Heteroskedasticity and Autocorrelation Consistent Standard Errors

Xiang Ao

January 27, 2009

1 Properties of OLS estimator

1.1 When is OLS estimator unbiased?

In a linear model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \tag{1}$$

the OLS estimator is

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Since $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$,

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}. \tag{2}$$

This makes

$$\mathrm{E}(\hat{\beta}) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{E}(\mathbf{u}|\mathbf{X}). \tag{3}$$

The condition that makes the OLS estimator unbiased is:

$$\mathrm{E}(\mathbf{u}|\mathbf{X}) = \mathbf{0}, \tag{4}$$

that is, the explanatory variables which form the columns of $\mathbf{X}$ are exogenous. This condition is weaker than the independence condition that u and X are independent.

In the context of cross-sectional data, this assumption is plausible. However, when we have time series data, the assumption becomes strong, because it assumes that the entire series of $\mathbf{X}$ has no relationship with the error term. In a time series context, this is hard to satisfy. The OLS estimator is biased if condition 4 is not satisfied.

For example, suppose we have a model

$$y_t = \beta_1 + \beta_2 y_{t-1} + u_t, \quad u_t \sim \mathrm{IID}(0, \sigma^2).$$

In this simple model, even if we assume that y_{t-1} and u_t are uncorrelated, OLS estimator is still biased. That is because condition 4 is not satisfied: y_{t-1} depends on u_{t-1}, u_{t-2} and so on. Assumption 4 does not hold for regressions with lagged dependent variables. Models with time series data are likely to violate assumption 4.

1.2 When is OLS estimator consistent?

For OLS estimator to be consistent, a much weaker condition is needed:

$$\mathrm{E}(u_t|X_t) = 0, \tag{5}$$

This condition is much weaker since it only assumes that the mean of current error term does not depend on the current predictors. Even a model with lagged dependent variable can easily satisfy this condition. Condition 5 is called predeterminedness condition, or say regressors are predetermined.

2 Estimation of Variance

If the OLS estimator is unbiased (that is, if $\mathbf{X}$ are exogenous), and suppose that data is generated by this model:

the linear regression models is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \quad \mathrm{E}(\mathbf{u}|\mathbf{X}) = \mathbf{0}, \quad \mathrm{E}(\mathbf{u}\mathbf{u}') = \mathbf{\Omega}, \tag{6}$$

that is, the error terms are independently but not identically distributed, then the following holds:

$$\begin{aligned} \mathrm{Var}(\hat{\beta}) &= \mathrm{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\ &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathrm{E}(\mathbf{u}\mathbf{u}'))\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \tag{7}$$

In the case of IID, Ω is identity matrix, and

$$\mathrm{var}(\hat{\beta}) = \sigma^{\mathbf{2}}(\mathbf{X}'\mathbf{X})^{-1}$$

In more general situations, the error terms are not IID.

2.1 White's estimator

In the case of heteroskedatic errors, $\mathbf{\Omega}$ is the error variance-covariance matrix with diagonal elements being σ_t^2 for t^{th} element, off-diagonal elements being zero.

If we know σ_t^2, then we would be able to estimate this "sandwich covariance matrix". But we don't.

$$\mathrm{Var}(\hat{\beta}) = \frac{\mathbf{1}}{\mathbf{n}}[\frac{\mathbf{1}}{\mathbf{n}}(\mathbf{X}'\mathbf{X})]^{-1}[\frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{\Omega}\mathbf{X}][\frac{\mathbf{1}}{\mathbf{n}}(\mathbf{X}'\mathbf{X})]^{-1}$$

Let y_t denote the tth observation on the dependent variable, and $x_t' = [1 \; x_{2t} \; \cdots \; x_{kt}]$ denote the tth row of the $\mathbf{X}$ matrix. Then

$$\mathbf{X}'\mathbf{\Omega}\mathbf{X} = \sum_{\mathbf{t=1}}^{\mathbf{n}} \sigma_{\mathbf{t}}^{\mathbf{2}}\mathbf{x_t}\mathbf{x_t}' \tag{8}$$

There are n distinct σ_t^2 to estimate, the problem seems hopeless: with n goes to infinity, the number of parameters need to estimate also goes to infinity. White (1980) shows that we don't need to do that: all we need to do is to get a consistent estimator for $\mathbf{X'\Omega X}$, which is $k \times k$ and symmetric. It has $\frac{1}{2}(k^2+k)$ distinct elements. The White estimator replaces the unknown σ_t^2 by $\hat{u}_t^2$, the estimated OLS residuals. This provides a consistent estimator of the variance matrix for the OLS coefficient vector and is particularly useful since it does not require any specific assumptions about the form of the heteroscedasticity. This type of estimator is also called heteroskedasticity-consistent covariance matrix estimator.

$$\begin{aligned} \hat{\text{Var}}(\hat{\beta}) &= \mathbf{\frac{1}{n}[\frac{1}{n}(X'X)]^{-1}[\frac{1}{n}X'\hat{\Omega}X][\frac{1}{n}(X'X)]^{-1}} \quad (9) \\ \hat{\mathbf{\Omega}} &= \text{diag}(\hat{\text{u}}_1^2, \hat{\text{u}}_2^2, \cdots, \hat{\text{u}}_\text{n}^2) \quad (10) \end{aligned}$$

2.2 Newey-West estimator

White's estimator deals with the situation that we have heteroskedasticity (a diagonal Σ) of unknown form. When we have serial correlation of unknown form (a non-diagonal Σ), we can estimate the variance-covariance matrix by a heteroskedasticity and autocorrelation consistent, or HAC, estimator. Newey-West estimator is the most popular HAC estimator. It's not as straightforward as White's estimator to illustrate, but I'll try to summarize.

