Lecture 5
Functional Form and Prediction

 1

OLS Estimation - Assumptions

• CLM Assumptions

(**A1**) DGP: $\boldsymbol{y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is correctly specified.

(**A2**) E[$\boldsymbol{\varepsilon}|\boldsymbol{X}$] = 0

(**A3**) Var[$\boldsymbol{\varepsilon}|\boldsymbol{X}$] = $\sigma^2 \mathbf{I}_T$

(**A4**) $\boldsymbol{X}$ has full column rank –rank($\boldsymbol{X}$) = k–, where $T \geq k$.

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

• But, it turns out that in the framework of OLS estimation, we can be more flexible with $f(\boldsymbol{X}, \beta)$.

Functional Form: Linearity in Parameters

• Linear in variables and parameters:

$$\boldsymbol{y} = \beta_1 + \beta_2 \boldsymbol{X}_2 + \beta_3 \boldsymbol{X}_3 + \beta_4 \boldsymbol{X}_4 + \boldsymbol{\varepsilon}$$

• 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 $\boldsymbol{y}$ and $\boldsymbol{X}$, but OLS still can be used.

7

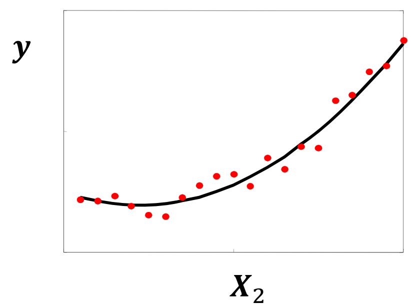
Functional Form: Linearity in Parameters

• Suppose we have:

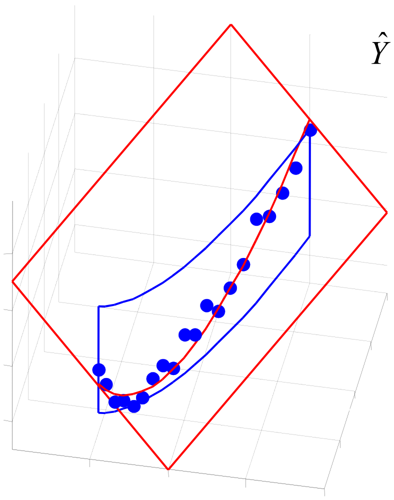
$$\boldsymbol{y} = \beta_1 + \beta_2 \boldsymbol{X}_2 + \beta_3 \boldsymbol{X}_2^2 + \boldsymbol{\varepsilon}$$

• The model is intrinsic linear, but it allows for a quadratic relation between $\boldsymbol{y}$ and $\boldsymbol{X}_2$:

:



Functional Form: Linearity in Parameters



Functional Form: Linearity in Parameters

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

$$\boldsymbol{r_i} - \boldsymbol{r_f} = \beta_0 + \beta_1 (\boldsymbol{r_{m,t}} - \boldsymbol{r_f}) + \beta_2\, \mathbf{SMB} + \beta_3\, \mathbf{HML} + \beta_4 (\boldsymbol{r_{m,t}} - \boldsymbol{r_f})^2 + \boldsymbol{\varepsilon}$$

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

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1^2 + \varepsilon$$

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
x0	-0.004765	0.002854	-1.670	0.0955 .	
xx1	0.906527	0.057281	15.826	<2e-16 ***	
xx2	-0.215128	0.084965	-2.532	0.0116 *	
xx3	-0.173160	0.085054	-2.036	0.0422 *	
xx4	-0.143191	0.617314	**-0.232**	0.8167	=> Not significant

7

Functional Form: Linearity in Parameters

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

$$\boldsymbol{y} = \beta_1 + \beta_2 \boldsymbol{X}_2 + \beta_3\ \boldsymbol{X}_2^2 + \cdots + \beta_k\ \boldsymbol{X}_2^k + \boldsymbol{\varepsilon}$$

• 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.

7

Functional Form: Linearity in Parameters

• Nonlinear in parameters:

$$\boldsymbol{y} = \beta_1 + \beta_2 \boldsymbol{X}_2 + \beta_3 \boldsymbol{X}_3 + \beta_2 \beta_3 \boldsymbol{X}_4 + \boldsymbol{\varepsilon}$$

This model is nonlinear in parameters since the coefficient of $\boldsymbol{X}_4$ is the product of the coefficients of $\boldsymbol{X}_2$ and $\boldsymbol{X}_3$.

• Some nonlinearities in parameters can be linearized by appropriate transformations, but not this one. This in not an intrinsic linear model.

7

Functional Form: Linearity in Parameters

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

• Suppose we start with a power function: $\boldsymbol{y} = \beta_1 \boldsymbol{X}^{\beta_2} \boldsymbol{\varepsilon}$

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

$$\log \boldsymbol{y} = \log \beta_1 \boldsymbol{X}^{\beta_2} \varepsilon = \log \beta_1 + \beta_2 \log \boldsymbol{X} + \log \boldsymbol{\varepsilon},$$

Define:

$$\boldsymbol{y}' = \log \boldsymbol{y}$$
$$\boldsymbol{X}' = \log \boldsymbol{X}$$
$$\beta_1' = \log \beta_1$$
$$\boldsymbol{\varepsilon}' = \log \boldsymbol{\varepsilon}$$

Then, we have an intrinsic linear model: $\boldsymbol{y}' = \beta_1' + \beta_2 X' + \boldsymbol{\varepsilon}'$,

25

Functional Form: Linearity in Parameters

• Similar intrinsic linear model can be obtained if:

$$\boldsymbol{y} = \boldsymbol{e}^{\beta_1 + \beta_2 \boldsymbol{X} + \boldsymbol{\varepsilon}}$$

$$\Rightarrow \qquad \log \boldsymbol{y} = \beta_1 + \beta_2 \boldsymbol{X} + \boldsymbol{\varepsilon}$$

Define: $\boldsymbol{y}' = \log \boldsymbol{y}$

Then, we have an intrinsic linear model:

$$\boldsymbol{y}' = \beta_1 + \beta_2 \boldsymbol{X} + \boldsymbol{\varepsilon}$$

• But, we cannot linearize all model. For example,

$$\boldsymbol{y} = \beta_1 + \beta_2 \boldsymbol{X}^{\beta_3} + \boldsymbol{\varepsilon}$$

We will have to use some nonlinear estimation technique.

25

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:

$$E[y_i|\boldsymbol{X}] = \beta_{00} + \beta_{01}\, x_i \qquad \text{if } x_i \leq t_0$$

$$E[y_i|\boldsymbol{X}] = \gamma_0 + \gamma_1\, x_i \qquad \text{if } x_i > t_0$$

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

$$E[y_i|\boldsymbol{X}] = \beta_{00} + \beta_{01}\, x_i + \beta_{10}\,(x_i - t_0)_+^0 + \beta_{11}\,(x_i - t_0)_+^1$$

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:

$$E[y_i|\boldsymbol{X}] = \beta_{00} + \beta_{01}\, x_i + \beta_{10}\,(x_i - t_0)_+^0 + \beta_{11}\,(x_i - t_0)_+^1$$

That is,

$$E[y_i|\boldsymbol{X}] = \beta_{00} + \beta_{01}\, x_i \qquad \text{if } x_i \leq t_0$$

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

• We have a linear model:

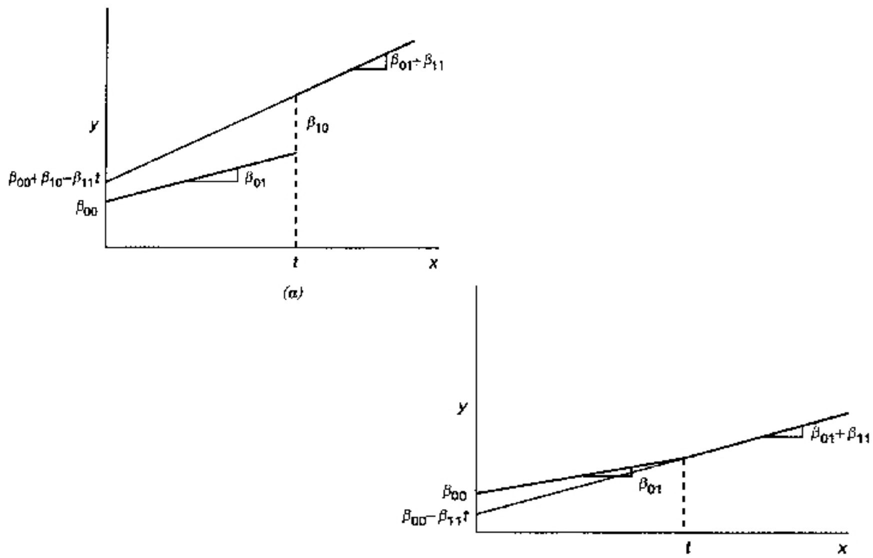
$$y_i = \beta_{00} + \beta_{01}\, x_i + \beta_{10}\,(x_i - t_0)_+^0 + \beta_{11}\,(x_i - t_0)_+^1 + \varepsilon_i$$

$\Rightarrow$ It can be estimated using OLS.

• If, in addition, we want the function to be continuous at the knot:

$$\beta_{00} + \beta_{01} t_0 = (\beta_{00} + \beta_{10} - \beta_{11}\, t_0) + (\beta_{01} + \beta_{11}) t_0 \;\Rightarrow \beta_{10} = 0$$

Functional Form: Linear Splines



Functional Form: Linear vs Log specifications

• Linear model: $\boldsymbol{y} = \beta_1 + \beta_2\, \boldsymbol{X} + \boldsymbol{\varepsilon}$

• (Semi-) Log model: $\log \boldsymbol{y} = \beta_1 + \beta_2\, \boldsymbol{X} + \boldsymbol{\varepsilon}$

• Box–Cox transformation: $\dfrac{\boldsymbol{y}^{\lambda} - \mathbf{1}}{\boldsymbol{\lambda}} = \beta_1 + \beta_2\, \boldsymbol{X} + \boldsymbol{\varepsilon}$

$$\frac{\boldsymbol{y}^{\lambda} - 1}{\lambda} = \boldsymbol{y} - 1 \qquad \text{when} \quad \lambda = 1$$

$$\frac{\boldsymbol{y}^{\lambda} - 1}{\lambda} = log(\boldsymbol{y}) \qquad \text{when} \quad \lambda \to 0$$

• Putting $\lambda = 0$ gives the (semi–)logarithmic model (think about the limit of λ tends to zero.). We can estimate λ. One would like to test if λ is equal to 0 or 1. It is possible that it is neither!

