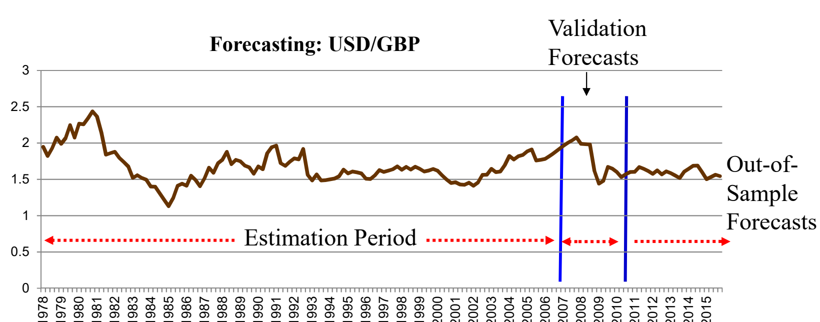
Lecture 6-d: Forecasting, Prediction and Model Selection

Brooks (4th edition): Chapter 5

Review: Forecasting – 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.

Review: Prediction Intervals – Point Estimate

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

- Given the CLM, we have:

  Expectation: $E[y|\mathbf{X}, \mathbf{x}^0] = \boldsymbol{\beta}'\mathbf{x}^0$;
  Predictor: $\hat{y}^0 = \mathbf{b}'\mathbf{x}^0$
  Realization: $y^0 = \boldsymbol{\beta}'\mathbf{x}^0 + \varepsilon^0$

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

- Associated with $\hat{y}^0$ (a point estimate), there is a forecast error, e^0:
$$e^0 = \hat{y}^0 - y^0 = \mathbf{b}'\mathbf{x}^0 - \boldsymbol{\beta}'\mathbf{x}^0 - \varepsilon^0 = (\mathbf{b} - \boldsymbol{\beta})'\mathbf{x}^0 - \varepsilon^0$$

and a variance
$$\Rightarrow \text{Var}[(\hat{y}^0 - y^0)|\mathbf{x}^0] = E[(\hat{y}^0 - y^0)'(\hat{y}^0 - y^0)|\mathbf{x}^0]$$
$$\text{Var}[e^0|\mathbf{x}^0] = \mathbf{x}^{0\prime}\,\text{Var}[(\mathbf{b} - \boldsymbol{\beta})|\mathbf{x}^0]\,\mathbf{x}^0 + \sigma^2$$

Review: Prediction Intervals – C.I. and Variance

- Assuming $\mathbf{x}^0$ is known, the variance of the forecast error is
$$\sigma^2 + \mathbf{x}^{0\prime}\,\text{Var}[\mathbf{b}|\mathbf{x}^0]\mathbf{x}^0 = \sigma^2 + \sigma^2[\mathbf{x}^{0\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^0]$$
If the model contains a constant term, this is

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

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

Note: Large σ^2, small N, and large deviations of driving variables from their means, decrease the precision of the forecasting error.

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

Review: Evaluation of Forecasts – MSE & MAE

- The most popular measures of out-of-sample forecast accuracy, after m forecasts are:

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

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

Review: Evaluation of forecasts – Testing MSEs

- Suppose two competing forecasting procedures produce a vector of errors: $e^{(1)}$ & $e^{(2)}$. We use the MSE to evaluate the models:

- 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, $E[(e^{(1)} + e^{(2)})(e^{(1)} - e^{(2)})] = 0.$

Review: Evaluation of forecasts – Testing MSEs

• Under H_0, $(e^{(1)} + e^{(2)})$ & $(e^{(1)} - e^{(2)})$ 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 error with higher MSE. Let
$z_t = e^{(1)} + e^{(2)}$ - $e^{(1)}$ is 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: $\beta = 0$ $\Rightarrow$ a simple *t-test*.

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

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

Forecasting Application: Fundamental Approach

• Based on how we select the "**driving**" variables X_t, we have different forecasting approaches:
- Fundamental (based on data considered fundamental, from theory)
- Technical analysis (based on data that incorporates only past prices)

• Fundamental Approach to Forecast Exchange Rates, S_t (USD/JPY)
Based on an economic model, we generate
$$E_t[S_{t+1}] = E_t[f(X_{t+1})] = g(X_t),$$
X_t: dataset with *fundamental* economic variables:
- GNP growth rate,
- Current Account,
- Interest rates,
- Inflation rates, etc.

Forecasting Application: Fundamental Approach

• The economic model usually incorporates:

- Statistical characteristics of data (seasonality, autocorrelation, etc.)
- Experience of the forecaster (what information to use, lags, etc.)

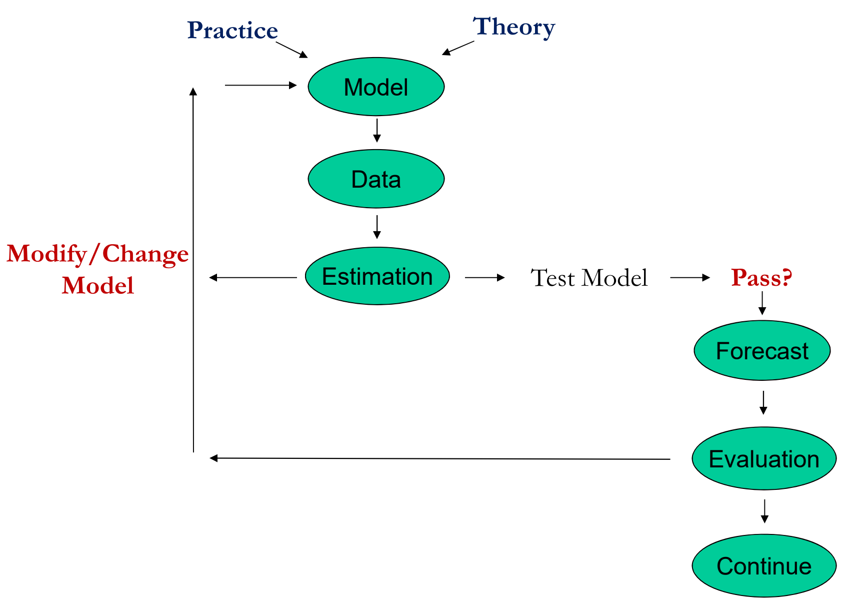
⇒ Mixture of art and science.

