Analysis of residuals in linear regressions [*]

Jean-Marie Dufour [†]
McGill University

First version: November 1983
Revised: March 2010, August 2011
This version: August 2011
Compiled: November 9, 2011, 13:47

[*] This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

[†] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

Contents

1. Graphical examination of the OLS residuals

After estimating a model, it is usually important to examine the residuals

$$\hat{\varepsilon}_i,\ i = 1, \ldots, T. \tag{1.1}$$

$\hat{\varepsilon}_i$ is an estimator of ε_i.
In principle, the residuals $\hat{\varepsilon}_i$ should behave approximately like i.i.d. random variables.
One should notice:

a) "very large" residuals;

b) systematic relations between residuals and certain variables;

c) heteroskedasticity in the errors;

d) autocorrelation in the errors.

2. Properties and standardization of OLS residuals

2.1. Basic structure of the residuals

$$
\begin{aligned}
y &= X\beta + \varepsilon \quad , \quad \varepsilon \sim N\left[0, \sigma^2 I_T\right] && (2.2) \\
y &: \ T \times 1, \quad X : T \times k, \quad \varepsilon : T \times 1 && (2.3)
\end{aligned}
$$

$$\hat{\varepsilon} = y - X\hat{\beta} = M_X \varepsilon \tag{2.4}$$

$$
\begin{aligned}
M_X &= I_T - X(X'X)^{-1}X' = I_T - H \\
H &= X(X'X)^{-1}X'
\end{aligned}
$$

$$E(\hat{\varepsilon}) = 0 \tag{2.5}$$

$$V(\hat{\varepsilon}) = \sigma^2 M_X \tag{2.6}$$

$$\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)' \tag{2.7}$$

$\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T$ do not have the same variance and are not independent.

$$X = \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_T' \end{bmatrix}$$

$$
\begin{aligned}
V(\hat{\varepsilon}_i) &= \sigma^2 \left[1 - X_i'(X'X)^{-1}X_i\right] = \sigma^2(1 - h_i) \le \sigma^2 \\
h_i &= X_i'(X'X)^{-1}X_i \\
Cov(\hat{\varepsilon}_i, \hat{\varepsilon}_j) &= \sigma^2(-h_{ij}) \ , \quad \text{for} \quad i \neq j \\
h_{ij} &= X_i'(X'X)^{-1}X_j
\end{aligned}
$$

Note $h_i = h_{ii}$ is the i-th diagonal element of H, hence

$$
\begin{aligned}
\sum_{i=1}^{T} h_i &= \mathrm{tr}[H] \\
&= \mathrm{tr}[X(X'X)^{-1}X'] \\
&= \mathrm{tr}[(X'X)^{-1}X'X] = \mathrm{tr}[I_K] = K, && (2.8)
\end{aligned}
$$

$$
\begin{aligned}
\sum_{i=1}^{T}(1 - h_i) &= \mathrm{tr}[I_T - H] \\
&= \mathrm{tr}(I_T) - \mathrm{tr}(H) = T - K, && (2.9)
\end{aligned}
$$

and the “average value” of h_i is

$$\frac{1}{T}\sum_{i=1}^{T} h_i = \frac{K}{T}\,. \tag{2.10}$$

Since

$$\hat{\varepsilon} = (I_T - H)\varepsilon,$$

we have

$$\hat{\varepsilon}_i = \varepsilon_i - \sum_{j=1}^{T} h_{ij}\varepsilon_j \quad , \quad i = 1, \ldots, T. \tag{2.11}$$

Each residual $\hat{\varepsilon}_i$ is the difference between the “true” error ε_i and a weighted average of all the errors.

2.2. Graphical methods

We usually proceed to a preliminary examination of the residuals by graphical methods.
A) For time series, we graph:

$$\hat{\varepsilon}_t \text{ against time } (t)\,. \tag{2.12}$$

B) More generally, we graph:

1. -$\hat{\varepsilon}_t$ against $\hat{y}_i$
2. $\hat{\varepsilon}_i$ against each explanatory variable

$$(x_{ki}, 1 \leq k \leq K) \tag{2.13}$$

   or against other variables.

