Applying the EM Algorithm to Multivariate Signal Extraction

2nd Annual SAPW

James Livsey

Joint work with: Tucker McElroy

Center for Statistical Research and Methodology
U.S. Census Bureau

April 26, 2018

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 author and not necessarily those of the U.S. Census Bureau.

The Problem

Multivariate signal extraction can be accomplished through the use of latent component models. e.g.

$$X_t = T_t + S_t + I_t.$$

- Typically the number of parameters increases quadratically with dimension. Linear filtering theory built on knowledge of variance and covariance matrices.
- Direct approaches to maximum likelihood estimation (MLE) for N-dimensional time series encounter the difficulty of numerical optimization over $\mathbb{R}^p$, where the number of parameters p is large, i.e., $p > 100$.
  - longer times to evaluate the objective function
  - long search times (large p)
  - termination at saddlepoints
  - results sensitive to initialization

United States™ Census Bureau

Research Goals

- Explore Expectation Maximization (EM) Algorithm as method to alleviate some computational burden without losing full MLE appeal.
  - implicitly compute MLEs
  - approximate the true MLEs
- We see the most promise in moderate dimensional signal extraction problems.
- Can be extremely beneficial as we look to analyze higher frequency series.
  - more exotic seasonal patterns (spectral peaks)
  - higher dimensional seasonal vector form

Quick Overview of EM Algorithm

- Start complete likelihood (as if we could observe the latent signals). This will be a random quantity.

$$L(\boldsymbol{\Theta}|X, T, S)$$

- Take conditional expectation to map to a deterministic quantity. (E-step)

$$E\left[-2\log L(\boldsymbol{\Theta}|X, T, S) \;\mid\; T, S, \boldsymbol{\Theta}^{k-1}\right]$$

- Maximize this conditional likelihood (M-step). Get updated parameter estimates $\boldsymbol{\Theta}^k$.

- Iterate until convergence.

United States™ Census Bureau

Model Setup

Consider an N-dimensional vector time series $\{X_t\}$ consisting of J signals denoted $\{S_t^{(j)}\}$ and an irregular $\{S_t^{(0)}\}$, which is stationary. These are additively related:

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

Example:

$$X_t = \underbrace{S_t^{(1)}}_{\text{trend}} + \underbrace{S_t^{(2)}}_{\text{seasonal}} + \underbrace{S_t^{(0)}}_{\text{irregular}}$$

Difference Stationary Process

We formulate our signal extraction for difference-stationary processes.

- There exists a scalar differencing polynomial for each component that maps it to stationarity. For $j = 1, 2, \ldots, J$,

$$\delta^{(j)}(B)\, S_t^{(j)} = \underline{S}_t^{(j)}$$

  where $\underline{S}_t^{(j)}$ is a covariance stationary and mean zero.
- It is assumed $\delta^{(0)}(B) = 1$ because $S^{(0)}$ is stationary.

Stationarity of Observed Data

- It follows that $\delta(B) = \prod_{j=1}^{J} \delta^{(j)}(B)$ is sufficient to reduce $\{X_t\}$ to stationarity.

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

  where $\delta^{(-j)}(B) = \prod_{k \neq j} \delta^{(k)}(B)$.

- Denote over-differenced stationary components as

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

Example

To demonstrate a difference stationary construction and each over-differenced component consider:

$$X_t = \underbrace{S_t^{(1)}}_{\text{trend}} + \underbrace{S_t^{(2)}}_{\text{seasonal}} + \underbrace{S_t^{(0)}}_{\text{irregular}}$$

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

Then our over-differenced series are:

$$\begin{aligned} (1-B^{12})X_t =& (1 + B + B^2 + \ldots + B^{11})\underline{S}^{(1)} \\ &+ (1-B)\underline{S}^{(2)} \\ &+ (1-B^{12})\underline{S}^{(0)} \\ &= \underline{S}^{(-1)} + \underline{S}^{(-2)} + \underline{S}^{(-0)} \end{aligned}$$

Likelihood Function (Incomplete)

- Each latent process is assumed to have the form of a scalar ARMA equation driven by multivariate white noise.
- The block Toeplitz covariance matrix of $\underline{S}_t^{(-j)}$ is given by $\Gamma^{(-j)}$.
- The divergence (-2 times the log Gaussian likelihood with constants removed) is given by:

$$\mathbf{x}'\Gamma^{-1}\mathbf{x} + \log \det \Gamma$$

where $\Gamma = \sum_{j=0}^{J} \Gamma^{(-j)}$ the covariance matrix of the over-differenced sample and $\mathbf{x}$ is the stacked observation vector.

Note, I omit discussion of differenced sample size being smaller than full sample size.

