Lecture 6-c: Forecasting, Prediction and Model Selection

Brooks (4th edition): Chapter 5

1

(for private use, not to be posted/shared online)

Review - 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):

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

2

Review - 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)

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

⇒ 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

3

Review - 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)
```

4

Review - 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
```

5

Review - 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
```

⇒ small p-values: Reject H_0 (No *structural change*).

3 6

Review - Modeling 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

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

7

Review - Modeling Structural Change

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 AR(1) model, say D_{73}:

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

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

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) :	$\beta_0 + \gamma_0$	$\boldsymbol{\beta}_1 + \boldsymbol{\gamma}_1$

• We estimate the above model and perform an F-test to test if H_0 (No *structural change*): $\gamma_0 = 0$ & $\boldsymbol{\gamma}_1 = 0$.

8

Review - Modeling Structural Change

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**)).

9

Review – Chow Test with Unknown Break

• Problem with Chow Test for structural break: It is conditional on the set date for the structural break, which in general is unknown. An easy solution? Run a Chow test for all possible dates.

• The problem with this approach is that the technical conditions under which the asymptotic distribution is derived are not met in this setting.

• Andrews (1993) showed that under appropriate conditions, the QLR statistic, also known as SupLR statistic, has a *non-standard limiting distribution* ("*non-standard*" = no existing table; needs a new one).

• Andrews (1993) tabulated the non-standard distribution for different number of parameters in model (k), trimming values (π_0), & significance level (α). Andrews' table is in the next slide.

3

Review – Chow 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

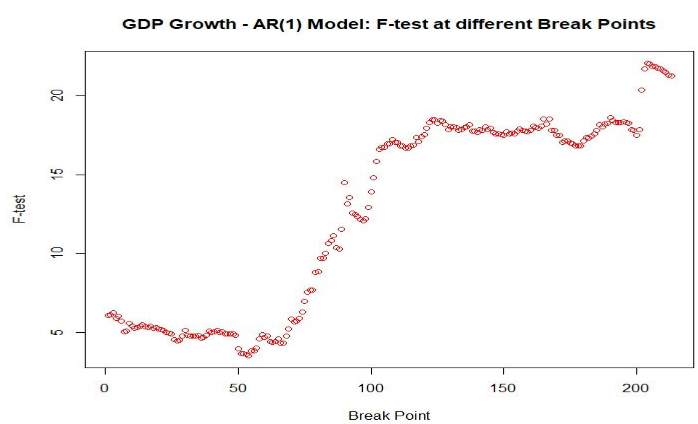
		$p=6$			$p=7$			$p=8$			$p=9$			$p=10$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.00	10.64	12.59	16.81	12.02	14.07	18.48	13.36	15.51	20.09	14.68	16.92	21.67	15.99	18.31	23.21
.49	1.08	11.81	13.74	18.32	13.27	15.52	19.93	13.29	15.63	20.53	16.17	18.56	23.05	17.35	19.79	24.62
.48	1.17	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

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

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

Review – Chow Test with Unknown Break

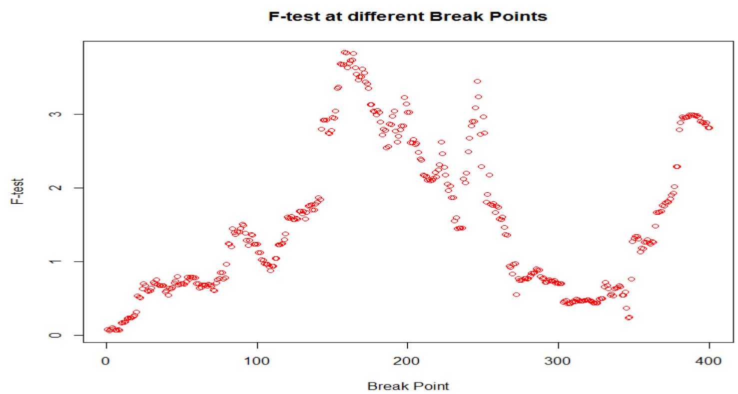
Example (continuation): We search for breaking points for GDP growth rate in AR(1) model. Below, we plot all F-tests starting at T*15:



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

Review – Chow Test with Unknown Break

Example: We search for breaking points for IBM returns in the 3-factor FF model. Below, we plot all F-tests starting at T*15:



- Maximum F is **3.83** occurs in May 1993 (observation #243). Then, $\widehat{QLR}$ = **3.83** < **16.45** $\Rightarrow$ Cannot reject H_0 at 5% level.

Chow Test: Structural Change – Remarks

- The results are *conditional* on the breaking point –say, **October 73** or **Dec 2001**.

- The breaking point is usually unknown. It needs to be estimated.

- It can deal only with one structural break –i.e., two categories!

- The number of breaks is also unknown. They need to be estimated.

- Characteristics of the data (heteroscedasticity –for example, regimes in the variance- and unit roots (high persistence) complicate the test.

- In general, only asymptotic (consistent) results are available.

- There are many modern tests that take care of these issues, but usually also with *non-standard* distributions.

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 $\boldsymbol{x}^0$ ⇒ predict y^0.

Forecast: Given $\boldsymbol{x}^0_{t+1}$ ⇒ predict y_{t+1}.

Forecasting and Prediction: Types

- Two types of predictions:

- *In sample* (**prediction**): The expected value of $\boldsymbol{y}$ (in-sample), given the estimates of the parameters. The in-sample prediction are the fitted values, $\hat{\boldsymbol{y}}$.

- *Out of sample* (**forecasting**): The value of a future $\boldsymbol{y}$ that is not observed by the sample.

Notation & Terminology:

Let T be the forecast origin and l is the forecast horizon.

- Prediction for T made at T: $\hat{Y}_T$.
- Forecast for $T+l$ made at T: $\hat{Y}_{T+l}$, $\hat{Y}_{T+l|T}$, $\hat{Y}_T(l)$.
- $\hat{Y}_T(l)$: l-*step ahead* forecast = Forecasted value Y_{T+l} at time T.

Forecasting and Prediction: Information Set, I_T

• 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, $I_T = \{\mathbf{x}^0\}$ to predict $\mathbf{y}^0$.
Or in time series, $I_T = \{\mathbf{x}^0_{T-1}, \mathbf{x}^0_{T-2}, ..., \mathbf{x}^0_{T-q}\}$ 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: Conditional E[.]

• The conditional expectation of Y_{T+l} is, in general, based on a model, the experience of the forecaster or a combination of both.

Example: We base the conditional expectation on the 3-factor FF model:

$$\hat{Y}_{T+l} = \mathrm{E}[(\beta_0 + \beta_1 (r_{m,T+l} - r_f) + \beta_2 SMB_{T+l} + \beta_3 HML_{T+l}) \,|\, \mathrm{I_T}]$$

• In the above equation, the forecast of Y_{T+l} also needs a forecast for the driving variables in the model. That is, we need a forecast for:
- $\mathrm{E}[(r_{m,T+l} - r_f) \,|\, \mathrm{I_T}]$
- $\mathrm{E}[SMB_{T+l} \,|\, \mathrm{I_T}]$
- $\mathrm{E}[HMM_{T+l} \,|\, \mathrm{I_T}]$

• In general, we will need a model for $\hat{X}_{T+l}$. Things can get complicated very quickly.

Forecasting and Prediction: Forecasts are RV

• 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**. to gauge the uncertainty in the forecast.

• 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.

Note: Later in this class, when we cover time series (Brooks, Chapter 6), we go deeper into forecasting.

Forecasting and Prediction: Variance-bias

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

• Given $\mathbf{x^0}$, we predict y^0, using the expectation: $E[y|\mathbf{X}, \mathbf{x^0}] = f(\mathbf{x^0}, \theta)$

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

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

• With y^0 observed, we compute the prediction error,

$$e^0 = \hat{y}^0 - y^0 = f(\mathbf{x^0}, \hat{\theta}) - f(\mathbf{x^0}, \theta) - \varepsilon^0$$

• The associated expected squared prediction error can be written as:

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

• We want to minimize this squared error, $E[(e^0)^2]$.

Forecasting and Prediction: Variance-bias

• The associated expected squared prediction error can be written as:

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

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

• Then, 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 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 $\mathbf{x}^0 \Rightarrow$ 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

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

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

Prediction Intervals: Point Estimate

Example: We estimated the 3 Factor FF Model for IBM returns:

> summary(fit_ibm_ff3)

	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 $\mathbf{x}^0$ = [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 (continuation): In R:

