CHAPTER 8
Regression Diagnostics and Specification Tests

8.1 Influential Observations[1]

Sources of influential observations include: (i) improperly recorded data, (ii) observational errors in the data, (iii) misspecification and (iv) outlying data points that are legitimate and contain valuable information which improve the efficiency of the estimation. It is constructive to isolate extreme points and to determine the extent to which the parameter estimates depend upon these desirable data.

One should always run descriptive statistics on the data, see Chapter 2. This will often reveal outliers, skewness or multimodal distributions. Scatter diagrams should also be examined, but these diagnostics are only the first line of attack and are inadequate in detecting multivariate discrepant observations or the way each observation affects the estimated regression model.

In regression analysis, we emphasize the importance of plotting the residuals against the explanatory variables or the predicted values $\widehat{y}$ to identify patterns in these residuals that may indicate nonlinearity, heteroskedasticity, serial correlation, etc, see Chapter 3. In this section, we learn how to identify significantly large residuals and compute regression diagnostics that may identify influential observations. We study the extent to which the deletion of any observation affects the estimated coefficients, the standard errors, predicted values, residuals and test statistics. These represent the core of diagnostic tools in regression analysis.

Accordingly, Belsley, Kuh and Welsch (1980, p. 11) define an influential observation as "..one which, either individually or together with several other observations, has demonstrably larger impact on the calculated values of various estimates (coefficients, standard errors, t-values, etc.) than is the case for most of the other observations."

First, what is a significantly large residual? We have seen that the least squares residuals of y on X are given by $e = (I_n - P_X)u$, see equation (7.7). y is $n \times 1$ and X is $n \times k$. If $u \sim \text{IID}(0, \sigma^2 I_n)$, then e has zero mean and variance $\sigma^2(I_n - P_X)$. Therefore, the OLS residuals are *correlated* and *heteroskedastic* with $\text{var}(e_i) = \sigma^2(1 - h_{ii})$ where h_{ii} is the i-th diagonal element of the *hat* matrix $H = P_X$, since $\widehat{y} = Hy$.

The diagonal elements h_{ii} have the following properties:

$$\sum_{i=1}^n h_{ii} = \text{tr}(P_X) = k \quad \text{and} \quad h_{ii} = \sum_{j=1}^n h_{ij}^2 \geq h_{ii}^2 \geq 0.$$

The last property follows from the fact that P_X is symmetric and idempotent. Therefore, $h_{ii}^2 - h_{ii} \leq 0$ or $h_{ii}(h_{ii} - 1) \leq 0$. Hence, $0 \leq h_{ii} \leq 1$, (see problem 1). h_{ii} is called the *leverage* of the i-th observation. For a simple regression with a constant,

$$h_{ii} = (1/n) + (x_i^2 / \textstyle\sum_{i=1}^n x_i^2)$$

where $x_i = X_i - \bar{X}; h_{ii}$ can be interpreted as a measure of the distance between X values of the i-th observation and their mean over all n observations. A large h_{ii} indicates that the i-th observation is distant from the center of the observations. This means that the i-th observation with large h_{ii} (a function only of X_i values) exercises substantial leverage in determining the fitted value $\widehat{y}_i$. Also, the larger h_{ii}, the smaller the variance of the residual e_i. Since observations

B.H. Baltagi, *Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-20059-5_8,

with high leverage tend to have smaller residuals, it may not be possible to detect them by an examination of the residuals alone. But, what is a large leverage? h_{ii} is large if it is more than twice the mean leverage value $2\bar{h} = 2k/n$. Hence, $h_{ii} \geq 2k/n$ are considered outlying observations with regards to X values.

An alternative representation of h_{ii} is simply $h_{ii} = d_i'P_X d_i = ||P_X d_i||^2 = x_i'(X'X)^{-1}x_i$ where d_i denotes the i-th observation's dummy variable, i.e., a vector of dimension n with 1 in the i-th position and 0 elsewhere. x_i' is the i-th row of X and $||.||$ denotes the Euclidian length. Note that $d_i'X = x_i'$.

Let us standardize the i-th OLS residual by dividing it by an estimate of its variance. A *standardized residual* would then be:

$$\widetilde{e}_i = e_i/s\sqrt{1-h_{ii}} \tag{8.1}$$

where σ^2 is estimated by s^2, the MSE of the regression. This is an *internal studentization* of the residuals, see Cook and Weisberg (1982). Alternatively, one could use an estimate of σ^2 that is independent of e_i. Defining $s_{(i)}^2$ as the MSE from the regression computed without the i-th observation, it can be shown, see equation (8.18) below, that

$$s_{(i)}^2 = \frac{(n-k)s^2 - e_i^2/(1-h_{ii})}{(n-k-1)} = s^2\left(\frac{n-k-\widetilde{e}_i^2}{n-k-1}\right) \tag{8.2}$$

Under normality, $s_{(i)}^2$ and e_i are independent and the *externally studentized* residuals are defined by

$$e_i^* = e_i/s_{(i)}\sqrt{1-h_{ii}} \sim t_{n-k-1} \tag{8.3}$$

Thus, if the normality assumption holds, we can readily assess the significance of any single studentized residual. Of course, the e_i^* will not be independent. Since this is a t-statistic, it is natural to think of e_i^* as large if its value exceeds 2 in absolute value.

Substituting (8.2) into (8.3) and comparing the result with (8.1), it is easy to show that e_i^* is a monotonic transformation of $\widetilde{e}_i$

$$e_i^* = \widetilde{e}_i\left(\frac{n-k-1}{n-k-\widetilde{e}_i^2}\right)^{\frac{1}{2}} \tag{8.4}$$

Cook and Wiesberg (1982) show that e_i^* can be obtained as a t-statistic from the following augmented regression:

$$y = X\beta^* + d_i\varphi + u \tag{8.5}$$

where d_i is the dummy variable for the i-th observation. In fact, $\widehat{\varphi} = e_i/(1-h_{ii})$ and e_i^* is the t-statistic for testing that $\varphi = 0$. (see problem 4 and the proof given below). Hence, whether the i-th residual is large can be simply determined by the regression (8.5). A dummy variable for the i-th observation is included in the original regression and the t-statistic on this dummy tests whether this i-th residual is large. This is repeated for all observations $i = 1, \ldots, n$.

This can be generalized easily to testing for a group of significantly large residuals:

$$y = X\beta^* + D_p\varphi^* + u \tag{8.6}$$

where D_p is an $n \times p$ matrix of dummy variables for the p-suspected observations. One can test $\varphi^* = 0$ using the Chow test described in (4.17) as follows:

$$F = \frac{[\text{Residual SS(no dummies)} - \text{Residual SS}(D_p \text{ dummies used})]/p}{\text{Residual SS}(D_p \text{ dummies used})/(n-k-p)} \tag{8.7}$$

This will be distributed as $F_{p,n-k-p}$ under the null, see Gentleman and Wilk (1975). Let

$$e_p = D_p' e, \quad \text{then} \quad E(e_p) = 0 \quad \text{and} \quad \text{var}(e_p) = \sigma^2 D_p' \bar{P}_X D_p \tag{8.8}$$

Then one can show, (see problem 5), that

$$F = \frac{[e_p'(D_p'\bar{P}_X D_p)^{-1}e_p]/p}{[(n-k)s^2 - e_p'(D_p'\bar{P}_X D_p)^{-1}e_p]/(n-k-p)} \sim F_{p,n-k-p} \tag{8.9}$$

Another refinement comes from estimating the regression without the i-th observation:

$$\widehat{\beta}_{(i)} = [X_{(i)}'X_{(i)}]^{-1}X_{(i)}'y_{(i)} \tag{8.10}$$

where the (i) subscript notation indicates that the i-th observation has been deleted. Using the updating formula

$$(A - a'b)^{-1} = A^{-1} + A^{-1}a'(I - bA^{-1}a')^{-1}bA^{-1} \tag{8.11}$$

with $A = (X'X)$ and $a = b = x_i'$, one gets

$$[X_{(i)}'X_{(i)}]^{-1} = (X'X)^{-1} + (X'X)^{-1}x_i x_i'(X'X)^{-1}/(1-h_{ii}) \tag{8.12}$$

Therefore

$$\widehat{\beta} - \widehat{\beta}_{(i)} = (X'X)^{-1}x_i e_i/(1-h_{ii}) \tag{8.13}$$

Since the estimated coefficients are often of primary interest, (8.13) describes the change in the estimated regression coefficients that would occur if the i-th observation is deleted. Note that a high leverage observation with h_{ii} large will be influential in (8.13) only if the corresponding residual e_i is not small. Therefore, high leverage implies a potentially influential observation, but whether this potential is actually realized depends on y_i.

Alternatively, one can obtain this result from the augmented regression given in (8.5). Note that $P_{d_i} = d_i(d_i'd_i)^{-1}d_i' = d_i d_i'$ is an $n \times n$ matrix with 1 in the i-th diagonal position and 0 elsewhere. $\bar{P}_{d_i} = I_n - P_{d_i}$, has the effect when post-multiplied by a vector y of deleting the i-th observation. Hence, premultiplying (8.5) by $\bar{P}_{d_i}$ one gets

$$\bar{P}_{d_i} y = \begin{pmatrix} y_{(i)} \\ 0 \end{pmatrix} = \begin{pmatrix} X_{(i)} \\ 0 \end{pmatrix} \beta^* + \begin{pmatrix} u_{(i)} \\ 0 \end{pmatrix} \tag{8.14}$$

where the i-th observation is moved to the bottom of the data, without loss of generality. The last observation has no effect on the least squares estimate of β^* since both the dependent and independent variables are zero. This regression will yield $\widehat{\beta}^* = \widehat{\beta}_{(i)}$, and the i-th observation's residual is clearly zero. By the Frisch-Waugh-Lovell Theorem given in section 7.3, the least squares estimates and the residuals from (8.14) are numerically identical to those from (8.5).

Therefore, $\widehat{\beta}^* = \widehat{\beta}_{(i)}$ in (8.5) and the i-th observation residual from (8.5) must be zero. This implies that $\widehat{\varphi} = y_i - x_i'\widehat{\beta}_{(i)}$, and the fitted values from this regression are given by $\widehat{y} = X\widehat{\beta}_{(i)} + d_i\widehat{\varphi}$ whereas those from the original regression (7.1) are given by $X\widehat{\beta}$. The difference in residuals is therefore

$$e - e_{(i)} = X\widehat{\beta}_{(i)} + d_i\widehat{\varphi} - X\widehat{\beta} \tag{8.15}$$

premultiplying (8.15) by $\bar{P}_X$ and using the fact that $\bar{P}_X X = 0$, one gets $\bar{P}_X(e - e_{(i)}) = \bar{P}_X d_i\widehat{\varphi}$. But, $\bar{P}_X e = e$ and $\bar{P}_X e_{(i)} = e_{(i)}$, hence $\bar{P}_X d_i\widehat{\varphi} = e - e_{(i)}$. Premultiplying both sides by d_i' one gets $d_i'\bar{P}_X d_i\widehat{\varphi} = e_i$ since the i-th residual of $e_{(i)}$ from (8.5) is zero. By definition, $d_i'\bar{P}_X d_i = 1 - h_{ii}$, therefore

$$\widehat{\varphi} = e_i/(1 - h_{ii}) \tag{8.16}$$

premultiplying (8.15) by $(X'X)^{-1}X'$ one gets $0 = \widehat{\beta}_{(i)} - \widehat{\beta} + (X'X)^{-1}X'd_i\widehat{\varphi}$. This uses the fact that both residuals are orthogonal to X. Rearranging terms and substituting $\widehat{\varphi}$ from (8.16), one gets

$$\widehat{\beta} - \widehat{\beta}_{(i)} = (X'X)^{-1}x_i\widehat{\varphi} = (X'X)^{-1}x_i e_i/(1 - h_{ii})$$

as given in (8.13).

Note that $s^2_{(i)}$ given in (8.2) can now be written in terms of $\widehat{\beta}_{(i)}$:

$$s^2_{(i)} = \textstyle\sum_{t \neq i}(y_t - x_t'\widehat{\beta}_{(i)})^2/(n - k - 1) \tag{8.17}$$

upon substituting (8.13) in (8.17) we get

$$\begin{aligned}(n-k-1)s^2_{(i)} &= \textstyle\sum_{t=1}^n \left(e_t + \dfrac{h_{it}e_i}{1 - h_{ii}}\right)^2 - \dfrac{e_i^2}{(1 - h_{ii})^2} \\ &= (n-k)s^2 + \dfrac{2e_i}{1 - h_i}\textstyle\sum_{t=1}^n e_t h_{it} + \dfrac{e_i^2}{(1 - h_{ii})^2}\textstyle\sum_{t=1}^n h_{it}^2 - \dfrac{e_i^2}{(1 - h_{ii})^2} \\ &= (n-k)s^2 - \dfrac{e_i^2}{1 - h_{ii}} \end{aligned} \tag{8.18}$$

which is (8.2). This uses the fact that $He = 0$ and $H^2 = H$. Hence, $\sum_{t=1}^n e_t h_{it} = 0$ and $\sum_{t=1}^n h_{it}^2 = h_{ii}$.

To assess whether the change in $\widehat{\beta}_j$ (the j-th component of $\widehat{\beta}$) that results from the deletion of the i-th observation, is large or small, we scale by the variance of $\widehat{\beta}_j$, $\sigma^2(X'X)^{-1}_{jj}$. This is denoted by

$$DFBETAS_{ij} = (\widehat{\beta}_j - \widehat{\beta}_{j(i)})/s_{(i)}\sqrt{(X'X)^{-1}_{jj}} \tag{8.19}$$

Note that $s_{(i)}$ is used in order to make the denominator stochastically independent of the numerator in the Gaussian case. Absolute values of *DFBETAS* larger than 2 are considered influential. However, Belsley, Kuh, and Welsch (1980) suggest $2/\sqrt{n}$ as a size-adjusted cutoff. In fact, it would be most unusual for the removal of a single observation from a sample of 100 or more to result in a change in any estimate by two or more standard errors. The size-adjusted cutoff tend to expose approximately the same proportion of potentially influential observations, regardless of sample size. The size-adjusted cutoff is particularly important for large data sets.

In case of Normality, it can also be useful to look at the change in the t-statistics, as a means of assessing the sensitivity of the regression output to the deletion of the i-th observation:

$$DFSTAT_{ij} = \frac{\widehat{\beta}_j}{s\sqrt{(X'X)^{-1}_{jj}}} - \frac{\widehat{\beta}_{j(i)}}{s_{(i)}\sqrt{(X'_{(i)}X_{(i)})^{-1}_{jj}}} \tag{8.20}$$

Another way to summarize coefficient changes and gain insight into forecasting effects when the i-th observation is deleted is to look at the change in fit, defined as

$$DFFIT_i = \widehat{y}_i - \widehat{y}_{(i)} = x_i'[\widehat{\beta} - \widehat{\beta}_{(i)}] = h_{ii}e_i/(1-h_{ii}) \tag{8.21}$$

where the last equality is obtained from (8.13).

We scale this measure by the variance of $\widehat{y}_{(i)}$, i.e., $\sigma\sqrt{h_{ii}}$, giving

$$DFFITS_i = \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} \frac{e_i}{s_{(i)}\sqrt{1-h_{ii}}} = \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} e_i^* \tag{8.22}$$

where σ has been estimated by $s_{(i)}$ and e_i^* denotes the externally studentized residual given in (8.3). Values of *DFFITS* larger than 2 in absolute value are considered influential. A size-adjusted cutoff for *DFFITS* suggested by Belsley, Kuh and Welsch (1980) is $2\sqrt{k/n}$.

In (8.3), the studentized residual e_i^* was interpreted as a t-statistic that tests for the significance of the coefficient φ of d_i, the dummy variable which takes the value 1 for the i-th observation and 0 otherwise, in the regression of y on X and d_i. This can now be easily proved as follows:

Consider the Chow test for the significance of φ. The $RRSS = (n-k)s^2$, the $URSS = (n-k-1)s^2_{(i)}$ and the Chow F-test described in (4.17) becomes

$$F_{1,n-k-1} = \frac{[(n-k)s^2 - (n-k-1)s^2_{(i)}]/1}{(n-k-1)s^2_{(i)}/(n-k-1)} = \frac{e_i^2}{s^2_{(i)}(1-h_{ii})} \tag{8.23}$$

The square root of (8.23) is $e_i^* \sim t_{n-k-1}$. These studentized residuals provide a better way to examine the information in the residuals, but they do not tell the whole story, since some of the most influential data points can have small e_i^* (and very small e_i).

One overall measure of the impact of the i-th observation on the estimated regression coefficients is Cook's (1977) distance measure D_i^2. Recall, that the confidence region for all k regression coefficients is $(\widehat{\beta}-\beta)'X'X(\widehat{\beta}-\beta)/ks^2 \sim F(k, n-k)$. Cook's (1977) distance measure D_i^2 uses the same structure for measuring the combined impact of the differences in the estimated regression coefficients when the i-th observation is deleted:

$$D_i^2(s) = (\widehat{\beta} - \widehat{\beta}_{(i)})'X'X(\widehat{\beta} - \widehat{\beta}_{(i)})/ks^2 \tag{8.24}$$

Even though $D_i^2(s)$ does not follow the above F-distribution, Cook suggests computing the percentile value from this F-distribution and declaring an influential observation if this percentile value $\geq 50\%$. In this case, the distance between $\widehat{\beta}$ and $\widehat{\beta}_{(i)}$ will be large, implying that the i-th observation has a substantial influence on the fit of the regression. Cook's distance measure can be equivalently computed as:

$$D_i^2(s) = \frac{e_i^2}{ks^2}\left(\frac{h_{ii}}{(1-h_{ii})^2}\right) \tag{8.25}$$

$D_i^2(s)$ depends on e_i and h_{ii}; the larger e_i or h_{ii} the larger is $D_i^2(s)$. Note the relationship between Cook's $D_i^2(s)$ and Belsley, Kuh, and Welsch (1980) $DFFITS_i(\sigma)$ in (8.22), i.e.,

$$DFFITS_i(\sigma) = \sqrt{k}D_i(\sigma) = (\widehat{y}_i - x_i'\widehat{\beta}_{(i)})/(\sigma\sqrt{h_{ii}})$$

Belsley, Kuh, and Welsch (1980) suggest nominating $DFFITS$ based on $s_{(i)}$ exceeding $2\sqrt{k/n}$ for special attention. Cook's 50 percentile recommendation is equivalent to $DFFITS > \sqrt{k}$, which is more conservative, see Velleman and Welsch (1981).

