Heteroskedasticity and Autocorrelation

Hannu Kahra

Helsinki School of Economics

Spring 2008, First Period

Linear Model

- Consider a linear statistical model

$$y_t = \beta_1 + \beta_2 x_{i2} + ... + \beta_K x_{iK} + \varepsilon_i$$

or

$$y_i = x_i' \beta + \varepsilon_i .$$

- We can write the model using matrix notation and stack all the observations to write

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

where y and ε are N-dimensional vectors and X is of dimension $N \times K$.

- The OLS estimator of β has some important properties (it is *unbiased* and the *best linear unbiased estimator (BLUE)*). To obtain these properties, we have to make some assumptions about the error term ε_i and the explanatory variables x_i:
  - the so-called Gauss-Markov assumptions, and
  - the assumption that the disturbances ε_i are normally distributed.

Assumptions of the Classical Linear Regression Model

1. *Linearity*: $y_t = \beta_1 + \beta_2 x_{i2} + ... + \beta_K x_{iK} + \varepsilon_i$.
2. *Full rank*: The $N \times K$ sample data matrix X has full column rank.
3. *Exogeneity of the independent variables*: $E\left[\varepsilon_i | x_{j1}, x_{j2}, ..., x_{jK}\right] = 0$, $i, j = 1, ..., N$. There is no correlation between the the disturbances and the independent variables.
4. *Homoskedasticity and nonautocorrelation*: Each disturbance, ε_i, has the same finite variance, σ^2, and is uncorrelated with every other disturbance, ε_j, conditional on x.
5. *Stochastic or nonstochastic data*: $\{x_{i1}, x_{i2}, ..., x_{NK}\}$, $i = 1, ..., N$.
6. *Normal distribution*: The disturbances are normally distributed.

Gauss-Markov Assumptions

1. $E[\varepsilon_i] = 0; i = 1, ..., N$: The expected value of the error term is zero, which means that, *on average*, the regression line should be correct.
2. $\{\varepsilon_1, ..., \varepsilon_N\}$ and $\{x_1, ..., x_N\}$ are independent.
3. $V[\varepsilon_i] = \sigma^2; \forall i = 1, ..., N$: All error terms have the same variance (*homoskedasticity*).
4. $Cov[\varepsilon_i, \varepsilon_j] = 0; \forall i, j = 1, ..., N; i \neq j$: Zero correlation between different error terms that excludes any form of autocorrelation.

- Taken together, assumptions 1, 3 and 4 imply that the error terms are uncorrelated drawings from a distribution with expectation zero and constant variance σ^2.
- The essential Gauss-Markov assumptions (1-4) can be summarized as

$$E[\varepsilon|X] = E[\varepsilon_i] = 0, \text{ and} \quad (2)$$
$$V[\varepsilon|X] = V[\varepsilon_i] = \sigma^2 I_N, \quad (3)$$

where I_N is the $N \times N$ identity matrix.

Violation of (3)

- Both heteroskedasticity and autocorrelation imply that condition (3) does no longer hold.
  - Heteroskedasticity arises if different error terms do not have identical variances, so that the diagonal elements of the covariance matrix are not identical.
  - Autocorrelation almost excessively arises in cases where the data have a time dimension. It implies that the covariance matrix is nondiagonal such that different error terms are correlated. The reason could be persistence in the unexplained part of the model.
- Let us assume that the error covariance matrix can more generally be written as

$$V[\varepsilon|X] = \sigma^2\Psi. \tag{4}$$

- We obtain (for a given matrix X)

$$\begin{aligned} V[b|X] &= V\left[(X'X)^{-1}X'\varepsilon|X\right] = (X'X)^{-1}X'V[\varepsilon|X]X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'\Psi X(X'X)^{-1}, \end{aligned}$$

which reduces to the simpler expression $\sigma^2(X'X)^{-1}$ if $\Psi = I_N$.

