Chapter 5

Classical linear regression model assumptions and diagnostics

I. Violation of the Assumptions of the CLRM

- Assumptions for the CLRM:

1. $E(u_t) = 0$
2. $\mathrm{Var}(u_t) = \sigma^2 < \infty$
3. $\mathrm{Cov}(u_i, u_j) = 0$
4. $\mathrm{Cov}(u_t, x_t) = 0$

   Stronger assumption: the ***X*** matrix is non-stochastic or fixed in repeated samples

5. $u_t \sim N(0, \sigma^2)$

Investigating Violations of the Assumptions of the CLRM

- We will now study these assumptions further, and in particular look at:
  - How we test for violations of the assumptions
  - Causes
  - Consequences

    in general we could encounter any combination of 3 problems:
    - the **coefficient estimates** are wrong
    - the associated **standard errors** are wrong
    - the **distribution** that we assumed for the test statistics will be inappropriate
  - Solutions
    - Operate such that the assumptions are no longer violated
    - Or we work around the problem so that the alternative techniques will still be valid

II. Statistical Distributions for Diagnostic Tests

- **Diagnostic tests** are based on the calculation of a a test statistic. Often, an *F*- and a χ^2- version of the test are available.

- The *F*-test version (**Wald Test**) involves estimating a restricted and an unrestricted version of a test regression and comparing the *RSS*. Degrees of freedom are *(m,T-k).*

- The χ^2- version is sometimes called an **"Lagrange Multiplier" test**, and only has one degree of freedom parameter: the number of restrictions being tested, *m*.

- Asymptotically, the 2 tests are equivalent since the χ^2 is a special case of the *F*-distribution:

$$\frac{\chi^2(m)}{m} = \lim F(m, T-k) \text{ as } T-k \to \infty$$

- For small samples, the ***F*-version is preferable**.

III. Assumption 1: $E(u_t) = 0$

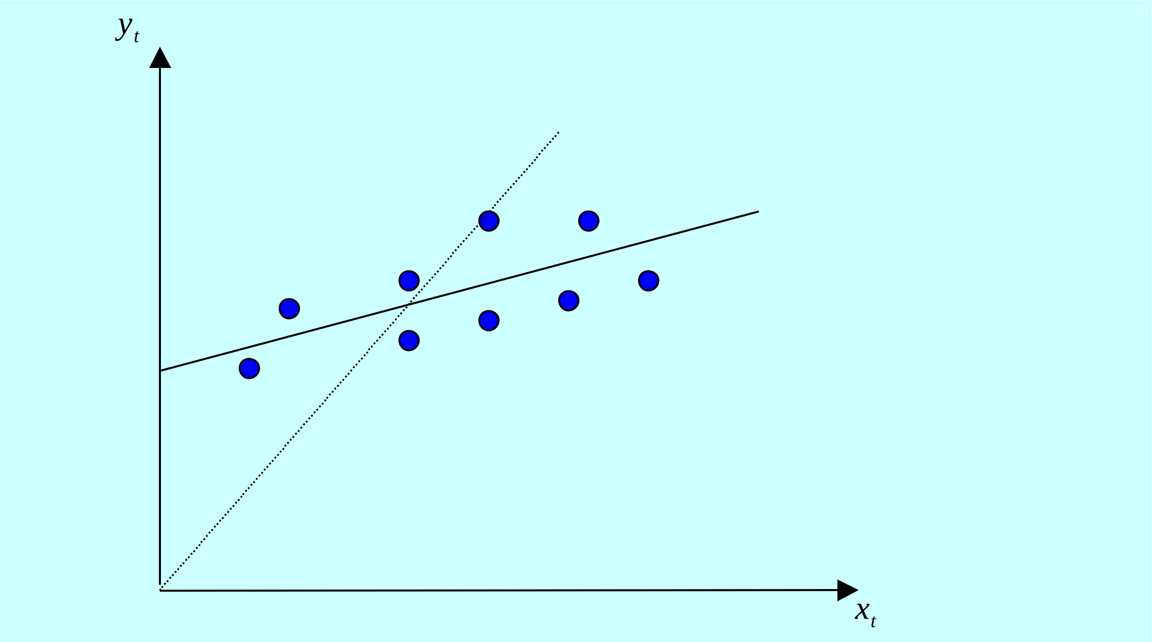
- Assumption that the mean of the disturbances is zero.

- For all diagnostic tests, we cannot observe the disturbances and so perform the **tests of the residuals**.

- Solution: **always introduce the constant term in the regression**.

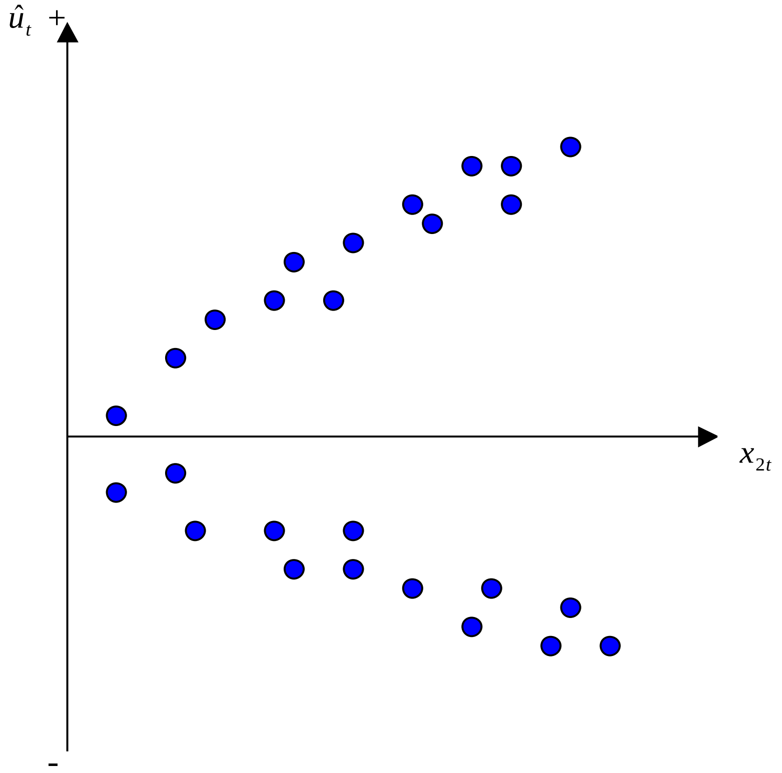
- Consequences if not:
  - R^2=ESS/TSS can be negative, then meaningless
  - Biased slope coefficient estimates

