Multivariate Time Series Regression: Recap and Modeling Example

This Lecture

- Review three scenarios of time series residuals
- Example with treasuries

Time Series Residuals

We want to know the relationship between time series variables y_t and x_t. How can lag dependence in e_t screw things up?

$$y_t = \alpha + \beta x_t + e_t$$

Examples:

- AR(p): x_t is a vector containing p lags of y_t.
- Market Model: y_t is a stock return, x_t market return
- Energy demand: y_t is electricity or natural gas consumption, or fuel in storage, x_t is a vector of weather variables (e.g., heating/cooling degree days, hurricane incidence, etc.)
- Commodity market linkages (cointegration):
  - y_t is copper price, x_t is gold price.
  - y_t is global LNG price, x_t is Brent crude price.
- Prediction: y_t is natural gas price returns, x_t is a vector of lags of gas returns, oil returns, production, and weather.

Three scenarios of interest for residuals e_t

1. y_t and x_t are stationary, x_t does NOT include lags of y, and e_t has lag dependence.
2. y_t and x_t are stationary, x_t DOES include lags of y, and e_t has lag dependence.
3. y_t and x_t are unit root nonstationary. Interpretation depends on whether e_t is stationary or not.
   - *cointegrated* vs. *spurious*

Scenario 1: y_t, x_t stationary, no lags of y_t, e_t lag dependent

- e.g., Market model, CAPM
- Solution 1: estimate the ARMA model for e_t jointly with the regression equation
- Solution 2: Use heteroskedasticity-autocorrelation robust (HAC) standard errors (sometimes called Newey-West std. errors)

Scenario 2: y_t, x_t stationary, x_t includes y_{t-p}, e_t lag dependent

- Solution: e_t has lag dependence because the ARMA behavior of y_t is misspecified.
- Add lags or MA terms to the model for y_t until residuals no longer have autocorrelation.

Scenario 3: y_t, x_t NOT stationary

- Interpretation depends on whether e_t is stationary or not.
- If e_t is stationary, then y_t and x_t are **cointegrated**, and α and β are consistently estimated.
- If e_t is not stationary, then y_t and x_t are NOT cointegrated.
  - The regression of y on x is **spurious**.
  - It may produce very high R-squared, and very high t-stats that are meaningless.
  - Need to model relationship between Δy_t and Δx_t.

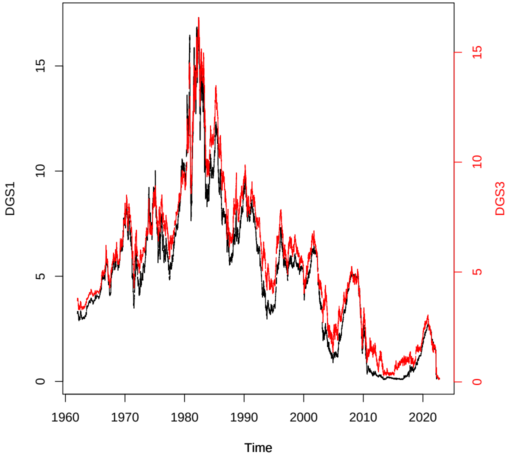
Example: 1- and 3-year Treasury rates

# 1-year Treasury rate
getSymbols("DGS1",src="FRED")
# 3-year Treasury rate

getSymbols("DGS3",src="FRED")

```
dgs1 = ts(DGS1$DGS1,freq=252,start=1962)
dgs3 = ts(DGS3$DGS3,freq=252,start=1962)
par(mar=c(5,4,4,5)+0.1)
plot(dgs1)
par(new=T)
plot(dgs3,axes=FALSE,ylab="",col="red")
mtext("dgs3",side=4,line=2.5,col="red")
axis(side=4,col="red",col.axis="red")
```

Example: 1- and 3-year Treasury rates



Example: 1- and 3-year Treasury rates

Interest rates are non-stationary, fail to reject the null of a unit root:

```
ar(diff(na.remove(dgs1)))
CADFtest(dgs1,max.lag.y=41,type="drift")
ar(diff(na.remove(dgs3)))
CADFtest(dgs3,max.lag.y=34,type="drift")
```

But estimates are consistent if e_t is stationary in

$$r_{3t} = \alpha + \beta r_{1t} + e_t$$

```
tbillreg1 = lm(dgs3 ~ dgs1)
summary(tbillreg1)
```

