

Applying the Expectation-Maximization Algorithm to Multivariate Signal Extraction

Journal of Official Statistics

2024, Vol. 40(4) 660–684

© The Author(s) 2024

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0282423X241271471

journals.sagepub.com/home/jof



James Livsey  and Tucker McElroy

Abstract

Understanding time series signal extraction is vital for federal agencies, especially as data collection accelerates. One natural extension is moving from univariate to multivariate signal extraction, which offers the promise of reducing extraction error by exploiting cross-sectional relationships. However, such an extension ushers in new computational challenges, viz. larger parameter spaces and more complicated objective functions. The Expectation-Maximization (EM) algorithm provides a methodology to implicitly compute (or approximate) maximum likelihood estimates (MLEs). This paper provides methodology for applying the EM algorithm to a class of latent component multivariate time series models that allow for a nuanced specification of the unobserved signal. We derive an explicit formula for the maximization step, which facilitates computation speed while also improving the stability of the algorithm. Numerical studies demonstrate EM's ability to efficiently compute MLEs in low-dimensional systems, while also providing feasible estimates in moderate-dimensional systems where MLEs are infeasible to compute. Applications to monthly housing starts and daily immigration are provided.

Keywords

time series, structural component models, seasonal adjustment

1. Introduction

Multivariate signal extraction can be accomplished through the use of latent component models (McElroy and Trimbur 2015), for which the number of parameters typically increases as a quadratic function of dimension. Heuristically, this is

U.S. Census Bureau, Washington, DC, USA

Corresponding author:

James Livsey, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233-0002, USA.

Email: james.a.livsey@census.gov



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

because the linear filtering theory is built upon a knowledge of variance and covariances, that is, for N random variables we need to know the covariance matrix, which has $\binom{N+1}{2}$ independent entries. Direct maximum likelihood (ML; i.e., numerical optimization of the likelihood), routines for an N -dimensional time series encounter the difficulty of numerical optimization over \mathbb{R}^p , where the number of parameters p is large, that is, $p > 100$. Some of the issues that arise here are: (i) longer times to evaluate the objective function (e.g., the log Gaussian likelihood), due to large N ; (ii) long search times for BFGS and other algorithms, because p is large; (iii) termination at saddle-points (i.e., the numerical Hessian has at least one negative eigenvalue); (iv) sensitivity of results to initialization (i.e., possibility of termination at local optima); (v) a weak form of non-identifiability, whereby distinct parameter values yield indistinguishable evaluations of the objective function.

From the perspective of signal extraction, obtaining parameter Maximum Likelihood Estimates (MLEs) is only a means to the end of constructing a model-based linear filter. Hence, premature terminations of the nonlinear optimization (resulting in saddle-points or local optima) may be adequate for the practitioner objectives. In the case of the Gaussian likelihood, the objective function is optimized by minimizing one-step ahead forecast error, and therefore the filters constructed from such premature terminations may still perform well. Nevertheless, there are some features of the model fitting that are absolutely vital to the performance of signal extraction filtering: (i) the signal-to-noise ratios (snr), which are determined by estimates of variability for different components; (ii) cross-correlations between series, with respect to a particular signal dynamic.

The failure to obtain accurate snr estimates results in either under- or over-smoothing. Under-smoothing refers to an incomplete extraction of the signal that is actually present, and by necessity some of this signal will be leached into other components—this is a critical failure in an application such as seasonal adjustment. Over-smoothing refers to poaching signal content that is spurious, or not really present, essentially from stationary oscillations in other components, thereby destroying structure that may be of interest. For example, seasonal adjustment filters that have a spectral trough that is too wide can potentially alter and disrupt the business cycle component (McElroy 2012). Over- and under-smoothing can also arise in trend extraction, whereby the use of a moving average trend filter that is either too long or too short yields, respectively, an over-smoothed or under-smoothed trend.

A danger of over-stating cross-correlation patterns is that collinear models may be fallaciously adopted (McElroy and Jach 2019), resulting in so-called common components (e.g., common trends, common seasonals, etc.) in the model Zuur et al. (2003). This results in filters that strongly rely upon cross-sectional filtering rather than temporal weighting. When such collinearity patterns are not truly present in the data, dynamics from one series are spuriously imposed upon another. Under-stating the correlations is a less serious offense—if there is no cross-

correlation, the problem reduces to univariate filtering, and we suffer efficiency losses, that is, less precise extraction of the signal.

Whereas straight likelihood maximization may avoid the pitfalls of snr and cross-correlation, the computational challenges mentioned above have motivated other approaches, such as the method of moment (MOM) estimators of McElroy (2017). Although these MOM estimators are consistent (McElroy and Roy 2023) and computationally feasible, being given by an analytical formula only involving linear combinations of sample autocovariances, they do suffer from deficiencies. In particular, these estimators can yield covariance matrix estimates that have negative eigenvalues; the closest positive definite approximation, obtained by setting any negative eigenvalues to zero and reconstructing the matrix with the same eigenvectors, yields a singular covariance matrix wherein spurious collinearity has been infused.

Another approach, based on work of Shumway and Stoffer, is to use the Expectation-Maximization (EM) algorithm (Dempster et al. 1977) to implicitly compute MLEs, or perhaps approximate the true MLEs (Shumway and Stoffer 1982). EM proceeds first by the concept of a complete data likelihood, which in this context amounts to considering the data jointly with the signals of interest (see Chapter 8 of Little and Rubin (2019)). Using log Gaussian likelihoods, the conditional expectation operator is applied (the E-step). At this point, we suppose this expectation operator is based upon prior knowledge of the parameters, that is, the values of the parameters obtained from an earlier iteration of the iterative algorithm. However, there are other quantities in the likelihood that depend upon the true, unknown parameters, and at this stage (the M-step) we optimize with respect to these, in the hope that a simple formula will emerge, or at worst that the numerical optimization will be easier than direct likelihood maximization. With these new updated parameters, one computes signal extraction estimates, along with their error covariances, as these are required in the next E-step.

In this article we study a special case of the component models driven by vector white noise considered by McElroy (2017), wherein all the white noise covariance matrices have full rank. We show the M-step yields an explicit formula for the white noise covariance matrices, and this formula can be computed from a knowledge of the extracted signal and the error covariances. This formula is fast to compute (no matrix inversions), and hence the speed of the method depends on our computational facility with signal extraction. In high-dimensional problems, the extraction of a signal takes an amount of time—using State Space (SS) methods, or the recursive approach of *Ecce Signum* (ES; an R package developed by the US Census Bureau to compute multivariate signal extraction and forecasting results for data with ragged edge missing values; see McElroy and Livsey (2022) and www.github.com/tuckermcelroy/sigex)—roughly comparable to, or a little longer than, a single likelihood evaluation. Whereas our main results (Theorems 2 and 2) can be applied under either an SS or ES paradigm, we have developed our numerical results under the latter computational framework because ES provides the requisite error covariances at all lags using efficient recursive methods described in McElroy (2022). These error covariances are not a standard output of a Kalman filter/smoothing algorithm (although the error variances are indeed calculated), and typically require

a custom implementation of the SS iterations; for example, Shumway and Stoffer (1982) developed a custom SS routine to generate the lag one error covariances, which was sufficient for their application. In our more general context, such error covariances are potentially needed at all lags. However, the choice of ES by the authors for implementation does not preclude the application of this paper's EM methods within an SS framework.

When estimating high dimensional multivariate models it is commonplace to employ Newton-Raphson (NR) type optimization routines. These routines are not guaranteed to improve the likelihood for each iteration; decreases to the likelihood in the short-term can be helpful for skipping over local optima. In contrast, successive iterations of the EM algorithm always improve the likelihood and converge to a stationary point (Shumway and Stoffer 1982; Wu 1983). Additionally, there has been considerable development to improve the stability and speed of EM (Jamshidian and Jennrich 1997; Liu et al. 1998). For high-dimensional parameter spaces, the EM method has a speed advantage over NR methods because fewer likelihood evaluations are required, and moreover there is no need to compute numerical derivatives. Furthermore, EM parameter estimates by construction inherit the relevant properties (e.g., estimates of covariance matrices given in Equation (11) below are symmetric and non-negative definite), whereas numerical optimization routines (such as NR) typically require a re-parameterization to enforce these properties. However, the rate of convergence of EM methods near an optimum can be slow; neither NR or EM can avoid the possibility of finding local optima, though the risks can be mitigated by utilizing informative initial conditions—such as MOM or univariate ML (assuming no cross-correlations) parameter estimates.

The rest of the article proceeds as follows: in Section 2 we outline the modeling paradigm and define structural models. The only parameters of these simple models are the covariance matrices of latent processes' innovations, and our main result in Section 2 provides the M-step for such covariance parameters in a closed form. Then in Section 3 we discuss the algorithmic details, including pseudo-code and convergence criteria specifics. Sections 4 and 5 give numerical studies, as well as applications to monthly housing starts and daily immigration. We conclude with final remarks in Section 6.

2. Methodology

Consider an N -dimensional vector time series $\{X_t\}$ that incorporates the dynamics of J latent signals, denoted by $\{S_t^{(j)}\}$ for $1 \leq j \leq J$, as well as an irregular process $\{S_t^{(0)}\}$, which is assumed to be stationary. These latent processes are additively related to the data process via

$$X_t = \sum_{j=0}^J S_t^{(j)}. \quad (1)$$

A leading example is furnished by $S^{(1)}$ corresponding to a stochastic trend, $S^{(2)}$ corresponding to a stochastic seasonal, and $S^{(0)}$ corresponding to a white noise process.

