Лекция №5.
Линейная модель множественной регрессии (базовая модель эконометрики)
План

1. Завершение обсуждения схемы построения эконометрических моделей;

2. **Линейная модель множественной и парной регрессии (базовая модель эконометрики);**

3. Уравнение наблюдения объекта (схема Гаусса-Маркова), компактная запись схемы Гаусса-Маркова и понятие статичсктической процедуры оценивания параметров модели.

На прошлой лекции присупили к обсуждению схемы построения эконометрических моделей:

1. Спецификая модели;

2. Сбор и проверка статистической информации;

3. Оценивание (найстройка) модели;

4. Проверка адекватности (верификации) модели

Мы разобрали **1 пункт.** Там же подчеркнули, что спецификация эконометрической модели непременно включает в себя неизвестные константы $a_0,\ a_1$, которые называют параметрами модели. В самом общем виде спецификация модели записывается так: $F\left(\vec{x}_t,\ \vec{y}_t;\ \vec{p}\right)\ =\ \vec{u}_t$. Здесь мы начинаем выявлять константы.

$$\begin{cases} Y\ =\ C\ +\ I \\ \ =\ a_0\ + a_1\ \cdot Y\ +\ u \\ 0\ <\ a_1\ <\ 1 \\ E(u)\ =\ 0,\ E\left(u^2\right)\ =\ \sigma_u^2 \end{cases}$$

Добавим, что спецификация временного ряда квартальных уровней ВВП России включает в себя 8 констант:

$$(a_0,\ a_1,\ a_2,\ a_3,\ b_1,\ b_2,\ b_3,\ \sigma_u)$$

Второй этап состоит в сборе и проверке статистической информации в виде конкретных реальных значений переменных входящих в модель. Примером второго этапа служат данные из таблицы № 1 из лекции № 1.

Замечание. Собранная статистическая информация разделяется на две части причём большая часть именуется обучаещей выборкой и предназначена для определения параметров модели; остальная информация именуется тестовой или контролирующей выборкой и используется для проверки адекватности модели.

На третьем этапе схемы методами математической статистики оцениваются параметры модели по обучающей выборке. 4 и 5 практическое занятие служит иллюстрация 3 этапа. Подчеркнём, что всегда удаётся вычислить только приближённые значения параметров $\widetilde{\vec{p}}$ (оценки). Причина приближённого значения оценок параметров состоит в наличии случайных возмущений, порождённых неучтёнными факторами.

Четвёртый этап состоит из проверки адекватности оценённой модели $F\left(\vec{x}_t,\ \vec{y}_t;\ \widetilde{\vec{p}}\right) = \vec{u}_t$ путём сопоставления прогнозов значений эндогенных переменных из контролирующей выборки $\tilde{y}_t = f\left(\vec{x}_t,\ \widetilde{\vec{p}}\right)$ (3.1.11) с реальными значениями. Модели признаётся адекватной, если ошибки прогнозов не превышают критические уровни $|\tilde{y}_t - y_t|$ (ошибка прогноза) $\leq e_{\text{крит}}\left(\tilde{y}_t\right)$ (3.1.12) (15% или $2-3\ \sigma$).

Вывод: схема построения эконометрических моделей состоит из 4 этапов и если модель признаётся не адекватной, то экономист возращается на первый этап и выявляет ошибки спецификации модели.

Линейная модель множественной регрессии (базовая модель эконометрики)

Модель со следующей спецификацией

$$\begin{cases} y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_k x_k + u \\ E(u) = 0;\ E\left(u^2\right) = \sigma_u^2; \end{cases} \tag{3.2.1}$$

является базовой моделью эконометрики и называется *линеной моделью множественной регресии*. Символом y обозначена единственная объясняемая переменная; символом $(x_1,\ x_2, \ldots,\ x_k)$(3.2.3) обозначены предопределённые (объясняющие переменные); символами $(a_0,\ a_1, \ldots,\ a_k)$(3.2.4) обозначены константы и носят название *коэффициенты модели* (более полно, коэффициентов *функции регрессии*).

$$\tilde{y} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_k x_k$$

Символом $\tilde{y}$ обозначена функция объясняющий перменных, имеющая смысл той части эндогенной переменной y, которая объясняется предопределёнными перменными модели; величина $\tilde{y}$ носит название *функции регрессии* и с точки зрения теории вероятностей является условным математическим ожидание велечины y; u обозначено случайное возмущение.

