Estimating Variances and Covariances in a Non-stationary Multivariate Time Series Using the K-matrix

Stephen P Smith, January 2019

Abstract. A second order time series model is described, and generalized to the multivariate situation. The model is highly flexible, and is suitable for non-parametric regression, coming with unequal time steps. The resulting K-matrix is described, leading to its possible factorization and differentiation using general purpose software that was recently developed. This makes it possible to estimate variance matrices in the multivariate model corresponding the signal and noise components of the model, by restricted maximum likelihood. A nested iteration algorithm is presented for conducting the maximization, and an illustration of the methods are demonstrated on a 4-variate time series with 89 observations.

1. Introduction

Multivariate time series analysis is well described in Lütkepohl (1993). Restricted maximum likelihood (REML) estimation of variances and covariances has also been worked out satisfactory for multiple records in models with equal design matrices and with one random effect in addition to the random residue (Meyer 1985, Taylor et al. 1985, Jensen and Mao 1988).

Smith (2001) describe state-space models in the context of a symmetric and indefinite matrix called the K-matrix that can be subjected to a Cholesky decomposition, even though it was not customary to factorize such a matrix with this tool. Not to dismiss any of the previous methods, the goal of the present paper is to specifically utilize the K-matrix with a particular non-stationary multivariate time series. The non-stationary model is found in Appendix A of Smith (2018b) where attention was given to non-parametric regression over a two-dimensional spatial lattice. With each dimension given a time-like orientation, the model is suitable with unequal time or space steps. Rather than testing the model as a spatial model, however, the present paper evaluates its applicability with a multivariate times series that comes with variance-covariance matrices representing signal and noise effects. Because the model accommodates unequal time steps, the approach competes with Smith's (1997, Section 5) illustrative model that depicts a univariate non-parametric regression for a one-dimensional lattice but with equal steps. The approach competes with, even makes obsolete, the stochastic spline function described by Smith (2018c) that was intended to treat unequal time steps.

The non-stationary model is described in Section 2. Its development from a univariate 2nd order stochastic differential equation is presented in Section 2.1, and its extension to

the multivariate case in presented in Section 2.2. The construction of the K-matrix is presented in Section 3. A nested iteration algorithm for estimating variances and covariances by REML is introduced in Section 4. In Section 5, the application of these tools are illustrated with four time series that coincide with yearly measurements.

2. Model Specifications

2.1 Univariate Second Order State-space model

The 2nd order stochastic differential equation $u(t)'' = \lambda \times \xi(t)$, where $\xi(t)$ represents white noise at time t, and its solution, are described in Appendix A of Smith(2018b). The solution is given in a 2×1 state space vector containing a Wiener process $u(t)' = W(t)$ and u(t), as indicated below.

$$W(t) = W(t_o) + \varepsilon_1$$
$$u(t) = u(t_o) + \Delta \times W(t_o) + \varepsilon_2$$

where

$$\Delta = t - t_o$$
$$E\{\varepsilon_1\} = E\{\varepsilon_2\} = 0$$
$$Var\begin{Bmatrix}\varepsilon_1 \\ \varepsilon_2\end{Bmatrix} = \lambda \times \begin{bmatrix}\Delta & \frac{\Delta^2}{2} \\ \frac{\Delta^2}{2} & \frac{\Delta^3}{3}\end{bmatrix}$$

The errors, ϵ_1 and ϵ_2 are multivariate normal, and λ is a baseline variance parameter for the bivariate process. The state-space equations can be applied over time intervals that may be unequal, where $\Delta_i = t_i - t_{i-1}$ for i=1, 2, ...N. In matrix notation, the state-space equations are given by the following[1].

[1] The function vech() that is introduced essentially takes the columns of a matrix and stacks them one on top of the other into a vector. See Harville (1997, Section 16.4) for more details.

$$\mathbf{0} = \mathbf{H}\mathbf{u} + \mathbf{r} \qquad (1)$$

where

$$\mathbf{u} = vech(\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_N)$$

$$\mathbf{u}_i = \begin{pmatrix} W(t_i) \\ u(t_i) \end{pmatrix}, \quad i = 0,1,\ldots N$$

$$\mathbf{r} = vech(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$$

$$\mathbf{r}_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix}, \quad i = 1,2,\ldots,N$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{C}_1 & -\mathbf{I}_{2\times 2} & & & \\ & \mathbf{C}_2 & -\mathbf{I}_{2\times 2} & & \\ & & \bullet & \bullet & \\ & & & \mathbf{C}_N & -\mathbf{I}_{2\times 2} \end{bmatrix}$$

$$\mathbf{C}_i = \begin{bmatrix} 1 & \\ \Delta_i & 1 \end{bmatrix}, \quad i = 1,2,\ldots,N$$

