Financial Econometrics

UNC CHARLOTTE

BELK COLLEGE *of* BUSINESS

Lecture 2
Linear Regression

2.1 Simple Linear Regression Model

2.1.1 Simple Regression: Constant and One Regressor

The simplest regression model is

$$y_t = \beta_0 + \beta_1 x_t + u_t, \text{ where } \mathrm{E}\, u_t = 0 \text{ and } \mathrm{Cov}(x_t, u_t) = 0, \tag{2.1}$$

where we can observe (have data on) the dependent variable y_t and the regressor x_t but not the residual u_t. In principle, the residual should account for all the movements in y_t that we cannot explain (by x_t).

Note the two very important assumptions: (*i*) the mean of the residual is zero; and (*ii*) the residual is not correlated with the regressor, x_t. If the regressor summarizes all the useful information we have in order to describe y_t, then the assumptions imply that we have no way of making a more intelligent guess of u_t (even after having observed x_t) than that it will be zero.

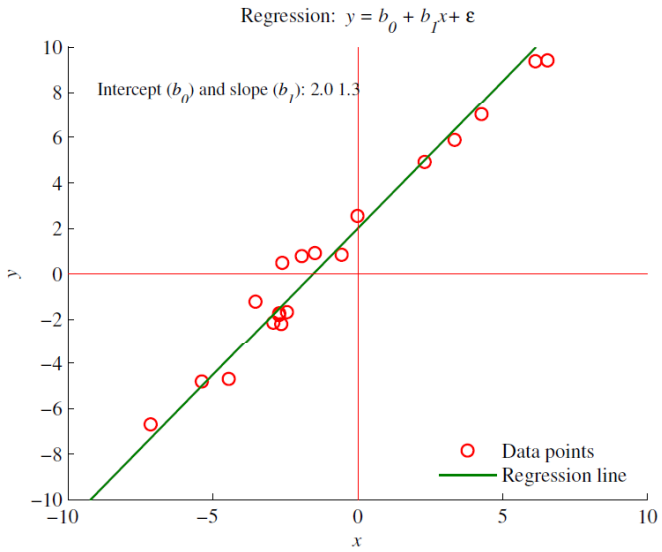
Least Squares Estimator

Suppose you do not know β_0 or β_1, and that you have a sample of data: y_t and x_t for $t = 1, ..., T$. The LS estimator of β_0 and β_1 minimizes the loss function

$$\sum_{t=1}^{T}(y_t - b_0 - b_1 x_t)^2 = (y_1 - b_0 - b_1 x_1)^2 + (y_2 - b_0 - b_1 x_2)^2 + ... \qquad (2.2)$$

by choosing b_0 and b_1 to make the loss function value as small as possible. The objective is thus to pick values of b_0 and b_1 in order to make the model fit the data as closely as possible—where close is taken to be a small variance of the unexplained part (the residual). See Figure 2.1 for an illustration.

Illustration of Least Squares Fit



First-Order Conditions

The first order conditions for a minimum are that the derivatives of this loss function with respect to b_0 and b_1 should be zero. Notice that

$$\frac{\partial}{\partial b_0}(y_t - b_0 - b_1 x_t)^2 = -2(y_t - b_0 - b_1 x_t)1 \tag{2.3}$$

$$\frac{\partial}{\partial b_1}(y_t - b_0 - b_1 x_t)^2 = -2(y_t - b_0 - b_1 x_t)x_t. \tag{2.4}$$

Let $(\hat{\beta}_0, \hat{\beta}_1)$ be the values of (b_0, b_1) where that is true

$$\frac{\partial}{\partial \beta_0}\sum_{t=1}^{T}(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)^2 = -2\sum_{t=1}^{T}(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)1 = 0 \tag{2.5}$$

$$\frac{\partial}{\partial \beta_1}\sum_{t=1}^{T}(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)^2 = -2\sum_{t=1}^{T}(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)x_t = 0, \tag{2.6}$$

which are two equations in two unknowns ($\hat{\beta}_0$ and $\hat{\beta}_1$), which must be solved simultaneously. These equations show that both the constant and x_t should be *orthogonal* to the fitted residuals, $\hat{u}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t$. This is indeed a defining feature of LS and can be seen as the sample analogues of the assumptions in (2.1) that $\mathrm{E}\, u_t = 0$ and $\mathrm{Cov}(x_t, u_t) = 0$.

