Lecture 5
Functional Form and Prediction

OLS Estimation - Assumptions

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

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

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

• Linear in variables and parameters:
  \[ Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \]

• Linear in parameters (intrinsic linear), nonlinear in variables:
  \[ Y = \beta_1 + \beta_2 X_2^2 + \beta_3 \sqrt{X_3} + \beta_4 \log(X_4) + \epsilon \]
  \[ Z_2 = X_2^2, \quad Z_3 = \sqrt{X_3}, \quad Z_4 = \log X_4 \]
  \[ Y = \beta_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + \epsilon \]

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

---

Functional Form: Linearity in Parameters

• Suppose we have:
  \[ Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + \epsilon \]

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

![Graph showing a quadratic relationship between y and X2](image-url)
Functional Form: Linearity in Parameters

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

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + \beta_3 X_3^3 + \cdots + \beta_{k+1} X_2^k + \epsilon$$

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

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

- $k$ may be too large or too small.
Functional Form: Linearity in Parameters

- Nonlinear in parameters:

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

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

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

Functional Form: Linearity in Parameters

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

- Suppose we start with a power function: \( Y = \beta_1 X^\beta_2 \varepsilon \)

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

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

- Now, we have an intrinsic linear model.

- To use the OLS estimates of \( \beta_1' \) and \( \beta_2' \), we need to say something about \( \varepsilon \). For example, \( \varepsilon = \exp(\xi) \), where \( \xi \mid X \sim \text{iid } D(\mathbf{0}, \sigma^2 I_n) \).
Functional Form: Linearity in Parameters

• Not all models are intrinsic linear. For example:

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

We cannot linearize the model by taking logarithms. There is no way of simplifying \(\log(b_1X^b + \varepsilon)\). We will have to use some nonlinear estimation technique.

---

Functional Form: Piecewise Linearity

• Sometimes non-linear relations in an interval can be linearized by splitting the interval. If this can be done, we say the relation is piecewise linear (a special case of a spline regression).

• Suppose we can linearized the data using two intervals –i.e., we have only one knot \((t_0)\). For example:

\[ E[y_i|X] = \beta_{00} + \beta_{01}x_i \quad \text{if } x_i \leq t_0 \]
\[ E[y_i|X] = \gamma_0 + \gamma_1x_i \quad \text{if } x_i > t_0 \]

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

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

where \((x_i-t_0)_+\) is the positive part of \((x_i-t_0)\) and zero otherwise.
Functional Form: Linear Splines

- We fit both equations into one single equation:

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

That is,

\[
E[y_i | X] = \begin{cases} 
\beta_{00} + \beta_{01}x_i & \text{if } x_i \leq t_0 \\
\gamma_0 + \gamma_1x_i = (\beta_{00} + \beta_{10} - \beta_{11}t_0) + (\beta_{01} + \beta_{11})x_i & \text{if } x_i > t_0 
\end{cases}
\]

- 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 \text{It can be estimated the model using OLS.}

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

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

- Linear model: \[ Y = \beta_1 + \beta_2 X + \epsilon \]
- (Semi-) Log model: \[ \log Y = \beta_1 + \beta_2 X + \epsilon \]
- Box–Cox transformation: \[ \frac{Y^{\lambda} - 1}{\lambda} = \beta_1 + \beta_2 X + \epsilon \]
  \[ \frac{Y^{\lambda} - 1}{\lambda} = Y - 1 \quad \text{when} \ \lambda = 1 \]
  \[ \frac{Y^{\lambda} - 1}{\lambda} = \log(Y) \quad \text{when} \ \lambda \to 0 \]
- Putting \( \lambda = 0 \) gives the (semi–)logarithmic model (think about the limit of \( \lambda \) tends to zero). We can estimate \( \lambda \). One would like to test if \( \lambda \) is equal to 0 or 1. It is possible that it is neither!