Equation (1) represents 2N equations in 2N+2 unknowns. The variance of **r** is the following block diagonal matrix or order 2N.

$$Var\{\mathbf{r}\} = \lambda \times \mathbf{V}$$

where

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & & & \\ & \mathbf{V}_2 & & \\ & & \bullet & \\ & & & \mathbf{V}_N \end{bmatrix}$$

$$\mathbf{V}_i = \begin{bmatrix} \Delta_i & \frac{\Delta_i^2}{2} \\ \frac{\Delta_i^2}{2} & \frac{\Delta_i^3}{3} \end{bmatrix}, \quad i = 1,2,\ldots,N$$

The observational equations are given by system (2) involving the column vector **y**

containing N observations.[2]

$$\mathbf{y} = \mathbf{X}\mathbf{u} + \mathbf{e} \qquad (2)$$

where

$$\mathbf{X} = \mathbf{I}_{N \times N} \otimes (0 \quad 1)$$

$$Var\{\mathbf{e}\} = \sigma^2 \mathbf{I}_{N \times N}$$

Models (1) and (2) complete the univariate specification, and this is enough to build the corresponding K-matrix. However, it is first desired to generalize the specifications for multivariate models.

2.2 Multivariate Model

The following notation is needed to superimpose on the univariate model, (1) and (2), to permit representation of the multivariate model with M time series.

$$\tilde{\mathbf{y}}_i \equiv \mathbf{y}$$

$$\tilde{\mathbf{u}}_i \equiv \mathbf{u}$$

$$\tilde{\mathbf{e}}_i \equiv \mathbf{e}$$

$$\tilde{\mathbf{r}}_i \equiv \mathbf{r}$$

$$where\ i = 1,2,\ldots,M$$

One the left side of the above equivalence statements, the tilde signifies a multivariate time series and the subscript indicates which one. And this is required to avoid confusion with the univariate time series coming with symbols on the right side of each equivalence that relate directly to (1) and (2).

It is also very convenient to define the following rectangular matrices.

$$\tilde{\mathbf{Y}} = \left[\tilde{\mathbf{y}}_1, \quad \tilde{\mathbf{y}}_2, \quad \ldots, \quad \tilde{\mathbf{y}}_M\right]$$

$$\tilde{\mathbf{U}} = \left[\tilde{\mathbf{u}}_1, \quad \tilde{\mathbf{u}}_2, \quad \ldots, \quad \tilde{\mathbf{u}}_M\right]$$

$$\tilde{\mathbf{E}} = \left[\tilde{\mathbf{e}}_1, \quad \tilde{\mathbf{e}}_2, \quad \ldots, \quad \tilde{\mathbf{e}}_M\right]$$

$$\tilde{\mathbf{R}} = \left[\tilde{\mathbf{r}}_1, \quad \tilde{\mathbf{r}}_2, \quad \ldots, \quad \tilde{\mathbf{r}}_M\right]$$

---

[2] The symbol "⊗" introduced with (2) denotes the Kronecker product. See Harville (1997, Section 16.1) for more details.

These can be turned into talk vectors with the subscript removed, as indicated below.

$$\tilde{\mathbf{y}} = vech\left(\tilde{\mathbf{Y}}\right)$$
$$\tilde{\mathbf{u}} = vech\left(\tilde{\mathbf{U}}\right)$$
$$\tilde{\mathbf{e}} = vech\left(\tilde{\mathbf{E}}\right)$$
$$\tilde{\mathbf{r}} = vech(\tilde{\mathbf{R}})$$

Therefore, model (1) corresponds to the following.

$$\mathbf{0} = \left[\mathbf{I}_{M\times M} \otimes \mathbf{H}\right]\tilde{\mathbf{u}} + \tilde{\mathbf{r}} \qquad (3)$$
$$Var\left\{\tilde{\mathbf{r}}\right\} = \mathbf{G}_{M\times M} \otimes \mathbf{V}$$

Note that **H** and **V** have the same definition as in (1), and **G** is a M×M variance-covariance matrix that substitutes for λ in (1).

Likewise, model (2) corresponds to (4).

$$\tilde{\mathbf{y}} = \left[\mathbf{I}_{M\times M} \otimes \mathbf{X}\right]\tilde{\mathbf{u}} + \tilde{\mathbf{e}} \qquad (4)$$
$$Var\left\{\tilde{\mathbf{e}}\right\} = \mathbf{W}_{M\times M} \otimes \mathbf{I}_{N\times N}$$

Note that **X** has the identical definition given with (2), and **W** is a M×M variance-covariance matrix that substitutes for σ^2 in (2).