In general, we formulate the signal extraction in terms of difference-stationary processes. Hence we adopt the following assumptions: there exist relatively prime scalar differencing polynomials $\delta^{(j)}(B)$ (for $1 \leq j \leq J$) such that $\underline{S}_t^{(j)} = \delta^{(j)}(B)S_t^{(j)}$ is covariance stationary and mean zero. The differencing polynomial $\delta^{(j)}$ is of order d_j , and is denoted by $\delta^{(j)}(B) = \sum_{k=0}^{d_j} \delta_k^{(j)} B^k$. Here the scalar polynomial in the backshift operator B is applied in the same way to each of the N components of the vector process. In the case of the irregular process, the differencing polynomial $\delta^{(0)}(B) = 1$ because $S^{(0)}$ is stationary (and hence $\underline{S}_t^{(0)} = S_t^{(0)}$).

It follows that $\delta(B) = \prod_{j=1}^J \delta^{(j)}(B)$ is sufficient to reduce $\{X_t\}$ to stationarity. In particular, let $\delta^{(-j)}(B) = \prod_{k \neq j} \delta^{(k)}(B)$, so that

$$\underline{X}_t = \delta(B)X_t = \delta(B)S_t^{(0)} + \sum_{j=1}^J \delta^{(-j)}(B)\underline{S}_t^{(j)} \tag{2}$$

is covariance stationary with mean zero. It is convenient to introduce the notation

$$\underline{S}_t^{(-0)} = \delta(B)S_t^{(0)} \quad \underline{S}_t^{(-j)} = \delta^{(-j)}(B)\underline{S}_t^{(j)}.$$

These denote stationary components that have been over-differenced.

Our goal is to develop an expression for the Gaussian likelihood in terms of these latent processes, as we endeavor to develop a so-called complete data likelihood. To proceed, we must make some assumptions about these latent processes. We suppose that all the differenced latent processes $\{\underline{S}_t^{(j)}\}$ are independent of one another, and each has dynamics governed by a scalar ARMA equation driven by multivariate white noise; hence, the spectral density of $\underline{S}^{(j)}$ can be written as $g^{(j)}(\lambda)\Sigma^{(j)}$, where $g^{(j)}$ is scalar and non-negative. Any constant (positive) scaling of $g^{(j)}$ can be incorporated into $\Sigma^{(j)}$, and hence without loss of generality we suppose that $g^{(j)}$ corresponds to the spectral density of an ARMA process of unit innovation variance, that is, $0 = \int_{-\pi}^{\pi} \log[g^{(j)}(\lambda)] d\lambda$ by Szëgo’s formula. Suppose that we have a sample of size T , denoted by $X_{1:T} = [X'_1, \dots, X'_T]'$. The differencing polynomial $\delta^{(j)}(B)$ is of order d_j , and hence the available stationary elements of $\underline{S}_t^{(j)}$ are $[\underline{S}_{d_j+1}^{(j)}, \dots, \underline{S}_T^{(j)}]$. The block Toeplitz covariance matrix of this stacked random vector is

$$\Gamma^{(j)} = \Gamma_{T-d_j}(g^{(j)}) \otimes \Sigma^{(j)}, \tag{3}$$

where $\Gamma_t(g)$ denotes a t -dimensional Toeplitz matrix with $k\ell$ th entry given by

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) e^{i\lambda(k-\ell)} d\lambda = \langle g \rangle_{k-\ell}.$$

(This defines the notation $\langle g \rangle_h$.) We remark that in some modeling scenarios it is possible that $g^{(j)}$ has zero values, or that $\Sigma^{(j)}$ is singular, in which case $\Gamma^{(j)}$ is non-invertible; for example, this can arise with common trend (or common seasonal) processes. However, we preclude such situations explicitly, since our methodology relies upon the invertibility of the $\Sigma^{(j)}$ matrices.

The block Toeplitz covariance matrices for the over-differenced vectors have an expression that is similar to (3): letting $d = \sum_{j=1}^J d_j$ and $h^{(-j)}(\lambda) = |\delta^{(-j)}(e^{-i\lambda})|^2$, we have

$$\Gamma^{(-j)} = \Gamma_{T-d}(h^{(-j)}g^{(j)}) \otimes \Sigma^{(j)} \tag{4}$$

for $1 \leq j \leq J$. Because $h^{(-j)}$ has zeroes, the matrix $\Gamma_{T-d}(h^{(-j)}g^{(j)})$ will be ill-conditioned for large values of T , but is indeed invertible for small to moderate values of T . For $j = 0$, we have the covariance matrix of the sample of $\underline{X}^{(-0)}$, which is

$$\Gamma^{(-0)} = \Gamma_{T-d}(h^{(0)}) \otimes \Sigma^{(0)},$$

where $h^{(0)}(\lambda) = |\delta(e^{-i\lambda})|^2$. It follows from Equation (2) and the independence of the differenced latent processes that the covariance matrix of the differenced sample $\underline{X}_{d+1:T} = [\underline{X}'_{d+1}, \dots, \underline{X}'_T]'$ is $\Gamma^{(*)} = \sum_{0 \leq j \leq J} \Gamma^{(-j)}$. In order to compute the divergence, or -2 times the log Gaussian likelihood with constants removed, we suppose the following relationship between the initial values X_1, \dots, X_d and the differenced latent processes:

Assumption \tilde{A} : the initial values X_1, \dots, X_d are independent of $\{\underline{S}^{(j)}\}$ for $0 \leq j \leq J$.

This assumption generalizes Assumption M_∞ of McElroy and Trimbur (2015) and Assumption \tilde{A} of McElroy (2022) to more than two latent processes, and implies that the divergence is

$$\mathcal{L} = \underline{x}'_{d+1:T} \Gamma^{(*)^{-1}} \underline{x}_{d+1:T} + \log \det \Gamma^{(*)}. \tag{5}$$

(We use lower case for realizations of random vectors.) So up to constants, this is the log of the likelihood of the differenced data, or $\log p_{\underline{X}_{d+1:T}}$. Under conventional assumptions (discussed in McElroy and Monsell (2015)), this also equals the log likelihood of the undifferenced data, up to some scaling constants that do not depend on model parameters. The complete likelihood of the EM literature is defined as $p_{\underline{X}_{d+1:T}, \underline{S}_{d_1+1:T}^{(1)}, \dots, \underline{S}_{d_J+1:T}^{(J)}}$ in our context, which is obtained by considering the joint distribution of the differenced data process together with the differenced signal processes. Here, we have excluded the joint distribution with the irregular. This joint probability density function (pdf) factors as follows:

$$P_{\underline{X}_{d+1:T}, \underline{S}_{d_1+1:T}^{(1)}, \dots, \underline{S}_{d_J+1:T}^{(J)}} = P_{\underline{X}_{d+1:T} | \underline{S}_{d_1+1:T}^{(1)}, \dots, \underline{S}_{d_J+1:T}^{(J)}} \prod_{j=1}^J P_{\underline{S}_{d_j+1:T}^{(j)}}.$$

The first pdf, that is, the data conditional on the signals, on the right hand side of the equation is equal to the pdf of the over-differenced irregular evaluated at $\underline{x}_t - \sum_{j=1}^J \underline{y}_t^{(-j)}$ for $d + 1 \leq t \leq T$. This fact uses the assumed independence of the irregular component from the differenced component signals. Assuming that each $\Gamma^{(-j)}$ is invertible, we find that the complete data divergence is

$$\sum_{j=0}^J \underline{y}_{d+1:T}^{(-j)'} \Gamma^{(-j)-1} \underline{y}_{d+1:T}^{(-j)} + \sum_{j=0}^J \log \det \Gamma^{(-j)}, \tag{6}$$

noting that $\underline{x}_{d+1:T} - \sum_{j=1}^J \underline{y}_{d+1:T}^{(-j)} = \underline{y}_{d+1:T}^{(-0)}$. Our main result shows that for a Gaussian process the E-step of the EM algorithm can be explicitly calculated by applying the conditional expectation operator to the complete data divergence Equation (6).

Theorem 1. *Let $\{X_t\}$ be a Gaussian process of form in Equation (1) that satisfies Assumption \tilde{A} , and such that the $\{\underline{S}_t^{(j)}\}$ are independent of one another with invertible $\Sigma^{(j)}$. Then the conditional expectation of the complete data divergence in Equation (6) is*

$$\sum_{j=0}^J \text{tr} \left\{ \Gamma^{(-j)-1} \left[M^{(j)} + \underline{\hat{y}}_{d+1:T}^{(-j)} \underline{\hat{y}}_{d+1:T}^{(-j)'} \right] \right\} + \sum_{j=0}^J \log \det \Gamma^{(-j)}, \tag{7}$$

where $M^{(j)}$ is the error covariance matrix for the j th over-differenced signal and $\underline{\hat{y}}_t^{(-j)}$ are the signal extraction estimates for the j th over-differenced signal.

