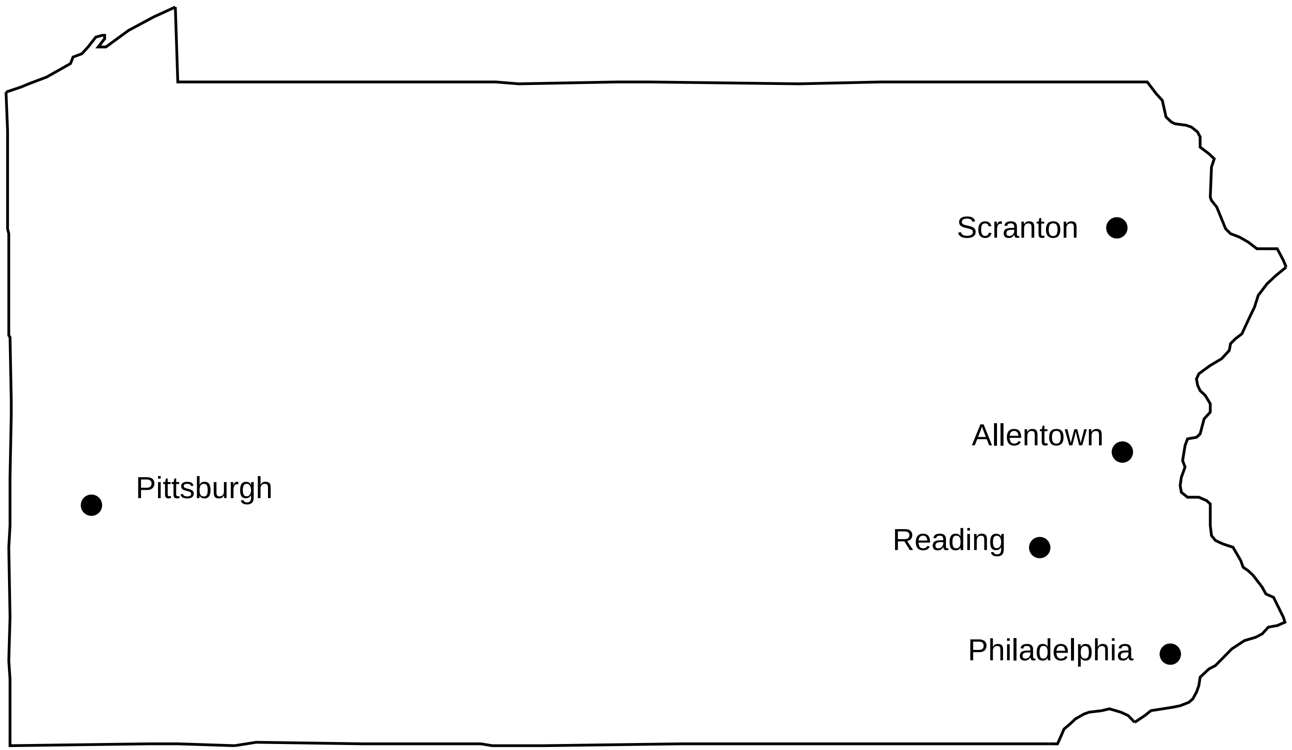
Spatio-Temporal Modeling for Biosurveillance

D.S. Stoffer
University of Pittsburgh

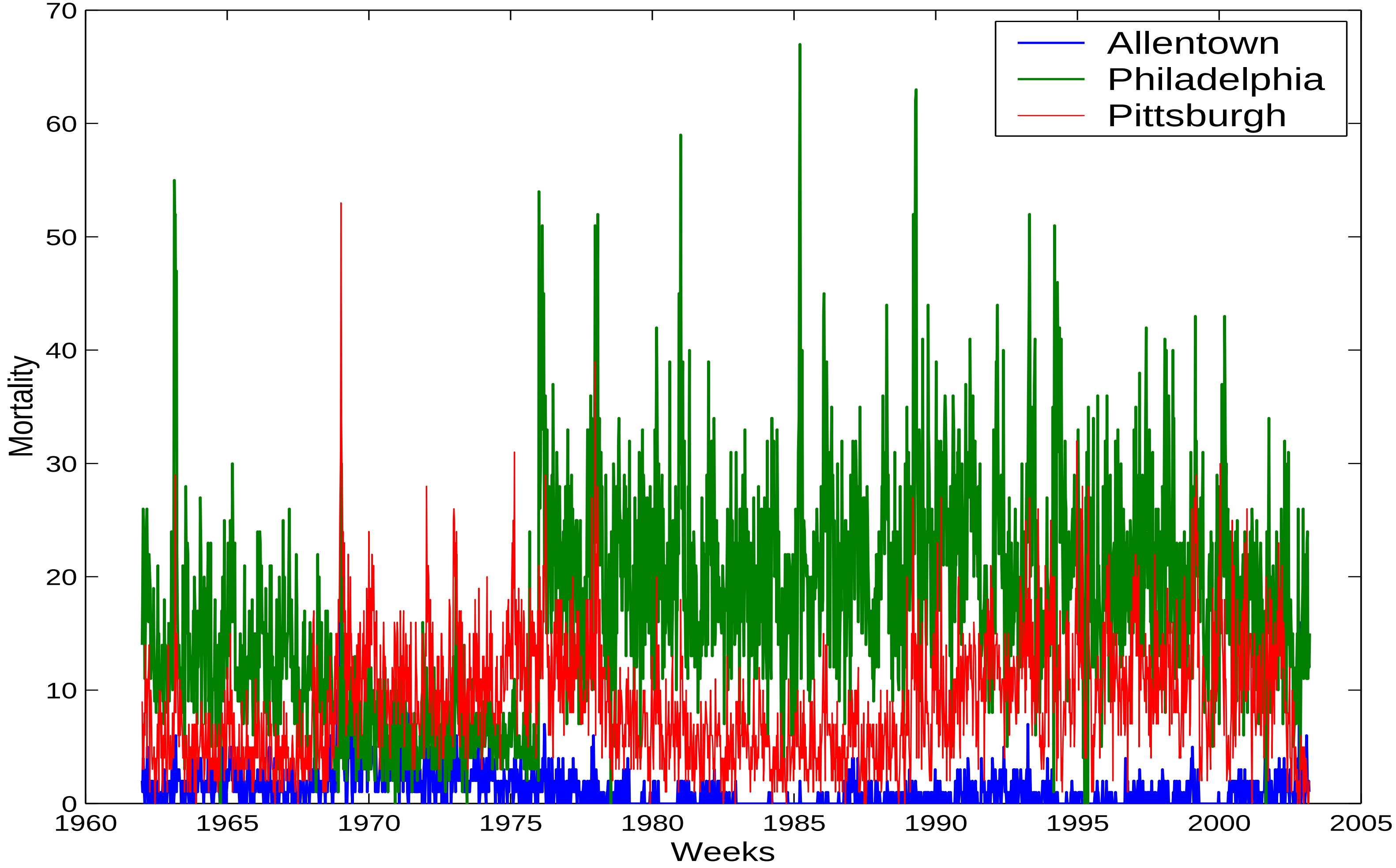
STARMAX Revisted

Stoffer (1986). Estimation and Identification of Space-Time ARMAX Models with Missing Data *JASA*, 762-772. *(An approach to space-time modeling using a spatially constrained state space model.)*

The Data: The Pennsylvania portion of CDCs National Influenza Surveillance Effort data set. These are weekly mortality reports for pneumonia and influenza in various locations, 1962–present.