Alternative BLUE Estimator

- We shall derive the best unbiased estimator (b) for β under assumption (4), assuming Ψ is completely known, by writing

$$\Psi^{-1} = P'P \tag{5}$$

  for some nonsingular square matrix P.
- Using (5) it is possible to write:

$$\begin{aligned} \Psi &= \left(P'P\right)^{-1} = P^{-1}\left(P'\right)^{-1} \\ P\Psi P' &= PP^{-1}\left(P'\right)^{-1}P' = I. \end{aligned}$$

- Consequently, it holds for the error term vector ε premultiplied by the transformation matrix P that

$$\begin{aligned} E\left[P\varepsilon|X\right] &= PE\left[\varepsilon|X\right] = 0, \\ V\left[P\varepsilon|X\right] &= PV\left[\varepsilon|X\right]P' = \sigma^2 P\Psi P' = \sigma^2 I. \end{aligned}$$

Alternative BLUE Estimator
(continued)

- Thus, $P\varepsilon$ satisfies the Gauss-Markov conditions. Consequently, we can transform the entire model by this P matrix to obtain

$$Py = PX\beta + P\varepsilon \text{ or } y^* = X^*\beta + \varepsilon^*,$$

where ε^* satisfies the Gauss-Markov conditions.

- We know that applying OLS in this transformed model produces the BLUE estimator for β.
- This, therefore is automatically the BLUE estimator for β in the original model with assumptions (2)-(4):

$$\widehat{\beta} = \left(X^{*\prime}X^*\right)^{-1} X^{*\prime}y^* = \left(X'\Psi^{-1}X\right)^{-1} X'\Psi^{-1}y. \qquad (6)$$

- It is referred to as the *generalized least squares* (**GLS**) *estimator*, that reduces to the OLS estimator if $\Psi = I$. The choice of P is irrelevant, only Ψ^{-1} matters.

Testing Strategy

- Heteroskedasticity poses potentially severe problems for inferences based on OLS.
- One can rarely be certain that the disturbances are heteroskedastic however, and unfortunately, what form the heteroskedasticity takes if they are.
- As such, it is useful to be able to test for homoskedasticity and, if necessary, modify our estimation procedures accordingly.
- Several types of tests have been suggested. They can be roughly grouped in descending order in terms of their generality and, as might be expected, in ascending order in terms of their power.
- The most commonly used tests (2) are based on the following strategy. OLS is a consistent estimator of β even in the presence of heteroskedasticity. As such, the OLS residuals will mimic, albeit imperfectly because of sampling variability, the heteroskedasticity of the true disturbances. Therefore, tests designed to detect heteroskedasticity will, in general, be applied to the OLS residuals.

White's Test

- To formulate most of the available tests, it is necessary to specify, at least in rough terms, the nature of the heteroskedasticity. It would be desirable to be able to test a general hypothesis of the form

$$\begin{aligned} H_0 &: \sigma_i^2 = \sigma^2 \text{ for all } i, \\ H_1 &: \text{Not } H_0. \end{aligned}$$

- Such a test has been suggested by White (1980). The correct covariance matrix for the least squares estimator is

$$V[b|X] = \sigma^2 \left[X'X\right]^{-1} \left[X'\Omega X\right] \left[X'X\right]^{-1},$$

which can be estimated using *White heteroskedasticity consistent estimator*

$$\begin{aligned} \text{Est. Asy. } V[b] &= \frac{1}{N}\left(\frac{1}{N}X'X\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^2 x_i x_i'\right)\left(\frac{1}{N}X'X\right)^{-1} \\ &= N\left(X'X\right)^{-1} S_0 (X'X)^{-1}. \end{aligned}$$

White's Test
(continued)

- The conventional estimator is $V[b] = s^2 [X'X]^{-1} = s^2 \left(\sum_{i=1}^{N} x_i x_i' \right)$.
- If there is no heteroskedasticity, then V will give a consistent estimator of $V[b|X]$, whereas if there is, then it will not.
- A simple operational version of the test is carried out by obtaining NR^2 in the regression of e_i^2 on a constant and all unique variables contained in x and all the squares and cross products of the variables in x.
- The statistic is asymptotically distributed as χ^2_{P-1}, where P-1 is the number of regressors in the equation, including the constant.
- The White test is extremely general. To carry it out, we need not make any specific assumptions about the nature of the heteroskedasticity.
- But, the White test is nonconstructive: if we reject the null hypothesis, then the result of the test gives no indication of what to do next.