• The economic model provides the structure for the forecasts (also called *structural model*).

• We compare the economic model's performance with the performance of a simpler model, for example, the **Random Walk** (**RWM**). For many assets, the **RWM** is found to be a good forecasting model, especially for S_t, in the short-run. The RWM forecasts are:

$$E_t[S_{t+1}] = S_t$$

Forecasting Application: Fundamental Approach



Forecasting Application: Fundamental Approach

• **Fundamental Forecasting**: We want to forecast the FX rate S_t = USD/JPY. We model percentage changes in S_t:

$$e_{f,t} = \log(S_t) - \log(S_{t-1})$$

(1) Select a Model: Based on Theory (IFE, & Asset Approach)

$$e_{f,t} = \beta_0 + \beta_1 (i_{US,t} - i_{JAP,t}) + \beta_2 (y_{US,t} - y_{JAP,t}) + \varepsilon_t$$

$$E_t[e_{f,t+1}] = \beta_0 + \beta_1 E_t[(i_{US} - i_{JAP})]_{t+1} + \beta_2 E_t[(y_{US} - y_{JAP})]_{t+1}$$

$$\Rightarrow E_t[S_{t+1}] = S^F_{t+1} = S_t * (1 + E_t[e_{f,t+1}])$$

(2) Collect data: S_t, $\boldsymbol{X}_t$ (Interest rates, i_t, & GDP growth rates, y_t).

(3) Estimation of Model (using *estimation period*): OLS $\Rightarrow$ get **b**.

Forecasting Application: Fundamental Approach

• **Fundamental Forecasting** (continuation)

(4) Generate forecasts. Assumptions about $\boldsymbol{X}_t$ are needed.

$$E_t[\boldsymbol{X}_{t+1}] = \delta_1 + \delta_2 (\boldsymbol{X}_t) \qquad \text{-an AR(1) model.}$$

$$\Rightarrow \quad E_t[e_{f,t+1}] = E_t[\boldsymbol{X}_{t+1}]' \, \mathbf{b}$$

$$\Rightarrow \quad E_t[S_{t+1}] = S_t * (1 + E_t[e_{f,t+1}])$$

<u>Note</u>: We estimate δ_1 & δ_2 using OLS and, then, we used them to forecast $\boldsymbol{X}_{t+1}$.

(5) Evaluation of Forecasts: MSE (& compare with **RWM**'s MSE).

Model's Forecast Error$_{t+1}$ = $E_t[S_{t+1}]$ - S_{t+1}

RWM's Forecast Error$_{t+1}$ = S_t - S_{t+1}

Compute: $MSE_j = \frac{1}{m}\sum_{i=T+1}^{T+m} e_{j,i}{}^2$ (j = Our Model, RWM)

Forecasting Application: Fundamental Approach

Example: **(1)** & **(2)** Based on the following model,

$e_{f,t} = \beta_0 + \beta_1 (i_{US,t} - i_{JAP,t}) + \beta_2 (y_{US,t} - y_{JAP,t}) + \beta_3 (m_{US,t} - m_{JAP,t}) + \varepsilon_t$

we collect quarterly data (FX_USA_JAP.csv) from 1978:II – 2020:II. we read the data and transform it to estimate model:

```
FX_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/FX_USA_JAP.csv", head=TRUE, sep=",")
us_I <- FX_da$US_INF          # Read US Inflation (I_US) data from file
us_mg <- FX_da$US_M1_c        # Read US Money growth (m_US) data from file
us_i <- FX_da$US_I3M          # Read US 3-mo Interest rate (i_US) data from file
us_y <- FX_da$US_GDP_g        # Read US GDP growth (y_US) data from file
us_tb <- FX_da$US_CA_c        # Read US Current account change (tb_US) data from file
jp_I <- FX_da$JAP_INF         # Read Japan Inflation (I_US) data from file
jp_mg <- FX_da$JAP_MI_c       # Read Japan Money growth (m_JP) data from file
jp_i <- FX_da$JAP_I3M         # Read Japan 3-mo Interest rate (i_JP) data from file
jp_y <- FX_da$JAP_GDP_g       # Read Japan GDP growth y_JP) data from file
jp_tb <- FX_da$JAP_CA_c       # Read Japan Current account change (tb_JP) data from file
e_f <- FX_da$JPY.USD_c        # Read changes in JPY/USD (e)
```

Forecasting Application: Fundamental Approach

Example (continuation):

```
inf_dif <- us_I - jp_I        # Define inflation rate differential (inf_dif)
int_dif <- us_i - jp_i
mg_dif <- us_mg - jp_mg
y_dif <- us_y - jp_y
tb_dif <- us_tb - jp_tb
xx <- cbind(int_dif, mg_dif,  y_dif)
T <- length(e_f)
T_est <- 161                  # Define final observation for estimation period.
e_f1 <- e_f[1:T_est]          # Adjust sample size to T_est
xx_1 <- xx[1:T_est,]          # Adjust sample size to T_est
```

(3) Estimation of model: OLS

```
fit_ef <- lm(e_f1 ~ xx_1)
summary(fit_ef)
```

Forecasting Application: Fundamental Approach

Example (continuation):

