathbf{v} & \overline{\mathbf{V}} \end{bmatrix}$$

Equation (1) implies that

(2) $\mathbf{V}=\mathbf{BVB}^T$ + Var{$\mathbf{z}_t$}

$$\mathbf{BVB}^T = \begin{bmatrix} \vec{\mathbf{w}}^T & w \\ \mathbf{I} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \overline{\mathbf{V}} & \mathbf{v} \\ \mathbf{v}^T & v \end{bmatrix}\mathbf{B}^T = \begin{bmatrix} \vec{\mathbf{w}}^T\overline{\mathbf{V}} + w\mathbf{v}^T & \vec{\mathbf{w}}^T\mathbf{v} + wv \\ \overline{\mathbf{V}} & \mathbf{v} \end{bmatrix}\begin{bmatrix} \vec{\mathbf{w}} & \mathbf{I} \\ w & \mathbf{0}^T \end{bmatrix}$$

$$= \begin{bmatrix} \vec{\mathbf{w}}^T \overline{\mathbf{V}} \vec{\mathbf{w}} + w\mathbf{v}^T \vec{\mathbf{w}} + w(\vec{\mathbf{w}}^T \mathbf{v} + wv) & \vec{\mathbf{w}}^T \overline{\mathbf{V}} + w\mathbf{v}^T \\ \overline{\mathbf{V}} \vec{\mathbf{w}} + \mathbf{v}w & \overline{\mathbf{V}} \end{bmatrix}$$

$$= \begin{bmatrix} \vec{\mathbf{w}}^T \overline{\mathbf{V}} \vec{\mathbf{w}} + 2w\mathbf{v}^T \vec{\mathbf{w}} + w^2 v & \vec{\mathbf{w}}^T \overline{\mathbf{V}} + w\mathbf{v}^T \\ \overline{\mathbf{V}} \vec{\mathbf{w}} + \mathbf{v}w & \overline{\mathbf{V}} \end{bmatrix}$$

Equation (2) implies that

(3) $\vec{\mathbf{w}} = (1-w)\,\overline{\mathbf{V}}^{-1}\mathbf{v}$

which can be plugged into the leading diagonal of $\mathbf{BVB}^T$, then (2) also implies the following:

$$\left[(1-w)^2 + 2w(1-w)\right]\mathbf{v}^T \overline{\mathbf{V}}^{-1}\mathbf{v} + (w^2-1)v + \sigma^2 = 0$$

This simplifies to the following, and is readily solved:

$$(1-w^2)\left[\mathbf{v}^T \overline{\mathbf{V}}^{-1}\mathbf{v} - v\right] + \sigma^2 = 0$$

The negative solution is,

$$w = -\sqrt{1 + \frac{\sigma^2}{v - \mathbf{v}^T \overline{\mathbf{V}}^{-1}\mathbf{v}}}$$

and this can be plugged back into equations (3); as well as the positive solution. Recognize that (2) is a linear function in the coefficients of **V**, and therefore, its also possible in principle to solve for **V** in terms of **B** and σ^2.

In general, the variance-covariance structure is provided by the following.

(4) $$Var\begin{Bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \mathbf{x}_{t-3} \end{Bmatrix} = \begin{bmatrix} \mathbf{V} & \mathbf{VB}^T & \mathbf{VB}^{2T} & \mathbf{VB}^{3T} \\ \mathbf{BV} & \mathbf{V} & \mathbf{VB}^T & \mathbf{VB}^{2T} \\ \mathbf{B}^2\mathbf{V} & \mathbf{BV} & \mathbf{V} & \mathbf{VB}^T \\ \mathbf{B}^3\mathbf{V} & \mathbf{B}^2\mathbf{V} & \mathbf{BV} & \mathbf{V} \end{bmatrix}$$

The matrix given by (4) is singular when time intervals are equal given the redundant use of the processes x_t (found in $\mathbf{x}_t$, $\mathbf{x}_{t+1}$ and $\mathbf{x}_{t+2}$ for a 3rd order process) in the standard formulation. Such singularities are non-problematic for K-matrix applications.

