Lecture 3 Example: Regression with Time Series Errors

US weekly interest rates: use 1-year constant maturity rate (X_t) to predict 3-year rate (Y_t).

Step 1

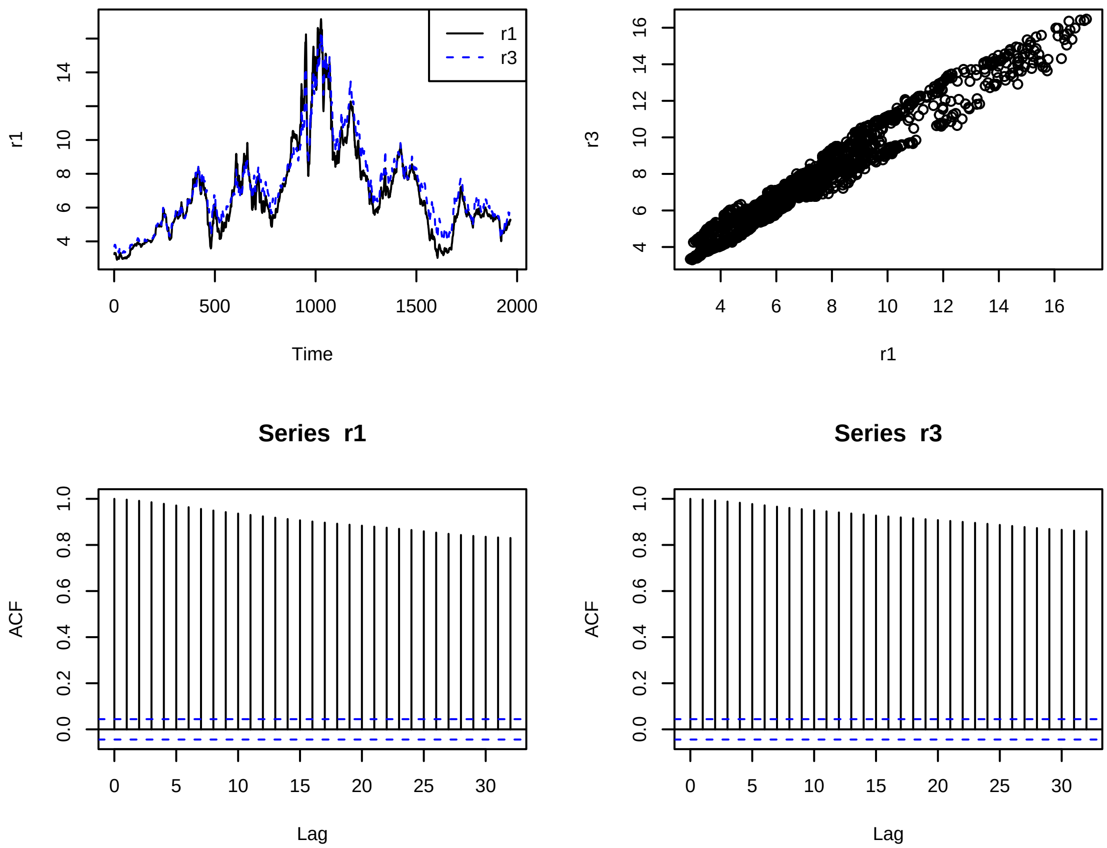
Fig 1 suggests $X_t =$ `r1` will be a good predictor of $Y_t =$ `r3` in a regression model (model `m1`):

$$Y_t = \beta_0 + \beta_1 X_t + a_t, \qquad a_t \sim \text{WN}(0, \sigma_a^2).$$

However, the ACF of both series is slowly decaying, suggesting they are ARIMAs.... (Confirmed by the ACF of the residuals).

```
> da=read.table("../Datasets/w-gs1n36299.txt", header=T)
> head(da)
    y1   y3     date
1 3.24 3.70 19620104
2 3.32 3.75 19620112
3 3.29 3.80 19620120

> r1=da[,1]          # 1-year rate
> r3=da[,2]          # 3-year rate

### Fig 1: Plot the data
#pdf(file="../Lectures/Plots/Lec3-Fig1.pdf", pointsize=9,width=6,height=5)
> par(mfrow=c(2,2))
> ts.plot(r1, ylab="")
> lines(1:length(r3), r3, lty=2, col="blue")
> legend("topright", c("r1","r3"), lty=1:2, col=c("black","blue"))
> plot(r1,r3)
> acf(r1)
> acf(r3)
> par(mfrow=c(1,1))
#dev.off()
###

### Fit a regression model to (r3,r1).
> m1=lm(r3~r1)
> summary(m1)
```