Biased slope coefficient estimate

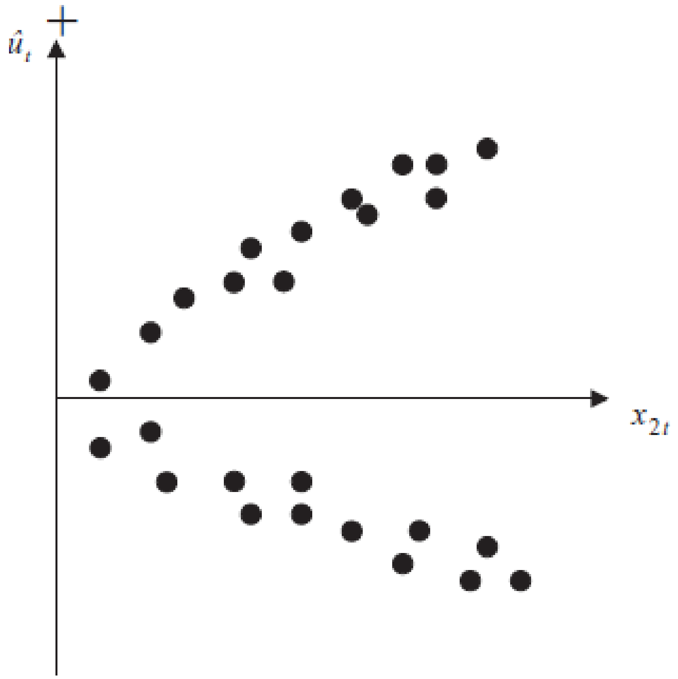


IV. Assumption 2: $\text{Var}(u_t) = \sigma^2 < \infty$

- We have so far assumed that the variance of the errors is constant, σ^2 - this is known as **homoscedasticity**. If the errors do not have a constant variance, we say that they are **heteroscedastic** e.g. say we estimate a regression and calculate the residuals, $\hat{u}_t$.



1. Detection of Heteroscedasticity: Graphical methods



2. Detection of Heteroscedasticity: The GQ Test

- Graphical methods: not always easy
- Formal tests: There are many of them: we will discuss Goldfeld-Quandt test and White's test

The **Goldfeld-Quandt (GQ) test** is carried out as follows.

1. Split the total sample of length T into two sub-samples of length T_1 and T_2. The regression model is estimated on each sub-sample and the two residual variances are calculated.

2. The null hypothesis is that the variances of the disturbances are equal,

$$H_0: \sigma_1^2 = \sigma_2^2$$

The GQ Test (Cont'd)

4. The test statistic, denoted by *GQ*, is simply the ratio of the two residual variances where the larger of the two variances must be placed in the numerator.

$$GQ = \frac{s_1^2}{s_2^2}$$

5. The test statistic is distributed as an **$F(T_1\text{-}k,\ T_2\text{-}k)$** under the null of homoscedasticity.

A problem with the test is that the choice of where to split the sample is that usually arbitrary and may crucially affect the outcome of the test.

3. Detection of Heteroscedasticity using White's Test

- **White's general test** for heteroscedasticity is one of the best approaches because it makes few assumptions about the form of the heteroscedasticity.

- The test is carried out as follows:

1. Assume that the regression we carried out is as follows

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

And we want to test whether Var(u_t) is a constant. We estimate the model, obtaining the residuals $\hat{u}_t$.

2. Then run the auxiliary regression:

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$$

Performing White's Test (cont'd)

3. Obtain R^2 from the auxiliary regression and multiply it by the number of observations, T. It can be shown that

$$T R^2 \sim \chi^2 (m)$$

where m is the number of regressors in the auxiliary regression excluding the constant term (m=5 in this context).

4. If the χ^2 test statistic from step 3 is greater than the corresponding value from the statistical table then reject the null hypothesis that the disturbances are homoscedastic.

- *Note: null-hypothesis*

Example 4.1

Suppose that the model (4.2) above has been estimated using 120 observations, and the R^2 from the auxiliary regression (4.3) is 0.234. The test statistic will be given by $TR^2 = 120 \times 0.234 = 28.8$, which will follow a $\chi^2(5)$ under the null hypothesis. The 5% critical value from the χ^2 table is 11.07. The test statistic is therefore more than the critical value and hence the null hypothesis is rejected. It would be concluded that there is significant evidence of heteroscedasticity, so that it would not be plausible to assume that the variance of the errors is constant in this case.

4. Consequences of Using OLS in the Presence of Heteroscedasticity

- OLS estimation still gives **unbiased coefficient estimates** , but they are **no longer BLUE (variance is not minimum)**.

- This implies that if we still use OLS in the presence of heteroscedasticity, our **standard errors could be inappropriate** and hence any inferences we make could be misleading.

- For example, standard errors for intercepts **are too big or too small** will depend upon the form of the heteroscedasticity.

5. How Do we Deal with Heteroscedasticity?

- **If the form (i.e. the cause) of the heteroscedasticity is known**, then we can use an estimation method which takes this into account (called **generalised least squares, GLS**).

- A simple illustration of GLS is as follows: Suppose that the error variance is related to another variable z_t by

$$\operatorname{var}(u_t) = \sigma^2 z_t^2$$

- To remove the heteroscedasticity, divide the regression equation by z_t

$$\frac{y_t}{z_t} = \beta_1 \frac{1}{z_t} + \beta_2 \frac{x_{2t}}{z_t} + \beta_3 \frac{x_{3t}}{z_t} + v_t$$

where $v_t = \frac{u_t}{z_t}$ is an error term.

- Now $\operatorname{var}(v_t) = \operatorname{var}\left(\frac{u_t}{z_t}\right) = \frac{\operatorname{var}(u_t)}{z_t^2} = \frac{\sigma^2 z_t^2}{z_t^2} = \sigma^2$ for known z_t.

6. Other Approaches to Dealing with Heteroscedasticity

- So the disturbances from the new regression equation will be homoscedastic.
- But the exact form of heteroscedasticity is usually unknown.