Three Series:



Poisson models: The basic model for time series of counts is $y_t \big| y_{t-1}, y_{t-2}, \ldots \sim \text{Poisson}(\mu_t)$, where

$$\log \mu_t = \Gamma \boldsymbol{u}_t + \sum_{i=1}^{p} \phi_i y_{t-i}.$$

Here, $\boldsymbol{u}_t$ are covariates (or inputs). Some difficulties:

- **not stationary except under restrictive conditions**
- **no obvious way to analyze multiple series**
- **interpretation difficult:** $E\{y_t|past\} = \exp(\Gamma \boldsymbol{u}_t) \exp(\sum_{i=1}^{p} \phi_i y_{t-i})$
- **including correlated errors is difficult (GLARMA) .**

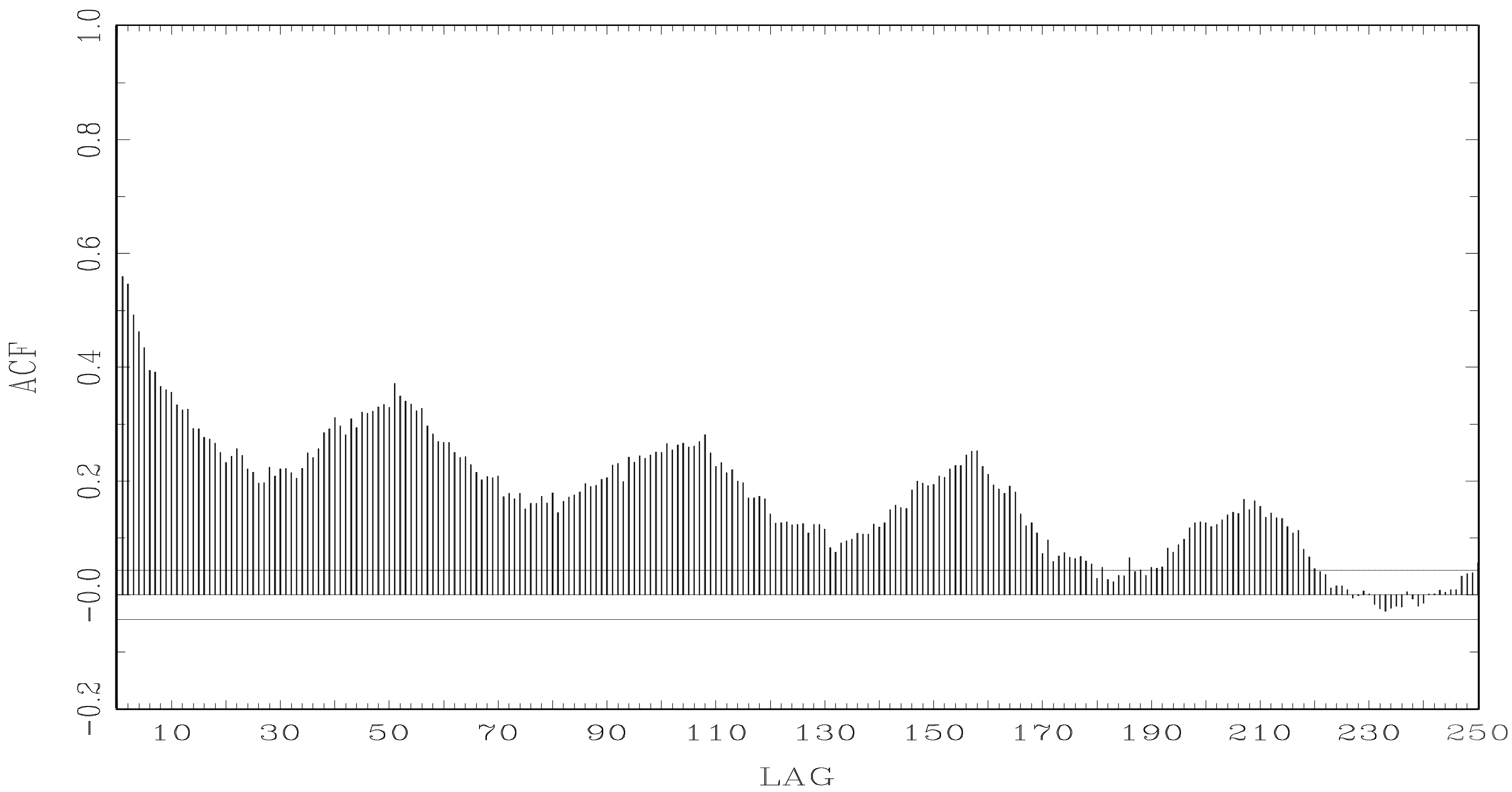
Davis, Dunsmuir & Wang (1999). Modelling Time Series of Count Data. *Asymptotics, Nonparametrics, and Time Series*, Marcel–Dekker, 63114.

Fahrmeir & Tutz (1994) *Multivariate Statistical Modeling Based on Generalised Linear Models*, Springer-Verlag.

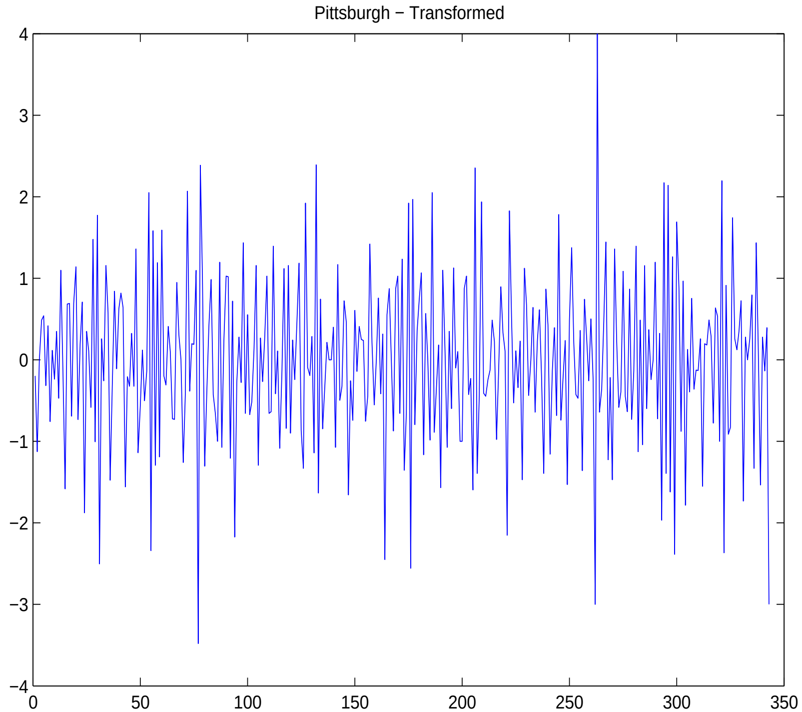
Zeger (1988). A regression model for time series of counts. *Biometrika*, 621- 629.

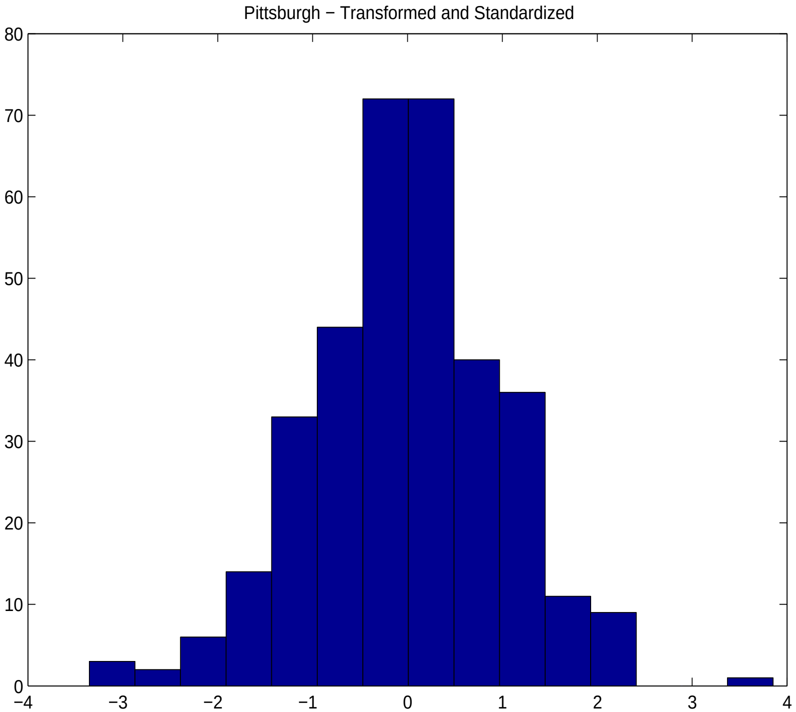
It would be difficult to model the original data (even under normality) without some transformation, which isn't allowed in Poisson models, because it would destroy the Poisonness.

