Multiple Regression Analysis: Inference

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

BS1802 Statistics and Econometrics

Jiahua Wu

382 Business School
j.wu@imperial.ac.uk

Roadmap

- Regression analysis with cross-sectional data
  - The multiple regression analysis
    - Basics: estimation, inference, analysis with dummy variables
    - More technically involved: asymptotics, heteroskedasticity, specification and data issues
- Advanced topics
  - Limited dependent variable models
  - Panel data analysis
  - Regression analysis with time series data

Outline (Wooldridge, Ch. 4.1 - 4.6)

- Sampling distribution of the OLS estimators
- Testing hypotheses about a single population parameter: t test
- Confidence intervals
- Testing hypotheses about a single linear combination of parameters
- Testing multiple linear restrictions: F test

Outline

- Sampling distribution of the OLS estimators
- Testing hypotheses about a single population parameter: t test
- Confidence intervals
- Testing hypotheses about a single linear combination of parameters
- Testing multiple linear restrictions: F test

Motivation

- The multiple regression model:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

- Goal is to gain knowledge about the population parameters (β's) in the model
- OLS provides the point estimates of the parameters; OLS will get it right on average (being unbiased)
- Knowing the mean and variance of $\hat{\beta}_j$ is not enough. We need the sampling distribution of the OLS estimators to answer questions, such as
  - what we can say about the "true values"?
  - how to decide if a hypothesis is supported or not?

Sampling Distribution of OLS

Theorem (4.1, Normal Sampling Distribution)

With a "good" model,

$$\hat{\beta}_j \sim Normal\left(\beta_j, Var(\hat{\beta}_j)\right),$$

where the variance is given by

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}, \qquad j = 1, \ldots, k.$$

It implies:

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim \text{Normal}(0,1), \qquad \text{where } sd(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$$

Sampling Distribution of OLS

In practice, σ^2 has to be estimated:

$$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{SST_j(1 - R_j^2)}}, \qquad j = 1, \ldots, k,$$

which is called the standard error of $\hat{\beta}_j$.

Theorem (4.2, t-Distribution)

With a "good" model,

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1},$$

where $k+1$ is the number of unknown parameters in the model, and $n-k-1$ is the degrees of freedom (df).

Outline

- Sampling distribution of the OLS estimators
- Testing hypotheses about a single population parameter: t test
- Confidence intervals
- Testing hypotheses about a single linear combination of parameters
- Testing multiple linear restrictions: F test

Testing Simple Null Hypothesis

- Some questions of interest may be formulated as a simple null hypothesis about a population parameter,

$$H_0 : \beta_j = 0$$

- Eg. In the log wage model

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u,$$

"$H_0 : \beta_1 = 0$", is economically interesting. If the null hypothesis is accepted, it implies that, holding *exper* and *tenure* fixed, a person's education level has no effect on wage.

Testing Simple Null Hypothesis

- To test a simple null hypothesis, the test statistic is usually called "the" *t* statistic or "the" *t* ratio

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

- Sampling distribution of t statistic when H_0 is true
  - By Theorem 4.2, $t_{\hat{\beta}_j}$ has the t-distribution with $n-k-1$ df
  - When df is large (> 120), the t distribution approaches the standard normal distribution

Testing Simple Null Hypothesis

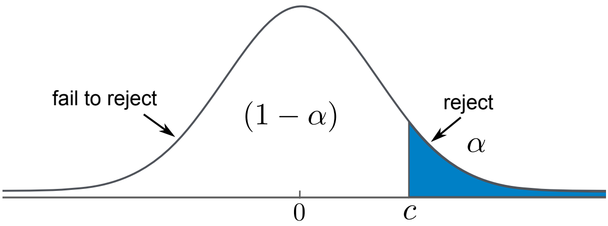
- t statistic along with a rejection rule (depends on alternative hypothesis and the chosen significance level) will be used to determine whether to accept the null hypothesis H_0
- Significance level
  - the probability of rejecting H_0 when it is true
  - typical values: 1%, 5%, 10%
- Alternative hypothesis
  - H_1 may be one-sided, or two-sided
  - $H_1 : \beta_j > 0$ or $H_1 : \beta_j < 0$ are one-sided
  - $H_1 : \beta_j \neq 0$ is a two-sided alternative

