3.2. ПОСТРОЕНИЕ ОДНОФАКТОРНЫХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ

Пусть в ходе корреляционного анализа прогнозисту удалось определить степень взаимосвязи между двумя случайными факторами Y_t и x_t и определить направление причинно-следственной связи между ними. После этого перед ним стоит задача найти форму зависимости и значения её коэффициентов.

Первая часть задачи довольно часто решается ещё на этапе корреляционного анализа. Действительно, если коэффициент парной корреляции близок по модулю к единице, то это, как следует из выводов предыдущего параграфа, свидетельствует о том, что предполагаемая зависимость между этими переменными близка к линейной. Для того чтобы у прогнозиста не оставалось сомнений в выборе вида модели, рекомендуется воспользоваться графическим способом – нанести на плоскость две перпендикулярные оси, по каждой из которых откладывать значения переменных Y_t и x_t. Визуальный анализ взаимосвязи может помочь прогнозисту сделать окончательный выбор модели.

Чаще всего в практике прогнозирования используют в виде моделей следующие элементарные функции: линейная, квадратичная, степенная, показательная, экспоненциальная. Их свойства и графики приведены во второй главе учебника, поэтому не будем их рассматривать вновь.

На основе визуального анализа графика, а также результатов корреляционного анализа делается вывод о применимости для прогнозирования одной или нескольких прогнозных моделей. При этом если прогнозист колеблется в выборе и считает, что две отличающиеся друг от друга модели на его взгляд в одинаковой степени подходят для целей прогнозирования, то следует отдать предпочтение более простой модели, так как именно этого требует общенаучный принцип простоты.

После выбора вида модели, возникает задача оценки коэффициентов этой модели применительно к тем статистическим данным, которые есть в распоряжении прогнозиста, т.е. – наполнения модели конкретным статистическим наполнением. Эта задача имеет довольно простое решение, но для того, чтобы понять суть процесса решения этой задачи, будем использовать другой общенаучный принцип: от простого – к сложному. Применительно к нашему случаю он означает необходимость рассмотреть решение задачи на простом примере, а затем, на основе полученных навыков и знаний, перейти к решению задачи оценки коэффициентов более сложных моделей. Поэтому рассмотрим вначале простую линейную модель:

$$\hat{Y}_t = a_0 + a_1 x_t \qquad (3.2.1)$$

Эта модель описывает исходные значения показателя Y_t с некоторой ошибкой аппроксимации ε_e:

$$\varepsilon_t = Y_t - \hat{Y}_t\,. \qquad (3.2.2)$$

Вычисляемое отклонение графически означает, что из точки каждого фактического наблюдения (на рис 3.2.1 одна из таких точек на плоскости обозначена $A\tau$) проводится линия, перпендикулярная оси $0x$ до пересечения с прямой линией, которая соответствует графическому расположению модели.

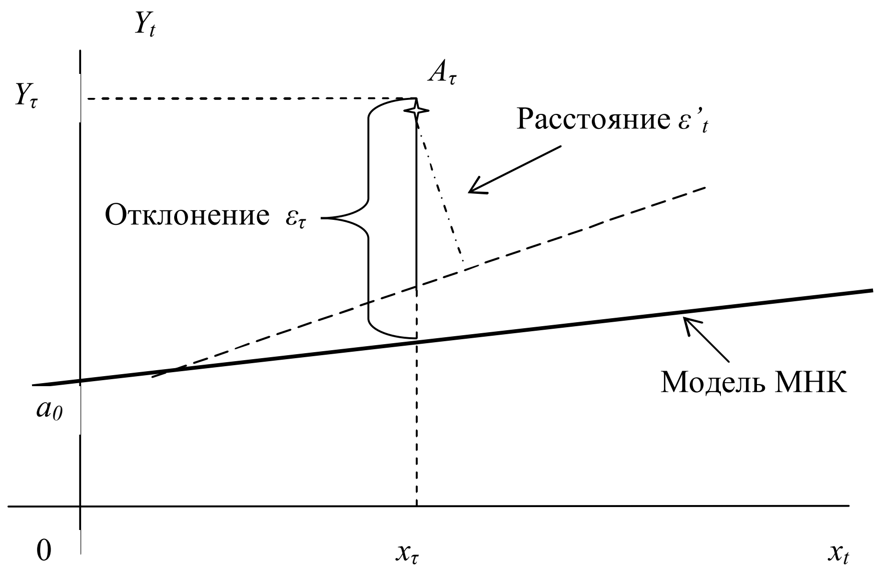

Рис. 3.2.1. Фактические точки, модель и отклонения между ними

Метод наименьших квадратов, предусматривающий нахождение таких коэффициентов модели, для которых сумма квадратов отклонений ε_t будет минимальной:

$$\sum_t \varepsilon_t^2 = \sum_t (Y_t - \hat{Y}_t)^2 \to \min, \qquad (3.2.3)$$

означает построение такой модели, которая проходит через все фактические точки так, чтобы дисперсия отклонений расчётных значений от фактических была бы минимально возможной.