The ACF for Pittsburgh (local trends, long memory, persistent seasonality):

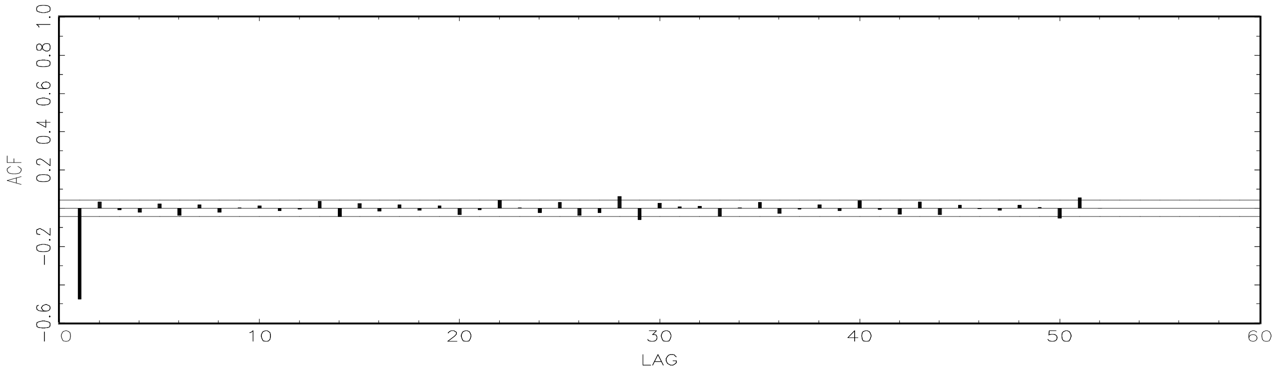


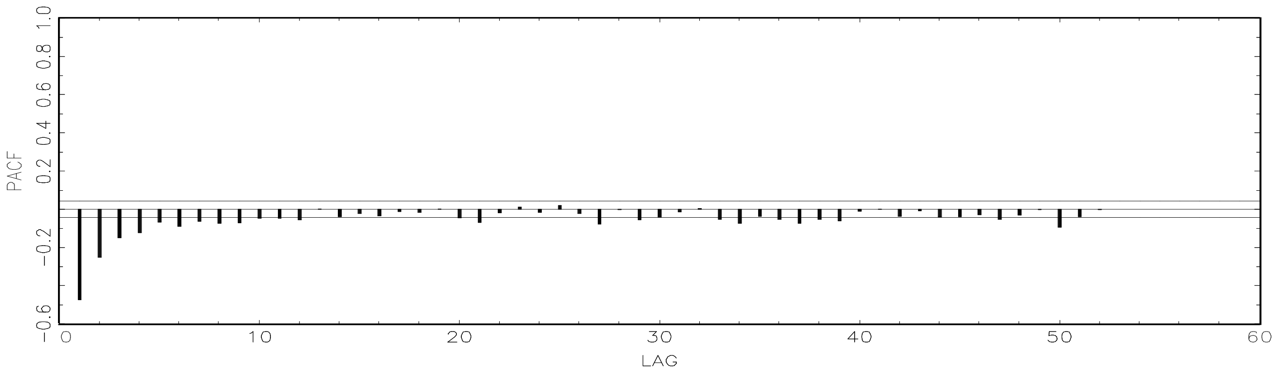
An approach is to transform the data. For example, for Pittsburgh, let z_t^* be the original observations. Let $z_t = \sqrt{z_t^* + 1}$ (variance stabilizing transformation), and finally, consider the weekly changes: $y_t = z_t - z_{t-1}$



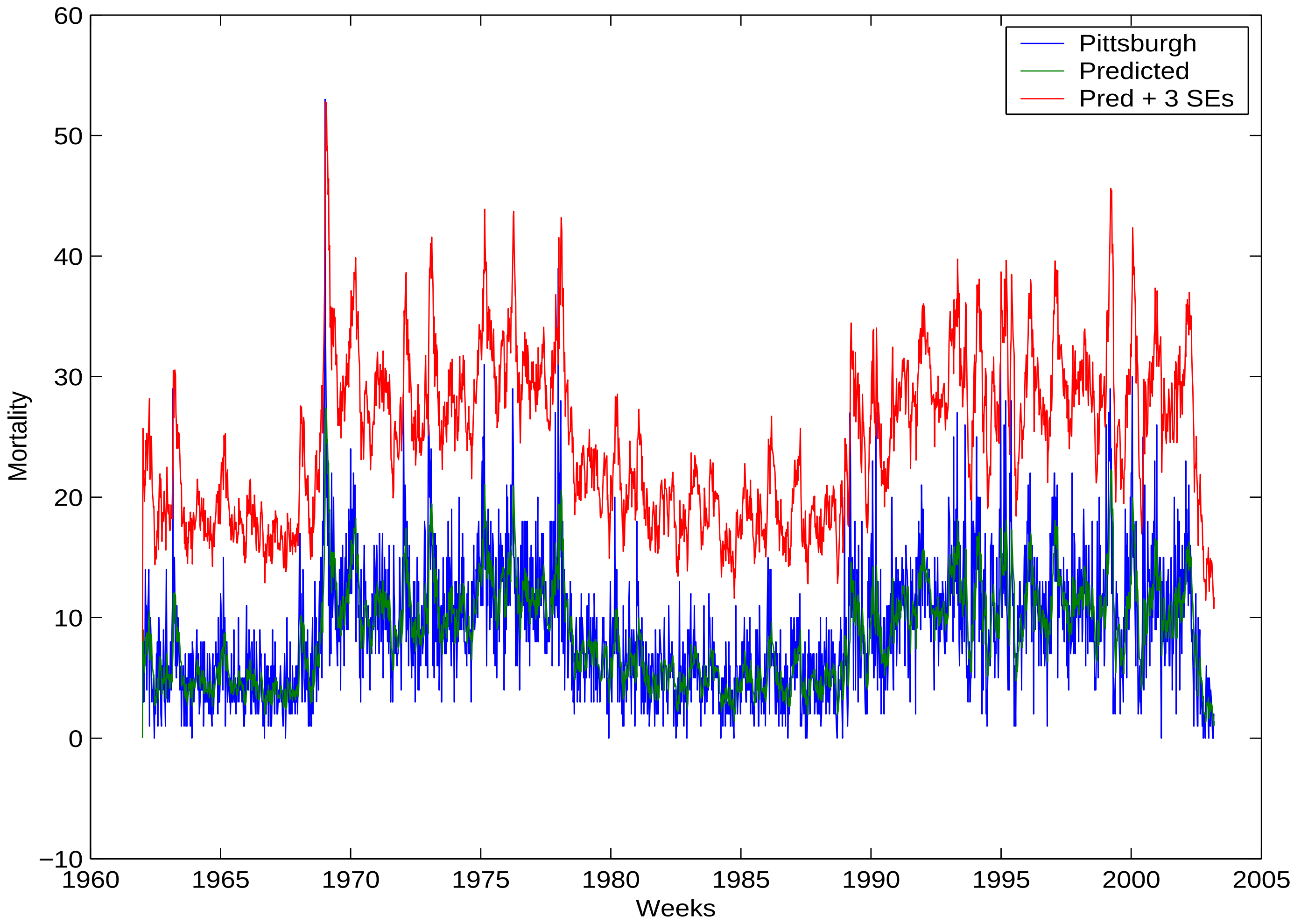


The ACF and PACF of y_t suggest a simple MA(1) model:
$y_t = w_t - \theta w_{t-1}$ where w_t is white noise (or perhaps ARCH). Similary, this is an IMA(1,1) for the transformed series; that is,
$z_t = z_{t-1} + w_t - \theta w_{t-1}$. (recall $y_t = z_t - z_{t-1}$ where "$z = \sqrt{\text{data} + 1}$")



