USC, Fall 2016, Economics 513

Lecture 5: Inference in the CLR model

The CLR model as a random experiment

The CLR assumptions specify a random experiment. In this experiment

1. The $n \times K$ matrix X is drawn from some distribution such that $\text{rank}(X) = K$.
2. The $n \times 1$ vector ε is drawn from a distribution that satisfies $\text{E}(\varepsilon|X) = 0$ and $\text{E}(\varepsilon\varepsilon'|X) = \sigma^2 I$.
3. $y = X\beta + \varepsilon$.

In econometrics we have two main types of data: time-series and cross-section data. Panel data, i.e. a cross-section of time-series, have some features of both types.

In cross-section data we have a random sample $y_i, x_i, i = 1, \ldots, n$ from a population joint distribution. Because the ε_i, x_i are independent for $i = 1, \ldots, n$ due to random sampling, we observed that if $\text{E}(\varepsilon_i|x_i) = 0$ and $\text{E}(\varepsilon_i^2|x_i) = \sigma^2$ then Assumptions 1 and 2 of the CLR model hold, so that we can think of the data being generated by the above random experiment.

For a cluster sample, e.g. a sample of students for a number of schools, independence holds between but not within schools, so that Assumption 2 may not hold and if it does not we cannot apply the CLR model.

In time-series data the observations $y_i, x_i, i = 1, \ldots, n$ are not a random sample from a population. The CLR model applies if Assumptions 1-3 are satisfied. Although $\text{E}(\varepsilon_i|x_i)$ often holds, the assumption $\text{E}(\varepsilon_i|x_1, \ldots, x_n)$ is problematic.

Example: AR(1) model

$$y_i = \beta_1 + \beta_2 y_{i-1} + \varepsilon_i$$

Here i indicates a time period. Here

$$X = \begin{pmatrix} 1 & y_0 \\ \vdots & \vdots \\ 1 & y_{n-1} \end{pmatrix}$$

Because $\varepsilon_i = y_i - \beta_1 - \beta_2 y_{i-1}$ we have

$$\text{E}(\varepsilon_i|X) = \text{E}(\varepsilon_i|y_0, \ldots, y_{n-1}) = \text{E}(\varepsilon_i|\varepsilon_1, \ldots, \varepsilon_{n-1}) = \varepsilon_i \neq 0$$

for $i = 1, \ldots, n-1$ and 0 if $i = n$ (and the random errors are independent).

The CLR assumptions can hold for time-series data. Example: linear time trend

$$y_i = \beta_1 + \beta_2 i + \varepsilon_i$$

If e.g. the random errors are uncorrelated and have constant variance over time, then the CLR assumptions hold.

In sequel we assume first that the data are generated by the CLR model. The next step will be to relax the assumptions.

The sampling distribution of $\hat{\beta}$ and $\hat{\sigma}^2$

In the CLR model the OLS solution $\hat{\beta}$ is an estimator for the parameter vector β. As usual we evaluate the quality of this estimator by studying its *sampling distribution*. The sampling distribution is the distribution of the OLS estimator in repeated samples $y_s, X_s, s = 1, \ldots S$ from the CLR model.

If β were known and also the distributions of ε and X we can simulate from the CLR model to obtain the sampling distribution of $\hat{\beta}$, i.e. we would have

$$y_s = X_s \beta + \varepsilon_s \qquad s = 1, \ldots S$$

and

$$\hat{\beta}_s = (X_s' X_s)^{-1} X_s' y_s \qquad s = 1, \ldots S$$

Even if we do not know β nor the distributions of ε and X we can use the rules of probability theory to derive (features) of the sampling distribution of $\hat{\beta}$. Substitute the CLR model in the expression for the OLS estimator

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

Because this is a linear expression in ε we can use Assumption 1 to find the conditional mean of $\hat{\beta}$:

$$\mathrm{E}(\hat{\beta}|X) = \beta + (X'X)^{-1} X' \mathrm{E}(\varepsilon|X) = \beta$$

By the Law of Iterated Expectations

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}[\mathrm{E}(\hat{\beta}|X)] = \beta$$

Therefore under the CLR assumptions the OLS estimator is unbiased for β.