10

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 from our (**A1**) model:

$\hat{\boldsymbol{y}} = \boldsymbol{X}\mathbf{b}$. (for example, $\hat{\boldsymbol{y}} = b_1 X_1 + b_2 X_2$)

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

$\boldsymbol{y} = \boldsymbol{X}\,\beta + \hat{\boldsymbol{y}}^2\ \gamma + \boldsymbol{\varepsilon}$ ($\hat{\boldsymbol{y}}^2 = (b_1 X_1)^2 + (b_2 X_2)^2 + 2 b_1 b_2 X_2 X_1$)

• If $\hat{\boldsymbol{y}}^2$ 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 $\boldsymbol{X}$ variables.

• We test H_0 (linear functional form): $\boldsymbol{\gamma} = 0$

H_1 (non linear functional form): $\boldsymbol{\gamma} \neq 0$

3

Functional Form: Ramsey's RESET Test

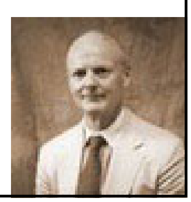
• We test H_0 (linear functional form): $\boldsymbol{\gamma} = 0$

H_1 (non linear functional form): $\boldsymbol{\gamma} \neq 0$

$\Rightarrow$ **t-test** on the OLS estimator of $\boldsymbol{\gamma}$.

• If the *t-statistic* for $\hat{\boldsymbol{y}}^2$ 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).



James B. Ramsey, England

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.

```
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)
```

```
> 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**	⇒ Not significant!

3

Qualitative Variables and Functional Form

• We want to model CEO compensation as a function of education. We have data on annual total CEO compensation ($Comp$), annual returns, annual sales, CEO's age, and CEO's **last degree** (education). We have qualitative data.

• We can estimate CEO compensation regressions for each last degree – i.e., BA/BS; MS/MA/MBA; Doctoral. We have three regressions:

Undergrad degree $\quad Comp_i = \beta_{0\text{-}u} + \boldsymbol{\beta}_{1\text{-}u}{}'\, \boldsymbol{z}_i + \varepsilon_{u,i}$

Masters degree $\quad Comp_i = \beta_{0\text{-}m} + \boldsymbol{\beta}_{1\text{-}m}{}'\, \boldsymbol{z}_i + \varepsilon_{m,i}$

Doctoral degree $\quad Comp_i = \beta_{0\text{-}d} + \boldsymbol{\beta}_{1\text{-}d}{}'\, \boldsymbol{z}_i + \varepsilon_{d,i}$

where the $\boldsymbol{z}_i$ is a vector of the CEO i's age and previous experience, and his/her firm's *annual* returns and annual sales.

18

Potential problem: We have 3 small samples –i.e, lose power & precision.

Qualitative Variables and Functional Form

• Alternatively, we can combine the 3 regressions in one, using the whole sample. We use a *dummy variable* (*indicator variable*) that points whether an observation belongs to a category or class or not. For example:

$D_{C,i}$ = 1 if observation i belongs to category C (say, male.)
= 0 otherwise.

• For CEO's education, we define two dummy variables:

$D_{m,i}$ = 1 if CEO i's has at least a Masters degree
= 0 otherwise.
$D_{d,i}$ = 1 if CEO i's has a Doctoral degree
= 0 otherwise.

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

$$Comp_i = \beta_0 + \boldsymbol{\beta}_1'\mathbf{z}_i + \beta_2 D_{m,i} + \beta_3 D_{d,i} + \boldsymbol{\gamma}_1'\mathbf{z}_i D_{m,i} + \boldsymbol{\gamma}_2'\mathbf{z}_i D_{d,i} + \varepsilon_i$$

19

Qualitative Variables and Functional Form

Our CEO Compensation model becomes:

$$Comp_i = \beta_0 + \boldsymbol{\beta}_1'\mathbf{z}_i + \beta_2 D_{m,i} + \beta_3 D_{d,i} + \boldsymbol{\gamma}_1'\mathbf{z}_i D_{m,i} + \boldsymbol{\gamma}_2'\mathbf{z}_i D_{d,i} + \varepsilon_i$$

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

- Model for undergrads only ($D_{m,i} = 0$ & $D_{d,i} = 0$):

$$Comp_i = \beta_0 + \boldsymbol{\beta}_1'\mathbf{z}_i + \varepsilon_i$$

- Model for Masters degree only ($D_{m,i} = 1$ & $D_{d,i} = 0$):

$$Comp_i = (\beta_0 + \beta_2) + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma}_1)'\mathbf{z}_i + \varepsilon_i$$

- Model for Doctoral degree only ($D_{m,i} = 1$ & $D_{d,i} = 1$):

$$Comp_i = (\beta_0 + \beta_2 + \beta_3) + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2)'\mathbf{z}_i + \varepsilon_i$$

20

Qualitative Variables and Functional Form

- Three models, encompassed by one regression:

$Comp_i = \beta_0 + \boldsymbol{\beta}_1'\mathbf{z}_i + \varepsilon_i$ Undergrad degree

$Comp_i = (\beta_0 + \beta_2) + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma}_1)'\mathbf{z}_i + \varepsilon_i$ Masters degree

$Comp_i = (\beta_0 + \beta_2 + \beta_3) + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2)'\mathbf{z}_i + \varepsilon_i$ Doctoral degree

- The parameters for the different categories are:

- Constant:

Constant for undergrad degree: β_0
Constant for Masters degree: $\beta_0 + \beta_2$
Constant for Doctoral degree: $\beta_0 + \beta_2 + \beta_3$

- Slopes:

Slopes for Masters degree: $\boldsymbol{\beta}_1 + \boldsymbol{\gamma}_1$
Slopes for Doctoral degree: $\boldsymbol{\beta}_1 + \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2$

21

Qualitative Variables and Functional Form

- We can test the effect of education on CEO compensation:
  (1) H_0: No effect of grad degree: $\beta_3 = \beta_2 = 0$ & $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \mathbf{0} \Rightarrow$ *F-test.*
  (2) H_0: No effect of Masters degree on constant: $\beta_2 = 0 \Rightarrow$ *t-test.*
  (3) H_0: No effect of doctoral degree: $\beta_3 = 0$ & $\boldsymbol{\gamma}_2 = \mathbf{0} \Rightarrow$ *F-test.*
  (4) H_0: No effect of Dr degree on marginal effect: $\boldsymbol{\gamma}_2 = \mathbf{0} \Rightarrow$ *t-test.*

- We may have more than one qualitative category (last degree above) in our data that we may want to introduce in our model.

Example: Suppose we also have data for CEO graduate school. Now, we can create another qualitative category, "quality of school", defined as Top 20 school, to test if a Top 20 school provides "more value." To do this, we use D_{TOP} to define if any schooling is in the Top 20.

$D_{TOP,i} = 1$ if CEO i's school is a Top 20 school
$\phantom{D_{TOP,i}} = 0$ otherwise.

22

Qualitative Variables and Functional Form

Example (continuation):
The model becomes:

$$Comp_i = \beta_0 + \boldsymbol{\beta}_1'\mathbf{z}_i + \beta_2\, D_{m,i} + \beta_3\, D_{d,i} + \beta_4\, D_{TOP,i} + \boldsymbol{\gamma}_1'\mathbf{z}_i\, D_{m,i} + + \gamma_2'\mathbf{z}_i\, D_{d,i} + \gamma_3'\mathbf{z}_i\, D_{TOP,i} + \varepsilon_i$$

In this setting, we can test the effect of a Top20 education on CEO compensation:

(1) H_0: No effect of Top20 degree: $\beta_4 = 0$ and $\boldsymbol{\gamma}_3 = \mathbf{0}$ $\Rightarrow$ *F-test.*

• The omitted category is the *reference* or *control category*. In our first example, with only educational degrees, the reference category is undergraduate degree. In the second example, with educational degrees and quality of school (Top20 dummy), the reference category is undergraduate degree with no Top 20 education.

23

Qualitative Variables and Functional Form

• *Dummy trap*.
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 variables equals the number of categories, you will create perfect multicollinearity –i.e., you fell on the **dummy trap**.

24

Dummy Variables as Seasonal Factors

• A popular use of dummy variables is in estimating seasonal effects. We may be interested in studying the January effect in stock returns or if the returns of oil companies (say, Exxon or BP) are affected by the seasons, since in the winter people drive less and in the summer more.

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

$D_{Sum,i}$ = 1 if observation i occurs in Summer
= 0 otherwise.
$D_{Fall,i}$ = 1 if observation i occurs in Fall
= 0 otherwise.
$D_{Win,i}$ = 1 if observation i occurs in Winter
= 0 otherwise.

Then, letting **Z** be the vector of the three FF factors, we have:

$$(r_i - r_f) = \beta_0 + \boldsymbol{\beta}_1'\mathbf{z}_i + \beta_2\, D_{Sum,i} + \beta_3\, D_{Fall,i} + \beta_4\, D_{Win,i} + \varepsilon_i$$

25

Dummy Variables as Seasonal Factors

Example: In the context of the 3-factor FF model, we test if Exxon's excess returns (XOM) are affected by seasonal (quarters) factors:

$$(r_{XOM,i} - r_f) = \beta_0 + \boldsymbol{\beta}_1'\mathbf{z}_i + \beta_2\, D_{Sum,i} + \beta_3\, D_{Fall,i} + \beta_4\, D_{Win,i} + \varepsilon_i$$

```
x_xom <- SFX_da$XOM                                   # Extract XOM prices
T <- length(x_xom)
lr_xom <- log(x_xom[-1]/x_xom[-T])
xom_x <- lr_xom - RF