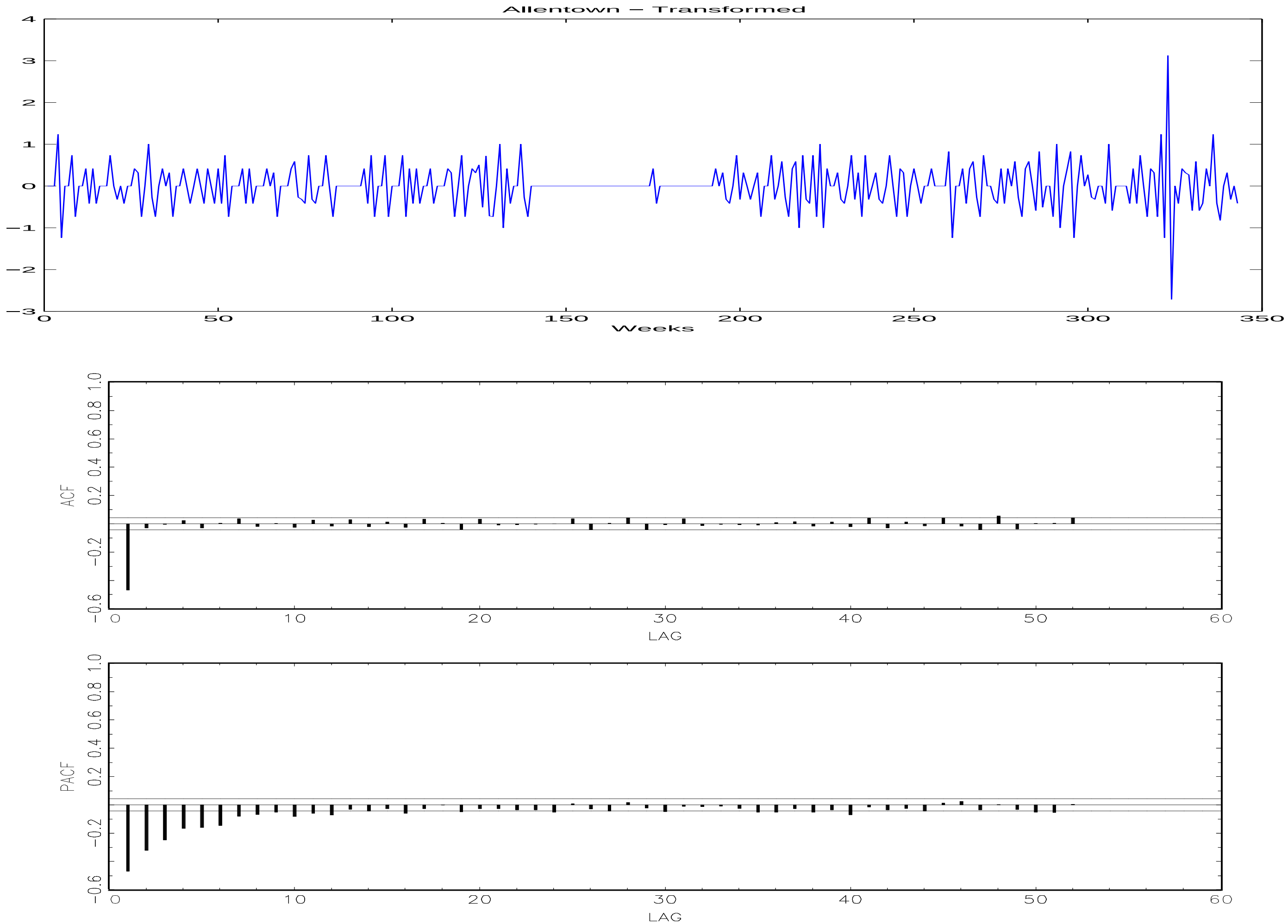


A simple estimate yields $\widehat{\theta} = .6$ and the fitted model (to the actual data) is:

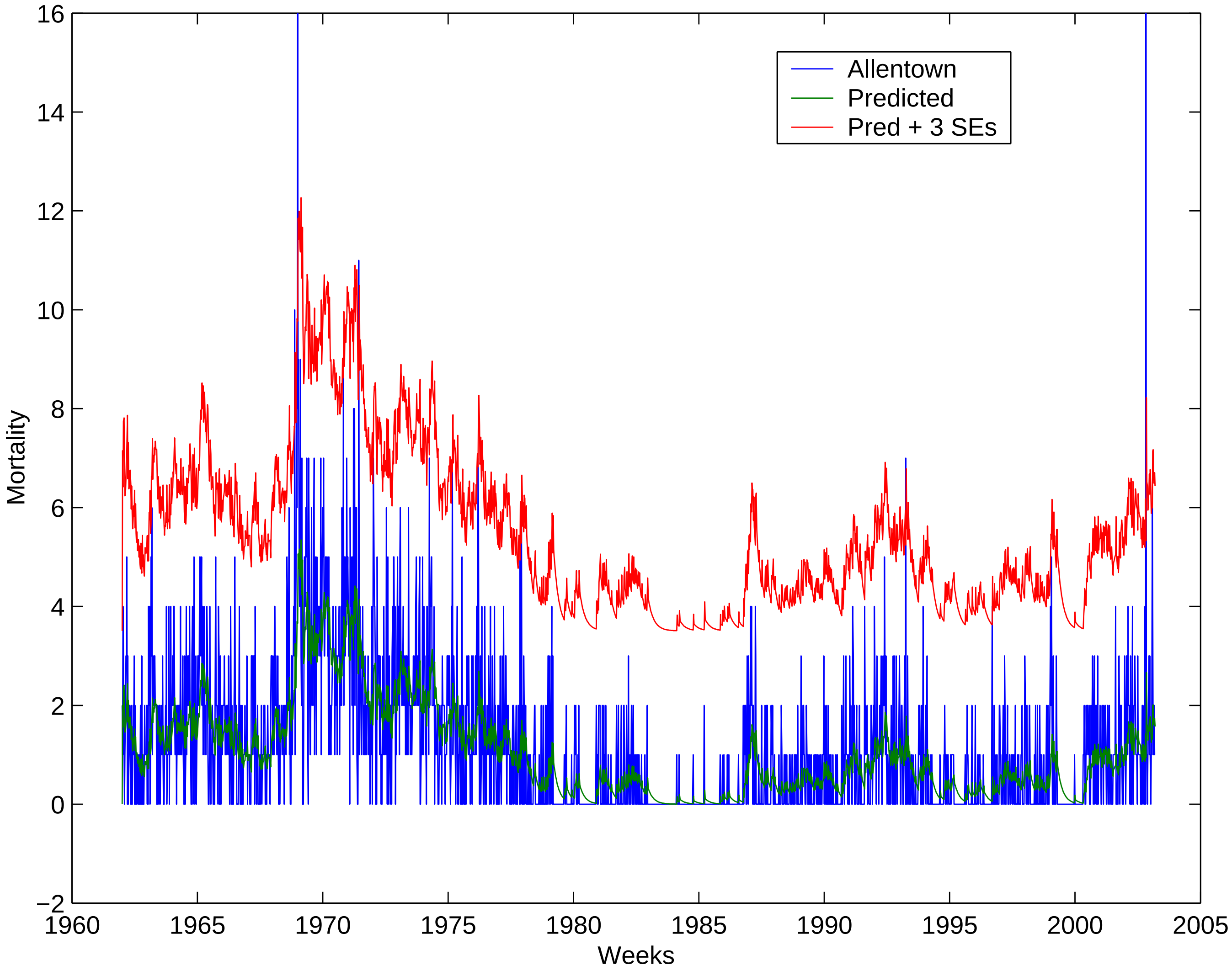


Recall $z_t = z_{t-1} + w_t - \theta w_{t-1}$ where "$z = \sqrt{\text{data} + 1}$".

How about Allentown?



Allentown? Same model as Pittsburgh with $\widehat{\theta} = .8$. 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,

$$y_t = \phi y_{t-1} + w_t - \theta w_{t-1}$$

where $y_t = z_t - z_{t-1}$ and "$z = \sqrt{\text{data} + 1}$". This model can be written in state-space form†:

$$x_{t+1} = \phi x_t + (\phi - \theta) w_t$$

$$y_t = x_t + w_t$$

x_t: **state equation** (unobserved - factor)
y_t: **observation equation** (observed)
w_t: **white noise** (unobserved - error).