Functional Form: Ramsey’s RESET Test

- To test the specification of the functional form, Ramsey designed a simple test. We start with the fitted values:
  \[ \hat{y} = Xb. \]
  Then, we add \( \hat{y}^2 \) to the regression specification:
  \[ y = X \beta + \hat{y}^2 \gamma + \epsilon \]
  - If \( \hat{y}^2 \) is added to the regression specification, it should pick up quadratic and interactive nonlinearity, if present, without necessarily being highly correlated with any of the \( X \) variables.
  - We test \( H_0 \) (linear functional form): \( \gamma = 0 \)
    \[ H_1 \] (non linear functional form): \( \gamma \neq 0 \)
Functional Form: Ramsey’s RESET Test

• We test $H_0$ (linear functional form): $\gamma = 0$
  
  $H_1$ (non linear functional form): $\gamma \neq 0$
  
  $\Rightarrow$ *t*-test on the OLS estimator of $\gamma$.

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

• The RESET test is intended to detect nonlinearity, but not be specific about the most appropriate nonlinear model (no specific functional form is specified in $H_1$).

James B. Ramsey, England

Qualitative Variables and Functional Form

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

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

  Undergrad degree  \[ \text{Comp}_i = \beta_{0,u} + \beta_{1,u} z_i + \varepsilon_{u,i} \]

  Masters degree  \[ \text{Comp}_i = \beta_{0,m} + \beta_{1,m} z_i + \varepsilon_{m,i} \]

  Doctoral degree  \[ \text{Comp}_i = \beta_{0,d} + \beta_{1,d} z_i + \varepsilon_{d,i} \]

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

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

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

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

\[
Comp_i = \beta_0 + \beta_1'z_i + \beta_2D_{m,i} + \beta_3D_{d,i} + \gamma_1'z_iD_{m,i} + \gamma_2'z_iD_{d,i} + \epsilon_i
\]

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

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

---

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

• Suppose we have data for CEO graduate school. We can include another indicator variable in the model. Say \( D_{T20} \) to define if a graduate school is in the Top 20.

\[
D_{T20} = 1 \text{ if grad school is a Top 20 school} \]
\[ = 0 \text{ otherwise.} \]

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

• The omitted category is the reference category. In our previous example, the reference category is undergraduate degree.
**Dummy Variable for One Observation**

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

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

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

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

**Functional Form: Chow Test**

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

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

• Suppose we are interested in the effect of the 1973 oil shock in growth rates. We can include a dummy variable in the model, say $D_{73}$:

$$D_{73,i} = \begin{cases} 1 & \text{if observation } i \text{ occurred after October 1973} \\ 0 & \text{otherwise.} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 D_{73,i} + \gamma_1 x_i D_{73,i} + \epsilon_i$$

• We can use the Chow test to check if the 1973 oil crash produced a structural change in the growth rate.

• This is the more popular use of the Chow test.

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

• One important consideration: $T$ may not be large enough.

---

Functional Form: Structural Change

• Variation of the Chow test: Chow Predictive Test

• When there is not enough data to do the regression on both subsamples, 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.

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

Gregory C. Chow (1929, USA)
Chow Test (Greene)

Structural Change Test – the CHOW test. Is the regression model the same up to, then after the 1973 oil shock? Use 3 regressions to find out.

\[
p = \frac{(SS_{pooled} - (SS_0 + SS_1)) / K}{(SS_0 + SS_1) / (N0 - N1 - 2K)}
\]

Test statistic:

\[
= \frac{(0.0006419301 + 003142439 - 0.08553135)/9}{(0.0006419301 + 003142439)/(14 + 22 - 2(9))}
\]

= 3.334654

The critical F for [9,18] degrees of freedom is 2.456. The hypothesis is rejected.

Residuals Show the Implication of the Restriction of Equal Coefficients (Greene)
Algebra for the Chow Test

Unrestricted regression is
\[
\begin{pmatrix}
\mathbf{y}_{1960-1973} \\
\mathbf{y}_{1974-1995}
\end{pmatrix} =
\begin{pmatrix}
\mathbf{X}_{1960-1973} \\
\mathbf{X}_{1974-1995}
\end{pmatrix}
\begin{pmatrix}
\mathbf{0} \\
\mathbf{0}
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2
\end{pmatrix}
+ \begin{pmatrix}
\mathbf{e}_{1960-1973} \\
\mathbf{e}_{1974-1995}
\end{pmatrix}
\]