Example: 1- and 3-year Treasury rates

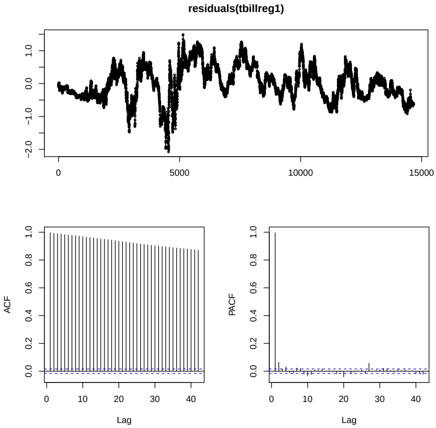
Notice extremely high R-squared, t-stat, and $\sigma_e \approx 0.5$

```
##
## Call:
## lm(formula = dgs3 ~ dgs1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.07663 -0.34625 -0.01676  0.35211  1.49098
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.653964   0.007390   88.49   <2e-16 ***
## dgs1        0.952219   0.001214  784.35   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5024 on 14670 degrees of freedom
##   (654 observations deleted due to missingness)
## Multiple R-squared:  0.9767,Adjusted R-squared:  0.9767
## F-statistic: 6.152e+05 on 1 and 14670 DF,  p-value: < 2.2e-16
```

Example: 1- and 3-year Treasury rates

However, residuals do not appear stationary:

```
tsdisplay(residuals(tbillreg1))
```



Example: 1- and 3-year Treasury rates

Need to work with differences. Notice regression intercept gets "differenced" out.

$$(r_{3t} - r_{3,t-1}) = (\alpha + \beta r_{1t} + e_t) - (\alpha + \beta r_{1,t-1} + e_{t-1})$$

$$\Delta r_{3t} = \beta \Delta r_{1t} + \Delta e_t$$

```
deldgs1 = diff(na.remove(dgs1))
deldgs3 = diff(na.remove(dgs3))
dtbillreg = lm(deldgs3 ~ -1 + deldgs1)
summary(dtbillreg)
```

Example: 1- and 3-year Treasury rates

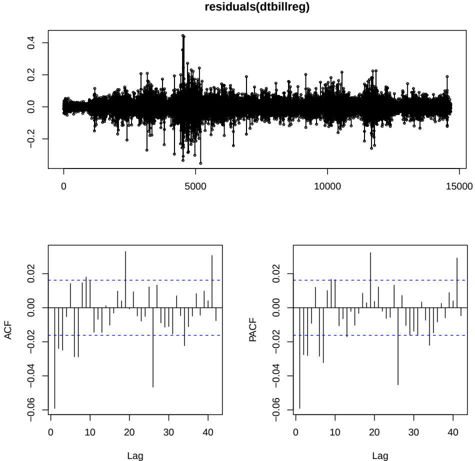
Notice R-squared, t-stat, σ_e much lower than before

```
##
## Call:
## lm(formula = deldgs3 ~ -1 + deldgs1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3530 -0.0200  0.0000  0.0200  0.4453
##
## Coefficients:
##         Estimate Std. Error t value Pr(>|t|)
## deldgs1 0.765628   0.004165   183.8   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04082 on 14670 degrees of freedom
## Multiple R-squared:  0.6973,Adjusted R-squared:  0.6972
## F-statistic: 3.379e+04 on 1 and 14670 DF,  p-value: < 2.2e-16
```

Example: 1- and 3-year Treasury rates

However, residuals are still autocorrelated, affects t-stats:

```
tsdisplay(residuals(dtbillreg))
```



Example: 1- and 3-year Treasury rates

- From previous slide, lots of little autocorrelations. Not sure we can get rid of it all.
- Let's ARMA(1,1) for sake of illustration

Example: 1- and 3-year Treasury rates

ARMA(1,1) model for residuals:

```
# Could use gls() or arima() functions
# dtbillregARMA = gls(deldgs3 ~ -1 + deldgs1,
#                     correlation=corARMA(p=1,q=1))
# summary(dtbillregARMA)
dtbillregARMAx = arima(deldgs3,order=c(1,0,1),
                     xreg=deldgs1,include.mean=F)
summary(dtbillregARMAx)
# arima doesn't produce R-squared, so calculate it
rsqARMA=(sum(deldgs3^2)-sum(dtbillregARMAx$residuals^2))/sum(deldgs3^2)
rsqARMA
tsdisplay(residuals(dtbillregARMAx),lag.max=40)
tsdiag(dtbillregARMAx,gof=40)
```