† $y_t - \phi y_{t-1} = x_t + w_t - \phi(x_{t-1} + w_{t-1}) = [(\phi - \theta) w_{t-1}] + w_t - \phi w_{t-1}$

The General State-Space Model[†]

NOTATION:

$$\boldsymbol{x}_{t+1} = \Phi \boldsymbol{x}_t + \Upsilon \boldsymbol{u}_t + G \boldsymbol{w}_t \qquad t = 0, 1, ..., n$$

$$\boldsymbol{y}_t = A_t \boldsymbol{x}_t + \Gamma \boldsymbol{u}_t + \boldsymbol{v}_t \qquad t = 1, ..., n$$

$\boldsymbol{x}_t$: p-dimensional **state** vector

$\boldsymbol{y}_t$: q-dimensional **observation** vector

$\boldsymbol{u}_t$: r-dimensional fixed **input** vector

$\text{var}(\boldsymbol{w}_t) = Q$, $\text{var}(\boldsymbol{v}_t) = R$, $\text{cov}(\boldsymbol{w}_t, \boldsymbol{v}_t) = S$

Model uniquely parameterized by Θ (k-dimensional):

$\Phi = \Phi(\Theta)$, $\Upsilon = \Upsilon(\Theta)$, $G = G(\Theta)$, $Q = Q(\Theta)$, $A_t = A_t(\Theta)$,
$\Gamma = \Gamma(\Theta)$, $R = R(\Theta)$, $S = S(\Theta)$.

[†]Shumway & Stoffer (2000, Ch 4). *Time Series Analysis and Its Applications.* New York: Springer.

The Kalman filter yields

Prediction: $\boldsymbol{y}_t^{t-1} = \mathrm{BLP}\{\boldsymbol{y}_t \,\big|\, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_1; \Theta\}$

Innovations: $\boldsymbol{\epsilon}_t(\Theta) = \boldsymbol{y}_t - \boldsymbol{y}_t^{t-1}$; $\mathrm{var}\{\boldsymbol{\epsilon}_t(\Theta)\} = \Sigma_t(\Theta)$

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

$$-\ln L_Y(\Theta) = \frac{1}{2}\sum_{t=1}^{n}\left\{\ln|\Sigma_t(\Theta)| + \boldsymbol{\epsilon}_t(\Theta)'\Sigma_t(\Theta)^{-1}\boldsymbol{\epsilon}_t(\Theta)\right\}$$

where $L_Y(\Theta)$ denotes the likelihood of Θ given the data $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ assuming normality.

Quasi-GML via Newton-Raphson: $\widehat{\Theta} = \mathrm{argmax}_\Theta L_Y(\Theta)$

Notes: Can use a mixture of normal if the data are markedly non-normal. Can include stochastic volatility. CUSUM possible on standardized innovations $\Sigma_t^{-1/2}\boldsymbol{\epsilon}_t$.

A Model for an Individual Location (e.g. Pittsburgh):

$$x_{t+1} = \phi x_t + (\phi - \theta) w_t, \qquad t = 0, 1, ..., n$$

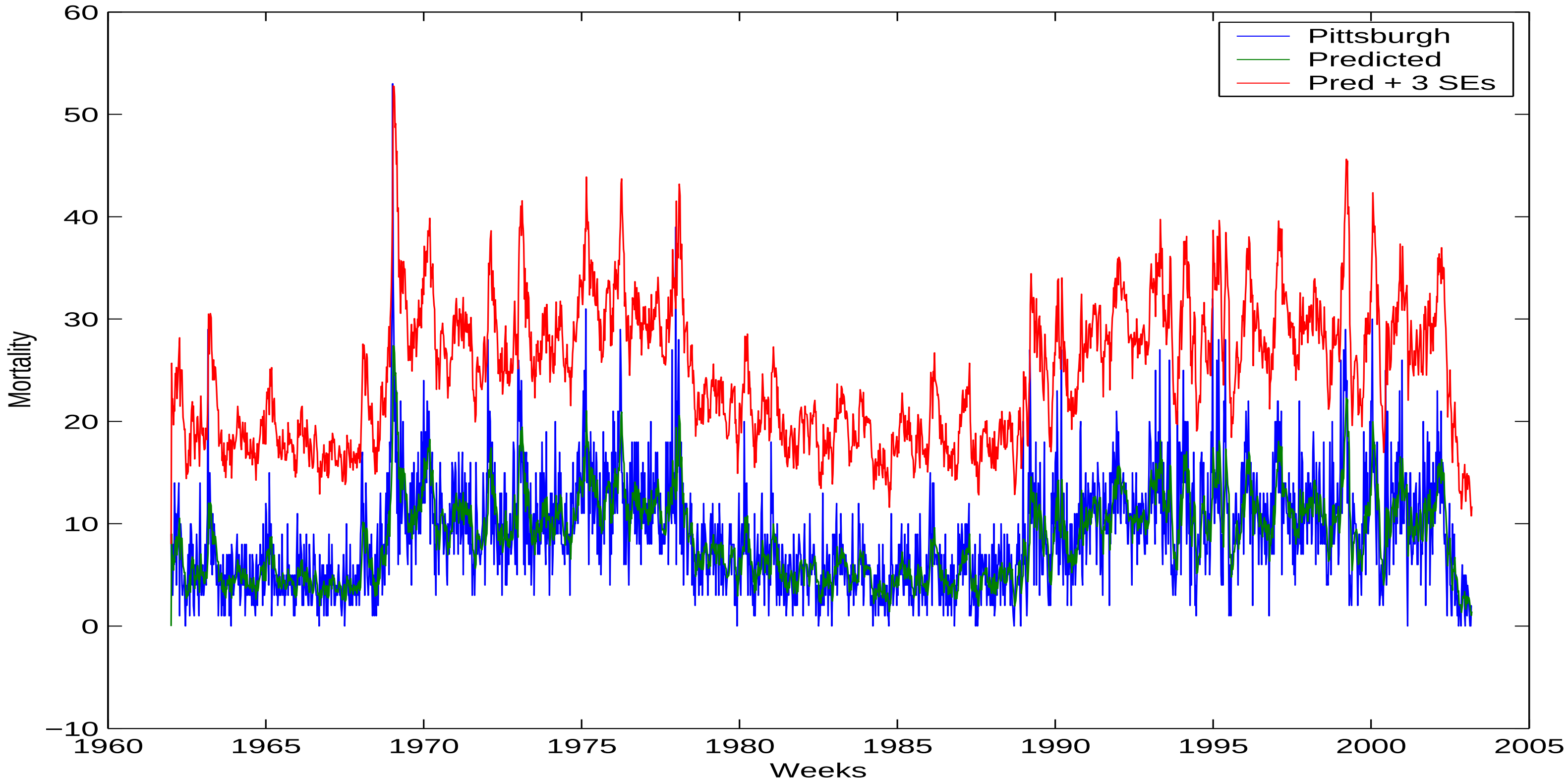
$$y_t = x_t + \Gamma \boldsymbol{u}_t + w_t \qquad t = 1, ..., n.$$

- $y_t = z_t - z_{t-1}$ is a univariate process ("$z = \sqrt{\text{data} + 1}$" ... could also try $y_t = z_t - z_{t-52}$),
- Γ is a $1 \times r$ vector of (constrained/unconstrained) regression parameters,
- $\boldsymbol{u}_t$ is an $r \times 1$ vector of inputs, **including** mortality rates from nearby locations at various time lags (contemporaneous values included).
- Here, $\Theta = (\phi, \theta, \gamma_1, \ldots, \gamma_r, \sigma_w^2)'$.