In particular, then

$\mathbf{x}_t = \mathbf{B}\,\mathbf{x}_{t-1} + \mathbf{z}_t = \mathbf{B}^2\,\mathbf{x}_{t-2} + \mathbf{B}\mathbf{z}_{t-1} + \mathbf{z}_t = \mathbf{B}^3\,\mathbf{x}_{t-3} + \mathbf{B}^2\mathbf{z}_{t-2} + \mathbf{B}\mathbf{z}_{t-1} + \mathbf{z}_t$

Define

$\mathbf{r}_t = \mathbf{B}^2\mathbf{z}_{t-2} + \mathbf{B}\mathbf{z}_{t-1} + \mathbf{z}_t$, then

$\mathbf{x}_t = \mathbf{B}^3\,\mathbf{x}_{t-3} + \mathbf{r}_t$

$\mathrm{Var}\{\mathbf{r}_t\} = \mathbf{B}^3\mathbf{D}\mathbf{B}^{3\,T} + \mathbf{B}^2\mathbf{D}\mathbf{B}^{2\,T} + \mathbf{B}\mathbf{D}\mathbf{B}^T + \mathbf{D}$

where $\mathrm{Var}\{\mathbf{z}_t\} = \mathbf{D}$, a diagonal matrix, but with only one non-zero diagonal.

In general, the following applies for N steps back.

$$\text{(5)} \qquad \mathbf{x}_t = \mathbf{B}^N \mathbf{x}_{t-N} + \mathbf{r}_t$$

$$\text{(6)} \qquad Var\{\mathbf{r}_t\} = \sum_{k=1}^{N} \mathbf{B}^k \mathbf{D} \mathbf{B}^{k\,T} \quad + \mathbf{D}$$

Equation (5), also implies that $\mathrm{Var}\{\mathbf{z}_t\} = \mathbf{V} - \mathbf{B}^N\mathbf{V}\mathbf{B}^{N\,T}$, whereas (6) provides an alternative that does not directly depend on **V**. This is enough to show how an autoregressive process can treat uneven time steps:

$$\text{(7)} \qquad \begin{aligned} &\mathbf{x}_{t(k)} = \mathbf{B}^{\Delta(k)}\mathbf{x}_{t(k-1)} + \mathbf{r}_{t(k)} \\ &Var\{\mathbf{r}_{t(k)}\} = \mathbf{V} - \mathbf{B}^{\Delta(k)}\mathbf{V}\mathbf{B}^{\Delta(k)\ T}, k > 1 \\ &Var\{\mathbf{x}_{t(1)}\} = \mathbf{V} \end{aligned}$$

where $\Delta(k)=t(k)-t(k-1)$, k=2, 3, ..., n. Because the processes are spaced evenly within each $\mathbf{x}_t$ alone, the process can easily be made into a rolling average process where g_t is an ARMA and $g_t=\mathbf{h}^T\mathbf{x}_t$ where **h** are the coefficients that define the rolling average.

What is needed is the calculation of the matrix power of **B**, which will necessitate (at least for continuous powers) an eigenvalue and eigenvector decomposition of **B** which may admit complex arithmetic. A univariate time series that represents a series of complex numbers has one serious formulation flaw if any of the hypothetical members of the series x_t represent observations, or possible observation, that can only be represented as real numbers when the model expects an imaginary component. This potential incoherence goes away when the time scale is adjusted such that $\Delta(k)$, $1 \le k \le n$, are always integers. Then $\mathbf{B}^{\Delta(k)}$ is always composed of real numbers. This reformulation of time redefines **V**, and therefore **B** changes accordingly. Moreover, changing the time scale in the direction of finer increments is moving closer to the stochastic differential equation described by Jones and Ackerson (1990). The stochastic differential equation is an equivalent model but comes with state-space formulation that's slightly different[1] and ends by needing a matrix exponential, rather than a matrix power, and also avoids the aforementioned incoherence.

