

1. Эконометрика как наука: предмет, цели, задачи

эконометрический стохастический бинарный кластерный

Эконометрикой называется наука, позволяющая анализировать связи между различными экономическими показателями на основании реальных статистических данных с применением методов теории вероятностей и математической статистики. С помощью эконометрики выявляют новые, ранее неизвестные связи, уточняют или отвергают гипотезы о существовании определенных связей между экономическими показателями, предлагаемые экономической теорией.

Основная цель эконометрики заключается в модельном описании конкретных количественных взаимосвязей, обусловленных общими качественными закономерностями, выявленными в экономической теории.

Основной предмет исследования эконометрики - это массовые экономические явления и процессы. Предметы исследования эконометрики и статистики являются весьма схожими, потому что эконометрика исследует массовые экономические явления и процессы, а статистика исследует массовые явления и процессы любой природы (в том числе и экономические).

С помощью эконометрики решается очень широкий круг задач. Наиболее общими задачами эконометрики являются:

- 1) обнаружение и анализ статистических закономерностей в экономике;
- 2) построение на базе выявленных эмпирических экономических зависимостей эконометрических моделей.

Данные задачи делятся на более конкретные подзадачи, которые можно классифицировать по трём признакам:

- 1) классификация задач по конечным прикладным целям:
  - а) прогноз социально-экономических показателей, определяющих состояние и развитие изучаемой системы;
  - б) моделирование возможных вариантов социально-экономического развития системы для выявления факторов, изменение которых оказывает наиболее мощное влияние на состояние системы в целом;
- 2) классификация задач по уровню иерархии:
  - а) задачи, решаемые на макроуровне (страна в целом);
  - б) задачи, решаемые на мезоуровне (уровень отраслей, регионов);
  - в) задачи, решаемые на микроуровне (уровень фирмы, семьи, предприятия);
- 3) классификация задач по профилю изучаемой экономической системы:
  - а) рынок; б) инвестиционная, социальная, финансовая политика; в) ценообразование;
  - г) распределительные отношения; д) спрос и потребление; е) отдельно выделенный комплекс проблем.

## 2. Подготовка статистической базы эконометрического исследования

Первым этапом при проведении эконометрического исследования является сбор статистических данных об анализируемом объекте или процессе в виде конкретных значений эндогенных переменных и predetermined переменных, входящих в спецификацию модели. Данная информация необходима для определения оценок неизвестных коэффициентов, входящих в эконометрическую модель.

Сбором статистических данных называется процесс получения исходных данных об элементах исследуемой совокупности и их свойствах, которые в дальнейшем становятся предметом статистической обработки и анализа.

В связи с многообразием статистических наблюдений, их принято классифицировать по следующим признакам:

- 1) по форме организации;
- 2) по времени регистрации фактов;
- 3) по признаку полноты охвата элементов изучаемой совокупности.

По форме организации выделяют отчётность и специально организованные статистические наблюдения.

Отчётностью называется основная организационная форма статистического наблюдения, которая состоит в сборе сведений от предприятий, учреждений и организаций о различных сторонах их деятельности на специальных бланках, называемых отчётами. В зависимости от продолжительности периода, относительно которого составляется отчётность, выделяют основную и текущую отчётность.

Основной отчётностью называется организационная форма статистического наблюдения, которая содержит наиболее широкий круг показателей, характеризующих все стороны деятельности предприятия. Основная отчётность также называется годовой.

Текущей отчётностью называется организационная форма статистического наблюдения, которая представляется предприятиями в течение года за различные по продолжительности промежутки времени.

По той причине, что существуют данные, которые принципиально невозможно получить на основе отчётности и данные, которые нецелесообразно включать в неё, используются специально организованные статистические наблюдения - различного рода обследования и переписи.

Статистическим обследованием называется такая форма специально организованного статистического наблюдения, при котором исследуемая совокупность явлений подвергается наблюдению в течение определённого периода времени.

Переписью называется такая форма специально организованного статистического наблюдения, при котором исследуемая совокупность явлений наблюдается на какую-либо дату.

По признаку времени регистрации фактов в эконометрике различают текущее (непрерывное) и дискретное (прерывное) статистическое наблюдение.

Текущим (непрерывным) статистическим наблюдением называется наблюдение,

которое осуществляется во времени непрерывно. При этом отдельные явления, факты, события регистрируются по мере их возникновения.

Дискретным (прерывным) статистическим наблюдением называется наблюдение, при котором наблюдаемые явления, факты, события регистрируются через периоды времени, равной или неравной продолжительности. Дискретное наблюдение может быть периодическим и единовременным.

Периодическим наблюдением называется такая форма прерывного наблюдения, которая осуществляется через периоды времени равной продолжительности.

Единовременным наблюдением называется такая форма прерывного наблюдения, которое осуществляется через периоды времени неравной продолжительности или имеющие разовый характер.

В соответствии с признаком полноты охвата элементов изучаемой совокупности явлений, фактов, событий статистические наблюдения делятся на сплошные и несплошные наблюдения.

Сплошным наблюдением называется такая форма статистического наблюдения, при использовании которой учитываются все без исключения явления, факты, события, входящие в исследуемую совокупность.

Несплошным наблюдением называется такая форма статистического наблюдения, при использовании которой учитывается только некоторая часть явлений, фактов, событий, входящих в исследуемую совокупность.

Объективные причины использования несплошного наблюдения:

- 1) физическая невозможность или нецелесообразность осуществления сплошного наблюдения;
- 2) ограниченность исследователей во времени или средствах.

