Spatio-Temporal Modeling for Biosurveillance

D.S. Stoffer

University of Pittsburgh

STARMAX Revisited

The Data: The Pennsylvania portion of CDC's National Influenza Surveillance Effort data set. These are weekly mortality reports for pneumonia and influenza in various locations, 1962–present.
Poison models: The basic model for time series of counts is

\[ y_t \sim \text{Poisson}(\eta_t) \]

where \( y_t \sim \text{Poisson}(\eta_t) \) and \( \eta_t = \log(\mu_t) \) with

\[ \mu_t = \exp(\sum_{i=1}^{d} \phi_i \eta_{t-i} + \sum_{i=1}^{d} \sum_{k=1}^{n_k} \exp(\eta_{t-k} \phi_i)) = \exp(\eta_{t-past}) \]

Here, \( \eta_t \) are covariates (or inputs). Some difficulties:

- Not stationary except under restrictive conditions
- No obvious way to analyze multiple series
- Including correlated errors is difficult (GLARMA)
- Interpretation difficult

Some difficulties:

- Linear models, Springer-Verlag
- Fahrmeir & Tutz (1994), Multivariate Statistical Modeling Based on Generalized
- Davis, Dunsmuir & Wang (1999), Modelling Time Series of Count Data
- Asymptotics, Nonparametrics, and Time Series, Marcel-Dekker, 63114
- Zeger (1988), A regression model for time series of counts, Biometrika, 621-629
- Fahrmeir & Tutz (1994), Multivariate Statistical Modeling Based on Generalized

Fahrmeir & Tutz (1994), Multivariate Statistical Modeling Based on Generalized
It would be difficult to model the original data (even under normality) without some transformation, which isn't allowed in Poisson models.

Because it would destroy the Poissonness.

The ACF for Pittsburgh (local trends, long memory, persistent seasonality):
An approach is to transform the data. For example, for Pittsburgh, let 
\[ z_t = \chi z_t \] 
be the original observations. Let \( z_t \) be the variance-stabilizing transformation, and finally, consider the weekly changes:

\[ y_t = z_t - z_{t-1} \]
The ACF and PACF of $y_t$ suggest a simple MA(1) model:

$$y_t = w_t + w_{t-1}$$

where $w_t$ is white noise (or perhaps ARCH). Similarly,

this is an IMA(1,1) for the transformed series; that is,

$$z_t = w_t + w_{t-1}$$

The ACF and PACF of $z_t$ suggest a simple MA(1) model.
A simple estimate yields $\tilde{\theta} = 0.6$ and the fitted model (to the actual data) is:

\[ z_t = z_{t-1} + w_t - \theta w_{t-1} \]

where $z_t = \sqrt{\text{data}_t + 1}$.
How about Allentown?
Allentown? Samemodelas Pittsburgh with $b = 8$. Not bad!

Weeks

Mortality

Pred + 3 SEs

Predicted

Allentown

Same model as Pittsburgh with $\theta = 8$. Not bad!
Toward a more general (spatial) model:

A model that fits the data better is an ARMA(1,1) for $y_t$, that is,

$$I^{-1}m\phi - I m + [I^{-1}m(\theta - \phi)] = (I^{-1}m + I^{-1}x)\phi - I m + I x = I^{-1}y\phi - y$$

$m$: white noise (unobserved - error).

$y_t$: observation equation (observed).

$x_t$: state equation (unobserved - factor).

$$I^{-1}m + I x = Iy$$

$$I^{-1}m(\theta - \phi) + I x\phi = I + Ix$$

State-space form:

$$I^{-1}m\theta - I m + I^{-1}y\phi = y$$

This model can be written in the form $I^{-1}z\land z = y$ and $I^{-1}\phi - I z = y$ where $I$ is a "data". Thus model can be written in the form $I^{-1}\phi = I y$.

