Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра Математической экономики

Определение характеристик экономических трендов с помощью метода наименьших квадратов

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 451 группы

направления 38.03.05 «Бизнес-информатика»

механико-математического факультета

Зайнышевой Дарьи Айратовны

Научный руководитель
д.ф.-м.н, доцент ____________________ А.Ю. Трынин
подпись, дата

Зав. кафедрой:
д.ф.-м.н., профессор ____________________ С.И. Дудов
подпись, дата

Саратов 2017

ВВЕДЕНИЕ

Согласно определению, приведенному в учебнике Е. А. Аникиной «Экономическая теория», экономический тренд – это длительная («вековая») тенденция изменения экономических показателей. Понятие экономического тренда чаще всего используется в техническом анализе.

При техническом подходе к анализу рынка понятие тенденции или тренда является ключевым. Тренд – это направление, в котором движется рынок. Динамика рынка представляет собой серию зигзагов, которые напоминают череду волн, подъемов и спадов. Направление динамики этих подъемов и падений и образует тенденцию рынка.

Экономический тренд можно рассматривать в виде парной линейной регрессионной зависимости среднего значения экономического показателя от времени. В этом случае необходимо выбрать способ оценивания параметров регрессии.

Классический подход к оцениванию параметров линейной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых отклонение фактических значений от тренда минимально.

Актуальность данной работы связана с использованием экономических трендов в техническом анализе, а также возможностью применения уравнения регрессии для прогнозирования возможных ожидаемых значений результативного признака.

Прогнозируемое значение результативного показателя получается при подстановке в уравнение регрессии ожидаемой величины факторного признака. Прогноз, полученный подстановкой в уравнение регрессии ожидаемого значения фактора, называют точечным прогнозом. Вероятность точной реализации такого прогноза крайне мала, поэтому особенно важно, как можно более точно оценить коэффициенты регрессии, а метод наименьших квадратов дает наилучшие (состоятельные, эффективные и несмещенные) оценки параметров уравнения регрессии. Но только в том случае, если выполняются определенные предпосылки.

Целью данной работы является изучение метода наименьших квадратов и использование его для построения тренда наблюдаемых экономических явлений.

Для достижения поставленное цели были выделены следующие задачи:

- определение понятия парной регрессионной модели;
- выведение формул для оценки коэффициентов парной регрессионной модели с помощью метода наименьших квадратов;
- изучение предпосылок использования метода наименьших квадратов;
- интерпретация параметров парной регрессионной модели;
- изучение способов оценки качества модели;
- построение экономического тренда с помощью метода наименьших квадратов в программе Microsoft Excel.

Работа состоит из двух разделов. Первый раздел носит теоретический характер и делится на три подраздела. В первом подразделе вводятся понятия функциональной зависимости, корреляционной зависимости, а также парной регрессионной модели, приводится регрессионное уравнение.

Второй подраздел содержит вывод формул расчета параметров парной регрессионной модели с помощью метода наименьших квадратов, предпосылки применения метода наименьших квадратов, анализ точности определения коэффициентов регрессии.

В третьем подразделе рассматривается алгоритм оценки качества модели.

Второй раздел работы состоит из двух подразделов. Первый подраздел содержит краткую информацию о программе Microsoft Excel и языке макрокоманд Visual Basic for Applications. Во втором подразделе приведен код макроса, который строит экономический тренд по введенным данным, и его подробное описание.

Основное содержание работы

Парная регрессия представляет собой регрессию между двумя переменными – y и x, т. е. модель вида:

$$y = \hat{f}(x), \tag{1}$$

где y – зависимая переменная (результативный признак); x – независимая или объясняющая переменная (признак-фактор). Знак «^» означает, что между переменными x и y нет строгой функциональной зависимости, поэтому практически в каждом отдельном случае величина y складывается из двух слагаемых:

$$y = \widehat{y_x} + \varepsilon, \tag{2}$$

где y – фактическое значение результативного признака; $\widehat{y_x}$ – теоретическое значение результативного признака, найденное исходя из уравнения регрессии; ε – случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии.

Случайная величина ε 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 зависит величина случайных ошибок: они тем меньше, чем в большей мере теоретические значения результативного признака $\widehat{y_x}$, подходят к фактическим данным y.

В парной регрессии выбор вида математической функции $\widehat{y_x} = f(x)$ может быть осуществлен тремя методами:

1) графическим;

2) аналитическим, т.е. исходя из теории изучаемой взаимосвязи;

3) экспериментальным.

