Н.В.Катаргин

Прикладная эконометрика.

Учебное пособие «Прикладная эконометрика» предназначено для подготовки бакалавров и магистров, переподготовки специалистов в области экономики, финансов и менеджмента. Изучаются различные методы построения и настройки экономико-математических моделей на основе статистических данных с использованием компьютеров. Рассмотрены линейные и нелинейные модели парной и множественной регрессии, прогноз цен и формирование оптимального портфеля ценных бумаг на фондовом рынке, макроэкономические модели из нескольких уравнений. Внимание уделено оценке погрешностей параметров моделей, в том числе методом Монте Карло.

Ключевые слова: эконометрика, регрессия, временные ряды, экономические модели, Excel.

The manual "Applied Econometrics" is intended for the training of bachelors and masters, retraining of experts in of economics, finance and management. Various methods of construction and estimation of economic-mathematical models using statistical data with use of computers are studied. Linear and nonlinear models of pair and multiple regression, the forecast of the prices and formation of an optimum portfolio in stock market, macroeconomic models with several equations are considered. The attention is paid to an assessment of errors of parameters of models, including Monte Carlo method.

Keywords: econometrics, regression, time series, economic models, Excel.

.

Катаргин Николай Викторович

Прикладная эконометрика. Учебное пособие.

Nikolai Katargin
Applied Econometrics. Manual.

Содержание

Введение

Эконометрика как наука развивается более 100 лет, включена в обязательный перечень дисциплин для подготовки экономистов и финансистов, написано много учебников. Зачем нужна эта наука, и зачем нужен ещё один учебник?

По определению Ю.М.Лужкова, бухгалтер должен уметь посчитать деньги, припрятать их и не дать растратить, а экономист и финансист должны уметь построить модель процесса, в соответствии с ней принять решение, вложить деньги и получить прибыль. Для принятия ответственного управленческого решения в экономике и управлении капиталом нужно обработать и осмыслить огромное количество информации. (Даже обычный магазин производит огромное количество данных о поставках, продажах, платежах и т.д., а инвестиционные проекты и управление капиталом…). Как это сделать? Современные информационные технологии обеспечивают сбор, хранение информации и доступ к ней. Но как осмыслить данные и принять решение? Для этого разработаны методики и программные продукты, обеспечивающие:

- ***Представление данных в наглядном виде.*** Наиболее удобная форма – графики и диаграммы, хотя в некоторых случаях можно применять анимацию. Графика позволяет наглядно отобразить большое количество информации. Человеческий глаз обладает хорошей аналитической способностью, он может уловить закономерности и аномалии на графиках и картинках, и это используется при мониторинге на электростанциях и производствах, в космонавтике, в военном деле и т.д.
- ***Статистическая обработка данных:*** расчёт средних значений, дисперсий, корреляций, законов распределения, то есть представление данных в обобщённом виде.
- ***Построение экономико-математических моделей и их настройка,*** то есть оценка параметров с помощью методов эконометрики.

- ***Построение на основе модели оптимальных планов и управленческих решений;*** на Западе и часто в России это именуют ***логистика.***

Основная цель изучения эконометрики – увидеть за множеством чисел модель, построить её, настроить и оценить её пригодность для принятия управленческого решения.

Данный курс имеет практическую направленность и рассчитан на будущих специалистов финансово-экономического профиля, конкретно – на обучение бакалавров и магистров, а также на переподготовку специалистов.

За основу взят курс эконометрики профессора К. Доугерти, Лондонская школа экономики [1], а также использованы учебники профессора В.А. Бывшева [2], профессора Л.О. Бабешко [3] и член-корр. РАН И.И. Елисеевой [4 , 5]. Курс адаптирован для подготовки бакалавров в связи с сокращением аудиторной, особенно лекционной нагрузки. Поэтому многие разделы, требующие для своего изучения много времени и хорошей математической подготовки, исключены или о них даётся общее представление. Теоретический материал сведён к минимуму, но его надо знать обязательно, чтобы грамотно строить и оценивать модели. Предполагается, что изучивший этот курс сможет решать практические задачи с использованием компьютера, но не будет профессиональным математиком, не будет доказывать теорем и выводить сложные формулы. Всё уже доказано, выведено и заложено в программное обеспечение компьютеров. Задачи решаются на компьютерах, но грамотное построение модели и понимание выдаваемых компьютером результатов остаётся за человеком.

Программирование сложных алгоритмов в качестве сервисов и функций электронных таблиц Excel и других прикладных программ (в эконометрике – Stata, EViews и др.) позволило их использовать широкому кругу научных сотрудников и практиков, от которых требуется:

- грамотная постановка задачи;
- подбор и оценка исходных данных;

- построение концептуальной модели: тексты, формулы, графики, рисунки, таблицы, описывающие систему или процесс;
- построение структурной модели: системы уравнений, тождеств и неравенств, и алгоритма решения задачи;
- подбор сервисов и функций для ее решения;
- построение и настройка модели в компьютере;
- проведение расчетов с помощью сервисов и функций, используя графические интерфейсы и подсказки;
- интерпретация результатов и оценка их надежности;
- оформление результатов работы.

1. Последовательность разработки математических моделей и решения задач

В данном разделе рассмотрены общие принципы выполнения любого проекта – от забивания гвоздя до строительства предприятия, с учётом особенностей разработки и использования экономико-математических, в частности эконометрических моделей. Если раньше предполагалось, что перечисленные этапы выполняются последовательно, то в настоящее время считается нормальным возврат к предыдущим этапам. Например, при разработке концептуальной модели может возникнуть потребность в новых данных, и вообще этот этап может предшествовать сбору данных.

1. ***Постановка задачи***. Необходимо понять потребности, сформулировать цель работы, предполагаемые результаты, имеющиеся ресурсы (денежные, технические, кадровые, юридические), объем работ, который предполагается выполнить; оценить имеющиеся разработки и программное обеспечение, стоимость закупки или разработки недостающего; решить вопрос о целесообразности разработки; разработать техническое задание, календарный план, соглашение о цене. Неверные постановка задачи или выходные параметры могут привести к большим потерям времени и денег! Одна из

основных задач данного курса – чтобы вы смогли грамотно сформулировать задачу по обработке данных и построению экономической модели.

2. ***Обследование предметной области, сбор и оценка качества информации.*** От качества исходной информации об объекте моделирования зависят как адекватность модели, так и достоверность результатов моделирования. Возможно, исследуемые объекты придётся разбивать на группы, и в каждой будут свои закономерности. Возможно наличие ошибочных или аномальных данных (например, о продажах перед праздниками), которые надо уметь выделить и рассмотреть отдельно. Обычно они видны на графиках. Теоретические предпосылки также являются информацией об объекте и способствуют целенаправленному сбору данных. В эконометрических исследованиях данные разбиваются на три группы: ***cross-sectional data***, в российских учебниках обозначаются как "***пространственные***"; ***временные ряды (time series); panel data,*** в российских учебниках "***панельные***", содержащие набор временных рядов.

3. ***Построение концептуальной модели***:

В зависимости от характера изучаемых процессов в системе все виды моделирования могут быть разделены на детерминированные и стохастические, статические и динамические, дискретные, непрерывные и дискретно-непрерывные. ***Детерминированное моделирование*** отображает детерминированные процессы, т.е. процессы, в которых предполагается отсутствие всяких случайных воздействий; ***стохастическое моделирование*** отображает вероятностные процессы и события; эконометрическое моделирование относится к этому виду. В этом случае анализируется ряд реализаций случайного процесса и оцениваются средние характеристики, т.е. набор однородных реализаций. ***Статическое моделирование*** служит для описания поведения объекта в какой-либо момент времени, в эконометрике такие модели называют пространственными. ***Динамическое моделирование*** отражает поведение объекта во времени. В эконометрике изучают временные ряды и их наборы (панельные данные). ***Дискретное моделирование*** служит

для описания процессов, которые предполагаются дискретными, соответственно непрерывное моделирование позволяет отразить непрерывные процессы в системах, а ***дискретно-непрерывное моделирование*** используется для тех случаев, когда хотят выделить наличие как дискретных, так и непрерывных процессов. Эконометрика базируется на дискретных данных, но результатом является непрерывная функция.

На этапе концептуальной модели можно использовать неформальное описание реального объекта (тексты, рисунки, схемы, диаграммы и т.д.), но используются и формализованные технологии: системный анализ, Универсальный Язык Моделирования UML и др.

Основные этапы построения концептуальной модели:

– выдвижение гипотез и предложений;

– определение параметров и переменных модели;

– обоснование выбора показателей и критериев эффективности системы;

– составление содержательного описания модели.

Гипотезы при построении модели системы служат для заполнения "пробелов" в понимании задачи исследователем. Предположения дают возможность провести упрощение модели. В процессе работы с моделью системы возможно многократное возвращение к этому подэтапу в зависимости от полученных результатов моделирования и новой информации об объекте.

При определении параметров и переменных составляется перечень входных, выходных и управляющих переменных, а также внешних и внутренних параметров системы.

Выбранные показатели и критерии эффективности системы должны отражать цель функционирования системы и представлять собой функции переменных и параметров системы. Основные виды переменных в эконометрике:

- ***эндогенные***, или зависимые переменные, прогнозирование которых является одной из основных задач эконометрики;

- ***экзогенные,*** или влияющие переменные; могут быть внешними по отношению к системе (курс доллара, учетная ставка, время), или мы можем ими управлять: расходы на разные цели;
- ***лаговые:*** переменные прошедших временных интервалов; вчера мы пытались их прогнозировать, а сегодня знаем.

Экзогенные и лаговые объединяют термином ***предопределённые.*** Кроме того, существуют фиктивные, замещающие, инструментальные переменные, которые мы обсудим далее.

Разработка концептуальной модели завершается составлением содержательного описания, которое может содержать тексты, формулы, графики, рисунки, таблицы, описывающие систему или процесс, и используется как основной документ для дальнейших действий.

4. ***Формальное описание*** задач, построение ***структурной модели***: системы уравнений, тождеств, ограничений-равенств и ограничений-неравенств.

5. Разработка алгоритма решения задачи. Алгоритм – это конечная последовательность точно определенных действий, однозначно определяющая процесс преобразования исходных и промежуточных данных, приводящий к решению задачи. В эконометрике – это преобразование структурной модели к приведённой форме: уравнению или системе равенств, в которых эндогенные (прогнозируемые) переменные будут в левой части, а экзогенные и лаговые – в правой. Этот этап требует большого количества вычислений. В настоящее время многие программы для решения таких задач оформлены в виде сервисов различных прикладных пакетов: Excel, MatCad, MatLab, Stata, EViews и др., решение задачи обычно заменяется ***выбором пакета, сервиса, его настройкой и стыковкой с используемыми данными***, обычно в интерактивном графическом режиме: ввод формул, установка ограничений и т.д. Правильный выбор метода решения, программного обеспечения (и программиста в реальном проекте) могут уменьшить время и стоимость проекта в десятки раз! Личный опыт автора это подтверждает, и предлагаемые далее технологии решения задач позволяют это сделать.

6. ***Тестирование***. Программа, не имеющая синтаксических ошибок, может иметь логические ошибки и выдавать неверные результаты. Поэтому как отдельные блоки, так и программа в целом должны быть проверены с помощью тестовых задач с известными решениями. 90% эконометрики – это методы оценки надёжности модели в целом и её параметров. ***Показатели качества эконометрической модели: коэффициент детерминации R^2, статистика Фишера F, t-статистики Стьюдента для коэффициентов уравнений, тест Дарбина-Уотсона на автокорреляцию DW, тест Голдфелда-Квандта на гетероскедастичность GQ, выявление мультиколлинеарности по матрице корреляции экзогенных переменных, а также оценка погрешности прогноза и проверка адекватности модели.***

7. ***Оформление и интерпретация результатов*** моделирования имеет целью переход от информации, полученной в результате машинного эксперимента с моделью, к выводам, касающимся процесса функционирования объекта-оригинала. Результаты моделирования могут быть представлены в виде таблиц, графиков, диаграмм, схем и т.п. В большинстве случаев наиболее простой формой считаются таблицы, хотя ***графики более наглядно иллюстрируют результаты моделирования системы***. Приносите начальнику хорошо оформленные графики, он сможет принять решение и будет вами доволен. И преподавателю удобно проверять решение задач.

Контрольные вопросы.

1. Последовательность разработки математических моделей и решения задач.
2. Назначение эконометрических моделей. Принципы их спецификации.
3. Типы данных в эконометрике.
4. Виды экономико-математических моделей.
5. Построение концептуальной модели.

2. Основы математической статистики

Предполагается, что приступая к изучению эконометрики студенты уже знают основы математической статистики. Практика показывает, что не знают.

Это результат реформ образования. Поэтому необходимо вспомнить основные понятия теории вероятностей и математической статистики. Разделы 2.1 и частично 2.2 основаны на учебнике В.А.Бывшева [2] , так как содержат стандартные определения и формулы.

2.1. Случайная переменная.

Пусть $q_1, q_2, \ldots, q_n$ – набор n чисел, формирующих множество $X = \{ q_1, q_2, \ldots, q_n\}$. Величина x называется переменной, а множество X – множеством её возможных значений, или областью изменения, если x может принимать любые значения q_i из множества X. Переменная величина, все возможные значения которой можно занумеровать, называется дискретной переменной. Если же возможные значения переменной x непрерывно заполняют собой некоторый интервал (a, b), то есть $X = (a, b)$, то такая переменная величина называется непрерывной.

Переменная величина x с областью изменения X называется случайной, если в результате некоторого опыта со случайными элементарными исходами она принимает значение из множества X, которое заранее невозможно предсказать. Случайная величина может быть дискретной или непрерывной.

Теория вероятностей, математическая статистика и эконометрика базируются на предположении о существовании вероятности события

$$p: \quad x = q_i$$

для дискретной случайной величины и

$$p: \quad x \in (q_i , \ q_i + \Delta q)$$

для непрерывной, то есть существует вероятность того, что значение x попадёт в интервал $(q_i , \ q_i + \Delta q)$. Вероятность каждого элементарного исхода пропорциональна Δq .

Полной характеристикой случайной переменной x служит ее ***дифференциальный закон распределения***. Так называется функция $P_x(q)$

скалярного аргумента q, определенная на всей числовой прямой, характеризующая объективную возможность появления в опыте значений q случайной переменной x. Если x – дискретная случайная переменная, то

$$P_x(q_i) = \begin{cases} 0 \text{ при } q_i \notin X, \\ p_i = p(x{=}q_i) \text{ при } q_i \in X. \end{cases}$$

Следовательно, $P_x(q_i)$ – это вероятность появления в опыте значения q_i случайной переменной x. Функция $P_x(q_i)$ именуется ***вероятностной функцией*** **(или *функцией частот, распределением частот)*** дискретной случайной переменной x. Нередко эту функцию задают таблицей, именуемой *таблицей распределения.* Значения функции $P_x(q_i)$ неотрицательны и обладают следующим свойством:

$$\sum_{i=1}^{n} p_i = 1.$$

Дифференциальный закон распределения $P_x(q)$ непрерывной случайной переменной x, если этот закон существует, имеет более сложный смысл:

$$P_x(q) = \lim_{\Delta q \to 0} \frac{P(q \le x < q + \Delta q)}{\Delta q}$$

и называется ***плотностью вероятности***. Как видите, это отношение вероятности попадания x в интервал Δq к величине этого интервала. Значения функции $P_x(q)$ неотрицательны и обладают свойством

$$\int_a^b P_x(q)\,dq = \int_{-\infty}^{+\infty} P_x(q)\,dq = 1,$$

то есть какое-то значение переменная x примет.

2.2. Ожидаемое значение случайной переменной, ее дисперсия и среднее квадратическое отклонение

Важную роль играют две количественные характеристики случайной переменной ***x:*** математическое ожидание (ожидаемое значение) и дисперсия. Ожидаемое значение, которое обычно обозначают μ, m или $E(x)$ находится по формуле

$$\mu = \sum_{i=1}^{n} q_i P_x(q_i) \qquad (2.1)$$

Подчеркнем, что μ – это константа, вокруг которой рассеяны возможные значения q случайной переменной x.

Дисперсия σ^2, $Var(x)$ – это математическое ожидание квадрата отклонения случайной переменной x от её ожидаемого значения:

$$\mathrm{Var}(x) = \sigma^2 = E(x-\mu)^2 = \begin{cases} \sum_{1}^{n} (q_i - \mu)^2 P_x(q_i); \\ \int_{a}^{b} (q-\mu)^2 P_x(q)\, dq. \end{cases} \qquad (2.2)$$

Положительный квадратный корень из дисперсии $\sigma = \sqrt{\mathrm{Var}(x)}$ именуется ***средним квадратическим отклонением (СКО),*** или ***стандартным отклонением.*** Размерности σ и x совпадают. Величина σ (как и σ^2) служит характеристикой неопределенности (изменчивости) x. Формула (2.2) может быть преобразована к виду

$$\sigma^2 = E(x^2) - \mu^2 \qquad (2.3)$$

который часто используется для расчётов вручную. Из формул (2.1) - (2.2) видно, что для отыскания величин μ, σ нужно знать закон распределения $P_x(q)$ случайной переменной x. Часто это закон неизвестен, и тогда можно оценить (приближенно определить) характеристики μ, σ^2 по результатам n независимых наблюдений (опытов) $\{ x_1, x_2, \ldots, x_n \}$. В этом наборе каждая

компонента x_i – это случайная переменная с одним и тем же законом распределения $P_x(q)$, при этом величины x_i являются *независимыми.*

Можно выделить три уровня параметров случайной величины:

1. Результаты замеров реально существующей константы. Примеры: масса протона, период полураспада (или вероятность распада) радиоактивного изотопа, вероятность падения монеты орлом кверху. Эти константы объективно существуют, и, проводя эксперименты, мы можем приближаться к ним, достигая заданной точности. Увеличивая число бросков монеты, мы можем сделать оценку вероятности выпадения орла сколь угодно близкой к 1/2. В экономике и социологии абсолютных констант не существует, нет абсолютно точных взаимозависимостей величин, как в физике. Существуют константы, устанавливаемые правительством, например, ставка налога, но они не являются фундаментальными, могут меняться, и их не оценивают с использованием статистики и эконометрики.

2. Роль абсолютных констант, характеризующих экономику и социальную сферу страны и региона играют параметры генеральных совокупностей – всех доступных значений по стране или региону. Примеры: средний доход домохозяйств, процент заболевших гриппом. В принципе, эти параметры можно измерить во время переписей населения или тотальных проверок (при условии достоверной информации), но такие технологии дороги, а исследуемые параметры непрерывно меняются. Поэтому для оценки параметров природных и социально-экономических объектов служат случайные выборки.

3. Случайные выборки. Было доказано, что если замеры x независимы, то наилучшая оценка математического ожидания $E(x)$ – среднее значение по выборке

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (2.4)$$

а наилучшая оценка дисперсии σ^2

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (2.5)$$

Почему *n-1*, а не *n*? Дело в том, что в формулах (2.2) и (2.3) используется не математическое ожидание *E(x)*, которое мы не знаем, а его оценка – среднее значение $\bar{x}$, вычисляемое по выборке X{ x_1, x_2, …, x_n}, поэтому смещённое относительно *E(x)* и расположенное ближе к центру значений множества { x_1, x_2, …, x_n}. Если делить на *n*, получим заниженную оценку дисперсии. *n* в формулах (2.2), (2.3) и *n-1* в формуле (2.5) – это число степеней свободы, независимых суммируемых переменных. Поскольку $\bar{x}$ вычислено по { x_1, x_2, …, x_n}, одно из выражений в скобках в формуле (2.5) мы можем вычислить, зная *n-1* значений *x*.

Что такое наилучшая оценка, или наилучшая технология оценки (estimator) математического ожидания случайной величины? Каковы её критерии?

1. ***Несмещенность.*** Применяя правильную технологию расчёта, мы не получим в результате обработки серии замеров статистически значимого отклонения от реального значения оцениваемого параметра.

2. ***Эффективность.*** Если в формуле (2.5) мы используем вместо $\bar{x}$ другую величину, полученную по другой формуле, то оценка дисперсии *S* будет больше. Значит, среднее значение обеспечивает наиболее эффективную оценку математического ожидания *E(x).* Эффективность может вступить в противоречие с несмещённостью. Например, исключение переменных из эконометрических моделей может привести к уменьшению дисперсий оцениваемых параметров и к их смещению относительно истинных значений.

3. ***Consistency.*** В российских учебниках это слово переводят как “состоятельность”, но правильнее говорить о ***сходимости.*** Это значит, что

увеличивая количество замеров в серии n, мы можем получить разность оценок исследуемого параметра меньше любого ε (вспомнили матанализ?), то есть наши оценки сходятся к какому-то пределу.

2.3. Законы распределения случайной величины

В технических вузах проводят лабораторную работу: дают студентам одинаковые детали и микрометр. Студенты измеряют размеры деталей и строят гистограммы частотных распределений, то есть считают количество деталей в каждом интервале размеров.

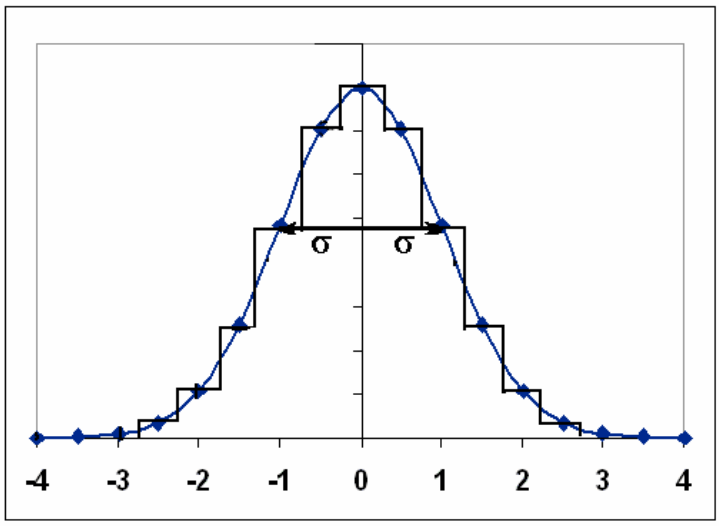


Рис.2.1. Гистограмма частотного распределения и кривая Гаусса с параметрами $E(x) = 0$ и $\sigma = 1$.

Инженеры считают, что размеры деталей подчиняются ***закону нормального распределения (ЗНР)***, выведенного К.Гауссом

$$p(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{(x_i - \mu_x)^2}{2\sigma^2}\right)$$

Как видите, в ***функции Гаусса*** всего два параметра: математическое ожидание μ_x и стандартное отклонение σ, которые сравнительно легко оценить по выборке, используя формулы (2.4) и (2.5). Эти формулы реализованы в Excel в функциях соответственно СРЗНАЧ, ДИСП и СТАНДОТКЛОН, категория

«Статистические». Зная параметры гауссианы, можно вычислить процент деталей в различных диапазонах x (***квантили***), используя таблицы или функцию НОРМРАСП Excel. Поэтому закон нормального распределения широко применяется при проектировании машин и механизмов. Например, можно вычислить количество событий (деталей) в диапазоне $\{E(x) -2\sigma, E(x) +2\sigma\}$. Это примерно 95%, то есть в "хвостах" останется по 2,5%. В данном случае $p = 0{,}95$ – ***доверительная вероятность***, а $\{E(x) -2\sigma, E(x) +2\sigma\}$ - соответствующий ***доверительный интервал***.

На Рисунке 2.2 показано применение функции НОРМРАСП. Площадь левого хвоста гауссианы (Рисунок 2.1) от $-\infty$ до -1,96 (почти 2) равна 0,024997895, то есть 2,5%.

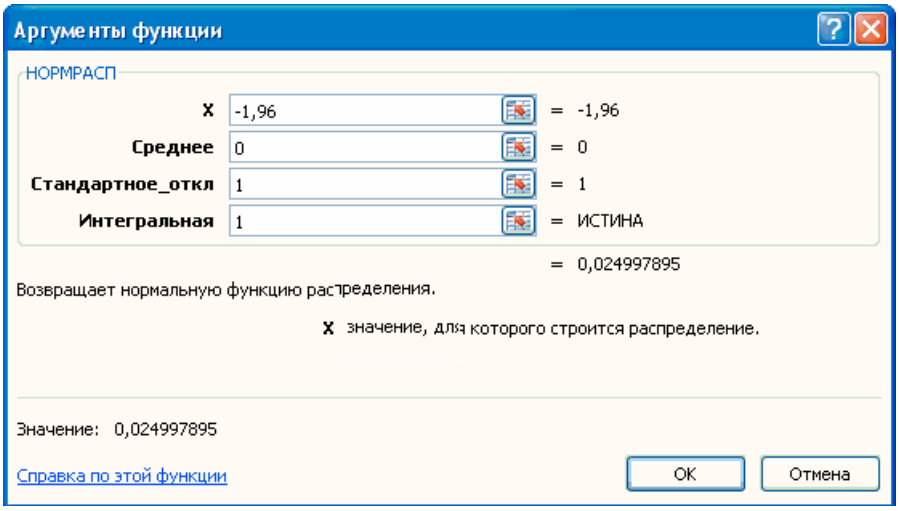


Рис.2.2.

В общем виде это утверждение выглядит следующим образом:

для уровня значимости $\alpha = 1 - p$ доверительный интервал равен $\{E(x) - t_{крит}\sigma;\ E(x) + t_{крит}\sigma\}$, где $t_{крит}$ – критические значения статистики Стьюдента $t = E(x)/\sigma$. В нашем примере α – доля деталей в одном или двух "хвостах". При уменьшении числа замеров надёжность оценки $E(x)$ и дисперсии падают, и доверительный интервал надо расширять. Поэтому критические значения статистики Стьюдента зависят от уровня значимости (доверительной вероятности) и количества замеров (степеней свободы). Распределение Стьюдента $t_{крит}(\alpha,\ n)$ приведено во всех учебниках и

практикумах по математической статистике и эконометрике. В Excel имеется функция СТЬЮДРАСП($t_{крит}$, n, число хвостов (1 или 2)), которая возвращает долю событий в одном или двух “хвостах”. Для практических целей достаточно запомнить, что при числе замеров больше 30 и p=95% $t_{крит}$ примерно равно 2 (при “бесконечном” числе замеров – 1,96). Инженеры используют правило, опирающееся на распределение Гаусса: “за тремя сигмами ничего нет”, то есть количество деталей с размерами, отклоняющимися от среднего более чем на 3σ, ничтожно мало, меньше 0,135% в каждом “хвосте” (сейчас переходят на шестисигмовый уровень надёжности). Разница экономики и техники состоит в том, что 5% невыгодных сделок – не страшно, а 5% или 2,5% (один хвост) заклиненных деталей – это много.

На рисунке 2.3 представлено окно функции СТЬЮДРАСП. Функция вычисляет площадь одного “хвоста” от - ∞ до -2 (или от 2 до ∞) : в данном случае 0,027312522, то есть 2,7%. В окне функции СТЬЮДРАСПОБР на рисунке 2.4 представлено значение t-статистики (отклонение от среднего значения в сигмах), равное 2,042; то есть площадь двух “хвостов” с границами ±2,042σ равна 5%.

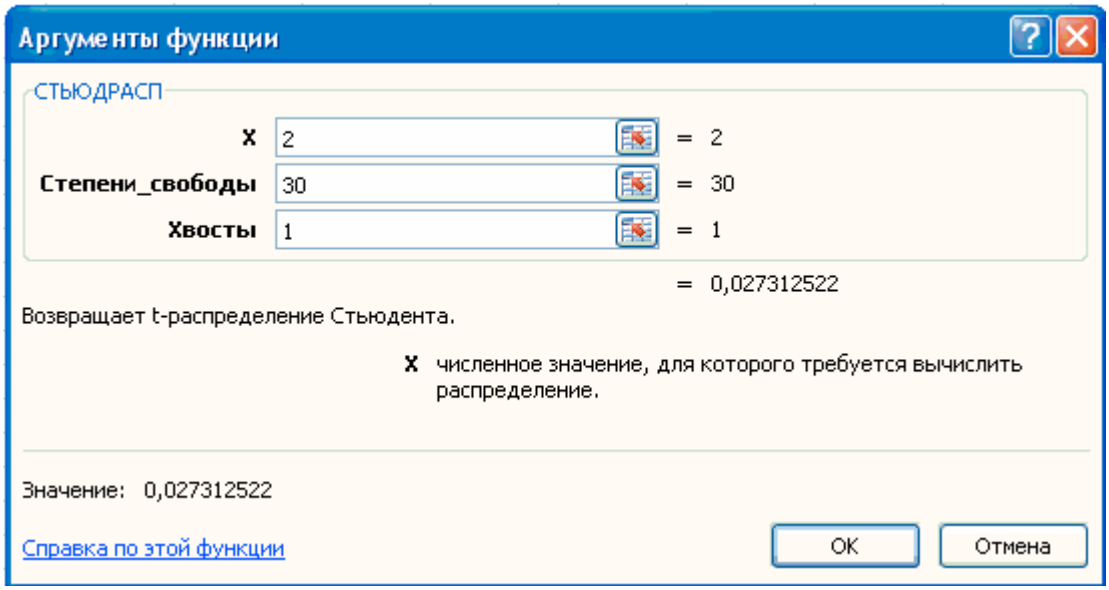


Рис.2.3.

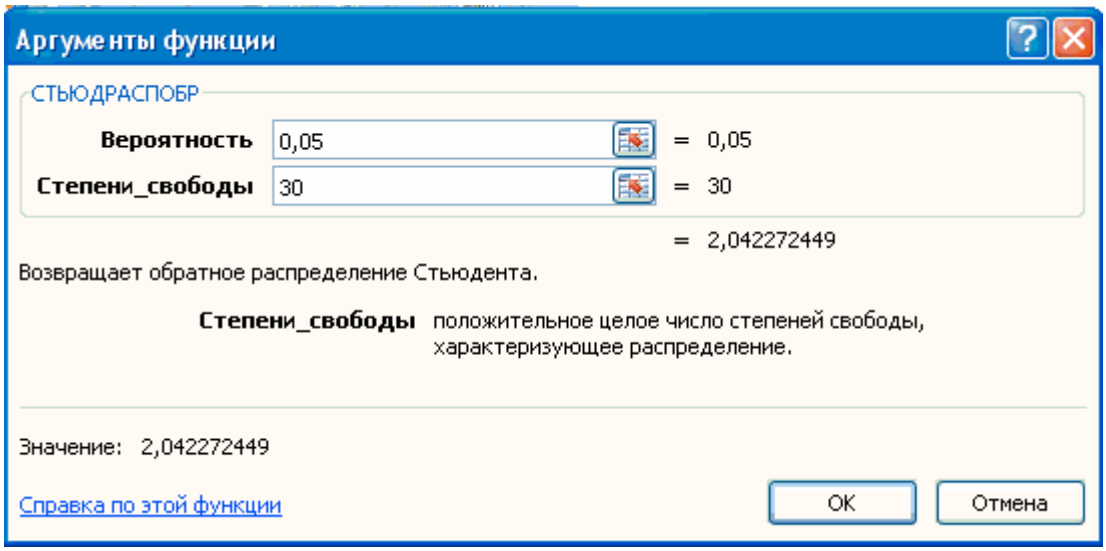


Рис.2.4.

В метеорологии, геохимии, биологии и экономике закон нормального распределения не работает, что связано с когерентностью, то есть взаимной зависимостью событий. Например, изъятие вкладов из банка может многократно превысить средний уровень из-за негативных публикаций или слухов. Для природы и экономики характерны распределения “с толстыми хвостами”, то есть количество аномальных замеров достаточно велико. Известно, что количество природных катастроф в зависимости от количества жертв подчиняется экспоненциальному закону. Успешно используется логнормальное распределение, сводимое к нормальному заменой $\boldsymbol{x_i}$ на $\mathbf{log}(\boldsymbol{x_i})$. Логнормальному распределению подчиняются, по данным автора, микроэлементы и чернобыльские радионуклиды в пробах, количество покупок в магазине в зависимости от их стоимости.