For the (conditional) sampling variance

$$\mathrm{Var}(\hat{\beta}|X) = \mathrm{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X]$$

we have because

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

upon substitution and using Assumption 2

$$\mathrm{Var}(\hat{\beta}|X) = \mathrm{E}\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}|X\right] =$$

$$(X'X)^{-1}X'\mathrm{E}\left[\varepsilon\varepsilon'|X\right]X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

Because

$$\mathrm{Var}(\hat{\beta}) = \mathrm{E}[\mathrm{Var}(\hat{\beta}|X)] + \mathrm{Var}(\mathrm{E}(\hat{\beta}|X))$$

and the second term is 0, we have

$$\mathrm{Var}(\hat{\beta}) = \sigma^2\mathrm{E}\left[(X'X)^{-1}\right]$$

Because we do not want to make an assumption on the distribution of X we use the unbiased estimator of the variance $\sigma^2(X'X)^{-1}$.

In special case of a constant and one regressor we have the unbiased variance estimator

$$\mathrm{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}$$

Note that this decreases with σ^2 and with the variation in x.

Optimality of OLS estimator in CLR model

Consider class of estimators for β that are linear in y , i.e.

$$\hat{\beta}_L = Cy$$

with C a $K \times n$ matrix that depends on X. For the OLS estimator $C = (X'X)^{-1}X'$

Gauss-Markov Theorem: In CLR model the OLS estimator is the Best Linear Unbiased (BLU) of β, i.e. it has the smallest variance of linear unbiased estimators for β.

Estimation of σ^2

The other parameter in CLR model is the variance of the random error σ^2.

It seems obvious that we should use the sample variance of the OLS residuals e. Instead we use (in model with K regressors including constant)

$$\hat{\sigma}^2 = \frac{1}{n-K}\sum_{i=1}^{n}e_i^2 = \frac{1}{n-K}e'e$$

The sampling distribution of this estimator can be derived from

$$e = My = M(X\beta + \varepsilon) = M\varepsilon \qquad M = I - X(X'X)^{-1}X'$$

so that

$$\hat{\sigma}^2 = \frac{1}{n-K}\varepsilon' M \varepsilon$$

We show that this is an unbiased estimator of σ^2.

The trace of an $n \times n$ matrix A is

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$$

The trace has an important property: for matrices A and B

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

if both AB and BA are well-defined. Also for a scalar a we have $\mathrm{tr}(a) = a$ and for matrices A and B of the same order $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$

Using these properties

$$\mathrm{tr}(\varepsilon' M \varepsilon) = \mathrm{tr}(M \varepsilon' \varepsilon)$$

so that

$$\mathrm{E}(\hat{\sigma}^2|X) = \frac{1}{n-K}\mathrm{E}\left[\mathrm{tr}(\varepsilon' M \varepsilon)|X\right] = \frac{1}{n-K}\mathrm{E}\left[\mathrm{tr}(M\varepsilon\varepsilon')|X\right] =$$

$$\frac{1}{n-K}\mathrm{tr}(M\mathrm{E}[\varepsilon\varepsilon'|X]) = \frac{\sigma^2}{n-K}\mathrm{tr}(M)$$

Finally

$$\mathrm{tr}(M) = \mathrm{tr}(I) - \mathrm{tr}(X(X'X)^{-1}X') = n - \mathrm{tr}((X'X)^{-1}X'X) = n - K$$

We conclude

$$\mathrm{E}(\hat{\sigma}^2|X) = \sigma^2$$

and by the Law of Iterated Expectations

$$\mathrm{E}(\hat{\sigma}^2) = \sigma^2$$

so that $\hat{\sigma}^2$ is an unbiased estimator or σ^2 (if the CLR assumptions hold).

The estimated variance of the OLS estimator

Combining the results we find that the variance matrix of the OLS estimator $\hat{\beta}$ can be estimated as

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

This is an unbiased estimator of the sampling variance of the OLS estimator.

The estimator of the variance of $\hat{\beta}_k$ is

$$\widehat{\text{Var}}(\hat{\beta}_k) = \hat{\sigma}^2 (X'X)^{-1}_{kk}$$

where $(X'X)^{-1}_{kk}$ denotes element kk of the matrix $(X'X)^{-1}$.

