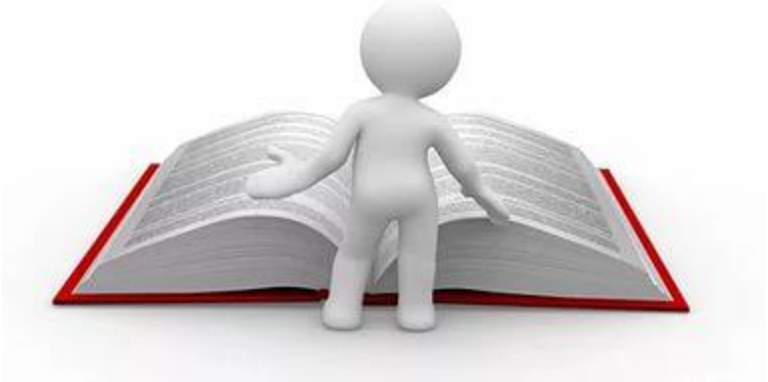
Моделирование социально-экономических процессов в экономике



Семёнычев Валерий Константинович
Д.т.н., Д.э.н., профессор по кафедре «Высшая математика»

– практика, лабораторные работы

– лекции

То, что видим мы – видимость только

одна.

Далеко от поверхности моря до дна.

Полагай несущественным явное в мире,

Ибо тайная сущность вещей – не видна.

Омар Хайям

Рекомендуемая литература

1. Аникин П.В., Королев В.А., Тороповцев Е.А Математические и инструментальные методы. Изд-во «Кнорус». 2014. Можно скачать.

2. Интеллектуальные информационные системы: учебное пособие / А.А. Смагин, С.В. Липатова, А.С. Мильченко. – Ульяновск: УлГУ, 2010. – 136 с. (можно скачать).

3. Семенычев В.К., Семенычев Е.В. Параметрическая идентификация рядов динамики: структуры, модели, эволюция. - Самара. Изд-во «СамНЦ РАН», 2011. – 346 с.

4. Семенычев В.К., Коробецкая А.А., Кожухова В.Н. Предложения эконометрического инструментария моделирования и прогнозирования эволюционных процессов. - Самара. САГМУ. – 384 с.

5. Конюховский П.В. Микроэкономическое моделирование в банковской деятельности. - Спб. Питер.-2001. - 224 с.

6. Эконометрика / Под. Ред. И.И. Елисеевой. – М.: Финансы и статистика, 2005. – 575 с. (и более поздние издания).

8. Бородич С.А. Эконометрика. - Минск: Новое знание. 2001. - 408 с.

9. Кондратьевские волны. Под редакцией Л.Е. Гринина, А.В. Коротаева. – Волгоград: Учитель. 2014. – 360 с.

10. Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. Айвазян А.А., Мхитарян В.С.-- М.; ЮНИТИ-ДАНА. 2001.- 656 с.
11. Дайитбегов Д.М. Компьютерные технологии анализа данных в Эконометрике. – М. ИНФРА-М. 2008.- 578 с.
12. Кожухова В.Н., Коробецкая А.А., Семенычев В.К. Свободная программная среда R. Практикум. Самара. Изд-во «САГМУ». 2016.-48 с.
13. Статистический анализ структуры социально-экономических процессов и явлений (Сивелькин В.А., Кузнецова В.Е.) (можно скачать).
14. Статистические методы и модели (Костин В.Н., Тишина Н.А.) (можно скачать).
15. Лукьянов Б.В., Лукьянов П.Б. Математические и инструментальные методы поддержки принятия решений. Кнорус. 2016.
16. Э. Колин Камерон, Правин К. Триведи. Микроэконометрика. Методы и их применение. Кн.1. 2015. Изд. дом. «Дело» РАНХ и ГС.
17. Э. Колин Камерон, Правин К. Триведи. Микроэконометрика. Методы и их применение. Кн.2. 2015. Изд. дом. «Дело» РАНХ и ГС. 1160 стр.
18. Шитиков В.К. Розенберг Г.С. Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R. Тольятти. 2013. Возможно и получение дополнительной интернет-версия от 15.11.2913.

19. Мастицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. ДМК Пресс. 2015. 496 стр.
20. Кабаков Р. R в действии. Анализ и визуализация данных на языке R. ДМК Пресс. 2013. 580 стр.

21. Д. Рутковская, М. Пилиньский, Л. Рутковский. Нейронная сеть, генетические алгоритмы. Горячая линия-Телеком. 2013.
22. Ширяев В.И. Финансовые рынки. Нейронные сети, хаос и нелинейная динамика. Либрокон. 2015.
23. Бородич С.А. Эконометрика. Практикум. Изд-Инфра-М. 2016. 336.

Приглашаю и на свой персональный сайт, набрав в поисковике «Семенычев В.К.»: монографии, методические пособия, статьи – мои и учеников.

Главный враг Знания

– не невежество, а иллюзия знаний.

Необходимость моделирования

- Каждое лицо принимающее на практике какие-либо решения (ЛПР) руководствуется правилами (моральными, юридическими, санитарными и т.п.), а также имеющимся у него опытом и сложившимися стереотипами: по сути индивидуальной моделью – ему понятной, как правило, более простой, характеризующей окружающий мир.
- Различают эндогенные, экзогенные факторы (характеристики). Примеры объектов анализа: гараж – вектор разной размерности, содержащий случайность – погрешности измерений и др., горячий чай – (нечеткая логика при оценке температуры).
- При взаимодействии нескольких ЛПР необходимо обмениваться моделями для однозначного определения явления (экономического объекта, системы, процесса, ситуации). Имманентны (всегда присутствуют) ошибки при принятии решений: 1) неточность информации; 2) неадекватная оценка полученной информации (соотношение цели моделирования, точности модели и адекватности); 3) неточная идентификация модели (оптимизационная задачи на max меры точности); 4) неверная оценка последствий принимаемых решений, присутствие нестационарности явления (его эволюции). Различают системы (объекты) – слабо структурированные (вероятностные), неструктурированные (хаотические) в отличие от многих курсов, где объекты - структурированные (детерминированные). Объект нашего курса - слабо структурированные системы, для которых применяют СППР (системы поддержки принятия решений), а из моделей формируют знания, на основании которых ЛПР должен сам принять решение.

Неопределенность

описывается

теорией вероятностей

и/или

теорией нечетких множеств

(fuzzy sets)

и

нечеткой логикой (fuzzy logic)

Нечеткая логика - раздел современной математики, позволяющий формализовать и перевести на компьютерный язык интуитивные знания и умения специалистов-практиков.

Например, напомню о уже приведенном выше примере с «горячем чае» - по разному люди оценивают насколько «горяч чай». Еще пример: если «давление высокое» и «температура низкая», а также «оборудование старое», то нужно «немного убавить обороты»: формализуются для данной ситуации понятия «высокое», «старое» и «немного». Затем эксперт-практик может на языке, близком к человеческому, задать правила, которыми он обычно руководствуется в своей деятельности.

Правила могут быть нестрогими, нечеткими, противоречащими друг-другу – почти как в жизни.

Классическая логика не позволяет этого, а нечеткая - вполне, потому что условие правила может быть не только «истинным» или ложным, но и истинным, например, «на половину» или «на треть» и т.п.

Понятия нечеткой логики

(нечеткие множества и высказывания) появились в середине 1960-годов в публикациях американского математика Лотфи А. Заде. К 1990-м годам нечеткая логика из математической игрушки превратилась в необычайно популярный прикладной метод.

Нечеткая логика начала применяться в фото-и видеокамерах (Sony, Canon, Minolta), стиральных машинах (Siemens, Samsung, Candy), автомобильных навигаторах (Opel, Porsche), автоматических коробках передач в автомобилях (Porsche, Renault, Peugeot, Hyundai, Skoda), аппаратах измерения кровяного давления (Omron), при анализе новых рынков, биржевой игре, оценке политических рейтингов, оптимальной ценовой стратегии, СППР и т.д.

Первый период характеризуется развитием теоретического аппарата нечетких множеств (Л. Заде, Э. Мамдани, Беллман).

Во втором периоде (70–80-е годы) появляются первые практические результаты в области нечеткого управления сложными техническими системами (парогенератор с нечетким управлением). Одновременно стало уделяться внимание вопросам построения экспертных систем, построенных на нечеткой логике, разработке нечетких контроллеров. Нечеткие экспертные системы для поддержки принятия решений находят широкое применение в медицине и экономике.

В третьем периоде, который длится с конца 80-х годов и продолжается в настоящее время, появляются пакеты программ для построения нечетких экспертных систем, а области применения нечеткой логики заметно (по сути искусственного интеллекта) расширяются.

Триумфальное шествие нечеткой логики по миру началось после доказательства в конце 80-х Бартоломеем Коско знаменитой теоремы FAT (Fuzzy Approximation Theorem) о связи теории вероятностей и нечеткой логики.

В бизнесе и финансах нечеткая логика получила признание после того как в 1988 году экспертная система на основе нечетких правил для прогнозирования финансовых индикаторов единственная предсказала биржевой крах.

И количество успешных фаззи-применений в настоящее время исчисляется тысячами.

Теория вероятностей – наука о закономерностях массовых случайных явлений

.

Лаплас, Пуассон, Гаусс, Бернулли, П.Л. Чебышев, А.М. Ляпунов А. А. Марков, А.Н. Колмогоров и др.

- социально - экономическая статистика; многомерные статистические методы;
- эконометрика; эконометрическое моделирование; методы социально-экономического прогнозирования; СППР;
- страхование и актуарные расчеты; теория риска и моделирования рисковых ситуаций;
- маркетинг; теория массового обслуживания; технический и фундаментальный анализ,
- теория планирования эксперимента; теория надежности; теория информации (статистическая радиотехника),
- выборочный контроль качества и др.

Семинары.

№1. Модели. Параметрические (аналитические) модели, виды, свойства, атлас моделей для их предложения к реальным временным и пространственным выборкам экономических объектов.

1.1. Математические модели, переменные и параметры, линейные и нелинейные модели, временные ряды, пространственные ряды, пространственно-временные ряды, эволюционные модели, виды зависимостей.

1.2. Функции и графики в экономическом моделировании, основные элементарные функции (линейная, параболическая, степенная, логарифмическая, показательная и обобщенная показательная, обратная и обобщенная обратная, гармоническая и их графики.

2.Выбор подхода при выборе методов моделирования и прогнозирования

- 2.1.Параметрический (аналитический) подход:
- Достоинства: относительно малые выборки (до 30 наблюдений), возможность для слабо структурированных реализации системного подхода (декомпозиции) для моделирования и последующего прогнозирования.
- Недостатки: сложность идентификации нелинейных моделей, в частности при мультипликативной структуре стохастической компоненты (гомоскедастичность, гетероскедастичность, условия получения оптимальных оценок Гаусса-Маркова проверка знаний из курса «Эконометрика»).
- 2.2.Альтернатива - алгоритмический подход (на примере сезонности, эволюции, «средней температуры по больнице»):

Достоинства: простота, универсальность.

- Недостатки: требования больших выборок, невысокая точность, трудности декомпозиции (сложных трендов (мультитредов) – проверить знания Ряда Тейлора и при представлении колебательных компонент в виде ряда Фурье – проверить знания), практическая невозможность прогнозирования (лишь при сложных процедурах адаптации и потере при этом универсальности).

2.2.Свойства аналитических функций при их выборе для моделирования трендов (временных и пространственных)

- Функция У=f(x) , где y – определяемая (зависимая) переменная, x-независимая или определяющая переменная.

  Область определения, область изменения.

  Способы задания функции аналитической формулой, таблицей, графиком.
- Четность f(x)= f(-x), сумма, разность, произведение и частное нечетных функций есть четная функция: y= x^{2n}, где n- натуральное число, y=|x|.
- Нечетность f(-x)= - f(x): y= x^{2n+1}, y= $\frac{x}{x^2+4}$.
- Не всякая функция является либо четной, либо нечетной: y = $\sqrt{x}$, y= x^2+3x,
- y= $(x+1)^2$, их называют аморфными.
- Сумма и разность нечетных функций есть функция нечетная, а произведение и частное нечетных функций – функция четная.
- График четной функции симметричен относительно оси ОУ, а нечетной – относительно центра 0.

Монотонность, асимптоты (вертикальные, горизонтальные, наклонные), ограниченные функции, обратная функция и ее график), сложная функция, неявная функция

- У=Аx + в, У= $^{к}/_{x^2}$, У= X^{2m}, У=X^{2m+1} , где m- натуральное число,
- У= $^{1}/_{(X-1)}$

- У= X^{α} , α - больше и меньше 0. Асимптоты, область определения, область существования, графики.

Преобразования графиков:

- f(x) ⇒ f(x)+b, f(x) ⇒ f(x + b),
- f(x) ⇒ k f(x),

  f(x) ⇒ f(k x),

- $Y = \dfrac{A_o}{1 + A_1 \exp(-\alpha x)}$ - построить графики функции при различных значениях

параметров, дать характеристики динамики траектории.

- Геометрический смысл производной.
- Экстремумы – минимумы и максимумы.
- Необходимые и достаточные условия экстремума функций одной и двух переменных.
- Периодичность (модели сезонности) и апериодичность (модели циклов) аналитических функций.

Например, в экономической литературе по курсу «методы принятия оптимальных решений» рассматривалась ситуации принятия решений на основе только структурированной (детерминированной) информации. Они являются повторяющимися, могут быть формализованы и автоматизированы. Каждая проблема имеет при этом, как правило, единственное решение возможно, получаемое и различными методами.

Реальная экономическая практика требует учитывать реальную вероятностную погрешность (или нечеткость) в используемой информации (в выборках, в наблюдениях). Такие слабоструктурированные проблемы – совмещают в себе знания ЛПР и возможности компьютера. Они выступают в качестве симбиоза человека и машины. Компьютер используется как вычислитель, но решение остается за человеком. При этом, их решения могут существенно отличаться друг от друга.

Неструктурированные проблемы (хаос) вообще предполагают решение лишь на основе человеческой интуиции и его рассуждений. Задача либо вообще не имеет решения, либо имеет их слишком много.

Точность моделирования характеризуется абсолютными или относительными значениями. Наиболее известен - коэффициент детерминации R^2 с диапазоном изменения [0, 1]. Допустим для принятия решений ЛПР диапазон значений > 0,7 (0,9).

Адекватность модели - соответствие данных с точки зрения цели моделирования. Адекватность, как более общее понятие, включает в себя точность. Не всякая точная модель адекватна, но если модель неточная, то она не будет адекватной. Адекватность определяется (суммой) точностью моделирования ($R^2 \geq 0{,}7$) и точности прогнозирования (< 20% - хорошая). Структурная идентификация обычно сложнее параметрической.

Актуальны для экономической практики не статические, а динамические модели (включающие в себя время): аналитические решения дифференциальных уравнений (рассматриваемых в курсе математической экономики) или аналитические феноменологические (полученные путем обработки только (без теории) статистических данных объекта) для реализации возможности моделирования и прогнозирования.

Архиактуален для современной практики также мониторинг (контроль) эволюции моделей динамики с возможностью использования относительно малых выборок (< 30 наблюдений) для идентификации. Такие модели позволяет получить современная эконометрика, классическая статистика не может быть использована.

Использование результатов моделирования объекта для прогнозирования его будущего состояния

Покажем суть прогнозирования на примере временных (можно и пространственных) траекторий. Моделирование (идентификацию) проводят на рабочей выборке. Объем выборки контрольной выборки и горизонт прогноза примерно равны, обычно менее трети от рабочей выборки.

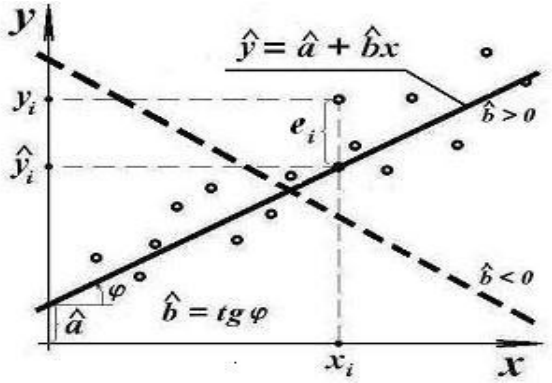
.

- Модель обычно имеет целью прогнозирование траектории объекта для последующего планирования действий по улучшению (м. б. коррекции) неблагоприятного состояния объекта, если оно определяется прогнозом. Идентификация модели в общем случае включает в себя определение аналитического вида модели (спецификацию – или структурную идентификацию) и параметров траектории (параметрическую идентификацию).
- Затем, тем или иным образом рассчитают разницу (ошибку или функцию потерь – например, применением МНК) между реальными и модельными значениями траектории.
- Считая условия функционирования объекта моделирования неизменными, принимают ошибку на горизонте прогноза такой же и в дальнейшем (после окончания использованной для моделирования исходной выборки).

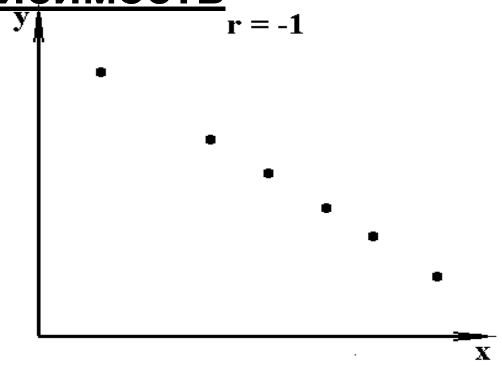
Нам кажется, что мы - заложники неведомого будущего…

На самом деле мы – пленники <u>хорошо</u> известного прошлого!

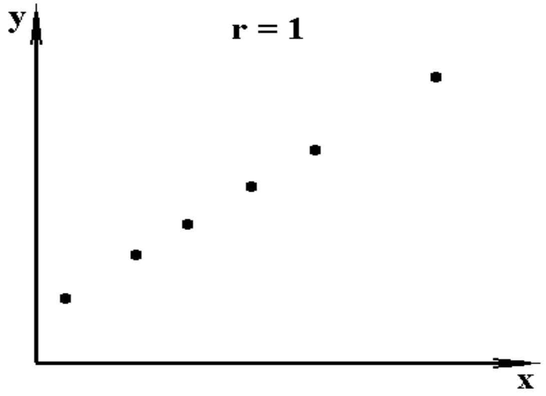
Самой простой и популярной моделью является парная линейная регрессия



Отрицательная функциональная зависимость



Положительная функциональная зависимость

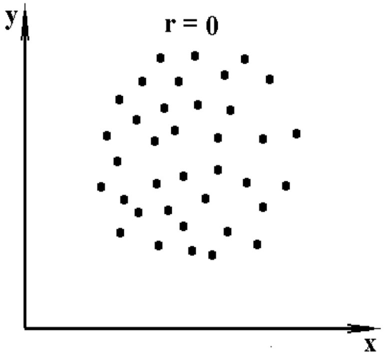


парная линейная регрессия

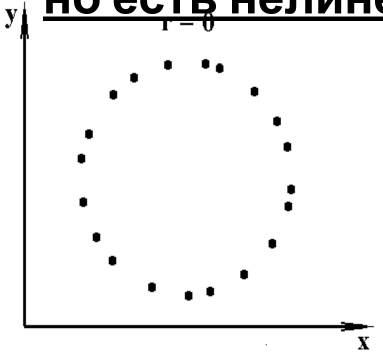
Задача состоит в получении таких оценок (случайных) коэффициентов (параметров) уравнения регрессии $\hat{a}$ и $\hat{b}$, чтобы обеспечить минимальное значение квадратической функции потерь:

$$min\left(\sum_{i=1}^{n}(e_i)^2\right) \Rightarrow \hat{a}, \hat{b}$$

полное отсутствие связи



отсутствует линейная связь,
но есть нелинейная



Ряды Тейлора и Фурье

Более общим решением аналитического моделирования сложных моделей трендов являются ряды. Например, если *аналитическая модель тренда непрерывна и* имеет *все производные* при $x = a$, то тренд можно представить в виде суммы нескольких первых членов достаточно простого степенного ряда Тейлора:

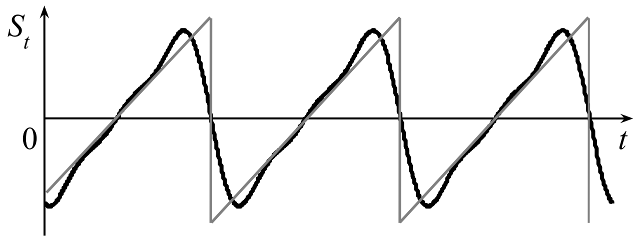
$$f(x) = f(a) + \frac{x-a}{1!} f^{(1)}(a) + \frac{(x-a)^2}{2!} f^{(2)}(a) + \ldots + \frac{(x-a)^n}{n!} f^{(n)}(a) + \ldots$$

При *a = 0* имеем, как частный случай, ряд Маклорена. Известно и разложение в ряд Тейлора функции многих (двух) переменных.

Периодическую функцию f(x) ($S(t)$) с известным периодом можно заменить суммой с определяемыми специальным образом коэффициентами a_n, b_n и кратными частотами. Связь угловой частоты с периодом $\omega = \frac{2\pi}{T}$, а ряд Фурье (1807г.) имеет вид:

$$s_n(x) = \frac{a_0}{2} + a_1 \mathrm{Cos}\,\omega x + a_2 \mathrm{Cos}\,2\omega x + \ldots + a_n \mathrm{Cos}\,n\omega x + b_1 \mathrm{Sin}\,\omega x + b_2 \mathrm{Sin}\,2\omega x + \ldots + b_n \mathrm{Sin}\,n\omega x + \ldots$$

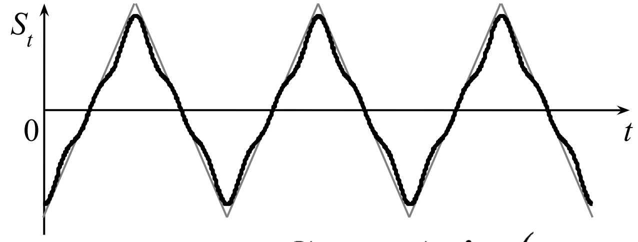
Аппроксимация периодических колебаний рядами Фурье



Колебания пилообразной формы

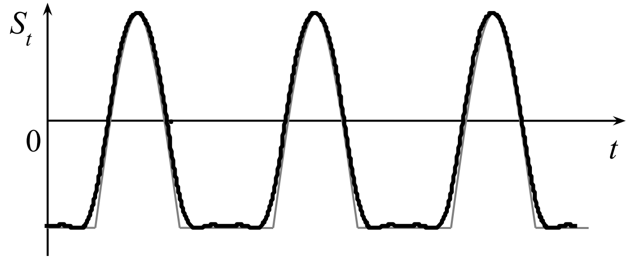
(см.теория управления запасами),

$$S_t = A\sin(\omega t+\psi) - 0{,}4A\sin(2(\omega t+\psi)) + 0{,}15A\sin(3(\omega t+\psi))$$



Колебания треугольной формы,

$$S_t = A\sin(\omega t+\psi) - 0{,}15A\sin(3(\omega t+\psi)) + 0{,}1A\sin(5(\omega t+\psi))$$



Колебания куполообразной формы
(моделирование сезонных отпусков)
и др.

Применяется модель циклов Е. Слуцкого из суммы гармоник с некратными частотами :

$$S_t = A_1\sin(\omega_1 t+\psi_1) + A_2\sin(\omega_2 t+\psi_2) + A_3\sin(\omega_3 t+\psi_3)$$

Мы сами убедились в глобальной нелинейности эволюционной динамики (при эволюции социально-экономическая система приобретает – или теряет - качественно новые свойства), моделируя тенденции роста отраслей экономики: только 3.3% из моделей роста оказались парными линейными регрессиями (из 9 рассмотренных более сложных нелинейных моделей) для 12 отраслей экономики в 78 регионах РФ за период времени в последние 12 лет. Возможно использование парных регрессий лишь на начальных участках роста систем.

Добавим, что погрешность прогнозирования будет еще больше: при принятии линейной модели динамики в макроэкономике будет не менее 50%, если реальная динамика нелинейна.

Нужен поиск и применение более сложных моделей роста с переменной динамикой (с изменением первой и второй производной).

Оценки параметров парной линейной регрессии решением «нормальной» системы алгебраических линейных уравнений (СЛАУ)

функция потерь:

$$Q(\hat{a},\hat{b})=\sum_{i=1}^{n}(y_i-\hat{a}-\hat{b}x_i)^2;$$

необходимые условия экстремума

$$\Rightarrow\begin{cases}\dfrac{\partial Q}{\partial\hat{a}}=-2\sum_{i=1}^{n}\left(y_i-\hat{a}-\hat{b}x_i\right)=0\\ \dfrac{\partial Q}{\partial\hat{b}}=-2\sum_{i=1}^{n}\left(y_i-\hat{a}-\hat{b}x_i\right)x_i=0\end{cases}$$

$$\begin{cases}\sum_{i=1}^{n}y_i-\hat{a}n-\hat{b}\sum_{i=1}^{n}x_i=0\\ \sum_{i=1}^{n}y_ix_i-\hat{a}\sum_{i=1}^{n}x_i-\hat{b}\sum_{i=1}^{n}x_i^2=0\end{cases}\Rightarrow\begin{cases}\dfrac{1}{n}\sum_{i}y_i-\hat{a}-\hat{b}\dfrac{1}{n}\sum_{i}x_i=0\\ \dfrac{1}{n}\sum_{i}y_ix_i-\hat{a}\sum_{i}x_i-\hat{b}\dfrac{1}{n}\sum_{i}x_i^2=0\end{cases}$$

$$\hat{b}=\frac{M(x\cdot y)-M(x)\cdot M(y)}{M(x^2)-M^2(x)}\qquad \hat{a}=M(y)-\hat{b}\cdot M(x)\qquad M(\,)=\sum_{i=1}^{n}(\,)$$

- **оценки матожидания,**

оценки коэффициента корреляции:

$$r=\frac{M(x\cdot y)-M(x)\cdot M(y)}{S_x\cdot S_y}$$

и среднеквадратических отклонений:

$$S_x=\sqrt{M(x^2)-M^2(x)}\qquad S_y=\sqrt{M(y^2)-M^2(y)}$$

центрированная с.в.: $\overset{\circ}{x}=X-m_x.$

Точечные оценки точности прогнозирования и моделирования

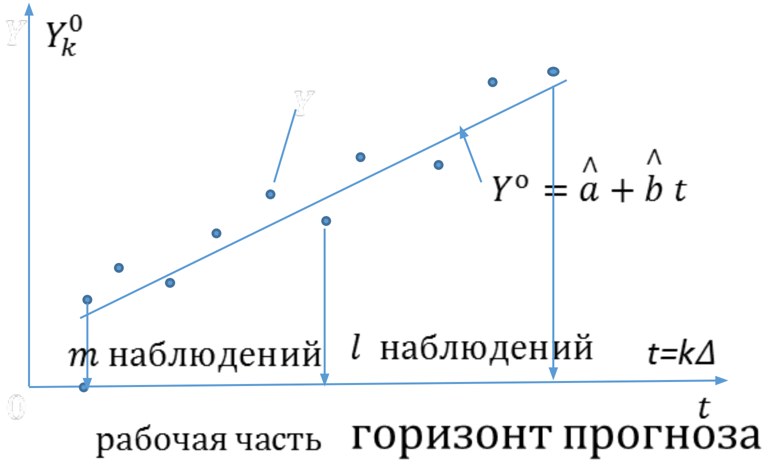
$$R^2 = \frac{\sum_{k=1}^{n}\left(Y_k^0 - \frac{1}{n}\sum_{k=1}^{n} Y_k\right)^2}{\sum_{k=1}^{n}\left(Y_k - \frac{1}{n}\sum_{k=1}^{n} Y_k\right)^2}$$

\- для оценки точности моделирования наиболее распространен коэффициент детерминации: отношение **двух дисперсий** (с предложеннной моделью и без нее).