Example: 1- and 3-year Treasury rates

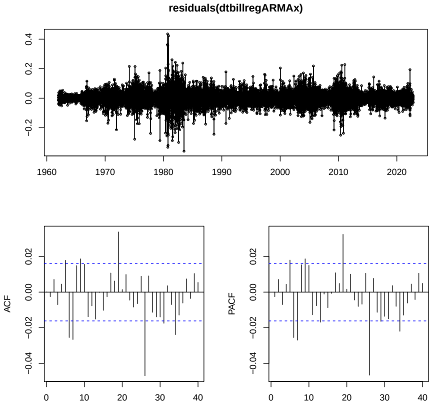
ARMA(1,1) model for residuals. $R^2 \approx 0.7$

```
##
## Call:
## arima(x = deldgs3, order = c(1, 0, 1), xreg = deldgs1, include.mean = F)
##
## Coefficients:
##          ar1      ma1    DGS1
##       0.5545  -0.6144  0.7688
## s.e.  0.0844   0.0801  0.0041
##
## sigma^2 estimated as 0.001657:  log likelihood = 26147.91,  aic = -52287.82
##
## Training set error measures:
##                         ME       RMSE        MAE MPE MAPE      MASE
## Training set -9.319435e-05 0.04071234 0.02798711 NaN  Inf 0.4361327
##                      ACF1
## Training set -0.002580592
## [1] 0.6988188
```

Example: 1- and 3-year Treasury rates

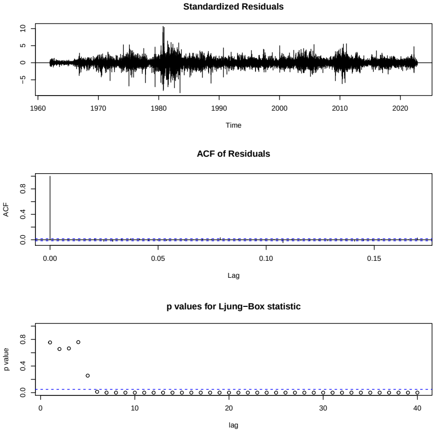
Estimated ARMA(1,1) model for residual improves, but still some residual autocorrelation

$$\Delta r_{3t} = 0.77\Delta r_{1t} + e_t, \quad e_t = 0.55e_{t-1} + a_t - 0.61a_{t-1}$$



Example: 1- and 3-year Treasury rates

Use Ljung-Box test to see flaws in ARMA(1,1)



Example: 1- and 3-year Treasury rates, HAC errors

- Getting the "right" ARMA here seems hard without a huge model.
- We know MA(1) and AR(1) aren't sufficient
- Let's use HAC standard errors on the basic model.

```
dtbillreg2 = lm(deldgs3 ~ -1 + deldgs1)
summary(dtbillreg2)
vcovHAC(dtbillreg2)
coeftest(dtbillreg2,vcov=vcovHAC(dtbillreg2))
```

Example: 1- and 3-year Treasury rates, no HAC errors

Basic model ignoring lag dependence in e_t:

```
##
## Call:
## lm(formula = deldgs3 ~ -1 + deldgs1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3530 -0.0200  0.0000  0.0200  0.4453
##
## Coefficients:
##         Estimate Std. Error t value Pr(>|t|)
## deldgs1 0.765628   0.004165   183.8   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04082 on 14670 degrees of freedom
## Multiple R-squared:  0.6973,Adjusted R-squared:  0.6972
## F-statistic: 3.379e+04 on 1 and 14670 DF,  p-value: < 2.2e-16
```

Example: 1- and 3-year Treasury rates, HAC errors

Now with HAC standard errors, notice big decline in *t*-stat, increase in standard error but coefficient is the same:

```
# HAC variance of coefficient estimate
vcovHAC(dtbillreg2)

##              deldgs1
## deldgs1 0.0001027968

# HAC standard errors & t-stats
coeftest(dtbillreg2,vcov=vcovHAC(dtbillreg2))

##
## t test of coefficients:
##
##         Estimate Std. Error t value  Pr(>|t|)
## deldgs1 0.765628   0.010139  75.514 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```