Эконометрика

Обратная и пропорциональная модель парной линейной регрессии.

Фиктивная линейная зависимость.

«Обратная» модель парной линейной связи

Задача состоит в оценивании модели прямолинейной связи между некоторыми переменными x и y на основе наблюдений *n* пар значений этих переменных. Вопрос об оценивании параметров модели наблюдений уже рассмотрен ранее. Встает вопрос о том, что изменится, если исходить из «обратной» модели:

Пусть – оценки параметров α и β в модели наблюдений , а – оценки параметров в модели наблюдений

. Тогда:

т.е.

или

Наилучшая прямая

По первой модели наблюдений мы получаем наилучшую прямую

а по второй – прямую

Первую прямую мы можем записать в виде

Сравнивая коэффициенты при x в двух последних уравнениях, находим, что эти коэффициенты равны в том и только в том случае, когда выполнено соотношение

т. е. или, с учетом предыдущего, когда .

Что касается отрезков на осях, то они будут совпадать тогда и только тогда, когда

или

Но так что

Наилучшая прямая

При получаем В то же время,

При совпадают и отрезки на осях, т. е. наилучшая прямая одна для обоих моделей, это прямая, на которой расположены все наблюдаемые точки

Наилучшие прямые, построенные по двум альтернативным моделям, совпадают в том и только в том случае, когда все точки , расположены на одной прямой (так что); при этом, **.** В противном случае , и подобранные «наилучшие» прямые имеют разные угловые коэффициенты.

Пропорциональная связь между переменными

Пусть получены наблюдения , такие, что гипотетическая линейная связь между переменными и имеет вид (**пропорциональная связь между переменными**), и ей соответствует модель наблюдений:

Применение метода наименьших квадратов в этой ситуации сводится к минимизации суммы квадратов

по всем возможным значениям β. Последняя сумма квадратов является функцией единственной переменной β (при известных значениях), и точка минимума этой функции легко находится. Для этого мы приравниваем нулю производную по β:

(нормальное уравнение)

Отсюда получаем,

Оценка единственного параметра пропорциональной модели будет:

И, следовательно, а точка не лежит на

полученной прямой . Более того, в такой ситуации

где

Коэффициент детерминации

выражение **не имеет смысла**. И можно

воспользоваться формулой . Однако, такой

подход к определению коэффициента детерминации не решает проблемы, поскольку, в принципе, при оценивании модели без постоянной составляющей **возможны ситуации, при которых возникают отрицательные значения** .

Преодолеть возникающие затруднения можно, если определить в модели наблюдений без постоянной составляющей формулой:

Здесь используется

сумма квадратов, не центрированных значений переменной *y* (отклонений значений переменной *y* от «нулевого уровня»). При таком определении, неотрицательность коэффициента гарантируется соотношением:

которое отражает геометрическую сущность метода наименьших квадратов (аналог знаменитой теоремы Пифагора для многомерного пространства) и выполняется как для модели без постоянной составляющей, так и для модели с наличием постоянной составляющей в правой части модели наблюдений. Деля обе части последнего

равенства на приходим к соотношению:

Следовательно,

Доказать заявленное равенство не сложно. Действительно,

Но,

что и приводит к искомому результату.

Фиктивная линейная связь

На практике часто встречаются ситуации, при которых существует заметный тренд (убывание или возрастание) в динамике изменений различных показателей. Однако, объективной связи между этими показателями не существует. В этом случае принято говорить о **фиктивной** (ложной, паразитной) линейной связи между показателями.

Для объяснения этого явления необходимо обратиться к уже полученному ранее равенству Из него вытекает, что близкие к единице значения коэффициента детерминации соответствуют близким по абсолютной величине к единице значениям коэффициента корреляции между переменными *x* и *y*. Но коэффициент корреляции

Равен где

При фиксированных значениях

и , значение будет тем ближе к 1, чем большим будет значение Последнее же обеспечивается совпадением знаков разностей и

для максимально возможной доли наблюдений переменных *x* и *y*, что как раз и имеет место, когда в процессе наблюдения обе переменные возрастают или обе переменные убывают по величине.

При этом превышение одной из переменных своего среднего значения сопровождается, как правило, и превышением второй переменной от своего среднего значения. Если же одна из переменных принимает значение, меньшее среднего значения этой переменной, то и вторая переменная, как правило, принимает значение, меньшее своего среднего.

Аналогичным образом,

значение будет тем ближе к -1, чем меньшим будет значение Последнее же обеспечивается несовпадением знаков разностей и для максимально возможной доли наблюдений переменных *x* и *y*, что имеет место, когда в процессе наблюдения одна из переменных возрастает, а вторая убывает. В этом случае, если одна из переменных принимает значение, меньшее среднего значения этой переменной, то вторая переменная, как правило, принимает значение, большее своего среднего.