$t = k\Delta$ $\qquad k = l\Delta$

Для точности прогноза используют несколько оценок:

$(m + l = n)$



1. MAPE – оценка:

$$MAPE = \frac{1}{l}\sum_{k=1}^{n}\left|\frac{Y_k - Y_k^o}{Y_k}\right| \cdot 100\%$$

не чувствительна к выбросам,

погрешность $\leq$ 10 – отличная,

$\leq$ 20 – хорошая

2. Второй коэффициент Тейла

удобен при малых значениях Y

$$k_{T2} = \sqrt{\left(\sum_{k=n}^{n+l-1}\left(Y_k - Y_k^o\right)^2\right) \Big/ \left(\sum_{k=n}^{n+l-1} Y_k^2 + \sum_{k=n}^{n+l-1}\left(Y_k^o\right)^2\right)} \cdot 100\%.$$

нелинейность доверительного интервала для парной линейной регрессии

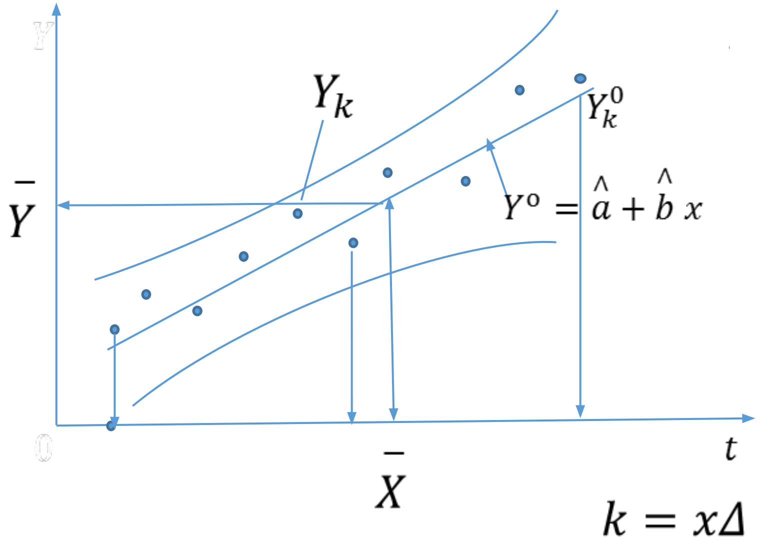
В знаменателе – мера разброса относительно среднего значения $\bar{Y}$. В числителе - мера разброса (квадрат сумм невязок или функция потерь) относительно модели.

$$R^2 = \frac{\sum\limits_{k=1}^{n}\left(Y_k^0 - \frac{1}{n}\sum\limits_{k=1}^{n} Y_k\right)^2}{\sum\limits_{k=1}^{n}\left(Y_k - \frac{1}{n}\sum\limits_{k=1}^{n} Y_k\right)^2}$$

$t = k\Delta$

$k = x\Delta$

Коэффициент детерминации характеризует долю разброса определяемой переменной, объясняемой регрессией Y на X.



Более универсальной характеристикой точности моделирования является не точечная, а интервальная. **Любой прогноз, не содержащий расчета доверительного интервала**, по сути не отличается от экспертной оценки, т.к. доверительный интервал может оказаться бесконечно большим, а вероятность значения траектории – близкой к нулю.

На рисунке показан доверительный интервал для парной линейной регрессии **(пояснить)**.

Точечные оценки точности оценок (статистик) генеральных числовых характеристик θ_x

***N*-генеральная совокупность**

дающая точную оценку параметра

$$\{x_1, x_2, \dots x_N\} \to \theta_x;$$

$$\{x_{11}, x_{21}, \dots x_{n1}\} \to \theta^*_{x1};$$

$$\{x_{12}, x_{22}, \dots x_{n2}\} \to \theta^*_{x2};$$

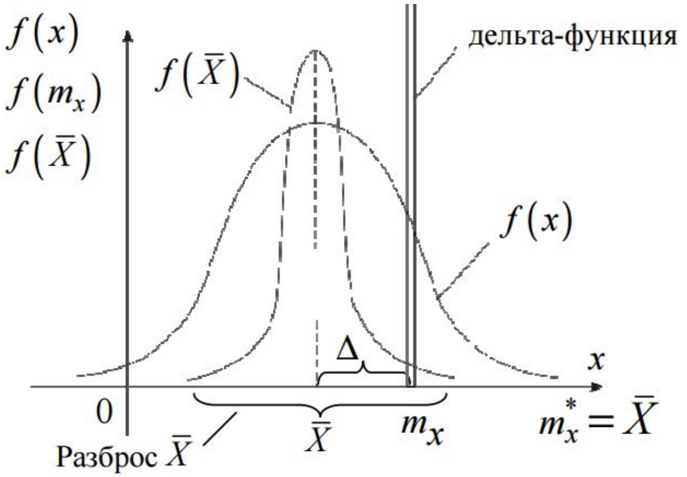
....

$$\{x_{1k}, x_{2k}, \dots x_{nk}\} \to \theta^*_{xk}.$$

По выборке объемом *n* найдем *k* оценок того же параметра

θ^*_{xi} - с. в. и поэтому характеризуются законами распределения (плотностью или интегральной функцией распределения), числовыми характеристиками (матожиданием, дисперсией, коэффициентом детерминации, квантилем и т.д.).

1. Оценка называется состоятельной, если она сходится по вероятности к оцениваемой характеристике $\lim\limits_{n\to\infty} P\left(\left|\theta_i^* - \theta\right| < \varepsilon\right) = 1,\ \forall \varepsilon > 0.$
2. Оценка параметра называется несмещенной, **если** $M[\theta_i^*] = \theta$.
   $M[\theta_i^*] - \theta = \Delta$ – смещение, систематическая погрешность.
   Асимптотическая несмещенная оценка: $\lim\limits_{n\to\infty} M[|\theta_i^* - \theta] = 0.$



Можно говорить **о законе распределении оценки матожидания, о матожидании оценки матожидания, о дисперсии оценки матожидания**…

МНК дает оптимальные (эффективные, состоятельные, несмещенные) оценки $\boldsymbol{\beta_0}, \boldsymbol{\beta_1}$ (т.е. $\boldsymbol{b_0}, \boldsymbol{b_1}$) лишь при соблюдении условий (Гаусса-Маркова):

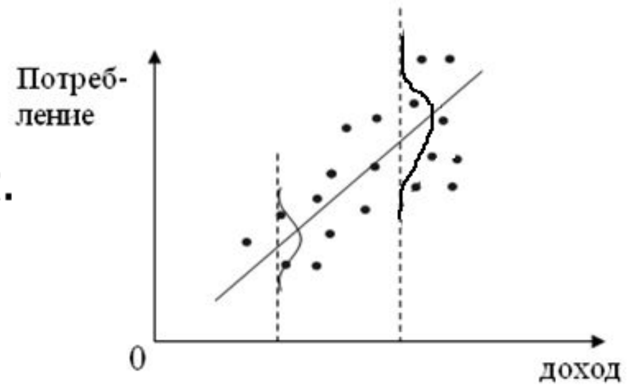
1. $\mathbf{M}[e] = \mathbf{0}$ – невязка (шум, стохастическая переменная) является центрированной с.в.;

2. $\mathbf{D}[e] = \mathbf{const} < \infty$ при выполнении этого условия невязка гомоскедастична, в противном случае – гетероскедастична.

Гетероскедастичность (неравноточность оценок по оси аргумента) является одним из наиболее нежелательных проблем при идентификации и, силу этого, компенсируется специальными приемами.

3. $\mathbf{M}[e_i\, e_j] = \mathbf{0}$ - некоррелированность невязок;

4. $e \sim N(0, S)$ – нормальный закон распределения невязок.



$$\begin{cases} M\left[\hat{a}\right] = \beta_0 \\ D\left[\hat{a}\right] = min \end{cases}; \qquad \begin{cases} M\left[\hat{b}\right] = \beta_1 \\ D\left[\hat{b}\right] = min \end{cases};$$

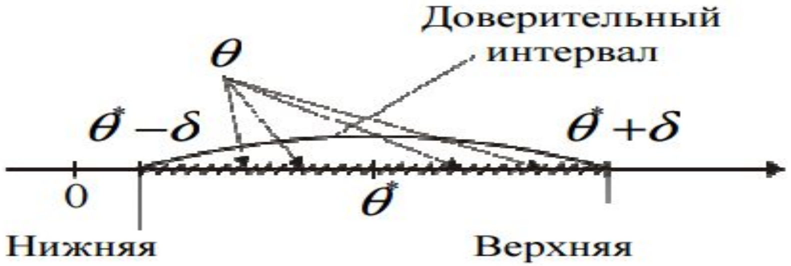
$\Uparrow$

<u>Желательна несмещенность и эффективность оценок параметров регрессии</u>

Интервальная оценка точности (надежность) генеральных математического ожидания и дисперсии

Доверительный интервал - интервал значений, в котором с заданной доверительной вероятностью (обычно назначают β = *0,9; 0,95; 0,99*) находится истинное значение оцениваемой статистической характеристики :

$$\theta_x : P[|\theta - \theta^*| < \delta] = \beta \rightarrow P(\theta^* - \delta < \theta < \theta^* + \delta) = \beta$$



Радиус доверительного интервала равен:

$$\delta = \frac{t\,\sigma_\theta}{\sqrt{n}} \approx \frac{t\,\sigma_{\theta^*}}{\sqrt{n}}$$

t- аргумент, соответствующий значению функции Лапласа, равной *β/2* : $\boxed{\Phi(t) = \beta/2}$; σ_{θ^*}

Доверительный интервал случаен (зависит от конкретных выборок): случайно его положение на числовой оси и случайна его длина.

При $n \rightarrow \infty \rightarrow \delta \rightarrow 0$, а при $\beta \rightarrow 1 \rightarrow \delta \rightarrow \infty$.

Скорректированный коэффициент детерминации при оценки точности моделирования

- В общем случае коэффициент детерминации лежит в пределах от нуля до единицы. Когда модель совершенно не объясняет реальные данные $\boldsymbol{R^2}$ может быть и отрицательным.
- Это может быть вызвано и вычислительными ошибками и неверным выбором модели. При усложнении модели $\boldsymbol{R^2}$ неизбежно возрастает, поэтому он не всегда может служить критерием выбора модели.
- Часто используют скорректированный коэффициент детерминации, учитывающий количество параметров в модели *m*:

$$\bar{\boldsymbol{R}}^2 = \mathbf{1} - (\mathbf{1} - \boldsymbol{R}^2)\frac{\boldsymbol{n} - \mathbf{1}}{\boldsymbol{n} - \boldsymbol{m} - \mathbf{1}} < \boldsymbol{R}^2.$$

- При $m > 5$ (и более) разница становится существенной. Поэтому $\bar{R}^2$ рекомендуется использовать не для оценки качества моделирования, а лишь для моделей с различным числом параметров.

Обоснование выбора вида моделей при моделировании и прогнозировании

Параметрический (аналитический) подход к выбору вида модели:

Достоинства: возможность использования на относительно малых выборках и для слабо структурированных объектов системный подход (декомпозицию) для моделирования и последующего прогнозирования. **Недостатки**: сложность идентификации нелинейных моделей, а также мультипликативных структур компонент траекторий (в частности, мультипликативной структуры вхождения стохастической компоненты, которая может обусловить гетероскедастичность оценок параметров моделей при применении МНК.

Алгоритмический подход *к выбору вида модели* не требует аналитических моделей, использует, например, укрупнение интервалов выборки, сглаживание методом скользящей средней, применения модели Хольта - Уинтерса и др.

Достоинства: простота, универсальность. **Недостатки**: требования больших объемов выборок траектории, малая точность, трудности декомпозиции (мультитредов и при представлении колебательных компонент в виде ряда Фурье), практическая невозможность прогнозирования (лишь при сложных процедурах адаптации с потерей при этом универсальности).

Эконометрика - «существует только то, что можно измерить»

Экономика Метрика (измерение)

Макроэкономика
Микроэкономика - **основы образования экономиста**
Эконометрика

Экономическая теория
Эконометрика: Математика Математическая экономика
Статистика Математическая, экономическая статистика

В централизованной экономике эконометрика не изучалась. Набор статистических методов, используемых для наблюдения за ходом развития экономики, её анализа (моделирования и прогнозирования) **называется эконометрикой.** Причинно-следственными связями занимается экономическая теория, а связями без выявления их причин - эконометрика.

Эконометрика, как и статистика, из данных формирует знания.

Пример иллюстрации задач СППР

- Предложена модель функции потребления:

$$lnC = \beta_0 + \beta_1\, lnY + \beta_2\, lnP,$$

где *C* - потребление некоторого пищевого продукта на душу населения в некотором году; *Y* - реальный доход на душу населения в том же году; *P* - индекс цен на этот продукт, скорректированный (дефлированный) на общий индекс стоимости жизни т.е. относительный уровень цен; β_i – параметры;

C, Y, P – измеряемые показатели.

Задача «параметризации» модели – определение β_i:

1. Измерение показателей (насколько корректно они измерены, представляют ли они то, что должны представлять по нашему мнению)?
2. Погрешность измерения ε показателей *Y, P, C* . Она аддитивна (суммируются с моделью) или мультипликативна (умножается на нее)?
3. Какой метод использовать при решении задачи параметрической идентификации – метод наименьших квадратов (МНК)?

Задача «спецификации» модели

- Нет ли переменных, которые следовало бы дополнительно включить в уравнение (например, цены на непродовольственные товары). Не следует ли исключить из уравнения некоторую переменную?
- Верно ли то, что модель линейная (верны ли теоретические предпосылки модели)?
- Является ли модель полной? Может быть необходимо учесть и уравнение предложения, кроме уравнения спроса?
- Достаточно ли изучать макроэкономическое уравнение или необходимы такие индивидуальные (микроэкономические) данные для поставленной задачи, может быть нужно выделить факторы, которые зависят непосредственно от принятия управленческих решений данным объектом хозяйствования (предприятием, регионом), и влияние факторов, которые от менеджмента на данном хозяйствующем объекте не зависят? Модель является статической. Возможно, более подходящей была бы динамическая модель? Можно предположить, что прошлогодний доход может влиять на текущий уровень потребления.

Конструкция и задачи СППР в эконометрике:

- **Конструкция:** предназначена для различного уровня агрегирования объекта анализа, может быть адаптирована для группового и индивидуального использования, зависит от вида задач, для решения которых она разрабатывается, от наличия доступных данных и знаний (знание, как правило, является результатом обработки данных);

•**Задачи**: поддерживать как взаимозависимые, так и последовательные решения на всех фазах процесса решения: интеллектуальной части, проектирования и выбора; может реализовывать разнообразные стили и методы решения, что может быть полезно при решении задачи группой ЛПР.

- должна быть гибкой, адаптироваться к изменениям как организации, так и ее окружения, должна быть проста в использовании и модификациях;
- должна улучшать эффективность процесса принятия решений, позволяя человеку управлять процессом принятия решений с помощью компьютера;
- должна поддерживать эволюционное использование и легко адаптироваться к изменяющимся требованиям, осуществлять и моделирование, и прогнозирование.

Итак, СППР - скоординированный набор данных, систем, инструментов и технологий, программного и аппаратного обеспечения, с помощью которого пользователь собирает и обрабатывает информацию о бизнесе и окружающей среде с целью обоснования управленческих решений.

Некоторые области применения СППР:

Моделирование и прогнозирование (наш курс и не только):

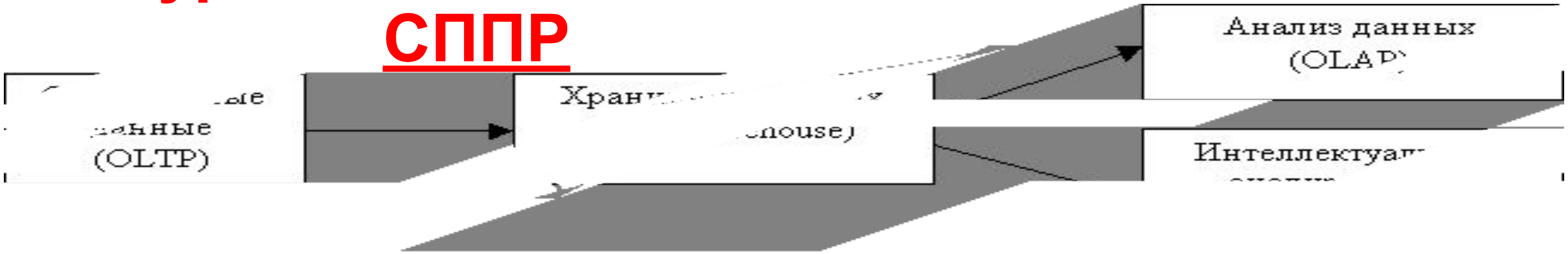
- Анализ каналов снабжения и распределения (логистика);
- Анализ производства; Анализ продаж;
- Маркетинговый анализ; Анализ клиентской базы;
- Анализ качества сервисного обслуживания; Финансовый анализ;
- Инвестиционный анализ; Анализ персонала;
- Анализ исследовательской и проектной деятельности;
- Анализ стратегии организации; Управление рисками;
- Многомерный анализ сложных систем
- и др.

Примеры решения конкретных задач с помощью СППР:

- обоснование направлений развития систем высшего образования;
- выбор методов завоевания рынка бытовой техники;
- оценка привлекательности регионов для трудоустройства людей, окончивших вуз, в ближайшие 10 лет;
- оценка перспективности видов альтернативного горючего для автомобилей;
- распределение средств между проектами социальной программы гуманитарной направленности;
- отбор научно-технических проектов в рамках конкурса;
- выбор перспективных направлений информатизации страны и пр.

В последнее время СППР начинают применяться и в интересах малого и среднего бизнеса (например, для выбора варианта размещения торговых точек, выбор кандидатуры на замещение вакантной должности, выбор варианта информатизации и т.д.).

Архитектурно-технологическая схема СППР



Первоначально информация хранится в оперативных базах данных (OLTP). Она используется в процедурах многомерного анализа (OLAP) и для интеллектуального анализа данных (ИАД), который реализует:

1) выявление закономерностей (моделирование);

2) использование выявленных закономерностей для прогноза;

3) анализ исключений (аномалий или выбросов) в найденных закономерностях, осуществление мониторинга эволюции.

ИАД реализуют экспертные и интеллектуальные системы, базы знаний и данных, компьютерное моделирование, нейронные сети, нечеткие системы, метод симуляции, основанный на идее метода Монте-Карло - ресамплинг, бутстреп и др.

СППР предназначены для выбора субъекта кредитования, исполнителя работы, назначения на должность, использования в торговых предприятиях, торгующих дорогими товарами длительного пользования, и пр. Такие системы используют информацию о результатах решений этой же задачи, принятых в прошлом. В большинстве реальных случаев СППР должна помочь ЛПР формализовать его собственные представления о ценности полученных результатов и затратах на их получение.

Главным в СППР является не вычислительная часть, а технологическая поддержка процедуры корректного извлечения и формализации субъективных требований и предпочтений специалистов, а также процедуры пошагового агрегирования информации под контролем аналитика. Примером слабо структурированного объекта являются, например, задачи моделирование и прогнозирование жизненного цикла (ЖЦ) продукта (товара, услуги и т. д), которые и рассмотрим далее.

Структуры траекторий определяемого параметра. Декомпозиция

- **Структуры траекторий** определяемых параметров объектов: пространственных (тренды и стохастические компоненты) и временных (тренды, колебательные и стохастические компоненты). **Тренд-сезонная декомпозиция** временных траекторий - форма реализации системного подхода.

Детерминированная часть траектории D_t, состоящая из тренда $T(t)$ и колебательной компоненты $S(t)$ учитывает наиболее значительные по величине и более стабильные, долговременные факторы, формирует **регулярную динамику**.

Колебательные компоненты, как правило, меньше по величине, чем тренд, но обладают большей динамикой (изменчивостью), могут быть **циклическими** и **сезонными.** Сезонная описывает - внутригодовые колебания, а циклическая — многогодовые.

Стохастическая компонента $\varepsilon(t)$ отличается от тренда существенно меньшей величиной траектории и существенно большей динамикой.

Для трендов используют **десятки** отдельных аналитических выражений и их взаимодействий, которое может быть также описано аналитическим выражением, или, в более общем случае, **разложением в ряд Тейлора**.

Периодические компоненты представимы гармоникой или **разложением в ряд Фурье**.

Компонента $\varepsilon(t)$ описывается законами распределения и числовыми характеристиками.

Ряды Тейлора и Фурье

Более общим решением аналитического моделирования сложных моделей трендов являются ряды.

Например, если *аналитическая модель тренда* ***непрерывна*** *и* имеет *все производные* при $x = a$, то его можно представить в виде суммы нескольких первых членов достаточно простого степенного ряда Тейлора:

$$f(x) = f(a) + \frac{x-a}{1!} f^{(|)}(a) + \frac{(x-a)^2}{2!} f^{(||)}(a) + \ldots + \frac{(x-a)^n}{n!} f^{(n)}(a) + \ldots$$

При *a* = *0* имеем, как частный случай, ряд Маклорена. Известно и разложение в ряд Тейлора функции многих переменных.

Периодическую функцию f(x) ($T(t)$ *или* $S(t)$ *) с известным периодом* можно заменить приближенно суммой с определяемыми специальным образом коэффициентами a_n, b_n и кратными частотами. Связь угловой частоты с периодом $\omega = \frac{2\pi}{T}$. Ряд Фурье имеет вид:

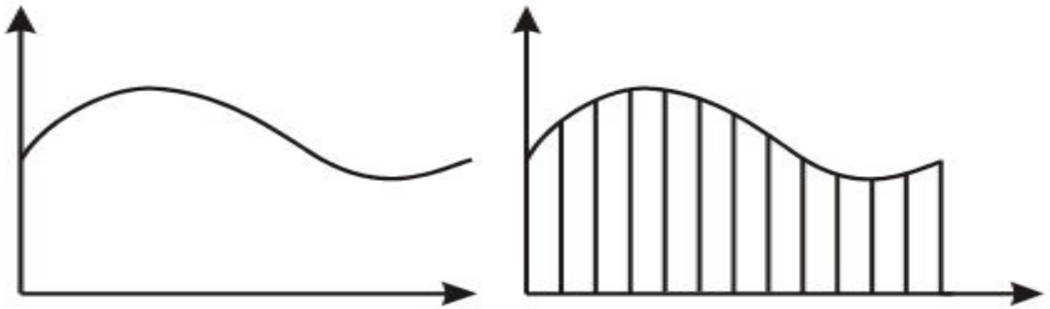
$$s_n(x) = \frac{a_0}{2} + a_1 \mathrm{Cos}\,\omega x + a_2 \mathrm{Cos}\,2\omega x + \ldots + a_n \mathrm{Cos}\,n\omega x + b_1 \mathrm{Sin}\,\omega x + b_2 \mathrm{Sin}\,2\omega x + \ldots + b_n \mathrm{Sin}\,n\omega x.$$

Условие формирования рядов экономической (пространственной и временной) динамики

Примеры пространственной динамики: спрос в функции цены; обьемы продаж в функции рекламного бюджета, обьем привлекаемых банком средств в функции нормы затрат на привлекаемые средства и т.д.

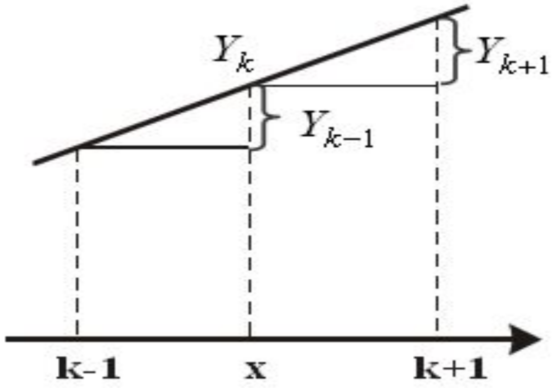
Примеры временной динамики: ежегодные, ежеквартальные, ежемесячные или ежедневные данные по прибыли, остаткам на счетах банка, данные на кв. метр жилья, обьемы перевозок авиакомпании, показатели качества добываемой бокситовой руды, данные по ВНП, ВВП, обьему экспорта, инфляции и т.д.