The square root of this estimator is called the *standard error* of the OLS estimator of β_k. This standard error is reported together with the OLS estimates of the regression coefficients.

It is common practice to report standard errors of OLS estimates of the regression coefficients. For the lottery data

	I	II	III
CONST	12.38 (1.07)	2347.58 (899.15)	-0.757 (0.932)
PRIZE	-0.0262 (0.0129)	-0.0420 (0.0098)	-0.0503 (0.0113)
AGEWON		-0.184 (0.047)	
EDUC		-0.0275 (0.280)	
MALE		1.212 (1.342)	
TIXBOT		-0.00406 (.177)	
WORKTHEN		1.900 (1.598)	
XEARN6		-0.143 (0.134)	
XEARN5		-0.118 (0.193)	
XEARN4		0.0396 (0.158)	
XEARN3		0.390 (0.163)	
XEARN2		0.0772 (0.157)	
XEARN1		0.286 (0.101)	
YEARWON		-1.176 (0.453)	

The CLR model with normally distributed random errors

To use the standard errors we make the additional Assumption 4 in the CLR model, i.e.

$$\varepsilon | X \sim N(0, \sigma^2 I)$$

This means that the random errors ε_i and ε_j are stochastically independent if $i \neq j$ (remember that two normal random variables are independent if and only if they have covariance 0/are uncorrelated) and that ε_i has a normal distribution with mean 0 and variance σ^2.

Because

$$\hat{\beta} = \beta + (X'X)^{-1} X' \varepsilon$$

is linear in ε we use the fact that linear expressions in normal random variables have a normal distribution to conclude that

$$\hat{\beta} | X \sim N(\beta, \sigma^2 (X'X)^{-1})$$

With Assumption 4 we get the exact sampling distribution of $\hat{\beta}$.

Next we consider the sampling distribution of $\hat{\sigma}^2$. Here we use the result that if $z_1, \ldots, z_d$ are independent and all have a standard normal distribution then

$$\sum_{j=1}^{d} z_j^2 \sim \chi^2(d)$$

i.e. has a chi-squared distribution with d degrees of freedom. Now consider

$$\frac{(n-K)\hat{\sigma}^2}{\sigma^2} = \left(\frac{\varepsilon}{\sigma}\right)' M \left(\frac{\varepsilon}{\sigma}\right)$$

with ε/σ a vector of independent standard normal random variables.

This is a quadratic expression in standard normal random variables. Remember

$$M = I - X(X'X)^{-1}X'$$

is the difference of two projection matrices (and itself a projection matrix) of ranks n and K respectively (the rank of a projection matrix is equal to its trace). This implies that

$$\frac{(n-K)\hat{\sigma}^2}{\sigma^2}|X \sim \chi^2(n-K)$$

It can also be shown that given X, $\hat{\beta}$ and $\hat{\sigma}^2$ are stochastically independent.

t-ratio

Because $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$

$$\frac{\hat{\beta}_k - \beta_k}{\sigma\sqrt{(X'X)^{-1}_{kk}}}|X \sim N(0,1)$$

Take the ratio with the independent (given X)

$$(n-K)\frac{\hat{\sigma}^2}{\sigma^2}|X \sim \chi^2(n-K)$$

to obtain

$$\frac{\frac{\hat{\beta}_k - \beta_k}{\sigma\sqrt{(X'X)^{-1}_{kk}}}}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}} = \frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}\sqrt{(X'X)^{-1}_{kk}}} \sim t(n-K)$$

Note that

$$\frac{\hat{\beta}_k - \beta_k}{\sigma\sqrt{(X'X)^{-1}_{kk}}} \sim N(0,1)$$

The t distribution has mean 0, but a larger variance than the standard normal. If $n - K \to \infty$ the t distribution converges to the standard normal. For $n - K \geq 100$ the difference is negligible.