One-Sided Alternatives

- Testing against $H_1 : \beta_j > 0$
  - Pick a significance level, α
  - Look up the $(1-\alpha)^{th}$ percentile in a t distribution with $n-k-1$ df and call this c, the critical value (use normal critical values when df > 120)
  - Reject the null hypothesis if the t statistic is greater than c

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$$

$$H_0 : \beta_j = 0 \qquad\qquad H_1 : \beta_j > 0$$

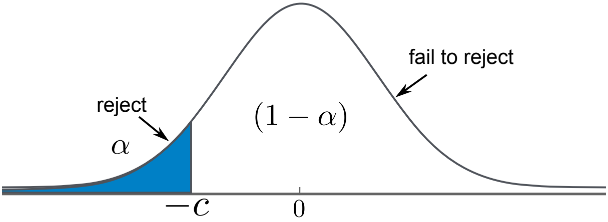


One-Sided Alternatives

- Testing against $H_1 : \beta_j < 0$
  - The critical value is just the negative of before because the t distribution is symmetric
  - Reject the null if $t_{\hat{\beta}_j} < -c$
  - If $t_{\hat{\beta}_j} \geq -c$ then we fail to reject the null

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$$

$$H_0 : \beta_j = 0 \qquad\qquad H_1 : \beta_j < 0$$

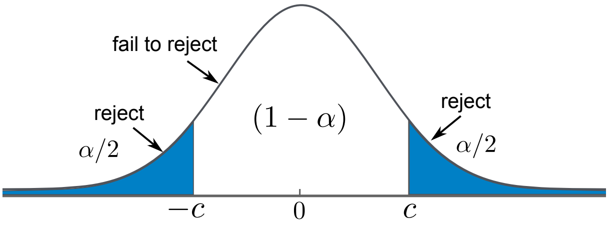


Two-Sided Alternatives

- For a two-sided test ($H_1 : \beta_j \neq 0$)
  - The critical value is based on $(1 - \alpha/2)$ percentile in a t distribution with $n - k - 1$ df
  - Reject $H_0 : \beta_j = 0$ if the absolute value of the t statistic is greater than c, i.e., $|t_{\hat{\beta}_j}| > c$

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$$

$$H_0 : \beta_j = 0 \qquad\qquad H_1 : \beta_j \neq 0$$



Terminology

- Unless otherwise stated, the alternative is assumed to be two-sided
- In the case of $H_0 : \beta_j = 0$ and $H_1 : \beta_j \neq 0$,
  - if we reject the null, we typically say "x_j is statistically significant (or different from 0) at the α level" or
  - if we fail to reject the null, we typically say "x_j is statistically insignificant at the α level"

Testing Simple Null Hypothesis: An Example

Example 4.1. Log wage model (standard errors are in brackets):

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$

$n = 526$, $R^2 = .316$

- Q. Is "returns to education" statistically significant at the 1% level, after controlling for experience and tenure?
- Hypotheses:

Testing Simple Null Hypothesis: An Example

Example 4.1. Log wage model (standard errors are in brackets):

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$

$n = 526$, $R^2 = .316$

- Q. Is "returns to education" statistically significant at the 1% level, after controlling for experience and tenure?
- Hypotheses: $H_0 : \beta_{educ} = 0$ vs $H_1 : \beta_{educ} \neq 0$
- Test statistic and decision rule:

Testing Simple Null Hypothesis: An Example

Example 4.1. Log wage model (standard errors are in brackets):

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$

$n = 526$, $R^2 = .316$

- Q. Is "returns to education" statistically significant at the 1% level, after controlling for experience and tenure?
- Hypotheses: $H_0 : \beta_{educ} = 0$ vs $H_1 : \beta_{educ} \neq 0$
- Test statistic and decision rule: reject H_0 if $|t_{\hat{\beta}_{educ}}| > c$
- Critical value (large df, normal):

Testing Simple Null Hypothesis: An Example

Example 4.1. Log wage model (standard errors are in brackets):

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$

$n = 526$, $R^2 = .316$

- Q. Is "returns to education" statistically significant at the 1% level, after controlling for experience and tenure?
- Hypotheses: $H_0 : \beta_{educ} = 0$ vs $H_1 : \beta_{educ} \neq 0$
- Test statistic and decision rule: reject H_0 if $|t_{\hat{\beta}_{educ}}| > c$
- Critical value (large df, normal): $c = 2.576$
- Conclusion:

Testing Simple Null Hypothesis: An Example

Example 4.1. Log wage model (standard errors are in brackets):