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

Значительный интерес представляет аналитический метод выбора типа уравнения регрессии. Он основан на изучении материальной природы связи исследуемых признаков.

При обработке информации на компьютере выбор вида уравнения регрессии обычно осуществляется экспериментальным методом, т. е. путем сравнения величины остаточной дисперсии $\sigma^2_{\text{ост}}$, рассчитанной при разных моделях.

$$\sigma^2_{\text{ост}} = \frac{1}{n}\sum(y - \widehat{y_x})^2. \qquad (3)$$

Чем меньше величина остаточной дисперсии, тем меньше влияние не учитываемых в уравнении регрессии факторов и тем лучше уравнение регрессии подходит к исходным данным.

Линейная регрессия сводится к нахождению уравнения вида

$$\widehat{y_x} = a + b * x \text{ или } y = a + b * x + \varepsilon. \qquad (4)$$

Уравнение вида $\widehat{y_x} = a + b * x$ позволяет по заданным значениям фактора x находить теоретические значения результативного признака, подставляя в него фактические значения фактора x.

Построение линейной регрессии сводится к оценке ее параметров – a и b. МНК позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\widehat{y_x}$ минимальна:

$$\sum_{i=1}^{n}\varepsilon_i^2 = \sum_{i=1}^{n}\left(y_i - \hat{y}_{x_i}\right)^2 = \sum_{i=1}^{n}(y_i - a - bx_i)^2 \to min. \qquad (5)$$

Как известно из курса математического анализа, чтобы найти минимум функции (5), надо вычислить частные производные по каждому из параметров a и b и приравнять их к нулю. Обозначим $\sum_{i=1}^{n}\varepsilon_i^2$ через $S(a,b)$, тогда:

$$S(a,b) = \sum (y - a - b * x)^2. \tag{6}$$

$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum (y - a - b * x) = 0 \\ \frac{\partial S}{\partial b} = -2 \sum (y - a - b * x) = 0 \end{cases} \tag{7}$$

После преобразований, получим искомые оценки параметров a и b.

$$a = \bar{y} - b\bar{x}, \qquad b = \frac{cov\,(x,y)}{\sigma_x^2}, \tag{8}$$

где $cov\,(x,y) = \overline{xy} - \bar{x} * \bar{y}$ – ковариация признаков x и y, $\sigma_x^2 = \overline{x^2} - \bar{x}^2$ – дисперсия признака x и

$$\bar{x} = \frac{\sum x}{n}, \qquad \bar{y} = \frac{\sum y}{n}, \qquad \overline{xy} = \frac{\sum x\,y}{n}, \qquad \overline{x^2} = \frac{\sum x^2}{n}.$$

Параметр b называется коэффициентом регрессии. Его величина показывает среднее изменение результата с изменением фактора на одну единицу.

Формально a – значение y при $x = 0$. Если признак-фактор x не может иметь нулевого значения, то вышеуказанная трактовка свободного члена a не имеет смысла, т.е. параметр a может не иметь экономического содержания.

После построения уравнения регрессии проводится проверка наличия у оценок ε_i (случайных остатков) тех свойств, которые предполагались. Связано это с тем, что 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.

Исследования остатков ε_i предполагают проверку наличия следующих пяти предпосылок МНК:

1) случайный характер остатков;

2) нулевая средняя величина остатков;

3) гомоскедастичность – дисперсия каждого отклонения ε_i одинакова для всех значений x;

4) отсутствие автокорреляции остатков – значения остатков ε_i распределены независимо друг от друга;

5) остатки подчиняются нормальному распределению.

Если распределение случайных остатков ε_i не соответствует некоторым предпосылкам МНК, то следует корректировать модель.

После того как найдено уравнение линейной регрессии, проводится оценка значимости как уравнения в целом, так и отдельных его параметров.

Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации:

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y - \widehat{y_x}}{y}\right| * 100\% \tag{9}$$

Средняя ошибка аппроксимации не должна превышать 8–10%.

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии в качестве такого показателя выступает линейный коэффициент корреляции r_{xy}, который можно рассчитать по следующим формулам:

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = \frac{cov\,(x, y)}{\sigma_x * \sigma_y}. \tag{10}$$

Линейный коэффициент корреляции находится в пределах: $-1 \leq r_{xy} \leq 1$. Чем ближе абсолютное значение r_{xy} к единице, тем сильнее линейная связь между факторами (при $r_{xy} = \pm 1$ имеем строгую функциональную зависимость). Но следует иметь в виду, что близость абсолютной величины линейного коэффициента корреляции к нулю еще не означает отсутствия связи между признаками. При другой (нелинейной) спецификации модели связь между признаками может оказаться достаточно тесной.