Proof. To the expression in Equation (6) is applied the conditional expectation with respect to the data (for this purpose, we consider random vectors rather than realizations in the above). Under Assumption \tilde{A} , the expectation conditional on either the data or the differenced data is identical. Also, we use the following fact: if S is a Gaussian random vector with $\hat{S} = \mathbb{E}[S|X]$, for X a Gaussian random vector, then the signal extraction error $S - \hat{S}$ is uncorrelated with all linear functions of X . As a result

$$\mathbb{E}[(S - \hat{S})(S - \hat{S})'|X] = \mathbb{E}[(S - \hat{S})(S - \hat{S})'],$$

which we denote by M ; this is the error covariance matrix, the diagonals of which are the signal extraction mean squared errors (MSEs). Next, because

$$SS' = (S - \hat{S})(S - \hat{S})' + S\hat{S}' + \hat{S}S' - \hat{S}\hat{S}',$$

applying the conditional expectation operator yields

$$\mathbb{E}[SS'|X] = M + \hat{S}\hat{S}'. \tag{8}$$

Next, we apply Equation (8) in the case of each over-differenced signal $\underline{\Sigma}_t^{(-j)}$ that occurs in Equation (6), rewriting each quadratic form in terms of the trace operator and applying the conditional expectation; this yields the desired result. \square

Theorem 1 produces the E-step of the EM algorithm, wherein all the signal extraction estimates $\hat{s}_t^{(j)}$ are computed (based on previous values of the parameters) and differenced, along with the error covariance matrices, and supplied to the objective function Equation (7). We remark that these matrices ($\Gamma^{(-j)}$ and $M^{(j)}$) are $N(T - d)$ -dimensional, which can be quite large in common applications; in our simulations the dimension is 1464, which is substantial. The M-step seeks to optimize with respect to the parameters present in each $\Gamma^{(-j)}$. If these are obtained, then the model is fitted, and new signal extraction estimates can be generated.

The matrices $\Gamma^{(-j)}$ depend on the ARMA parameters through $g^{(j)}$, as well as the covariance parameters present in $\Sigma^{(j)}$. At this point we focus upon the case where the ARMA parameters are known, so that inference is focused only upon the covariance parameters—this case can arise when each $\{\underline{\Sigma}_t^{(-j)}\}$ is a white noise process, for example. (However, the EM algorithm could be used in combination with NR methods of likelihood optimization in the case that ARMA parameters are unknown.) From (4) we immediately obtain

$$\Gamma^{(-j)^{-1}} = \Gamma_{T-d}(h^{(-j)}g^{(j)})^{-1} \otimes \Sigma^{(j)^{-1}} \tag{9}$$

for $1 \leq j \leq J$, and a similar expression for $j = 0$. Let $D^{(j)} = \Gamma_{T-d}(h^{(-j)}g^{(j)})^{-1}$, which is $(T - d)$ -dimensional with entries denoted $D_{k\ell}^{(j)}$ for $d + 1 \leq k, \ell \leq T$. The error covariance matrix consists of blocks of dimension N , and we let $M_{k\ell}^{(j)}$ denote the $N \times N$ block that equals the error covariance of $\underline{\Sigma}_k^{(-j)}$ and $\underline{\Sigma}_\ell^{(-j)}$, for $d + 1 \leq k, \ell \leq T$. Then the following result provides a formula for the covariance matrices estimated in the M-step.

Theorem 2. *Under the assumptions of Theorem 1, the conditional expectation of the complete data divergence can be rewritten as*

$$\sum_{j=0}^J \text{tr} \left\{ \Sigma^{(j)^{-1}} \sum_{k, \ell = d+1}^T D_{k\ell}^{(j)} \left[M_{k\ell}^{(j)} + \hat{\underline{\Sigma}}_\ell^{(-j)} \hat{\underline{\Sigma}}_k^{(-j)'} \right] \right\} - \sum_{j=0}^J \log \det D^{(j)} + \sum_{j=0}^J \log \det \Sigma^{(j)}. \tag{10}$$

The M-step estimators of $\Sigma^{(j)}$ are given by the formula

$$\hat{\Sigma}^{(j)} = \frac{1}{T-d} \sum_{k, \ell = d+1}^T D_{k\ell}^{(j)} \left[M_{k\ell}^{(j)} + \hat{\underline{\Sigma}}_\ell^{(-j)} \hat{\underline{\Sigma}}_k^{(-j)'} \right]. \tag{11}$$

These estimators are symmetric, non-negative definite, and unbiased.

Proof. The expression for the conditional expectation of the complete data divergence Equation (10) follows at once from Equation (7) and Equation (9).

In seeking to optimize the expected complete data divergence with respect to one of the matrices $\Sigma^{(j)}$, we proceed without constraint on the form of the matrix, and by results in Magnus and Neudecker (2019) obtain the critical point Equation (11). To show the properties of the estimators we omit the superscript j on all components, for clarity. The symmetry property follows at once by interchanging k and ℓ in Equation (11), and

$$\begin{aligned} \mathbb{E}[\Sigma] &= \frac{1}{T} \sum_{k=d+1}^T \sum_{\ell=d+1}^T D_{k\ell}(M_{\ell k} + \mathbb{E}[\hat{s}_\ell \hat{s}'_k]) \\ &= \frac{1}{T} \sum_{k=d+1}^T \sum_{\ell=d+1}^T D_{k\ell} \underbrace{\mathbb{E}[s_\ell s'_k]}_{\gamma^{(\ell-k)\Sigma}} \\ &= \Sigma \cdot \frac{1}{T} \text{tr}\{D\Gamma\} \\ &= \Sigma. \end{aligned}$$

This shows the estimator is unbiased. Next, let $\mathbf{a} \in \mathbb{R}^N$ such that $\mathbf{a} \neq \mathbf{0}$. Then

$$\begin{aligned} \mathbf{a}'\Sigma\mathbf{a} &= \frac{1}{T} \sum_{k=d+1}^T \sum_{\ell=d+1}^T D_{k\ell}[\mathbf{a}'M_{\ell k}\mathbf{a} + (\mathbf{a}'\hat{\underline{s}}_\ell)(\mathbf{a}'\hat{\underline{s}}_k)'] \\ &= \frac{1}{T} \{\text{tr}(DA) + \hat{\mathbf{c}}'D\hat{\mathbf{c}}\}, \end{aligned}$$

where A is a $(T-d) \times (T-d)$ -dimensional matrix with (ℓ, k) th element $A_{\ell k} = \mathbf{a}'M_{\ell k}\mathbf{a}$, and $\hat{\mathbf{c}}$ is a length $T-d$ vector with k th element $\hat{c}_k = \mathbf{a}'\hat{\underline{s}}_k$. Since D is positive definite, $\hat{\mathbf{c}}'D\hat{\mathbf{c}} \geq 0$, and it remains to show that $\text{tr}(DA)$ is non-negative. We know M is non-negative definite (because it is a covariance matrix), and hence

$$A = (\mathbf{a}' \otimes \mathbf{1}_{T-d})M(\mathbf{a} \otimes \mathbf{1}_{T-d})$$

is also non-negative definite. Thus $\text{tr}(DA) \geq 0$. □

Remark 1. We will subsequently need to evaluate $M_{ts}^{(-j)}$, the cross covariance between over-differenced signals, that is,

$$\begin{aligned} M_{ts}^{(-j)} &= \text{Cov}(\underline{s}_t^{(-j)} - \hat{\underline{s}}_t^{(-j)}, \underline{s}_s^{(-j)} - \hat{\underline{s}}_s^{(-j)}) \\ &= \sum_{k=0}^d \sum_{\ell=0}^d \delta_k^{(-j)} \delta_\ell^{(-j)} M_{t-k, s-\ell}^{(j)}. \end{aligned}$$

For example, for a lag 12 seasonal difference $\delta^{(-j)}(B) = 1 - B^{12}$, we obtain $M_{ts}^{(-j)} = M_{ts}^{(j)} - M_{t-12, s}^{(j)} - M_{t, s-12}^{(j)} + M_{t-12, s-12}^{(j)}$.

With the methodology developed by Theorems 1 and 2, we can construct an EM algorithm for signal extraction—this is discussed in Section 3 below.

3. Computation

We now discuss the implementation of the EM algorithm, which involves calculation of signal extraction estimates (the E-step) followed by updating the covariance matrix estimates (the M-step) using the formulas in Equation (11). Theorem 2 guarantees that each $\Sigma^{(j)}$ estimate is a symmetric non-negative definite matrix, and this is vital for computing the next iteration of signal extraction estimates. We also remark that the quantity $D^{(j)}$ does not impact computational complexity, since it can be determined once outside of the main iteration loop. In contrast, calculation of $M^{(j)}$ can be burdensome for large T and/or N , because it involves signal extraction error covariances at lags of up to order T . In the ES implementation these are computed through a combination of multi-step ahead forecasting error covariances with the Wiener-Kolmogorov signal extraction filter; see Proposition 3.1 of McElroy (2022) for details.

If adopting an SS approach to the calculation of these quantities, one could utilize a diffuse smoother with the original data for signal extraction; see Kailath et al. (2000) and Gómez (2016) for details. A so-called “square root” version of likelihood and smoother recursions is available, and has been successful in avoiding numerical difficulties in other settings (Marczak and Gómez 2017). Alternatively, with the ES approach (which we have adopted in our implementation) a faster, approximate approach to the calculation of $M^{(j)}$ is to first determine the error spectral density f for the Wiener-Kolmogorov filter of each signal (see McElroy and Livsey (2022), Subsection 2.4). Then $M_{k\ell}^{(j)}$ is approximately given by $\langle f \rangle_{k-\ell}$, which can be approximated using a Riemannian mesh.

