OLS Estimation - Assumptions

- CLM Assumptions
  - (A1) DGP: \( y = X \beta + \epsilon \) is correctly specified.
  - (A2) \( E[\epsilon | X] = 0 \)
  - (A3) \( \text{Var}[\epsilon | X] = \sigma^2 I_T \)
  - (A4) \( X \) has full column rank \( \text{rank}(X) = k \) – where \( T \geq k \).

- In this lecture, again, we will look at assumption (A1). So far, we have restricted \( f(X, \beta) \) to be a linear function: \( f(X, \beta) = X \beta \).

- But, it turns out that in the framework of OLS estimation, we can be more flexible with \( f(X, \beta) \).
Functional Form: Linearity in Parameters

• Linear in variables and parameters:

\[ Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \]

• Linear in parameters (intrinsic linear), nonlinear in variables:

\[ Y = \beta_1 + \beta_2 X_2^2 + \beta_3 \sqrt{X_3} + \beta_4 \log X_4 + \varepsilon \]

\[ Z_2 = X_2^2, \quad Z_3 = \sqrt{X_3}, \quad Z_4 = \log X_4 \]

\[ Y = \beta_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + \varepsilon \]

Note: We get some nonlinear relation between y and X, but OLS still can be used.

---

Functional Form: Linearity in Parameters

• Suppose we have:

\[ Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + \varepsilon \]

• The model is intrinsic linear, but it allows for a quadratic relation between y and X_2:

![Graph](image-url)
Functional Form: Linearity in Parameters

\[ \hat{Y} = b_1 + b_2 X_2 + b_3 X_2^2 = b_1 + b_2 X_2 + b_3 X_3 \]

Example: We want to test if a measure of market risk \((\text{MktRet} - r_f)^2\) is significant in the 3 FF factors (SMB, HML) for IBM returns. The model is non-linear in \((\text{MktRet} - r_f)\), but still intrinsic linear:

\[
\text{IBM}_{\text{Ret}} - r_f = \beta_0 + \beta_1 (\text{MktRet} - r_f) + \beta_2 \text{SMB} + \beta_3 \text{HML} + \beta_4 (\text{MktRet} - r_f)^2 + \epsilon
\]

We can do OLS, by redefining the variables: Let \(X_1 = (\text{MktRet} - r_f); X_2 = \text{SMB}; X_3 = \text{HML}; X_4 = X_1^2\). Then,

\[
Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4^2 + \epsilon
\]

| Coefficients: | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------|----------|------------|---------|----------|
| \(x_0\)       | -0.004756| 0.002854   | -1.670  | 0.0955   |
| \(xx_1\)     | 0.906527 | 0.057281   | 15.826  | <2e-16   ***|
| \(xx_2\)     | -0.215128| 0.084965   | -2.532  | 0.0116   *|
| \(xx_3\)     | -0.173160| 0.085054   | -2.036  | 0.0422   *|
| \(xx_4\)     | -0.143191| 0.617314   | -0.232  | 0.8167   => Not significant |
Functional Form: Linearity in Parameters

- We can approximate very complex non-linearities with polynomials of order $k$:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + \beta_3 X_2^3 + \ldots + \beta_{k+1} X_2^k + \epsilon$$

- Polynomial models are also useful as approximating functions to unknown nonlinear relationships. You can think of a polynomial model as the Taylor series expansion of the unknown function.

- Selecting the order of the polynomial — i.e., selecting $k$ — is not trivial.

- $k$ may be too large or too small.

Functional Form: Linearity in Parameters

- Nonlinear in parameters:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_2 \beta_3 X_4 + \epsilon$$

This model is nonlinear in parameters since the coefficient of $X_4$ is the product of the coefficients of $X_2$ and $X_3$.

- Some nonlinearities in parameters can be linearized by appropriate transformations, but not this one. This is not an intrinsic linear model.
Functional Form: Linearity in Parameters

• Intrinsic linear models can be estimated using OLS. Sometimes, transformations are needed.

• Suppose we start with a power function: \( Y = \beta_1 X^{\beta_2} \epsilon \)

• The errors enter in multiplicative form. Then, using logs:

\[
\log Y = \log \beta_1 X^{\beta_2} \epsilon = \log \beta_1 + \beta_2 \log X + \log \epsilon
\]

\[
Y' = \beta_1' + \beta_2 X' + \epsilon' \text{ where } Y' = \log Y, X' = \log X, \beta_1' = \log \beta_1, \epsilon' = \log \epsilon
\]

• Now, we have an intrinsic linear model.

• To use the OLS estimates of \( \beta_1' \) and \( \beta_2' \), we need to say something about \( \epsilon \). For example, \( \epsilon = \exp(\xi) \), where \( \xi | X \sim \text{iid } D(0, \sigma^2 I_1) \).

---

Functional Form: Linearity in Parameters

• Not all models are intrinsic linear. For example:

\[
Y = \beta_1 X^{\beta_2} + \epsilon
\]

\[
\log Y = \log(\beta_1 X^{\beta_2} + \epsilon)
\]

We cannot linearize the model by taking logarithms. There is no way of simplifying \( \log(b_1 X^b + \epsilon) \). We will have to use some nonlinear estimation technique.
**Functional Form: Piecewise Linearity**

- Sometimes non-linear relations in an interval can be linearized by splitting the interval. If this can be done, we say the relation is *piecewise linear* (a special case of a *spline regression*).
- Suppose we can linearized the data using two intervals – i.e., we have only one *knot* ($t_0$). For example:

  \[
  E[y_i | X] = \beta_{00} + \beta_{01} x_i \quad \text{if } x_i \leq t_0 \\
  E[y_i | X] = \gamma_0 + \gamma_1 x_i \quad \text{if } x_i > t_0
  \]

  Note: We can fit both equations into one single equation using a linear approximation:

  \[
  E[y_i | X] = \beta_{00} + \beta_{01} x_i + \beta_{10} (x_i - t_0)_+ + \beta_{11} (x_i - t_0)_+^1
  \]

  where $(x_i - t_0)_+$ is the positive part of $(x_i - t_0)$ and zero otherwise.

**Functional Form: Linear Splines**

- We fit both equations into one single equation:

  \[
  E[y_i | X] = \beta_{00} + \beta_{01} x_i + \beta_{10} (x_i - t_0)_+ + \beta_{11} (x_i - t_0)_+^1
  \]

  That is,

  \[
  E[y_i | X] = \beta_{00} + \beta_{01} x_i \quad \text{if } x_i \leq t_0 \\
  E[y_i | X] = \gamma_0 + \gamma_1 x_i = (\beta_{00} + \beta_{10} - \beta_{11} t_0) + (\beta_{01} + \beta_{11}) x_i \quad \text{if } x_i > t_0
  \]

- We have a linear model:

  \[
  y_i = \beta_{00} + \beta_{01} x_i + \beta_{10} (x_i - t_0)_+ + \beta_{11} (x_i - t_0)_+^1 + \varepsilon_i
  \]

  \[\Rightarrow\] It can be estimated the model using OLS.