Выделяют несколько основных разновидностей несплошного наблюдения:

- 1) обследование основного массива характеризуется тем, что та часть исследуемой совокупности, которая подлежит наблюдению, устанавливается заранее. При этом отобранная часть единиц является преобладающей в объеме исследуемого объекта;
- 2) выборочное наблюдение характеризуется тем, что отбор той части единиц исследуемой совокупности, которая подлежит обследованию, производится строго в случайном порядке в соответствии с требованиями, установленными в теории вероятности;
- 3) анкетное наблюдение характеризуется тем, что лицам, от которых необходимо получить сведения, рассылают анкеты с просьбой заполнить их и вернуть обратно;
- 4) монографическое наблюдение характеризуется тем, что в составе исследуемой совокупности выделяются типические группы. В каждой подлежащей обследованию группе подвергают наблюдению одну (иногда две, три) типичную единицу. Установленные при наблюдении величины признаков рассматривают как типические (средние) величины для группы в целом. Программа наблюдения при монографическом наблюдении обычно бывает достаточно широкой, т. е. охватывает большое число признаков.

### 3. Критерии и принципы эконометрики

Эконометрические методы строятся на синтезе трех областей знаний: экономики, математики и статистики. Основой является экономическая модель, под которой понимается схематическое представление экономического явления или процесса с помощью научной абстракции, отражения только характерных черт. Наибольшее распространение в современной экономике получил метод анализа экономики “затраты - выпуск”. Это матричные (балансовые) модели, строящиеся по шахматной схеме и позволяющие в наиболее компактной форме представить взаимосвязь затрат и результатов производства. Таким образом, объектом эксперимента стали не только многократно воспроизводимые явления и процессы, но и системы и изменения в них, реально в практике трудно либо вообще неосуществимые.

Описание экономических систем математическими методами, или эконометрика, дает заключение о реальных объектах и связях по результатам выборочного обследования или моделирования. Вместе с тем, чтобы сделать вывод о том, какие из полученных результатов являются достоверными, а какие сомнительными или просто необоснованными, необходимо уметь оценивать их надежность и величину погрешности. Все перечисленные аспекты и составляют содержание эконометрики как науки.

В эконометрике, как и в любой научной дисциплине, познание развивается в соответствии с общим научным методом, предполагающим: 1) формулировку гипотезы с учетом соотношений между наблюдаемыми данными; 2) сбор статистических данных и представление гипотезы в сжатой или математической форме; 3) модификацию или улучшение гипотезы. Т.е., сердцевиной познания в экономике является эксперимент, предполагающий либо непосредственное наблюдение (измерение), либо математическое моделирование. Область применения эконометрических моделей и методов достаточно обширна. Это все сферы экономической теории и практики, где есть возможность сбора и обработки статистических данных, проведения наблюдений и экспериментов с целью учета воздействия случайных факторов, выявления качественных и количественных взаимосвязей между экономическими величинами и прогнозирования их поведения.

#### 4. Этапы эконометрического моделирования

Можно выделить шесть основных этапов эконометрического моделирования: постановочный, априорный, этап параметризации, информационный, этапы идентификации и верификации модели.

1-й этап (постановочный). Формируется цель исследования, набор участвующих в модели экономических переменных. В качестве цели эконометрического моделирования обычно рассматривают анализ исследуемого экономического объекта (процесса); прогноз его экономических показателей, имитацию развития объекта при различных значениях экзогенных переменных (отражая их случайный характер, изменение во времени), выработку управленческих решений. При выборе экономических переменных необходимо теоретическое обоснование каждой переменной (их число должно быть не очень большим и, в несколько раз меньше

числа наблюдений). Объясняющие переменные не должны быть связаны функциональной или тесной корреляционной зависимостью, так как это может привести к невозможности оценки параметров модели или к получению неустойчивых, не имеющих реального смысла оценок, т. е. к явлению мультиколлинеарности. При включении в модель тех или иных переменных является экономический (качественный) анализ исследуемого объекта. 2-й этап (априорный). Проводится анализ сущности изучаемого объекта, формирование и формализация априорной (известной до начала моделирования) информации.

3-й этап (параметризация). Осуществляется непосредственно моделирование, т.е. выбор общего вида модели, выявление входящих в нее связей. Основная задача -- выбор вида функции  $f(X)$  в эконометрической модели, в частности, возможность использования линейной модели как наиболее простой и надежной. Весьма важной проблемой на этом этапе эконометрического моделирования является проблема спецификации модели, в частности: выражение в математической форме обнаруженных связей и соотношений. От того, насколько удачно решена проблема спецификации модели, в значительной степени зависит успех всего эконометрического моделирования.

4-й этап (информационный). Осуществляется сбор необходимой статистической информации -- наблюдаемых значений экономических переменных: наблюдения, полученные как с участием исследователя, так и без его участия (в условиях активного или пассивного эксперимента).

5-й этап (идентификация модели). Осуществляется статистический анализ модели и оценка ее параметров. Также существует проблема идентифицируемости, т. е. проблема возможности получения однозначно определенных параметров модели, заданной системой одновременных уравнений (точнее, параметров структурной формы модели, раскрывающей механизм формирования значений эндогенных переменных, по параметрам приведенной формы модели, в которой эндогенные переменные непосредственно выражаются через predetermined переменные).

6-й этап (верификация модели). Проводится проверка истинности, адекватности модели. Выясняется, насколько удачно решены проблемы спецификации, идентификации и идентифицируемости модели, какова точность расчетов по данной модели, в конечном счете, насколько соответствует построенная модель моделируемому реальному экономическому объекту или процессу.