- **Other solutions may include when the form of heteroscedasticity is unknown:**

1. Transforming the variables into logs or reducing by some other measure of "size".
2. Use **White's heteroscedasticity consistent standard error** estimates

V. Assumption 3:
Testing Non-Autocorrelation Cov $(u_i, u_j) = 0$

The Concept of a Lagged Value

	Current value	*Lagged value*	*Difference*
t	y_t	y_{t-1}	Δy_t
1989M09	0.8	-	-
1989M10	1.3	0.8	1.3-0.8=0.5
1989M11	-0.9	1.3	-0.9-1.3=-2.2
1989M12	0.2	-0.9	0.2--0.9=1.1
1990M01	-1.7	0.2	-1.7-0.2=-1.9
1990M02	2.3	-1.7	2.3-(-1.7)=4.0
1990M03	0.1	2.3	0.1-2.3=-2.2
1990M04	0.0	0.1	0.0-0.1=-0.1
.	.	.	.

Non-Autocorrelation

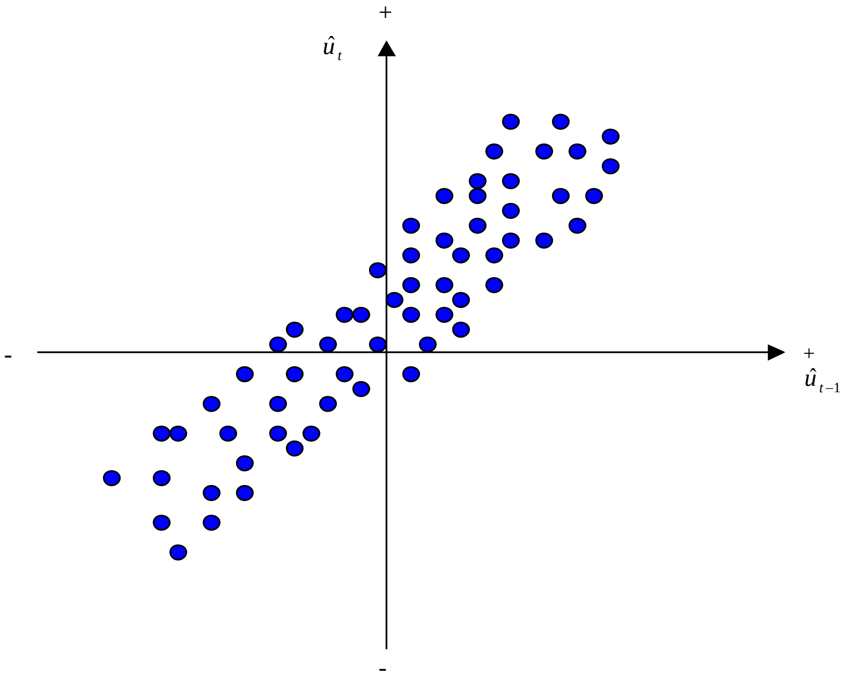
- We assumed of the CLRM's errors that Cov (u_i , u_j) = 0 for $i \neq j$, i.e. **errors are uncorrelated** with each other

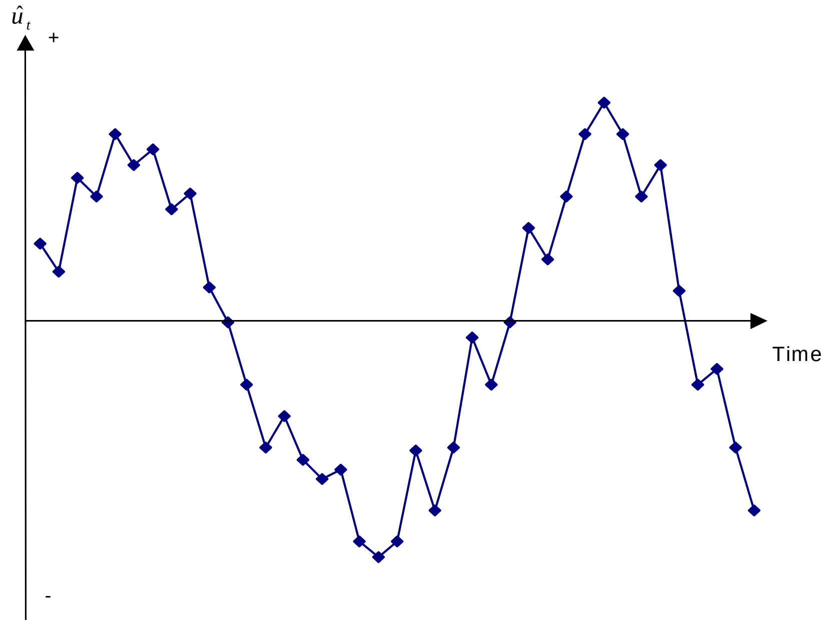
- Obviously we never have the actual u's, so we use their sample counterpart, the residuals (the $\hat{u}_t$).

- If the errors are not uncorrelated, we say that they are "**autocorrelated", or "serially correlated"**