2.2.1 Consistent Estimation of the Variance of the Sample Mean

Given a time series data set, suppose we are interested in estimating the mean vector (suppose we have more than one variable) and its variance. We know that given IID data, we can apply central limit theorem: sample mean is a consistent estimator of the population mean and it's variance can be calculated since asymptotically the sample mean conforms to a normal distribution and the variance can be estimated, relatively easily. However, in the case of time series data, autocorrelation usually exists. We may be concerned the CLT may not work in this case.

Fortunately, as proved in Hamilton (1994), if $\mathbf{y_t}$ is a covariance-stationary (meaning that the covariance is not a function of time) vector process, then the sample mean satisfies:

1. $$\bar{\mathbf{y}}_\mathbf{t} \to \mu,$$

2. $$\mathbf{S} = \lim_{T\to\infty} \mathbf{T} \cdot \mathbf{E}[(\bar{\mathbf{y}}_\mathbf{T} - \mu)(\bar{\mathbf{y}}_\mathbf{T} - \mu)'] = \sum_{\mathbf{v}=-\infty}^{\infty} \mathbf{\Gamma_v}.$$

   where $\mathbf{\Gamma_v}$ is the variance-covariance matrix for $\mathbf{y_t}$ and $\mathbf{y_{t-v}}$.

The first one says for a covariance-stationary vector process, the law of large numbers still holds. The second one is used to calculate the standard error.

If the data were generated by a vector MA(q) process, then

$$\mathbf{S} = \sum_{v=-q}^{q} \Gamma_v.$$

A natural estimate is

$$\hat{\mathbf{S}} = \hat{\mathbf{\Gamma}}_\mathbf{0} + \sum_{\mathbf{v=1}}^{\mathbf{q}} (\hat{\mathbf{\Gamma}}_\mathbf{v} + \hat{\mathbf{\Gamma}}'_\mathbf{v}),$$

where

$$\hat{\mathbf{\Gamma}}_\mathbf{v} = (\mathbf{1/T}) \sum_{\mathbf{t=v+1}}^{\mathbf{T}} (\mathbf{y_t} - \bar{\mathbf{y}})(\mathbf{y_{t-v}} - \bar{\mathbf{y}}).$$

This gives a consistent estimate of $\mathbf{S}$; however, it sometimes is not positive semidefinite.

Newey-West (1987) suggested putting in a weight:

$$\hat{\mathbf{S}} = \hat{\mathbf{\Gamma}}_\mathbf{0} + \sum_{v=1}^{q} (1 - \frac{v}{q+1})(\hat{\mathbf{\Gamma}}_\mathbf{v} + \hat{\mathbf{\Gamma}}'_\mathbf{v}),$$

where q is from the MA(q) process.

2.2.2 Newey-West estimator for linear regressions

Consider a linear regression model:

$$y_t = \mathbf{x}'_\mathbf{t} \beta + u_t$$

Suppose we have the OLS estimator $\mathbf{b_T}$, then

$$\sqrt{T}(\mathbf{b_t} - \beta) = [(1/T)\sum_{t=1}^{T} \mathbf{x_t x'_t}]^{-1}[(\sqrt{T})\sum_{t=1}^{T} \mathbf{x_t u'_t}]$$

The first term converges in probability to some constant. The second term is the sample mean of the vector $\mathbf{x_t u_t}$.

Under general conditions,

$$\sqrt{T}(\mathbf{b_t} - \beta) \rightarrow^L N(0, Q^{-1}SQ^{-1})$$

where S can be estimated by

$$\hat{\mathbf{S}}_\mathbf{T} = \hat{\mathbf{\Gamma}}_\mathbf{0T} + \sum_{v=1}^{q} (1 - \frac{v}{q+1})(\hat{\mathbf{\Gamma}}_{\mathbf{v},\mathbf{T}} + \hat{\mathbf{\Gamma}}'_{\mathbf{v},\mathbf{T}}),$$

where

$$\hat{\Gamma}'_{v,T} = (1/T) \sum_{t=v+1}^{T} (x_t \hat{u}_{t,T} \hat{u}_{t-v,T} x'_{t-v}),$$

where $\hat{u}_{t,T}$ is the OLS residual for data t in a sample of size T.

Overall, the variance of $\mathbf{b_T}$ is approximated by

$$\hat{\Sigma}_{NW} = [\sum_{t=1}^{T} x_t x'_t][\sum_{t=1}^{T} \hat{u}_t^2 x_t x'_t + \sum_{v=1}^{q} (1 - \frac{v}{q+1}) \sum_{t=v+1}^{T} (x_t \hat{u}_{t,T} \hat{u}_{t-v,T} x'_{t-v} + x_{t-v} \hat{u}_{t-v,T} \hat{u}_{t,T} x'_t)][\sum_{t=1}^{T} x_t x'_t]^{-1}$$

This estimation obviously depends on the selection of q, the lag length beyond which we are willing to assume that the autocorrelation of $x_t u_t$ and $x_{t-v} u_{t-v}$ is essentially zero. The rule of thumb for the selection of q is $0.75 \cdot T^{\frac{1}{3}}$. Newey-West (1994) has suggested a way to automatically select the bandwidth q. Here we omitted the discussion. Both Stata and R now also implement Newey-West (1994) estimator, with no need to specify q.

2.2.3 Implementation

Stata has *newey* and *newey*2 implemented for cross-sectional data. For panel data, it has *xtivreg*2 which implements Newey-West (1994) estimator with automatic bandwidth selection.

R has a library called *sandwich* which implements different *HAC* estimators, including Newey-West.