Равномерная дискретизация $\Rightarrow$ Решетчатые функции



$Y_{t-1},\ Y_t,\ Y_{t+1},\ Y_{t+2}$

$x_{t-1},\ x_t,\ x_{t+1},\ x_{t+2}$



Левая первая разность

$$\Delta Y_k = Y_k - Y_{k-1}$$

Правая первая разность

$$\Delta Y = Y_{k+1} - Y_k$$

Симметричная первая разность

$$\Delta Y_k = Y_{k+1} - Y_{k-1}$$

$$Y^{(1)} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}, \quad \Delta_k^2 = \Delta y_k - \Delta y_{k-1}$$ — вторая разность

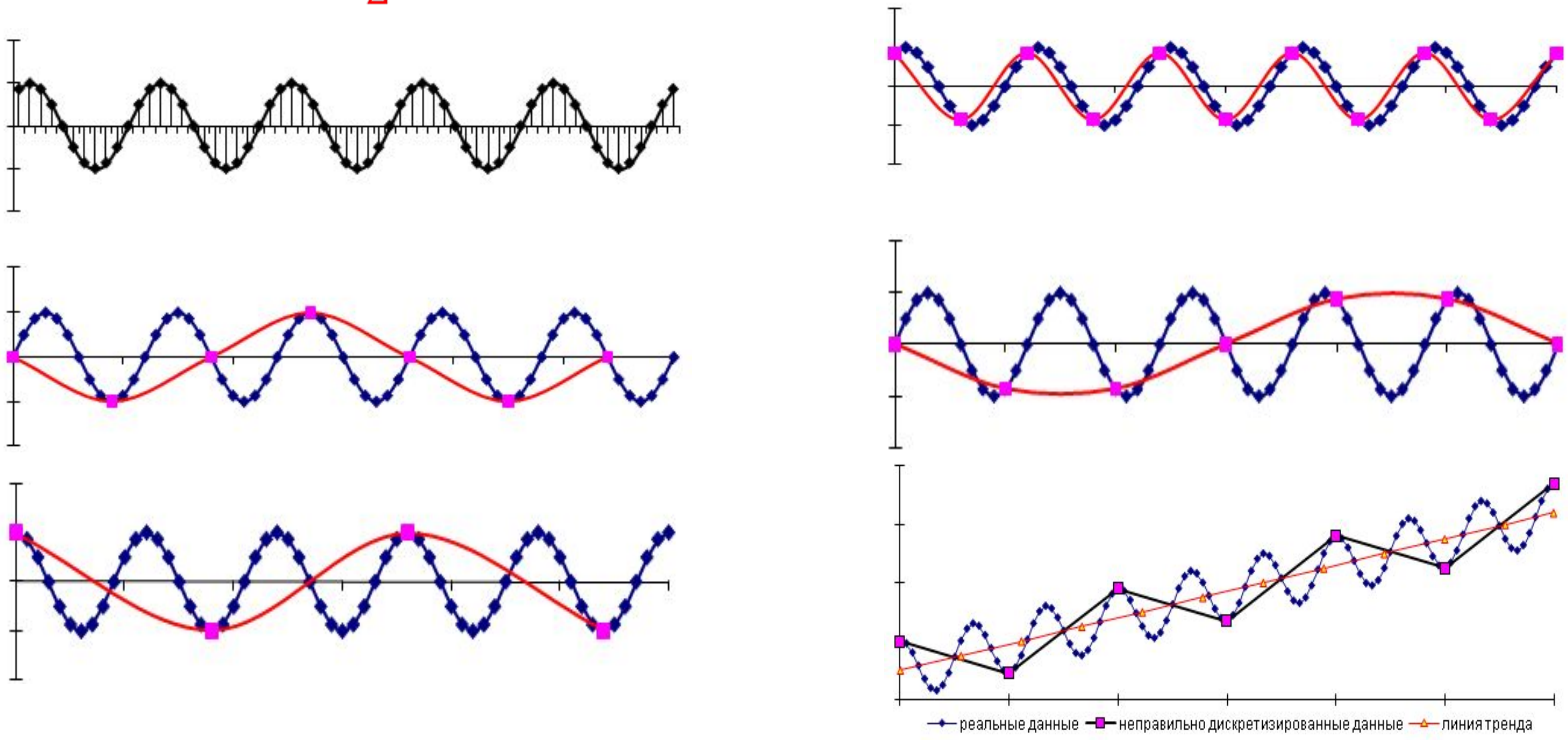
Вместо дифференциальных уравнений широко используют разностные модели.

Спектр временной или пространственной траектории показывает все частоты гармоник, присутствующих в модели детерминированной компоненты.

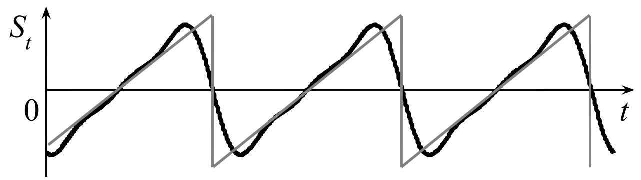
Формирование эквидистантных временных рядов:

$f_{\max}$ - максимальная частота спектра траектории, интервал дискретизации . Δ
Условие Найквиста (Котельникова) выбора периода дискретизации для правильной передачи частоты и фазы самой высокочастотной гармоники в спектре с частотой

$f_{\max}\,(\omega_{\max}, T_{\min})$: $5 < \dfrac{T}{\Delta} < 10.$

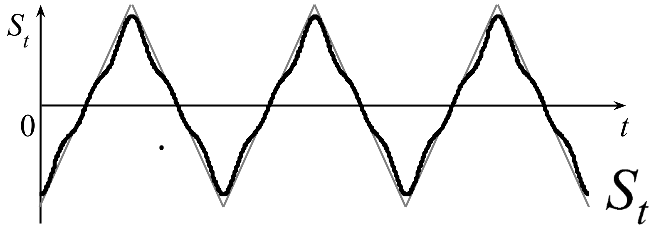


Три члена разложения в ряд Фурье передает многообразие моделей колебательной компоненты



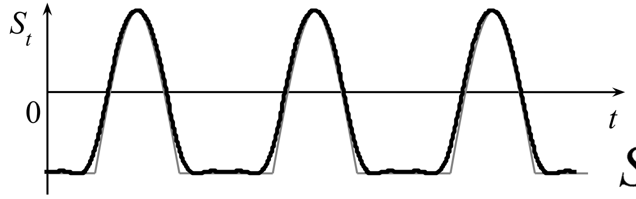
Колебания пилообразной формы,

$$S_t = A\sin(\omega t+\psi) - 0{,}4A\sin(2(\omega t+\psi)) + 0{,}15A\sin(3(\omega t+\psi))$$



Колебания треугольной формы,

$$S_t = A\sin(\omega t+\psi) - 0{,}15A\sin(3(\omega t+\psi)) + 0{,}1A\sin(5(\omega t+\psi))$$



Колебания куполообразной формы

$$S_t = A\cos(\omega t+\psi) + 0{,}4A\cos(2(\omega t+\psi)) - 0{,}05A\cos(4(\omega t+\psi))$$

Используют и суммы некратных частот:

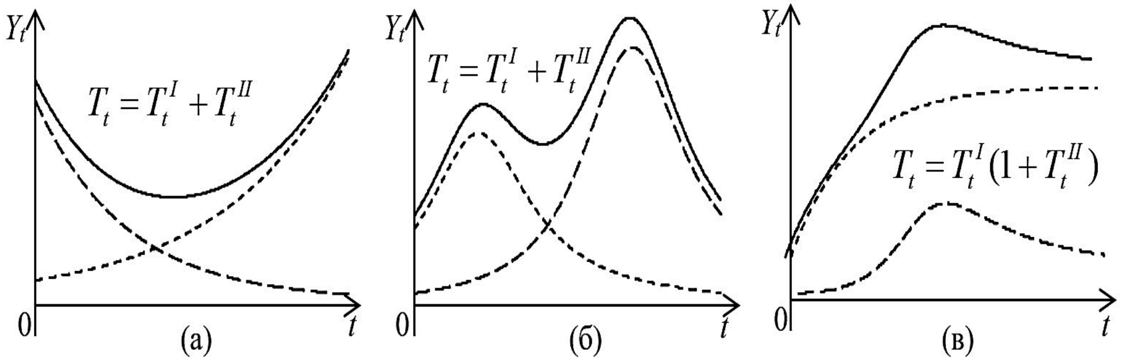
$$S_t = A_1\sin(\omega_1 t+\psi_1) + A_2\sin(\omega_2 t+\psi_2) + A_3\sin(\omega_3 t+\psi_3) \quad \omega_1, \omega_2, \omega_3$$

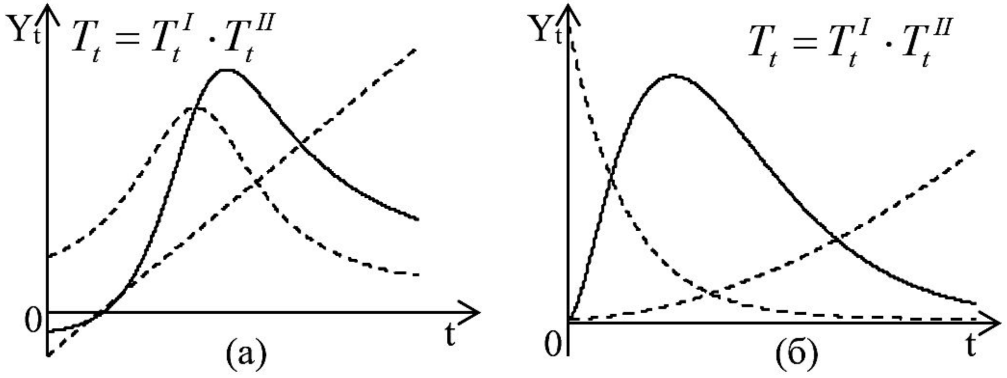
Мультитренды, как более сложное и гибкое представление тенденции

Наряду с десятками аналитических моделей трендов, начиная с линейной, в последнее время аналитики часто рассматривают и более сложные выражения, называемые мультитрендами, для моделирования тенденции.

Наиболее простым вариантом формирования мультитрендов является алгебра взаимодействия двух (и более) трендов: аддитивное, мультипликативное и пропорционально-мультипликативное, определяемое следующими формулами:

$$T(t) = T_1(t) + T_2(t), \qquad T(t) = T_1(t)(1 + T_2(t)), \qquad T(t) = T_1(t) \cdot T_2(t).$$





Структуры (канонические) трендовой модели с детерминированной компонентой

Аддитивная структура трендовой модели (условного математического ожидания от времени) – наиболее простой и наиболее часто используемый в экономической практике случай, предполагает независимость детерминированной и стохастической компонент ε_k, их размерность в натуральных единицах (в штуках, тоннах и т.п.):

$$Y_k = D_k + \varepsilon_k$$

Мультипликативная структура трендовой модели со стохастической компонентой

$$Y_k = D_k \cdot \mu_k$$

существенно сложнее в идентификации (практически не используется), но в большей мере соответствует стоимостным показателям.

Детерминированная компонента в натуральных показателях при этом зависит от стохастической μ_k, которая изменяется в долях (в процентах) в диапазоне от 0 до 1, а закон ее распределения «для удобства» принимается обычно логнормальным, чтобы потом можно было применить для «линеаризации» операцию логарифмирования и получить тем самым нормальное распределение

При реализации системного подхода для слабоагрегированного объекта анализа выполняется тренд-колебательная (неоднозначная) декомпозиция, которая включает в себя условия:

$\varepsilon(t) \ll T(t),$ $|S(t)| < |T(t)|$ $|\varepsilon(t)| < |S(t)|$

1. **Тренд** (вековая компонента);
2. **колебательная компонента: апериодические циклы, периодическая сезонная компонента**;
3. календарная компонента;
4. выбросы (аномальные значения);
5. инфляционная компонента.

Неоднозначность тренд-колебательной декомпозиции требует сравнения: выполнения условия адекватности моделей, устранения выбросов, величинами точности моделирования и/или прогнозирования и учетом предыдущих (исторических) результатов.

Многообразие возможных структур декомпозиции: при взаимодействии тренда и колебательных (сезонной и циклической) компонент

$$Y_k = T_k + C_k + \varepsilon_k \qquad Y_k = T_k \cdot C_k \cdot \mu_k$$

- наиболее известные (аддитивная и мультипликативная) или «канонические» структуры взаимодействия компонент (свойства стохастических компонент в них различны!).

Предложены:

$$Y_k = T_k\,(1+S_k) + C_k + \varepsilon_k = T_k + T_k\,S_k + C_k + \varepsilon_k$$

$$Y_k = T_k + S_k + C_k(1+\varepsilon_k) = T_k + S_k + C_k(1+\varepsilon_k)$$

$$Y_k = T_k\,(1+C_k) + S_k + \varepsilon_k = T_k + T_k\,C_k + S_k + \varepsilon_k$$

$$Y_k = T_k\,(1+C_k + S_k) + \varepsilon_k = T_k + T_k\,(C_k + S_k) + \varepsilon_k$$

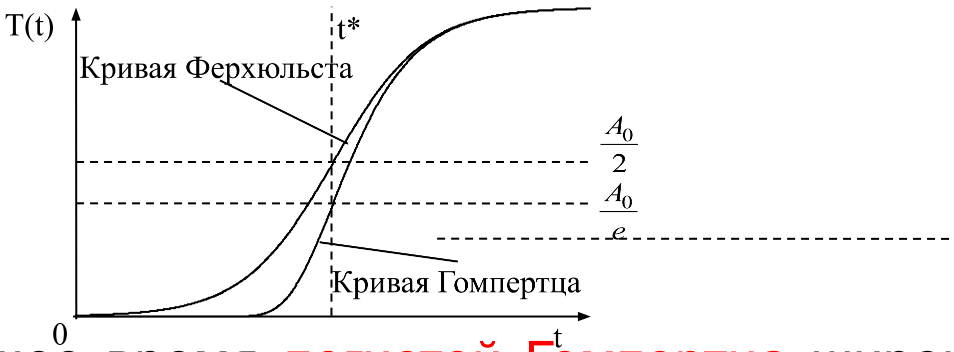
и другие параметрические модели. Видим возможность существования в модели одновременно и аддитивных и пропорциональных тренду колебательных компонент.

Все приведенные здесь примеры моделей компонент могут быть параметрическими (в отдельных случаях – гетероскедастические). Почему компонента ε_k одинакова в них?

Пример тренда: дифференциальное уравнение для тренда Гомпертца:

$$\frac{dT}{dt} = \alpha T \ln\left(\frac{A_0}{T}\right)$$ **где** $T(0) = A_0\, e^{-\exp\,(\alpha t_0)}$

Аналитическое решение данного уравнения – известная «кумулятивная» или «накопленная» логиста, точка перегиба и ее ордината будут, соответственно :



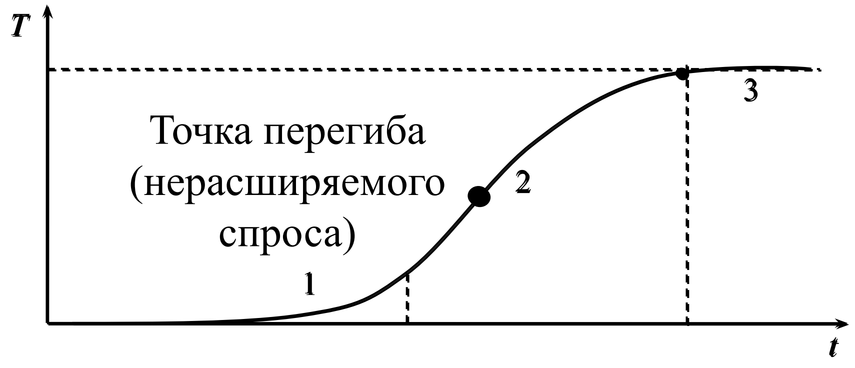
$$T(t) = A_0 e^{-\exp(-\alpha(t - t_0))}$$

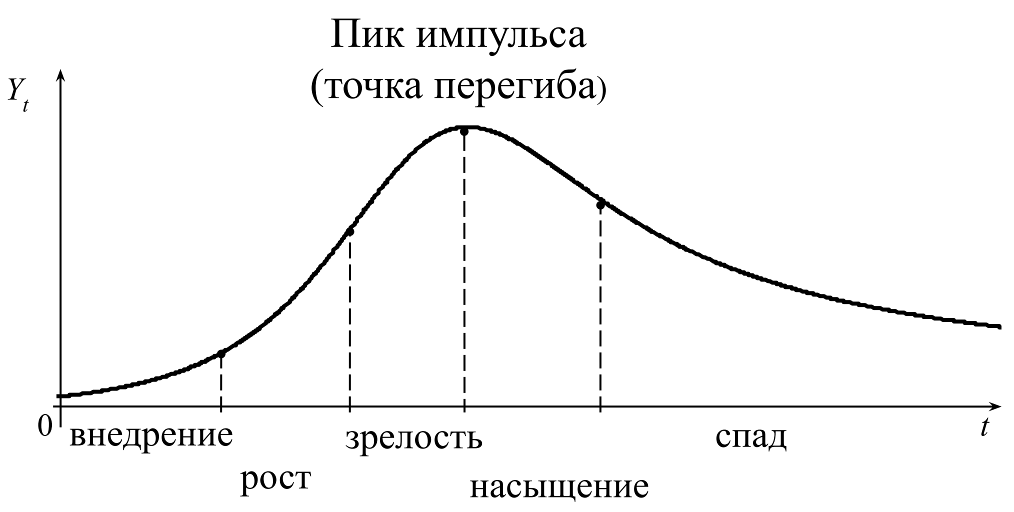
$$T(t^*) = \frac{A_0}{e} \quad t^* = t_0$$

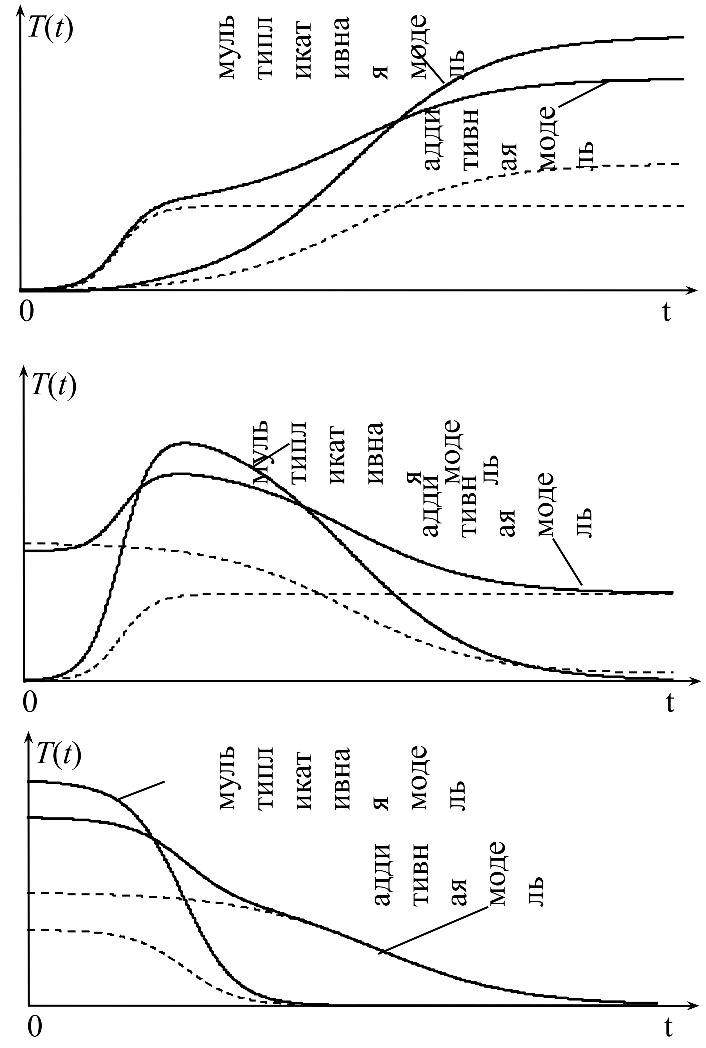
В настоящее время логистой Гомпертца широко пользуются при моделировании динамики роста опухолей, числа абонентов сотовой телефонии, численности населения, потребительских товаров длительного пользования, инноваций в сельском хозяйстве и т.д.

Логиста Гомпертца асимметрична по отношению к половине уровня насыщения в отличие от другой, также наиболее известной, логисты Ферхюльста, т.е. обладает левой асимметрией. Важны для многочисленных приложений логисты с произвольной (настраиваемой) асимметрией.

Производная от кумулятивной (накопленной) логисты – «импульсная логиста». Используют и формирование импульсных и кумулятивных логист алгеброй двух логист



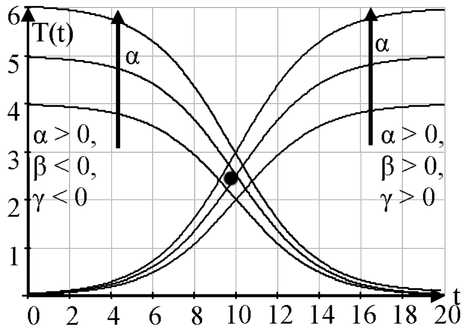


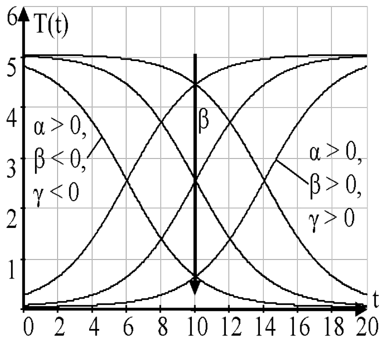


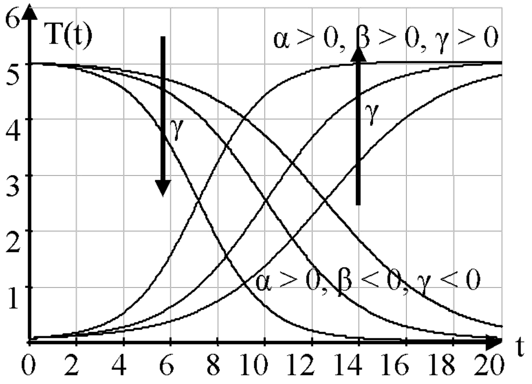
Выбор модели путем сравнения разных функций и атласы функций с разными параметрами для ВИЗУАЛЬНОГО предложения той или иной модели для выборки

Альбом параметров симметричной логистической функции Верхулста (Ферхульста)

$$T = \frac{\alpha}{1 + e^{(\beta - \gamma t)}}\ .$$



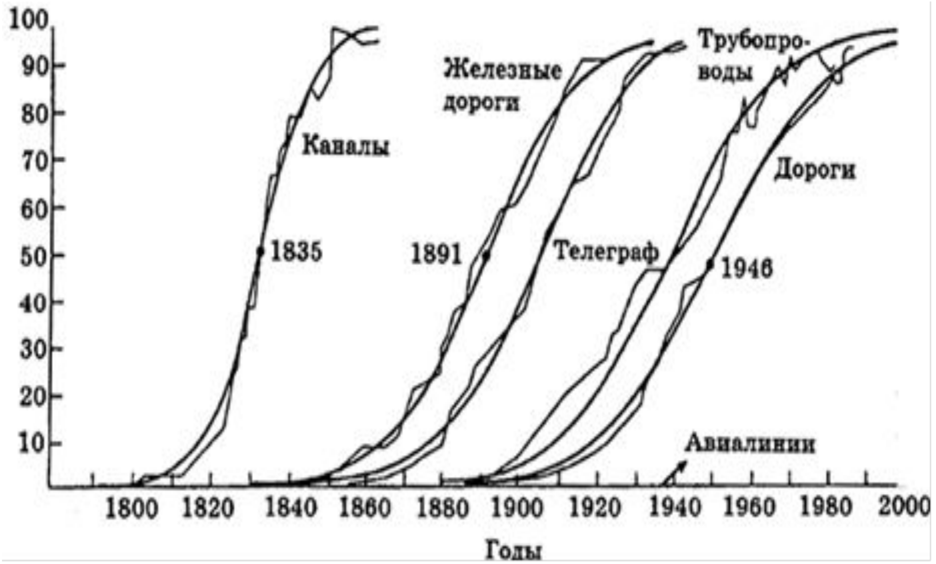




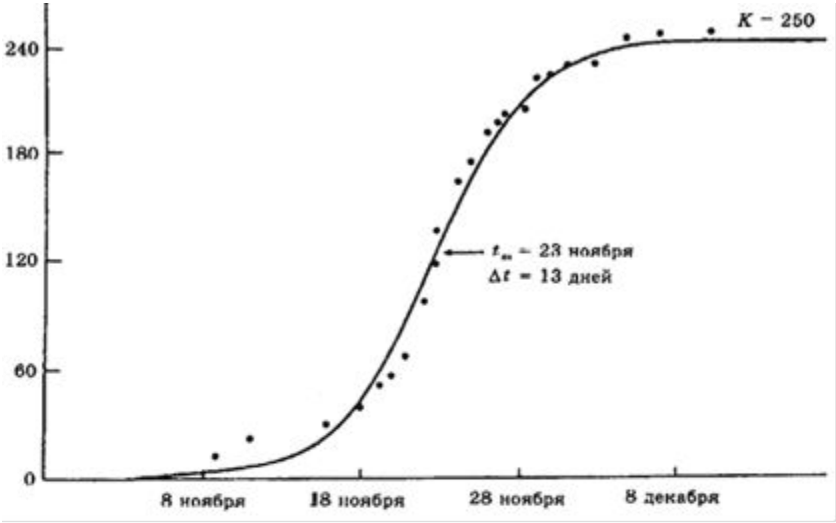
Для вычислительной устойчивости идентификации рекомендуем возможно простые модели, разбиение выборки на участки (что позволяет учет эволюции), учет истории прежних моделирований, анализ механизма исследуемого экономического явления, использование асимптот.

Логисты (их известно более 20) описывают динамику объектов с разным характером динамики в условиях разных экономических ограничений, возможны их взаимодействия.

<u>Многообразие (универсальность) форм (и далеко не всех приложений) логистических моделей (основных моделей эволюции)</u>



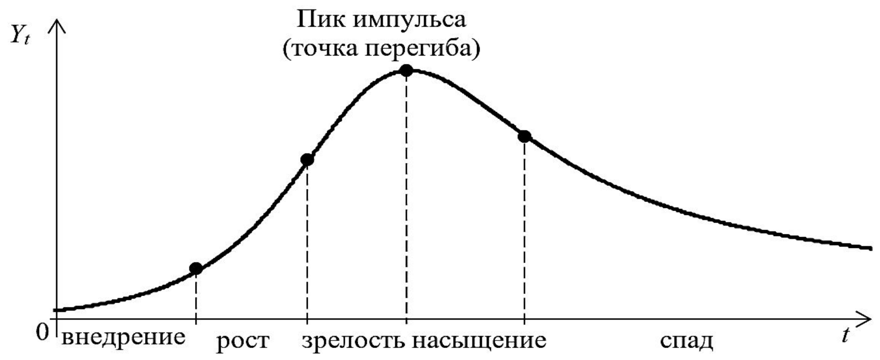
<u>Динамика развития инфраструктуры США</u>



<u>Динамика движения протеста (количество разрушенных с/х машин в Англии в 1830г.)</u>

Распространение демократических форм правления (к 2100г.- 90% населения), наркотиков, численности биологических видов, добыча невосполняемых ресурсов, эффективность рекламы и т.д. – моделируют кумулятивные и импульсные логисты.