3. Formulation of the K-matrix

3.1 Time Series

The state-space model, given by (7), requires an augmentation to connect with actual observations and to accommodate the rolling average formulation. This leads to (8) that depicts the set of observation equations.

(8) $$y_{t(k)} = \mathbf{f}_{t(k)}\mathbf{b} + \mathbf{h}\,\mathbf{x}_{t(x)} + \mathbf{e}_{t(k)}$$

where **b** is a column vector of fixed effect included here to expand generality, $\mathbf{f}_{t(k)}$ is a row vector that combines the fixed effect to depict the t(k)-th observation, **h** is a row vector that defines the rolling average process that applies over all time steps, and $\mathbf{e}_{t(k)}$ is a random residual at time t(k). The collection of error terms are assumed to be IID normal $(0,\sigma_e^2)$.

Following Smith (2001), the K-matrix is constructed directly from (7) and (8) as found below.

[1] As the unit time becomes finer, it becomes possible to approximate one set of parameters in terms of the other, with exact agreement achieved when the unit time represents the infinitesimal.

$$
\mathbf{K} = \left[\begin{array}{ccccccccccccccccc}
\sigma_e^2 & & & & & & & & & & \mathbf{f}_1 & \mathbf{h} & & & & & y_1 \\
& \mathbf{V} & & & & & & & & & & \mathbf{I} & & & & & \\
& & \sigma_e^2 & & & & & & & & \mathbf{f}_2 & & \mathbf{h} & & & & y_2 \\
& & & \mathbf{V}-\mathbf{B}^{\Delta 2}\mathbf{V}\mathbf{B}^{\Delta 2\,T} & & & & & & & & -\mathbf{B}^{\Delta 2} & \mathbf{I} & & & & \\
& & & & \sigma_e^2 & & & & & & \mathbf{f}_3 & & & \mathbf{h} & & & y_3 \\
& & & & & \mathbf{V}-\mathbf{B}^{\Delta 3}\mathbf{V}\mathbf{B}^{\Delta 3\,T} & & & & & & & -\mathbf{B}^{\Delta 3} & \mathbf{I} & & & \\
& & & & & & \bullet & & & & \bullet & & & & & & \bullet \\
& & & & & & & \bullet & & & \bullet & & & & & & \bullet \\
& & & & & & & & \sigma_e^2 & & \mathbf{f}_n & & & & & \mathbf{h} & y_n \\
& & & & & & & & & \mathbf{V}-\mathbf{B}^{\Delta n}\mathbf{V}\mathbf{B}^{\Delta n\,T} & & & & & -\mathbf{B}^{\Delta n} & \mathbf{I} & \\
\mathbf{f}_1^T & & \mathbf{f}_2^T & & \mathbf{f}_3^T & & \bullet & \bullet & \mathbf{f}_n^T & & & & & & & & \\
\mathbf{h}^T & \mathbf{I} & & -\mathbf{B}^{\Delta 2\,T} & & & & & & & & & & & & & \\
& & \mathbf{h}^T & \mathbf{I} & & -\mathbf{B}^{\Delta 3\,T} & & & & & & & & & & & \\
& & & & \mathbf{h}^T & \mathbf{I} & & & & & & & & & & & \\
& & & & & & & & & & & & & & & & \\
& & & & & & & & & -\mathbf{B}^{\Delta n\,T} & & & & & & & \\
& & & & & & & & \mathbf{h}^T & \mathbf{I} & & & & & & & \\
y_1 & & y_2 & & y_3 & & \bullet & \bullet & y_n & & & & & & & &
\end{array}\right]
$$

Smith (2001) provided the permutation matrix (**P**) that was needed for the K-matrix given the ARAM(1,1) model, such that **PKP**^{T} can be factored by Cholesky's decomposition in linear time. This led to parameter estimation by restricted maximum likelihood. For general applications represented by the K-matrix above, a similar permutation matrix can be found where the run time and space requirements are again found proportional to the number of time steps (n), even with unequal spacing. For multiple dimensions the analysis returns to vector processes and all the above can be generalized, and run time is again proportional to the number of time steps.