2.3. Standardized and Studentized residuals

If one wishes to obtain residuals with the same variance, we can consider:

$$\tilde{\varepsilon}_i = \hat{\varepsilon}_i / [1 - h_i]^{1/2} \,, \quad i = 1, \ldots, T \,, \tag{2.14}$$

$$Var(\tilde{\varepsilon}_i) = \sigma^2 \,. \tag{2.15}$$

If we wish to make them more easily interpretable, we can divide by $s = \left[\hat{\varepsilon}\hat{\varepsilon} \big/ (T - K)\right]^{1/2}$:

$$r_i = \tilde{\varepsilon}_i / s = \frac{\hat{\varepsilon}_i}{s\,[1 - h_i]^{1/2}} \,, \quad i = 1, \ldots, T$$

"Internally Studentized residuals"

We wish to determine whether r_i is "large".
r_i does not follow a Student law.

Let

$$y_{(i)} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_T)', \quad i = 1, \ldots, T$$
$$X_{(i)} = [X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_T]'$$
$$\hat{\beta}_{(i)} = \left[X'_{(i)} X_{(i)}\right]^{-1} X'_{(i)} y_{(i)} \qquad \text{OLS estimator of } \beta \text{ based on } y \text{ without } y_i$$
$$\varepsilon_{(i)} = y_{(i)} - X_{(i)} \hat{\beta}_{(i)}$$
$$s^2_{(i)} = \varepsilon'_{(i)} \varepsilon_{(i)} / (T - K - 1)$$
$$d_i = X'_i \left[X'_{(i)} X_{(i)}\right]^{-1} X_i$$
$$v_i = y_i - X'_i \hat{\beta}_{(i)}$$

One can check easily that

$$\begin{aligned} Var(v_i) &= \sigma^2 [1 + d_i] \\ t_i &\equiv \frac{v_i}{s_{(i)} [1 + d_i]^{1/2}} \sim t(T - K - 1) \qquad \text{Externally Studentized residuals} \end{aligned}$$

We can also show that

$$\begin{aligned} h_i &\equiv X'_i (X'X)^{-1} X_i = \frac{d_i}{1 + d_i} \\ \hat{\varepsilon}_i &= \frac{v_{(i)}}{1 + d_i} \\ (T - K) s^2 &= (T - K - 1) s^2_{(i)} + (1 + d_i) t_i^2 \end{aligned}$$

hence

$$t_i = (T - K - 1)^{1/2} \frac{r_i}{(T - K - r_i^2)^{1/2}}$$

t_i is a monotonic nondecreasing transformation of r_i and

$$t_i \sim t(T-K-1). \tag{2.16}$$

To test whether a given residual $\hat{\varepsilon}_i$ is large, it is sufficient to compute

$$r_i = \hat{\varepsilon}_i / s\,[1-h_i]^{1/2} \tag{2.17}$$

$$t_i = (T-K-1)^{1/2} \frac{r_i}{\left[T-K-r_i^2\right]^{1/2}} \tag{2.18}$$

and see whether

$$|t_i| \geq t_{\alpha/2}(T-K-1)$$

This test is however only applicable for a given single residual.

3. Test for an outlier

If we observe one or several residuals which appear “large”, we may wish to declare that these correspond to “outlying observations”.

If we make a tests at level α on a residual $\hat{\varepsilon}_i$, we can reject the latter if

$$|t_i| \geq t_{\alpha/2}(T-K-1)\,.$$

Problem: If we make T tests, the probability of rejecting at least one observation as “outlying” (even if there is none) is larger than α.
To control the level, we adopt a rule of the following type:

$$\underset{1\leq i\leq T}{Max}\,|t_i| \geq c_\alpha$$

or

$$\underset{1\leq i\leq T}{Max}\,\left|t_i'\right| \geq c_\alpha^2$$

The observations which are declared “outlying” are those such that

$$|t_i| \geq c_\alpha \quad \text{or} \quad t_i^2 \geq c_\alpha^2\,.$$

Difficulty: The distribution of $Max\,|t_i|$ is difficult to determine.
However, we can show (using the Boole-Bonferroni inequality) that

$$c_\alpha^2 \leq F_{\alpha/T}\left(1, T-K-1\right) = \left[t_{\alpha/2T}\left(1, T-K-1\right)\right]^2\,.$$