It is better to re-parameterize both **W** and **G**, in terms of diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ and the non-singular matrix **Q**, such that $\mathbf{QWQ}^T=\mathbf{D}_1$ and $\mathbf{QGQ}^T=\mathbf{D}_2$; this is referred to as the transformation to the canonical scale by Meyer (1985). This offers advantages because with **Q** specified, the multivariate system can be transformed and turned into M univariate systems. From equation (4) the transformation is indicated below.

$$
\begin{aligned}
\left[\mathbf{Q}\otimes\mathbf{I}_{N\times N}\right]\tilde{\mathbf{y}} &= \left[\mathbf{Q}\otimes\mathbf{I}_{N\times N}\right]\left[\mathbf{I}_{M\times M}\otimes\mathbf{X}\right]\tilde{\mathbf{u}}+\left[\mathbf{Q}\otimes\mathbf{I}_{N\times N}\right]\tilde{\mathbf{e}}\\
&=\left[\mathbf{Q}\otimes\mathbf{X}\right]\tilde{\mathbf{u}}+\left[\mathbf{Q}\otimes\mathbf{I}_{N\times N}\right]\tilde{\mathbf{e}}\\
&=\left[\mathbf{I}_{M\times M}\otimes\mathbf{X}\right]\left[\mathbf{Q}\otimes\mathbf{I}_{(2N+2)\times(2N+2)}\right]\tilde{\mathbf{u}}+\left[\mathbf{Q}\otimes\mathbf{I}_{N\times N}\right]\tilde{\mathbf{e}}\\
&=\left[\mathbf{I}_{M\times M}\otimes\mathbf{X}\right]\bar{\mathbf{u}}+\bar{\mathbf{e}} \qquad (5)
\end{aligned}
$$

where

$$
\bar{\mathbf{u}}=\left[\mathbf{Q}\otimes\mathbf{I}_{(2N+2)\times(2N+2)}\right]\tilde{\mathbf{u}}
$$

$$
\bar{\mathbf{e}}=\left[\mathbf{Q}\otimes\mathbf{I}_{N\times N}\right]\tilde{\mathbf{e}}
$$

$$
Var\{\bar{\mathbf{e}}\}=\left[\mathbf{Q}\otimes\mathbf{I}_{N\times N}\right]\left[\mathbf{W}\otimes\mathbf{I}_{N\times N}\right]\left[\mathbf{Q}^{T}\otimes\mathbf{I}_{N\times N}\right]=\mathbf{Q}\mathbf{W}\mathbf{Q}^{T}\otimes\mathbf{I}=\mathbf{D}_{1}\otimes\mathbf{I}
$$

From equation (3) the transformation is presented below.

$$
\begin{aligned}
\mathbf{0} &= \left[\mathbf{Q}\otimes\mathbf{I}_{2N\times 2N}\right]\mathbf{0}=\left[\mathbf{Q}\otimes\mathbf{I}_{2N\times 2N}\right]\left[\mathbf{I}_{M\times M}\otimes\mathbf{H}\right]\tilde{\mathbf{u}}+\left[\mathbf{Q}\otimes\mathbf{I}_{2N\times 2N}\right]\tilde{\mathbf{r}}\\
&=\left[\mathbf{Q}\otimes\mathbf{H}\right]\tilde{\mathbf{u}}+\left[\mathbf{Q}\otimes\mathbf{I}_{2N\times 2N}\right]\tilde{\mathbf{r}}\\
&=\left[\mathbf{I}_{M\times M}\otimes\mathbf{H}\right]\left[\mathbf{Q}\otimes\mathbf{I}_{(2N+2)\times(2N+2)}\right]\tilde{\mathbf{u}}+\left[\mathbf{Q}\otimes\mathbf{I}_{2N\times 2N}\right]\tilde{\mathbf{r}}\\
&=\left[\mathbf{I}_{M\times M}\otimes\mathbf{H}\right]\bar{\mathbf{u}}+\bar{\mathbf{r}} \qquad (6)
\end{aligned}
$$

where

$$
\bar{\mathbf{r}}=\left[\mathbf{Q}\otimes\mathbf{I}_{2N\times 2N}\right]\tilde{\mathbf{r}}
$$

$$
Var\{\bar{\mathbf{r}}\}=\left[\mathbf{Q}\otimes\mathbf{I}_{2N\times 2N}\right]\left[\mathbf{G}\otimes\mathbf{V}\right]\left[\mathbf{Q}^{T}\otimes\mathbf{I}_{2N\times 2N}\right]=\mathbf{Q}\mathbf{G}\mathbf{Q}^{T}\otimes\mathbf{V}=\mathbf{D}_{2}\otimes\mathbf{V}
$$

To estimate variances and covariance, a possible approach hinted at by Lin and Smith (1990) is to estimate **Q**, $\mathbf{D}_1$ and $\mathbf{D}_2$ directly from the likelihood given by the M univariate models provided by (5) and (6). However, this can only be a constrained optimization, as it is clear that **Q** must be non-singular. Moreover, the number of free parameters should[3] not exceed the number of free parameters in **W** and **G**, being M(M+1). A suitable set of constrains is given by (7), that has no change in the number of free parameters.

$$
\begin{aligned}
&|\mathbf{Q}|=1 \qquad (7)\\
&\mathbf{D}_{1}=\alpha\mathbf{I},\quad \alpha>0
\end{aligned}
$$

[3] In some cyclic assent algorithms, however, maximizing over an under determined set of parameters can be feasible, and come with good convergence properties, even if the solutions are not uniquely defined.