3.2 Separable Time Series as Spatial Model

Spatial models in two and three dimensions may use two or three ARMA models that are made separable over the spatial coordinates (Martin 1979). This can result in tremendous savings in run-time performance, when a sparse matrix application is used with the K-matrix; if only because unequal time steps is much less burdensome than with many equal time steps with missing information. The variance matrices and terms in the model (including the K-matrix) are all generalized by the matrix function, the Kronecker product, see Galecki (1994). The run time is approximately bounded below the number of space steps for each dimension, multiplied together, but is realizes as a matrix factorization that preserves sparsity. Gone is any need to define uniform spatial steps on a lattice as described by Smith (1997).

In two dimensions, define the matrix equations (8) and (9), that treats the spatial variability uniform in all directions; with C columns and R rows, and **x** and **e** are CR×1 column vectors, and the symbol, ⊗, is the Kronecker product.

(8) $$(\mathbf{I} - \mathbf{B}_c) \otimes (\mathbf{I} - \mathbf{B}_r)\, \mathbf{x} = (\mathbf{I} - \mathbf{B}_c \otimes \mathbf{B}_r)\, \mathbf{x} = \mathbf{e}$$

where

$$\mathbf{I} - \mathbf{B}_c = \begin{bmatrix} \mathbf{I} & & & & & & & \\ -\mathbf{B}^{\omega 2} & \mathbf{I} & & & & & & \\ & -\mathbf{B}^{\omega 3} & \mathbf{I} & & & & & \\ & & -\mathbf{B}^{\omega 4} & \mathbf{I} & & & & \\ & & & & \bullet & & & \\ & & & & & \bullet & & \\ & & & & & & \bullet & \\ & & & & & & -\mathbf{B}^{\omega C} & \mathbf{I} \end{bmatrix}$$

$$\mathbf{I} - \mathbf{B}_r = \begin{bmatrix} \mathbf{I} & & & & & & & \\ -\mathbf{B}^{\upsilon 2} & \mathbf{I} & & & & & & \\ & -\mathbf{B}^{\upsilon 3} & \mathbf{I} & & & & & \\ & & -\mathbf{B}^{\upsilon 4} & \mathbf{I} & & & & \\ & & & & \bullet & & & \\ & & & & & \bullet & & \\ & & & & & & \bullet & \\ & & & & & & -\mathbf{B}^{\upsilon R} & \mathbf{I} \end{bmatrix}$$

Where ω2, ω3, ω4, ..., ωC, are the distances between C spatial grid lines representing columns, and υ2, υ3, υ4, ..., υR, are the distances between R spatial grid lines representing rows. The variance matrix of **e**, or Var(**e**), is given by Var(**e**)=$\mathbf{V}_c \otimes \mathbf{V}_r$, where

$$\mathbf{V}_c = \begin{bmatrix} \mathbf{V} & & & & & & \\ & \mathbf{V}-\mathbf{B}^{\omega 2}\mathbf{V}\mathbf{B}^{\omega 2\,T} & & & & & \\ & & \mathbf{V}-\mathbf{B}^{\omega 3}\mathbf{V}\mathbf{B}^{\omega 3\,T} & & & & \\ & & & \bullet & & & \\ & & & & \bullet & & \\ & & & & & \bullet & \\ & & & & & & \mathbf{V}-\mathbf{B}^{\omega C}\mathbf{V}\mathbf{B}^{\omega C\,T} \end{bmatrix}$$

$$\mathbf{V}_R = \begin{bmatrix} \mathbf{V} & & & & & & \\ & \mathbf{V}-\mathbf{B}^{\nu 2}\mathbf{V}\mathbf{B}^{\nu 2\,T} & & & & & \\ & & \mathbf{V}-\mathbf{B}^{\nu 3}\mathbf{V}\mathbf{B}^{\nu 3\,T} & & & & \\ & & & \bullet & & & \\ & & & & \bullet & & \\ & & & & & \bullet & \\ & & & & & & \mathbf{V}-\mathbf{B}^{\nu R}\mathbf{V}\mathbf{B}^{\nu R\,T} \end{bmatrix}$$

