Empirical Methods in Finance
A note on asymptotic standard errors

Lecturer: Lars A. Lochstoer*
UCLA Anderson School of Management

January 6, 2020

*Let me know of any errors, please. Contact email: lars.lochstoer@anderson.ucla.edu

1 Departing from the OLS assumptions: Asymptotic Theory

Standard OLS makes strong assumptions. In particular, we assume the residuals are normally distributed and i.i.d. This is not a reasonable assumption for financial data. Can we make less assumptions and still maintain the conclusions and intuition from the strict OLS tests? The answer is yes, if we are willing to accept statistics based on asymptotic theory as a good approximation to finite sample statistics.

The model:

$$y_t = x_t' \beta + \varepsilon_t \tag{1}$$

where we still need the identifying OLS assumption: $E\left[x_t' \varepsilon_t\right] = 0$. Let x_t and β be $k \times 1$. We will soon make some assumptions regarding the second moments, but we do not make any distributional assumptions. We want to find the asymptotic distribution of the OLS estimate of β, $\hat{b}$. Define the $k \times 1$ OLS moment conditions for an arbitrary choice of b

$$f_t(b) = x_t\left(y_t - x_t' b\right) \tag{2}$$

Define the sample mean of these moment conditions as

$$g_T(b) = \frac{1}{T} \sum_{t=1}^{T} f_t(b) \tag{3}$$

The least squares minimization finds $\hat{b}$ such that $g_T\left(\hat{b}\right) = 0$.

Sample means are key in asymptotic analysis because we can take advantage of two huge results: The Law of Large Numbers and the Central Limit Theorem. Therefore we need to look more closely at the properties of the sample mean.

1.1 The Sample Mean

Consider a random $n \times 1$ vector y. The sample mean is

$$\overline{y}_T = \frac{1}{T} \sum_{t=1}^{T} y_t \tag{4}$$

Assume that the variable y in fact has a constant unconditional mean, i.e., $E\left[y_t\right] = \mu$ (which is an $n \times 1$ vector). In this case, the sample mean is an unbiased estimate of the true mean: $E\left[\overline{y}_T\right] = E\left[\frac{1}{T} \sum_{t=1}^{T} y_t\right] = \frac{1}{T} \sum_{t=1}^{T} E\left[y_t\right] = \mu$.

In our tests we use sample means and we need their test statistics in order to perform hypothesis testing and in order to say anything about the efficiency of an estimate and the power of the tests.

Therefore, we need the variance of the sample mean estimate. In order for the coming statistics to be nicely behaved, we assume that y follows a *stationary process*. In fact, we'll go one step further and assume that y_t is *covariance-stationary*:

$$E\left[\left(y_t-\mu\right)\left(y_{t-j}-\mu\right)'\right]=\underset{n\times x}{\gamma_j}\quad\forall\ j \tag{5}$$

In words, the unconditional covariance matrix of observations j periods apart is only a function of the distance j and not of time t. In this case the variance of the sample mean is given by

$$\begin{aligned} Var\left(\overline{y}_T\right) &= E\left[\left(\frac{1}{T}\sum_{t=1}^{T}y_t-\mu\right)\left(\frac{1}{T}\sum_{t=1}^{T}y_t-\mu\right)'\right] \\ &= \frac{1}{T}\sum_{j=-T}^{T}\frac{T-|j|}{T}\gamma_j \end{aligned} \tag{6}$$

(Work out the last equality yourself!).

1.2 Some definitions

Asymptotic analysis is about what happens when the number of observations goes to infinity: $T\to\infty$. In particular, what will happen to the mean and variance of the sample mean estimate as $T\to\infty$? We would like the estimate of the sample mean to be consistent, which means that if we had infinite amounts of data the estimated sample mean would equal the true sample mean with probability one.