- If in addition, we want the function to be continuous at the knot. Then,

  \[
  \beta_{00} + \beta_{01} t_0 = (\beta_{00} + \beta_{10} - \beta_{11} t_0) + (\beta_{01} + \beta_{11}) t_0 \Rightarrow \beta_{10} = 0
  \]
Functional Form: Linear vs Log specifications

- Linear model: $Y = \beta_1 + \beta_2 X + \varepsilon$
- (Semi-) Log model: $\log Y = \beta_1 + \beta_2 X + \varepsilon$

- Box–Cox transformation: $\frac{Y^\lambda - 1}{\lambda} = \beta_1 + \beta_2 X + \varepsilon$

  $\frac{Y^\lambda - 1}{\lambda} = Y - 1$ when $\lambda = 1$

  $\frac{Y^\lambda - 1}{\lambda} = \log(Y)$ when $\lambda \to 0$

- Putting $\lambda = 0$ gives the (semi-)logarithmic model (think about the limit of $\lambda$ tends to zero). We can estimate $\lambda$. One would like to test if $\lambda$ is equal to 0 or 1. It is possible that it is neither!
Functional Form: Ramsey’s RESET Test

• To test the specification of the functional form, Ramsey designed a simple test. We start with the fitted values:
  \[ \hat{y} = Xb. \]

Then, we add \( \hat{y}^2 \) to the regression specification:

\[ y = X \beta + \hat{y}^2 \gamma + \epsilon \]

• If \( \hat{y}^2 \) is added to the regression specification, it should pick up quadratic and interactive nonlinearity, if present, without necessarily being highly correlated with any of the \( X \) variables.

• We test \( H_0 \) (linear functional form): \( \gamma = 0 \)
  
  \[ H_1 \ ( \text{non linear functional form}): \gamma \neq 0 \]

• If the \( t \)-statistic for \( \hat{y}^2 \) is significant \( \Rightarrow \) evidence of nonlinearity.

• The RESET test is intended to detect nonlinearity, but not be specific about the most appropriate nonlinear model (no specific functional form is specified in \( H_1 \)).
Functional Form: Ramsey’s RESET Test

**Example:** We want to test the functional form of the 3 FF Factor Model for IBM returns, using monthly data 1973-2020.

```r
fit <- lm(ibm_x ~ Mkt_RF + SMB + HML)
y_hat <- fitted(fit)
y_hat2 <- y_hat^2
fit_ramsey <- lm(ibm_x ~ Mkt_RF + SMB + HML + y_hat2)
summary(fit_ramsey)
```

```r
> summary(fit_ramsey)

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | -0.004547  | 0.002871 | -1.584  | 0.1137  |
| Mkt_RF    | 0.903783   | 0.058003 | 15.582   | <2e-16 *** |
| SMB       | -0.217268  | 0.085128 | -2.552   | 0.0110 *  |
| HML       | -0.173276  | 0.084875 | -2.042   | 0.0417 *  |
| y_hat2    | -0.289197  | 0.763526 | -0.379   | 0.7050 \[
\Rightarrow \text{Not significant!}\]
```

Qualitative Variables and Functional Form

- Suppose that you want to model CEO compensation. You have data on annual total CEO compensation, annual returns, annual sales, and the CEO’s last degree (education). We have qualitative data.

- We can run individual regressions for each last degree --i.e., BA/BS; MS/MA/MBA; Doctoral-- but we will have three small samples:

  - Undergrad degree: $\text{Comp}_i = \beta_{0,u} + \beta_{1,u}'z_i + \epsilon_{u,i}$
  - Masters degree: $\text{Comp}_i = \beta_{0,m} + \beta_{1,m}'z_i + \epsilon_{m,i}$
  - Doctoral degree: $\text{Comp}_i = \beta_{0,d} + \beta_{1,d}'z_i + \epsilon_{d,i}$

- Alternatively, we can combine the regressions in one. We can use a variable (a dummy or indicator variable) that points whether an observation belongs to a category or class or not. For example:

  $$D_{Cj} = \begin{cases} 
  1 & \text{if observation } i \text{ belongs to category } C \text{ (say, male.)} \\
  0 & \text{otherwise.} 
\end{cases}$$
**Qualitative Variables and Functional Form**

- Define dummy/indicator variables for Masters & doctoral degrees:
  
  \[ D_m = 1 \quad \text{if at least Masters degree} \]
  \[ = 0 \quad \text{otherwise.} \]
  \[ D_d = 1 \quad \text{if doctoral degree} \]
  \[ = 0 \quad \text{otherwise.} \]

Then, we introduce the dummy/indicator variables in the model:

\[ Comp_i = \beta_0 + \beta_1 z_i + \beta_2 D_{m,i} + \beta_3 D_{d,i} + \gamma_1 z_i D_{m,i} + \gamma_2 z_i D_{d,i} + \varepsilon_i \]

This model uses all the sample to estimate the parameters. It is flexible:

- Constant for undergrad degree: \( \beta_0 \)
- Constant for Masters degree: \( \beta_0 + \beta_2 \)
- Constant for Doctoral degree: \( \beta_0 + \beta_2 + \beta_3 \)
- Slopes for Masters degree: \( \beta_1 + \gamma_1 \)
- Slopes for Doctoral degree: \( \beta_1 + \gamma_1 + \gamma_2 \)

---

**Qualitative Variables and Functional Form**

- Now, you can test the effect of education on CEO compensation. Say (1) \( H_0: \) No effect of doctoral degree: \( \beta_3 = 0 \) and \( \gamma_2 = 0 \) \( \Rightarrow F\)-test.

- Suppose we have data for CEO graduate school. We can include another indicator variable in the model. Say \( D_{T20} \) to define if a graduate school is in the Top 20.
  \[ D_{T20} = 1 \quad \text{if grad school is a Top 20 school} \]
  \[ = 0 \quad \text{otherwise.} \]

- If there is a constant, the numbers of dummy variables per qualitative variable should be equal to the number of categories minus 1. If you put the number of dummies per qualitative variable equal to the number of categories, you will create perfect multicollinearity (dummy trap).

- The omitted category is the reference category. In our previous example, the reference category is undergraduate degree.
### Dummy Variables as Seasonal Factors

- A popular use of dummy variables is in estimating seasonal effects. We may be interested in estimating the January effect for stock returns or in studying if the returns of power companies (CNP) are affected by the seasons, since in the winter and summer the power demand increases.

In this case, we define dummy/indicator variables for Summer, Fall and Winter (the base case is, thus, Spring):

- \( \text{DSum}_{i} = 1 \) if observation \( i \) occurs in Summer
  \( = 0 \) otherwise.
- \( \text{DFall}_{i} = 1 \) if observation \( i \) occurs in Fall
  \( = 0 \) otherwise.
- \( \text{DWin}_{i} = 1 \) if observation \( i \) occurs in Winter
  \( = 0 \) otherwise.