(3) Estimation of model (using only *estimation period* (T=161): Get **b**.

```
> summary(fit_ef)
Call:
lm(formula = e_f1 ~ xx_1)

Coefficients:
               Estimate   Std. Error t value Pr(>|t|)
(Intercept)    1.7246     0.6971   2.474   0.0144 *
xx_1int_dif  -0.5281     0.2478  -2.131   0.0346 *
xx_1mg_dif   0.1104      0.1912   0.577   0.5647
xx_1y_dif    -0.2034     0.4538  -0.448   0.6546
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.293 on 157 degrees of freedom
Multiple R-squared:  0.04673,   Adjusted R-squared:  0.02851
F-statistic: 2.565 on 3 and 157 DF,  p-value: 0.05661
```

Forecasting Application: Fundamental Approach

Example (continuation): (4) Generate Forecasts. Need first to estimate model for **X** variables. (using *estimation period* data only)

- AR(1) for $(i_{US,t} - i_{JAP,t})$

```
int_dif_lag1 <- int_dif[1:T_est-1]              # Lag (i_US,t – i_JAP,t)
int_dif_lag0 <- int_dif[2:T_est]                # Adjust sample size (lost one observation above)
fit_int <- lm(int_dif_lag0 ~ int_dif_lag1)      # Fit AR(1) model
> summary(fit_int)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.22774   0.11074   2.057   0.0414 *
int_dif_lag1  0.87537   0.03772  23.210   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.045 on 158 degrees of freedom
Multiple R-squared:  0.7732,    Adjusted R-squared:  0.7718
F-statistic: 538.7 on 1 and 158 DF,  p-value: < 2.2e-16
```

Forecasting Application: Fundamental Approach

Example (continuation): (4 continuation)

- AR(1) for $(m_{US,t} - m_{JAP,t})$

```
mg_dif_lag1 <- mg_dif[1:T_est-1]            # Lag (m_US,t – m_JAP,t)
mg_dif_lag0 <- mg_dif[2:T_est]              # Adjust sample size (lost one observation above)
fit_mg <- lm(mg_dif_lag0 ~ mg_dif_lag1)     # Fit AR(1) model
> summary(fit_mg)

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.008708   0.216621  -0.040 0.967986
mg_dif_lag1  0.296597   0.076124   3.896 0.000144 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.74 on 158 degrees of freedom
Multiple R-squared:  0.08766,   Adjusted R-squared:  0.08188
F-statistic: 15.18 on 1 and 158 DF,  p-value: 0.000144
```

Forecasting Application: Fundamental Approach

Example (continuation): (4 continuation)

- AR(1) for $(y_{US,t} - y_{JAP,t})$

```
y_dif_lag1 <- y_dif[1:T_est-1]              # Lag (y_US,t – y_JAP,t)
y_dif_lag0 <- y_dif[2:T_est]                # Adjust sample size (lost one observation above)
fit_y <- lm(y_dif_lag0 ~ y_dif_lag1)        # Fit AR(1) model
> summary(fit_y)

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.166258   0.086575   1.920   0.0566 .
y_dif_lag1  -0.008828   0.077255  -0.114   0.9092
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.08 on 158 degrees of freedom
Multiple R-squared:  8.263e-05, Adjusted R-squared:  -0.006246
F-statistic: 0.01306 on 1 and 158 DF,  p-value: 0.9092
```

Forecasting Application: Fundamental Approach

Example (continuation): (4 continuation)

• Now, we can do *one-step-ahead* forecast for the **X** variables:

```
T_val <- T_est+1                                                    # start of Validation period
xx_cons <- rep(1,T-T_val+1)                                          # create the constant vector
int_dif_0 <- cbind(xx_cons,xx[T_val:T,1]) %*% fit_int$coeff          # 8 forecasts for (i_US,t – i_JAP,t)
mg_dif_0 <- cbind(xx_cons,xx[T_val:T,2]) %*% fit_mg$coeff            # 8 forecasts for (m_US,t – m_JAP,t)
y_dif_0 <- cbind(xx_cons,xx[T_val:T,3]) %*% fit_y$coeff              # 8 forecasts for (y_US,t – y_JAP,t)
```

• Finally, we compute the *one-step-ahead* forecast for **e** and MSE:

```
e_Mod_0 <- cbind(xx_cons,int_dif_0,mg_dif_0,y_dif_0)%*%fit_ef$coeff     # Model's forecast
f_e_Mod <- e_f[T_val:T] - e_Mod_0                                       # Model's forecast error
mse_e_f <- sum(f_e_Mod^2)/(T-T_val+1)                                   # Model's MSE
> mse_e_f
[1] 3.974203
```

Forecasting Application: Fundamental Approach

Example (continuation): (5) Evaluation of Forecasts

```
mse_e_f <- sum(f_e_Mod^2)/(T-T_val+1)                    # Model's MSE
> mse_e_f
[1] 3.974203
```

• Compute the *one-step-ahead* forecast for **RW Model** and, then, its MSE:

```
e_f_RW_0 <- rep(0,T-T_val+1)                             # RW forecast = 0 (always 0, for all t+T!)
f_e_RW <- e_f[T_val:T] - e_f_RW_0                        # RW's forecast error
mse_e_RW <- sum(f_e_RW^2)/(T-T_val+1)                    # RW's MSE
> mse_e_RW
[1] 3.381597                  ⇒ Lower MSE than Model. Not good for Model.
```

• Compare MSEs: The RW model has a better MSE (usual finding).

• A MGN test is usually done. But, we have only m=8 observations, we can do the test, but the results are very likely not to be taken seriously.

Forecasting Application: Fundamental Approach

Example (continuation): (5) Evaluation of Forecasts

• MGN/HLN test:

```
z_mgn <- e_Mod + e_RW
x_mgn <- e_Mod - e_RW
fit_mgn <- lm(z_mgn ~ x_mgn)
> summary(fit_mgn)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.355     2.680   0.506    0.631
x_mgn         1.798     2.759   0.651    0.539
```

⇒ not significant, but unreliable (a very small sample).

```
Residual standard error: 3.026 on 6 degrees of freedom
Multiple R-squared:  0.05322,   Adjusted R-squared:  -0.1046
F-statistic: 0.3373 on 1 and 6 DF,  p-value: 0.5826
```

⇒ very small # for df to make inferences.