Figure 1: The series $r3 = Y_t$ and $r1 = X_t$.




```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.910687   0.032250   28.24   <2e-16 ***
r1          0.923854   0.004389  210.51   <2e-16 ***
---
Residual standard error: 0.538 on 1965 degrees of freedom
Multiple R-squared:  0.9575,Adjusted R-squared:  0.9575
F-statistic: 4.431e+04 on 1 and 1965 DF,  p-value: < 2.2e-16
```

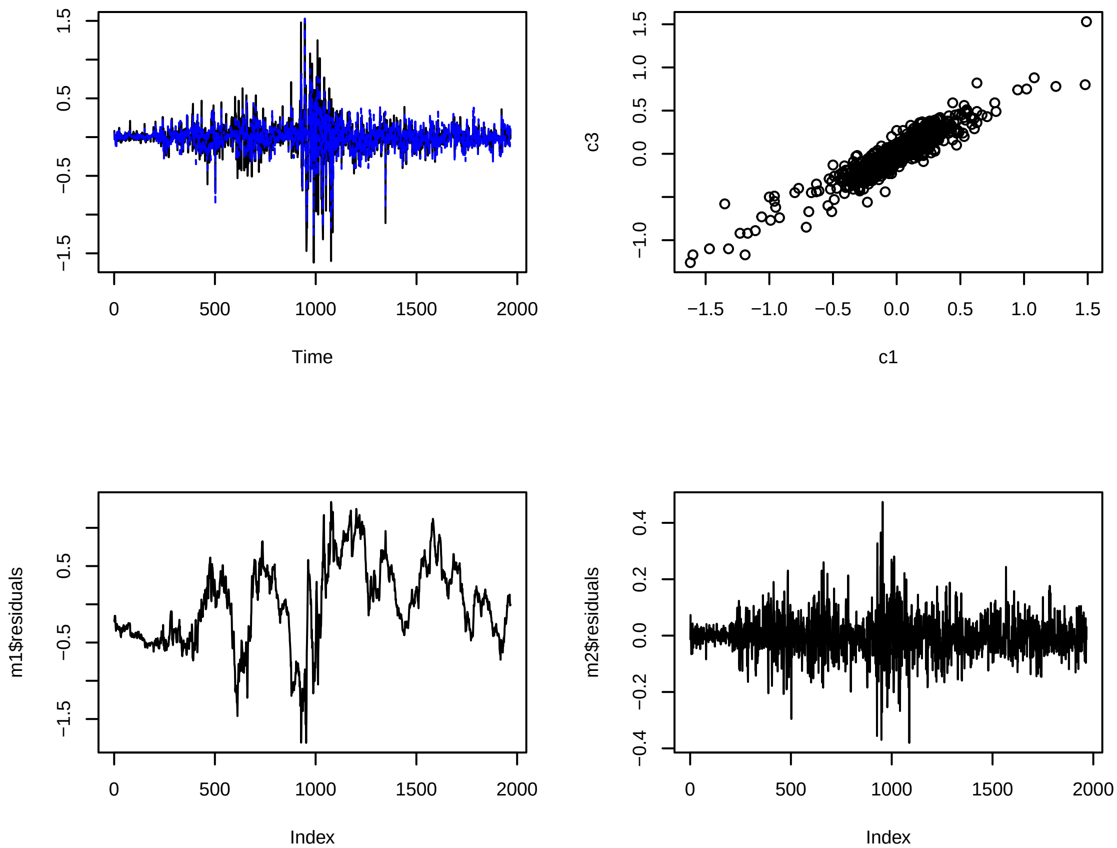
Step 2

Difference each series, producing $y_t = (1 - B)Y_t =$ `c3` and $x_t = (1 - B)X_t =$ `c1`, and refit the regression (model `m2`):

$$y_t = \beta_0 + \beta_1 x_t + a_t, \qquad a_t \sim \text{WN}(0, \sigma_a^2).$$

Fig 2 suggests the resids from `m2` look stationary, but are they WN?

Figure 2: The series y_t and x_t, and the resids from `m1` and `m2`.