If we declare an observation as outlying when

$$Max\,t_i^2 \geq F_{\alpha/T}\left(1, T-K-1\right)$$

or

$$Max\;|t_i| \geq t_{\alpha/2T}(t-K-1)\,.$$

4. Tests for heteroskedasticity

$$y_t = x_t'\beta + \varepsilon_t \quad , \quad t = 1, \dots, T \tag{4.19}$$

$$\sigma_t^2 = V(\varepsilon_t) = E(\varepsilon_t^2) \tag{4.20}$$

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_T^2 = \sigma^2 \qquad \text{(Homoskedasticity)} \tag{4.21}$$

Suppose we have reasons to believe that the variance increases with time.

$$Var(\varepsilon_t) > Var(\varepsilon_{t-1})$$

This can be informally checked by plotting the residuals $\hat{\varepsilon}_t$.

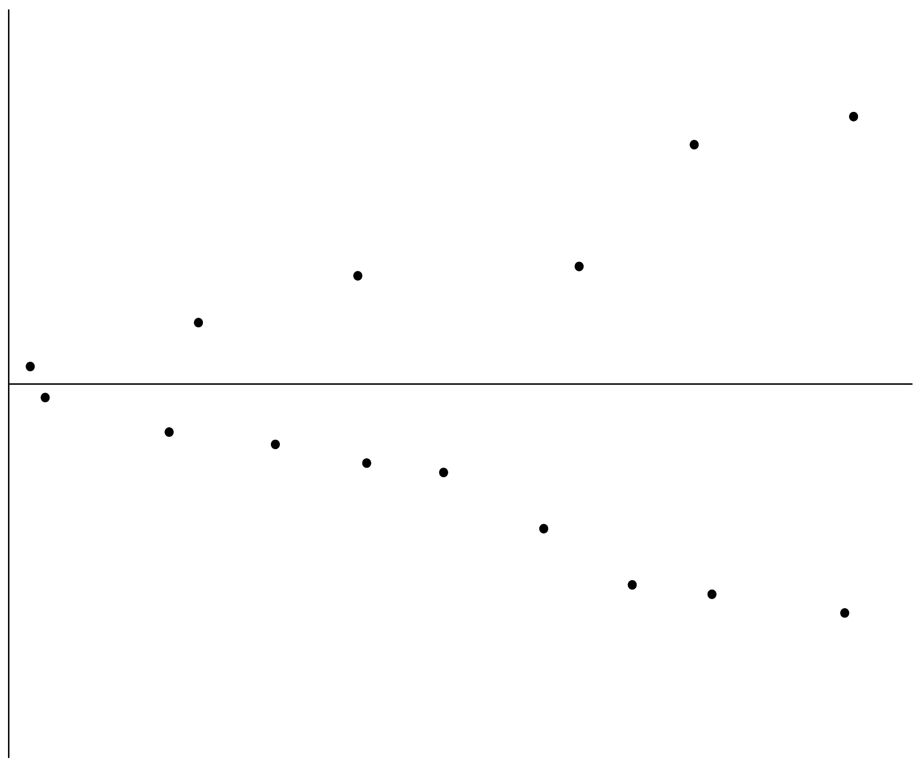

Figure 1. Residuals with increasing variance

Let us divide the sample in two parts:

$$\underbrace{t=1,\ldots,T_1}_{T_1 \text{ obs.}} \quad , \quad \underbrace{t=T_1+1,\ldots,T}_{T_2 \text{ obs.}} \qquad\qquad T_1+T_2=T \tag{4.22}$$
$$(e.g.\ T_1 = T/2 = T_2)$$