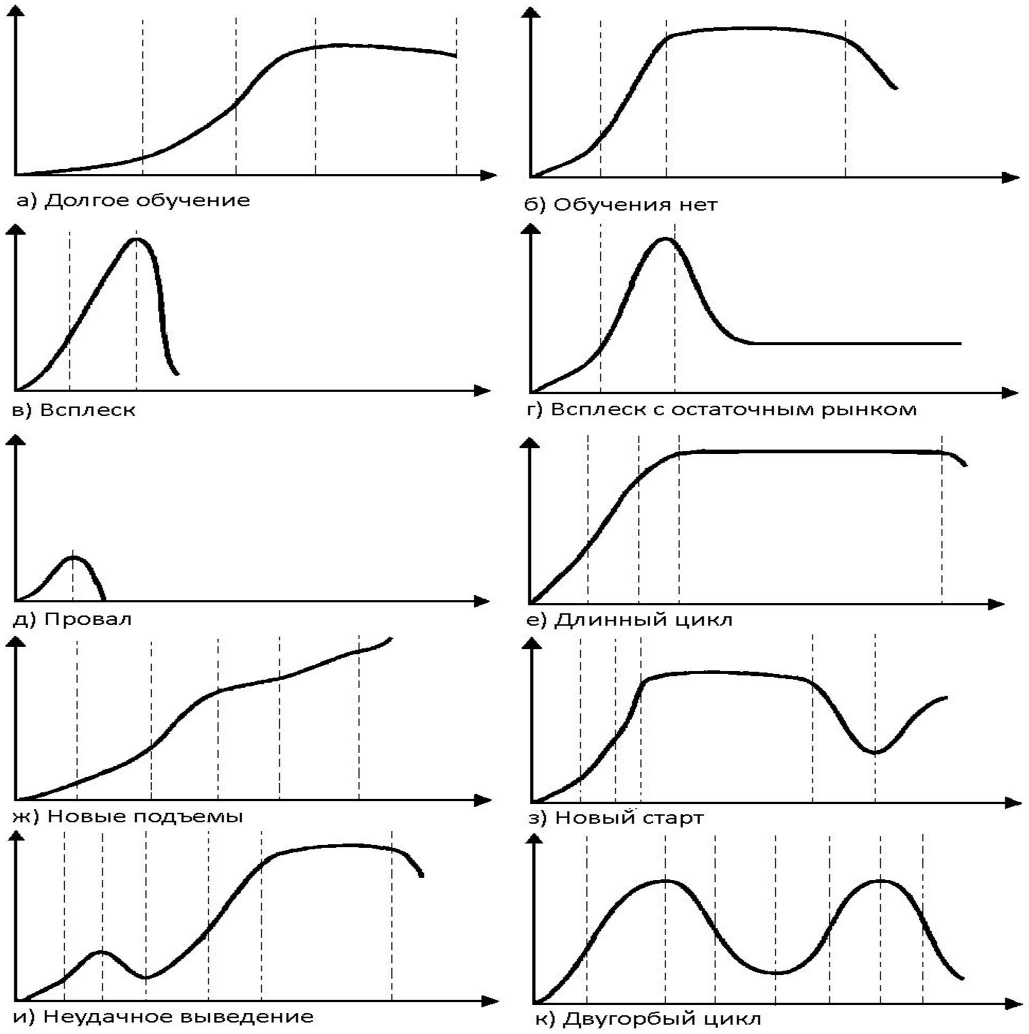
Жизненные цикла продукта, как примеры эволюции экономических объектов



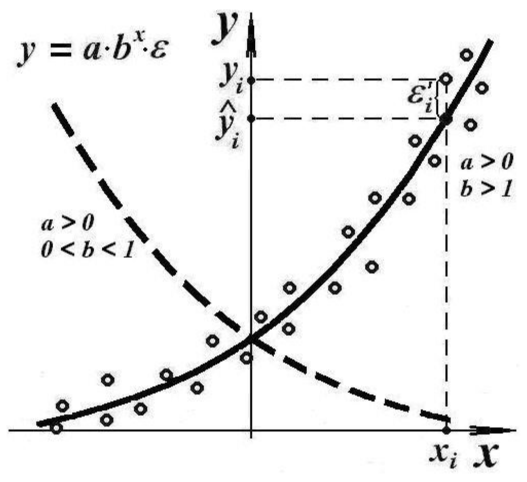
Этапы модели ЖЦП. Ранее были известны были лишь графические и вербальные описания траектории. Эконометрика предлагает и аналитические модели (алгеброй более 250 л

⇒

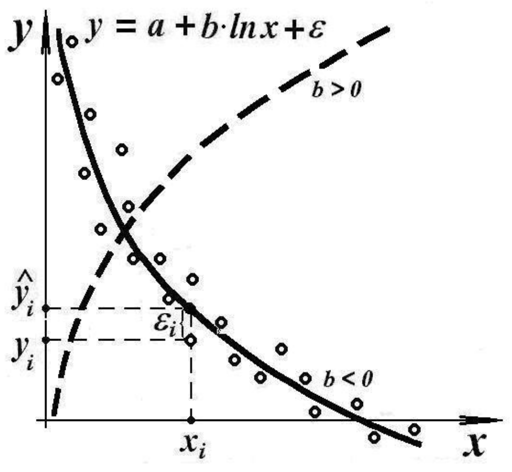
Сегментация потенциальных потребителей нового продукта по признаку индивидуальной предрасположенности к восприятию инновации



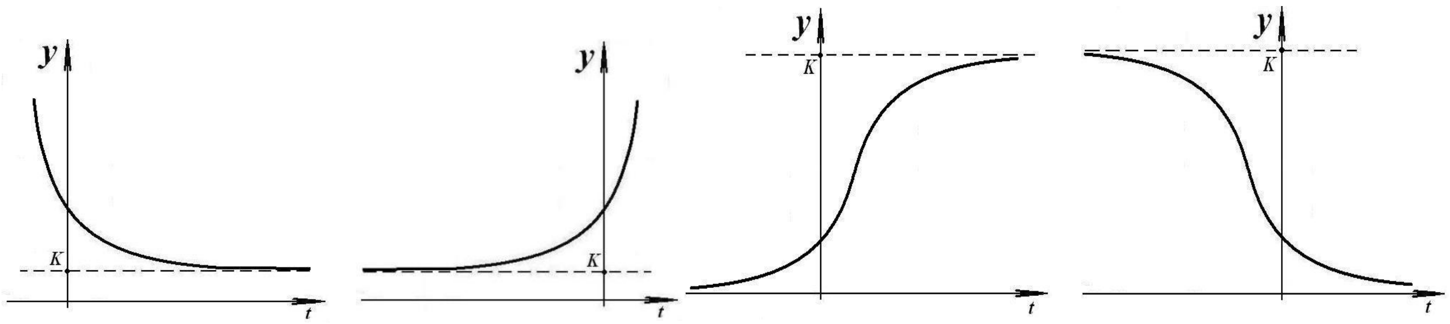
Другие функции (линейные и нелинейные) по параметрам и определяющим переменным.



Показательная функция



логарифмическая функция



•

Функции Гомпертца при различных значениях параметров (атлас видов моделей)

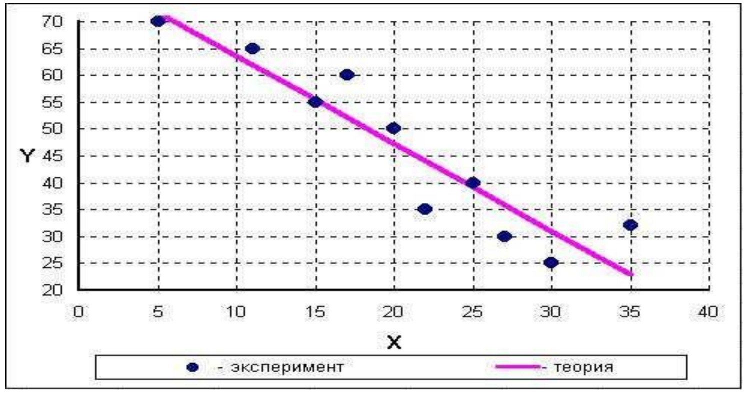
$Y = K \cdot a^{b^t}$ – логистическая функция Гомпертца широко используется в банковском и страховом деле, демографии, для товаров длительного пользования и т.д. Она левую асимметрию относительно точки перегиба.

Рисунки показывают ее вид при различных сочетаниях параметров *K*, *a*, *b*.

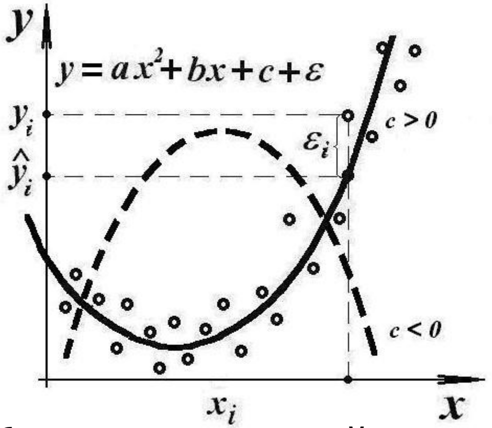
Функция может быть записана и в другой форме:

$$Y = A \cdot e^{\left(-\frac{\alpha}{\gamma} e^{-\gamma t}\right)}.$$

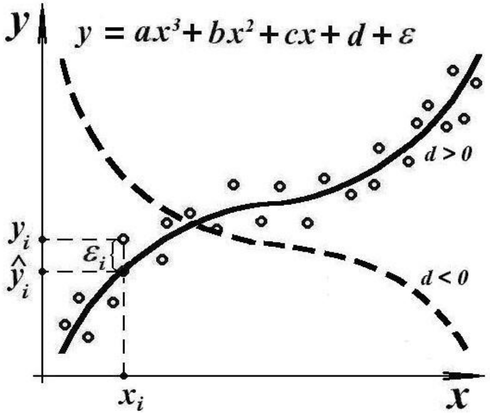
Популярные виды (линейные и нелинейные по параметрам и переменным) парные модели



Парная линейная регрессия



Квадратический полином



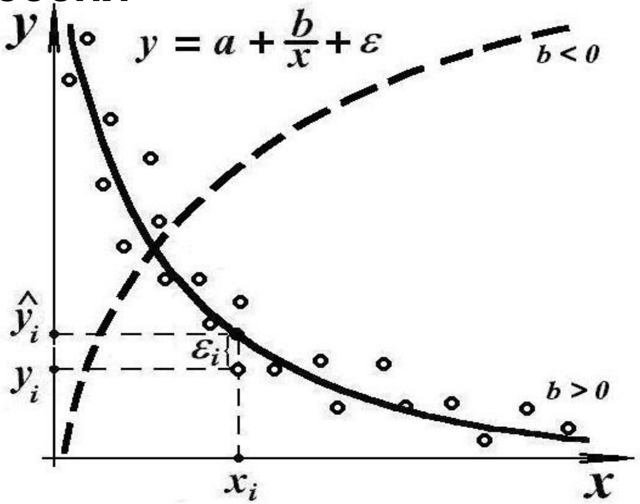
Кубический полином

(применяют

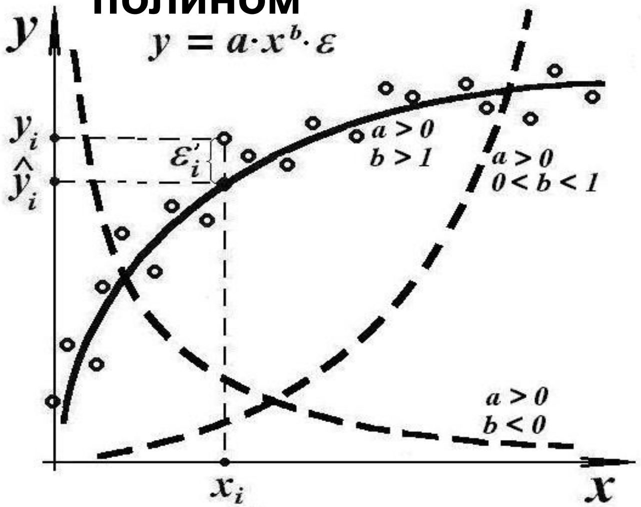
для моделей с двумя

точками

экстремума)



Гиперболический полином



Степенная функция

Формирование многообразия видов логистических трендов

- 1. Решение дифференциальных уравнений, описывающих динамику объектов.
- 2. Дифференцирование кумулятивных логист дает импульсные логисты.
- 3. Назначение тех или иных параметров моделей может трансформировать вид логисты (например, делать из кумулятивной логисты импульсную: например, трансформация логисты Гомпертца и логисты Рамсея).
- 4. Применением алгебры из двух и более импульсных и/или кумулятивных логист.
- 5. Феноменологические («мягкие» или статистические, не формируемые детерминированно) модели логист.

Преимущества феноменологических моделей, получаемых обработкой реальных данных:

-возможность выделения компонент ряда в структуре модели (одного или нескольких трендов, циклической компоненты, сезонной компоненты при различных способах их взаимодействия (аддитивного, мультипликативного, смешанного) и тем самым найти их экономическую интерпретацию;

-удобство количественного анализа динамики комплекса моделей через исследования нулей функции, асимптот, экстремумов, точек перегиба;

-возможность построения в этом случае как краткосрочных, так и средне- и долгосрочных прогнозов;

-возможность прогнозирования неоднородных по быстроте эволюционных изменений, в том числе при отсутствии аналогов на имеющейся рабочей (ретровыборке), но заложенных в нелинейном виде модели;

-учет особенностей конкретной реальной ситуации (исторических данных и теоретических конструкций) при выборе модели из возможного их комплекса на этапе спецификации.

Канонические (аддитивная и мультипликативная) структуры временных рядов

- Аддитивная структура $\boxed{Y_k = T_k + S_k + \varepsilon_k}$

- Мультипликативная структура $\boldsymbol{Y_k = T_k(1 + S_k)(1 + \varepsilon_k)}$

$\boldsymbol{T_k}$, $\boldsymbol{S_k}$, $\boldsymbol{\varepsilon_k}$ – ненаблюдаемые (unobserved) компоненты ряда.

Задача: выделить ненаблюдаемые компоненты ряда динамики. Возможны неоднозначные предложения, требующие количественного подтверждения и, по возможности, теоретического толкования.

Подходы:

- непараметрический (Census I, Census II);
- новый итерационный параметрический.

Метод Census I

Непараметрический метод, разработан американским агентством Census Bureau в 1954г.

В результате получают значения рядов ненаблюдаемых компонент в виде таблиц.

Этапы:

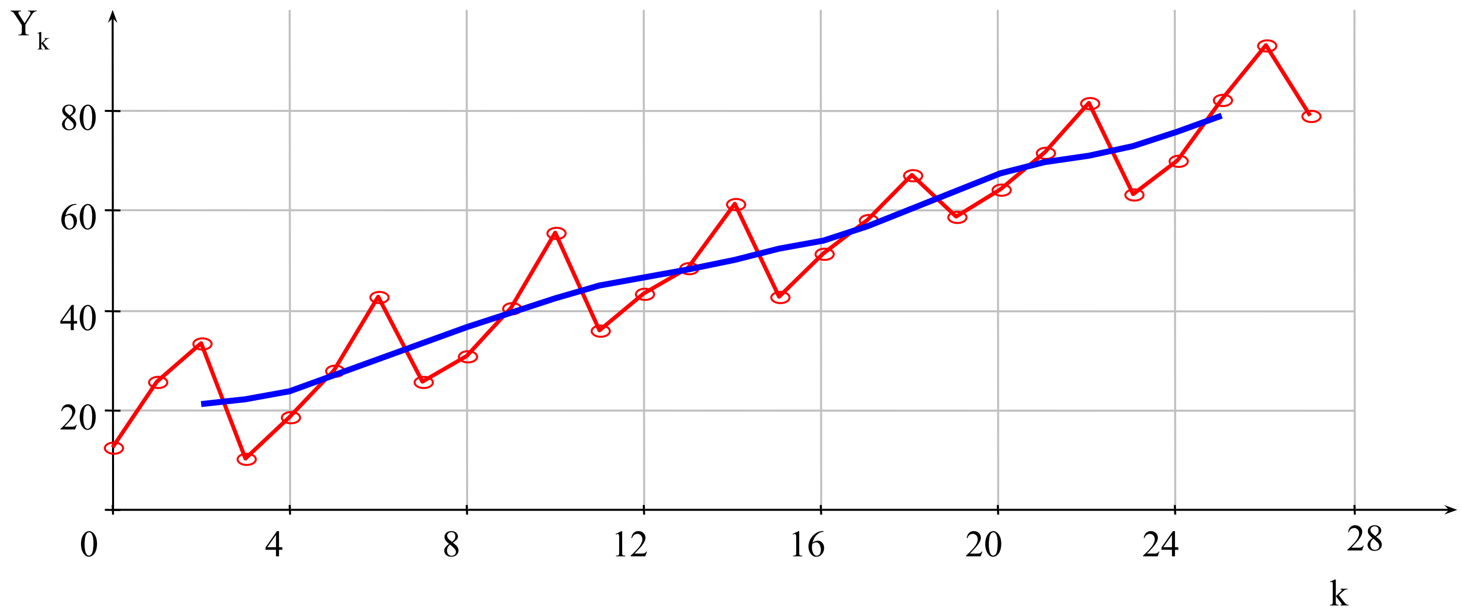
1. Выделение тренда.
2. Детрендирование.
3. Выделение сезонной компоненты.
4. Десезонализация.
5. Повторное выделение тренда.
6. Выделение стохастической компоненты.

Выделение тренда

В начале выполняется сглаживание исходного ряда (простое скользящее среднее - материал уже был):

$$\hat{T}_k = \frac{Y_{k-2}}{8} + \frac{Y_{k-1}}{4} + \frac{Y_k}{4} + \frac{Y_{k+1}}{4} + \frac{Y_{k+2}}{8}$$

Результат: предварительный ряд значений тренда.



Детрендирование-

удаление полученного тренда из исходного ряда:

$$Y_k = T_k + S_k + \varepsilon_k$$

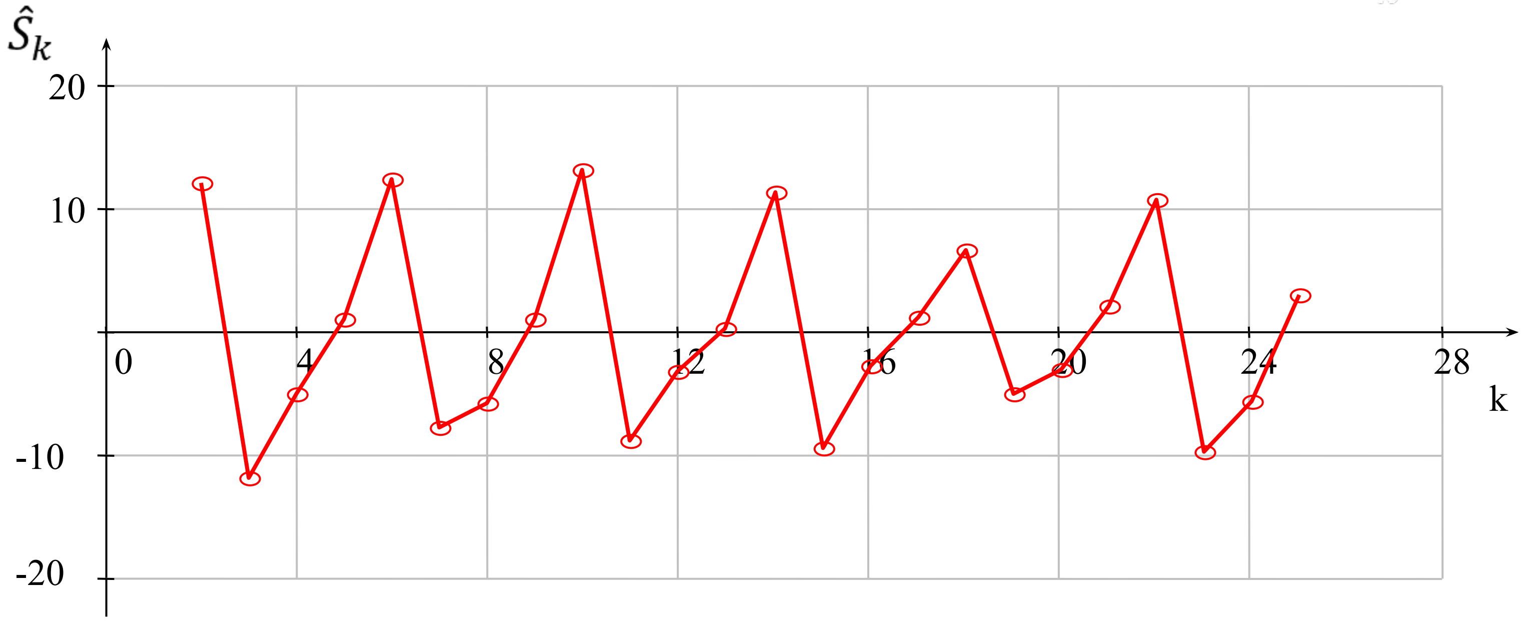
A: $\boxed{Y_k = \hat{T}_k (1+S_k)(1+\varepsilon_k)}$

M:

$$Y_k = \hat{T}_k (1+S_k)(1+\varepsilon_k)$$

$$\hat{S}_k = S_k(1+\varepsilon_k) = \frac{Y_k}{\hat{T}_k} - 1$$

Результат: зашумленный ряд значений сезонной компоненты.



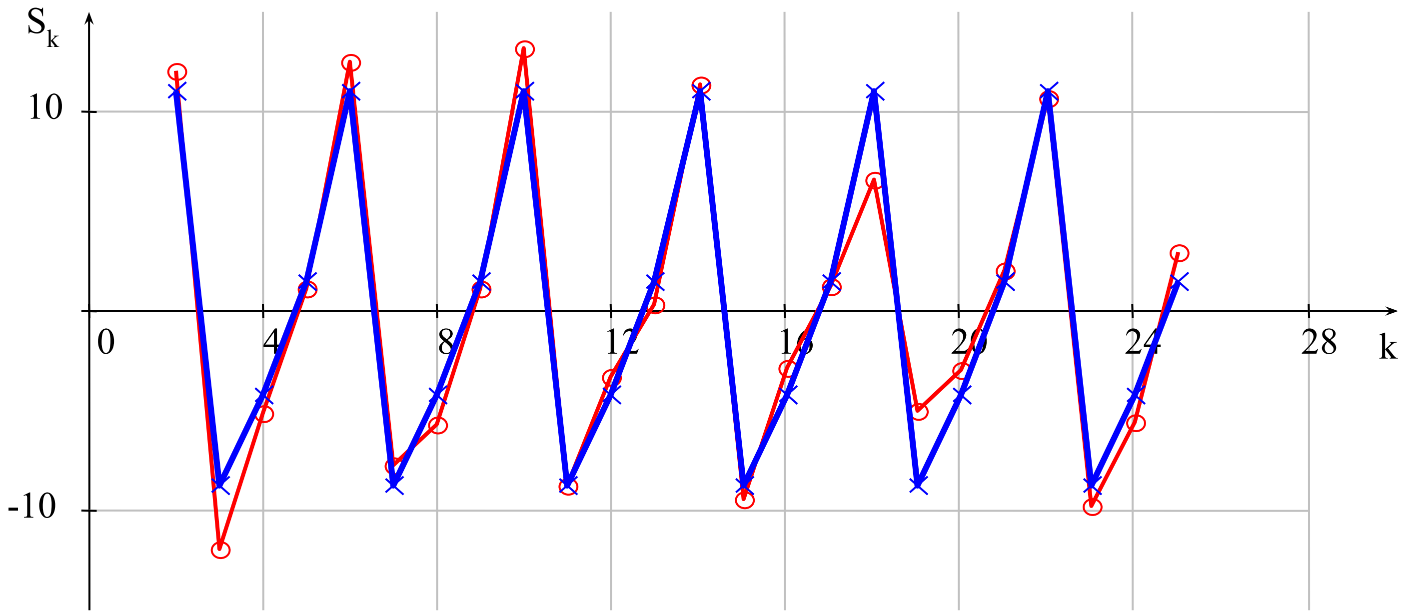
Выделение сезонной компоненты

Берутся средние значения сезонных колебаний для каждого квартала (месяца):

$$S^{I}=\frac{\hat{S}_1+\hat{S}_5+\ldots+\hat{S}_{n-3}}{n/4},\; S^{II}=\frac{\hat{S}_2+\hat{S}_6+\ldots+\hat{S}_{n-3}}{n/4},\ldots$$

Результат: таблица сезонных корректировок.

Квартал	Значение
I	-4.26
II	1.47
III	11.08
IV	-8.75



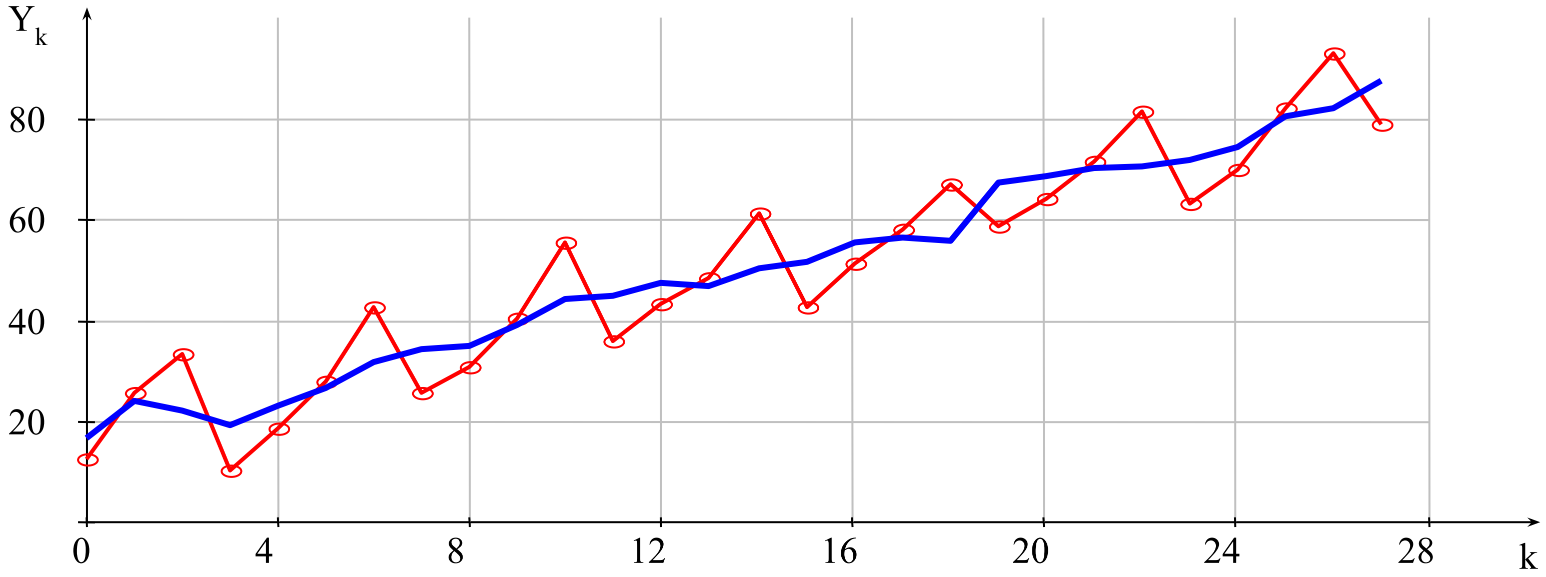
Десезонализация-

удаление сезонной компоненты из исходного ряда:

A: $$\hat{T}_k = T_k + \varepsilon_k = Y_k - S_k$$

M: $$\hat{T}_k = T_k \cdot (1 + \varepsilon_k) = \frac{Y_k}{S_k}$$

Результат: «зашумленный» ряд значений тренда.