A more parsimonious parameterization is available, requiring **Q** to be an orthogonal matrix, indicated by the restriction (8), where the diagonals of $\mathbf{D}_1$ and $\mathbf{D}_2$ are free to vary in the positive space. This restriction corresponds to the case where it is possible to simultaneously diagonalize both **W** and **G** (Harville, 1997, Section 21.13), but this preposition is enforced during optimization. The advantage of using (8), rather than (7), is that it comes with a smaller number of free parameters to estimate, and is, therefore, more suitable with smaller data sets; plus there is a useful greedy algorithm available to carry out the optimization as describe in Section 4.

$$\mathbf{Q}^T\mathbf{Q}=\mathbf{Q}\mathbf{Q}^T=\mathbf{I} \qquad (8)$$

3. Building the K-matrix

The standard construction of the K-matrix is immediate from (3) and (4), and is presented below.

$$\mathbf{K}=\begin{bmatrix} \mathbf{G}_{M\times M}\otimes\mathbf{V} & & \mathbf{I}_{M\times M}\otimes\mathbf{H} & \\ & \mathbf{W}_{M\times M}\otimes\mathbf{I}_{N\times N} & \mathbf{I}_{M\times M}\otimes\mathbf{X} & \tilde{\mathbf{y}} \\ \mathbf{I}_{M\times M}\otimes\mathbf{H}^T & \mathbf{I}_{M\times M}\otimes\mathbf{X}^T & & \\ & \tilde{\mathbf{y}}^T & & \end{bmatrix}$$

While holding that last row and column of **K** fixed in the last position, a permutation matrix **P** can be found such that $\mathbf{PKP}^T=\mathbf{LDL}^T$, where **L** is lower triangular with positive diagonals and **D** is a diagonal matrix with diagonals 1 or -1. Because $\mathbf{PKP}^T$ is found nearly banded as outputted from general purpose software that computes **P**, the factorization of $\mathbf{PKP}^T$ requires only linear time in N. If **K** has order K, then the log-likelihood suitable for REML is given by (9), where L_{ii} is the i-th diagonal of **L**.

$$\log-L=-\frac{L_{KK}^2}{2}-\sum_{i<K}\log\left(L_{ii}\right) \qquad (9)$$

Finding **G** and **W** that maximizes of (9) is a very feasible goal, given the differentiation algorithms described in Smith (2018a). However, its better to reparameterize the model in terms of unknown parameters **Q**, $\mathbf{D}_1$ and $\mathbf{D}_2$, and evaluate the appropriate K-matrix for M univariate models given by (5) and (6). Moreover, its better to re-structure the likelihood function, and by implication the K-matrix, by concentrating σ^2 out of each univariate likelihood function; where σ^2 represents one of the diagonals of $\mathbf{D}_1$.

To define the suitable K-matrix, first consider one univariate time series defined from the i-th row of **Q**, or i-th column of $\mathbf{Q}^T$ denoted by $\mathbf{q}_i$. The linear transformation of the data is given by this post-multiplication, $\tilde{\mathbf{Y}}\mathbf{q}_i$.

Its better to represent this transformation implicitly in the K-matrix, however, effectively defining a linear model with M right-hand sides. With the i-th diagonal of $\mathbf{D}_1$, or σ^2, concentrated out of the likelihood function, that leaves a new parameter to replace the i-th diagonal of $\mathbf{D}_2$, or ρ as the ratio of i-th diagonal of $\mathbf{D}_2$ over σ^2. This K-matrix, suitable for (5) and (6), is presented below.

$$\mathbf{K} = \begin{bmatrix} \rho \times \mathbf{V} & & \mathbf{H} & \\ & \mathbf{I}_{N\times N} & \mathbf{X} & \tilde{\mathbf{Y}} \\ \mathbf{H}^T & \mathbf{X}^T & & \\ & \tilde{\mathbf{Y}}^T & & \end{bmatrix}$$

Now when a permutation matrix **P** is computed, such that $\mathbf{PKP}^T = \mathbf{LDL}^T$ where **L** is lower triangular with positive diagonals and **D** is a diagonal matrix with diagonals 1 or -1, the last M rows of columns of **K** remain fixed in the last positions. The last M rows and columns of **L**, represent a smaller lower triangular matrix denoted by $\tilde{\mathbf{L}}$ in the lower corner of **L**.

