Эконометрика

Ивашина Наталья Викторовна, к. э.н., доцент Департамента экономических наук ШЭМ

Эконометрика – наука о применении статистических и математических методов в экономическом анализе для проверки правильности экономических теоретических моделей и способов решения экономических проблем [Демидова, Малахов «Эконометрика», 2018]

Задачи эконометрики

- Оценка влияния различных факторов на интересующий нас показатель
- Прогноз экономических показателей

Основные этапы эконометрического исследования

1. Утверждение экономической теории
2. Предложение соответствующей математической модели
3. Предложение соответствующей эконометрической модели
4. Поиск данных для эмпирической проверки эконометрической модели
5. Оценка параметров эконометрической модели
6. Верификация модели (проверка того, является ли выбранная эконометрическая модель и способ ее оценивания подходящими)
7. Использование результатов оценки (если модель признана подходящей)

   а) проверка гипотез

   б) создание прогнозов

   в) policy implication (использование модели для экономико-политических предложений)

Пример

1)Утверждение экономической теории

Доходность конкретной акции зависит от состояния финансового рынка в целом

2) Построение соответствующей математической модели

Модель CAPM для России :

$Y = \beta_0 + \beta_1 X$, X – изменение индекса РТС, Y- доходность акции.

3) Построение соответствующей статистической (или эконометрической) модели

$Y = \beta_0 + \beta_1 X + \varepsilon$, где ε – стохастическая составляющая (возмущение)