The Breusch-Pagan/Godfrey LM Test

- The White test is a generalization of the test proposed by Breusch and Pagan (1980), who have devised a *Lagrange multiplier* test of the hypothesis that $\sigma_i^2 = \sigma^2 f\left(\alpha_0 + \alpha' z_i\right)$, where z_i is a vector of independent variables. The model is homoskedastic if $\alpha = 0$.
- The test can be carried out with a simple regression:

$$\text{LM} = \frac{1}{2} \text{ explained sum of squares in the OLS of } e_i^2 / \left(e'e/N\right) \text{ on } z_i.$$

- For computational purposes, let Z be the $N \times P$ matrix of observations on $(1, z_i)$, and let g be the vector of observations of $g_i = e_i^2 / \left(e'e/N\right) - 1$.
- Then

$$\text{LM} = \frac{1}{2}\left[g'Z\left(Z'Z\right)^{-1}Z'g\right].$$

- Under the null hypothesis of homoskedasticity, LM has a limiting χ^2 distribution with degrees of freedom equal to the variables in z_i.

Autocorrelation

- Next, we will have a look at another case where $V[\varepsilon] = \sigma^2 I$ is violated, viz. when the covariances between different error terms are not all equal to zero.
- The most relevant example of this occurs when two or more consecutive error are correlated, and we say that the error term is subject to *autocorrelation* or *serial correlation*.
- As long as it can be assumed that $E[\varepsilon|X] = 0$, the consequences of autocorrelation are similar to those of heteroskedasticity: OLS remains unbiased, but it becomes inefficient and its standard errors are estimated in the wrong way.
- Autocorrelation normally occurs only when using time series data. To stress this, we shall follow the literature and index the observations from $t = 1, 2, ..., T$ rather than from $i = 1, 2, ..., N$. The important difference is that now the order of the observations does matter and the index reflects a natural ordering.

Autocorrelation
(continued)

- Economic time series often display a "memory" in that variation around the regression function is not independent from one period to the next.
- The seasonally adjusted price and quantity series published by government agencies are examples.
- Time series data are usually homoskedastic, so $E\left[\varepsilon\varepsilon'|X\right] = V\left[\varepsilon|X\right] = \sigma^2\Omega = \Sigma$ might be

$$\sigma^2\Omega = \Sigma = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \cdots & \rho_{T-2} \\ & & & \\ \rho_{T-1} & \rho_{T-2} & \cdots & 1 \end{bmatrix}.$$

- The values that appear off the diagonal depend on the model used for the disturbances. In most cases, consistent with the notion of fading memory, the values decline as we move away from the diagonal.

Autocorrelation
(continued)

- In general, the error term ε_t picks up the influence of those variables affecting the dependent variables that have not been included in the model.
- There are alternative estimation approaches that can make better use of the characteristics (e.g., autocorrelation) of the models. In some cases only minimal assumption about Ω are needed.
- Tests for autocorrelation are very often interpreted as misspecification tests. Incorrect functional forms, omitted variables and an inadequate dynamic specification of the model may all lead to findings of autocorrelation.
- It is also possible to formulate parametric models that make specific assumptions about Ω. Estimators in this setting are some form of generalized least squares or maximum likelihood.
- There are many forms of autocorrelation and each one leads to a different structure for the error covariance matrix $V[\varepsilon]$.

First Order Autocorrelation

- The most popular form is known as the first-order autoregressive process. In this case the error term in

$$y_t = x_t'\beta + \varepsilon_t \tag{7}$$

is assumed to depend upon its predecessor as follows

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t, \tag{8}$$

where v_t is an error term with mean zero and constant variance σ_v^2 that exhibits no serial correlation.
- The parameters ρ and σ_v^2 are typically unknown, and, along with β we may wish to estimate them.
- To derive the covariance matrix of the error term vector ε, we need to make an assumption about the distribution of the initial period error, ε_1. Most commonly, it is assumed that ε_1 is mean zero with the same variance as all other ε_ts.
- This is consistent with the idea that the process has been operating for a long period in the past and that $|\rho| < 1$.

First Order Autocorrelation
(continued)

- When the condition $|\rho| < 1$ is satisfied we say that the first-order autoregressesive process is *stationary*.

