Econ 140 - Spring 2016
Section 7

GSI: Fenella Carpena

March 10, 2016

1 Hypothesis Testing in MRM: Overview

Type	Example	Test Statistic
1. Individual		
2. Joint		
3. Linear		

2 Hypothesis Testing in MRM: Single Coefficients and Joint Tests

Suppose we have the following model that explains baseball players' salaries.

$$salary_i = \beta_0 + \beta_1 years_i + \beta_2 gamesyr_i + \beta_3 bavg_i + \beta_4 hrunsyr_i + \beta_5 rbisyr_i + u_i \tag{1}$$

where for each player i, *salary* is the salary in 1993, *years* is years in the league, *gamesyr* is average games played per year, *bavg* is the career batting average, *hrunsyr* is the number of home runs per year, *rbisyr* is runs batted in per year. Further, suppose that we estimated the above equation using data we have on hand, and that we obtained the following regression results

$$\widehat{salary} = \underset{(0.29)}{11.10} + \underset{(0.0121)}{0.0689} \cdot years + \underset{(0.0026)}{0.0126} \cdot gamesyr + \underset{(0.0010)}{0.00098} \cdot bavg + \underset{(0.0161)}{0.0144} \cdot hrunsyr + \underset{(0.0072)}{0.0108} \cdot rbisyr$$

$$N = 353,\ SSR = 183.186,\ R^2 = 0.6278$$

Example 2.1. What test statistic would we use to test the hypothesis $H_0 : \beta_4 = 0, H_1 : \beta_4 \neq 0$? Carry out this test at the 5% level?

Example 2.2. What test statistic would we use to test the hypothesis $H_0 : \beta_2 = 0.1, H_1 : \beta_2 \neq 0.1$? Carry out this test at the 10% level.

Example 2.3. A sports analyst hypothesizes that once years in the league and games per year have been controlled for, the variables *bavg*, *hrunsyr*, and *rbisyr* (which we can think of as measure of performance) have no effect on salary.

(a) What is the null and alternative hypothesis?

(b) What test statistic would we use to test the hypothesis in part (a)? Assuming that the population errors are homoskedastic, what is the formula for this test statistic?

(c) What is the restricted regression?

(d) What is the unrestricted regression?

(e) What is q?

(f) What is n?

(g) What is k?

(h) Suppose that a regression of *salary* on *years* and *gamesyr* yielded an SSR of 198.311. Calculate the F-statistic.

(i) Given a significance level of 5%, what is the critical value from the F-distribution.

(j) What is the conclusion of your hypothesis test?

Example 2.4. Let us consider again the joint hypothesis

$$H_0 : \beta_3 = 0,\ \beta_4 = 0,\ \beta_5 = 0\ vs.\ H_1 : \text{at least one of } \beta_3, \beta_4, \beta_5 \text{ is not equal to } 0$$

that we tested in Example 2.3 using an F-test. Is it possible to carry out this joint hypothesis test using the 3 t-statistics from the following 3 individual tests: (1) $H_0 : \beta_3 = 0, H_1 : \beta_3 \neq 0$; (2) $H_0 : \beta_4 = 0, H_1 : \beta_4 \neq 0$; (3) $H_0 : \beta_5 = 0, H_1 : \beta_5 \neq 0$?

Example 2.5. Looking back at Example 2.3, we found that we rejected the joint hypothesis that *bavg*, *hrunsyr*, *rbisyr* have no effect on salary. But if we had looked at each of these variables individually, we would have failed to reject each null hypothesis separately because the individual t-stats are less than 1.96. What might explain the difference in these results?

3 Hypothesis Testing in MRM: Linear Restrictions

Example 3.1. (Adapted from Stock and Watson, Exercise 7.9) Consider the regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$, and the hypothesis test $H_0 : \beta_1 = \beta_2$, $H_1 : \beta_1 \neq \beta_2$.

(a) What test statistics can we use to carry out the above hypothesis test?

(b) Describe how you would calculate the F-statistic (under the assumption of homoskedasticity). What is the restricted and unrestricted regression?

(c) Describe how you can use a t-statistic to test $H_0 : \beta_1 = \beta_2$, $H_1 : \beta_1 \neq \beta_2$. Specifically, transform the regression so that you can use a t-statistic to carry out the test.

Example 3.2. In the same regression as in in Example 3.1, transform the regression so that you can use a t-statistic to test $\beta_1 + 2\beta_2 = 0$.