Given that two degrees of freedom are used to estimate $W(t_0)$ and $u(t_0)$, these being treated as fixed, the REML estimate of σ^2 is provided by the following.

$$\hat{\sigma}^2 = \frac{\mathbf{q}_i^T \tilde{\mathbf{L}} \tilde{\mathbf{L}}^T \mathbf{q}_i}{N-2} \qquad (10)$$

The concentrated log-likelihood, corresponding to $\mathbf{q}_i$, is given by:

$$\log - L = -\frac{(N-2)\log\left(\mathbf{q}_i^T \tilde{\mathbf{L}} \tilde{\mathbf{L}}^T \mathbf{q}_i\right)}{2} - \sum_{i \le K-M} \log(L_{ii}) \qquad (11)$$

With $\mathbf{q}_i$, held fixed, its easy to maximize (11) and estimate ρ given the derivative algorithms presented in Smith(2018a). To estimate $\mathbf{q}_i$, and the rest of **Q**, note that log-likelihoods in the form given by (11) will need to be added together over the index-i, and sided conditions like (7) or (8) become important which will have a collective impact on the rows of **Q**. Regarding side condition (8), a nested interaction is feasible as described next.

4. Nested Iteration

In this section a greedy algorithm is described that will involve a nested iteration that's maximizes (11) each step of the way, and therefore, its guarantee to converge to a local maximum if not a global maximum while enforcing (8).

The vector $\mathbf{q}_i$ belongs to a vector space that constitutes the span of a set of basis vectors making the columns of $\mathbf{B}$, such that $\mathbf{q}_i = \mathbf{B}\mathbf{s}$ for some vector $\mathbf{s}$. The basis vectors in $\mathbf{B}$ will change during iteration, and moreover, the number of columns of $\mathbf{B}$ will become progressively smaller. Therefore, a subscript is added to describe $\mathbf{B}_i$ as a set of basis vectors used during a i-th step in the algorithm.

Initially set $\mathbf{B}_1=\mathbf{I}_{M\times M}$, and i=j=1. The steps of the algorithm follow.

4.1 If j=1, set ρ to an initial value, usually ρ=1 is good. Otherwise, j≠1 and a nested iteration is progressing on changes of ρ. Reconstitute $\mathbf{L}$. The objective is to find $\mathbf{q}_i \in \mathbb{C}(\mathbf{B}_i)$[4] such that (11) is maximum where $\mathbf{q}_i^T\mathbf{q}_i = 1$. If i≠M, the maximum is found by diagonalizing the following matrix, i.e., finding Eigen values and orthogonal Eigen vectors.

$$\mathbf{B}_i^T\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^T\mathbf{B}_i = \mathbf{S}\Delta\mathbf{S}^T$$
$$\mathbf{S}\mathbf{S}^T = \mathbf{S}^T\mathbf{S}^T = \mathbf{I}$$

The matrix Δ is diagonal, and Eigen values are located on the diagonals. The maximum is found by location $\mathbf{s}_k$ corresponding to the column of $\mathbf{S}$ where the k-th diagonal of Δ is the smallest Eigen value, then setting $\mathbf{q}_i=\mathbf{B}_i\mathbf{s}_k$. If i=M, the case is degenerate, $\mathbf{B}_M$ is a column vector and set $\mathbf{q}_i=\mathbf{B}_M$.

4.2 Now continue to maximize (11) by holding $\mathbf{q}_i$ fixed and letting ρ change. A supervise Newton-Raphson iteration performs well, given the differentiation algorithms presented in Smith (2018a) and where $\mathbf{L}$ is reconstituted at each step. Occasionally update $\mathbf{q}_i$ for the current value of ρ, by setting j←j+1 and returning to step 4.1. If both ρ and $\mathbf{q}_i$ have converged, then the i-th step is complete, these two quantities and σ^2 estimated from (10) are saved as partial solutions. If i=M then the algorithm is done. Otherwise remove the k-th column of $\mathbf{S}$, i.e., removing $\mathbf{s}_k$ to get $\underline{\mathbf{S}}$, using the most current evaluation of $\mathbf{S}$. Then set $\mathbf{B}_{i+1}=\mathbf{B}_i\,\underline{\mathbf{S}}$, and set i←i+1 and j=1. Return to step 4.1.

Owing to the heuristic quality of the above algorithm, it is not surprising that other algorithms may do better at targeting a local maximum that is larger than what comes from the above. One such possibility is to replace the matrix diagonalization contained in step 4.1 with the following diagonalization but when i>1, once $\mathbf{L}$ is reconstituted.