Автор не располагает данными о количестве льготников – пассажиров на городском и пригородном транспорте, но предполагает, что именно незнание законов частотных распределений в социальной сфере привело к бунтам и блокированию трасс при монетизации льгот. Предположим, что количество льготников $\boldsymbol{N}$ в зависимости от стоимости проезда распределено по логнормальному закону (Рис.2.5). По оси абсцисс указано количество поездок на городском транспорте в день.

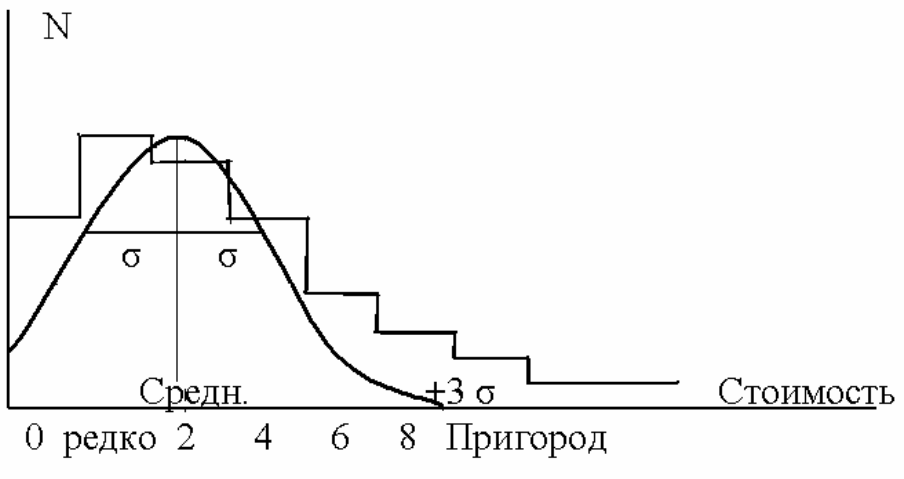


Рис.2.5. Количество льготников ***N*** в зависимости от стоимости проезда.

Видимо, при расчетах компенсаций был использован закон нормального распределения (плавная кривая), компенсировали средние затраты, но больше половины льготников были недовольны. Даже когда добавили σ, потом 2σ, может быть 3σ, то осталось много недовольных: бывшие военные, полярники, милиционеры, которые ездят из пригородов в Москву на заработки. В результате – огромные траты из казны, а льготный проезд из пригородов пришлось оставить.

В математической статистике используются также распределения Пирсона (хи-квадрат), Фишера, Пуассона.

2.4. Взаимосвязь случайных величин

Одна из основных задач эконометрики – выявление взаимосвязи переменных. Количественными оценками взаимосвязи служат ковариация и коэффициент корреляции. Ковариация переменных ***x*** и ***y*** – это ожидаемое значение произведения их отклонений от ожидаемых значений:

$$cov(x,y) = E((x-E(x))*(y-E(y)))$$

Для оценки ковариации по выборке используется формула, аналогичная формуле дисперсии

$$\mathrm{cov}(X,Y)=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

Cov(x,x) – это дисперсия *x*. Коэффициент корреляции – это ковариация, нормированная на стандартные отклонения *x* и *y*:

$$r=\frac{\mathrm{cov}(x,y)}{\sigma_x\sigma_y}$$

Коэффициент корреляции – безразмерная величина, изменяется от –1 до +1; близость к нулю означает отсутствие связи переменных.

Практическое задание

Проведите обработку простого массива данных ***X*** и ***Y***. Вычислите количество данных, используя функцию Excel СЧЁТ(). До 11 мы считать умеем, но реальные таблицы экономических данных могут быть огромными. Вычислите суммы *X* и *Y*, используя функцию Σ, и их средние значения, используя формулу и функцию СРЗНАЧ(). Вычислите квадраты отклонений *X* и *Y* от их средних значений, просуммируйте. Обратите внимание на фиксацию адресов *Xcp* и *Ycp* знаком $. Вычислите дисперсии и среднеквадратические отклонения (СКО) по формулам и через функции ДИСП и СТАНДОТКЛОН. Сравните результаты. Вычислите ковариацию и корреляцию по формулам и через функции КОВАР и КОРРЕЛ.

Таблица 2.1

X	*Y*	*(X-$Xcp)^2*	*(Y-$Ycp)^2*	*(X-$Xcp)*(Y-$Ycp)*
10	12	25	109,2	52,2
11	15	16	55,5	29,8
12	18	9	19,8	13,3
13	16	4	41,6	12,9
14	24	1	2,38	-1,54
15	22	0	0,20	0
16	27	1	20,6	4,54
17	28	4	30,7	11,0
18	25	9	6,47	7,63
19	32	16	91,1	38,1
20	28	25	30,7	27,7

11	11			
165	247	110	408,7	196
15	22,45			
15	22,45			Ковариация
	Sum(X-$Xcp)^2 /(N-1)	11	40,8	19,6
	КОРЕНЬ	3,31	6,39	
	СТАНДОТКЛОН	3,31	6,39	Корреляция
			Cov/Sx/Sy	0,924
			КОРРЕЛ()	0,924

Контрольные вопросы

1. Дифференциальный и интегральный закон распределения случайной величины, виды функций распределения. Что такое “толстые хвосты”?
2. Параметры случайной величины: ожидаемое значение, дисперсия и среднее квадратическое отклонение, коэффициенты ковариации и корреляции.
3. Проверка статистических гипотез, t-статистика Стьюдента, доверительная вероятность и доверительный интервал, критические значения статистики Стьюдента.

3. Регрессионный анализ

Понятие ожидаемого значения случайной переменной позволяет дать точное определение понятия функции регрессии. Пусть случайная переменная y принимает свои значения в опыте вместе с переменной x (случайной или детерминированной — неважно).

Простая (парная) регрессия представляет собой модель, где ожидаемое значение зависимой (объясняемой, эндогенной) переменной y рассматривается как функция одной объясняющей (независимой или управляемой, предопределённой) переменной x, то есть модель вида

$$E(y) = f(x)$$

Множественная регрессия представляет собой модель, где ожидаемое значение зависимой переменной y рассматривается как функция многих объясняющих переменных, то есть модель вида

$$E(y) = f(x_1, x_2, \ldots, x_n)$$

Случайную переменную *y* формируют функция $f(x)$ и случайная величина u (uncertainty, disturbance term, возмущение) с ожидаемым значением, равным нулю:

$$y = f(x) + u$$

Такое разложение случайной переменной y именуется ***регрессионным анализом переменной y.***

Предполагается, что $f(x)$ отражает идеальную закономерность, на которую накладываются неучтённые факторы или ошибки измерения. В физике это так, а в экономике – нет. В физике параметрами функции $f(x)$ являются константы, которые надо оценить по результатам измерений (скорость света, масса протона, период полураспада радиоактивного изотопа). В экономике измеряемые величины (ВВП, количество населения) и их взаимосвязи постоянно меняются, поэтому нет фундаментальных констант. Тем не менее, эконометрика переняла математический аппарат, разработанный для физики, и мы его будем использовать.

Регрессионные модели, которые наиболее часто используются в эконометрике:

1) ***Линейная*** $y = a + bx + u$; употребляется наиболее часто, остальные функции стараются преобразовать к линейному виду, линеаризовать.

Регрессии, нелинейные относительно включённых в анализ объясняющих переменных:

2) ***Полином*** второй, редко третьей степени $y = a + bx + cx^2 + u$.

3) Равносторонняя ***гипербола*** $y = a + b/x + u$.

Эти модели сводятся к линейным заменой переменных: $z = x^2$ для полинома и $z = 1/x$ для гиперболы.

К нелинейным регрессиям по оцениваемым параметрам относятся:

4) ***Степенная*** $y = ax^b\varepsilon$;

5) ***Показательная*** $y = ab^x\varepsilon$;

6) ***Экспоненциальная*** $y = e^{a+bx}\varepsilon$.

Здесь $\varepsilon = 1 + u$. Эти модели могут быть линеаризованы логарифмированием.

Следует отметить разницу между идеальной закономерностью, которую для линейной модели обычно записывают

$$y = \alpha + \beta x + u$$

и оценённой регрессионной моделью

$$y = a + bx + e,$$

а также возмущением u и отклонением, или ошибкой e. Предполагается, что α и β являются реальными константами, а a и b служат их оценками. В экономике констант нет, но математический аппарат сохраняется. Возмущение u – это отклонение реального замера от идеальной закономерности $\alpha+\beta x$, которую мы не знаем. Значит, u мы тоже не знаем, но можем делать предположение о его свойствах. Ошибка e – это разность между реальным y и его значением, оценённым по формуле $a + bx$; она служит оценкой u.

Коэффициенты b и a можно вычислить по формулам

$$b = \frac{Cov(X,Y)}{Var(X)}; \qquad a = \bar{Y} - b\,\bar{X}$$

3.1. Метод наименьших квадратов

Для оценки параметров линейной или линеаризованной модели применяется ***метод наименьших квадратов (МНК).*** Суть метода состоит в следующем: к реальным данным подбирается функция и её параметры, чтобы разности (отклонения, остатки) между реальными и вычисленными значениями y были минимальны. Но разностей много, поэтому минимизируется сумма квадратов этих разностей:

$$L = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

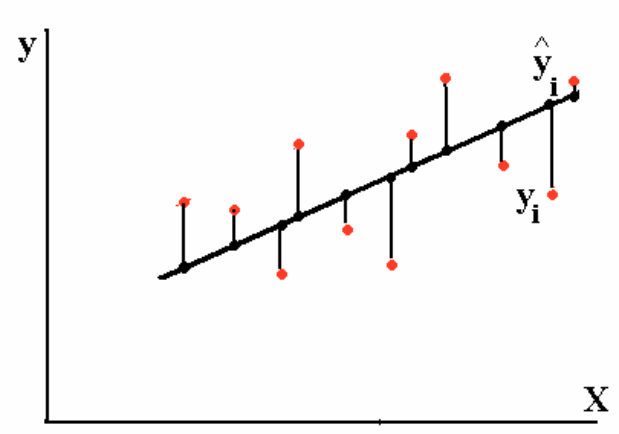


Рис.3.1. Отклонения реальных y от оценённой функции регрессии.

Как правило, вычисления проводятся на компьютере с использованием различных сервисов и программ. Далее мы рассмотрим технологию МНК, которую использовали при ручном вычислении параметров парной линейной регрессии.

Сумма квадратов остатков, зависящая от параметров a и b

$$L(a,b) = \sum_{i=1}^{n} (a + bx_i - y_i)^2 \to \min$$

где n – количество измерений. Эта функция достигает минимума в точке, где её частные производные по a и по b равны нулю:

$$\frac{\partial L}{\partial a} = 2\sum_{i=1}^{n} (a + bx_i - y_i) = 2na + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} y_i = 0$$

$$\frac{\partial L}{\partial b} = 2\sum_{i=1}^{n} (a + bx_i - y_i)x_i = 2a\sum_{i=1}^{n} x_i + 2b\sum_{i=1}^{n} x^2 - 2\sum_{i=1}^{n} x_i y_i = 0$$

или

$$an + b\Sigma x = \Sigma y$$
$$a\Sigma x + b\Sigma x^2 = \Sigma xy$$

Это называется ***система нормальных уравнений***. В ней два уравнения и два неизвестных ***a*** и ***b***, а коэффициенты получаются суммированием ***x, y*** и т.д.

Решать её можно разными способами. В данном случае использован сервис Excel *Поиск решения* для настройки линейной модели по данным ***X*** и ***Y***, представленным в Таблице 3.1. Коэффициенты системы нормальных уравнений расположены в виде матрицы (верхние строки таблицы 3.2), неизвестные ***a*** и ***b*** задаются произвольно и умножаются на коэффициенты (нижние строки). В окне *Поиска решения* задаются: *Целевая ячейка* – первая сумма, *Значение* равно 247 (Σy), *Изменяя ячейки* – ***a*** и ***b***, *Ограничения*: вторая сумма равна 3901 (Σxy). Исходные данные *X* и *Y* приведены в Таблице 3.1. результаты расчёта в Таблице 3.2.

Таблица 3.1.

X	*Y*	X^2	*XY*
10	12	100	120
11	15	121	165
12	18	144	216
13	16	169	208
14	24	196	336
15	22	225	330
16	27	256	432
17	28	289	476
18	25	324	450
19	32	361	608
20	28	400	560
Суммы 165	247	2585	3901

Таблица 3.2.

11	165	247
165	2585	3901
a	b	
-4,27	1,78	
		Суммы по строкам
-47,00	294,00	246,9999
-705,00	4606,00	3901

Теперь можно построить функцию регрессии $\hat{\boldsymbol{Y}}$, сравнить её с ***Y*** и использовать для прогноза.

В принципе, МНК с *Поиском решения* можно использовать непосредственно. Для этого надо задать произвольные коэффициенты *a* и *b*, построить по ним функцию $\hat{\boldsymbol{Y}} = \boldsymbol{a} + \boldsymbol{bX}$, вычислить остатки $\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$ и их квадраты, сумму e^2.

В окне *Поиска решения* установить *Целевая ячейка* $\Sigma \boldsymbol{e}^2$ *минимум*, *Изменяя ячейки* *a* и *b*, ограничений нет.

Таблица 3.3.

X	***Y***	$\hat{Y}$	***Остатки e***	e^2
10	12	13,545	-1,545	2,388
11	15	15,327	-0,327	0,107
12	18	17,109	0,890	0,793
13	16	18,890	-2,890	8,357
14	24	20,672	3,327	11,070
15	22	22,454	-0,454	0,206
16	27	24,236	2,763	7,637
17	28	26,018	1,981	3,927
18	25	27,8	-2,8	7,840
19	32	29,581	2,418	5,847
20	28	31,363	-3,363	11,314
		Суммы	1E-06	59,490
Дисперсии	40,872	34,923	5,949	
R^2	0,854		***a***	***b***
F	52,833		-4,27	1,78

Этот метод описан более подробно в разделе 4.4.

3.2. Оценка значимости параметров линейной регрессии

Критерии качества модели: коэффициент детерминации и статистика Фишера.

Коэффициент детерминации

$$R^2 = 1 - \frac{\Sigma e^2}{\Sigma (Y - \bar{Y})^2} \qquad (3.1)$$

Для линейной модели он совпадает с квадратом коэффициента корреляции, но пригоден и для нелинейных моделей. На Рисунке 3.2. показана аппроксимация параболой. Коэффициент корреляции близок к нулю, а коэффициент детерминации – к единице, так как дисперсия остатков существенно меньше дисперсии ***Y***. Это говорит о высоком качестве модели.

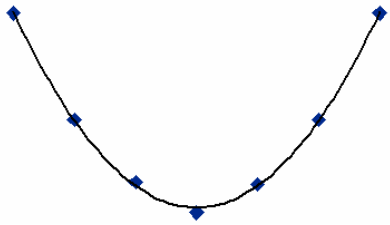


Рис.3.2.

Формула (3.1) легко преобразуется

$$R^2 = 1 - \frac{\Sigma e^2}{\Sigma(Y-\bar{Y})^2} = 1 - \frac{\Sigma(e-\bar{e})^2/(n-1)}{\Sigma(Y-\bar{Y})^2/(n-1)} = 1 - \frac{ДИСП(e)}{ДИСП(Y)} \qquad (3.2)$$

где *ДИСП* – функция Excel *Дисперсия.* Вообще говоря, несмещённой оценкой дисперсии остатков парной регрессии является

$$S_e^{\,2} = \frac{\Sigma e^2}{n-2}$$

но функция *ДИСП.В* делит на *(n-1)*, и в данном случае всё получается правильно. В данном случае R^2 = 0,854, что соответствует коэффициенту корреляции 0,924, то есть имеет место сильное влияние переменной ***X*** на ***Y***.

Дисперсия суммы двух независимых переменных равна сумме их дисперсий. В Таблице вы видите, что $ДИСП(Y)=ДИСП(\hat{Y}) + ДИСП(e)$.
Надо сказать, что $\Sigma(Y - Ycp)^2$ обозначают *TSS (Total Squared Sum)*; в российских учебниках $\Sigma(\hat{Y} - \hat{Y}cp)^2$ обозначают *RSS*, а Σe^2 *ESS (Error Squared Sum;* в английских учебниках $\Sigma(\hat{Y} - \hat{Y}cp)^2$ обозначают *ESS (Explained Squared Sum)* а Σe^2 *RSS (Residual Squared Sum)*. Поэтому мы не будем пользоваться этими обозначениями.

Оценка значимости уравнения регрессии в целом даётся с помощью ***F***-критерия Фишера. При этом проверяется нулевая гипотеза, что коэффициент регрессии β равен нулю и, следовательно, фактор ***X*** не оказывает влияния на результат ***Y***. Давно составлены таблицы критических значений ***F***-статистики в зависимости от числа измерений *n*, числа степеней свободы, или количества независимых переменных *m* и уровня значимости α.

Статистика Фишера равна частному от деления дисперсии $\hat{Y}$, или факторной дисперсии, и дисперсии остатков, вычисленных с учётом числа степеней свободы: 1 для $\hat{Y}$ и n-2 для остатков.

Для множественной регрессии и полиномиальной, которую можно преобразовать в множественную, число степеней свободы $\hat{Y}$ равно числу

независимых переменных *m*, а число степеней свободы остатков равно *n-m-1*. Статистику Фишера удобно вычислять через коэффициент детерминации:

$$F = \frac{R^2}{1-R^2} \times \frac{n-m-1}{m} \qquad (3.3)$$

Чем больше статистика Фишера, тем лучше прогнозы, сделанные с использованием модели. Из формулы (3.3) следует, что *F* возрастает с ростом R^2 и числа измерений, но уменьшается при увеличении числа влияющих переменных, то есть надо аккуратно подходить к включению в модель новых влияющих переменных, а также не использовать для аппроксимации полиномы высоких степеней. Полезно помнить, что при уровне значимости α=0,05, то есть при доверительной вероятности 95% и количестве замеров более 15 критическое значение *F* для парной регрессии около 4,2 , а при *m=4* около 3. Начиная с этих значений *F* можно говорить о существовании влияния регрессоров на эндогенную переменную. Таблицы критических значений *F* есть во всех книгах по мат.статистике и эконометрике, поэтому в этой книге они не приводятся. Их можно вычислить в Excel с помощью функции FРАСПОБР с аргументами: уровень значимости (здесь α=0,05); число регрессоров *m; N-m-1;* где *N* число измерений.

Коэффициенты линейного уравнения регрессии $\boldsymbol{b_i}$ имеют экономический смысл: это предельные функции, или производные эндогенной переменной по влияющим:

$$b_i = \frac{\Delta Y}{\Delta X_i}$$

В случае парной регрессии это однозначно, в множественной регрессии всё сложнее из-за взаимного влияния регрессоров.

Для оценки погрешностей коэффициентов уравнения парной линейной регрессии $\boldsymbol{\hat{Y} = a + bx}$ используются выражения

$$S^2_{ост} = \frac{\Sigma ост^2}{n-2}; \qquad S_b = \frac{S_{ост}}{\sqrt{\Sigma(X-\bar{X})^2}} = \frac{S_{ост}}{\sqrt{(n-1)S_X^2}}; \qquad S_a = S_{ост}\frac{\sqrt{\Sigma X^2}}{(n-1)S_X}$$

где S – выборочные оценки стандартных отклонений σ. Для принятия гипотезы о влиянии регрессора на эндогенную переменную используются таблицы критических значений t-статистики Стьюдента. Для b $t=b/S_b$. Предполагается, что при числе измерений больше 20 истинные значения коэффициентов уравнения регрессии α и β лежат в интервалах $\{a-2S_a, b+2S_b\}$ и $\{b-2S_b, b+2S_a\}$ с доверительной вероятностью 95%.

3.3. Матричный метод МНК

Матричный метод МНК основан на представлении множеств **X, Y,** остатков **E** и параметров линейной модели **B** в виде векторов, над которыми затем проводятся операции. Векторное представление модели

$$\mathbf{Y} = \mathbf{B} * \mathbf{X} + \mathbf{E}$$

где

Y	**B**	**X**	**E**
y_1		$1 \quad x_1$	e_1
y_2		$1 \quad x_2$	e_2
.	a	.	.
.	b	.	.
.		.	.
y_n		$1 \quad x_n$	e_n

Эту модель, записанную в векторном виде или в виде системы линейных уравнений, называют ***схемой Гаусса-Маркова***.

Условие МНК Σe^2 **-> min** , или в матричном виде $(\mathbf{Y-XB})^T(\mathbf{Y-XB})$ **-> min**.

T означает транспонирование, то есть преобразование столбца в строку.

Решением является вектор **B**:

$$\mathbf{B} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Здесь **-1** означает обращение матрицы. Транспонирование и обращение матриц можно выполнять в Excel, используя функции ТРАНСП и МОБР.

3.4. Теорема Гаусса-Маркова

Согласно ***теореме Гаусса-Маркова***, Метод наименьших квадратов, приведённый к линейному преобразованию матриц или к системе линейных уравнений, обеспечивает наилучшую несмещенную, эффективную и сходящуюся к пределу (“состоятельную”) оценку вектора параметров, т.е. наилучшее качество линейной модели, если соблюдаются условия (по [1]):

1. Линейная модель соответствует действительности.
2. Существует дисперсия регрессора.
3. Математическое ожидание возмущения равно нулю: $E(u_i) = 0$.
4. Возмущение имеет нормальное распределение.
5. Равенство ожидаемых значений дисперсий возмущений в разных диапазонах ***X:*** $E(u^2) = Const$. Это свойство называется ***гомоскедастичность***, его несоблюдние – ***гетероскедастичность***. Отклонение от гомоскедастичности проверяется по тесту ***Голдфелда-Квандта***

$$GQ = \Sigma e_1^2 / \Sigma e_2^2$$

где Σe_1^2 **и** Σe_2^2 – суммы квадратов остатков (отклонений) в первой и последней трети (или в половинах) диапазона ***X*****; *большая сумма делится на меньшую!!!; GQ*** сравнивают с критерием Фишера для заданных уровня значимости и количества измерений; гипотеза о гомоскедастичности принимается при $GQ < 4{,}35$.

6. Отсутствие ***автокорреляции***, т.е. взаимозависимости возмущений. Её оценивают, вычисляя ***статистику Дарбина-Уотсона остатков e:***

$$DW = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=2}^{n} e^2}$$

для которой вычислены критические значения при различных уровнях значимости и числе измерений. Приблизительно DW=0…1 означает положительную автокорреляцию, 3…4 отрицательную автокорреляцию, DW=1,5…2,5 позволяет принять гипотезу об отсутствии автокорреляции,

DW=1…1,5 и *DW*=2,5…3 не позволяют принять гипотезу о наличии или отсутствии автокорреляции. Наличие автокорреляции означает, что аппроксимирующая функция подобрана неверно, или же требуется применение других методов и моделей. Автокорреляция разобрана в главе 8.

Статистику Дарбина-Уотсона можно вычислить по формуле

$$DW = 2(1-R_{авт}),$$

где ***Rавт*** - коффициент автокорреляции, вычисляемый с помощью функции КОРРЕЛ: задать в окне *Массив1* диапазон остатков с номерами 1 : *n*-1, а в окне *Массив2* диапазон 2 : *n.*

Понятия “гетероскедастичность” и “автокорреляция” актуальны, если массивы данных упорядочены, что имеет место для временных рядов. “Пространственные” данные можно искусственно упорядочить, например, отсортировав их по возрастанию какой-либо переменной; при этом можно выявить кластеры с аномальной дисперсией остатков, что может означать неоднородность выборки или неадекватность модели.

Считается, что гетероскедастичность может привести к снижению эффективности оценок коэффициентов, и надо её искусственно подавлять: делить остатки в таблице 3.3 на их стандартные отклонения в диапазонах, а затем минимизировать сумму их квадратов. Эта технология называется Взвешенный метод наименьших квадратов (ВМНК) и обычно используется в матричном варианте МНК (раздел 3.3). При обнаружении автокорреляции остатков применяется ***Обобщённый метод наименьших квадратов ОМНК***, основанный на преобразовании матриц, но с учётом корреляций остатков.
Целесообразность применения ВМНК и ОМНК обсуждается в разделе 5.1.

Контрольные вопросы.

1. Общий вид уравнений парной и множественной регрессии.
2. Нелинейные уравнения регрессии.
3. Формулы для вычисления коэффициентов парной линейной регрессии и их погрешностей.

4. Метод наименьших квадратов (МНК) и система нормальных уравнений парной линейной регрессии.

5. Схема Гаусса-Маркова и Матричный метод МНК.

6. Теорема Гаусса-Маркова: формулировка и условия.

7. Показатели качества эконометрической модели: коэффициент детерминации R^2, статистика Фишера *F, t*-статистики Стьюдента для коэффициентов уравнений.

8. Показатели качества эконометрической модели: тест Дарбина-Уотсона на автокорреляцию *DW*, тест Голдфелда-Квандта на гетероскедастичность *GQ*.

9. Гетероскедастичность случайного возмущения. Причины, последствия.

10. Что такое ВМНК и ОМНК, и когда они применяется.

4. Оценка коэффициентов парной регрессии на компьютере

Для решения задачи прогнозирования требуется по экспериментальным точкам провести гладкую кривую (или, в общем случае, многомерную поверхность), то есть выявить функциональную зависимость (***аппроксимация***), продлить ее в неизученную область (***экстраполяция***) и оценить надежность прогноза.

Для конкретизации задачи будем исследовать продажу мороженого в зависимости от температуры воздуха в диапазоне $0^{o} – 30^{o}$: Постановка задачи: по имеющимся данным о продажах в диапазоне температур $0^{o}…20^{o}$ спрогнозировать уровень и диапазон продаж при температуре 30^{o}. Обобщённая модель:

$$y = f(x) + u$$

где x – независимая (***экзогенная***) переменная: температура,

y - зависимая (***эндогенная***) переменная: продажа мороженого.

В таблице 4.1 и на рисунке 4.1 приведены известные нам объемы продаж при различных температурах в диапазоне $0^{o} – 20^{o}$, требуется дать прогноз на диапазон $21^{o} – 30^{o}$. Рассмотрим две модели и 4 способа решения задачи.

Таблица 4.1

Температура *X*	Продажи *Y*
0	11
1	8
2	9
3	13
4	6
5	10
6	11
7	6
8	7
9	13
10	12
11	15
12	18
13	16
14	24
15	22
16	27
17	28
18	25
19	32
20	28

Рис.4.1.

В данной задаче температуры упорядочены изначально. В противном случае обычно требуется провести предварительную сортировку данных по независимой (экзогенной) переменной: выделить таблицу, в меню *Данные* – *Сортировка* – указать столбец независимой переменной – *ОК*.

4.1. Построение диаграмм и спецификация моделей.

Решение эконометрической задачи начинается со ***спецификации модели***, то есть выявления функции регрессии и особенностей возмущений. Для этого надо построить диаграмму: выделить оба столбца данных и построить диаграмму *Точечная*, что позволит сразу правильно расположить данные по оси абсцисс и оценить корреляцию ***X*** и ***Y*** или ее отсутствие. Диаграмма позволяет оценить вид аппроксимирующей функции, может быть, различной в разных диапазонах ***X***, и увидеть точки, выпадающие из закономерности. Эти точки надо удалить из выборки и рассматривать отдельно. В данном примере

аномальные значения ***Y*** могут быть связаны с праздничными днями и премиями.

Если функция $\hat{Y} = f(X)$ нелинейная, то ее, как правило, надо линеаризовать, заменив значения ***X*** и ***Y*** на их логарифмы, квадраты, квадратные корни или более сложные функции, а после решения задачи провести обратное преобразование полученной аппроксимирующей функции. Если точки уплотняются в левой части диаграммы, то целесообразно заменить значения ***X***, а может и ***Y*** их логарифмами. Целесообразно построить гистограммы частотных распределений ***X*** и *Y*, и при распределении с "толстым хвостом" провести логарифмирование.

В данном случае можно увидеть на диаграмме, что от 0^o до 10^o продажи не возрастают, а после 10^o – возрастают. Можно построить две модели с расчетом коэффициентов по различным диапазонам:

1) Линейная $Y = a + b\,X + u$ в диапазоне от 10^o до 20^o**;**
2) Парабола $Y = a + b\,X + c\,X^2 + u$ в диапазоне от 0^o до 20^o**.**

(Любую функцию в некотором диапазоне можно достаточно точно представить многочленом).

В первом случае мы отбрасываем половину исходных данных и считаем, что до 10^o одна закономерность, а свыше 10^o – другая. Во втором случае мы используем для настройки модели все данные. Выбор модели зависит от теоретических закономерностей и от личного опыта, а также от результатов эконометрических тестов, например, теста Чоу.

Параметры моделей и прогноз можно получить, построив диаграммы.

Модель1:

- выделите оба столбца в диапазоне температур 10^o – 20^o, постройте диаграмму типа *Точечная*;
- щелкните правой клавишей мыши по любой точке, в появившемся окне щёлкните *Добавить линию тренда*, установите *Тип – Линейная*, выберите *Параметры*, установите *Прогноз вперед на 10 единиц* и флажки *Показывать уравнение на диаграмме, Поместить на диаграмму величину достоверности*

(R^2) – ОК. На диаграмме появятся уравнение регрессии и коэффициент детерминации.

Модель 2:

- выделите оба столбца в диапазоне температур 0^{o} – 20^{o}, постройте диаграмму типа *Точечная*;
- щелкните правой клавишей мыши по любой точке, в появившемся окне щёлкните *Добавить линию тренда*, установите *Тип – Полиномиальная*, *Степень 2*, выберите *Параметры*, установите *Прогноз вперед на 10 единиц* и флажки *Показывать уравнение на диаграмме, Поместить на диаграмму величину достоверности (R^2) – ОК*.

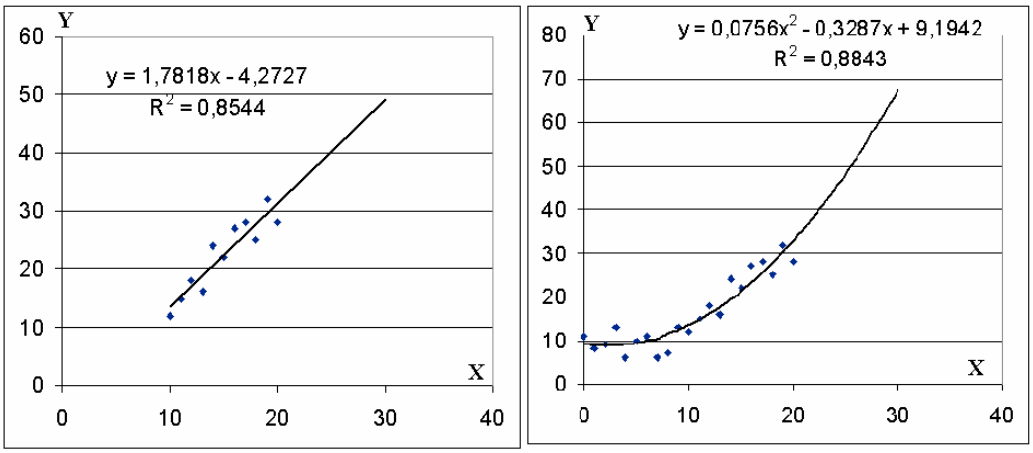


Рис.4.1. Диаграммы линейной и параболической моделей.