4 Hypothesis Testing in MRM: Stata

Example 4.1. Suppose we have 1980 census data on the 50 states recording the population size in each state (`pop`), the median age (`medage`), the number of deaths (`death`), the number of marriages, (`marriage`), and the number of divorces (`divorce`). We estimate the following regression:

```
. reg pop medage death marriage divorce

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  4,    45) = 1299.46
       Model |  1.0800e+15     4  2.7000e+14           Prob > F      =  0.0000
    Residual |  9.3500e+12    45  2.0778e+11           R-squared     =  0.9914
-------------+------------------------------           Adj R-squared =  0.9907
       Total |  1.0893e+15    49  2.2232e+13           Root MSE      =  4.6e+05

------------------------------------------------------------------------------
         pop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      medage |  -181303.8   43749.97    -4.14   0.000    -269420.8   -93186.87
       death |   91.30243   4.137673    22.07   0.000     82.96873    99.63613
    marriage |    1.80206   4.303597     0.42   0.677    -6.865829    10.46995
     divorce |   39.80303   8.146704     4.89   0.000     23.39473    56.21134
       _cons |    5241295    1272002     4.12   0.000      2679350     7803239
------------------------------------------------------------------------------
```

What Stata commands would you use to test the following hypothesis?

(a) Test the individual hypothesis that the coefficient on `medage` is zero.

(b) Test the joint hypothesis that the coefficients on all four regressors are simultaneously zero.

(c) Test the linear hypothesis that the coefficient on `death` minus the coefficient on `marriage` is zero.

5 Additional Exercise: Spring 2014 MT2, Question 4

In this exercise we use a dataset containing information on 269 NBA basketball players including their salaries. Table 1 below (see next page) shows the results from 3 OLS regressions where heteroskedasticity-robust standard errors are given in square brackets. The dependent variable is `salary` (annual salary in thousands $), and the explanatory variables are `points` (average points per game), `rebounds` (average rebounds per game), and `assists` (average assists per game).

(a) [6] What is the interpretation of the coefficient on `points` in all three regressions? Give the meaning of the OLSE of the points coefficient from the third regression.

Table 1: Regressions on NBA Salaries

Variables	Model 1	Model 2	Model 3
points	111.67	87.72	80.67
	[8.18]**	[9.43]**	[11.20]**
rebounds		86.44	93.36
		[20.62]**	[21.25]**
assists			26.08
			[21.56]
_cons	278.10	137.56	115.02
	[83.62]**	[83.09]	[86.09]
R2	0.43	0.47	0.48
N	269	269	269

Notes: * p-value<0.05; ** p-value<0.01.

(b) [8] When the `rebounds` variable is added, both the R2 and the Adjusted R2 increase whereas when additionally `assists` variable is added, only the R2 increases. Does this indicate that `rebounds` should be in the regression and that `assists` should not? Explain.

(c) [5] We wonder whether a player's position matters, thinking that different positions may be more valuable to a team than others. To investigate, we regresses `salary` on dummy indicators of each of the three possible court positions a player can have, `center`, `forward`, and `guard` (and a constant), with no other explanatory variables. Stata refuses to report OLSEs for all three regressor. Why does this happen? How would you solve this problem?

(d) [3] Based on the results, it seems that the marginal effect of `rebounds` on salaries is higher than the marginal effect of `points`. So immediately after running the regression for model 3 we issue the command: `test rebounds=points`. What does this Stata command do?

(e) [7] Stata generates output from the command in (d) that includes an F-statistic. Suppose that it takes the value: 4.60. Decide whether you reject the null at the 1% level of significance.

TABLE 5A Critical Values for the F_{n_1,n_2} Distribution—10% Significance Level

Denominator Degrees of Freedom (n_2)	Numerator Degrees of Freedom (n_1) 1	2	3	4	5	6	7	8	9	10
1	39.86	49.50	53.59	55.83	57.24	58.20	58.90	59.44	59.86	60.20
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65
∞	**2.71**	**2.30**	**2.08**	**1.94**	**1.85**	**1.77**	**1.72**	**1.67**	**1.63**	**1.60**

This table contains the 90th percentile of the F_{n_1,n_2} distribution, which serves as the critical values for a test with a 10% significance level.

TABLE 5B Critical Values for the F_{n_1,n_2} Distribution—5% Significance Level

Denominator Degrees of Freedom (n_2)	Numerator Degrees of Freedom (n_1) 1	2	3	4	5	6	7	8	9	10
1	161.40	199.50	215.70	224.60	230.20	234.00	236.80	238.90	240.50	241.90
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.39	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
∞	**3.84**	**3.00**	**2.60**	**2.37**	**2.21**	**2.10**	**2.01**	**1.94**	**1.88**	**1.83**

This table contains the 95th percentile of the distribution F_{n_1,n_2} which serves as the critical values for a test with a 5% significance level.

TABLE 5C Critical Values for the F_{n_1,n_2} Distribution—1% Significance Level

Denominator Degrees of Freedom (n_2)	Numerator Degrees of Freedom (n_1) 1	2	3	4	5	6	7	8	9	10
1	4052.00	4999.00	5403.00	5624.00	5763.00	5859.00	5928.00	5981.00	6022.00	6055.00
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	**6.63**	**4.61**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.41**	**2.32**

This table contains the 99th percentile of the F_{n_1,n_2} distribution, which serves as the critical values for a test with a 1% significance level.