T <- length(xom_x)
Summ <- rep(c(0,0,0,0,0,0,1,1,1,0,0,0), round(T/12))  # Create Summer dummy
Fall <- rep(c(0,0,0,0,0,0,0,0,0,1,1,1), round(T/12))  # Create Fall dummy
Wint <- rep(c(1,1,1,0,0,0,0,0,0,0,0,0), round(T/12))  # Create Winter dummy
T1 <- T+1
Fall_1 <- Fall[2:T1]                                  # Adjust sample (starts in Feb)
Wint_1 <- Wint[2:T1]
Summ_1 <- Summ[2:T1]

fit_xom_s <- lm(xom_x ~ Mkt_RF + SMB + HML + Fall_1 + Wint_1 + Summ_1)
```

26

Dummy Variables as Seasonal Factors

Example (continuation):

```
fit_xom_s <- lm(xom_x ~ Mkt_RF + SMB + HML + Fall_1 + Wint_1 + Summ_1)
> summary(fit_xom_s)

Coefficients:
             Estimate   Std. Error t value Pr(>|t|)
(Intercept) 0.002445   0.003485  0.702   0.4832       ⇒ constant for reference category (Spring) ≈ 0.
Mkt_RF      0.761816   0.040602  18.763  < 2e-16 ***
SMB        -0.261925   0.060575  -4.324 1.81e-05 ***
HML         0.370623   0.060049   6.172 1.29e-09 ***
Fall_1     -0.006609   0.004947  -1.336   0.1822
Wint_1     -0.011283   0.004928  -2.290   0.0224 *    ⇒ significant. Reject H0: No Winter effect.
Summ_1     -0.007100   0.004944  -1.436   0.1515
```

Interpretation: In the Winter quarter, Exxon excess returns decrease, relative to the Spring, by 1.13%. But since Spring's (& Fall's & Summer's) effect is non-significant, the decrease is in absolute terms.

27

Dummy Variables as Seasonal Factors

Example (continuation): We can test if all quarters *jointly* matter. That is, H_0: $\beta_2 = \beta_3 = \beta_4 = 0$.

We do an F-test:

```
fit_u <- lm(xom_x ~ Mkt_RF + SMB + HML + Fall_1 + Wint_1 + Summ_1)
fit_r <- lm(xom_x ~ Mkt_RF + SMB + HML)

resid_u <- fit_u$residuals
RSS_u <- sum((resid_u)^2)

resid_r <- fit_r$residuals
RSS_r <- sum((resid_r)^2)

f_test <- ((RSS_r - RSS_u)/2)/(RSS_u/(T-4))
> f_test
[1] 2.706574
>
p_val <- 1 - pf(f_test,df1=3, df2=T-3)              # p-value of F-test
>  p_val
[1] 0.05504357
```

Conclusion: p-value is "marginal." At 5% level, cannot reject H_0: No joint seas effect. 28

Dummy Variables: Is There a January Effect?

Example: We want to test the January effect on IBM stock returns, where because of tax reasons/window dressing, stocks go down in December and recover in January. The test can be done by adding a dummy variable to the 3-factor FF model:

$D_{J,t} = 1$ if observation t occurs in January
$= 0$ otherwise.

Then, we estimate the expanded model:

$$(r_{i,t} - r_f) = \beta_0 + \beta_1 (r_{m,t} - r_f) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 D_{J,t} + \varepsilon_{i,t}$$

We test H_0(No January effect): $\beta_4 = 0$ $\Rightarrow$ *t-test.*

We create a January dummy:

```
T <- length(ibm_x)
Jan <- rep(c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0), (round(T)/12+1))        # Create January dummy
T2 <- T+1
```

29

Dummy Variables: Is There a January Effect?

Example (continuation):

```
>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)

Coefficients:
             Estimate   Std. Error t value Pr(>|t|)
(Intercept) -0.007302  0.002561  -2.851  0.00452 **
Mkt_RF       0.905182  0.056405  16.048  < 2e-16 ***
SMB         -0.247691  0.084063  -2.946  0.00335 **
HML         -0.154093  0.083606  -1.843  0.06584 .
Jan_1        0.026966  0.008906   3.028  0.00258 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.058 on 565 degrees of freedom
Multiple R-squared:  0.3499,    Adjusted R-squared:  0.3453
F-statistic: 76.01 on 4 and 565 DF,  p-value: < 2.2e-16
```

3

Dummy Variables: Is There a January Effect?

Example (continuation):
Interpretation: We have two constants (excess return, Jensen's alpha):
Feb - Dec: **-0.7302%** (significant).
January: **-0.7302%** + **2.6966%** = **1.9664%** (significant).

When the January dummy was not in the model, we had: **-0.005191**, which is close to an average of the constants (= **-0.007302** *11 + **0.019664**)/12 = -0.00505).

Interpretation: During January IBM has an additional **2.6966%** excess returns. This is a big number. Today, the evidence for the January effect is much weaker than in this case.

• Note that in the FF model we expect the constant to be very small (≈0). In this case, it is not zero. Maybe we have a misspecified (**A1**).

31

Dummy Variable for One Observation

• We can use a dummy variable to isolate a single observation.

$D_j \quad = 1 \quad$ for observation j.
$\quad\;\; = 0 \quad$ otherwise.

• Define **d** to be the dummy variable in question.

$\boldsymbol{Z}$ = all other regressors. $\boldsymbol{X} = [\boldsymbol{Z}, \boldsymbol{D}_j]$

• Multiple regression of $\boldsymbol{y}$ on $\boldsymbol{X}$. We know that

$\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$ where $\boldsymbol{e}$ = the column vector of residuals.

$\Rightarrow D_j'\boldsymbol{e} = 0 \quad \Rightarrow e_j = 0$ (perfect fit for observation j).

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

Dummy Variable for One Observation

Example: In Dec 1992, IBM reported record losses and gave a very bleak picture of its future. The stock tumbled **-30.64%** that month. We check the effect of that extreme observation, a potential outlier, on the 3-factor FF model + January dummy:

```
dec_1992 <- rep(0,T)                                  # Define Dec 1992 dummy
dec_1992[239] <- 1                                    # Define Dec 1992 dummy (=1 if Dec 1992)
fit_d92 <- lm (ibm_x ~ Mkt_RF + SMB + HML + Jan_1 + dec_1992)
> summary(fit_d92)

Coefficients:
             Estimate   Std. Error t value Pr(>|t|)
(Intercept) -0.006772   0.002502  -2.707  0.00699 **
Mkt_RF       0.908775   0.055054  16.507  < 2e-16 ***
SMB         -0.239213   0.082059  -2.915  0.00370 **
HML         -0.138629   0.081647  -1.698  0.09008 .
Jan_1        0.026163   0.008694   3.009  0.00273 **
dec_1992    -0.306202   0.056710  -5.399 9.86e-08 ***          (same value of observation)
```

Note: Potential "Outlier" has no major effect on coefficients. 33

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 & 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):

$$F = \frac{(RSS_R - RSS_U)/(k_U - k_R)}{(RSS_U)/(T - k_U)} = \frac{(RSS_R - [RSS_1 + RSS_2])/k}{(RSS_1 + RSS_2)/(T - 2k)}$$

32

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.)

Gregory C. Chow (1929, USA)

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

DOCVIS = number of visits to the doctor in the observation period

HHNINC = household nominal monthly net income in German marks / 10000.
(4 observations with income=0 were dropped)
HHKIDS = children under age 16 in the household = 1; otherwise = 0
EDUC = years of schooling
AGE = age in years
MARRIED= marital status (1 = if married)
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

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=HHNINC   Mean                 =    .3590541     |
|              Standard deviation   =    .1735639     |
|              Number of observs.   =       14243     |
| Model size   Parameters           =           5     |
|              Degrees of freedom   =       14238     |
| Residuals    Sum of squares       =    379.8470     |
|              Standard error of e  =    .1633352     |
| Fit          R-squared            =    .1146423     |
|              Adjusted R-squared   =    .1143936     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|Constant|     .04169***       .00894        4.662   .0000           |
|AGE     |     .00086***       .00013        6.654   .0000    42.6528|
|EDUC    |     .02044***       .00058       35.528   .0000    11.7287|
|MARRIED |     .03825***       .00341       11.203   .0000     .76515|
|WHITEC  |     .03969***       .00305       13.002   .0000     .29994|
+--------+------------------------------------------------------------+
```

Chow Test: Males or Females visit doctors more?

- OLS Estimation for Women only. Keep RSS_W = 363.8789

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=HHNINC   Mean                 =    .3444951     |
|              Standard deviation   =    .1801790     |
|              Number of observs.   =       13083     |
| Model size   Parameters           =           5     |
|              Degrees of freedom   =       13078     |
| Residuals    Sum of squares       =    363.8789     |
|              Standard error of e  =    .1668045     |
| Fit          R-squared            =    .1432098     |
|              Adjusted R-squared   =    .1429477     |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|Constant|     .01191          .01158        1.029   .3036           |
|AGE     |     .00026*         .00014        1.875   .0608    44.4760|
|EDUC    |     .01941***       .00072       26.803   .0000    10.8764|
|MARRIED |     .12081***       .00343       35.227   .0000     .75151|
|WHITEC  |     .06445***       .00334       19.310   .0000     .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***      .919581D-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 = F = [(752.4767 – (379.847 + 363.8789))/5] /
[(379.847 + 363.8789)/(27,326 – 10)] = 64.281

$F(5, 27311) = 2.214100$ $\Rightarrow$ reject H_0

Wald Test (Greene)

```
--> Matrix   ; zero=init(k,k,0) ; Ik = Iden(k) $
--> Matrix   ; bwald = [bm/bf] $  Column vector
--> matrix   ; vwald = [Vm/zero,Vf] $
--> Matrix   ; Mik = -1*Ik ; R = [Ik,MIk]  ; q = init(k,1,0) $
--> Matrix   ; M = R*bwald - q
             ; VM = R*vwald*R'
      ; List ; Wald = m'<vm>m
             ; JF = k*ChowTest$
Matrix WALD     has  1 rows and  1 columns.
               1
        +-------------+
       1|  321.00313
        +-------------+
Matrix JF       has  1 rows and  1 columns.
               1
        +-------------+
       1|  321.40815
        +-------------+
```

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,

$$y_i = \beta_0^1 + \beta_1^1\, X_{1,i} + \beta_2^1\, X_{2,i} + \beta_3^1\, X_{3,i} + \varepsilon_i \qquad \text{for } i \leq T_{SB}$$
$$y_i = \beta_0^2 + \beta_1^2\, X_{1,i} + \beta_2^2\, X_{2,i} + \beta_3^2\, X_{3,i} + \varepsilon_i \qquad \text{for } i > T_{SB}$$

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:

$$y_i = \beta_0 + \beta_1\, X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i \qquad \text{for all } i$$

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

3

Functional Form: Structural Change

• 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.

