

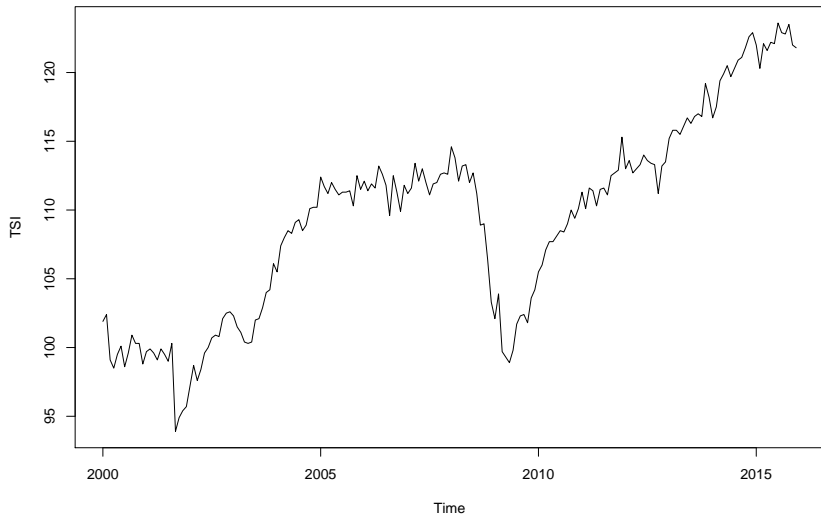
Outliers in Time Series

Amirhosein “Emerson” Azarbakht, Michael Dumelle,
Camden Lopez & Tadesse Zemicheal

March 10, 2016

Motivating Example

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



Motivating Example

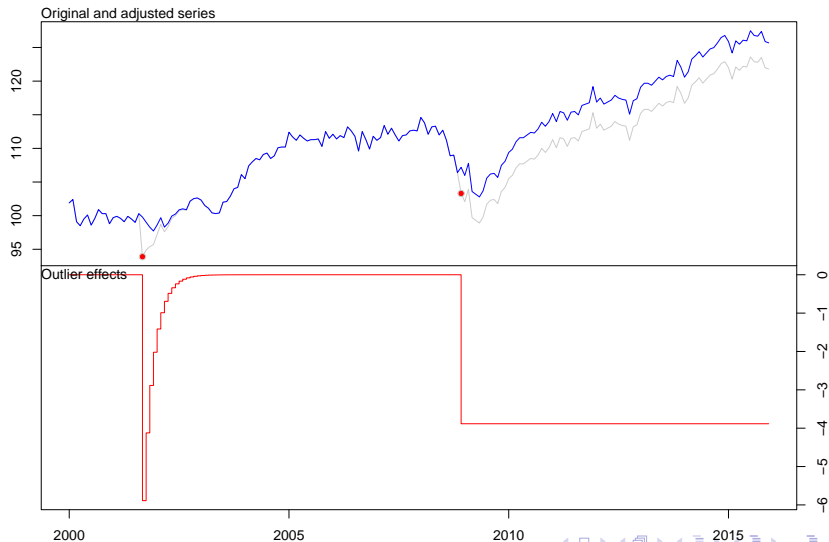
- ▶ The `tso` function in the `tsoutliers` package automatically detects and fits a model to the series with outlier effects removed
 - ▶ We'll describe how this works
- ▶ Two outliers detected:
 - ▶ Temporary Change outlier in Sept. 2001
 - ▶ Level Shift outlier in Dec. 2008

```
tso_output$outliers
```

##	type	ind	time	coefhat	tstat
## 1	TC	21	2001:09	-5.889364	-5.928143
## 2	LS	108	2008:12	-3.884195	-3.633127

Motivating Example

```
plot(tso_output)
```



Motivating Example

- ▶ Fitting ARIMA(1, 1, 0) model...
- ▶ Model fit without adjusting for outliers:
 - ▶ $\hat{\alpha}_1 = -0.165952$
 - ▶ $\hat{\sigma}_2 = 1.425164$
- ▶ Model fit after adjusting for outliers:
 - ▶ $\hat{\alpha}_1 = -0.2159957$
 - ▶ $\hat{\sigma}_2 = 1.1362642$
- ▶ Failing to adjust for outliers can result in
 - ▶ Wrong model or biased parameter estimates
 - ▶ Increased forecasting error