Из этого следует, что **близость к единице наблюдаемого значения коэффициента детерминации не обязательно означает наличие причинной связи между двумя рассматриваемыми переменными**, а может являться лишь **следствием тренда значений обеих переменных.**

Последнее обстоятельство часто наблюдается при анализе различных экономических показателей, вычисленных без поправки на инфляцию (**недефлированные данные**).

Эконометрика

Свойства МНК-оценок параметров регрессии.
Показатели качества регрессии

Свойства МНК-оценок параметров регрессии

- Способ оценивания дает **состоятельные оценки,** если при бесконечно большом объеме выборки значение статистической оценки стремится к искомому значению параметра (характеристики) генеральной совокупности.

- Способ оценивания дает **несмещенные оценки,** если математическое ожидание оценки при данном способе оценивания тождественно искомому параметру генеральной совокупности (при любом объеме выборки).

- Оценка называется **эффективной,** если ее дисперсия минимальна (при заданном объеме выборки n).

Оценки по методу наименьших квадратов являются наилучшими, то есть несмещенными, состоятельными и эффективными!!!

Докажем это:

Докажем, что является *несмещенной оценкой* β. Если выполнены предпосылки нормальной линейной модели регрессии, то *x* – неслучайная величина, а является известной константой, а математическое ожидание $E(Cov(x, \varepsilon))=0$. Тогда

Что и требовалось доказать. Аналогично доказывается несмещенность α.

Эффективность МНК–оценок параметров доказывается с помощью теоремы Гаусса–Маркова, которая гласит, что МНК дает оценки, имеющие наименьшую дисперсию в классе всех линейных несмещенных оценок, если выполняются предпосылки нормальной линейной модели.

Состоятельность МНК-оценок параметров является следствием закона больших чисел теории вероятностей.

Оценка дисперсии случайной составляющей

Дисперсии МНК–оценок параметров регрессии будут

где – дисперсия случайной составляющей, – дисперсия фактора *x*. А так как неизвестна, то пользуются оценкой .

Оценка дисперсии случайной составляющей применяется в эконометрических задачах для анализа качества полученной модели регрессии. В случае парной линейной регрессии несмещенная оценка дисперсии случайной составляющей будет:

где e_i – остаток, равный разности между фактическим и рассчитанным по уравнению регрессии значениям *y*.

Качество модели регрессии

связывают с адекватностью модели по наблюдаемым (эмпирическим) данным. Проверка адекватности (или соответствия) модели регрессии наблюдаемым данным проводится на основе анализа остатков – e_i.

Если значения остатков для всех наблюдений $e_i=0$, то фактическое значение значимой (результирующей)

переменной совпадает с расчетным: , но на практике это, зачастую, не верно.

При анализе качества модели используется **теорема о разложении дисперсии,** согласно которой общая дисперсия результативного признака может быть разложена на две составляющие – объясненную и необъясненную уравнением регрессии.

Теорема о разложении общей дисперсии

где – **объясненная уравнением регрессии**

дисперсия результирующего признака, а

– **необъясненная уравнением регрессии дисперсия результирующего признака.** На основе этой теоремы рассчитываются **показатели качества модели регрессии.**

Общая дисперсия при этом будет:

Показатели качества модели регрессии

1. **Теоретический коэффициент детерминации (индекс для нелинейных форм связей)**

. Этот коэффициент характеризует долю

вариации (дисперсии) результирующего признака,

объясняемую регрессией в общей дисперсии и,

соответственно, 1- характеризует долю вариации необъясненную уравнением регрессии, вызванную влиянием прочих неучтенных в модели факторов. В случае парной регрессии: . Коэффициент детерминации изменяется на промежутке [0; 1], близость его к 0 означает отсутствие линейной связи, а близость к 1 наличие тесной линейной зависимости.

Показатели качества модели регрессии

2. **Коэффициент (индекс) множественной корреляции** рассчитывается как корень квадратный из коэффициента

детерминации .

Он тоже изменяется на промежутке [0; 1]. Для случая парной регрессии .

3. **Средняя квадратическая ошибка уравнения регрессии**

,

где h – число параметров модели. Величину S_e можно сравнить с дисперсией результирующего признака

. Если S_e< , то

использование данной модели регрессии является целесообразным.

Показатели качества модели регрессии

4. **Средняя ошибка аппроксимации**

Чем меньше рассеяние эмпирических точек вокруг теоретической линии регрессии, тем меньше средняя ошибка аппроксимации. Ошибка аппроксимации менее 7% свидетельствует о хорошем качестве модели.