Restricted regression is
\[
\begin{pmatrix}
\mathbf{y}_{1960-1973} \\
\mathbf{y}_{1974-1995}
\end{pmatrix} =
\begin{pmatrix}
\mathbf{X}_{1960-1973} \\
\mathbf{X}_{1974-1995}
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2
\end{pmatrix}
+ \begin{pmatrix}
\mathbf{e}_{1960-1973} \\
\mathbf{e}_{1974-1995}
\end{pmatrix}
\]

In the unrestricted model, \( R = [1, -1], q = 0 \).
\( Rb - q = b_1 - b_2 \); 
\( R[\text{Var}(b_1, b_2)]R' = \text{Var}[b_1] + \text{Var}[b_2] \) (no covariance)

Application (Greene) – Health and Income

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

WHITEC = 1 if has “white collar” job
### Chow and Wald Tests for Gender (Greene) : Men

<table>
<thead>
<tr>
<th>Ordinary least squares regression</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LHS=HHNINC</td>
<td>Mean</td>
<td>.3590541</td>
<td>Standard deviation</td>
<td>.1735639</td>
</tr>
<tr>
<td>Model size</td>
<td>Parameters</td>
<td>5</td>
<td>Degrees of freedom</td>
<td>14238</td>
</tr>
<tr>
<td>Residuals</td>
<td>Sum of squares</td>
<td>379.8470</td>
<td>Standard error of e</td>
<td>.1633352</td>
</tr>
<tr>
<td>Fit</td>
<td>R-squared</td>
<td>.1146423</td>
<td>Adjusted R-squared</td>
<td>.1143936</td>
</tr>
</tbody>
</table>

| 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 and Wald Tests for Gender (Greene) : Female

<table>
<thead>
<tr>
<th>Ordinary least squares regression</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LHS=HHNINC</td>
<td>Mean</td>
<td>.3444951</td>
<td>Standard deviation</td>
<td>.1801790</td>
</tr>
<tr>
<td>Model size</td>
<td>Parameters</td>
<td>5</td>
<td>Degrees of freedom</td>
<td>13078</td>
</tr>
<tr>
<td>Residuals</td>
<td>Sum of squares</td>
<td>363.8789</td>
<td>Standard error of e</td>
<td>.1668045</td>
</tr>
<tr>
<td>Fit</td>
<td>R-squared</td>
<td>.1432098</td>
<td>Adjusted R-squared</td>
<td>.1429477</td>
</tr>
</tbody>
</table>

| 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 and Wald Tests for Gender (Greene) : All

<table>
<thead>
<tr>
<th>Ordinary least squares regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHS=HHNINC Mean = .3520836</td>
</tr>
<tr>
<td>Standard deviation = .1769083</td>
</tr>
<tr>
<td>Number of observs. = 27326</td>
</tr>
<tr>
<td>Model size Parameters = 5</td>
</tr>
<tr>
<td>Degrees of freedom = 27321</td>
</tr>
<tr>
<td>Residuals Sum of squares = 752.4767</td>
</tr>
<tr>
<td>Residuals Sum of squares = 379.8470</td>
</tr>
<tr>
<td>Residuals Sum of squares = 363.8789</td>
</tr>
</tbody>
</table>

Chow Test = F = \[(752.4767 - (379.847+363.8789))/5\] / \[(379.847+363.8789)/*(27321-10)\] = 64.269136
F(5,27311) = 2.214100 => reject H0

Wald Test (Greene)

```plaintext
--> 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 
+------------+ 
| 321.00313  |
+------------+ 
Matrix JF has 1 rows and 1 columns.  
1 
+------------+ 
| 321.40815  |
+------------+ 
```
Functional Form: Issues with Chow Test

• Issues with Chow tests
  - The results are conditional on the breaking point —say, October 73.
  - 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.

Functional Form: Unknown break

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

\[ QLR_T = \max_{\tau \in \{\tau_{\min}, \ldots, \tau_{\max}\}} F_T(\tau) \]

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

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

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

\[ \lim_{T \to \infty} QLR_T \convergeinlaw \sup_{r \in [r_{min}, r_{max}]} \left( \frac{B_k(r) - r B_k(1)}{(1-r)} \right) \]

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

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

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

Forecasting and Prediction

“There are two kinds 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 vs. model validation.
  - Within sample prediction
  - Hold out sample

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

• Prediction: Given \( x^0 \) => predict \( y^0 \).
  