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$

$n = 526,\ R^2 = .316$

- Q. Is "returns to education" statistically significant at the 1% level, after controlling for experience and tenure?
- Hypotheses: $H_0 : \beta_{educ} = 0$ vs $H_1 : \beta_{educ} \neq 0$
- Test statistic and decision rule: reject H_0 if $|t_{\hat{\beta}_{educ}}| > c$
- Critical value (large df, normal): $c = 2.576$
- Conclusion: reject H_0 at the 1% level because

$$|t_{\hat{\beta}_{educ}}| = .092/.007 = 13.149 > c$$

p-Values

- An alternative to the classical approach is to ask, "what is the smallest significance level at which the null would be rejected?"
  - Compute the t statistic
  - p-value is the probability that we'd observe a more extreme test statistic in the direction of the alternative hypothesis than we did, if the null is true
  - Smaller the p-value, stronger the evidence against H_0

p-Values and Testing Other Hypotheses

- p-values for t tests
  - Most computer packages compute the p-value for you, assuming a two-sided test
  - If you want a one-sided alternative, just divide the two-sided p-value by 2
  - R provides the t statistic, p-value for $H_0 : \beta_j = 0$ in columns labeled "`t value`", and "`Pr(>|t|)`", respectively
- Testing other hypotheses
  - A more general form of the t statistic: $H_0 : \beta_j = a_j$
  - In this case, the appropriate t statistic is

$$t = \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)},$$

where $a_j = 0$ for the standard test

Testing Simple Null Hypothesis: An Example

Example 4.5.(hprice2.RData) Housing Prices and Air Pollution:

$$\begin{aligned}\widehat{log(price)} &= \underset{(.32)}{11.08} - \underset{(.117)}{.954}\, log(nox) - \underset{(.043)}{.134}\, log(dist) \\ &\quad + \underset{(.019)}{.255}\, rooms - \underset{(.006)}{.052}\, stratio,\end{aligned}$$

$n = 506,\ R^2 = .581$

- Variable description
  - *nox*: the amount of nitrogen oxide in the air
  - *dist*: distance of the community from employment centers
  - *rooms*: the average number of rooms in houses
  - *stratio*: the average student-teacher ratio of schools in the community
- Q. Can we reject $H_0 : \beta_{nox} = -1$ at the 5% level?

Economical/Statistical Significance

- An explanatory variable is statistically significant when the size of the t-ratio $t_{\hat{\beta}_j}$ is sufficiently large (beyond the critical value c)
- An explanatory variable is economically (practically) significant when the size of the estimate $\hat{\beta}_j$ is sufficiently large (in comparison to the size of y)
- An important x should be both statistically and economically significant

Outline

- Sampling distribution of the OLS estimators
- Testing hypotheses about a single population parameter: t test
- Confidence intervals
- Testing hypotheses about a single linear combination of parameters
- Testing multiple linear restrictions: F test

Confidence Intervals

- The confidence interval (CI) for β_j is based on

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

- A $(1-\alpha)$% CI is defined as

$$\hat{\beta}_j \pm c \cdot se(\hat{\beta}_j) = \left[\hat{\beta}_j - c \cdot se(\hat{\beta}_j), \hat{\beta}_j + c \cdot se(\hat{\beta}_j)\right] = [L, U],$$

  where c is the $(1-\alpha/2)$ percentile in a t_{n-k-1} distribution
- The interpretation of 95% CI
  - if many random samples are drawn and $[L, U]$ is computed for each sample, then 95% of these $[L, U]$ will cover the true population parameter β_j

Confidence Intervals and Two-Sided Tests

- When *df* is large (> 120), the t_{n-k-1} distribution is very close to the normal distribution and we use $N(0,1)$ critical values
  - eg. For large *df*, the 95% CI is about $\hat{\beta}_j \pm 1.96 \cdot se(\hat{\beta}_j)$
- The width of CI depends on the standard error $se(\hat{\beta}_j)$ and the critical value c
  - high confidence level $\rightarrow$ large c $\rightarrow$ wide CI
  - large standard error $\rightarrow$ wide CI
- CI and two-sided test
  - test "$H_0 : \beta_j = a_j$" against "$H_1 : \beta_j \neq a_j$"
  - reject H_0 at the α% significant level if (and only if) the $(1-\alpha)$% CI does not contain a_j

Confidence Intervals: An Example

- Example 4.1. Log wage model (standard errors are in brackets):