Note that the t distribution does not depend on X.
The result that

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}\sqrt{(X'X)^{-1}_{kk}}} \sim t(n-K)$$

can be used for

1. Confidence interval for β_k
2. Hypothesis test for β_k

Confidence interval for β_k

From the table of the t distribution we find t_α such that

$$\Pr\left(-t_\alpha \leq \frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}\sqrt{(X'X)^{-1}_{kk}}} \leq t_\alpha\right) = \alpha$$

Hence t_α is the $\left(1 - \frac{1-\alpha}{2}\right)$-th quantile of the $t(n-K)$ distribution, i.e.

$$\Pr(t(n-K) \leq t_\alpha) = 1 - \frac{1-\alpha}{2}$$

For $n = 40, K = 5, \alpha = .95$ we find $t_\alpha = 2.030$. Also for $n = 237$ the quantiles are those of the standard normal.

A 100α % confidence interval for β_k is

$$\left[\hat{\beta}_k - t_\alpha\hat{\sigma}\sqrt{(X'X)^{-1}_{kk}}, \hat{\beta}_k + t_\alpha\hat{\sigma}\sqrt{(X'X)^{-1}_{kk}}\right]$$

For lottery data a 95% confidence interval for the coefficient on PRIZE is (for specification III)

$$[-.0503 - 1.96 \times .01127, -.0503 + 1.96 \times .01127] = [-0.0724, -0.0282108]$$

Hypothesis tests on β_k

Statistical hypothesis

$$H_0 : \beta_k = \beta_{k0}$$
$$H_1 : \beta_k \neq \beta_{k0}$$

TABLE G.2 Percentiles of the Student's t Distribution. Table Entry Is x Such that $\text{Prob}[t_n \leq x] = P$

n	.750	.900	.950	.975	.990	.995
1	1.000	3.078	6.314	12.706	31.821	63.657
2	.816	1.886	2.920	4.303	6.965	9.925
3	.765	1.638	2.353	3.182	4.541	5.841
4	.741	1.533	2.132	2.776	3.747	4.604
5	.727	1.476	2.015	2.571	3.365	4.032
6	.718	1.440	1.943	2.447	3.143	3.707
7	.711	1.415	1.895	2.365	2.998	3.499
8	.706	1.397	1.860	2.306	2.896	3.355
9	.703	1.383	1.833	2.262	2.821	3.250
10	.700	1.372	1.812	2.228	2.764	3.169
11	.697	1.363	1.796	2.201	2.718	3.106
12	.695	1.356	1.782	2.179	2.681	3.055
13	.694	1.350	1.771	2.160	2.650	3.012
14	.692	1.345	1.761	2.145	2.624	2.977
15	.691	1.341	1.753	2.131	2.602	2.947
16	.690	1.337	1.746	2.120	2.583	2.921
17	.689	1.333	1.740	2.110	2.567	2.898
18	.688	1.330	1.734	2.101	2.552	2.878
19	.688	1.328	1.729	2.093	2.539	2.861
20	.687	1.325	1.725	2.086	2.528	2.845
21	.686	1.323	1.721	2.080	2.518	2.831
22	.686	1.321	1.717	2.074	2.508	2.819
23	.685	1.319	1.714	2.069	2.500	2.807
24	.685	1.318	1.711	2.064	2.492	2.797
25	.684	1.316	1.708	2.060	2.485	2.787
26	.684	1.315	1.706	2.056	2.479	2.779
27	.684	1.314	1.703	2.052	2.473	2.771
28	.683	1.313	1.701	2.048	2.467	2.763
29	.683	1.311	1.699	2.045	2.462	2.756
30	.683	1.310	1.697	2.042	2.457	2.750
35	.682	1.306	1.690	2.030	2.438	2.724
40	.681	1.303	1.684	2.021	2.423	2.704
45	.680	1.301	1.679	2.014	2.412	2.690
50	.679	1.299	1.676	2.009	2.403	2.678
60	.679	1.296	1.671	2.000	2.390	2.660
70	.678	1.294	1.667	1.994	2.381	2.648
80	.678	1.292	1.664	1.990	2.374	2.639
90	.677	1.291	1.662	1.987	2.368	2.632
100	.677	1.290	1.660	1.984	2.364	2.626
∞	.674	1.282	1.645	1.960	2.326	2.576