• Suppose you are happy with the Model, you believe the difference in MSEs is not significant, now you generate out-of-sample forecasts.

Forecasting Application: Fundamental Approach

Example (continuation):

(6) Out-of-sample one-step-ahead forward forecast for S_t:

$$E_{t=2020:II}[S_{t+1=2020:III}] = S_{t=2020:II}\ (1 + E_{t=2020:II}[e_{f,t+1=2020:III}])$$

We observe S_t today (2020:II): $S_{2020:II}$ = 100.77 JPY/USD, which we invert since we work with direct quotes: $S_{2020:II}$ = 0.009279 USD/JPY.

We need to forecast the independent variables, based on AR(1) results,

$$\boldsymbol{X}_t = \{(i_{US,t} - i_{JAP,t}), (y_{US,t} - y_{JAP,t}), (m_{US,t} - m_{JAP,t})\ \}$$

• Forecasting $(i_{US,t+1} - i_{JAP,t+1})$: $E_{t=2020:II}[(i_{US} - i_{JAPUS})_{t+1=2020:III}]$

```
int_dif_p1 <- cbind(1,int_dif[T]) %*% fit_int$coeff       # int_dif_p1 = E_t=2020:II[(i_US,t – i_JAP)_t+1=2020:III]
> int_dif_p1
        [,1]
[1,] 0.4684645
```

Forecasting Application: Fundamental Approach

Example (continuation): (6) Out-of-sample forecast for S_t:

- Forecasting $(m_{US,t} - m_{JAP,t})$: $E_{t=2020:II}[(m_{US,t} - m_{JAP})_{t+1=2020:III}]$

```
mg_dif_p1 <- cbind(1,m_dif[T]) %*% fit_m$coeff        # mg_dif_p1 = E_{t=2020:II}[(m_US,t – m_JAP)_{t+1=2020:III}]
> mg_dif_p1
      [,1]
[1,] 4.921977
```

- Forecasting $(y_{US,t} - y_{JAP,t})$: $E_{t=2020:II}[(y_{US,t} - y_{JAP})_{t+1=2020:III}]$

```
y_dif_p1 <- cbind(1,y_dif[T]) %*% fit_y$coeff          # y_dif_p1 = E_{t=2020:II}[(y_US,t – y_JAP)_{t+1=2020:III}]
> y_dif_p1
      [,1]
[1,] 0.176617
```

- Forecasting $E_{t=2020:II}[S_{t+1=2020:III}]$

```
S <- 0.009279                                                # Today's value of S_{t=2020:II}
e_f_p1 <- cbind(1,int_dif_p1,mg_dif_p1,y_dif_p1) %*% fit_ef$coeff  # Today's forecast for e_{t=2020:III}
S_p1 <- S * (1+e_f_p1/100)                                   # Today's forecast for S_{t=2020:III}
```

Forecasting Application: Fundamental Approach

Example (continuation): (6) Out-of-sample forecast for S_t:

- Forecasting $E_{t=2020:II}[S_{t+1=2020:III}]$ (=S_p1 in the R script below)

```
> S <- 0.00927902                                             # Today's value of S_{t=2020:II}
> e_f_p1 <- cbind(1,inf_dif_p1,mg_dif_p1,y_dif_p1)%*%fit_ef$coeff  # Today's forecast for e_{t=2020:III}
> e_f_p1                                                      # Print forecast for e_{t=2020:III}
      [,1]
[1,] 1.984401                  ⇒ 1.98% depreciation of USD against JPY in 3rd Quarter.
> S_p1 <- S * (1+e_f_p1/100)        # e is in %, we divide by 100 to put it decimal from
> S_p1                                                        # Print forecast for S_{t=2020:III}
      [,1]
[1,] 0.009463133               ⇒ Model's forecast for S_{t+1=2020:III} = 0.009463133 USD/JPY.
```

- $E_{t=2020:II}[S_{t+1=2020:III}]$ = **0.009463133 USD/JPY**.

(using the indirect quote, $E_{t=2020:II}[S_{t+1=2020:III}]$ = 105.6732 JPY/USD).

Forecasting Application: Fundamental Approach

Example (continuation): (6) Out-of-sample forecast for S_t:

• We can use the one-step-ahead forecasts to generate *two-step-ahead* forecasts. That is, we forecast $E_{t=2020:II}[S_{t+1=2020:IV}]$ (=S_p2 below)

```
> S1 <- S_p1                                          # Today's forecast for S_t+1=2020:III
> int_dif_p2 <- cbind(1,int_dif_p1)%*%fit_int$coeff   # Today's forecast for (i_US – i_JP)_t+2
> mg_dif_p2 <- cbind(1,mg_dif_p1)%*%fit_mg$coeff      # Today's forecast for (m_US – m_JP)_t+2
> y_dif_p2 <- cbind(1,y_dif_p1)%*%fit_y$coeff         # Today's forecast for (y_US – y_JP)_t+2
> e_f_p2 <- cbind(1,int_dif_p2,mg_dif_p2,y_dif_p2)%*%fit_ef$coeff    # Today's forecast for e_t=2020:IV
> e_f_p2
     [,1]
[1,] 1.514363                  ⇒ 1.11% depreciation of USD against JPY in 4th Quarter.
> S_p2 <- S1*(1+e_f_p2/100)
> S_p2
        [,1]
[1,] 0.009606439               ⇒ Model's forecast for S_t+1=2020:III = 0.009606439USD/JPY.
```

• $E_{t=2020:II}[S_{t+1=2020:III}]$ = **0.009606439 USD/JPY.**

Forecasting Application: Fundamental Approach

Example (continuation): (6) Out-of-sample forecast for S_t:

• We can use the two-step-ahead forecast to generate *three-step-ahead* forecasts. Obviously, we can continue this process to generate *l-step-ahead* forecasts for S_t (a simple do loop will do it).

