Estimation of Parameterized Spatio-Temporal Dynamic Models

(Bill) Ke Xu
Joint work with Dr. Chris Wikle

Department of Statistics
University of Missouri – Columbia
Spatio-temporal dynamic Models

- A state-space model setup
- Data model $z_t = K_t y_t + \varepsilon_t$ where $\varepsilon_t \sim MVN(0, R_t)$ and $K_t$ is known 0-1 matrix relating data $z_t$ to underlying process $y_t$.
- Process model $y_t = H_\theta y_{t-1} + \eta_t$ where $\eta_t \sim MVN(0, Q)$ and $H_\theta$ is unknown parameterized transition matrix.
- Parameter of Interest: $\Theta = \{R_t, Q_t, H_\theta\}$ to be estimated by General Expectation-Maximization (GEM) algorithms.
- Ultimately want to know $y^s_t = E(y_t|z_1 \ldots z_s)$ and its associated variance $P^s_t = var(y_t|z_1 \ldots z_s)$. In particular, $y_{t-1}^t$, $y^t_t$, $y^T_t$ are called predicted, filtered and smoothed values, respectively.
Kalman Filter and Smoother

- For an overview, see Shumway and Stoffer 2000
- Used for state-space models
- Given $\Theta$, Kalman Filter is an iterative algorithm to obtain $\mathbf{y}_t^{t-1}, \mathbf{y}_t^t$ and the associated variance matrices $\mathbf{P}_t^{t-1}, \mathbf{P}_t^t$, respectively for $t = 1 : T$.
- Given $\Theta$, Kalman Smoother is an iterative algorithm to obtain $\mathbf{y}_t^T$ and its variance matrix $\mathbf{P}_t^T$. 
EM for state-space models (Shumway and Stoffer 1982)

- Provides an iterative method of estimating $\Theta$ for state-space models
- Consider $\{z_1, \ldots, z_T; y_0, y_1, \ldots, y_T\}$ as ”complete data”
- Consists of two steps : E-step and M-step
  - E-step : Compute $g(\Theta)$, the expected value of log likelihood of complete data given $\{z_1 \ldots z_T\}$ and $\Theta^{(j-1)}$
  - M-step : Obtain $\Theta^{(j)}$ by maximizing $g(\Theta)$
- Continue until $\Theta^{(0)}$, $\Theta^{(1)}$, $\Theta^{(2)}$ . . . converges.
EM continued

• Shumway and Stoffer 1982

• Provides explicit updating formula (for unrestricted parameters):
  at $j^{th}$ iteration,
  
  \[- \mathbf{H}(j) = \mathbf{S}_{10}\mathbf{S}_{00}^{-1} \]
  
  \[- \mathbf{R}(j) = \frac{1}{T}\mathbf{B} \]
  
  \[- \mathbf{Q}(j) = \frac{1}{T} (\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}\mathbf{S}'_{10}) \]

where $\mathbf{S}$, $\mathbf{B}$ are computed from Kalman filter and smoothing output.
Standard Error (SE) of EM Estimates

- Boostrapping state-space model e.g., Stoffer and Wall 1991, 2004
- The numerical approach. The Hessian can be approximated by
  \[
  \frac{\partial^2 \log L(z; \Theta)}{\partial \theta_i^2} \approx \log(z; \tilde{\Theta} + \Delta_i) - 2 \times \log L(z; \tilde{\Theta}) + \log L(z; \tilde{\Theta} - \Delta_i)
  \]
  where \( \Delta_i \) is a zero matrix (or vector) except for the \( i^{th} \) position which is 0.01.
- The SE of \( \theta_i \) can be approximated by
  \[
  SE(\theta_i) = \sqrt{\left(-\frac{\partial^2 \log L(z; \Theta)}{\partial \theta_i^2}\right)^{-1}}
  \]
General EM (GEM)

• More general than EM; Same E-step; M-step is different
• At M-step, only requires $g(\Theta)$ increases in value, thereby increasing likelihood value
• Useful when closed form M-step update is hard to obtain.
• Cost to be paid: longer iteration time.
• Two algorithms particular useful for us: ECM and ”GEM based on One Newton-Raphson Step”
GEM: Expectation-Conditional Maximization (ECM)

- Partition $\Theta$ into, say, two parts $\Theta_1$ and $\Theta_2$.
- Update $\Theta_1$ and $\Theta_2$ sequentially or conditionally.
- First, update $\Theta_1$ while holding $\Theta_2 = \Theta_2^{(j-1)}$ by maximizing $g(\Theta_1, \Theta_2^{(j-1)})$ over $\Theta_1$.
- Second, update $\Theta_2$ by maximizing $g(\Theta_1^{(j)}, \Theta_2)$ over $\Theta_2$. 
GEM based on one Newton-Raphson Step