3

Functional Form: Structural Change

- We structure the Chow test to test H_0 (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 RSS_R.

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

Before Date T_{SB}. Keep RSS_1.

After Date T_{SB}. Keep RSS_2 $\Rightarrow RSS_U = RSS_1 + RSS_2$.

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

$$F = \frac{(RSS_R - RSS_U)/(k_U - k_R)}{(RSS_U)/(T - k_U)} = \frac{(RSS_R - [RSS_1 + RSS_2])/k}{(RSS_1 + RSS_2)/(T - 2k)}$$

3

Functional Form: Structural Change

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

$$y_t = \beta_0 + \beta_1\, y_{t-1} + \varepsilon_t$$

```
GDP_da <- read.csv("http://www.bauer.uh.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_s <- 108                              # T_SB = 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]
t_s <- t_s -1                           # Adjust t_s (we lost the first observation)
```

44

Functional Form: Structural Change

Example (continuation):

```
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)  0.011406   0.001118    10.200    < 2e-16 ***
lr_gdp1      0.262234   0.055543    4.721     3.59e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01248 on 302 degrees of  freedom.
```

45

Functional Form: Structural Change

Example (continuation):

```
# 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] 4.391997
p_val <- 1 - pf(F, df1 = 2, df2 = T - 2*k)  # p-value of F_test
> p_val
[1] 0.0131817
```

$\Rightarrow$ small p-values: Reject H_0 (No *structural change*).

3 46

Functional Form: Structural Change

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

Q: Did the dot.com bubble (end of 2001) affect the structure of the FF Model? Sample: Jan 1973 – June 2020 (T = 569).

Pooled RSS = **1.9324**

Jan 1973 – Dec 2001 RSS = RSS_1 = **1.33068** (T = 342)

Jan 2002 – June 2020 RSS = RSS_2 = **0.57912** (T = 227)

$$F = \frac{[RSS_R - (RSS_1 + RSS_2)]/k}{(RSS_1 + RSS_2)/(T-k)} = \frac{[1.9324 - (1.3307 + 0.57911)]/4}{(1.3307 + 0.57911)/(569 - 2*4)} = \mathbf{1.6627}$$

$\Rightarrow$ Since $F_{4,565,.05}$ = **2.39**, we cannot reject H_0

	Constant	Mkt – rf	SMB	HML	RSS	T
1973-2020	-0.0051	0.9083	-0.2125	-0.1715	**1.9324**	569
1973-2001	-0.0038	0.8092	-0.2230	-0.1970	**1.3307**	342
2002 – 2020	-0.0073	1.0874	-0.1955	-0.3329	**0.5791**	227

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)

$$F = \frac{[RSS_R - (RSS_1 + RSS_2)]/k}{(RSS_1 + RSS_2)/(T-k)} = \frac{[1.5700 - (0.5456 + 0.9348)/4}{(0.5456 + 0.9348)/570 - 2*4)} = \mathbf{8.499996}$$

$\Rightarrow$ Since $F_{4,562,.05}$ = **2.39**, we reject H_0

Conclusion: At the 5% level, we have evidence for a Dot.com bubble structural change.

Functional Form: Structural Change

- Under the H_0 (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 t, for example:

$$y_t = \beta_0 + \beta_1\, y_{t-1} + \varepsilon_t$$

- If the Chow test rejects H_0, we need to reformulate the model. A typical reformulation includes a dummy variable ($D_{SB,t}$). For example, with vector $\boldsymbol{x}_t$ of explanatory variables:

$$y_t = \beta_0 + \boldsymbol{\beta}_1'\boldsymbol{x}_t + \beta_2\, D_{SB,t} + \boldsymbol{\gamma}_1'\boldsymbol{x}_t\, D_{SB,t} + \varepsilon_t$$

where

$$D_{SB,t} \quad = 1 \qquad \text{if observation } t \text{ occurred after } T_{SB}$$
$$\quad = 0 \qquad \text{otherwise.}$$

3

Structural Change: Specification with Dummies

Example: We are interested in modelling the effect of the Oct 1973 oil shock in GDP growth rates. We include a dummy variable in the model, say D_{73}:

$$\boldsymbol{D}_{73,t} = 1 \text{ if observation } t \text{ occurred after October 1973}$$
$$= 0 \text{ otherwise.}$$

Then, $\quad y_t = \beta_0 + \boldsymbol{\beta}_1'\boldsymbol{x}_t + \beta_2\, \boldsymbol{D}_{73,t} + \boldsymbol{\gamma}_1'\boldsymbol{x}_t\, \boldsymbol{D}_{73,t} + \varepsilon_t$

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

	Constant	Slopes:
Before oil shock ($D_{73} = 0$):	β_0	$\boldsymbol{\beta}_1$
After oil shock ($D_{73} = 1$) :	$\beta_0 + \beta_2$	$\boldsymbol{\beta}_1 + \boldsymbol{\gamma}_1$

- We estimate the above model and perform an F-test to test:

H_0 (No *structural change*): $\beta_2 = 0$ & $\boldsymbol{\gamma}_1 = 0$.

50

Structural Change: Specification with Dummies

Example: We add an Oct 1973 dummy in the AR(1) GDP model.

```
T1 <- T - t_s                        # Number of Observations after SB
D73_0 <- rep(0,t_s)                  # Dummy_t = 0 if t <= t_s
D73_1 <- rep(1,T1)                   # Dummy_t = 1 of t > t_s
D73 <- c(D73_0,D73_1)                # SB Dummy variable t_s <- 108
lr_gdp1_D73 <- lr_gdp1 * D73         # interactive dummy (effect on slope)
fit_ar1_d_2 <- lm(lr_gdp0 ~ lr_gdp1 + D73 + lr_gdp1_D73)
summary(fit_ar1_d_2)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)		
(Intercept)	0.009139	0.001939	4.712	3.75e-06	***	
lr_gdp1	0.457011	0.090716	5.038	8.15e-07	***	
D73	0.003499	0.002362	1.482	0.13947		⇒ no significant effect on constant
lr_gdp1_D73	-0.316005	0.114197	-2.767	0.00601	**	⇒ significant effect of oil shock on slope.

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

Conclusion: After the oil shock the slope significantly changed from 0.457011 to **0.141006** (= 0.457011 + (-0.316005)).

51

Chow Test: Structural Change in Constant

Example (continuation): Suppose you suspect the dot.com bubble only affected **GE**'s constant (create dummy) and decide to model it:

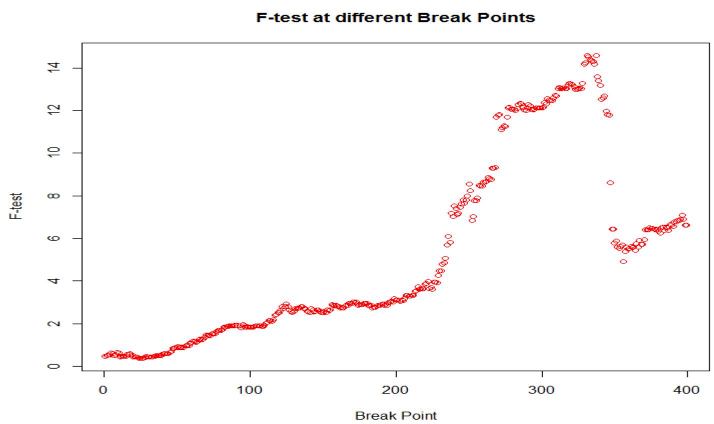
```
T <- length(ge_x)
x_break <- 342
dot_0 <- rep(0, x_break)             # 0 up to Dec 2001
dot_1 <- rep(1, T - x_break)         # 1 after Dec 2001
dot <- c(dot_0,dot_1)                # Doc.com dummy

fit_ge_dot <- lm(ge_x ~ Mkt_RF + SMB + HML + dot)
> summary(fit_ge_dot)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)		
(Intercept)	-0.003273	0.002877	-1.138	0.25566		
Mkt_RF	1.226412	0.050868	24.110	< 2e-16	***	
SMB	-0.308411	0.075433	-4.089	4.97e-05	***	
HML	0.341709	0.075755	4.511	7.86e-06	***	
dot	**-0.013052**	0.004502	**-2.899**	**0.00388**	**	⇒ significant effect on constant.

Chow Test: Structural Change - Example

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



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

Chow Test: Structural Change - Issues

• Issues with Chow tests
  - The results are *conditional* on the breaking point, T_{SB} –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.

3

Structural Change: Unknown Break

• For an unknown break date, Quandt (1958, 1960) proposed a likelihood ratio test statistics, called Supremum (Max)-Test,

$$QLR_T = \max_{\tau \varepsilon \{\tau_{\min},\ldots,\tau_{\max}\}} F_T(\tau)$$

The max (supremum) is taken over all potential breaks in (τ_{min}, τ_{max}). For example, τ_{min} = T*.15; τ_{max} = T*.85).

Easy to calculate QLR_T with a do loop.

The assumptions that make the LR-statistic asymptotically χ_J^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 τ is not identified under H_0 (no structural break) ⇒ regularity conditions are violated!

3

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*:

$$QLR_T \xrightarrow{d} \sup_{r \varepsilon [r_{\min}, r_{\max}]} \left(\frac{B_k(r)' B_k(r)}{r(1-r)} \right)$$

where 0< r_{min}< r_{max}<1 & $B_k(.)$ is a "*Brownian Bridge*" process defined on [0,1]. Percentiles of this distribution as functions of r_{max}, r_{min} & k are tabulated in Andrews (1993). (Critical values much larger than χ_J^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: $B(t) = W(t) - t\,W(1)$ is a Brownian Bridge.