Next, we study the influence of the i-th observation deletion on the covariance matrix of the regression coefficients. One can compare the two covariance matrices using the ratio of their determinants:

$$COVRATIO_i = \frac{\det(s_{(i)}^2[X_{(i)}'X_{(i)}]^{-1})}{\det(s^2[X'X]^{-1})} = \frac{s_{(i)}^{2k}}{s^{2k}}\left(\frac{\det[X_{(i)}'X_{(i)}]^{-1}}{\det[X'X]^{-1}}\right) \tag{8.26}$$

Using the fact that

$$\det[X_{(i)}'X_{(i)}] = (1-h_{ii})\det[X'X] \tag{8.27}$$

see problem 8, one obtains

$$COVRATIO_i = \left(\frac{s_{(i)}^2}{s^2}\right)^k \times \frac{1}{1-h_{ii}} = \frac{1}{\left(\frac{n-k-1}{n-k} + \frac{e_i^{*2}}{n-k}\right)^k (1-h_{ii})} \tag{8.28}$$

where the last equality follows from (8.18) and the definition of e_i^* in (8.3). Values of *COVRATIO* not near unity identify possible influential observations and warrant further investigation. Belsley, Kuh and Welsch (1980) suggest investigating points with $|COVRATIO - 1|$ near to or larger than $3k/n$. The *COVRATIO* depends upon both h_{ii} and e_i^{*2}. In fact, from (8.28), *COVRATIO* is large when h_{ii} is large and small when e_i^* is large. The two factors can offset each other, that is why it is important to look at h_{ii} and e_i^* separately as well as in combination as in *COVRATIO*.

Finally, one can look at how the variance of $\widehat{y}_i$ changes when an observation is deleted.

$$\text{var}(\widehat{y}_i) = s^2 h_{ii} \qquad \text{and} \qquad \text{var}(\widehat{y}_{(i)}) = \text{var}(x_i'\widehat{\beta}_{(i)}) = s_{(i)}^2(h_{ii}/(1-h_{ii}))$$

and the ratio is

$$FVARATIO_i = s_{(i)}^2/s^2(1-h_{ii}) \tag{8.29}$$

This expression is similar to *COVRATIO* except that $[s_{(i)}^2/s^2]$ is not raised to the k-th power. As a diagnostic measure it will exhibit the same patterns of behavior with respect to different configurations of h_{ii} and the studentized residual as described for *COVRATIO*.

Table 8.1 Cigarette Regression

Dependent Variable: LNC
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	0.50098	0.25049	9.378	0.0004
Error	43	1.14854	0.02671		
C Total	45	1.64953			

Root MSE	0.16343	R-square	0.3037
Dep Mean	4.84784	Adj R-sq	0.2713
C.V.	3.37125		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	4.299662	0.90892571	4.730	0.0001
LNP	1	–1.338335	0.32460147	–4.123	0.0002
LNY	1	0.172386	0.19675440	0.876	0.3858

Example 1: For the cigarette data given in Table 3.2, Table 8.1 gives the SAS least squares regression for $\log C$ on $\log P$ and $\log Y$.

$$\underset{}{\log C} = \underset{(0.909)}{4.30} - \underset{(0.325)}{1.34}\ \log P + \underset{(0.197)}{0.172}\ \log Y + \text{residuals}$$

The standard error of the regression is $s = 0.16343$ and $\bar{R}^2 = 0.271$. Table 8.2 gives the data along with the predicted values of $\log C$, the least squares residuals e, the internal studentized residuals $\widetilde{e}$ given in (8.1), the externally studentized residuals e^* given in (8.3), the Cook statistic given in (8.25), the leverage of each observation h, the *DFFITS* given in (8.22) and the *COVRATIO* given in (8.28).

Using the leverage column, one can identify four potential observations with high leverage, i.e., greater than $2\bar{h} = 2k/n = 6/46 = 0.13043$. These are the observations belonging to the following states: Connecticut (CT), Kentucky (KY), New Hampshire (NH) and New Jersey (NJ) with leverage $0.13535, 0.19775, 0.13081$ and 0.13945, respectively. Note that the corresponding OLS residuals are $-0.078, 0.234, 0.160$ and -0.059, which are not necessarily large. The internally studentized residuals are computed using equation (8.1). For KY this gives

$$\widetilde{e}_{KY} = \frac{e_{KY}}{s\sqrt{1-h_{KY}}} = \frac{0.23428}{0.16343\sqrt{1-0.19775}} = 1.6005$$

From Table 8.2, two observations with a high internally studentized residuals are those belonging to Arkansas (AR) and Utah (UT) with values of 2.102 and -2.679 respectively, both larger than 2 in absolute value.

The externally studentized residuals are computed from (8.3). For KY, we first compute $s^2_{(KY)}$, the MSE from the regression computed without the KY observation. From (8.2), this is given by

$$s^2_{(KY)} = \frac{(n-k)s^2 - e^2_{KY}/(1-h_{KY})}{(n-k-1)}$$
$$= \frac{(46-3)(0.16343)^2 - (0.23428)^2/(1-0.19775)}{(46-3-1)} = 0.025716$$

From (8.3) we get

$$e^*_{(KY)} = \frac{e_{KY}}{s_{(KY)}\sqrt{1-h_{KY}}} = \frac{0.23428}{0.16036\sqrt{1-0.19775}} = 1.6311$$

This externally studentized residual is distributed as a t-statistic with 42 degrees of freedom. However, e^*_{KY} does not exceed 2 in absolute value. Again, e^*_{AR} and e^*_{UT} are 2.193 and −2.901 both larger than 2 in absolute value. From (8.13), the change in the regression coefficients due to the omission of the KY observation is given by

$$\widehat{\beta} - \widehat{\beta}_{(KY)} = (X'X)^{-1}x_{KY}e_{KY}/(1-h_{KY})$$

Using the fact that

$$(X'X)^{-1} = \begin{bmatrix} 30.929816904 & 4.8110214655 & -6.679318415 \\ 4.81102114655 & 3.9447686638 & -1.177208398 \\ -6.679318415 & -1.177208398 & 1.4493372835 \end{bmatrix}$$

and $x'_{KY} = (1, -0.03260, 4.64937)$ with $e_{KY} = 0.23428$ and $h_{KY} = 0.19775$ one gets

$$(\widehat{\beta} - \widehat{\beta}_{(KY)})' = (-0.082249, -0.230954, 0.028492)$$

In order to assess whether this change is large or small, we compute DFBETAS given in (8.19). For the KY observation, these are given by

$$DFBETAS_{KY,1} = \frac{\widehat{\beta}_1 - \widehat{\beta}_{1(KY)}}{s_{(KY)}\sqrt{(X'X)^{-1}_{11}}} = \frac{-0.082449}{0.16036\sqrt{30.9298169}} = -0.09222$$

Similarly, $DFBETAS_{KY,2} = -0.7251$ and $DFBETAS_{KY,3} = 0.14758$. These are not larger than 2 in absolute value. However, $DFBETAS_{KY,2}$ is larger than $2/\sqrt{n} = 2/\sqrt{46} = 0.2949$ in absolute value. This is the size-adjusted cutoff recommended by Belsley, Kuh and Welsch (1980) for large n.

The change in the fit due to the omission of the KY observation is given by (8.21). In fact,

$$DFFIT_{KY} = \widehat{y}_{KY} - \widehat{y}_{(KY)} = x'_{KY}[\widehat{\beta} - \widehat{\beta}_{(KY)}]$$
$$= (1, -0.03260, 4.64937)\begin{pmatrix} -0.082249 \\ -0.230954 \\ -0.028492 \end{pmatrix} = 0.05775$$

or simply

$$DFFIT_{KY} = \frac{h_{KY}e_{KY}}{(1-h_{KY})} = \frac{(0.19775)(0.23428)}{1-0.19775} = 0.05775$$

Scaling it by the variance of $\widehat{y}_{(KY)}$ we get from (8.22)

$$DFFITS_{KY} = \left(\frac{h_{KY}}{1-h_{KY}}\right)^{1/2} e^*_{KY} = \left(\frac{0.19775}{1-0.19775}\right)^{1/2} (1.6311) = 0.8098$$

This is not larger than 2 in absolute value, but it is larger than the size-adjusted cutoff of $2\sqrt{k/n} = 2\sqrt{3/46} = 0.511$. Note also that both $DFFITS_{AR} = 0.667$ and $DFFITS_{UT} = -0.888$ are larger than 0.511 in absolute value.

Cook's distance measure is given in (8.25) and for KY can be computed as

$$D^2_{KY}(s) = \frac{e^2_{KY}}{ks^2}\left(\frac{h_{KY}}{(1-h_{KY})^2}\right) = \left(\frac{(0.23428)^2}{3(0.16343)^2}\right)\left(\frac{0.19775}{(1-0.19775)^2}\right) = 0.21046$$

The other two large Cook's distance measures are $DR^2_{AR}(s) = 0.13623$ and $D^2_{UT}(s) = 0.22399$, respectively. $COVRATIO$ omitting the KY observation can be computed from (8.28) as

$$COVRATIO_{KY} = \left(\frac{s^2_{(KY)}}{s^2}\right)^k \frac{1}{1-h_{KY}} = \left(\frac{0.025716}{(0.16343)^2}\right)^3 \left(\frac{1}{(1-0.019775)}\right) = 1.1125$$

which means that $COVRATIO_{KY} - 1/ = 0.1125$ is less than $3k/n = 9/46 = 0.1956$.

Finally, $FVARATIO$ omitting the KY observation can be computed from (8.29) as

$$FVARATIO_{KY} = \frac{s^2_{(KY)}}{s^2(1-h_{KY})} = \frac{0.025716}{(0.16343)^2(1-0.19775)} = 1.2001$$

By several diagnostic measures, AR, KY and UT are influential observations that deserve special attention. The first two states are characterized with large sales of cigarettes. KY is a producer state with a very low price on cigarettes, while UT is a low consumption state due to its high percentage of Mormon population (a religion that forbids smoking). Table 8.3 gives the predicted consumption along with the 95% confidence band, the OLS residuals, and the internalized student residuals, Cook's D-statistic and a plot of these residuals. This last plot highlights the fact that AR, UT and KY have large studentized residuals.

8.2 Recursive Residuals

In Section 8.1, we showed that the least squares residuals are heteroskedastic with non-zero covariances, even when the true disturbances have a scalar covariance matrix. This section studies recursive residuals which are a set of linear unbiased residuals with a scalar covariance matrix. They are independent and identically distributed when the true disturbances themselves are independent and identically distributed.[2] These residuals are natural in time-series regressions and can be constructed as follows:

1. Choose the first $t \geq k$ observations and compute $\widehat{\beta}_t = (X'_tX_t)^{-1}X'_tY_t$ where X_t denotes the $t \times k$ matrix of t observations on k variables and $Y'_t = (y_1, \ldots, y_t)$. The recursive residuals are basically standardized one-step ahead forecast residuals:

$$w_{t+1} = (y_{t+1} - x'_{t+1}\widehat{\beta}_t)/\sqrt{1 + x'_{t+1}(X'_tX_t)^{-1}x_{t+1}} \tag{8.30}$$

Table 8.2 Diagnostic Statistics for the Cigarettes Example

OBS	STATE	LNC	LNP	LNY	PREDICTED	e	$\tilde{e}$	$e^{\star}$	Cook's D	Leverage	DFFITS	COVRATIO
1	AL	4.96213	0.20487	4.64039	4.8254	0.1367	0.857	0.8546	0.012	0.0480	0.1919	1.0704
2	AZ	4.66312	0.16640	4.68389	4.8844	–0.2213	–1.376	–1.3906	0.021	0.0315	–0.2508	0.9681
3	AR	5.10709	0.23406	4.59435	4.7784	0.3287	2.102	2.1932	0.136	0.0847	0.6670	0.8469
4	CA	4.50449	0.36399	4.88147	4.6540	–0.1495	–0.963	–0.9623	0.033	0.0975	–0.3164	1.1138
5	CT	4.66983	0.32149	5.09472	4.7477	–0.0778	–0.512	–0.5077	0.014	0.1354	–0.2009	1.2186
6	DE	5.04705	0.21929	4.87087	4.8458	0.2012	1.252	1.2602	0.018	0.0326	0.2313	0.9924
7	DC	4.65637	0.28946	5.05960	4.7845	–0.1281	–0.831	–0.8280	0.029	0.1104	–0.2917	1.1491
8	FL	4.80081	0.28733	4.81155	4.7446	0.0562	0.352	0.3482	0.002	0.0431	0.0739	1.1118
9	GA	4.97974	0.12826	4.73299	4.9439	0.0358	0.224	0.2213	0.001	0.0402	0.0453	1.1142
10	ID	4.74902	0.17541	4.64307	4.8653	–0.1163	–0.727	–0.7226	0.008	0.0413	–0.1500	1.0787
11	IL	4.81445	0.24806	4.90387	4.8130	0.0014	0.009	0.0087	0.000	0.0399	0.0018	1.1178
12	IN	5.11129	0.08992	4.72916	4.9946	0.1167	0.739	0.7347	0.013	0.0650	0.1936	1.1046
13	IA	4.80857	0.24081	4.74211	4.7949	0.0137	0.085	0.0843	0.000	0.0310	0.0151	1.1070
14	KS	4.79263	0.21642	4.79613	4.8368	–0.0442	–0.273	–0.2704	0.001	0.0223	–0.0408	1.0919
15	KY	5.37906	–0.03260	4.64937	5.1448	0.2343	1.600	1.6311	0.210	0.1977	0.8098	1.1126
16	LA	4.98602	0.23856	4.61461	4.7759	0.2101	1.338	1.3504	0.049	0.0761	0.3875	1.0224
17	ME	4.98722	0.29106	4.75501	4.7298	0.2574	1.620	1.6527	0.051	0.0553	0.4000	0.9403
18	MD	4.77751	0.12575	4.94692	4.9841	–0.2066	–1.349	–1.3624	0.084	0.1216	–0.5070	1.0731
19	MA	4.73877	0.22613	4.99998	4.8590	–0.1202	–0.769	–0.7653	0.018	0.0856	–0.2341	1.1258
20	MI	4.94744	0.23067	4.80620	4.8195	0.1280	0.792	0.7890	0.005	0.0238	0.1232	1.0518
21	MN	4.69589	0.34297	4.81207	4.6702	0.0257	0.165	0.1627	0.001	0.0864	0.0500	1.1724
22	MS	4.93990	0.13638	4.52938	4.8979	0.0420	0.269	0.2660	0.002	0.0883	0.0828	1.1712
23	MO	5.06430	0.08731	4.78189	5.0071	0.0572	0.364	0.3607	0.004	0.0787	0.1054	1.1541
24	MT	4.73313	0.15303	4.70417	4.9058	–0.1727	–1.073	–1.0753	0.012	0.0312	–0.1928	1.0210
25	NE	4.77558	0.18907	4.79671	4.8735	–0.0979	–0.607	–0.6021	0.003	0.0243	–0.0950	1.0719
26	NV	4.96642	0.32304	4.83816	4.7014	0.2651	1.677	1.7143	0.065	0.0646	0.4504	0.9366
27	NH	5.10990	0.15852	5.00319	4.9500	0.1599	1.050	1.0508	0.055	0.1308	0.4076	1.1422
28	NJ	4.70633	0.30901	5.10268	4.7657	–0.0594	–0.392	–0.3879	0.008	0.1394	–0.1562	1.2337
29	NM	4.58107	0.16458	4.58202	4.8693	–0.2882	–1.823	–1.8752	0.076	0.0639	–0.4901	0.9007
30	NY	4.66496	0.34701	4.96075	4.6904	–0.0254	–0.163	–0.1613	0.001	0.0888	–0.0503	1.1755
31	ND	4.58237	0.18197	4.69163	4.8649	–0.2825	–1.755	–1.7999	0.031	0.0295	–0.3136	0.8848
32	OH	4.97952	0.12889	4.75875	4.9475	0.0320	0.200	0.1979	0.001	0.0423	0.0416	1.1174
33	OK	4.72720	0.19554	4.62730	4.8356	–0.1084	–0.681	–0.6766	0.008	0.0505	–0.1560	1.0940
34	PA	4.80363	0.22784	4.83516	4.8282	–0.0246	–0.153	–0.1509	0.000	0.0257	–0.0245	1.0997
35	RI	4.84693	0.30324	4.84670	4.7293	0.1176	0.738	0.7344	0.010	0.0504	0.1692	1.0876
36	SC	5.07801	0.07944	4.62549	4.9907	0.0873	0.555	0.5501	0.008	0.0725	0.1538	1.1324
37	SD	4.81545	0.13139	4.67747	4.9301	–0.1147	–0.716	–0.7122	0.007	0.0402	–0.1458	1.0786
38	TN	5.04939	0.15547	4.72525	4.9062	0.1432	0.890	0.8874	0.008	0.0294	0.1543	1.0457
39	TX	4.65398	0.28196	4.73437	4.7384	–0.0845	–0.532	–0.5271	0.005	0.0546	–0.1267	1.1129
40	UT	4.40859	0.19260	4.55586	4.8273	–0.4187	–2.679	–2.9008	0.224	0.0856	–0.8876	0.6786
41	VT	5.08799	0.18018	4.77578	4.8818	0.2062	1.277	1.2869	0.014	0.0243	0.2031	0.9794
42	VA	4.93065	0.11818	4.85490	4.9784	–0.0478	–0.304	–0.3010	0.003	0.0773	–0.0871	1.1556
43	WA	4.66134	0.35053	4.85645	4.6677	–0.0064	–0.041	–0.0404	0.000	0.0866	–0.0124	1.1747
44	WV	4.82454	0.12008	4.56859	4.9265	–0.1020	–0.647	–0.6429	0.011	0.0709	–0.1777	1.1216
45	WI	4.83026	0.22954	4.75826	4.8127	0.0175	0.109	0.1075	0.000	0.0254	0.0174	1.1002
46	WY	5.00087	0.10029	4.71169	4.9777	0.0232	0.146	0.1444	0.000	0.0555	0.0350	1.1345

