Эконометрика

Модели с переменной структурой (фиктивные переменные)

Как правило независимые переменные имеют непрерывные области измерения (возраст, стаж, денежные доходы, уровень безработицы).

Однако, существуют переменные которые могут принимать два значения или в общем случае дискретное множество значений.

Необходимость в таких переменных возникает в тех случаях, когда требуется учесть влияние качественных признаков (пол, национальность, уровень образования и т.д).

Для того чтобы вести такие переменные в регрессионную модель, им должны быть присвоены те или иные цифровые метки, т.е. качественные переменные необходимо преобразовать в количественные.

Такого вида сконструированные переменные в эконометрике принято называть ***фиктивными переменными.***

Например, рассмотрим модель формирования заработной платы (**Y**) от количества отработанных часов (X_1) и стажа работы (X_2).

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + U$$

Зависит ли заработная плата от пола работника?

На практике используется два метода моделирования:

1. Регрессия строится для каждой качественно отличной группы единиц совокупности, т.е. для каждой группы в отдельности;

2. Общая регрессионная модель строится для совокупности в целом. В этом случае в регрессионную модель вводятся фиктивные переменные, т.е. строится модель с переменной структурой.

В английской литературе такие переменные называют dummy – фиктивная переменная (косвенным образом придает количественное значение качественным признакам).

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \delta_1 d_1 + U$$

Ведем переменную d_1, присвоив ей значения по следующему правилу:

$d_1 = 1$, если работник мужчина;

$d_1 = 0$, если работник женщина;

Тогда ожидаемое значение заработной платы при одинаковых значениях количества отработанных часов и стажа будет:

Для мужчин

$$\hat{Y} = a + b_1 x_1 + b_2 x_2 + \gamma_1$$

Для женщин:

$$y = a + b_1 x_1 + b_2 x_2$$

Заработная плата мужчин и женщин отличается на величину **γ.**

Проверив с помощью t-статистики значимость коэффициентов регрессии, можно определить, имеет ли место дискриминация по половому признаку.

Если коэффициент γ статистически значим, то очевидно, что есть различия в оплате труда мужчин и женщин при прочих равных условиях. Если этот коэффициент положителен, то дискриминация в пользу мужчин, если отрицателен – в пользу женщин.

Стандартные гипотезы в данном случае имеют следующий смысл:

$$H_0: \quad \gamma_1 = 0$$

– на рынке труда нет дискриминации.

$$H_0: \quad \gamma_1 \neq 0$$

– дискриминация присутствует.

Переменные такого типа во всем остальном не отличаются от обычных непрерывных регрессоров для оценивания уравнения с фиктивными переменными МНК коэффициент при фиктивной переменной интерпретируются также как и при остальных регрессорах.

Способ задания значений переменной не влияет на результаты оценивания, т.к. направление влияния данного признака отражает значение коэффициента.

Такая модель называется «Модель с переменной структурой».

Качественные различия можно формализовать с помощью любой переменной принимающей два значения не обязательно 0 и 1.

Однако, в эконометрической практики почти всегда используют фиктивные переменные типа 0 и 1 т.к. в этом случае интерпретация выглядит наиболее наглядно.

Введем в первоначальную модель еще одну фиктивную переменную, отражающую влияние образования на заработную плату:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \delta_1 d_1 + \delta_2 d_2 + U$$

$d_2=1$ – высшее образование;

$d_2=2$ – среднее специальное образование;

$d_2=3$ – бакалавр;

$d_2=4$ – магистр;

$d_2 =0$ - общее среднее образование.

Если включаемый в рассмотрение качественный признак имеет не два, а несколько значений, то можно было бы ввести дискретную переменную, принимающее такое же значение, но в этом случае трудно дать содержательную интерпретацию соответствующему коэффициенту.

На практике в таких случаях используют набор бинарных фиктивных переменных.

Рассмотрим пример: необходимо оценить влияние времени года на потребление некоторого товара.

y – объем потребления некоторого продукта в месяц, кг.

d_1=1, если зима;

d_1=0, в противном случае (любое другое время года);

d_2 =1, если весна;

d_2 = 0, в противном случае;

d_3= 1, если лето;

d_3 = 0, в противном случае.

$$y = \alpha + \delta_1 d_1 + \delta_2 d_2 + \delta_3 d_3 + U$$

- Одна категория должна отсутствовать потому что она эталонная.

- Мы не вводим 4-у бинарную переменную для осени потому что в этом случае выполнялось бы тождество $d1+d2+d3+d4=1$ что означает линейную зависимость регрессоров и невозможность нахождения оценок по МНК.