- Some patterns we may find in the residuals are given on the next 3 slides

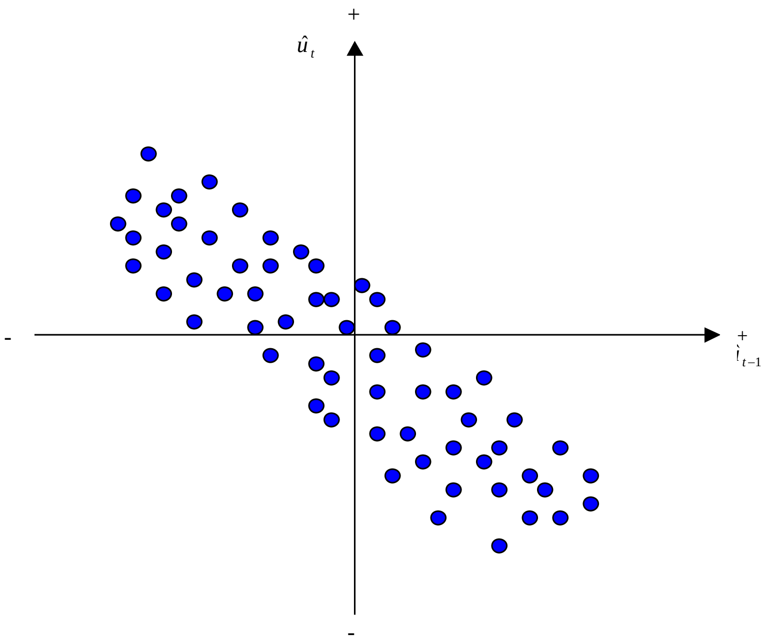
Positive Autocorrelation

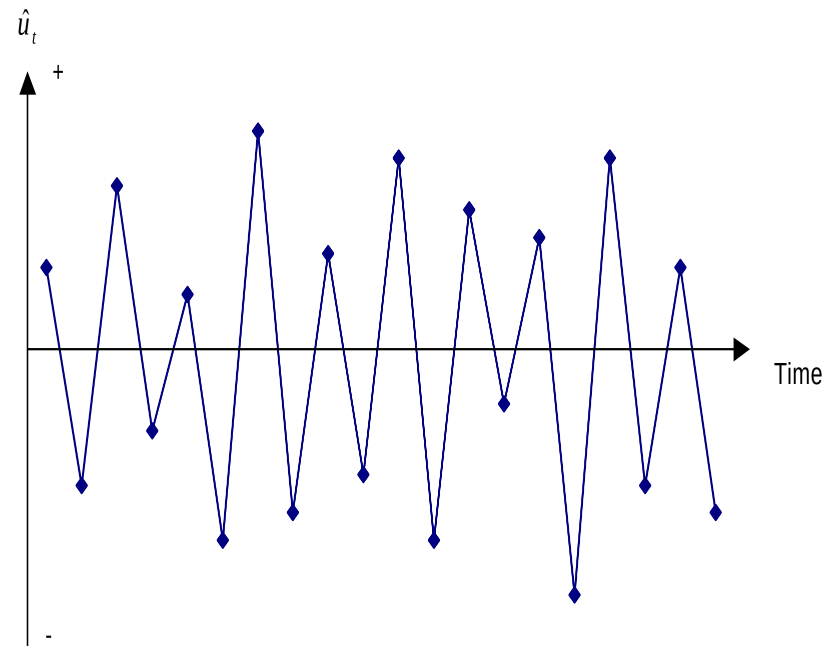




Positive Autocorrelation is indicated by a cyclical residual plot over time, where the residuals do not cross the time axis frequently

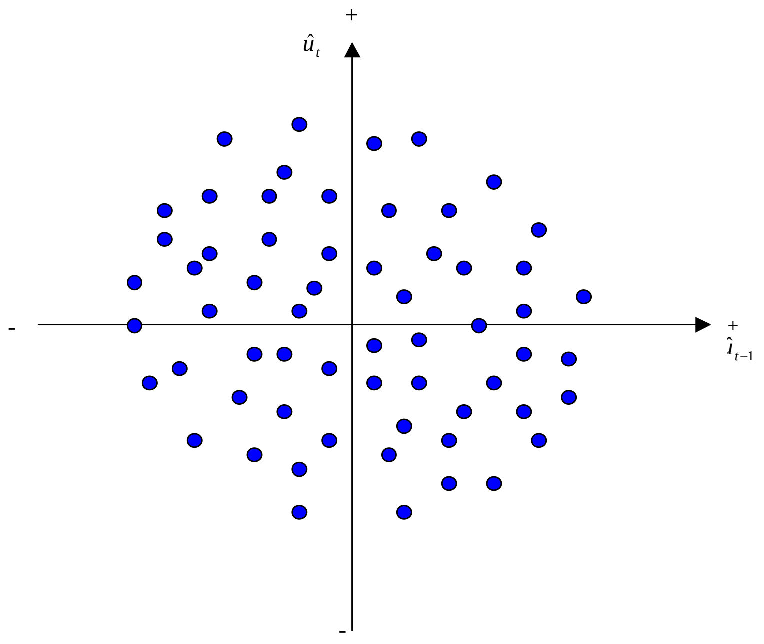
Negative Autocorrelation

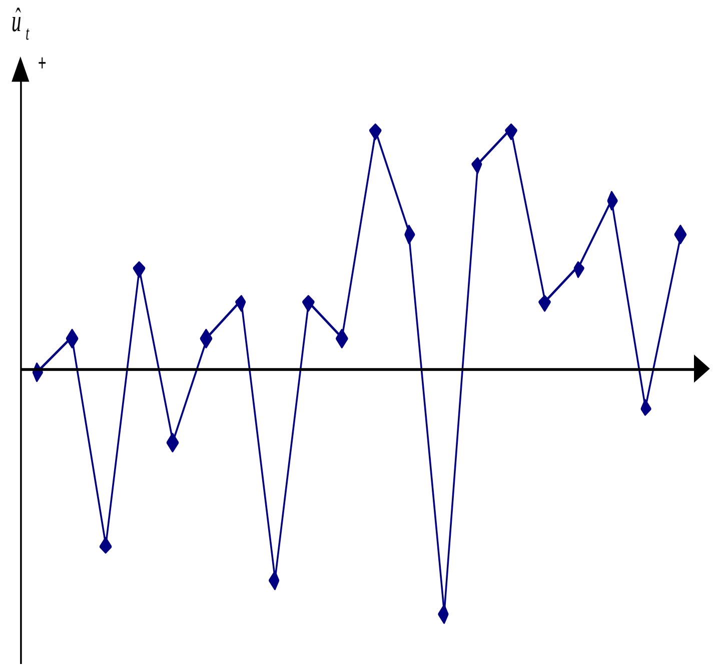




Negative autocorrelation is indicated by an alternating pattern where the residuals cross the time axis more frequently than if they were distributed randomly

No pattern in residuals –
No autocorrelation





No pattern in residuals at all: this is what we would like to see

1. Detecting Autocorrelation: The Durbin-Watson Test

The Durbin-Watson (DW) is a test for first order autocorrelation - i.e. it assumes that the relationship is between an error and the previous one

$$u_t = \rho u_{t\text{-}1} + v_t \qquad (1)$$

where $v_t \sim N(0, \sigma_v^2)$.