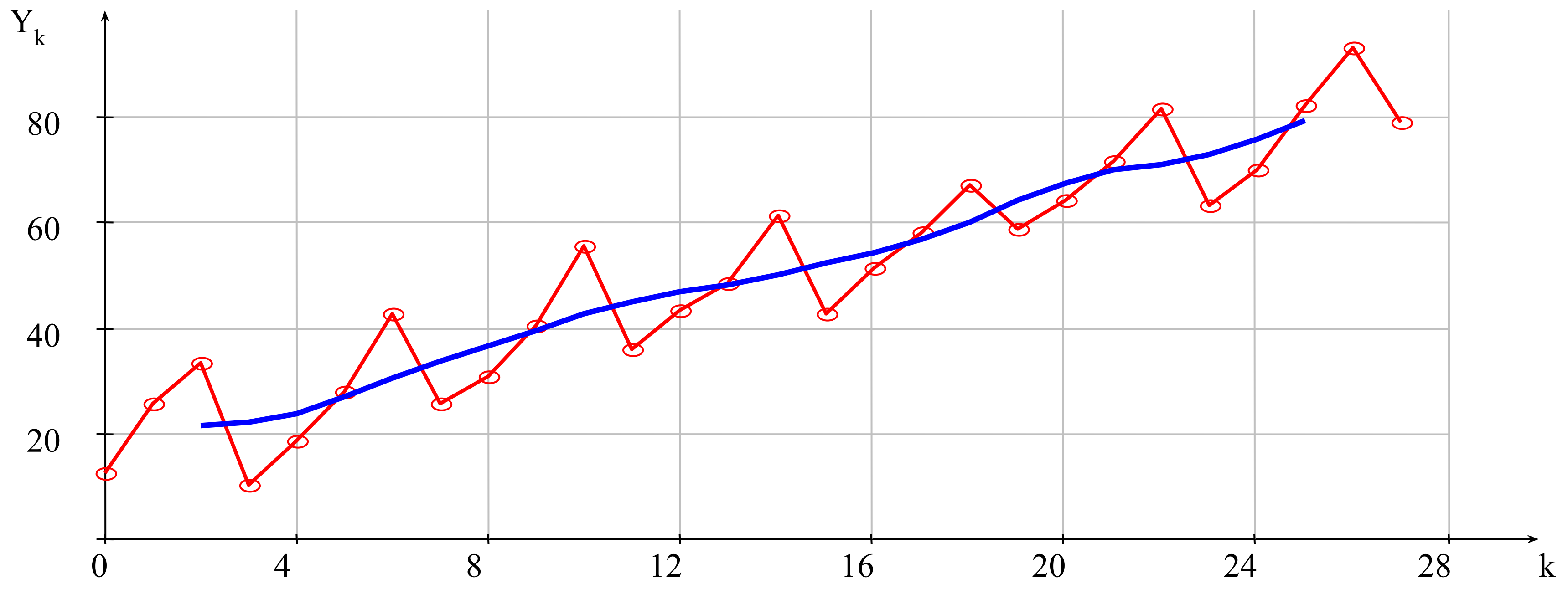


Повторное выделение тренда

Используется взвешенное сглаживание глубиной 1-3 значения:

$$T_k = \frac{\hat{T}_{k-2} + 2\hat{T}_{k-1} + 3\hat{T}_k + 2\hat{T}_{k+1} + \hat{T}_{k+1}}{9}$$

Результат: окончательный ряд значений тренда.



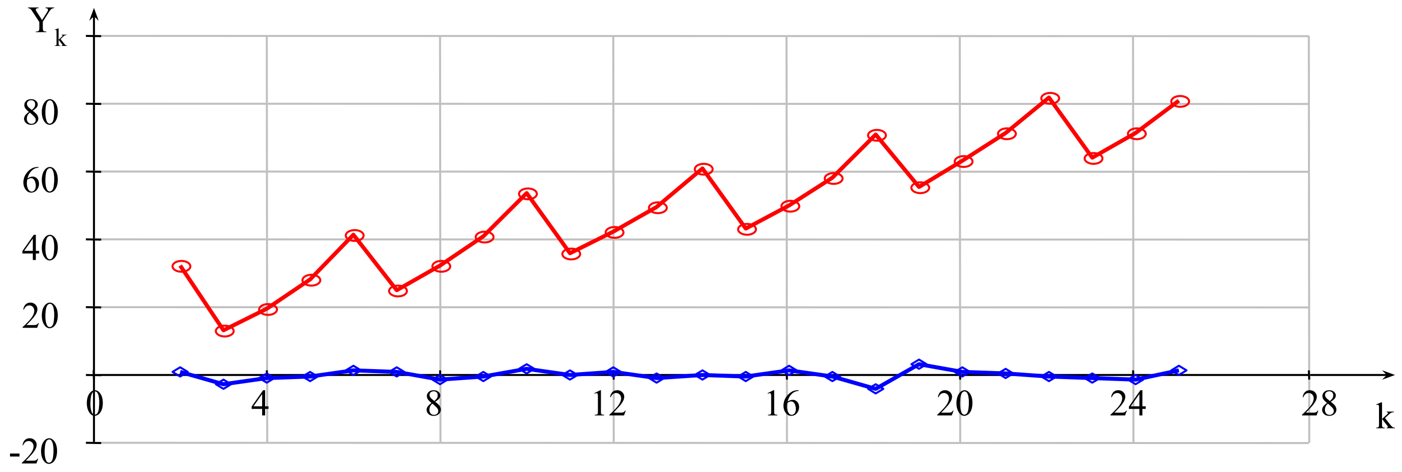
Выделение стохастической компоненты

Удаление сглаженных значений из исходного ряда:

А: $$\varepsilon_k = Y_k - (T_k + S_k)$$

М: $$\varepsilon_k = \frac{Y_k}{T_k(1 + S_k)} - 1$$

Результат: ряд случайных остатков и сглаженный ряд.



Метод Census II

Census II (1967г.) объединяет различные приемы и улучшения метода Census I. Наиболее известные варианты – **X-11**, опубликованный в 1978г., X-11-ARIMA, разработанный в Канаде в 1980г., **X-12-ARIMA**, наиболее популярный и продолжающий развиваться в настоящее время (www.census.gov.), среди которых:

1. Поправка на число рабочих дней.
2. Устранение случайных выбросов.
3. Последовательное уточнение компонент ряда.
4. Изменение шага 3 – для выделения сезонной компоненты используется взвешенное сглаживание, как для тренда на шаге 5.
5. Построение ARIMA-модели ряда для прогнозирования.
6. Вычисление дополнительных характеристик помимо самих компонент.

X-12-ARIMA реализован в виде бесплатной программы Census Bureau (www.census.gov).

Метод итерационной параметрической декомпозиции

Результат:

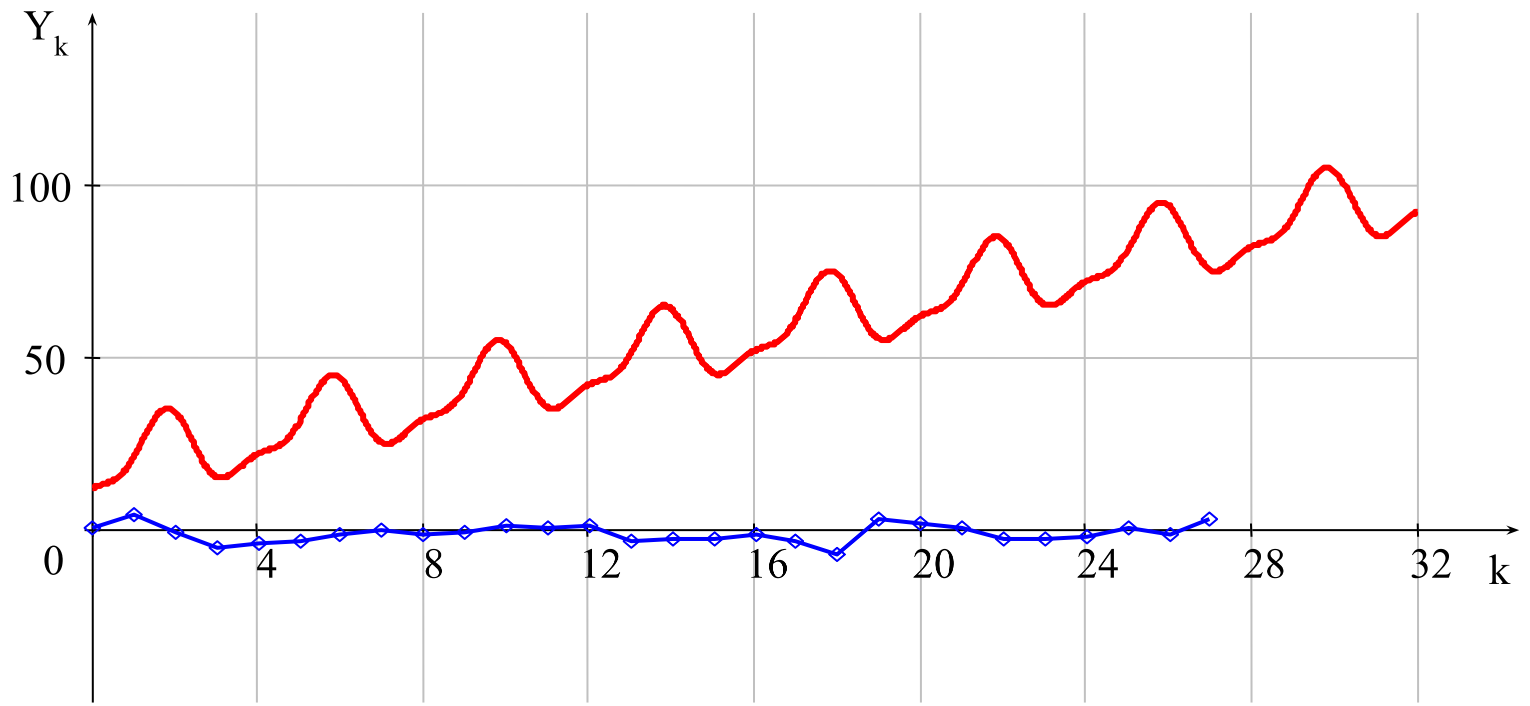
математические модели тренда и сезонной компоненты, ряд значений стохастической компоненты (достоинство – работа на относительно малых выборках, возможность применения AR-MA моделей и ГА).

Этапы:

1. Построение первичной модели тренда (МНК или генетический алгоритм (ГА).
2. Детрендирование.
3. Построение модели сезонной компоненты (МНК+ARMA модели).
4. Десезонализация.
5. Построение модели тренда (МНК или генетический алгоритм).
6. Выделение стохастической компоненты

Метод параметрической декомпозиции

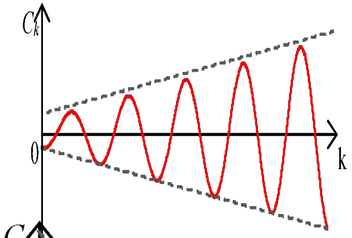
- Тренд (**идентификация** с помощью МНК): $\boldsymbol{T_k = 17 + 2,5\,k}$
- Сезонная компонента: $$S_k = 10\sin\left(\frac{\pi}{2}k - 1\right) + 4\sin(\pi k + 2)$$



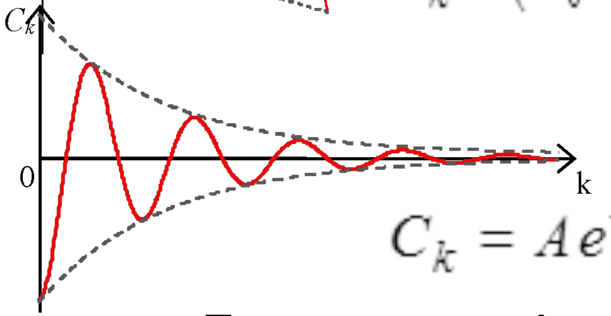
Эволюция компонент моделей

Для реальной экономической практики актуальна идентификация моделей на относительно коротких выборках – для мониторинга эволюции тренда и колебательной компоненты. Последняя чаще, чем тренд, демонстрируют эволюцию: в первую очередь – амплитуд, реже – фаз и еще реже – частот.

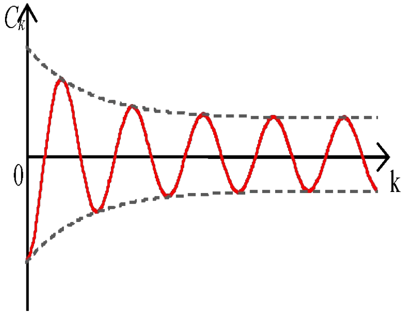
Некоторые примеры законов изменения амплитуд колебательных компонент:



$$C_k = \left(A_0 + A_1 k\Delta\right)\sin\left(\omega k\Delta + \psi\right)$$



$$C_k = A e^{-\alpha k\Delta} \sin\left(\omega k\Delta + \psi\right)$$



$$C_k = \left(A_0 + A_1 e^{-\alpha k\Delta}\right)\sin\left(\omega k\Delta + \psi\right)$$

Для идентификации эволюции гармоник высокие результаты по точности дает применение обобщенных параметрических моделей авторегрессии - скользящего среднего (будут показаны далее).

Интерполяция, экстраполяция, аппроксимация

Интерполяция – метод восстановление тех значений **определяемой переменной *Y*** объекта, которые находятся «между» соседними известными дискретными значениями (ее узлами интерполяции).

1. За промежуточные значения можно принимать те известные, к которым узлы интерполяции находятся ближе.

2. Можно моделировать расстояние между соседними известными значениями линейной функцией и брать ее значение при нужном нам аргументе определяемой переменной внутри интервала из двух значений (см. формулы в линейной алгебре).

3. Можно через «*n*» точек строить полином (*n+1*) порядка (например, Лагранжа, Ньютона и т.п., использовать члены ряда Фурье для периодических траекторий) для ***Y*** и находить его значение при нужном значении определяющей переменной.

4. Можно разбивать интервал интерполяции на меньшие отрезки и строить для них соответствующие полиномы: кусочно-линейная, кусочно-квадратичная, кубическая сплайн-интерполяция, сохраняющая непрерывными и производные, и т.п.).

.

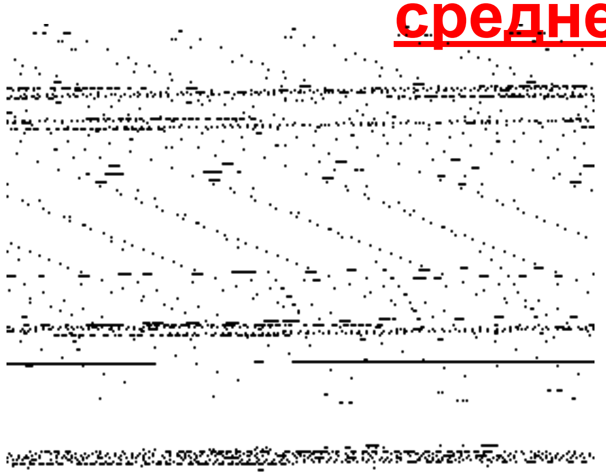
Экстраполяция - определение значений модели вне известных узлов интервала значений (больших или меньших) аргумента модели. Она практически не применима для прогнозирования.

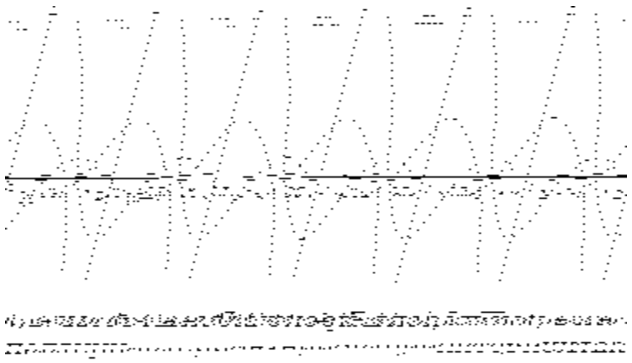
Реальные данные всегда содержат ошибки: систематические (постоянны или закономерно изменяются (если причины и характер известны, то их можно устранить введением поправок); случайные (не систематичны, повторение эксперимента может быть невозможно или дорого, тогда оправдана статистическая обработка); грубые (следует отбросить). Аппроксимация – «приближенное» выражение определяемой переменной в виде более простых моделей, которые должны «сгладить» реальные данные, уменьшить влияние случайных ошибок. Заменяют функцию $f(x)$, представляемую выборкой, другой более простой функцией $\varphi(x)$, которую можно вычислить при любом значении аргумента (и вне интервала), т.е. можно осуществлять и экстраполяцию, и прогнозирование. Аппроксимация может быть непрерывной (равномерной), если обеспечить условие $|f(x) - \varphi(x)| < \varepsilon,\quad a \le x \le b,$ что выполнить на трудно.

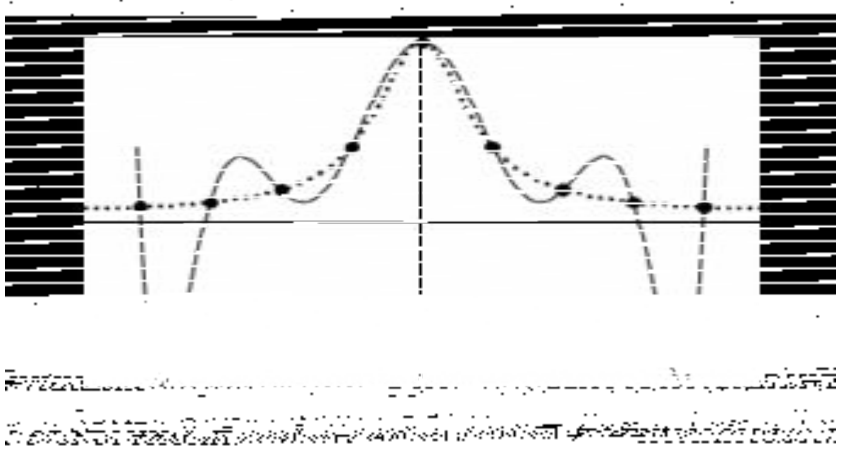
Среднеквадратическая аппроксимация (приближение) проще, минимизирует (сумму или интеграл) на параметрах модели $\varphi(x)$ (МНК):

$$\Rightarrow \min\left[\int_a^b \{f(x) - \varphi(x)\}^2 dx\right].$$

Графическая иллюстрация интерполяции, экстраполяции и среднеквадратического приближения (аппроксимации)







На рис. 4.1 показано сравнение интерполяционного полинома 6–й степени и среднеквадратического приближения функции на тех же точках внутри диапазона анализа. Погрешность интерполяции вне диапазона анализа (рис. 4.2) увеличивается существенно. Рис. 4.4 показывает несоответствие между аналитической функцией и ее интерполяцией на 9 узлах интерполяции при точных значениях используемых аналитических функций. В экономической практике интерполяция может собрать, (накопить) все ошибки, которые могут присутствовать в узлах интерполяционного полинома в отличие от метода приближения («сглаживания») ошибок.

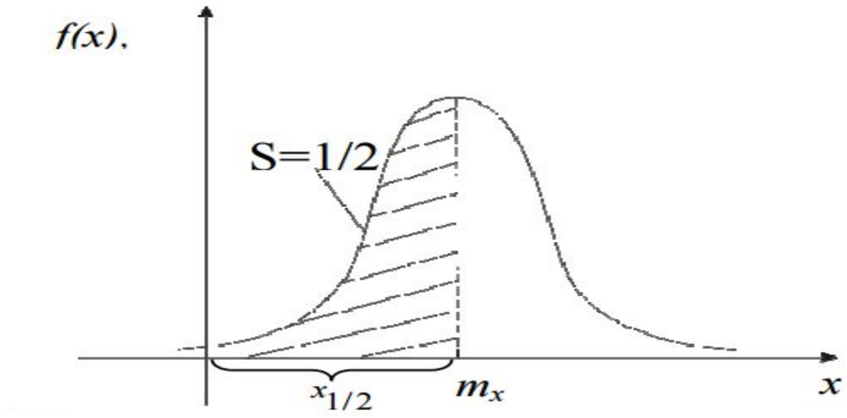
Квантили распределения, как характеристика формы
распределения, и возможность оценки ее параметров

x_p – квантиль уровня (вероятности) p определяется как корень уравнения $\boldsymbol{F(x_p) = p}$ или, сответственно, $\boldsymbol{P(X < x_p) = p}$.

- Через квантили может быть реализована робастная (устойчивая к тяжелым хвостам распределения – выбросам в виде очень малых и очень больших значений выборки) статистика: например, рассчитываться устойчивые к вариации законов распределения стохастической компоненты оценки матожидания, дисперсия, эксцесс и др.

Квантили:

$x_{1/2} = \mu e$; (медиана) квартили; децили; процентили.



Квартили

<u>Квартили</u> - $x_{1/4}, x_{2/4}, x_{3/4}$; <u>Децили</u> - $x_{1/10}, x_{2/10}, \ldots, x_{9/10}$;

<u>Процентили -</u> $x_{1/100}, x_{2/100}, \ldots, x_{99/100}$.

Квантили широко используется <u>в робастной статистике</u>:

178; 187; 1655, 175; 178 $\Rightarrow \boxed{m_x = 474{,}6}$. Наблюдение 1655 – <u>выброс (промах)</u>.

175; 178; 178; 187; 1655 $\Rightarrow \boxed{\mu e = 178}$

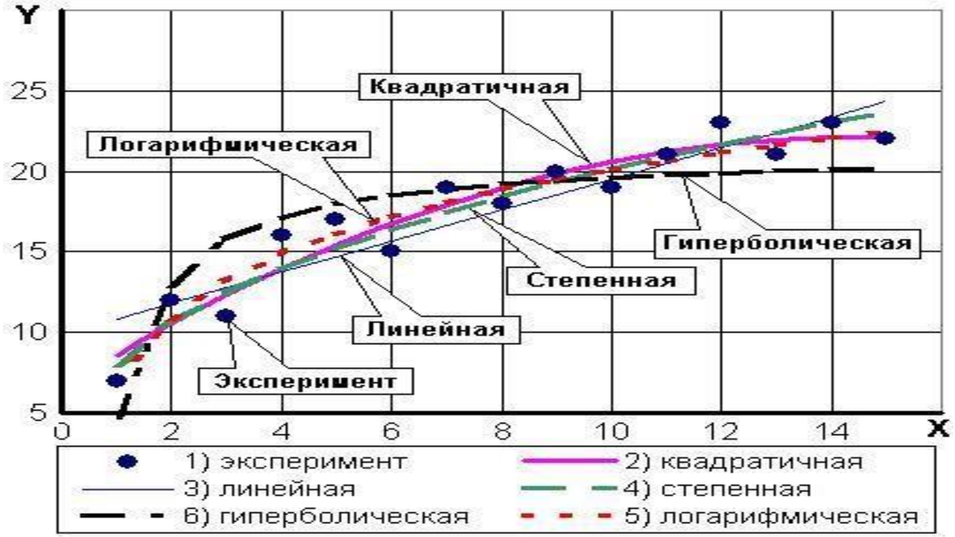
Сравнение и обоснование выбора модели трендов на выборке

Для обоснованного выбора модели могут быть необходимы дополнительные исследования: например, оценка точности и **значимости модели в целом**, а **также значимости её коэффициентов регрессии**.

Чтобы модель адекватно и точно описывала экономическую зависимость, необходимо выполнение дополнительных требований к объему, способу формирования выборки, в том числе и учета истории предыдущих исследований. Практически всегда более простые модели имеют преимущество по вычислительной устойчивости идентификации.

Выбор модели по значению моделей ряда :

Квадратическая	R^2	0,9202
Линейная		0,8464
Степенная		0,9170
Логарифмическая		0,9319
Гиперболическая		0,7742



О моделировании в случаях, когда размерности переменных существенно различаются

- Целесообразно использовать при больших различиях определяющей и определяемой переменных функции

$$LgY_k = aX_k, \qquad LgY_k = a_0 + a_1 X_k .$$

- В случае очень больших значений определяющей переменной по сравнению с значениями определяемой переменной рекомендуют полулогарифмические использовать функции:

$$Y_k = a\, LgX_k, \qquad Y_k = a_0 + a_1\, LgX_k .$$

Линеаризация модели по переменным

Квадратичный полином:

(применяют для моделей с одной точкой экстремума)

$$Y = ax^2 + bx + c + \varepsilon,$$

линеаризация

$$z = x^2 \Rightarrow Y = az + bx + c + \varepsilon \quad \Rightarrow$$

$$\text{МНК} \Rightarrow \min_{a,b,c} Q(a,\ b,\ c)$$

МНК, нормальная СЛАУ третьего порядка, для ее решения – формулы Крамера

Гиперболический полином: $Y = a + \dfrac{b}{x} + \varepsilon,$ линеаризация $z = \dfrac{b}{x} \Rightarrow Y = a + bz + \varepsilon,$

нормальная СЛАУ первого порядка, МНК

Эта модель известна, как **кривая Филлипса**. При отрицательном знаке у параметра ***b*** она трансформируется в **кривую Энгеля**: модель взаимосвязи расходов на товары длительного пользования и общих сумм расходов (доходов): доля расходов на питание уменьшается, тогда доля расходов на непродовольственные товары будет расти. К ее обобщению вернемся позже.

Моделирование неслучайной компоненты *D* обратной функцией

При аддитивной стохастической компоненте можно предложить два преобразования

$$Y = A + \frac{B}{x} + \varepsilon = \frac{Ax+B}{x} + \varepsilon x\,, \qquad \frac{1}{x} = z \rightarrow Y = A + Bz + \varepsilon\ .$$

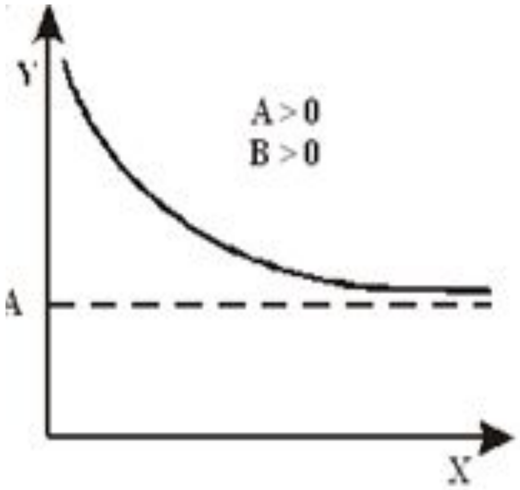
После данного линеаризирующего преобразования модель является линейной по параметрам. Стохастическая компонента –только в одном случае гомоскедастична и можно применять МНК.

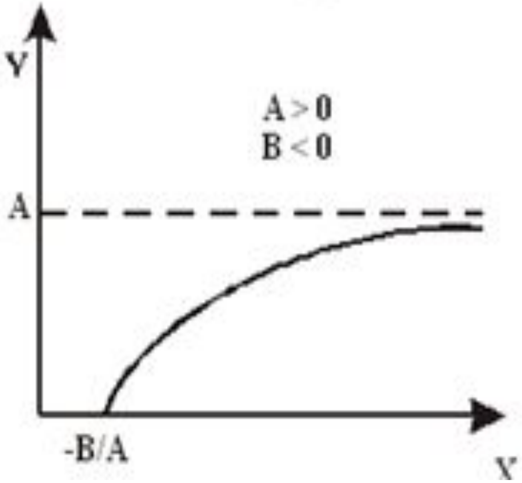
При еще одной возможной структуре - пропорционально-мультипликативной стохастической компоненте - будем иметь

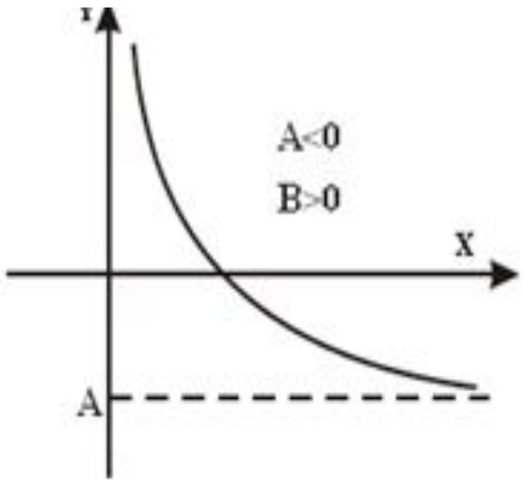
$$Y = \left(A + \frac{B}{x}\right)(1+\varepsilon) = A + \frac{B}{x} + A\,\varepsilon + \frac{B}{x}\,\varepsilon\,.$$

Здесь стохастическая компонента $\varepsilon = A\,\varepsilon + \frac{B}{x}$ является гетероскедастичной, нужно использовать специальные (например, указанные выше, методы сглаживания).