The t statistic that is used to test this hypothesis is

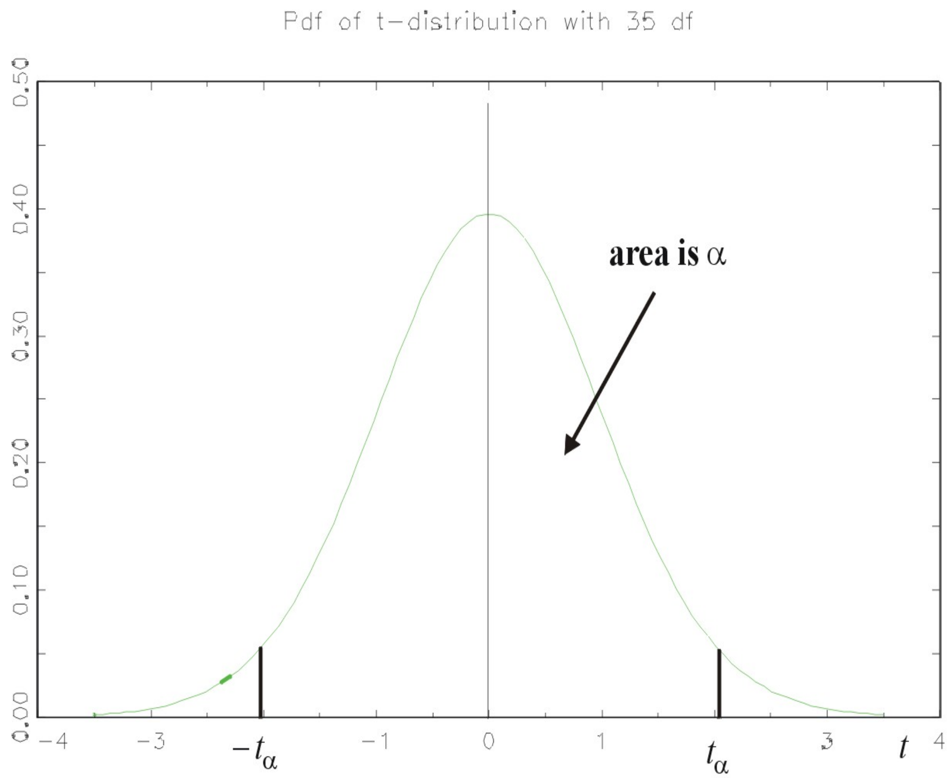
$$t = \frac{\hat{\beta}_k - \beta_{k0}}{s\sqrt{(X'X)^{-1}_{kk}}}$$

If H_0 is true, then $t \sim t(n-K)$, so that $\mathrm{E}(t) = 0$.

If H_0 is false, i.e. $\beta_k = \beta_{k1} \neq \beta_{k0}$, then

$$\begin{aligned} \mathrm{E}(t) > 0 \quad &\text{if} \quad \beta_k = \beta_{k1} > \beta_{k0} \\ \mathrm{E}(t) < 0 \quad &\text{if} \quad \beta_k = \beta_{k1} < \beta_{k0} \end{aligned}$$

See figure.



This suggests that we reject H_0/choose H_1 if $|t| > c$ (mean of $t(n-K)$ distribution is 0) and reject H_1/choose H_0 if $|t| \leq c$.

We can make the wrong choice: False rejection of H_0/type I error or false choice of H_0/type II error.

As usual we control the size of the type I error, i.e. we choose the cut-off value c such that

$$\Pr(|t| > c) = \alpha$$

with $t \sim t(n-K)$, so that c is $\left(1-\frac{\alpha}{2}\right)$-th quantile of the $t(n-K)$ distribution.

This test is called the t test of H_0. α is called the size or significance level of the test and is typically chosen as .01 (1 %), .05 (5%) or .1 (10%). Sometimes the p value is reported. If the test statistic has the value $\tilde{t}$, then