Table 8.3 Regression of Real Per-Capita Consumption of Cigarettes

Dep Obs	Var LNC	Predict Value	Std Err Predict	Lower95% Mean	Upper95% Mean	Lower95% Predict	Upper95% Predict	Std Err Residual	Student Residual	Residual	−2 −1 0 1 2	Cook's D
1	4.9621	4.8254	0.036	4.7532	4.8976	4.4880	5.1628	0.1367	0.159	0.857	*	0.012
2	4.6631	4.8844	0.029	4.8259	4.9429	4.5497	5.2191	−0.2213	0.161	−1.376	**	0.021
3	5.1071	4.7784	0.048	4.6825	4.8743	4.4351	5.1217	0.3287	0.156	2.102	****	0.136
4	4.5045	4.6540	0.051	4.5511	4.7570	4.3087	4.9993	−0.1495	0.155	−0.963	*	0.033
5	4.6698	4.7477	0.060	4.6264	4.8689	4.3965	5.0989	−0.0778	0.152	−0.512	*	0.014
6	5.0471	4.8458	0.030	4.7863	4.9053	4.5109	5.1808	0.2012	0.161	1.252	**	0.018
7	4.6564	4.7845	0.054	4.6750	4.8940	4.4372	5.1318	−0.1281	0.154	−0.831	*	0.029
8	4.8008	4.7446	0.034	4.6761	4.8130	4.4079	5.0812	0.0562	0.160	0.352		0.002
9	4.9797	4.9439	0.033	4.8778	5.0100	4.6078	5.2801	0.0358	0.160	0.224		0.001
10	4.7490	4.8653	0.033	4.7983	4.9323	4.5290	5.2016	−0.1163	0.160	−0.727	*	0.008
11	4.8145	4.8130	0.033	4.7472	4.8789	4.4769	5.1491	0.00142	0.160	0.009		0.000
12	5.1113	4.9946	0.042	4.9106	5.0786	4.6544	5.3347	0.1167	0.158	0.739	*	0.013
13	4.8086	4.7949	0.029	4.7368	4.8529	4.4602	5.1295	0.0137	0.161	0.085		0.000
14	4.7926	4.8368	0.024	4.7876	4.8860	4.5036	5.1701	−0.0442	0.162	−0.273		0.001
15	5.3791	5.1448	0.073	4.9982	5.2913	4.7841	5.5055	0.2343	0.146	1.600	***	0.210
16	4.9860	4.7759	0.045	4.6850	4.8668	4.4340	5.1178	0.2101	0.157	1.338	**	0.049
17	4.9872	4.7298	0.038	4.6523	4.8074	4.3912	5.0684	0.2574	0.159	1.620	***	0.051
18	4.7775	4.9841	0.057	4.8692	5.0991	4.6351	5.3332	−0.2066	0.153	−1.349	**	0.084
19	4.7388	4.8590	0.048	4.7625	4.9554	4.5155	5.2024	−0.1202	0.156	−0.769	*	0.018
20	4.9474	4.8195	0.025	4.7686	4.8703	4.4860	5.1530	0.1280	0.161	0.792	*	0.005
21	4.6959	4.6702	0.048	4.5733	4.7671	4.3267	5.0137	0.0257	0.156	0.165		0.001
22	4.9399	4.8979	0.049	4.8000	4.9959	4.5541	5.2418	0.0420	0.156	0.269		0.002
23	5.0643	5.0071	0.046	4.9147	5.0996	4.6648	5.3495	0.0572	0.157	0.364		0.004
24	4.7331	4.9058	0.029	4.8476	4.9640	4.5711	5.2405	−0.1727	0.161	−1.073	**	0.012
25	4.7756	4.8735	0.025	4.8221	4.9249	4.5399	5.2071	−0.0979	0.161	−0.607	*	0.003
26	4.9664	4.7014	0.042	4.6176	4.7851	4.3613	5.0414	0.2651	0.158	1.677	***	0.065
27	5.1099	4.9500	0.059	4.8308	5.0692	4.5995	5.3005	0.1599	0.152	1.050	**	0.055
28	4.7063	4.7657	0.061	4.6427	4.8888	4.4139	5.1176	−0.0594	0.152	−0.392		0.008
29	4.5811	4.8693	0.041	4.7859	4.9526	4.5293	5.2092	−0.2882	0.158	−1.823	***	0.076
30	4.6650	4.6904	0.049	4.5922	4.7886	4.3465	5.0343	−0.0254	0.156	−0.163		0.001
31	4.5824	4.8649	0.028	4.8083	4.9215	4.5305	5.1993	−0.2825	0.161	−1.755	***	0.031
32	4.9795	4.9475	0.034	4.8797	5.0153	4.6110	5.2840	0.0320	0.160	0.200		0.001
33	4.7272	4.8356	0.037	4.7616	4.9097	4.4978	5.1735	−0.1084	0.159	−0.681	*	0.008
34	4.8036	4.8282	0.026	4.7754	4.8811	4.4944	5.1621	−0.0246	0.161	−0.153		0.000
35	4.8469	4.7293	0.037	4.6553	4.8033	4.3915	5.0671	0.1176	0.159	0.738	*	0.010
36	5.0780	4.9907	0.044	4.9020	5.0795	4.6494	5.3320	0.0873	0.157	0.555	*	0.008
37	4.8155	4.9301	0.033	4.8640	4.9963	4.5940	5.2663	−0.1147	0.160	−0.716	*	0.007
38	5.0494	4.9062	0.028	4.8497	4.9626	4.5718	5.2406	0.1432	0.161	0.890	*	0.008
39	4.6540	4.7384	0.038	4.6614	4.8155	4.4000	5.0769	−0.0845	0.159	−0.532	*	0.005
40	4.4086	4.8273	0.048	4.7308	4.9237	4.4839	5.1707	−0.4187	0.156	−2.679	*****	0.224
41	5.0880	4.8818	0.025	4.8304	4.9332	4.5482	5.2154	0.2062	0.161	1.277	**	0.014
42	4.9307	4.9784	0.045	4.8868	5.0701	4.6363	5.3205	−0.0478	0.157	−0.304		0.003
43	4.6613	4.6677	0.048	4.5708	4.7647	4.3242	5.0113	−0.00638	0.156	−0.041		0.000
44	4.8245	4.9265	0.044	4.8387	5.0143	4.5854	5.2676	−0.1020	0.158	−0.647	*	0.011
45	4.8303	4.8127	0.026	4.7602	4.8653	4.4790	5.1465	0.0175	0.161	0.109		0.000
46	5.0009	4.9777	0.039	4.9000	5.0553	4.6391	5.3163	0.0232	0.159	0.146		0.000

Sum of Residuals	0
Sum of Squared Residuals	1.1485
Predicted Resid SS (Press)	1.3406

2. Add the $(t+1)$-th observation to the data and obtain $\widehat{\beta}_{t+1} = (X'_{t+1}X_{t+1})^{-1}X'_{t+1}Y_{t+1}$. Compute w_{t+2}.

3. Repeat step 2, adding one observation at a time. In time-series regressions, one usually starts with the first k-observations and obtain $(T-k)$ forward recursive residuals. These recursive residuals can be computed using the updating formula given in (8.11) with $A = (X'_tX_t)$ and $a = -b = x'_{t+1}$. Therefore,

$$(X'_{t+1}X_{t+1})^{-1} = (X'_tX_t)^{-1} - (X'_tX_t)^{-1}x_{t+1}x'_{t+1}(X'_tX_t)^{-1}/[1+x'_{t+1}(X'_tX_t)^{-1}x_{t+1}] \quad (8.31)$$

and only $(X'_tX_t)^{-1}$ have to be computed. Also,

$$\widehat{\beta}_{t+1} = \widehat{\beta}_t + (X'_tX_t)^{-1}x_{t+1}(y_{t+1} - x'_{t+1}\widehat{\beta}_t)/f_{t+1} \quad (8.32)$$

where $f_{t+1} = 1 + x'_{t+1}(X'_tX_t)^{-1}x_{t+1}$, see problem 13.

Alternatively, one can compute these residuals by regressing Y_{t+1} on X_{t+1} and d_{t+1} where $d_{t+1} = 1$ for the $(t+1)$-th observation, and zero otherwise, see equation (8.5). The estimated coefficient of d_{t+1} is the numerator of w_{t+1}. The standard error of this estimate is s_{t+1} times the denominator of w_{t+1}, where s_{t+1} is the standard error of this regression. Hence, w_{t+1} can be retrieved as s_{t+1} multiplied by the t-statistic corresponding to d_{t+1}. This computation has to be performed sequentially, in each case generating the corresponding recursive residual. This may be computationally inefficient, but it is simple to generate using regression packages.

It is obvious from (8.30) that if $u_t \sim \text{IIN}(0, \sigma^2)$, then w_{t+1} has zero mean and $\text{var}(w_{t+1}) = \sigma^2$. Furthermore, w_{t+1} is linear in the y's. Therefore, it is normally distributed. It remains to show that the recursive residuals are independent. Given normality, it is sufficient to show that

$$\text{cov}(w_{t+1}, w_{s+1}) = 0 \qquad \text{for} \qquad t \neq s;\ t, s = k, \ldots, T-1 \quad (8.33)$$

This is left as an exercise for the reader, see problem 13.

Alternatively, one can express the $T-k$ vector of recursive residuals as $w = Cy$ where C is of dimension $(T-k) \times T$ as follows:

$$C = \begin{bmatrix} -\frac{x'_{k+1}(X'_kX_k)^{-1}X'_k}{\sqrt{f_{k+1}}} & \frac{1}{\sqrt{f_{k+1}}} & & & 0....0 \\ \vdots & & \ddots & & \\ -\frac{x'_t(X'_{t-1}X_{t-1})^{-1}X'_{t-1}}{\sqrt{f_t}} & & & \frac{1}{\sqrt{f_t}} & 0....0 \\ \vdots & & & & \ddots \\ -\frac{x'_T(X'_{T-1}X_{T-1})^{-1}X'_{T-1}}{\sqrt{f_T}} & & & & \frac{1}{\sqrt{f_T}} \end{bmatrix} \quad (8.34)$$

Problem 14 asks the reader to verify that $w = Cy$, using (8.30). Also, that the matrix C satisfies the following properties:

$$(i)\ CX = 0 \qquad (ii)\ CC' = I_{T-k} \qquad (iii)\ C'C = \bar{P}_X \quad (8.35)$$

This means that the recursive residuals w are (LUS) linear in y, unbiased with mean zero and have a scalar variance-covariance matrix: $\text{var}(w) = CE(uu')C' = \sigma^2 I_{T-k}$. Property (iii) also means that $w'w = y'C'Cy = y'\bar{P}_X y = e'e$. This means that the sum of squares of $(T-k)$ recursive residuals is equal to the sum of squares of T least squares residuals. One can also show from (8.32) that

$$RSS_{t+1} = RSS_t + w_{t+1}^2 \qquad \text{for} \qquad t = k, \ldots, T-1 \tag{8.36}$$

where $RSS_t = (Y_t - X_t\widehat{\beta}_t)'(Y_t - X_t\widehat{\beta}_t)$, see problem 14. Note that for $t = k$; $RSS = 0$, since with k observations one gets a perfect fit and zero residuals. Therefore

$$RSS_T = \sum_{t=k+1}^{T} w_t^2 = \sum_{t=1}^{T} e_t^2 \tag{8.37}$$

Applications of Recursive Residuals

Recursive residuals have been used in several important applications:

(1) Harvey (1976) used these recursive residuals to give an alternative proof of the fact that Chow's post-sample *predictive test* has an F-distribution. Recall, from Chapter 7, that when the second sample n_2 had fewer than k observations, Chow's test becomes

$$F = \frac{(e'e - e_1'e_1)/n_2}{e_1'e_1/(n_1-k)} \sim F(n_2, n_1-k) \tag{8.38}$$

where $e'e = RSS$ from the total sample ($n_1 + n_2 = T$ observations), and $e_1'e_1 = RSS$ from the first n_1 observations. Recursive residuals can be computed for $t = k+1, \ldots, n_1$, and continued on for the extra n_2 observations. From (8.36) we have

$$e'e = \sum_{t=k+1}^{n_1+n_2} w_t^2 \qquad \text{and} \qquad e_1'e_1 = \sum_{t=k+1}^{n_1} w_t^2 \tag{8.39}$$

Therefore,

$$F = \frac{\sum_{t=n_1+1}^{n_1+n_2} w_t^2/n_2}{\sum_{t=k+1}^{n_1} w_t^2/(n_1-k)} \tag{8.40}$$

But the w_t's are $\sim$IIN$(0, \sigma^2)$ under the null, therefore the F-statistic in (8.38) is a ratio of two independent chi-squared variables, each divided by the appropriate degrees of freedom. Hence, $F \sim F(n_2, n_1-k)$ under the null, see Chapter 2.

(2) Harvey and Phillips (1974) used recursive residuals to test the null hypothesis of homoskedasticity. If the alternative hypothesis is that σ_i^2 varies with X_j, the proposed test is as follows:

1) Order the data according to X_j and choose a base of at least k observations from among the central observations.

2) From the first m observations compute the vector of recursive residuals w_1 using the base constructed in step 1. Also, compute the vector of recursive residuals w_2 from the last m observations. The maximum m can be is $(T-k)/2$.

3) Under the null hypothesis, it follows that

$$F = w_2'w_2/w_1'w_1 \sim F_{m,m} \tag{8.41}$$

Harvey and Phillips suggest setting m at approximately $(n/3)$ provided $n > 3k$. This test has the advantage over the Goldfeld-Quandt test in that if one wanted to test whether σ_i^2 varies with some other variable X_s, one could simply regroup the existing recursive residuals according to low and high values of X_s and compute (8.41) afresh, whereas the Goldfeld-Quandt test would require the computation of two new regressions.

(3) Phillips and Harvey (1974) suggest using the recursive residuals to test the null hypothesis of no serial correlation using a modified von Neuman ratio:

$$MVNR = \frac{\sum_{t=k+2}^{T}(w_t - w_{t-1})^2/(T-k-1)}{\sum_{t=k+1}^{T} w_t^2/(T-k)} \tag{8.42}$$

This is the ratio of the mean-square successive difference to the variance. It is arithmetically closely related to the DW statistic, but given that $w \sim N(0, \sigma^2 I_{T-k})$ one has an exact test available and no inconclusive regions. Phillips and Harvey (1974) provide tabulations of the significance points. If the sample size is large, a satisfactory approximation is obtained from a normal distribution with mean 2 and variance $4/(T-k)$.

(4) Harvey and Collier (1977) suggest a test for functional misspecification based on recursive residuals. This is based on the fact that $w \sim N(0, \sigma^2 I_{T-k})$. Therefore,

$$\bar{w}/(s_w/\sqrt{T-k}) \sim t_{T-k-1} \tag{8.43}$$

where $\bar{w} = \sum_{t=k+1}^{T} w_t/(T-k)$ and $s_w^2 = \sum_{t=k+1}^{T}(w_t - \bar{w})^2/(T-k-1)$. Suppose that the true functional form relating y to a single explanatory variable X is concave (convex) and the data are ordered by X. A simple linear regression is estimated by regressing y on X. The recursive residuals would be expected to be mainly negative (positive) and the computed t-statistic will be large in absolute value. When there are multiple X's, one could carry out this test based on any single explanatory variable. Since several specification errors might have a self-cancelling effect on the recursive residuals, this test is not likely to be very effective in multivariate situations. Wu (1993) suggested performing this test using the following augmented regression:

$$y = X\beta + z\gamma + v \tag{8.44}$$

where $z = C'\iota_{T-k}$ is one additional regressor with C defined in (8.34) and ι_{T-k} denoting a vector of ones of dimension $T-k$. In fact, the F-statistic for testing H_0; $\gamma = 0$ turns out to be the square of the Harvey and Collier (1977) t-statistic given in (8.43), see problem 15.

Alternatively, a Sign test may be used to test the null hypothesis of no functional misspecification. Under the null hypothesis, the expected number of positive recursive residuals is equal to $(T-k)/2$. A critical region may therefore be constructed from the binomial distribution. However, Harvey and Collier (1977) suggest that the Sign test tends to lack power compared with the t-test described in (8.43). Nevertheless, it is very simple and it may be more robust to non-normality.

