Applying the EM Algorithm to Multivariate Signal Extraction

JSM 2019

James Livsey
Joint work with: Tucker McElroy

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

July 30, 2019

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.

Interest in signal extraction

The Census Bureau has a long history of signal extraction and seasonal adjustment.

- 1954 released first “X-11” software for seasonal adjustment.
- 2012 released X-13ARIMA-SEATS our most recent seasonal adjustment variant.

We currently maintain and research new methods to implement in X-13ARIMA-SEATS.

United States™ Census Bureau

Overview of our SA methodology

- Currently we seasonally adjust only **monthly** and **quarterly** time series.

$$X_t = T_t + S_t + I_t$$

  → Can also include exogenous regressors e.g. trading day and moving holidays

- We estimate the latent components $\widehat{T}_t,\ \widehat{S}_t,\ \widehat{I}_t$

- The Census Bureau then publishes a seasonally adjusted series:

$$X_t^{\mathrm{SA}} = \widehat{T}_t + \widehat{I}_t = X_t - \widehat{S}_t$$

United States Census Bureau

Higher Frequency Time Series

There is more and more interest in publishing series at a higher frequency than monthly and quarterly.

However, higher frequency series bring many more seasonal patterns that are not present in monthly/quarterly data.

Consider a daily time series:

$$X_t = \underbrace{S_t^{(1)}}_{\text{trend}} + \underbrace{S_t^{(2)}}_{\text{day of year}} + \underbrace{S_t^{(3)}}_{\text{week of year}} + \underbrace{S_t^{(4)}}_{\text{day of month}} + \underbrace{S_t^{(5)}}_{\text{day of week}} + \underbrace{S_t^{(0)}}_{\text{irregular}}$$

United States™ Census Bureau

Higher Frequency Time Series

There is more and more interest in publishing series at a higher frequency than monthly and quarterly.

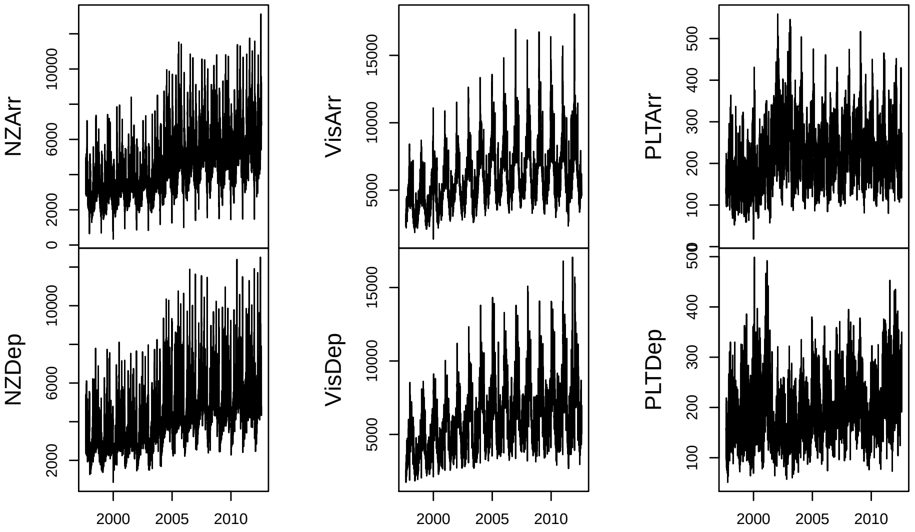
However, higher frequency series bring many more seasonal patterns that are not present in monthly/quarterly data.

Consider a daily time series:

$$X_t = \underbrace{S_t^{(1)}}_{\text{trend}} + \underbrace{S_t^{(2)}}_{\text{day of year}} + \underbrace{S_t^{(3)}}_{\text{week of year}} + \underbrace{S_t^{(4)}}_{\text{day of month}} + \underbrace{S_t^{(5)}}_{\text{day of week}} + \underbrace{S_t^{(0)}}_{\text{irregular}}$$

Publish SA series $X_t^{\text{SA}} = S_t^{(1)} + S_t^{(0)}$?

Daily data


NZArr
NZDep
VisArr
VisDep
PLTArr
PLTDep
0
2000
6000
10000
5000
15000
100
200
300
400
500
2000
2005
2010

New Zealand Immigration Data

NZArr New Zealand residents arriving in New Zealand after an absence of less than 12 months.