Разделение эконометрического моделирования на отдельные этапы носит условный характер, так как эти этапы могут пересекаться, взаимно дополнять друг друга и т. п.

## 5. Общее представление о детерминированных и стохастических процессах

Детерминированные процессы характеризуются тем, что знание их в некотором интервале времени позволяет полностью определить поведение этих процессов вне этого интервала. Для детерминированного процесса заранее задан критерий оптимальности, а ограничения первого и второго рода известны. Стохастические процессы характеризуются тем, что знание их на некотором интервале времени

позволяет определить лишь вероятностные характеристики поведения этих процессов вне этого интервала. Если эти вероятностные характеристики, например плотности распределения, заранее заданы, то и в этом случае можно определить в явной форме критерий оптимальности и ограничения, которые представляют собой, как это упоминалось выше, некоторые условные математические ожидания.

Детерминированные процессы можно рассматривать как частный случай стохастических процессов, плотность распределения которых представляет собой импульсную функцию Дирака, т. е.  $\delta(x)$ -функцию. При этом условные математические ожидания, входящие в критерий оптимальности и ограничения, превращаются просто в детерминированные функции, не зависящие от случайного вектора.

Стохастические процессы отличаются друг от друга и, в частности, от детерминированных процессов видом вероятностных характеристик -- плотностей распределения.

Объем априорной информации для детерминированных процессов обычно больше, чем для стохастических, поскольку для детерминированных процессов плотность распределения заранее известна, тогда как для стохастического процесса, как правило, ее еще нужно определить.

Однако если плотность распределения тем или иным способом предварительно определена и нам удалось записать функционал и уравнения ограничений в явной форме, то, несмотря на существенные идейные различия между детерминированными и стохастическими процессами, трудно установить сколь-нибудь заметные расхождения в формулировке и решении проблемы оптимальности для этих процессов.

Мы хорошо понимаем, что для детерминированных процессов часто можно получить оптимальность для каждого процесса в отдельности, как это имеет место, например, в оптимальных по быстродействию системах. В то же время для статистических процессов можно обеспечить лишь оптимальность в среднем, по это скорее относится к области идейных различий, нежели к формулировке и решению проблемы оптимальности.

## 6. Понятие эконометрических моделей, классификация и типы

Эконометрическая модель имеет следующий вид:

$$Y=f(X) + e$$

где  $Y$  - наблюдаемое значение переменной (объясняемая переменная);

$f(X)$  - объясненная часть, зависящая от значений объясняющих переменных;

$$X=\{x_1,x_2,\dots,x_n\}$$

$e$  - случайная составляющая (возмущения).

Выделяют три основных класса эконометрических моделей:

- 1) модель временных рядов;
- 2) модели регрессии с одним уравнением;
- 3) системы одновременных уравнений.

Моделью временных рядов называется зависимость результативной переменной от переменной времени или переменных, относящихся к другим моментам времени.

К моделям временных рядов, характеризующих зависимость результативной переменной от времени, относятся:

- а) модель зависимости результативной переменной от трендовой компоненты или модель тренда;
- б) модель зависимости результативной переменной от сезонной компоненты или модель сезонности;
- в) модель зависимости результативной переменной от трендовой и сезонной компонент или модель тренда и сезонности.

К моделям временных рядов, характеризующих зависимость результативной переменной от переменных, датированных другими моментами времени, относятся:

- а) модели с распределённым лагом, объясняющие вариацию результативной переменной в зависимости от предыдущих значений факторных переменных;
- б) модели авторегрессии, объясняющие вариацию результативной переменной в зависимости от предыдущих значений результативных переменных;
- в) модели ожидания, объясняющие вариацию результативной переменной в зависимости от будущих значений факторных или результативных переменных.

Кроме рассмотренной классификации, модели временных рядов делятся на модели, построенные по стационарным и нестационарным временным рядам.

Стационарным временным рядом называется временной ряд, который характеризуется постоянными во времени средней, дисперсией и автокорреляцией, т. е. данный временной ряд не содержит трендовой и сезонной компонент.

Нестационарным временным рядом называется временной ряд, который содержит трендовую и сезонную компоненты.

Моделью регрессии с одним уравнением называется зависимость результативной переменной, обозначаемой как  $y$ , от факторных (независимых) переменных, обозначаемых как  $x_1, x_2, \dots, x_n$ . Данную зависимость можно представить в виде функции регрессии или модели регрессии:

$$y = f(x, v) = f(x_1, x_2, \dots, x_n, v_1 \dots v_k)$$

где  $v_1 \dots v_k$  - параметры модели регрессии.

Можно выделить две основных классификации моделей регрессии::

- а) классификация моделей регрессии на парные и множественные регрессии в зависимости от числа факторных переменных;
- б) классификация моделей регрессии на линейные и нелинейные регрессии в зависимости от вида функции  $f(x, v)$ .

Системой одновременных уравнений называется модель, которая описывается системами взаимозависимых регрессионных уравнений.

Системы одновременных уравнений могут включать в себя тождества и регрессионные уравнения, в каждое из которых могут входить не только факторные переменные, но и результативные переменные из других уравнений системы.

Регрессионные уравнения, входящие в систему одновременных уравнений, называются поведенческими уравнениями. В поведенческих уравнениях значения параметров являются неизвестными и подлежат оцениванию.

Основное отличие тождеств от регрессионных уравнений заключается в том, что их

вид и значения параметров известны заранее.

Общая классификация эконометрических или экономико-математических моделей включает более десяти основных признаков.