A fully specified model will have J structural components of the form in Equation (1). Each individual component has a specified $\delta^{(j)}(B)$ that reduces the process to stationarity. In our implementation and applications, we assume that the functions $g^{(j)}$ are trivial, corresponding to each differenced process being a multivariate white noise. With these operators $\delta^{(j)}(B)$ being known, quantities such as $D^{(j)}$ can be computed immediately, and then the model parameters can be estimated with the EM methodology, using the pseudo-code in Algorithm 3. Specifically, the first step is to define for each component the structure of the over-differenced covariance matrices $\Gamma^{(-j)-1}$, as defined in Equation (9). At each step of the EM procedure, the conditional

Algorithm 1 Expectation-Maximization Algorithm. Signal estimates and associated errors are calculated using *Ecce Signum*

Procedure EXPECTATION-MAXIMIZATION

Initialization:

- Model specification: Specify $\delta^{(j)}(B)$ for all components $S_t^{(j)}, j = 1, 2, \dots, J$
- Compute $\Gamma^{(-j)}$ via (3)
- Initialize parameter estimates $\widehat{\Sigma}^{(j)}, \widehat{M}^{(j)}, \widehat{S}^{(j)}$ for all $j = 1, 2, \dots, J$

Main loop:

- while not converged do**
- Update $\Sigma^{(j)} = (T - d)^{-1} \sum_{k,\ell=d+1}^T D_{k\ell}^{(j)} \left[M_{\ell k}^{(j)} + \widehat{s}_\ell^{(-j)} \widehat{s}_k^{(-j)} \right]$.
- Run *Ecce Signum* with updated covariance structure $\Sigma^{(j)}, j = 0, 1, \dots, J$.
- Get updated $M^{(j)}, \widehat{S}^{(j)}$

expectation of the complete data divergence is fully specified through $M^{(j)}$ and $\underline{\Sigma}_{d+1:T}^{(-j)}$. The maximization point has the closed form in Equation (11), so each successive iteration is a simple updating procedure. The EM method is initialized at informative parameter values; in Algorithm 1 we utilize MOM estimates.

Termination of the algorithm occurs at convergence. Numerically, this amounts to deciding on an objective distance between successive parameter estimates. Since we are studying covariance matrices, the choice of matrix norm and termination distance can have an impact on the run time of the procedure. In this article we utilize the absolute distance between successive divergence values as our criterion. At the i th step of the main loop in Algorithm 1 the divergence \mathcal{L}_i , given by Equation (5), is calculated; then the absolute reduction in divergence from step $i - 1$ to step i , viz. $|\mathcal{L}_i - \mathcal{L}_{i-1}|$, is computed. The algorithm is said to have converged if the relative reduction in divergence is less than some pre-specified value.

The Gaussian divergence \mathcal{L} can be computed using the algorithms of Sections 2 and 3 of McElroy (2022), which are implemented in ES. (Recall that the Gaussian divergence given in Equation (5) is not the same as the expected complete data divergence given in Equation (10).) To compute the model parameters via ML, we minimize the Gaussian divergence using the BFGS algorithm (Golub and Van Loan 2013). This includes a reparameterization into an unconstrained Euclidean space, as advocated by Pinheiro and Bates (1996). (The ES software allows users to specify other methods of optimization, such as simulated annealing and the conjugate gradient method, utilizing the standard R optim function.)

4. Numerical Studies

We consider a series with two non-stationary components ($J = 2$), which is defined via

$$X_t = S_t^{(1)} + S_t^{(2)} + S_t^{(0)}, \tag{12}$$

where $S_t^{(1)}$ is a trend component, $S_t^{(2)}$ is a seasonal component, and $S_t^{(0)}$ is an irregular term. Furthermore, each latent component has the following structure:

$$\begin{aligned} (1 - B)S_t^{(1)} &= \underline{\Sigma}^{(1)} \sim \text{WN}(0, \Sigma^{(1)}) \\ (1 + B + B^2 + \dots + B^{11})S_t^{(2)} &= \underline{\Sigma}^{(2)} \sim \text{WN}(0, \Sigma^{(2)}) \end{aligned} \tag{13}$$

Here, $\text{WN}(0, \Sigma)$ represents an uncorrelated process with constant mean zero and covariance Σ . The component definitions Equation (13) imply that our full differencing operator is $\delta(B) = (1 - B^{12})$, and our over-differenced series are given as $\underline{\Sigma}^{(-1)}$, $\underline{\Sigma}^{(-2)}$, and $\underline{\Sigma}^{(-0)}$, such that

$$\begin{aligned} (1 - B^{12})X_t &= (1 + B + B^2 + \dots + B^{11})\underline{\Sigma}_t^{(1)} \\ &\quad + (1 - B)\underline{\Sigma}_t^{(2)} \\ &\quad + (1 - B^{12})\underline{\Sigma}_t^{(0)} \\ &= \underline{\Sigma}_t^{(-1)} + \underline{\Sigma}_t^{(-2)} + \underline{\Sigma}_t^{(-0)}. \end{aligned}$$

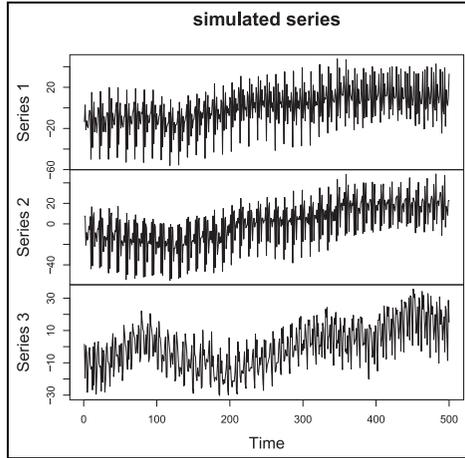


Figure 1. A sample path from a three-dimensional process. The trend and seasonal components of the first two series are positively correlated. The third series is independent of the first two.

We assume the dimension of each component is $N = 3$. The length of each series is $T = 500$. Each of the white noise covariance matrices are such that the first two elements are correlated in the trend and seasonal, whereas the third element of the vector time series is uncorrelated with the first two.

$$\Sigma^{(1)} = \begin{bmatrix} 1 & .75 & 0 \\ .75 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Sigma^{(2)} = \begin{bmatrix} 1 & .75 & 0 \\ .75 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Sigma^{(0)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

A simulated path from this model is displayed in Figure 1. Notice the similar movement in the trend and seasonal pattern for Series 1 and Series 2.

The EM algorithm proposed in Section 2 was applied to two hundred realizations of this process. The modest dimension of the process facilitates direct numerical optimization of the likelihood, allowing us to make direct comparisons to the ML results. (Note that for a higher-dimensional processes such comparisons are less feasible, due to computational challenges in obtaining the MLEs.) The Gaussian likelihood of the model given by Equation (12) and Equation (13) is described in McElroy (2017), and is calculated as described in Section 3; hence we are also able to calculate the Gaussian divergence at both the true parameters and the MOM estimators. The value of the Gaussian divergence at the true parameters is used as a basis for comparison, and its value at the MOM estimators is used to understand how far the Gaussian divergence has been reduced by both the ML and EM procedures (since both the ML and EM routines are initialized at the MOM estimates). The results show that both the ML and EM procedures yield significant improvements over the simple MOM estimator. Not only are the EM estimates

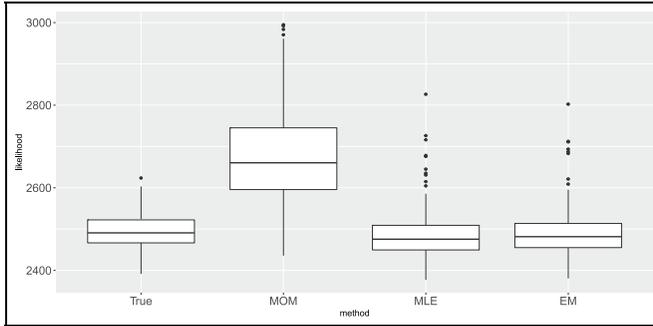


Figure 2. Boxplot of Gaussian divergence values for all two hundred replications of $N = 3$ dimensional process. ML and EM routines were initialized at the MOM estimates.

close to the MLEs in most of the cases, but both also yield values of the likelihood that are in the vicinity of the likelihood evaluated at the true parameters.

Box plots of these four Gaussian divergence values for all replicates are shown in Figure 2. As for the parameter estimates themselves, the results are quite promising. Boxplots of all twenty-seven scalar entries of the three covariance matrices $\Sigma^{(1)}$, $\Sigma^{(2)}$, and $\Sigma^{(0)}$ are shown in Figures 3 and 4, which display the results for the EM and ML procedures, respectively. The horizontal red line in each box plot represents the true parameter value, and we see that all the estimates appear to be unbiased. The x -axis indicates what scalar entry of the covariance matrix the corresponding box-plot is displaying. For example, the label $T11$ indicated the top-left entry of the trend covariance matrix $\Sigma^{(1)}$. For both methods the estimates of the irregular ($\Sigma^{(0)}$) have the most variation.