NZDep New Zealand resident departures for an intended period of less than 12 months.

VisArr Overseas residents arriving in New Zealand for a stay of less than 12 months.

VisDep Overseas residents departing New Zealand after a stay of less than 12 months.

PLTArr Permanent and Long Term Arrivals includes overseas migrants who arrive in New Zealand intending to stay for a period of 12 months or more (or permanently), plus New Zealand residents returning after an absence of 12 months or more.

PLTDep Permanent and Long Term Departures includes New Zealand residents departing for an intended period of 12 months or more (or permanently), plus overseas visitors departing New Zealand after a stay of 12 months or more.

MLE issues. 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

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

Overview of EM Algorithm

- Start with **complete likelihood**. Act as if we could observe the latent signals. Not observable.

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

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

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

- 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.

Let B denote the backshift operator. $BX_t = X_{t-1}$

- 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:

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

United States Census Bureau

- 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 $\Gamma^{(-j)}$.
- Start with the full data likelihood $L(\boldsymbol{\Theta}|X, S^{(1)}, \ldots, S^{(J)})$.
- Rewrite as a divergence:

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

E and M Steps

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

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

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

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

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

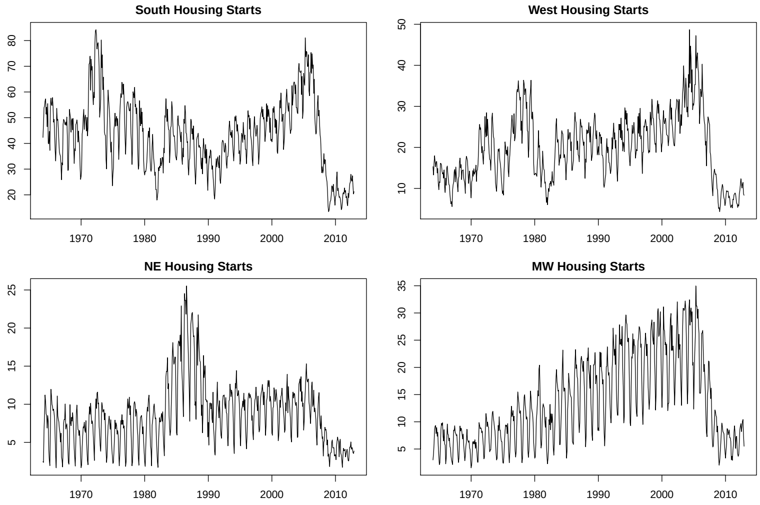
  where $\Gamma^{(j)}$ is inverse covariance matrix of over-differenced portion of $\Gamma^{(-j)}$.

United States™ Census Bureau

Pseudo Code for EM Procedure

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

Housing Starts Data



Proposed Model

A proposed model to the housing starts data:

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

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

Then our full-differencing operator:

