

# Эконометрика

Доктор экономических наук, профессор

**Коречков Юрий Викторович**

Развитие экономики, усложнение экономических процессов и повышение требований к принимаемым управленческим решениям в области макро и микроэкономики потребовало более тщательного и объективного анализа реально протекающих процессов на основе привлечения современных математических и статистических методов. С другой стороны, проблема нарушения предпосылок классических статистических методов при решении реальных экономических задач привели к необходимости развития и совершенствования классических методов математической статистики и уточнения постановок соответствующих задач. В результате этих процессов осуществилось выделение и формирование новой отрасли знания под названием **Эконометрика**, связанной с разработкой и применением методов количественной оценки экономических явлений и процессов и их взаимосвязей.

**Основным методом исследования в эконометрике является экономико-математическое моделирование.** Правильно построенная модель должна давать ответ на вопрос о количественной оценке величины изменения изучаемого явления или процесса в зависимости от изменений внешней среды. Например, как скажется увеличение или уменьшение уровня инвестиций на совокупном валовом продукте, какие дополнительные ресурсы понадобятся для запланированного увеличения выпуска продукции и т. п. Практическая значимость эконометрики определяется тем, что применение ее методов позволяет выявить реально существующие связи между явлениями, дать обоснованный прогноз развития явления в заданных условиях, проверить и численно оценить экономические последствия принимаемых управленческих решений.

Построение эконометрических моделей приходится осуществлять в условиях, когда нарушаются предпосылки классических статистических методов, и учитывать наличие таких явлений, как:

- мультиколлинеарность объясняющих переменных;
- закрытость механизма связи между переменными в изолированной регрессии;
- эффект гетероскедастичности, т. е. отсутствия нормального распределения остатков для регрессионной функции;
- автокорреляция остатков;
- ложная корреляция.

Основным инструментом математической статистики, используемым для построения эконометрических моделей, являются методы корреляционного и регрессионного анализа.

**Корреляционный анализ** ставит своей целью проверку наличия и значимости линейной зависимости между переменными без деления переменных на зависимые и объясняющие. Ответ на эти вопросы дается с помощью вычисления показателей (коэффициентов) корреляции.

**Регрессионный анализ** направлен на выражение изучаемой зависимости в виде аналитической формулы с предварительным выделением зависимых и объясняющих переменных.

Регрессионный анализ призван ответить на такие вопросы, как:

- какие переменные определяют поведение других величин и, следовательно, могут использоваться как объясняющие переменные?
- какова формула зависимости и каков экономический смысл ее коэффициентов?

Результатом проведения регрессионного анализа является построение, так называемого, уравнения регрессии.

После построения уравнения регрессии осуществляется проверка его статистического качества, включающая:

- проверку статистической значимости коэффициентов уравнения регрессии;
- проверку общего качества уравнения регрессии;
- проверку наличия свойств данных, предполагавшихся при оценивании уравнения регрессии.

Рассматривая **эконометрическое исследование** в целом, в нем можно выделить следующие **этапы**:

1. **Постановка проблемы**, т. е. определение цели и задач исследования, выделение зависимых ( $y_j$ ) и независимых ( $x_k$ ) *экономических переменных на основе* качественного анализа изучаемых взаимосвязей методами экономической теории.
2. **Сбор необходимых исходных данных.**
3. **Построение эконометрической модели и оценка** ее адекватности и степени соответствия исходным данным.
4. **Использование модели для целей анализа и прогнозирования** параметров исследуемого явления.
5. **Качественная и количественная интерпретация** полученных на основе модели **результатов.**
6. **Практическое использование результатов.**

В процессе экономической интерпретации результатов необходимо ответить на следующие вопросы:

- являются ли статистически значимыми объясняющие факторы, важные с теоретической точки зрения?
- соответствуют ли оценки параметров модели качественным представлениям?

Примером эконометрической модели может служить аналитическое выражение взаимосвязи показателей инфляции и безработицы, записанное без учета инфляционных ожиданий и с учетом последних (1.1) :

$$\pi = -\beta (u - u^*) \quad (1.1)$$

где  $\pi$  – фактический темп инфляции (в процентах),  $u$  – фактический и  $u^*$  – естественный уровни безработицы (в процентах),  $\beta$  – постоянный параметр. При проведении исследования определяется, какая из этих зависимостей лучше соответствует реальной взаимосвязи между уровнями инфляции и безработицы, а также оценивается значение величины естественного уровня безработицы.