Then, letting \( Z \) be the three FF factors, we have:

\[
\text{CNP}_i = \beta_0 + \beta_1' z_i + \beta_2 \text{DSum}_i + \beta_3 \text{DFall}_i + \beta_4 \text{DWin}_i + \epsilon_i
\]

### Dummy Variables: Is There a January Effect?

#### Example (continuation):

```r
> Jan <- rep(c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0), (length(zz)/12+1)) # Create January dummy
> T2 <- T+1
> Jan_1 <- Jan[2:T2]
> fit_Jan <- lm(y ~ Mkt_RF + SMB + HML + Jan_1)
> summary(fit_Jan)
```

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | -0.007195 | 0.002566 | -2.804 0.00522 ** |
| Mkt_RF | 0.902968 | 0.056345 | 16.026 < 2e-16 *** |
| SMB | -0.240186 | 0.084013 | -2.859 0.00441 ** |
| HML | -0.190710 | 0.084317 | -2.262 0.02409 * |
| Jan_1 | 0.026993 | 0.008923 | 3.025 0.00260 ** |

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05807 on 564 degrees of freedom
Multiple R-squared: 0.3495, Adjusted R-squared: 0.3449
F-statistic: 75.75 on 4 and 564 DF, p-value: < 2.2e-16
Dummy Variable for One Observation

- We can use a dummy variable to isolate a single observation.
  \[ D_j = 1 \text{ for observation } j. \]
  \[ = 0 \text{ otherwise.} \]

- Define \( d \) to be the dummy variable in question.
  \[ Z = \text{all other regressors}. \]
  \[ X = [Z, D_j] \]

- Multiple regression of \( y \) on \( X \). We know that
  \[ X'e = 0 \text{ where } e = \text{the column vector of residuals}. \]
  \[ \Rightarrow D_j'e = 0 \Rightarrow e_j = 0 \text{ (perfect fit for observation } j). \]

- This approach can be used to deal with (eliminate) outliers.

Functional Form: Chow Test

- It is common to have a qualitative variable with two categories, say
  education (Top 20 school or not). Before modelling the data, we can
  check if only one regression model applies to both categories.

- Chow Test (an F-test) —Chow (1960, *Econometrica*):
  1. Run OLS with all the data, with no distinction between categories
     (*Restricted regression* or Pooled regression). Keep RSS_R.
  2. Run two separate OLS, one for each category (*Unrestricted regression*).
     Keep RSS_1 and RSS_2 \( \Rightarrow RSS_U = RSS_1 + RSS_2. \)
     (Alternative, we can run just one regression with the dummy variable).
  3. Run a standard F-test (testing Restricted vs. Unrestricted models):
     \[
     F = \frac{(RSS_R - RSS_U)/(k_U - k_R)}{(RSS_U)/(T-k_U)} = \frac{(RSS_R - [RSS_1 + RSS_2])}{RSS_1 + RSS_2}/(T - 2k)
     \]
Functional Form: Chow Test

- A Wald Test can also be used to compare the coefficient estimates, in the two samples (regimes 1 & 2), with \( T_1 \) and \( T_2 \) observations, respectively:

\[
W = T (\hat{\beta}_1 - \hat{\beta}_2)'Var[(\hat{\beta}_1 - \hat{\beta}_2)]^{-1}(\hat{\beta}_1 - \hat{\beta}_2)
\]

- This test is a bit more flexible, since it is easy to allow for different formulations for \( Var[(\hat{\beta}_1 - \hat{\beta}_2)] \). (In econometrics, violations of (A3) are common, for example, different variances in regimes 1 & 2.)

Chow Test: Males or Females visit doctors more?

- Taken from Greene

**German Health Care Usage Data, 7,293 Individuals, Varying Numbers of Periods**

**Variables in the file are**

Data downloaded from Journal of Applied Econometrics Archive. This is an unbalanced panel with 7,293 individuals. There are altogether 27,326 observations. The number of observations ranges from 1 to 7 per family. (Frequencies are: 1=1525, 2=2158, 3=825, 4=926, 5=1051, 6=1000, 7=987). The dependent variable of interest is

- \( \text{DOCVIS} = \) number of visits to the doctor in the observation period
- \( \text{HHNINC} = \) household nominal monthly net income in German marks / 10000.
  (4 observations with income=0 were dropped)
- \( \text{HHKIDS} = \) children under age 16 in the household = 1; otherwise = 0
- \( \text{EDUC} = \) years of schooling
- \( \text{AGE} = \) age in years
- \( \text{MARRIED} = \) marital status (1 = if married)
- \( \text{WHITEC} = 1 \) if has “white collar” job
Chow Test: Males or Females visit doctors more?

- OLS Estimation for Men only. Keep RSS_M = 379.8470

| Variable | Coefficient | Standard Error | \(b/\text{St.Er.}\) | \(P[|Z|>z]\) | Mean of X |
|----------|-------------|----------------|------------------|--------------|-----------|
| Constant | 0.04169***  | 0.00894        | 4.662            | .0000        |            |
| AGE      | 0.00086***  | 0.00013        | 6.654            | .0000        | 42.6528   |
| EDUC     | 0.02044***  | 0.00058        | 35.528           | .0000        | 11.7287   |
| MARRIED  | 0.03825***  | 0.00341        | 11.203           | .0000        | 0.76515   |
| WHITEC   | 0.03969***  | 0.00305        | 13.002           | .0000        | 0.29994   |

Chow Test: Males or Females visit doctors more?

- OLS Estimation for Women only. Keep RSS_W = 363.8789

| Variable | Coefficient | Standard Error | \(b/\text{St.Er.}\) | \(P[|Z|>z]\) | Mean of X |
|----------|-------------|----------------|------------------|--------------|-----------|
| Constant | 0.01191     | 0.01158        | 1.029            | .3036        |           |
| AGE      | 0.00026*    | 0.00014        | 1.875            | .0608        | 44.4760   |
| EDUC     | 0.01941***  | 0.00072        | 26.803           | .0000        | 10.8764   |
| MARRIED  | 0.12081***  | 0.00343        | 35.227           | .0000        | 0.75151   |
| WHITEC   | 0.06445***  | 0.00334        | 19.310           | .0000        | 0.29924   |
### Chow Test: Males or Females visit doctors more?

Ordinary least squares regression

LHS=HHNINC  Mean = .3520836
Standard deviation = .1769083
Number of observs. = 27326
Model size Parameters = 5
Degrees of freedom = 27321
Residuals Sum of squares = 752.4767  All
Residuals Sum of squares = 379.8470  Men
Residuals Sum of squares = 363.8789  Women

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
|----------|-------------|----------------|---------|---------|-----------|
| Constant | .04186***   | .00704         | 5.949   | .0000   |           |
| AGE      | .00030***   | .91958D-04     | 3.209   | .0013   | 43.5257   |
| EDUC     | .01967***   | .00045         | 44.180  | .0000   | 11.3206   |
| MARRIED  | .07947***   | .00239         | 33.192  | .0000   | .75862    |
| WHITEC   | .04819***   | .00225         | 21.465  | .0000   | .29960    |

Chow Test: Males or Females visit doctors more?

\[
\text{Chow Test} = F = \frac{(752.4767 - (379.847 + 363.8789))/5}{(379.847 + 363.8789)/(27326 - 10)} = 64.281 \\
F(5, 27311) = 2.214100 \Rightarrow \text{reject } H_0
\]
Functional Form: Structural Change

• Suppose there is an event that we think had a big effect on the behaviour of our model. Suppose the event occurred at time \( T_{SB} \).
  