- Consider $\Theta = [\theta; \bar{\Theta}]$. Goal: to update $\theta$.
- $\theta(j) = \theta(j-1) + a(j-1)\delta(j-1)$ where $\delta(j-1) = [\frac{\partial^2 g(\theta)}{\partial \theta^2}]_{\theta=\theta(j-1)}^{-1} \left[ \frac{\partial g(\theta)}{\partial \theta} \right]_{\theta=\theta(j-1)}$
  and $0 < a(j-1) \leq 1$
- When near the maximum, $a(j-1) = 1$ will suffice (also known as EM gradient by Lange 1995).
- This algorithm can be embedded in ECM.
Parameterization of $\Theta$

- Updating formula are derived using symbolic matrix calculus. See "Vector Differential Calculus in Statistics" by M. P. Wand 2002
- See my paper for update formula.
- $R_t$ (variance matrix of measurement error)
  - Simple measurement error: $R_t = \sigma^2 \varepsilon I_t$
  - Measurement error plus some structure: $R_c = c I_t + \sum_{i=1}^{I} \lambda_i A_i$
    where $c > 0$, $A_i A_j = 0$, $\lambda_i \geq 0$ and $A_i$ are known symmetric and idempotent matrices. Only unknown parameter is the scalar $c$. This parameterization is suitable for dimension reduction case.
Parameterization Cont’d

• **Q** (variance matrix of model error)
  – Unrestricted case
  – Diagonal matrix: \( Q(\theta) = \text{diag}(\theta_1, \ldots, \theta_n) \)
  – Spatial Lattice, Conditional Autoregressive (CAR) Model, \( Q(\delta, \rho) = \delta(I - \rho C)^{-1} \)
  – Exponential Covariogram \( Q(\sigma^2_\eta, \theta) = \sigma^2_\eta C(\theta) \)

• **H** (transition matrix)
  – unrestricted case
  – \( H = H(0, \lambda) \), where \( \lambda = (\lambda_1, \ldots, \lambda_m) \) , i.e., the entry of \( H \) is either 0 or \( \lambda_i \)
Example: Palmer Drought Severity Index (PDSI)

- Monthly data 1/1900 - 12/1997
- 107 locations in central USA
- values: -6 (dry) to +6 (wet)
- Left figure: dark circle = dry; open circle = wet; size $\propto$ magnitude
- Left figure: 1st column = data for two months; 2nd column = 1-month ahead prediction for these two months based on our model
EOFs of PDSI data

- Empirical Orthogonal Function (EOF)
- Essentially Principal Component Analysis
- Eigenvalue Decomposition of the sample covariance matrix
- First 10 EOF accounts for about 80% variability
- The next 10 EOF explains additional 10%.
Model for PDSI

• Dimension reduction: let $y_t = \Phi a_t$, where $a_t$ is a low-dimension vector.

• $z_t = K_t \Phi a_t + \varepsilon_t$, where $\Phi$ is known $n \times K$ (EOF) basis matrix and $\varepsilon_t \sim MVN(0, R_c)$, $R_c = c \mathbf{I} + \sum_{i=K+1}^{K+k} \lambda_i \phi \phi'$. 

• $a_t = Ha_{t-1} + \eta_t$, where $\eta_t \sim MVN(0, \sigma_{\eta}^2 Q(\theta))$ and $Q$ is a diagonal matrix.

• Parameters : $c, H, \theta$

• Use regular EM to update all parameters.
PDSI Time Series for 3 locations

- 1-month ahead prediction
- Prediction mean = red line
- 95% band = green dashed line
- Cross = data
Example: Advection-diffusion PDE

- Consider a 1-dimensional $u_t(x)$ process at spatial location $x$ and time $t$, $\frac{\partial u}{\partial t} + \alpha \frac{\partial u}{\partial x} = \beta \frac{\partial^2 u}{\partial x^2}$

- Discretize: $u_t = H(\gamma) u_{t-1} + \eta_t$, where $\gamma = (\gamma_1, \gamma_2, \gamma_3)$. Furthermore, let $\eta_t \sim MVN(0, \sigma^2_\eta C(\theta))$ and $\theta$ is the (scalar) spatial dependence parameter for exponential covariogram.

- Let $z_t$ be the observed the data vector and $z_t = K_t u_t + \varepsilon_t$, where $\varepsilon_t \sim MVN(0, \sigma^2_\varepsilon I)$

- Parameters: $\gamma_1, \gamma_2, \gamma_3, \sigma^2_\eta, \theta, \sigma^2_\varepsilon$

- Use ”One Newton-Raphson Step” to update $\theta$ and use ECM to update all parameters.
Simulated Diffusion Data

- 20 spatial locations;
- 100 time units
- 10% missing
- Goal: to recover the underlying process; so we want $u_t^T$
Same Diffusion Data

- Time Series Plots for two locations
- Data = Square
- Dot = Estimation
- Line = True process
Thank You!