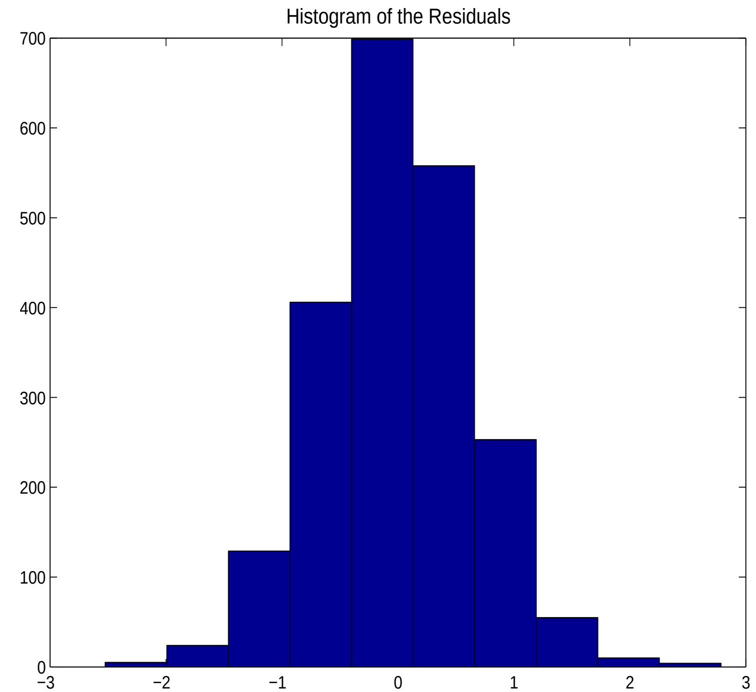
The previous model was fit to the Pittsburgh data.

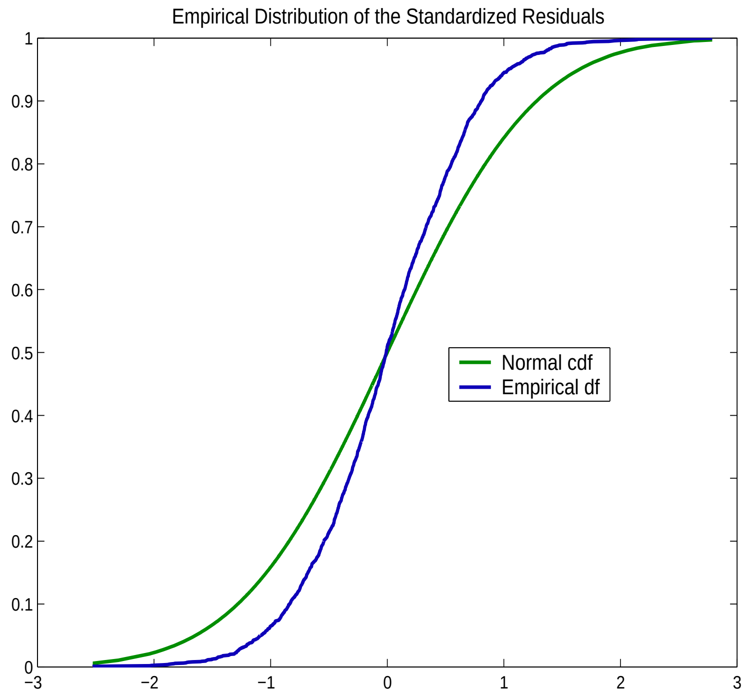
$$\widehat{\Gamma \boldsymbol{u}_t} = (.04\,\text{al}_t, .04\,\text{ph}_t, .09\,\text{rd}_t, .11\,\text{sc}_t, .07\,\text{al}_{t-1}, .02\,\text{ph}_{t-1}, .01\,\text{rd}_{t-1}, .02\,\text{sc}_{t-1},$$
$$-\,.03\,\text{al}_{t-2}, .01\,\text{ph}_{t-2}, .06\,\text{rd}_{t-2}, .03\,\text{sc}_{t-2})' \quad \text{[largest SE} = .003]$$

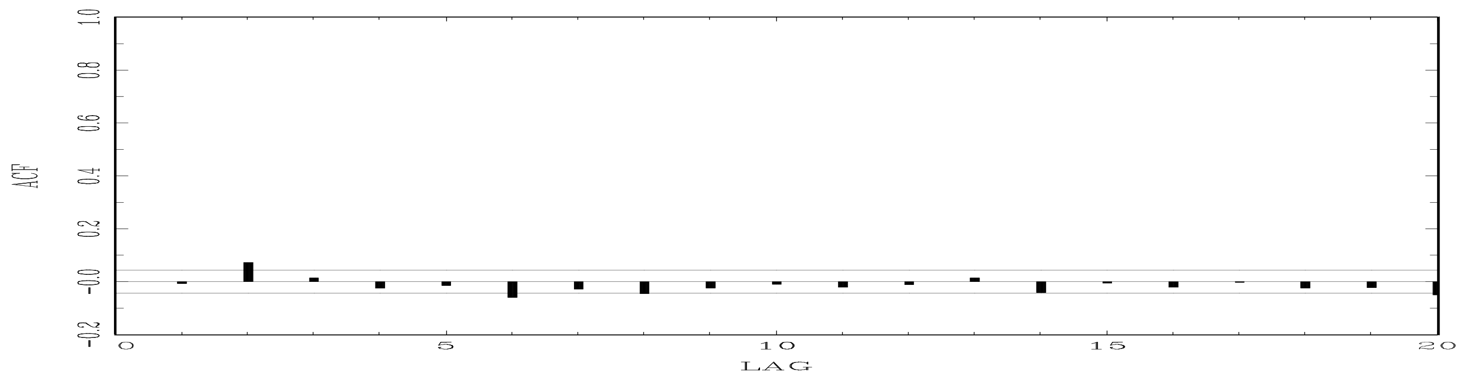
And $\widehat{\phi} = .1_{(.05)}, \widehat{\theta} = .7_{(.02)}, \widehat{\sigma}_w = .7$



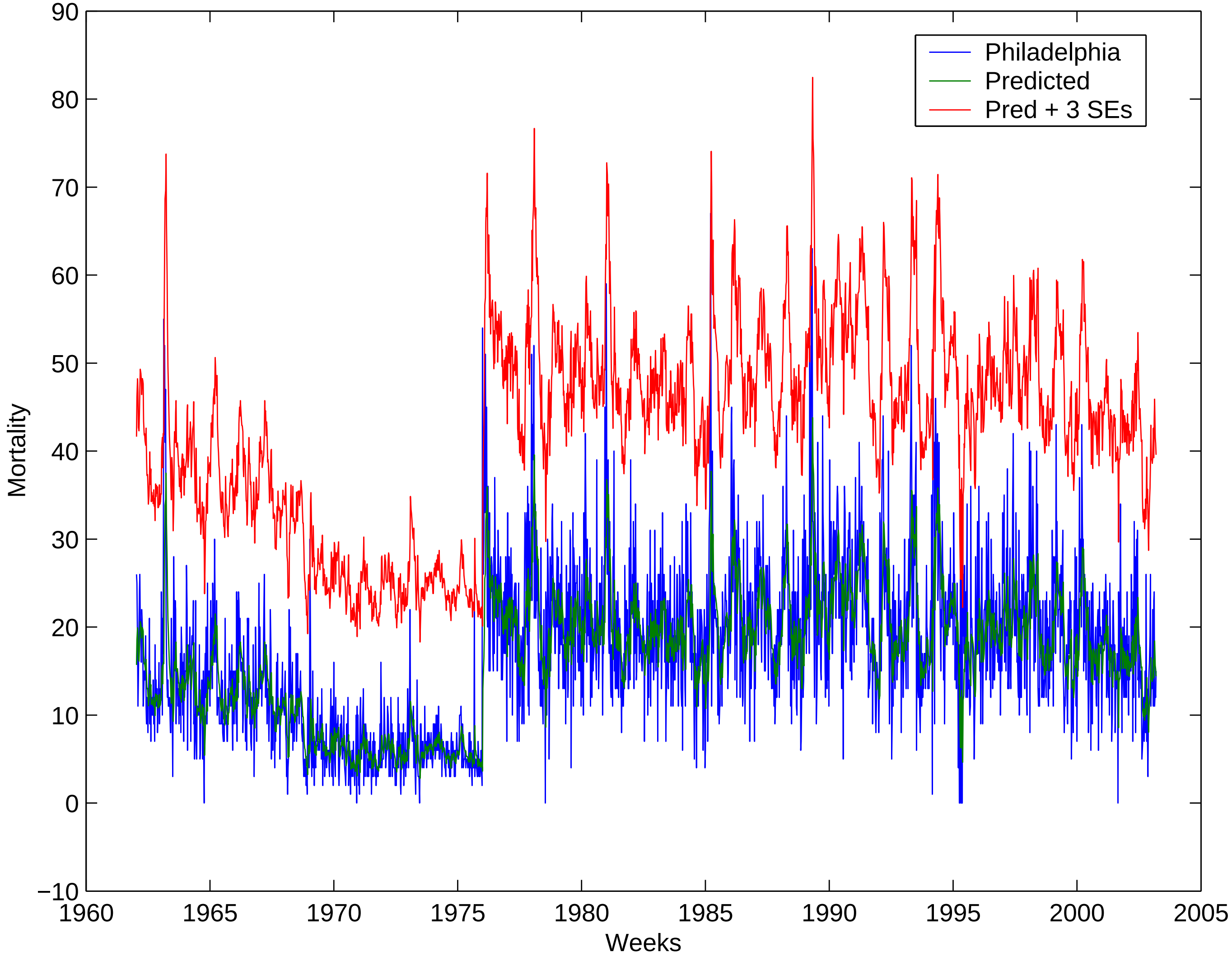
The Innovations (residuals):