---

[4] The function $\mathbb{C}(\mathbf{B})$ signifies the column space of $\mathbf{B}$.

$$\tilde{\mathbf{L}}\left\{\mathbf{I}-\tilde{\mathbf{L}}^T\tilde{\mathbf{B}}_i\left[\tilde{\mathbf{B}}_i^T\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T\tilde{\mathbf{B}}_i\right]^{-1}\tilde{\mathbf{B}}_i^T\tilde{\mathbf{L}}\right\}\tilde{\mathbf{L}}^T=\mathbf{S}\Delta\mathbf{S}^T \qquad (12)$$

$$\mathbf{S}\mathbf{S}^T=\mathbf{S}^T\mathbf{S}^T=\mathbf{I}$$

$$\tilde{\mathbf{B}}_i=\left[\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_{i-1}\right]$$

Note that $\tilde{\mathbf{B}}_i$ gets larger by one column with the incremental change i←i+1, whereas $\mathbf{B}_i$ gets smaller by one column. The matrix define by (12) will have i-1 zero Eigen values, and so when selecting $\mathbf{s}_k$ it is necessary to find the smallest Eigen value that is not zero. Then set $\mathbf{q}_i = \mathbf{s}_k$.

5. Example

For the purpose of illustration only, econometric data were collected for each year, beginning in 1929 and ending in 2017. There were no missing observations, making N=89. The data has equal time steps by year, even as the model does not require this. The data represented four time series (M=4): real GDP provided by *The Balance*; US national debt and federal tax receipts provided by the *Office of Management and Budget*; and the US population size provided the *US Census Bureau*.

The four times series came in different units of measure, as indicated by Table 1. A pictorial view of the 4 time series is presented by Figure 1, with the x-axis representing years but leaving the y-axis arbitrary for each time series and adjusted so to permit viewing all four in one figure.

Table 1. Summary Statistics for Four Time Series.

	Average	Standard Deviation S.D.	Units
GDP	6.977743	5.22278	Trillions $
Debt	3307.348	5125.92	Billions $
Taxes	783810.8	960135.4	Thousands $
US Pop	213.580	63.03553	Millions

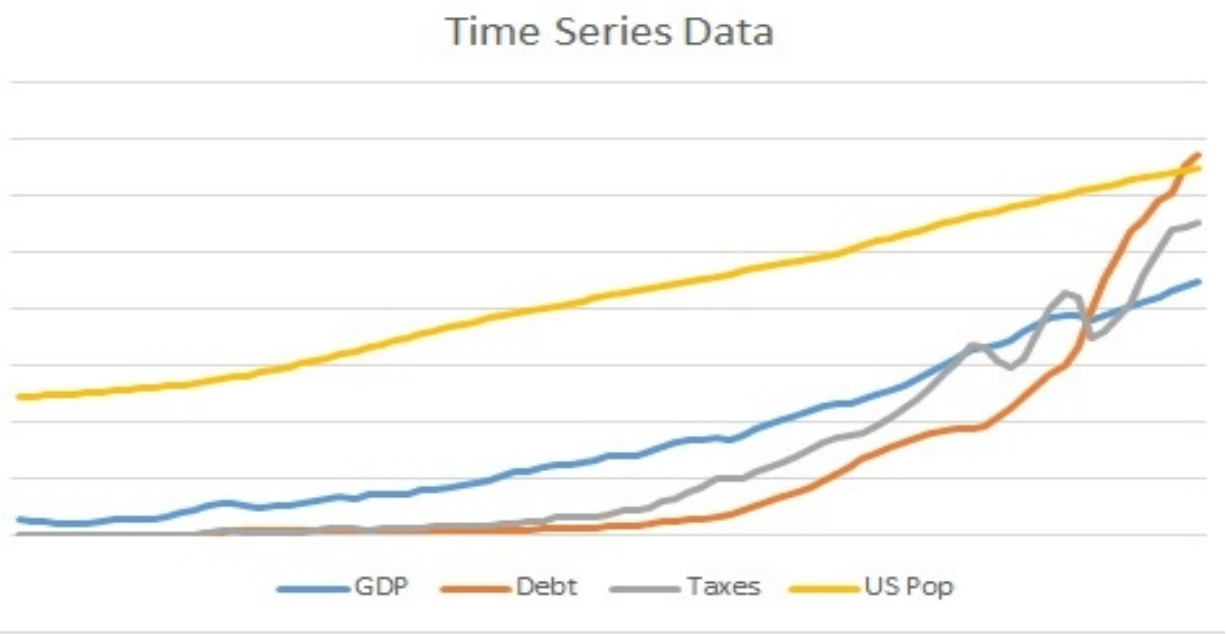

Figure 1. Time series data for the years 1929 to 2017, showing real GNP, the US national debt, federal tax receipts and the US population side.