```
b_ibm <- fit_ibm_ff3$coefficients                    # regression coefficients, b
x_0 <- rbind(1.0000, -0.0189, -0.0142, -0.0027)      # x0
 y_0 <- 0.1555214
y_f0 <-  t(b_ibm)%*% x_0
> y_f0
          [,1]
[1,] -0.01877582
ef_0 <- y_f0 - y_0
> ef_0
          [,1]
[1,] -0.1742973
```

Prediction Intervals: C.I.

• We estimate the uncertainty behind the forecast with the Var[e^0].

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

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

⇒ Form a (1- α)% C.I. as usual.

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

Note: In out-of-sample forecasting, $\mathbf{x}^0$ is unknown, it has to be estimated.

(2) If $\mathbf{x}^0$ has to be estimated, then we use a random variable. What is the variance of the product? One possibility: Use bootstrapping.

Prediction Intervals: C.I. and Forecast Variance

• Assuming $\mathbf{x}^0$ is known, the variance of the forecast error is

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

If the model contains a constant term, this is

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

(where $\mathbf{Z}$ is $\mathbf{X}$ without $\mathbf{x}_1 = \boldsymbol{i}$). 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: $[\hat{y}^0 \pm t_{T\text{-}k,\alpha/2} * \text{sqrt}(\text{Var}[e^0])]$

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

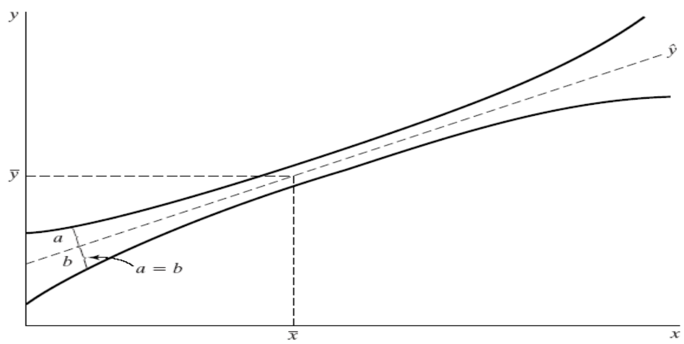


FIGURE 6.1 Prediction Intervals.

Prediction Intervals: C.I. and Forecast Variance

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

Then,

$$\text{Estimated Var}[e^0|\mathbf{x}^0] = \mathbf{x}^{0\prime}\,\text{Var}[\mathbf{b}|\mathbf{x}^0]\,\mathbf{x}^0 + s^2 = \mathbf{0.003429632}$$