Suppose Sample Means are Zero

When the means of y and x are zero, then we can disregard the constant. In this case, (2.6) with $\hat{\beta}_0 = 0$ immediately gives

$$\sum_{t=1}^{T} y_t x_t = \hat{\beta}_1 \sum_{t=1}^{T} x_t x_t \text{ or}$$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{T} y_t x_t / T}{\sum_{t=1}^{T} x_t x_t / T}. \tag{2.7}$$

In this case, the coefficient estimator is the sample covariance (recall: means are zero) of y_t and x_t, divided by the sample variance of the regressor x_t (this statement is actually true even if the means are not zero and a constant is included on the right hand side—just more tedious to show it).

Goodness of Fit

The quality of a regression model is often measured in terms of its ability to explain the movements of the dependent variable.

Let $\hat{y}_t$ be the fitted (predicted) value of y_t. For instance, with (2.1) it would be $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$. If a constant is included in the regression (or the means of y and x are zero), then a check of the *goodness of fit* of the model is given by

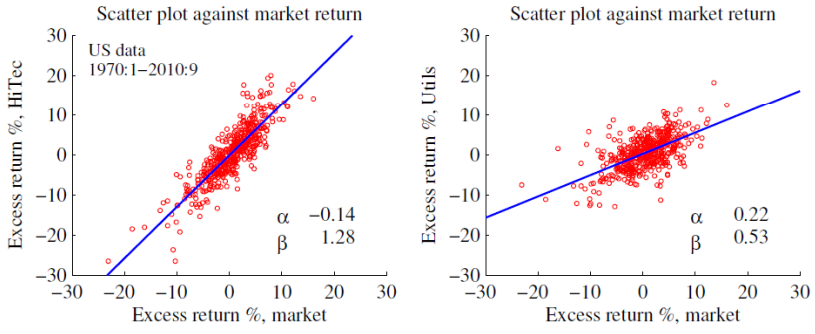
$$R^2 = \text{Corr}(y_t, \hat{y}_t)^2. \tag{2.8}$$

This is the squared correlation of the actual and predicted value of y_t.

To understand this result, suppose that x_t has no explanatory power, so R^2 should be zero. How does that happen? Well, if x_t is uncorrelated with y_t, then the numerator in (2.7) is zero so $\hat{\beta}_1 = 0$. As a consequence $\hat{y}_t = \hat{\beta}_0$, which is a constant—and a constant is always uncorrelated with anything else (as correlations measure comovements around the means).

Example Using Tech and Utility Stocks

	HiTec	Utils
constant	-0.14	0.22
	(-0.87)	(1.42)
market return	1.28	0.53
	(32.27)	(12.55)
R2	0.75	0.35
obs	489.00	489.00



Outliers

There is of course nothing sacred about the quadratic loss function. Instead of (2.2) one could, for instance, use a loss function in terms of the absolute value of the error $\Sigma_{t=1}^{T} |y_t - \beta_0 - \beta_1 x_t|$. This would produce the Least Absolute Deviation (LAD) estimator. It is typically less sensitive to outliers. This is illustrated in Figure 2.7. However, LS is by far the most popular choice. There are two main reasons: LS is very easy to compute and it is fairly straightforward to construct standard errors and confidence intervals for the estimator. (From an econometric point of view you may want to add that LS coincides with maximum likelihood when the errors are normally distributed.)

OLS versus LAD

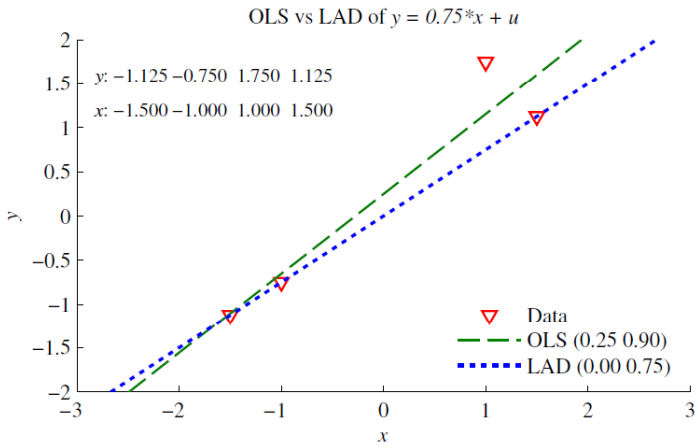


Figure 2.7: Data and regression line from OLS and LAD

OLS Estimator is a Random Variable

Note that the estimated coefficients are random variables since they depend on which particular sample that has been "drawn." This means that we cannot be sure that the estimated coefficients are equal to the true coefficients (β_0 and β_1 in (2.1)). We can calculate an estimate of this uncertainty in the form of variances and covariances of $\hat{\beta}_0$ and $\hat{\beta}_1$. These can be used for testing hypotheses about the coefficients, for instance, that $\beta_1 = 0$.