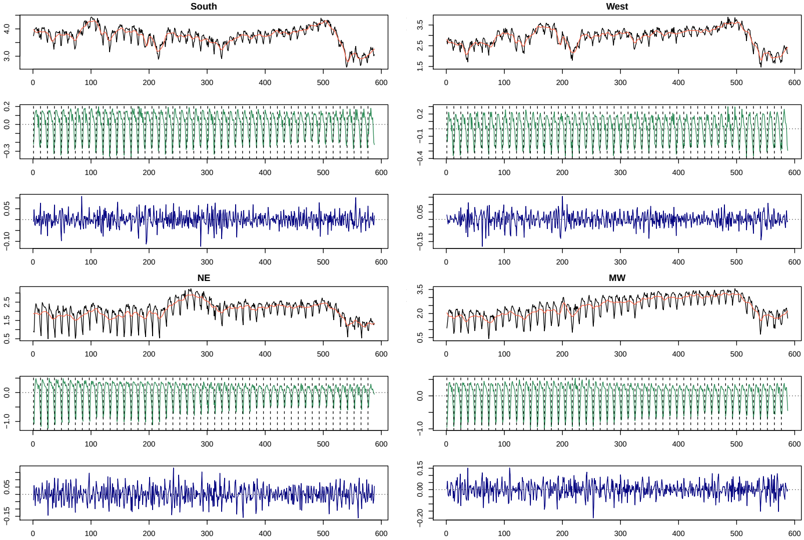
$$\delta(B) = (1-B)(1-B^{12})$$

Housing Starts EM Fit

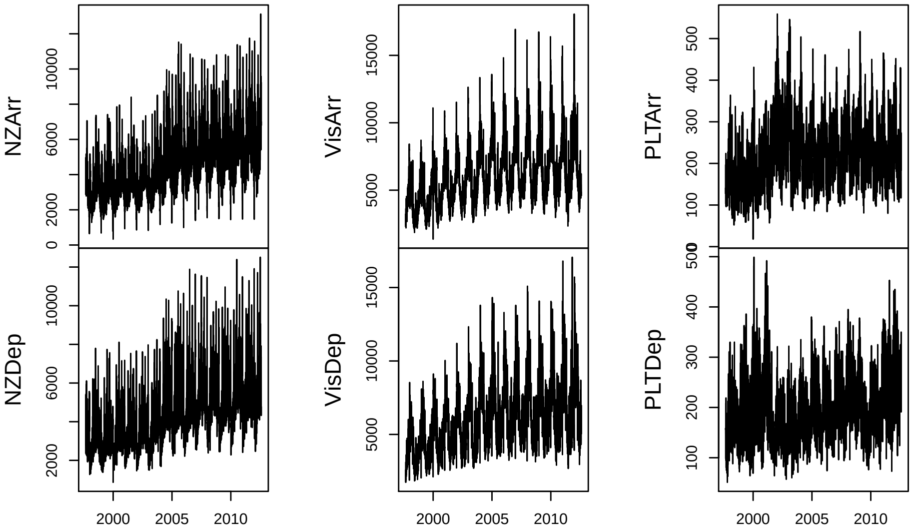
- Likelihood is improved over method of moments estimator
- Strongest correlation among signals was between south and west regions