  For example, the parameters are different before and after \( T_{SB} \). That is,

  \[
  y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \epsilon_i \quad \text{for } i \leq T_{SB}
  \]

  \[
  y_i = \beta'_0 + \beta'_1 X_{1,i} + \beta'_2 X_{2,i} + \beta'_3 X_{3,i} + \epsilon_i \quad \text{for } i > T_{SB}
  \]

  The event caused structural change in the model. \( T_{SB} \) separates the behaviour of the model in two regimes/categories ("before" & "after").

• A Chow test tests if one model applies to both regimes:

  \[
  y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \epsilon_i \quad \text{for all } i
  \]

• Under \( H_0 \) (No structural change), the parameters are the same for all \( i \).

• We test \( H_0 \) (No structural change): \( \beta_0^1 = \beta_0^2 = \beta_0 \)
  \( \beta_1^1 = \beta_1^2 = \beta_1 \)
  \( \beta_2^1 = \beta_2^2 = \beta_2 \)
  \( \beta_3^1 = \beta_3^2 = \beta_3 \)

  \( H_1 \) (structural change): For at least one \( k \) (= 0, 1, 2, 3): \( \beta_k^1 \neq \beta_k^2 \)

• What events may have this effect on a model? A financial crisis, a big recession, an oil shock, Covid-19, etc.

• Testing for structural change is the more popular use of Chow tests.

• Chow tests have many interpretations: tests for structural breaks, pooling groups, parameter stability, predictive power, etc.

• One important consideration: \( T \) may not be large enough.
Functional Form: Structural Change

- We structure the Chow test to test H₀ (No structural change), as usual.

- Steps for Chow (Structural Change) Test:

  1. Run OLS with all the data, with no distinction between regimes. (Restricted or pooled model). Keep RSSᵣ.

  2. Run two separate OLS, one for each regime (Unrestricted model):

     Before Date Tₜᵢₜᵢ: Keep RSS₁.

     After Date Tₜᵢₕᵢᵢ: Keep RSS₂. ⇒ RSSᵤ = RSS₁ + RSS₂.

  3. Run a standard F-test (testing Restricted vs. Unrestricted models):

     \[ F = \frac{(RSSᵣ - RSSᵤ)/(kᵤ - kᵣ)}{(RSSᵤ)/(T - kᵤ)} \]

     \[ = \frac{(RSSᵣ - [RSS₁ + RSS₂]) / k}{(RSS₁ + RSS₂) / (T - 2k)} \]

Example: We test if the Oct 1973 oil shock in quarterly GDP growth rates had an structural change on the GDP growth rate model.

We model GDP the growth rate with an AR(1) model, that is, GDP growth rate depends only on its own lagged growth rate:

\[ yₜ = \beta₀ + \beta₁ yₜ₋₁ + \epsilonᵢ \]

GDP_da <- read.csv("http://www.bauerca.edu/rsusmel/4397/GDP_q.csv", head=TRUE, sep=";")
x_date <- GDP_da$DATE
x_gdp <- GDP_da$GDP
x_dummy <- GDP_da$D73
T <- length(x_gdp)
T <- T - 108 # Tₛᵢᵢ = Oct 1973
lr_gdp <- log(x_gdp[-1]/x_gdp[-T])
T <- length(lr_gdp)
lr_gdp0 <- lr_gdp[-1]
lr_gdp1 <- lr_gdp[-T]
tₜₛᵢᵢ <- tₛᵢᵢ - 1 # Adjust tₛ (we lost the first observation)
Example (continuation):

\begin{verbatim}
y <- lr_gdp0
x1 <- lr_gdp1
T <- length(y)
x0 <- matrix(1,T,1)
x <- cbind(x0,x1)
k <- ncol(x)

# Restricted Model (Pooling all data)
fit_ar1 <- lm(lr_gdp0 ~ lr_gdp1) # Fitting AR(1) (Restricted) Model
e_R <- fit_ar1 $residuals # regression residuals, e
RSS_R <- sum(e_R^2) # RSS Restricted

> summary(fit_ar1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.011565   0.001145  10.097  < 2e-16 ***
lr_gdp1     0.244846   0.056687   4.319 2.14e-05 ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.01296 on 294 degrees of freedom

# Unrestricted Model (Two regimes)
y_1 <- y[1:t_s]
x_u1 <- x[1:t_s,]
fit_ar1_1 <- lm(y_1 ~ x_u1 - 1) # AR(1) Regime 1
e1 <- fit_ar1_1$residuals # Regime 1 regression residuals, e
RSS1 <- sum(e1^2) # RSS Regime 1

kk = t_s+1 # Starting date for Regime 2
y_2 <- y[kk:T]
x_u2 <- x[kk:T,]
fit_ar1_2 <- lm(y_2 ~ x_u2 - 1) # AR(1) Regime 2
e2 <- fit_ar1_2$residuals # Regime 2 regression residuals, e
RSS2 <- sum(e2^2) # RSS Regime 2

F <- ((RSS_R - (RSS1+RSS2))/k)/((RSS1+RSS2)/(T - 2*k))

> F
[1] 4.877371

p_val <- 1 - pf(F, df1 = 2, df2 = T - 2*k) # p-value of F_test

> p_val
[1] 0.00824892 => small p-values: Reject H0 (No structural change).
\end{verbatim}
Example: 3 Factor Fama-French Model for IBM (continuation)


Pooled RSS = 1.9324

Jan 1973 – Dec 2001 RSS = RSS₁ = 1.33068 (T = 342)
Jan 2002 – June 2020 RSS = RSS₂ = 0.57912 (T = 227)

\[ F = \frac{(RSS_R - (RSS₁ + RSS₂))/k}{(RSS₁ + RSS₂)/(T-k)} = \frac{1.9324 - (1.3307 + 0.57911))}{4} = 1.6627 \]

⇒ Since \( F_{4,565.05} = 2.39 \), we cannot reject \( H₀ \)

<table>
<thead>
<tr>
<th></th>
<th>Constant</th>
<th>Mkt – rf</th>
<th>SMB</th>
<th>HML</th>
<th>RSS</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973-2020</td>
<td>-0.0051</td>
<td>-0.2125</td>
<td>0.9083</td>
<td>-0.1715</td>
<td>1.9324</td>
<td>569</td>
</tr>
<tr>
<td>1973-2001</td>
<td>-0.0038</td>
<td>-0.2230</td>
<td>0.8092</td>
<td>-0.1970</td>
<td>1.3307</td>
<td>342</td>
</tr>
<tr>
<td>2002 – 2020</td>
<td>-0.0073</td>
<td>-0.3329</td>
<td>1.0874</td>
<td>-0.1955</td>
<td>0.5791</td>
<td>227</td>
</tr>
</tbody>
</table>

Functional Form: Structural Change

• Under the \( H₀ \) (No structural change), we pool the data into one model. That is, the parameters are the same under both regimes. We fit the same model for all \( i \), for example:

\[ y_i = β_0 + β_1'x_i + ε_i \]