Convergence in Probability: Consider a sequence of random variables x_T and an arbitrarily small, positive ε. Then if

$$\lim_{T\to\infty}\Pr\left(|x_T-c|>\varepsilon\right)=0 \tag{7}$$

we say that x_T converges in probability to c: plim $x_T=c$, or $x_T\xrightarrow{p}c$. We say that $\hat{b}$ is a *consistent estimator* of β if $\hat{b}\xrightarrow{p}\beta$.

Law of Large Numbers There are several of these, but here's one that we will use: If x_T has mean μ_T and variance σ_T^2 and the ordinary limits of μ_T and σ_T^2 are c and 0, respectively, x_T is a consistent estimator of c.

So, is the sample mean a consistent estimator of the true mean? We need to check the limits of both its expected value and variance. We already found that the expected value was μ for any

T. The limit of the variance of the sample mean is

$$\begin{aligned}
\lim_{T\to\infty} E\left[\left(y_T-\mu\right)\left(y_T-\mu\right)'\right] &= \frac{1}{T}\left(\gamma_0+2\gamma_1+2\gamma_2+2\gamma_3+...\right) \\
&= \frac{1}{T}\sum_{j=-\infty}^{\infty}\gamma_j \\
&\equiv \frac{1}{T}S
\end{aligned} \tag{8}$$

The last equality defines the matrix S as the infinite sum of autocovariance matrices. This is an important concept. It is sometimes referred to as the *spectral density matrix.* For the variance to go to zero as required, S must be finite. *Ergodicity* ensures this by requiring that as the distance between two observations gets very large, the covariance between the two goes to zero. In the scalar case, it is sufficient that

$$\sum_{j=0}^{\infty}\left|\gamma_j\right|<\infty \tag{9}$$

Another important implication of the Law of Large Numbers is that the sample estimate of the covariance matrix is a consistent estimate of the true covariance matrix:

$$\hat{\gamma}_j=\frac{1}{T}\sum_{t=j+1}^{\infty}\left(y_{t-j}-\overline{y}_T\right)\left(y_t-\overline{y}_T\right)' \xrightarrow{p} \gamma_j \tag{10}$$

We only need fairly weak restrictions on the process of y_t for this to hold (e.g., finite fourth moments). Essentially, this estimator is a sample mean of a function $f_T\left(y_t\right)$. In fact, *any* function which satisfies the law of large numbers has the property that the sample mean is a consistent estimator of the true mean. This is a key property that underlies the Generalized Method of Moments.

Convergence in Distribution Let $F_{x_T}\left(x\right)$ be the cumulative density function for the random vector x_T. Then if

$$\lim_{T\to\infty} F_{x_T}\left(x\right)=F_x\left(x\right) \tag{11}$$

x_T converges in distribution to x, $x_T \xrightarrow{d} x$. If the distribution is well known, like the normal distribution, we usually write $x_T \xrightarrow{d} N\left(\mu,\sigma^2\right)$. Now, we're ready for the punchline!

Central Limit Theorem

$$\sqrt{T}\left(\overline{y}_T-\mu\right) \xrightarrow{d} N\left(0,S\right) \tag{12}$$

The (scaled) sample mean estimate converges in distribution to a normal variable with mean zero and variance equal to the infinite sum of autocovariances. This is what makes asymptotic analysis beautiful!

We can extend this result to any function of $\overline{y}_T$ using the Delta Method: Let $c(\overline{y}_T)$ be a $j \times 1$ continuous and differentiable function of the $k \times 1$ vector $\overline{y}_T$. Then

$$\sqrt{T}\left(c\left(\overline{y}_T\right) - c\left(\mu\right)\right) \xrightarrow{d} N\left[0, C\left(\mu\right) SC\left(\mu\right)'\right] \tag{13}$$

where $C(\mu) \equiv \text{plim} \frac{\partial c}{\partial \overline{y}_T}|_{\overline{y}_T=\mu}$. For brief intuition on this result, consider a Taylor-expansion of $c(\cdot)$. As T gets large, the first order term dominates. An affine function of a normally distributed variable ($\overline{y}_T$, in this case) is also normal. The Delta method is very useful. If we can estimate parameters based on the sample mean of moments that perhaps are nonlinear in the parameters, the Delta method gives us the limiting distribution of the parameters. We use the distribution of the estimated parameters for hypothesis testing.