Complete Likelihood Function

Start with full data likelihood

$$L(\boldsymbol{\Theta} | X, S^{(1)}, \ldots, S^{(J)}).$$

This joint pdf factors as follows:

$$p_{\underline{X}, \underline{S}^{(1)}, \ldots, \underline{S}^{(J)}} = p_{\underline{X} | \underline{S}^{(1)}, \ldots, \underline{S}^{(J)}} \prod_{j=1}^{J} p_{\underline{S}^{(j)}}.$$

Rewriting as a divergence gives:

$$\sum_{j=0}^{J} \underline{s}^{(-j)\prime} \, {\Gamma^{(-j)}}^{-1} \, \underline{s}^{(-j)} + \sum_{j=0}^{J} \log \det \Gamma^{(-j)},$$

E and M Steps

- The conditional expectation of this divergence (E-step) can then be computed:

$$\sum_{j=0}^{J} \mathrm{tr} \left\{ {\Gamma^{(-j)}}^{-1} \left[M^{(j)} + \widehat{\underline{s}}^{(-j)} \widehat{\underline{s}}^{(-j)} \right] \right\} + \sum_{j=0}^{J} \log \det \Gamma^{(-j)},$$

where $M^{(j)}$ is error covariance matrix for the jth over-differenced signal.

- Which has a critical point (M-step) at:

$$\Sigma^{(j)} = (T-d)^{-1} \sum_{k,\ell=d+1}^{T} D_{k\ell}^{(j)} \left[M_{\ell k}^{(j)} + \widehat{\underline{s}}_{\ell}^{(-j)} \widehat{\underline{s}}_{k}^{(-j)} \right].$$

where $D^{(j)}$ is inverse covariance matrix of over-differenced portion of $\Gamma^{(-j)}$.

United States Census Bureau

Pseudo Code for EM Procedure

1. Setup your model components.
   This gives global $\Gamma^{(-j)}$ you fix once.
2. Initialize $\Sigma^{(j)}$, $M^{(j)}$, $S^{(j)}$ for all $j = 1, 2, \ldots, J$
3. Update $\Sigma^{(j)} \longrightarrow (T-d)^{-1} \sum_{k,\ell=d+1}^{T} D_{k\ell}^{(j)} \left[M_{\ell k}^{(j)} + \underline{\widehat{s}}_{\ell}^{(-j)} \underline{\widehat{s}}_{k}^{(-j)} \right]$.
4. Run `sigex` with updated covariance structure.
   Get updated $M^{(j)}$, $S^{(j)}$
5. Iterate 3-4 until convergence.

United States Census Bureau

Simulation

To demonstrate a difference stationary construction and each over-differenced component consider:

$$X_t = \underbrace{S_t^{(1)}}_{\text{trend}} + \underbrace{S_t^{(2)}}_{\text{seasonal}} + \underbrace{S_t^{(0)}}_{\text{irregular}}$$

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

Then our over-differenced series are:

$$\begin{aligned}(1-B^{12})X_t =& (1+B+B^2+\ldots+B^{11})\underline{S}^{(1)} \\ &+ (1-B)\underline{S}^{(2)} \\ &+ (1-B^{12})\underline{S}^{(0)} \\ =& \underline{S}^{(-1)} + \underline{S}^{(-2)} + \underline{S}^{(-0)}\end{aligned}$$

Simulation

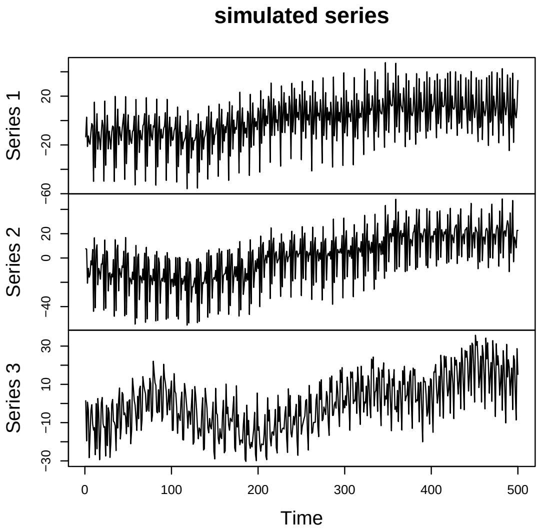
This simulation has values:

- $N = 3$
- $T = 500$
- $\Sigma^{(1)} = \begin{bmatrix} 1 & .75 & 0 \\ .75 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\Sigma^{(2)} = \begin{bmatrix} 1 & .75 & 0 \\ .75 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$,

  $\Sigma^{(0)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

United States Census Bureau

Simulation


simulated series
Series 1
Series 2
Series 3
20
−20
−60
20
0
−40
30
10
−10
−30
0
100
200
300
400
500
Time

Simulation Fit

- EM algorithm converges after 22 iterations (.01 criteria)