Эконометрика

ПРОВЕРКА ГИПОТЕЗ О ЗНАЧИМОСТИ ПАРАМЕТРОВ РЕГРЕССИИ, КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ И УРАВНЕНИЯ РЕГРЕССИИ В ЦЕЛОМ

Проверка гипотез о значимости параметров регрессии

- С помощью метода наименьших квадратов можно получить лишь оценки параметров модели, чтобы проверить, значимы ли они (значимо ли отличаются от нуля) в «истинном» уравнении (в генеральной совокупности), значим ли коэффициент корреляции, требуется обратиться к проверке гипотез.
- В качестве основной гипотезы (H_0) выдвигают гипотезу о незначительном отличии от нуля «истинного» параметра или коэффициента корреляции. Альтернативной гипотезой (H_1) является обратная – о неравенстве нулю «истинного» параметра или коэффициента.

Проверка гипотез о значимости параметров регрессии

- Для проверки гипотез используют *t*–статистику распределения Стьюдента, критическое значение которой определяется по таблице (или с помощью функции СТЬЮДРАСПОБР пакета Excel) в зависимости от уровня значимости α и числа степеней свободы равного ($n–h$), где n – число наблюдений, а h – число параметров модели.

- Если **фактическое значение *t*–статистики (по модулю) меньше критического, то нет оснований отвергать основную гипотезу**, то есть «истинный» параметр регрессии (либо коэффициент корреляции) незначительно отличается от нуля при заданном уровне значимости.

Проверка гипотез о значимости параметров регрессии

Для проверки гипотезы: β=0 статистика критерия будет

где – оценка параметра β, полученная по наблюдаемым данным, а – стандартная ошибка оценки коэффициента регрессии .

Для линейной парной регрессии:

Проверка гипотез о значимости параметров регрессии

Для проверки гипотезы: α=0 статистика критерия будет

В случае парной линейной регрессии:

Проверка гипотез о значимости коэффициента корреляции

Для проверки гипотез о незначительном отклонении от нуля «истинного» коэффициента линейной парной корреляции используют статистику

при этом:

В случае парной линейной регрессии

Существует взаимосвязь

Рассмотренная формула статистики критерия проверки гипотез применима, если:

1. Оценки делаются по большому числу наблюдений *n*.
2. Величина   не близка к единице.

В случае парной линейной регрессии

Если же величина выборочного коэффициента корреляции близка к 1, то распределение его оценок отличается от распределения Стьюдента. В данном случае используют статистику Фишера. Вводится вспомогательная величина

при этом можно пользоваться готовыми таблицами преобразований. Эта величина изменяется на промежутке (–∞; +∞), что соответствует нормальному распределению. Для проверки гипотезы о незначительном отклонении от нуля «истинного» значения коэффициента парной линейной корреляции используется статистика

где Критическое значение находят по таблице стандартного нормального распределения с доверительной вероятностью (1–α). Основную гипотезу отвергают, если

Проверка гипотез о значимости уравнения регрессии в целом

Оценка значимости уравнения регрессии производится для того, чтобы узнать, пригодно ли уравнение для практического применения. Основная гипотеза утверждает, что уравнение регрессии незначительно в целом, что формально сводится к равенству нулю параметров модели или коэффициента детерминации. Для ее проверки используют *F*–статистику:

Данная статистика имеет распределение Фишера–Снедоккора, критическое значение которой можно найти в таблицах этого распределения для α=0,05 и двух степеней свободы $k_1=h-1$ и $k_1=n-h$. **В случае парной линейной регрессии существует взаимосвязь** :

Эконометрика

ПРОГНОЗ ОЖИДАЕМОГО ЗНАЧЕНИЯ РЕЗУЛЬТИРУЮЩЕГО ПРИЗНАКА.

Прогноз ожидаемого значения результирующего признака

Предположим, необходимо определить для заданного значения фактора x^p с доверительной вероятностью $(1-\alpha)$ прогнозируемое значение результирующего признака. Тогда, прогнозируемое значение результата должно принадлежать интервалу $(y^p - t \cdot \mu_p;\ y^p + t \cdot \mu_p)$, где y^p – точечный прогноз, t – коэффициент доверия (определяется по таблицам распределения Стьюдента в зависимости от уровня значимости α и числа степеней свободы $(n-2)$), а μ_p – средняя ошибка прогноза.