Outline

- ▶ ~~Motivating example~~
- ▶ Models to describe four types of outliers
- ▶ Estimating outlier effects using linear regression
- ▶ Detecting outliers
- ▶ Using `tsoutliers::tso`

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
- ▶ $\alpha(B) = (1 - B)^d$
- ▶ $Z_t \sim_{iid} \text{Normal}(0, \sigma^2)$

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

$$\text{AO:} \quad X_t^* = X_t + \omega l_t(t_1)$$

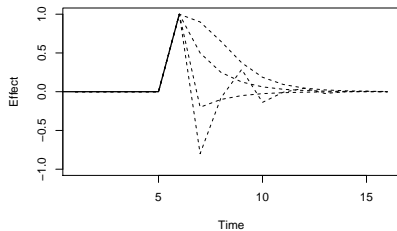
$$\text{LS:} \quad X_t^* = X_t + \frac{1}{1-B} \omega l_t(t_1)$$

$$\text{TC:} \quad X_t^* = X_t + \frac{1}{(1-\delta B)} \omega l_t(t_1)$$

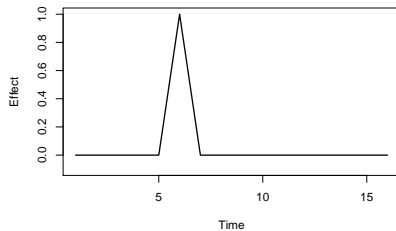
$$\text{IO:} \quad X_t^* = \frac{\theta(B)}{\alpha(B)\phi(B)} [Z_t + \omega l_{t_1}(t)]$$

Outlier Models

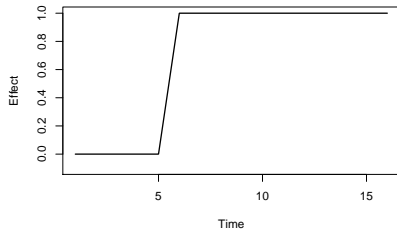
IO



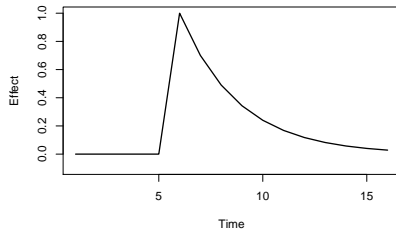
AO



LS



TC



Outlier Estimation

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

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

- ▶ (Remember $X_t = \frac{\theta(B)}{\alpha(B)\phi(B)}$)
- ▶ If there were no outliers, result is white noise: $\pi(B)X_t = Z_t$
- ▶ When outlier present, residuals $\hat{e}_t = \pi(B)X_t^*$ reveal outlier effect

Outlier Estimation

- Residuals for each type of outlier:

$$\text{IO:} \quad \hat{e}_t = \omega I_t(t_1) + Z_t$$

$$\text{AO:} \quad \hat{e}_t = \omega \pi(B) I_t(t_1) + Z_t$$

$$\text{LS:} \quad \hat{e}_t = \omega \frac{\pi(B)}{1 - B} I_t(t_1) + Z_t$$

$$\text{TC:} \quad \hat{e}_t = \omega \frac{\pi(B)}{1 - \delta B} I_t(t_1) + Z_t$$

- 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 $\sim \text{Normal}(0, 1)$

Outlier Detection

- ▶ At each $t = 1, \dots, n$, for each outlier type (AO, LS, TC, IO),
 - ▶ Estimate outlier effect $\hat{\omega}$ and calculate $\hat{\tau}$
 - ▶ Large $|\hat{\tau}|$ indicates an outlier
- ▶ Once outlier is detected, 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
- ▶ Implemented in `tso` function in `tsoutliers` R package

Outlier Detection

```
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)
```

- ▶ Sometimes Often fails or takes a long time
- ▶ Might need to increase maxit.iloop
- ▶ Might help to specify model to fit using arima instead of letting auto.arima choose the model

The End

- ▶ Thank you
- ▶ “Please clap” — Jeb Bush