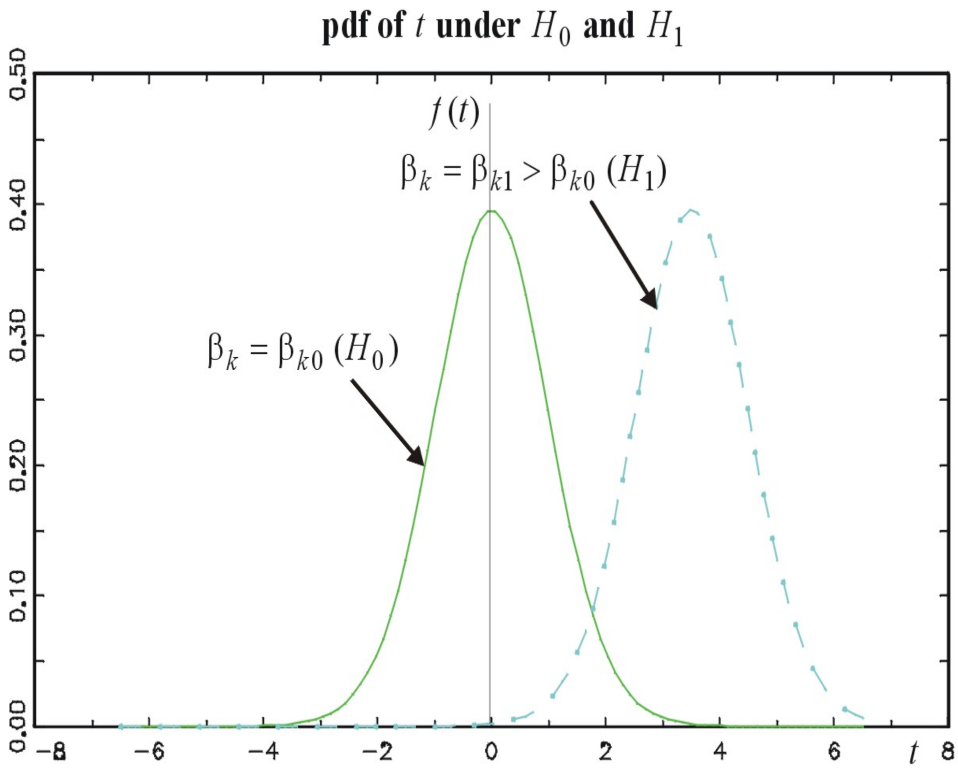
$$p = \Pr(|t| > |\tilde{t}|)$$

so that p can be compared to α. It is a a measure of the evidence for H_0.

For lottery data we test in specification III the hypothesis that the coefficient on PRIZE is 0.

$$t = \frac{-.0503}{.01127} = -4.463$$

The cutoff-value is 1.645 (10%), 1.96 (5%), 2.576 (1 %), so that the null is rejected at all levels of significance considered. Also $p = 0.0000125$.



If we want to test

$$H_0 : \beta_k < \beta_{k0}$$
$$H_1 : \beta_k \geq \beta_{k0}$$

then the probability of a type II error is smaller if we use a one-sided test, i.e. we reject H_0 if $t > c$ where

$$\sup_{\beta_k \leq \beta_{k0}} \Pr(t > c) = \alpha$$

so that c is the $(1-\alpha)$-th quantile of the $t(n-K)$ distribution. Note that the size of the test is the maximal rejection probability if H_0 is true. This probability is largest if $\beta_k = \beta_{k0}$. The p-value is

$$p = \Pr(t > \tilde{t})$$

If we test the hypothesis that the effect of PRIZE is negative against the alternative that it is positive we have the cut-off values 1.282 (10%), 1.645 (5%) and 2.326 (1%) so that we cannot reject the null. The p-value is .9999

If we interchange the null and the alternative we have cut-off values -1.282 (10%), -1.645 (5%) and -2.326 (1%) and we reject the hypothesis that the effect is positive at these levels. Now $p = 0.0000063$.

Tests on subvectors of β

Partition

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

The hypothesis is

$$\begin{aligned} H_0 &: \beta_2 = \beta_{20} \\ H_1 &: \beta_2 \neq \beta_{20} \end{aligned}$$

with β_2 a $K_2 \times 1$ vector. In most cases the intercept is in β_1.

We have

$$\hat{\beta}_2 | X \sim N(\beta_2, \sigma^2 (X'X)_2^{-1})$$

with $(X'X)_2^{-1}$ the submatrix of the final K_2 rows and columns of $(X'X)^{-1}$.