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции r_{xy}^2, называемый коэффициентом детерминации. Коэффициент детерминации характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака:

$$r_{xy}^2 = 1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}, \tag{11}$$

где $\sigma_{\text{ост}}^2 = \frac{1}{n}\sum(y - \widehat{y_x})^2$, $\sigma_y^2 = \frac{1}{n}\sum(y - \bar{y})^2 = \overline{y^2} - \bar{y}^2$.

Соответственно величина $1 - r_{xy}^2$ характеризует долю дисперсии y, вызванную влиянием остальных, не учтенных в модели, факторов.

Оценка значимости уравнения регрессии в целом производится на основе F -критерия Фишера. В математической статистике дисперсионный анализ рассматривается как самостоятельный инструмент статистического анализа. В эконометрике он применяется как вспомогательное средство для изучения качества регрессионной модели.

Согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной y от среднего значения $\bar{y}$ раскладывается на две части – «объясненную» и «необъясненную»:

$$\sum(y - \bar{y})^2 = \sum(\widehat{y_x} - \bar{y})^2 + \sum(y - \widehat{y_x})^2, \tag{12}$$

где $\sum(y - \bar{y})^2$ – общая сумма квадратов отклонений; $\sum(\widehat{y_x} - \bar{y})^2$ – сумма квадратов отклонений, объясненная регрессией (или факторная сумма квадратов отклонений); $\sum(y - \widehat{y_x})^2$ – остаточная сумма квадратов отклонений, характеризующая влияние неучтенных в модели факторов.

Схема дисперсионного анализа имеет вид, представленный в таблице 1 (n – число наблюдений, m – число параметров при переменной x).

Таблица 1

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Общая	$\sum(y-\bar{y})^2$	$n-1$	$S^2_{\text{общ}} = \dfrac{\sum(y-\bar{y})^2}{n-1}$
Факторная	$\sum(\widehat{y_x}-\bar{y})^2$	m	$S^2_{\text{факт}} = \dfrac{\sum(\widehat{y_x}-\bar{y})^2}{m}$
Остаточная	$\sum(y-\widehat{y_x})^2$	$n-m-1$	$S^2_{\text{ост}} = \dfrac{\sum(y-\widehat{y_x})^2}{n-m-1}$

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F -критерия Фишера:

$$F = \frac{S^2_{\text{факт}}}{S^2_{\text{ост}}}. \tag{13}$$

Фактическое значение F -критерия Фишера сравнивается с табличным значением $F_{\text{табл}}(\alpha;\ k_1;\ k_2)$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n-m-1$. При этом, если фактическое значение F -критерия больше табличного, то признается статистическая значимость уравнения в целом.

В парной линейной регрессии оценивается значимость не только уравнения в целом, но и отдельных его параметров. С этой целью по каждому из параметров определяется его стандартная ошибка: m_b и m_a.

Стандартная ошибка коэффициента регрессии определяется по формуле:

$$m_b = \sqrt{\frac{S^2_{\text{ост}}}{\sum(x-\bar{x})^2}} = \frac{S_{\text{ост}}}{\sigma_x\sqrt{n}}, \tag{14}$$

где $S^2_{\text{ост}} = \frac{\sum(y-\widehat{y_x})^2}{n-2}$ – остаточная дисперсия на одну степень свободы.

Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т.е. определяется фактическое значение t -критерия Стьюдента: $t_b = \frac{b}{m_b}$ которое затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы $(n-2)$. Доверительный интервал для коэффициента регрессии определяется как $b \pm t_{\text{набл}} * m_b$.

Стандартная ошибка параметра a определяется по формуле:

$$m_a = \sqrt{S_{\text{ост}}^2 * \frac{\sum x^2}{n * \sum (x - \bar{x})^2}} = S_{\text{ост}} * \frac{\sqrt{\sum x^2}}{\sigma_x n}, \tag{15}$$

Процедура оценивания существенности данного параметра не отличается от рассмотренной выше для коэффициента регрессии. Вычисляется t -критерий: $t_a = \frac{a}{m_a}$, его величина сравнивается с табличным значением при $(n-2)$ степенях свободы.