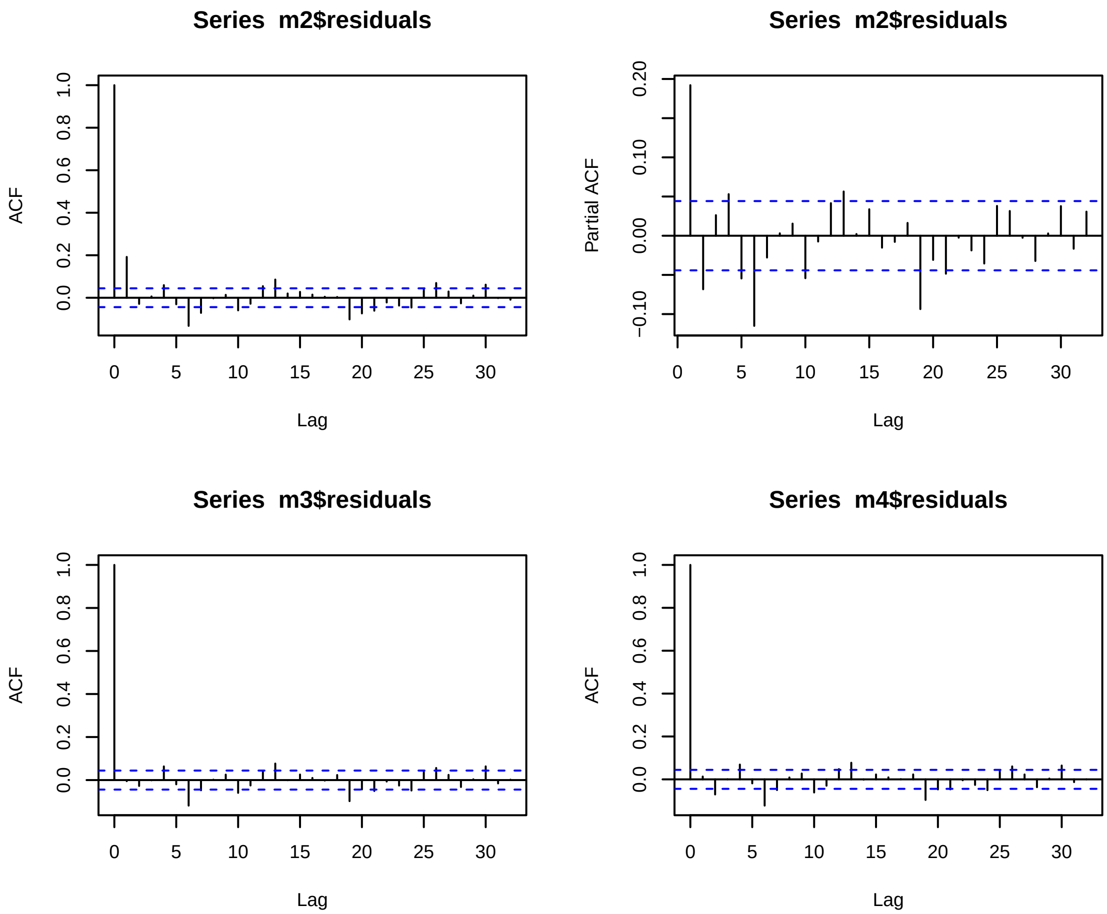



```
### ACF of m1 resids confirms we need to difference the series before fitting a reg.
> acf(m1$residuals)
> c3=diff(r3)
> c1=diff(r1)
> m2=lm(c3~c1)
> summary(m2)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.0002475  0.0015380   0.161    0.872
c1          0.7810590  0.0074651 104.628   <2e-16 ***
---
Residual standard error: 0.06819 on 1964 degrees of freedom
Multiple R-squared:  0.8479,Adjusted R-squared:  0.8478
F-statistic: 1.095e+04 on 1 and 1964 DF,  p-value: < 2.2e-16
```

Step 3

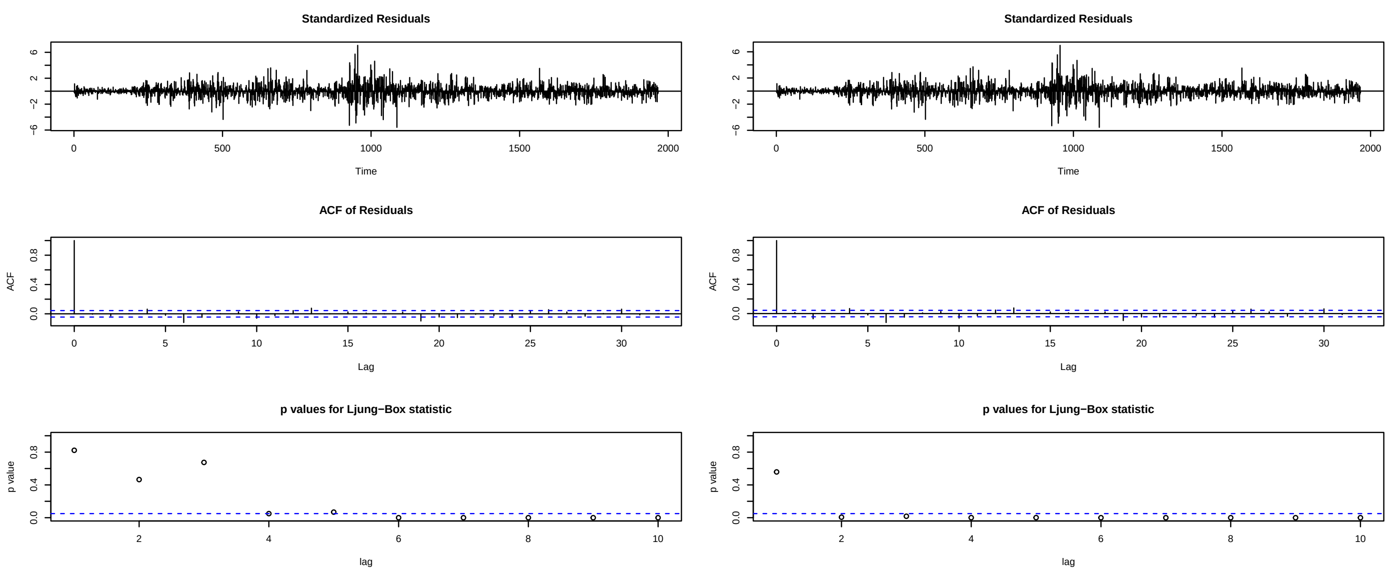
Examine ACF/PACF of resids from `m2` to find suitable ARMA: suggests MA(1) or AR(1). (See Fig 3.) In practice one would use `auto.arima` to search for a "best" model to fit to the resids from `m2`.