3

Structural Change: Test with Unknown Break

Critical values of the QLR test Distribution, taken from Andrews (1993). Note: p = # of parameters (k), π_0= trimming value. (Ignore λ.)

840 DONALD W. K. ANDREWS

TABLE I
ASYMPTOTIC CRITICAL VALUES

		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.00	2.71	3.84	6.63	4.61	5.99	9.21	6.25	7.81	11.34	7.78	9.49	13.28	9.24	11.07	15.09
.49	1.08	3.47	4.73	7.82	5.42	6.86	10.30	7.19	8.83	12.58	8.93	10.63	14.64	10.39	12.28	16.34
.48	1.17	3.79	5.10	8.26	5.80	7.31	10.71	7.64	9.29	13.05	9.42	11.17	15.17	10.96	12.88	16.83
.47	1.27	4.02	5.38	8.65	6.12	7.67	11.01	7.98	9.62	13.39	9.82	11.63	15.91	11.40	13.27	17.32
.45	1.49	4.38	5.91	9.00	6.60	8.11	11.77	8.50	10.15	14.23	10.35	12.27	16.64	12.05	14.00	18.06
.40	2.25	5.10	6.57	9.82	7.45	9.02	12.91	9.46	11.17	14.88	11.39	13.32	17.66	13.09	15.16	19.23
.35	3.45	5.59	7.05	10.53	8.06	9.67	13.53	10.16	12.05	15.71	12.10	14.12	18.54	13.86	15.93	19.99
.30	5.44	6.05	7.51	10.91	8.57	10.19	14.16	10.76	12.58	16.24	12.80	14.79	19.10	14.58	16.48	20.67
.25	9.00	6.46	7.93	11.48	9.10	10.75	14.47	11.29	13.16	16.60	13.36	15.34	19.78	15.17	17.25	21.39
.20	16.00	6.80	8.45	11.69	9.59	11.26	15.09	11.80	13.69	17.28	13.82	15.84	20.24	15.63	17.88	21.90
.15	32.11	7.17	8.85	12.35	10.01	11.79	15.51	12.27	14.15	17.68	14.31	16.45	20.71	16.20	18.35	22.49
.10	81.00	7.63	9.31	12.69	10.50	12.27	16.04	12.81	14.62	18.28	14.94	16.98	21.04	16.87	18.93	23.34
.05	361.00	8.19	9.84	13.01	11.20	12.93	16.44	13.47	15.15	19.06	15.62	17.56	21.54	17.69	19.61	24.18

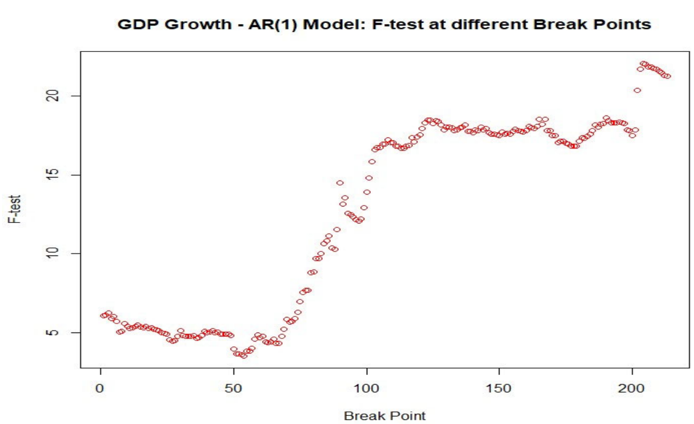
		$p=6$			$p=7$			$p=8$			$p=9$			$p=10$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.00	10.64	12.59	16.81	12.02	14.07	18.48	13.36	15.51	20.09	14.68	16.92	21.67	15.99	18.31	23.21
.49	1.08	11.81	13.74	18.32	13.27	15.52	19.93	13.29	15.63	20.53	16.17	18.56	23.05	17.35	19.79	24.62
.48	1.17	12.42	14.45	19.12	13.92	16.14	20.64	13.89	16.31	21.14	16.82	19.25	23.83	18.08	20.35	25.75
.47	1.27	12.90	14.86	19.64	14.32	16.63	21.14	14.43	16.74	21.72	17.26	19.74	24.80	18.67	20.92	26.43
.45	1.49	13.53	15.59	20.45	14.97	17.38	22.32	15.05	17.53	22.28	18.10	20.59	25.52	19.39	21.78	27.30
.40	2.25	14.71	16.91	21.60	16.23	18.41	23.35	16.26	18.73	23.63	19.56	22.12	26.86	20.74	23.15	28.86
.35	3.45	15.56	17.75	22.33	17.09	19.34	24.10	17.06	19.46	24.64	20.49	22.93	27.77	21.87	24.17	29.76
.30	5.44	16.32	18.46	23.06	17.74	20.01	24.86	17.90	20.36	25.64	21.27	23.65	28.50	22.73	25.05	30.74
.25	9.00	17.00	19.07	23.65	18.38	20.63	25.11	18.61	20.95	26.10	21.93	24.31	29.23	23.32	25.80	31.32
.20	16.00	17.56	19.64	24.27	19.04	21.07	25.72	19.17	21.47	26.76	22.54	24.91	29.92	24.00	26.42	31.98
.15	32.11	18.12	20.26	24.79	19.69	21.84	26.23	19.82	22.13	27.25	23.15	25.47	30.52	24.62	27.03	32.33
.10	81.00	18.78	20.82	25.21	20.32	22.51	26.91	20.45	22.87	27.69	23.77	26.16	31.15	25.39	27.87	32.95
.05	361.00	19.49	21.56	25.96	21.02	23.22	27.53	21.23	23.60	28.77	24.64	26.94	31.61	26.24	28.63	33.86

Critical value for test for k=2, π_0 = .15 and α = .05.

Critical value for test for k=4, π_0 = .15 and α = .05.

Structural Change: Unknown Break - Example

Example: We search for breaking points for **GDP growth** rate in AR(1) model. Below, we plot all F-tests starting at $T * 0.15$:



• Maximum F is **22.08** occurs at t = Jan 2009 (observation 250). Then, $\widehat{QLR}$ = **22.08** > **11.79** ⇒ Reject H_0 at 5% level & break is not Oct 73!.

Structural Change: Unknown Break - Example

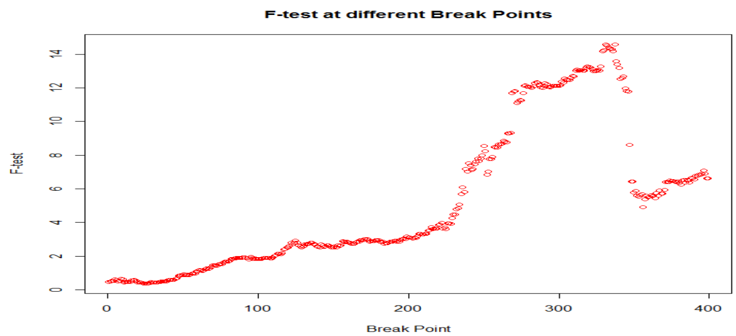
Example: 3 Factor Fama-French Model for **GE** excess returns

Andrews' (1993) test with τ_{min} = 50 (T * .15), τ_{max} = 286 (T * .85)

$\widehat{QLR}$ = **14.5936** at t = 433 (April 2008)

Critical value (k = 4, $\pi_1 = \tau_{min}/T$ = (1- τ_{max}/T) = .15, & α=.05) = **16.45**

⇒ cannot reject H_0



- Q: Multiple breaks?

Structural Change: Unknown Break - Example

```
b <- solve(t(x)%*% x)%*% t(x)%*%y                 # b = (X'X)-1 X' y  (OLS regression)
e <- y - x%*%b                                    # regression residuals, e
RSS_R <- as.numeric(t(e)%*%e)                     # RSS R

T1 <- round(T*.15)
T2 <- round(T*.85)
All_F <- matrix(0,T2-T1,1)

t <- T1
while (t <= T2) {
y_1 <- y[1:T1]
x_u1 <- x[1:T1,]
b_1 <- solve(t(x_u1)%*% x_u1)%*% t(x_u1)%*%y_1
e1 <- y_1 - x_u1%*%b_1
RSS1 <- as.numeric(t(e1)%*%e1)                              # RSS 1

kk = t+1
y_2 <- y[kk:T]
x_u2 <- x[kk:T,
b_2 <- solve(t(x_u2)%*% x_u2)%*% t(x_u2)%*%y_2
e2 <- y_2 - x_u2%*%b_2
RSS2 <- as.numeric(t(e2)%*%e2)                              # RSS 2

F <- ((RSS_R - (RSS1+RSS2)/k)/((RSS1+RSS2)/(T1-k))
All_F = rbind(All_F,F)
}
plot(All_F, col="red",ylab ="F-test", xlab ="Break Point")
title("F-test at different Break Points")
```

3

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 $\boldsymbol{x}^0 \Rightarrow$ predict $\boldsymbol{y}^0$.

Forecast: Given $\boldsymbol{x}_{t+1}^0 \Rightarrow$ predict $\mathbf{y}_{t+1}$.

Forecasting and Prediction

- Two types of predictions:

- In-sample (IS, prediction): The value of a future $\boldsymbol{y}$ (& $\boldsymbol{X}$) is observed by the sample. The expected value of $\boldsymbol{y}$ (in-sample), given the estimates of the parameters, is what we called fitted values.

- Out-of-sample (OOS, forecasting): The value of a future $\boldsymbol{y}$ that is not observed by the sample. The expected value of $\boldsymbol{y}$ (out-of-sample), given the estimates of the parameters, is what we called forecast value.

<u>Notation</u>:

- 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, $I_T = \{\boldsymbol{x}^0\}$ to predict $\boldsymbol{y}^0$.
Or in time series, $I_T = \{\boldsymbol{x}^0_{T-1}, \boldsymbol{x}^0_{T-2}, ..., \boldsymbol{x}^0_{T-q}\}$ to predict y_{T+l}.