```
Var_b <- vcov(fit_ibm_ff3)
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 $\mathbf{x}^0$ = colMeans(x)?

Prediction Intervals: C.I. and Forecast Variance

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. A wide interval, which makes clear the uncertainty surrounding the point forecast: $\hat{y}^0$ = **-0.01877587**

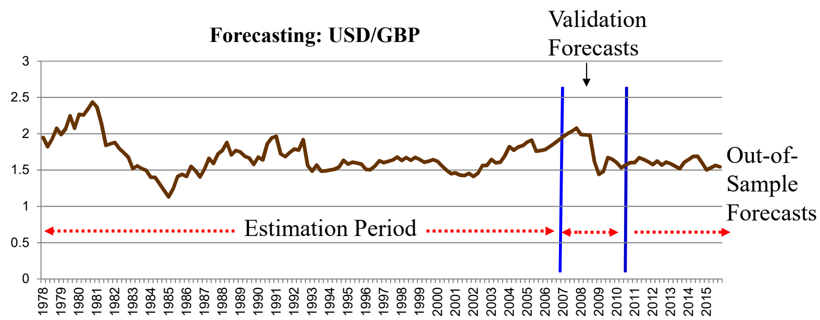
Forecasting and Prediction: Model Validation

• *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 $\hat{\mathbf{y}}^0$ with actual $\mathbf{y}^0$.

Forecasting and Prediction: Model Validation



Steps to measure forecast accuracy:

1) Select a (long) part of the sample (***estimation period***) to estimate the parameters of the model. (Get in-sample forecasts, $\hat{y}$.)

2) Keep a (short) part of the sample to check the model's forecasting skills. This is the ***validation step***. You can calculate true MSE or MAE

3) If happy with **Step 2**), proceed to do out-of-sample forecasts.

Forecasting and Prediction: Model Validation

Details:

1) ***Estimation period***. Use the first T_1 observations to estimate the parameters of the model. This step produces in-sample forecasts, $\hat{y}$. In-sample evaluation of model is usually performed here.