Similar model for Philadelphia:



STARMAX:
A Spatially Constrained Multivariate Approach:

$$\boldsymbol{x}_{t+1} = D\Phi\boldsymbol{x}_t + D(\Phi - \Theta)\boldsymbol{w}_t \qquad t = 0, 1, ..., n$$

$$\boldsymbol{y}_t = \boldsymbol{x}_t + \Gamma\boldsymbol{u}_t + \boldsymbol{w}_t \qquad t = 1, ..., n$$

$\boldsymbol{x}_t$ is the p-dimensional state vector

$\boldsymbol{y}_t$ is the p-dimensional observation vector

$\boldsymbol{u}_t$ is the r-dimensional vector of eXogenous variables

$\boldsymbol{w}_t$ is the p-dimensional noise vector

Φ and Θ are diagonal $p \times p$ parameter matrices

Γ is a $p \times r$ matrix of regression parameters

D is a $p \times p$ matrix of specified spatial constraints

This state-space model implies

$$\boldsymbol{y}_t = D\Phi\boldsymbol{y}_{t-1} + \Gamma\boldsymbol{u}_t + \boldsymbol{w}_t - D\Theta\boldsymbol{w}_{t-1}$$

The model is easily generalized to arbitrary orders and spatial constraints. For example:

$$\begin{aligned} \boldsymbol{x}_{t+1} &= \begin{bmatrix} D_1\Phi_1 & D_2\Phi_2 \\ I & 0 \end{bmatrix} \boldsymbol{x}_t + \begin{bmatrix} D_1(\Phi_1 - \Theta_1) \\ D_2\Phi_2 \end{bmatrix} \boldsymbol{w}_t \\ \boldsymbol{y}_t &= [\, I, 0 \,]\, \boldsymbol{x}_t + \Gamma \boldsymbol{u}_t + \boldsymbol{w}_t \end{aligned}$$

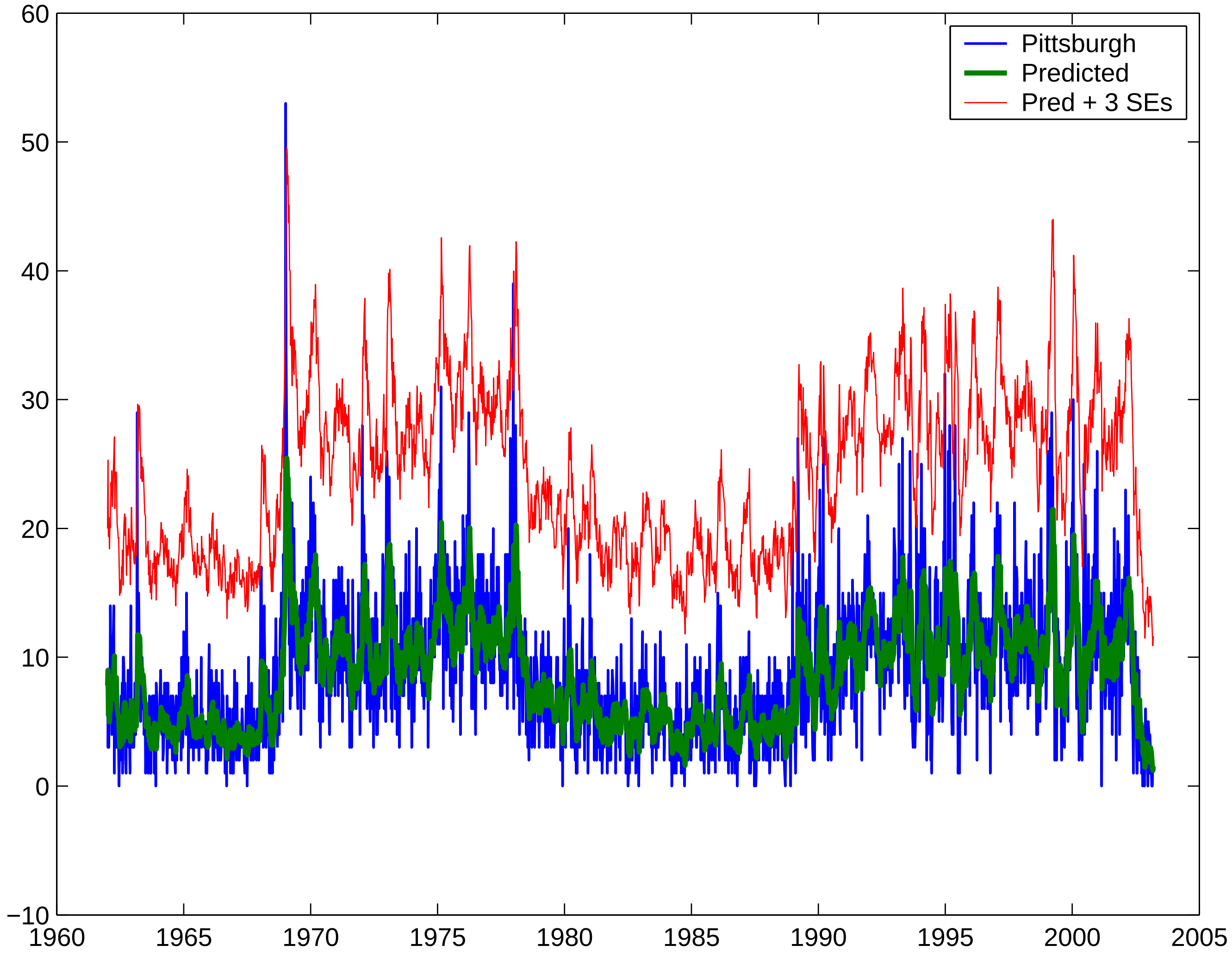
yields the STARMAX$(2,1)$ model:

$$\boldsymbol{y}_t = D_1\Phi_1\boldsymbol{y}_{t-1} + D_2\Phi_2\boldsymbol{y}_{t-2} + \Gamma\boldsymbol{u}_t + \boldsymbol{w}_t - D_1\Theta_1\boldsymbol{w}_{t-1}$$

where D_1 and D_2 are first order and second order spatial constraint matrices and Φ_1, Φ_2, Θ_1 are diagonal matrices, as before. The eXogenous variables [inputs] are $\boldsymbol{u}_t$.