• Eventually, we will collect *m* of out-of-sample forecasts (*m* one-step-ahead forecasts, *m* two-step-ahead forecasts, *m* three-step-ahead forecasts, etc.) to get an MSE and run a MGN/HLN test on them.

• It is possible that one model is the best in the short-term (say, up to 3 steps ahead); other is better in the medium-term (say, from 4 to 6 steps ahead); and another is best for longer-term. For example, the RW model is very good ("*unbeatable*") up to 3 months ahead. Then, other models start to produce better forecasts, especially after 6 months.

Forecasting Application: Fundamental Approach

• Practical Issues in Fundamental Forecasting

- Are we using the "right model?"
- Estimation of the model (OLS, MLE, other methods).
- Some explanatory variables (X_{t+T}) are contemporaneous.

⇒ We also need a model to forecast the X_{t+T} variables.

• Does Forecasting Work?

For exchange rates, in the short-run (up to 6 months), **RW models** tend to do very well. They beat structural (and other) models: Lower MSE, MAE.

Many argue that the structural models used are not the "right models."

Model Selection Strategies

• Specifying the DGP in (**A1**) is the most important step in applied work. We have assumed "*correct specification*," which, in practice, is an unrealistic assumption, since we do not really observed the true DGP.

• A bad model can create many problems: biases, wrong inferences, bad forecasts, etc.

• So far, we have implicitly used a simple strategy:

(1) We started with a DGP, which we assumed to be true.

(2) We tested some H_0 (from economic theory).

(3) We used the model (restricted, if needed) for prediction & forecasting.

Model Selection Strategies

- Q: How do we propose and select a model (a DGP)?

- Potentially, we have a huge number of possible models with:

- Different functional form: *f(.)*, *g(.)*, *h(.)*, etc.
- Different explanatory variables: **X**, **Z**, **W**, dummy variables, **D**, etc.

Suppose, we have 4 different models to choose from:

Model 1 $\boldsymbol{y} = \mathbf{X}\beta + \varepsilon$

Model 2 $\boldsymbol{y} = \mathbf{Z}\gamma + \xi$

Model 3 $\boldsymbol{y} = (\mathbf{W}\gamma)^{\lambda} + \eta$

Model 4 $\boldsymbol{y} = \exp(\mathbf{Z}\,\mathbf{D}\,\delta) + \epsilon$

- We want to select the best model, the one that is closest to the true and unobserved DGP. In practice, we aim for a "good" model.

Model Selection Strategies – Views

- A model is a simplification. Many approaches to build a model:

- "**Pre-eminence of theory**." Economic theory should drive a model. Data is only used to quantify theory. Econometric methods offer sophisticated ways 'to bring data into line' with a particular theory.

- **Purely data driven models**. Success of ARIMA models (late 60s – early 70s), discussed in Parts 8 & 9: No theory, only exploiting the time-series characteristics of the data to build models.

- **Modern (LSE) view**. A compromise: theory and the characteristics of the data are used to build a model.

Model Selection Strategies – Modern View

• Theory and practice play a role in deriving a good model. David Hendry (2009) emphasizes:

"This implication is not a tract for mindless modeling of data in the absence of economic analysis, but instead suggests formulating more general initial models that embed the available economic theory as a special case, consistent with our knowledge of the institutional framework, historical record, and the data properties."

"Applied econometrics cannot be conducted without an economic theoretical framework to guide its endeavors and help interpret its findings. Nevertheless, since economic theory is not complete, correct, and immutable, and never will be, one also cannot justify an insistence on deriving empirical models from theory alone."

Model Selection Strategies – A Good Model

• According to David Hendry, a good model should be:

- Data admissible	-i.e., modeled and observed **y** should have the same properties.
- Theory consistent	-our model should "make sense"
- Predictive valid	-we should expect out-of-sample validation
- Data coherent	-all information should be in the model. Nothing left in the errors (*white noise errors*).
- Encompassing	-our model should explain earlier models.

• That is, we are searching for a statistical model that can generate the observed data (**y**, **X**), this is usually referred as *statistical adequacy*, makes theoretical sense and can explain other findings.

Model Selection Strategies – FAQ

- FAQ in practice:
  - Should I include all the variables in the database in my model?
  - How many explanatory variables do I need in my model?
  - How many models do I need to estimate?
  - What functional form should I be using?
  - Should the model allow for structural breaks?
  - Should I include dummies & interactive dummies ?
  - Which regression model will work best and how do I arrive at it?

Model Selection Strategies – Important Concepts

- *Diagnostic testing:* We test assumptions behind the model. In our case, assumptions (**A1**)-(**A5**) in the CLM.

Example: Test $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$ -i.e., the residuals are zero-mean, uncorrelated with anything (that is, white noise distributed errors).

In selecting a model, this is a very important step. We run a lot of test to check the residuals are acceptable or the model is not misspecified: Ramsey's reset test, tests for autocorrelation, etc.

- *Parameter testing:* We test economic H_0's.

Example: Test $\beta_k = 0$ -for example, there is no size effect on the expected return equation.

Model Selection Strategies: Two Methods

- There are several *model-selection methods*. We will consider two:
  - *Specific to General*
  - *General to Specific*

- Specific to General. Start with a small "restricted model," do some testing and make model bigger model in the direction indicated by the tests (for example, add variable x_k when test reject H_0: β_k=0).

- General to Specific. Start with a big "general unrestricted model," do some testing and reduce model in the direction indicated by the tests (for example, eliminate variable x_k when test cannot reject H_0: β_k=0).

Model Selection Strategies: Specific to General

- Steps:

(1) Begin with a small theoretical model – for example, the CAPM

$$\boldsymbol{y} = \mathbf{X}\beta + \varepsilon.$$

(2) Estimate the model – say, using OLS

(3) Do some diagnostic testing – are residuals white noise?