We also investigated the timing results for ML and EM estimation. Specifically, we simulated data from different sample sizes $T = 50, 100, 200, 400, 800$ and dimensions $N = 2, 3, 4, 5$, focusing upon a single realization of Equation (12). For each such dataset the parameter estimates for model in Equation (12) were estimated. To make equitable comparisons, each method had the same convergence criteria. Small dimensionality ($N = 2$) and small sample size ($T = 100, 200, 400$) is the only situation where the ML routine was faster than the EM algorithm, and in these situations the difference was small. For all other cases the EM algorithm was faster, and in many cases the difference was substantial. Most importantly, when dimension and sample size increased there were three situations in this timing study where ML failed to produce estimates. In contrast, the EM procedure was able to produce estimates in these situations ($N = 4$ with $T = 800$, and $N = 5$ with $T = 400$ or 800). Conclusions from such timing exercises require caution, because results are dependent upon coding efficiencies and programming vagaries—facets that are especially pertinent with an interpreted language such as R. Our limited investigation indicates that EM speed is on par with (or better) than the ML approach.

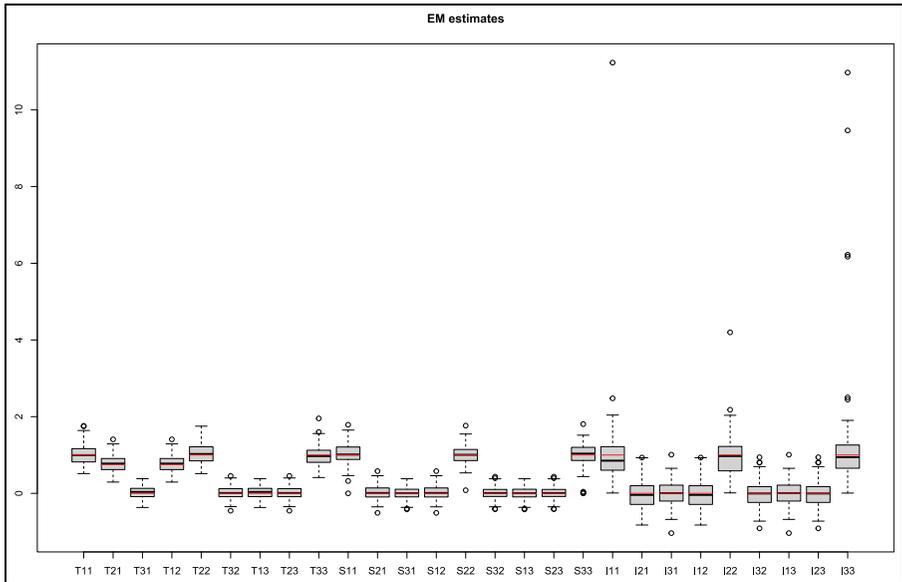


Figure 3. Boxplot of EM estimates for all two hundred replicates and all twenty-seven scalar entries of $\Sigma^{(1)}$ (trend covariance), $\Sigma^{(2)}$ (seasonal covariance), and $\Sigma^{(0)}$ (irregular covariance). The x-axis label indicate which component and entry is being displayed.

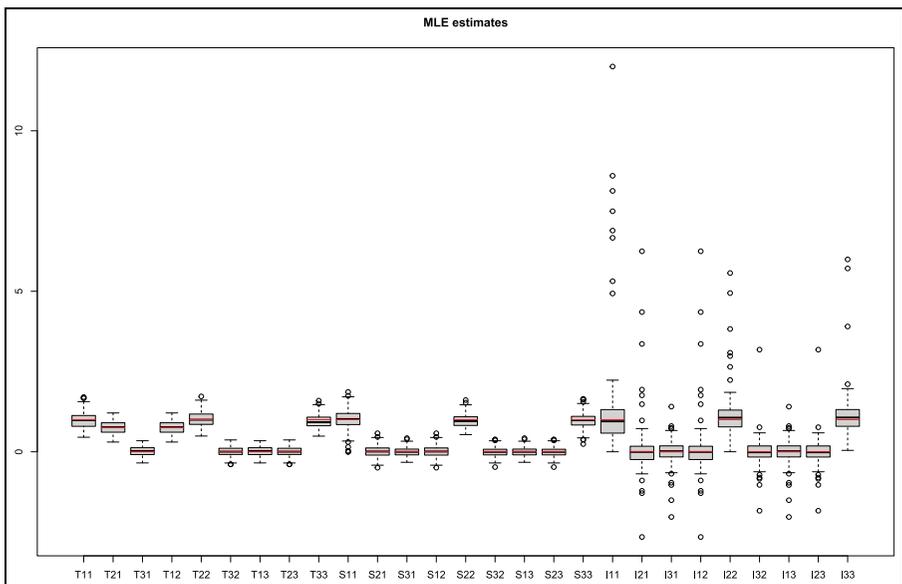


Figure 4. Boxplot of MLEs for all two hundred replicates and all twenty-seven scalar entries of $\Sigma^{(1)}$ (trend covariance), $\Sigma^{(2)}$ (seasonal covariance), and $\Sigma^{(0)}$ (irregular covariance). The x-axis label indicate which component and entry is being displayed.

5. Application

We investigate two data examples of our methodology. We first consider a four-dimensional monthly housing starts example from the Survey of Construction at the U.S. Census Bureau. Second, we examine a six-dimensional daily immigration series provided by Statistics New Zealand. The housing starts data is of small enough dimension and length that it can be fit via ML, so that we can make a comparison to the parameter estimates of our proposed EM procedure. In contrast, the daily immigration data is only feasibly fitted by EM, being too large to permit calculation of the MLEs. (In McElroy and Livsey (2022) this same immigration series was fit with ML to a reduced data span; this took weeks and multiple restarts of BFGS from new initializations to converge.)

5.1. Housing Starts

The U.S. Census Bureau's Survey of Construction provides current national and regional statistics on housing starts, completions, and characteristics of single-family housing units and on sale new single-family houses. The four series are from the Survey of Construction of the U.S. Census Bureau, available at https://www.census.gov/construction/nrc/historical_data/index.html. Data was retrieved from this site in June 2021. Information on the the Survey of Construction can be found at <https://www.census.gov/construction/soc>. Data collected includes start date, completion date, sales date, sales price (single-family houses only), and physical characteristics of each housing unit, such as square footage and number of bedrooms. Each month, housing starts, completions, and sales estimates derived from this survey are adjusted by the total numbers of authorized housing units (obtained from the Building Permits Survey) to develop national and regional estimates. Estimates are adjusted to reflect variations by region and type of construction, and to account for late reports and houses started or sold before a permit has been issued.

Here we focus on single-family housing starts for each of the major four regions in the United States: South, West, Northeast (NE), and Midwest (MW). The housing starts time series for these four regions is plotted in Figure 5. Visually, there is clear monthly seasonality as well as similar trends. Seasonal adjustment for production release of this data is accomplished with a standard univariate ($N = 1$) three component model of the form in Equation (12). In a production environment the signal extraction, and subsequent seasonal adjustment, is performed with X-13ARIMA-SEATS (Time Series Staff, Center for Statistical Research, and Methodology 2017) in a univariate fashion, applied individually to each of the four series.

In the multivariate setting we proceed with the same model setup as in Equation (12). Preliminary analysis of the data leads to a trend component requiring a second difference and seasonal sum for the seasonal component. Hence we end up with $J = 2$ components, following

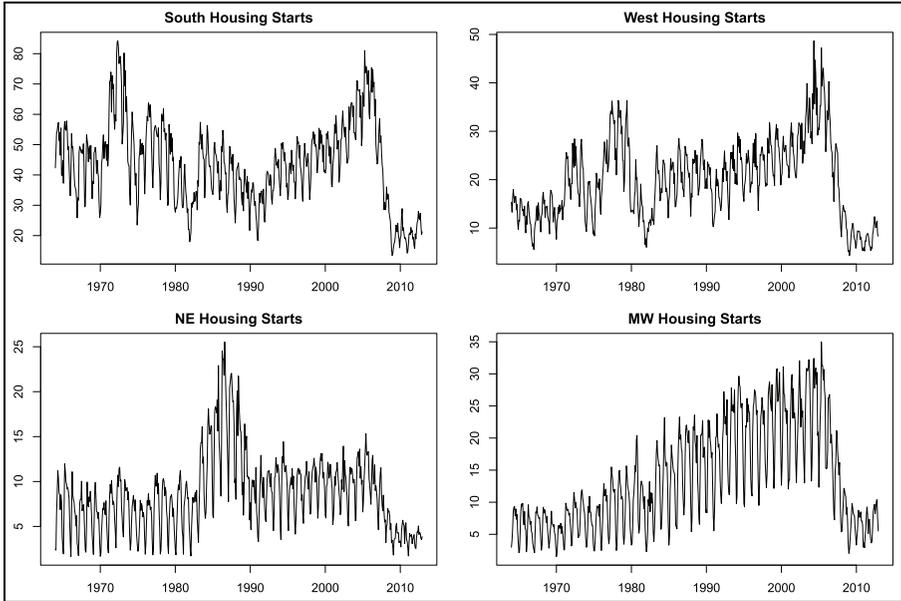


Figure 5. Single family home housing starts broken up by region within the United States. The four regions are the Northeast (NE), South, Midwest (MW), and West. Data comes from the US Census Bureau’s Survey of Construction.

$$(1 - B)^2 S_t^{(1)} = \underline{\Sigma}^{(1)} \sim \text{WN}(0, \Sigma^{(1)})$$

$$(1 + B + B^2 + \dots + B^{11}) S_t^{(2)} = \underline{\Sigma}^{(2)} \sim \text{WN}(0, \Sigma^{(2)}).$$

This implies that our full differencing operator is $\delta(B) = (1 - B)(1 - B^{12})$.