Но, очевидно, что МНК – не единственный метод оценивания коэффициентов прогнозных моделей, а лишь один из многих. Прямая линия может пройти через фактические точки на плоскости и по-другому. Выбор того, как должна пройти через исходные точки модель, чтобы с её помощью и получить наиболее точный прогноз, как раз и составляет основную задачу прогнозирования. Использование МНК имеет очевидное преимущество в том, что прогнозируемый показатель Y_t моделируется так, что дисперсия описания прошлого минимальная из всех возможных, а для стационарных рядов – это гарантия того, что и в будущем при прогнозе именно такая модель даст наименьшее отклонение прогнозируемого показателя от реального значения.

Но если задача заключается в том, чтобы наилучшим образом описать взаимосвязь между переменными Y_t и x_t, с тем, чтобы прогнозировать не один показатель, а пару, то МНК будет не лучшим способом оценивания. В этом случае в качестве критерия оценивания коэффициентов линейной модели следует принять минимизацию квадратов расстояний ε'_t между прямой и фактическими наблюдениями (на рисунке 3.2.1 это расстояние показано штрих пунктирной стрелкой). Модель, коэффициенты которой оцениваются с помощью такого критерия, в общем случае пройдёт иначе, чем модель с оценками МНК. На рисунке эта модель показана жирной пунктирной линией.

В любом случае, при разных способах оценивания коэффициентов модели, получаются разные значения оценок коэффициентов, что приводит к тому, что каждая прямая на плоскости факторов будет иметь и свой оригинальный угол наклона, и отсекать будет на оси $0Y_t$ отличный от других моделей отрезок. Иначе говоря, на плоскость рисунка 3.2.1 можно нанести бесконечное множество прямых, каждая из которых будет тем или иным образом описывать фактические значения и характеризоваться различными значениями дисперсии, суммы квадратов расстояний, средними абсолютными отклонениями и т.п.

Параметр a_0 линейной модели характеризует тот отрезок, который отсекает прямая на оси показателя Y_t, когда фактор x_t принимает значение, равное нулю; параметр a_1 характеризует тангенс угла наклона прямой линии к оси $0x_t$.

Дополнительную информацию о сути процесса оценивания коэффициентов модели даёт иная интерпретации задачи, ведь она сводится по сути к поиску единственной пары оптимальных значений a_0 и a_1 из бесконечного множества возможных значений этих параметров. Действительно, варьируя, например, значения свободного члена a_0 модели (3.2.1), мы тем самым приводим к тому, что прямая линия МНК будет пересекать ось $0Y_t$ в разных точках (рис. 3.2.1), модель будет отстоять от фактических точек по разному, а значит, и сумма (3.2.3) будет различной. Именно неизвестные значения a_0 и a_1 являются параметрами задачи, а не значения Y_t и x_t. Поэтому логично рассматривать задачу оценивания коэффициентов модели не на плоскости исходных переменных, а на плоскости неизвестных параметров, то есть – на плоскости с осями $0a_0$ и $0a_1$ (Рис. 3.2.2).

Обозначим функцию, соответствующую критерию оптимизации коэффициентов модели (минимум суммы квадратов отклонений, минимум суммы абсолютных отклонений и т.п.) буквой Q. Применительно к задаче МНК это означает:

$$Q = \sum_t \varepsilon_t^2 = \sum_t (Y_t - \hat{Y}_t)^2$$

В случае нахождения оценок коэффициентов линейной модели это – функция двух переменных, то есть – $Q(a_0, a_1)$.

На рисунке 3.2.2 оценке МНК соответствует одна точка с координатами $(\hat{a}_0, \hat{a}_1)$. В этой точке значение функции Q достигает своей минимальной величины:

Q=Q*.