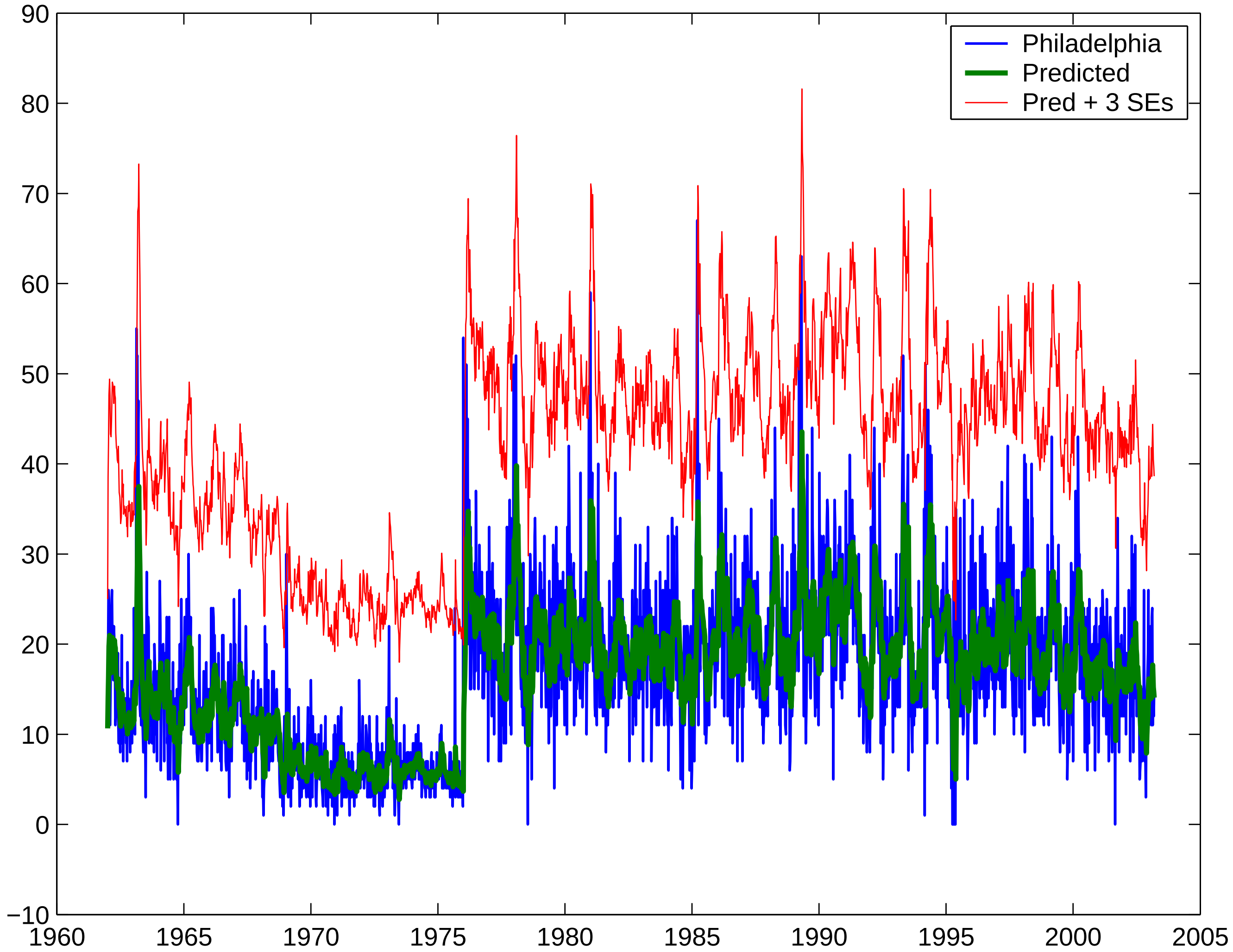
Note: $\boldsymbol{x}_t$ is $2p \times 1$ and $\boldsymbol{y}_t$ is $p \times 1$

Pittsburgh

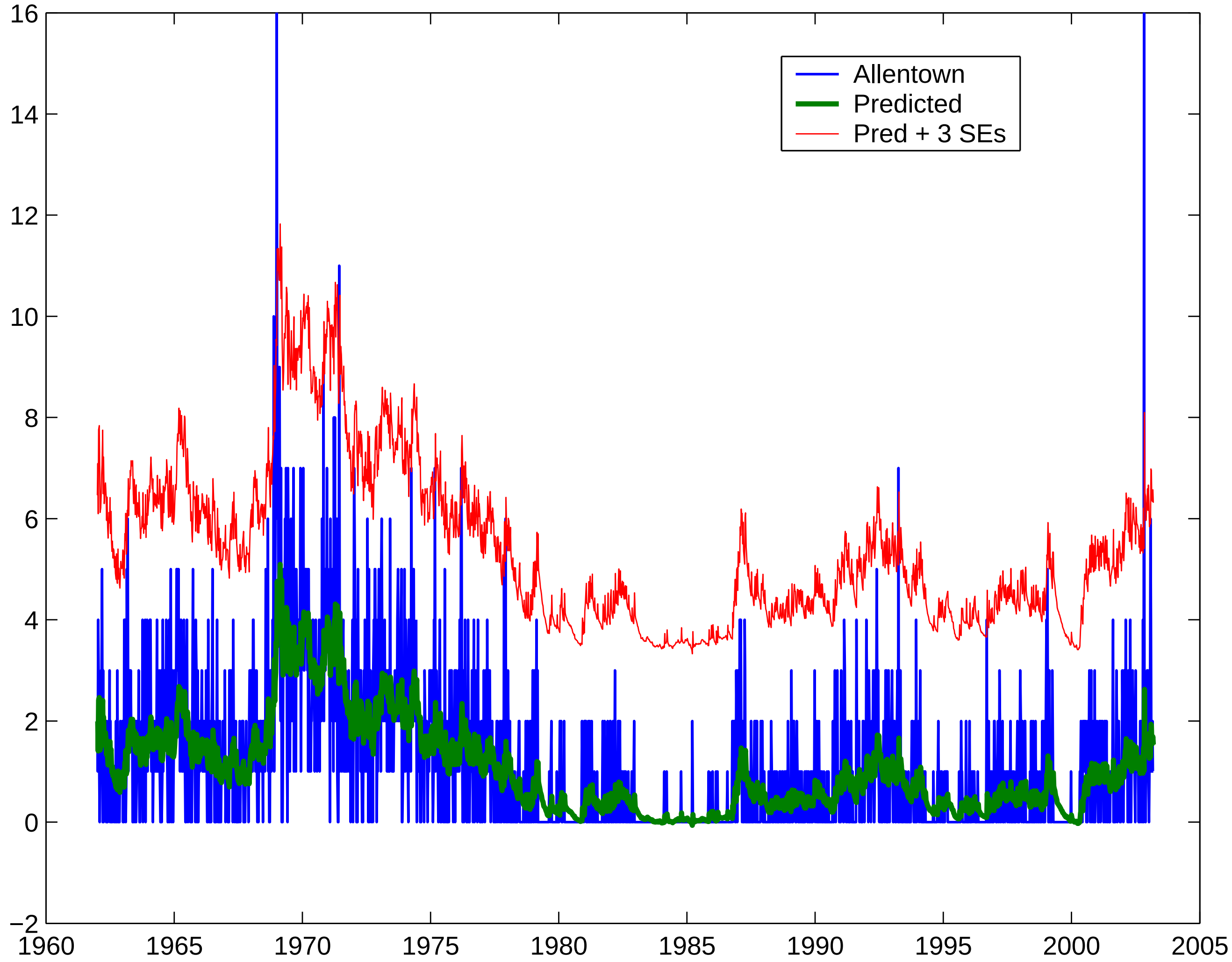


60
50
40
30
20
10
0
−10
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
Pittsburgh
Predicted
Pred + 3 SEs

Philadelphia



Allentown



Allentown
Predicted
Pred + 3 SEs
16
14
12
10
8
6
4
2
0
−2
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005

D Matrix $[D_{ij} = \widehat{\text{corr}}(\boldsymbol{y}_t, \boldsymbol{y}_{t-1})]$:

	Alln(t-1)	Phil(t-1)	Pitt(t-1)	Read(t-1)	Scrn(t-1)
Alln(t)	1	.016	.007	-.016	.027
Phil(t)	.042	1	-.047	-.022	.040
Pitt(t)	.043	-.010	1	-.051	-.029
Read(t)	.013	-.011	-.038	1	.002
Scrn(t)	.017	-.016	-.035	-.011	1

Estimates & Errors:

					original data units		
	phi	se	theta	se	rmspe	stdev	improvement
Allentown	0.09	0.02	0.83	0.02	1.3	1.6	19%
Philadelphia	0.07	0.03	0.72	0.02	7.0	9.5	26%
Pittsburgh	0.09	0.03	0.73	0.02	4.3	5.5	22%
Reading	0.04	0.02	0.80	0.02	1.9	2.3	17%
Scranton	0.13	0.02	0.81	0.02	1.5	1.6	6%