A model that fits the data better is an ARMA(1,1) for $y_t$, that is,
The General State-Space Model

Notation:

\[ (\Theta)S = S, \quad (\Theta)Y = Y, \quad (\Theta)I = I \]
\[ (\Theta)V = V, \quad (\Theta)\Phi = \Phi \]

Model uniquely parameterized by \( \Theta \) (\( \gamma \)-dimensional):

\[ S = (\eta, \eta) \text{cov}, \quad Y = (\eta) \text{var}, \quad \Phi = (\eta) \text{var} \]
\[ \text{Input vector} \quad : \eta \]
\[ \text{Observation vector} \quad : \eta \]
\[ \text{State vector} \quad : \eta \]
\[ u, \ldots, I, \eta = \eta \]
\[ 0, \ldots, I, \eta = \eta \]

\[ \text{Input vector} \quad : \eta \]
\[ \text{Observation vector} \quad : \eta \]
\[ \text{State vector} \quad : \eta \]

\[ u = \eta, \quad \eta = \eta \]
\[ \eta = \eta \]

NOTATION:

The General State-Space Model
The Kalman filter yields the prediction:

\[ y_{t+1} = B(\Theta) L_{t} \Theta \]

Innovations:

\[ u_{t} = \epsilon_{t} \]

Estimation of \( \Theta \): The innovations form of the Gaussian likelihood (ignoring a constant) is

\[ \ln L_{\Theta}(y) = \frac{1}{2} n \sum_{t=1}^{n} \ln |\Sigma_{t}(\Theta)| + \frac{1}{2} (y_{t} - \theta_{t})(\Sigma_{t}(\Theta)^{-1} (y_{t} - \theta_{t})) \]

where \( L_{\Theta}(y) \) denotes the likelihood of \( y_{1}, \ldots, y_{n} \) given the data \( \Theta \) assuming normality.

Quasi-GML via Newton-Raphson:

\[ \Theta^{\text{argmax}} = \Theta \]

Notes: Can use a mixture of normal if the data are markedly non-normal. Can include stochastic volatility. Can use a mixture of normal if the data are markedly non-normal. Can include a mixture of normal if the data are markedly non-normal.

\[ \{ (\Theta)^{2} \Sigma_{t-1}(\Theta)^{2} \Sigma_{t}, (\Theta)^{2} \epsilon + (\Theta)^{2} \epsilon \Sigma_{t} \} \]

\[ \sum_{u}^{T} \frac{1}{2} = (\Theta)^{L_{T} \epsilon} - \ln \]

\[ \{ (\Theta)^{2} \Sigma_{t-1}(\Theta)^{2} \Sigma_{t}, (\Theta)^{2} \epsilon \Sigma_{t} \} \]

\[ \{ \Theta, \theta_{1}, \ldots, \theta_{T-1}, \} \mid \theta_{T} \} \]

\[ \text{BTLF} = \text{BTLF} \]

The Kalman filter yields

\[ \text{BTLF} \]

\[ \text{BTLF} \]

\[ \text{BTLF} \]
A Model for an Individual Location (e.g. Pittsburgh):

\[ (\gamma, \eta, \cdots, \gamma, \eta, \theta, \phi) = \Theta \]

Here, \( \gamma \) is a 1 \( \times \) 1 vector of \( \eta \) parameters, including mortality rates from nearby locations at various time lags (contemporaneous values).

Including inputs, \( \gamma \) is an \( \eta \) \( \times \) 1 vector of inputs. \( \gamma \) is a univariate process (\( \eta \) is a vector of (constrained/unconstrained) regression parameters).

Also try \( \gamma \) is a 1 \( \times \) 1 vector of \( \eta \) \( \times \). \( \gamma \) \( \gamma \) is a univariate process (data \( \gamma \) data + \( \gamma \) data = \( \gamma \) could)

\[ u = \gamma \]
\[ \eta + \gamma \eta + \gamma = \gamma \]
\[ u = \gamma \]
\[ \frac{\gamma}{\gamma}(\theta - \phi) + \gamma \phi = \gamma + \gamma \]

\[ y = \gamma \]
The previous model was fit to the Pittsburgh data.

\[ L' = m \theta_0, \theta_1 = \zeta, L = \theta_0(z - 2), \theta_1 = \zeta \]

And

\[ \text{largest SE} = 0.03 \]

The previous model was fit to the Pittsburgh data.
The Innovations (residuals):
Similar model for Philadelphia:

- Predicted
- Predicted + 3 SEs

Weeks

Mortality

Philadelphia Predicted + 3 SEs
\[ \begin{align*}
\mathbf{1}^{-1} \mathbf{m} \mathcal{D} - \mathbf{1} \mathbf{n} + \mathbf{1} \mathbf{f} + \mathbf{1}^{-1} \mathbf{f} \mathcal{D} = \mathbf{1} \hat{\mathbf{y}}
\end{align*} \]