(5) Brown, Durbin and Evans (1975) used recursive residuals to test for structural change over time. The null hypothesis is

$$H_0;\begin{cases} \beta_1 = \beta_2 = .. = \beta_T = \beta \\ \sigma_1^2 = \sigma_2^2 = .. = \sigma_T^2 = \sigma^2 \end{cases} \tag{8.45}$$

where β_t is the vector of coefficients in period t and σ_t^2 is the disturbance variance for that period. The authors suggest a pair of tests. The first is the CUSUM test which computes

$$W_r = \sum_{t=k+1}^{r} w_t/s_w \quad \text{for } r = k+1, \ldots, T \tag{8.46}$$

where s_w^2 is an estimate of the variance of the w_t's, given below (8.43). W_r is a cumulative sum and should be plotted against r. Under the null, $E(W_r) = 0$. But, if there is a structural break, W_r will tend to diverge from the horizontal line. The authors suggest checking whether W_r cross a pair of straight lines (see Figure 8.1) which pass through the points $\{k, \pm a\sqrt{T-k}\}$ and $\{T, \pm 3a\sqrt{T-k}\}$ where a depends upon the chosen significance level α. For example, $a = 0.850, 0.948$, and 1.143 for $\alpha = 10\%, 5\%$, and 1% levels, respectively.

If the coefficients are not constant, there may be a tendency for a disproportionate number of recursive residuals to have the same sign and to push W_r across the boundary. The second test is the cumulative sum of squares (CUSUMSQ) which is based on plotting

$$W_r^* = \sum_{t=k+1}^{r} w_t^2 / \sum_{t=k+1}^{T} w_t^2 \quad \text{for } t = k+1, \ldots, T \tag{8.47}$$

against r. Under the null, $E(W_r^*) = (r-k)/(T-k)$ which varies from 0 for $r = k$ to 1 for $r = T$. The significance of the departure of W_r^* from its expected value is assessed by whether W_r^* crosses a pair of lines parallel to $E(W_r^*)$ at a distance c_o above and below this line. Brown, Durbin and Evans (1975) provide values of c_0 for various sample sizes T and levels of significance α.

The CUSUM and CUSUMSQ should be regarded as *data analytic* techniques; i.e., the value of the plots lie in the information to be gained simply by inspecting them. The plots contain more information than can be summarized in a single test statistic. The significance lines constructed are, to paraphrase the authors, best regarded as 'yardsticks' against which to assess the observed plots rather than as formal tests of significance. See Brown et al. (1975) for various examples. Note that the CUSUM and CUSUMSQ are quite general tests for structural change in that they do not require a prior determination of where the structural break takes place. If this is known, the Chow-test will be more powerful. But, if this break is not known, the CUSUM and CUSUMSQ are more appropriate.

Example 2: Table 8.5 reproduces the consumption-income data, over the period 1959–2007, taken from the Economic Report of the President. In addition, the recursive residuals are computed as in (8.30) and exhibited in column 5, starting with 1961 and ending in 2007. The CUSUM given by W_r in (8.46) is plotted against r in Figure 8.2. The CUSUM crosses the upper 5% line in 1998, showing structural instability in the latter years. This was done using EViews 6.

The post-sample predictive test for 1998, can be obtained from (8.38) by computing the RSS from 1950–1997 and comparing it with the RSS from 1950-2007. The observed F-statistic is 5.748 which is distributed as $F(10, 37)$. Using EViews, one clicks on *stability diagnostics* and then selects *Chow forecast test*. You will be prompted to enter the break point period which

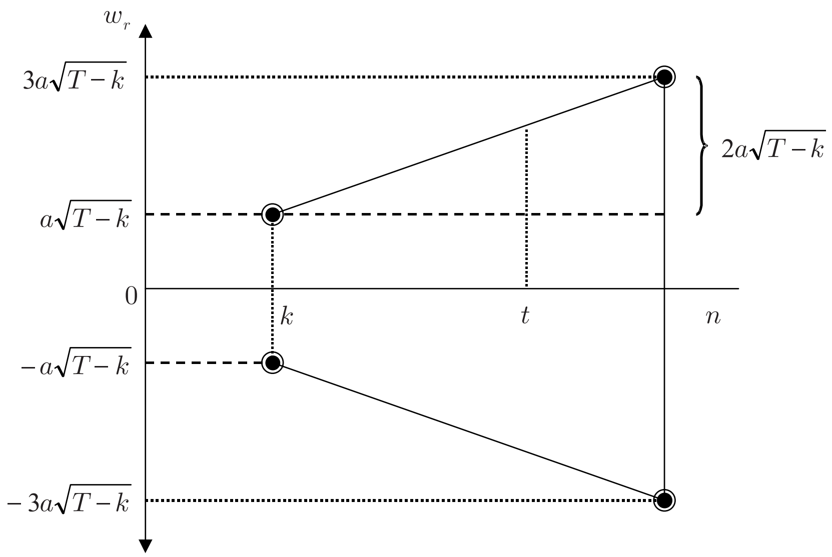

Figure 8.1 CUSUM Critical Values

Table 8.4 Chow Forecast Test

Specification: CONSUM C Y
Test predictions for observations from 1998 to 2007

	Value	df	Probability
F-statistic	5.747855	(10,37)	0.0000
Likelihood ratio	45.93529	10	0.0000
F-test summary:			
	Sum of Sq.	df	Mean Squares
Test SSR	5476210.	10	547621.0
Restricted SSR	9001348.	47	191518.0
Unrestricted SSR	3525138.	37	95273.99
LR test summary:			
	Value	df	
Restricted LogL	–366.4941	47	
Unrestricted LogL	–343.5264	37	

in this case is 1998. EViews gives the back up regression which is not shown here, and also performs a likelihood ratio test, see Table 8.4.

The reader can verify that the same F-statistic can be obtained from (8.40) using the recursive residuals in Table 8.5. In fact,

$$F = (\sum_{t=1998}^{2007} w_t^2/10)/(\sum_{t=1961}^{1997} w_t^2/37) = 5.748$$

Table 8.5 Recursive Residuals for the Consumption Regression

Year	CONSUM	Income	RESID	Recursive RES
1959	8776	9685	635.4909	NA
1960	8837	9735	647.5295	NA
1961	8873	9901	520.9776	–30.06109
1962	9170	10227	498.7493	53.63333
1963	9412	10455	517.4853	57.07454
1964	9839	11061	351.0732	–14.42043
1965	10331	11594	321.1447	40.23840
1966	10793	12065	321.9283	72.59054
1967	10994	12457	139.0709	–58.72718
1968	11510	12892	229.1068	88.63871
1969	11820	13163	273.7360	125.0883
1970	11955	13563	17.04481	–88.54736
1971	12256	14001	–110.8570	–123.0740
1972	12868	14512	0.757470	68.23355
1973	13371	15345	–311.9394	–118.2972
1974	13148	15094	–289.1532	–100.8288
1975	13320	15291	–310.0611	–72.86693
1976	13919	15738	–148.7760	148.9270
1977	14364	16128	–85.67493	231.2810
1978	14837	16704	–176.7102	178.9840
1979	15030	16931	–205.9950	147.8067
1980	14816	16940	–428.8080	–80.37207
1981	14879	17217	–637.0542	–229.1660
1982	14944	17418	–768.8790	–296.0910
1983	15656	17828	–458.3625	86.49899
1984	16343	19011	–929.7892	–205.6594
1985	17040	19476	–688.1302	111.3357
1986	17570	19906	–579.1982	251.5306
1987	17994	20072	–317.7500	479.8759
1988	18554	20740	–411.8743	405.8181
1989	18898	21120	–439.9809	366.8060
1990	19067	21281	–428.6367	347.8156
1991	18848	21109	–479.2094	243.0261
1992	19208	21548	–549.0905	195.0177
1993	19593	21493	–110.2330	588.3097
1994	20082	21812	66.39330	731.2551
1995	20382	22153	32.47656	660.7508
1996	20835	22546	100.6400	696.2055
1997	21365	23065	122.4207	689.6197
1998	22183	24131	–103.4364	474.3981
1999	23050	24564	339.5579	870.8977
2000	23862	25472	262.4189	751.2861
2001	24215	25697	395.0926	808.6041
2002	24632	26238	282.3303	639.0555
2003	25073	26566	402.1435	700.0686
2004	25750	27274	385.8501	633.1310
2005	26290	27403	799.5297	970.8717
2006	26835	28098	663.9663	760.6385
2007	27319	28614	642.6847	673.7335

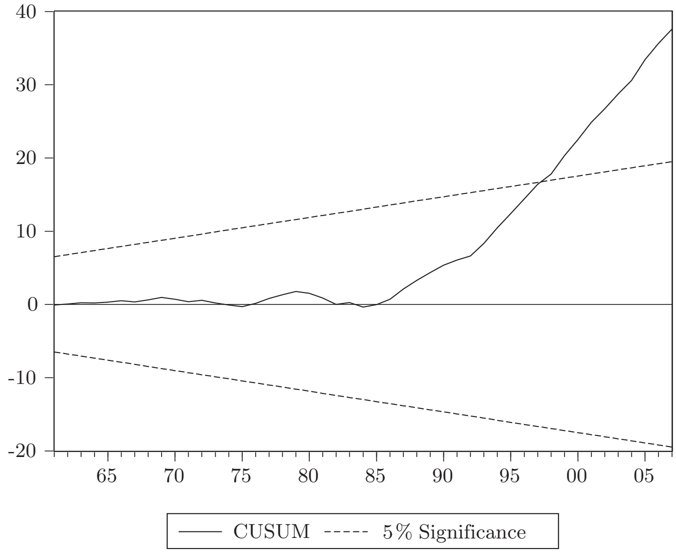

Figure 8.2 CUSUM Plot of the Consumption Regression

8.3 Specification Tests

Specification tests are an important part of model specification in econometrics. In this section, we only study a few of these diagnostic tests. For an excellent summary on this topic, see Wooldridge (2001).

(1) Ramsey's (1969) RESET (Regression Specification Error Test)

Ramsey suggests testing the specification of the linear regression model $y_t = X_t'\beta + u_t$ by augmenting it with a set of regressors Z_t so that the augmented model is

$$y_t = X_t'\beta + Z_t'\gamma + u_t \tag{8.48}$$

If the Z_t's are available then the specification test would reduce to the F-test for H_0; $\gamma = 0$. The crucial issue is the choice of Z_t variables. This depends upon the true functional form under the alternative, which is usually unknown. However, this can be often well approximated by higher powers of the initial regressors, as in the case where the true form is quadratic or cubic. Alternatively, one might approximate it with higher moments of $\hat{y}_t = X_t'\widehat{\beta}_{OLS}$. The popular Ramsey RESET test is carried out as follows:

(1) Regress y_t on X_t and get $\hat{y}_t$.

(2) Regress y_t on X_t, $\hat{y}_t^2$, $\hat{y}_t^3$ and $\hat{y}_t^4$ and test that the coefficients of all the powers of $\hat{y}_t$ are zero. This is an $F_{3,T-k-3}$ under the null.

Note that $\hat{y}_t$ is not included among the regressors because it would be perfectly multicollinear with X_t.[3] Different choices of Z_t's may result in more powerful tests when H_0 is not true. Thursby and Schmidt (1977) carried out an extensive Monte Carlo and concluded that the test based on $Z_t = [X_t^2, X_t^3, X_t^4]$ seems to be generally the best choice.

(2) Utts' (1982) Rainbow Test

The basic idea behind the Rainbow test is that even when the true relationship is nonlinear, a good linear fit can still be obtained over subsets of the sample. The test therefore rejects the null hypothesis of linearity whenever the overall fit is markedly inferior to the fit over a properly selected sub-sample of the data, see Figure 8.3.

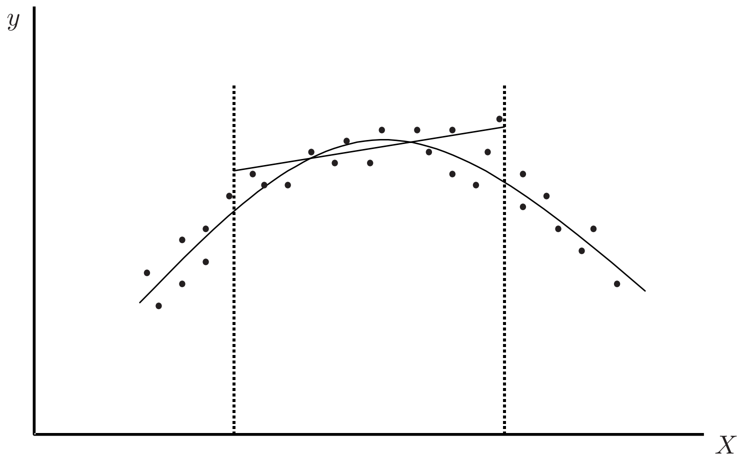

Figure 8.3 The Rainbow Test

Let $e'e$ be the OLS residuals sum of squares from all available n observations and let $\widetilde{e}'\widetilde{e}$ be the OLS residual sum of squares from the middle half of the observations $(T/2)$. Then

$$F = \frac{(e'e - \widetilde{e}'\widetilde{e})/\left(\frac{T}{2}\right)}{\widetilde{e}'\widetilde{e}/\left(\frac{T}{2}-k\right)} \quad \text{is distributed as} \quad F_{\frac{T}{2},\left(\frac{T}{2}-k\right)} \text{ under } H_0 \tag{8.49}$$

Under H_0; $E(e'e/(T-k)) = \sigma^2 = E\left[\widetilde{e}'\widetilde{e}/\left(\frac{T}{2}-k\right)\right]$, while in general under H_A; $E(e'e/(T-k)) > E\left[\widetilde{e}'\widetilde{e}/\left(\frac{T}{2}-k\right)\right] > \sigma^2$. The RRSS is $e'e$ because *all* the observations are forced to fit the straight line, whereas the URSS is $\widetilde{e}'\widetilde{e}$ because only a *part* of the observations are forced to fit a straight line. The crucial issue of the Rainbow test is the proper choice of the subsample (the middle $T/2$ observations in case of one regressor). This affects the power of the test and not the distribution of the test statistic under the null. Utts (1982) recommends points close to $\bar{X}$, since an incorrect linear fit will in general not be as far off there as it is in the outer region. Closeness to $\bar{X}$ is measured by the magnitude of the corresponding diagonal elements of P_X. Close points are those with low leverage h_{ii}, see section 8.1. The optimal size of the subset depends upon the alternative. Utts recommends about 1/2 of the data points in order to obtain some robustness to outliers. The F-test in (8.49) looks like a Chow test, but differs in the selection of the sub-sample. For example, using the post-sample predictive Chow test, the data are arranged according to time and the first T observations are selected. The Rainbow test arranges the data according to their distance from $\bar{X}$ and selects the first $T/2$ of them.

(3) Plosser, Schwert and White (1982) (PSW) Differencing Test

The differencing test is a general test for misspecification (like Hausman's (1978) test, which will be introduced in the simultaneous equation chapter) but for time-series data only. This test compares OLS and First Difference (FD) estimates of β. Let the differenced model be

$$\dot{y} = \dot{X}\beta + \dot{u} \tag{8.50}$$

where $\dot{y} = Dy$, $\dot{X} = DX$ and $\dot{u} = Du$ where

$$D = \begin{bmatrix} 1 & -1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & -1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & -1 & \ldots & 0 & 0 \\ . & . & . & . & \ldots & . & . \\ 0 & 0 & 0 & 0 & \ldots & 1 & -1 \end{bmatrix}$$ is the familiar $(T-1) \times T$ differencing matrix.

Wherever there is a constant in the regression, the first column of X becomes zero and is dropped. From (8.50), the FD estimator is given by

$$\widetilde{\beta}_{FD} = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y} \tag{8.51}$$

with $\text{var}(\widetilde{\beta}_{FD}) = \sigma^2(\dot{X}'\dot{X})^{-1}\dot{X}'DD'\dot{X}(\dot{X}'\dot{X})^{-1}$ since $\text{var}(\dot{u}) = \sigma^2(DD)'$ and

$$DD' = \begin{bmatrix} 2 & -1 & 0 & \ldots & 0 & 0 \\ -1 & 2 & -1 & \ldots & 0 & 0 \\ : & : & : & : & : & : \\ 0 & 0 & 0 & \ldots & 2 & -1 \\ 0 & 0 & 0 & \ldots & -1 & 2 \end{bmatrix}$$

The differencing test is based on

$$\widehat{q} = \widetilde{\beta}_{FD} - \widehat{\beta}_{OLS} \quad \text{with} \quad V(\widehat{q}) = \sigma^2[V(\widetilde{\beta}_{FD}) - V(\widehat{\beta}_{OLS})] \tag{8.52}$$

A consistent estimate of $V(\widehat{q})$ is

$$\widehat{V}(\widehat{q}) = \widehat{\sigma}^2 \left[\left(\frac{\dot{X}'\dot{X}}{T}\right)^{-1} \left(\frac{\dot{X}'DD'\dot{X}}{T}\right) \left(\frac{\dot{X}'\dot{X}}{T}\right)^{-1} - \left(\frac{X'X}{T}\right)^{-1} \right] \tag{8.53}$$

where $\widehat{\sigma}^2$ is a consistent estimate of σ^2. Therefore,

$$\Delta = T\widehat{q}'[\widehat{V}(\widehat{q})]^{-1}\widehat{q} \sim \chi^2_k \text{ under } H_0 \tag{8.54}$$

where k is the number of slope parameters if $\widehat{V}(\widehat{q})$ is nonsingular. $\widehat{V}(\widehat{q})$ could be singular, in which case we use a generalized inverse $\widehat{V}^{-}(\widehat{q})$ of $\widehat{V}(\widehat{q})$ and in this case is distributed as χ^2 with degrees of freedom equal to the rank$(\widehat{V}(\widehat{q}))$. This is a special case of the general Hausman (1978) test which will be studied extensively in Chapter 11.