For this model we initialized parameters at default values of $\underline{\Sigma}^{(j)} = I$ (the 4×4 identity matrix) for all j , as well as MOM estimators. Additionally, as a robustness check, we tried multiple perturbations of the MOM estimators. With all initializations, estimates obtained from both the EM and ML methods converged to the same parameter values. Moreover, when initializing at the default values, the first few iterations of the algorithm produced parameter values that were proximate to the MOM estimates in the parameter space. This phenomenon—large first iterations followed by smaller steps as the algorithm approaches convergence—has been noticed in other applications of the EM algorithm (Shumway and Stoffer 1982). The convergence threshold was set based on the relative reduction of the divergence at each iteration as described at the end of Section 3. The timing results were consistent with those presented in Section 4.

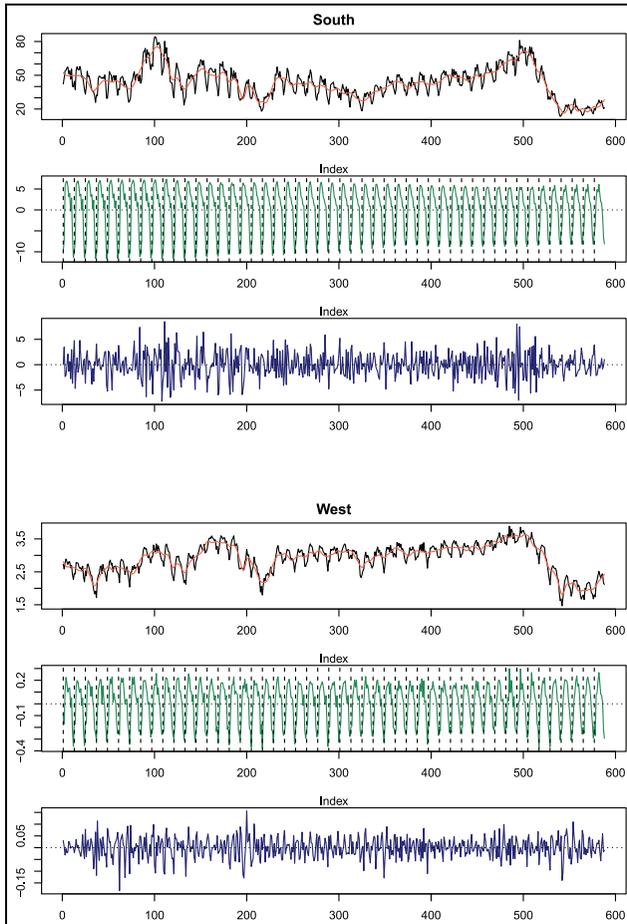


Figure 6. Signal extraction estimates for parameters obtained using EM algorithm. The top three plots are for the South region and the bottom three are for the West region. The original series is in black, trend estimates in red, seasonal component in green, and the irregular in blue.

Results from the EM model fit for the South and West are shown in Figure 6. The NE and MW regions have similar plots. It should be noted that the MLEs were obtainable for this series; the dimension of the housing starts data is $N = 4$ and the sample size is $T = 588$, which is on the boundary of the numerical limits for computing the MLEs with current algorithms and code. The EM algorithm performed without hindrance, and converged to the same location on the likelihood surface as (the more canonical but much longer procedure) ML. The parameter estimates for the model were

$$\hat{\Sigma}^{(1)} = \begin{bmatrix} 2.5963 & \cdot & \cdot & \cdot \\ 1.1831 & 0.9450 & \cdot & \cdot \\ 0.2988 & 0.1837 & 0.2306 & \cdot \\ 0.6882 & 0.4975 & 0.2319 & 0.5588 \end{bmatrix},$$

$$\hat{\Sigma}^{(2)} = \begin{bmatrix} 0.7764 & \cdot & \cdot & \cdot \\ 0.0500 & 0.2079 & \cdot & \cdot \\ 0.0148 & -0.0017 & 0.0965 & \cdot \\ -0.1020 & 0.0403 & 0.0258 & 0.2310 \end{bmatrix},$$

$$\hat{\Sigma}^{(0)} = \begin{bmatrix} 6.3136 & \cdot & \cdot & \cdot \\ 0.0451 & 2.3581 & \cdot & \cdot \\ 0.1352 & 0.0217 & 0.7472 & \cdot \\ 0.2393 & 0.1279 & 0.1443 & 1.2136 \end{bmatrix}.$$

Residual diagnostics and properties of the signal extractions are not further investigated here, as the focus of this manuscript is on presenting and implementing Algorithm 1. The code and methodology used in ES provide signal extraction uncertainty; Figure 6 displays the data, trend, and seasonal components.

5.2. Daily Immigration Data

Here we investigate NZ immigration data. The series consists of New Zealand residents arriving in New Zealand after an absence of less than twelve months. These are public use data produced by Stats New Zealand via a customized extract, and correspond to a portion of the “daily border crossings—arrivals” tab (Total) of the Travel category in the Covid-19 portal: <https://www.stats.govt.nz/experimental/covid-19-data-portal>. The data was downloaded in June 2021. The six-variate series is broken down by the type of person entering or leaving the country. Specifically, the six variables are:

NZArr New Zealand residents arriving in New Zealand after an absence of less than twelve months.

NZDep New Zealand resident departures for an intended period of less than twelve months.

VisArr Overseas residents arriving in New Zealand for a stay of less than twelve months.

VisDep Overseas residents departing New Zealand after a stay of less than twelve months.

PLTArr Permanent and Long Term Arrivals includes overseas migrants who arrive in New Zealand intending to stay for a period of twelve months or more (or permanently), plus New Zealand residents returning after an absence of twelve months or more.

PLTDep Permanent and Long Term Departures includes New Zealand residents departing for an intended period of twelve months or more (or

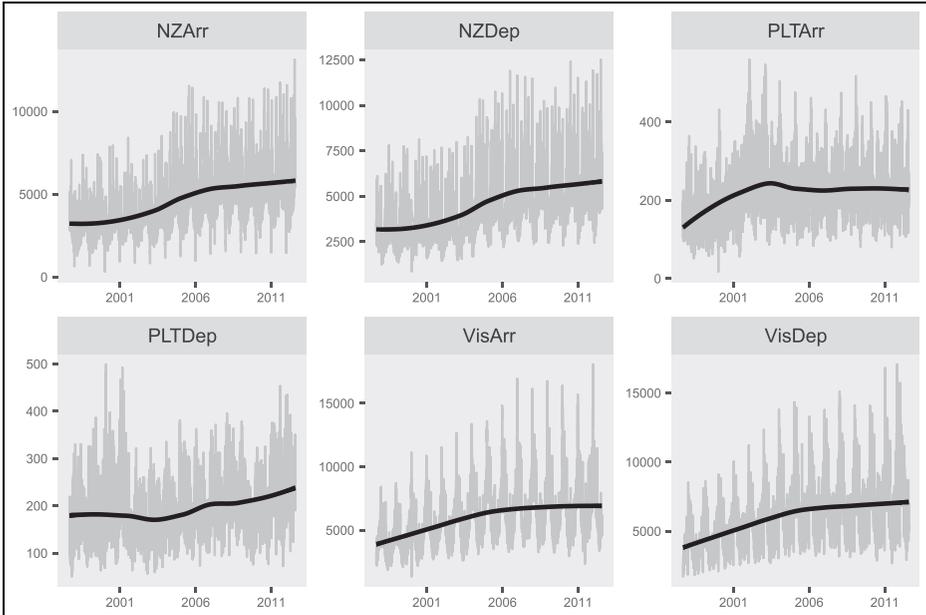


Figure 7. Daily immigration data shown in gray for full span available. The black line is a local polynomial regression fit superimposed to show the long-term trend.

permanently), plus overseas visitors departing New Zealand after a stay of twelve months or more.

The data is plotted in Figure 7 along with a trend line superimposed. The full dataset consists of 5,114 total observations, beginning on September 1, 1997; additionally, a three-year subspan is plotted in Figure 8 to highlight the within-year seasonal patterns present. A more nuanced model is needed for this daily series beyond those presented for the housing starts data in Subsection 5.1. This is primarily due to the larger number of seasonal patterns present in daily time series, as compared with monthly series. Naturally, the aggregation from daily to monthly will mask many seasonal patterns. For example, the Christmas holiday always occurs in December for a monthly series, so its signal can be directly attributed the the seasonal component in a model of the type in Equation (12). However, Christmas does not always occur on the same day of the week and hence the exact dynamics associated with the holiday season immigration change from year to year (Cleveland and Scott 2007; McElroy et al. 2018; Ollech 2021).