Asymptotic Distribution We use asymptotic theory by assuming that the asymptotic distribution is a good proxy for the true finite sample distribution. If

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, V\right) \tag{14}$$

Then the asymptotic distribution of $\hat{\theta}$ is

$$\hat{\theta} \xrightarrow{a} N\left(\theta, \frac{1}{T}V\right) \tag{15}$$

The asymptotic covariance matrix is then $\frac{1}{T}V$, which must be estimated for instance by the sample covariance matrix.

1.3 Asymptotic OLS continued

Now, let's apply this to the OLS example. Remember, the model is

$$y_t = x_t'\beta + \varepsilon_t \tag{16}$$

where $E\left[x_t\varepsilon_t\right] = 0$, and x_t is $k \times 1$. We can rewrite this model as

$$E\left[f_t\left(\beta\right)\right] = 0, \qquad f_t\left(b\right) \equiv x_t\left(y_t - x_t'b\right) \tag{17}$$

Define sample mean of f_t evaluated at an arbitrary b as

$$g_T\left(b\right) \equiv \frac{1}{T}\sum_{t=1}^{T} f_t\left(b\right) \tag{18}$$

Minimizing the sum of squared errors is equivalent to finding the $\hat{b}$ that sets the sample mean to zero:

$$g_T\left(\hat{b}\right) = 0 \tag{19}$$

We know that the true, unobserved sample mean converges to zero:

$$\sqrt{T}\left(g_T\left(\beta\right) - 0\right) \xrightarrow{d} N\left(0, S\right), \text{ where } S = \sum_{j=-\infty}^{\infty} E\left[f_t\left(\beta\right) f_{t-j}\left(\beta\right)'\right] \tag{20}$$

We want the distributional properties of the estimator so we can perform hypothesis tests. Apply the Delta method by applying a Taylor expansion to g :

$$g_T\left(\hat{b}_T\right) = g_T\left(\beta\right) + d_T\left(\beta\right)\left(\hat{b}_T - \beta\right) + \textit{higher order terms} \tag{21}$$

where $d_T\left(\beta\right) \equiv \frac{\partial g_T(b)}{\partial b}|_{b=\beta}$. Drop the higher order terms and solve for $\hat{b} - \beta$:

$$\hat{b}_T - \beta = -\left[d_T\left(\beta\right)\right]^{-1} g_T\left(\beta\right) \tag{22}$$

Now, we're ready for the Delta method, which says that

$$\sqrt{T}\left(\hat{b}_T - \beta\right) \xrightarrow{d} N\left(0, d\left(\beta\right)^{-1} S d\left(\beta\right)^{-1}\right) \tag{23}$$

where $d\left(\beta\right) = \text{plim } \frac{\partial g(b)}{\partial b}|_{b=\beta}$ and $S = \sum_{j=-\infty}^{\infty} E\left[f_t\left(\beta\right) f_{t-j}\left(\beta\right)'\right]$. From the definition of $f_t\left(b\right)$ we have that

$$d_T\left(b\right) = -\sum_{t=1}^{T} x_t x_t', \text{ so plim } d_T\left(b\right) = -E\left[x_t x_t'\right] \tag{24}$$

which gives

$$\sqrt{T}\left(\hat{b}_T - \beta\right) \xrightarrow{d} N\left(0, E\left[x_t x_t'\right]^{-1} S E\left[x_t x_t'\right]^{-1}\right) \tag{25}$$

so

$$\hat{b}_T - \beta \xrightarrow{a} N\left(0, \frac{1}{T} E\left[x_t x_t'\right]^{-1} S E\left[x_t x_t'\right]^{-1}\right) \tag{26}$$

1.4 Estimating the Variance-covariance matrix

What is S in the OLS case?