Индекс детерминации R^2 даёт представление о надёжности модели. Критерии качества модели и её параметров мы обсудим позднее.

Возможно использование других функций: степенной, экспоненты, логарифма. Обратите внимание, что модели дают различные прогнозы на 30^{o}.

Данный пример служит иллюстрацией того, что в разных диапазонах переменных могут действовать разные закономерности. Из житейских соображений следует, что параболу нельзя продлевать в область отрицательных значений температуры: продажи мороженого зимой не вырастут. Обе функции нельзя экстраполировать на 40 и более градусов. Для каждой закономерности существует диапазон значений, или область определения экзогенной переменной.

Возникает вопрос: можно ли повысить качество модели, увеличив степень полинома? Это возможно, но только при очень большом количестве измерений и высоком (>0,95) коэффициенте детерминации. Попробуем увеличить степень полинома в нашем примере до 4, а последние 4 замера (17^o - 20^o) используем не для настройки модели, а для проверки её адекватности. Получим результат: Рисунок 4.2А. Получился высокий $\boldsymbol{R^2}$, последние 4 значения $\boldsymbol{Y}$ предсказаны неплохо. Решим, что модель качественная и адекватная, используем её для прогнозирования. Получим абсолютно безобразный прогноз: Рисунок 4.2 Б.

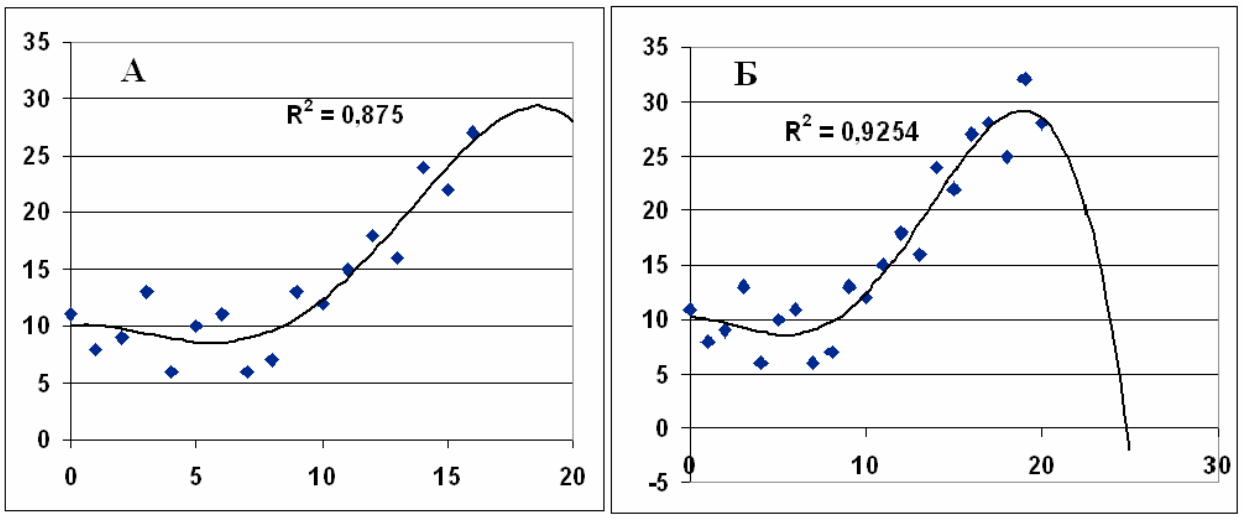


Рис. 4.2.

Это произошло из-за большого количества тесно связанных влияющих переменных – $\boldsymbol{x}$ в различных степенях, что привело к высокой дисперсии (большим ошибкам) коэффициентов. В диапазоне настройки эти ошибки друг друга компенсируют, а в области прогноза – нет.

4.2. Функция ЛИНЕЙН

Параметры линейной регрессии можно определить с помощью встроенной статистической функции ЛИНЕЙН. Порядок вычисления следующий:

- Ввод исходных данных;
- Выделите область пустых ячеек 5x2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики или область 1x2 – для получения только оценок коэффициентов регрессии;

- Активизируйте *Мастер функций* – щелкните *fx* на панели инструментов или в главном меню выберите *Вставка – Функция*;
- В окне *Категория* выберите *Статистические,* в окне *Функция* – ЛИНЕЙН. Щелкните ОК.
- Заполните аргументы функции:
  - *Известные значения y* – диапазон, содержащий данные результативного признака;
  - *Известные значения_x* – диапазон, содержащий данные факторов независимого признака;
  - *Константа* – логическое значение, которое указывает на наличие или отсутствие свободного члена в уравнении: если *Константа* = 1, то свободный член рассчитывается обычным способом, если *Константа* = 0, то свободный член равен 0;
  - *Статистика* – логическое значение, которое указывает, выводить дополнительную информацию по регрессионному анализу (= 1) или нет (=0);
- Нажмите комбинацию клавиш CTRL – SHIFT – ENTER. Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей таблице:

Таблица 4.2.

Коэффициент *b*	Коэффициент *a*
Среднеквадратическое отклонение *b*	Среднеквадратическое отклонение *a*
Индекс детерминации R^2	Среднеквадратическое отклонение остатков
F – статистика	Число степеней свободы остатков
Регрессионная сумма квадратов $\Sigma(\hat{Y} - \hat{Y}_{средн.})^2$	Сумма квадратов остатков $\Sigma(Y - \hat{Y})^2$

Полученный результат:

1,7818	-4,2727
0,2451	3,7578
0,8544	2,5710
52,833	9
349,23	59,490

Если случайно щёлкнули ОК, нажмите на клавишу F2, а затем – на комбинацию клавиш CTRL – SHIFT – ENTER.

Для вычисления параметров показательной функции $\boldsymbol{Y = ab^x}$ в Excel применяется встроенная статистическая функция ЛГФПРИБЛ. Порядок вычислений аналогичен применению функции ЛИНЕЙН.

Как видите, полученные коэффициенты ***a***, ***b*** и индекс детерминации $\boldsymbol{R^2}$ совпадают с результатами их оценки с помощью диаграммы. Кроме того, получены погрешности коэффициентов ***a***, ***b***, стандартное отклонение ***Y***, число степеней свободы остатков ($\boldsymbol{n}$-2 = 9), сумма квадратов остатков, регрессионная сумма квадратов = $\boldsymbol{\Sigma(\hat{Y} - \hat{Y}средн.)^2}$ и статистика Фишера.

4.3. Сервис *Регрессия*

Ещё больше информации даёт сервис *Регрессия* из *Пакета анализа* Excel. Для его запуска надо щелкнуть в Меню Excel 2003 и более ранних версий *Сервис – Анализ данных – Регрессия*. (Если *Анализ данных* в меню *Сервиса* не появится, щелкните *Надстройки* и установите флажок *Пакет анализа*). В Excel 2007 и 2010 *Пакет анализа* вызывается в разделе Меню *Данные*. Если *Анализ данных* не виден, установить его: *Файл – Параметры – Надстройки – Параметры Excel применить – Пакет анализа*. Укажите диапазоны ячеек *Y* и *X* и на какой лист выводить результаты – на новый или на тот же. В этом случае надо указать достаточно большой диапазон ячеек для вывода. Поставьте флажок *Метка*, если выделили X и Y с заголовками.

Сервис *Регрессия* выводит все статистические характеристики модели с соответствующими надписями. Сервис *Регрессия* может применяться для линейных или линеаризованных моделей.

Оценка параметров Модели 1 с использованием сервиса *Регрессия:*

Таблица 4.3.

Регрессионная статистика	
Множественный R	0,9243
R-квадрат	0,8544
Нормированный R-квадрат	0,8382
Стандартная ошибка	2,5710
Наблюдения	11

Квадратный корень из коэффициента детерминации. Для линейной модели – коэффициент корреляции

Коэффициент детерминации

$$R^2_{норм} = 1 - \frac{\Sigma ост.^2}{\Sigma (Y - \bar{Y})^2} \times \frac{n-1}{n-m-1}$$

Стандартное отклонение остатков

Количество наблюдений ***n***

Дисперсионный анализ	Число степеней свободы сумм	Суммы квадратов	Дисперсия на одну степень свободы		
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия (Y^)	1	349,23	(349,23/1=) 349,23	52,83	4,72E-05
Остаток	9	59,490	(59,49 / 9=) 6,610		
Итого (Y)	10	408,72	(Var(Y) =) 40,87		

	Коэффи-циенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	-4,2727	3,7578	-1,1370	0,2849	-12,773	4,228
X	1,7818	0,2451	7,2686	4,72051E-05	1,227	2,336

Стандартные надписи и дополнительные пояснения позволяют быстро разобраться в таблице результатов сервиса *Регрессия*. *Коэффициент детерминации* (здесь R-квадрат), статистика Фишера *F* и *t-статистика* Стьюдента разобраны в разделах 2.3 и 3.2. Осталось добавить про *Значимость F* и *P-Значение*. Соответствующие числа в таблице означают вероятности принятия неверных гипотез относительно наличия влияния всех переменных на *Y* (*Значимость F*) и каждой экзогенной переменной в отдельности (*P-Значение*). В данном случае имеется одна влияющая переменная, поэтому значимости *F* и *b* совпадают. Погрешность *b* равна 14%, *t*-статистика *b* высокая, вероятность того, что $b \le 0$, то есть продажи не зависят от температуры, ничтожно мала (*P-Значение* = 4,72051E-05). Погрешность *a* равна 88%, *t*-статистика низкая, и вероятность того, что *a* окажется больше

нуля, равна 28,5% (*P*-значение). В разделе *Дисперсионный анализ* выведены: регрессионная сумма квадратов $\Sigma(\hat{Y} - \hat{Y}средн.)^2$, здесь равная 349,23 , и соответствующая дисперсия для одной степени свободы (один *x*), а также сумма квадратов остатков, здесь 59,49 , дисперсия остатков 6,61 и соответствующие величины для эндогенной переменной *Y*.

Сервис *Регрессия* можно применять к линеаризованным моделям, а также считая ***x*** в разных степенях в полиноме как самостоятельные экзогенные переменные, то есть сводя полиномиальную модель к модели множественной регрессии, которая рассмотрена далее. Пример: оценка параметров Модели 2 (парабола) с помощью сервиса *Регрессия:*

Таблица 4.4.

	Температура	Продажи	ВЫВОД ИТОГОВ				
x^2	*x*	*y*					
0	0	11	*Регрессионная статистика*				
1	1	8	Множественный R	0,940			
4	2	9	R-квадрат	0,884			
9	3	13	Нормированный R-квадрат	0,871			
16	4	6	Стандартная ошибка	2,961			
25	5	10	Наблюдения	21			
36	6	11					
49	7	6	Дисперсионный анализ				
64	8	7		*df*	*SS*	*MS*	*F*
81	9	13	Регрессия	2	1205	602	68,7
100	10	12	Остаток	18	157,8	8,7	
121	11	15	Итого	20	1363		
144	12	18					
169	13	16					
....	....	....					
361	19	32					
400	20	28					

	Коэффициенты	*СКО*	*t*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	9,1942	1,7674	5,2020	0,0001	5,48	12,90
x^2	0,0756	0,0198	3,8235	0,0012	0,034	0,117
x	-0,3287	0,4095	-0,8025	0,4327	-1,189	0,531

4.4. Сервис *Поиск решения (Solver)*

Использование сервиса *Поиск решения* позволяет наглядно продемонстрировать суть метода наименьших квадратов (МНК). Вызывается он так же, как и *Анализ данных*: в Excel-2003 и более ранних версиях через меню Сервис (если не вызывается, то Сервис-Надстройки) ; в Excel-2007 и 2010 в меню Данные (если не вызывается, то Пуск – Параметры Excel – Надстройки – Перейти). Схема расчетов та же, что и в задачах математического программирования:

- задать произвольные коэффициенты аппроксимирующей функции $f(X)$,
- построить функцию $\hat{Y} = f(X)$ в заданном диапазоне X,
- вычислить отклонения $Y - \hat{Y}$ для диапазона, в котором значения Y используются для настройки модели, то есть оценки коэффициентов,
- вычислить все $(Y - \hat{Y})^2$ и их сумму $\Sigma(Y - \hat{Y})^2$ (сумма квадратов отклонений (остатков) ,
- вызвать *Поиск решения*, целевая ячейка $\Sigma(Y - \hat{Y})^2$, *Изменяя ячейки* коэффициенты, ограничений нет, *Выполнить*.

Применение *Поиска решения* к линейной модели, представлена в таблице 4.5.

Метод наименьших квадратов с *Поиском решения* может применяться для настройки нелинейных моделей. Его использование для настройки Модели 2 – параболической – показано в Таблице 4.6.

Показатель качества линейной модели – коэффициент корреляции X и Y R_{xy} и его квадрат – ***коэффициент детерминации*** R^2.

Вычисленные для обеих моделей R^2, *DW, GQ* представлены в таблицах, а также показаны графики остатков. Видно, что качество обеих моделей высокое, применение МНК правомерно. Применение для прогноза одной из двух моделей зависит от дополнительной информации и личного опыта.

4.5. Вычисление эластичности

Важная характеристика экономических процессов – ***эластичность***, которая показывает, на сколько процентов изменится зависимая переменная ***Y*** при увеличении влияющей переменной ***X*** на 1 % :

$$Э = (\Delta Y / Y) / (\Delta X / X)$$

Применение компьютера позволяет вычислить эластичность по всему диапазону ***X***, а не только средние значения, как при ручном счете.

В качестве ***X*** и ***Y*** берутся их средние значения на соответствующих интервалах ΔX и ΔY, расчет ведется по аппроксимирующей функции $\hat{Y}$:

$$Э = (\hat{Y}1 - \hat{Y}0)/(\hat{Y}1 + \hat{Y}0)/(X1 - X0)*(X1+X0)$$

где индексы 0 и 1 относятся к первым двум значениям переменных ***X*** и $\hat{Y}$. Затем формула копируется на весь диапазон, кроме последней ячейки; в Модели1 расчет начинается с температуры 10^o. Графики показывают, что расчет эластичности по разным моделям приводит к различным результатам. Обычно экономисты используют среднюю эластичность

$$\bar{Э} = \frac{\Delta Y}{\Delta X} \times \frac{\bar{X}}{\bar{Y}} \quad ,$$

Где $\Delta Y / \Delta X$ – средний наклон функции $\hat{Y} = f(X)$. Применение функции эластичности позволяет изучать влияние добавок ***X*** на изменение ***Y*** при различных значениях влияющей переменной.

Далее представлены результаты расчетов по двум моделям.

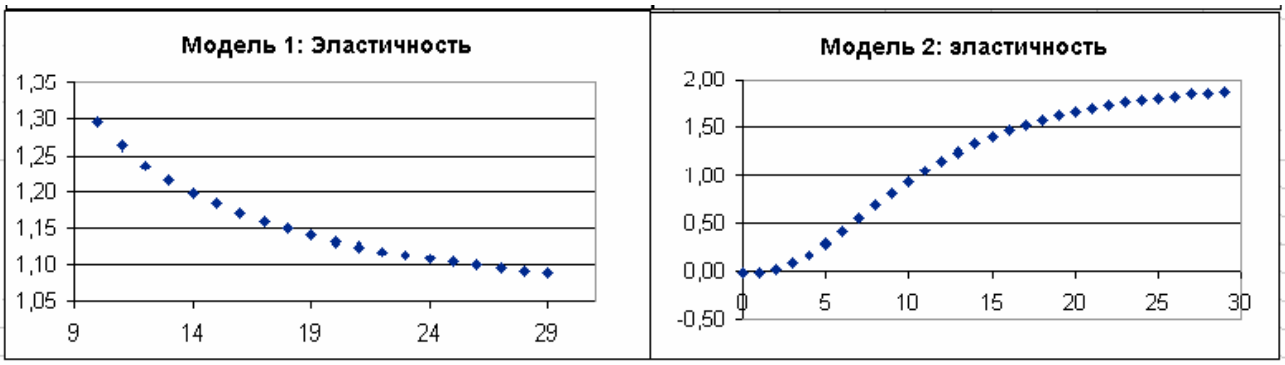


Рис.4.3.

Таблица 4.5.

		Модель 1	**a**	**b**	
			-4,2727	**1,7818**	
Температура X	Продажи Y	$\hat{Y}$.	Остатки: $Y-\hat{Y}$	$(Y-\hat{Y})^2$	Эластичность
10	12	13,55	-1,55	2,39	1,30
11	15	15,33	-0,33	0,11	1,26
12	18	17,11	0,89	0,79	1,24
13	16	18,89	-2,89	8,36	1,22
14	24	20,67	3,33	11,07	1,20
15	22	22,45	-0,45	0,21	1,18
16	27	24,24	2,76	7,64	1,17
17	28	26,02	1,98	3,93	1,16
18	25	27,80	-2,80	7,84	1,15
19	32	29,58	2,42	5,85	1,14
20	28	31,36	-3,36	11,31	1,13
21		33,15			1,13
22		34,93	**Σост^2**	**59,49**	1,12
23		36,71	**Корреляция**	**0,92**	1,11
24		38,49	**Индекс детерминации**	**0,85**	1,11
25		40,27			1,10
26		42,05	**Автокор-реляция**	**-0,58**	1,10
27		43,84	**DW**	**3,17**	1,10
28		45,62	**GQ**	**1,61**	1,09
29		47,40	**Дисп.ост.1**	**4,55**	1,09
30		49,18	**Дисп.ост.2**	**7,34**	
	ДИСП Y	**ДИСП $\hat{Y}$**	**ДИСП остатков**		
	40,87	**34,90**	**5,95**		

Таблица 4.6.

		Модель 2	**a**	**b**	**c**
			9,1942	**-0,3286**	**0,075588**
Температура *X*	Продажи *Y*	$\hat{Y}$	Остатки: $Y-\hat{Y}$	$(Y-\hat{Y})^2$	Эластичность
0	11	9,19	1,81	3,26	-0,01
1	8	8,94	-0,94	0,89	-0,02
2	9	8,84	0,16	0,03	0,01
3	13	8,89	4,11	16,90	0,08
....	**....**	**....**	**....**	**....**	**....**
18	25	27,77	-2,77	7,67	1,57
19	32	30,24	1,76	3,10	1,62
20	28	32,86	-4,86	23,59	1,66
21		*35,63*			1,69
22		*38,55*			1,72
23		*41,62*	**Σост^2**	**157,82**	1,75
24		*44,85*	**Индекс детерминации**	**0,88**	1,78
25		*48,22*	**F**	**61,0**	1,80
26		*51,75*	**Автокор-реляция**	**-0,13**	1,82
27		*55,43*	**DW**	**2,26**	1,84
28		*59,25*	**GQ**	**1,05**	1,85
29		*63,23*	**Дисп.ост.1**	**7,86**	1,87
30		*67,37*	**Дисп.ост.2**	**8,28**	
	ДИСП *Y*	**ДИСП $\hat{Y}$**	**ДИСП остатков**		
	68,19	**60,31**	**7,89**		

Некоторые комментарии к таблицам.

Индексы детерминации и *F*-статистики вычислены по формулам (3.2) и (3.3) на стр. 29 с использованием функции ДИСП. Коэффициент автокорреляции остатков *Rавт* вычислен с помощью функции КОРРЕЛ($\boldsymbol{e_1:e_{n-1}}$; $\boldsymbol{e_2:e_n}$), то есть в первом окне указан диапазон остатков с первого до ($\boldsymbol{n}$-*1*)-го, во втором – со второго до $\boldsymbol{n}$-го. Тест Дарбина-Уотсона осуществлён по формуле *DW=2(1–Rавт)*. В линейной модели *DW=3,17* , то есть попадает в интервал 3,07…4, соответствующий отрицательной автокорреляции. Этот пример объясняет секрет процветания казино. Исходные данные для этой задачи автор придумал сам, и тест Дарбина-Уотсона выявил, что эти числа не являются случайными. Человек не может создать абсолютно случайный ряд чисел, а

рулетка его создаёт. Из теории игр следует, что отклонение от оптимальной смешанной стратегии, в данном случае ряда случайных чисел, приводит к проигрышу игрока и выигрышу казино.

Тест Голдфелда-Квандта *GQ* проведён по первой и второй половинам диапазона остатков: данных слишком мало, чтобы исключать середину, как положено по правилам.

4.6. Нелинейные модели

Довольно часто приходится использовать нелинейные функции регрессии двух видов:

1. Регрессии, нелинейные относительно включённых в анализ объясняющих переменных:

Полином второй, редко третьей степени $y = a + bx + cx^2 + u$.

Гипербола $y = a + b/x + u$.

Эти модели сводятся к линейным заменой переменных: $z = x^2$ для полинома и $z = 1/x$ для гиперболы. После этого можно использовать функцию ЛИНЕЙН и сервис Регрессия, выделяя в качестве влияющих переменных x и z для полинома и z для гиперболы.

2. Регрессии, нелинейные по оцениваемым параметрам относятся:

Степенная $y = ax^b \varepsilon$;

Показательная $y = ab^x \varepsilon$;

Экспоненциальная $y = e^{a+bx} \varepsilon$.

Здесь $\varepsilon = 1 + u$. Эти модели могут быть линеаризованы логарифмированием, после чего можно использовать функцию ЛИНЕЙН и сервис *Регрессия*.

Например, показательная функция преобразуется в $ln(y) = ln(a) + xln(b) + ln(\varepsilon)$, или, *после переименования* $z = A + cx + v$.

После нахождения коэффициентов A и c можно вычислить $z\hat{} = A + cx$ и $y\hat{} = exp(z\hat{})$.

Самостоятельная работа

По данным Таблицы 4.7 определите по графику вид каждой функции регрессии, оцените её коэффициенты, используя ЛИНЕЙН или *Регрессия* с линеаризацией, или *Поиск решения*. По вектору остатков вычислите $\boldsymbol{R^2}$***, F, GQ, DW***. Сделайте выводы о качестве модели.

Таблица 4.7.

x	***y***	***y***	***y***	***y***	***y***	***y***	***y***	***y***	***y***	***y***	***y***	***y***	***y***	***y***
1	55	3	55	1	88	4	9	55	177	33	2	45	444	144
2	50	5	55	3	77	5	6	66	88	22	4	65	222	133
3	40	9	55	6	66	12	4	44	88	8	7	47	100	122
4	22	9	66	11	44	22	9	99	77	7	7	99	88	99
5	12	22	33	19	33	25	12	122	55	5	9	124	77	99
6	33	44	33	22	22	17	11	111	66	6	9	117	66	122
7	38	33	22	11	25	17	18	188	33	3	8	188	55	133
8	55	77	11	6	16	12	22	222	28	2	5	229	54	144
9	77	99	11	2	15	4	27	277	27	3	5	366	48	166
10	77	222	1	2	15	5	27	555	27	2	2	555	47	188
x	***y***	***x***	***y***	***x***	***y***	***x***	***y***	***x***	***y***	***x***	***y***	***x***	***y***	
55	3	55	1	66	4	9	55	111	33	1	45	220	55	
50	5	50	3	55	5	6	66	88	22	4	228	170	62	
40	9	40	6	66	12	4	44	88	8	9	47	100	122	
22	9	22	11	44	22	9	99	77	7	7	99	88	99	
12	22	12	19	33	25	12	122	55	5	9	124	77	99	
33	44	33	22	22	17	11	111	66	6	9	117	66	122	
38	33	38	11	25	17	18	188	33	3	8	188	55	133	
55	77	55	6	16	12	22	222	28	2	5	229	54	144	
77	99	77	2	15	4	27	277	27	3	5	298	48	166	

Контрольные вопросы

1. Метод наименьших квадратов (МНК) и работа с функцией ЛИНЕЙН.
2. Метод наименьших квадратов (МНК) и смысл выходной статистической информации сервиса Регрессия
3. Метод наименьших квадратов (МНК) и его реализация с использованием сервиса “Поиск решения”
4. Оценка погрешности прогноза и проверка адекватности модели
5. Экономический смысл коэффициентов линейного и степенного уравнений регрессии .

5. Оценка погрешностей параметров модели методом Монте-Карло

При работе на компьютере проще многократно проделать простые вычисления, чем один раз решить сложную аналитическую задачу. Поэтому для исследования стохастических моделей удобен метод Монте-Карло, позволяющий, в частности, оценивать погрешности параметров сложных моделей. Основные этапы реализации метода Монте-Карло:

1. Построение модели с “идеальными” параметрами.
2. Изменение значений переменных случайным образом в соответствии с дисперсией и законом распределения.
3. Расчет по проверяемой методике и сохранение параметров модели.
4. Возврат к п.2.

Пункты 2 и 3 выполняются заданное число раз – десятки, сотни, тысячи. В результате накапливаются массивы параметров, которые можно статистически обработать и установить надежность их оценок. В принципе, это можно сделать по аналитическим формулам дисперсионного анализа, но для сложной системы с внутренними связями такие расчеты становятся сложными и неустойчивыми.

В качестве примера используем эконометрическую модель парной регрессии, рассмотренную в предыдущем разделе. Этапы работы:

1. Задать коэффициенты линейной модели ***Yидеал = a + bX*** и стандартное отклонение остатков *(Sост)*. В данном случае a= - 4,27, b=1,78, *Sост* =2,44. Полученные результаты представлены в таблице 5.1. в столбце ***Yидеал***.

Таблица 5.1.

X	*Y*	*Yидеал*	*Yимит.*	$\hat{Y}$	**остатки**
10	12	13,55	13,55	13,13	0,42
11	15	15,33	13,50	14,91	-1,41
…					
19	32	29,58	34,34	29,11	5,23
20	28	31,36	27,34	30,89	-3,55
30				48,48	

2. Ввести в ячейки формулы и функции для расчета коэффициента детерминации R^2, коэффициента автокорреляции остатков *Rавт* и статистики Дарбина-Уотсона *DW = 2(1- Rавт)*, дисперсий остатков по первой и второй половинам диапазона и теста Голдфелда-Квандта *GQ=МАКС(ДИСП1; ДИСП2)/МИН(ДИСП1; ДИСП2)*; кроме того, в данном примере вычисляется прогнозное значение для *X*=30. *Ŷ(30), GQ* и *DW* размещаются в той же строке таблицы Excel, что и коэффициенты ***b*** и ***a***, что упрощает их сохранение.

3. Расчёт параметров модели с использованием функции ЛИНЕЙН.

Таблица 5.2.

b	***a***	***Y(30)***	***GQ***	***DW***
1,77	-4,62	48,48	8,05	3,11
0,25	3,94			
0,84	2,70		Автокорреляция	-0,55
47,53	9		Дисп.ост.1	1,76
346,88	65,68		Дисп.ост.2	14,20

5. Сохранение в таблице Excel вычисленных параметров модели (сотни и тысячи имитаций) и статистическая обработка. В Таблице 5.2 представлена часть массива результатов. Вычислено среднее значение каждого параметра, что позволяет оценить несмещённость, стандартное отклонение и относительную погрешность.

Таблица 5.3.

	b	***a***	***Y(30)***	***GQ***	***DW***
	0,95	7,60	36,22	3,88	4,11
	1,69	-3,91	46,91	1,40	3,31
	1,71	-3,55	47,90	9,47	2,69
	2,08	-10,59	51,70	1,55	3,34
	1,74	-5,20	47,14	1,93	2,86
	2,08	-8,47	53,85	7,99	2,25
Среднее	1,78	-4,42	48,96	4,37	3,09
СКО	0,2	3,14	3,113	3,52	0,64
%	11,4	71,2	6,358	80,58	20,82

В представленных таблицах не предусмотрено сохранение коэффициента детерминации, вычисляемого функцией ЛИНЕЙН. Включите его в рассмотрение.

Процедура и программный модуль для создания имитаций и сохранения результатов, а также упрощенная технология создания имитаций, позволяющая обойтись без программирования, представлены в Приложении 1.

5. После завершения набора результатов (не меньше 100 циклов) вычислите стандартные отклонения *a, b, Y(30), GQ, DW,* R^2, сравните полученные значения с вычисленными по аналитическим формулам

$$Sa = \frac{Socm\sqrt{\sum X^2}}{NSx} \qquad S_b = \frac{Socm}{Sx\sqrt{N}} \qquad S_{Rxy} = \sqrt{\frac{1-Rxy}{N-2}}$$

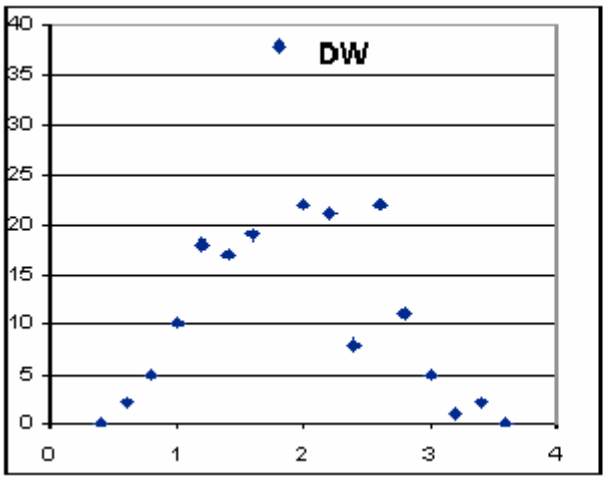
$$S_Y = Socm\sqrt{\frac{1}{N} + \frac{(X-\bar{X})^2}{\sum(X_i-\bar{X})^2}}$$

6. Постройте корреляционные графики и вычислите коэффициенты корреляции *a(b),* $a(R^2)$, $b(R^2)$.

7. Постройте гистограммы частотных распределений *a, b,* R^2, *Y(30), DW, GQ.* Для этого введите в таблицу Excel границы интервалов значений параметров (карманы) и запустите *Сервис (или Данные) – Анализ данных – Гистограмма.*

Исследования сравнительно простой модели – парной линейной регрессии – приводят к интересным результатам.

1. На рисунках представлены графики частотных распределений ***DW*** и ***GQ***. Тесты показывают наличие автокорреляции для 7,5 % имитаций и гетероскедастичности для 8,5 % имитаций, причём график *GQ* имеет длинный хвост. Имитации создавались на основе нормального распределения возмущений, значит, автокорреляцию и гетероскедастичность можно обнаружить, если для них нет никаких предпосылок.



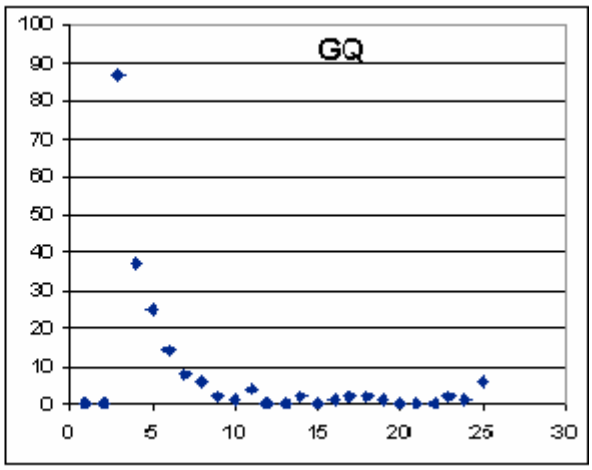


Рис.5.1.

2. В данном случае гипотеза $a = 0$ является приемлемой. Одна серия испытаний проведена с исключением a из модели, другая – без исключения. При ислючении a Sb уменьшается втрое, а СКО $\hat{Y}прогн(30)$ вдвое, но появляется систематическая погрешность (смещение).

Таблица 5.4.

	b	a	$\hat{Y}прогн(30)$
Среднее	1,509	0,000	45,26
СКО	0,053	0,000	1,6
СКО /среднее %	3,536		3,535

Похожие результаты получаются и при исследовании нелинейной зависимости $\hat{Y} = a + bX + cX^2$. При исключении слагаемого с коэффициентом b, имеющим погрешность 157%, погрешности a, c и $\hat{Y}(30)$ уменьшились вдвое, но появилось смещение $\hat{Y}(30)$ примерно на 10%.