• Then, the forecast is just the conditional expectation of Y_{T+l}, given the observed sample:

$$\hat{Y}_{T+l} = E[Y_{T+l} \,|\, X_T, X_{T-1}, \dots, X_1]$$

Example: If $X_T = Y_T$, then, the one-step ahead forecast is:

$$\hat{Y}_{T+1} = E[Y_{T+1} \,|\, Y_T, Y_{T-1}, \dots, Y_1]$$

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.

• Q: What is a good forecast? We need metrics to evaluate the forecasting performance of different models.

• In general, the evaluation of forecasts relies on MSE.

Forecasting and Prediction: Variance-bias

• We start with general model (DGP):
(**A1**) DGP: $\boldsymbol{y} = f(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$.

• Given $\boldsymbol{x}^0$, we predict y^0, using: $E[y|\mathbf{X}, \boldsymbol{x}^0] = f(\boldsymbol{X}, \boldsymbol{\theta})$
• We estimate $E[y|\mathbf{X}, \boldsymbol{x}^0]$ with $\hat{\boldsymbol{y}}^0 = f(\boldsymbol{x}^0, \hat{\boldsymbol{\theta}})$.

• The realization y^0 is just: $y^0 = f(\boldsymbol{x}^0, \boldsymbol{\theta}) + \varepsilon^0$

• With y^0 observed, we compute the prediction error: $\hat{y}^0 - y^0$ and its associated expected squared error, which can be written as:

$$E[(\hat{y}^0 - y^0)^2] = \text{Var}[\hat{y}^0] + [\text{Bias}(\hat{y}^0)]^2 + \text{Var}[\boldsymbol{\varepsilon}]$$

• We want to minimize this squared error. Note that there is nothing a forecaster can do regarding the last term, called the **irreducible error**.

Forecasting and Prediction: Variance-bias

• Since there is nothing to do regarding the *irreducible error*, all efforts are devoted to minimize the sum of a variance and a squared bias. This creates the **variance-bias trade-off** in forecasting.

• It is possible that a biased forecast can produce a lower MSE than an unbiased one. In this lecture, we based our forecasts on OLS estimates, which under the CLM assumptions, produce unbiased forecasts.

Note: The variance-bias trade-off is always present in forecasting. In general, more flexible models have less bias and more variance. The key is to pick an "optimal" mix of both.

Prediction Intervals: Point Estimate

- Prediction: Given $\boldsymbol{x}^0 \Rightarrow$ predict $\boldsymbol{y}^0$.

- Given the CLM, we have:

  Expectation: $\quad E[\boldsymbol{y}|\mathbf{X}, \boldsymbol{x}^0] = \beta' \boldsymbol{x}^0$;

  Predictor: $\quad \hat{\boldsymbol{y}}^0 = \mathbf{b}' \boldsymbol{x}^0$

  Realization: $\quad \boldsymbol{y}^0 = \beta' \boldsymbol{x}^0 + \boldsymbol{\varepsilon}^0$

Note: The predictor includes an estimate of $\boldsymbol{\varepsilon}^0$:

$\hat{\boldsymbol{y}}^0 = \mathbf{b}' \boldsymbol{x}^0$ + estimate of $\boldsymbol{\varepsilon}^0$. (Estimate of $\boldsymbol{\varepsilon}^0$=0, but with variance.)

- Associated with the prediction (a point estimate), there is a forecast error, e^0:

$$e^0 = \hat{\boldsymbol{y}}^0 - \boldsymbol{y}^0 = \mathbf{b}' \boldsymbol{x}^0 - \beta' \boldsymbol{x}^0 - \boldsymbol{\varepsilon}^0 = (\mathbf{b} - \beta)' \boldsymbol{x}^0 - \boldsymbol{\varepsilon}^0$$

$$\Rightarrow \text{Var}[(\hat{\boldsymbol{y}}^0 - \boldsymbol{y}^0)|\boldsymbol{x}^0] = E[(\hat{\boldsymbol{y}}^0 - \boldsymbol{y}^0)' (\hat{\boldsymbol{y}}^0 - \boldsymbol{y}^0)|\boldsymbol{x}^0]$$

$$\text{Var}[e^0|\boldsymbol{x}^0] = \boldsymbol{x}^{0\prime}\, \text{Var}[(\mathbf{b} - \beta)|\boldsymbol{x}^0]\, \boldsymbol{x}^0 + \sigma^2$$

Prediction Intervals: Point Estimate

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 $\boldsymbol{x}^0$ = [1.0000 -0.0189 -0.0142 -0.0027]

Then,

$\hat{y}^0$ = -0.005089 + 0.908299 * (-0.0189) -0.212460 * -0.0142 -
- 0.171500 * (-0.0027) = **-0.01877582**

Suppose we observe $\mathbf{y}^0$ = **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:

```
> x_0 <- rbind(1.0000, -0.0189, -0.0142, -0.0027)
> y_0 <- 0.1555214
> y_f0 <-  t(b)%*% x_00
> y_f0
          [,1]
[1,] -0.01877582
> ef_0 <- y_f0 - y_00
> ef_0
          [,1]
[1,] -0.1742973
```

Prediction Intervals: C.I.

- How do we estimate the uncertainty behind the forecast? Form a (1-α)% confidence interval, as usual:

$$[\hat{y}^0 \pm t_{T-k,1-\alpha/2} * \text{sqrt}(\text{Var}[e^0])]$$

Two cases:

(1) If $\boldsymbol{x}^0$ is given –i.e., constants. Then,

$$\text{Var}[\hat{y}^0 - y^0 \mid \boldsymbol{x} = \boldsymbol{x}^0] = \boldsymbol{x}^{0\prime}\,\text{Var}[\mathbf{b} \mid \boldsymbol{x}^0]\,\boldsymbol{x}^0 + \sigma^2$$

$\Rightarrow$ Form C.I. as usual.

Note: In OOS forecasting, $\boldsymbol{x}^0$ is unknown, it has to be estimated.

(2) If $\boldsymbol{x}^0$ 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 $\boldsymbol{x}^0$ is known, the variance of the forecast error is

$$\sigma^2 + \boldsymbol{x}^{0\prime}\,\mathrm{Var}[\mathbf{b}\mid\boldsymbol{x}^0]\,\boldsymbol{x}^0 = \sigma^2 + \sigma^2[\boldsymbol{x}^{0\prime}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}^0]$$

If the model contains a constant term, this is

$$\mathrm{Var}[e^0] = \sigma^2\left[1 + \frac{1}{N} + \sum_{j=1}^{K-1}\sum_{k=1}^{K-1}(x_j^0 - \bar{x}_j)(x_k^0 - \bar{x}_k)(\boldsymbol{Z}'\mathrm{M}^0\boldsymbol{Z})^{jk}\right]$$

(where $\boldsymbol{Z}$ is $\boldsymbol{X}$ without $\boldsymbol{x}_1 = \boldsymbol{\iota}$). In terms squares and cross products of deviations from means.

Note: Large σ^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:

$$[\hat{y}^0 \pm t_{T-k,1-\alpha/2} * \mathrm{sqrt}(\mathrm{Var}[e^0])]$$

- As $\boldsymbol{x}^0$ moves away from , the C.I increases, this is known as the "*butterfly effect*."

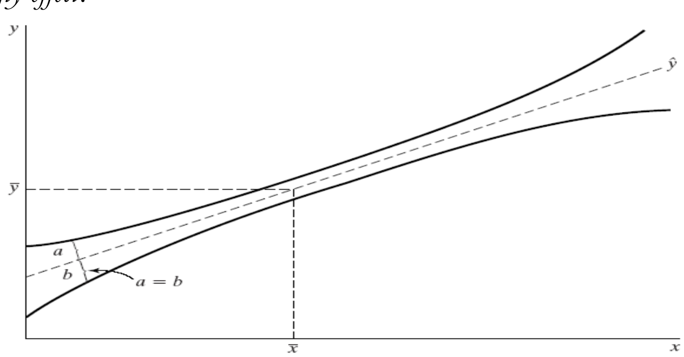


FIGURE 6.1 Prediction Intervals.

Prediction Intervals

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

Then,
Estimated Var[$\hat{y}^0 - y^0 \,|\, \boldsymbol{x}^0$] = $\boldsymbol{x}^{0\prime}$ Var[$\mathbf{b} \,|\, \boldsymbol{x}^0$] $\boldsymbol{x}^0 + s^2$ =
= **0.003429632**

```
> var_ef_0 <- t(x_0)%*% Var_b%*% x_0 + Sigma2
> var_ef_0
          [,1]
[1,] 0.003429632
> sqrt(var_ef_0)
         [,1]
[1,] 0.05856306
```