Definition

A stationary process is such that the mean, variance and covariances of ε_t do not change over time.

- Imposing stationarity it easily follows from

$$E[\varepsilon_t] = \rho E[\varepsilon_{t-1}] + E[v_t]$$

that $E[\varepsilon_t] = 0$. Further from

$$V[\varepsilon_t] = V[\rho\varepsilon_{t-1} + v_t] = \rho^2 V[\varepsilon_{t-1}] + \sigma_v^2,$$

we obtain that the variance of ε_t, denoted as σ_ε^2, is given by

$$\sigma_\varepsilon^2 = V[\varepsilon] = \frac{\sigma_v^2}{1-\rho^2}.$$

First Order Autocorrelation

(continued)

- The nondiagonal elements in the variance-covariance matrix of ε follow from

$$\text{cov}\left[\varepsilon_t, \varepsilon_{t-1}\right] = E\left[\varepsilon_t \varepsilon_{t-1}\right] = \rho E\left[\varepsilon_{t-1}^2\right] + E\left[\varepsilon_{t-1} v_t\right] = \rho \frac{\sigma_v^2}{1-\rho^2}.$$

- The covariance between error terms two periods apart is

$$\text{cov}\left[\varepsilon_t, \varepsilon_{t-2}\right] = E\left[\varepsilon_t \varepsilon_{t-2}\right] = \rho E\left[\varepsilon_{t-1} \varepsilon_{t-2}\right] + E\left[\varepsilon_{t-2} v_t\right] = \rho^2 \frac{\sigma_v^2}{1-\rho^2}.$$

- In general we have, for non-negative values of s,

$$E\left[\varepsilon_t \varepsilon_{t-s}\right] = \rho^s \frac{\sigma_v^2}{1-\rho^2}.$$

First Order Autocorrelation
(continued)

- Looking at (7) and (8), it is immediately apparent which transformation is appropriate.
- Because $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$, where v_t satisfies the Gauss-Markov condition, it is obvious that a transformation like $\varepsilon_t - \rho\varepsilon_{t-1}$ will generate homoskedastic non-autocorrelated errors.
- That is, all observations should be transformed as $y_t - \rho y_{t-1}$ and $x_t - \rho x_{t-1}$. Consequently, the transformed model is given by

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\beta + v_t, \quad t = 2, 3, ..., T.$$

- Because the transformed model satisfies the Gauss-Markov conditions, estimation with OLS yields the GLS estimator, assuming ρ is known.

Unknown Correlation

- In practise it is uncommon that the value of ρ is known. In that case we have to estimate it. Starting from the AR(1) model

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t,$$

  where v_t satisfies the usual assumptions, it seems natural to estimate ρ from a regression of the OLS residual e_t on e_{t-1}.
- The resulting OLS estimator for ρ is given by

$$\widehat{\rho} = \left(\sum_{t=2}^{T} e_{t-1}^2 \right)^{-1} \left(\sum_{t=2}^{T} e_t e_{t-1} \right).$$

- While this estimator for ρ is typically biased, it is a consistent estimator for ρ under weak regularity conditions. If we use $\widehat{\rho}$ instead of ρ to compute the feasible GLS (FGLS) estimator $\widehat{\beta}^*$, the BLUE property is no longer retained.

Unknown Correlation
(continued)

- Under the same conditions as before, it holds that the FGLS estimator $\widehat{\beta}^*$ is asymptotically equivalent to the GLS estimator $\widehat{\beta}$. This means that for large sample sizes we can ignore the fact that ρ is estimated.
- A related estimation procedure is the so-called iterative Cochrane-Orcutt (1949) procedure.
  - In this procedure ρ and β are recursively estimated under convergence, i.e. having estimated β with FGLS (by $\widehat{\beta}^*$), the residuals are recomputed and ρ is estimated again using the residuals from the FGLS step.
  - With this new estimate of ρ, FGLS is applied again and one obtains a new estimate of β.
  - This procedure goes on until convergence, i.e. until both the estimate for ρ and the estimate for β do not change anymore.
- Unlike the heteroskedastic model, iterating when there is autocorrelation does not, however, produce the *maximum likelihood* (ML) estimator.