Все остальные точки этой плоскости будут давать в качестве своих координат такие значения коэффициентов a_0 и a_1, подставляя которые в (3.2.1) и вычисляя (3.2.3), будут получены значения функции Q (суммы квадратов отклонений), которые всегда будут больше Q*. Множество точек, в каждой из которых функция Q>Q* имеет одно и то же значение, можно представить как замкнутые гладкие выпуклые кривые. По мере приближения точек к их оптимальным с позиций критерия МНК значениям, размер этих кривых уменьшается, как уменьшается и значение функции (Q_3>Q_2>Q_1>Q*). При достижении оптимума кривая превращается в точку, а значение критерия отбора принимает оптимальное значение Q*. Именно это и происходит в ситуации, когда прогнозист оценивает значения коэффициентов нелинейной по параметрам прогнозной модели с помощью численных методов. Задавая на плоскости рис. 3.2.2 начальную точку, определяется значение функции Q в этой точке и определённым каждым численным методом подходом выбирается направление оптимизации – определяется следующая точка на плоскости коэффициентов модели. Для новой точки вновь высчитывается значение функции Q, после чего сравнивается это значение функции с предыдущим и определяется направление следующего шага. И так продолжается до тех пор, пока очередная точка на плоскости не окажется в заранее определённых допустимых окрестностях оптимальной точки Q*.

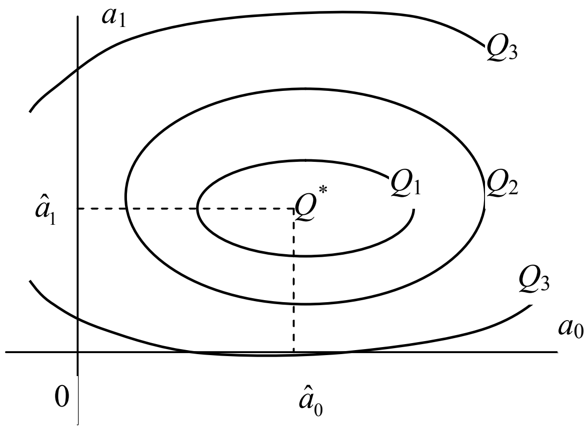

Рис. 3.2.2. Плоскость коэффициентов линейной модели и значения функции

Однако в простом случае оценивания коэффициентов линейной однофакторной модели с помощью МНК никаких многоитеративных процедур осуществлять не нужно, довольно просто решить систему нормальных уравнений:

$$\begin{cases} \sum_t Y_t = a_0 T + a_1 \sum_t x_t \\ \sum_t Y_t x_t = a_0 \sum_t x_t + a_1 \sum_t x_t^2 \end{cases}. \qquad (3.2.4)$$

Но что означает эта система уравнений, если её рассматривать как задачу на плоскости коэффициентов (рис. 3.2.2)? Первое уравнение системы МНК представляет собой уравнение прямой линии, проходящей на плоскости коэффициентов a_0 и a_1. Его можно представить для удобства в виде, называющемся «уравнением в отрезках»:

$$\frac{a_0}{\frac{\sum_t Y_t}{T}} + \frac{a_1}{\frac{\sum_t Y_t}{\sum_t x_t}} = 1. \qquad (3.2.5)$$

Значение, вычисляемое в знаменателе под каждым коэффициентом представляет собой отрезок, который отсекает данная прямая линия на оси этого коэффициента. Или, иначе говоря, запись (3.2.5) даёт возможность сразу же определить две точки, лежащие на прямой МНК плоскости коэффициентов:

- первая точка с координатами $(\frac{\sum_t Y_t}{T}, 0)$ - она лежит на оси $0a_0$,

- вторая точка имеет координаты $(0, \frac{\sum_t Y_t}{\sum_t x_t})$ - она лежит на оси $0a_1$.

Точно также и второе уравнение системы МНК (3.2.4) можно рассматривать как уравнение прямой линии, расположенной на плоскости коэффициентов модели. И её удобнее всего изучать, представив в форме «уравнения в отрезках»:

$$\frac{a_0}{\frac{\sum_t Y_t x_t}{\sum_t x_t}} + \frac{a_1}{\frac{\sum_t Y_t x_t}{\sum_t x_t^2}} = 1. \qquad (3.2.6)$$

Эта линия, как можно увидеть из отрезков (3.2.6) также пересекает оси координат, причём точки пересечения легко определить из этой записи:

- ось $0a_0$ модель пересекает в точке $(\frac{\sum_t Y_t x_t}{\sum_t x_t}, 0)$,

- ось $0a_1$ – в точке $(0, \frac{\sum_t Y_t x_t}{\sum_t x_t^2})$.

Нанесём эти прямые линии на плоскость коэффициентов (рис. 3.2.3). Их пересечение, как следует из логики решения систем уравнений, и будет представлять собой на плоскости точку, координаты которой соответствуют оценкам МНК.

Такое понимание задачи оценивания значений коэффициентов линейной модели позволит нам в дальнейшем более подробно изучить задачи оценивания доверительных границ выборочных значений коэффициентов прогнозных моделей, найденных с помощью МНК, а также исследовать проблему построения многофакторных моделей в условиях мультиколлинеарности.