If the assumptions do not hold, then use:

- More advanced econometrics – GLS instead of OLS?
- A more general model – More regressors? Lags?

(4) Test economic H_0 on the parameters – Is HML significant?

(5) Modify model in (1) in the direction of rejections of H_0.

- This strategy is known as *specific to general*. In the machine learning literature, this strategy is also called *forwards selection*.

Model Selection Strategies: Specific to General

Example: Specific-to-general strategy to model IBM returns:

(1) We start with the 3-factor FF model for IBM:

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

(2) Estimate the 3-factor FF model for IBM:

```
fit_ibm_ff3 <- lm (ibm_x ~ Mkt_RF + SMB + HML)
> summary(fit_ibm_ff3)
Coefficients:
              Estimate    Std. Error t value Pr(>|t|)
(Intercept)  -0.005191   0.002482  -2.091   0.0369 *
Mkt_RF        0.910379   0.056784  16.032   <2e-16 ***
SMB          -0.221386   0.084214  -2.629   0.0088 **
HML          -0.139179   0.084060  -1.656   0.0983 .
---

Residual standard error: 0.05842 on 566 degrees of freedom
Multiple R-squared:  0.3393,    Adjusted R-squared:  0.3358
F-statistic:  96.9 on 3 and 566 DF,  p-value: < 2.2e-16
```

Model Selection Strategies: Specific to General

Example (continuation):

(3) Diagnostic tests: Check *t-stats* & R^2, F-test goodness of fit, etc.

(4) LM Test to test if there is a January Effect (H_0: No January effect):

```
> LM_test
[1] 9.084247
```

⇒ LM_test > **3.84** ⇒ Reject H_0: No January effect.

(5) Given this result, we modify the 3-factor FF and add the January Dummy to the FF model:

```
fit_ibm_new <- lm (ibm_x ~ Mkt_RF + SMB + HML + Jan_1)
> summary(fit_ibm_new)
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 **
```

Model Selection Strategies: Specific to General

• The specific-to-general method makes assumptions along the way Some remarks based on the previous example:

(1) Very likely the starting model is based on theory and experience (HML is not significant at the usual 5% level). Not clear how to proceed from there to a more general model.

(2) We tested for a January effect and then added to the model. However, we could have tested for a Dot.com effect or for an interactive Dot.com/January effect with the 3 FF factors. Not clear when to stop the search.

(3) Select a p-value to add variables to the model. In this case, we use the standard 5% for the tests.

Model Selection Strategies: Specific to General

• Note that in the previous example, we started with a model. What happens if are skeptical regarding models?

• A popular implementation of the specific-to-general model selection is the *stepwise regression*, where we start with only a set of potential explanatory variables and let the data, based on some criteria (R^2, AIC, etc.), determine which variables to keep.

Model Selection Strategies: Stepwise Regression

• Overall structure:

- The method begins with a k potential regressors.
- Do k one-variable regressions. Pick the one that shows the biggest t-stat or maximizes a goodness of fit measure, say, Adjusted-R^2, $\bar{R}^2$. Suppose x_j is selected.
- Then, do $(k-1)$-variable regressions all with x_j. Select the regressor (in addition to x_j) that has the highest t-stat or that maximizes $\bar{R}^2$.
- Continue. But, when we start adding regressors, we usually check if the added regressor(s) change the significance of previous steps. (Note: at each step, we remove or add a regressor(s) based on t- or F-tests.)
- Stop: Additional regressors do not have *significant* t-stats/increase $\bar{R}^2$.

Decisions: Selection of k variables, α for tests (α = 5%, 10%, 20%?) and goodness of fit statistic.

Model Selection Strategies: Stepwise Regression

• Decisions: Selection of k initial variables, α for tests (α = 5%, 10%, 30%?) and goodness of fit statistic.

Remark: Always keep in mind that the selected (final) model is not necessarily better than others. Type I and Type II errors are likely to occur, thus the final model may have irrelevant and/or omitted variables.

Technical Note: Though popular in practice, in general, selecting variables based on *p-values* is not advised, since the distribution of the OLS coefficients is affected. (Pre-testing problem, due to accumulation of Type I/Type II errors.)

Model Selection Strategies: Stepwise Regression

Example: Stepwise regression strategy to model IBM returns. We start with the 5 FF factors as candidates for IBM. We use the function *ols_step_forward_p* in the *olsrr* package, which uses *p-values* to select:

```
library(olsrr)
ff_step_data <- data.frame(Mkt_RF, SMB, HML, RMW, CMA)
ibm_ff_model <- lm(ibm_x ~ ., data = ff_step_data)        # default p-value (penter) is 0.3
ols_step_forward_p(ibm_ff_model , details = TRUE)        #  long final output
```

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	-0.005	0.002		-1.999	0.046	-0.010	0.000
Mkt_RF	0.887	0.055	0.574	16.227	0.000	0.780	0.995
SMB	-0.261	0.088	-0.111	-2.960	0.003	-0.435	-0.088
RMW	-0.128	0.114	-0.042	-1.122	0.262	-0.351	0.096

Model Selection Strategies: Stepwise Regression

Example (continuation):

Selection Summary

Step	Variable Entered	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	Mkt_RF	0.3087	0.3075	7.7108	-1665.5551	0.0594
2	SMB	0.3174	0.3151	2.2117	-1671.0548	0.0590
3	RMW	0.3188	0.3154	2.9552	-1670.3207	0.0590

Conclusion: The Stepwise Regression method selects Market excess returns, SMB & RMW as the drivers of IBM excess returns.