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

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

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

3) ***True OOS forecast period***. Produce OSS $\hat{y}^0$, but since 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, *validation (or "single-split" validation)*. This step can be used to "*tune (hyper-)parameters*." In our CLM, we can "tune" the model for departures of (**A1**)-(**A4**), for example, by including more or different variables (**A1**) and re-estimating the model accordingly using "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:

$$CV_n = \frac{1}{n}\sum_{i=1}^{n} MSE_{(-i)}$$

• 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 3-factor Fama-French Model for IBM returns with K=5:

```
y <- ibm_x
ff_cv_data <- data.frame(Mkt_RF, SMB, HML)
###### 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_cv_data, 5)
> mean(CV_ff_5)
[1] 0.00346262
```

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 $= \dfrac{\sqrt{\frac{1}{m}\sum_{i=T+1}^{T+m} {e_i}^2}}{\sqrt{\frac{1}{T}\sum_{i=1}^{T} {y_i}^2}}$

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

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

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

• Again, cross-validation measures can be used 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** (***m*** = **30** observations) for validation of predictions.

```
T0 <- 1
T1 <- 539                          # End of Estimation Period (Dec 2017)
T2 <- T1+1                         # Start of Validation Period (Jan 2018)
y1 <- y[T0:T1]
x1 <- x[T0:T1,]
fit_ibm_2 <- lm(y1~ x1 - 1)        # Estimation Period Regression From T0 to T1
b1 <- fit_ibm_2$coefficients       # Extract OLS coefficients from regression
> summary(fit_ibm_2)
           Estimate Std. Error t value Pr(>|t|)
x1         -0.003848   0.002571  -1.497  0.13510
x1Mkt_RF    0.865579   0.059386  14.575  < 2e-16 ***
x1SMB      -0.224914   0.085505  -2.630  0.00877 **
x1HML      -0.230838   0.090251  -2.558  0.01081 *
```


Evaluation of Forecasts: Measures of Accuracy

Example (continuation): We condition on the observed data (no model to predict FF factors used) from 2018: Jan to 2020: Jun.

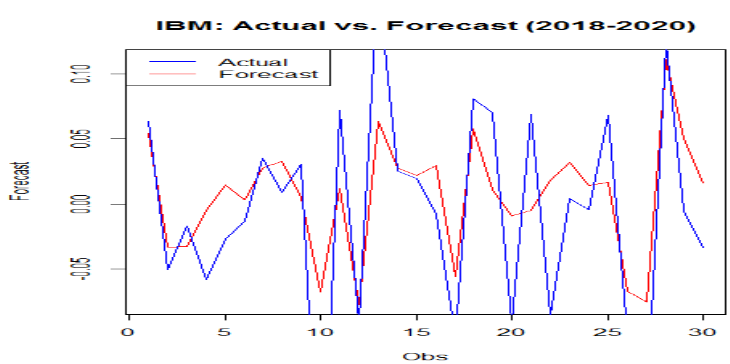
```
x_0 <- x[T2:T,]                    # Validation data
y_0 <- y[T2:T]                     # Validation data
y_f0 <-  x_0%*% b_ibm              # Forecast
ef_0 <- y_f0 - y_0                 # Forecasat error
mse_ef_0 <- sum(ef_0^2)/nrow(x_0)  # MSE
> mse_ef_0
[1] 0.003703207
mae_ef_0 <- sum(abs(ef_0))/nrow(x_0)   # MAE
> 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: Measures of Accuracy

• So far, we have judged the model with the better (usually, lower) measure of accuracy as the better forecasting mode.

But, measures of accuracy are RV. Then, we cannot look at these measures and establish that Model 1 is "more accurate" than Model 2. Statistical error ("luck") can create problems.

• The most popular measure of accuracy is the MSE.

Q: How do we know the MSE for model 1 is significantly better than the MSE for model 2? We need a test for

H_0: MSE(1) = MSE(2)
H_1: MSE(1) ≠ MSE(2).

.

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).

<u>Assumptions</u>: 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, $(e^{(1)} + e^{(2)})$ & $(e^{(1)} - e^{(2)})$ 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 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: $\beta = 0$ $\Rightarrow$ 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 & Newbold (1997).

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

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:

$$(r_i - r_f)_t = \beta_0 + \beta_1\,(r_m - r_f)_t + \beta_2 SMB_t + \beta_3\, HML_t + \varepsilon_t$$

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

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

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

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

Now, replacing the β by the estimated **b**, we have our one-step-ahead forecasts. We produce one forecast at a time.

Evaluation of Forecasts: Testing Accuracy

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

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

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