Рассмотрим несколько ключевых классификаций эконометрических моделей:

1) классификация эконометрических моделей по целевому назначению:

- а) теоретико-аналитические модели, которые используются при исследовании общих свойств и закономерностей экономических процессов;
- б) прикладные модели, которые используются при решении конкретных экономических задач (модели экономического анализа, прогнозирования, управления);

Также эконометрические модели могут быть использованы при исследовании различных сторон народного хозяйства и его отдельных частей.

2) классификация эконометрических моделей по исследуемым экономическим процессам и содержательной проблематике. При этом выделяются:

- а) модели народного хозяйства в целом и его отдельных подсистем-отраслей, регионов и т. д.;
- б) комплексы моделей производства и потребления;
- в) комплексы моделей формирования и распределения доходов;
- г) комплексы моделей трудовых ресурсов;
- д) комплексы моделей ценообразования;
- е) комплексы моделей финансовых связей и др.

3) классификация эконометрических моделей на дескриптивные и нормативные модели:

а) дескриптивные модели предназначены для объяснения наблюдаемых фактов или для построения вероятностного прогноза.

б) нормативные модели отвечают на вопрос «как это должно быть», т. е. предполагают целенаправленную деятельность;

4) классификация эконометрических моделей по характеру отражения причинно-следственных связей. При этом выделяют:

- а) модели жестко детерминистские;
- б) модели, в которых учитываются факторы случайности и неопределенности.

5) Классификация эконометрических моделей по способам отражения фактора времени. При этом выделяют:

- а) статические модели, характеризующие исследуемую зависимость между переменными на определённый момент времени;
- б) динамические модели, характеризующие изменение экономических процессов во времени.

7. Последовательность построения эконометрического моделирования

Выделяют семь основных этапов эконометрического моделирования:

1) постановочный этап, в процессе осуществления которого определяются конечные цели и задачи исследования, а также совокупность включённых в модель факторных и результативных экономических переменных. При этом включение в



эконометрическую модель той или иной переменной должно быть теоретически обоснованно и не должно быть слишком большим. Между факторными переменными не должно быть функциональной или тесной корреляционной связи, потому что это приводит к наличию в модели мультиколлинеарности и негативно сказывается на результатах всего процесса моделирования;

2) априорный этап, в процессе осуществления которого проводится теоретический анализ сущности исследуемого процесса, а также формирование и формализация известной до начала моделирования (априорной) информации и исходных допущений, касающихся в частности природы исходных статистических данных и случайных остаточных составляющих в виде ряда гипотез;

3) этап параметризации (моделирования), в процессе осуществления которого выбирается общий вид модели и определяется состав и формы входящих в неё связей, т. е. происходит непосредственно моделирование.

К основным задачам этапа параметризации относятся:

а) выбор наиболее оптимальной функции зависимости результативной переменной от факторных переменных. При возникновении ситуации выбора между нелинейной и линейной функциями зависимости, предпочтение всегда отдаётся линейной функции, как наиболее простой и надёжной;

б) задача спецификации модели, в которую входят такие подзадачи, как аппроксимация математической формой выявленных связей и соотношений между переменными, определение результативных и факторных переменных, формулировка исходных предпосылок и ограничений модели.

4) информационный этап, в процессе осуществления которого происходит сбор необходимых статистических данных, а также анализируется качество собранной информации;

5) этап идентификации модели, в ходе осуществления которого происходит статистический анализ модели и оценивание неизвестных параметров. Данный этап непосредственно связан с проблемой идентифицируемости модели, т. е. ответа на вопрос «Возможно ли восстановить значения неизвестных параметров модели по имеющимся исходным данным в соответствии с решением, принятым на этапе параметризации». После положительного ответа на этот вопрос решается проблема идентификации модели, т. е. реализуется математически корректная процедура оценивания неизвестных параметров модели по имеющимся исходным данным;

6) этап оценки качества модели, в ходе осуществления которого проверяется достоверность и адекватность модели, т. е. определяется, насколько успешно решены задачи спецификации и идентификации модели, какова точность расчётов, полученных на её основе. Построенная модель должна быть адекватна реальному экономическому процессу. Если качество модели является неудовлетворительным, то происходит возврат ко второму этапу моделирования;

7) этап интерпретации результатов моделирования.

К наиболее распространённым эконометрическим моделям относятся:

1) модели потребительского и сберегательного потребления;

2) модели взаимосвязи риска и доходности ценных бумаг;

- 3) модели предложения труда;
- 4) макроэкономические модели (модель роста);
- 5) модели инвестиций;
- 6) маркетинговые модели;
- 7) модели валютных курсов и валютных кризисов и др.

Эконометрическое исследование связано с решением следующих проблем:

- 1) качественный анализ связей экономических переменных, т. е. определение зависимых ( $y_i$ ) и независимых ( $x_i$ ) переменных;
  - 2) изучение соответствующего раздела экономической теории;
  - 3) подбор данных;
  - 4) спецификация формы связи между  $y_i$  и  $x_i$ ;
  - 5) оценка неизвестных параметров модели;
  - 6) проверка ряда гипотез о свойствах распределения вероятностей для случайной компоненты (гипотезы о средней дисперсии и ковариации);
  - 7) анализ мультиколлинеарности объясняющих переменных, оценка ее статистической значимости, определение переменных, ответственных за мультиколлинеарность;
  - 8) введение фиктивных переменных;
  - 9) выявление автокорреляции;
  - 10) выявление тренда, циклической и случайной компонент;
  - 11) проверка остатков модели на гетероскедастичность;
  - 12) анализ структуры связей и построения системы одновременных уравнений;
  - 13) проверка условия идентификации;
  - 14) оценка параметров системы одновременных уравнений;
  - 15) проблемы моделирования на основе системы временных рядов;
  - 16) построение рекурсивных моделей, авторегрессионных моделей;
  - 17) выработка управленческих решений
  - 18) прогноз экономических показателей, характеризующих изучаемый процесс;
  - 19) моделирование поведения процесса при различных значениях независимых (факторных) переменных.
8. Цели и задачи спецификации эконометрических моделей

Любое эконометрическое исследование начинается со спецификации модели, т.е. с формулировки вида модели, исходя из соответствующей теории связи между переменными. Иными словами, исследование начинается с теории, устанавливающей связь между явлениями.