$$\begin{aligned} S &= \sum_{j=-\infty}^{\infty} E\left[f_t(\beta) f_{t-j}(\beta)'\right] \\ &= \sum_{j=-\infty}^{\infty} E\left[x_t x_{t-j}' \varepsilon_t \varepsilon_{t-j}\right] \end{aligned} \tag{27}$$

To put this into use, we need an estimate of the asymptotic covariance matrix. Here are three commonly used:

I.i.d. residuals with constant variance σ_ε^2 This strong assumption allows us to drop all terms where $j \neq 0$.

$$S_{iid} = E\left[x_t x_t'\right] \sigma^2 \tag{28}$$

and

$$Asy.Var\left(\hat{b}\right) = \frac{1}{T} E\left[x_t x_t'\right]^{-1} \sigma^2 \tag{29}$$

A consistent estimator of this matrix is

$$Est.Asy.Var\left(\hat{b}\right) = \frac{1}{T}\left(X_T' X_T\right)^{-1} \hat{\varepsilon}_T' \hat{\varepsilon}_T \tag{30}$$

Independent residuals, but squared value correlated with x's (White (1980)) This case allow for time-varying volatility (heteroskedasticity) of the residuals

$$S = E\left[x_t x_t' \varepsilon_t^2\right] \tag{31}$$

and

$$Asy.Var\left(\hat{b}\right) = \frac{1}{T} E\left[x_t x_t'\right]^{-1} E\left[x_t x_t' \varepsilon_t^2\right] E\left[x_t x_t'\right]^{-1} \tag{32}$$

A consistent estimator of this matrix is

$$Est.Asy.Var\left(\hat{b}\right) = \left(X_T' X_T\right)^{-1} \left(\sum_{t=1}^{T} x_t x_t' \hat{\varepsilon}_t^2\right) \left(X_T' X_T\right)^{-1} \tag{33}$$

Non-zero correlations between $f_t(b)$ and $f_{t-j}(b)$ if $|j| \le q$.

$$Asy.Var\left(\hat{b}\right) = \frac{1}{T} E\left[x_t x_t'\right]^{-1} \left(\sum_{q=-j}^{j} E\left[x_t x_{t-j}' \varepsilon_t \varepsilon_{t-j}\right] \right) E\left[x_t x_t'\right]^{-1} \tag{34}$$

where the estimate of this is the sample analogue. This estimate is due to Hansen and Hodrick (1983). However, in finite samples the estimate of the variance-covariance matrix is not always positive-definite (i.e., we sometimes get negative variance). Newey-West (1987) has proposed a much used correction that alleviates this problem:

$$R_T(v; b) = \frac{1}{T} \sum_{t=1+v}^{T} f_t(b) f_{t-v}(b)' \tag{35}$$

where the estimate of the spectral density matrix is

$$\hat{S}_T = R_T\left(0; \hat{b}_T\right) + \sum_{v=1}^{q} \frac{q+1-v}{q+1} \left(R_T\left(v; \hat{b}_T\right) + R_T\left(v; \hat{b}_T\right)' \right) \tag{36}$$

The estimate covariance matrix is then

$$Est.Asy.Var\left(\hat{b}\right) = T\left(X_T' X_T\right)^{-1} \hat{S}_T \left(X_T' X_T\right)^{-1} \tag{37}$$

Correcting for autocorrelation is especially important when using overlapping observations (e.g., annual overlapping at the monthly frequency).

1.5 Standard OLS vs. robust OLS

Note that in this case, the empirical implementation of the asymptotic (robust) OLS is the same as the finite sample OLS, but without having to make the normality assumption! Since the robust OLS estimates all are normal, the joint test statistics are distributed χ^2: This is the same as the finite sample distribution with *known* covariance matrix. Of course, we never have infinite amounts of data as we implicitly assume when we use asymptotic theory in practice. Since we do not, we cannot know the covariance matrix. We know, however, that we need quite a few datapoints to be able to estimate a covariance matrix with reasonable accuracy.