Prior to the multivariate time series analysis, the data was centered and standardized, to help avoid rounding errors; the average reported in Table 1 was subtracted from each time series, and the result was divided by the standard deviation (S.D., also reported in Table 1). With the small number of data points (N=89), constraint (8) is preferred to (7). It turns out that the implementation of constraint (8) is not invariant to standardization, but depends on the innate variation presented to the analysis. Given that the units of measure are arbitrary, and the standard deviations therefore vary, pre-treating the data so that each time series has unit standard deviation is also preferred to present the data for analysis in a more even handed way. It becomes necessary to apply an inverse transformation to present the results on the original scale.

The results from using (12) as a modification to the nested algorithm of Section 4 came with a log-likelihood that was slightly larger, and so those are the results that are reported in Table 2; showing the estimates of σ^2 and ρ followed by the associated transformation $\mathbf{q}_i^T$. The matrix $\mathbf{Q}^T$ is the 4×4 block of number in the lower-left corner of Table 2. The diagonals of $\mathbf{D}_1$ are given by the first column of numbers, the diagonals of $\mathbf{D}_2$ are found by the multiplying the first column of numbers by the second column of numbers (row by row). The estimate of **W** is $\mathbf{Q}^T\mathbf{D}_1\mathbf{Q}$ and of **G** it is $\mathbf{Q}^T\mathbf{D}_2\mathbf{Q}$. A result of condition (8) is that **W** and **G** commute (Harville 1997, Section 21.13), i.e., **WG=GW**. As soon as **W** and **G** are turned into correlation matrices, or if returned to the original scale represented by Table 1, the matrices no longer commute when multiplied together; therefore, the analysis is not

invariant to the scale of measure.

Table 2. Results by Minimum Principal Components.						
Component	Fit Statistics		Linear coefficients for each component			
	σ^2	$\rho=\frac{\sigma_s^2}{\sigma^2}$	GDP	Debt	Taxes	US Pop
1st	4.67×10^{-7}	34.58	-1.97×10^{-2}	-2.18×10^{-3}	3.17×10^{-3}	.999798
2nd	7.64×10^{-5}	5.369	.8289837	-.46715	-.30707	1.63×10^{-2}
3rd	1.29×10^{-4}	3.922	.5221395	.843201	.127456	1.17×10^{-2}
4th	5.11×10^{-4}	19.67	.1994173	-.26605	.943107	3.63×10^{-4}

The algorithms performed well. While the initial factorization of **K** depends on the theoretical selection of ρ, once factorization is complete for the first time and the sparse structure of **L** is found, reconstituting **L** for different selections of ρ never became difficult; the same sparse structure was used over and over. Moreover, the specialized software was able to permute the rows and columns of **K** to permit a linear time factorization. While prior knowledge can be used to find a permutation that permits factorization as demonstrated by Smith (2001), this knowledge was not needed with the general purpose software that was used.

The results of Table 2 are returned to the original scale as shown in Table 3. The signal variation is given by variation associated with u(t), or the variation in the numerator of ρ. The noise variation is associated with σ^2. The noise variation was small relative to the signal variation. Moreover, the signal variation was tiny compared to the raw standard deviations reported in Table 1, implying that the second order model did a good[5] job explaining most of the variation for each time series. While there is a positive correlation between the signals given by GDP and tax receipts, there was a negative correlation between tax receipts and debt, and a negative correlation between GDP and dept, perhaps because of the antagonism of taxation on economic growth. The correlations involving the population size were all small.

[5] Perhaps this is unsurprising because this particular 2nd order model mimics non-parametric regression, and because it represents a non-stationary time series.

Table 3. Estimated Parameters in Multivariate Time Series Model.							
	$S.D.^{A} \times \sigma$	$S.D.^{A} \times \sigma_s$	$\frac{\sigma_s^2}{\sigma^2}$	Correlations [B]			
				GDP	Debt	Taxes	US Pop
GDP	.05426	.14952	7.593	1	-.4821	.6707	.0782
Debt	61.603	174.62	8.036	.0001	1	-.7462	.0063
Taxes	20686.	91058.	19.37	.3811	-.3998	1	.0057
US Pop	.04480	.25474	32.33	.2505	.0752	-.0009	1
Notes: A - S.D. taken from Table 1. B - Correlations above the diagonal correspond to the signal variation, correlations below the diagonal correspond to noise variation.							