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

Prediction Intervals

Example (continuation):

```
># (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.

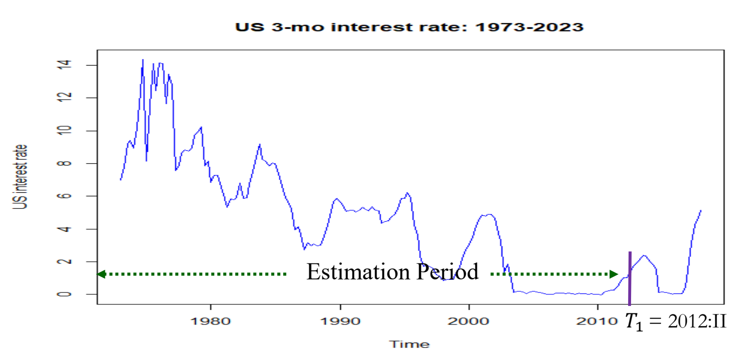
Forecasting and Prediction: Model Validation

• **Model validation** refers to establishing the statistical adequacy of the assumptions behind the model –i.e., (**A1**)-(**A5**) in this lecture. Predictive power or forecast accuracy can be used to do model validation.

• In the context of prediction and forecasting, model validation is done by fitting a model in-sample, but keeping a small part of the sample, the *hold-out-sample*, to check the accuracy of OOS forecasts.

• **Hold out sample**: We estimate the model using only a part of the sample (say, up to time T_1). The rest of the observations, the hold out sample, ($T - T_1$ observations) are used to check the predictive power of the model –i.e., the accuracy of predictions, by comparing $\hat{\boldsymbol{y}}^0$ with actual $\boldsymbol{y}^0$.

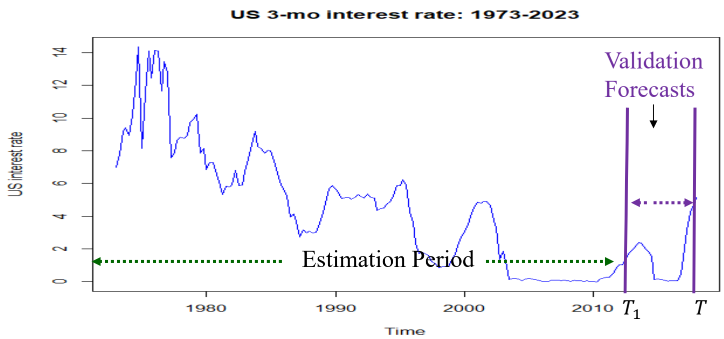
Review: Forecasting - Model Validation



• For model validation, we keep a small part of the sample for checking the forecasting skills (or accuracy) of the model. Steps:

Step 1. Estimate the model using all the observation up to T_1 (above from 1973:I to 2012:II). The period used is called "**estimation period** or **estimation sample**." (Get in-sample forecasts, $\hat{y}$.)

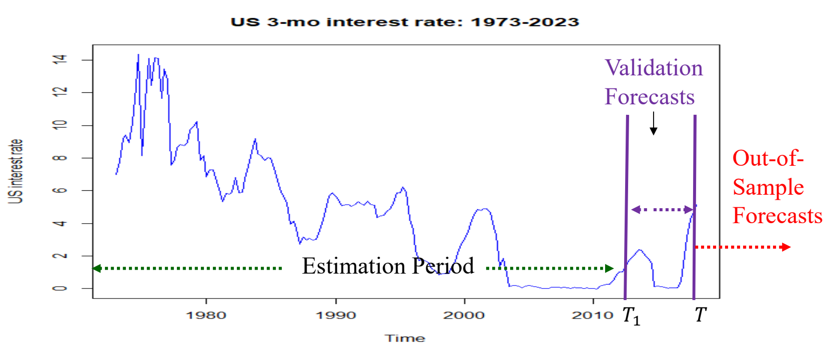
Review: Forecasting - Model Validation



Step 2. Keep a (short) part of the sample, $(T-T_1)$, to check the model's forecasting skills. Using the estimates from **Step 1**, we produce forecasts, $\hat{y}$, for the period $(T-T_1)$. Since in the period $(T-T_1)$ we know y, we can compute true MSE (MAE). This is the ***validation step***.

For example, we compute: $\text{MSE} = \frac{1}{(T-T_1)}\sum_{i=T_1+1}^{T}(\hat{y}_i - y_i)^2$

Review: Forecasting - Model Validation



Step 3. If happy with **Step 2**, we proceed to do true out-of-sample forecasts. In general, for the out-of-sample forecast, we re-estimate the model using all the sample –i.e., all T observations.

To evaluate the true OOS forecasts, we have to wait, say m periods, to compute an MSE : $\text{MSE} = \frac{1}{m}\sum_{i=T+1}^{m}(\hat{y}_i - y_i)^2$

Forecasting and Prediction: Model Validation

Details:

1) ***Estimation period***. Use the first T_1 observations to estimate the parameters of the model. Get in-sample forecasts, $\hat{y}$. In-sample evaluation of model (R^2, t- & F-tests) is usually performed here.

2) ***Validation period***. Use $(T - T_1)$ observations to check the model's forecasting skills. Given estimates in (1), OLS **b**, & using $\boldsymbol{x}^0$, get OSS $\hat{\boldsymbol{y}}^0 = \mathbf{b}'\, \boldsymbol{x}^0$. Since $\boldsymbol{y}^0$ is known, calculate true MSE or MAE. For example:

$$MSE = \frac{1}{(T-T_1)} \sum_{i=(T_1+1)}^{T} (\hat{y}_i^0 - y_i^0)^2$$

Note: It is common to set $(T - T_1)$ close to 10% of sample.

3) ***True OOS forecast period***. Re-estimate model Produce OSS $\hat{\boldsymbol{y}}^0$, but since $\boldsymbol{y}^0$ is not known now, it will take time to evaluate the true OOS forecasts.

Forecasting and Prediction: Model Validation

Note: In the **Machine Learning** literature, the terminology used for model validation is slightly different.

Step 1 is called "*training*," the data used (say, first T_1 observations) are called *training data/set*. In this step, we estimate the parameters of the model, subject to the assumptions, for example, (**A1**)-(**A4**).

Step 2 has the same name, the *validation step*. This step is used to "*tune (hyper-)parameters*." In our CLM, we can "tune" the model for departures of (**A1**)-(**A4**), for example, by including more variables (**A1**) and re-estimating the model accordingly using the "training data" alone. We choose the model with lower MSE or MAE

Remark: The idea of this step is to **simulate** out-of-sample accuracy. But, the "tuned" parameters selected in Step 2 are fed back to Step 1.

Step 3 *tests* the true out-of-sample forecast accuracy of model selected by **Step 1** & **Step 2**. This last part of the sample is called "*testing sample*."

Forecasting and Prediction: Cross Validation

• **Step 2** is used as a testing ground of the model before performing OOS forecasting. There are many ways to approach the validation step.

• Instead of a single split, split the data in K parts. This is called *K-fold cross-validation*. For j = 1, 2, …, K, use all folds but fold j to estimate model; use fold j to check model's forecasting skills by computing MSE, MSE_j. The K-fold CV estimate is an average of each fold MSE's:

$$CV_K = \frac{1}{K}\sum_{j=1}^{K} MSE_j$$

Usual choices for K are 5 & 10. (These are arbitrary choices.)

Random and non-random splits of data can be used. The non-random splits are used for some special cases, such as qualitative data, to make sure the splits are "representative."

Forecasting and Prediction: Cross Validation

• Use a single observation for validation. This is called **leave-one-out cross-validation** (**LOOCV**). A special case of **K-fold cross-validation** with $K = T$. That is, use $(T-1)$ observations for estimation, and, then, use the observation left out, i = 1, …, T, to compute $MSE_{(-i)}$, which is just $(\hat{y}_{(-i)} - y_i)^2$, where $\hat{y}_{(-i)}$ is the prediction for observation i based on the full sample but observation i. Then, compute:

$$CV_T = \frac{1}{T}\sum_{i=1}^{T} MSE_{(-i)}$$

• Instead of just one, it is possible to leave p observations for validation. This is called **leave-p-out cross-validation** (**LpOCV**).

Remark: In time series, since the order of the data matters, cross validation is more complicated. In general, rolling windows are used.

Forecasting and Prediction: Cross Validation

Example: We do cross-validation on the 5-factor Fama-French Model for IBM returns with $K = 5$:

```
y <- ibm_x
ff_cv_data <- data.frame(Mkt_RF, SMB, HML, RMW, CMA)
###### CV: Cross-Validation K-fold Code Function ######
CV<- function(dats, n.folds){
  folds <- list()                    # flexible object for storing folds
  fold.size <- nrow(dats)/n.folds
  remain <- 1:nrow(dats)             # all obs are in
  for (i in 1:n.folds){
    select <- sample(remain, fold.size, replace = FALSE)    #randomly sample fold_size from remaining obs)
    folds[[i]] <- select             # store indices ( write a special statement for last fold if 'leftover points')
      if (i == n.folds){
      folds[[i]] <- remain
    }
remain <- setdiff(remain, select)    #update remaining indices to reflect what was taken out
remain
  }
```

Forecasting and Prediction: Cross Validation

Example (continuation):

```
  results <- matrix(0,1,n.folds)
  for (i in 1:n.folds){
    # fold i
    indis <- folds[[i]]                                   # unpack into a vector
    estim <- dats[-indis, ]                               #split into estimation (train) & validation (test) sets
    test <- dats[indis, ]
    lm.model <- lm(y[-indis] ~ ., data = estim)           # OLS with estimation data
    pred <- predict(lm.model, newdata = test)             # predicted values for fold not used
    MSE <- mean((y[indis] - pred)^2)                      # MSE (any other evaluation measure can be used)
    results[[i]]<- MSE                                    # Accumulate MSE in vector
  }
  return(results)
}
CV_ff_5 <- CV(ff_step_data, 5)
> mean(CV_ff_5)
[1] 0.003532592
```

Evaluation of Forecasts: Measures and Tests

• We want to evaluate the forecast accuracy of a model:
- For individual (in-sample and out-of-sample) observations.
- For a group of (in-sample and out-of-sample) observations.

• Since squared loss functions are easy to work with, the traditional in-sample model evaluation has been based on MSE or R^2. For example,

$$\text{MSE} = \frac{1}{T}\sum_{i=1}^{T}(\hat{y}_i - y_i)^2$$

• Problem: In sample, models tend to overfit. The usual solution is to include penalties for model complexity, say, higher k. For example, use AIC or Adjusted R^2 to judge a model.

• Another solution is to use cross-validation.

Evaluation of Forecasts: Measures and Tests

• For OOS forecast, there are many measures, but it is common to adapt the traditional measures, MSE or MAE. For example, with m out of sample forecasts:

$$\text{MSE} = \frac{1}{m}\sum_{i=T+1}^{T+m}(\hat{y}_i - y_i)^2 = \frac{1}{m}\sum_{i=T+1}^{T+m} {e_i}^2$$

Note: Always keep in mind that all measures to evaluate forecasts are RV. We need a test to do any statistical comparison of measures.

Evaluation of Forecasts: Measures of Accuracy

• Popular measures of OOS forecast accuracy, after *m* forecasts:

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

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

Root Mean Square Error (RMSE) = $\sqrt{\frac{1}{m}\sum_{i=T+1}^{T+m} {e_i}^2}$

Mean Absolute Percentage Error (MAPE) = $\frac{1}{m}\sum_{i=T+1}^{T+m} |\frac{\hat{y}_i - y_i}{y_i} * 100|$

Theil's U statistics: $U = \dfrac{\sqrt{\frac{1}{m}\sum_{i=T+1}^{T+m} {e_i}^2}}{\sqrt{\frac{1}{T}\sum_{i=1}^{T} {y_i}^2}}$

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.

• An OOS R^2 can be computed as:

$$R^2_{OOS} = 1 - \frac{MSE_A}{MSE_N}$$

with $MSE_A = \sum_{t=1}^{Q} (y_{t+\tau} - \hat{y}_{t+\tau})^2$

$MSE_N = \sum_{t=1}^{Q} (y_{t+\tau} - \bar{y}_t)^2$

where τ is the forecasting horizon. (See Goyal and Welch (2008) for a well-known finance application.)

• We can also use cross-validation measures that use the whole (or almost all the) sample to evaluate forecasting performance.

Evaluation of forecasts: Measures of Accuracy

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.