Davidson, Godfrey, and MacKinnon (1985) show that, like the Hausman test, the PSW test is equivalent to a much simpler omitted variables test, the omitted variables being the sum of the lagged and one-period ahead values of the regressors.

Thus if the regression equation we are considering is

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \tag{8.55}$$

the PSW test involves estimating the expanded regression equation

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \gamma_1 z_{1t} + \gamma_2 z_{2t} + u_t \tag{8.56}$$

where $z_{1t} = x_{1,t+1} + x_{1,t-1}$ and $z_{2t} = x_{2,t+1} + x_{2,t-1}$ and testing the hypothesis $\gamma_1 = \gamma_2 = 0$ by the usual F-test.

If there are lagged dependent variables in the equation, the test needs a minor modification. Suppose that the model is

$$y_t = \beta_1 y_{t-1} + \beta_2 x_t + u_t \tag{8.57}$$

Now the omitted variables would be defined as $z_{1t} = y_t + y_{t-2}$ and $z_{2t} = x_{t+1} + x_{t-1}$. There is no problem with z_{2t} but z_{1t} would be correlated with the error term u_t because of the presence of y_t in it. The solution would be simply to transfer it to the left hand side and write the expanded regression equation in (8.56) as

$$(1 - \gamma_1)y_t = \beta_1 y_{t-1} + \beta_2 x_t + \gamma_1 y_{t-2} + \gamma_2 z_{2t} + u_t \tag{8.58}$$

This equation can be written as

$$y_t = \beta_1^* y_{t-1} + \beta_2^* x_t + \gamma_1^* y_{t-2} + \gamma_2^* z_{2t} + u_t^* \tag{8.59}$$

where all the starred parameters are the corresponding unstarred ones divided by $(1 - \gamma_1)$.

The PSW now tests the hypothesis $\gamma_1^* = \gamma_2^* = 0$. Thus, in the case where the model involves the lagged dependent variable y_{t-1} as an explanatory variable, the only modification needed is that we should use y_{t-2} as the omitted variable, not $(y_t + y_{t-2})$. Note that it is only y_{t-1} that creates a problem, not higher-order lags of y_t, like y_{t-2}, y_{t-3}, and so on. For y_{t-2}, the corresponding z_t will be obtained by adding y_{t-1} to y_{t-3}. This z_t is not correlated with u_t as long as the disturbances are not serially correlated.

(4) Tests for Non-nested Hypothesis

Consider the following two competing non-nested models:

$$H_1;\ y = X_1\beta_1 + \epsilon_1 \tag{8.60}$$

$$H_2;\ y = X_2\beta_2 + \epsilon_2 \tag{8.61}$$

These are non-nested because the explanatory variables under one model are not a subset of the other model even though X_1 and X_2 may share some common variables. In order to test H_1 versus H_2, Cox (1961) modified the LR-test to allow for the non-nested case. The idea behind Cox's approach is to consider to what extent Model I under H_1, is capable of predicting the performance of Model II, under H_2.

Alternatively, one can artificially nest the 2 models

$$H_3;\ y = X_1\beta_1 + X_2^*\beta_2^* + \epsilon_3 \tag{8.62}$$

where X_2^* excludes from X_2 the common variables with X_1. A test for H_1 is simply the F-test for $H_0;\ \beta_2^* = 0$.

Criticism: This tests H_1 versus H_3 which is a (Hybrid) of H_1 and H_2 and not H_1 versus H_2. Davidson and MacKinnon (1981) proposed (testing $\alpha = 0$) in the linear combination of H_1 and H_2:

$$y = (1 - \alpha)X_1\beta_1 + \alpha X_2\beta_2 + \epsilon \tag{8.63}$$

where α is an unknown scalar. Since α is not identified, we replace β_2 by $\widehat{\beta}_{2,OLS} = (X_2'X_2/T)^{-1}$ $(X_2'y/T)$ the regression coefficient estimate obtained from running y on X_2 under H_2, i.e., (1) Run y on X_2 get $\hat{y}_2 = X_2\widehat{\beta}_{2,OLS}$; (2) Run y on X_1 and $\widehat{y}_2$ and test that the coefficient of $\widehat{y}_2$ is zero. This is known as the J-test and this is asymptotically $N(0,1)$ under H_1.

Fisher and McAleer (1981) suggested a modification of the J-test known as the JA test.

$$\text{Under } H_1; \text{plim}\widehat{\beta}_2 = \text{plim}(X_2'X_2/T)^{-1}\text{plim}(X_2'X_1/T)\beta_1 + 0 \tag{8.64}$$

Therefore, they propose replacing $\widehat{\beta}_2$ by $\widetilde{\beta}_2 = (X_2'X_2)^{-1}(X_2'X_1)\widehat{\beta}_{1,OLS}$ where $\widehat{\beta}_{1,OLS} = (X_1'X_1)^{-1}X_1'y$. The steps for the JA-test are as follows:

1. Run y on X_1 get $\widehat{y}_1 = X_1\widehat{\beta}_{1,OLS}$.
2. Run $\widehat{y}_1$ on X_2 get $\widetilde{y}_2 = X_2(X_2'X_2)^{-1}X_2'\widehat{y}_1$.
3. Run y on X_1 and $\widetilde{y}_2$ and test that the coefficient of $\widetilde{y}_2$ is zero. This is the simple t-statistic on the coefficient of $\widetilde{y}_2$. The J and JA tests are asymptotically equivalent.

Criticism: Note the asymmetry of H_1 and H_2. Therefore one should reverse the role of these hypotheses and test again.

In this case one can get the four scenarios depicted in Table 8.6. In case both hypotheses are not rejected, the data are not rich enough to discriminate between the two hypotheses. In case both hypotheses are rejected neither model is useful in explaining the variation in y. In case one hypothesis is rejected while the other is not, one should remember that the non-rejected hypothesis may still be brought down by another challenger hypothesis.

Small Sample Properties: (i) The J-test tends to reject the null more frequently than it should. Also, the JA test has relatively low power when K_1, the number of parameters in H_1 is larger than K_2, the number of parameters in H_2. Therefore, one should use the JA test when K_1 is about the same size as K_2, i.e., the same number of non-overlapping variables. (ii) If both H_1 and H_2 are false, these tests are inferior to the standard diagnostic tests. In practice, use higher significance levels for the J-test, and supplement it with the artificially nested F-test and standard diagnostic tests.

Table 8.6 Non-nested Hypothesis Testing

		$\alpha = 0$: Not Rejected	$\alpha = 0$: Rejected
$\alpha = 1$	Not Rejected	Both H_1 and H_2 are not rejected	H_1 rejected H_2 not rejected
	Rejected	H_1 not rejected H_2 rejected	Both H_1 and H_2 are rejected

Note: J and JA tests are one degree of freedom tests, whereas the artificially nested F-test is not.

For a recent summary of non-nested hypothesis testing, see Pesaran and Weeks (2001). Examples of non-nested hypothesis encountered in empirical economic research include linear versus log-linear models, see section 8.5. Also, logit versus probit models in discrete choice, see Chapter 13 and exponential versus Weibull distributions in the analysis of duration data. In the logit versus probit specification, the set of regressors is most likely to be the same. It is only the form of the distribution functions that separate the two models. Pesaran and Weeks (2001, p. 287) emphasize the differences between hypothesis testing and model selection:

> *The model selection process treats all models under consideration symmetrically, while hypothesis testing attributes a different status to the null and to the alternative hypotheses and by design treats the models asymmetrically. Model selection always ends in a definite outcome, namely one of the models under consideration is selected for use in decision making. Hypothesis testing on the other hand asks whether there is any statistically significant evidence (in the Neyman-Pearson sense) of departure from the null hypothesis in the direction of one or more alternative hypotheses. Rejection of the null hypothesis does not necessarily imply acceptance of any one of the alternative hypotheses; it only warns the investigator of possible shortcomings of the null that is being advocated. Hypothesis testing does not seek a definite outcome and if carried out with due care need not lead to a favorite model. For example, in the case of nonnested hypothesis testing it is possible for all models under consideration to be rejected, or all models to be deemed as observationally equivalent.*

They conclude that the choice between hypothesis testing and model selection depends on the primary objective of one's study. Model selection may be more appropriate when the objective is decision making, while hypothesis testing is better suited to inferential problems.

> *A model may be empirically adequate for a particular purpose, but of little relevance for another use... In the real world where the truth is elusive and unknowable both approaches to model evaluation are worth pursuing.*

(5) White's (1982) Information-Matrix (IM) Test

This is a general specification test much like the Hausman (1978) specification test which will be considered in details in Chapter 11. The latter is based on two different estimates of the regression coefficients, while the former is based on two different estimates of the Information Matrix I(θ) where $\theta' = (\beta', \sigma^2)$ in the case of the linear regression studied in Chapter 7. The first estimate of $I(\theta)$ evaluates the expectation of the second derivatives of the log-likelihood at the MLE, i.e., $-E(\partial^2 \text{log}L/\partial\theta\partial\theta')$ at $\widehat{\theta}_{mle}$ while the second sum up the outer products of the score vectors $\sum_{i=1}^{n}(\partial \text{log}L_i(\theta)/\partial\theta)(\partial \text{log}L_i(\theta)/\partial\theta)'$ evaluated at $\widehat{\theta}_{mle}$. This is based on the fundamental identity that

$$I(\theta) = -E(\partial^2 \text{log}L/\partial\theta\partial\theta') = E(\partial \text{log}L/\partial\theta)(\partial \text{log}L/\partial\theta)'$$

If the model estimated by MLE is not correctly specified, this equality will not hold. From Chapter 7, equation (7.19), we know that for the linear regression model with normal disturbances, the first estimate of $I(\theta)$ denoted by $I_1(\widehat{\theta}_{mle})$ is given by

$$I_1(\widehat{\theta}_{MLE}) = \begin{bmatrix} X'X/\widehat{\sigma}^2 & 0 \\ 0 & n/2\widehat{\sigma}^4 \end{bmatrix} \tag{8.65}$$

where $\widehat{\sigma}^2 = e'e/n$ is the MLE of σ^2 and e denotes the OLS residuals.

Similarly, one can show that the second estimate of $I(\theta)$ denoted by $I_2(\theta)$ is given by

$$\begin{aligned} I_2(\theta) &= \textstyle\sum_{i=1}^n \left(\frac{\partial \log L_i(\theta)}{\partial \theta}\right)\left(\frac{\partial \log L_i(\theta)}{\partial \theta}\right)' \\ &= \textstyle\sum_{i=1}^n \begin{bmatrix} \frac{u_i^2 x_i x_i'}{\sigma^4} & \frac{-u_i x_i}{2\sigma^4} + \frac{u_i^3 x_i}{2\sigma^6} \\ -\frac{u_i x_i'}{2\sigma^4} + \frac{u_i^3 x_i'}{2\sigma^6} & \frac{1}{4\sigma^4} - \frac{u_i^2}{2\sigma^6} + \frac{u_i^4}{4\sigma^8} \end{bmatrix} \end{aligned} \tag{8.66}$$

where x_i is the i-th row of X. Substituting the MLE we get

$$I_2(\widehat{\theta}_{MLE}) = \begin{bmatrix} \frac{\sum_{i=1}^n e_i^2 x_i x_i'}{\widehat{\sigma}^4} & \frac{\sum_{i=1}^n e_i^3 x_i}{2\widehat{\sigma}^6} \\ \frac{\sum_{i=1}^n e_i^3 x_i'}{2\widehat{\sigma}^6} & -\frac{n}{4\widehat{\sigma}^4} + \frac{\sum_{i=1}^n e_i^4}{4\widehat{\sigma}^8} \end{bmatrix} \tag{8.67}$$

where we used the fact that $\sum_{i=1}^n e_i x_i = 0$. If the model is correctly specified and the disturbances are normal then

$$\text{plim } I_1(\widehat{\theta}_{MLE})/n = \text{plim } I_2(\widehat{\theta}_{MLE})/n = I(\theta)$$

Therefore, the Information Matrix (IM) test rejects the model when

$$[I_2(\widehat{\theta}_{MLE}) - I_1(\widehat{\theta}_{MLE})]/n \tag{8.68}$$

is too large. These are two matrices with $(k+1)$ by $(k+1)$ elements since β is $k \times 1$ and σ^2 is a scalar. However, due to symmetry, this reduces to $(k+2)(k+1)/2$ unique elements. Hall (1987) noted that the first $k(k+1)/2$ unique elements obtained from the first $k \times k$ block of (8.68) have a typical element $\sum_{i=1}^n (e_i^2 - \widehat{\sigma}^2) x_{ir} x_{is}/n\widehat{\sigma}^4$ where r and s denote the r-th and s-th explanatory variables with $r, s = 1, 2, \ldots, k$. This term measures the discrepancy between the OLS estimates of the variance-covariance matrix of $\widehat{\beta}_{OLS}$ and its robust counterpart suggested by White (1980), see Chapter 5. The next k unique elements correspond to the off-diagonal block $\sum_{i=1}^n e_i^3 x_i/2n\widehat{\sigma}^6$and this measures the discrepancy between the estimates of the $\text{cov}(\widehat{\beta}, \widehat{\sigma}^2)$. The last element correspond to the difference in the bottom right elements, i.e., the two estimates of $\widehat{\sigma}^2$. This is given by

$$\left[-\frac{3}{4\widehat{\sigma}^4} + \frac{1}{n}\textstyle\sum_{i=1}^n e_i^4/4\widehat{\sigma}^8\right]$$

These $(k+1)(k+2)/2$ unique elements can be arranged in vector form $D(\theta)$ which has a limiting normal distribution with zero mean and some covariance matrix $V(\theta)$ under the null. One can show, see Hall (1987) or Krämer and Sonnberger (1986) that if $V(\theta)$ is estimated from the sample moments of these terms, that the IM test statistic is given by

$$m = nD'(\theta)[V(\theta)]^{-1}D(\theta) \overset{H_0}{\to} \chi^2_{(k+1)(k+2)/2} \tag{8.69}$$

In fact, Hall (1987) shows that this statistic is the sum of three asymptotically independent terms

$$m = m_1 + m_2 + m_3 \tag{8.70}$$

where m_1 = a particular version of White's heteroskedasticity test; $m_2 = n$ times the explained sum of squares from the regression of e_i^3 on x_i divided by $6\widehat{\sigma}^6$; and

$$m_3 = \frac{n}{24\widehat{\sigma}^8}\left(\sum_{i=1}^{n} e_i^4/n - 3\widehat{\sigma}^4\right)^2$$

which is similar to the Jarque-Bera test for normality of the disturbances given in Chapter 5.

It is clear that the IM test will have power whenever the disturbances are non-normal or heteroskedastic. However, Davidson and MacKinnon (1992) demonstrated that the IM test considered above will tend to reject the model when true, much too often, in finite samples. This problem gets worse as the number of degrees of freedom gets large. In Monte Carlo experiments, Davidson and MacKinnon (1992) showed that for a linear regression model with ten regressors, the IM test rejected the null at the 5% level, 99.9% of the time for $n = 200$. This problem did not disappear when n increased. In fact, for $n = 1000$, the IM test still rejected the null 92.7% of the time at the 5% level.

These results suggest that it may be more useful to run individual tests for non-normality, heteroskedasticity and other misspecification tests considered above rather than run the IM test. These tests may be more powerful and more informative than the IM test. Alternative methods of calculating the IM test with better finite-sample properties are suggested in Orme (1990), Chesher and Spady (1991) and Davidson and MacKinnon (1992).

Example 3: For the consumption-income data given in Table 5.3, we first compute the RESET test from the consumption-income regression given in Chapter 5. Using EViews, one clicks on *stability tests* and then selects RESET. You will be prompted with the option of the number of fitted terms to include (i.e., powers of $\widehat{y}$). Table 8.7 shows the RESET test including $\widehat{y}^2$ and $\widehat{y}^3$. The F-statistic for their joint-significance is equal to 94.94. This is significant and indicates misspecification.

Table 8.7 Ramsey RESET Test

F-statistic	94.93796	Prob. F(2,45)		0.00000
Log likelihood ratio	80.96735	Prob. Chi-Square(2)		0.00000

Test Equation:
Dependent Variable: CONSUM
Method: Least Squares
Sample: 1959 2007
Included observations: 49

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3519.599	1141.261	3.083956	0.0035
Y	0.421587	0.173597	2.428540	0.0192
FITTED^2	1.99E-05	1.09E-05	1.834317	0.0732
FITTED^3	–1.18E-10	2.10E-10	–0.560377	0.5780

R-squared	0.998789	Mean dependent var	16749.10
Adjusted R-squared	0.998708	S.D. dependent var	5447.060
S.E. of regression	195.7648	Akaike info criterion	13.46981
Sum squared resid	1724573.	Schwarz criterion	13.62425
Log likelihood	–326.0104	Hannan-Quinn criter.	13.52840
F-statistic	12372.26	Durbin-Watson stat	1.001605
Prob(F-statistic)	0.000000		

Table 8.8 Consumption Regression 1971–1995

Dependent Variable: CONSUM
Method: Least Squares
Sample: 1971 1995
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	–1410.425	371.3812	–3.797783	0.0009
Y	0.963780	0.020036	48.10199	0.0000

R-squared	0.990157	Mean dependent var	16279.48
Adjusted R-squared	0.989730	S.D. dependent var	2553.097
S.E. of regression	258.7391	Akaike info criterion	14.02614
Sum squared resid	1539756.	Schwarz criterion	14.12365
Log likelihood	–173.3267	Hannan-Quinn criter.	14.05318
F-statistic	2313.802	Durbin-Watson stat	0.613064
Prob(F-statistic)	0.000000		

Next, we compute Utts (1982) Rainbow test. Table 8.8 gives the middle 25 observations of our data, i.e., 1971-1995, and the EViews 6 regression using this data. The RSS of these middle observations is given by $\widetilde{e}'\widetilde{e} = 1539756.14$, while the RSS for the entire sample is given by $e'e = 9001347.76$ so that the observed F-statistic given in (8.49) can be computed as follows:

$$F = \frac{(9001347.76 - 1539756.14)/25}{1539756.14/23} = 4.46$$

This is distributed as $F_{25,23}$ under the null hypothesis and rejects the hypothesis of linearity.