- The DW test statistic actually tests

$$H_0 : \rho=0 \text{ and } H_1 : \rho\neq 0$$

- The test statistic is calculated by

$$DW = \frac{\sum_{t=2}^{T} \left(\hat{u}_t - \hat{u}_{t-1}\right)^2}{\sum_{t=2}^{T} \hat{u}_t^2}$$

The Durbin-Watson Test: Critical Values

- We can also write

$$DW \approx 2(1 - \hat{\rho}) \qquad (2)$$

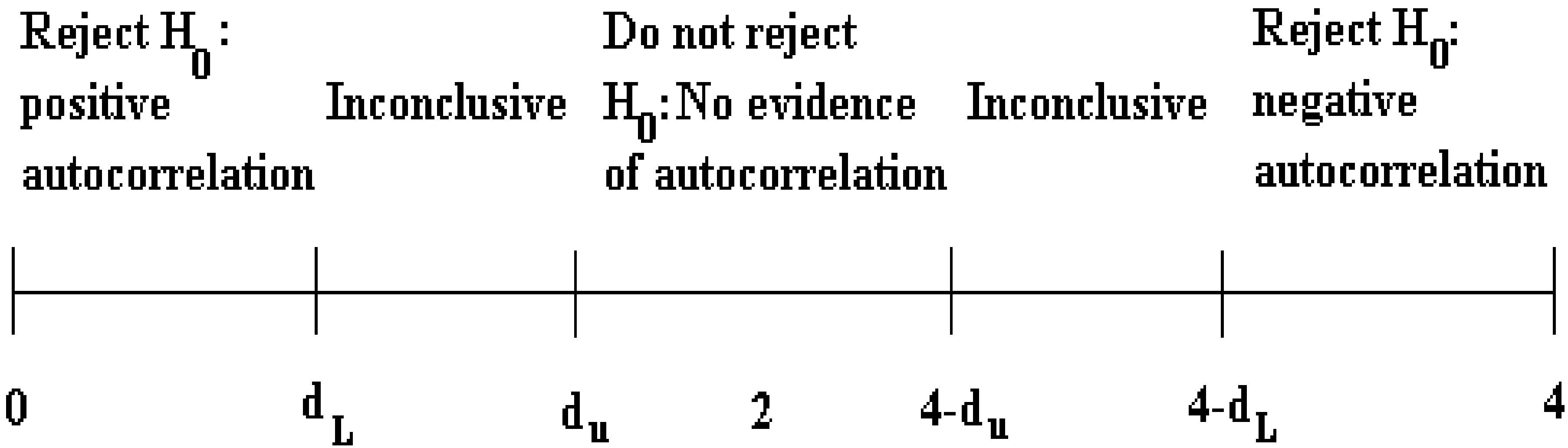
where $\hat{\rho}$ is the **estimated correlation coefficient**. Since $\hat{\rho}$ is a correlation, it implies that $-1 \leq \hat{p} \leq 1$.

- Rearranging for *DW* from (2) would give $\mathbf{0 \leq \mathit{DW} \leq 4}$.

- If $\hat{\rho} = 0$, $DW = 2$. So roughly speaking, do not reject the null hypothesis if *DW* is near 2 $\rightarrow$ i.e. there is little evidence of autocorrelation

- Unfortunately, *DW* has 2 critical values, an **upper critical value** (d_u) and a **lower critical value** (d_L), and there is also an intermediate region where we can neither reject nor not reject H_0.

The Durbin-Watson Test: Interpreting the Results



<u>Conditions which Must be Fulfilled for DW to be a Valid Test</u>

1. There must be a constant term in regression
2. Regressors must be non-stochastic
3. No lags of dependent variables in regression

Example 4.2

A researcher wishes to test for first order serial correlation in the residuals from a linear regression. The *DW* test statistic value is 0.86. There are 80 quarterly observations in the regression, and the regression is of the form

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t \quad (4.15)$$

The relevant critical values for the test (see table A2.6 in the appendix of statistical distributions at the end of this book), are $d_L = 1.42$, $d_U = 1.57$, so $4 - d_U = 2.43$ and $4 - d_L = 2.58$. The test statistic is clearly lower than the lower critical value and hence the null hypothesis of no autocorrelation is rejected and it would be concluded that the residuals from the model appear to be positively autocorrelated.

2. Another Test for Autocorrelation: The Breusch-Godfrey Test

- It is a more general test for r^{th} order autocorrelation:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \ldots + \rho_r u_{t-r} + v_t \quad , \quad v_t \sim \mathrm{N}(0, \sigma_v^2)$$

- The null and alternative hypotheses are:

$$\mathrm{H}_0 : \rho_1 = 0 \text{ and } \rho_2 = 0 \text{ and } \ldots \text{ and } \rho_r = 0$$

$$\mathrm{H}_1 : \rho_1 \neq 0 \text{ or } \rho_2 \neq 0 \text{ or } \ldots \text{ or } \rho_r \neq 0$$

- The test is carried out as follows:

1. Estimate the linear regression using OLS and obtain the residuals, $\hat{u}_t$
2. Regress $\hat{u}_t$ on all of the regressors from stage 1 (the x's) plus $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-r}$
   Obtain R^2 from this regression.
3. It can be shown that **$(T\text{-}r)R^2 \sim \chi^2(r)$**

- If the test statistic exceeds the critical value from the statistical tables, reject the null hypothesis of non-autocorrelation.

3. Consequences of Ignoring Autocorrelation if it is Present

- The coefficient estimates derived using OLS are still unbiased, but they are inefficient, i.e. they are not BLUE, even in large sample sizes.
- Thus, if the standard error estimates are inappropriate, there exists the possibility that we could make the wrong inferences.

4. “Remedies” for Autocorrelation

- Using **Newey-West Standard Errors** (Heteroskedasticity and Autocorrelation Consistent Standard Errors)
- Or Using **Dynamic Models** (see next)

5. Dynamic Models

- All of the models we have considered so far have been **static**, e.g.

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t$$

- But we can easily extend this analysis to the case where the current value of y_t depends on previous values of y or one of the x's, e.g.

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + \gamma_1 y_{t-1} + \gamma_2 x_{2t-1} + \ldots + \gamma_k x_{kt-1} + u_t$$

- We could extend the model even further by adding extra **lags**, e.g.

x_{2t-2} , y_{t-3} .

Why Might we Want/Need To Include Lags in a Regression?

- Inertia of the dependent variable: delay of change effect
- Over-reactions: effects of new announcements (psychological factors)

However, other problems with autocorrelation, and could not be remedied by adding lagged values:

  - Omission of (important) relevant variables, which are themselves autocorrelated.
  - Autocorrelation resulting from unparameterized seasonality (seasonal pattern implying positive correlation, not captured by model).
  - If "misspecification" error has been committed by using an inappropriate functional form.

Models in First Difference Form

- Another way to sometimes deal with the problem of autocorrelation is to switch to a model in first differences.

- Denote the first difference of y_t, i.e. y_t - y_{t-1} as Δy_t; similarly for the x-variables, $\Delta x_{2t} = x_{2t} - x_{2t-1}$ etc.