The test statistic is

$$F = \frac{(\hat{\beta}_2 - \beta_{20})'((X'X)_2^{-1})^{-1}(\hat{\beta}_2 - \beta_{20})}{K_2 \hat{\sigma}^2}$$

If H_0 is true then $F \sim F(K_2, n-K)$ and we reject the null if $F > c$, so that c is the $1-\alpha$-th quantile of this distribution.

The 95% quantiles of the F distribution are in the table. If $n - K \to \infty$, then the $F(K_2, n-K_2)$ converges to the $\chi^2(K_2)/K_2$ distribution, so that if $n - K_2$ is large we can use the test statistic

$$C = \frac{(\hat{\beta}_2 - \beta_{20})'((X'X)_2^{-1})^{-1}(\hat{\beta}_2 - \beta_{20})}{\hat{\sigma}^2}$$

and obtain the critical/cut-off value from the $\chi^2(K_2)$ distribution.

Application to lottery data: To check the random assignment/random response of PRIZE we regress PRIZE on the other independent variables. We test

TABLE G.4 95th Percentiles of the F Distribution. Table Entry is f such that Prob[$F_{n1,n2} \leq f$] = .95

n_1 = *Degrees of Freedom for the Numerator*

n_2	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97
∞	3.84	3.00	2.60	2.37	2.21	2 .10	2.01	1.94	1.88

n_2	10	12	15	20	30	40	50	60	∞
1	241.88	243.91	245.95	248.01	250.10	251.14	252.20	252.20	254.19
2	19.40	19.41	19.43	19.45	19.46	19.47	19.48	19.48	19.49
3	8.79	8.74	8.70	8.66	8.62	8.59	8.57	8.57	8.53
4	5.96	5.91	5.86	5.80	5.75	5.72	5.69	5.69	5.63
5	4.74	4.68	4.62	4.56	4.50	4.46	4.43	4.43	4.37
6	4.06	4.00	3.94	3.87	3.81	3.77	3.74	3.74	3.67
7	3.64	3.57	3.51	3.44	3.38	3.34	3.30	3.30	3.23
8	3.35	3.28	3.22	3.15	3.08	3.04	3.01	3.01	2.93
9	3.14	3.07	3.01	2.94	2.86	2.83	2.79	2.79	2.71
10	2.98	2.91	2.85	2.77	2.70	2.66	2.62	2.62	2.54
15	2.54	2.48	2.40	2.33	2.25	2.20	2.16	2.16	2.07
20	2.35	2.28	2.20	2.12	2.04	1.99	1.95	1.95	1.85
25	2.24	2.16	2.09	2.01	1.92	1.87	1.82	1.82	1.72
30	2.16	2.09	2.01	1.93	1.84	1.79	1.74	1.74	1.63
40	2.08	2.00	1.92	1.84	1.74	1.69	1.64	1.64	1.52
50	2.03	1.95	1.87	1.78	1.69	1.63	1.58	1.58	1.45
70	1.97	1.89	1.81	1.72	1.62	1.57	1.50	1.50	1.36
100	1.93	1.85	1.77	1.68	1.57	1.52	1.45	1.45	1.30
∞	1.83	1.75	1.67	1.57	1.46	1.39	1.34	1.31	1.30

whether the regression coefficients (except the intercept) are equal to 0. These are 12 coefficients. We find

$$F = 3.129$$

The critical value is 1.75 (5%) so that we reject the null. Also $p = 0.000392$.

For this type of H_0 we have

$$F = \frac{(n-K)R^2}{K_2(1-R^2)}$$

Functional form and partial effects

Multiple linear regression model

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i$$

In the derivations until know nothing prevents us from defining

$$\begin{aligned} x_{i2} &= x_i \\ \vdots &= \vdots \\ x_{iK} &= x_i^{K-1} \end{aligned}$$

The regression equation is linear in $\beta_1, \ldots, \beta_K$ but non-linear in x_i.

We can also transform the dependent and/or independent variables:

- Log-linear relation: Define $x_{i2} = \ln x_i$ and take $\ln y_i$ as the dependent variable. Note that
$$\beta_2 = \frac{\partial \ln y}{\partial \ln x}$$
i.e. β_2 is the elasticity of y with respect to x.