Прежде всего из всего круга факторов, влияющих на результативный признак, необходимо выделить наиболее существенно влияющие факторы.

Парная регрессия достаточна, если имеется доминирующий фактор, который и используется в качестве объясняющей переменной. Уравнение простой регрессии характеризует связь между двумя переменными, которая проявляется как некоторая закономерность лишь в среднем в целом по совокупности наблюдений.

Случайная величина называется также возмущением. Она включает влияние не

учтенных в модели факторов, случайных ошибок и особенностей измерения. Ее присутствие в модели порождено тремя источниками: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных. От правильно выбранной спецификации модели во многом зависит величина случайных ошибок: они тем меньше, чем в большей мере теоретические значения результативного признака ( $\hat{y}$ ) соответствует фактическому ( $y$ ).

К ошибкам спецификации будут относиться не только неправильный выбор той или иной математической функции для  $\hat{y}$ , но и недоучет в уравнении регрессии какого-либо существенного фактора, т. е. использование парной регрессии вместо множественной.

Наряду с ошибками спецификации могут иметь место ошибки выборки, поскольку исследователь чаще всего имеет дело с выборочными данными при установлении закономерной связи между признаками. Ошибки выборки имеют место и в силу неоднородности данных в исходной статистической совокупности, что, как правило, бывает при изучении экономических процессов. Если совокупность неоднородна, то уравнение регрессии не имеет практического смысла. Для получения хорошего результата обычно исключают из совокупности единицы с аномальными значениями исследуемых признаков. И в этом случае результаты регрессии представляют собой выборочные характеристики.

В парной регрессии выбор вида математической функции может быть осуществлен тремя методами:

- графическим;
- аналитическим, т. е. исходя из теории изучаемой взаимосвязи;
- экспериментальным.

При изучении зависимости между двумя признаками графический метод подбора вида уравнения регрессии достаточно нагляден. Он основан на поле корреляции.

## 9. Методы отбора факторов эконометрических моделей

“Оптимальный” состав факторов, включаемых в эконометрическую модель, является одним из основных условий ее “хорошего” качества, понимаемого и как соответствие формы модели теоретической концепции, выражающей содержание взаимосвязей между рассматриваемыми переменными, и как точность предсказания на рассматриваемом интервале времени  $t=1, 2, \dots, T$  наблюдаемых значений переменной  $y_t$  уравнением  $f(\hat{y}_t, x_t)$ .

Проблема выбора “оптимальных” факторов обычно решается на основе содержательного и количественного (статистического) анализа тенденций рассматриваемых процессов.

На этапе содержательного анализа решается вопрос о целесообразности включения в модель тех или иных факторов, исходя из “здорового” смысла. В макроэкономических исследованиях состав факторов, как правило, определяется на основании допущений экономической теории.

Здесь следует иметь в виду, что на этапе содержательного анализа обычно решается проблема установления самого факта наличия взаимосвязей между явлениями.

Однако, каждое из явлений может быть выражено разными факторами и даже их комбинациями. Поэтому в ряде исследований на основании содержательного анализа однозначно состав независимых переменных модели определить практически невозможно. Могут существовать их альтернативные наборы. Несложно заметить, что факторы, выражающие одну и ту же причину, могут быть тесно взаимосвязаны между собой.

Вследствие этого, одновременное включение таких факторов в модель вряд ли целесообразно, поскольку таким образом одна и та же причина будет учтена дважды. В результате в общем случае на этапе обоснования эконометрической модели исследователи могут столкнуться с проблемой выбора наиболее предпочтительного состава независимых факторов среди ряда альтернативных вариантов.

Можно выделить два основных подхода к решению этой проблемы. Первый из них предполагает априорное (до построения модели) исследование характера и силы взаимосвязей между рассматриваемыми переменными, по результатам которого в модель включаются факторы, наиболее значимые по своему “непосредственному” влиянию на зависимую переменную  $Y_t$ . И, наоборот, из модели исключаются факторы, которые, либо малозначимы с точки зрения силы своего влияния на переменную  $Y_t$ , либо их сильное влияние на нее можно трактовать как индуцированное взаимосвязями с другими экзогенными переменными.

Второй подход к отбору независимых факторов можно назвать апостериорным. Он предполагает первоначально включить в модель все отобранные на этапе содержательного анализа факторы. Уточнение их состава в этом случае производится на основе анализа характеристик качества построенной модели, одной из групп которых являются и показатели, выражающие силу влияния каждого из факторов на зависимую переменную  $Y_t$ .

#### 10. Априорные и апостериорные подходы к отбору факторов

В основе “априорного” подхода лежат следующие предположения.

1. Сильное влияние фактора на зависимую переменную должно подтверждаться и определенными количественными характеристиками, важнейшей из которых является их парный линейный коэффициент корреляции, выборочное значение которого рассчитывается на основании имеющейся информации по формуле:  
где  $\bar{x}$  - средние значения соответствующих переменных,  $\sigma_x$  - их среднеквадратические отклонения.

