Outliers in Time Series

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

March 10, 2016

Intro

Time series affected by external/interruptive events. Examples:

- holidays
- strikes
- policy changes
- outbreak of war
- unexpected political or economic crises

=> Anomaly in data (observations inconsistent with the rest of the series)

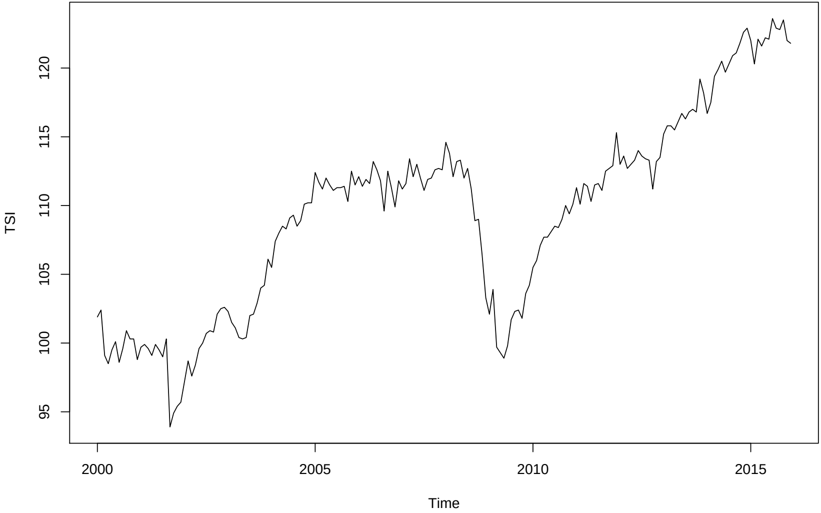
- measurement error
- anomalous behavior

Outline

- Motivating example
- Models to describe four types of outliers
- Estimating outlier effects using linear regression
- Using estimated effects to detect outliers
- R function `tsoutliers::tso`

Motivating Example

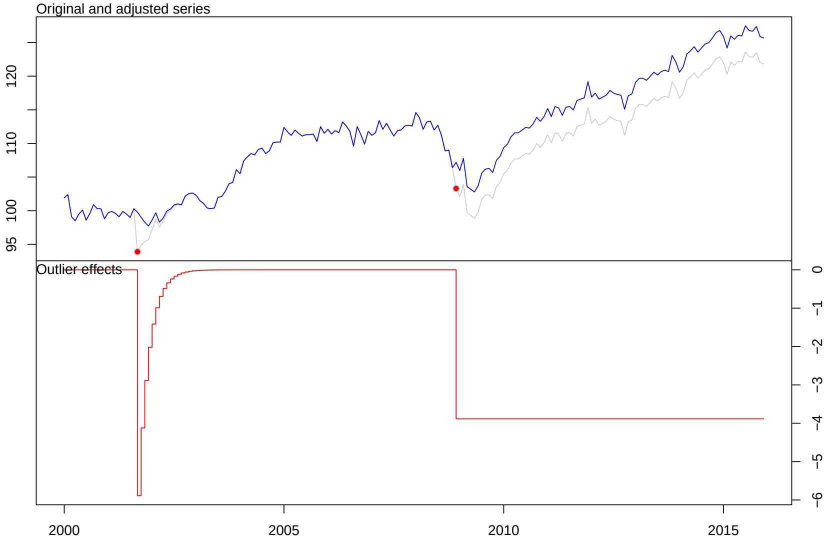
- Transportation Services Index (TSI): monthly measure of volume of services provided by for-hire transportation sector



Motivating Example

Motivating Example

```r
plot(tso_output)
```



Motivating Example

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

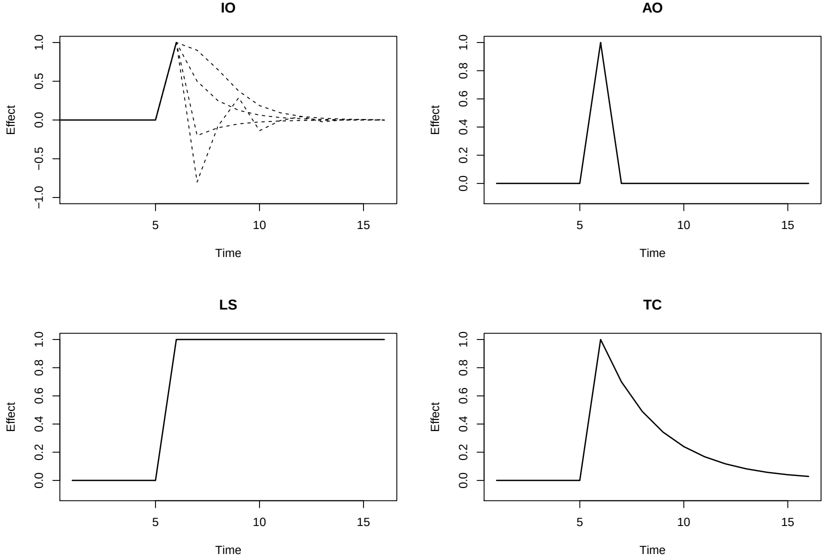
$$X_t^* = X_t + \text{ outlier effect}$$

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

Outlier Models

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

Outlier Models



Outlier Estimation

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

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 (w/o intercept):

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

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 $|\hat{\tau}|$ $(> C)$ indicates an outlier
- Once outlier is detected, estimated effect can be subtracted to obtain adjusted series

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 with many iterations of outlier detection, series adjustment, and re-fitting the model
  - Necessary to deal with masking, other issues with estimating outlier effects one-at-a-time when multiple outliers are present

Outlier Detection

- Implemented in `tso` function in `tsoutliers` R package

```
tso(y, cval = NULL, delta = 0.7,
    types = c("AO", "LS", "TC"),
    maxit = 1, maxit.iloop = 4,
    tsmethod = c("auto.arima", "arima", "stsm"),
    args.tsmethod = NULL)
```

The End

- Thank you
- "Please clap" — Jeb Bush