Обобщим моделирование неслучайной компоненты обратной функцией







Зависимость между объемом выпуска X и средними фиксированными издержками Y.

Зависимость между доходом X и спросом на блага Y (например, на товары первой необходимости либо на товары относительной роскоши). При этом в модели Торнквиста B/A - минимально необходимый уровень дохода.

Кривая Филлипса, определяет зависимость между уровнем безработицы X и процентным изменением заработной платы Y. Точка пересечения с осью OX - естественный уровень безработицы (5-6%).

Моделирование неслучайной компоненты обобщенной обратной функцией

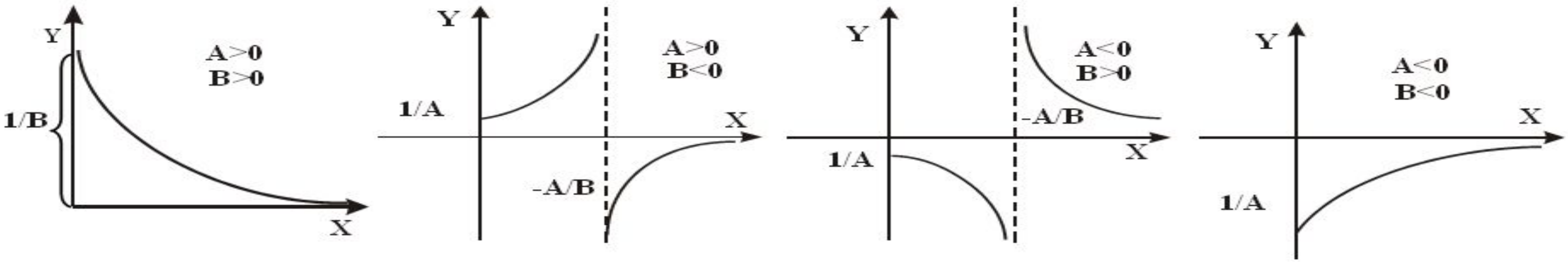
При аддитивной структуре стохастической компоненты будем иметь

$$Y = \frac{1}{A+Bx} + \varepsilon \rightarrow z = \frac{1}{Y} = A + Bx + (A + Bx)\,\varepsilon .$$

Обычно предлагают использовать линеаризирующее преобразование и применять МНК к выражению **$Z=A + B\,x+\varepsilon$**, что неверно, т.к. означает по сути принятие не канонической (а просто «удобной») структуры $Y = \frac{1}{A+Bx+\varepsilon}$.

При пропорционально-мультипликативном вхождении стохастической компоненты $Y = \frac{1}{A+Bx}(1+\varepsilon) = \frac{1}{A+Bx} + \frac{\varepsilon}{A+Bx}$,

получим гетероскедастическую стохастическую компоненту (нужно применять методы ее компенсации, указанные выше).



Часто моделируют неслучайную компоненту дробно-рациональными функциями

$$D = A_1 t + A_0 t + \frac{A_3}{t},$$

$$D = \frac{1}{A_2 t^2 + A_1 t + A_0},$$

$$D = \frac{A_0}{A_1 + t} + A_2$$

- в случае «пространственной» динамики моделирует спрос на определенный вид товаров или услуг от уровня доходов потребителей: **функция Торнквиста на товары относительной роскоши** (на товары второй необходимости), например, на дорогие продукты питания;

$$D = \frac{A_2 t^2 + A_1 t}{t^2 + A_0}$$

- **функция Торнквиста на малоценные товары;**

$$D = \frac{A_0 t}{t + A_1}$$

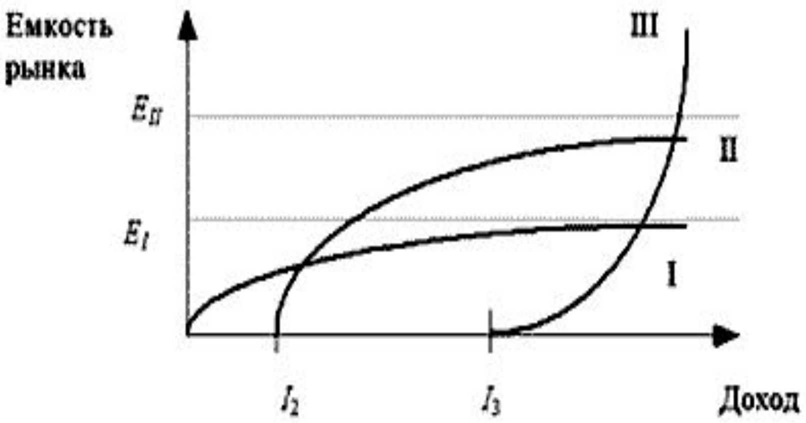
- **функция Торнквиста на товары первой необходимости** (например, на основные продукты питания);

$$D = \frac{A_0 t^2 + A_1 t}{t + A_2}$$

- **функция Торнквиста на предметы роскоши.**

Все функции Торнквиста являются упрощенными моделями спроса.

Вид моделей Торквиста и другие модели спроса



Для некоторых **товаров длительного пользования** используют и **логарифмические модели**, где *P*- цена единицы товара или услуги, *Q* – среднедушевой доход:

$$\ln D = a_0 + a_1 \ln Q + a_2 \ln P, \qquad D = a_0 + a_1 \ln P + a_2 \ln Q + a_3 \ln P \ln Q .$$

Отметим, что данные модели спроса **линейны** относительно параметров, но в общем случае **нелинейны** относительно переменных (факторов).

Используют и **полиномиальные модели спроса** (П.Л. Чебышева и др.), запись которых обобщим следующим выражением: $D = \sum_i a_i \varphi_i(t)$.

- **Товар Гиффена** — товар, потребление которого (при прочих равных условиях) увеличивается при повышении цены (то есть, эффект замещения от изменения цены перевешивается действием эффекта дохода). При соблюдении прочих равных условий (особенно, стабильного уровня дохода) потребление товаров Гиффена отражает положительный наклон кривой спроса.

Для большинства товаров при повышении цены снижается потребление: при повышении цен на мясо население покупает меньше мяса, заменяя его, например, рыбой, грибами и т. д.

- **У товара Гиффена наоборот** — при повышении цен на картофель люди начинают покупать больше картофеля, но меньше, например, мяса.

Все товары Гиффена — малоценные товары, которые занимают в потребительском бюджете значительное место и для которых отсутствует равнозначный товар-заменитель. Так, например, товарами Гиффена в России являются соль, чай, табак и другие.

«**Парадокс Гиффена**»: при повышении цен на определённые виды их потребление повышается за счёт экономии на других товарах.

Подходы и методы прогнозирования емкости рынка:

Подходы к прогнозированию емкости рынка	Методы прогнозирования емкости рынка в рамках соответствующего подхода
1. Эвристический подход	1.1. Методы средней оценки по индивидуальным оценкам экспертов
1. Эвристический подход	1.2. Метод оптимистических, пессимистических и вероятностных мнений экспертов
1. Эвристический подход	1.3. Метод комиссии
1. Эвристический подход	1.4. Метод Дельфи
1. Эвристический подход	1.5. Метод сводного индекса готовности приобретения продукции целевыми потребителями
2. Экономико - математический подход	2.1. Трендовые модели
2. Экономико - математический подход	2.2. Через кривые жизненного цикла
2. Экономико - математический подход	2.3. Факторные модели: 2.3.1.Однофакторные модели: - через коэффициенты эластичности - через кривые Энгеля и Торнквиста 2.3.2. Многофакторные модели

Моделирование неслучайной компоненты степенной функцией (например, при изучении уровня оплаты труда от его производительности)

$$T = Ax^{\beta}$$

$$Y = Ax^{\beta} \cdot e^{\varepsilon} \quad (1);$$
$$Y = Ax^{\beta} \cdot \varepsilon \quad (2);$$
$$Y = Ax^{\beta} + \varepsilon \quad (3).$$

Для (1) после логарифмирования будем иметь:

$$lnY = lnA + \beta lnx + \varepsilon\,; \qquad Y^* = A^* + \beta x^* + \varepsilon\,;$$
$$Y^* = lny; \quad A^* = \ln A; \quad x^* = \ln x.$$

Одно из условий Гаусса-Маркова для оптимальности МНК-оценок параметров получаемой парной линейной регрессии является нормальный закон распределения стохастической компоненты ε , следовательно, e^{ε} должна иметь логнормальное распределение. Данное условие является существенным ограничением для возможности применения операции логарифмирования.

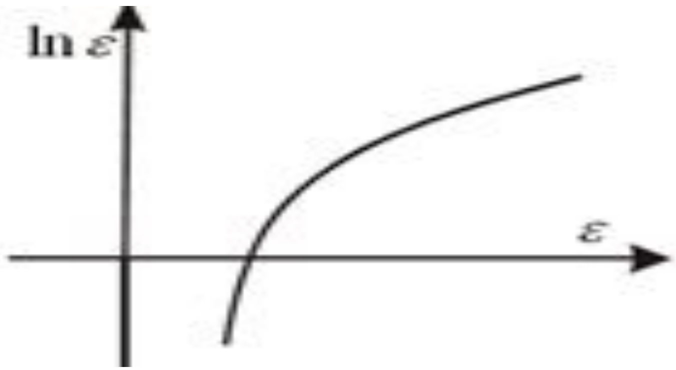
Для модели (2) логарифмирование дает:

$$\ln Y = \ln A + \beta \ln x + \ln \varepsilon.$$

При этом значения ε должно быть неотрицательны, т.е. также имеется существенное ограничение для возможности применения операции логарифмирования.

Для модели (3) логарифмирование дает:

$$\ln Y = \ln\left(Ax^{\beta} + \varepsilon\right).$$



В этом случае прием логарифмирования для линеаризации модели и сведения задачи к парной линейной регрессии вообще не работоспособен (не разделяет параметры в скобках). Необходимо обращение к другим методам (например, к АР-СС моделям - покажем далее).

Выявление детерминированных компонент D_k ряда динамики Y_k «сглаживанием» (уменьшением стохастической компоненты)

$$\boxed{Y_k = D_k + \varepsilon_k} \quad \begin{cases} D_k = T_k + S_k + C_k & \text{– временные ряды} \\ D_k = T_k & \text{– пространственные ряды} \end{cases}$$

Сглаживание

Аналитическое

МНК

СЛАУ

АР-СС

НМНК

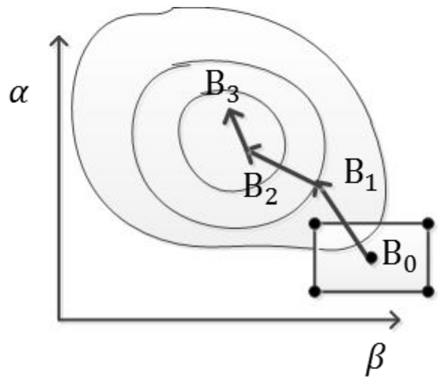
Градиентные методы (Гаусса, симплекса, покоординатного спуска)

Алгоритмическое

текущее (скользящее), простое сглаживание

взвешенное (чаще экспоненциальное) сглаживание

Алгоритмические методы поиска экстремума функции потерь для двух параметрической нелинейной по параметрам модели



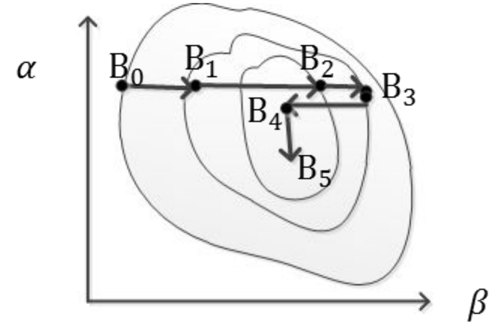
Градиент $\varphi(..)$ – целевая функция (потерь)

$$grad\,\varphi(\alpha,\beta)=\frac{\partial\varphi}{\partial\alpha}\cdot\vec{i}+\frac{\partial\varphi}{\partial\beta}\cdot\vec{k}$$

$$\frac{\partial\varphi}{\partial\alpha}=\frac{\Delta\varphi(\alpha+\Delta\alpha,\beta)}{\Delta\alpha}\;;\quad\frac{\partial\varphi}{\partial\beta}=\cdots$$

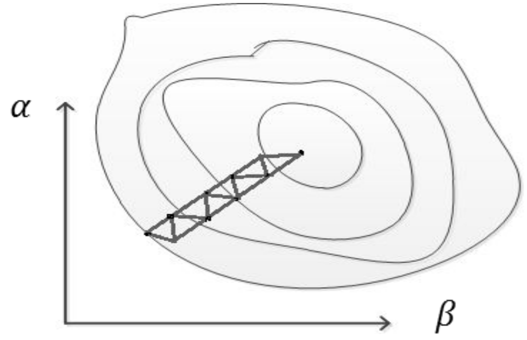
, $\vec{i},\vec{j}$ - единичные ортогональные векторы

Движение по координатам



Одна координата «заморожена», движемся по другой до уменьшения (увеличения) функции потерь. Возвращаемся, движемся по другой координате (первая- «заморожена»).

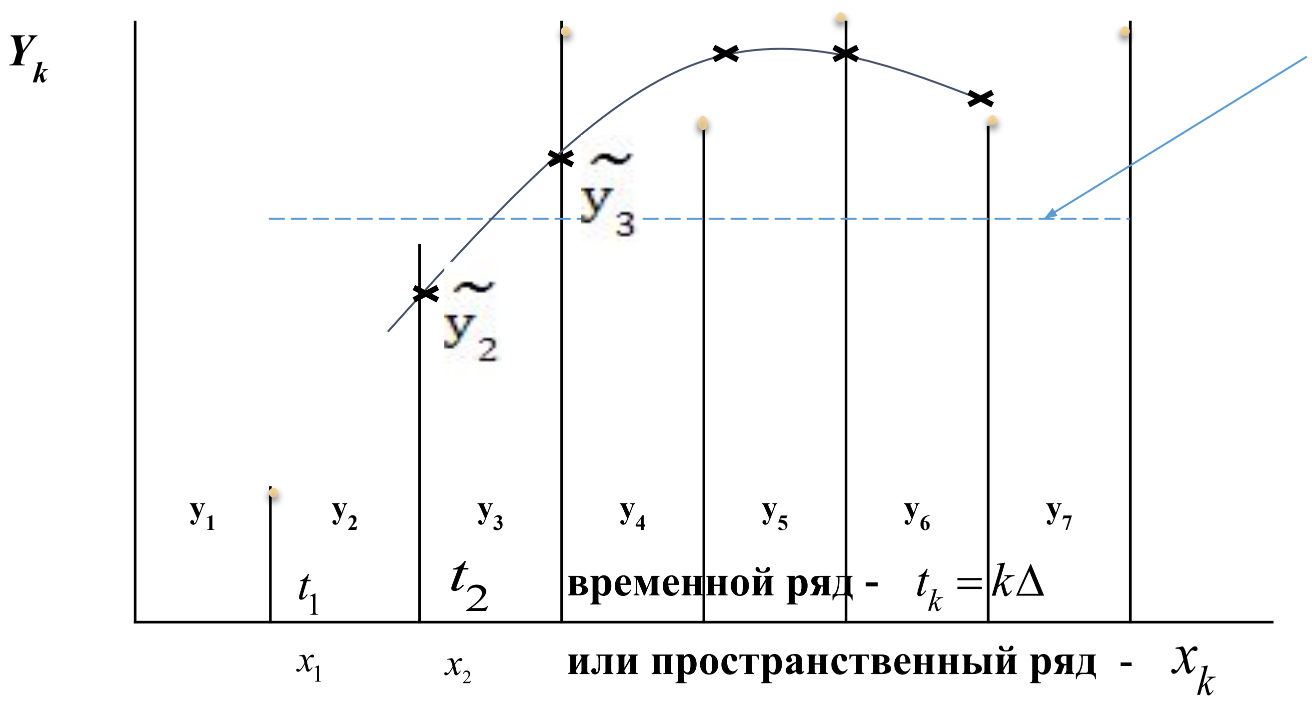
Метод симплекса («равностороннего треугольника»)



Поворот вокруг стороны правильного треугольника, противолежащей вершине с минимальным (максимальным) значением функции потерь.

Не зная аналитического вида функции потерь, можем вычислять ее значения при наборах параметров.

Возможности алгоритмического сглаживания: простого и текущего



$$\bar{Y}_k = \frac{1}{n}\sum_{k=1}^{n} y_k$$

Простое сглаживание (среднее значение)

n- объем выборки

Текущее простое сглаживание:

$$\boxed{\widetilde{y_k} = \frac{\sum_{k-1}^{k+1} y_i}{3}}$$

k=2, i=1, 2, 3
k=3, i=2, 3, 4
……………

~ - символ сглаживания

Равные веса при суммировании:

$$\widetilde{y_2} = \frac{y_1 + y_2 + y_3}{3} = \frac{y_1}{3} + \frac{y_2}{3} + \frac{y_3}{3}$$

Оценка $\bar{X}$ среднего или ($\bar{Y}_k$ в обозначениях предыдущего слайда) несмещенная, а ее дисперсия уменьшается при усреднении в n раз:

$$M[\bar{X}] = M\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n} n M_x = m_x, \qquad \mathrm{D}[\bar{X}] = D\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{x^2} n D_x = \frac{D_x}{n}$$

Если $x \sim N(m_x, \sigma_x)$, $\Rightarrow$ $\bar{X}$ – эффективная оценка.

Оценка дисперсии: $D_x^* = S^2 = M\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2\right]$. Доказано, что $S^2 = M\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2\right] = \frac{n-1}{n} D_x n$, т.е. данный алгоритм дает смещенную оценку дисперсии: $\Delta = \frac{D_x}{n}$.

Исправленная (несмещенная) оценка дисперсии (поправка Шеппарда) будет следующей $\frac{nD_x^*}{n-1} = S^2_{\text{испр.}} = \frac{1}{n-1}\sum_{n-1}^{n}(x_i - \bar{X})^2$.

На практике исправленной оценкой дисперсии пользуются при $n < 30$.

Моделирование тренда алгоритмическим простым симметричным скользящим сглаживанием

Простые скользящие средние (simple moving average, SMA) – значения ряда заменяются средними по соседним наблюдениям.

$$\tilde{Y}_k = \frac{1}{m}\sum_{i=-r}^{r} Y_{k+i}, \quad m = 2r+1$$

$\tilde{Y}_k$ - сглаженные значения

r - глубина сглаживания (насколько далеко берутся значения).

m - интервал сглаживания (сколько всего берется значений для сглаживания).

Примеры:

$$\mathrm{SMA1:} \quad r=1 \quad m=3 \quad \tilde{Y}_k = \frac{Y_{k-1}+Y_k+Y_{k+1}}{3}$$

$$\mathrm{SMA2:} \quad r=2 \quad m=5 \quad \tilde{Y}_k = \frac{Y_{k-2}+Y_{k-1}+Y_k+Y_{k+1}+Y_{k+2}}{5}$$

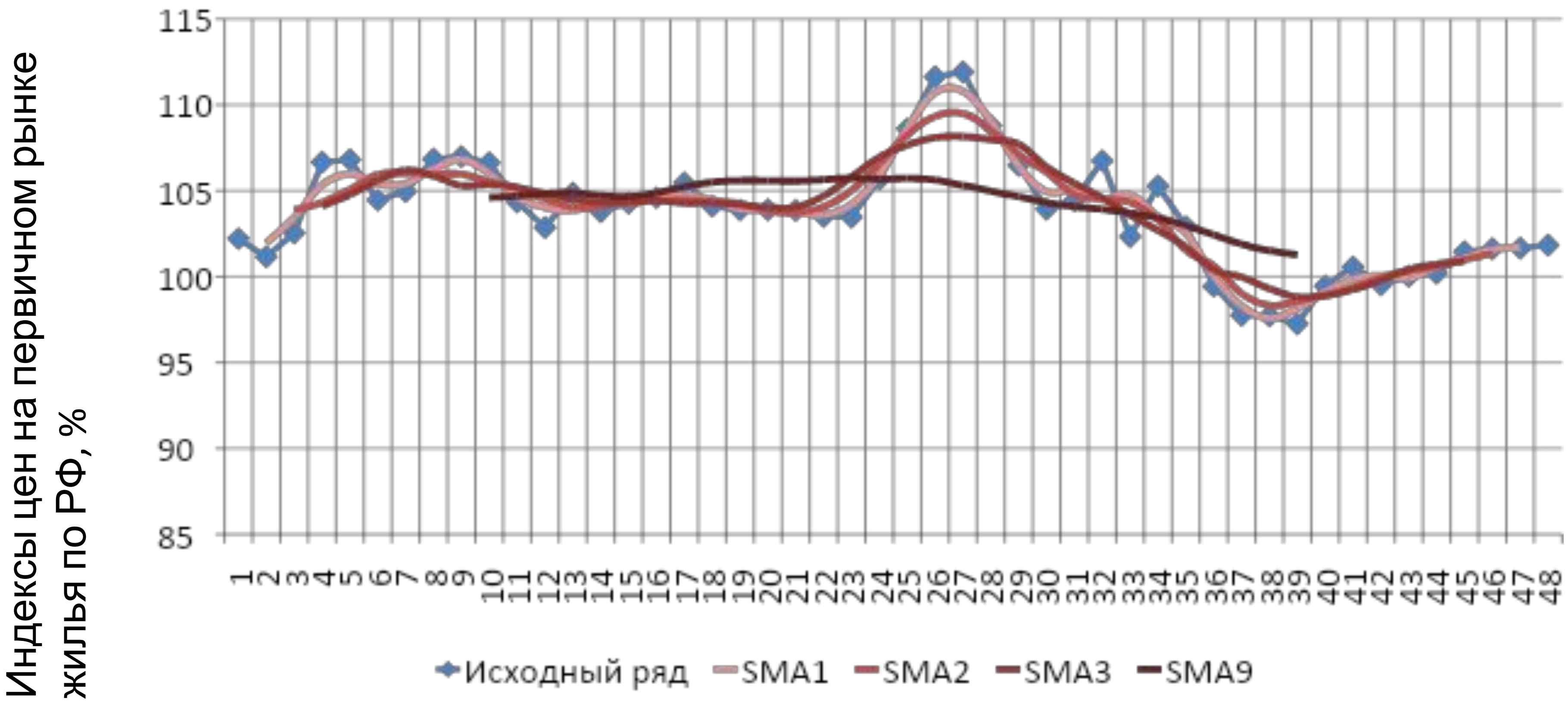
Пример для четного $m = 4$: пример годового сглаживания квартальных наблюдений

$$\tilde{Y}_k = \frac{\frac{Y_{k-2}}{2} + Y_{k-1} + Y_k + Y_{k+1} + \frac{Y_{k+2}}{2}}{4}$$

Пример простого (симметричного) сглаживания

Чем больше глубина сглаживания *r*, тем более гладким получается ряд.

Сглаженный ряд содержит на (*m*-1) значений меньше, чем исходный – теряются *r* значений по краям ряда.



$$\widetilde{y_k} = \frac{\sum_{k-2}^{k+2} y_i}{5}$$

k=3, i=1, 2, 3, 4, 5
k=4, i=2, 3, 4, 5, 6
……………………

$$\widetilde{y_3} = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5} = \frac{y_1}{5} + \frac{y_2}{5} + \frac{y_3}{5} + \frac{y_4}{5} + \frac{y_5}{5}$$

$$\widetilde{y_4} = \frac{y_2 + y_3 + y_4 + y_5 + y_6}{5}$$

равные веса

Чем больше (шире) интервал сглаживания, тем в большей мере поглощаются колебания, а тенденция развития носит более плавный, сглаженный характер. Чем сильнее колебания, тем шире должен быть интервал сглаживания.

Интервал сглаживания скользит по ряду динамики с шагом, равным единице.

Первое и последнее значения ряда не используются. У них нулевые веса.

Экспоненциальное сглаживание

$\boxed{(1) S_k = \alpha y_k + \beta S_{k-1}}$ - рекуррентная формула Брауна

где $s_k = \widetilde{y_k}$ - значение экспоненциальной средней (сглаженное $\boldsymbol{y_k}$ значение исходного ряда динамики в момент наблюдения *«k»* , *α = const* – параметр экспоненциального сглаживания, *0 < α <1, β=1 - α,* y_k - фактическое значение (уровень) исходного ряда динамики в момент наблюдения *«k».*

Принимают, что $S_{-1} = 0$ - значение «экспоненциальной средней» в момент наблюдения *«k-1»:*

$$S_{k-1} = \tilde{y}_{k-1} \cdot$$

Для устранения сезонных колебаний на практике часто приходится использовать **скользящее значение среднее с длиной интервала сглаживания**, равной 4 или 12 (ежеквартальные или ежемесячные наблюдения). При этом не будет выполняться условие нечетности числа наблюдений. Можно брать первое и последнее наблюдение с половинными весами:

$$\widetilde{y_k} = \frac{\frac{1}{2}y_{k-2} + y_{k-1} + y_k + y_{k+1} + \frac{1}{2}y_{k+2}}{4} \quad \text{e=4}$$

$$\widetilde{y_k} = \frac{\frac{1}{2}y_{k-6} + y_{k-5} + \cdots + y_k + \cdots + y_{k+5} + \frac{1}{2}y_{k+6}}{12} \quad \text{e=12}$$

Если для тренда характерно динамическое **нелинейное развитие**, то простая скользящая средняя может привести к существенным искажениям. Если траектория имеет **изгибы и мелкие волны**, то целесообразно использовать взвешенную скользящую среднюю:

$$\widetilde{y_k} = \frac{\sum_{i=k-p}^{k+p} \omega_i y_i}{\sum_{i=k-p}^{k+p} \omega_i}$$

ω_i *– весовые коэффициенты*

обычно $\sum_i \omega_i = 1$

Из взвешенных скользящих средних наиболее популярно экспоненциальное.

Представим (1) в виде (раскроем β)

$$(2)\ S_k = \alpha y_k + (1-\alpha)S_{k-1} = S_{k-1} + \alpha(y_k - S_{k-1})$$

т.е. S_k является суммой S_{k-1} и доли ***α*** от разности $(y_k - S_{k-1})$

Последовательно подставляя *«к»* в (1) получим

0

$$S_k = \alpha \sum_{i=0}^{n-1} \beta^i y_{k-i} + \beta^4\ S_{0,}$$

где ***n*** - объём выборки, $n \to \infty \Rightarrow \beta^n \to 0$

S_0 - некоторая величина, характеризующая начальные условия для первого изменения (*k*=1) формулы (1). Так как *β*<1, то при .

