

Estimation Methods in Cases of Heteroskedastic or Autocorrelated
Disturbance Terms

I Possible Causes of Heteroskedasticity:

1. The values of the variables in the regression equation vary substantially in different observations. The variation in the omitted variables and the measurement errors which are jointly responsible for the disturbance term will be relatively small when y and x are small and relatively large when they are large (economic variables tending to move in size together).
2. In time series analysis if y and x are growing over time, then it may well happen that the variance of the disturbance term is also growing over time.

II Effects of Heteroskedasticity:

1. OLS is no longer efficient, since it treats all observations as being of equal importance, while there is much more information about the line to be had from some observations than from others, and when estimating the line more attention should be paid to the observations having small variances than those with larger variances.
2. The standard errors are biased, and thus hypothesis tests and confidence intervals based on the t -distribution are likely to be invalid.

III Correcting Heteroskedasticity (Weighted Least Squares)

The way to correct heteroskedasticity is to postulate a model for the disturbances, and on the basis of this model seek transformations of the data so that with **the transformed data** OLS is an efficient procedure.

Let us assume that the variance of $\varepsilon(t)$ is directly proportional to the variable $Z(t)$ and the model for heteroskedasticity is

$$\sigma_{\varepsilon}^2 = \sigma^2 \cdot Z(t), \quad (1)$$

where σ^2 is an **unknown** parameter and $Z(t)$ is a variable on which we have observations.

The full heteroskedastic model may now be written as

$$y(t) = a + bx(t) + \varepsilon(t), \quad (2)$$

where

$$E(\varepsilon(t)) = 0,$$

$$E(\varepsilon^2(t)) = \sigma_{\varepsilon(t)}^2 = \sigma^2 \cdot Z(t),$$

and $E(\varepsilon(t), \varepsilon(t - \tau)) = 0$ when $t \neq \tau$.

Now divide equation (2) by $\sqrt{Z(t)}$ to obtain

$$\frac{y(t)}{\sqrt{Z(t)}} = a \cdot \frac{1}{\sqrt{Z(t)}} + b \cdot \frac{x(t)}{\sqrt{Z(t)}} + \varepsilon^*, \quad (3)$$

where $\tilde{\varepsilon}(t) = \frac{\varepsilon(t)}{\sqrt{Z(t)}}$.

If we introduce the notation

$$\tilde{y}(t) = \frac{y(t)}{\sqrt{Z(t)}}, \hat{x}_o(t) = \frac{1}{\sqrt{Z(t)}}, \tilde{x}(t) = \frac{x(t)}{\sqrt{Z(t)}},$$

then the transformed model is

$$\tilde{y}(t) = a \cdot \tilde{x}_o(t) + b \cdot \tilde{x}(t) + \tilde{\varepsilon}(t). \quad (4)$$

The OLS estimates of the unknown parameters a and b in the transformed model are BLUE, because the disturbance term of the transformed model satisfies the classical disturbance assumptions:

$$(a) \quad E(\tilde{\varepsilon}(t)) = E\left(\frac{\varepsilon(t)}{\sqrt{Z(t)}}\right) = \frac{1}{\sqrt{Z(t)}} \cdot E(\varepsilon(t)) \text{ (because } Z(t) \text{ is not stochastic)} = 0 \text{ (because}$$

$$E(\varepsilon(t)) = 0),$$

$$(b) \quad E(\tilde{\varepsilon}^2(t)) = E\left(\frac{\varepsilon^2(t)}{Z(t)}\right) = \frac{1}{Z(t)} \cdot E(\varepsilon^2(t)) = \frac{1}{Z(t)} \cdot \sigma^2 \cdot Z(t) = \sigma^2,$$

$$(c) \quad E(\tilde{\varepsilon}(t), \tilde{\varepsilon}(t-\tau)) = E\left(\frac{\varepsilon(t)}{\sqrt{Z(t)}} \cdot \frac{\varepsilon(t-\tau)}{\sqrt{Z(t-\tau)}}\right) = \frac{1}{Z(t) \cdot Z(t-\tau)} \cdot E(\varepsilon(t), \varepsilon(t-\tau)) = 0.$$

Note: dividing by $\sqrt{Z(t)}$ actually means **weighting** the observations according to their relative importance (\Rightarrow the estimating procedure is termed **weighted least squares**).

IV Possible Causes of Residual (Serial) Autocorrelation:

Autocorrelation normally occurs in regression analysis using time series data. The disturbance $\varepsilon(t)$ may be regarded as summarizing all the influences on Y not accounted for by the explanatory variables X . As successive values of economic variables are nearly always highly correlated, it is reasonable to assume that the disturbances are autocorrelated.