4. Обработка расчетов методом Монте-Карло показала правомерность расчета погрешности точечной оценки прогноза $\hat{Y}$ по формуле

$$S^2(\hat{Y})=(Sa)^2 + X^2(Sb)^2+2XCov(a,b).$$

Особый интерес представляет полученный коэффициент корреляции a и b, равный $-0,98$, что неудивительно, т.к. $a = Ycp - bXcp$. Из этого следует важнейший вывод: ***погрешность прогнозного значения*** $\hat{Y}$ ***меньше***

относительных погрешностей a и b, и вдали от средних X и Y близка к разности погрешностей слагаемых в уравнении регрессии:

$$S(\hat{Y}прог) = | Sb * Xпрог - Sa |,$$ здесь 0,2*30 – 3,14 = 2,86

что совпадает с результатами расчетов по формуле (=2,92) и методом Монте-Карло (=3,11).

5.1. Результаты воздействия гетероскедастичности и автокорреляции, оценённые методом Монте Карло

Во всех учебниках по эконометрике написано, что, в соответствии с теоремой Гаусса-Маркова, линейное преобразование $\mathbf{B}=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ обеспечивает несмещённую, эффективную и состоятельную оценку компонент вектора **B**, состоящего из коэффициентов линейного уравнения регрессии

$$\mathbf{Y} = \mathbf{BX} + \mathbf{U}$$

если возмущения $\mathbf{u_i} \in \mathbf{U}$ подчиняются закону нормального распределения, их ожидаемые величины равны нулю, отсутствуют гетероскедастичность и автокорреляции. Здесь **X** – матрица значений влияющих переменных, **Y** – вектор зависимых переменных. Но насколько изменятся коэффициенты уравнения регрессии и прогнозируемые величины, если возмущения **U** будут гетероскедастичны и коррелированны? В этом случае рекомендуется применять взвешенный и обобщённый методы наименьших квадратов, но оправдано ли усложнение методов решения задачи? Ответ может дать оценка погрешностей коэффициентов уравнения регрессии и прогнозных значений методом Монте Карло. Проделайте дальнейшие расчёты самостоятельно, это существенно улучшит ваше понимание эконометрики.

Расчёты методом Монте Карло проводились следующим образом.

1. Задана “идеальная” зависимость $Y = 5 + x;\ x$=1 . . . 20.
2. Созданы массивы ожидаемых значений возмущений V{v*(x)*}:
   1. v*(x)*=*Const* = 4.
   2. V= 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6.
   3. V= 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2.

4. V= 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 10, 10, 9, 3, 3, 3.

5. V= 3, 3, 10, 10, 9, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3.

6. V= 4, сдвиг 4, -4,- 4, 4, 4, -4, -4, 4, 4, -4, - 4, 4, 4, -4, -4, 4, 4, -4, -4, 4.

7. V= 4, сдвиг 4, 4, 4, -4, -4, -4, -4, -4, 4, 4, 4, 4, 4, -4, -4, -4, -4, -4, 4, 4.

8. V= 4, сдвиг 4, -4, 4, -4, 4, -4, 4, -4, 4, -4, 4, -4, 4, -4, 4, -4, 4, -4, 4, -4.

Массивы 2-5 обеспечивают гетероскедастичность, массивы 6-8 автокорреляцию.

3. Разработан программный модуль на языке Visual Basic, обеспечивающий создание случайных величин q, имеющих нормальное распределение с $E(q)=0$, $E(\sigma(q))=1$, а также случайных величин $\boldsymbol{Y_{имит\ i} = 5+x_i+qv_i+сдвиг.}$ Кроме того, программный модуль обеспечивает сохранение вычисляемых параметров модели при каждой имитации.

4. Имитация $\boldsymbol{Y_{имит\ i}}$, вычисление с использованием функции ЛИНЕЙН() коэффициентов уравнения регрессии $\boldsymbol{\hat{Y} = a+bx}$, коэффициента детерминации $\boldsymbol{R^2}$ и статистики Фишера F, прогнозного значения $\boldsymbol{\hat{Y}(30)}$. По вычисленным $\boldsymbol{a}$ и $\boldsymbol{b}$ строится вектор оценённых значений $\boldsymbol{\hat{Y}(x)}$ и вектор остатков $\boldsymbol{e = Y - \hat{Y}}$, по которому вычисляются тесты Голдфелда-Квандта

$$\boldsymbol{GQ = МАКС(\sigma^2(E1),\sigma^2(E2))/МИН(\sigma^2(E1),\sigma^2(E2))}$$

и Дарбина-Уотсона $DW = 2(1-Rавт)$. Здесь $\sigma^2(E1)$, $\sigma^2(E2)$ – дисперсии остатков в диапазонах $e(1)…e(10)$ и $e(11)…e(20)$, $Rавт$ = КОРРЕЛ($e(1):e(19); e(2):e(20)$). Используются функции Excel ДИСП() и КОРРЕЛ().

5. Сохранение вычисленных $\boldsymbol{a, b, \hat{Y}(30), R^2, F, GQ, DW}$.

6. Повторение п.п.3-5 много раз. В данном примере каждый опыт (имитации и расчёты с заданными возмущениями) повторялись 10000 раз.

7. Статистическая обработка и интерпретация накопленных результатов с использованием функций Excel СРЗНАЧ (средние значения), СТАНДОТКЛОН (стандартные отклонения). Использован также сервис “Гистограмма” из пакета “Анализ данных” для построения гистограмм частотных распределений.

На рисунках 1-2 представлены примеры $Y_{идеал} = 5+x$, $Y_{имит\ i}$, $\hat{Y}=a+bx$, соответствующие восьми массивам ожидаемых значений возмущений п.2.

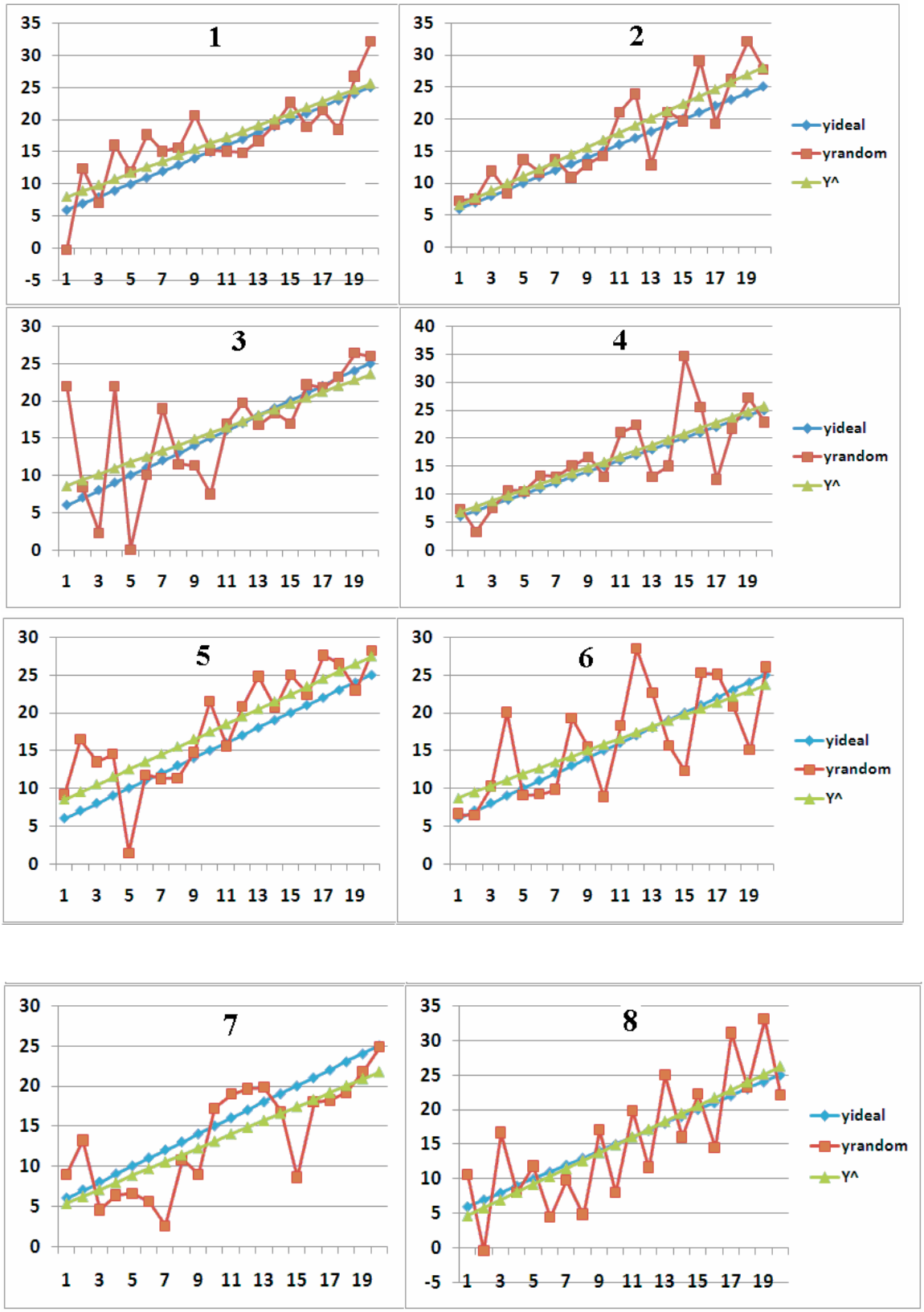


Рис.5.2.

Результаты статистической обработки накопленных данных представлены в Таблице 5.5: средние значения (Ср), стандартные отклонения S, относительные погрешности (%) и процент аномальных значений. Аномальными считались значения параметров: F>4,35; GQ>5; DW<1,2 - DW>2,8.

Таблица 5.5.

№	Возмущения		b	a	$\hat{Y}(30)$	R^2	F	GQ	DW
1	4 4 4. . . 4 4 4	Ср	0,9988	5,018	34,98	0,699	48,38	1,937	2,207
	(нет искажений)	S	0,152	1,833	3,098	0,093	23,62	1,196	0,43
		%	15	36	8,8	13	48	61	19
	% аномальных						0	2,42	1,04-8,4
2	2 2 2 . . . 6 6 6	Ср	1,001	5,002	35,031	0,65	42,05	10,75	2,204
		S	0,172	1,328	4,121	0,118	25,55	8,68	0,50
		%	17	26	11,7	18	60	80	22
	% аномальных						0,02	77	2,35-12
3	6 6 6 . . . 2 2 2	Ср.	0,9977	5,049	34,982	0,65	41,61	10,86	2,203
		S	0,172	2,596	2,758	0,119	25,07	9,34	0,506
		%	17	51	7,88	18	60	85	22,9
	% аномальных						0,05	78	2,4-13
4	3 3 ... 3 10 10 9 3 3 3	Ср.	0,9992	5,0144	34,990	0,64	41,57	4,966	2,206
		S	0,174	1,510	4,077	0,137	27,01	5,06	0,498
		%	17	30	11,6	21	65	102	22
	% аномальных						0,23	32,7	1,9-13
5	3 3 10 10 9 3 3 … 3 3	Ср	0,9981	5,0322	34,975	0,645	42,48	4,775	2,231
		S	0,193	2,784	3,261	0,145	28,20	4,78	0,48
		%	19	55	9,32	22	66	100	21,9
	% аномальных						0,46	30,91	1,4-14
6	4 -4 -4 4 4 -4 -4 4 4 -4 -4 4 4 -4 -4 4 4-4-4 4	Ср.	1,0005	5,0314	35,046	0,527	22,10	1,738	2,159
		S	0,153	1,835	3,118	0,102	9,81	0,81	0,23
		%	15	36	9	19	44	46	10,7

№	Возмущения		*b*	*a*	$\hat{Y}$ *(30)*	R^2	*F*	*GQ*	*DW*
7	4 4 4 -4 -4 -4 -4 -4 4 4 4 4 4 4 -4-4-4-4-4 4 4	Ср.	0,9379	5,6602	33,797	0,497	19,63	1,748	1,513
		S	0,154	1,840	3,138	0,106	9,127	0,864	0,364
		%	16	32	9,3	21	46	49	24
	% аномальных						0,24	1,07	20-0,15
8	4-4 4 -4 4 4 -4 -4 4-4 4-4 4-4 -4 4-4 4 -4 4	Ср.	0,9398	5,6531	33,8226	0,497	19,60	1,739	3,122
		S	0,151	1,834	3,080	0,106	8,998	0,857	0,327
		%	16	32	9,1	21	46	49	10,5
	% аномальных						0,21	0,98	0-83,7

Гистограмма *DW* и частоты *GQ* по опыту 1 *(u(x)=4)* представлены на рисунке 5.3 и в таблице 5.6.

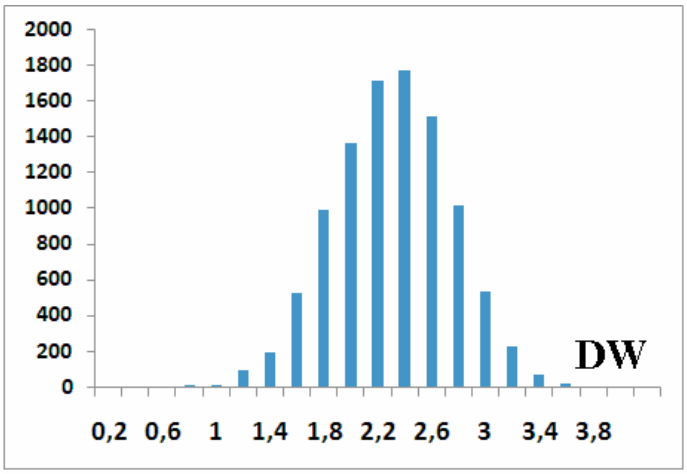


Рис. 5.3.

Таблица 5.6.

Интервал *GQ*	Частота
1-2	6897
2-3	1973
3-4	653
4-5	235
5-6	109
6-7	61
7-8	25
8-9	14
9-10	13
10-11	3
11-12	3
12-13	3
13-14	3
14-15	2
15-16	1
>16	5

ВЫВОДЫ. 1. Теоретическое значение погрешности коэффициента ***b*** равно 0,155; точечная оценка статистической погрешности прогнозного значения $\hat{Y}$***(30)*** равна 3,22; интервальная 5,14. При 10000 имитаций погрешность среднего значения ***b*** должна быть 0,00155, а $\hat{Y}$***(30)*** 0,0322. В Опытах 1, 2, 4 погрешности средних значений ***b*** укладываются в интервал 1 СКО, в Опытах 3 и 5 в 2 СКО. В Опытах 1-5 погрешности средних значений $\hat{Y}$***(30)*** укладываются

в интервал 1 СКО, в Опыте 6 в 2 СКО. Можно сказать, что при разнице СКО остатков в 3 раза гетероскедастичность не приводит к значимым ошибкам, которые остаются в пределах статистических погрешностей оценок средних значений параметров. СКО коэффициентов и ***Ŷ(30) существенно не возрастает.*** Результаты Опытов 2-5 показали, что погрешность прогноза зависит не от ***GQ***, а близости больших возмущений к точке прогноза. Значит, СКО $\hat{Y}$ без корректировки правильнее отражает истинную погрешность прогноза. В связи с этим теряет смысл ***Взвешенный метод наименьших квадратов*** (ВМНК), предполагающий искусственную корректировку остатков путём деления на их ожидаемые СКО.

2. Во всех опытах обнаружена нулевая корреляция ***Ŷ(30)*** и ***b*** с ***GQ*** **и** ***DW***. Это видно и на Рисунке 5.4 (Опыт 2).

3. В Опытах 7 и 8 с положительной и отрицательной автокорреляцией обнаружено существенное смещение средних значений ***b*** и ***Ŷ(30)***, в отличие от Опыта 6, где положительные и отрицательные сдвиги чередуются через 2 и *DW* близок к 2. Сильная отрицательная автокорреляция – экзотика, не характерная для реальной жизни. Положительная автокорреляция отклонений от тренда проявляется во временных рядах цен на фондовом рынке и означает, что надо применять другие методы и модели: авторегрессии (см. Раздел 8.2), технический анализ фондового рынка и др. Возникает вопрос о целесообразности изучения и применения ***Обобщённого метода наименьших квадратов***, основанного на преобразовании матриц (см. раздел 3.3), но с учётом корреляций остатков.

4. Относительная погрешность ***Ŷ (30)*** меньше погрешностей ***a*** и ***b***. Это связано с тем, что

$\sigma^2(\hat{Y}(30) = \sigma^2(a)+30^2\sigma^2(b)+2Cov(a,bx) = \sigma^2(a)+30^2\sigma^2(b)+2*30*\sigma(a)*\sigma(b)*Rab$

где *Rab* коэффициент корреляции *a* и *b*. В наших опытах *Rab = - 0,82… -0,96*, формула близка к формуле квадрата разности, и $\sigma(\hat{Y}(30)$ близок к модулю разности $\sigma(a)$ и $30\sigma(b)$.

5. Случайное сочетание результатов измерений может имитировать гетероскедастичность и автокорреляцию, даже если нет порождающей их закономерности: в Опыте 1 2,42% тестов *GQ* превысили 5; 1% *DW* показал положительную автокорреляцию (<1,2) и 8,4% отрицательную (>2,8). На Рис.5.4 (Опыт 2) видны большие значения *GQ*. Были обнаружены огромные величины *GQ*, не коррелирующие с $\hat{Y}$ *(30):* 90 ($\hat{Y}$ =40), 93 ($\hat{Y}$ =39), 97 ($\hat{Y}$ =34), 101 ($\hat{Y}$ =33), 111 ($\hat{Y}$ =35), 135 ($\hat{Y}$ =35).

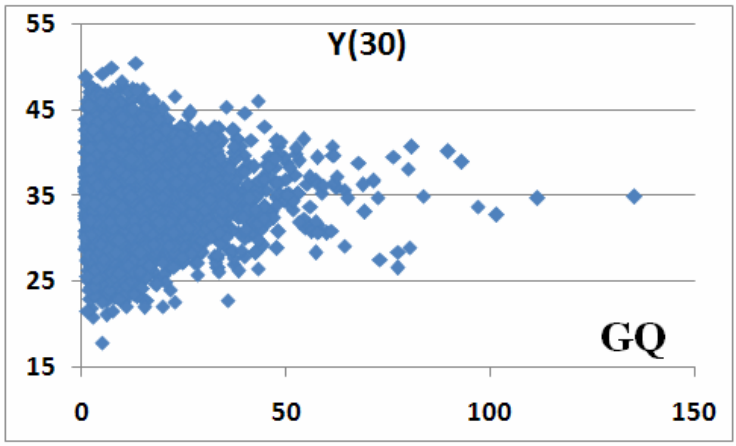


Рис. 5.4.

Контрольные вопросы.

1. Оценка погрешностей параметров модели по формулам.
2. Оценка интервального и точечного среднеквадратичного отклонения прогнозного значения в парной линейной регрессии.
3. Проверка адекватности модели.
4. Оценка погрешностей параметров модели методом Монте-Карло.

6. Некоторые методы регрессионного анализа

6.1. Тест Чоу

Тест Чоу позволяет количественно оценить выгоду от усложнения модели, например, замены одной прямой линии двумя или кривой. Результаты расчета по формуле Чоу сравнивается с критическими значениями статистики

Фишера, что даёт основание принять или отвергнуть гипотезу об улучшении модели.

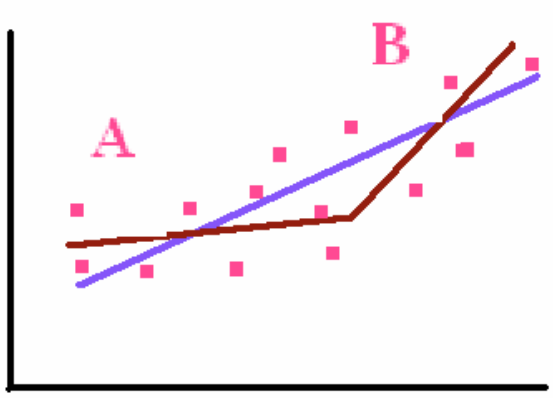


Рис.6.1.

Рассмотрим диаграмму рассеяния на рисунке 6.1. Какую регрессионную модель использовать? Стоит ли заменять прямую линию на ломаную? Согласно тесту Чоу

$$F(k, n-2k) = \frac{\text{УЛУЧШЕНИЕ / Доп.Степени Свободы}}{\text{Остаточное } \Sigma ост^2 \text{ /Остаточное число степеней свободы}}$$

или

$$F(k, n-2k) = \frac{(\ \Sigma ост^2\ \ все\ - \Sigma ост^2\ A - \Sigma ост^2\ B\ ) / k}{(\Sigma ост^2\ A + \Sigma ост^2\ B\ ) / (n - 2k)}$$

Попробуем применить тест Чоу к нашей задаче Раздела 3. Улучшается ли модель при переходе от прямой к параболе и от 11 замеров (10^o – 20^o) к 21 замеру (0^o – 20^o) ? В данном случае $\boldsymbol{\Sigma ост^2\ A + \Sigma ост^2\ B}$ = *157, 82* (парабола), $\boldsymbol{\Sigma ост^2\ B}$ = *59,5* (прямая в области 10^o – 20^o), $\boldsymbol{n}$=*21*, $\boldsymbol{k}$=*1*, ($\boldsymbol{n - 2k}$) = *19*.

Проводить ломаную линию нецелесообразно, лучше нормировать $\Sigma ост^2\ B$ на 21 точку: $\boldsymbol{\Sigma ост^2 все}$ = 59,5 * 21 /11 = 113,6. При этом величина ($\boldsymbol{\Sigma ост^2\ \ все\ - \Sigma ост^2\ A - \Sigma ост^2\ B}$) отрицательна, то есть применение теста Чоу невозможно.

6.2. Тобит-анализ

Рассмотрим ситуацию, касающуюся покупки антиквариата, драгоценностей и т.п. в зависимости от дохода семьи. Ясно, что при малом доходе семьи это не покупают или почти не покупают, около порогового значения дохода количество покупок незначительно, а затем начинается закономерный рост. Как построить эконометрическую модель? Можно разбить массив данных на кластеры и обрабатывать отдельно, как в линейной модели торговли мороженым. Это сделано в модели, представленной на Рисунке 6.2. Джеймс Тобин предложил технологию решения таких задач, основанную на методе наибольшего правдоподобия: предполагается нормальное распределение остатков с обрезанным левым краем. Метод Тобина, или Тобит-анализ, включён в статистические пакеты (Stata, EViews и др.) и позволяет оценивать параметры таких моделей.

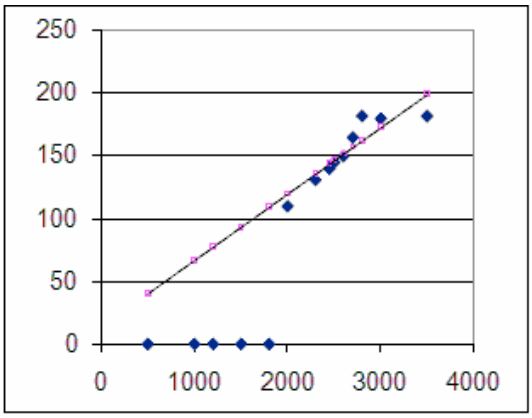


Рис. 6.2.

6.3. Модели двоичного выбора, логит и пробит

Существует класс задач, где экзогенная переменная может принимать только два значения: например, сдача экзамена на права. Вектор зависимой переменной представляет из себя набор нулей и единиц. Требуется оценить вероятность положительного исхода в зависимости от потраченного времени (или денег). Ясно, что линейная модель не годится. Функция должна иметь очень малые значения при $\boldsymbol{x}$ меньше порогового, затем возрастать и иметь асимптоту на уровне единицы, так как вероятность больше единицы быть не может при любых затратах времени. Таким требованиям удовлетворяют две функции: логистическая

$$\hat{Y} = \frac{1}{1+\exp(-Z)}$$

и кумулятивная функция нормального распределения

$$\hat{Y} = \int \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{(Z-\mu)^2}{2\sigma^2})$$

где ***Z = a+bx***

Параметры обеих функций можно оценить методом наименьших квадратов, используя *Поиск решения*. Результат показан на рисунке 6.3. Разумеется, требования теоремы Гаусса-Маркова о нормальном распределении остатков и об отсутствии автокорреляции и гетероскедастичности здесь не соблюдаются: Рисунок 6.4. В программах статистического анализа имеются функции Logit и Probit, основанные на методе наибольшего правдоподобия и позволяющие решать такие задачи.

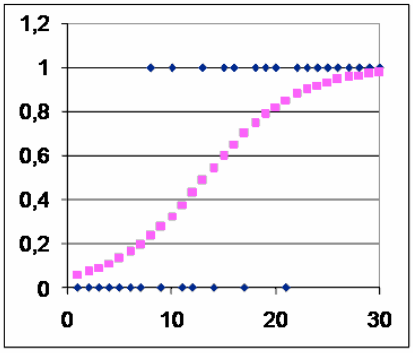


Рис.6.3.

Рис.6.4.

6.4. Метод максимального правдоподобия

Метод максимального правдоподобия позволяет работать при несоблюдении требований теоремы Гаусса-Маркова, а именно: при наличии гетероскедастичности и несоответствии остатков закону нормального распределения. Метод основан на максимизации совместной вероятности появления реальных значений переменной (в данном случае – остатков) при подборе параметров их функции распределения.

Предварительно приведем пример (по [1]). Требуется оценить математическое ожидание ***E(X)*** переменной ***X*** по выборке из двух чисел: ***X1***=4 и ***X2***=6. Можно в качестве оценки ***E(X)*** использовать среднее значение***:***

Xcp. = (4 + 6) / 2 . Если знать, что закон распределения переменной ***X*** – нормальный с известным стандартным отклонением ***σ***, то можно подобрать ***E(X),*** максимизирующую функцию

$$L(X1, X2, \mu) = p(X1, X2, \mu)) = p(X1, \mu) * p(X2, \mu) =$$

$$\frac{Exp(-(X1-\mu)^2/(2\sigma^2))}{\sigma\sqrt{2\pi}} * \frac{Exp(-(X2-\mu)^2/(2\sigma^2))}{\sigma\sqrt{2\pi}}$$

которая называется функцией правдоподобия.

Ее логарифм

$$l(X1, X2, \mu) = ln(L(X1, X2, \mu)) = \Sigma(-(Xi-\mu)^2/2\sigma^2) - ln(\sigma) - ln(\sqrt{2\pi})$$

т.е. от произведения мы переходим к сумме. При σ=1, отброшенных константах и различных μ:

$$\mu = 4 : l(4,6,4) = -(4-4)^2 - (6-4)^2 = -4$$

$$\mu = 5 : l(4,6,5) = -(4-5)^2 - (6-5)^2 = -2$$

$$\mu = 6 : l(4,6,6) = -(4-6)^2 - (6-6)^2 = -4$$

т.е. максимум функции правдоподобия достигается при μ=5.

При построении эконометрической модели вместо μ происходит оценка коэффициентов функции регрессии. В случае парной регрессии $\hat{Y} = a + bX$, и происходит подбор коэффициентов ***a*** и ***b,*** обеспечивающих максимизацию функции правдоподобия для совокупностей значений (векторов) Y и $\hat{Y}$ при заданной функции распределения плотности вероятности их разностей (остатков).

Далее представлено построение парной регрессии двумя методами: МНК и Методом максимального правдоподобия (МНП). Распределение остатков предполагается нормальным с оценкой стандартного отклонения **s**. Технология МНП аналогична МНК с использованием *Поиска решения*: задаются произвольные начальные значения ***a, b,*** но, кроме того, ***s;*** по ***a, b*** строится функция $\hat{Y} = a + bX$, вычисляются остатки $Y_i - \hat{Y}_i$, но затем вычисляются не квадраты остатков, а функции $ocm^2/2s^2 + ln\ s$ и их сумма, равная $-ln(L)$. *Поиск решения* ищет ее минимум, изменяя ячейки ***a , b*** и, может быть, ***s.*** Квадраты остатков МНП приведены для сравнения: видно, что их сумма меньше, чем для МНК, что говорит о высоком качестве подгонки модели.

Таблица 6.1.

		МНК			Метод максимального правдоподобия			
	a	**6,80**			***a***	**6,45**		
	b	**2,09**			***b***	**2,14**		
					s	**3,53**		
X	***Y***	Ŷ	**остатки**	**ост^2**		**остатки**	**ост^2/2s^2 + ln *s***	**ост^2**
1	12	8,90	3,10	9,63	8,59	3,41	1,73	11,60
2	8	10,99	-2,99	8,95	10,73	-2,73	1,56	7,47
3	15	13,09	1,91	3,66	12,87	2,13	1,44	4,53
4	18	15,18	2,82	7,95	15,01	2,99	1,62	8,93
5	14	17,28	-3,28	10,73	17,15	-3,15	1,66	9,93
6	22	19,37	2,63	6,92	19,29	2,71	1,56	7,34
7	18	21,46	-3,46	12,00	21,43	-3,43	1,73	11,77
8	17	23,56	-6,56	43,01	23,57	-6,57	2,99	43,16
9	28	25,65	2,35	5,51	25,71	2,29	1,47	5,25
10	25	27,75	-2,75	7,55	27,85	-2,85	1,59	8,11
11	35	29,84	5,16	26,61	29,99	5,01	2,27	25,12
12	33	31,94	1,06	1,13	32,13	0,87	1,29	0,76
		Σ	**0,00**	**143,6**		**0,00**	**17,62**	**124,90**

Для демонстрации возможностей МНП были проведены расчеты в предположении, что распределение остатков имеет форму равностороннего треугольника с основанием 2***s***. Отрицательные вероятности заменяются на небольшую положительную величину (здесь 0,1) при помощи функции ЕСЛИ. При работе *Поиска решения* изменялись ***a*** и ***b***, ***s*** не изменялась. Полученная сумма квадратов остатков в этом случае аналогична результату МНК. Начальные значения ***a*** и ***b*** в данном случае требуется задавать достаточно близко к истинным значениям, иначе *Поиск решения* выдает неверные решения.

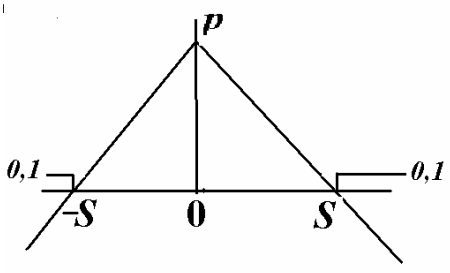


Рис 6.5.

Таблица 6.2.

	a	**6,66**						
	b	**2,10**		Распределение возмущений - треугольник				
	s	**8,00**						
X	***Y***	$\hat{Y}$	**остатки**	**ост^2**	**1-ABS(ост)/s**	**ЕСЛИ(L >0; L ; 0,1)**		
1	12	8,76	3,24	10,48	0,60	0,60		
2	8	10,86	-2,86	8,19	0,64	0,64		
…								
10	25	27,66	-2,66	7,07	0,67	0,67		
11	35	29,76	5,24	27,47	0,34	0,34		
12	33	31,86	1,14	1,30	0,86	0,86		
			Σ	**Σ**				
			1,27	143,78				

6.5. Фиктивные и замещающие переменные

Фиктивные (dummy) переменные позволяют ввести в модель и учесть качественные характеристики, например пол покупателей, расположение магазина и т.п. Например, задача раздела 3 (продажи пива в зависимости от температуры) может быть модифицирована

$$Y = a + bX + cz + u \quad \text{или}$$

$$Y = a + b(1+cz)X + u,$$

где переменная **z** принимает значения 0 или 1.