```
> 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_RF0.865579   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.

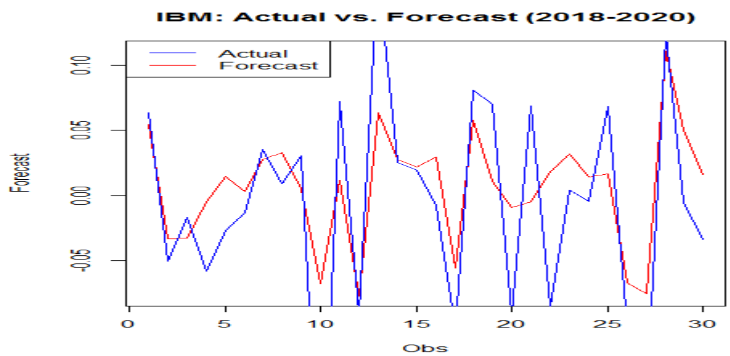
```
> x_0 <- x[T2:T,]
> y_0 <- y[T2:T]
> y_f0 <-  x_0%*% b1
> ef_0 <- y_f0 - y_0
> mes_ef_0 <- sum(ef_0^2)/nrow(x_0)
> mes_ef_0
[1] 0.003703207
> mae_ef_0 <- sum(abs(ef_0))/nrow(x_0)
> mae_ef_0
[1] 0.04518326
```

That is, MSE = **0.003703207**
MAE = **0.04518326**

Evaluation of forecasts: Measures of Accuracy

Example (continuation): Plot of actual IBM returns and forecasts.

```
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: Testing Accuracy

• We have measures of accuracy, which are RV, a function of the data. Given usual sampling variation, per se, measures are difficult to compare.

Q: We have two models, how do we know one forecast significantly better than the other? We need a test for this.

Evaluation of forecasts: Testing Accuracy

• Suppose two competing forecasting procedures produce a vector of errors: $e^{(1)}$ & $e^{(2)}$. Then, if expected MSE is the criterion used, the procedure with the lower MSE will be judged superior.

• We want to test H_0: MSE(1) = MSE(2)
H_1: MSE(1) ≠ MSE(2).

Assumptions: forecast errors are unbiased, normal, and uncorrelated. If forecasts are unbiased, then MSE = Variance.

• Consider, the pair of RVs: $(e^{(1)} + e^{(2)})$ & $(e^{(1)} - e^{(2)})$. Now,

$$E[(e^{(1)} + e^{(2)})(e^{(1)} - e^{(2)})] = \sigma_1^2 - \sigma_2^2$$

• That is, we test H_0 by testing that the two RVs are not correlated!
Under H_0, $E[(e^{(1)} + e^{(2)})(e^{(1)} - e^{(2)})] = 0.$

Evaluation of forecasts: Testing Accuracy

• Under H_0, $\left(e^{(1)} + e^{(2)}\right)$ & $(e^{(1)} - e^{(2)})$ are not correlated. This idea is due to Morgan, Granger and Newbold (MGN, 1977).

• There is a simpler way to do the MGN test. Steps:

1. Define $e^{(1)}$ & $e^{(2)}$, where $e^{(1)}$ is the error with the higher MSE. Let
$z_t = e^{(1)} + e^{(2)}$ – $e^{(1)}$: the error with the higher MSE.
$x_t = e^{(1)} - e^{(2)}$
2. Do a regression: $z_t = \beta\, x_t + \varepsilon_t$
3. Test H_0: β = 0 ⇒ a simple *t-test*.

The MGN test statistic is exactly the same as that for testing H_0: β = 0. This is the approach taken by Harvey, Leybourne and Newbold (1997).

• Non-parametric: Spearman's rank test for zero x_t & z_t correlation.

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:

$$(r_{i=IBM} - r_f)_t = \beta_0 + \beta_1 (r_m - r_f)_t + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_t$$

Taking expectations at time $t+1$, conditioning on time t information set, $I_t = \{(r_m - r_f)_t, \text{SMB}_t, \text{HML}_t\}$:

$$E[(r_{i=IBM} - r_f)_{t+1} | I_t] = \beta_0 + \beta_1 E[(r_m - r_f)_{t+1} | I_t] + \\ + \beta_2 E[SMB_{t+1} | I_t] + \beta_3 E[HML_{t+1} | I_t]$$

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

$$E[(r_{i=IBM} - r_f)_{t+1} | I_t] = \beta_0 + \beta_1 (r_m - r_f)_t + \beta_2 SMB_t + \beta_3 HML_t$$

Now, replacing the β by the estimated **b**, we have our one-step-ahead forecasts.

Evaluation of forecasts: Testing Accuracy

Example: We compare the forecast accuracy relative to a random walk model for IBM returns. That is,

$$E[(r_{i=IBM} - r_f)_{t+1} | I_t] = (r_{i=IBM} - r_f)_t$$

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

```
> 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 $z_t = e^{(1)} + e^{(2)}$, & $x_t = e^{(1)} - e^{(2)}$.
Then, regress: $z_t = \beta\, x_t + \varepsilon_t$ and test H_0: $\beta = 0$.

```
> 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 ⇒ MSE of RW is higher.

Evaluation of forecasts: Testing Accuracy – DM

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

• In practice, these tests are only applied to one-step predictions and the MSE is the loss function.

• Diebold and Mariano (DM, 1995) generalized the MGN approach to any loss function, g(.), and can be applied to forecast errors that are biased, non-normal and correlated.

• The test is based on the loss differential between two forecasts:

$$d_t = g(e^{(1)}) - g(e^{(2)})$$

• Then, we test the null hypotheses of equal predictive accuracy:

$$H_0: \mathrm{E}[d_t] = 0$$
$$H_1: \mathrm{E}[d_t] = \mu \neq 0.$$

Evaluation of forecasts: Testing Accuracy – DM

• Then, we test the null hypotheses of equal predictive accuracy:

H_0: $\mathrm{E}[d_t] = 0$

H_1: $\mathrm{E}[d_t] = \mu \neq 0$.

- Diebold and Mariano (1995) assume $\{e^{(1)}\}$ & $\{e^{(2)}\}$ is covariance stationarity and other regularity conditions (finite Var[d_t], independence of forecasts after ℓ periods) needed to apply CLT. Then,

$$\frac{\bar{d} - \mu}{\sqrt{Var[\bar{d}]/T}} \xrightarrow{d} N(0,1), \qquad \bar{d} = \frac{1}{m}\sum_{i=T+1}^{T+m} d_i$$

• Then, under H_0, the DM test is a simple *z-test*:

$$DM = \frac{\bar{d}}{\sqrt{\hat{V}ar[\bar{d}]/T}} \xrightarrow{d} N(0,1)$$

Evaluation of forecasts: Testing Accuracy – DM

where $\hat{V}ar[\bar{d}]$ is a consistent estimator of the variance, usually based on sample autocovariances of d_t:

$$\hat{V}ar[\bar{d}] = \gamma(0) + 2\sum_{j=k}^{\ell} \gamma(j)$$

• There are some suggestion to calculate small sample modification of the DM test. For example, :

$$\mathrm{DM}^* = \mathrm{DM}/\{[T + 1 - 2\,\ell + \ell\,(\ell - 1)/T]/T\}^{1/2} \sim t_{T-1}.$$

where ℓ-step ahead forecast. If ARCH is suspected, replace ℓ with $[0.5\,\sqrt{(T)}] + \ell$.

Note: If $\{e^{(1)}\}$ & $\{e^{(2)}\}$ are perfectly correlated, the numerator and denominator of the DM test are both converging to 0 as $T \to \infty$.

⇒ Avoid DM test when this situation is suspected (say, two nested models.) Though, in small samples, it is OK.

Evaluation of forecasts: Testing Accuracy – DM

Example: Code in R

```
 dm.test <- function (e1, e2, h = 1, power = 2) {
d <- c(abs(e1))^power - c(abs(e2))^power
  d.cov <- acf(d, na.action = na.omit, lag.max = h - 1, type = "covariance", plot = FALSE)$acf[, , 1]
  d.var <- sum(c(d.cov[1], 2 * d.cov[-1]))/length(d)
  dv <- d.var                                   #max(1e-8,d.var)
  if(dv > 0)
    STATISTIC <- mean(d, na.rm = TRUE) / sqrt(dv)
  else if(h==1)
    stop("Variance of DM statistic is zero")
  else
  {
    warning("Variance is negative, using horizon h=1")
    return(dm.test(e1,e2,alternative,h=1,power))
  }
    n <- length(d)
  k <- ((n + 1 - 2*h + (h/n) * (h-1))/n)^(1/2)
  STATISTIC <- STATISTIC * k
  names(STATISTIC) <- "DM"
}
```

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.

$$F = \frac{(RSS_R - RSS_1)/T_2}{RSS_1/(T_1 - k)} \sim F_{T_2, T_1 - k}$$

3

Out-of-sample predictions and prediction errors: Chow Test Revisited

Example: 3 Factor Fama-French Model for IBM (continuation)
We have $T = 336$ observations. We set $T_1 = 252$ & $T_2 = 86$. Then,
$RSS_{252} = 8.063611$.
$RSS_{336} = 12.92964$.

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

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