The PSW differencing test is computed using the artificial regression given in (8.56) with $Z_t = Y_{t+1} + Y_{t-1}$. The results are given in Table 8.9 using EViews 6. The t-statistic for Z_t is 1.19 and has a p-value of 0.24 which is insignificant.

Now consider the two competing non-nested models:

$$H_1;\ C_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + u_t \quad H_2;\ C_t = \gamma_0 + \gamma_1 Y_t + \gamma_2 C_{t-1} + v_t$$

The two non-nested models share Y_t as a common variable. The artificial model that nests these two models is given by:

$$H_3;\ C_t = \delta_0 + \delta_1 Y_t + \delta_2 Y_{t-1} + \delta_3 C_{t-1} + \epsilon_t$$

Table 8.10, runs regression (1) given by H_2 and obtains the predicted values $\widehat{C}_2(C2HAT)$. Regression (2) runs consumption on a constant, income, lagged income and $C2HAT$. The coefficient of this last variable is 1.18 and is statistically significant with a t-value of 16.99. This is the Davidson and MacKinnon (1981) J-test. In this case, H_1 is rejected but H_2 is not rejected. The JA-test, given by Fisher and McAleer (1981) runs the regression in H_1 and keeps the predicted values $\widehat{C}_1(C1HAT)$. This is done in regression (3). Then $C1HAT$ is run on a constant, income and lagged consumption and the predicted values are stored as $\widetilde{C}_2(C2TILDE)$. This is done in regression (5). The last step runs consumption on a constant, income, lagged income and $C2TILDE$, see regression (6). The coefficient of this last variable is 97.43 and is statistically significant with a t-value of 16.99. Again H_1 is rejected but H_2 is not rejected.

Table 8.9 Artificial Regression to compute the PSW Differencing Test

Dependent Variable: CONSUM
Method: Least Squares
Sample (adjusted): 1960 2006
Included observations: 47 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	–1373.390	226.1376	–6.073251	0.0000
Y	0.596293	0.321464	1.854930	0.0703
Z	0.191494	0.160960	1.189700	0.2405

R-squared	0.993678	Mean dependent var	16693.85
Adjusted R-squared	0.993390	S.D. dependent var	5210.244
S.E. of regression	423.5942	Akaike info criterion	14.99713
Sum squared resid	7895011.	Schwarz criterion	15.11523
Log likelihood	–349.4326	Hannan-Quinn criter.	15.04157
F-statistic	3457.717	Durbin-Watson stat	0.119325
Prob(F-statistic)	0.000000		

Reversing the roles of H_1 and H_2, the J and JA-tests are repeated. In fact, regression (4) runs consumption on a constant, income, lagged consumption and $\widehat{C}_1$ (which was obtained from regression (3)). The coefficient on $\widehat{C}_1$ is -15.20 and is statistically significant with a t-value of -6.5. This J-test rejects H_2 but does not reject H_1. Regression (7) runs $\widehat{C}_2$ on a constant, income and lagged income and the predicted values are stored as $\widetilde{C}_1$(C1TILDE). The last step of the JA test runs consumption on a constant, income, lagged consumption and $\widetilde{C}_1$, see regression (8). The coefficient of this last variable is -1.11 and is statistically significant with a t-value of -6.5. This JA test rejects H_2 but not H_1. The artificial model, given in H_3, is also estimated, see regression (9). One can easily check that the corresponding F-tests reject H_1 against H_3 and also H_2 against H_3. In sum, all evidence indicates that both C_{t-1} and Y_{t-1} are important to include along with Y_t. Of course, the true model is not known and could include higher lags of both Y_t and C_t.

Stata 11 performs White's (1982) Information matrix test by issuing the command *estat imtest* after running the regression of consumption on income. The results yield:

```
. estat imtest
Cameron & Trivedi's decomposition of IM-test
----------------------------------------------
              Source |   chi2    df        p
---------------------+------------------------
  Heteroskedasticity |   2.64     2     0.2677
            Skewness |   0.45     1     0.5030
            Kurtosis |   4.40     1     0.0359
---------------------+------------------------
               Total |   7.48     4     0.1124
```

This does not reject the null even though Kurtosis seems to be a problem. Note that the IM test is split into its components following Hall (1987) as described above.

Table 8.10 Non-nested J and JA Tests for the Consumption Regression

Regression 1

Dependent Variable: CONSUM
Method: Least Squares
Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	–254.5241	155.2906	–1.639019	0.1082
Y	0.211505	0.068310	3.096256	0.0034
CONSUM(–1)	0.800004	0.070537	11.34159	0.0000

R-squared	0.998367	Mean dependent var	16915.21
Adjusted R-squared	0.998294	S.D. dependent var	5377.825
S.E. of regression	222.1108	Akaike info criterion	13.70469
Sum squared resid	2219995.	Schwarz criterion	13.82164
Log likelihood	–325.9126	Hannan-Quinn criter.	13.74889
F-statistic	13754.09	Durbin-Watson stat	0.969327
Prob(F-statistic)	0.000000		

Regression 2

Dependent Variable: CONSUM
Method: Least Squares
Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	144.3306	125.5929	1.149194	0.2567
Y	0.425354	0.090692	4.690091	0.0000
Y(–1)	–0.613631	0.094424	-6.498678	0.0000
C2HAT	1.184853	0.069757	16.98553	0.0000

R-squared	0.999167	Mean dependent var	16915.21
Adjusted R-squared	0.999110	S.D. dependent var	5377.825
S.E. of regression	160.4500	Akaike info criterion	13.07350
Sum squared resid	1132745.	Schwarz criterion	13.22943
Log likelihood	–309.7639	Hannan-Quinn criter.	13.13242
F-statistic	17585.25	Durbin-Watson stat	1.971939
Prob(F-statistic)	0.000000		

8.4 Nonlinear Least Squares and the Gauss-Newton Regression[4]

So far we have been dealing with linear regressions. But, in reality, one might face a nonlinear regression of the form:

$$y_t = x_t(\beta) + u_t \quad \text{for } t = 1, 2, \ldots, T \tag{8.71}$$

where $u_t \sim \text{IID}(0, \sigma^2)$ and $x_t(\beta)$ is a scalar nonlinear regression function of k unknown parameters β. It can be interpreted as the expected value of y_t conditional on the values of the inde-

Table 8.10 (continued)

REGRESSION 3

Dependent Variable: CONSUM
Method: Least Squares
Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	–1424.802	231.2843	–6.160393	0.0000
Y	0.943371	0.232170	4.063283	0.0002
Y(–1)	0.040368	0.234363	0.172244	0.8640

R-squared	0.993702	Mean dependent var	16915.21
Adjusted R-squared	0.993423	S.D. dependent var	5377.825
S.E. of regression	436.1488	Akaike info criterion	15.05431
Sum squared resid	8560159.	Schwarz criterion	15.17126
Log likelihood	–358.3033	Hannan-Quinn criter.	15.09850
F-statistic	3550.327	Durbin-Watson stat	0.174411
Prob(F-statistic)	0.000000		

REGRESSION 4

Dependent Variable: CONSUM
Method: Least Squares
Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	–21815.80	3319.691	–6.571637	0.0000
Y	15.01623	2.278648	6.589974	0.0000
CONSUM(–1)	0.947887	0.055806	16.98553	0.0000
C1HAT	–15.20110	2.339106	–6.498678	0.0000

R-squared	0.999167	Mean dependent var	16915.21
Adjusted R-squared	0.999110	S.D. dependent var	5377.825
S.E. of regression	160.4500	Akaike info criterion	13.07350
Sum squared resid	1132745.	Schwarz criterion	13.22943
Log likelihood	–309.7639	Hannan-Quinn criter.	13.13242
F-statistic	17585.25	Durbin-Watson stat	1.971939
Prob(F-statistic)	0.000000		

pendent variables. Nonlinear least squares minimizes $\sum_{t=1}^{T}(y_t - x_t(\beta))^2 = (y - x(\beta))'(y - x(\beta))$. The first-order conditions for minimization yield

$$X'(\widehat{\beta})(y - x(\widehat{\beta})) = 0 \tag{8.72}$$

where $X(\beta)$ is a $T \times k$ matrix with typical element $X_{tj}(\beta) = \partial x_t(\beta)/\partial \beta_j$ for $j = 1, \ldots, k$. The solution to these k equations yield the Nonlinear Least Squares (NLS) estimates of β denoted by $\widehat{\beta}_{NLS}$. These normal equations given in (8.72) are similar to those in the linear case in that they

Table 8.10 (continued)

Regression 5

Dependent Variable: C1HAT
Method: Least Squares
Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	–1418.403	7.149223	–198.3996	0.0000
Y	0.973925	0.003145	309.6905	0.0000
CONSUM(–1)	0.009728	0.003247	2.995785	0.0044

R-squared	0.999997	Mean dependent var	16915.21
Adjusted R-squared	0.999996	S.D. dependent var	5360.865
S.E. of regression	10.22548	Akaike info criterion	7.548103
Sum squared resid	4705.215	Schwarz criterion	7.665053
Log likelihood	–178.1545	Hannan-Quinn criter.	7.592298
F-statistic	6459057.	Durbin-Watson stat	1.678118
Prob(F-statistic)	0.000000		

Regression 6

Dependent Variable: CONSUM
Method: Least Squares
Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	138044.4	8211.501	16.81111	0.0000
Y	–94.21814	5.603155	–16.81519	0.0000
Y(–1)	–0.613631	0.094424	–6.498678	0.0000
C2TILDE	97.43471	5.736336	16.98553	0.0000

R-squared	0.999167	Mean dependent var	16915.21
Adjusted R-squared	0.999110	S.D. dependent var	5377.825
S.E. of regression	160.4500	Akaike info criterion	13.07350
Sum squared resid	1132745.	Schwarz criterion	13.22943
Log likelihood	–309.7639	Hannan-Quinn criter.	13.13242
F-statistic	17585.25	Durbin-Watson stat	1.971939
Prob(F-statistic)	0.000000		

require the vector of residuals $y - x(\widehat{\beta})$ to be orthogonal to the matrix of derivatives $X(\widehat{\beta})$. In the linear case, $x(\widehat{\beta}) = X\widehat{\beta}_{OLS}$ and $X(\widehat{\beta}) = X$ where the latter is independent of $\widehat{\beta}$. Because of this dependence of the fitted values $x(\widehat{\beta})$ as well as the matrix of derivatives $X(\widehat{\beta})$ on $\widehat{\beta}$, one in general cannot get explicit analytical solution to these NLS first-order equations. Under fairly general conditions, see Davidson and MacKinnon (1993), one can show that the $\widehat{\beta}_{NLS}$ has asymptotically

Table 8.10 (continued)

Regression 7

Dependent Variable: C2HAT
Method: Least Squares
Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	–1324.328	181.8276	–7.283424	0.0000
Y	0.437200	0.182524	2.395306	0.0208
Y(–1)	0.551966	0.184248	2.995785	0.0044

R-squared	0.996101	Mean dependent var	16915.21
Adjusted R-squared	0.995928	S.D. dependent var	5373.432
S.E. of regression	342.8848	Akaike info criterion	14.57313
Sum squared resid	5290650.	Schwarz criterion	14.69008
Log likelihood	–346.7551	Hannan-Quinn criter.	14.61732
F-statistic	5748.817	Durbin-Watson stat	0.127201
Prob(F-statistic)	0.000000		

Regression 8

Dependent Variable: CONSUM
Method: Least Squares
Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	–1629.522	239.4806	–6.804403	0.0000
Y	1.161999	0.154360	7.527865	0.0000
CONSUM(–1)	0.947887	0.055806	16.98553	0.0000
C1TILDE	–1.111718	0.171068	–6.498678	0.0000

R-squared	0.999167	Mean dependent var	16915.21
Adjusted R-squared	0.999110	S.D. dependent var	5377.825
S.E. of regression	160.4500	Akaike info criterion	13.07350
Sum squared resid	1132745.	Schwarz criterion	13.22943
Log likelihood	–309.7639	Hannan-Quinn criter.	13.13242
F-statistic	17585.25	Durbin-Watson stat	1.971939
Prob(F-statistic)	0.000000		

a normal distribution with mean β_0 and asymptotic variance $\sigma_0^2(X'(\beta_0)X(\beta_0))^{-1}$, where β_0 and σ_0 are the true values of the parameters generating the data. Similarly, defining

$$s^2 = (y - x(\widehat{\beta}_{NLS}))'(y - x(\widehat{\beta}_{NLS}))/(T - k)$$

we get a feasible estimate of this covariance matrix as $s^2(X'(\widehat{\beta})X(\widehat{\beta}))^{-1}$. If the disturbances are normally distributed then NLS is MLE and therefore asymptotically efficient as long as the model is correctly specified, see Chapter 7.

Table 8.10 (continued)

Regression 9

Dependent Variable: CONSUM
Method: Least Squares
Sample (adjusted): 1960 2007
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	–157.2430	113.1743	–1.389389	0.1717
Y	0.675956	0.086849	7.783091	0.0000
Y(–1)	–0.613631	0.094424	–6.498678	0.0000
CONSUM(–1)	0.947887	0.055806	16.98553	0.0000

R-squared	0.999167	Mean dependent var	16915.21
Adjusted R-squared	0.999110	S.D. dependent var	5377.825
S.E. of regression	160.4500	Akaike info criterion	13.07350
Sum squared resid	1132745.	Schwarz criterion	13.22943
Log likelihood	–309.7639	Hannan-Quinn criter.	13.13242
F-statistic	17585.25	Durbin-Watson stat	1.971939
Prob(F-statistic)	0.000000		

Taking the first-order Taylor series approximation around some arbitrary parameter vector β^*, we get

$$y = x(\beta^*) + X(\beta^*)(\beta - \beta^*) + \text{higher-order terms} + u \tag{8.73}$$

or

$$y - x(\beta^*) = X(\beta^*)b + \text{residuals} \tag{8.74}$$

This is the simplest version of the Gauss-Newton Regression, see Davidson and MacKinnon (1993). In this case the higher-order terms and the error term are combined in the residuals and $(\beta - \beta^*)$ is replaced by b, a parameter vector that can be estimated. If the model is linear, $X(\beta^*)$ is the matrix of regressors X and the GNR regresses a residual on X. If $\beta^* = \widehat{\beta}_{NLS}$, the unrestricted NLS estimator of β, then the GNR becomes

$$y - \widehat{x} = \widehat{X}b + \text{residuals} \tag{8.75}$$

where $\widehat{x} \equiv x(\widehat{\beta}_{NLS})$ and $\widehat{X} \equiv X(\widehat{\beta}_{NLS})$. From the first-order conditions of NLS we get $(y - \widehat{x})'\widehat{X} = 0$. In this case, OLS on this GNR yields $\widehat{b}_{OLS} = (\widehat{X}'\widehat{X})^{-1}\widehat{X}'(y - \widehat{x}) = 0$ and this GNR has no explanatory power. However, this regression can be used to (i) check that the first-order conditions given in (8.72) are satisfied. For example, one could check that the t-statistics are of the 10^{-3} order, and that R^2 is zero up to several decimal places; (ii) compute estimated covariance matrices. In fact, this GNR prints out $s^2(\widehat{X}'\widehat{X})^{-1}$, where $s^2 = (y - \widehat{x})'(y - \widehat{x})/(T - k)$ is the OLS estimate of the regression variance. This can be verified easily using the fact that this GNR has no explanatory power. This method of computing the estimated variance-covariance matrix is useful especially in cases where $\widehat{\beta}$ has been obtained by some method other than NLS.

For example, sometimes the model is nonlinear only in one or two parameters which are known to be in a finite range, say between zero and one. One can then search over this range, running OLS regressions and minimizing the residual sum of squares. This search procedure can be repeated over finer grids to get more accuracy. Once the final parameter estimate is found, one can run the GNR to get estimates of the variance-covariance matrix.

Testing Restrictions (GNR Based on the Restricted NLS Estimates)

The best known use for the GNR is to test restrictions. These are based on the LM principle which requires only the restricted estimator. In particular, consider the following competing hypotheses:

$$H_0; y = x(\beta_1, 0) + u \qquad H_1; y = x(\beta_1, \beta_2) + u$$

where $u \sim \text{IID}(0, \sigma^2 I)$ and β_1 and β_2 are $k \times 1$ and $r \times 1$, respectively. Denote by $\widetilde{\beta}$ the restricted NLS estimator of β, in this case $\widetilde{\beta}' = (\widetilde{\beta}_1', 0)$.

The GNR evaluated at this restricted NLS estimator of β is

$$(y - \widetilde{x}) = \widetilde{X}_1 b_1 + \widetilde{X}_2 b_2 + \text{residuals} \tag{8.76}$$

where $\widetilde{x} = x(\widetilde{\beta})$ and $\widetilde{X}_i = X_i(\widetilde{\beta})$ with $X_i(\beta) = \partial x / \partial \beta_i$ for $i = 1, 2$.