To see where the uncertainty comes from consider the simple case in (2.7). Use (2.1) to substitute for y_t (recall $\beta_0 = 0$)

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{t=1}^{T} x_t \left(\beta_1 x_t + u_t\right)/T}{\sum_{t=1}^{T} x_t x_t/T} \\ &= \beta_1 + \frac{\sum_{t=1}^{T} x_t u_t/T}{\sum_{t=1}^{T} x_t x_t/T},\end{aligned} \tag{2.9}$$

so the OLS estimate, $\hat{\beta}_1$, equals the true value, β_1, plus the sample covariance of x_t and u_t divided by the sample variance of x_t.

Distribution of the OLS Estimator

There are three main routes to learn more about the distribution of $\hat{\beta}$: *(i)* set up a small "experiment" in the computer and simulate the distribution (Monte Carlo or bootstrap simulation); *(ii)* pretend that the regressors can be treated as fixed numbers and then assume something about the distribution of the residuals; or *(iii)* use the asymptotic (large sample) distribution as an approximation. The asymptotic distribution can often be derived, in contrast to the exact distribution in a sample of a given size. If the actual sample is large, then the asymptotic distribution may be a good approximation.

Fixed Regressors and Normal Errors

The assumption of fixed regressors makes a lot of sense in controlled experiments, where we actually can generate different samples with the same values of the regressors (the heat or whatever). It makes much less sense in econometrics. However, it is easy to derive results for this case—and those results happen to be very similar to what asymptotic theory gives.

Suppose $u_t \sim N\left(0, \sigma^2\right)$, then (2.9) shows that $\hat{\beta}_1$ is normally distributed. The reason is that $\hat{\beta}_1$ is just a constant (β_1) plus a linear combination of normally distributed residuals (with fixed regressors $x_t / \sum_{t=1}^{T} x_t x_t$ can be treated as constant). It is straightforward to see that the mean of this normal distribution is β_1 (the true value), since the rest is a linear combination of the residuals—and they all have a zero mean.

Variance of the Estimator

Finding the variance of $\hat{\beta}_1$ is just slightly more complicated. First, write (2.9) as

$$\hat{\beta}_1 = \beta_1 + \frac{1}{\sum_{t=1}^{T} x_t x_t} \left(x_1 u_1 + x_2 u_2 + \ldots x_T u_T\right). \tag{2.10}$$

Second, remember that we treat x_t as fixed numbers ("constants"). Third, assume that the residuals are iid: they are uncorrelated with each other (independently distributed) and have the same variances (identically distributed). The variance of (2.10) is then

$$\begin{aligned}
\operatorname{Var}(\hat{\beta}_1) &= \frac{1}{\sum_{t=1}^{T} x_t x_t} \operatorname{Var}\left(x_1 u_1 + x_2 u_2 + \ldots x_T u_t\right) \frac{1}{\sum_{t=1}^{T} x_t x_t} \\
&= \frac{1}{\sum_{t=1}^{T} x_t x_t} \left(x_1^2 \sigma_1^2 + x_2^2 \sigma_2^2 + \ldots x_T^2 \sigma_T^2\right) \frac{1}{\sum_{t=1}^{T} x_t x_t} \\
&= \frac{1}{\sum_{t=1}^{T} x_t x_t} \left(x_1^2 \sigma^2 + x_2^2 \sigma^2 + \ldots x_T^2 \sigma^2\right) \frac{1}{\sum_{t=1}^{T} x_t x_t} \\
&= \frac{1}{\sum_{t=1}^{T} x_t x_t} \left(\sum_{t=1}^{T} x_t x_t\right) \sigma^2 \frac{1}{\sum_{t=1}^{T} x_t x_t} \\
&= \frac{1}{\sum_{t=1}^{T} x_t x_t} \sigma^2.
\end{aligned} \tag{2.11}$$