To describe the set of observation equations, first consider the artificial situation where all C×R grid points have reordered observations given by the vector **ÿ**. The statistical model is represented by **ÿ**= $\mathbf{Z}_c \otimes \mathbf{Z}_r$ **x** + ř, where ř is a column vector of random residuals, and where $\mathbf{Z}_c$ is a C×Cm matrix and $\mathbf{Z}_r$ is a R×RCm matrix and m is the number of parameters in a 1×m row vector **h** representing the rolling average processes. Both $\mathbf{Z}_c$ and $\mathbf{Z}_r$ have the following form.

$$\mathbf{Z}_c \; or \; \mathbf{Z}_r \equiv \begin{bmatrix} \mathbf{h} & & & & & \\ & \mathbf{h} & & & & \\ & & \mathbf{h} & & & \\ & & & \bullet & & \\ & & & & \bullet & \\ & & & & & \mathbf{h} \end{bmatrix}$$

In practice, not all of $\ddot{\mathbf{y}}$ will be observed, rather a subset of these, **y**, will be observed. Therefore, the observation equations are given by (9).

(9) $$\mathbf{y}=\mathbf{Z}\mathbf{x} + \mathbf{r}$$

where **Z** represents the selected rows of $\mathbf{Z}_c \otimes \mathbf{Z}_r$ that are included in the model, and Var(**r**)=**R**. Combining (9) and (8) returns the following K-matrix.

$$\mathbf{K} = \begin{bmatrix} \mathbf{R} & & \mathbf{Z} & \mathbf{y} \\ & \mathbf{V}_c \otimes \mathbf{V}_r & (\mathbf{I}-\mathbf{B}_c)\otimes(\mathbf{I}-\mathbf{B}_r) & \\ \mathbf{Z}^T & (\mathbf{I}-\mathbf{B}_c^T)\otimes(\mathbf{I}-\mathbf{B}_r^T) & & \\ \mathbf{y}^T & & & \end{bmatrix}$$

As was done with time series, a continuous space formulation based on stochastic differential equations can be developed (Jones and Vecchia 1993). The main difference is to substitute the matrix exponential with the matrix power, with the K-matrix formulation remaining.

It is also important to note that the variance matrix, **V**, represents the grounding variance matrix for each of two separable processes, rather than the grounding variance matrix, **W**, for the spatial process. Nevertheless, the two interpretations are easy to adjust for the default case of **h**, with the first entry 1 and 0 elsewhere, thereby transforming one Toeplitz matrix into the other:

$$\mathbf{V} = \frac{1}{\sqrt{W_{11}}}\mathbf{W}$$

References

Galecki, A.T., 1994, General class correlation strutures for two or more repeated factors in logitudinal data analysis, *Communication In Statistics-Theory and Methods*, 23, 3105-3119.

Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press. Princeton, New Jersey.

Jones, R.H., and L.M. Ackerson, 1990, Serial correlation in unequally spaced longitudinal data, *Biometrika*, 77, 4, 721-731

Jones, R.H., and A.V. Vecchia, 1993, Fitting continuous ARMA models to unequally spaced spatial data, *Journal of the American Statistical Association*, 88, 423, 947-954.

Martin, R., 1979, A subclass of lattices processes applied to a problem in planar sampling, *Biometrika*, 66. 209-217.

Smith, S.P, 1997, Sparse matrix tools for Gaussian models on lattices, *Computational Statistics & Data Analysis*, 26, 1-15.

Smith, S.P., 2001, Likelihood-based analysis of linear state-space models using the Cholesky decomposition, *Journal of Computational and Graphical Statistics*, 10, 2, 350-369.

Smith, J.R., M. Nikolic and S.P. Smith, 2012, Hunting the Higgs Boson using the Cholesky Decomposition of an Indefinite Matrix, Memo, vixRa archived in High Energy Particle Physics.