By the FWL Theorem this yields the same estimate of b_2 as

$$\bar{P}_{\widetilde{X}_1}(y - \widetilde{x}) = \bar{P}_{\widetilde{X}_1} \widetilde{X}_2 b_2 + \text{residuals} \tag{8.77}$$

But $\bar{P}_{\widetilde{X}_1}(y - \widetilde{x}) = (y - \widetilde{x}) - P_{\widetilde{X}_1}(y - \widetilde{x}) = (y - \widetilde{x})$ since $\widetilde{X}_1'(y - \widetilde{x}) = 0$ from the first-order conditions of restricted NLS. Hence, (8.77) reduces to

$$(y - \widetilde{x}) = \bar{P}_{\widetilde{X}_1} \widetilde{X}_2 b_2 + \text{residuals} \tag{8.78}$$

Therefore,

$$b_{2,OLS} = (\widetilde{X}_2' \bar{P}_{\widetilde{X}_1} \widetilde{X}_2)^{-1} \widetilde{X}_2' \bar{P}_{\widetilde{X}_1}(y - \widetilde{x}) = (\widetilde{X}_2' \bar{P}_{\widetilde{X}_1} \widetilde{X}_2)^{-1} \widetilde{X}_2'(y - \widetilde{x}) \tag{8.79}$$

and the residual sums of squares is $(y - \widetilde{x})'(y - \widetilde{x}) - (y - \widetilde{x})' \widetilde{X}_2 (\widetilde{X}_2' \bar{P}_{\widetilde{X}_1} \widetilde{X}_2)^{-1} \widetilde{X}_2'(y - \widetilde{x})$.

If $\widetilde{X}_2$ was excluded from the regression in (8.76), $(y - \widetilde{x})'(y - \widetilde{x})$ would be the residual sum of squares. Therefore, the reduction in the residual sum of squares brought about by the inclusion of $\widetilde{X}_2$ is

$$(y - \widetilde{x})' \widetilde{X}_2 (\widetilde{X}_2' \bar{P}_{\widetilde{X}_1} \widetilde{X}_2)^{-1} \widetilde{X}_2'(y - \widetilde{x})$$

This is also equal to the explained sum of squares from (8.76) since $\widetilde{X}_1$ has no explanatory power. This sum of squares divided by a consistent estimate of σ^2 is asymptotically distributed as χ_r^2 under the null.

Different consistent estimates of σ^2 yield different test statistics. The two most common test statistics for H_0 based on this regression are the following: (1) TR_u^2 where R_u^2 is the *uncentered* R^2 of (8.76) and (2) the F-statistic for $b_2 = 0$. The first statistic is given by $TR_u^2 = T(y - \widetilde{x})' \widetilde{X}_2 (\widetilde{X}_2' \bar{P}_{\widetilde{X}_1} \widetilde{X}_2)^{-1} \widetilde{X}_2'(y - \widetilde{x}) / (y - \widetilde{x})'(y - \widetilde{x})$ where the *uncentered* R^2 was defined in the

Appendix to Chapter 3. This statistic implicitly divides the explained sum of squares term by $\widetilde{\sigma}^2$ = (restricted residual sums of squares)/T. This is equivalent to the LM-statistic obtained by running the artificial regression $(y - \widetilde{x})/\widetilde{\sigma}$ on $\widetilde{X}$ and getting the explained sum of squares. Regression packages print the *centered* R^2. This is equal to the *uncentered* R_u^2 as long as there is a constant in the restricted regression so that $(y - \widetilde{x})$ sum to zero.

The F-statistic for $b_2 = 0$ from (8.76) is

$$\frac{(RRSS - URSS)/r}{URSS/(T-k)} = \frac{(y-\widetilde{x})'\widetilde{X}_2(\widetilde{X}_2'\bar{P}_{\widetilde{X}_1}\widetilde{X}_2)^{-1}\widetilde{X}_2'(y-\widetilde{x})/r}{[(y-\widetilde{x})'(y-\widetilde{x}) - (y-\widetilde{x})'\widetilde{X}_2(\widetilde{X}_2'\bar{P}_{\widetilde{X}_1}\widetilde{X}_2)^{-1}\widetilde{X}_2'(y-\widetilde{x})]/(T-k)} \quad (8.80)$$

The denominator is the OLS estimate of σ^2 from (8.76) which tends to σ_0^2 as $T \to \infty$. Hence (rF-statistic $\to \chi_r^2$). In small samples, use the F-statistic.

Diagnostic Tests for Linear Regression Models

Variable addition tests suggested by Pagan and Hall (1983) consider the additional variables Z of dimension $(T \times r)$ and test whether their coefficients are zero using an F-test from the regression

$$y = X\beta + Z\gamma + u \quad (8.81)$$

If H_0; $\gamma = 0$ is true, the model is $y = X\beta + u$ and there is no misspecification. The GNR for this restriction would run the following regression:

$$\bar{P}_X y = Xb + Zc + \text{residuals} \quad (8.82)$$

and test that c is zero. By the FWL Theorem, (8.82) yields the same residual sum of squares as

$$\bar{P}_X y = \bar{P}_X Zc + \text{residuals} \quad (8.83)$$

Applying the FWL Theorem to (8.81) we get the same residual sum of squares as the regression in (8.83). The F-statistic for $\gamma = 0$ from (8.81) is therefore identical to the F-statistic for $c = 0$ from the GNR given in (8.82). Hence, "Tests based on the GNR are equivalent to variable addition tests when the latter are applicable," see Davidson and MacKinnon (1993, p. 194).

Note also, that the nR_u^2 test statistic for H_0; $\gamma = 0$ based on the GNR in (8.82) is exactly the LM statistic based on running the restricted least squares residuals of y on X on the unrestricted set of regressors X and Z in (8.81). If X has a constant, then the uncentered R^2 is equal to the centered R^2 printed by the regression.

Computational Warning: It is tempting to base tests on the OLS residuals $\widehat{u} = \bar{P}_X y$ by simply regressing them on the test regressors Z. This is equivalent to running the GNR *without* the X variables on the right hand side of (8.82) yielding test-statistics that are *too small*.

Functional Form

Davidson and MacKinnon (1993, p. 195) show that the RESET with $y_t = X_t\beta + \widehat{y}_t^2 c +$ residual which is based on testing for $c = 0$ is equivalent to testing for $\theta = 0$ using the nonlinear model $y_t = X_t\beta(1 + \theta X_t\beta) + u_t$. In this case, it is easy to verify from (8.74) that the GNR is

$$y_t - X_t\beta(1 + \theta X_t\beta) = (2\theta(X_t\beta)X_t + X_t)b + (X_t\beta)^2 c + \text{residual}$$

At $\theta = 0$ and $\beta = \widehat{\beta}_{OLS}$, the GNR becomes $(y_t - X_t\widehat{\beta}_{OLS}) = X_t b + (X_t\widehat{\beta}_{OLS})^2 c +$ residual. The t-statistic on $c = 0$ is equivalent to that from the RESET regression given in section 8.3, see problem 25.

Testing for Serial Correlation

Suppose that the null hypothesis is the nonlinear regression model given in (8.71), and the alternative is the model $y_t = x_t(\beta) + \nu_t$ with $\nu_t = \rho\nu_{t-1} + u_t$ where $u_t \sim \text{IID}(0, \sigma^2)$. Conditional on the first observation, the alternative model can be written as

$$y_t = x_t(\beta) + \rho(y_{t-1} - x_{t-1}(\beta)) + u_t$$

The GNR test for H_0; $\rho = 0$, computes the derivatives of this regression function with respect to β and ρ evaluated at the restricted estimates under the null hypothesis, i.e., $\rho = 0$ and $\beta = \widehat{\beta}_{NLS}$ (the nonlinear least squares estimate of β assuming no serial correlation). Those yield $X_t(\widehat{\beta}_{NLS})$ and $(y_{t-1} - x_{t-1}(\widehat{\beta}_{NLS}))$ respectively. Therefore, the GNR runs $\widehat{u}_t = y_t - x_t(\widehat{\beta}_{NLS}) = X_t(\widehat{\beta}_{NLS})b + c\widehat{u}_{t-1} +$ residual, and tests that $c = 0$. If the regression model is linear, this reduces to running ordinary least squares residuals on their lagged values in addition to the regressors in the model. This is exactly the Breusch and Godfrey test for first-order serial correlation considered in Chapter 5. For other applications as well as benefits and limitations of the GNR, see Davidson and MacKinnon (1993).

8.5 Testing Linear Versus Log-Linear Functional Form[5]

In many economic applications where the explanatory variables take only positive values, econometricians must decide whether a linear or log-linear regression model is appropriate. In general, the linear model is given by

$$y_i = \sum_{j=1}^{k} \beta_j X_{ij} + \sum_{s=1}^{\ell} \gamma_s Z_{is} + u_i \quad i = 1, 2, \ldots, n \tag{8.84}$$

and the log-linear model is

$$\log y_i = \sum_{j=1}^{k} \beta_j \log X_{ij} + \sum_{s=1}^{\ell} \gamma_s Z_{is} + u_i \quad i = 1, 2, \ldots, n \tag{8.85}$$

with $u_i \sim \text{NID}(0, \sigma^2)$. Note that, the log-linear model is general in that only the dependent variable y and a subset of the regressors, i.e., the X variables are subject to the logarithmic transformation. Of course, one could estimate both models and compare their log-likelihood values. This would tell us which model fits best, but not whether either is a valid specification.

Box and Cox (1964) suggested the following transformation

$$B(y_i, \lambda) = \begin{cases} \frac{y_i^{\lambda} - 1}{\lambda} & \text{when} \quad \lambda \neq 0 \\ \log y_i & \text{when} \quad \lambda = 0 \end{cases} \tag{8.86}$$

where $y_i > 0$. Note that for $\lambda = 1$, as long as there is constant in the regression, subjecting the linear model to a Box-Cox transformation is equivalent to not transformation yields the log-linear regression. Therefore, the following Box-Cox model regression. Therefore, the following Box-Cox model

$$B(y_i, \lambda) = \sum_{j=1}^{k} \beta_j B(X_{ij}, \lambda) + \sum_{s=1}^{\ell} \gamma_s Z_{is} + u_i \tag{8.87}$$

encompasses as special cases the linear and log-linear models given in (8.84) and (8.85), respectively. Box and Cox (1964) suggested estimating these models by ML and using the LR test to test (8.84) and (8.85) against (8.87). However, estimation of (8.87) is computationally burdensome, see Davidson and MacKinnon (1993). Instead, we give an LM test involving a Double Length Regression (DLR) due to Davidson and MacKinnon (1985) that is easier to compute. In fact, Davidson and MacKinnon (1993, p. 510) point out that "everything that one can do with the Gauss-Newton Regression for nonlinear regression models can be done with the DLR for models involving transformations of the dependent variable." The GNR is not applicable in cases where the dependent variable is subjected to a nonlinear transformation, so one should use a DLR in these cases. Conversely, in cases where the GNR is valid, there is no need to run the DLR, since in these cases the latter is equivalent to the GNR.

For the linear model (8.84), the null hypothesis is that $\lambda = 1$. In this case, Davidson and MacKinnon suggest running a regression with 2n observations where the dependent variable has observations $(e_1/\widehat{\sigma}, \ldots, e_n/\widehat{\sigma}, 1, \ldots, 1)'$, i.e., the first n observations are the OLS residuals from (8.84) divided by the MLE of σ, where $\widehat{\sigma}^2_{mle} = e'e/n$. The second n observations are all equal to 1. The $2n$ observations for the regressors have typical elements:

for β_j: $X_{ij} - 1$	for $i = 1, \ldots, n$	and	0	for the second n elements
for γ_s: Z_{is}	for $i = 1, \ldots, n$	and	0	for the second n elements
for σ: $e_i/\widehat{\sigma}$	for $i = 1, \ldots, n$	and	-1	for the second n elements
for λ: $\sum_{j=1}^{k} \widehat{\beta}_j (X_{ij}\log X_{ij} - X_{ij} + 1) - (y_i \log y_i - y_{i+1})$				for $i = 1, \ldots, n$
		and	$\widehat{\sigma}\log y_i$	for the second n elements

The explained sum of squares for this DLR provides an asymptotically valid test for $\lambda = 1$. This will be distributed as χ^2_1 under the null hypothesis.

Similarly, when testing the log-linear model (8.85), the null hypothesis is that $\lambda = 0$. In this case, the dependent variable of the DLR has observations $(\widetilde{e}_1/\widetilde{\sigma}, \widetilde{e}_2/\widetilde{\sigma}, \ldots, \widetilde{e}_n/\widetilde{\sigma}, 1, \ldots, 1)'$, i.e., the first n observations are the OLS residuals from (8.85) divided by the MLE for σ, i.e., $\widetilde{\sigma}$ where $\widetilde{\sigma}^2 = \widetilde{e}'\widetilde{e}/n$. The second n observations are all equal to 1. The $2n$ observations for the regressors have typical elements:

for β_j: $\log X_{ij}$	for $i = 1, \ldots, n$	and	0	for the second n elements
for γ_s: Z_{is}	for $i = 1, \ldots, n$	and	0	for the second n elements
for σ: $\widetilde{e}_i/\widetilde{\sigma}$	for $i = 1, \ldots, n$	and	-1	for the second n elements
for λ: $\frac{1}{2}\sum_{j=1}^{k} \widetilde{\beta}_j (\log X_{ij})^2 - \frac{1}{2}(\log y_i)^2$		for $i = 1, \ldots, n$		
		and	$\widetilde{\sigma}\log y_i$	for the second n elements

The explained sum of squares from this DLR provides an asymptotically valid test for $\lambda = 0$. This will be distributed as χ^2_1 under the null hypothesis.

For the cigarette data given in Table 3.2, the linear model is given by $C = \beta_0 + \beta_1 P + \beta_2 Y + u$ whereas the log-linear model is given by $\log C = \gamma_0 + \gamma_1 \log P + \gamma_2 \log Y + \epsilon$ and the Box-Cox model is given by $B(C, \lambda) = \delta_0 + \delta_1 B(P, \lambda) + \delta_2 B(Y, \lambda) + \nu$, where $B(C, \lambda)$ is defined in (8.86). In this case, the DLR which tests the hypothesis that H_0; $\lambda = 1$, i.e., the model is linear, gives an explained sum of squares equal to 15.55. This is greater than a $\chi^2_{1,0.05} = 3.84$ and is therefore significant at the 5% level. Similarly the DLR that tests the hypothesis that H_0; $\lambda = 0$, i.e., the model is log-linear, gives an explained sum of squares equal to 8.86. This is also greater than $\chi^2_{1,0.05} = 3.84$ and is therefore significant at the 5% level. In this case, both the linear and log-linear models are rejected by the data.

Finally, it is important to note that there are numerous other tests for testing linear and log-linear models and the interested reader should refer to Davidson and MacKinnon (1993).

Notes

1. This section is based on Belsley, Kuh and Welsch (1980).
2. Other residuals that are linear unbiased with a scalar covariance matrix (LUS) are the BLUS residuals suggested by Theil (1971). Since we are explicitly dealing with time-series data, we use subscript t rather than i to index observations and T rather than n to denote the sample size.
3. Ramsey's (1969) initial formulation was based on BLUS residuals, but Ramsey and Schmidt (1976) showed that this is equivalent to using OLS residuals.
4. This section is based on Davidson and MacKinnon (1993, 2001).
5. This section is based on Davidson and MacKinnon (1993, pp. 502–510).

Problems

1. We know that $H = P_X$ is idempotent. Also, $(I_n - P_X)$ is idempotent. Therefore, $b'Hb \geq 0$ for any arbitrary vector b. Using these facts, show for $b' = (1, 0, \ldots, 0)$ that $0 \leq h_{11} \leq 1$. Deduce that $0 \leq h_{ii} \leq 1$ for $i = 1, \ldots, n$.
2. For the *simple regression with no constant* $y_i = x_i\beta + u_i$ for $i = 1, \ldots, n$
   (a) What is h_{ii}? Verify that $\sum_{i=1}^{n} h_{ii} = 1$.
   (b) What is $\widehat{\beta} - \widehat{\beta}_{(i)}$, see (8.13)? What is $s^2_{(i)}$ in terms of s^2 and e_i^2, see (8.18)? What is *DFBETAS*$_{ij}$, see (8.19)?
   (c) What are $DFFIT_i$ and $DFFITS_i$, see (8.21) and (8.22)?
   (d) What is Cook's distance measure $D_i^2(s)$ for this simple regression with no intercept, see (8.24)?
   (e) Verify that (8.27) holds for this simple regression with no intercept. What is $COVRATIO_i$, see (8.26)?
3. From the definition of $s^2_{(i)}$ in (8.17), substitute (8.13) in (8.17) and verify (8.18).
4. Consider the augmented regression given in (8.5) $y = X\beta^* + d_i\varphi + u$ where φ is a scalar and $d_i = 1$ for the i-th observation and 0 otherwise. Using the Frisch-Waugh Lovell Theorem given in section 7.3, verify that
   (a) $\widehat{\beta}^* = (X'_{(i)}X_{(i)})^{-1}X'_{(i)}y_{(i)} = \widehat{\beta}_{(i)}$.
   (b) $\widehat{\varphi} = (d_i'\bar{P}_X d_i)^{-1}d_i'\bar{P}_X y = e_i/(1 - h_{ii})$ where $\bar{P}_X = I - P_X$.
   (c) Residual Sum of Squares from (8.5) = (Residual Sum of Squares with d_i deleted) $- e_i^2/(1 - h_{ii})$.
   (d) Assuming Normality of u, show that the t-statistic for testing $\varphi = 0$ is $t = \widehat{\varphi}/s.e.(\widehat{\varphi}) = e_i^*$ as given in (8.3).