- Semi-log relation: Take $\ln y$ as the dependent variable.
$$\beta_2 = \frac{\partial \ln y}{\partial x} = \frac{\partial y / y}{\partial x}$$
i.e. β_2 is a semi-elasticity.

Application to lottery data

With PRIZE and PRIZE squared we find

$$\begin{array}{lcccccc} \widehat{\text{PEARN}}_i & = & 1.956 & - & 0.137 \cdot \text{PRIZE}_i & + & 0.000299 \cdot \text{PRIZE}_i^2 \\ & & (1.125) & & (0.0239) & & (0.000074) \end{array}$$

Note that the effect of PRIZE is non-linear. The effect is largest for small winners and decreases with the size of PRIZE. The partial derivative with respect to PRIZE is

$$\frac{\partial \text{PEARN}}{\partial \text{PRIZE}} = \beta_2 + 2\beta_2 \text{PRIZE}$$

so that we have to chose a point where to evaluate this derivative. We take PRIZE equal to 0 or to the median.

The latter partial effect is estimated by

$$\hat{\beta}_2 + 2\hat{\beta}_3 \text{med(PRIZE)}$$

This is a linear combination of the estimates of the regression coefficients. If c is a $K \times 1$ vector of constants then such a linear combination can be expressed as $c'\hat{\beta}$.

We derive the variance of the linear combination

$$\mathrm{Var}(c'\hat{\beta}) = \mathrm{E}[(c'\hat{\beta} - c'\beta)^2] = \mathrm{E}[(c'\hat{\beta} - c'\beta)(\hat{\beta}'c - \beta'c)] =$$

$$\mathrm{E}[c'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'c] = c'\mathrm{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']c = c'\mathrm{Var}(\hat{\beta})c$$

For the lottery data we find that the partial effect at 0 is -0.137 with standard error 0.0239 and that at the median is -0.118 with standard error 0.0199.

There is another way to find the effect at the median. Define the 1-1 transformation of the parameters

$$\begin{aligned} \gamma_1 &= \beta_1 \\ \gamma_2 &= \beta_2 + 2\mathrm{med}(\mathrm{PRIZE})\beta_3 \\ \gamma_3 &= \beta_3 \end{aligned}$$

with inverse transformation

$$\begin{aligned} \beta_1 &= \gamma_1 \\ \beta_2 &= \gamma_2 - 2\mathrm{med}(\mathrm{PRIZE})\gamma_3 \\ \beta_3 &= \gamma_3 \end{aligned}$$

Substitution gives

$$\mathrm{PEARN}_i = \gamma_1 + (\gamma_2 - 2\mathrm{med}(\mathrm{PRIZE})\gamma_3)\mathrm{PRIZE}_i + \gamma_3\mathrm{PRIZE}_i^2 + \varepsilon_i =$$

$$\gamma_1 + \gamma_2\mathrm{PRIZE}_i + (\mathrm{PRIZE}_i^2 - 2\mathrm{med}(\mathrm{PRIZE})\mathrm{PRIZE}_i)\gamma_3 + \varepsilon_i$$

The OLS estimator of the coefficient on PRIZE is -0.118 with standard error 0.0199.

If for the semi-log relation

$$\ln \mathrm{PEARN}_i = \beta_1 + \beta_2\mathrm{PRIZE}_i + \varepsilon_i$$

we want the partial effect of PRIZE on E(PEARN|PRIZE) then we have to evaluate that conditional expectation. This is possible if we assume

$$\varepsilon_i|\mathrm{PRIZE}_i \sim N(0, \sigma^2)$$

PEARN has a lognormal distribution and

$$\mathrm{E}(\mathrm{PEARN}_i|\mathrm{PRIZE}_i) = e^{\beta_1 + \beta_2\mathrm{PRIZE}_i + \frac{\sigma^2}{2}}$$

We have to take the derivative with respect to PRIZE and choose a value at which to evaluate that derivative. The resulting expression is nonlinear in the parameters and we have to delay the computation of the standard error of such a partial effect. Note that we also need the variance of $\hat{\sigma}^2$ (also the covariance of $\hat{\sigma}^2$ and the OLS estimates of the regression coefficients?).