Тогда получим следующее взвешенное среднее:

$$S_k = \alpha \sum_{i=0}^{\infty} \beta^i y_{k-i} = \alpha y_k + \alpha(1-\alpha)y_{k-1} + \alpha(1-\alpha)^2 y_{k-2} + \cdots$$

Экспоненциальное сглаживание сглаживает весь ряд в целом (выступающие значения (выбросы) D_k и ε_k).

Оператор экспоненциального сглаживания можно применить к уже сглаженным значениям – тогда реализуем двойное экспоненциальное сглаживание.

«Деликатные» вопросы:

1. выбор S_0 (например, иногда рекомендуют использование среднего арифметического всех точек выборки);

2. выбор α . Обычно $\alpha\ (0{,}1 < \alpha < 0{,}3)$.

Если $\alpha \to 0$, то веса, по которым взвешиваются уровни ряда динамики, убывают медленно и прогнозирование идёт с учётом всех прошлых значений.

Если $\alpha \to 1$, то учитываются, в основном, последние значения.

Некоторые компьютерные программы имеют возможность автоматического изменения $\boldsymbol{\alpha}$: если R^2 мало или *MAPE* - оценка прогноза слишком велика.

<u>Компенсация гетероскедастичности (ведет к неэффективности оценок регресии)</u>

- **<u>Тестирование</u>** (не только визуальным наблюдением корреляционного поля) **<u>гетероскедастичности</u>**: тесты Бреуша-Пагана, Голдфелда-Квандта, ранговой корреляции Спирмена.
- **<u>Компенсация гетероскедастичности</u>** путем:

\- **логарифмирования данных** (неприменимо при отрицательных значениях ряда);

\- **перехода к безразмерным величинам** делением данных на некоторые известные величины той же размерности (неприменимо при нулевых значениях ряда), например, **выбора одного из** имеющихся **наблюдений** или **переходом к цепным индексам**

$$Z_{i+1} = \frac{Y_{i+1}}{Y_i},\ i = 1,\ 2, \ldots, n\ ;$$

\- **предположения** для дисперсии остатков некоторого закона изменения $D(\varepsilon_i) = \sigma^2 \phi_i$,

затем деление каждого члена регрессионного уравнения на ϕ_i , тогда остатки будут гомоскедастическими с дисперсией σ^2 .

\- **применения** обобщенного метода наименьших квадратов (ОМНК) или взвешенного МНК (требуются априорные знания об ошибках наблюдений, поэтому область практического применения ограничена).

\- **использования** робастной статистики.

Проблемы при выполнении приема логарифмирования при реализации линеаризации

- Степенную функцию (и ряд других) удобно прологарифмировать

$$Y = a\,x^b\,e^{\varepsilon} \Rightarrow \ln Y = \ln a + b\ln x + \varepsilon$$

Ограничения: определяющий фактор и стохастическая компонента должны быть > 0, а компонента, кроме этого, должна быть в диапазоне значений [0,1] и должна иметь логнормальный закон распределения для того, чтобы получить, после логарифмирования

e^{ε}, нормальный закон распределения для ε, выполнив тем условие Гаусса-Маркова, и оправдать применение МНК. При этом функция потерь будет формироваться

на невязках $\ln Y$ и $\ln x$, а не на невязках Y и x, т.е. она будет другой.

Другими будут и оценки параметров модели (они не могут быть отрицательными, уменьшая тем самым область применения рассматриваемой степенной функции).

Возникает и вопрос о том, как поступать при аддитивной (независимой) структуре вхождения стохастической компоненты (так же ограничение приема логарифмирования):

$Y = a\,x^b + \varepsilon \Rightarrow ?$ Позже покажем корректность использования моделей авторегрессии.

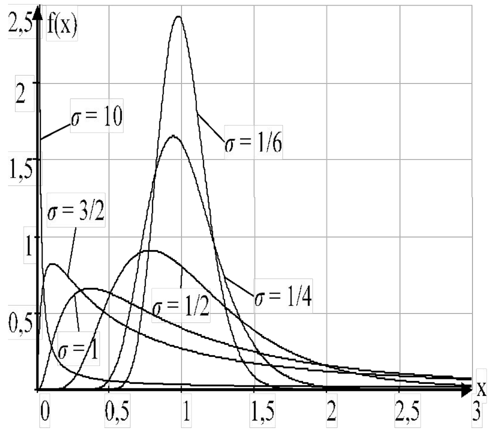
Роль логнормального распределения при мультипликативной структуре стохастической компоненты в модели

У многих моделей мультипликативная структура взаимодействия компонент, в частности стохастической, затрудняет реализацию моделирования и прогнозирования параметрическими моделями, требует специальных решений.

При мультипликативной стохастической компоненте μ она может принимать значения лишь в диапазоне от 0 до 1, что нарушает условие Гаусса-Маркова для МНК, не позволяя получать оптимальные оценки. Для «удобства», принимают ее закон распределения логнормальным (ее логарифм имеет нормальное распределение):

$$f(\mu) = \frac{\exp\left(-\frac{1}{2}\left[\frac{\ln\mu-\beta}{\sigma}\right]^2\right)}{\mu\sigma^2\sqrt{2\beta}},$$

$$\ln Y = \ln A + \alpha\cdot\ln K + \beta\cdot\ln L + \varepsilon,$$



с параметрами σ и β, но с ненулевым математическим ожиданием, что опять же нарушает условие Гаусса-Маркова.

После логарифмирования получим, что стохастическая компонента будет иметь уже нормальный закон распределения. Казалось бы оправдана МНК-идентификация парной линейной регрессии. Но при этом функция потерь будет минимальна, не на данных модели, а на значениях их логарифмах. Разных величин функции потерь обычно «стыдливо» не замечают.

О упрощениях при линеаризации: «дело не в том, люди зачастую не могут найти решение, а в том, что они не могут увидеть проблему….»

- МНК (для функций потерь) реализуется на выборках Y_k (и $\boldsymbol{X_k}$) , а классические (канонические) структуры имеют аддитивную: $Y_k = D_k(X_k) + \varepsilon_k$ и мультипликативную структуры $Y_k = D_k(X_k) \cdot \varepsilon_k$.

- Справедлива линеаризация модели $Y_k = a + \frac{b}{X_k} + \varepsilon_k$ заменой

$$Y_k = a + bZ_k + \varepsilon_k \quad Z_k = \frac{1}{X_k}$$

и модели $Y_k = a + b\ln X_k + \varepsilon_k \Rightarrow Y_k = a + bZ_k + \varepsilon_k \quad Z_k = \ln X_k$.

Однако некорректно выполнение операции логарифмирования для линеаризации следующих трех функций

$$Y_k = a\, b^{X_k} + \varepsilon_k \Rightarrow (\ln Y_k = \ln a + X_k \ln b) + \varepsilon_k\,, \quad Y_k = a X_k^{b} + \varepsilon_k \Rightarrow (\ln a + b \ln X_k) + \varepsilon_k\,,$$

$$Y_k = e^{a + bX_k} + \varepsilon_k \Rightarrow (\ln Y_k = \ln a + b \ln X_k) + \varepsilon_k\,,$$

потому, что будем иметь другие метрики и для других функций потерь, и другие ограничения на параметры моделей, используемых в этих функциях.

Модели роста и эволюции

- Логистическую тенденцию тренда (например, модели роста Гомпертца, Ферхульста и другие) упрощенно можно считать объединением трёх разных по типу тенденций: параболической с ускоряющимся ростом на первом этапе, линейной – на втором этапе и гиперболической с замедляющимся ростом – на третьем этапе.
- Однако весомее доводы в пользу рассмотрения всего цикла логистического развития как особого единого типа тенденции со сложными переменными свойствами, но с постоянным направлением изменений в сторону увеличения (или уменьшения) уровней, т.е. **она является едва ли** не основной моделью эволюции.

Особое внимание отведем технологическим инновациям и моделирования их логистическими функциями (в том числе возможна и алгеброй их взаимодействия, как при формировании мультитрендов).

Основной характеристикой процесса эволюции служит так называемый «технологический разрыв». Он характеризует различие в эффективности *P* новой и старой технологий, а также в объемах средств, требуемых для вложения в новую технологию, с целью достижения ею результативности, которую не имеет на сегодня старая технология.

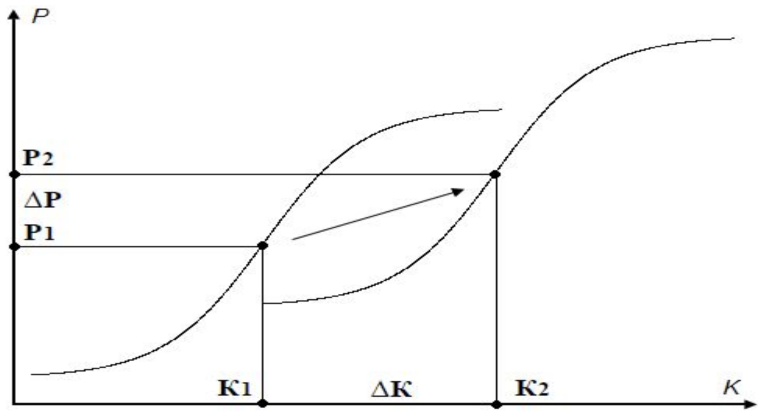
Определение технологического уклада

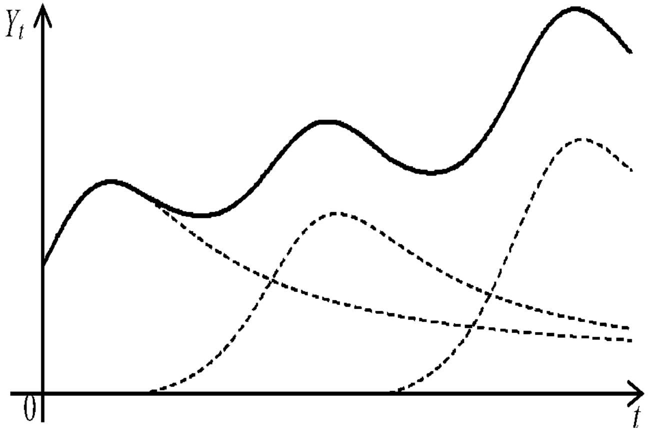
- **Технологические уклады** — группы технологических совокупностей, связанные друг с другом однотипными технологическими цепями и образующие воспроизводящие целостности.

- **Каждый уклад** представляет собой целостное и устойчивое образование, в рамках которого осуществляется замкнутый цикл, включающий добычу и получение первичных ресурсов, все стадии их переработки и выпуск набора конечных продуктов, удовлетворяющих соответствующему типу общественного потребления.

<u>Демонстрация технологического разрыва</u>

и объемов средств, требуемых для вложения в новую технологию





Обычно в экономике одновременно действуют несколько (как правило, два) технологических укладов с периодом жизни 100-150 лет (кривые на рисунках носят в значительной мере качественный и иллюстративный характер, справа – сумма кривых ЖЦП).

Зарождение нового технологического уклада по времени совпадает с началом падения эффективности доминирующего уклада, в результате суммарная траектория экономической эволюции испытывает колебания вокруг повышающегося тренда.

Диффузия инноваций

- Исследователи склоняются к тому, что именно **на периоды депрессий приходятся основные инновации** – технологические и организационные новшества.
- В условиях благоприятной конъюнктуры предприниматели предпочитают избегать чрезмерного риска, связанного с коренной перестройкой производства, пытаются ограничиться **рационализацией и усовершенствованием** существующих технологических процессов.
- В периоды депрессий, когда само существование огромного количества хозяйствующих единиц ставится под угрозу, предприниматели вынуждены рисковать, понимая, что незначительные усовершенствования не приведут к кардинальному улучшению ситуации. Через 10-15 лет после базисных нововведений начинается повышение экономической конъюнктуры, создаются благоприятные условия для дополняющих нововведений.
- Вторичные инновации, частичные усовершенствования доводят до совершенства то фундаментально новое, что возникло в фазе депрессии. Формируется новый технологический уклад, жизненный цикл которого составляет от 100 до 130 лет.
- Технологические уклады доминируют в экономике, последовательно сменяя друг друга, вызывая тем самым колебания траектории экономического развития (т.е. **одну из причин кризисов**).

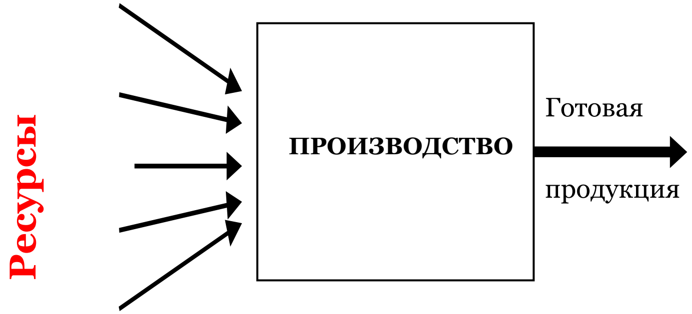
Диффузия инноваций вдоль подъемов циклов экономической активности Кондратьева

Циклы Кондратьева (живем и «учимся» в 5-ом). Известны и другие циклы: Китчина, Кузнеца, Джанглера, …

Модели производственных функций (ПФ) (известное количество ПФ - более 10)

Модель производства можно представить, как некоторую **системy, перерабатывающую различные виды ресурсов в готовую продукцию.**

В качестве ресурсов выступают:



- сырье;
- трудовые затраты;
- энергозатраты;
- научно-исследовательские ресурсы;
- технологические ресурсы;
- транспортные ресурсы и др.

Производственной функцией называется зависимость между объемом произведенной продукции ***y*** и затратами различных видов ресурсов $x_1, x_2, \ldots, x_n$ необходимых для выпуска этой продукции

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

Прогнозирование с использованием производственных функций

- Очевидно, что изучив структуру объекта анализа (общую динамику показателей и их взаимосвязь, реакции на изменение эндогенных и экзогенных факторов, осуществив мониторинг закономерностей эволюции), сможем точнее предсказать его развитие, т.е. точнее выполнить его прогнозирование его траектории.
- Кобб-Дуглас (1922г.), Солоу, В. Леонтьев (50-ые годы 20 века).
- Поскольку ПФ не учитывают всех факторов производства, то они будут стохастическими и должны быть регрессионными.
- Выделяют два вида ПФ: аддитивные и мультипликативные.
- Аддитивные обладают серьезным недостатком: в них факторы абсолютно взаимозаменяемы, т.е. производство возможно без одного из них (при сборке мебели шурупы можно заменить клеем), поэтому их используют довольно редко.

Простейшая производственная функция – **один продукт** из **одного ресурса**:

$$y = f(x)$$

В реальности продуктов из одного единственного вида ресурсов не бывает, но ее можно принять как упрощение:

а) если есть один основной ресурс, а затраты остальных жестко с ним связаны;

Пример: сыр производится из молока. Другие ингредиенты (закваска, соль, специи) задаются рецептурой.

б) если количество одного ресурса можно изменять, а другие зафиксированы;

Пример: фотоателье – на количество напечатанных фотографий влияют затраты фотоматериалов (бумага, краска, пленка). Затраты на покупку фотоаппарата, принтера, зарплату фотографа, арендную плату фиксированы, не зависят от числа напечатанных фотографий.

в) если разные виды ресурсов можно объединить в один.

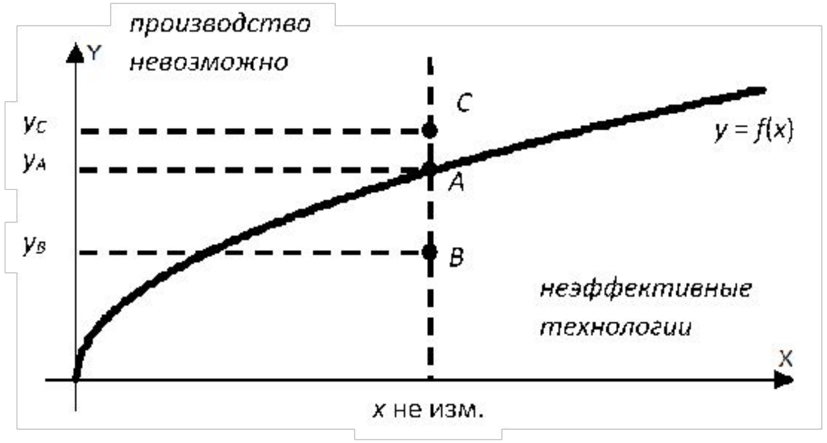
Пример: любые ресурсы можно объединить в общую сумму потраченных денег.

Но из одного и того же количества ресурса можно произвести разное количество продукта.

Примеры: из 5 м ткани можно сшить две юбки, часть ткани уйдет в обрезки. Но если правильно раскроить эту ткань, то из тех же 5 м ткани можно сшить 3 юбки. Но 4 уже нельзя;

в печи хлебозавода можно выпекать один батон, можно 5, а можно сразу 100 (но нельзя 150 или 200). Причем затраты электроэнергии на разогрев печи будут одинаковыми.

Очевидно, что выгоднее производить максимально возможное количество продукта. Другие варианты (технологии) считаются *неэффективными*. ПФ учитывает *эффективные технологии производства*, т.е. такие, для которых нельзя увеличить объем производства без увеличения затрат ресурса.



A, B, C – технологии производства (из одного и того же количества ресурса получаем разное количество продукта $y_C > y_A > y_B$). *A* – эффективная технология, *B* – неэффективная, *C* – невозможная (технически нереализуемая).

Произвести продукции больше y_A технически невозможно.

Свойства неоклассической производственной функции (для любого числа ресурсов):

Необходимость всех ресурсов – без любого из них производств невозможно:

$$f(0) = 0$$

$$f\ (0,\ x_2,\ x_3,...) = 0,$$

$$f\ (x_1,\ 0,\ x_3,...) = 0 \text{ и т.д.}$$

При увеличении затрат любого ресурса выпуск продукции увеличивается:

$$x_A > x_B \Rightarrow f(x_A) > f(x_B)$$

$$f\ '\ (x) > 0$$

Иными словами, функция $f(x)$ всюду возрастает, ее производна больше 0.

Закон убывания эффективности: с ростом затрат любого из ресурсов прирост объёма производства снижается.

Другими словами, производственная функция должна быть выпуклой вверх, а ее вторая производная отрицательна:

$$f\ ''\ (x) < 0.$$

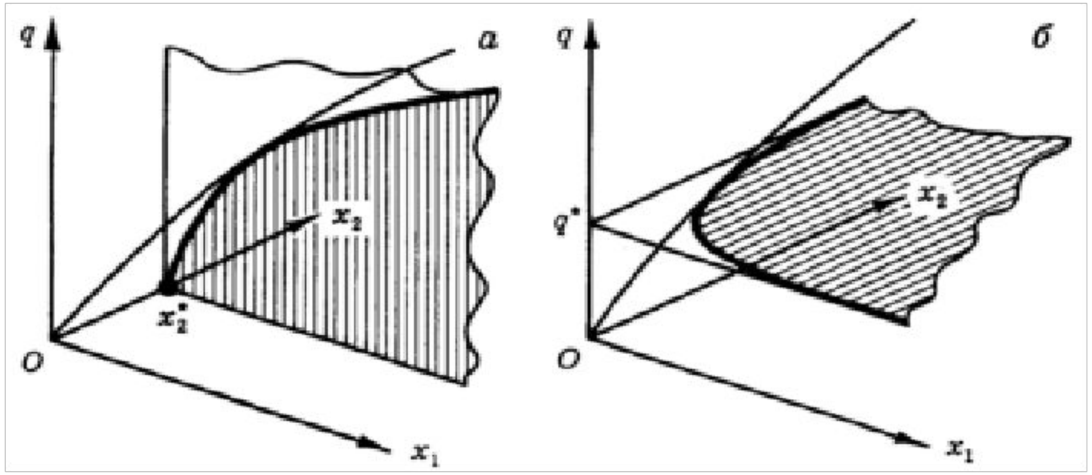
Если увеличить зарплату работников в два раза, объем производства в два раза не вырастет. Если закупить слишком много сырья, то не хватит производственных мощностей для его переработки.

На практике чаще рассматривают производственные функции с двумя ресурсами.

$$y = f(x_1, x_2)$$

<u>Например</u>, x_1 – материальные затраты; x_2 – трудовые.

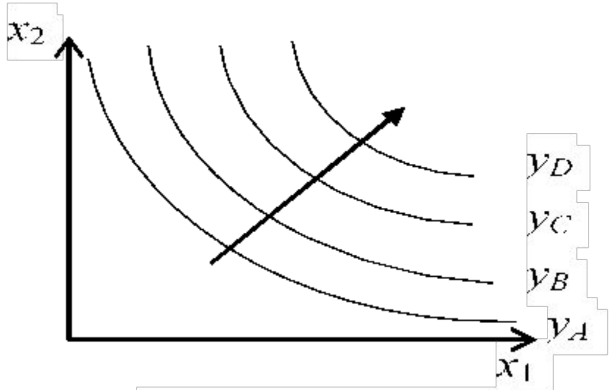
График такой функции будет трехмерным:



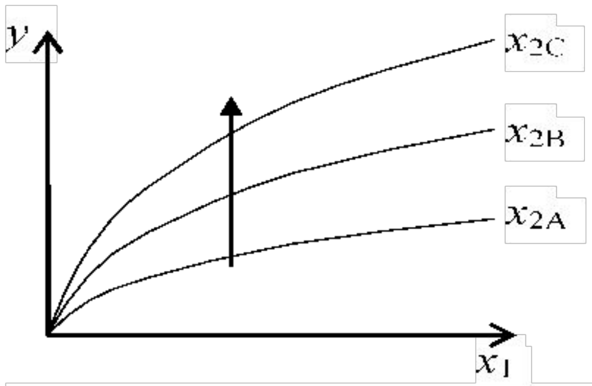
Если затраты одного ресурса зафиксировать (рисунок *а*), то получим *кривую выпуска*.

Изокванта **(греч. isoz - одинаковый и лат. quantum - сколько) – линия, обозначающая разные сочетания ресурсов, которые обеспечивают одинаковый выпуск продукции. Горизонтальный срез трехмерного графика (рисунок *б*).**

Трехмерные графики строить неудобно, поэтому используют *карту изоквант*, реже – *семейство кривых выпуска*.



Карта изоквант



Семейство кривых выпуска

Свойства изокванты:

1. Не пересекает оси при $y > 0$.
2. Изокванты не пересекаются между собой.
3. Чем больше выпуск, тем дальше изокванта от начала координат.

Основные характеристики производственной функции

Средняя производительность по каждому ресурсу – сколько в среднем получается продукта из 1ед. ресурса:

$$A_1 = \frac{f(x_1, x_2)}{x_1}, \quad A_2 = \frac{f(x_1, x_2)}{x_2},$$

если x_1 – материальные затраты (капитал), то A_1 – капиталоотдача, если x_2 – трудовые затраты, то A_2 – средняя производительность труда.

Предельная или маржинальная производительность – производные от ПФ: $M_1 = \frac{\partial f\left(x_1, x_2\right)}{\partial x_1}$,$M_2 = \frac{\partial f\left(x_1, x_2\right)}{\partial x_2}$

Они показывают приближённо, на сколько единиц изменится выпуск, если затраты того или иного ресурса изменятся на единицу.

Частная эластичность: $E_1 = \frac{M_1}{A_1}$, $E_2 = \frac{M_2}{A_2}$. Эластичности приближенно показывают, на сколько процентов изменится выпуск, если затраты того или иного ресурса изменятся на один процент.
Величина $E = E_1 + E_2$ называется полной эластичностью или *эластичностью производства*. *Технологическая норма замены* $R_{12} = \frac{E_1 x_2}{E_2 x_1}$ показывает, как изменится выпуск, если единицу 1-го ресурса заменить единицей 2-го. *Предельная норма замены* $MRS_{12} = \frac{M_1}{M_2}$ – сколько нужно добавить ресурса 2 вместо ресурса 1, чтобы объем выпуска не изменился.

1. Линейная производственная функция

Строится в случаях, когда объем выпуска пропорционален затратам:

$$y = a_0 + a_1x_1 + a_2x_2$$

Нарушаются 1 и 3 свойство ПФ, т.е. она не является неоклассической. На практике используется, когда два ресурса свободно заменяемы между собой, или для приближенного описания производственного процесса.

Пример У фирмы имеется два завода, производящих одинаковую продукцию. Если x_1 – производственные мощности 1-го завода, а x_2 – производственные мощности 2-го завода, то можно использовать линейную производственную функцию.

Изокванты и кривые выпуска – прямые линии.

$$A_1 = \frac{a_0 + a_1x_1 + a_2x_2}{x_1},\ A_2 = \frac{a_0 + a_1x_1 + a_2x_2}{x_2};\ M_1 = a_1,\ M_2 = a_2;$$

$$E_1 = \frac{a_1x_1}{a_0 + a_1x_1 + a_2x_2},\ E_2 = \frac{a_2x_2}{a_0 + a_1x_1 + a_2x_2},\ E = \frac{a_1x_1 + a_2x_2}{a_0 + a_1x_1 + a_2x_2},$$

$$R_{12} = \frac{a_1}{a_2},\ MRS_{12} = \frac{a_1}{a_2}$$

Таким образом, коэффициенты a_1 и a_2 имеют смысл предельных производительностей

1. Линейная ПФ

Строится в случаях, когда объем выпуска пропорционален затратам:

$$y = a_0 + a_1x_1 + a_2x_2$$

Нарушаются 1 и 3 свойство ПФ, т.е. она не является неоклассической. На практике используется, когда два ресурса свободно заменяемы между собой, или для приближенного описания производственного процесса.

<u>Пример.</u> У фирмы имеется два завода, производящих одинаковую продукцию. Если x_1 – производственные мощности 1-го завода, а x_2 – производственные мощности 2-го завода, то можно использовать линейную производственную функцию.

Изокванты и кривые выпуска – прямые линии:

$$A_1 = \frac{a_0 + a_1x_1 + a_2x_2}{x_1},\ A_2 = \frac{a_0 + a_1x_1 + a_2x_2}{x_2};\ M_1 = a_1,\ M_2 = a_2;$$

$$E_1 = \frac{a_1x_1}{a_0 + a_1x_1 + a_2x_2},\ E_2 = \frac{a_2x_2}{a_0 + a_1x_1 + a_2x_2},\ E = \frac{a_1x_1 + a_2x_2}{a_0 + a_1x_1 + a_2x_2},$$

$R_{12} = \frac{a_1}{a_2}$, $MRS_{12} = \frac{a_1}{a_2}$. Коэффициенты a_1 и a_2 имеют смысл предельных производительностей.

2. Степенная однофакторная производственная функция:

$$Y = \beta_0 X^{\beta_1}.$$

При положительных значениях β_i, производственная функция с ростом затрат ресурса объем производства возрастает без ограничений.