Model Selection Strategies: General to Specific

• Begin with a *general unrestricted model* (GUM), which nests restricted models and, thus, allows any restrictions to be tested. Say:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{W}^{\lambda}\boldsymbol{\delta} + (\mathbf{X} * \mathbf{W})\boldsymbol{\zeta} + (\mathbf{Z} * \mathbf{D})\boldsymbol{\psi} + \boldsymbol{\varepsilon}.$$

• Then, reduction of the GUM starts. Mainly using *t-tests*, and *F-tests*, we move from the GUM to a smaller, more parsimonious, specific model. If competing models are selected, encompassing tests or information criteria (AIC, BIC) can be used to select a final model. This is the *discovery stage*. After this reduction, we keep a final (restricted GUM) model:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

• Creativity is needed for the specification of a GUM. Theory and empirical evidence play a role in designing a GUM.

Model Selection Strategies: General to Specific

• General-to-Specific Method:

Step 1 - First ensure that the GUM does not suffer from any diagnostic problems. Check residuals in the GUM to ensure that they possess acceptable properties. (For example, test for white noise in residuals, incorrect functional form, autocorrelation, etc.).

Step 2 - Test the restrictions implied by the specific model against the general model – either by exclusion tests or other tests of linear restrictions.

Step 3 - If the restricted model is accepted, test its residuals to ensure that this more specific model is still acceptable on diagnostic grounds.

• This strategy is called *general to specifics* ("*gets*"), *LSE*, *TTT* (Test, test, test). In the machine learning literature, this strategy is also called *backwards selection*.

Model Selection Strategies: General to Specific

- The role of diagnostic testing is two-fold.

- In the *discovery steps* (**Steps 1** & **2**), the tests are being used as design criteria. Testing plays the role of checking that the original GUM was a good starting point after the GUM has been simplified.

- In the context of model evaluation (**Step 3**), the role of testing is clear cut. Suppose you use the model to produce forecasts. These forecasts can be evaluated with a test. This is the critical evaluation of the model.


John Dennis Sargan (1924 – 1996, England)

Model Selection Strategies: Properties

- A modeling strategy is *consistent* if its probability of finding the true model tends to 1 as T -the sample size- increases.

- Properties for strategies

(1) Specific to General
- It is not consistent if the original model is incorrect.
- It need not be predictive valid, data coherent, & encompassing.
- No clear stopping point for an unordered search.

(2) General to Specific
- It is consistent under some circumstances. But, it needs a large T.
- It uses data mining, which can lead to incorrect models for small T.
- The significance levels are incorrect. This is the problem of *mass significance*.

Model Selection Strategies: General to Specific

Example: General-to-specific strategy to model IBM returns:

Step 1 - Start with a GUM: the 3-factor FF model for IBM + January (Jan_t) & Dot.com (Dot_t) Dummy + non-linear & interactive effects:

$$(\text{IBM}_{\text{Ret}} - \text{r}_f)_t = \beta_0 + \beta_1 (r_{m,t} - r_f) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 Jan_t + \beta_5 Dot_t + \beta_6 (r_{m,t} - r_f)^2 + \beta_7 SMB_t{}^2 + \beta_8 HML_t{}^2 + \beta_9 (r_{m,t} - r_f) * SMB_t + \beta_{10} (r_{m,t} - r_f) * HML_t + \beta_{11} (r_{m,t} - r_f) * Jan_t + \beta_{12} SMB_t * Jan_t + \beta_{13} HML_t * Jan_t + \beta_{14} (r_{m,t} - r_f) * Dot_t + \beta_{15} HML_t * Dot_t + \beta_{16} SMB_t * Dot_t + \varepsilon_t$$

Step 1 - Estimate GUM:

```
Mkt_Jan <- Mkt_RF * Jan_1
HML_Jan <- HML * Jan_1
Mkt_Dot <- Mkt_RF * Dot_com
HML_Dot <- HML * Dot_com
SMB_Dot <- SMB * Dot_com
```

Model Selection Strategies: General to Specific

Example (continuation):

```
fit_ibm_gum <- lm (ibm_x ~ Mkt_RF + SMB + HML + Jan_1 + Mkt_RF_2 + SMB_2 + HML_2 +
Mkt_HML + Mkt_SMB + SMB_HML + Mkt_Jan + HML_Jan + Mkt_Dot + HML_Dot + SMB_Dot)
> summary(fit_ibm_gum)
Coefficients:
```

	Estimate	Std. Error	t value	Pr(>|t|)		
(Intercept)	-0.007836	0.003063	-2.559	0.010772	*	
Mkt_RF	0.791866	0.090474	8.752	< 2e-16	***	
SMB	-0.295790	0.110655	-2.673	0.007738	**	
HML	-0.233942	0.135146	-1.731	0.084004	.	⇒ practice says "keep it." Judgement call.
Jan_1	0.031769	0.009349	3.398	0.000727	***	
Mkt_RF_2	-0.433762	0.850899	-0.510	0.610417		
SMB_2	-0.927271	1.470645	-0.631	0.528615		
HML_2	2.707992	1.670366	1.621	0.105545		⇒ almost 10%, I keep it. Judgement call.
Mkt_HML	0.628721	1.557090	0.404	0.686531		
Mkt_SMB	0.791625	1.746939	0.453	0.650618		
SMB_HML	-1.044806	2.029091	-0.515	0.606819		
Mkt_Jan	-0.069413	0.189309	-0.367	0.714008		
HML_Jan	-0.259697	0.255484	-1.016	0.309841		

Model Selection Strategies: General to Specific

Example (continuation):

	Estimate	Std. Error	t value	Pr(>|t|)	
Mkt_Dot	0.323382	0.130645	**2.475**	0.013612	*
HML_Dot	0.059742	0.208277	0.287	0.774342	
SMB_Dot	0.076998	0.198964	0.387	0.698910	

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

Residual standard error: 0.05788 on 553 degrees of freedom
Multiple R-squared: 0.3663, Adjusted R-squared: 0.3491
F-statistic: 21.31 on 15 and 553 DF, p-value: < 2.2e-16