Рассмотрев задачу оценки коэффициентов линейной модели с помощью МНК, мы понимаем, что аналогичные процессы происходит при оценивании коэффициентов нелинейных моделей. Если эти модели представлены в мультипликативном виде, то их линеаризуют и к линеаризованному виду модели применяют МНК.

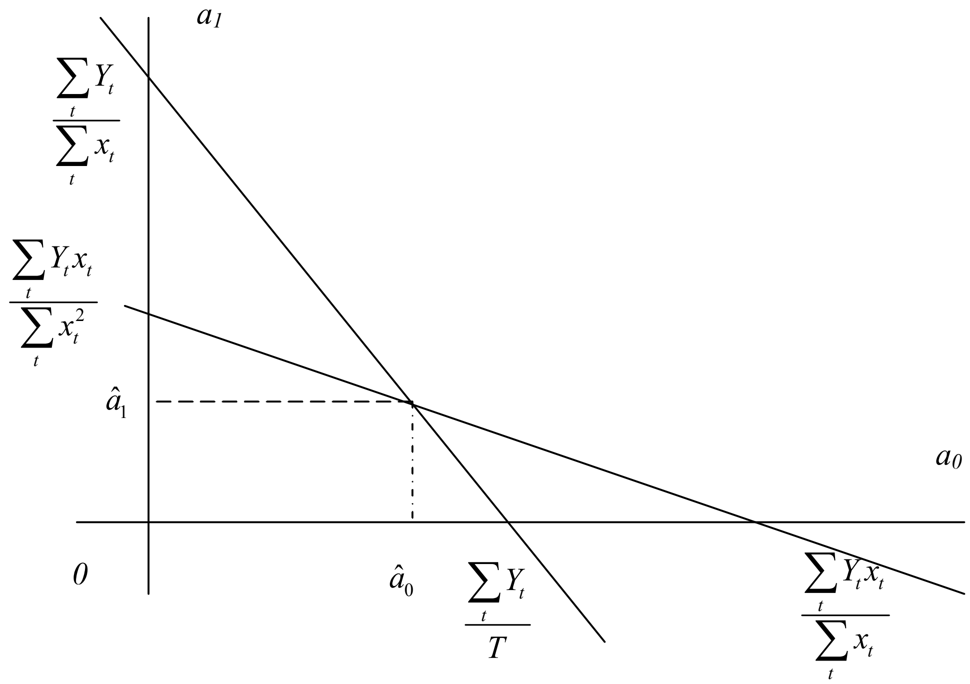

Рис. 3.2.3. Уравнения системы МНК на плоскости коэффициентов модели и оценки МНК этих коэффициентов

Покажем это на примере степенной однофакторной модели:

$$\hat{Y}_t = a_0 x_t^{a_0} \,. \qquad (3.2.7)$$

Прежде всего, необходимо линеаризовать модель, то есть, привести её у виду (3.2.1). Для этого прологарифмируем левую и правую части равенства (3.2.7) по любому основанию, например, десятичному:

$$\lg \hat{Y}_t = \lg a_0 + a_1 \lg x_t \,. \qquad (3.2.8)$$

Получена линейная модель, которая описывает десятичные логарифмы исходного ряда с некоторой ошибкой:

$$\varepsilon_t = \lg Y_t - \lg \hat{Y}_t = \lg Y_t - (\lg a_0 + a_1 \lg x_t) \,. \qquad (3.2.9)$$

Если теперь просуммировать квадраты этих отклонений и потребовать найти такие значения коэффициентов lga_0 и a_1, чтобы эта сумма была минимально возможная, то тем самым мы приходим к необходимости решить систему нормальных уравнений МНК следующего вида:

$$\begin{cases} \sum_t \lg Y_t = T \lg a_0 + a_1 \sum_t \lg x_t \\ \sum_t \lg Y \lg_t x_t = \lg a_0 \sum_t \lg x_t + a_1 \sum_t (\lg x_t)^2 \end{cases} . \qquad (3.2.10)$$

После решения этой системы и вычисления коэффициентов lga_0 и a_1 можно найти коэффициент a_0:

$$a_0 = 10^{\lg a_0} \,. \qquad (3.2.11)$$

Тем самым находятся коэффициенты исходной модели (3.2.7). Не стоит при этом всё же забывать, что найденный таким образом коэффициент a_0 модели будет смещённым, а его оценка – неэффективной и не состоятельной. О причине этого писалось в предыдущей главе, поэтому здесь повторять аналогичные рассуждения не имеет смысла.