• If the Chow test rejects \( H₀ \), we need to reformulate the model. A typical reformulation includes a dummy variable \( (D_{SB,i}) \). For example:

\[ y_i = β_0 + β_1'x_i + β_2D_{SB,i} + γ_1'D_{SB,i} + ε_i \]

where

\[ D_{SB,i} = \begin{cases} 1 & \text{if observation } i \text{ occurred after } T_{SB} \\ 0 & \text{otherwise} \end{cases} \]
Example: We are interested in the effect of the Oct 1973 oil shock in GDP growth rates. We include a dummy variable in the model, say \( D_{73,i} \):

\[
D_{73,i} = 1 \quad \text{if observation } i \text{ occurred after October 1973}
\]

\[
D_{73,i} = 0 \quad \text{otherwise.}
\]

\[
y_i = \beta_0 + \beta_1'x_i + \beta_2 D_{73,i} + \gamma_1'x_i D_{73,i} + \varepsilon_i
\]

In the model, the oil shock affects the constant and the slopes.

<table>
<thead>
<tr>
<th>Before oil shock (( D_{73} = 0 )):</th>
<th>Constant</th>
<th>Slopes:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>( \beta_1' )</td>
<td></td>
</tr>
</tbody>
</table>

| After oil shock (\( D_{73} = 1 \)): | \( \beta_0 + \beta_2 \) | \( \beta_1' + \gamma_1' \) |

- We can estimate the above model and do an F-test to test if \( H_0 \) (No structural change): \( \beta_2 = 0 \) & \( \gamma_1 = 0 \).

Chow Test: Structural Change - Example

Example: 3-Factor Fama-French Model for GE

Q: Did the dot.com bubble (end of 2001) affect the structure of the FF Model?

Sample: Jan 1973 – July 2020 (\( T = 570 \)).

Pooled RSS = 1.569956

Jan 1973 – Dec 2001 RSS = \( RSS_1 = 0.5455917 \) (\( T = 342 \))

Jan 2002 – July 2020 RSS = \( RSS_2 = 0.9348033 \) (\( T = 228 \))

\[
F = \frac{[RSS_{p} - (RSS_1 + RSS_2)]/k}{(RSS_1 + RSS_2)/(T-k)} = \frac{[1.5700 - (0.5456 + 0.9348)/4]}{(0.5456 + 0.9348)/570 - 2\times 4} = 8.499996
\]

\( \Rightarrow \) Since \( \chi^2_3 = 9.4877 \), we cannot reject \( H_0 \) at 5% level. But, \( p-value = .07488 \) (close enough? F distribution good?)

Conclusion: At the 5% level, we do not have evidence for a Dot.com bubble structural change.
Chow Test: Structural Change - Example

• But, we can try different breaking points, starting at \( T=85 \):

\[
\text{Chow Test: Structural Change - Example}
\]

\[
\text{Note: Recall that the Chow test is an F-test, we are testing a joint hypothesis, all coefficients are subject to structural change.}
\]

\[
\text{Example (continuation): Now, we check the effect of the dot.com bubble (create dummy) on the constant using a t-test:}
\]

\[
T \leftarrow \text{length(ge_x)}
\]

\[
x_\text{break} \leftarrow 342
\]

\[
dot_0 \leftarrow \text{rep}(0, x_\text{break}) \quad \# \text{0 up to Dec 2001}
\]

\[
dot_1 \leftarrow \text{rep}(1, T - x_\text{break}) \quad \# \text{1 after Dec 2001}
\]

\[
dot \leftarrow c(dot_0, dot_1) \quad \# \text{Doc.com dummy}
\]

\[
\text{fit_ge_dot} \leftarrow \text{lm(ge_x} \sim \text{Mkt_RF + SMB + HML + dot)}
\]

\[
> \text{summary(fit_ge_dot)}
\]

\[
\begin{array}{lrrrr}
\text{Coefficients:} & \text{Estimate} & \text{Std. Error} & t value & \text{Pr(>|t|)} \\
(Intercept) & -0.003273 & 0.002877 & -1.138 & 0.25566 \\
\text{Mkt_RF} & 1.226412 & 0.050868 & 24.110 & < 2e-16 *** \\
\text{SMB} & -0.308411 & 0.075433 & -4.089 & 4.97e-05 *** \\
\text{HML} & 0.341709 & 0.075755 & 4.511 & 7.86e-06 *** \\
\text{dot} & -0.013052 & 0.004502 & -2.899 & 0.00388 **
\end{array}
\]

\[
\Rightarrow \text{significant effect on constant.}
\]
Chow Test: Structural Change - Issues

• Issues with Chow tests
  - The results are conditional on the breaking point—say, October 73 or Dec 2001.
  - The breaking point is usually unknown. It needs to be estimated.
  - It can deal only with one structural break—i.e., two categories!
  - The number of breaks is also unknown.
  - Heteroscedasticity—for example, structural breaks in the variance—and unit roots (high persistence) complicate the test.
  - In general, only asymptotic (consistent) results are available.

Structural Change: Unknown Break

• For an unknown break date, Quandt (1958, 1960) proposed a likelihood ratio test statistics, called Supremum (Max)-Test,
  \[ QLR_T = \max_{\tau \in [\tau_{\text{min}}, \tau_{\text{max}}]} F_T(\tau) \]
  The max (supremum) is taken over all potential breaks in \((\tau_{\text{min}}, \tau_{\text{max}})\).
  For example, \(\tau_{\text{min}} = T^*.15; \tau_{\text{max}} = T^*.85\).
  Easy to calculate \(QLR_T\) with a do loop.

  The assumptions that make the LR-statistic asymptotically \(\chi^2\) do not apply in this setting. (Quandt was aware of the problem, but did not know how to derive the asymptotic null distribution of \(QLR_T\).)

  Problem: The (nuisance) parameter \(\tau\) is not identified under \(H_0\) (no structural break) \(\Rightarrow\) regularity conditions are violated!
Structural Change: Unknown Break

• Andrews (1993) showed that under appropriate regularity conditions, the QLR statistic, also referred to as a SupLR statistic, has a nonstandard limiting distribution:

\[ QLR_T \overset{d}{\to} \sup_{r \in [r_{\min}, r_{\max}]} \frac{B_k(r)'}{r(1-r)} \]

where \(0 < r_{\min} < r_{\max} < 1\) and \(B_k(\cdot)\) is a “Brownian Bridge” process defined on \([0,1]\). Percentiles of this distribution as functions of \(r_{\max}, r_{\min}\) and \(k\) are tabulated in Andrews (1993). (Critical values much larger than \(\chi^2\).)

Note: A Brownian bridge is a continuous-time stochastic process \(B(t)\) whose probability distribution is the conditional probability distribution of a Wiener process \(W(t)\) given the condition that \(B(0) = B(1) = 0\). The increments in a Brownian bridge are not independent. 

Example: \(B(t) = W(t) - tW(1)\) is a Brownian Bridge.