4) Получение данных

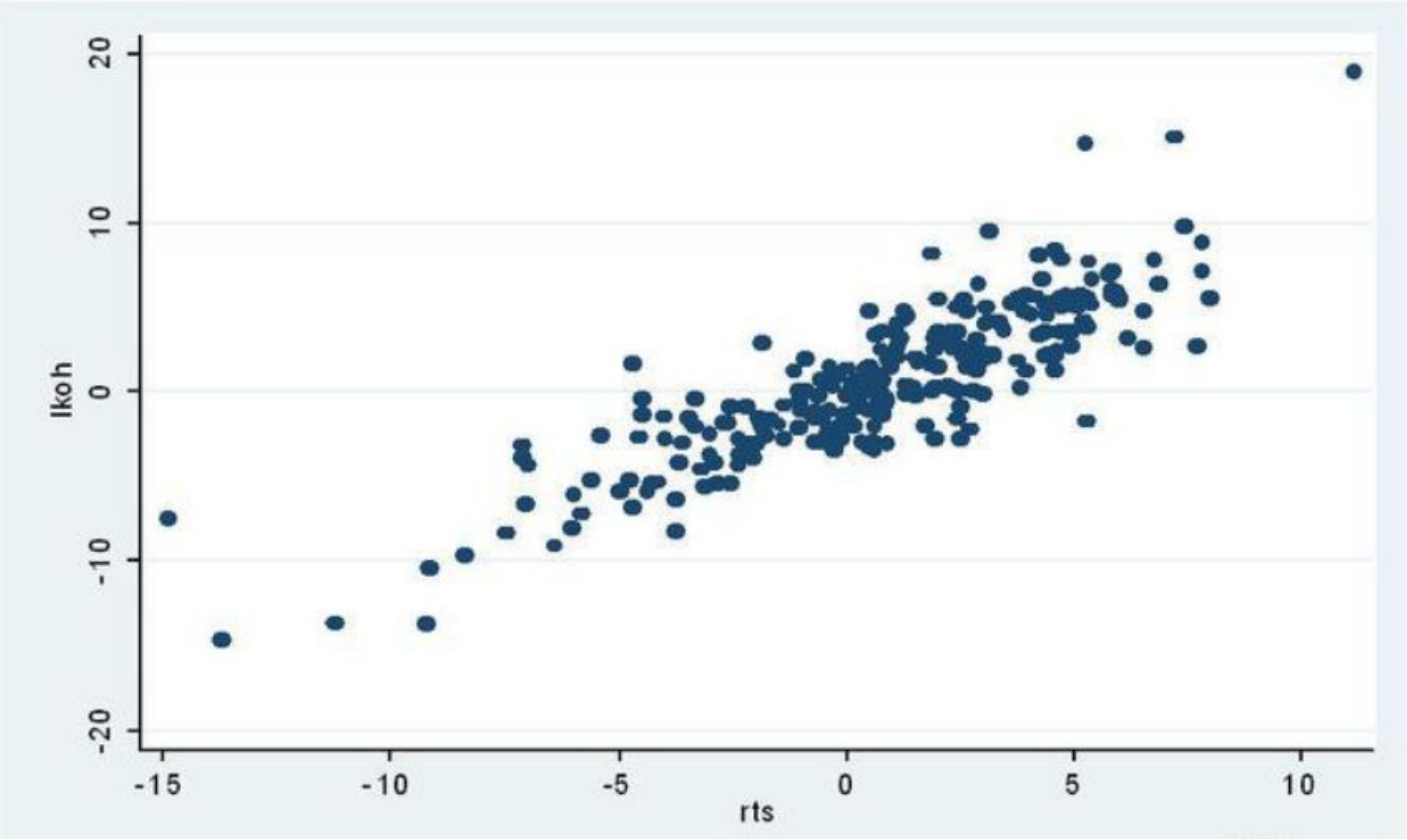
Данные

Y – доходность акций ЛУКойла (в %), X – изменение РТС (в %),

30.08.2002 – 24.08.2007

	LKOH	RTS
date	last price	last price
30.08.2002	15.800	332.900
06.09.2002	15.780	335.470
13.09.2002	15.620	337.110
20.09.2002	15.550	333.560
...	...	...
27.07.2007	80.000	1967.060
03.08.2007	79.300	1970.750
10.08.2007	74.300	1897.200
17.08.2007	73.200	1860.700
24.08.2007	73.500	1864.740

Диаграмма рассеяния



5) Оценка параметров эконометрической модели

Оценка параметров с помощью метода наименьших квадратов (МНК)

$$\hat{Y} = -0.0471 + 1.002X$$

6) Верификация модели

Проверка адекватности модели

7) Выбор другой модели или способа оценивания при отрицательном ответе

8) Проверка гипотез

Проверка гипотезы: 1.002 статистически больше 1

$H_0: \beta_{1} = 1$

$H_1: \beta_{1} > 1$ (акции ЛУКойла доходнее, чем рынок в среднем)

9) Создание прогнозов при положительном ответе

Прогнозы о доходности акций ЛУКойла

10) Использование модели для контроля (или регулирования)

Принятие инвестором решения о покупке акций ЛУКойла

Типы экономических данных

- **Перекрестные данные (пространственные данные, cross-sectional data)** – наблюдения для различных однородных объектов (индивиды, фирмы, регионы, страны) в фиксированный момент времени.

- **Временные ряды (time series)** – совокупность наблюдений одного и того же показателя в различные моменты времени (дни, недели, месяцы, годы). Например – цены акций одной компании, уровень безработицы в Приморском крае, обменный курс рубль/доллар.

- **Панельные данные** - множество наблюдений за некоторыми однородными объектами в различные моменты времени (например, данные о потреблении молока 200 индивидов за 5 лет).

Источники данных

1. Много Российских временных рядов на сайте Росстата www.gks.ru
2. RLMS (РМЭЗ – Российский мониторинг экономического положения и здоровья населения, см. сайт НИУ ВШЭ)
3. WVS (World Value Survey, http://www.worldvaluessurvey.org)
4. ESS (European Social Survey, http://www.europeansocialsurvey.org)
5. BEEPS (Business Environment & Enterprise Performance Surveys, http://web.worldbank.org/)

Основные понятия математической статистики

Генеральная совокупность - совокупность всех значений случайной величины

Выборочная совокупность (выборка) – статистически представленный ряд наблюдений из генеральной совокупности

Репрезентативная выборка – выборка, характеристики которой обладают свойствами генеральной совокупности

Основная задача математической статистики – оценивание характеристик генеральной совокупности по выборке.

Гипотезы – обо всей генеральной совокупности мы как правило, не имеем информации, поэтому можем строить лишь предположения (гипотезы) о ее параметрах

Проверка гипотез – для проверки гипотез используют независимую случайную выборку из генеральной совокупности и на основании выборки строят выборочные оценки неизвестных теоретических параметров

γ

Статистические оценки параметров распределения

Задача математической статистики: оценить количественный признак генеральной совокупности по данным выборки.

Оценкой числового параметра θ называется функция выборочных значений $\theta^*(x_1, x_2, \dots x_n)$

Например: оценкой математического ожидания (генеральной средней) для генеральной совокупности является выборочная средняя

Оценки бывают точечные (число) и интервальными (интервал)

Оценку θ^* можно рассматривать как случайную величину.

Почему?

Статистические оценки параметров распределения

Для того, чтобы чтобы оценки давали «хорошие» приближения оцениваемых параметров, они должны обладать свойствами:

1) Несмещенность. Оценка называется несмещенной, если математическое ожидание оценки равно истинному значению параметра (если математическое ожидание оценки не равно истинному значению параметра, то оценка называется смещенной).
2) Эффективность. Оценка называется эффективной, если при заданном объеме выборки n она имеет наименьшую возможную дисперсию.
3) Состоятельность. Оценка называется состоятельной, если она при $n \to \infty$ она стремится по вероятности к истинному значению параметра.