United States Census Bureau

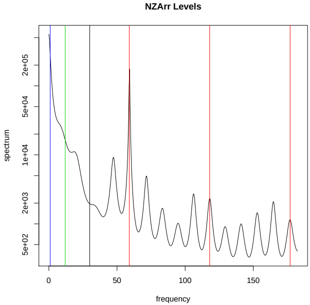
Housing Starts Fit

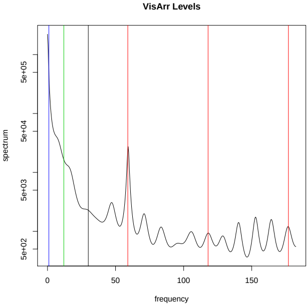


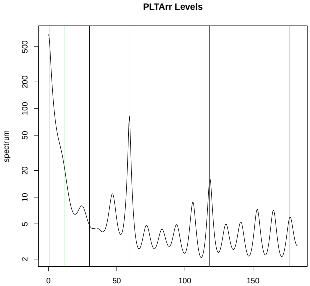
Back to Immigration Series



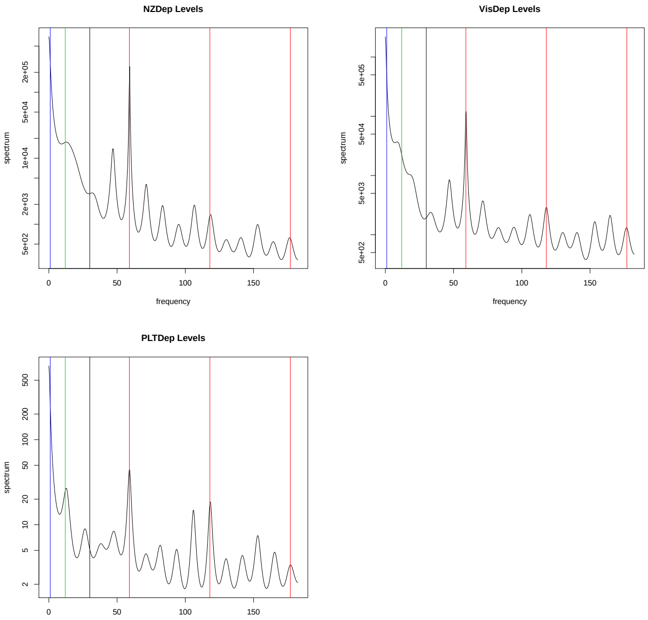
Spectrum of Immigration Series







Spectrum of Immigration Series



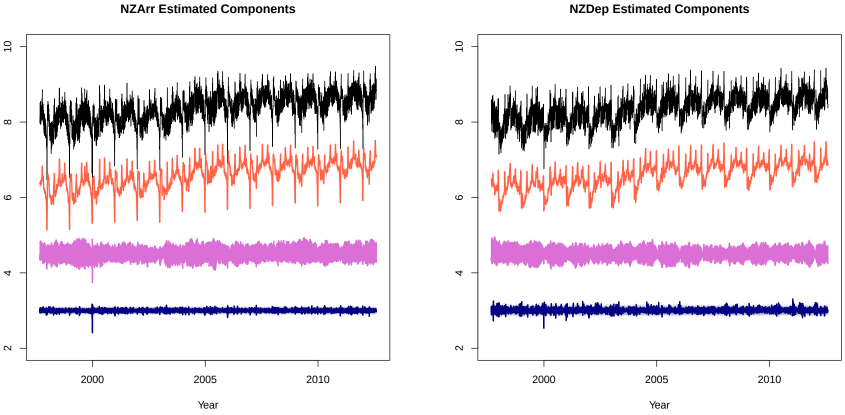
Model for Immigration Series

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

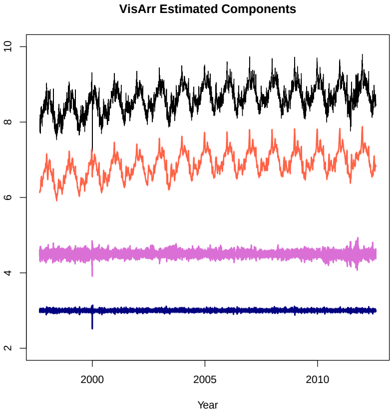
It is advantageous to entertain a more nuanced weekly specification.

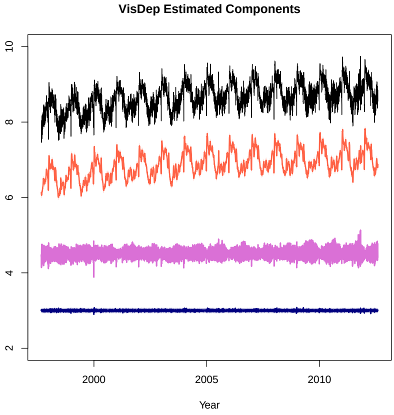
$$S_t^{(2)} = \underbrace{S_t^{(2,1)}}_{\text{1st weekly peak}} + \underbrace{S_t^{(2,2)}}_{\text{2nd weekly peak}} + \underbrace{S_t^{(2,3)}}_{\text{3rd weekly peak}}$$