Multiple Regression

Consider the linear model

$$\begin{aligned} y_t &= x_{1t}\beta_1 + x_{2t}\beta_2 + \cdots + x_{kt}\beta_k + u_t \\ &= x_t'\beta + u_t, \end{aligned} \tag{2.13}$$

where y_t and u_t are scalars, x_t a $k \times 1$ vector, and β is a $k \times 1$ vector of the true coefficients (see Appendix A for a summary of matrix algebra). Least squares minimizes the sum of the squared fitted residuals

$$\sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} (y_t - x_t'\hat{\beta})^2, \tag{2.14}$$

by choosing the vector β. The first order conditions are

$$\mathbf{0}_{kx1} = \sum_{t=1}^{T} x_t (y_t - x_t'\hat{\beta}) \text{ or } \sum_{t=1}^{T} x_t y_t = \sum_{t=1}^{T} x_t x_t'\hat{\beta}, \tag{2.15}$$

which can be solved as

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t. \tag{2.16}$$

FOCs in Matrix Notation

$$\begin{aligned} SSE &= (Y - X\beta)'(Y - X\beta) \\ SSE &= (Y' - \beta'X')(Y - X\beta) \\ SSE &= Y'Y - \underbrace{\beta'X'Y - Y'X\beta}_{\text{middle terms are both scalars}} + \beta'X'X\beta \\ SSE &= Y'Y - 2\beta'X'Y + \beta'X'X\beta \end{aligned}$$

- *The FOC*

$$\frac{\partial SSE}{\partial \beta} = -2X'Y + 2X'X\beta \doteq 0$$

yields the normal equations

$$\begin{aligned} X'X\hat{\beta} &= X'Y \\ \Rightarrow \quad \hat{\beta} &= (X'X)^{-1}X'Y \end{aligned}$$

Multiple Regression in MATLAB

```
% Performs an OLS regression and returns the coefficient estimates, standard errors
% t-values, fitted values, residuals, estimated innovation variance, and coefficient
% variance-covariance matrix.

function [bHat,bHatStd,tbHat,yHat,eHat,sigma2,bHatCov]=ols(y,X)

[T,K]=size(X);                      % Let T=sample size and K=number of coefficients
bHat=(X'*X)\(X'*y);                 % Regression coefficients and residuals
yHat=X*bHat;
uHat=y-yHat;
sigma2=uHat'*uHat/(T-K);            % Estimated variance of the disturbance term
bHatCov=sigma2*inv(X'*X);           % Variance-covariance matrix of bHat
bHatStd=sqrt(diag(bHatCov));        % Read off the standard errors of bHat
tbHat=bHat./bHatStd;                % Calculate the t-statistics of bHat for H(0): b=0
end
```

2.2 Hypothesis Testing

2.2.1 Testing a Single Coefficient

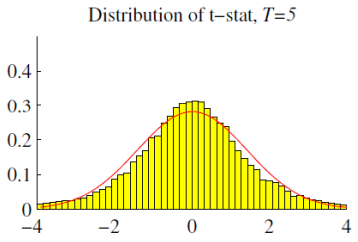
We assume that the estimates are normally distributed. To be able to easily compare with printed tables of probabilities, we transform to a $N(0, 1)$ variable as

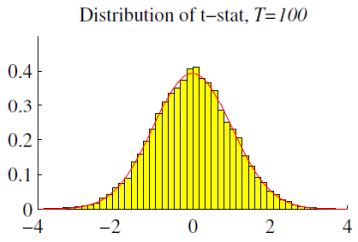
$$t = \frac{\hat{\beta} - \beta}{\text{Std}(\hat{\beta})} \sim N(0, 1), \tag{2.17}$$

where $\text{Std}(\hat{\beta})$ is the standard error (deviation) of $\hat{\beta}$.

Clearly, a 5% significance level gives the critical values -1.96 and 1.96,which would be *really* unusual under H_0. We sometimes compare with a t-distribution instead of a $N(0, 1)$, especially when the sample is short. For samples of more than 30–40 data points, the difference is trivial—see Table A.1. The *p-value* is a related concept. It is the lowest significance level at which we can reject the null hypothesis.