# 1. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ ПЕРЕКРЁСТНЫХ ДАННЫХ

**Основная цель любого научного (в том числе и эконометрического) исследования** – количественное описание взаимосвязей между переменными, которые характеризуют изучаемый процесс. Например, полезно знать, как цена товара связана с объёмом его реализации. Знание формы взаимосвязи этих переменных позволяет прогнозировать сбыт товара при изменении в будущем его цены в ту или иную сторону.

Приступая к решению подобных задач, следует в первую очередь ответить на вопросы: **есть ли вообще связь между изучаемыми переменными, какова сила этой связи?**

Интуитивно ясно, что чем теснее связь между переменными, тем больше информации содержит одна переменная относительно другой, тем точнее можно спрогнозировать неизвестное значение одной переменной по заданной величине другой.

Измерение тесноты взаимосвязи переменных составляет содержание корреляционного анализа. Форму, или формулу, взаимосвязи переменных устанавливают методами регрессионного анализа. Обозначим некоторые принципиальные моменты корреляционного и регрессионного анализа на простом примере взаимосвязи двух переменных  $Y$  и  $X$ . В специальной литературе зависимую переменную  $Y$  называют **результатирующей**, **объясняемой**, а независимую переменную  $X$ , которая определяет изменение  $Y$ , называют **объясняющей**, **факторной переменной**.

В зависимости от тесноты взаимосвязи между переменными  $Y$  и  $X$  различают следующие варианты.

Полная, точная взаимосвязь, при которой каждому значению величины  $X$  соответствует строго определенное значение  $Y$ . Такую связь называют **функциональной**. Например, площадь круга функционально зависит от его радиуса. Функциональную связь изучают в классических естественных науках (физике, математике, механике и т. д.).



Другой предельный вариант – полное отсутствие связи между независимыми переменными.

Промежуточный вариант неполной связи характерен для всех реально наблюдаемых величин. В частности, неполные, неточные связи между экономическими переменными обусловлены целым рядом причин: **во-первых**, трудно выявить все факторы, влияющие на экономический показатель; **во-вторых**, это влияние может быть не прямым, а опосредованным, т. е. проявляться через цепочку других факторов; **в-третьих**, многие факторы в экономике носят случайный характер и т. д.

В силу названных причин экономические явления характеризуются случайными переменными, взаимосвязь которых всегда статистическая, т. е. неполная.

**Статистической** называют зависимость, при которой изменение одной из величин влечет изменение распределения другой.

В частном случае, когда при изменении одной из величин изменяется только среднее значение другой, а закон распределения сохраняется, статистическая зависимость называется **корреляционной**.

Приведем пример. Пусть  $Y$  – урожай зерна, а  $X$  – количество удобрений на единицу площади. С одинаковых по площади участков земли при равных количествах внесенных удобрений снимают разный по величине урожай, то есть связь  $Y$  и  $X$  не является полной (функциональной). Разница в значениях  $Y$  при одном и том же  $X$  объясняется влиянием других, в том числе и случайных факторов (осадки, температура воздуха и др.).

Вместе с тем, как показывает опыт, средний урожай является функцией от количества удобрений, т. е.  $Y$  связан с  $X$  корреляционной зависимостью.

При корреляционной связи каждому значению независимой переменной  $X$  соответствует не одно значение  $Y$ , а совокупность значений с некоторым средним по совокупности  $M(Y|x)$ .

Именно это среднее значение  $M(Y|x)$  меняется в зависимости от  $X$ .

Математически этот факт записывается в следующем виде:

$$f(x) = M(Y|x), \quad (1)$$

где  $M(Y|x)$  – условное математическое ожидание величины  $Y$ , соответствующее фиксированному значению  $x$  случайной величины  $X$ ;  $f(x)$  – некоторая функция.

Уравнение (1) называется **уравнением регрессии  $Y$  на  $X$ , функция  $f(x)$  – функцией регрессии, а ее график – линией регрессии.**

При рассмотрении зависимости двух переменных говорят о парной регрессии. Уравнение парной регрессии определяет, каким в среднем будет значение результативного признака  $Y$  при каждом фиксированном значении факторного признака  $X$ .

В зависимости от вида функции регрессии  $f(x)$  говорят о той или иной форме зависимости между изучаемыми величинами: линейной, квадратичной, экспоненциальной и т. д.