Фиктивные переменные позволяют объединить в одной модели выборки, имеющие отличия. В принципе, их можно рассматривать отдельно, но объединение может дать более качественную модель. Повышение качества модели можно оценить, используя тест Чоу.

Если при проведении статистических исследований какую-либо переменную сложно или невозможно измерить, но существует и известна её зависимость от другой, доступной переменной, то применяются замещающие (proxy) переменные: измеряют доступную переменную, строят с ней регрессионную модель, а затем прогнозируют значения недоступной переменной. Конечно, точность прогноза падает, но в некоторых случаях иначе не получается.

Контрольные вопросы

1. Что такое фиктивные переменные и тест Чоу
2. Фиктивные переменные: определение, назначение, типы
3. Тест Чоу на наличие структурных изменений в регрессионной модели
4. Что такое и где применяется Тобит-анализ
5. Логит, пробит и модели двоичного выбора
6. Что такое и где применяется Метод максимального правдоподобия

7. Множественная регрессия

7.1. Зависимость валового дохода от основных фондов и оборотных средств

В моделях множественной регрессии зависимая переменная является функцией многих факторов. Далее приведен пример решения задачи из практикума [5], в которой требуется определить зависимость валового дохода за год ***Y*** от основных фондов ***X1*** и оборотных средств ***X2***.

Таблица 7.1.

	Среднегодовая стоимость,млн.руб		
Номер	основных фондов *X1*	оборотных средств *X2*	Валовый доход за год,млн.руб. *Y*
1	118	105	203
2	28	56	63
3	17	54	45
4	50	63	113
5	56	28	121
6	102	50	88
7	116	54	110
8	124	42	56
9	114	36	80
10	154	106	237
11	115	88	160
12	98	46	75

На результаты расчета коэффициентов в моделях множественной регрессии негативное влияние оказывает взаимозависимость влияющих

факторов (***коллинеарность***, ***мультиколлинеарность***), поэтому изучение зависимости ***Y*** от различных факторов следует начинать с расчета коэффициентов корреляции ***Y*** от всех ***X*** и факторов ***X*** между собой. Для этого удобно использовать сервис *Корреляция*, входящий в *Пакет анализа* Excel. Результаты представлены в таблице 7.2. и на графиках (диаграмма *Точечная*).

Таблица 7.2.

	X1	*X2*	*Y*
X1	1		
X2	0,4130	1	
Y	0,5708	0,8328	1

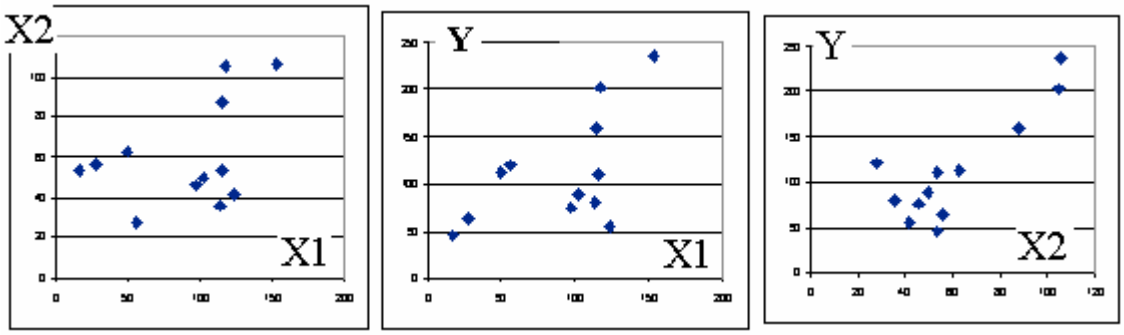


Рис. 7.1 .

Видна слабая зависимость факторов *X1* и *X2* между собой (отсутствие коллинеарности векторов **X1** и **X2**) и зависимость *Y* от фактора *X2*.

Мы исследуем две модели: линейную (аддитивную) модель и степенную (мультипликативную). Линейная модель:

$$\hat{Y} = a + b1*X1 + b2*X2$$

Для оценки параметров модели можно использовать функцию ЛИНЕЙН, *Сервис Регрессия* или *Поиск решения*. Расчёт коэффициентов модели с использованием сервиса *Регрессия:*

Таблица 7.3.

	Коэффициенты	*Стандартная ошибка*	*t-статистика*
a	-24,02	28,05	-0,856
b1	0,3829	0,253	1,51
b2	1,677	0,421	3,97

Для вычисления эластичности по *X1* надо предварительно провести сортировку таблицы по этому фактору. Вычислите эластичность $\hat{Y}$ по *X1* по формуле

$$\boldsymbol{Э = (\hat{Y}1 - \hat{Y}0)/(\hat{Y}1 + \hat{Y}0)/(X1 - X0)*(X1+X0)}$$

Как видите, результат в таблице 7.4 и на Рисунке 7.2 *все X2* получился безобразный, т.к. $\hat{Y}$ зависит от двух факторов. Для получения "срезов" по поверхности $\hat{Y}(X1,X2)$ надо фиксировать *X2*, т.е. заполнить этот столбец одинаковыми значениями. Ниже представлены графики эластичности $\hat{Y}$ по *X1* при *X2 = 28*, *X2* = 56 и *X2* = 106. Результаты весьма информативны и позволяют судить о целесообразности вложений в основные фонды и оборотные средства при их различных значениях, в отличие от обычно применяемого среднего значения эластичности, вычисляемого по формуле

$$Э(Y, X1) = b1*X1ср/Yср$$

Здесь *Э(Y, X1)* = 0,31. Также вычислен и представлен в таблице 7.4. коэффициент детерминации $R^2 = 1 - ДИСП(ост.) / ДИСП(Y)$, здесь равный 0,755.

Таблица 7.4

				a	*b1*	*b2*
	Среднегодовая стоимость,млн.руб			-24,02	0,38	1,68
Номер	Основных фондов *X1*	Оборотных средств *X2*	Валовый доход за год,млн.руб. *Y*	$\hat{Y}$	$(Y- \hat{Y})$	Эластичность
3	17	54	45	73,07	-28,16	0,20
2	28	56	63	80,63	-17,7	0,39
4	50	63	113	100,80	12,18	-6,86
5	56	28	121	44,39	76,7	1,26
12	98	46	75	90,66	-15,5	2,17
6	102	50	88	98,90	-10,74	-1,90
9	114	36	80	80,01	0,22	81,01
11	115	88	160	167,62	-7,52	-46,97
7	116	54	110	110,97	-0,78	32,76
1	118	105	203	197,29	5,78	-14,32
8	124	42	56	93,91	-37,66	3,59

10	154	106	237	212,75	24,42	
Средн.	91,00	60,67	112,58			
ДИСП			3571		872	
Детерминация		0,755		Эластичность =*b1*X1cp/Ycp*		0,31

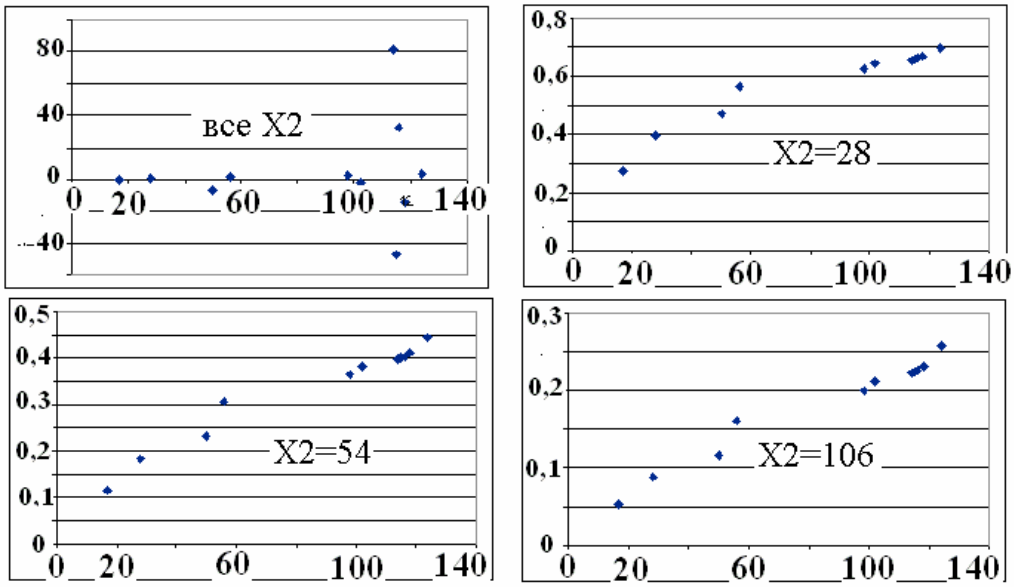


Рис.7.2. Эластичность *Y* по *X2*

Отсортируйте таблицу по столбцу *X2* и постройте графики эластичности *Y* по *X2* при малых, средних и больших значениях *X1*.

Аддитивные модели часто используются для изучения эконометрики, но насколько они соответствуют реальной экономике? В нашем случае нет: если имеются только основные фонды (здания, станки), но нет оборотных активов, то нет и производства. Именно это произошло в России в 1992 году, когда в результате “шоковой терапии” предприятия остались без средств и были захвачены или уничтожены. Поэтому более реальной представляется мультипликативная модель, предложенная Коббом и Дугласом для описания макроэкономики. Мы её применим к микроэкономике, а потом воспользуемся данными, с которыми работали Кобб и Дуглас.

Рассмотрим ***мультипликативную модель***

$$Y = A*X1^{b1}*X2^{b2}(1+\varepsilon) \qquad (7.1)$$

Обратите внимание, что возмущение входит в выражение (7.1) как часть сомножителя. После логарифмирования получим

$$ln(Y) = ln(A) + b1*ln(X1) + b2*ln(X2) + ln(\mathbf{1}+\boldsymbol{\varepsilon}),$$

или, после переопределения переменных

$$z = a + b_1V_1 + b_2V_2 + u$$

т.е. в результате логарифмирования модель стала линейной (выполнена ***линеаризация***) и задача сведена к предыдущей.

Соответствующие функции регрессии

$$\hat{Y} = A*X1^{b1}*X2^{b2}$$

$$ln(\hat{Y}) = ln(A) + b1*ln(X1) + b2*ln(X2),$$

$$\hat{\mathbf{z}} = \boldsymbol{a} + \boldsymbol{b_1V_1} + \boldsymbol{b_2V_2}$$

Этапы решения задачи:

1. Отсортируйте таблицу исходных данных по *X1* или по X2 (если хотите вычислить эластичность как функцию).

2. Постройте таблицу натуральных логарифмов *X1, X2* и *Y*,

3. Постройте корреляционную матрицу логарифмов, используя сервис *Корреляция.*

	V1	*V2*
V1	1	
V2	0,232	1
z	0,580	0,634

4. Проведите вычисления коэффициентов модели *a, b1, b2,* используя функцию *ЛИНЕЙН,* сервис *Регрессия* или *Поиск решения*. В качестве зависимой переменной используйте *z=ln(Y),* В качестве влияющих переменных выделяйте оба столбика *V1* и *V2*.

Таблица 7.5.

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
a	0,456	1,158	0,394	0,703
b1	0,343	0,162	2,112	0,064
b2	0,659	0,271	2,434	0,038

Обратите внимание, что *t*-статистики, характеризующие значимость влияющих переменных, изменились. В линейной модели они были 0,856; 1,51 и 3,97.

5. Вычислите и $\hat{Y} = exp(z\text{^})$. В данной модели коэффициенты *b1* и *b2* являются средними эластичностями *Y* по *X1* и *X2*. Обратите внимание, что их сумма почти равна единице, что предполагали Кобб и Дуглас. При этом если основные фонды и оборотные активы номинируются в денежных единицах, то и *Y* будет иметь размерность денег.

6. Постройте точечную диаграмму $\hat{Y}$, *Y*. Обратите внимание на хорошую линейную зависимость этих величин и выпадающие точки: фирмы № 5 и № 8. Фирма № 5 имеет высокий доход при малых оборотных активах. Возможно, там платят зарплату “в конвертах” или держат нелегалов-гастарбайтеров. Фирма № 8 показывает малый доход при высоких основных фондах и нормальных оборотных активах. Значит, или средства там используются неэффективно, или занижают доход. Эти фирмы представляют особый интерес для аудиторов и налоговиков.

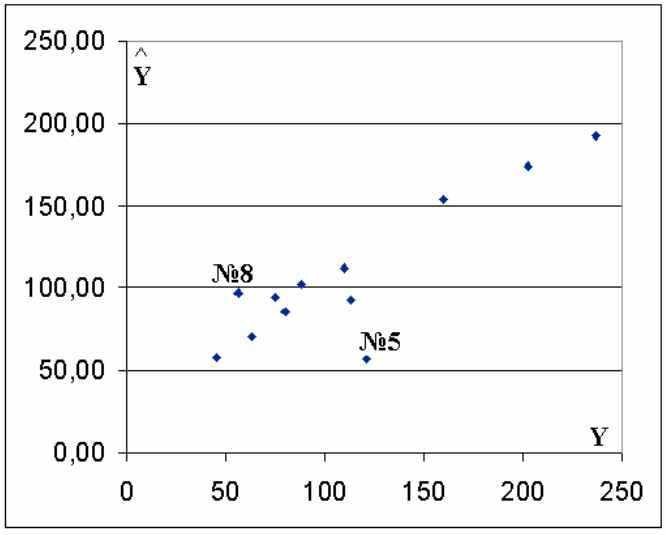


Рис. 7.3. Диаграмма *Точечная* зависимости *Y^* и *Y*.

7. Вычислите эластичность *Y* по *X1*, используя формулу

$$Э = \frac{\Delta Y}{Y} \Big/ \frac{\Delta X1}{X1}$$

В Excel используется расчётная формула

$$Э = (\hat{Y}2 - \hat{Y}1)/(\hat{Y}1 + \hat{Y}2)/(X12 - X11)*(X11 + X12),$$

Где $\hat{Y}1$ и $\hat{Y}2$ – первое и второе значения $\hat{Y}$, *X11* и *X12* – первое и второе значения *X1*. Скопируем формулу вниз, получим хаотичный набор чисел. Почему? На стр.68-69 указано, что на $\hat{Y}$ и на эластичность влияет не только *X1*, но и *X2*, и движение от точки к точке по поверхности $\hat{Y}(X1, X2)$ представляет собой пилообразную линию. Для получения "срезов" по поверхности $\hat{Y}(X1,X2)$ надо фиксировать *X2*, т.е. заполнить этот столбец одинаковыми значениями. Эластичности становятся одинаковыми и близкими к *b1*.

8. Оцените качество модели по индексу детерминации, статистике Фишера и *t*-статистикам коэффициентов. Уберите из данных фирмы № 5 и № 8, повторите настройку модели и оценку её качества.

9. Используйте модель для оптимизации плана инвестиций в основные фонды и оборотные активы. Для этого надо задать опорный план – начальные значения *X1* и *X2*, вычислить их сумму и зависимый от них $\hat{Y}$. Вызовите *"Поиск решения"*, установите целевую ячейку $\hat{Y}$*(X1план, X2план), изменяя ячейки X1план, X2план, ограничения X1план, X2план* ≥ 0, *X1план+X2план* $\leq$ заданной величины, здесь 500.

Таблица 7.6.

	Среднегодовая стоимость, млн. руб				
Номер	основных фондов *X1*	оборотных средств *X2*	Доход за год, млн.руб. *Y*	$\hat{Y}$ = *exp(z^)*	Эластичность по *X1*
3	17	54	45	57,83	0,398
2	28	56	63	70,29	0,487
4	50	63	113	92,68	-4,291
5	56	28	121	56,45	0,931
12	98	46	75	94,88	1,716
6	102	50	88	101,62	-1,602
9	114	36	80	85,02	65,895
11	115	88	160	153,71	-36,531
7	116	54	110	111,73	25,568
1	118	105	203	174,22	-11,510
8	124	42	56	96,86	3,054
10	154	106	237	192,07	
План	100	100			
Бюджет	500		R-квадрат	0,6001	
	X1+X2	200	F	6,7549	

Таблица 7.6. Продолжение

Номер	*V1= ln(X1)*	*V2= ln(x2)*	*z= ln(Y)*	*z^*
3	2,83	3,99	3,81	4,06
2	3,33	4,03	4,14	4,25
4	3,91	4,14	4,73	4,53
5	4,03	3,33	4,80	4,03
12	4,58	3,83	4,32	4,55
6	4,62	3,91	4,48	4,62
9	4,74	3,58	4,38	4,44
11	4,74	4,48	5,08	5,04
7	4,75	3,99	4,70	4,72
1	4,77	4,65	5,31	5,16
8	4,82	3,74	4,03	4,57
10	5,04	4,66	5,47	5,26
План	4,61	4,61		5,07

Аналогичным образом исследуйте модель Кобба-Дугласа

$$\hat{Y} = A_0 * K^a * L^b, \qquad (7.2)$$

где $\hat{Y}$ – выпуск продукции, K – затраты на основные фонды (капитал), L – затраты на труд. В таблице приведены данные в процентах к 1899 году.

Таблица 7.7.

Год	*K*	*L*	*Y*		Год	*K*	*L*	*Y*
1899	100	100	100		1911	216	145	153
1900	107	105	101		1912	226	152	177
1901	114	110	112		1913	236	154	184
1902	122	118	122		1914	244	149	169
1903	131	123	124		1915	266	154	189
1904	138	116	122		1916	298	182	225
1905	149	125	143		1917	335	196	227
1906	163	133	152		1918	366	200	223
1907	176	138	151		1919	378	193	218
1908	185	121	126		1920	407	193	231
1909	198	140	155		1921	417	147	179
1910	208	144	159		1922	431	161	240

Исследуйте модель по всему временному интервалу 1899-1922 г.г. и по его первой, второй и третьей части, по интервалу 1899 – 1914 г.г. и сравните

полученный прогноз на 1915-22 г.г. с реальными значениями *Y*, исследуйте в предположении $b = 1 - a$. Постройте графики *K, L, Y, Ŷ*.

Исследуйте модель с использованием *Поиска решения* и нелинеаризованной функции (7.2). Используйте разные начальные значения коэффициентов, и вы получите разные решения; при этом графики $\hat{Y}$ будут приблизительно совпадать. Это связано с ***коллинеарностью***, то есть взаимной зависимостью *K* и *L,* а также с алгоритмами, используемыми в *Поиске решения* (Ньютона и др.), которые ищут минимум функции $L=\Sigma e^2$, двигаясь от начальных значений. Но у нелинейной функции может быть несколько минимумов, и компьютер находит решение, ближайшее к начальным значениям.

7.2. Задача с высокой мультиколлинеарностью

Следующая задача – одна из первых эконометрических задач. В ней исследуется зависимость потребления бройлеров в Англии в 20-е – 30-е годы в зависимости от среднедушевого дохода и цены курятины, говядины и свинины. Данные можно считать “панельными” (panel data), так как все переменные фактически зависят от времени. Было предложено и исследовано несколько моделей:

1. Функция спроса $\hat{Y} = b0 * X2^{b1}$
2. Функция потребления $\hat{Y} = b0 * X1^{b1}$
3. Функция спроса-потребления $\hat{Y} = b0 * (X2/X1)^{b1}$
4. Модель спроса на несколько товаров $\hat{Y} = b0*X2^{b2}*X3^{b3}*X4^{b4}$

Мы используем мультипликативную модель, как в предыдущих задачах:
$\hat{Y} = b0* X1^{b1} *X2^{b2}*X3^{b3}*X4^{b4}$. Последние 4 строки не используйте для проведения вычислений. Мы их используем для оценки адекватности модели.

Этапы исследования модели:

1. Построить корреляционную матрицу по всем переменным, включая время. Построить графики всех переменных в зависимости от времени. Выбрать вид модели.

2. Выбрать мультипликативную модель и линеаризовать её логарифмированием:

$Ln\ \hat{Y} = Ln\ b0 + b1*LnX1 + b2*LnX2 + b3*LnX3 + b4*LnX4$

после переобозначения

$Z^{\wedge} = a + b1V1 + b2V2 + b3V3 + b4V4$

3. Построить корреляционную матрицу

Таблица 7.8.

	t	*V1*	*V2*	*V3*	*V4*	*Z*
t	1					
V1	0,995	1				
V2	0,879	0,882	1			
V3	0,926	0,932	0,968	1		
V4	0,983	0,973	0,898	0,938	1	
Z	0,924	0,912	0,661	0,774	0,877	1

Обратите внимание на высокие коэффициенты корреляции всех переменных. Это называется ***мультиколлинеарность*** и приводит к существенному росту погрешности коэффициентов модели. Если вспомнить, что эти коэффициенты являются эластичностями результата по влияющим переменным, то становится понятно, что мультиколлинеарность может привести к существенным ошибкам при планировании.

4. Постройте графики логарифмов всех переменных. Как видите, для логарифмов можно использовать линейную модель.

5. Получить коэффициенты *a, b1, b2, b3, b4* , R^2 , *F* используя функцию ЛИНЕЙН, сервис *Регрессия* или *Поиск решения*. Расчёты проводить по логарифмам, как в прошлой задаче. 4 последних строки не использовать!

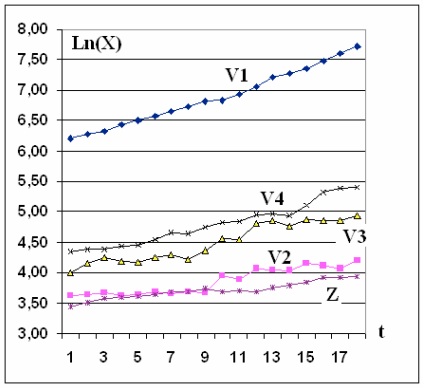


Рис.7.4. Логарифмы переменных.

Таблица 7.9.

№ п/п	Средне-душевой доход	Стоимость 1 фунта цыплят	Стоимость 1 фунта свинины	Стоимость 1 фунта говядины	Потреб-ление цыплят	
t	***X1***	***X2***	***X3***	***X4***	***Y***	$\hat{Y}$
1	492,9	37,3	54,7	77,4	31,2	33,20
2	528,6	38,1	63,7	80,2	33,3	34,62
3	560,3	39,3	69,8	80,4	35,6	34,88
4	624,6	37,8	65,9	83,9	36,4	35,76
5	666,4	38,4	64,5	85,5	36,7	35,62
6	717,8	40,1	70	93,7	38,4	37,09
7	768,2	38,6	73,2	106,1	40,4	40,43
8	843,3	39,8	67,8	104,8	40,3	38,89
9	911,6	39,7	79,1	114	41,8	41,98
10	931,1	52,1	95,4	124,1	40,4	39,97
11	1021,5	48,9	94,2	127,6	40,7	41,63
12	1165,9	58,3	123,5	142,9	40,1	42,92
13	1349,6	57,9	129,9	143,6	42,7	43,67
14	1449,4	56,5	117,6	139,2	44,1	42,56
15	1575,5	63,7	130,9	165,5	46,7	44,59
16	1759,1	61,6	129,8	203,3	50,6	49,71
17	1994,2	58,9	128	219,6	50,1	52,47
18	2258,1	66,4	141	221,6	51,7	50,85

Таблица 7.9. Продолжение

Л о г а р и ф м ы					
V1	***V2***	***V3***	***V4***	***Z***	***Z^***
6,20	3,62	4,00	4,35	3,44	3,50
6,27	3,64	4,15	4,38	3,51	3,54
6,33	3,67	4,25	4,39	3,57	3,55
6,44	3,63	4,19	4,43	3,59	3,58
6,50	3,65	4,17	4,45	3,60	3,57
6,58	3,69	4,25	4,54	3,65	3,61
6,64	3,65	4,29	4,66	3,70	3,70
6,74	3,68	4,22	4,65	3,70	3,66
6,82	3,68	4,37	4,74	3,73	3,74
6,84	3,95	4,56	4,82	3,70	3,69
6,93	3,89	4,55	4,85	3,71	3,73
7,06	4,07	4,82	4,96	3,69	3,76
7,21	4,06	4,87	4,97	3,75	3,78
7,28	4,03	4,77	4,94	3,79	3,75
7,36	4,15	4,87	5,11	3,84	3,80
7,47	4,12	4,87	5,31	3,92	3,91
7,60	4,08	4,85	5,39	3,91	3,96
7,72	4,20	4,95	5,40	3,95	3,93

Последовательно исключайте из модели цены на говядину, свинину, а затем и курятину. Должны получиться следующие результаты:

Таблица 7.10.

		1		2		3		4	
	R^2 частн	***Коэф***	***t***	***Коэф.***	***t***	***Коэф.***	***t***	***Коэф***	***t***
a		2,377	8,36	2,406	8,807	2,153	12,3	1,898	*8,32*
b1	0,39	0,313	2,86	0,373	5,648	0,424	8,39	0,261	*7,69*
b2	0,28	-0,55	-2,92	-0,544	-2,94	-0,357	-3,6		
b3	0,03	0,168	1,05	0,183	1,189				
b4	0,07	0,115	0,68						
R^2		0,937		0,933		0,924		0,832	
R^2 норм		0,909		0,914		0,91		0,818	
F		*33,32*		*46,72*		*66,86*		*59,26*	

Обратите внимание, что коэффициент корреляции *Cor(Z,V2)* = 0,661, то есть положительный, а коэффициент *b2* – отрицательный, что правильнее отражает взаимосвязь потребления курятины и её цены. Здесь проявилась ***ложная корреляция***, связанная с ***коинтеграцией:*** все переменные модели растут со временем, и только регрессионный анализ позволяет выделить истинное взаимное влияние переменных. *t*-статистики коэффициентов *b3* и *b4* незначительны, и мы не можем принять гипотезу о влиянии цен на свинину и говядину на потребление цыплят. Последовательное исключение из модели говядины и свинины приводит к росту *F*-статистики, то есть качества модели, а исключение цен на цыплят приводит к уменьшению *F*-статистики и коэффициента детерминации R^2 .

В таблицу результатов 7.10 включены нормированные, или скорректированные коэффициенты детерминации R^2*норм*, учитывающие поправку на число степеней свободы суммы квадратов остатков. Если это не

учитывать, то получится систематическое завышение коэффициента детерминации.

$$R^2_{норм.} = 1 - (1 - R^2)\frac{n-1}{n-m-1}$$

В таблицу также включены частные коэффициенты детерминации, характеризующие тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включённых в уравнение регрессии. Расчётная формула частного коэффициента детерминации

$$R^2 частнi = 1 - \frac{1 - R^2}{1 - R^2 безXi}$$

где $R^2безXi$ – коэффициент детерминации, вычисленный при исключённом из модели факторе *Xi*. Мы выяснили, что основным фактором, влияющим на продажу цыплят, является среднедушевой доход. Цены на цыплят также влияют на их потребление, причём негативно.

Исключение свинины и говядины приводит к смещению оценок эластичностей по доходам и цене цыплят, но погрешность прогнозов, оценённая методом Монте-Карло, уменьшается в среднем на 35%. Корреляции некоторых коэффициентов модели, полученные методом Монте-Карло, велики и, как правило, отрицательны:

Таблица 7.11. Корреляционная матрица коэффициентов уравнения регрессии

	b4	*b3*	*b2*	*b1*	*a*
b4	1				
b3	-0,118	1			
b2	-0,109	**-0,837**	1		
b1	**-0,790**	-0,312	0,234	1	
a	-0,139	**0,790**	**-0,767**	-0,242	1

В таблице 7.12 показано, как влияет на ошибки включение в модель незначимого фактора и исключение значимого.

Таблица 7.12.

	Истинная модель		
Оценка модели		$Y=\alpha+\beta_1X_1+u$	$Y=\alpha+\beta_1X_1+\beta_2X_2+u$
	$Y\hat{}=a+bX_1$	Верно	Коэф.смещены, σ коэф. неверны
	$Y=a+b_1X_1+b_2X_2$	Коэффиц. не смещены, но неэффективны σ коэф. верны	Верно

Проверка модели на адекватность осуществляется следующим образом. Ряд измерений не используются при настройке модели, затем проводится прогноз соответствующих эндогенных переменных и сравнение прогнозных и реальных значений. В случае парной регрессии можно оценить интервальное среднеквадратичное отклонение *Yпрогноз* по формуле

$$S_Y = Sост\sqrt{1+\frac{1}{N}+\frac{(X-\bar{X})^2}{\sum(X_i-\bar{X})^2}}$$

и посмотреть, попадают ли реальные значения Y в интервал $\hat{Y} \pm 2S_{Yпрогноз}$. В случае множественной регрессии, особенно при наличии мультиколлинеарности, оценить $S_{Yпрогноз}$ достаточно сложно, и лучше сравнивать графики Y и $\hat{Y}$.

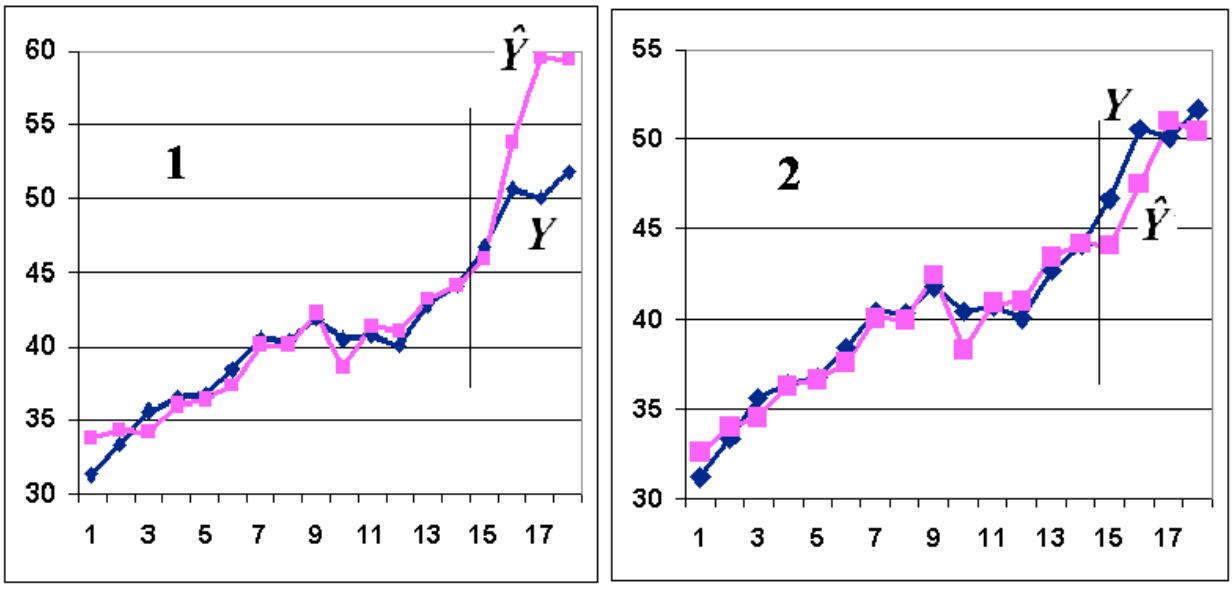


Рис.7.5. Проверка на адекватность аддитивной (1) и мультипликативной модели (2).

7.3. Мультиколлинеарность

Мультиколлинеарность – это взаимная зависимость влияющих переменных. Проблема состоит в том, что при её наличии становится сложно или невозможно разделить влияние регрессоров на зависимую переменную, и коэффициенты теряют экономический смысл предельной функции или эластичности. Дисперсии коэффициентов растут, сами коэффициенты, оценённые по различным выборкам или методом Монте-Карло, коррелируют между собой. Это приводит к тому, что в области настройки модели графики Y и $\hat{Y}$ прекрасно совпадают, R^2 и F высокие, а в области прогноза графики могут совпасть, как на Рисунке 7.5.2, что можно объяснить взаимным подавлением погрешностей (см. раздел 5), или расходятся, как на Рисунке 7.5.1, то есть модель оказывается неадекватной.