Under the hypothesis of an increasing variance, we have:

$$\begin{aligned} \frac{1}{T_1}E\left(\varepsilon_1^2+\cdots+\varepsilon_{T_1}^2\right) &< \frac{1}{T_2}E\left(\varepsilon_{T_1+1}^2+\cdots+\varepsilon_T^2\right) \\ E\left[\frac{1}{T_1}\sum_{t=1}^{T_1}\varepsilon_t^2\right] &< E\left[\frac{1}{T_2}\sum_{t=T_1+1}^{T_2}\varepsilon_t^2\right] \end{aligned}$$

If we knew $\varepsilon_1, \ldots, \varepsilon_T$, we could compute:

$$F = \frac{\sum\limits_{t=T_1+1}^{T}\varepsilon_t^2/T_2}{\sum\limits_{t=1}^{T_1}\varepsilon_t^2/T_1} = \frac{T_1}{T_2}\frac{\sum\limits_{t=T_1+1}^{T}\varepsilon_t^2}{\sum\limits_{t=1}^{T_1}\varepsilon_t^2} \sim F\left(T_2,\, T_1\right)$$

1. One-sided tests

   (a) Against σ_t^2 increasing, we reject H_0 when

   $$F > F_\alpha\left(T_2, T_1\right). \tag{4.23}$$

   (b) Against σ_t^2 decreasing, we reject H_0 when

   $$F \le F_{1-\alpha}\left(T_\alpha, T_1\right). \tag{4.24}$$

2. Two-sided test – We reject H_0 when

$$F \ge F_{\frac{\alpha}{2}}\left(T_2, T_1\right) \text{ or } F \le F_{1-\frac{\alpha}{2}}\left(T_2, T_1\right). \tag{4.25}$$

It is tempting to replace ε_t by $\hat{\varepsilon}_t$ in F.
Difficulty: the $\hat{\varepsilon}_t$ are not independent.
Goldfeld-Quandt solution:

$$\underset{T_1\times 1}{y_A} = X_A\beta + \varepsilon_A \Rightarrow \hat{\varepsilon}_A = y_A - X_A\hat{\beta}_A \quad , \quad \hat{\beta}_A = \left(X_A'X_A\right)^{-1} X_A y_A \tag{4.26}$$

$$\underset{T_2\times 1}{y_B} = X_B\beta + \varepsilon_B \Rightarrow \hat{\varepsilon}_B = y_B - X_B\hat{\beta}_B \quad , \quad \hat{\beta}_B = \left(X_B'X_B\right)^{-1} X_B y_B \tag{4.27}$$

$$\hat{\varepsilon}_A'\hat{\varepsilon}_A/\sigma^2 \sim \mathscr{X}^2(T_1 - K) \tag{4.28}$$

$$\hat{\varepsilon}_B'\hat{\varepsilon}_B/\sigma^2 \sim \mathscr{X}^2(T_2 - K) \tag{4.29}$$

$$F = \frac{\hat{\varepsilon}_B'\hat{\varepsilon}_B/(T_2-K)}{\hat{\varepsilon}_A'\hat{\varepsilon}_A/(T_1-K)} = \frac{T_1-K}{T_2-K_1}\frac{\hat{\varepsilon}_B'\hat{\varepsilon}_B}{\hat{\varepsilon}_A'\varepsilon_A'} \sim F(T_2-K, T_1, -K) \qquad \text{Goldfeld-Quandt test}$$

We reject H_0 when:

$$\left.\begin{array}{c} F \geq F_\alpha \\ F \leq F_{1-\alpha} \end{array}\right\} \text{One-sided tests}$$

$$\left. F \geq F_{\alpha/2} \ \text{ or } \ F \leq F_{1-\frac{\alpha}{2}} \right\} \text{Two-sided test}$$

Notes:

1. If we think that
$$E(\varepsilon_t^2) = \sigma^2 X_{tk}^2 \qquad t = 1, \dots, T\,,$$
we can reorder the observations according to the order of X_{tk}^2.