5. Consider the augmented regression $y = X\beta^* + \bar{P}_X D_p \varphi^* + u$, where D_p is an $n \times p$ matrix of dummy variables for the p suspected observations. Note that $\bar{P}_X D_p$ rather than D_p appear in this equation. Compare with (8.6). Let $e_p = D_p' e$, then $E(e_p) = 0$, $\text{var}(e_p) = \sigma^2 D_p' \bar{P}_X D_p$. Verify that

   (a) $\widehat{\beta}^* = (X'X)^{-1}X'y = \widehat{\beta}_{OLS}$ and

   (b) $\widehat{\varphi}^* = (D_p'\bar{P}_X D_p)^{-1} D_p' \bar{P}_X y = (D_p'\bar{P}_X D_p)^{-1} D_p' e = (D_p'\bar{P}_X D_p)^{-1} e_p$.

   (c) Residual Sum of Squares = (Residual Sum of Squares with D_p deleted) $- e_p'(D_p'\bar{P}_X)D_p^{-1}e_p$. Using the Frisch-Waugh Lovell Theorem show this residual sum of squares is the same as that for (8.6).

   (d) Assuming normality of u, verify (8.7) and (8.9).

   (e) Repeat this exercise for problem 4 with $\bar{P}_X d_i$ replacing d_i. What do you conclude?

6. Using the updating formula in (8.11), verify (8.12) and deduce (8.13).

7. Verify that Cook's distance measure given in (8.25) is related to $DFFITS_i(\sigma)$ as follows: $DFFITS_i(\sigma) = \sqrt{k} D_i(\sigma)$.

8. Using the matrix identity $\det(I_k - ab') = 1 - b'a$, where a and b are column vectors of dimension k, prove (8.27). **Hint:** Use $a = x_i$ and $b' = x_i'(X'X)^{-1}$ and the fact that $\det[X_{(i)}'X_{(i)}] = \det[\{I_k - x_i x_i'(X'X)^{-1}\}X'X]$.

9. For the cigarette data given in Table 3.2

   (a) Replicate the results in Table 8.2.

   (b) For the New Hampshire observation (NH), compute $\widetilde{e}_{NH}$, e^*_{NH}, $\widehat{\beta} - \widehat{\beta}_{(NH)}$, $DFBETAS_{NH}$, $DFFIT_{NH}$, $DFFITS_{NH}$, $D^2_{NH}(s)$, $COVRATIO_{NH}$, and $FVARATIO_{NH}$.

   (c) Repeat the calculations in part (b) for the following states: AR, CT, NJ and UT.

   (d) What about the observations for NV, ME, NM and ND? Are they influential?

10. For the Consumption-Income data given in Table 5.3, compute

    (a) The internal studentized residuals $\widetilde{e}$ given in (8.1).

    (b) The externally studentized residuals e^* given in (8.3).

    (c) Cook's statistic given in (8.25).

    (d) The leverage of each observation h.

    (e) The *DFFITS* given in (8.22).

    (f) The *COVRATIO* given in (8.28).

    (g) Based on the results in parts (a) to (f), identify the observations that are influential.

11. Repeat problem 10 for the 1982 data on earnings used in Chapter 4. This data is provided on the Springer web site as EARN.ASC.

12. Repeat problem 10 for the Gasoline data provided on the Springer web site as GASOLINE.DAT. Use the gasoline demand model given in Chapter 10, section 5. Do this for Austria and Belgium separately.

13. *Independence of Recursive Residuals.*

    (a) Using the updating formula given in (8.11) with $A = (X_t'X_t)$ and $a = -b = x_{t+1}'$, verify (8.31).

(b) Using (8.31), verify (8.32).

(c) For $u_t \sim \text{IIN}(0, \sigma^2)$ and w_{t+1} defined in (8.30) verify (8.33). **Hint:** define $v_{t+1}=\sqrt{f_{t+1}}w_{t+1}$. From (8.30), we have

$$v_{t+1} = \sqrt{f_{t+1}}w_{t+1} = y_{t+1} - x'_{t+1}\widehat{\beta}_t = x'_{t+1}(\beta - \widehat{\beta}_t) + u_{t+1} \quad \text{for} \quad t = k, \ldots, T-1$$

Since f_{t+1} is fixed, it suffices to show that $\text{cov}(v_{t+1}, v_{s+1}) = 0$ for $t \neq s$.

14. *Recursive Residuals are Linear Unbiased With Scalar Covariance Matrix (LUS).*

(a) Verify that the $(T-k)$ recursive residuals defined in (8.30) can be written in vector form as $w = Cy$ where C is defined in (8.34). This shows that the recursive residuals are linear in y.

(b) Show that C satisfies the three properties given in (8.35) i.e., $CX = 0$, $CC' = I_{T-k}$, and $C'C = \bar{P}_X$. Prove that $CX = 0$ means that the recursive residuals are unbiased with zero mean. Prove that the $CC' = I_{T-k}$ means that the recursive residuals have a scalar covariance matrix. Prove that $C'C = \bar{P}_X$ means that the sum of squares of $(T-k)$ recursive residuals is equal to the sum of squares of T least squares residuals.

(c) If the true disturbances $u \sim N(0, \sigma^2 I_T)$, prove that the recursive residuals $w \sim N(0, \sigma^2 I_{T-k})$ using parts (a) and (b).

(d) Verify (8.36), i.e., show that $RSS_{t+1} = RSS_t + w_{t+1}^2$ for $t = k, \ldots, T-1$ where $RSS_t = (Y_t - X_t\widehat{\beta}_t)'(Y_t - X_t\widehat{\beta}_t)$.

15. *The Harvey and Collier (1977) Misspecification t-Test as a Variable Additions Test.* This is based on Wu (1993).

(a) Show that the F-statistic for testing $H_0; \gamma = 0$ versus $\gamma \neq 0$ in (8.44) is given by

$$F = \frac{y'\bar{P}_X y - y'\bar{P}_{[X,z]}y}{y'\bar{P}_{[X,z]}y/(T-k-1)} = \frac{y'P_z y}{y'(\bar{P}_X - P_z)y/(T-k-1)}$$

and is distributed as $F(1, T-k-1)$ under the null hypothesis.

(b) Using the properties of C given in (8.35), show that the F-statistic given in part (a) is the square of the Harvey and Collier (1977) t-statistic given in (8.43).

16. For the Gasoline data for Austria given on the Springer web site as GASOLINE.DAT and the model given in Chapter 10, section 5, compute:

(a) The recursive residuals given in (8.30).

(b) The CUSUM given in (8.46) and plot it against r.

(c) Draw the 5% upper and lower lines given below (8.46) and see whether the CUSUM crosses these boundaries.

(d) The post-sample predictive test for 1978. Verify that computing it from (8.38) or (8.40) yields the same answer.

(e) The modified von Neuman ratio given in (8.42).

(f) The Harvey and Collier (1977) functional misspecification test given in (8.43).

17. *The Differencing Test in a Regression with Equicorrelated Disturbances.* This is based on Baltagi (1990). Consider the time-series regression

$$Y = \iota_T \alpha + X\beta + u \tag{1}$$

where ι_T is a vector of ones of dimension T. X is $T \times K$ and $[\iota_T, X]$ is of full column rank. $u \sim (0, \Omega)$ where Ω is positive definite. Differencing this model, we get

$$DY = DX\beta + Du \tag{2}$$

where D is a $(T-1) \times T$ matrix given below (8.50). Maeshiro and Wichers (1989) show that GLS on (1) yields through partitioned inverse:

$$\widehat{\beta} = (X'LX)^{-1}X'LY \tag{3}$$

where $L = \Omega^{-1} - \Omega^{-1}\iota_T(\iota_T'\Omega^{-1}\iota_T)^{-1}\iota_T'\Omega^{-1}$. Also, GLS on (2) yields

$$\widetilde{\beta} = (X'MX)^{-1}X'MY \tag{4}$$

where $M = D'(D\Omega D')^{-1}D$. Finally, they show that $M = L$, and GLS on (2) is equivalent to GLS on (1) as long as there is an intercept in (1).

Consider the special case of equicorrelated disturbances

$$\Omega = \sigma^2[(1-\rho)I_T + \rho J_T] \tag{5}$$

where I_T is an identity matrix of dimension T and J_T is a matrix of ones of dimension T.

(a) Derive the L and M matrices for the equicorrelated case, and verify the Maeshiro and Wichers result for this special case.

(b) Show that for the equicorrelated case, the differencing test given by Plosser, Schwert, and White (1982) can be obtained as the difference between the OLS and GLS estimators of the differenced equation (2). **Hint:** See the solution by Koning (1992).

18. For the 1982 data on earnings used in Chapter 4, provided as EARN.ASC on the Springer web site, (a) compute Ramsey's (1969) RESET. (b) Compute White's (1982) information matrix test given in (8.69) and (8.70).

19. Repeat problem 18 for the Hedonic housing data given on the Springer web site as HEDONIC.XLS.

20. Repeat problem 18 for the cigarette data given in Table 3.2.

21. Repeat problem 18 for the Gasoline data for Austria given on the Springer web site as GASOLINE.DAT. Use the model given in Chapter 10, section 5. Also compute the PSW differencing test given in (8.54).

22. Use the 1982 data on earnings used in Chapter 4, and provided on the Springer web site as EARN.ASC. Consider the two competing non-nested models

$$\begin{aligned} H_0; \log(\text{wage}) &= \beta_0 + \beta_1 ED + \beta_2 EXP + \beta_3 EXP^2 + \beta_4 WKS \\ &\quad + \beta_5 MS + \beta_6 FEM + \beta_7 BLK + \beta_8 UNION + u \end{aligned}$$

$$\begin{aligned} H_1; \log(\text{wage}) &= \gamma_0 + \gamma_1 ED + \gamma_2 EXP + \gamma_3 EXP^2 + \gamma_4 WKS \\ &\quad + \gamma_5 OCC + \gamma_6 SOUTH + \gamma_7 SMSA + \gamma_8 IND + \epsilon \end{aligned}$$

Compute:

(a) The Davidson and MacKinnon (1981) J-test for H_0 versus H_1.

(b) The Fisher and McAleer (1981) JA-test for H_0 versus H_1.

(c) Reverse the roles of H_0 and H_1 and repeat parts (a) and (b).

(d) Both H_0 and H_1 can be artificially nested in the model used in Chapter 4. Using the F-test given in (8.62), test for H_0 versus this augmented model. Repeat for H_1 versus this augmented model. What do you conclude?

23. For the Consumption-Income data given in Table 5.3,

(a) Test the hypothesis that the Consumption model is linear against a general Box-Cox alternative.

(b) Test the hypothesis that the Consumption model is log-linear against a general Box-Cox alternative.

24. Repeat problem 23 for the Cigarette data given in Table 3.2.

25. *RESET as a Gauss-Newton Regression.* This is based on Baltagi (1998). Davidson and MacKinnon (1993) showed that Ramsey's (1969) regression error specification test (RESET) can be derived as a Gauss-Newton Regression. This problem is a simple extension of their results. Suppose that the linear regression model under test is given by:

$$y_t = X_t'\beta + u_t \qquad t = 1, 2, \ldots, T \tag{1}$$

where β is a $k \times 1$ vector of unknown parameters. Suppose that the alternative is the nonlinear regression model between y_t and X_t:

$$y_t = X_t'\beta[1 + \theta(X_t'\beta) + \gamma(X_t'\beta)^2 + \lambda(X_t'\beta)^3] + u_t, \tag{2}$$

where θ, γ, and λ are unknown scalar parameters. It is well known that Ramsey's (1969) RESET is obtained by regressing y_t on X_t, $\widehat{y}_t^2$, $\widehat{y}_t^3$ and $\widehat{y}_t^4$ and by testing that the coefficients of all powers of $\widehat{y}_t$ are jointly zero. Show that this RESET can be derived from a Gauss-Newton Regression on (2), which tests $\theta = \gamma = \lambda = 0$.

References

This chapter is based on Belsley, Kuh and Welsch (1980), Johnston (1984), Maddala (1992) and Davidson and MacKinnon (1993). Additional references are the following:

Baltagi, B.H. (1990), "The Differencing Test in a Regression with Equicorrelated Disturbances," *Econometric Theory*, Problem 90.4.5, 6: 488.

Baltagi, B.H. (1998), "Regression Specification Error Test as A Gauss-Newton Regression," *Econometric Theory*, Problem 98.4.3, 14: 526.

Belsley, D.A., E. Kuh and R.E. Welsch (1980), *Regression Diagnostics* (Wiley: New York).

Box, G.E.P. and D.R. Cox (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Series B, 26: 211–252.

Brown, R.L., J. Durbin, and J.M. Evans (1975), "Techniques for Testing the Constancy of Regression Relationships Over Time," *Journal of the Royal Statistical Society* 37:149–192.

Chesher, A. and R. Spady (1991), "Asymptotic Expansions of the Information Matrix Test Statistic," *Econometrica* 59: 787–815.

Cook, R.D. (1977), "Detection of Influential Observations in Linear Regression," *Technometrics* 19:15–18.

Cook, R.D. and S. Weisberg (1982), *Residuals and Influences in Regression* (Chapman and Hall: New York).

Cox, D.R. (1961), "Tests of Separate Families of Hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1: 105–123.

Davidson, R., L.G. Godfrey and J.G. MacKinnon (1985), "A Simplified Version of the Differencing Test," *International Economic Review*, 26: 639–47.

Davidson, R. and J.G. MacKinnon (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, 49: 781–793.

Davidson, R. and J.G. MacKinnon (1985), "Testing Linear and Loglinear Regressions Against Box-Cox Alternatives," *Canadian Journal of Economics*, 18: 499–517.

Davidson, R. and J.G. MacKinnon (1992), "A New Form of the Information Matrix Test," *Econometrica*, 60: 145–157.

Davidson, R. and J.G. MacKinnon (2001), "Artificial Regressions," Chapter 1 in Baltagi, B.H. (ed.) *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Fisher, G.R. and M. McAleer (1981), "Alternative Procedures and Associated Tests of Significance for Non-Nested Hypotheses," *Journal of Econometrics*, 16: 103–119.

Gentleman, J.F. and M.B. Wilk (1975), "Detecting Outliers II: Supplementing the Direct Analysis of Residuals," *Biometrics*, 31: 387–410.

Godfrey, L.G. (1988), *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches* (Cambridge University Press: Cambridge).

Hall, A. (1987), "The Information Matrix Test for the Linear Model," *Review of Economic Studies*, 54: 257–263.

Harvey, A.C. (1976), "An Alternative Proof and Generalization of a Test for Structural Change," *The American Statistician*, 30: 122–123.

Harvey, A.C. (1990), *The Econometric Analysis of Time Series* (MIT Press: Cambridge).

Harvey, A.C. and P. Collier (1977), "Testing for Functional Misspecification in Regression Analysis," *Journal of Econometrics*, 6: 103–119.

Harvey, A.C. and G.D.A. Phillips (1974), "A Comparison of the Power of Some Tests for Heteroskedasticity in the General Linear Model," *Journal of Econometrics*, 2: 307–316.

Hausman, J. (1978), "Specification Tests in Econometrics," *Econometrica*, 46: 1251–1271.

Koning, R.H. (1992), "The Differencing Test in a Regression with Equicorrelated Disturbances," *Econometric Theory*, Solution 90.4.5, 8: 155–156.

Krämer, W. and H. Sonnberger (1986), *The Linear Regression Model Under Test* (Physica-Verlag: Heidelberg).

Krasker, W.S., E. Kuh and R.E. Welsch (1983), "Estimation for Dirty Data and Flawed Models," Chapter 11 in *Handbook of Econometrics*, Vol. I, eds. Z. Griliches and M.D. Intrilligator, Amsterdam, North-Holland.

Maeshiro, A. and R. Wichers (1989), "On the Relationship Between the Estimates of Level Models and Difference Models," *American Journal of Agricultural Economics*, 71: 432–434.

Orme, C. (1990), "The Small Sample Performance of the Information Matrix Test," *Journal of Econometrics*, 46: 309–331.

Pagan, A.R. and A.D. Hall (1983), "Diagnostic Tests as Residual Analysis," *Econometric Reviews*, 2: 159–254.

Pesaran, M.H. and M. Weeks (2001), "Nonnested Hypothesis Testing: A Overview," Chapter 13 in Baltagi, B.H. (ed.) *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Phillips, G.D.A. and A.C. Harvey (1974), "A Simple Test for Serial Correlation in Regression Analysis," *Journal of the American Statistical Association*, 69: 935–939.

Plosser, C.I., G.W. Schwert, and H. White (1982), "Differencing as a Test of Specification," *International Economic Review*, 23: 535–552.

Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis," *Journal of the Royal Statistics Society*, Series B, 31: 350–371.

Ramsey, J.B. and P. Schmidt (1976), "Some Further Results in the Use of OLS and BLUS Residuals in Error Specification Tests," *Journal of the American Statistical Association*, 71: 389–390.

Schmidt, P. (1976), *Econometrics* (Marcel Dekker: New York).

Theil, H. (1971), *Principles of Econometrics* (Wiley: New York).

Thursby, J. and P. Schmidt (1977), "Some Properties of Tests for Specification Error in a Linear Regression Model," *Journal of the American Statistical Association*, 72: 635–641.

Utts, J.M. (1982), "The Rainbow Test for Lack of Fit in Regression," *Communications in Statistics*, 11: 2801–2815.

Velleman, P. and R. Welsch (1981), "Efficient Computing of Regression Diagnostics," *The American Statistician*, 35: 234–242.

White, H. (1980), "A Heteroskedasticity–Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48: 817–838.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50: 1–25.

Wooldridge, J.M. (2001), "Diagnostic Testing," Chapter 9 in B.H. Baltagi (ed.) *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Wu, P. (1993), "Variable Addition Test," *Econometric Theory*, Problem 93.1.2, 9: 145–146.