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure,$$

$n = 526$, $R^2 = .316$

  - The 95% CI for β_{educ}: $n - k - 1 = 522$ (large sample, use normal), $c = 1.96$,

$$.092 \pm 1.96 \cdot (.007) = [.078, .106]$$

  - reject "$H_0 : \beta_{educ} = 0$" in favor of the two-sided H_1 at the 5% significant level

Outline

- Sampling distribution of the OLS estimators
- Testing hypotheses about a single population parameter: t test
- Confidence intervals
- Testing hypotheses about a single linear combination of parameters
- Testing multiple linear restrictions: F test

Testing A Linear Combination of Parameters

- In the log wage model,

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u.$$

  Suppose we wish to see whether or not *educ* has the same effect on $\log(wage)$ as *exper*, i.e., to test

$$H_0 : \beta_1 - \beta_2 = 0 \quad \text{vs} \quad H_1 : \beta_1 - \beta_2 \neq 0,$$

  which involves a combination of 2 parameters

- Conceptually, we could use

$$t_{\hat{\beta}_1 - \hat{\beta}_2} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)}.$$

  However, $se(\hat{\beta}_1 - \hat{\beta}_2)$ is not usually provided by software

Testing A Linear Combination of Parameters

- Alternatively, we can re-parameterise the log wage model

$$\log(wage) = \beta_0 + \theta educ + \beta_2(exper + educ) + u,$$

  where $\theta = \beta_1 - \beta_2$

- The hypotheses become

$$H_0 : \theta = 0 \quad \text{vs} \quad H_1 : \theta \neq 0,$$

  which can easily be tested by regressing $\log(wage)$ on *educ* and $(exper + educ)$

- The idea here is to isolate the parameter of interest $\theta = \beta_1 - \beta_2$ by re-parameterisation. The OLS output provides both $\hat{\theta}$ and $se(\hat{\theta})$

Testing A Linear Combination of Parameters

- Eg. Log wage model (standard errors are in brackets)

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$

$n = 526,\ R^2 = .316$

  - Hypotheses: $H_0 : \beta_{educ} - \beta_{exper} = 0$ vs $H_1 : \beta_{educ} - \beta_{exper} \neq 0$

- Re-parameterised model

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.0070)}{.0879}\, educ + \underset{(.0017)}{.0041}\,(educ + exper) + \underset{(.003)}{.022}\, tenure$$

$n = 526,\ R^2 = .316$

  - Hypotheses: $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$
  - Test statistic $t_{\hat{\theta}} = .0879/.0070 = 12.59$

Outline

- Sampling distribution of the OLS estimators
- Testing hypotheses about a single population parameter: t test
- Confidence intervals
- Testing hypotheses about a single linear combination of parameters
- Testing multiple linear restrictions: F test

Testing Multiple Linear Restrictions

- Everything we have done so far has involved testing a single linear restriction (eg, $\beta_1 = 0$ or $\beta_1 = \beta_2$)
- We may want to check whether or not a group of x variables has a joint effect on y (with the rest of x variables as controls)
  - i.e., testing exclusion restrictions - whether a group of parameters are all equal to zero

Testing Exclusion Restrictions

- The unrestricted model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

- q restrictions under the null hypothesis

$$H_0 : \beta_{k-q+1} = 0, \ldots, \beta_k = 0$$

- The alternative is just $H_1 : H_0$ is not true
- Under H_0, the restricted model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-q} x_{k-q} + u_{(r)}$$

Testing Exclusion Restrictions

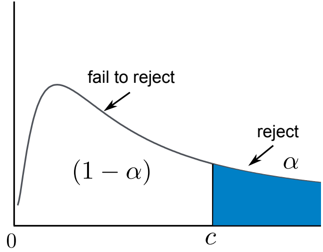
- To do the test, we need to estimate the restricted model without $x_{k-q+1}, \dots, x_k$, as well as the unrestricted model with all x's included
- Intuitively, we want to know if the change in SSR is big enough to warrant inclusion of $x_{k-q+1}, \dots, x_k$
- Test statistic

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \sim F_{q,n-k-1} \qquad \text{under } H_0$$

  - $q =$ number of restrictions, or $df_r - df_{ur}$
  - $n - k - 1 = df_{ur}$

F Statistic

- The F statistic is always positive, since the SSR from the restricted model cannot be less than the SSR from the unrestricted
- Reject H_0 if the increase in SSR when we move from the unrestricted to the restricted model is "big enough"
- Decision rule: reject H_0 if $F > c$ ($F_{q,n-k-1}$ critical value)