- The model would now be

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{2t} + \ldots + \beta_k \Delta x_{kt} + u_t$$

- Sometimes the change in y is purported to depend on previous values of y or x_t as well as changes in x:

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{2t} + \beta_3 x_{2t-1} + \beta_4 y_{t-1} + u_t$$

The Long Run Static Equilibrium Solution

- Dynamic models are difficult to interpret
- One interesting property of a dynamic model is its long run or static **equilibrium solution**

- "Equilibrium" implies that the variables have reached some steady state and are no longer changing, i.e. if y and x are in equilibrium, we can say

  $$y_t = y_{t+1} = \ldots = y \text{ and } x_t = x_{t+1} = \ldots = x$$

  Consequently, $\Delta y_t = y_t - y_{t-1} = y - y = 0$ etc.

- So the way to obtain a long run static solution is:
  1. Remove all time subscripts from variables
  2. Set error terms equal to their expected values, $E(u_t)=0$
  3. Remove first difference terms altogether
  4. Gather terms in x together and gather terms in y together.

- These steps can be undertaken in any order

The Long Run Static Equilibrium Solution: Example 4.3

If our model is

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{2t} + \beta_3 x_{2t\text{-}1} + \beta_4 y_{t\text{-}1} + u_t$$

then the static solution would be given by

$$0 = \beta_1 + \beta_3 x_{2t\text{-}1} + \beta_4 y_{t\text{-}1}$$

$$\beta_4 y_{t\text{-}1} = - \beta_1 - \beta_3 x_{2t\text{-}1}$$

$$y = \frac{-\beta_1}{\beta_4} - \frac{\beta_3}{\beta_4} x_2$$

Problems with Adding Lagged Regressors to "Cure" Autocorrelation

- In many cases, dynamic model will result in a removal of residual autocorrelation.

However, additional problems may be:

- Violating the assumption that the RHS variables are non-stochastic.
- The meaning of an equation with a large number of lags

VI. Assumption 4: Independent variables are non-stochastic

Since: $$\hat{\beta} = (X'X)^{-1}X'y \text{ with } y = X\beta + u,$$

$$\text{so } \hat{\beta} = (X'X)^{-1}X'\left(X\beta + u\right)$$

$$= \beta + (X'X)^{-1}X'u$$

If X and u are uncorrelated:

$$E\left(\hat{\beta}\right) = E\left(\beta\right) + E\left((X'X)^{-1}X'u\right)$$

$$= \beta + E\left[(X'X)^{-1}X'\right]E(u)$$

$$\text{so } E\left(\hat{\beta}\right) = \beta \text{ if } E(u) = 0.$$

Then OLS estimator is unbiased, consistent

VII. Assumption 5: Testing Normality Assumption for disturbances

- For conducting single and multiple joint hypothesis test about model parameters

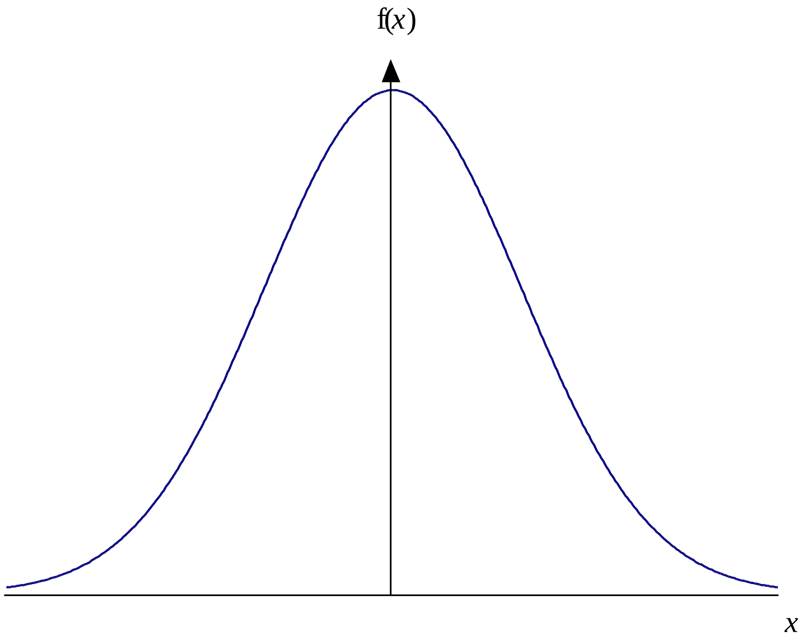
Testing for Departures from Normality

- *The Bera-Jarque normality test*

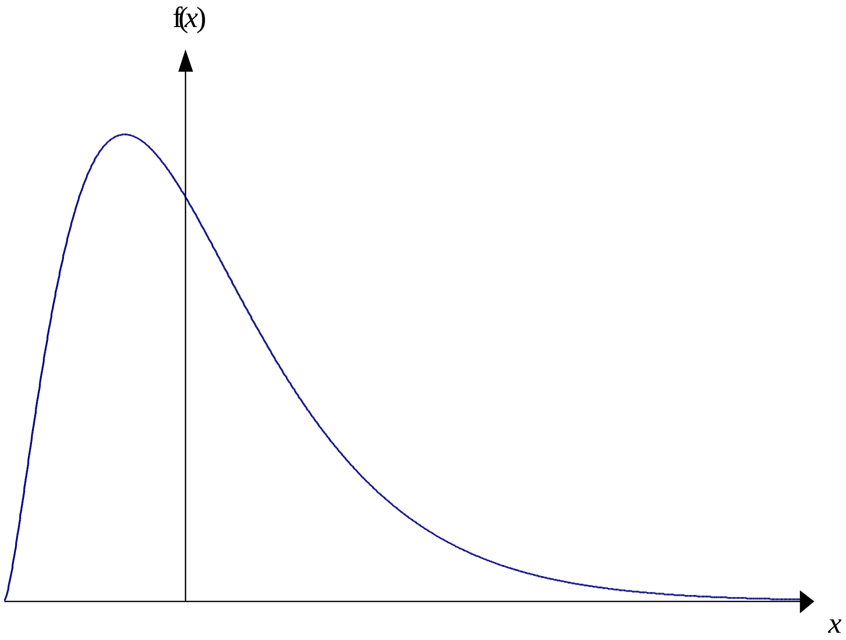
- A normal distribution is not skewed (b_1=0) and is defined to have a coefficient of kurtosis of 3 (b_2=3) .