2. Если два и более факторов выражают одно и то же явление (см. рассмотренные выше примеры), то, как правило, между ними также должна существовать достаточно сильная взаимосвязь. На это может указать выборочное значение их парного коэффициента корреляции

Здесь следует еще раз подчеркнуть, что при таком отборе, основанном на эмпирике и интуиции, обычно не принимается во внимание точность оценки выборочных коэффициентов корреляции, которая растет с увеличением выборки, т. е. значения  $T$ . Значительно усложняет проблему отбора факторов явление ложной корреляции, которое характеризуется достаточно высокими по абсолютной величине значениями

коэффициентов парной корреляции у процессов, с содержательной точки зрения между собой никак не связанных.

При апостериорном подходе уточнение состава факторов эконометрической модели осуществляется на основе анализа значений ряда качественных характеристик уже построенного ее варианта. Одну из групп таких характеристик, являющихся наиболее важными при отборе факторов, образуют значения критерия Стьюдента, рассчитываемые для коэффициентов при каждом из факторов модели. С помощью этого критерия проверяется гипотеза о значимости влияния фактора на зависимую переменную  $y$ .

Здесь следует отметить, что окончательное решение о целесообразности оставления фактора или его удаления из модели принимается на основе анализа всего комплекса ее характеристик качества с учетом содержательной стороны проблемы взаимосвязей между зависимой и независимыми переменными.

с помощью критерия Стьюдента может быть проверена гипотеза о равенстве найденного выборочного среднего предполагаемому значению математического ожидания.

Таким образом, можно предложить следующую поэтапную процедуру построения окончательно варианта модели на основе апостериорного подхода:

1. В исходный вариант модели включаются все факторы, отобранные в ходе содержательного анализа проблемы. Для этого варианта рассчитываются значения оценок коэффициентов модели, их среднеквадратические ошибки и значения критериев Стьюдента (выражение (1.25)).
2. Из модели удаляют незначимый фактор, характеризующийся наименьшим значением  $t$  (при условии, что  $t < t^*$ ), и таким образом формируют новый вариант модели с уменьшенным на один числом факторов.
3. Процесс отбора факторов можно считать законченным, когда остающиеся в модели факторы являются значимыми, Если полученный вариант модели удовлетворяет и другим критериям ее качества, то процесс построения модели можно считать завершенным в целом.

“Априорный” путь отбора факторов не обладает достаточной обоснованностью. Вместе с тем использование априорного подхода часто позволяет уточнить некоторые предварительные альтернативные варианты наборов независимых факторов, проверить исходные предпосылки модели относительно правильности выбора формы взаимосвязей между ними.

“Апостериорный” подход к отбору факторов на первый взгляд является более предпочтительным как раз из-за того, что целесообразность включения каждого из факторов в эконометрическую модель определяется на основании всего комплекса взаимосвязей между вошедшими в модель переменными. Однако, когда общее количество факторов достаточно велико, то нет никаких гарантий, что множество несущественных, а то и ложных взаимосвязей между ними не будет превалировать над основными.

Согласно этим рекомендациям с помощью методов “априорного” отбора, используя при этом и содержательный анализ, формируются альтернативные варианты

включаемых в модель наборов факторов. Далее с помощью методов “апостериорного” отбора эти наборы уточняются и соответствующие им варианты моделей сопоставляются по ряду характеристик их качества. Предполагается, что лучший из вариантов модели содержит и “оптимальный” набор факторов.

## 11. Методы выбора формы уравнения регрессии

Как и в парной зависимости, возможны разные виды уравнений множественной регрессии: линейные и нелинейные.

### 1. линейная регрессия

### 2. нелинейные регрессии

- степенная регрессия

- показательная - ;

- экспоненциальная - .

- гиперболическая регрессия

Ввиду четкой интерпретации параметров наиболее широко используются линейная и степенная функции. В линейной множественной регрессии параметры при  $x$  называются коэффициентами “чистой регрессии”. Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

Стандартные компьютерные программы обработки регрессионного анализа позволяют перебирать различные функции и выбрать ту из них, для которой остаточная дисперсия и ошибка аппроксимации минимальны, а коэффициент детерминации максимален.

Если исследователя не устраивает предлагаемый стандартный набор функций регрессии, то можно использовать любые другие, приводимые путем соответствующих преобразований к линейному виду.

Однако чем сложнее функция, тем менее интерпретируемы ее параметры.

При сложных полиномиальных функциях с большим числом факторов необходимо помнить, что каждый параметр преобразованной функции является средней величиной, которая должна быть подсчитана по достаточному числу наблюдений.

Если число наблюдений невелико, что, как правило, имеет место в эконометрике, то увеличение числа параметров функции приведет к их статистической незначимости и соответственно потребует упрощения вида функции. Если один и тот же фактор вводится в регрессию в разных степенях, то каждая степень рассматривается как самостоятельный фактор.

В эконометрике регрессионные модели часто строятся на основе макроуровня экономических показателей, когда ставится задача оценки влияния наиболее экономически существенных факторов на моделируемый показатель при ограниченном объеме информации. Поэтому полиномиальные модели высоких порядков используются редко.