Distribution of t-Statistic





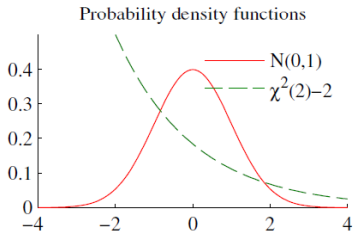
Model: $R_t = 0.9 f_t + \varepsilon_t$, $\varepsilon_t = v_t - 2$ where v_t has a $\chi^2(2)$ distribution

Results for $T=5$ and $T=100$:

Kurtosis of t–stat: 71.9 3.1

Frequency of abs(t–stat)>1.645: 0.25 0.10

Frequency of abs(t–stat)>1.96: 0.19 0.06



Can Also Construct Joint Tests

For instance, suppose we have estimated a model with three coefficients and the null hypothesis is

$$H_0 : \beta_1 = 1 \text{ and } \beta_3 = 0. \tag{2.18}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ or more generally} \tag{2.19}$$

$$R\beta = q, \tag{2.20}$$

where q has J (here 2) rows. Notice that the covariance matrix of these linear combinations is then

$$\text{Var}(R\hat{\beta}) = RV(\hat{\beta})R', \tag{2.21}$$

where $V(\hat{\beta})$ denotes the covariance matrix of the coefficients. Putting together these results we have the test static (a scalar)

$$(R\hat{\beta} - q)'[RV(\hat{\beta})R']^{-1}(R\hat{\beta} - q) \sim \chi^2_J. \tag{2.22}$$

This test statistic is compared to the critical values of a χ^2_J distribution—see Table A.2.

A Frequently Used Joint Test

A particularly important case is the test of the joint hypothesis that all slope coefficients in the regression (that is, excluding the intercept) are zero. It can be shown that the test statistics for this hypothesis is

$$TR^2 \sim \chi^2_{\#\text{slopes}}. \tag{2.23}$$

See Tables 2.1 and 2.2 for examples of this test.

2.3 Heteroskedasticity

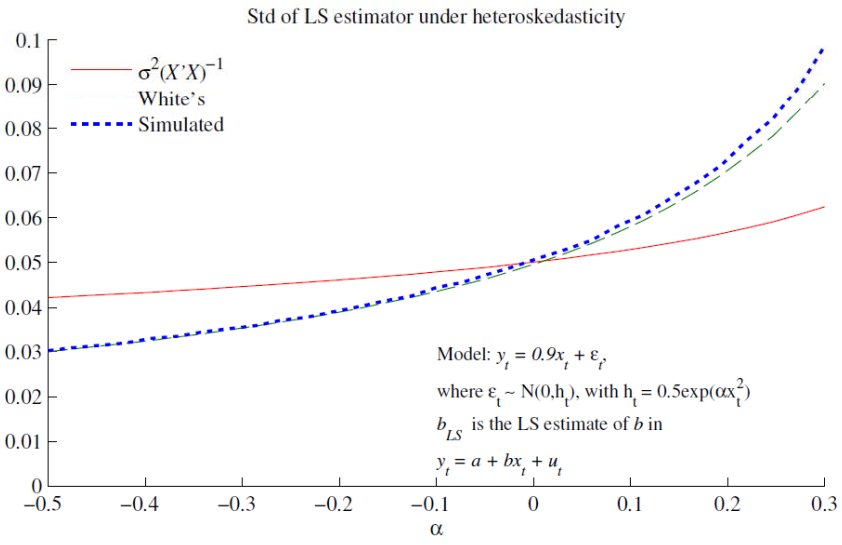
Suppose we have a regression model

$$y_t = x_t'b + u_t, \text{ where } \mathrm{E}\, u_t = 0 \text{ and } \mathrm{Cov}(x_{it}, u_t) = 0. \tag{2.24}$$

In the standard case we assume that u_t is iid (independently and identically distributed), which rules out heteroskedasticity.

In case the residuals actually are heteroskedastic, least squares (LS) is nevertheless a useful estimator: it is still consistent (we get the correct values as the sample becomes really large)—and it is reasonably efficient (in terms of the variance of the estimates). However, the standard expression for the standard errors (of the coefficients) is (except in a special case, see below) not correct. This is illustrated in Figure 2.13.

Impact of Heteroskedasticity



Testing for Heteroskedasticity

To test for heteroskedasticity, we can use *White's test of heteroskedasticity*. The null hypothesis is homoskedasticity, and the alternative hypothesis is the kind of heteroskedasticity which can be explained by the levels, squares, and cross products of the regressors—clearly a special form of heteroskedasticity. The reason for this specification is that if the squared residual is uncorrelated with these squared regressors, then the usual LS covariance matrix applies—even if the residuals have some other sort of heteroskedasticity (this is the special case mentioned before).