```
T1 <- 539                                # End of Estimation Period (Dec 2017)
x_0f <- x[T1:(T-1),]                     # By assumption on the X, it starts at T1.
y_0 <- y[T2:T]                           # T2 = T1 + 1 (Jan 2018)
y_f0 <-  x_0f %*% b1                     # b1 coefficients from fit_ibm_2
ef_0 <- y_f0 - y_0                       # e_t^(2)
mse_ef_0 <- sum(ef_0^2)/nrow(x_0)
> mse_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: $\beta = 0$.

```
# Step 1. Define errors and z & x
z_mgn <- ef_rw_0 + ef_0
x_mgn <- ef_rw_0 - ef_0

# Step 2. Regress x on z
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 ***       ⇒ significant!

# Step 3. t-test on β
> coef(summary(fit_mgn))[, "t value"]
1.619       4.762
```

Conclusion: We reject that both MSEs are equal ⇒ MSE of RW is higher.

Evaluation of Forecasts: MSE/MAE?

• MSE and MAE are very popular criteria to judge the forecasting power of a model. However, they may not be the best measure for everybody.

• Richard Levich's textbook compares forecasting services to the freely available forward rate. He finds that forecasting services may have some ability to predict direction (appreciation or depreciation).

For some investors, the direction is what really matters, since direction determines potential profits, not the error.

Forecasting Application: MSE/MAE?

Example: Two forecasts: Forward Rate ($F_{t,T}$) and Forecasting Service (FS)
$F_{t,1\text{-month}}$ = .7335 USD/CAD
$E_{FS,t}[S_{t+1\text{-month}}]$= **.7342** USD/CAD. (Assume S_t = **.7330** USD/CAD).
(Investor's strategy: buy CAD forward if FS forecasts CAD appreciation.)
Based on the FS forecast, Ms. Sternin decides to buy CAD forward at $F_{t,1\text{-m}}$.

(A) Suppose that the CAD appreciates to S_{t+1} = .7390 USD/CAD.
MAE_{FS} = |.7390 - **.7342**| = .0052 USD/CAD.
Investor makes a profit of .7390 - .7335 = **USD .055** USD.

(B) Suppose that the CAD depreciates to S_{t+1} = .7315 USD/CAD.
MAE_{FS} = |.7315 - **.7342**| = .0027 USD/CAD. ⇒ smaller MAE!
Investor takes a loss of .7315 - .7335 = **USD -.0020**. ¶

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)
- 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

$$E_t[S_{t+1}] = E_t[f(X_{t+1})] = g(X_t),$$

where X_t is a dataset regarded as *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.)

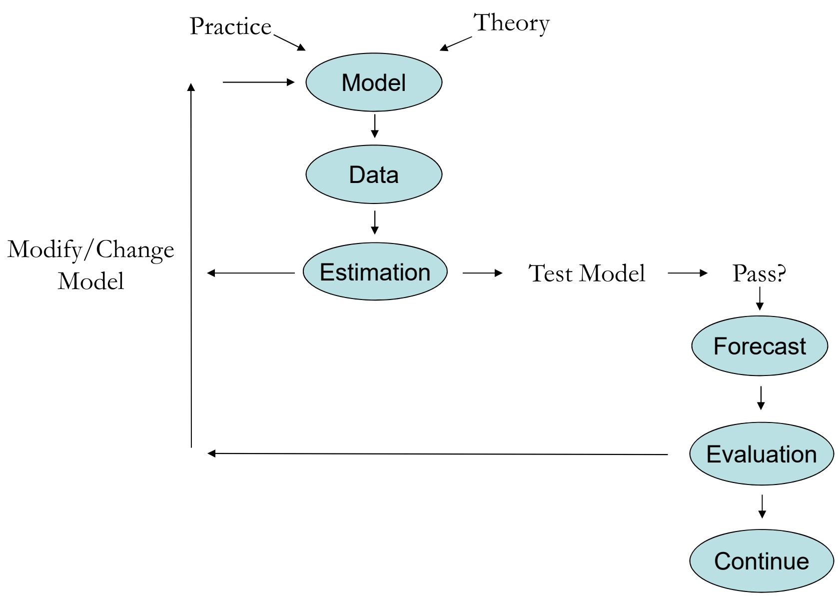
⇒ Mixture of art and science.

• 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, the **Random Walk** (**RW** model), which is found to be very good model for S_t in the short-run. The forecasts for the RW are given by:

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

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.

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

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

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

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

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

RW'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 = Model, RW)

Then, run MGN to compare MSEs.