The spectral densities of the six immigration series are plotted in Figure 9. Peaks in the spectrum give a natural starting point for the identification of models for the components. Vertical lines are shown to indicate frequencies that have historically been shown to impact economic time series. The red vertical lines are at all

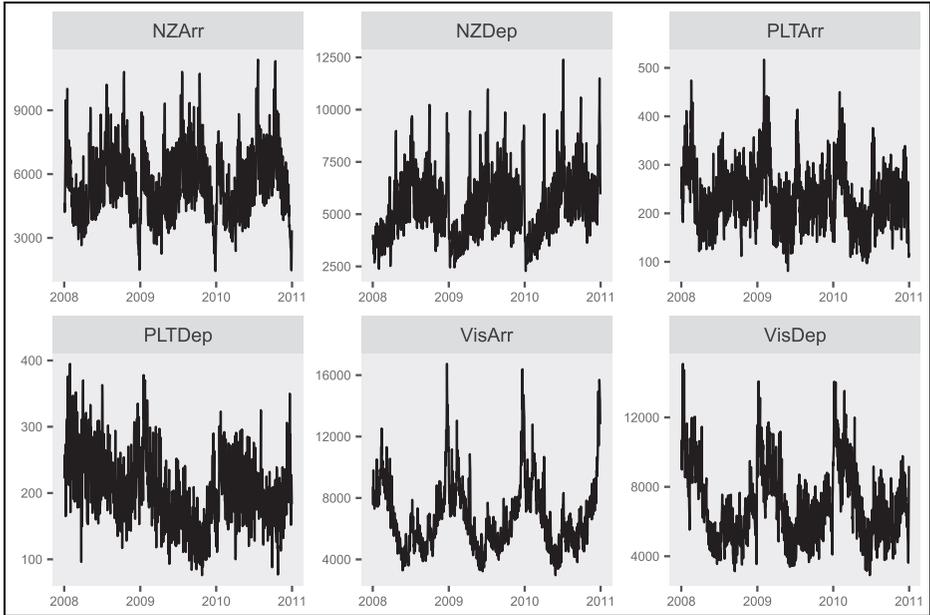


Figure 8. Daily immigration data for 2008 through 2011. The full dataset spans from September 1997 until mid 2011, but the limited window is utilized to better display the daily seasonality.

multiples of weekly frequency. This can be seen as a sinusoidal cycle every $365/7$ time points. These peaks correspond to activity due to the day of the week. The green vertical line is the monthly frequency of twelve occurrences per year, and the annual frequency of 1 is given as a blue vertical line. It should be noted here that it is a difficult signal extraction problem to distinguish the trend and the annual component. This is due to the proximity of frequency $1/365$ (the annual frequency) to frequency zero (corresponding to the trend).

These time series were analyzed in McElroy and Jach (2019), which argued that seasonal differencing, that is, application of the differencing polynomial $1 - B^7$, is appropriate to render the data stationary. For the six-variate series we consider latent components defined by

$$X_t = S_t^{(1)} + S_t^{(2,1)} + S_t^{(2,2)} + S_t^{(2,3)} + S_t^{(0)}, \tag{14}$$

where $S_t^{(1)}$ is a trend component with corresponding difference operator $\delta^{(1)}(B) = 1 - B$. The weekly component $S_t^{(2)}$ with corresponding difference operator $\delta^{(2)}(B) = 1 + B + \dots + B^6$ is further decomposed such that

$$S_t^{(2)} = S_t^{(2,1)} + S_t^{(2,2)} + S_t^{(2,3)}. \tag{15}$$

Each component on the right hand side of Equation (15) is an atomic weekly signal corresponding to one of the three weekly frequencies ($2\pi/7$, $4\pi/7$, and $6\pi/7$)

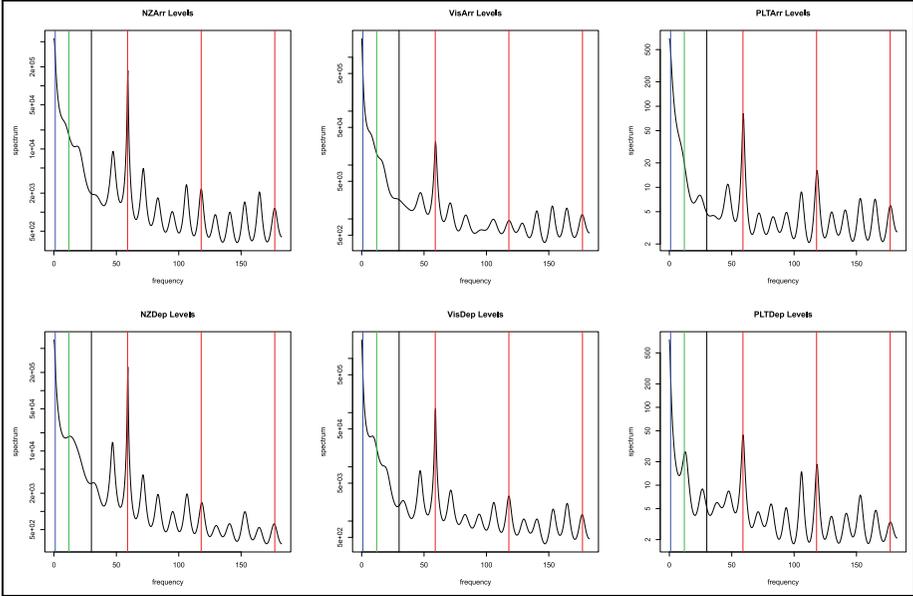


Figure 9. Spectrum for daily immigration data. Vertical lines indicate weekly (red), monthly (green), annual (blue), and trading day (black) effects.

displayed in red in Figure 9. Decomposing $\delta^{(2)}(B)$ down by these three weekly frequencies yields

$$\delta^{(2)}(B) = \delta^{2\pi/7}(B) \cdot \delta^{4\pi/7}(B) \cdot \delta^{6\pi/7}(B), \tag{16}$$

where $\delta^\omega(B) = 1 - 2 \cos(\omega)B + B^2$.

Algorithm 1 was run with the specification Equation (14), using the convergence criteria discussed in Section 4 (and used in Subsection 5.1). The moderate dimension and large sample size together result in a long run-times. The ML method has proven to be, at present, infeasible for a model of this complexity, so only MOM and EM are viable. The estimated signals from the fitted model are shown in Figure 10. The y-axis of these figures has been omitted since the extracted components have been off-set for visual appearance. We can see the VisDep series has the smallest variance in the irregular component, while the PLTDep has the largest variance in the irregular component. The weekly component broken up into the three atomic weekly components is plotted in Figure 11. Again, the y-axis is omitted since all components are off-set to avoid over-plotting.

6. Conclusion

In this manuscript we provide a new computational technique that expands the practical applicability of difference stationary latent component models. This

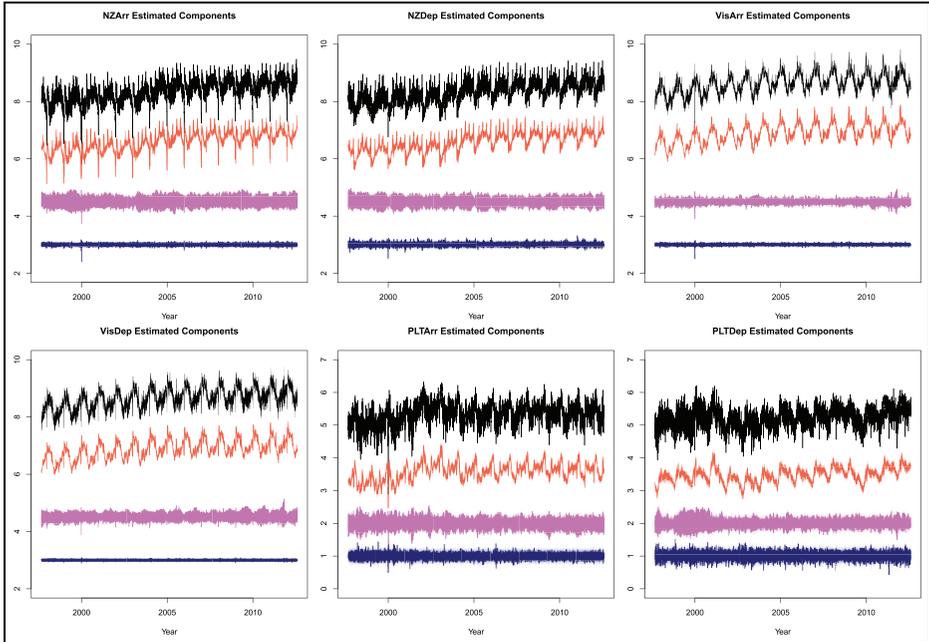


Figure 10. Estimated components for daily immigration data. The original data is in black, estimated trend component in red, weekly seasonal estimate in pink, and irregular in navy.

flexible class of models can be used for a wide range of applications, including multivariate signal extraction. At the present time, estimating parameters with ML estimation has proven numerically infeasible for even moderate-dimensional problems. In this paper we improve upon the available computational methods by deriving an EM algorithm with an explicit formula for the M-step. This derivation allows for the white noise covariance matrix to be computed from a knowledge of the extracted signal and error covariances. Moreover, the formula is fast to compute (no matrix inversions), and hence the speed of the algorithm is only bounded by the numerical efficiency of a particular implementation of signal extraction (e.g., by SS or ES methods). This new methodology renders feasible the fitting of a six-dimensional daily time series, which previously was intractable.