To implement White's test, let w_i be the squares and cross products of the regressors. The test is then to run a regression of squared fitted residuals on w_t

$$\hat{u}_t^2 = w_t'\gamma + v_t, \tag{2.25}$$

and to test if all the slope coefficients (not the intercept) in γ are zero. (This can be done be using the fact that $TR^2 \sim \chi_p^2$, $p = \dim(w_i) - 1$.)

Taking Heteroskedasticity into Account

There are two ways to handle heteroskedasticity in the residuals. First, we could use some other estimation method than LS that incorporates the structure of the heteroskedasticity. For instance, combining the regression model (2.24) with an ARCH structure of the residuals—and estimate the whole thing with maximum likelihood (MLE) is one way. As a by-product we get the correct standard errors—provided the assumed distribution (in the likelihood function) is correct. Second, we could stick to OLS, but use another expression for the variance of the coefficients: a heteroskedasticity consistent covariance matrix, among which "*White's covariance matrix*" is the most common.

White's Robust Standard Errors

To understand the construction of White's covariance matrix, recall that the variance of $\hat{\beta}_1$ is found from

$$\hat{\beta}_1 = \beta_1 + \frac{1}{\sum_{t=1}^{T} x_t x_t} (x_1 u_1 + x_2 u_2 + \ldots x_T u_T) . \tag{2.26}$$

Assuming that the residuals are uncorrelated gives

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \frac{1}{\sum_{t=1}^{T} x_t x_t} \text{Var} (x_1 u_1 + x_2 u_2 + \ldots x_T u_t) \frac{1}{\sum_{t=1}^{T} x_t x_t} \\ &= \frac{1}{\sum_{t=1}^{T} x_t x_t} \left(x_1^2 \text{Var} (u_1) + x_2^2 \text{Var} (u_2) + \ldots x_T^2 \text{Var} (u_T) \right) \frac{1}{\sum_{t=1}^{T} x_t x_t} \\ &= \frac{1}{\sum_{t=1}^{T} x_t x_t} \sum_{t=1}^{T} x_t^2 \sigma_t^2 \frac{1}{\sum_{t=1}^{T} x_t x_t} . \end{aligned} \tag{2.27}$$

This expression cannot be simplified further since σ_t is not constant—and also related to x_t^2. The idea of White's estimator is to estimate $\sum_{t=1}^{T} x_t^2 \sigma_t^2$ by $\sum_{t=1}^{T} x_t x_t' u_t^2$ (which also allows for the case with several elements in x_t, that is, several regressors).

2.4 Autocorrelation

Autocorrelation of the residuals ($\text{Cov}(u_t u_{t-s}) \neq 0$) is also a violation of the iid assumptions underlying the standard expressions for the variance of $\hat{\beta}_1$. In this case, LS is (typically) still consistent (exceptions: when the lagged dependent variable is a regressor), but the variances are (again) wrong. In particular, not even the the first line of (2.27) is true, since the variance of the sum in (2.26) depends also on the covariance terms.

There are several straightforward tests of autocorrelation—all based on using the fitted residuals. The null hypothesis is no autocorrelation. First, estimate the autocorrelations of the fitted residuals as

$$\rho_s = \text{Corr}(\hat{u}_t, \hat{u}_{t-s}), s = 1, ..., L. \quad (2.28)$$

Second, test the autocorrelation s by using the fact that $\sqrt{T}\hat{\rho}_s$ has a standard normal distribution (in large samples)

$$\sqrt{T}\hat{\rho}_s \sim N(0, 1). \quad (2.29)$$

Durbin-Watson Statistic

An alternative for testing the first autocorrelation coefficient is the Durbin-Watson. The test statistic is (approximately)

$$DW \approx 2 - 2\hat{\rho}_1, \quad (2.30)$$

and the null hypothesis is rejected in favour of positive autocorrelation if DW<1.5 or so (depending on sample size and the number of regressors). To extend (2.29) to higher-order autocorrelation, use the Box-Pierce test

$$Q_L = T\sum_{s=1}^{L}\hat{\rho}_s^2 \rightarrow^d \chi_L^2. \quad (2.31)$$

If there is autocorrelation, then we can choose to estimate a fully specified model (including how the autocorrelation is generated) by MLE or we can stick to OLS but apply an autocorrelation consistent covariance matrix—for instance, the "*Newey-West covariance matrix*."