## 12. Метод наименьших квадратов

Метод наименьших квадратов (МНК) является одним из наиболее разработанных и

распространенных вследствие своей относительной простоты и эффективности методов оценки параметров линейных эконометрических моделей. Он не предъявляет жестких требований к закону распределения ошибок моделей. Вследствие этого оценки коэффициентов моделей, полученные на основе МНК, не зависят от фактического (или предполагаемого) закона распределения. Хотя обычно закон распределения ошибки, если его знание необходимо для проверки качества модели, свойств ее параметров и т. п., предполагается нормальным. При этом в "классическом" варианте МНК, как это будет показано далее, в отношении свойств ошибки модели  $t$  выдвигаются следующие предположения:

- ошибка имеет нулевое математическое ожидание,  $M[t]=0$ ;
- ее дисперсия конечна и постоянна,  $\sigma^2=\text{const}$ ;
- автокорреляционные связи в ряду ошибки отсутствуют, т. е.  $\rho_1=\rho_2=\dots=0$ , где  $\rho_i$  - коэффициент автокорреляции рядов  $t$  и  $t-i$ ,  $i=1,2,\dots$ ;
- ряд значений ошибки статистически не связан с рядами значений независимых переменных модели.

Рассмотренные предположения определяют ошибку модели как процесс белого шума с ковариационной матрицей ее вектора ошибки, имеющей следующий вид:  $\text{Cov}(t)=\sigma^2 E$ .

Рассмотрим общую схему процедуру оценки параметров линейной эконометрической модели на основе МНК более подробно. Такая модель в общем виде была представлена уравнением (1.2):

$$y_t = \theta_0 + \theta_1 x_{1t} + \dots + \theta_n x_{nt} + \epsilon_t$$

Исходными данными при оценке параметров  $\theta_0, \theta_1, \dots, \theta_n$  являются измеренные (наблюдаемые) значения зависимой переменной, которые можно представить в виде вектора-столбца,

Наблюдаемые значения независимых переменных объединим в матрицу следующего вида:

$$X =$$

Свое название МНК получил, исходя из смыслового содержания критерия, которому должны удовлетворять полученные на его основе оценки параметров эконометрической модели:

сумма квадратов значений фактической ошибки модели должна быть минимальной. Иными словами, найденные с помощью МНК оценки  $\theta_0, \theta_1, \dots, \theta_n$ , обеспечивают минимум следующей квадратичной формы на множестве всех других комбинаций значений таких оценок:

где  $\epsilon_t$  - значение фактической ошибки модели в момент  $t=1,2,\dots, T$ , полученное после подстановки в выражение (1.2) вместо неизвестных истинных значений параметров  $\theta_0, \theta_1, \dots, \theta_n$  их оценок  $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_n$ .

Векторно-матричная форма записи линейной эконометрической модели (1.2) имеет следующий вид:

$$y = X\theta, \quad (2.4)$$

где  $y$  - вектор-столбец, состоящий из  $T$  компонент;  $X$  - матрица размера  $T \times (n+1)$  (если в

модели присутствует "свободный" коэффициент 0);  $\beta = (\beta_0, \beta_1, \dots, \beta_n)$  - вектор-столбец параметров, состоящий из  $n+1$ -й компоненты;  $\epsilon$  - вектор-столбец ошибки модели, состоящий, как и вектор  $y$ , из  $T$  компонент.

### 13. Классификация регрессионных моделей

Регрессионная модель есть прежде всего гипотеза, которая должна быть подвергнута статистической проверке, после чего она принимается или отвергается. Регрессионная модель -- это параметрическое семейство функций, задающее отображение

где  $\beta$  -- пространство параметров,  $X$  -- пространство свободных переменных,  $Y$  -- пространство зависимых переменных.

Модель является настроенной (обученной) когда зафиксированы её параметры, то есть модель задаёт отображение для фиксированного значения  $\beta$ .

Различают математическую модель и регрессионную модель.

Регрессионная модель объединяет широкий класс универсальных функций, которые описывают некоторую закономерность. При этом для построения модели в основном используются измеряемые данные, а не знание свойств исследуемой закономерности. Такая модель часто неинтерпретируема, но более точна. Это объясняется либо большим числом моделей-претендентов, которые используются для построения оптимальной модели, либо большой сложностью модели.

Нахождение параметров регрессионной модели называется обучением модели.

Примеры регрессионных моделей: линейные функции, алгебраические полиномы, ряды Чебышёва, нейронные сети без обратной связи, например, однослойный перцептрон Розенблатта, радиальные базисные функции и прочее.

### 14. Предпосылки метода наименьших квадратов

Известно, что для получения по МНК наилучших результатов требуется, чтобы выполнялся ряд предпосылок относительно случайного отклонения. (условия Гаусса-Маркова)

1. Математическое ожидание случайного отклонения  $e_i$  равно нулю:  $M(e_i) = 0$  для всех наблюдений. Данное условие означает, что случайное отклонение в среднем не оказывает влияния на зависимую переменную. В определенном наблюдении случайный член может быть положительным или отрицательным, но он не должен иметь систематического смещения. Выполнимость  $M(e_i) = 0$  влечет выполнимость:

2. Дисперсия случайных отклонений  $\epsilon_i$  постоянна:  $D(e_i) = D(e_j) = \sigma^2 = \text{const}$  для любых наблюдений  $i$  и  $j$ . Условие независимости дисперсии ошибки от номера наблюдения называется гомоскедастичностью. Невыполнимость этой предпосылки называется гетероскедастичностью. Поскольку  $D(e) = M((e_j - M e_j)^2) = M(e^2)$ , то эту предпосылку можно переписать в форме:  $M(e^2) = \sigma^2$ . Причины невыполнимости данной предпосылки и проблемы, связанные с этим, подробно рассматриваются ниже.

3. Случайные отклонения  $e_i$  и  $e_j$  являются независимыми друг от друга для  $i \neq j$ .