Unknown Correlation
(continued)

- Maximum likelihood estimators can be obtained by maximizing the log-likelihood with respect to β, σ_e^2 and ρ.
- The log-likelihood function may be written

$$\ln \mathcal{L} = -\frac{\sum_{t=1}^{T} e_t^2}{2\sigma_e^2} + \frac{1}{2} \ln \left(1 - \rho^2\right) - \frac{T}{2} \left(\ln 2\pi + \ln \sigma_e^2\right).$$

- In practice, maximum likelihood estimators are probably the most common choices.

Testing for First Order Autocorrelation

- When $\rho = 0$ no autocorrelation is present and OLS is BLUE. If $\rho \neq 0$ inferences based on the OLS estimator will be misleading because standard errors will be based on the wrong formula.
- Therefore, it is common practice with time series data to test for autocorrelation in the error term.
- Suppose we want to test for the first order autocorrelation indicated by $\rho \neq 0$ in (8).
- Next, we will examine alternative tests for autocorrelation. The first set of tests are relatively simple and based on asymptotic approximations, while the last test has a known small sample distribution.

Asymptotic Tests

- The OLS results from (7) provide useful information about the possible presence of serial correlation in the equation's error term.
- An intuitively appealing starting point is to consider the regression of the OLS residual e_t upon its lag e_{t-1}. This regression may be done with or without an intercept term.
- The auxiliary regression not only produces an estimate for the first order autocorrelation coefficient, $\widehat{\rho}$, but also routinely provides a standard error to this estimate.
- In the absence of lagged dependent variables in (7), the corresponding t-test is asymptotically valid. In fact, the resulting test statistic can be shown to be approximately equal to

$$t \approx \sqrt{T}\widehat{\rho},$$

which provides an alternatively way of computing the test statistic.

Asymptotic Tests
(continued)

- Consequently, at the 5% significance level we reject the null hypothesis of no autocorrelation against a two sided alternative if $|t| > 1.96$.
- If the alternative hypothesis is positive autocorrelation $(\rho > 0)$, which is often expected a priori, the null hypothesis is rejected at the 5% level if $t > 1.64$.
- Another alternative is based upon the R^2 of the auxiliary regression (including an intercept term). If we take the R^2 of this regression and multiply it by the effective number of observations $T - 1$ we obtain a test statistic that, under the null hypothesis, has a χ^2 distribution with one degree of freedom.
- Clearly an R^2 close to zero in this regression implies that lagged residuals are not explaining current residuals and a simple way to test $\rho = 0$ is by computing $(T - 1)R^2$.

Asymptotic Tests
(continued)

- This test is a special case of the Breusch (1978)–Goodfrey (1978) Lagrange multiplier test and is easily extended to higher orders of autocorrelation (by including additional lags of the residuals and allowing for ether an AR(p) or an MA(q) process in the residuals.
- The test is a Lagrange multiplier test of H_0: no autocorrelation versus H_1: $\varepsilon_1 = AR(p)$ or $\varepsilon_1 = MA(p)$. The same test is used for either structure. The test statistic is

$$\text{LM} = T\left(\frac{e'X_0\left(X_0'X_0\right)^{-1}X_0'e}{e'e}\right) = TR_0^2,$$

where X_0 is the original X matrix augmented by P additional columns containing the lagged OLS residuals, $e_{t-1}, ..., e_{t-p}$.

- The test can be carried out simply by regressing the OLS residuals e_t on x_{t0} (filling in missing values for lagged residuals with zeros) and referring TR_0^2 to the tabled critical value for the χ^2 distribution with p degrees of freedom.