Structural Change: Unknown Break - Example

Example: 3 Factor Fama-French Model for GE

Andrews’ (1993) test with \(r_{\min} = 50 (T^{*.15})\), \(r_{\max} = 286 (T^{*.85})\)

\(\sup F = 14.5936\) at \(t = 433\) (April 2008)

Critical value \((k=4, \pi_1=r_{\min}/T=(1-r_{\max}/T)=.15, \& \alpha=.05) = 16.45 \Rightarrow \text{reject } H_0\)

• Q: Multiple breaks?
Structural Change: Unknown Break - Example

\[ b <- \text{solve}(t(x) %*% x) %*% t(x) %*% y \]  
\# \( b = (X'X)^{-1} X'y \) (OLS regression)

\[ e <- y - x %*% b \]  
\# regression residuals, e

\[ \text{RSS}_R <- \text{as.numeric}(t(e) %*% e) \]  
\# RSS R

\[ T_1 <- \text{round}(T * .15) \]
\[ T_2 <- \text{round}(T * .85) \]
\[ \text{All}_F <- \text{matrix}(0, T_2 - T_1, 1) \]
\[ t <- T_1 \]
\[ \text{while } (t <= T_2) \{ \]
  \[ y_1 <- y[1:T_1] \]
  \[ x_{u1} <- x[1:T_1,] \]
  \[ b_1 <- \text{solve}(t(x_{u1}) %*% x_{u1}) %*% t(x_{u1}) %*% y_1 \]
  \[ e_1 <- y_1 - x_{u1} %*% b_1 \]
  \[ \text{RSS}_1 <- \text{as.numeric}(t(e_1) %*% e_1) \]  
\# RSS 1

  \[ k k = t + 1 \]
  \[ y_2 <- y[k:k:T] \]
  \[ x_{u2} <- x[k:k:T,] \]
  \[ b_2 <- \text{solve}(t(x_{u2}) %*% x_{u2}) %*% t(x_{u2}) %*% y_2 \]
  \[ e_2 <- y_2 - x_{u2} %*% b_2 \]
  \[ \text{RSS}_2 <- \text{as.numeric}(t(e_2) %*% e_2) \]  
\# RSS 2

\[ F <- (\text{RSS}_R - (\text{RSS}_1 + \text{RSS}_2) / k) / ((\text{RSS}_1 + \text{RSS}_2) / (T_1 - k)) \]
\[ \text{All}_F = \text{rbind}(\text{All}_F, F) \]
\[ \} \]
\[ \text{plot}(\text{All}_F, xlab = "k", ylab = "F-test", main = "F-test at different Break Points") \]

Forecasting and Prediction

“There are two kind of forecasters: those who don’t know and those who don’t know they don’t know”

*John Kenneth Galbraith (1993)*

- **Objective:** Forecast
- **Distinction:** Ex post vs. Ex ante forecasting
  - Ex post: RHS data are observed
  - Ex ante (true forecasting): RHS data must be forecasted

- **Prediction and Forecast**
  Prediction: Explaining an outcome, which could be a future outcome.
  Forecast: A particular prediction, focusing in a future outcome.

**Example:** Prediction: Given \( x^0 \) \( \Rightarrow \) predict \( y^0 \).  
Forecasts: Given \( x^0_{t+1} \) \( \Rightarrow \) predict \( y_{t+1} \).
Forecasting and Prediction

- Two types of predictions:
  - In sample (prediction): The expected value of $y$ (in-sample), given the estimates of the parameters.
  - Out of sample (forecasting): The value of a future $y$ that is not observed by the sample.

Notation:
- Prediction for $T$ made at $T$: $\hat{Y}_T$.
- Forecast for $T+l$ made at $T$: $\hat{Y}_{T+l}, \hat{Y}_{T+l|T}, \hat{Y}_T(l)$.

where $T$ is the forecast origin and $l$ is the forecast horizon. Then, $\hat{Y}_T(l)$: $l$-step ahead forecast = Forecasted value $Y_{T+l}$ at time $T$.

Forecasting and Prediction

- Any prediction or forecast needs an information set, $I_T$. This includes data, models and/or assumptions available at time $T$. The predictions and forecasts will be conditional on $I_T$.

For example, in-sample, $I_T = \{x^0\}$ to predict $y^0$. Or in a time series context, $I_T = \{x_{T-1}^0, x_{T-2}^0, ..., x_{T-q}^0\}$ to predict $y_{T+1}$.

- Then, the forecast is just the conditional expectation of $Y_{T+1}$ given the observed sample:
  $$\hat{Y}_{T+l} = E[Y_{T+l} | X_T, X_{T-1}, ..., X_1]$$

Example: If $X_T = Y_1$, then, the one-step ahead forecast is:
  $$\hat{Y}_{T+1} = E[Y_{T+1} | Y_T, Y_{T-1}, ..., Y_1]$$
Forecasting and Prediction

• Keep in mind that the forecasts are a random variable. Technically speaking, they can be fully characterized by a pdf.

• In general, it is difficult to get the pdf for the forecast. In practice, we get a point estimate (the forecast) and a C.I.

Forecasting and Prediction

• Prediction vs. model validation.
  – Within sample prediction: Using the whole sample (T observations), we predict y as usual, with ŷ.

  – Hold out sample: We estimate the model only using a part of the sample (say, up to time T₁). The rest of the sample (T - T₁ observations) are used to check the predictive power of the model –i.e., the accuracy of predictions, by comparing yᵯ with actual y).

*Model validation* refers to establishing the statistical adequacy of the assumptions behind the model –i.e., (A1)-(A5) in this lecture. Predictive power can be used to do model validation.
Forecasting and Prediction: Model Validation

Steps to measure forecast accuracy:
1) Select a (long) part of the sample (estimation period) to estimate the parameters of the model. (Get in-sample forecasts, \( \hat{y} \).)
2) Keep a (short) part of the sample to check the model's forecasting skills. This is the validation step. You can calculate true MSE or MAE.
3) If happy with Step 2), proceed to do out-of-sample forecasts.

Prediction Intervals: Point Estimate

- Prediction: Given \( x^0 \) ⇒ predict \( y^0 \).

- Given the CLM, we have:
  
  \[ \begin{align*}
  \text{Expectation:} & \quad E[y | X, x^0] = \beta'x^0; \\
  \text{Predictor:} & \quad \hat{y}^0 = b'x^0; \\
  \text{Realization:} & \quad \hat{y}^0 = \beta'x^0 + \epsilon^0
  \end{align*} \]

  Note: The predictor includes an estimate of \( \epsilon^0 \):
  \( \hat{y}^0 = b'x^0 \) + estimate of \( \epsilon^0 \). (Estimate of \( \epsilon^0 = 0 \), but with variance.)