V Main Effects of Autocorrelation:

1. A loss of efficiency. If disturbances are correlated, then previous values of the disturbance have some information to convey about the current disturbance. If this information is ignored, it is clear that the sample data are not being used with maximum efficiency.
2. If there is a positive autocorrelation, the standard errors of the coefficients will be biased downward. In the results from a particular sample, this is likely to show up in confidence intervals that are much narrower than they should be and values of R^2 that give an overly optimistic impression of the worth of the model for prediction. Parameter estimates associated with irrelevant variables may be highly significant.

VI Overcoming Autocorrelation (Generalized Least Squares)

Suppose that $y(t) = a + b \cdot x(t) + \varepsilon(t)$, (5)

where

$$\varepsilon(t) = \rho \cdot \varepsilon(t-1) + u(t). \quad (6)$$

The value of u in each observation is assumed to be independent of its value in all other observations. Assume for the moment that ρ is known, although in practice this is never the case. Our strategy is to transform (5) so that the disturbance for the transformed model has no autocorrelation. OLS is then used on the new model to produce parameters estimators that are BLUE.

We lag (5) by one period and multiply by ρ :

$$\rho \cdot y(t-1) = \rho \cdot a + b \cdot \rho \cdot x(t-1) + \rho \cdot \varepsilon(t-1), \quad t = \overline{2, T}. \quad (7)$$

Subtracting (7) from (5) produces

$$y(t) - \rho \cdot y(t-1) = a \cdot (1 - \rho) + b \cdot (x(t) - \rho \cdot x(t-1)) + \varepsilon(t) - \rho \cdot \varepsilon(t-1), \quad (8)$$

and since from (6) we may write

$$u(t) = \varepsilon(t) - \rho \cdot \varepsilon(t-1), \quad (9)$$

equation (8) can be rewritten as

$$y(t) - \rho \cdot y(t-1) = a \cdot (1 - \rho) + b \cdot (x(t) - \rho \cdot x(t-1)) + u(t).$$

(10)

If we introduce the notation

$$\tilde{y}(t) = y(t) - \rho \cdot y(t-1), \hat{x}_o(t) = 1 - \rho, \tilde{x}(t) = x(t) - \rho \cdot x(t-1),$$

then the transformed model is

$$\tilde{y}(t) = a \cdot \tilde{x}_o(t) + b \cdot \tilde{x}(t) + u(t). \quad (11)$$

The disturbance term for this model was originally assumed to exhibit no autocorrelation. But the number of observations to be used for fitting the line is only $T-1$ instead of T . This might cause the loss of efficiency that overweighs the gain due to overcoming autocorrelation.

Therefore it may be useful to regain the first observation $y(1)$, $x(1)$ with its residual variance σ_ε^2 .

The problem is that the latter is different from the residual variance in the remaining observations, σ_u^2 . So we have solved the problem of autocorrelation at the expense of introducing a special case of heteroskedasticity.

Using (6), we can express σ_ε^2 in terms of σ_u^2 :

$$\begin{aligned} \sigma_\varepsilon^2 &= \text{var}(\varepsilon) = \text{var}(\rho \cdot \varepsilon(t-1) + u(t)) = \rho^2 \cdot \text{var}(\varepsilon(t-1)) + \text{var}(u(t)) + 2 \cdot \rho \cdot \text{cov}(\varepsilon(t-1), u(t)) = \\ &= \rho^2 \cdot \sigma_\varepsilon^2 + \sigma_u^2, \end{aligned} \quad (12)$$

since $\varepsilon(t-1)$ and $u(t)$ are independent and since $\text{var} \varepsilon(t) = \text{var}(\varepsilon(t-1)) = \sigma_\varepsilon^2$. Hence

$$\sigma_\varepsilon^2 = \frac{\sigma_u^2}{1 - \rho^2}. \quad (13)$$

This allows us to eliminate heteroskedasticity by weighting the 1st observation by $\sqrt{1 - \rho^2}$.

The disturbance term in that observation becomes $\sqrt{1 - \rho^2} \cdot \varepsilon(1)$. Its variance is

$$\sigma_\varepsilon^2 = (1 - \rho^2) \cdot \sigma_\varepsilon^2 = \sigma_u^2 \quad (\text{see (13)}).$$

So after regaining the first observation and eliminating heteroskedasticity we obtain a classical linear regression model, which means that applying OLS will produce parameters estimates which are BLUE (Best Linear Unbiased Estimates).