D-W Test

- The Durbin-Watson (1950) statistic was the first formal procedure developed for testing for autocorrelation using the OLS residuals. The test statistic is

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} = 2(1-\rho) - \frac{e_1^2 + e_T^2}{\sum_{t=1}^{T} e_t^2}.$$

- If the sample is reasonably large, then the last term will be negligible, leaving $d \approx 2\,(1-\rho)$.
- Usable critical values depend on T (sample size) and K (number of variables in the regression) and the tables are presented at the end of many econometrics books.
- The true critical value d_{crit} is between the bounds that are tabulated, that is $d_L < d_{crit} < d_U$. Under H_0 we thus have that (at the 5% level)

$$P\,(d < d_L) \leq P\,(d < d_{crit}) = 0.05 \leq P\,(d < d_U)\,.$$

D-W Test
(continued)

- As a rule of thumb, a value $d \approx 2$ is an indication of $\rho = 0$ in the OLS residuals. Values substantially less than 2 are an indication of positive autocorrelation, corresponding to $\widehat{\rho} > 0$.
- For a one-sided test against positive autocorrelation, there are three possibilities:
  1. d is less than d_L. In this case, it is certainly lower than the true critical value d_{crit}, so you would reject H_0.
  2. d is larger than d_U. In this case, it is certainly larger than d_{crit} and you would not reject H_0.
  3. d lies between d_L and d_U. In this case it might be larger or smaller than the critical value. Because you cannot tell which, you are unable to accept or reject H_0. This is the so-called 'inconclusive region'.
- The existence of the inconclusive region and the requirement that the Gauss-Markov conditions, including normality of the error terms, are satisfied are important drawbacks of the Durbin-Watson test.

Box-Pierce and Ljung-Box Tests

- An alternative test that is asymptotically equivalent to the LM test when the null hypothesis, $\rho = 0$, is true and when X does not contain lagged values of y is due to Box and Pierce (1970).
- The Q *test* is carried out by referring

$$Q = T \sum_{j=1}^{p} r_j^2,$$

where $r_j = \left(\sum_{t=j+1}^{T} e_t e_{t-j}\right) / \left(\sum_{t=1}^{T} e_t^2\right)$, to the critical values of the χ^2 table with p degrees of freedom.

- A refinement suggested by Ljung and Box (1979) is

$$Q' = T\,(T+2) \sum_{j=1}^{p} \frac{r_j^2}{T-j}.$$

Newey-West Autocorrelation Consistent Covariance Estimator

- Let us consider the basic model

$$y_t = x_t'\beta + \varepsilon_t,$$

  where ε_t is subject to autocorrelation.
- We can choose to apply the GLS approach or apply ordinary OLS while adjusting its standard errors. The latter is particularly useful when the correlation between ε_t and ε_{t-s} can be argued to be (virtually) zero after some lag length H and/or when the conditions for consistency of the GLS estimator happen to be violated.
- If $E\left[x_t\varepsilon_t\right] = 0$ and $E\left[\varepsilon_t\varepsilon_{t-s}\right] = 0$ for $s = H, H+1, ...,$ the OLS estimator is consistent and its covariance matrix can be estimated by

$$\widehat{V}^*[b] = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} TS^* \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}. \tag{9}$$

Newey-West Autocorrelation Consistent Covariance Estimator

(continued)

- S^* is defined by

$$S^* = \frac{1}{T}\sum_{t=1}^{T} e_t^2 x_t x_t' + \frac{1}{T}\sum_{j=1}^{H-1} w_j \sum_{s=j+1}^{T} e_s e_{s-j}\left(x_s x_{s-j}' + x_{s-j} x_s'\right).$$

- We will obtain the so-called White covariance matrix if $w_j = 0$, so that (9) is a generalization. In the standard case $w_j = 1$, but this may lead to an estimated covariance matrix in finite samples that is not positive definite.
- To prevent this, it is common to use Bartlett weights, as suggested by Newey and West (1987). These weights decrease linearly with j as $w_j = 1 - j/H$.
- The use of such as set of weights is compatible with the idea that the impact of the autocorrelation of order j diminishes with $|j|$.

Newey-West Autocorrelation Consistent Covariance Estimator

(continued)

- Standard errors computed from (9) are referred to as *heteroskedasticity-and-autocorrelation-consistent (HAC) standard errors* or simply *Newey-West standard errors*.
- These kind of estimators are now standard features in modern statistical and econometrics computer programs.[1]
- There is a final problem to be solved. It must be determined in advance how large H is to be. In general, there is little theoretical guidance. Current practice specifies $H \approx T^{1/4}$.

[1] The *sandwich* package in R contains procedures for heteroskedasticity-consistent (HC) and heteroskedasticity-and-autocorrelation-consistent (HAC) covariance matrix estimation.