Как обнаружить мультиколлинеарность? Проще всего – по корреляционной матрице. Если коэффициенты корреляции регрессоров больше 0,7, значит они взаимосвязаны. Числовой характеристикой мультиколлинеарности может служить определитель корреляционной матрицы. Если он близок к 1, то регрессоры независимы; если к 0, значит они связаны сильно.

Как бороться с мультиколлинеарностью?

1. Смириться, принять во внимание и ничего не делать.
2. Увеличить объём выборки: дисперсии коэффициентов обратно пропорциональны количеству замеров.
3. Удалять из модели регрессоры, слабо коррелирующие с зависимой переменной, или коэффициенты которых имеют малую t-статистику. Как видно из таблицы 7.10, при этом происходит смещение коэффициентов при значимых регрессорах, и возникает вопрос об их экономическом смысле. (А смысл такой: если регрессоры коррелируют и вы можете ими управлять, например, расходы на станки и рабочих, то придётся изменять

их пропорционально). *F*-статистика, то есть качество модели, при этом растёт.

4. Использовать в уравнении регрессии агрегаты из коррелирующих переменных: линейные комбинации с коэффициентами, обратно пропорциональными стандартным отклонениям переменных и выравнивающими их масштабы. Такие агрегаты обычно не имеют экономического смысла, но могут повысить адекватность модели.
5. ***Факторный анализ***, или ***Метод главных компонент***. Используется, если переменных много, но они являются линейными комбинациями небольшого количества независимых факторов, может быть, не имеющих экономического смысла. На Рисунке 7.6 приведён пример: имеется три ортогональных вектора **Z1, Z2, Z3** и пять векторов **X1, X2, X3, X4, X5**, которые можно представить как линейные комбинации из **Z1, Z2, Z3.**

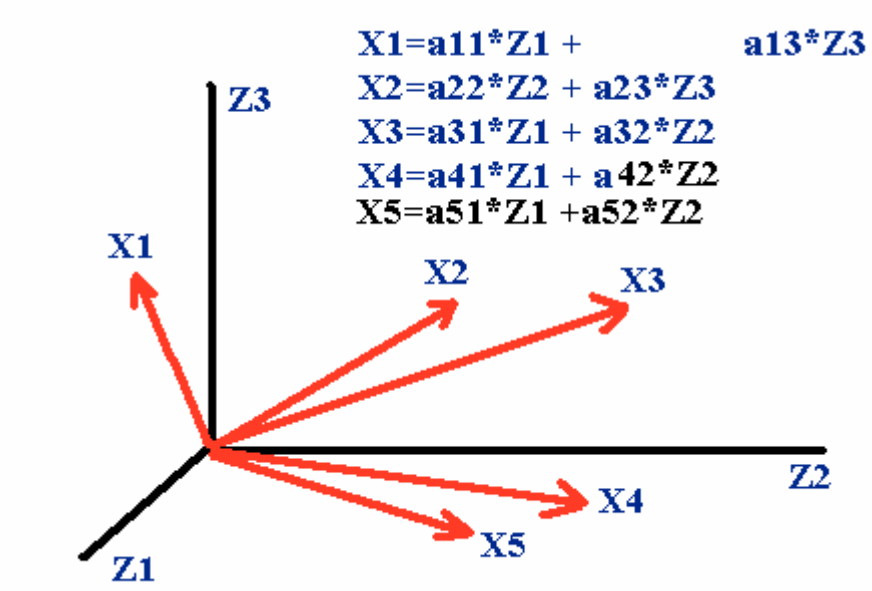


Рис.7.6. Представление векторов **X** через ортогональные векторы **Z.**

7.4. Использование Метода главных компонент для подавления мультиколлинеарности

Рассмотрим применение Метода главных компонент к задаче о торговле цыплятами. Предположим, что на экономический процесс влияют два

независимых фактора **M1** и **M2**, а компоненты векторов **V** являются линейными комбинациями их компонент:

$$\mathbf{V1^{\wedge}} = a1 + b11\ \mathbf{M1} + b12\ \mathbf{M2}$$
$$\mathbf{V2^{\wedge}} = a2 + b21\ \mathbf{M1} + b22\ \mathbf{M2}$$
$$\mathbf{V3^{\wedge}} = a3 + b31\ \mathbf{M1} + b32\ \mathbf{M2}$$
$$\mathbf{V4^{\wedge}} = a4 + b41\ \mathbf{M1} + b42\ \mathbf{M2}$$

Для оценивания коэффициентов *ai, bik* и компонент векторов **M1** и **M2** использован сервис *Поиск решения*. Для этого были заданы произвольные значения коэффициентов и компонент **M1** и **M2**, вычислено скалярное произведение (**M1, M2**), векторы **V^** и квадраты остатков $(V - V^{\wedge})^2$, которые в таблицах не показаны. Целевая ячейка – сумма квадратов остатков, которая минимизируется, изменяемые ячейки – коэффициенты и компоненты векторов **M1** и **M2**, ограничение: скалярное произведение (**M1**, **M2**) = 0, что означает ортогональность этих векторов. *Поиск решения* был запущен несколько раз с разными начальными значениями **M1** и **M2**, чтобы добиться минимума $\Sigma (V - V^{\wedge})^2$. Затем был запущен сервис *Регрессия* для оценки коэффициентов уравнения

$$\mathbf{Z^{\wedge}} = a + b1\ \mathbf{M1} + b2\ \mathbf{M2}$$

Полученные результаты:

$$\mathbf{Z^{\wedge}} = 3{,}18 + 0{,}93\ \mathbf{M1} + 1{,}26\ \mathbf{M2}$$
$$t \quad 55 \quad 9{,}7 \quad 8{,}25$$

Коэффициент детерминации R^2 = 0,9; F = 49,4. t-статистики коэффициентов существенно выше, чем при использовании обычной множественной регрессии: 8,36; 2,86; 2,92; 1,05; 0,68 , или при исключении *V3* и *V4* : 12,3; 8,39; 3,6 (см. Таблицу 7.10). Вычисленные методом Монте-Карло относительные стандартные отклонения (в процентах) четырёх прогнозных значений *Z^* для линеаризованной модели множественной регрессии: 0,64; 1,02; 1,16; 1,07; то же для метода главных компонент: 0,50; 0,66; 0,86; 0,82, то есть на 27% меньше, но

всё же хуже, чем даёт регрессия по доходу и цене цыплят: 0,52; 0,56; 0,71; 0,72 (35%).

Таблица 7.13

V1	*V2*	*V3*	*V4*	*Z*	*Z^*
6,20	**3,62**	**4,00**	**4,35**	**3,44**	**3,52**
6,27	**3,64**	**4,15**	**4,38**	**3,51**	**3,53**
6,33	**3,67**	**4,25**	**4,39**	**3,57**	**3,53**
6,44	**3,63**	**4,19**	**4,43**	**3,59**	**3,58**
6,50	**3,65**	**4,17**	**4,45**	**3,60**	**3,60**
6,58	**3,69**	**4,25**	**4,54**	**3,65**	**3,62**
6,64	**3,65**	**4,29**	**4,66**	**3,70**	**3,67**
6,74	**3,68**	**4,22**	**4,65**	**3,70**	**3,69**
6,82	**3,68**	**4,37**	**4,74**	**3,73**	**3,71**
6,84	**3,95**	**4,56**	**4,82**	**3,70**	**3,68**
6,93	**3,89**	**4,55**	**4,85**	**3,71**	**3,72**
7,06	**4,07**	**4,82**	**4,96**	**3,69**	**3,73**
7,21	**4,06**	**4,87**	**4,97**	**3,75**	**3,76**
7,28	**4,03**	**4,77**	**4,94**	**3,79**	**3,79**
7,36	**4,15**	**4,87**	**5,11**	**3,84**	**3,82**
7,47	**4,12**	**4,87**	**5,31**	**3,92**	**3,90**
7,60	**4,08**	**4,85**	**5,39**	**3,91**	**3,95**
7,72	**4,20**	**4,95**	**5,40**	**3,95**	**3,96**

Таблица 7.13. Продолжение

		M1	*M2*	*M1*M2*	*V1^*	*V2^*	*V3^*	*V4^*
		-0,13	**0,364**	**-0,05**	6,23	3,56	4,05	4,31
		-0,24	**0,443**	**-0,11**	6,29	3,63	4,16	4,35
		-0,32	**0,501**	**-0,16**	6,33	3,69	4,23	4,38
		-0,08	**0,37**	**-0,03**	6,42	3,65	4,18	4,45
		0,006	**0,326**	**0,002**	6,47	3,64	4,18	4,49
		0,014	**0,34**	**0,005**	6,56	3,69	4,25	4,56
a1	**4,769**	**0,165**	**0,268**	**0,044**	6,66	3,69	4,26	4,64
b11	**3,128**	**0,304**	**0,193**	**0,059**	6,71	3,67	4,23	4,68
b12	**5,141**	**0,265**	**0,235**	**0,062**	6,81	3,74	4,33	4,75
a2	**2,675**	**-0,12**	**0,485**	**-0,06**	6,87	3,92	4,58	4,78
b21	**1,409**	**0,06**	**0,386**	**0,023**	6,94	3,89	4,54	4,83
b22	**2,931**	**-0,21**	**0,581**	**-0,12**	7,09	4,08	4,80	4,93
a3	**2,764**	**-0,13**	**0,547**	**-0,07**	7,18	4,10	4,84	5,00
b31	**2,105**	**0,055**	**0,441**	**0,024**	7,21	4,04	4,77	5,03
b32	**4,286**	**0,038**	**0,478**	**0,018**	7,35	4,13	4,89	5,13
a4	**3,267**	**0,358**	**0,316**	**0,113**	7,51	4,11	4,87	5,26
b41	**2,302**	**0,599**	**0,192**	**0,115**	7,63	4,08	4,85	5,36
b42	**3,705**	**0,492**	**0,273**	**0,134**	7,71	4,17	4,97	5,41
			Сумма	**0**				

Проверка модели на адекватность представлена на рисунке 7.7, корреляционная матрица коэффициентов *a, b1* и *b2,* полученная методом Монте-Карло – в таблице 7.14. Таблица 7.14 объясняет, почему статистические ошибки прогнозов существенно меньше, чем у коэффициентов уравнений регрессии: отклонения одних коэффициентов компенсируется противоположными отклонениями других.

Таблица 7.14.

	b2	*b1*
b2	1	
b1	0,849	1
a	-0,987	-0,824

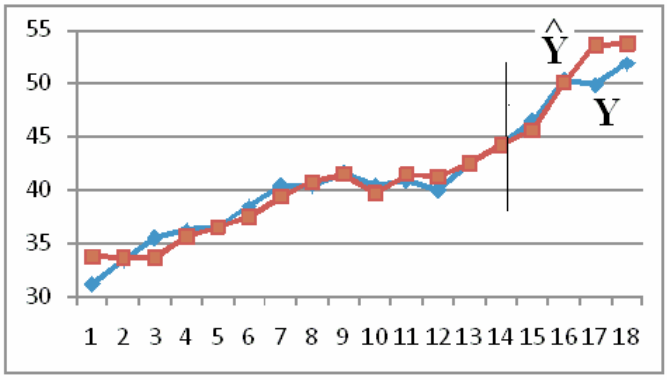


Рис.7.7. Проверка адекватности модели,
использующей Метод главных компонент

Таким образом, Метод главных компонент позволяет улучшить качество эконометрических моделей: уменьшить статистические ошибки коэффициентов и прогнозов. Его несложно реализовать в среде Excel и использовать для студенческих лабораторных работ, но здесь он представлен для ознакомления, а в реальной работе вы воспользуетесь специализированными программами.

7.5. Сложная мультипликативная модель с фиктивными переменными

Рассмотрим модель, содержащую степенные и показательные функции, а также ***фиктивные (dummy)*** переменные, позволяющие учитывать качественные показатели модели. Модель построена по реальным данным совместно со студенткой Финансового университета Е.С.Лукашенко. Исходные данные представлены в Приложении 2.

В модели учитываются следующие числовые переменные: **S** – площадь квартиры, кв.м.; **SK** – площадь кухни, кв.м.;

R – количество комнат; **M** – расстояние от метро (минут пешком);

E – количество этажей в доме.

Используются также фиктивные переменные:

Z – зона. При настройке модели использовались данные по квартирам в Москве: 1- район метро “Пушкинская”; 2 - район метро “Алексеевская”;

3 - район метро “Речной вокзал”.

H – тип дома: 1 – панельный, 2 – блочный, 3 – кирпичный, 4 – сталинский, 5 – монолитный.

F – этаж (первый или последний этаж – 0, остальные этажи 1);

C – состояние квартиры: 1 – евроремонт, 2 – косметический ремонт,
3 – новостройка (без ремонта).

Для оценки квартир в Москве построим мультипликативную модель: произведение степенных и показательных функций:

$$Y= b_0 * S^{b1*} SK^{b2*} R^{b3} *M^{b4} * E^{b5} * b_6^{Z} * b_7^{H*} b_8^{F} * b_9^{C} (1+u)$$

где Y – стоимость квартиры (в условных единицах), u – случайная величина.

Модель линеаризована логарифмированием:

$$lnY= lnb_0 + b_1 ln S + b_2 ln SK + b_3 ln R + b_4 ln M + b_5 ln E + Z ln b_6 + H ln b_7 + F ln b_8 + C ln b_9 + ln (1+u)$$

Для простоты дальнейших расчетов переименуем некоторые из наших переменных следующим образом:

$$lnY = Z ; \quad lnb_0 = a_0 ; \quad ln S = x_1; \quad ln SK = x_2; \quad ln R = x_3; \quad ln M= x_4;$$

$$ln E=x_5; \quad ln b_6 = a_6; \quad ln b_7= a_7; \quad ln b_8 = a_8 ; \quad ln b_9 = a_9; \quad ln (1+u)= e$$

После переименования переменных модель принимает вид

$$Z = a_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 + a_6 Z + a_7 H + a_8 F + a_9 C + e$$

где e – случайная величина.

Используя Сервис *“Регрессия”* или функцию ЛИНЕЙН, оцените коэффициенты, их погрешности и t-статистики, коэффициент детерминации R^2 и статистику Фишера F, а также стоимости квартир $\hat{Y}$ **=exp(Z^).** В связи с большой разницей в стоимости квартир, целесообразно вычислять относительную погрешность СТАНДОТКЛОН((Y- $\hat{Y}$)/Y) или СРЗНАЧ(ABS(Y-Y^)/Y). Постройте корреляционный график $\hat{Y}/Y$ и гистограмму частотных распределений относительных ошибок.

Контрольные вопросы

1. Мультиколлинеарность: чем плоха, как обнаружить и как бороться.
2. Выявление мультиколлинеарности по матрице корреляции экзогенных переменных
3. Что такое и почему возникает ложная корреляция и коинтеграция?

4. Расчётная формула частного коэффициента детерминации

5. Применение статистик Стьюдента и Фишера в процедуре подбора переменных в модели множественной регрессии

6. Ошибки от включения в модель незначимых переменных или исключения значимых.

8. Исследование временных рядов

Эконометрическую модель можно построить, используя три типа исходных данных:

- данные, характеризующие совокупность различных объектов в определенный момент (период) времени: ***cross sectional data***, “пространственные”;
- данные, характеризующие один объект за ряд последовательных моментов (периодов) времени: ***временные ряды, time series;***
- данные, характеризующие совокупность различных объектов за ряд последовательных моментов времени: ***panel data***, “панельные”.

Временной ряд – это совокупность значений какого-либо показателя за несколько последовательных моментов (периодов) времени. Он формируется под воздействием большого числа факторов, которые можно условно подразделить на три группы:

- факторы, формирующие тенденцию (***тренд***) ряда;
- факторы, формирующие ***циклические*** колебания ряда, например сезонный, недельный; для рядов цен на фондовом рынке характерны ***непериодические колебания;***
- ***случайные*** факторы.

В большинстве случаев значения временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент.

В таблице 8.1 и на рисунке 8.1 приведены уровни розничной торговли в России, млрд. руб., в 2002-03 г.г. Наиболее простой и достаточно точный способ прогноза – использование *автозаполнения* ячеек Excel. Для этого надо выделить оба столбца данных, поставить курсор на черный квадратик в правом

нижнем углу выделенной зоны, чтобы курсор превратился в черный тонкий крест, нажать левую клавишу мыши и протащить курсор на требуемое количество ячеек вправо.

Таблица 8.1.

			Прогноз протяжкой		
Месяц	2002	2003	2004	2005	2006
Январь	270,1	324	378,3	432,4	486,5
Февраль	267,1	322	376,9	431,8	486,7
Март	288,2	351	413,8	476,6	539,4
Апрель	292,7	353,8	414,9	476	537,1
Май	291	352,6	414,2	475,8	537,4
Июнь	297,8	356,5	415,2	473,9	532,6
Июль	310,1	368	425,9	483,8	541,7
Август	324,3	379	433,7	488,4	543,1
Сентябрь	326,1	386	445,9	505,8	565,7
Октябрь	339,4	403,4	467,4	531,4	595,4
Ноябрь	346,1	409,7	473,3	536,9	600,5
Декабрь	400,7	477,3	553,9	630,5	707,1

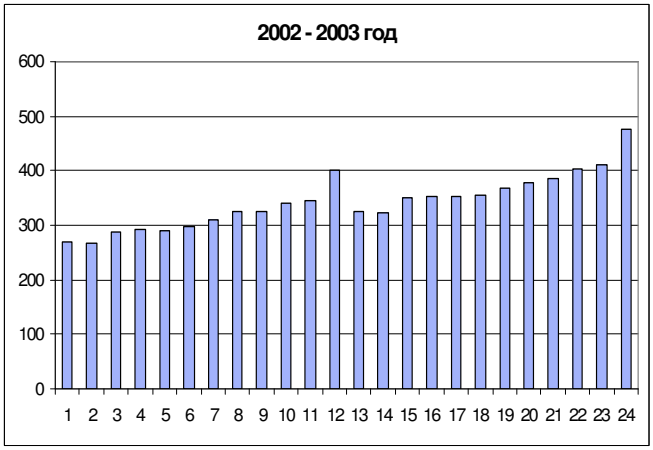


Рис.8.1

Обычно применяется более сложная технология: построение аналитической функции для моделирования тенденции (тренда) временного ряда, или аналитическое выравнивание временного ряда. Для этого чаще всего применяют следующие функции:

- Линейная $Y(t) = a + b * t$;
- Гипербола $Y(t) = a + b / t$;
- Экспонента $Y(t) = exp(a + b*t)$;
- Степенная функция $Y(t) = a * t^{b}$;
- Парабола $Y(t) = a + b_1*t + b_2*t^{2}$

Гиперболу можно линеаризовать заменой $z=1/t$, экспоненту и степенную функцию – логарифмированием, в параболе t и t^2 рассматривать как отдельные переменные множественной регрессии. Тогда параметры трендов можно оценивать обычными средствами МНК: функция ЛИНЕЙН, сервис *Регрессия*. В качестве независимой переменной выступает время $t = 1, 2, \ldots, n$, а в качестве зависимой переменной – уровни (значения) временного ряда $Y(t)$.

Критерием отбора наилучшей формы тренда является наибольшее значение коэффициента детерминации

$$R^2 = 1 - \text{ДИСП ост.} / \text{ДИСП } Y$$

и соответствующей статистики Фишера.

8.1. Временной ряд с сезонными колебаниями

Далее рассмотрена технология расчетов с использованием метода отклонений от тренда, предполагающего вычисление трендовых значений и расчет отклонений от трендов. Для дальнейшего анализа используют не исходные данные, а отклонения от тренда.

В таблице 8.2 и на графике 8.2 представлены месячные значения реального располагаемого денежного душевого дохода в России в 2001 – 2003г.г. в процентах к декабрю 2000 г. Требуется дать прогноз по месяцам 2004 – 2005 г.г. Для этого надо расположить значения по годам в одну строку (или столбец), построить по ним график с линией тренда (щелкнуть правой клавишей мыши по точке графика, *Добавить линию тренда, Параметры, Показывать уравнение на диаграмме)*. Ввести строку со сплошной нумерацией месяцев: $t = 1, 2, 3, \ldots, 60$ и по коэффициентам a и b уравнения на диаграмме построить линейный тренд $\hat{Y} = a + b*t$. Для 2001 – 2003 г.г. вычислить относительные отклонения $\eta = (Y - \hat{Y}) / \hat{Y}$. Чтобы получить средние η по месяцам 2001 – 2003г.г., надо скопировать эти значения за 2002 год и вставить их в строку под значениями η 2001 года, используя *Специальная вставка – Значения*, затем так же скопировать значения η 2003 года и вставить их в строку под значениями η 2002 года, чтобы получились три строки η за три года. Вычислите средние значения η по месяцам, используя функцию СРЗНАЧ. Скопируйте полученную строку под тренд 2004 и под тренд 2005 г.г., используя *Специальная вставка – Значения.* Прогнозные значения $\hat{Y}$ по месяцам 2004 – 2005 г.г. получим по формуле $\hat{Y}прогноз = \hat{Y} *(1 + \eta)$. Технология решения задачи несложная, но требует внимания.

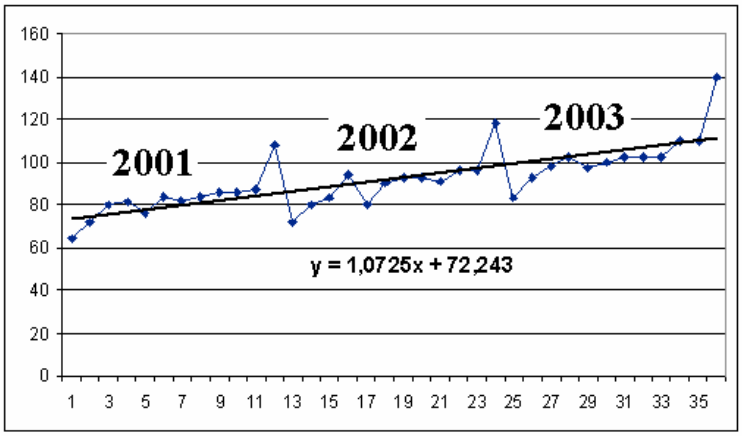


Рис.8.2

Возможны другие технологии прогнозирования с использованием трендов. В приведенном примере линия тренда проводится по трем суммарным значениям *Y* за 2001, 2002 и 2003 годы. Затем вычислены вклады каждого месяца в годовую сумму (в процентах) и соответствующие средние значения за 3 года. Эти же значения можно получить по-другому: просуммировать значения по месяцам за 2001 – 2003 годы и разделить их на сумму за 3 года. Результаты получаются примерно одинаковые. Прогнозные значения по месяцам 2004 – 2005 г.г. получаем, умножая полученные средние η по месяцам на полученные по трендам суммарные значения за 2004 и 2005 годы. При решении задачи автозаполнением результаты получаются примерно те же.

Таблица 8.2.

Год	**январь**	**февр**	**март**	**апрель**	**май**	**июнь**	**июль**	**август**	**сентяб.**	**октяб.**	**нояб**	**дек.**		
2000												100	Сум-ма	Тренд
2001	64	72	80	81	76	84	82	84	86	86	87	108	991	980,5
2002	72	80	83	94	80	90	93	93	91	96	96	118	1088	1105
2003	83	93	98	102	97	100	102	102	102	110	110	140	1242	1230
2004	89,18	99,77	106,5	112,9	103	112	113	114	114	119	119	149		1354
2005	97,38	108,9	116,3	123,3	112,5	122	123	124	124	130	130	162,7		1479
	Суммы по месяцам 2001-2003 г.г.													
	219	245	261	277	253	274	277	279	279	292	293	366	3321	
	% от суммы за год													
	6,458	7,265	8,073	8,174	7,669	8,48	8,27	8,48	8,68	8,68	8,78	10,9		
	6,618	7,353	7,629	8,64	7,353	8,27	8,55	8,55	8,36	8,82	8,82	10,85		
	6,683	7,488	7,89	8,213	7,81	8,05	8,21	8,21	8,21	8,86	8,86	11,2		
	Средние по месяцам													

	6,586	7,369	7,864	8,342	7,611	8,27	8,34	8,41	8,42	8,79	8,82	11,0		

Во временных рядах часто последующие значения зависят от предыдущих, т.е. имеет место ***автокорреляция в остатках.*** В разделе 3.4 была рассмотрена процедура расчета коэффициента автокорреляции как корреляции рядов значений с номерами 1, 2, ..., *n* - 1 и 2, 3, ... , *n*. Коэффициенты автокорреляции более высоких уровней *m* вычисляют как коэффициенты корреляции рядов значений с номерами 1, 2, ..., *n* - *m* и *m*, *m* + 1, *m* + 2, ... , *n*. Последовательность коэффициентов автокорреляции уровней первого, второго и т. д. порядков называют ***автокорреляционной функцией временного ряда.*** График зависимости ее значений от величины ***лага*** (порядка коэффициента автокорреляции) называется ***коррелограммой***. Анализ автокорреляционной функции и коррелограммы позволяет определить ***лаг*** (временной интервал), при котором автокорреляция наиболее высокая, следовательно, лаг, при котором связь между текущим и предыдущими уровнями ряда наиболее тесная, т. е. при помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда. Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка ***m***, ряд содержит циклические колебания с периодичностью в *m* моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать предположение относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний, либо ряд содержит сильно нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

Технология построения коррелограммы такова. Если мы имеем ***n*** уровней ряда (данных) и хотим построить коррелограмму до уровня ***m***, то надо вызвать функцию КОРРЕЛ, в верхнем окне указать диапазон ячеек 1 : ***n-m-1*** и “задолларить” его, нажав горячую клавишу F4. В нижнем окне указать диапазон 2: ***n-m***. Скопировать функцию на ***m*** ячеек и построить диаграмму.

Данные Таблицы 8.2 надо предварительно скопировать в одну строку. Функция выглядит =КОРРЕЛ($C$7:$Z$7 ; D7:AA7), данные расположены в ячейках C7 : AL7. Автокорреляция 12-го порядка близка к 1, то есть через год всё повторяется.

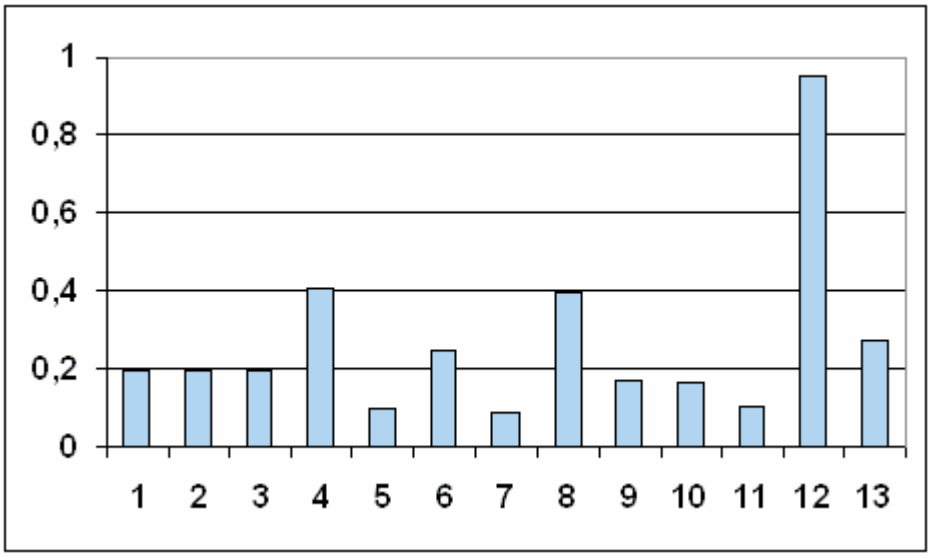


Рис.8.3

Автокорреляционный анализ используется для анализа временных рядов цен биржевых инструментов (акций и т.п.) и валют.

8.2. Ряды цен на фондовом рынке

В настоящее время используются различные методы прогнозирования на фондовом рынке, основанные на анализе временных рядов цен биржевых инструментов и индексов. Известны закономерности их формирования: тренды, непериодические колебания и статистические флуктуации. Предполагается, что после вычитания трендов остаётся зависимость между историческими данными и будущими уровнями временного ряда, то есть автокорреляции различных порядков в остатках. Это значит, что предпосылка теоремы Гаусса-Маркова нарушена, и требуются другие методы: технический анализ фондового рынка. В частности, для прогноза по волнообразным колебаниям уровней цен пытались применить ряды Фурье, существуют ассоциации любителей торговать

по волнам Эллиотта, на них основан метод Брауна: аппроксимация части временного ряда прямой линией, синусом и косинусом.

Но возникает вопрос: насколько все эти методы обоснованы, каковы предпосылки их использования, какова вероятность правильного прогноза? Ответ на этот вопрос может дать изучение автокорреляций высоких порядков во временных рядах после вычитания трендов, иначе автокорреляции будут близки к 1.

Проанализируем временные ряды цен на фондовом рынке с использованием автокорреляций высоких порядков с целью обоснования методов прогнозирования с применением волнообразных колебаний. Применяемый алгоритм:

1. Считывание временных рядов из торговой системы, например, FINAM.
2. Оценка целесообразности работы с рядом, отбраковка рядов с резкими скачками цен и с отсутствием волнообразных колебаний.
3. Вычитание трендов.
4. Проверка остатков на стационарность (см.раздел 8.3).
5. Расчёт автокорреляций и построение коррелограмм
6. Классификация рядов и коррелограмм.
7. Оценка целесообразности применения методов технического анализа с использованием волнообразных колебаний.
8. Аппроксимация участка ряда функцией с синусоидой.
9. Оценка точности аппроксимации и прогноза.

Этапы обработки одного из рядов представлены на Рисунках 8.4 и 8.5.

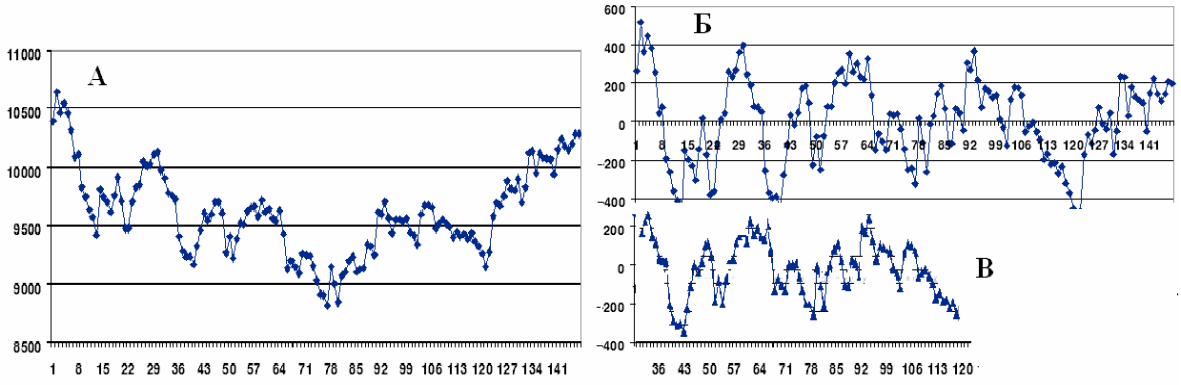


Рис.8.4. График индекса NIKKEI, (первое значение - 6 мая 2010 г.) **А**,

график индекса NIKKEI с вычтенными трендами **Б**,
график индекса NIKKEI с вычтенными трендами и со смещением на 30 дней **В**.