This state-space model implies

- \( \mathcal{D} \) is a matrix of specified spatial constraints
- \( \mathbf{I} \) is a matrix of regression parameters
- \( \Theta \) and \( \Phi \) are diagonal matrices
- \( \mathbf{m} \) is the \( d \)-dimensional noise vector
- \( \mathbf{n} \) is the \( r \)-dimensional vector of exogenous variables
- \( \mathbf{x} \) is the \( p \)-dimensional state vector
- \( \hat{\mathbf{y}} \) is the \( p \)-dimensional observation vector

\[ \begin{align*}
u' \cdots I' = I' \quad & \quad \mathbf{1} \mathbf{m} + \mathbf{1} \mathbf{n} \mathbf{I} + \mathbf{1} \mathbf{x} = \mathbf{1} \hat{\mathbf{y}} \\
u' \cdots I' = I' \quad & \quad \mathbf{1} \mathbf{m} (\Theta - \Phi) \mathcal{D} + \mathbf{1} \mathbf{x} \Phi \mathcal{D} = \mathbf{1} \mathbf{x}
\end{align*} \]

A Spatally Constrained Multivariate Approach

STARMAX:
Themodeliseasilygeneralizedtoarbitraryordersandspatial
constraints.

\[ \dot{1} \times d \ \text{is} \ \dot{1} \times d \ \text{and} \ \dot{\mathbf{f}} \ \text{is} \ ]

\[ \dot{\mathbf{n}} \ \text{are exogenous variables} \ \text{inputs,} \ 1 \ \Theta, \ 1 \ \Phi, \ 1 \ \Phi \ \text{are diagonal matrices, as before. The
matrixes} \ 1 \ \Phi \ \text{are first order and second order spatial constraints.} \]

\[ \dot{1} \mathbf{n} + \dot{1} \mathbf{n} \ 1 + \dot{1} \mathbf{n} \ 0 + \dot{1} \mathbf{n} \ 1 = \ 1 \dot{\mathbf{f}} \]

\[ \text{Yields the} \ \text{STARMAX(2)} \ \text{model:} \]

\[ \dot{1} \mathbf{n} + \dot{1} \mathbf{n} \ 1 + \dot{1} \mathbf{x} \ [0 \ 1] = \ 1 \dot{\mathbf{f}} \]

\[ \dot{1} \mathbf{n} \left[ \begin{array}{c} \dot{1} \Phi \ 0 \\ \left(1 \Theta - 1 \Phi \right) \ 1 \\ \Phi \ 1 \end{array} \right] + \dot{1} \mathbf{x} \left[ \begin{array}{cc} 0 & 1 \\ \Phi \ 1 \end{array} \right] = \ 1 + \dot{1} \mathbf{x} \]

\[ \text{constrants. For example,} \]

\[ \text{The model is easily generalized to arbitrary orders and spatial
constraints.} \]
<table>
<thead>
<tr>
<th>Scranition</th>
<th>0.13</th>
<th>0.02</th>
<th>1.5</th>
<th>1.6</th>
<th>6%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>0.04</td>
<td>0.02</td>
<td>1.9</td>
<td>2.3</td>
<td>17%</td>
</tr>
<tr>
<td>Pittsburg</td>
<td>0.09</td>
<td>0.03</td>
<td>0.73</td>
<td>0.72</td>
<td>22%</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.07</td>
<td>0.03</td>
<td>0.72</td>
<td>0.72</td>
<td>26%</td>
</tr>
<tr>
<td>Allentown</td>
<td>0.09</td>
<td>0.03</td>
<td>0.83</td>
<td>0.82</td>
<td>19%</td>
</tr>
</tbody>
</table>

Estimates & Errors:

<table>
<thead>
<tr>
<th>Original data units</th>
<th>phi</th>
<th>theta</th>
<th>rmspest</th>
<th>dev</th>
<th>improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allentown</td>
<td>0.09</td>
<td>0.03</td>
<td>0.83</td>
<td>0.82</td>
<td>19%</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.07</td>
<td>0.03</td>
<td>0.72</td>
<td>0.72</td>
<td>26%</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>0.09</td>
<td>0.03</td>
<td>0.73</td>
<td>0.72</td>
<td>22%</td>
</tr>
<tr>
<td>Reading</td>
<td>0.04</td>
<td>0.02</td>
<td>0.77</td>
<td>0.77</td>
<td>17%</td>
</tr>
<tr>
<td>Scranton</td>
<td>0.13</td>
<td>0.02</td>
<td>0.81</td>
<td>0.81</td>
<td>6%</td>
</tr>
</tbody>
</table>

\[
\begin{bmatrix}
\text{corr}^{-1}(\tilde{f})^{(t-1)} D \\
\end{bmatrix}
\]