Выполнимость этой предпосылки предполагает, что отсутствует систематическая связь между любыми случайными отклонениями. Величина и определенный знак любого случайного отклонения не должны быть причинами величины и знака любого другого отклонения. Выполнимость данной предпосылки влечет следующее соотношение:

Если данное условие выполняется, то можно говорить об отсутствии автокорреляции. С учетом выполнимости предпосылки 1 данное соотношение можно переписать в виде:

4. Случайное отклонение должно быть независимо от объясняющих переменных. Обычно это условие выполняется автоматически, если объясняющие переменные не являются случайными в модели. Данное условие предполагает выполнимость следующего соотношения:

5. Модель является линейной относительно параметров. Для случая множественной линейной регрессии существенными являются еще две предпосылки.

6. Отсутствие мультиколлинеарности. Между объясняющими переменными отсутствует сильная линейная зависимость.

7. Случайные отклонения  $e_i$ ,  $i = 1, 2, \dots, n$ , имеют нормальное распределение. Выполнимость данной предпосылки важна для проверки статистических гипотез и построения интервальных оценок.

Наряду с выполнимостью указанных предпосылок при построении классических линейных регрессионных моделей делаются еще некоторые предположения.

Например:

- объясняющие переменные не являются случайными величинами;
- число наблюдений намного больше числа объясняющих переменных (числа факторов уравнения);
- отсутствуют ошибки спецификации, т. е. правильно выбран вид уравнения и в него включены все необходимые переменные.

Зачастую полагают, что число наблюдений должно быть как минимум в 5-6 раз больше числа параметров уравнения (числа объясняющих переменных).

15. Несмещённость, эффективность и состоятельность оценок параметров регрессии

Оценки параметров регрессии должны отвечать определенным критериям. Они должны быть несмещенными, состоятельными и эффективными. Эти свойства оценок, полученных по МНК, имеют чрезвычайно важное практическое значение в использовании результатов регрессии и корреляции.

Несмещенность оценки означает, что математическое ожидание остатков равно нулю. Если оценки обладают свойством несмещенности, то их можно сравнивать по разным исследованиям.

Оценки считаются эффективными, если они характеризуются наименьшей дисперсией. В практических исследованиях это означает возможность перехода от точечного оценивания к интервальному.

Состоятельность оценок характеризует увеличение их точности с увеличением объема выборки. Большой практический интерес представляют те результаты

регрессии, для которых доверительный интервал ожидаемого значения параметра регрессии  $b_i$  имеет предел значений вероятности, равный единице. Иными словами, вероятность получения оценки на заданном расстоянии от истинного значения параметра близка к единице.

Указанные критерии оценок (несмещенность, состоятельность и эффективность) обязательно учитываются при разных способах оценивания.

#### 16. Гомоскедастичность и гетероскедастичность остатков

Дисперсия - среднее арифметическое из квадратов отклонений наблюдаемых значений ( $x_1, x_2, \dots, x_n$ ) случайной величины от их среднего арифметического.

Гомоскедастичность остатков означает, что для каждого значения фактора  $x_j$  остатки имеют одинаковую дисперсию. Если это условие применения МНК не соблюдается, то имеет место гетероскедастичность. Наличие гетероскедастичности можно наглядно видеть из поля корреляции.

Наличие гомоскедастичности или гетероскедастичности можно видеть и по графику зависимости остатков от теоретических значений результативного признака.

Гетероскедастичность будет сказываться на уменьшении эффективности оценок. В частности, становится затруднительным использование формулы стандартной ошибки коэффициента регрессии предполагающей единую дисперсию остатков для любых значений фактора.

При малом объеме выборки для оценки нарушения гомоскедастичности можно использовать метод Гольдфельда-Квандта, который включает:

1. Упорядочение наблюдений  $n$  по мере возрастания переменной  $x$ .
2. Исключения из рассмотрения центральных наблюдений  $C$ ; при этом где  $p$  - число оцениваемых параметров.
3. Разделение совокупности на две группы (с малыми и большими значениями фактора  $x$ ) и определение по каждой из групп уравнений регрессии.
4. Определение остаточной суммы квадратов для обеих групп и нахождение их отношения: При выполнении нулевой гипотезы о гомоскедастичности отношение  $R$  будет соответствовать  $F$ -критерию с степенями свободы для каждой остаточной суммы квадратов. Чем больше величина  $R$  превышает табл. значение  $F$ -критерия, тем больше нарушена предпосылка о равенстве дисперсий остаточных величин.

#### 17. Тестирование моделей на гетероскедастичность (тест Гольдфельда-Квандта)

Для проверки на гетероскедастичность существует большое число тестов. Мы остановимся на тесте Гольдфельда-Квандта.

Тест Гольдфельда-Квандта применяется в том случае, когда имеются предположения:

1. о прямой зависимости дисперсии  $u_t$ , ошибки регрессии  $e_t$  от величины некоторой независимой переменной  $X$  в наблюдении  $t$ ;
2. случайный член  $e_t$ , распределен нормально и не подвержен автокорреляции.

Алгоритм теста:

1. Упорядочивание  $n$  данных в выборке по величине независимой переменной, относительно которой есть подозрение на гетероскедастичность.

2. Исключение с средних наблюдений в этом упорядочении в целях построения двух независимых "частных" регрессий по данным  $n' = (n-c)/2$  в начале выборки и по данным  $n' = (n - c)/2$  в конце выборки
3. Проведение двух независимых "частных" регрессий - первых  $n'$  и последних  $n'$  наблюдений и построение соответствующих остатков  $e_1$  и  $e_2$ ;...