- Associated with the prediction (a point estimate), there is a forecast error:
  \[ \hat{y}^0 - y^0 = b'x^0 - \beta'x^0 - \epsilon^0 = (b - \beta)'x^0 - \epsilon^0 \]
  \[ \Rightarrow \text{Var}(\hat{y}^0 - y^0 | x^0) = E[(\hat{y}^0 - y^0)'(\hat{y}^0 - y^0) | x^0] \\
  = x^0 \text{Var}(b - \beta) | x^0 | x^0 + \sigma^2 \]
Example: We have already estimated the 3 Factor Fama-French Model for IBM returns:

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | -0.005089| 0.002488   | -2.046  | 0.0412 * |
| Mkt_RF         | 0.908299 | 0.056722   | 16.013  | <2e-16 ***|
| SMB            | -0.212460| 0.084112   | -2.526  | 0.0118 * |
| HML            | -0.171500| 0.084682   | -2.025  | 0.0433 * |

Suppose we are given \( x^0 = [1.0000, -0.0189, -0.0142, -0.0027] \)
Then,

\[
\hat{y}_0 = -0.005089 + 0.908299 \times (-0.0189) -0.212460 \times -0.0142 - 0.171500 \times (-0.0027) = -0.01877582
\]

Suppose we observe \( y_0 = 0.1555214 \). Then, the forecast error is

\[
\hat{y}_0 - y_0 = -0.01877582 - 0.1555214 = -0.1742973
\]

Prediction Intervals: Point Estimate

Example: In R:

```r
> x_0 <- rbind(1.0000, -0.0189, -0.0142, -0.0027)
> y_0 <- 0.1555214
> y_f0 <- t(b)%*% x_0
> y_f0
  [,1]
[1,] -0.01877582
> ef_0 <- y_f0 - y_0
> ef_0
  [,1]
[1,] -0.1742973
```
Prediction Intervals: C.I.

- How do we estimate the uncertainty behind the forecast? Form a confidence interval.

Two cases:
(1) If \( x^0 \) is given – i.e., constants. Then,
\[
\text{Var}[\hat{y}^0 - y^0| x^0] = x^0' \text{Var}[b|x^0] x^0 + \sigma^2
\]
\[\implies \text{Form C.I. as usual.}\]

Note: In out-of-sample forecasting, \( x^0 \) is unknown, it has to be estimated.

(2) If \( x^0 \) has to be estimated, then we use a random variable. What is the variance of the product? One possibility: Use a bootstrap to form a C.I.

Prediction Intervals: C.I. and Forecast Variance

- Assuming \( x^0 \) is known, the variance of the forecast error is
\[
\sigma^2 + x^0' \text{Var}[b|x^0] x^0 = \sigma^2 + \sigma^2[ x^0' (X'X)^{-1} x^0]
\]
If the model contains a constant term, this is
\[
\text{Var}[e^0] = \sigma^2 \left[ 1 + \frac{1}{n} + \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} (x_j^0 - \bar{x}_j)(x_k^0 - \bar{x}_k) (Z'M^0Z)^j_k \right]
\]
(where \( Z \) is \( X \) without \( x_1 \)). In terms squares and cross products of deviations from means.

Note: Large \( \sigma^2 \), small \( n \), and large deviations from the means, decrease the precision of the forecasting error.

Interpretation: Forecast variance is smallest in the middle of our “experience” and increases as we move outside it.
Prediction Intervals: C.I. and Forecast Variance

• Then, the \( (1 - \alpha)\% \) C.I. is given by: 
\[
\hat{y}_0 \pm t_{\alpha,k,q/2} \cdot \sqrt{\text{Var}(e^0)}
\]

• As \( x^0 \) moves away from , the C.I increases, this is known as the “butterfly effect.”

Prediction Intervals

**Example (continuation):** We want to calculate the variance of the forecast error: for the given \( x^0 = [1.0000 -0.0189 -0.0142 -0.0027] \)
Recall we got \( \hat{y}_0 = b'x_0 = -0.01877587 \)
Then,

\[
\text{Estimated } \text{Var}(\hat{y}_0 - y^0 \mid x^0) = x^0' \text{Var}(b \mid x^0) x^0 + \sigma^2 = 0.003429632
\]

Check: What is the forecast error if \( x^0 = \text{colMeans}(x) \)?

\[
\text{var} \left( \hat{y} - \hat{y}_0 \right) = \text{var}(\hat{y}) + \sigma^2
\]

\[
\text{var} \left( \hat{y} - \hat{y}_0 \right) = \text{var}(\hat{y}) + \sigma^2
\]

\[
\text{var}(\hat{y}) = x_0' \text{Var}(\hat{y}) x_0 + \sigma^2
\]

\[
\text{var}(\hat{y}) = x_0' \text{Var}(\hat{y}) x_0 + \sigma^2
\]
Prediction Intervals

Example (continuation):

```r
> # (1-alpha)% C.I. for prediction (alpha = .05)
> CI_lb <- y_f0 - 1.96 * sqrt(var_ef_0)
> CI_lb
> [1] -0.1335594
> CI_ub <- y_f0 + 1.96 * sqrt(var_ef_0)
> CI_ub
> [1] 0.09600778
```

That is, CI for prediction: \([-0.13356; 0.09601]\) with 95% confidence.

Forecasting performance of a model: Tests and measures of performance

- Evaluation of a model’s predictive accuracy for individual (in-sample and out-of-sample) observations
- Evaluation of a model’s predictive accuracy for a group of (in-sample and out-of-sample) observations
- Chow prediction test
Evaluation of forecasts: Measures of Accuracy

- Summary measures of out-of-sample forecast accuracy

Mean Error = \[ \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i) = \frac{1}{m} \sum_{i=1}^{m} e_i \]

Mean Absolute Error (MAE) = \[ \frac{1}{m} \sum_{i=1}^{m} |\hat{y}_i - y_i| = \frac{1}{m} \sum_{i=1}^{m} |e_i| \]

Mean Squared Error (MSE) = \[ \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 = \frac{1}{m} \sum_{i=1}^{m} e_i^2 \]

Root Mean Square Error (RMSE) = \[ \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} e_i^2} \]

Theil’s U-stat = \[ U = \frac{1}{m} \sum_{i=1}^{m} e_i^2 \]
\[ \sqrt{T} \sum y_i^2 \]

Evaluation of forecasts: Measures of Accuracy

- Theil’s U statistics has the interpretation of an \( R^2 \). But, it is not restricted to be smaller than 1.

- The lower the above criteria, say MSE, the better the forecasting ability of our model.