Step 1 – Check GUM residuals for departures of (**A2**)-(**A3**). A Ramsey's reset test can be done (using the *resettest* in the *lmtest* library).

```
> resettest(fit_ibm_gum, type="fitted")
      RESET test
data:  fit_gum
RESET = 1.2645, df1 = 2, df2 = 551, p-value = 0.2832
```

Model Selection Strategies: General to Specific

Example (continuation):

Step 2 – Reduce Model with t-test and F-tests. Say, we keep all the variables with a *p-value* close to 10% (we still keep HML, using previous experience). We estimate a restricted GUM:

```
fit_ibm_gum_r <- lm (ibm_x ~ Mkt_RF + SMB + HML + Jan_1 + HML_2 + Mkt_Dot)
> summary(fit_ibm_gum_r)
Coefficients:
```

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-0.008696	0.002788	**-3.119**	0.00191	**
Mkt_RF	0.779336	0.072453	**10.756**	< 2e-16	***
SMB	-0.280018	0.083891	**-3.338**	0.00090	***
HML	-0.250480	0.088504	**-2.830**	0.00482	**
Jan_1	0.028499	0.008937	**3.189**	0.00151	**
HML_2	1.676011	1.331161	1.259	0.20853	
Mkt_Dot	0.344030	0.116685	**2.948**	0.00333	**

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

Model Selection Strategies: General to Specific

Example (continuation):

Step 2 – Test the restrictions implied by the specific model against the general model. Using an F-test, we test *J*=9 restrictions:

$$H_0: \beta_5 = \beta_6 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = \beta_{14} = \beta_{15}.$$

```
e_u <- fit_ibm_gum$residuals                          # GUM residuals
RSS_u <- t(e_u)%*%e_u
e_r <- fit_ibm_gum_r$residuals                        # Restricted GUM residuals
RSS_r <- t(e_r)%*%e_r
f_test_gum <- ((RSS_r - RSS_u)/9)/(RSS_u/(T-16))      # F-test
> f_test_gum
      [,1]
[1,] 0.4299497          ⇒ we cannot reject H0 (f_test_gum < qchisq(.95, 9, 553) = 1.896801)
> qf(.95, df1=9, df2=T-16)
[1] 1.896801
p_val <- 1 - pf(f_test_gum, df = 9 , df2=T-16)        # p-value of F-test
>  p_val
[1,] 0.919105           ⇒ p-value is almost 1. No evidence for H0.
```

Model Selection Strategies: General to Specific

Example (continuation):

Step 2 – Further specification checks of Restricted GUM, for example, perform a Ramsey's reset test (using the *resettest* in the lmtest library).

```
> resettest(fit_gum_r, type="fitted")

      RESET test

data:  fit_ibm_gum_r
RESET = 1.0998, df1 = 2, df2 = 561, p-value = 0.3337
```

Step 3 - Test if Restricted GUM residuals are acceptable –i.e., do diagnostic tests (mainly, make sure they are white noise). If Restricted GUM passes all the diagnostic tests, it becomes the "final model."

Note: With the final model, we use it to justify/explain financial theory and features, and do forecasting.

Model Selection Strategies: General to Specific

• The general-to-specific method makes assumptions along the way. Some remarks based on the previous example:

(1) Select a *p-value* for the tests of significance in the discovery stage (we use **10%**). Given that we performed **15** *t-tests*, we should not be surprised we rejected the GUM, since we had an overall significance, $\alpha^* = .79$ [= 1 – (1 - **.10**)^**15**]. *Mass significance* is an issue.

(2) Judgement calls are also made.

(3) The reduction of the GUM involves "*pre-testing*" –i.e., data mining. We are likely rejecting a true H_0 (false positives) & not rejecting a true H_1 (false negatives), along the way. This increases the probability that the final model is not a good approximation. It is common to ignore (or not even acknowledge) pre-testing issues.

Model Selection Strategies: Best Subset

• Begin with a big model, with k regressors:

$$\boldsymbol{y} = \mathbf{X}\beta + \varepsilon.$$

The idea is to select the "best" subset of the k regressors in **X**, where "best" is defined by the researcher, say MSE, Adjusted-R^2, etc.

• In theory, it requires 2^k regressions. It can take a while if k is big (k < 40 is no problem).

• Many tricks are used to reduce the number of regressions.

• In practice, we use best subset to reduce the number of models to consider. For example, from the regressions with one-variable, keep the best one-variable model, from the regression with two-variables, keep the best two-variable model, etc.

Model Selection Strategies: Best Subset

Example: We want to select a model for IBM excess returns, using the k=3 Fama-French factors: Market excess returns (Mkt_RF), SMB, & HML. We have 8 (=2^3) models and, thus, regressions:

1) Constant;
2) Mkt_RF (CAPM)
3) SMB
4) HML
5) Mkt_RF & SMB
6) Mkt_Rf & HML
7) SMB & HML
8) Mkt_RF, SMB, & HML (the 3-factor F-F Model).

• We select the model with the lower MSE. Or, we can carry two or three models of the best models to do *cross-validation*.

Model Selection Strategies: Best Subset

Example (continuation): Suppose we selected three model: CAPM (M1); Mkt_RF & SMB (M2); and the 3-factor F-F Model (M3).

Now, we use K*-fold cross-validation*, with K = 5.

CV_5 M1: 0.003542756

CV_5 M2: **0.003505873**

CV_5 M3: 0.003556918

Note: Models look very similar. Practitioners compute a SE for CV_K and use a one SE rule. If within one SE, keep simplest model (M1).

Model Selection Strategies: Judgement Calls

- In the end, judgment must be used in weighing up various criteria:

- **The Economic Criterion** –are the estimated parameters plausible? (Economic Significance).

- **The First Order Statistical Criterion** –does the model provide a good fit (in-sample) with statistically significant parameter estimates?

- **The Second Order Statistical Criterion** –is the model generally free of misspecification problems – as evidenced in the diagnostic tests?

- **The Out of Sample Predictive Criterion** –does the model provide good out of sample predictions?