- $$\hat{\Sigma^{(1)}} = \begin{bmatrix} 1.10 & \cdot & \cdot \\ 0.83 & 1.08 & \cdot \\ 0.03 & 0.01 & 1.18 \end{bmatrix},$$
  $$\hat{\Sigma^{(2)}} = \begin{bmatrix} 0.90 & \cdot & \cdot \\ 0.56 & 0.90 & \cdot \\ -0.09 & 0.01 & 1.00 \end{bmatrix},$$
  $$\hat{\Sigma^{(3)}} = \begin{bmatrix} 1.22 & \cdot & \cdot \\ 0.57 & 1.45 & \cdot \\ 0.05 & 0.04 & 0.92 \end{bmatrix}$$

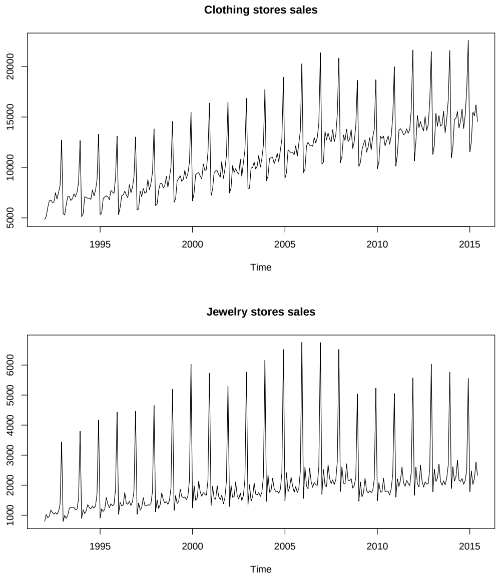
Simulation Fit

- EM algorithm converges after 22 iterations (.01 criteria)

- $$\hat{\Sigma^{(1)}} = \begin{bmatrix} 1.10 & \cdot & \cdot \\ 0.83 & 1.08 & \cdot \\ 0.03 & 0.01 & 1.18 \end{bmatrix},$$
  $$\hat{\Sigma^{(2)}} = \begin{bmatrix} 0.90 & \cdot & \cdot \\ 0.56 & 0.90 & \cdot \\ -0.09 & 0.01 & 1.00 \end{bmatrix},$$
  $$\hat{\Sigma^{(3)}} = \begin{bmatrix} 1.22 & \cdot & \cdot \\ 0.57 & 1.45 & \cdot \\ 0.05 & 0.04 & 0.92 \end{bmatrix}$$

- **The EM estimates converged to the MLE estimates**

Retail Series



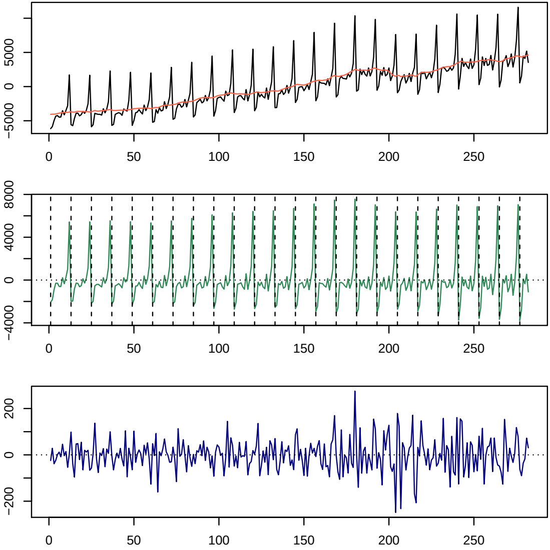
Retail Series Fit

- EM algorithm converges after 38 iterations (.01 criteria)

- $\hat{\Sigma^{(1)}} = \begin{bmatrix} 15050.2 & \cdot \\ 5344.9 & 4226.7 \end{bmatrix}$, $\hat{R^{(1)}} = \begin{bmatrix} 1 & \cdot \\ .67 & 1 \end{bmatrix}$,
  $\hat{\Sigma^{(2)}} = \begin{bmatrix} 22026.4 & \cdot \\ 7298.1 & 6715.5 \end{bmatrix}$, $\hat{R^{(2)}} = \begin{bmatrix} 1 & \cdot \\ .60 & 1 \end{bmatrix}$,
  $\hat{\Sigma^{(3)}} = \begin{bmatrix} 22026.4 & \cdot \\ -2628.3 & 315.3 \end{bmatrix}$, $\hat{R^{(3)}} = \begin{bmatrix} 1 & \cdot \\ -.92 & 1 \end{bmatrix}$

United States Census Bureau

Clothing Series Fit



Future Work

- Further investigate convergence criteria
- Sensitivity analysis
- Dimension/parameter size feasibility

Thank you.

Email: James.A.Livsey@census.gov