Once the i-th step is complete (refer to step 4.2 in Section 4 again), and with **L** computed, its easy to calculate predictions of effects shown in models (3) and (4) using backward substitution (see Smith 2001, Smith 2018b). Furthermore, with the matrix **Q** estimated, and with the numbers in Table 1, all transformations can be reversed and predictions of u(t) can be found for each time series on their original scale. Those predictions are presented in Figure 2. Curves in Figure 2 are a little smoother than those in Figure 1, but Figure 1 and 2 look almost identical. This shows that most of the variation was explained by the 2nd order model, that the 2nd order model is almost like a multivariate non-parametric regression, and that the fit did not suffer from using the more parsimonious constrains (8) rather than the more general constraints (7).

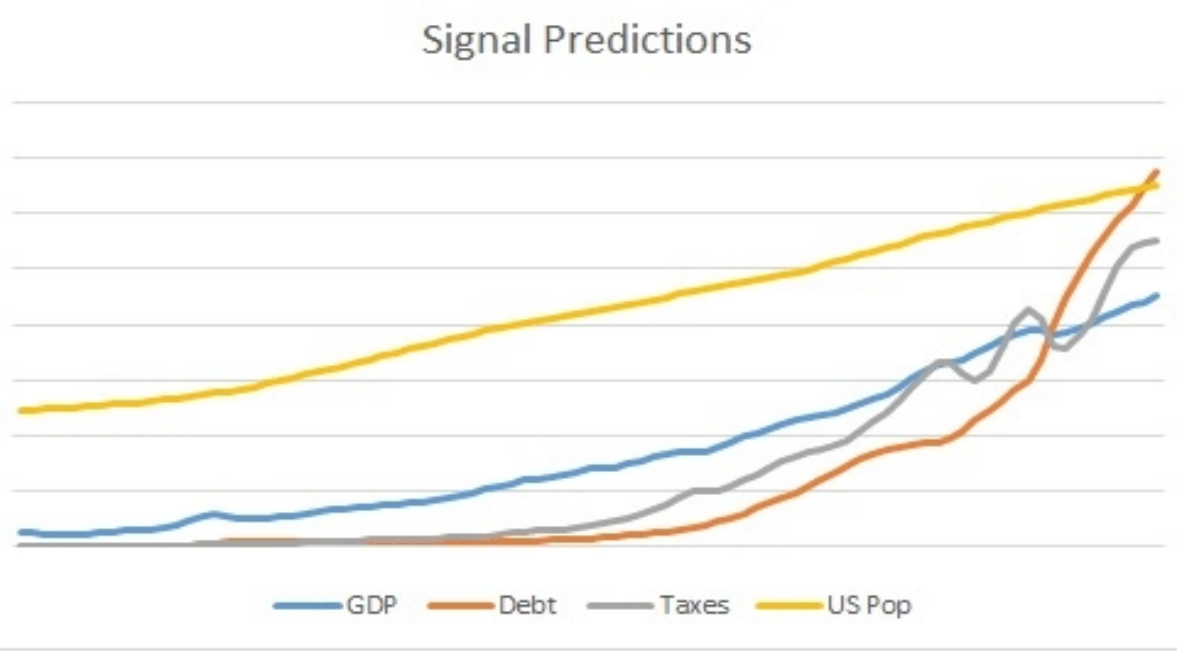

Figure 2. Prediction of 2nd order effect, or u(t), for the years 1929 to 2017, showing real GNP, the US national debt, federal tax receipts and the US population side.

References

Harville, D.A., 1997, *Matrix Algebra from a Statistician's Perspective*, Springer.

Jensen, J., and I.L. Mao, 1988, Transformation algorithms in analysis of single trait and of multitrait models with equal design matrices and one random factor per trait: a review, *Journal of Animal Science*, 66, 1750-.2761

Lütkepohl, H., 1993, *Introduction to Multiple Time Series Analysis*, 2nd Edition, Springer-Verlag.

Lin, C.Y., and S.P. Smith, 1990, Transformation of multitrait to unitrait mixed model analysis of data with multiple random effects, *Journal of Dairy Science*, 73, 2494-2502.

Meyer, K., 1985, Maximum likelihood estimation of variance components for multivariate mixed model with equal design matrices, *Biometrics*, 41, 153-165.

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, S.P., 2018a, The backward differentiation of the bordering algorithm for an indefinite Cholesky factorization, Data Structures and Algorithms, viXra archived #1707.0239.

Smith, S.P., 2018b, Non-parametric regression or smoothing on a two dimensional lattice using the K-matrix, Statistics, viXra archived #1812.0044.

Smith, S.P., 2018c, Stochastic spline functions with unequal time steps, Statistics, viXra archived #1809.0044.

Taylor, J.F., B. Bean, C.E. Marshal and J.J. Sillivan, 1985, Genetic and environmental components of semen production traits of artificial insemination of Holstein bulls, *Journal of Dairy Science*, 68, 2703-2722.