Практическая часть

В качестве практического применения полученных знаний была поставлена задача написать макрос, выполняющий следующие действия:

1) Построить график японские свечи для исходных данных.

2) Произвести все необходимые промежуточные вычисления для оценки параметров регрессии методом наименьших квадратов.

3) Произвести оценку параметров методом наименьших квадратов.

4) Построить график, отображающий полученный экономический тренд.

Исходные данные были взяты с сайта https://www.finam.ru. Они представляют собой котировки акций компаний Публичное акционерное общество «Акционерная нефтяная компания "Башнефть"» и Публичное акционерное общество «Нефтяная компания "ЛУКОЙЛ"» за период 1.02.2017 – 2.03.2017. Цены фиксировались один раз в день.

ЗАКЛЮЧЕНИЕ

В работе изучен метод наименьших квадратов и его использование для построения тренда наблюдаемых экономических явлений.

В работе приведено определение парной регрессивной модели; выведены формулы для оценки коэффициентов парной регрессионной модели с помощью метода наименьших квадратов; изучены предпосылки использования метода наименьших квадратов, интерпретация параметров парной регрессионной модели, способы оценки качества модели.

В качестве практического задания был написан макрос в программе Microsoft Excel, который производит оценку параметров линейной регрессии методом наименьших квадратов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Швагер, Джек. Технический анализ. Полный курс. М.: Альпина Паблишер, 2001. 768 с.

2 Аникина, Е.А. Экономическая теория: учебник / Е.А. Аникина, Л.И. Гавриленко. Томск: Изд-во Томского политехнического университета, 2014. 413 с.

3 Иконникова, И.А. Эконометрика: учебно-методическое пособие / И.А. Иконникова, Н.А. Вихорь. Томск: Изд-во Том. гос. архит.-строит. ун-та, 2012. 88 с.

4 Эконометрика: Учебник / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002. 344 с.

5 Елисеева, И.И. Общая теория статистики: Учебник / Под ред. чл.-корр. РАН И.И. Елисеевой. - 4-е изд. М.: Финансы и статистика, 2001. 480 с.

6 Эконометрика: Учебник / Тихомиров Н.П., Дорохина Е.Ю. М.: Издательство «Экзамен», 2003. 512 с.

7 Носко, В.П. Эконометрика: учебник для студентов высш. учеб. заведений, обучающихся по экон. специальностям. В 2 ч. М.: Дело, 2011.

8 Гмурман, В.Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В.Е. Гмурман. М. : Высшая школа, 2009. 480 с.

9 Кремер, Н.Ш. Эконометрика: Учебник для вузов / Под ред. проф. Н.Ш. Кремера. М.: ЮНИТИ-ДАНА, 2002. 311 с.

10 Эконометрика: Учебн. пособие для вузов / А.И. Орлов М.: Издательство «Экзамен», 2002. 576 с.

11 Мардас, А.Н. Эконометрика. СПб: Питер, 2001. 144 с.

12 Бородич, С.А. Эконометрика / С.А. Бородич. Минск : Новое знание, 2006. 407 с.

13 Кулинич, Е.И. Эконометрия. М.: Финансы и статистика, 2001. 304 с.

14 Просветов, Г.И. Теория вероятностей и математическая статистика. Задачи и решения / Г.И. Просветов. М. : Альфа-Пресс, 2009. 272 с.

15 Афанасьев, В.Н. Эконометрика: учебник / В.Н. Афанасьев, М.М. Юзбашев, Т.И. Гуляева ; под ред. В.Н. Афанасьева. М. : Финансы и статистика, 2006. 256 с.
16 Лебедев, А. Понятный самоучитель Excel 2013. СПб.: Питер, 2014. 128 с.
17 Винстон Уэйн Л. Microsoft Excel 2013. Анализ данных и бизнес-моделирование. СПб.: «БХВ-Петербург», 2015. 864с.
18 Серогодский, В.В. Excel 2013. Полное руководство. СПб. : Наука и Техника, 2015. 416 с.
19 Уокенбах, Джон. Excel 2013. Трюки и советы Джона Уокенбаха. СПб.: Питер, 2014. 336 с.
20 Уокенбах, Джон. Excel 2010: Профессиональное программирование на VBA. М. : ООО «И.Д. Вильямс», 2012. 944 с.
21 Уокенбах, Джон. Excel 2013. Библия пользователя. М. : ООО «И.Д. Вильямс», 2015. 928 с.