- The kurtosis of the normal distribution is 3 so its excess kurtosis (b_2-3) is zero.

Normal versus Skewed Distributions

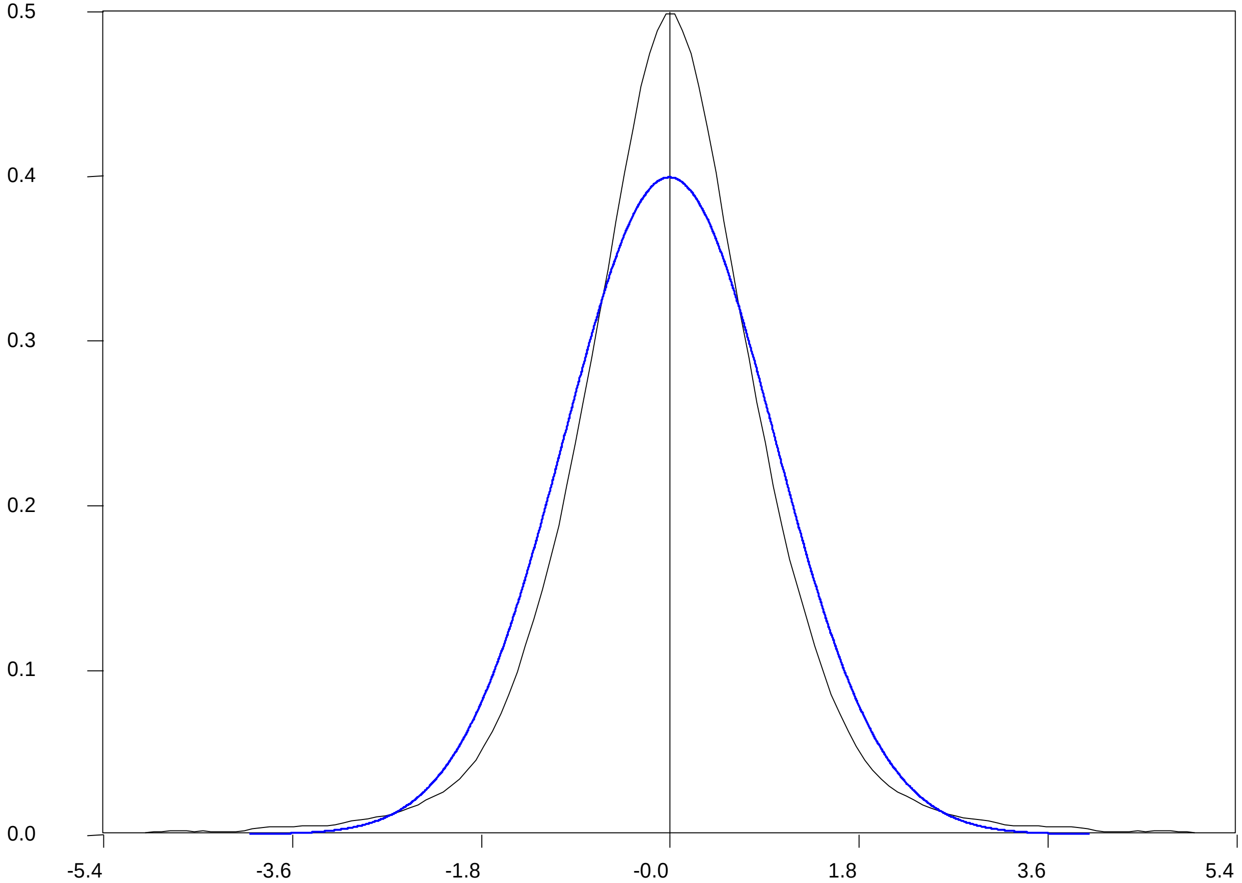


A normal distribution



A skewed distribution

Leptokurtic versus Normal Distribution



Testing for Normality

- Bera and Jarque formalise this by testing the residuals for normality by testing whether the coefficient of skewness and the coefficient of excess kurtosis are jointly zero.

- It can be proved that the coefficients of skewness and kurtosis can be expressed respectively as:

$$b_1 = \frac{E[u^3]}{\left(\sigma^2\right)^{3/2}} \quad \text{and} \quad b_2 = \frac{E[u^4]}{\left(\sigma^2\right)^2}$$

- The **Bera Jarque test statistic** is given by

$$W = T\left[\frac{b_1^2}{6} + \frac{(b_2 - 3)^2}{24}\right] \sim \chi^2(2)$$

- We estimate b_1 and b_2 using the residuals from the OLS regression, $\hat{u}$.

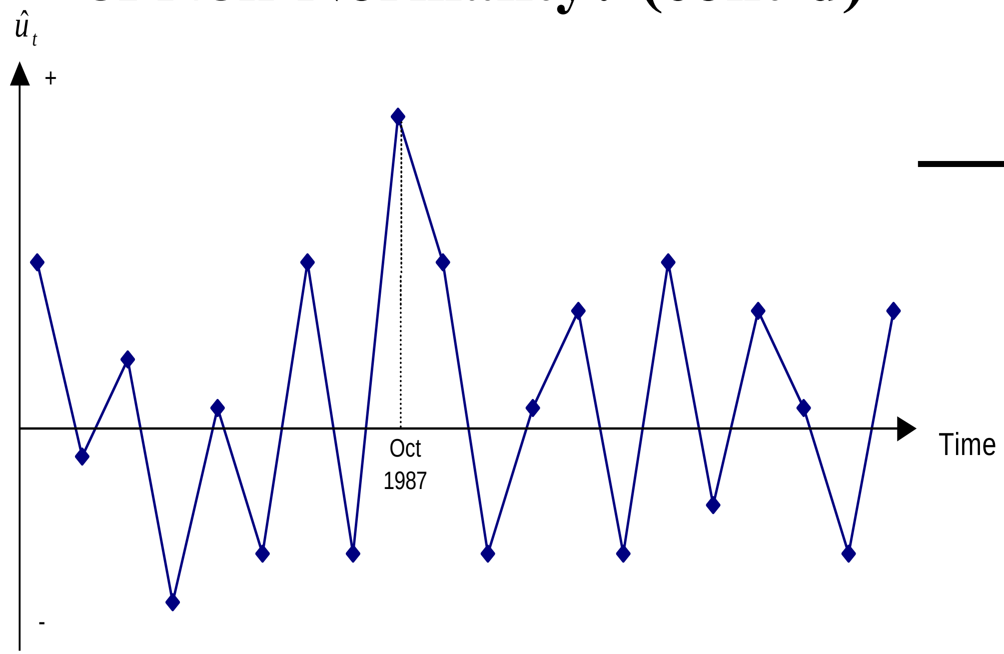
What do we do if we find evidence of Non-Normality?