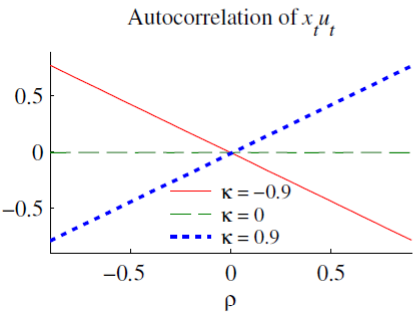
Newey-West Robust Standard Errors

To understand the Newey-West covariance matrix, consider the first line of (2.27). The middle term is a variance of a sum—which should (in principle) involve all covariance terms. For instance, with $T = 3$, we have

$$\text{Var}(x_1u_1 + x_2u_2 + x_3u_3) = x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + x_3^2\sigma_3^2 + 2x_1x_2\,\text{Cov}(u_1, u_2) + 2x_1x_3\,\text{Cov}(u_1, u_3) + 2x_2x_3\,\text{Cov}(u_2, u_3) \quad (2.32)$$

The idea of the Newey-West estimator is to first estimate all variances and a user-determined number of lags (covariances) and then sum them up as in (2.32). With only one lag the calculation is (with several regressors) $\sum_{t=1}^{T} x_t x_t' \hat{u}_t^2 + \sum_{t=2}^{T} \left(x_t x_{t-1}' + x_{t-1} x_t'\right) \hat{u}_t \hat{u}_{t-1}$. It is clear from this expression that what really counts is not so much the autocorrelation in u_t per se, but the autocorrelation of $x_t u_t$. If this is positive, then the standard expression underestimates the true variance of the estimated coefficients (and vice versa).

Autocorrelation of the Composite Term



Model: $y_t = 0.9x_t + \varepsilon_t$,

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

where u_t is iid N(0,h) such that $Std(\varepsilon_t) = 1$, and

$x_t = \kappa x_{t-1} + \eta_t$

b_{LS} is the LS estimate of b in

$y_t = a + bx_t + u_t$

Figure 2.15: Autocorrelation of $x_t u_t$ when u_t has autocorrelation ρ

Impact of Autocorrelated Composite Term

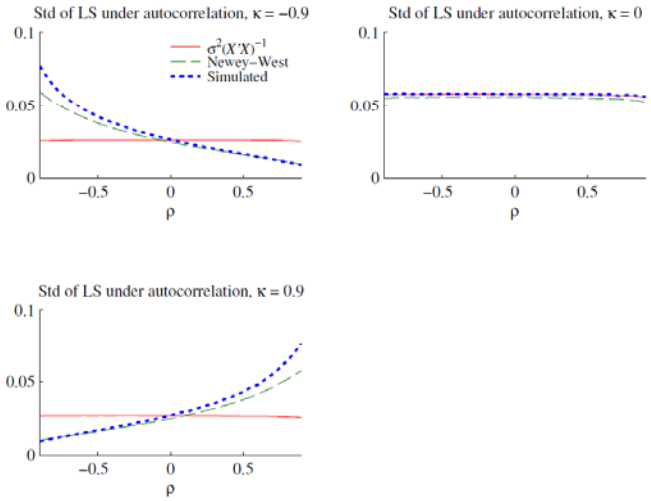


Figure 2.16: Variance of OLS estimator, autocorrelated errors

Example Using Overlapping Stock Returns

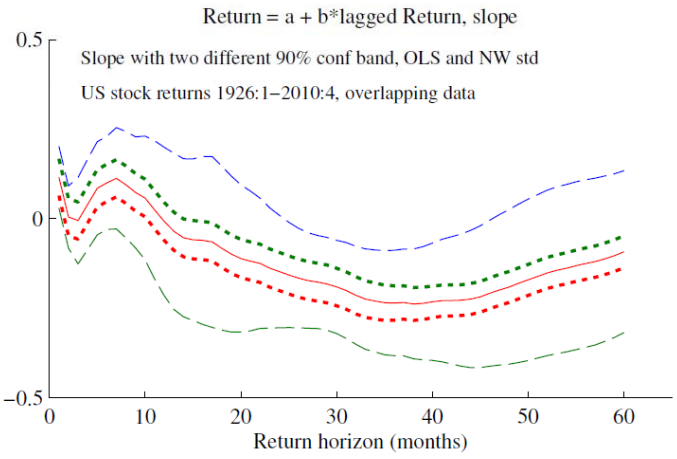


Figure 2.17: Slope coefficient, LS vs Newey-West standard errors