Смысл коэффициента a_g при переменной x_g

a_j имеет смысл ожидаемого предельного значения переменной y по переменной x_j.

$$E(\triangle\, y) \;=\; a_j \cdot \triangle\, x_j \tag{3.2.4}$$

Пример (линейной модели множественной модели). Вернёмся к нашей предшествующей лекции и рассмотрим модель квартальных уровней ВВП России с кубическим трендом. Объясняемые переменные - это квартальные уровни датированные кварталами или t, связанный с календарём следующим правилом $t \;=\; 1,$ для первого квартала 2011 года.

$$\begin{cases} Y_t \;=\; a_0 \;+\; a_1 \;\cdot t \;+\; b_1 \;\cdot\; d_1(t) \;+\; b_2 \;\cdot d_2(t) \;+\;\; b_3 \;\cdot d_3(t) \;+ u_t \\ E(u) \;=\; 0;\;\; E\left(u^2\right) \;=\; \sigma_u^2; \\ t \;= 1,\; 2,\; 3, \ldots \\ t \;=\; 1 \;\Rightarrow 1 \text{ квартал 2011 года} \end{cases} \tag{3.2.5}$$

Объясняющими переменными служит 6 переменных: $(t,\; t^2,\; t^3,\; d_1(t),\; d_2(t),\; d_3(t))$.
Замечание. Обратим внимание, что в линейной модели множесвтенной регрессии среди объясящих переменных $d_1(t),\;\; d_2(t),\;\; d_3(t)$, только одная является независимой - переменная t. Это значит, что в ЛММР объясняющие переменные могут быть как независимыми друг от друга, так и являться известными функциями каких-то других переменных велечин. Запомним, что аргумент - независим, а экзогенная переменная может быть зависимой.

Линейная модель парной регрессии

Простейшим случаем ЛММР служит модель парной регрессии с одной объясняющей переменной.

$$\begin{cases} y \;=\; a_0 \;+\; a_1 \;\cdot x \;+\; u \\ E(u) \;=\; 0;\;\; E\left(u^2\right) \;=\; \sigma_u^2; \end{cases} \tag{3.2.6}$$

Приведём важный для инвестиционного анализа пример инвестиционной модели. Является рыночная модель ценной бумаги.

$$\begin{cases} r \;=\; \alpha \;+\; \beta \;\cdot r_I \;+\; u \\ E(u) \;=\; 0;\;\; E\left(u^2\right) \;=\; \sigma_u^2; \end{cases} \tag{3.2.7}$$

$\boxed{\text{ДЗ}}$ Рыночная модель ценной бумаги У. Шарп, Г. Александер, Д. Бэйли. Объясняющие - это доходность на акцию за 1 месяц r. Доходность на рыночный актив r_I.

Добавим в следующей таблице приведены значения переменных r b r_I рыночной модели компании Лукойл.

Вывод: линейная модель имеет спецификацию (3.2.1), которая имеет следующие параметры ($a_0,\ a_1,\ \ldots,\ a_k,\ \sigma_u$). Приступаем к статистической процедуре оценивания этих параметров.

Разместим обучающую выборку при построении ЛММР в следующей таблице: Подставляем каждую строчку в уравнение модели (2.1)

№ наблюдений	y	x_1	x_2	$\ldots$	x_k
1	y_1	$x_{1,1}$	$x_{2,1}$	$\ldots$	$x_{k,1}$
2	y_2	$x_{1,2}$	$x_{2,2}$	$\ldots$	$x_{k,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
n	y_n	$x_{1,n}$	$x_{2,n}$	$\ldots$	$x_{k,n}$

В итоге получим следующую систему уравнений наблюдений в рамках (2.1):

$$\begin{cases} y_1 \ = \ a_0 \ + \ a_1\, x_{1,1} \ + \ a_2\, x_{2,1} \ + \ \ldots \ + \ a_k\, x_{k,1} \ + \ u \\ y_2 \ = \ a_0 \ + \ a_1\, x_{1,2} \ + \ a_2\, x_{2,2} \ + \ \ldots \ + \ a_k\, x_{k,2} \ + \ u \\ \ldots \\ y_n \ = a_0 \ + \ a_1\, x_{1,n} \ + \ a_2\, x_{2,n} \ + \ \ldots \ + \ a_k\, x_{k,n} \ + \ u \end{cases}$$

Её принято называть схемой Гаусса-Маркова.

Компактная запись:

$$\vec{y} \ = \ X \ \cdot \vec{a} \ + \ \vec{u} \tag{3.3.4}$$

$\vec{y}$ это схема левых частей. X – это матрица значений объяняющих перменных, расширенная столбцом 1 (если есть свободный член a_0), $\vec{a}$ –вектор коэффициентов модели $k+1$. $\vec{u}$ – веутор случайных возмущейний.