Estimated Components

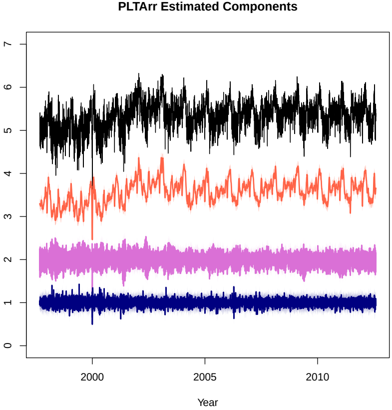


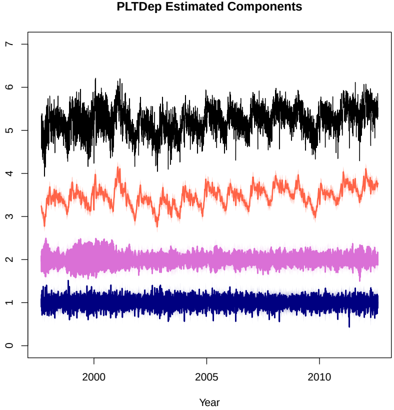
Estimated Components



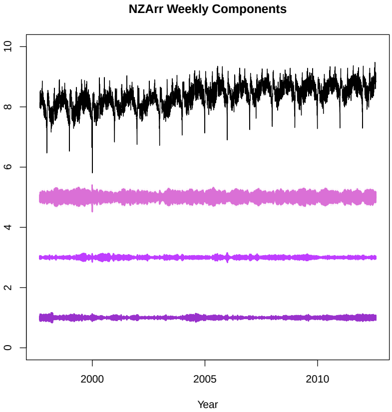


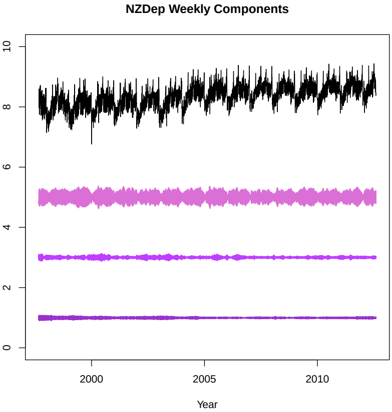
Estimated Components



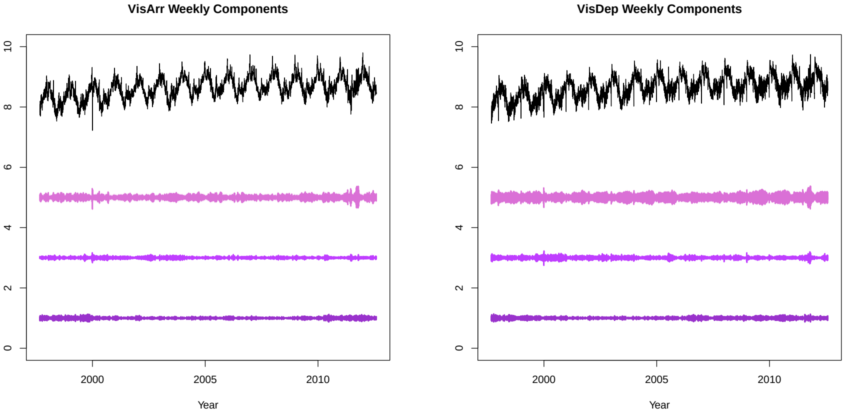


Estimated Components

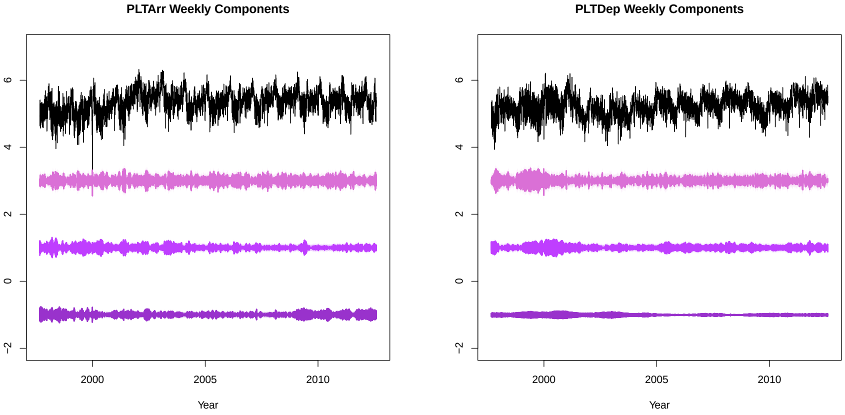




Estimated Components

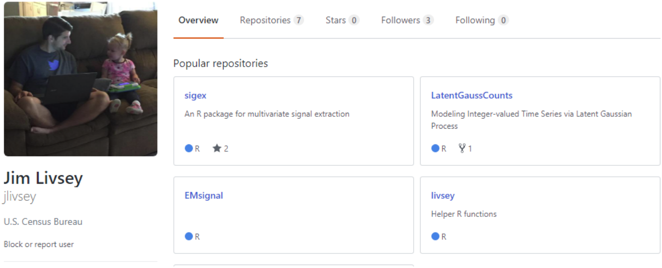


Estimated Components



Conclusions

- The M-step yields an explicit formula for the white noise covariance matrices, which 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 facility with signal extraction.
- Feel free to use this R package. It is up on my Github page



Future Work

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

Thank you.

Email: James.A.Livsey@census.gov