- Q: How do we know an MSE for model 1 is better than the MSE for model 2?
Example: We want to check the forecast accuracy of the 3 FF Factor Model for IBM returns. We estimate the model using only 1973 to 2017 data (T=539), leaving 2018-2020 (30 observations) for validation of predictions.

```r
T0 <- 1
T1 <- 539
T2 <- T1+1
y1 <- y[T0:T1]
x1 <- x[T0:T1,]
fit2 <- lm(y1~ x1-1)
summary(fit2)
```

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| x1       | -0.003848  | 0.002571| -1.497   | 0.13510   |
| x1Mkt_RF | 0.865579   | 0.059386| 14.575   | < 2e-16 *** |
| x1SMB    | -0.224914  | 0.085505| -2.630   | 0.00877 ** |
| x1HML    | -0.230838  | 0.090251| -2.558   | 0.01081 *  |

Evaluation of forecasts: Measures of Accuracy

Example (continuation): We condition on the observed data from 2018: Jan to 2020: Jun.

```r
x_0 <- x[T2:T,]
y_0 <- y[T2:T]
y_f0 <- x_0%*% b1
ef_0 <- y_f0 - y_0
mes_ef_0 <- sum(ef_0^2)/nrow(x_0)
mae_ef_0 <- sum(abs(ef_0))/nrow(x_0)
```

That is, \( \text{MSE} = 0.003703207 \)
\( \text{MAE} = 0.04518326 \)
Example (continuation): Plot of actual IBM returns and forecasts.

```r
plot(y_f0, type="l", col="red", main = "IBM: Actual vs. Forecast (2018-2020)", xlab = "Obs", ylab = "Forecast")
lines(y_0, type = "l", col = "blue")
legend("topleft", legend = c("Actual", "Forecast"), col = c("blue", "red"), lty = 1)
```

Evaluation of forecasts: Measures of Accuracy

- Suppose two competing forecasting procedures produce errors $e_t^{(1)}$ and $e_t^{(2)}$ for $t=1, \ldots, T$. Then, if expected MSE is the criterion used, the procedure with the lower MSE will be judged superior.

- We want to test $H_0: \text{MSE}(1) = \text{MSE}(2)$
  $H_1: \text{MSE}(1) \neq \text{MSE}(2)$.

Assumptions: forecast errors are unbiased, normal, and uncorrelated.

- Consider, the pair of RVs $e_t^{(1)} + e_t^{(2)}$ and $e_t^{(1)} - e_t^{(2)}$. Now,
  $$E[(e_t^{(1)} + e_t^{(2)})(e_t^{(1)} - e_t^{(2)})] = \sigma_1^2 - \sigma_2^2$$

- That is, we test $H_0$ by testing that the two RVs are not correlated! This idea is due to Morgan, Granger and Newbold (MGN, 1977).
Evaluation of forecasts: Testing Accuracy

- There is a simpler way to do the MGN test. Let,
  \[ x_t = e_t^{(1)} + e_t^{(2)} \]
  \[ z_t = e_t^{(1)} - e_t^{(2)} \]
  
  (1) Do a regression: \[ x_t = \beta z_t + \varepsilon_t \]
  (2) Test \( H_0: \beta = 0 \) \( \Rightarrow \) a simple \( t \)-test.

  The MGN test statistic is exactly the same as that for testing the null hypothesis that \( \beta = 0 \) in this regression (recall: \( b = (XX')^{-1}Xy \)). This is the approach taken by Harvey, Leybourne and Newbold (1997).

- If the assumptions are violated, these tests have problems.

- A non-parametric HLN variation: Spearman’s rank test for zero correlation between \( x_t \) and \( z_t \).

---

Evaluation of forecasts: Testing Accuracy

**Example:** We produce IBM returns one-step-ahead forecasts for 2018-2020 using the 3 FF Factor Model for IBM returns:

\[
(IBM_{Ret} - r_{\ell})_t = \beta_0 + \beta_1 (Mkt_{Ret} - r_{\ell})_t + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_t
\]

Taking expectations at time \( t+1 \), conditioning on time \( t \) information set, \( I_t = \{Mkt_{Ret} - r_{\ell}, SMB_t, HML_t\} \)

\[
E[(IBM_{Ret} - r_{\ell})_{t+1} | I_t] = \beta_0 + \beta_1 E[(Mkt_{Ret} - r_{\ell})_{t+1} | I_t] + \beta_2 E[SMB_{t+1} | I_t] + \beta_3 E[HML_{t+1} | I_t]
\]

In order to produce forecast, we will make a naive assumption: The best forecast for the FF factors is the previous observation. Then,

\[
E[(IBM_{Ret} - r_{\ell})_{t+1} | I_t] = \beta_0 + \beta_1 (Mkt_{Ret} - r_{\ell})_t + \beta_2 SMB_t + \beta_3 HML_t.
\]

Now, replacing the \( \beta \) by the estimated \( b \), we have our one-step-ahead forecasts.
Example: We compare the forecast accuracy relative to a random walk model for IBM returns. That is,
\[ E[(\text{IBM}_{\text{Ret}_{t+1}} - \text{rf}_{t+1}) | I_t] = (\text{IBM}_{\text{Ret}_{t+1}} - \text{rf}_{t}) \]

Using R, we create the forecasting errors for both models and MSE:

```r
> x_01 <- x[T1:(T-1),]
> y_0 <- y[T2:T]
> y_f0 <- x_01%*% b1
> ef_0 <- y_f0 - y_0 # \(e_t^{(2)}\)
> mes_ef_0 <- sum(ef_0^2)/nrow(x_0)
> mes_ef_0 # MSE(2)
[1] 0.01106811
> ef_rw_0 <- y[T1:(T-1)] - y_0 # \(e_t^{(1)}\)
> mse_ef_rw_0 <- sum(ef_rw_0^2)/nrow(x_0)
> mse_ef_rw_0 # MSE(1) <= (1) is the higher MSE.
[1] 0.02031009
```

Evaluation of forecasts: Testing Accuracy

Example: Now, we create
\[ x_t = e_t^{(1)} + e_t^{(2)}, \quad \zeta_t = e_t^{(1)} - e_t^{(2)}. \]

Then, regress: \(x_t = \beta \zeta_t + \epsilon_t\), and test \(H_0: \beta = 0.\)

```r
> z_mgn <- ef_rw_0 + ef_0
> x_mgn <- ef_rw_0 - ef_0
> fit_mgn <- lm(z_mgn ~ x_mgn)
> summary(fit_mgn)

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 0.05688 | 0.03512 | 1.619 | 0.117 |
| x_mgn | 2.77770 | 0.58332 | 4.762 | 5.32e-05 *** |

Conclusion: We reject that both MSE are equal \(\Rightarrow\) MSE of RW is higher.
Out-of-sample predictions and prediction errors: Chow Test Revisited (Greene)

- Variation of the Chow test: Chow Predictive Test

- When there is not enough data to do the regression on both sub-samples, we can use an alternative formulation of the Chow test.
  1. We estimate the regression over a (long) sub-period, with \( T_1 \) observations –say 3/4 of the sample. Keep RSS_1.
  2. We estimate the regression for the whole sample (restricted regression). Keep RSS_R.
  3. Run an \( F \)-test, where the numerator represents a “predicted” RSS for the \( T_2 \) (=\( T - T_1 \)) left out observations.

\[
F = \frac{(RSS_R - RSS_1)/T_2}{RSS_1/(T_1 - k)} \sim F_{T_2 \cdot (T_1 - k)}
\]

Example: 3 Factor Fama-French Model for IBM (continuation)

We have \( T = 336 \) observations. We set \( T_1 = 252 \) & \( T_2 = 86 \). Then, \( \text{RSS}_{252} = 8.063611 \).
\( \text{RSS}_{336} = 12.92964 \).

\( \Rightarrow F_{FF} = \frac{(12.92964 - 8.063611)/86}{8.063611/(336-4)} = 2.329618 \)

Since \( F_{86,332,05} = 1.308807 < F_{FF} \Rightarrow \text{reject } H_0 \) (constant parameters).