Before this EM method for parameter estimation can be used in the production of official statistics, a number of important issues should be resolved. Firstly, fixed effects such as outliers and moving holiday effects need to be estimated concurrently with the time series model parameters; we suggest an iterative approach (as is used in modeling software such as X-13ARIMA-SEATS), whereby fixed effects are first estimated using regression (e.g., by generalized least squares) and subtracted from the data, followed by the EM iterations, and with this two-step dance repeated unto convergence. These fixed effects could be specified through univariate regressors, which is probably better than a multivariate approach when

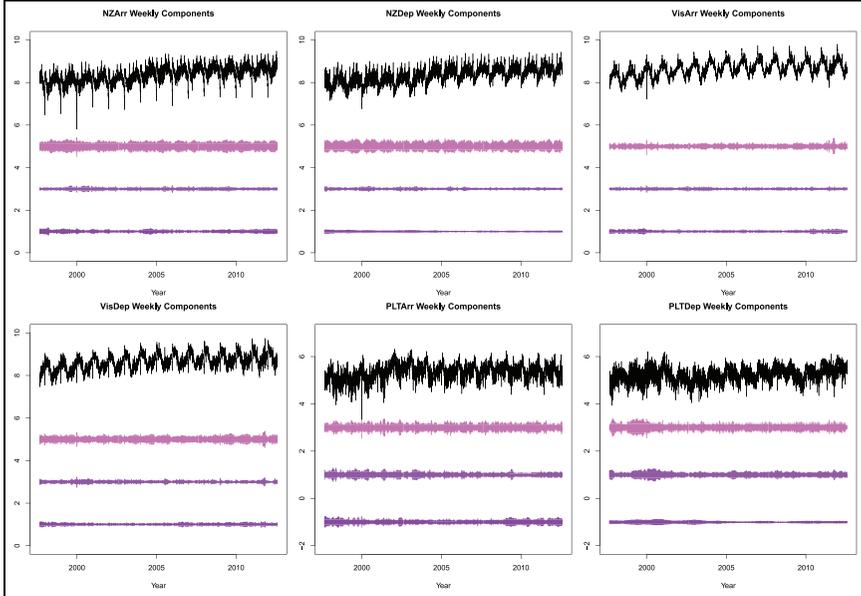


Figure 11. Weekly components for daily immigration data. Top pink line is the estimated once per week effect $\hat{S}_t^{(2,1)}$, the second line the estimated twice per week effect $\hat{S}_t^{(2,2)}$, and the bottom purple line the estimated three times per week effect $\hat{S}_t^{(2,3)}$.

considering extreme values and holiday/calendrical phenomena. Secondly, the differencing polynomial needs to be identified as a crucial aspect of the time series model specification; this could be pursued in a univariate manner, assuming that the differencing polynomials are scalar, perhaps along the lines of model identification used in X-13ARIMA-SEATS. Thirdly, missing data (with potential ragged edges) should be integrated into the proposed algorithm, although this is already possible in ES; ES; McElroy (2022) discusses how ragged edge missing values can be accommodated in likelihood and signal extraction calculations, and such algorithms have been implemented (McElroy and Livsey 2022). Fourthly, as log transformations are commonly used in the analysis of time series data in official statistics, it is of interest to modify seasonal adjustments restated in the original data scale so as to be compatible with annual averages—this might be done by incorporating the proposed methods with benchmarking techniques.

This paper is focused on computation and applications, but does not delve deeply into the specifics of difference-stationary latent component models. The impact and applicability of such models can be greatly extended with the aid of the results presented in the manuscript. Moreover, our results can facilitate a hybrid estimation procedure, wherein EM is run for a few iterations, followed by a likelihood-based estimation that is initialized at those parameter estimates. Currently the latent component covariance matrices $\Sigma(j)$ are assumed to be full rank, but in McElroy (2017) these matrices can have reduced rank, which

corresponds to co-integration; a future research goal is to extend the EM method to the case of a singular $\Sigma(j)$, which would allow us to treat cases of common trends or seasonals.

Funding

The author(s) declared that they received no financial support for the research, authorship, and/or publication of this article.

Disclaimer

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not those of the U.S. Census Bureau.

ORCID iD

James Livsey  <https://orcid.org/0000-0002-4033-4828>

References

- Cleveland, W. P., and S. Scott. 2007. "Seasonal Adjustment of Weekly Time Series with Application to Unemployment Insurance Claims and Steel Production." *Journal of Official Statistics* 23 (2): 209.
- Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data Via the EM Algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (1): 1–22. DOI: <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.
- Golub, G. H., and C. F. van Loan. 2013. *Matrix Computations*. Baltimore, MD: Johns Hopkins University Press. DOI: <https://doi.org/10.56021/9781421407944>.
- Gómez, V. 2016. *Multivariate Time Series with Linear State Space Structure*. Cham: Springer. DOI: https://doi.org/10.1007/978-3-319-28599-3_5.
- Jamshidian, M., and R. I. Jennrich. 1997. "Acceleration of the EM Algorithm by Using Quasi-Newton Methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59 (3): 569–87. DOI: <https://doi.org/10.1111/1467-9868.00083>.
- Kailath, T., A. H. Sayed, and B. Hassibi. 2000. *Linear Estimation*. Upper Saddle River, NJ: Prentice Hall.
- Little, R. J. A., and D. B. Rubin. 2019. *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley & Sons. DOI: <https://doi.org/10.1002/9781119482260>.
- Liu, C., D. B. Rubin, and Y. N. Wu. 1998. "Parameter Expansion to Accelerate EM: The PX-EM Algorithm." *Biometrika* 85 (4): 755–70. DOI: <https://doi.org/10.1093/biomet/85.4.755>.
- Magnus, J. R., and H. Neudecker. 2019. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Hoboken, NJ: John Wiley & Sons. DOI: <https://doi.org/10.1002/9781119541219>.
- Marczak, M., and V. Gómez. 2017. "Monthly US Business Cycle Indicators: A New Multivariate Approach Based on a Band-Pass Filter." *Empirical Economics* 52: 1379–408. DOI: <https://doi.org/10.1007/s00181-016-1108-2>.

- McElroy, T. 2012. "An Alternative Model-Based Seasonal Adjustment That Reduces Over-Adjustment." *Taiwan Economic Forecast and Policy* 43: 33–70.
- McElroy, T. 2017. "Multivariate Seasonal Adjustment, Economic Identities, and Seasonal Taxonomy." *Journal of Business & Economic Statistics* 35 (4): 611–25. DOI: <https://doi.org/10.1080/07350015.2015.1123159>.
- McElroy, T. 2022. "Casting Vector Time Series: Algorithms for Forecasting, Imputation, and Signal Extraction." *Electronic Journal of Statistics* 16 (2): 5534–69. DOI: <https://doi.org/10.1214/22-ejs2068>.
- McElroy, T., and A. Jach. 2019. "Testing Collinearity of Vector Time Series." *The Econometrics Journal* 22 (2): 97–116. DOI: <https://doi.org/10.1093/ectj/uty002>.
- McElroy, T., and J. Livsey. 2022. "Ecce Signum: An R Package for Multivariate Signal Extraction and Time Series Analysis." *arXiv preprint arXiv:2201.02148*. DOI: <https://doi.org/10.48550/arXiv.2201.02148>.
- McElroy, T., and B. Monsell. 2015. "Model Estimation, Prediction, and Signal Extraction for Nonstationary Stock and Flow Time Series Observed at Mixed Frequencies." *Journal of the American Statistical Association (Theory and Methods)* 110: 1284–303. DOI: <https://doi.org/10.1080/01621459.2014.978452>.
- McElroy, T., B. Monsell, and R. J. Hutchinson. 2018. "Modeling of Holiday Effects and Seasonality in Daily Time Series." Technical Report RR2018-01, U.S. Census Bureau, Center for Statistical Research and Methodology, Washington, DC.
- McElroy, T., and A. Roy. 2023. "Model Identification Via Total Frobenius Norm of Multivariate Spectra." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 85 (2): 454–73. DOI: <https://doi.org/10.1093/jrsss/qqad012>.
- McElroy, T., and T. Trimbur. 2015. "Signal Extraction for Nonstationary Multivariate Time Series with Illustrations for Trend Inflation." *Journal of Time Series Analysis* 36: 209–227. DOI: <https://doi.org/10.1111/jtsa.12102>.
- Ollech, D. 2021. "Seasonal Adjustment of Daily Time Series." *Journal of Time Series Econometrics* 13 (2): 235–64. DOI: <https://doi.org/10.1515/jtse-2020-0028>.
- Pinheiro, J. C., and D. M. Bates. 1996. "Unconstrained Parametrizations for Variance-Covariance Matrices." *Statistics and Computing* 6: 289–96. DOI: <https://doi.org/10.007/bf00140873>.
- Shumway, R. H., and D. S. Stoffer. 1982. "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm." *Journal of Time Series Analysis* 3 (4): 253–64. DOI: <https://doi.org/10.1111/j.1467-9892.1982.tb00349.x>.
- Time Series Staff, Center for Statistical Research, and Methodology. 2017. *X-13ARIMASEATS Reference Manual*. US Census Bureau. <http://www.census.gov/ts/x13as/docX13ASHTML.pdf>.
- Wu, C. F. J. 1983. "On the Convergence Properties of the EM Algorithm." *The Annals of Statistics* 11 (1): 95–103. DOI: <https://doi.org/10.1214/aos/1176346060>.
- Zuur, A. F., R. J. Fryer, I. T. Jolliffe, R. Dekker, and J. J. Beukema. 2003. "Estimating Common Trends in Multivariate Time Series Using Dynamic Factor Analysis." *Environmetrics: The Official Journal of the International Environmetrics Society* 14 (7): 665–85. DOI: <https://doi.org/10.1002/env.611>.

Received: September 28, 2023

Accepted: June 7, 2024