  Two cases:
  
  Estimate: \( \mathbb{E}[y | X, x^0] = \beta'x^0 \); 
  Prediction: \( y^0 = \beta'x^0 + \varepsilon^0 \)
  
  Predictor: \( \hat{y}^0 = b'x^0 + \text{estimate of } \varepsilon^0 \). (Est. \( \varepsilon^0 = 0 \), but with variance)

• Forecast error. We predict \( y^0 \) with \( \hat{y}^0 = b'x^0 \).
  \[
  \hat{y}^0 - y^0 = b'x^0 - \beta'x^0 - \varepsilon^0 = (b - \beta)'x^0 - \varepsilon^0
  \]
  \[
  \Rightarrow \text{Var}[(\hat{y}^0 - y^0) | x^0] = \mathbb{E}[(\hat{y}^0 - y^0)'(\hat{y}^0 - y^0) | x^0] = x^0'\text{Var}(b - \beta) | x^0)x^0 + \sigma^2
  \]

• How do we estimate this?
  Two cases:
  (1) If \( x^0 \) is a vector of constants,
  (2) If \( x^0 \) has to be estimated.
Forecast Variance

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

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

- Interpretation: Forecast variance is smallest in the middle of our “experience” and increases as we move outside it.

Butterfly Effect

![Butterfly Effect Diagram](image)

**Figure 6.1** Prediction Intervals.

5.1 in the 6th edition
Forecasting performance of a model: Tests and measures of performance

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

Evaluation of forecasts: Measures of Accuracy

- Summary measures of out-of-sample forecast accuracy

Mean Error = \( \frac{1}{m} \sum_{t=1}^{T_m} (\hat{y}_t - y_t) \) = \( \frac{1}{m} \sum_{t=1}^{T_m} e_t \)

Mean Absolute Error (MAE) = \( \frac{1}{m} \sum_{t=1}^{T_m} |\hat{y}_t - y_t| \) = \( \frac{1}{m} \sum_{t=1}^{T_m} |e_t| \)

Mean Squared Error (MSE) = \( \frac{1}{m} \sum_{t=1}^{T_m} (\hat{y}_t - y_t)^2 \) = \( \frac{1}{m} \sum_{t=1}^{T_m} e_t^2 \)

Root Mean Square Error (RMSE) = \( \sqrt{\frac{1}{m} \sum_{t=1}^{T_m} (\hat{y}_t - y_t)^2} \) = \( \sqrt{\frac{1}{m} \sum_{t=1}^{T_m} e_t^2} \)

Theil’s U-stat = \( U = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^{T_m} e_t^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^{T_m} y_t^2}} \)
Evaluation of forecasts: Measures of Accuracy

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

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

• Q: How do we know an MSE for model 1 is better than the MSE for model 2?

Evaluation of forecasts: Testing Accuracy

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

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

Assumptions: Individual forecast errors are unbiased, normal, and not autocorrelated.

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

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

- There is a simpler way to do the MGN test. Let,
  \[ x_t = \epsilon_t(1) + \epsilon_t(2) \]
  \[ z_t = \epsilon_t(1) - \epsilon_t(2) \]

1. Do a regression: \[ x_t = \beta z_t + \varepsilon_t \]
2. Test \( H_0: \beta = 0 \) => a simple \( t \)-test.

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

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

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

---

**Out-of-sample predictions and prediction errors: Chow Test Revisited (Greene)**

- Chow’s prediction test
  \[ CHOW = \frac{(RSS_{T+m} - RSS_T)}{RSS_T / (T - k)} \sim F(m, T-k) \]

**Example:** A wage equation based on \( T=500 \) observations and 3 regressors plus a constant intercept gives an \( RSS_{500} = 78.8769257 \). Hold out sample: \( m=95 \) observations => \( RSS_{595} = 94.4958041 \)

So \( F_{cal} = (94.4958041-78.8769257)/95 = 1.03385097 \]
\[ 78.8769257/(500-4) = 1.03385097 \]

Since \( F(95, 496) = 1.0339 \) [\( p \text{-value: 0.4026} \)] => we cannot reject \( H_0 \) (parameter constancy).