- Среднемесячный объем потребления в осенние месяцы есть величина α

- Для зимних месяцев объем потребления составляет $\alpha + \delta 1$, для весенних $\alpha + \delta 2$, для летних $\alpha + \delta 3$

- Т.о. оценки коэффициент δ показывают среднее отклонение в объеме потребления по сравнению с осенними месяцами

- Ho:$\alpha = \delta 1$ потребление осенью равно зимой или Ho: $\delta 1 = \delta 2$

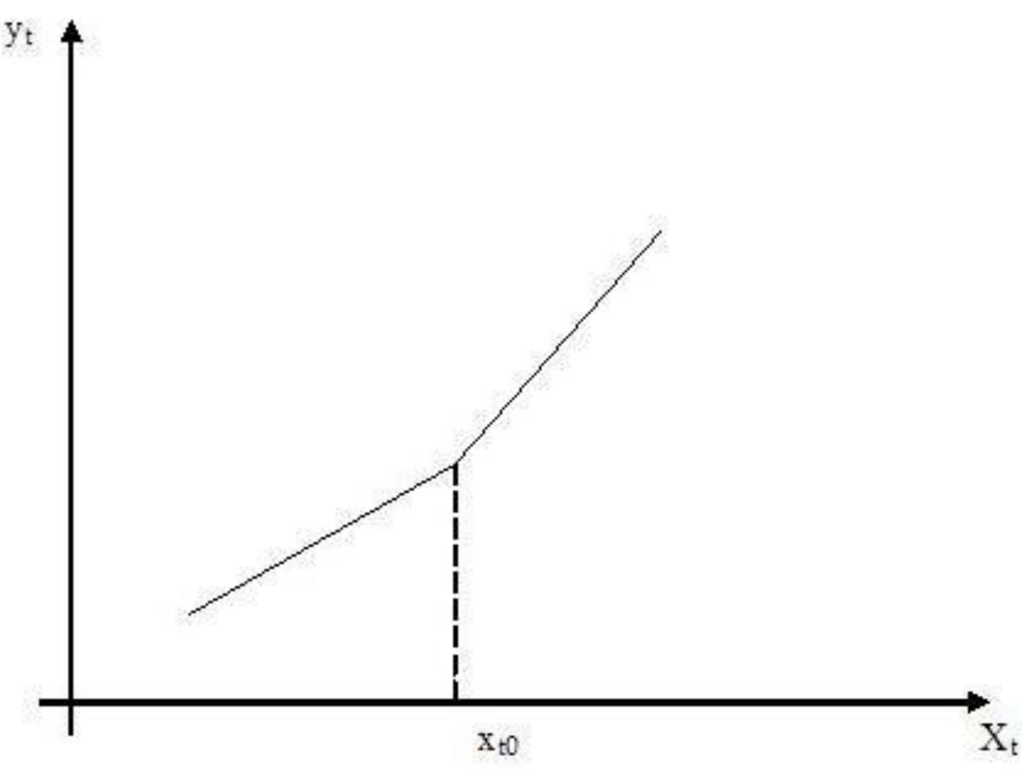
Фиктивные переменные позволяют строить и оценивать так называемые кусочно-линейные модели, которые можно применять для исследования структурных изменений.

Рассмотрим пример.

Пусть y – зависимая перееменная и для простоты в модель включена только 1 независимая переменная x. x и y представлены в виде временных рядов.

xt – размер ОПФ в период времени t,
yt – объем продукции в t.

Из некоторых априорных соображений исследователь считает, что в момент времени t_0 произошла структурная перестановка и линия регрессии будет отличаться от той которая была до момента t_0, но общая регрессия будет непрерывна.



Введем дискретную переменную $r_t = 0$, если $t \leq t_o$ и $r_t = 1$, если $t > t_o$

$$y = \alpha + \beta_1 x_t + \beta_2 (x_t - x_{t0}) r_t + U_t$$

отсюда следует, что регрессионная линия (рис) имеет коэффициент наклона β_1 для $t \leq t_o$ и наклон $\beta_1 + \beta_2$ для $t > t_o$. При этом разрыва в точке t_o не происходит.

- Тестируя стандартную гипотезу $\beta_2 = 0$ мы проверяем предположение о том, что фактически структурные изменения не повлияли на объем выпуска продукции.

- В зависимости от способа включения фиктивной переменной в модель регрессии интерпретация оценок коэффициента при ней будет различной.