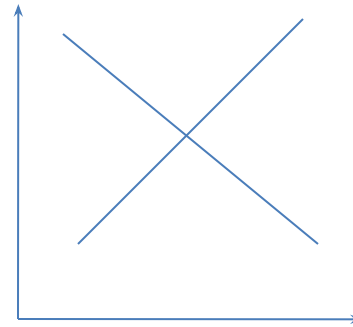
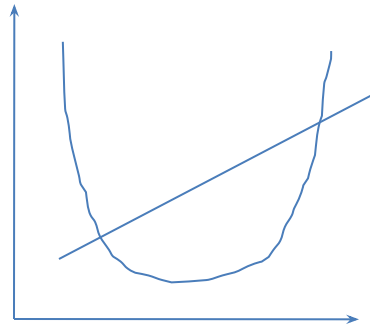
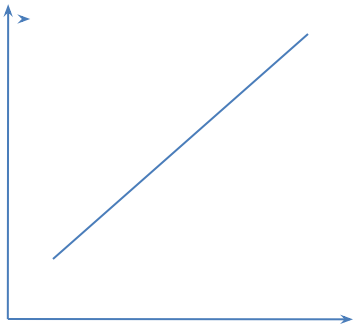
Таким образом, **процедура решения основной задачи эконометрики – описание взаимосвязи экономических переменных – состоит из двух последовательных этапов. Сначала в ходе корреляционного анализа исследуют наличие и силу связи между переменными. При положительном заключении (связь есть и она достаточно тесная) переходят ко второму этапу по определению формы связи между переменными средствами регрессионного анализа.**

## Парная корреляция

Пусть в результате наблюдений изучаемого экономического показателя  $Y$  удалось выявить один определяющий (доминирующий) фактор  $X$ . В этом случае говорят об изучении парной корреляционной связи.

Для изучения корреляции двух случайных величин  $X$  и  $Y$  необходимо заручиться эмпирическими данными следующего вида.

$X$	$x_1$	$x_2$	...	$x_n$
$Y$	$y_1$	$y_2$	...	...



При первичном анализе все пары наблюдений  $(x_i, y_i)$ ,  $i = 1, \dots, n$  изображают на координатной плоскости. В результате получают так называемое корреляционное поле (диаграмму рассеяния). Взаимное положение точек визуально характеризует форму взаимосвязи  $X$  и  $Y$  (рис. 1).

Прямая линия 1 (рис. 1, а) достаточно хорошо соответствует эмпирическим данным. Поэтому для описания формы зависимости между  $X$  и  $Y$  целесообразно выбрать линейную функцию

$$Y = \beta_0 + \beta_1 x$$

Форма взаимосвязи  $X$  и  $Y$  (рис. 1, б) близка к квадратичной

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2$$

(кривая 2) и никак не напоминает линейную (кривая 1).

Между  $X$  и  $Y$  отсутствует какая-либо взаимосвязь (рис. 1, в). В частности, прямые 1 и 2, проведенные через центр «облака» наблюдений, одинаково плохи для представления точек наблюдений.

Рассмотрим подробнее вариант, представленный на рис. 1, а. Пусть статистическая связь между переменными  $X$  и  $Y$  близка к линейной.

Мерой тесноты линейной связи двух случайных переменных является коэффициент линейной корреляции Пирсона  $r_{xy}$ , который для выборочных данных  $(x_i, y_i)$   $i = 1, \dots, n$  оценивают по формуле:

, (2)

где  $\bar{x}$ ,  $\bar{y}$  – выборочные средние;  $S_x$ ,  $S_y$  – стандартные ошибки.

В расчётах по формуле (2) довольно часто ошибаются, считая, что  $x^2 \neq x^2$ . Для того чтобы убедиться в обратном, рекомендуем расписать вручную несколько первых слагаемых при вычислении упомянутых выше средних величин.



Основные свойства выборочного коэффициента корреляции:

1. Коэффициент корреляции принимает значения на отрезке  $[-1; 1]$ , т. е.  $-1 \leq r_{xy} \leq 1$ .
2. Если между величинами  $X$  и  $Y$  существует положительная линейная корреляция, то  $r_{xy} > 0$ , если отрицательная, то  $r_{xy} < 0$ .
3. Если между величинами  $X$  и  $Y$  отсутствует линейная корреляционная связь, то  $r_{xy} = 0$ .
4. Чем ближе  $r_{xy}$  по модулю к 1, тем теснее линейная корреляционная связь между  $X$  и  $Y$ .
5. Если  $|r_{xy}| = 1$ , то величины  $X$  и  $Y$  связаны функциональной линейной зависимостью.