2. It is recommended to suppress a small group of observations in the middle to make the contrast more visible.

5. Tests against autocorrelation

Let $X_1, \ldots, X_T$ be i.i.d. random variables with distribution $N\left[\mu, \sigma^2\right]$.
We wish to test whether $X_1, \ldots, X_T$ are i.i.d. against

$$C(X_t, X_{t-1}) > 0 \quad , \; t = 2, \ldots, T \qquad \text{(positive autocorrelation)} \tag{5.30}$$

or

$$C(X_t, X_{t-1}) > 0 \quad , \; t = 2, \ldots, T \qquad \text{(negative autocorrelation).} \tag{5.31}$$

An alternative would be:

$$\text{e.g. } X_t = \rho X_{t-1} + \mu_\tau$$

The von Neumann statistic for testing the absence of serial dependence is:

$$VN = \frac{\sum_{t=2}^{T} (X_t - X_{t-1})^2/(T-1)}{\sum_{t=1}^{N} (X_t - \bar{X})^2/T} = \frac{\delta^2}{\hat{\sigma}^2}$$

where $\bar{X} = \sum_{t=1}^{T} X_t/T$.

If there positive (negative) autocorrelation, VN will tend take small (large) values.

One-sided tests:

reject H_0 (against positive autocorrelation) if $VN \leq C_\alpha^L$

reject H_0 (against negative autocorrelation) if $VN \geq C_\alpha^U$

Two-sided test:

reject H_0 if $VN \leq C_{\alpha/2}^L$ or $VN \geq C_{\alpha/2}^U$

Tables in Theil (1971, pp. 726-727).

If we knew $\varepsilon_1, \ldots, \varepsilon_T$,we could replace X_t by ε_τ and test whether the errors are autocorrelated.

$$VN = \frac{\sum_{t=2}^{T} (\varepsilon_t - \varepsilon_{t-1})^2/(T-1)}{\sum_{t=1}^{T} (\varepsilon_t - \overline{\varepsilon})^2/T}$$

Difficulty: the ε_τ are unknown.

Durbin-Watson proposed to use instead:

$$DW = \frac{\sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2} \quad \begin{array}{l} \text{vs. positive adductorrelation: } DW \leq d_\alpha \\ \text{vs. negative autocorrelation: } DW \geq d_\alpha \end{array}$$

$\hat{\varepsilon}_\tau, t = 1, \ldots, T$ are not independent (even under H_0):

$$\hat{\varepsilon} = \left[I - X(X'X)^{-1}X\right] \varepsilon = M \;\; \varepsilon$$

Problem: the distribution of DW depends on the matrix X. However, Durbin-Watson could establish bounds for the critical values.

For α given, we have (d_L, d_U) such that

$$\begin{array}{rcll} \text{if } DW & \leq & d_L & \text{we reject } H_0 \\ \text{if } DW & \geq & d_U & \text{we accept } H_0 \\ d_L & < & DW < d_U & \text{the test is inconclusive} \end{array}$$

Against an alternative of negative autocorrelation, we can compute $4 - DW$ and use the same test.

Generalizations to other lags

$$d_j = \sum_{t=j+1}^{T} (\hat{e}_t - \hat{e}_{t-j})^2 \,/\, \sum_{t=1}^{T} \hat{e}_t^2$$

1. $j = 4$; see Wallis (1972).

2. $j = 2, 3, 4$, with binary variables; seeVinod (1973).

3. Tests with a trend and seasonal dummies: King (1981).

References

King, M. L. (1981), 'The Durbin-Watson test for serial correlation: Bounds for regressions with trend and/or seasonal dummy variables', *Econometrica* **49**, 1571–1582.

Theil, H. (1971), *Principles of Econometrics*, John Wiley & Sons, New York.

Vinod, H. D. (1973), 'Generalisation of the Durbin-Watson statistic for higher order autoregressive processes', *Communication in Statistics, Theory and Methods* **2**, 115–144.

Wallis, K. J. (1972), 'Testing for fourth order autocorrelation in quarterly regression equation', *Econometrica* **40**, 617–636.