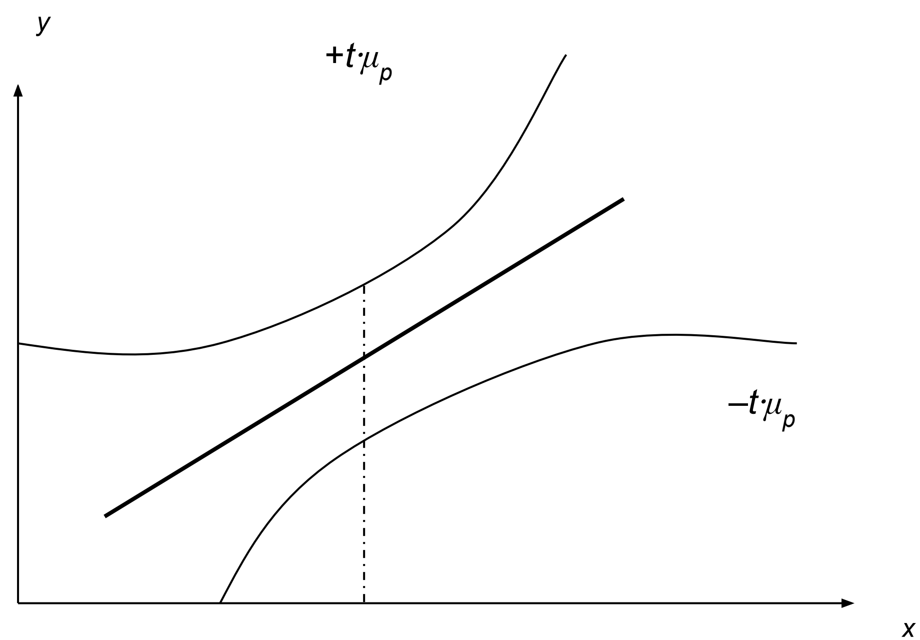
Точечный прогноз рассчитывается по линейному уравнению регрессии:

$$y^p = \alpha + \beta \cdot x^p,$$

а **средняя ошибка прогноза** будет:

Доверительный интервал

Вокруг линии регрессии образуется **доверительный интервал** («коридор», в который попадет прогнозируемое значение результирующего признака):



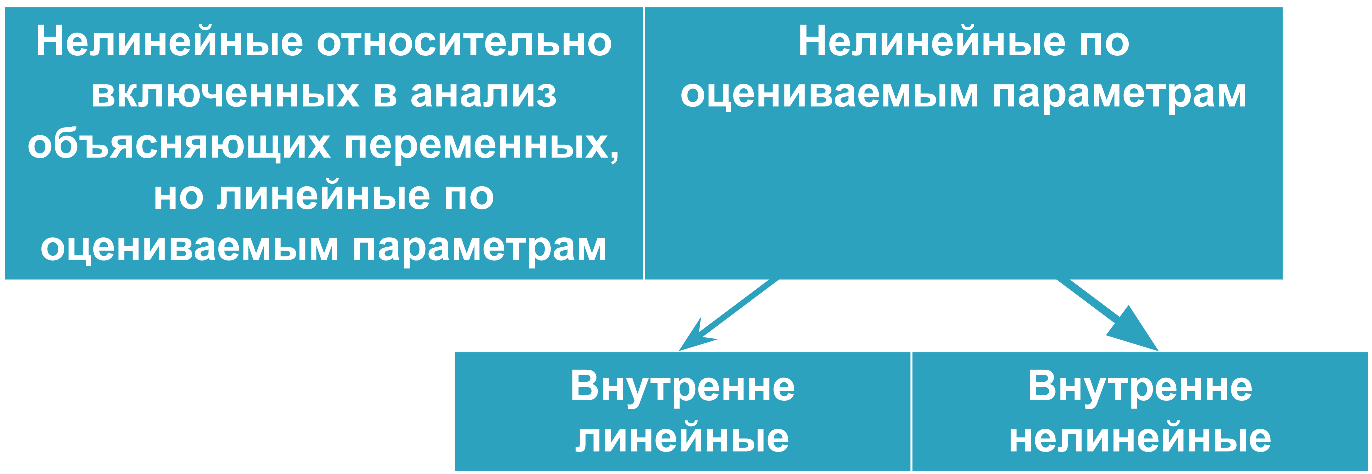
Точечная и интегральная оценка прогноза

Эконометрика

НЕЛИНЕЙНАЯ РЕГРЕССИЯ

Нелинейная регрессия

Большинство экономических ситуаций характеризуются нелинейной зависимостью между результирующим и факторными признаками.



Примеры нелинейных регрессий относительно включенных в анализ объясняющих переменных, но линейных по оцениваемым параметрам

- Полиномы различных степеней –

$$y_i = a_0 + a_1 \cdot x_i + \ldots + a_m \cdot x_i^m + \varepsilon_i.$$

- Равносторонння гипербола –

- Степенная функция –

;

- Показательная функция –

;

- Экспоненциальная функция –

.

Модель описывается уравнением:

y – спрашиваемое количество товара, x – цена, ε_i – случайная составляющая.

Преобразуем данную модель к линейному виду. Так как данная модель нелинейна относительно оцениваемых параметров, то необходимо произвести преобразование переменных. Логарифмическое преобразование приведет данное уравнение к линейному виду:

Введя новые переменные и введя новое обозначение для параметра a и ошибки, получим линейную регрессию.