Коэффициент корреляции  $r_{xy}$  вычисляется по выборочным данным и может существенно отличаться от значения коэффициента корреляции  $\rho_{xy}$  в генеральной совокупности.

Например, ненулевое значение выборочного коэффициента корреляции  $r_{xy}$  может быть целиком обусловлено случайным колебанием выборки, на основе которой он вычислен. В такой ситуации  $r_{xy}$  может неверно представлять характер взаимосвязи  $X$  и  $Y$  в генеральной совокупности.

Таким образом, необходима проверка гипотезы о статистической значимости коэффициента корреляции  $\rho_{xy}$ .

Проверяется основная гипотеза  $H_0: \rho_{xy} = 0$  – между  $X$  и  $Y$  корреляционная связь отсутствует, т. е. выборочный коэффициент корреляции  $r_{xy}$  статистически не отличим от нуля.

$H_1: \rho_{xy} \neq 0$  – альтернативная гипотеза, соответствующая наличию линейной корреляционной зависимости между  $X$  и  $Y$ .

По выборке  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  объёма  $n$  вычисляется статистика:

, (3)  
имеющая при справедливости  $H_0$  распределение Стьюдента с числом степеней свободы  $\nu = n - 2$ .

# Парная регрессия

После того как между изучаемыми переменными установлена линейная корреляция достаточной тесноты, регрессионный анализ позволяет найти формулу связи  $X$  и  $Y$ .

Модель парной линейной регрессии предполагает, что случайные значения  $Y$  складываются из значения закономерной (неслучайной) составляющей  $M(Y|x)$  и значения чисто случайной добавки  $\epsilon$ .

Определить точный вид  $M(Y|x)$  довольно сложно, поэтому для нее подбирают подходящую аппроксимацию (приближение). В случае тесной линейной связи переменных таким приближением служит функция регрессии  $f(x) = \beta_0 + \beta_1 x$ .

Складывая закономерную и случайную составляющие для линейно связанных  $X$  и  $Y$ , получим

$$Y = M_x(Y) + \varepsilon = \beta_0 + \beta_1 X + \varepsilon \quad (4)$$

Соотношение (4) называют *моделью парной линейной регрессии*.

- **Линейная регрессия** ([англ. \*Linear regression\*](#)) — используемая в [статистике регрессионная модель](#) зависимости одной (объясняемой, зависимой) переменной  $\{ \displaystyle y \}$  от другой или нескольких других переменных (факторов, регрессоров, независимых переменных)  $\{ \displaystyle x \}$  с [линейной функцией](#) зависимости.
- Модель линейной регрессии является часто используемой и наиболее изученной в [эконометрике](#). А именно изучены свойства оценок параметров, получаемых различными методами при предположениях о вероятностных характеристиках факторов, и случайных ошибок модели. Предельные (асимптотические) свойства оценок нелинейных моделей также выводятся исходя из аппроксимации последних линейными моделями. Необходимо отметить, что с эконометрической точки зрения более важное значение имеет [линейность по параметрам](#), чем линейность по факторам модели.

- Регрессионная модель
- $\{ \displaystyle y=f(x,b)+\varepsilon, \sim E(\varepsilon)=0 \}$ , где  $\{ \displaystyle b \}$  — параметры модели,  $\{ \displaystyle \varepsilon \}$  — случайная ошибка модели; называется линейной регрессией, если функция регрессии  $\{ \displaystyle f(x,b) \}$  имеет вид
- $\{ \displaystyle f(x,b)=b_0+b_1x_1+b_2x_2+\dots+b_kx_k \}$ , где  $\{ \displaystyle b_j \}$  — параметры (коэффициенты) регрессии,  $\{ \displaystyle x_j \}$  — регрессоры (факторы модели),  $k$  — количество факторов модели.
- Коэффициенты линейной регрессии показывают скорость изменения зависимой переменной по данному фактору, при фиксированных остальных факторах (в линейной модели эта скорость постоянна):
- $\{ \displaystyle \forall j \sim b_j = \frac{\partial f}{\partial x_j} = \text{const} \}$