Figure 3: ACF/PACF for the resids from `m2`, and ACF for the resids from `m3` and `m4`.




```
### Look at ACF/PACF of resids from m2 to find suitable ARMA: MA(1) or AR(1)?
> acf(m2$residuals)
> pacf(m2$residuals)

### Fit reg with MA(1) errors
> m3=arima(c3, xreg=c1, order=c(0,0,1))
> m3
Coefficients:
         ma1  intercept      c1
      0.2115     0.0002  0.7824
s.e.  0.0224     0.0018  0.0077

sigma^2 estimated as 0.004456:  log likelihood = 2531.84,  aic = -5055.69
```

```
### Fit reg with AR(1) errors
> m4=arima(c3, xreg=c1, order=c(1,0,0))
> m4
Coefficients:
         ar1  intercept      c1
      0.1922     0.0003  0.7829
s.e.  0.0221     0.0019  0.0077

sigma^2 estimated as 0.004474:  log likelihood = 2527.86,  aic = -5047.72

### Proper GoF checks on both models, looks like MA(1) is better!
> tsdiag(m3)
> c(AIC(m3),BIC(m3))
[1] -5055.688 -5033.353
> tsdiag(m4)
> c(AIC(m4),BIC(m4))
[1] -5047.718 -5025.383
```



(a) GoF for `m3`.

(b) GoF for `m4`.

Figure 4: GoF for the resids from the time series models with: (a) MA(1) errors, (b) AR(1) errors.

Step 4

The final model is m3:

```
### Final model: reg with MA(1) errors
> m3=arima(c3, xreg=c1, order=c(0,0,1))
> m3
Coefficients:
         ma1  intercept      c1
      0.2115     0.0002  0.7824
s.e.  0.0224     0.0018  0.0077

sigma^2 estimated as 0.004456:  log likelihood = 2531.84,  aic = -5055.69
```

With $y_t = (1-B)Y_t$ and $x_t = (1-B)X_t$, the final model is:

$$
\begin{aligned}
y_t &= 0.0002 + 0.7824x_t + a_t, \qquad a_t \sim \text{MA}(1) \\
a_t &= 0.2115e_{t-1} + e_t, \qquad e_t \sim \text{WN}(0, 0.004456).
\end{aligned}
$$

Since the intercept is not significant, we can trim the model:

```
m5=arima(c3, xreg=c1, order=c(0,0,1), include.mean = FALSE)
m5
Coefficients:
         ma1      c1
      0.2115  0.7824
s.e.  0.0224  0.0077

sigma^2 estimated as 0.004456:  log likelihood = 2531.84,  aic = -5057.67

> c(AIC(m5),BIC(m5))
[1] -5057.670 -5040.919
```

Final model is:

$$
\begin{aligned}
y_t &= 0.7824x_t + a_t, \qquad a_t \sim \text{MA}(1) \\
a_t &= 0.2115e_{t-1} + e_t, \qquad e_t \sim \text{WN}(0, 0.004456).
\end{aligned}
$$

Appendix: A proper model search for the noise a_t using `auto.arima`.

```
> library(forecast)
> auto.arima(m2$res, max.p = 15, max.q = 15, stationary=T, seasonal=T, ic="bic",
                     approximation=F, stepwise=F, trace=T)
Best model: ARIMA(3,0,2) with zero mean
Coefficients:
         ar1      ar2     ar3      ma1     ma2
      0.4647  -1.0316  0.2108  -0.2662  0.9513
s.e.  0.0250   0.0200  0.0230   0.0117  0.0333

sigma^2 = 0.00437:  log likelihood = 2553.29
AIC=-5094.58   AICc=-5094.54   BIC=-5061.08
```