- Stay with OLS and:
  - Increase sample size (Central Limit Theorem -> statistics follows appropriate distribution)
  - Often: 1 or 2 extreme residuals causes us to reject the normality assumption: outliers. An alternative is to use dummy variables to remove outliers.

- **Note: this assumption can be relaxed**

- **Example**: We estimate a monthly model of asset returns from 1980-1990, and we plot the residuals, and find a particularly large outlier for October 1987:

What do we do if we find evidence of Non-Normality? (cont'd)



- Create a new variable: $D87M10_t$ = 1 during October 1987 , = 0 otherwise.
  This effectively knocks out that observation.

Reasons: remove outliers -> decrease SE -> decrease RSS -> increase R -> better fit

But: lose information

How to detect outliers? By plotting y against x, or residuals u over time t

Often: stock market crash, financial panics, governmental crisis…

VIII. Multicollinearity problem (important)

- This problem occurs when the explanatory variables are very highly correlated with each other.

- **Perfect multicollinearity:** exact relationship between 2 or more variables

  Then we cannot estimate all the coefficients.

  Example: suppose $x_3 = 2x_2$ and the model is $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$. Then the matrix (X'X) is not invertible in the calculation of beta.

- **Near Multicollinearity**

- Problems if **Near Multicollinearity** is Present but Ignored

  - R^2 **will be high** but individual coefficients will have **high standard errors**.

  - The regression becomes **very sensitive to small changes** in the specification.

  - Thus confidence intervals for the parameters will be very wide, and significance tests might therefore give inappropriate conclusions.

Detect near multicollinearity using Correlation Matrix

- The easiest way to measure the extent of multicollinearity is simply to look at the matrix of correlations between the individual variables. e.g.

Corr	x_2	x_3	x_4
x_2	-	0.2	<u>0.8</u>
x_3	0.2	-	0.3
x_4	<u>0.8</u>	0.3	-

- <u>Question</u>: how to detect if 3 or more variables are linear:

  e.g. $x_{2t} + x_{3t} = x_{4t}$

- <u>Note</u> that high correlation between y and one of the x's is not muticollinearity.

Detect near multicollinearity using Variance Inflation Factor

- Run linear regression model of x_1 against $x_1, x_2, x_3, \ldots x_k$ and get R_1^2
- Run linear regression model of x_2 against $x_1, x_3, x_4, \ldots x_k$ and get R_2^2

  ….
- Run linear regression model of x_k against $x_1, x_2, x_3, \ldots$ and get R_k^2
- Set $VIF_i = 1/(1-R_i^2)$
- If $VIF_i > 10$ then multicollinearity is high
- If $VIF_i \le 10$ then multicollinearity is low

Solutions to the Problem of Multicollinearity

- Some econometricians argue that if the model is otherwise OK, just **ignore it**

- The easiest ways to "cure" the problems are:
  - **drop** one of the collinear variables
  - **transform** the highly correlated variables into a ratio
  - go out and **collect more data** e.g.
    - a longer run of data
    - switch to a higher frequency

XIV. Regression Analysis In Practice – A Further Example: Determinants of Sovereign Credit Ratings

- Cantor and Packer (1996)
- Sovereign credit ratings: assessment of riskiness of debt issued by government
- Two ratings agencies (Moody's and Standard and Poor's) provide credit ratings for many governments.
- Each possible rating is denoted by a grading:

Moody's	Standard and Poor's
Aaa	AAA
……	…..
B3	B-

Purposes of the Paper

- to attempt to explain and model how the ratings agencies arrived at their ratings.

- to use the same factors to explain the spreads of sovereign yields above a risk-free proxy

- to determine what factors affect how the sovereign yields react to ratings announcements

Determinants of Sovereign Ratings

- Data
  Quantifying the **ratings** (**dependent variable**): Aaa/AAA=16, ... , B3/B-=1

- Explanatory variables (units of measurement):
  - Per capita **income** in 1994 (thousands of dollars)
  - Average annual **GDP growth** 1991-1994 (%)
  - Average **annual inflation** 1992-1994 (%)
  - Fiscal balance: Average annual government budget surplus as a proportion of GDP 1992-1994 (%)
  - External balance: Average annual current account surplus as a proportion of GDP 1992-1994 (%)
  - External debt Foreign currency debt as a proportion of exports 1994 (%)
  - Dummy for **economic development**
  - Dummy for **default history**

The model: Linear and estimated using OLS

		Dependent Variable			
Explanatory Variable	Expected sign	Average Rating	Moody's Rating	S&P Rating	Moody's / S&P Difference
Intercept	?	1.442	3.408	-0.524	3.932**
		(0.663)	(1.379)	(-0.223)	(2.521)
Per capita income	+	1.242***	1.027***	1.458***	-0.431***
		(5.302)	(4.041)	(6.048)	(-2.688)
GDP growth	+	0.151	0.130	0.171**	-0.040
		(1.935)	(1.545)	(2.132)	(0.756)
Inflation	-	-0.611***	-0.630***	-0.591***	-0.039
		(-2.839)	(-2.701)	(2.671)	(-0.265)
Fiscal Balance	+	0.073	0.049	0.097*	-0.048
		(1.324)	(0.818)	(1.71)	(-1.274)
External Balance	+	0.003	0.006	0.001	0.006
		(0.314)	(0.535)	(0.046)	(0.779)
External Debt	-	-0.013***	-0.015***	-0.011***	-0.004***
		(-5.088)	(-5.365)	(-4.236)	(-2.133)
Development dummy	+	2.776***	2.957***	2.595***	0.362
		(4.25)	(4.175)	(3.861)	(0.81)
Default dummy	-	-2.042***	-1.63**	-2.622***	1.159***
		(-3.175)	(-2.097)	(-3.962)	(2.632)
Adjusted R^2		0.924	0.905	0.926	0.836

Notes: *t*-ratios in parentheses; *, **, and *** indicate significance at the 10%, 5% and 1% levels respectively. Source: Cantor and Packer (1996). Reprinted with permission from *Institutional Investor*.