

Семинар № 10

Исследование качества спецификации эконометрической модели

План

1. Коэффициент детерминации модели и F-тест качества её спецификации
2. Скорректированный коэффициент детерминации как инструмент отбора в модель объясняющих переменных (как инструмент модификации)
3. ДЗ

В процессе оценивания эконометрической мод. модели Самуэльсона-Хиггса расходов России методом наименьших квадратов полезно исследовать в качестве

$$\begin{cases} C_t = a_0 + a_1 \cdot Y_{t-1} + a_2 \cdot Cr_t + a_3 \cdot San_t + u_t; \\ E(u_t) = 0; Var(u_t) = \sigma_u^2 \end{cases} \quad (1)$$

Коэффициент детерминации, который вычисляется по следующему правилу:

$$R^2 = 1 - \frac{\sum_{i=1} \tilde{u}_i^2}{\sum_{i=1} (y_i - \bar{y})^2} \quad (2)$$

имеет смысл *доли текущих эндогенных переменных*, которые объясняются predetermined переменными. Этот коэффициент автоматически вычисляется в статистических пакетах например в протоколе функции ЛИНЕЙН размещается в 3 строчке 1 столбца этого протокола. Этот коэффициент также определяется функцией *ml* пакета R.

F текст

F тест формализованная процедура проверки гипотезы о неудовлетворительной спецификации эконометрической модели, то есть гипотезы, что ни одна объясняющая переменная модели не несёт в себе какую либо информации об эндогенной переменной модели, то есть этот тест проверяет гипотезу, что все коэффициенты при объясняющих коэффициентах равны 0. Например приминительно к модели (1) этот тест проверяет следующую гипотезу

$$H_0 : a_1 = a_2 = a_3 = 0 \quad (3)$$

Если эта гипотеза отвергается, то качество выбора объясняющих переменных считается удовлетворительным.

Задача. Исследовать качество спецификации модели (1) анализируя коэффициент детерминации этой модели и осуществляя F-тест. Статистика в критерии гипотезы (3) в F-тесте автоматически рассчитывается пакетами и в протоколе функции ЛИНЕЙН находится в 4 строчке 1 столбца протокола.

Решение:

Коэффициент детерминации

	a3	a2	a1	a0
	-2009.150407	-3606.31029	0.839684	-12106.6
	306.1664607	437.9424085	0.023019	851.7167
R2=	0.992798771	406.519065	#Н/Д	#Н/Д
	505.5056988	11	#Н/Д	#Н/Д
	250616203.5	1817835.253	#Н/Д	#Н/Д

В нашем примере коэффициент детерминации равен 0.99 и это значит, что в модели (1) лаговый доход, индикаторы кризиса и санкции объясняют на 99% текущий уровень потребления домохозяйств. Добавим, что статистика критерия (3) о неудовлетворительной спецификации модели находится в 4 строке 1 столбца и в нашем примере равна 505.5

Отметим правило по которому рассчитано значение статистики:

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

Гипотеза (3) о неудовлетворительной спецификации должна быть отвергнута, если величина F превышает критический уровень $F_{\text{крит}}$. Этот уровень имеет смысл квантили F распределения уровня $1 - \alpha$ с количеством степеней свободы $k = 3$ (экзогенные переменные) и n (кол-во ур-ний наблюдений) $- k + 1$; величина $n - k + 1$ автоматически рассчитывается функцией ЛИНЕЙН и всегда расположено в 4 строке 2 столбца протокола. И у нас она равна 11. Рассчитаем $F_{\text{крит}}$ с помощью функции F.ОБР

	a3	a2	a1	a0		
	-2009.150407	-3606.31029	0.839684	-12106.6	DW=	1.3920078
	306.1664607	437.9424085	0.023019	851.7167	Fкрит =	3.5874337
R2=	0.992798771	406.519065	#Н/Д	#Н/Д		
F=	505.5056988	11	#Н/Д	#Н/Д		
	250616203.5	1817835.253	#Н/Д	#Н/Д		

Если $F > F_{\text{крит}}$ то мы отвергаем гипотезы.

Теперь осуществим F тест в *R Studio*.

Шаг 1. Копируем базу данных копируя из *Excel* в блокнот и сохранить в рабочей директории.

Шаг 2. Открываем в *RStudio*

```
# Оценивание множественной регресси
getwd()
file.show("dataRStudio.txt")
C <- read.table("dataRStudio.txt", sep=" ", dec=".", header = TRUE)
C
Cmodel <- lm(data = C, Ct~Yt+Crt+Sant)
summary(Cmodel)
```

Гипотеза (3) о неудовлетворительной спецификации отвергается, если в протоколе *RStudio* величина $p - value$ меньше чем 0.05. В нашем случае $p - value : 0.5$. Если же $p - value$ больше чем 0.5, то гипотеза не может быть отвергнута.

Скорректированный коэффициент детерминации как инструмент модификации модели (как инструмент отбора в модели дополнительных объясняющих переменных)

В протоколе функции *RStudio* скорректированный коэффициент детерминации находится в предпоследней строчке протокола справа и в нашем примере он имеет значение 0.9908 обсудим методику отбора объясняющих переменных в эконометрическую модель.

Шаг 1. Сначала включаются все объясняющие переменные и фиксируется значение R^{-2} .

Шаг 2. Удаляется сомнительная переменная например индикатор санкции и фиксируется R^{-2} .

После удаления из модели San_t скорректированный коэффициент детерминации уменьшился 0.9587 является значащей её нельзя удалять из модели

```
# Оценивание множественной регрессии
getwd()
file.show("dataRStudio.txt")
C <- read.table("dataRStudio.txt", sep=" ", dec=".", header = TRUE)
C
# Модификация модели "От общего к частному"
# Шаг 1
Cmodel<-lm(data = C, Ct~Yt+Crt+Sant)
summary(Cmodel)

# Шаг 2
Cmodel2<-lm(data = C, Ct~Yt+Crt)
summary(Cmodel2)
```

Шаг 3. Удалим следующую сомнительную переменную индикатор кризиса.

```
# Оценивание множественной регрессии
getwd()
file.show("dataRStudio.txt")
C <- read.table("dataRStudio.txt", sep=" ", dec=".", header = TRUE)
C
# Модификация модели "От общего к частному"
# Шаг 1
Cmodel<-lm(data = C, Ct~Yt+Crt+Sant)
summary(Cmodel)
```

```
# Шаг 2
Cmodel2<-lm(data = C, Ct~Yt+Crt)
summary(Cmodel2)

# Шаг 3
Cmodel3<-lm(data = C, Ct~Yt+Sant)
summary(Cmodel3)
```

После удаления из модели Cr_t скорректированный коэффициент детерминации уменьшился 0.9398 является значащей её нельзя удалять из модели.

ДЗ Исследовать качество модификации в двух оставшихся моделей.