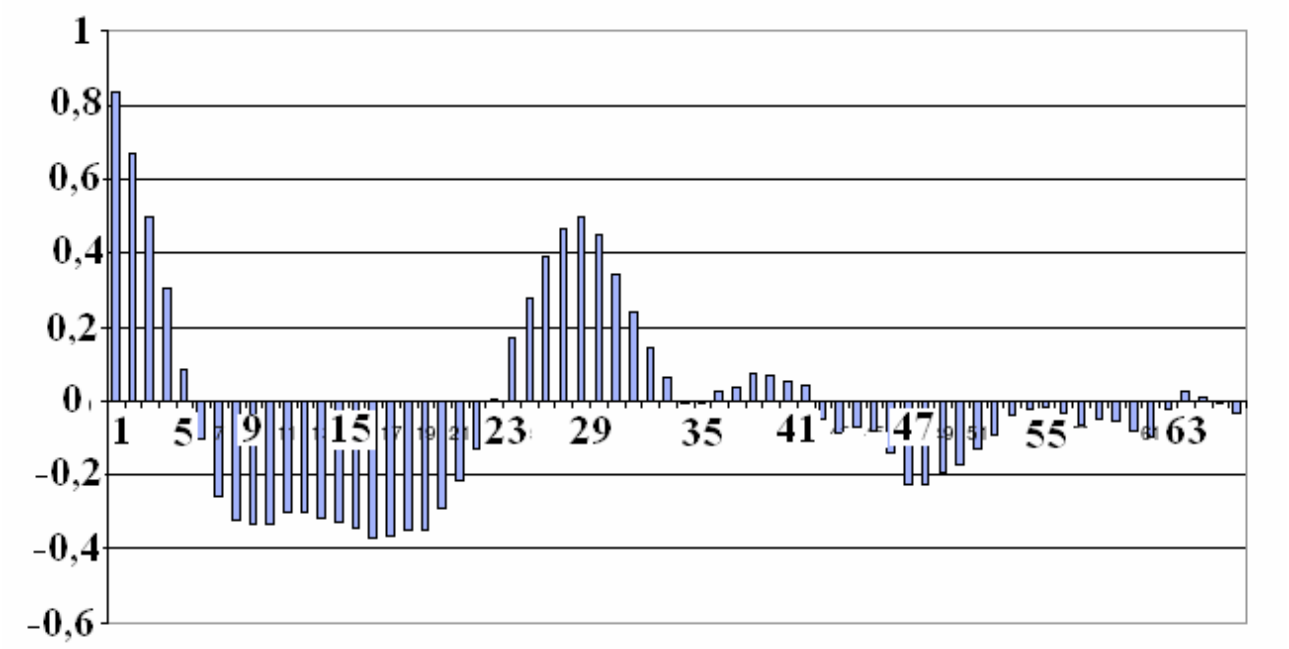


Рис.8.5. Коррелограмма ряда индекса NIKKEI с вычтенными трендами.

На Рисунках 8.4Б и 8.4В представлены графики индекса NIKKEI с вычтенными трендами: исходный и со смещением на 30 дней. Коэффициент автокорреляции *R(30) = 0,495*. Видно, что графики до 85 дня довольно хорошо совпадают.

Исследования [6] показали, что полученные по разным временным рядам коррелограммы можно классифицировать, разбив на три группы, причём коррелограммы внутри групп 1 и 2 имеют примерно одинаковый вид. Основным признаком деления служит первый и второй нули коррелограммы. Выделены три группы с порядками нулевой автокорреляции:

1) 6-8 и 16-25;

2) 11-13 и 21-25;

3) второй ноль больше 30.

Вторые нули, а также минимумы и максимумы группируются не так кучно. Некоторые типичные коррелограммы представлены на Рис. 8.6.

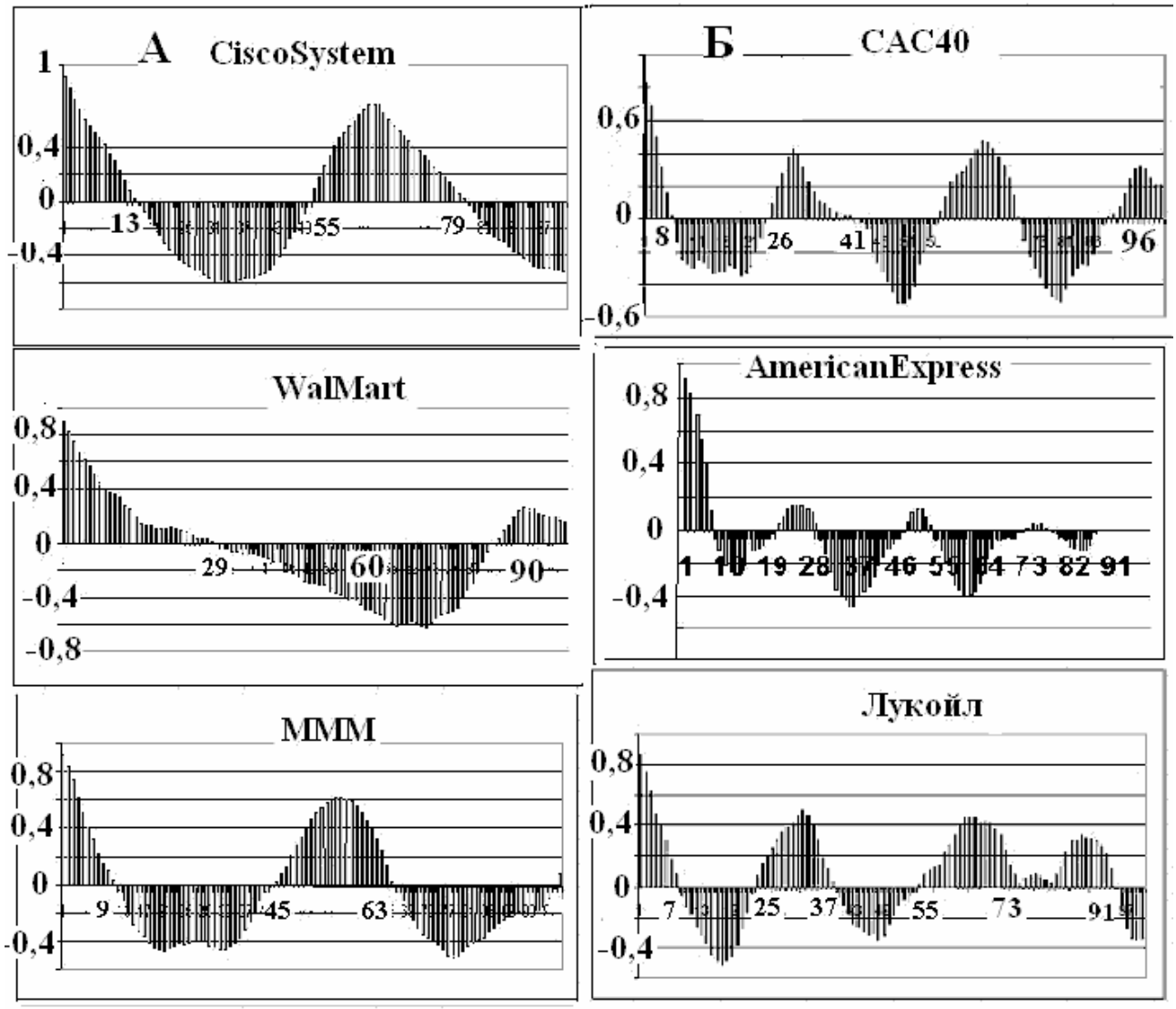


Рис.8.6. Типичные коррелограммы, полученные при обработке временных рядов: с широкими волнами **А** и с узкими **Б**.

Из всего вышеизложенного вытекает алгоритм прогнозирования цен на фондовом рынке:

1) отбросить ряды с резкими бросками цен;

2) вычесть из ряда тренды, используя средства Excel;

3) построить график остатков и оценить стационарность этого ряда;

4) построить коррелограмму по ряду остатков;

5) проанализировать вид коррелограммы; если первый ноль в районе 5-8 и второй 16-25, можно применить синусоидальную аппроксимацию;

6) если перед сегодняшним днем на графике цен или остатков видны 1,5 - 3 волны, целесообразно применить синусоидальную аппроксимацию: постройте для области настройки функцию

$$\hat{Y}(t) = a + b\,t + d\,Sin(\omega t+\varphi)$$

(модифицированная модель Брауна) ,

где $\hat{Y}(t)$- значение аппроксимирующей функции

t - время (день, час и др.)

a, b, d, ω, φ – коэффициенты аппроксимирующей функции.

Для оценки коэффициентов используется метод наименьших квадратов с применением сервиса Excel “Поиск решения”. Далее приведён пример аппроксимации части временного ряда индекса NIKKEI:

Таблица 8.3

Время	**Цена**	***Y^ (t)***	**(*Y^(t)*-Цена)^2**		**Изменяемые ячейки**
23	9473,48	9851,9	37550,6	**a**	**10097,08**
24	9476,78	9839,1	21763,6	**b**	**-11,55**
25	9701,32	9826,2	6545,97	**d**	**-263,92**
26	9822,98	9813,4	28529,7	**w**	**0,217**
27	9845,65	9800,6	16518,7	**f**	**4,24**
28	10045,95	9787,7	62372,1		
29	10009,25	9774,9	17992,2		
57	9618,24	9415,747			
58	9653,51	9402,918			
		Целевая функция:	**Сумма 1120790,6**		

Вначале коэффициенты задаются произвольно (“опорный план”) и проводится вычисление функции $\hat{Y}(t)$ в разумном диапазоне значений цен и на прогнозируемый период времени. Под “разумным диапазоном” следует понимать временной диапазон, в котором не было резких скачков цен и изменений тренда, и можно увидеть 1,5 – 2 волны. Обычно это 30-50 точек независимо от Δt. Затем вычисляется сумма квадратов отклонений ($\hat{Y}(t)$-Цена)^2,

которая является целевой минимизируемой функцией изменяемых коэффициентов. Скорее всего, первая итерация даст плохой результат для коэффициента **ω** (видно на графике: волны или мелкие, или очень длинные), и его надо изменять вручную, запуская затем *"Поиск решения"*. Это связано с тем, что временной ряд представляет собой суперпозицию непериодических колебаний, в которых можно найти широкий спектр частот, и компьютер находит частоту, ближайшую к исходному значению.

Рис. 8.7. Нелинейная аппроксимация участка ряда NIKKEI.

Методика была проверена на графиках цен, взятых с сайта rts.ru . Для аппроксимации использовались 65–100 точек (на Рисунке 8.8 – до вертикальной черты), а сопоставление графика функции с реальными ценами в правой части диаграммы дает представление о точности прогноза. При значении R^2 в интервале настройки более 0,7 его значения в интервале прогноза также достаточно велики, обычно более 0,5

В целом, метод позволяет угадывать движение цены до 10 периодов с вероятностью более 50 %, но фаза третьей, а тем более четвертой волны обычно сдвигается, что приводит к ошибочным прогнозам.

Проведены эксперименты с включением в модель члена ct^2. В некоторых случаях он может повысить точность прогноза при наличии явных изгибов, но

велика вероятность сильного расхождения прогнозных и реальных цен. Это связано с резким ростом дисперсий коэффициентов из-за увеличения количества регрессоров и корреляции t и t^2.

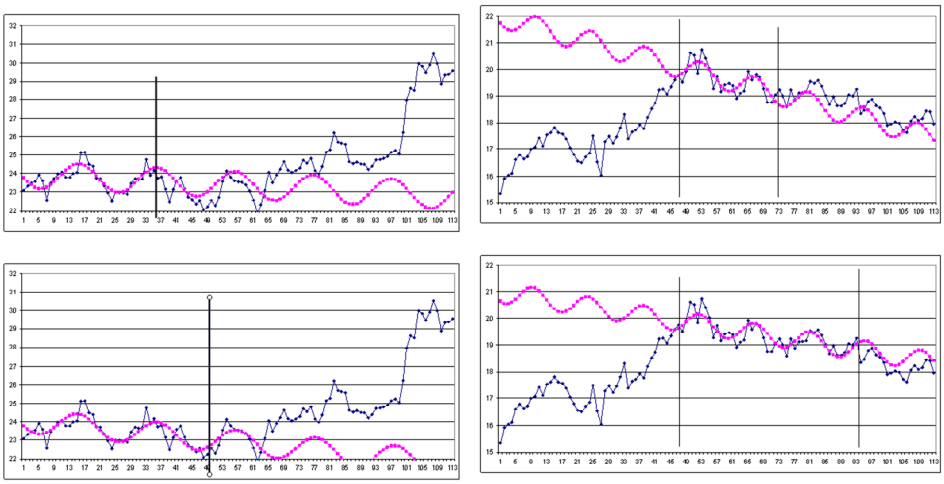


Рис. 8.8 Примеры использования синусоидальной аппроксимации для прогноза цен на фондовом рынке.

Очень интересный результат – группировка угловых частот синусоидальной аппроксимации **ω**. Они чётко разделяются на три группы: **ω** = 0,112 ± 0,013, **ω** =0,25 ± 0,03 , **ω** =0,48 ± 0,04. При первом нуле коррелограммы 6-7 и втором не более 25 на графике цен обычно прослеживаются волны с **ω** = 0,25, то есть с периодом волны 25. При первом нуле коррелограммы 11-13 и втором не более 25 на графике цен также прослеживаются волны, но их периоды могут существенно различаться. В частности, такое сочетание может указывать на наличие длинных волн, с **ω** = 0,1 то есть с периодом волны 63. Совсем непредсказуемыми становятся периоды волн и целесообразность их использования, если первый ноль расположен >15, второй ноль > 40. При этом компьютер обычно выделяет высокочастотные колебания, соответствующие не волнам, а статистическим флуктуациям, что непригодно для прогнозирования.

Автокорреляции высоких порядков в остатках связаны с ***фрактальной структурой*** рядов цен на фондовом рынке, то есть с их ***подобием в разных масштабах времени.*** Фрактальная природа рынков капитала порождает циклы,

тренды и множество справедливых (равновесных) цен [7]. Изучение фракталов не входит в нашу задачу, но проявление их и автокорреляций высоких порядков позволяют говорить о принципиальной возможности прогнозирования на фондовом рынке. Заметим, что на валютном рынке показатель фрактальности (Хёрста) существенно ниже [7], соответственно, прогнозы колебаний цен менее достоверны.

8.3. Стационарные и нестационарные стохастические процессы

Рассмотрим некоторые теоретические понятия и модели, применяемые для анализа рядов цен на фондовом рынке.

Временной ряд – это конечная реализация ***стохастического процесса***: генерации набора случайных переменных *Y(t).*

Стохастический процесс может быть стационарным и нестационарным.

Процесс является ***стационарным***, если

1. Математическое ожидание значений переменных не меняется.
2. Математическое ожидание дисперсий переменных не меняется.
3. Нет периодических флуктуаций.

Распознавание стационарности:

1. График: систематический рост или убывание, волны и зоны высокой волатильности (дисперсии) в длинном ряде сразу видны.
2. Автокорреляция (убывает при росте лага)
3. Тесты тренда: проверка гипотезы о равенстве нулю коэффициента при *t*.
4. Специальные тесты, включённые в пакеты компьютерных программ Stata, EViews и др., например, тест Дики-Фуллера (Dickey-Fuller) на единичный корень (Unit root).

Чисто случайный процесс, стационарный с отсутствием автокорреляции ($\mathrm{Cor}(u_i/u_k) = 0$) называется ***Белый шум.***

Пример нестационарного процесса – ***случайное блуждание***

$$Y(t) = Y(t-1) + a(t)$$

где $a(t)$ – белый шум.

Интересно, что процесс

$$Y(t) = 0{,}999*Y(t-1) + a(t)$$

является стационарным. Воспроизведите эти процессы на компьютере, добавляя к предыдущему члену ряда случайную величину с нулевым математическим ожиданием, например:

$$Y(t) = 0{,}999*Y(t-1) + (СЛЧИС() – 0{,}5)$$

Функция СЛЧИС создаёт случайные числа в диапазоне 0…1. Постройте коррелограммы и сравните их с коррелограммами цен на фондовом рынке.

Принципиальную возможность избавиться от нестационарности почему-то называют ***интегрируемость.*** Применяют различные способы избавления от нестационарности:

1. Вычитание тренда, что мы и делали в предыдущем разделе;
2. Использование разностей 1-го, 2-го и т.д. порядков, что можно делать только после сглаживания временного ряда (или энергетического спектра), иначе все эффекты будут подавлены статистическими флуктуациями: дисперсия разности равна сумме дисперсий.

Для исследования рядов цен на фондовом рынке применяются модели, использующие белый шум и авторегрессию, то есть взаимную зависимость уровней временного ряда.

Модель MA(q) (moving average) – линейная комбинация последовательных элементов белого шума

$$X(t) = a(t) – K(1)*a(t-1) – …. – K(q)*a(t-q)$$

Модель AR(p) (авторегрессия): линейная комбинация лаговых переменных

$$X(t) = b_0 + b_1*X(t-1) + …. + b_p*X(t-p)$$

Особенно популярны их комбинации

$$ARMA(p,q) = AR(p) + MA(q)$$

и $ARIMA(p, i ,q)$: то же, с интегрируемостью i –го порядка.

Особый интерес представляет модель ***векторная авторегрессия VAR***, состоящая из многих уравнений, в которой левые (эндогенные) переменные зависят и от своих, и от чужих лаговых значений. Используя этот метод, профессор Кристофер Симз получил в 2012 году Нобелевскую премию за изучение влияния на экономику эффектов от единовременных потрясений и действий регуляторов, в частности, изменения процентных ставок центробанков. Согласно его модели, негативные эффекты от повышения ставок (снижение экономической активности) проявляются почти сразу же, тогда как положительных результатов, например сокращения инфляции, приходится ждать порой несколько лет. Вместе с ним Нобелевскую премию получил Томас Сарджент, который наблюдал за реакцией банков, компаний и индивидов при повышении и понижении инфляции. Базируясь на этих исследованиях, Томас Сарджент сформулировал теорию, согласно которой на действия людей влияют не шаги правительства как таковые, а их ожидание. В результате эффект от той или иной стратегии может оказаться не совсем таким, какого ожидали власти. На этих же принципах основана рефлексивная модель Джорджа Сороса: поведение людей, в том числе биржевых игроков, зависит от подаваемой им информации. Управляя потоками информации, можно управлять и “толпой” биржевых игроков, а значит и ценами на бирже.

8.4. Формирование портфеля ценных бумаг

Одна из интереснейших задач экономико-математического моделирования – формирование портфеля ценных бумаг на основе рядов цен на фондовом рынке. При этом часто от цен P переходят к доходностям d. ***Доходность ценных бумаг*** определяется как разность между стоимостью ценной бумаги в настоящий и начальный моменты, отнесенная к стоимости в начальный момент

$$d_i = \frac{P_i - Pнач}{Pнач}$$

Доходности позволяют сопоставлять бумаги, сильно отличающиеся по цене. Часто используют ***логдоходности***, вычисляемые не по ценам, а по их логарифмам. Это связано с негауссовским распределением отклонений цен от трендов и наличием в них “толстых хвостов”, то есть гетероскедастичности. Доходность портфеля определяется как сумма доходностей ценных бумаг, его составляющих, взвешенных на их доли в портфеле:

$$D = \Sigma\, d_i\, x_i$$

Требуется составить оптимальный портфель ценных бумаг по известным доходностям ценных бумаг за некоторый промежуток времени, имеющий максимальную доходность при заданном риске (портфель Марковица), или заданную доходность при минимальном риске. Мерой риска доходности одной бумаги является стандартное отклонение значений доходностей за некоторый промежуток времени. Если имеется тренд, то есть ряд не является стационарным, то его надо вычесть, а потом вычислять риск (СКО).

При отсутствии взаимной зависимости доходностей ценных бумаг (т.е. при нулевых коэффициентах корреляции) суммарная дисперсия равна сумме дисперсий $S^2 = \Sigma\, x_i^{\,2} * S_i^{\,2}$, где x_i – количество (или процент) закупаемых ценных бумаг i-ой фирмы. При коэффициентах корреляции, равных ±1 суммарное стандартное отклонение (риск портфеля) S равно сумме стандартных отклонений S_i с соответствующими знаками. При составлении портфеля из ценных бумаг двух фирм квадрат риска равен

$$S^2 = x_1^{\,2} * S_1^{\,2} + x_2^{\,2} * S_2^{\,2} + 2x_1 * x_2 * Cov(d1,d2),$$

где ковариация $Cov(d1,d2) = (\Sigma(d1i - d1cp) * (d2i - d2cp))/(N - 1)$

Если портфель составляется из ценных бумаг большего количества n фирм, то дисперсия портфеля (квадрат риска) вычисляется по формуле

$S^2 = \Sigma\Sigma\, x_i * x_j * Cov(di,dj)$. Обозначим $b_{ij} = Cov(di,dj)$, тогда

$$S^2 = x_1 * x_1 * b_{11} + x_1 * x_2 * b_{12} + \ldots + x_1 * x_n * b_{1n}$$

$$+ x_2 * x_1 * b_{21} + x_2 * x_2 * b_{22} + \ldots + x_2 * x_n * b_{2n}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$+ x_n*x_1*b_{n1} + x_n*x_2*b_{n2} + \ldots + x_n*x_n*b_{nn}$$

Далее приведен пример решения задачи составления портфеля с заданным доходом и минимальным риском. Заданы доходности четырех ценных бумаг за 16 периодов времени. Тренды в данном случае невелики и не вычитаются перед вычислением ковариационной матрицы.

Ковариационную матрицу можно вычислить с помощью *Анализ данных – Ковариация*. Программа выдаст только часть ковариационной матрицы, заполните ее целиком: матрица должна быть симметричной относительно диагонали. Начальные значения x_i заданы в столбце и продублированы с помощью формулы в строке x_j. Вычислите средние значения доходностей d_icp с помощью функции СРЗНАЧ и *Доход* = $\Sigma\ x_i*d_icp$. Вычислите матрицу $x_i*x_j*Cov(d_i\,d_j)$. Для этого перемножьте x_1 из столбца на x_1 из строки и на b_{11}, фиксируя знаком $ столбец в первом сомножителе x_1 и строку во втором сомножителе x_1, затем скопируйте формулу вправо и вниз. Просуммируйте полученную матрицу. Вызовите *Поиск решения*. *Целевая ячейка* – сумма по матрице $x_i*x_j*b_{ij}$, ее надо минимизировать, изменяемые ячейки – x_i в столбце, они ≥ 0, *Доход* ≥ заданной величины (здесь 300). Можно установить ограничение на *Сумму x*, т.е. на расходы. Изменяя заданный доход, постройте график зависимости риска от дохода. Можно действовать по-другому: максимизировать доход при заданном риске. Учтите, что при некоторых сочетаниях дохода и расхода решения не существует.

	d1	**d2**	**d3**	**d4**		Корреляционная матрица
1	1,02	3,64	5,90	5,76		
2	-1,06	0,67	4,37	4,39		
3	0,66	-2,12	-1,59	12,64		
4	2,49	4,24	4,56	5,17		
5	-0,80	-0,54	3,64	10,21		
6	1,92	6,51	8,39	2,58		

	d1	**d2**	**d3**	**d4**
d1	1,00			
d2	0,52	1,00		
d3	-0,08	0,79	1,00	
d4	0,11	-0,68	-0,86	1,00

7	1,29	4,94	6,06	3,91		Ковариационная матрица						
8	0,15	5,87	9,57	3,94			**d1**	**d2**	**d3**	**d4**	**x_i**	
9	1,13	1,93	4,20	8,68		**d1**	1,59	1,79	-0,34	0,43	**x1**	**0,00**
10	1,90	2,85	3,45	9,40		**d2**	1,79	7,40	6,81	-5,80	**x2**	**0,00**
11	-1,20	3,64	10,87	2,47		**d3**	-0,34	6,81	10,0	-8,44	**x3**	**25,4**
12	-1,88	-2,11	3,45	5,18		**d4**	0,43	-5,80	-8,44	9,68	**x4**	**26,3**
13	-0,83	2,42	7,48	4,80								
14	0,13	0,26	3,04	7,23		**x_j**	**=x1**	**=x2**	**=x3**	**=x4**	**Сумма x**	
15	0,74	4,74	8,37	4,17			**0**	**0**	**25,4**	**26,37**	**51,82**	
16	0,54	-0,11	1,80	10,84	Матрица $x_i*x_j*Cov(di,dj)$							
Сред-нее	0,39	2,30	5,22	6,34		0,00	0	0,00	0,00			
	x_1*d1	**x_2*d2**	**x_3*d3**	**x_4*d4**	***Доход***	0,00	0	0,00	0,00	**Сумма по матрице**		
	0	0	132,9	167,0	**300,0**	0,00	0	6515	-5667			
						0,00	0	-5667	6734	**1915**		
	Заданный доход				300	**Риск: корень из суммы по матрице**				**43,76071**		

Контрольные вопросы

1. Свойства временных рядов экономических переменных
2. Прогноз по временному ряду с сезонными колебаниями
3. Свойства рядов цен на фондовом рынке. Что такое портфель Марковица?
4. Автокорреляция случайного возмущения. Причины. Последствия
5. Стационарные и нестационарные стохастические процессы
6. Модели AR, MA и ARIMA, что такое векторная авторегрессия VAR

9. Системы эконометрических уравнений

9.1. Модель спроса и предложения

Многие экономические модели требуют для своего описания систем взаимосвязанных уравнений. Для настройки этих моделей обычно используют временные ряды уровней различных переменных, часть которых принимают за эндогенные, а часть за экзогенные. Выбор переменных определяется

исследователем, но обычно экзогенные переменные или не зависят от нас (температура воздуха, курс доллара, цена нефти), или мы можем ими управлять (инвестиции, выпуск продукции).

В качестве примера рассмотрим Модель спроса и предложения на конкурентном рынке, а также рыночной цены (***p***) в зависимости от величины дохода (***x***) на душу населения. Её называют “паутинной моделью”, так как движение спроса и предложения к равновесию в соответствующей системе координат напоминает паутину. Пример взят из учебника В.А.Бывшева [2].

Изменение во времени спроса, предложения и цены на конкурентном рынке закреплено в следующих утверждениях экономической теории:

1) Текущий уровень спроса объясняется текущей ценой товара и текущим располагаемым доходом на душу населения, причём спрос падает с ростом цены и растёт с ростом дохода.
2) Текущее предложение объясняется ценой товара в предшествующем периоде и возрастает с ростом этой цены.
3) Текущее значение рыночной цены устанавливается при балансе текущего спроса и текущего предложения товара.

Кратко это можно записать, с учётом случайных возмущений:

Спрос =a0 + a1· цена + a2· доход +Возмущение1

Предложение = b0 + b1· цена вчера + Возмущение2

Спрос = Предложение (тождество)

Соответствующая система уравнений и тождеств:

$$\begin{aligned} &d = a0 + a1\cdot p + a2\cdot x + u1 \\ &s = b0 + b1\cdot p(t-1) + u2 \\ &d = s \\ &a1<0,\quad a2>0,\quad b1>0 \end{aligned} \qquad (9.1)$$

В данном случае ***d, s, p*** эндогенные,

x, p(t-1) предопределённые (экзогенная и лаговая)

Второе уравнение является обычным уравнением регрессии, и его можно настраивать, используя обычный метод наименьших квадратов. А с первым уравнением так поступить нельзя, так как в него входят две эндогенных переменных. Такая модель называется ***структурной,*** она возникает непосредственно из экономических предпосылок. Требуется преобразовать модель к ***приведённому виду***, где в левой части будут стоять эндогенные переменные, а в правой – предопределённые. Можно решать эту задачу путём последовательной замены эндогенных переменных. Мы рассмотрим метод, основанный на преобразовании матриц. Объединим эндогенные переменные в вектор **Y**, а предопределённые – в вектор **X**:

$$\mathbf{Y}= (d,\ s,\ p)\ ; \qquad \mathbf{X} = (1,\ p(t-1),\ x)$$

Единица в векторе **X** появилась, чтобы работать с коэффициентами *a0* и *b0.* В матричном виде система уравнений и тождеств выглядит

$$\mathbf{AY} + \mathbf{BX} = \mathbf{0}$$

или

$$\begin{aligned}
&1\cdot d + 0\cdot s - a1\cdot p + (-a0)\cdot 1 + 0\cdot p(t-1) + (-a2)\cdot x = 0\\
&0\cdot d + 1\cdot s + 0\cdot p + (-b0)\cdot 1 + (-b1)\cdot p(t-1) + 0\cdot x = 0\\
&1\cdot d + (-1)\cdot s + 0\cdot p + 0\cdot 1 + 0\cdot p(t-1) + 0\cdot x = 0
\end{aligned}$$

Здесь матрицы **A** и **B**:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -a1 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} -a0 & 0 & -a2 \\ -b0 & -b1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Приведённая форма модели $\mathbf{Y} = \mathbf{M}\,\mathbf{X}$. Компоненты матрицы **M**

$$\mathbf{M} = \begin{pmatrix} \alpha 0 & \alpha 1 & \alpha 2 \\ \alpha 0 & \alpha 1 & \alpha 2 \\ \beta 0 & \beta 1 & \beta 2 \end{pmatrix}$$

получим преобразованием $\mathbf{M = A^{-1}B}$, где $\mathbf{A^{-1}}$ означает обратную к **A** матрицу. Приведённая форма модели:

$$d = \alpha 0 + \alpha 1 \cdot p(t-1) + \alpha 2 \cdot x$$

$$s = \alpha 0 + \alpha 1 \cdot p(t-1) + \alpha 2 \cdot x$$

$$p = \beta 0 + \beta 1 \cdot p(t-1) + \beta 2 \cdot x$$

где

$$\alpha 0 = b0 \; ; \qquad \alpha 1 = b1 \; ; \qquad \alpha 2 = 0 \; ;$$

$$\beta 0 = (b0-a0)/a1 \; ; \qquad \beta 1 = b1/a1; \qquad \beta 2 = -\, a2/a1$$

9.2. Идентифицируемость системы

Интерес представляют коэффициенты не приведённой модели, а структурной, которая имеет экономический смысл. Поэтому после настройки по статистическим данным приведённой модели и оценки её коэффициентов требуется вычислить по ним структурные коэффициенты. Но это не всегда получается: возникает проблема идентификации – единственности соответствия между приведённой и структурной формами модели. Структурные модели можно подразделить на три вида:

- идентифицируемые;
- неидентифицируемые;
- сверхидентифицируемые.

Модель ***идентифицируема***, если структурные коэффициенты определяются однозначно, единственным образом по коэффициентам приведённой формы модели, то есть число параметров структурной модели равно числу параметров приведённой формы модели.

Модель ***неидентифицируема***, если число приведённых коэффициентов меньше числа структурных коэффициентов, и структурные коэффициенты не могут быть оценены через коэффициенты приведённой формы модели.

Модель ***сверхидентифицируема,*** если число приведённых коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведённой формы можно получить несколько

значений каждого структурного коэффициента, число структурных коэффициентов меньше числа коэффициентов приведённой формы. Модель может быть практически решена при применении специальных методов.

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых необходимо проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

Чтобы уравнение было идентифицируемо, нужно, чтобы число предопределённых переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного. Условие идентифицируемости модели может быть записано в виде следующего правила:

Предопределённых + 1 = Эндогенных **идентифицируемо**

Предопределённых + 1 < Эндогенных **неидентифицируемо**

Предопределённых + 1 > Эндогенных **сверхидентифицируемо**

Если обозначить число эндогенных переменных в ***j***-м уравнении системы через ***H***, а число предопределённых переменных, которые содержатся в системе, но не входят в данное уравнение, через ***D***, то

D + 1 = H **идентифицируемо**

D + 1 < H **неидентифицируемо**

D + 1 > H **сверхидентифицируемо**

В исследуемой модели ***d, p, s*** эндогенные; ***x, p(t-1)*** предопределённые, во всей системе их 2. Исследуем модель:

	Предопр.	***D***	***D+1***	***H***	
$d = a0 + a1 \cdot p + a2 \cdot x$	1	1	2	2	идент.

s = b0 + b1 · p(t-1)	1	1	2	1	сверх.
d = s	0	2	2	2	идент.