Точечные и интервальные оценки параметров

Точечной называют оценку, которая определяется одним числом. При выборках малого объема точечная оценка может значительно отличаться от оцениваемого параметра. Поэтому при небольшом объеме выборки пользуются интервальными оценками.

Интервальной называют оценку, которая определяется двумя числами – концами интервала. Интервальные оценки позволяют установить точность и надежность оценок.

Метод доверительных интервалов

Метод доверительных интервалов был разработан американским статистиком Ю.Нейманом.

Пусть по данным выборки найдена оценка θ^* неизвестного параметра θ генеральной совокупности. θ^* тем точнее определяет θ, чем меньше величина абсолютной разности $|\theta-\theta^*|$. Т.е. если найдется $\delta>0$ и $|\theta-\theta^*|<\delta$, то чем меньше δ, тем оценка точнее. Положительное число δ характеризует точность оценки.

Метод доверительных интервалов

Но мы не можем со 100% вероятностью утверждать, что оценка θ^* удовлетворяет неравенству $|\theta-\theta^*|<\delta$. Мы можем говорить лишь о вероятности γ, с которой это неравенство осуществляется, то есть

$$P\left(|\theta-\theta^*|<\delta\right)=\gamma .$$

Вероятность γ называют надежностью (доверительной вероятностью) оценки θ по θ^*. Обычно надежность оценки задается исследователем заранее, наиболее часто задают надежность равную 0,95; 0,9 или 0,99.

Метод доверительных интервалов

Заменим неравенство $\left|\theta - \theta^*\right| < \delta$ равносильным ему двойным неравенством $-\delta < \theta - \theta^* < \delta$, или $\theta^* - \delta < \theta < \theta^* + \delta$, получим

$$P\left(\theta^* - \delta < \theta < \theta^* + \delta\right) = \gamma.$$

Т.е. вероятность того, что интервал $\left(\theta^* - \delta; \theta^* + \delta\right)$ содержит в себе неизвестный параметр генеральной совокупности θ, равна γ.

Доверительным называют интервал $\left(\theta^* - \delta; \theta^* + \delta\right)$, который содержит неизвестный параметр генеральной совокупности с заданной надежностью γ.

Значение t для различных уровней значимости (двухсторонний критерий)

Число средней свободы *df*	$\alpha = 1 - \gamma$			Число средней свободы *df*	$\alpha = 1 - \gamma$		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	32,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Элементы теории корреляции

Для двух выборок $x_1, x_2, \dots x_n$ и $y_1, y_2, \dots y_n$ несмещенная оценка ковариации случайных величин X и Y имеет вид:

$$\operatorname{cov}(X,Y)_{\text{выб}} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right).$$

Эту величину называют выборочной ковариацией случайных величин X и Y.

Ковариация нужна для характеристики связи между величинами X и Y. Если ковариация = 0, то X и Y – независимые случайные величины, если ≠ 0, то X и Y – зависимые случайные величины

Недостаток!!! Ковариация зависит от выбранных единиц измерения!!!

Элементы теории корреляции

Для двух выборок x_1, x_2,…x_n и y_1, y_2, …y_n оценка коэффициента корреляции случайных величин X и Y имеет вид:

$$r^{выб}{}_{x,y} = \frac{\operatorname{cov}(X,Y)_{выб}}{\sqrt{s^2(X)s^2(Y)}}.$$

Эту величину называют выборочным коэффициентом корреляции случайных величин X и Y.

Здесь

$s^2(X)$ – исправленная выборочная дисперсия для выборки x_1, x_2,…x_n

$s^2(Y)$ – исправленная выборочная дисперсия для выборки y_1, y_2, …y_n

Элементы теории корреляции

Выборочный коэффициент корреляции - безразмерная величина!!!! Не зависит от величин измерения СВ X и Y!!!!

$$-1 \le r^{выб}{}_{x,y} \le 1$$

Если $r^{выб}_{x,y} > 0$ – связь между X и Y прямая;

Если $r^{выб}_{x,y} < 0$ – связь между X и Y обратная;

Если $\left|r^{выб}_{x,y}\right| > 0,7$ – связь между X и Y сильная;

Если $0,4 < \left|r^{выб}_{x,y}\right| < 0,7$ – связь между X и Y средняя;

Если $\left|r^{выб}_{x,y}\right| < 0,4$ – связь между X и Y слабая.