- F and t statistics
  - when $q = 1$, H_0 can be tested with either t stat or F stat

The R^2 Form of the F Statistic

- Using the fact that $SSR_r = SST(1 - R_r^2)$ and $SSR_{ur} = SST(1 - R_{ur}^2)$, we have

$$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n - k - 1)},$$

  where r is restricted and ur is unrestricted

- This is called the R-squared form of the F statistic

Testing Exclusion Restrictions

- If H_0 is rejected, we say that $x_{k-q+1}, \ldots, x_k$ are jointly statistically significant
- If H_0 is not rejected, we say that $x_{k-q+1}, \ldots, x_k$ are jointly insignificant, which justifies dropping them from the model
- The p-value for F test is the probability of F distribution beyond observed F statistics

F Tests: An Example

- Example 4.9. Child birth weight and parents' education

$$\begin{aligned} bwght &= \beta_0 + \beta_1 cigs + \beta_2 parity + \beta_3 faminc \\ &\quad + \beta_4 motheduc + \beta_5 fatheduc + u \end{aligned}$$

  - *bwght*: birth weight
  - *cigs*: average cigarettes per day by mother
  - *parity*: birth order
  - *faminc*: family income
  - *motheduc*: years of education for mother
  - *fatheduc*: years of education for father

- Hypotheses: $H_0 : \beta_4 = 0$ and $\beta_5 = 0$ vs $H_1 : H_0$ is false

F Test: An Example

- Unrestricted model (ur)

$$\begin{aligned} bwght &= \beta_0 + \beta_1 cigs + \beta_2 parity + \beta_3 faminc \\ &\quad + \beta_4 motheduc + \beta_5 fatheduc + u \end{aligned}$$

$\rightarrow SSR_{ur}$

- Restricted model (r)

$$bwght = \beta_0 + \beta_1 cigs + \beta_2 parity + \beta_3 faminc + u_{(r)}$$

$\rightarrow SSR_r$

F Test: An Example

- The F statistic is the relative difference between SSR_r and SSR_{ur}
$$F = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(n-6)}$$
- Under H_0, F has the F-distribution
$$F = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(n-6)} \sim F_{2,n-6},$$
with $(2, n-6)$ degrees of freedom
- Decision rule: reject if $F > c$, where c is the $F_{2,n-6}$ critical value

F Test: An Example

- Use the data in bwght.RData: $n = 1191$, $R_r^2 = .0364$ and $R_{ur}^2 = .0387$.

$$F = \frac{(R_{ur}^2 - R_r^2)/2}{(1 - R_{ur}^2)/(n-6)} \approx 1.42$$

- The 5% $F_{2,n-6}$ critical value is $c = 3.00$
- According to the decision rule, H_0 is not rejected at the 5% level because $F < c$

F Test for Overall Significance of a Regression

- When $q = k$, the null "$H_0 : \beta_1 = 0, \ldots, \beta_k = 0$" is routinely tested by most regression packages, known as the F test for overall significance
- The null is that none of the explanatory variables has an effect on y. The restricted model is simply

$$y = \beta_0 + u$$

- The F stat under the null has an $F_{k,n-k-1}$ distribution. As the R-squared is zero under null, this F stat is

$$F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)},$$

where R^2 is from the unrestricted model

A Caveat: Cannot just check each t statistic separately!

- We cannot test exclusion restrictions by checking each t statistic separately!

- ```
  > x1 <- rnorm(100, mean = 1, sd = 2)
  > x2 <- x1 + rnorm(100, mean = 1, sd = 1)
  > y <- x1 + x2 + rnorm(100, mean = 1, sd = 8)
  > m1 <- lm(y ~ x1 + x2)
  > stargazer(m1, align = TRUE, no.space = TRUE)
  ```
  ```

A Caveat: Cannot just check each *t* statistic separately!

	Dependent variable:
	y
x1	-0.073
	(0.898)
x2	1.500^{*}
	(0.780)
Constant	0.563
	(1.197)
Observations	100
R^2	0.153
Adjusted R^2	0.136
Residual Std. Error	7.812 (df = 97)
F Statistic	8.759^{***} (df = 2; 97)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

It is possible that a group variables are jointly significant but individually insignificant. This is a symptom of a group of highly correlated variables