Значит, модель сверхидентифицируема.

9.3. Методы решения систем эконометрических уравнений

Идентифицируемую систему эконометрических уравнений можно решить ***Косвенным методом наименьших квадратов (КМНК).*** Суть метода в следующем:

- Преобразование структурной формы модели в приведённую;
- Оценка коэффициентов уравнений приведённой формы обычным методом наименьших квадратов;
- Преобразовать их в коэффициенты структурной модели. Если при вычислении коэффициентов структурной формы (9.1) учитывать возмущения, то они войдут в оба уравнения модели, и принцип независимости эндогенных переменных и остатков будет нарушен. Это приведёт к смещению (отсутствию состоятельности) параметров модели. Для подавления этого эффекта, а также для настройки сверхидентифицируемых моделей используется ***двухшаговый метод наименьших квадратов ДМНК***.

Основная идея ДМНК – на основе приведённой формы модели получить для сверхидентифицируемого уравнения оценённые значения эндогенных переменных, содержащихся в правой части уравнения. Далее, подставив их в правые части уравнений вместо фактических значений, можно применить обычный МНК к структурной форме сверхидентифицируемого уравнения. Метод получил название “двухшаговый метод наименьших квадратов”, ибо МНК используется дважды: на первом шаге при определении коэффициентов приведённой формы модели и нахождении на её основе оценок оценённых значений эндогенных переменных $\hat{Y}$ и на втором шаге применительно к структурному сверхидентифицируемому уравнению при определении структурных коэффициентов модели с использованием оценённых значений

эндогенных переменных. Оценённые значения играют роль так называемых ***инструментальных переменных (instrumental variables, IV, instruments)*** – переменных, которые применяются, если обычные переменные коррелируют с возмущениями. Инструментальные переменные коррелируют с обычными переменными, но не коррелируют с возмущениями, что приводит к состоятельности (consistency) модели. Расчёты с использованием инструментальных переменных включены в статистические пакеты, так что не удивляйтесь, увидев ***IV*** на распечатке.

Алгоритмы и краткие замечания по КМНК и ДМНК:

Косвенный МНК:

1) Структурная => Приведенная

2) Коэффициенты по МНК

3) Преобразовать их в коэффициенты структурной модели

Нарушение предпосылки независимости факторов приводит к несостоятельности оценок структурных коэффициентов, они могут оказаться бессмысленными

Двухшаговый МНК

Применяется для сверхидентифицируемых систем уравнений

1) Структурная => Приведенная.

2) Коэффициенты по МНК.

3) Получить оценённые значения эндогенных переменных.

4) Подставить их в правые части структурной формы.

5) Применить МНК к структурной форме сверхидентифицируемых уравнений.

Сверхидентифицируемую модель можно превратить в идентифицируемую путем добавления некоторых переменных или отбрасывания некоторых ограничений на параметры

9.4. Настройка макроэкономических моделей с использованием итерационных градиентных методов

Данный раздел выполнен на основе статьи, опубликованной совместно с Е.А.Филиппович [8].

Модели макроэкономических процессов представляют собой, как правило, системы одновременных уравнений, переменными в которых являются валовой внутренний продукт *Y*, потребление *C*, инвестиции *I*, государственные расходы *G* и другие показатели, зависящие от времени и связанные между собой. Существует множество таких моделей, но основная проблема – их настройка, т.е. оценка коэффициентов в уравнениях. Для этого используются различные методы, наиболее известные – косвенный метод наименьших квадратов и двухшаговый (или трехшаговый) метод наименьших квадратов, которые позволяют свести задачу к подгонке коэффициентов регрессионных уравнений.

Мы опробуем настройку макроэкономических моделей с использованием итерационных градиентных методов – Ньютона, сопряженных градиентов, в версиях 2007 и выше – ОПГ (что хуже), заложенных в сервис *"Поиск решения" (Solver)* электронных таблиц Excel. Технология такова: исследователь строит в Excel структурную модель и задает произвольный набор её коэффициентов, а компьютер их варьирует, минимизируя сумму квадратов отклонений реальных и оцененных значений эндогенных переменных. Далее представлен пример таблицы для проведения расчетов. Здесь ***Y*** – валовый внутренний продукт (ВВП), ***C*** – расходы на потребление, *I* – чистые инвестиции, ***G*** – государственные расходы, ***Y^, C^, I^, G^*** – соответствующие расчетные значения, ΔC^2 , ΔI^2 , ΔG^2 – квадраты разностей истинных и расчетных значений, $\Sigma\Delta C^2$, $\Sigma\Delta I^2$, $\Sigma\Delta G^2$ - их суммы. Использованы данные по экономике США с 1946 по 2007 год, причем по данным 1946-2002 г.г. проводилась настройка модели, а по 2003-07 г.г. – проверка адекватности модели. Данные, представленные в Приложении 3, взяты из [9, 10]

Таблица 9.1.

Пример проведения расчетов в Excel

Год	*Y*	*C*	***I***	*G*	*Y^*	*C^*	*I^*	*G^*	ΔC^2	ΔI^2	ΔG^2
1946	211	147	29	31	211	147	29	31	0	0	0
1947	233	166	30	29	233	166	30	29	0	0	0
1948	259	178	43	37	135	69	32	34	11893	117	6,3
1949	258	181	34	44	164	88	37	40	8631	11	13,7
….											
2002	10469	7385	1582	1961	10963	7325	1647	1990	59,51	-65,77	-29,08
….											
2007	13807	9710	2130	2674	14456	9563	2241	2651			
									$\Sigma\Delta C^2$	$\Sigma\Delta I^2$	$\Sigma\Delta G^2$

Попробуем настроить, исследовать и модифицировать модель Самуэльсона–Хикса и одну из версий модели Кейнса.

Экономическим объектом служит закрытая экономика. Ее состояние в текущем периоде ***t*** описывается переменными ***(Y_t, C_t, I_t, G_t).*** Концептуальная модель Самуэльсона-Хикса:

1) Текущее потребление объясняется уровнем ВВП в предыдущем периоде, возрастая вместе с ним, но с меньшей скоростью;

2) Величина инвестиций прямо пропорциональна приросту ВВП за предшествующий период (прирост ВВП за предшествующий период – это разность $\boldsymbol{Y_{t-1} - Y_{t-2}}$);

3) Государственные расходы возрастают с постоянным темпом роста;

4) текущее значение ВВП есть сумма текущих уровней потребления, инвестиций и государственных расходов (тождество системы национальных счетов).

В сокращённом виде:

Потребление $= a_0 + a_1 *$ *ВВП(t-1)*

*Инвестиции = b * (ВВП(t-1) – ВВП(t-2)) = b * (рост ВВП в прошлом году)*

*Госрасходы = g * Госрасходы (t-1)*

ВВП = Потребление + Инвестиции + Госрасходы

Модели Самуэльсона-Хикса (слева) и Кейнса (справа) похожи, но значения C и I вычисляются с использованием разных Y.

$$C = a_0 + a_1 Y_{t-1} \qquad C = a_0 + a_1 Y_t + a_2 Y_{t-1}$$

$$I = b_0 + b_1 (Y_{t-1} - Y_{t-2}) \qquad I = b_0 + b_1 Y_t + b_2 Y_{t-1}$$

$$G = g G_{t-1}$$

$$Y = C + I + G \qquad Y = C + I + G$$

$$0 < a1 < 1, \; b > 0, \; g > 0$$

Особую сложность представляет построение уравнения для инвестиций I, так как их значения не аппроксимируются гладкой функцией и их колебания коррелируют с ($\boldsymbol{Y_{t-1} - Y_{t-2}}$), что видно на Рисунке 9.1, на котором также представлены оцененные значения I^ :

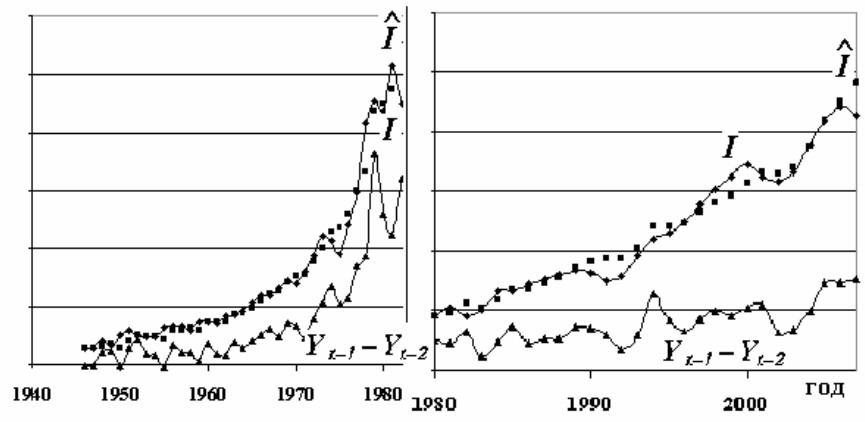


Рис. 9.1. Зависимость $\boldsymbol{I}$ и $\boldsymbol{I}$^ от $\boldsymbol{Y_{t-1} - Y_{t-2}}$.

Модель Самуэльсона–Хикса в исходном виде подогнать не удалось, т.к. прямая пропорциональность $\boldsymbol{I}$ разности $\boldsymbol{Y}$ привела к большим колебаниям $\boldsymbol{I}$^. Инвестиции вычислялись по формуле, аналогичной модели Кейнса, но со сдвигом Y назад на 1 год:

$$I = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2}$$

Государственные расходы G также не удалось аппроксимировать по исходной формуле, в уравнение была добавлена константа g_0. Расчет коэффициентов для оценки G проводился отдельно. Использован Сервис *"Поиск решения"* Excel, целевая ячейка $\Sigma \Delta G^2$, т.е. сумма квадратов отклонений истинных и расчетных значений за период 1946-2002 г.г., изменяемые значения – коэффициенты g_0 и

g. При расчете коэффициентов для C и I целевая ячейка $(\Sigma\Delta C^2 + \Sigma\Delta I^2)$. В результате получена модифицированная модель Самуэльсона–Хикса:

	R^2 прогноза
$C = -28,9 + 0,721Y_{t-1}$	0,86
$I = -10,34 + 0,424Y_{t-1} - 0,268Y_{t-2}$	0,71
$G = 3,85 + 1,057G_{t-1}$	0,995
$Y = C + I + G$	0,84

и модель Кейнса:

$C = -35,67 + 0,157Y_t + 0,556Y_{t-1}$	0,67
$I = -16,9 + 0,3535Y_t - 0,202 \quad Y_{t-1}$	0,74
$Y = C + I + G$	0,74

Точность и адекватность моделей оценена по индексам детерминации

$$R^2 = 1 - \Sigma ост^2 / TSS ,$$

где $TSS = \Sigma (X - Xсредн.)^2$, $X = Y, C, I, G.$ В диапазоне настройки $R^2 > 0,99$; в диапазоне прогноза (2003-07 г.г.) индексы детерминации для моделей Самуэльсона–Хикса и Кейнса представлены выше рядом с формулами. Точность прогноза нельзя назвать хорошей, и завышенные прогнозы Y и C видны на Рисунке 9.2. Возможно, это связано со спадом темпа роста инвестиций в 2002-03 г.г., что видно на Рисунке 1 и с уменьшением темпов роста всех показателей с начала 80-х.

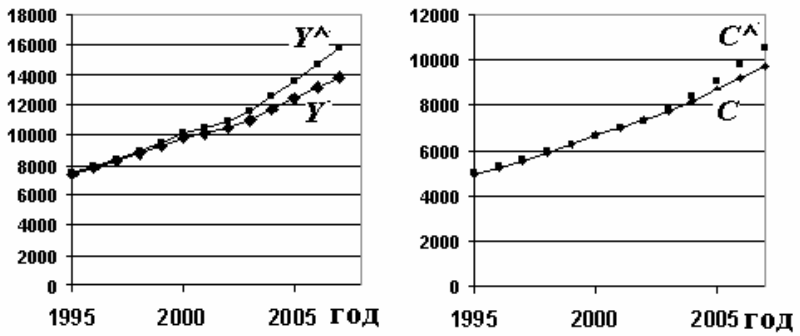


Рис.9.2. Реальные и прогнозные значения ВВП (Y) и потребления (C) по модели Самуэльсона–Хикса.

Мы опробовали также настройку моделей, существенно отличающихся от классических, в которых функции Y, C, I, G близки к экспонентам, за

исключением пилообразных отклонений *I*. Мы провели линеаризацию этих функций логарифмированием, а затем строили модели. Графики натуральных логарифмов *Y, C, G, I* представлены на Рисунке 9.3.

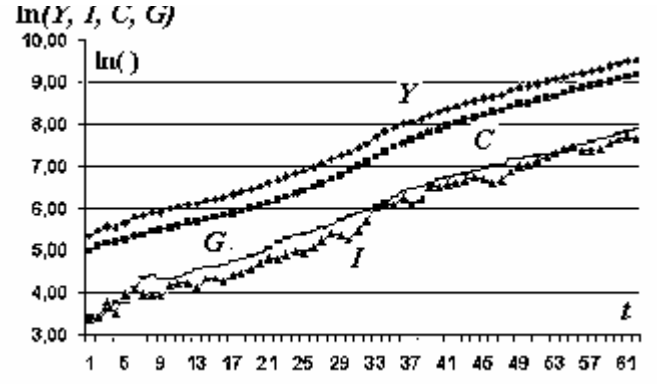


Рис. 9.3. Графики натуральных логарифмов *Y, C, G, I*

Были опробованы несколько моделей. Наиболее точные прогнозы на 2003-07 г.г. получены по следующей модели:

	R^2 прогноза
$\ln C\hat{} = -0{,}429 + 1{,}0096 \ln Y_t$	0,987
$\ln I\hat{} = -1{,}60 + \ln (2{,}167\, Y_t - 1{,}502\, Y_{t-1})$	0,945
$\ln G\hat{} = -1{,}13 + 0{,}946\, Y_{t-1}$	0,925
$\ln Y\hat{} = \ln (C\hat{} + I\hat{} + G\hat{})$	0,986

Графики реальных и прогнозных величин в 2003-07 г.г. практически совпали.

Как видим, регрессия по ВВП и его приросту дает хорошие результаты. А как спрогнозировать ВВП? Линейная регрессия ВВП по времени дала волнообразный график остатков, представленный на Рисунке 9.4А. Его период совпал с периодом циклов Кондратьева – примерно 40 лет. Поэтому была проведена аппроксимация зависимости ВВП от времени функцией

$$Y(t) = a + bt + d\,\mathrm{Sin}(\omega t + \varphi).$$

Настройка этой модели проведена с помощью сервиса “Поиск решения” Excel, изменяемые ячейки – пять коэффициентов *a, b, d, ω, φ*. Остатки стали случайными (Рис. 9.4В), кроме последнего участка, где реальный ВВП на самом деле несколько упал.

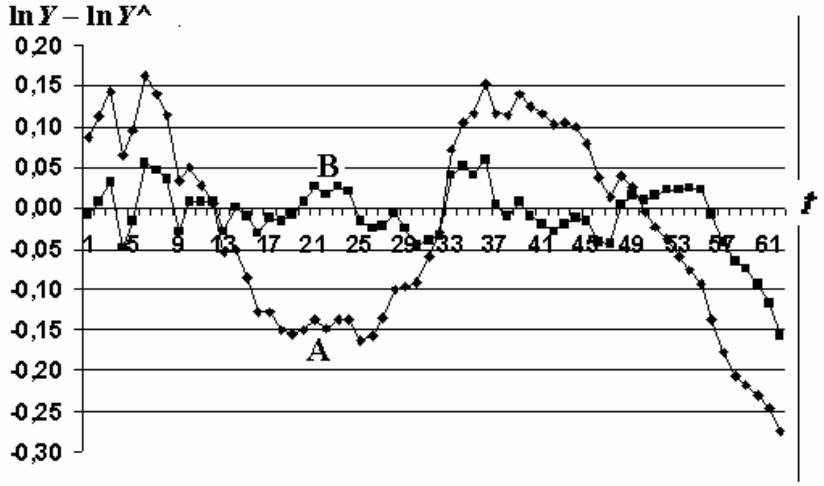


Рис. 9.4. Разности логарифмов реального и оцененного ВВП при линейной (А) и линейно-синусоидальной (В) регрессии.

ВЫВОД: В целом, применение итерационных градиентных методов, заложенного в сервис *“Поиск решения”* Excel, позволяет достаточно просто, быстро и наглядно настраивать и исследовать системы одновременных уравнений, описывающие экономические процессы. Точность аппроксимации в диапазоне настройки получается весьма высокой, даже при колебаниях переменных. Остается открытым вопрос об устойчивости решений и о надежности прогноза, особенно при колебаниях переменных в конце диапазона настройки.

Исследуйте модель, увеличив обучающую выборку, а также используя дополнительные статистические данные по экономикам разных стабильных стран. Исследуйте аналогичным образом другие макроэкономические модели. В принципе, таким образом можно сделать дипломную работу или даже кандидатскую диссертацию, но следует быть очень осторожным: поисковые работы в этих случаях опасны, можно ожидать критики в связи с недостаточным теоретическим обоснованием применяемых технологий.

Контрольные вопросы

1. Системы эконометрических уравнений: Модель спроса и предложения.
2. Компактная (матричная) запись структурной и приведённой формы динамической модели из одновременных линейных уравнений

3. Что такое идентифицируемость системы одновременных уравнений.

4. Условие идентифицируемости системы одновременных уравнений.

5. Методы решения систем эконометрических уравнений.

6. Эконометрическая инвестиционная модель Самуэльсона-Хикса, её настройка.

Литература

1. C. Dougherty. Introduction to Econometrics. Oxford, University Press, 2007.

2. В.А. Бывшев. Эконометрика. М.: Финансы и статистика, 2008.

3. Л.О. Бабешко. Основы эконометрического моделирования. – М.: Комкнига, 2001.

4. Эконометрика. Под редакцией И.И.Елисеевой. – М.: Финансы и статистика, 2005.

5. Практикум по эконометрике. Под редакцией И.И.Елисеевой. – М.: Финансы и статистика, 2005.

6. Н.В.Катаргин, А.В.Цветков. Исследование автокорреляций высоких порядков в рядах цен активов на фондовом рынке. Международный научный журнал № 4, 2011, стр. 38-42

7. Э.Петерс. Порядок и хаос на рынках капитала. М. – Мир, 2000

8. Н.В.Катаргин, Е.А.Филиппович. Настройка макроэкономических моделей с использованием метода Ньютона. Международный научный журнал № 1, 2010, стр.15-18.

9. Economic Indicators OCTOBER 2008 (Includes data available as of November 7, 2008), Economic Indicators DECEMBER 2003 (Includes data available as of December 31, 2003), http://www.gpoaccess.gov/indicators/

10. Economic Indicators DECEMBER 1995, Economic Indicators DECEMBER 1986, Economic Indicators DECEMBER 1978, Economic Indicators DECEMBER 1968, Economic Indicators DECEMBER 1959, Economic Indicators DECEMBER 1953. http://fraser.stlouisfed.org/publications/ei/

Приложение 1. Использование метода Монте-Карло для исследования ошибок регрессионной модели.

Создать на рабочем листе кнопку с программным модулем на языке Visual Basic for Applications (VBA) для сохранения получаемых параметров модели и создания псевдореального массива *Yимит*, значения которого будут распределены вокруг "идеальных" значений *Y* по закону нормального распределения (Гаусса) со стандартным отклонением *3СигмаОст.идеал*. В программном модуле используется генератор случайных чисел RND(), который выдает случайные числа, равномерно распределенные в диапазоне 0 – 1. Программа пересчитывает их в числа, распределенные по нормальному (Гауссову) закону в диапазоне от –3 до +3, используя заранее внесённую в одну из ячеек Excel функцию НОРМСТОБР. В данном примере функция внесена в ячейку N3, её аргумент формируется в предыдущей ячейке M3 функцией Бейсика Rnd(). Процедура пересчета представлена на рисунке. Случайное значение *Y* складывается из *Yидеал* и случайной величины, распределенной по нормальному закону в диапазоне от *–3СигмаОст.идеал* до *+3СигмаОст.идеал*. Технология создания на рабочем листе кнопки с программным модулем в среде Excel 2003:

- в Меню щелкните *Вид – Панели инструментов* – Visual Basic 
- при нажатой левой клавиши мыши растяните на рабочем листе контур кнопки;
- после появления кнопки дважды быстро щелкните по ней мышью, и вы войдете в окно программного модуля;
- впишите в программный модуль текст программы, контролируя соответствие адресов переменных на рабочем листе и параметров объектов типа "Массив-диапазон ячеек" (Range);
- перейдите из режима конструктора в рабочий режим, щелкнув по 

В Excel 2007 и более поздних версиях надо войти в Файл – Параметры, включить в Настройку ленты Разработчик, так же создать кнопку и сохранить рабочую книгу как файл Excel с поддержкой модулей VBA.

Для расчёта коэффициентов, R^2 и F следует применять функцию ЛИНЕЙН, которая автоматически срабатывает при каждой новой имитации. Все вычисленные параметры модели следует разместить в одну строку, используя копирование по формуле: так проще запрограммировать их сохранение. Целесообразно создать две кнопки: вторая для многократного вызова программного модуля главной кнопки.

Метод Монте-Карло можно применять только с использованием программных модулей, но его принципы можно усвоить в Excel в ручном режиме. Для этого надо заменить нормальное распределение возмущений в диапазоне от *–3СигмаОст.идеал* до *+3СигмаОст.идеал* на равномерное распределение в диапазоне от *–1,5СигмаОст.идеал* до *+1,5СигмаОст.идеал*, как показано на рисунке П1.2. и реализуется по формуле

$$Yимит = Yидеал + (СЛЧИС() - 0{,}5)*3*СигмаОст.идеал$$

Сохранение вычисленных параметров модели в этом случае реализуется вручную, с использованием *Копирования* и *Специальной вставки – Значения.*

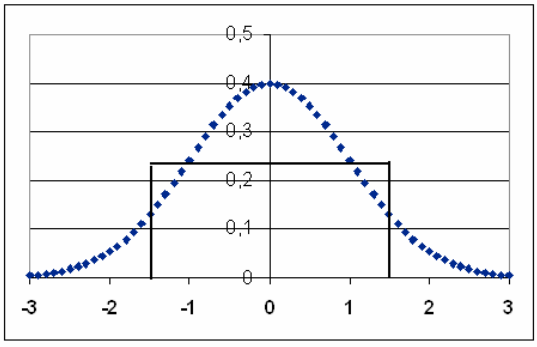


Рис.П1.1. Замена нормального распределения на равномерное.

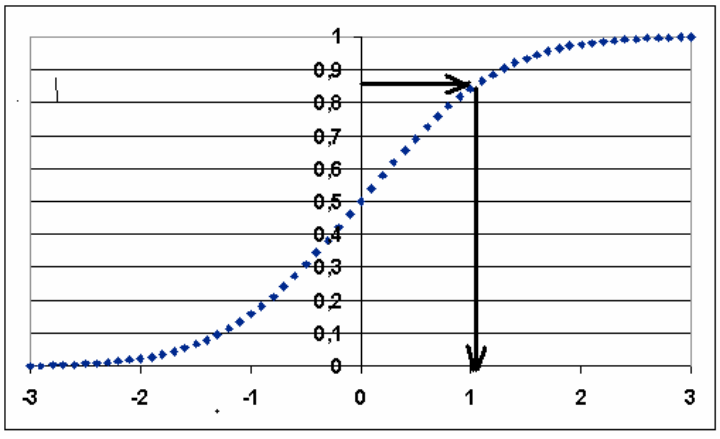


Рис.П1.2.Преобразование равномерного случайного распределения в нормальное

4. Программный модуль:

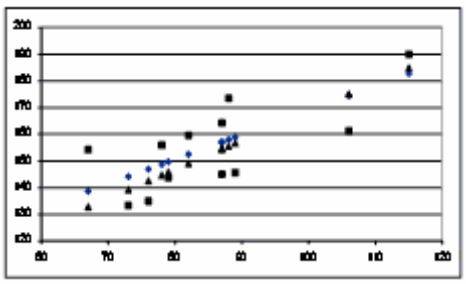
```
Private Sub CommandButton1_Click()
Dim A, B,  Gauss, Y   As Range  ‘Создание объектов типа массив-диапазон
                                ‘ Настройка массивов-диапазонов (Range) на
                                 соответствующие ячейки рабочего листа:
Set A = Range("H4")             ‘   a, b, Yпрогноз, GQ, DW и др.
Set B = Range("H20")            ‘Подготовка сохранения результатов
                                 ‘начиная с ячейки H20
Set Gauss  = Range("M3")   ‘ в ячейку N3 внести функцию НОРМСТОБР(“M3”)
Set Y = Range("C4")          ‘Yideal с ячейки C4, Yимит с D4

Randomize

For i = 1 To 12                 ‘  Создание массива из 12 чисел Yидеал+RND
 Gauss(1)= Rnd()
M=Gauss(1,2)
Sigma = Range("G19")                    ‘σ остатков в "G19"
Y(i,2) = Y(i) + M * Sigma               ‘ Yимит
Next i

N=Range(“N1”)     ‘ счётчик строк массива сохраняемых данных
For i=1 to 5
B(N,i) = A(N,1)      ‘ сохранение a, b, Yпрогноз, DW, GQ
Next i
Range(“N1”)  =   Range(“N1”)  + 1

End Sub
```



5. Выделите столбцы *X, Yideal,* Yrandom*, Y^*. Постройте диаграмму типа *Точечная*. На рисунке ромбики *Yideal,* квадраты *Yимит*, треугольники *Y^*.

6. Последовательно нажимайте кнопку и следите за изменениями на диаграмме и за накоплением результатов.

Приложение 2. Исходные данные для оценки стоимости квартир.

Общая площадь (S)	Площадь кухни (SK)	Коли-чество комнат- (R)	Расстоя-ние от метро (M)	Кол-во этажей в доме (E)	Зона (Z)	Тип дома (H)	Этаж (F)	Состоя-ние квартиры (C)	Цена (Y) тыс. $
50	15	1	15	9	1	1	1	3	525
50	15	1	15	9	1	2	1	3	525
50	15	1	15	9	1	3	1	3	557
50	15	1	15	9	1	4	1	3	557
50	15	1	15	9	1	5	1	3	594
50	15	1	15	15	1	1	1	3	525
50	15	1	15	3	1	1	0	3	525
50	15	1	15	3	1	1	1	3	525
50	15	1	15	3	1	1	1	2	581
50	15	1	15	3	1	4	1	2	609
70	16	2	15	17	1	3	0	1	828
70	16	2	15	17	1	3	1	1	828
70	16	2	15	17	1	3	0	1	789
70	16	2	15	17	1	3	1	1	873
70	16	2	15	17	1	4	1	1	828
70	16	2	15	17	1	5	1	1	873
70	16	2	15	17	1	5	1	2	838
70	16	2	10	17	1	5	1	2	855
70	16	2	10	17	1	5	1	3	783
70	9	2	10	17	1	5	1	3	740
70	20	2	10	17	1	5	1	3	783
70	20	2	10	17	1	4	1	3	729
70	6	2	10	17	1	4	1	3	619
70	6	2	10	17	1	1	1	3	574
70	6	2	10	17	1	1	1	2	656
70	6	2	10	17	1	1	1	1	701
70	6	2	20	17	1	1	1	1	665
85	15	3	15	17	1	1	1	3	753
85	15	3	15	17	1	1	1	1	926
85	15	3	15	17	1	5	1	1	1 036
85	15	3	15	17	1	4	1	1	978
85	15	3	15	17	1	4	1	2	928
100	15	3	15	17	1	4	1	2	1 092
100	15	3	15	17	1	4	0	2	1 029
100	15	3	3	17	1	4	0	2	1 106
100	15	3	3	17	1	4	0	2	1 164
100	15	3	5	17	1	4	0	2	1 164
100	15	3	8	17	1	4	0	2	1 120
100	15	3	10	17	1	4	0	2	1 120
120	15	3	10	17	1	4	0	2	1 344
120	15	3	10	17	1	5	0	2	1 430
120	15	3	10	17	1	5	0	1	1 490
120	15	3	10	9	1	5	0	1	1 491

Приложение 3. Исходные данные для настройки макроэкономических моделей.

Год	*Y*	*C*	*I*	*G*	*Год*	*Y*	*C*	*I*	*G*
1946	211,10	146,90	28,70	30,90	1977	1887,20	1206,50	297,80	394,00
1947	233,30	165,60	30,20	28,60	1978	2249,70	1403,50	416,80	425,20
1948	259,00	177,90	42,70	36,60	1979	2508,20	1566,80	454,80	467,80
1949	258,20	180,60	33,50	43,60	1980	2732,00	1732,60	437,00	530,30
1950	286,80	194,60	52,50	42,00	1981	3052,60	1915,10	515,50	588,10
1951	329,80	208,10	58,60	62,90	1982	3166,00	2050,70	447,30	641,70
1952	348,00	218,10	52,50	77,50	1983	3405,70	2234,50	502,30	675,00
1953	365,40	232,60	50,30	82,80	1984	3765,00	2428,20	662,10	733,40
1954	363,10	238,00	48,90	75,30	1985	3998,10	2600,50	662,10	815,40
1955	397,50	256,90	63,80	75,60	1986	4268,60	2850,60	717,60	833,00
1956	419,20	269,90	67,40	79,00	1987	4539,90	3052,20	749,30	881,50
1957	441,10	281,40	67,80	86,10	1988	4900,40	3296,10	793,60	918,70
1958	447,30	290,10	60,90	94,20	1989	5250,80	3523,10	832,30	975,20
1959	483,70	311,20	75,30	97,00	1990	5546,10	3761,20	808,90	1047,40
1960	503,70	325,20	74,80	99,60	1991	5724,80	3902,40	744,80	1097,40
1961	520,10	335,20	71,70	107,60	1992	6020,20	4136,90	788,30	1125,30
1962	560,30	355,10	83,00	117,10	1993	6657,40	4477,90	953,40	1291,20
1963	590,50	375,00	87,10	122,50	1994	7072,20	4743,30	1097,10	1325,50
1964	632,40	401,20	94,00	128,70	1995	7397,70	4975,80	1144,00	1369,20
1965	684,90	432,80	108,10	137,00	1996	7816,90	5256,80	1240,30	1416,00
1966	747,60	465,50	120,80	156,20	1997	8304,30	5547,40	1389,80	1468,70
1967	796,30	490,40	120,80	180,20	1998	8747,00	5879,50	1509,10	1518,30
1968	868,50	535,90	131,50	198,70	1999	9268,40	6282,50	1625,70	1620,80
1969	935,50	579,70	146,20	207,90	2000	9817,00	6739,40	1735,50	1721,60
1970	982,40	618,80	140,80	218,90	2001	10128,00	7045,40	1614,30	1825,60
1971	1063,40	668,20	160,00	233,70	2002	10469,60	7385,30	1582,10	1961,10
1972	1171,10	733,00	188,30	253,10	2003	10960,80	7703,60	1664,10	2092,50
1973	1306,60	809,90	220,00	269,50	2004	11685,90	8195,90	1888,60	2216,80
1974	1412,90	889,60	214,60	302,70	2005	12421,90	8694,10	2086,10	2355,30
1975	1528,80	979,10	190,90	338,40	2006	13178,40	9207,20	2220,40	2508,10
					2007	13807,50	9710,20	2130,40	2674,80