Outliers in Time Series

Amirhosein "Emerson" Azarbakht, Michael Dumelle,
Camden Lopez & Tadesse Zemicheal

March 10, 2016

Outlier Models

- ARIMA(p, d, q) process

$$X_t = \frac{\theta(B)}{\alpha(B)\phi(B)} Z_t$$

- Roots of $\theta(B), \phi(B)$ outside unit circle
- $\alpha(B) = (1 - B)^d$
- $Z_t \sim_{iid}$ Normal$(0, \sigma^2)$

Outlier Models

- Observed series

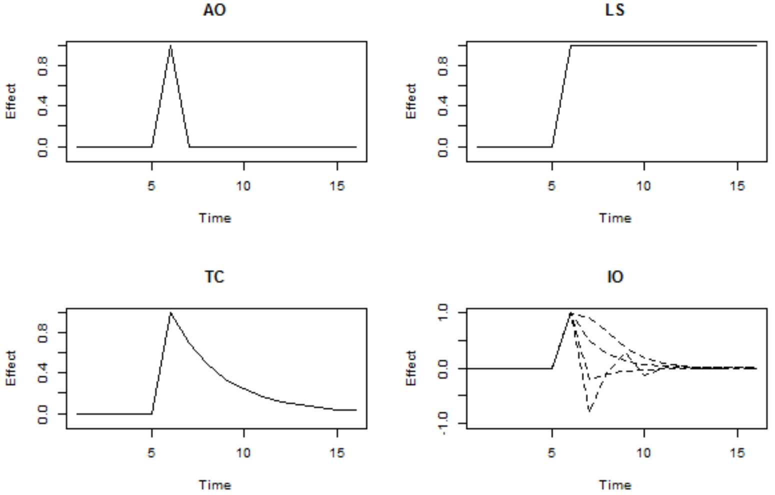
$$X_t^* = X_t + \text{ outlier effect}$$

- Four models for outlier effect:
  - Additive outlier (AO)
  - Level shift (LS)
  - Temporary change (TC)
  - Innovational outlier (IO)

Outlier Models

$$\begin{aligned}
\text{AO:} \qquad & X_t^* = X_t + \omega I_t(t_1) \\
\text{LS:} \qquad & X_t^* = X_t + \frac{1}{1-B}\omega I_t(t_1) \\
\text{TC:} \qquad & X_t^* = X_t + \frac{1}{(1-\delta B)}\omega I_t(t_1) \\
\text{IO:} \qquad & X_t^* = \frac{\theta(B)}{\alpha(B)\phi(B)}\left[Z_t + \omega I_{t_1}(t)\right]
\end{aligned}$$

Outlier Models



Outlier Estimation

- Obtain residuals $\hat{e}_t$ from the observed series X_t^* by applying

$$\pi(B) = \frac{\alpha(B)\phi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \ldots$$

- If no outliers, what's left is Z_t: $\pi(B)X_t = Z_t$
- When outlier present, residuals $\hat{e}_t = \pi(B)X_t^*$ reveal outlier effect

Outlier Estimation

- Residuals for each type of outlier:

$$
\begin{aligned}
\text{IO:} \qquad & \hat{e}_t = \omega I_t(t_1) + Z_t \\
\text{AO:} \qquad & \hat{e}_t = \omega \pi(B) I_t(t_1) + Z_t \\
\text{LS:} \qquad & \hat{e}_t = \omega \frac{\pi(B)}{1-B} I_t(t_1) + Z_t \\
\text{TC:} \qquad & \hat{e}_t = \omega \frac{\pi(B)}{1-\delta B} I_t(t_1) + Z_t
\end{aligned}
$$

- All have the form of simple linear regression:

$$\hat{e}_t = \omega x_t + Z_t$$

Outlier Estimation

- Least-squares estimate:

$$\hat{\omega} = \frac{\sum_{t=t_1}^{n} \hat{e}_t x_t}{\sum_{t=t_1}^{n} x_t^2}$$

- Divide by standard error:

$$\hat{\tau} = \frac{\hat{\omega}}{\hat{\sigma} / \sqrt{\sum_{t=t_1}^{n} x_t^2}}$$

  - Approximately $\sim$ Normal(0, 1)

Outlier Detection

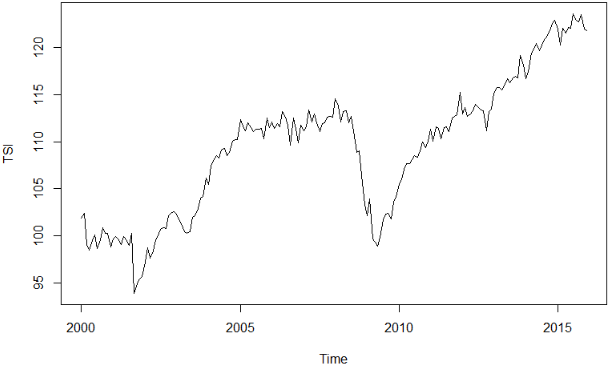
- At each $t = 1, \ldots, n$,
- For each outlier type (AO, LS, TC, IO),
  - Estimate outlier effect $\hat{\omega}$ and calculate $\hat{\tau}$
  - Large $|\hat{\tau}|$ indicates an outlier
- When multiple outliers present, can mask one another, cause biased estimates of effects
  - Need to repeatedly adjust series, re-estimates effects

Outlier Detection

- Iterative procedure for detecting outliers, adjusting series, and fitting (seasonal) ARIMA model:
  - Chen, C. and Liu, Lon-Mu (1993), “Joint Estimation of Model Parameters and Outlier Effects in Time Series,” *Journal of the American Statistical Association*, 88, 284–297.
- Three stages:
  1. Locate outliers in order of descending magnitude ($|\hat{\tau}|$)
  2. Drop outliers that are now insignificant after accounting for the others
  3. Make final estimates of model parameters and obtain final set of outliers
- Implemented in `tso` function in `tsoutliers` R package

Illustrative Example

- We applied `tso` to time series data from the US Bureau of Transportation Statistics
- Transportation Services Index (TSI), monthly measure of volume of services provided by for-hire transportation sector



Illustrative Example

- `tso` found two outliers:
  - Temporary Change outlier in Sept. 2001
  - Level Shift outlier in Dec. 2008