3. Показательная однофакторная производственная функция:

$$Y = \beta_0 - K\beta_1^x.$$

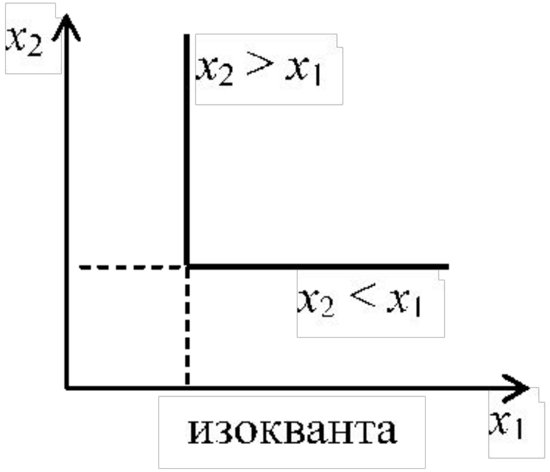
Характеризуется тем, что с ростом затрат ресурса *X* объем продукции *Y* также растет, стремясь при этом к значению параметра β_0 .

4.Функция Леонтьева (с фиксированными пропорциями)

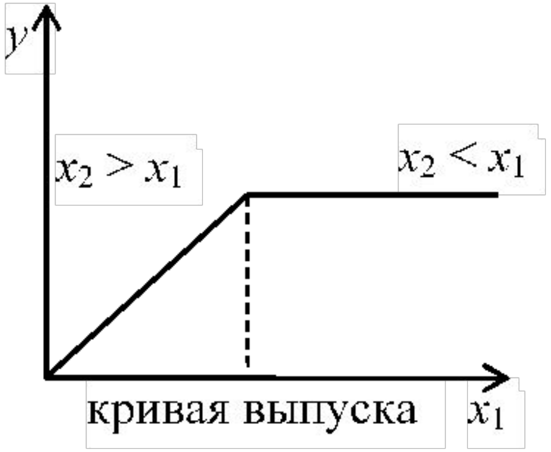
Строится в случаях строгой технологии, когда один ресурс нельзя заменить другим: $y = \min\left\{\frac{x_1}{c_1}; \frac{x_2}{c_2}\right\} = \begin{cases} \frac{x_1}{c_1}, & x_1 \le x_2 \\ \frac{x_2}{c_2}, & x_2 < x_1 \end{cases}$

c_1, c_2 – сколько каждого ресурса нужно на 1цу продукции.

<u>Пример.</u> Для производства велосипеда нужно 2 колеса и одна рама. Если в наличии имеется 3 рамы и 1000 колес, все равно можно собрать только 3 велосипеда.



изокванта



кривая выпуска

Если оба ресурса расходуются полностью (оптимальные затраты, без остатков сырья): $y = \frac{x_1}{c_1} = \frac{x_2}{c_2}, x_1 = \frac{c_1}{c_2} x_2,$

$$A_1 = \begin{cases} \frac{1}{c_1}, & x_1 \le x_2 \\ \frac{x_2}{c_2 x_1}, & x_2 < x_1 \end{cases}, \quad M_1 = \begin{cases} \frac{1}{c_1}, & x_1 \le x_2 \\ 0, & x_2 < x_1 \end{cases}, \quad E_1 = \begin{cases} 1, & x_1 \le x_2 \\ 0, & x_2 < x_1 \end{cases}$$

$$A_2 = \begin{cases} \frac{x_1}{c_2 x_2}, & x_1 \le x_2 \\ \frac{1}{c_2}, & x_2 < x_1 \end{cases}, \quad M_2 = \begin{cases} 0, & x_1 \le x_2 \\ \frac{1}{c_2}, & x_2 < x_1 \end{cases}, \quad E_2 = \begin{cases} 0, & x_1 \le x_2 \\ 1, & x_2 < x_1 \end{cases}$$

E = 1,

$MRS_{12} \to \infty$ (деление на 0).

В ПФ Леонтьева может быть лишь единственная рациональная структура производственных ресурсов.

ПФ Леонтьева предназначена для моделирования строго детерминированных технологий, не допускающих отклонений от технологических норм использования ресурсов на единицу продукции, т.е. для мелкомасштабных и полностью автоматизированных производственных объектов.

Функция Кобба-Дугласа

- Пусть **Y**-объем выпущенной продукции (в стоимостном или натуральном выражении), **K**-объем основного капитала или основных фондов, **L**- объем трудовых ресурсов или трудовых затрат(измеряемое количеством рабочих или количеством человеко-дней), тогда пространственный ряд будет иметь вид: $\boldsymbol{Y = A \cdot K^{\alpha} \cdot L^{\beta}}$, где A, α, β – идентифицируемые параметры (причем $A > 0, \le \alpha \le 1$, $0 \le \beta \le 1$, довольно часто используют условие $\alpha + \beta = 1$.
- ПФ Кобба-Дугласа применяют для описания среднемасштабных объектов (от промышленного объединения до отрасли), характеризующихся устойчивым, стабильным функционированием. Стохастическую компоненту почти всегда считают мультипликативной и имеющей логнормальное распределение, чтобы использовать «для удобства» операцию логарифмирования и применить МНК.
- Для перехода к временным рядам порой используют множители $e^{\gamma t}$ (для отражения Н-Т прогресса или ряд Фурье – для отражения циклов):

$$\boldsymbol{Y_t = A_t K_t^{\alpha} L_t^{\beta} e^{\varepsilon_t} \Rightarrow \ln Y_t = \ln A_t + \alpha \ln K_t + \beta\, L_t + e_t,}$$

где $\boldsymbol{e_t}$ - распределена по нормальному закону.

Средняя производительность:

$A_K = AK^{\alpha-1}L^{\beta}$ — *капиталоотдача,*

$A_L = AK^{\alpha}L^{\beta-1}$— *производительность труда.*

Предельная производительность:

$M_K = A\alpha K^{\alpha-1}L^{\beta}$,

$M_L = AK^{\alpha}\beta L^{\beta-1}$.

Частная эластичность:

$E_K = \alpha$,

$E_L = \beta$.

Таким образом, параметры α и β равны частным эластичностям модели и определяют отдачу при изменении масштаба производства:

$E = \alpha + \beta = 1$ – постоянная отдача от масштаба,

$E = \alpha + \beta > 1$ – возрастающая отдача от масштаба,

$E = \alpha + \beta < 1$ – убывающая отдача от масштаба.

Технологическая норма замены $R_{KL} = \dfrac{\alpha L}{\beta K}$

Предельная норма замены $MRS_{KL} = \dfrac{\alpha L}{\beta K}$

Пример.

Производственная функция небольшой фирмы имеет вид $Q=5L^{0,5}K^{0,5}$

а) каким является эффект масштаба для данной фирмы?

E = 0,5+0,5 = 1 – постоянный эффект от масштаба.

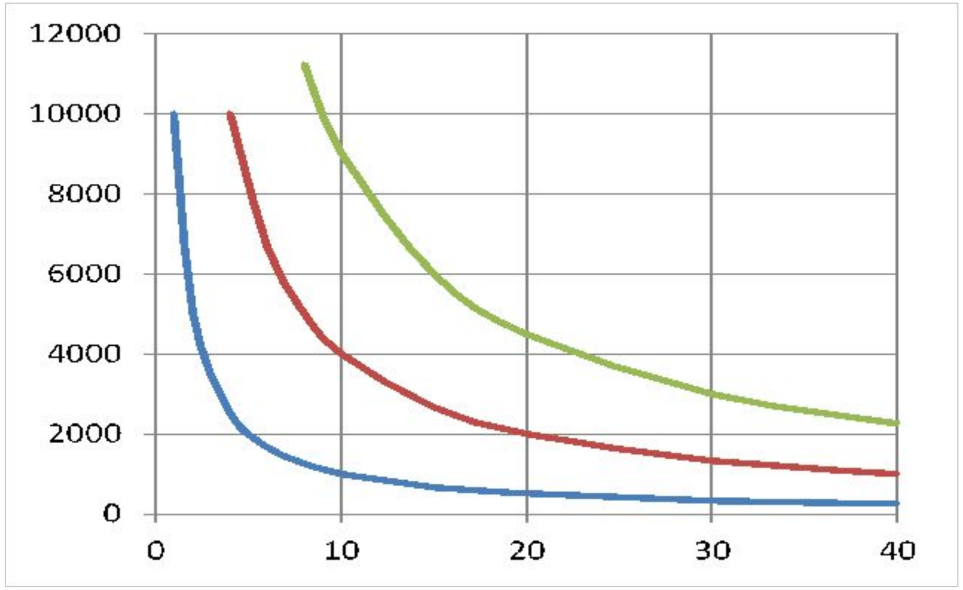
б) каков выпуск при затратах труда 9 чел/ч., и затратах капитала 9 маш./ч.? $Q=5L^{0,5}K^{0,5}$ = 5 3 3 = 45.

в) как изменится выпуск при удвоении каждого из факторов?

Удвоится (постоянная отдача от масштаба.

Изокванты: Выразим из функции Кобба-Дугласа L через K:

$$Q = AK^{\alpha}L^{\beta}, \; L^{\beta} = \frac{Q}{AK^{\alpha}}, \; L = \left(\frac{Q}{AK^{\alpha}}\right)^{\frac{1}{\beta}}.$$



Изокоста — линия, демонстрирующая комбинации объемов затраченных ресурсов, которые можно купить за одинаковую общую сумму денег. Изокосту иначе называют линией равных издержек.

Изокосты являются параллельными прямыми, если фирма может приобрести любое желаемое количество факторов производства по неизменным ценам.

Наклон изокосты выражает относительные цены факторов производства. Каждая точка на линии изокосты характеризуется одними и теми же общими издержками.

Совместив изокванты и изокосты, можно определить оптимальную позицию фирмы.

Точка *E*, в которой изокванта касается (но не пересекает) изокосту, означает наиболее дешёвую по стоимости комбинацию ресурсов, необходимых для выпуска определённого объёма продукта.

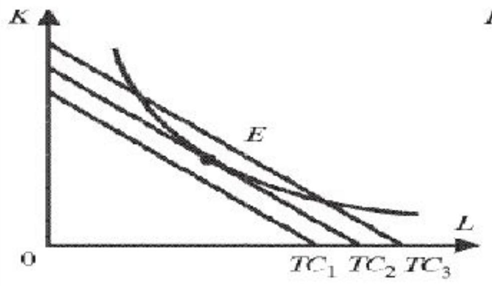


Рис. 1 Минимизация издержек при заданном объеме производства

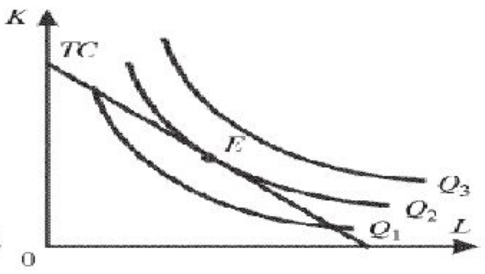


Рис. 2. Максимизаиция выпуска продукции при заданных издержках производства

ПФ постоянной эластичности замены факторов (CES) и с линейной эластичностью замены факторов (LES)

CES применяется в случаях, когда отсутствует точная информация об уровне взаимозаменяемости производственных факторов и есть основания предполагать, что он существенно не изменится при изменении объемов вовлекаемых ресурсов.

CES может быть использована для моделирования объектов любого уровня (при наличии соответствующих методов идентификации):

$$Y = \left(\alpha K^{\beta} + \beta L^{\beta}\right)^{\alpha_1}.$$

LES рекомендуется для моделирования объектов, у которых возможность замещения вовлекаемых факторов существенно зависит от их пропорций:

$$Y = K^{\alpha_0}(\alpha K + \beta L)^{\alpha_1} .$$

ПФ Солоу и ПФ с большим числом факторов

- ПФ Солоу может применяться примерно в тех же ситуациях, что и ПФ CES, однако предпосылки, лежащие в ее основе, слабее чем в ПФ CES.
- Рекомендуется в тех случаях, когда предположение об однородности представляется неоправданным. ПФ Солоу может моделировать системы любого масштаба:

$$Y = (\alpha K^{\alpha_1} + \beta L^{\alpha_2})^{\alpha_3} .$$

Известны и более сложные результаты по моделированию ПФ, включающие три, четыре и пять факторов.

В пятифакторную модель входят, например, труд, физический капитал, человеческий капитал, индекс институтов (формальных и неформальных договоров, кодексов и др. с оценкой эффективности первых трех показателей) и показатель обеспеченности инфраструктурой (плотность ж/д дорог, производство э/энергии и др.).

Интересна (для России) попытка исключение в ПФ части доходов от нефти и газа.

В качестве показателя человеческого капитала принималось - число лет обучения, которым располагает население в возрасте от 15 до 64 лет.

Исследовались при этом тренды и влияние кризисов. Сейчас реальное место России в мире от 12-20 - при учете тех или иных особенностей ПФ.

Задача агрегирования производственных функций на различных уровнях экономики

- ПФ – это технологическое соотношение, стоящее перед фирмой. Именно предприниматель выбирает нужные пропорции и уровни объема продукции. Возникает и вопрос о возможности построения производственных функций для отрасли или для промышленного, или сельскохозяйственного сектора в целом.
- Сразу видна одна трудность: те факторы, которые считались фиксированными для отдельной фирмы, вовсе не обязательно будут фиксированными для отрасли, например, предпринимательская способность. Другие факторы, такие как количество квалифицированного труда, которые не были фиксированы для отдельной фирмы, вполне могут быть значительно ограничены для отрасли. Даже если бы у всех фирм была растущая отдача от масштаба, это еще не означало бы, что отрасль целом также получает экономию от масштаба.

Расширение в отрасли в часто сталкивается с менее удобными вопросами, например, ограниченным предложением сырья и т.д.

Преимущества пространственных выборок для моделирования ПФ

- Значения переменных в пространственных выборках имеют, как правило, значительно больший размах колебаний по сравнению с макроэкономическими динамическими рядами, дают об изучаемом объекте более полное представление.
- Это свойство может быть использовано в прогнозировании на основе макроэкономических моделей для уточнения характеристик процесса за пределами наблюдений динамического ряда.
- Пространственные выборки содержат более разнообразные, по сравнению с динамическими рядами, сочетания объясняющих переменных, описывают более широкие области замещения факторов, повышая надежность получаемых оценок параметров ПФ, дают более реальное представление об эластичности замещения для моделируемого объекта.
- Уровень мультиколлинеарности данных пространственных выборок, как правило, заметно ниже, чем временных рядов.

- Когда исследователь работает с временными рядами агрегированных показателей выпуска и затрат ресурсов, то полученная ПФ отражает усредненную за рассматриваемый период времени зависимость.
- Если в качестве исходного материала используется совокупность данных по различным предприятиям (или отраслям), полученными в один и тот же момент времени (пространственная выборка), то искомая ПФ – это усредненная для данной группы предприятий (отраслей) зависимость выпуска от затрат ресурсов.
- При идентификации ПФ и многомерных регрессиях возникает ряд проблем, не получивших еще в современной удовлетворительного решения по точности моделирования и прогнозирования, по области возможного применения, поэтому далее обсудим их более подробно.

Множественная линейная регрессия

$$Y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_jx_j + \cdots + b_kx_k \ + \varepsilon_K \ ,$$

где x_j - факторы (объясняющие переменные); Y - показатель (объясняемая переменная); b_j ($j = \overline{1, k}$)- коэффициенты регрессии, определяемые при помощи МНК (СЛАУ k +1 - порядка).

Коэффициенты b_j имеют при этом разную размерность и поэтому не сравнимы друг с другом. Для сравнения степени влияния различных факторов используют стандартизированный коэффициент регрессии β_j: $\beta_j = b_j \frac{S_x}{S_y}$.

Как и в ПФ, здесь часто исследуют коэффициент эластичности $v_j = b_j \frac{\overline{x \cdot y}}{\overline{y}}$,

который говорит о том, что при отклонении данного фактора от средней величины на 1% (при постоянстве других факторов), параметр Y отклоняется от своего среднего значения на v_j % . Возможна точечная и интервальная оценки точности b_j.

Недостатки: «проклятие размерности», ее уменьшение (редукция), мнимая точность (линейность связи по параметрам и факторам, статичность, минимизация на разных метриках стохастических переменных после редукции), почему одна ε_K и т.д.

Мультиколлинеарность при идентификации ПФ

- Большой проблемой при идентификации может быть мультиколлинеарность: все или некоторые независимые переменные модели часто линейно связаны между собой, что выражается в статистической незначимости оценок параметров регрессионных моделей. Для ее выявления необходимо рассчитать коэффициенты парной корреляции между независимыми переменными. Наиболее простой метод «борьбы» с мультиколлинеарностью - исключение части связанных между собой независимых переменных (точность?). Тогда в модель включается минимальное количество факторов, ни один из которых не может быть исключен.
- Если оценки параметров ПФ статистически значимы и имеют сравнительно узкие доверительные интервалы, то их можно считать удовлетворительными. Доказательство надежности оценок - и их устойчивость, несущественность различий их величин, полученных для выборки в целом и отдельных ее частей.

Генетический алгоритм (ГА) оптимизации функции потерь при идентификации моделей

ГА является одним из наиболее универсальных и точных методов идентификации, применимым с высокой точностью при числе параметров нелинейных параметров до 5-6 при относительно высокой дисперсии шума.

ГА работает с совокупностью популяцией, каждая из которых представляет возможное решение данной проблемы и оценивается мерой ее «приспособленности»: насколько «хорошо» соответствующее ей решение задачи.

Мерой приспособленности могло бы быть отношение силы/веса для примера проектирования моста (в природе это эквивалентно оценке того, насколько эффективен организм при конкуренции за ресурсы.)

Наиболее приспособленные особи получают возможность "воспроизводить" потомство с помощью "перекрестного скрещивания" с другими особями популяции.

Это приводит к появлению популяций, которые сочетают в себе некоторые характеристики, наследуемые ими от «родителей».

Так и воспроизводится вся **новая** популяция, выбирая лучших представителей предыдущего поколения, скрещивая их и получая множество новых особей.

Скрещивание наиболее приспособленных особей приводит к тому, что исследуются наиболее перспективные участки пространства поиска. В конечном итоге, популяция будет сходиться к оптимальному решению задачи.

Схематично работу генетического алгоритма можно представить в следующем виде:

В процессе «селекции» отбирают только несколько лучших «пробных» решений, на основании заданного критерия (например, МНК). «Скрещивание» вместо пары решений создаёт другую пару решений (к примеру, путем нахождения среднего).

В результате серии «скрещиваний» размер прореженной исходной выборки (она называется обычно «популяцией») увеличивается до исходного размера. «Мутация» случайным образом изменяет коэффициенты решений, выводя алгоритм из состояний определения локальных экстремумов.

Для привнесения в ГА способности к настройке собственной структуры и/или параметров разработаны так называемые адаптивные алгоритм: изменение параметров скрещивания, изменение размера популяции, введение нечеткости в блок управления ГА.

ГА применяются для решения следующих задач: оптимизация функций; разнообразные задачи на графах (задача коммивояжера, раскраска вершин или ребер и т.д.); настройка и обучение искусственной нейронной сети; задачи компоновки; составление расписаний; игровые стратегии; аппроксимация функций; биоинформатика и т.д.

Преимущества ГА: универсальность; высокая обзорность поиска; нет ограничений на целевую функцию; любой способ задания функции.

Недостатки ГА: относительно высокая вычислительная стоимость; квазиоптимальность.

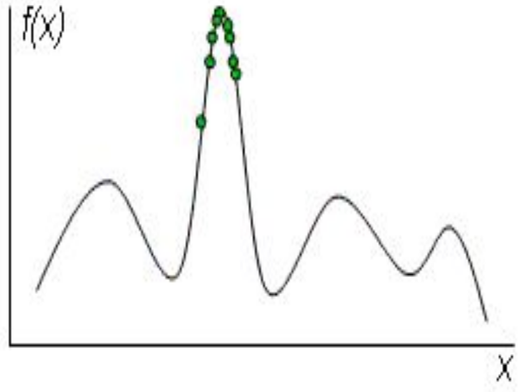
Когда надо использовать ГА: много параметров, плохая целевая функция, комбинаторные задачи.

Когда не надо использовать ГА: если задача хорошо решается традиционными методами и требуется высокая точность решения (например, модели авторегрессии – скользящего среднего при малой дисперсии шума).

Критерии останова ГА

Важный момент функционирования алгоритма ГА – определение критериев останова. В качестве критериев останова применяются или ограничение на максимальное число эпох функционирования алгоритма, или определение его сходимости, обычно путем сравнивания приспособленности популяции на нескольких этапах и остановки при стабилизации этого параметра.

Схождением называется такое состояние популяции, когда все строки популяции почти одинаковы и находятся в области некоторого экстремума (см. рис.). В такой ситуации «кроссовер» практически не изменяет популяции, а вышедшие из этой области за счет мутации особи склонны вымирать, имеют меньшую приспособленность.



Особенно если данный экстремум является глобальным максимумом. Таким образом, схождение популяции обычно означает, что найдено лучшее или близкое к нему решение.

Лаги в экономических моделях

При анализе многих экономических показателей (особенно в макроэкономике) часто используются ежегодные, ежеквартальные, ежемесячные и ежедневые данные. В этом случае следует упорядочить данные по времени их получения и построить временные ряды.

Динамические модели

Модели с распределенными лагами

Авторегрессионные модели

Причин наличия лагов в экономике достаточно много, и среди них можно выделить следующие:

Психологические причины – обычно выражаются через инерцию в поведении людей. Например, люди тратят свой доход постепенно, а не мгновенно. Привычка к определенному образу жизни приводит к тому, что люди приобретают те же блага в течении некоторого времени даже после падания реального дохода.

Технологические причины – например, изобретение персональных компьютеров не привело к мгновенному вытеснению ими больших ЭВМ в силу необходимости замены соответствующего программного обеспечения, которое потребовало продолжительного времени.

Институциональные причины – например, контракты между фирмами, трудовые договоры требуют определенного постоянства в течение времени контракта.

Механизм формирования экономических показателей – например, инфляция во многом является инерционным процессом, денежный мультипликатор (коэффициент, который показывает, во сколько раз увеличится (сократится) денежная масса при увеличении (сокращении) денежной базы на единицу на монетарном рынке) также проявляет себя на определенном временном интервале.

Модели авторегрессии

(применяют более 100 видов авторегрессий)

Наиболее известны следующие виды моделей:

- собственно модели авторегрессии (AR - auto regressive) с бесконечным числом слагаемых:

$$Y_k = \sum_{i=1}^{\infty} \lambda_i Y_{k-i} + \varepsilon_k \,;$$

- модели скользящего среднего (MA- auto average) с бесконечным числом слагаемых:

$$Y_k = \sum_{i=1}^{\infty} \gamma_i \varepsilon_{k-i} \,;$$

- модели авторегрессии - скользящего среднего (ARMA - autoregressive moving average   или

$$Y_k = \sum_{i=1}^{p} \lambda_i Y_{k-i} + \sum_{i=1}^{g} \gamma_i \varepsilon_{k-i} \,.$$

о слагаемых, p + q , что удобно для практическ

Частные виды авторегрессий

Авторегрессионные модели (АР-модели):

–первого порядка $$Y_k = \lambda_0 + \lambda_1 Y_{k-1} + \varepsilon_k$$

–порядка P $$Y_k = \lambda_0 + \lambda_1 Y_{k-1} + \ldots + \lambda_P Y_{k-P} + \varepsilon_k$$

Авторегрессионная модель с распределенными лагами порядка (L, P):

$$Y_k = \alpha + \beta_0 X_k + \beta_1 X_{k-1} + \ldots + \beta_L X_{k-L} + \gamma_1 Y_{k-1} + \gamma_2 Y_{k-2} + \ldots + \gamma_P Y_{k-P} + \varepsilon_k$$

Модель парной пространственной линейной регрессии:

$$Y_k = \alpha + \beta X_k + \varepsilon_k$$

Пространственная модель авторегрессии:

$$Y_k = \alpha + \beta X_k + \gamma Y_{k-1} + \varepsilon_k$$

- авторегрессии - проинтегрированного скользящего среднего (ARIMA - autoregressive integrated moving average) – они оправданы в первую очередь, для тренд-сезонных временных рядов (1970г. Бокс, Дженкинс);

- авторегрессии с условной гетероскедастичностью (ARCH -autoregressive conditional heteroscedasticity) - характерно для валового национального продукта (ВНП), индексов цен, денежной массы, урожайности и т.д.;

- расширения указанных выше авторегрессионных моделей: обобщенная авторегрессионная условно гетероскедастическая модель (GARCH - generalized autoregressive conditional heteroscedasticity model), интегрированная обобщенная авторегрессионная условно гетероскедастическая модель (IGARCH - integrated generalized autoregressive conditional heteroscedasticity model) и др.

Конструирование моделей авторегрессии

Применение моделей авторегрессии-скользящего среднего (AR-CC-модели) было известно в теории управления, но там «разладка» (изменение параметров модели) не была связано с параметрами (моделями) системы управления, не использовалось для идентификации параметров модели.

$$Y_k = \sum_{i=1}^{p} \lambda_i Y_{k-i} + \sum_{i=1}^{g} \gamma_i \varepsilon_{k-i}$$

Дискретные наблюдения действительной переменной (времени) становятся функцией комплексной переменной Z, свойства преобразования, использование таблицы соответствий или программы «Maple».

Прямое Z-преобразование (и его свойства):

$$Z\left[D_k\right] = \sum_{k=0}^{\infty} D_k\ z^{-k}$$

Обратное Z-преобразование:

$$Z^{-1}\left[Z\left[D_k\right]\right] = \frac{1}{2\pi i} \oint Z\left[D_k\right] z^{k-1} dz = D_k$$

Экспоненциальная функция, обобщенная экспоненциальная функция, экспоненциальные полиномы

- Экспонента относится к числу самых распространенных моделей в экономике, входит в состав и многих других моделей (полиномов).
- Экспоненциальный тренд характерен для явлений, развивающихся в среде, не создающей никаких ограничений, поэтому он может быть адекватным лишь на ограниченном интервале аргумента, так как любая среда рано или поздно создаст ограничения, а любые ресурсы будут исчерпаны.
- Например, при создании и освоении новых видов продукции рост оправдана именно такая простая, но позволяющая определить максимально возможный рост увеличения дохода.
- При наполнении рынка экспоненциальный рост прекращается.

Используют, например, модели

$D_k = A\,e^{-\lambda k\Delta}$, $D_k = A_0 + A\,e^{-\lambda k\Delta}$ - обобщенная экспонента,

$D_k = (A_0 + A\,k\Delta)\Delta\,e^{-\lambda k\Delta} + A_2\,\mathrm{Cos}\,(\omega k\Delta и + \phi)$.

Через экспоненты можно определить и такое явление, как структурный сдвиг.

Для экспоненты $D_k = A\,e^{-\alpha k \Delta}$ Z – преобразование при $k \geq 1$ дает разностную схему первого порядка:

$$Z[D_k] = {}^{1}\!/_{(1-\lambda_1 z^{-1})}$$

$$\lambda_1 = e^{-\alpha\Delta} \Rightarrow D_k - \lambda_1 D_{k-1} = 0.$$

Запишем теперь разностную схему через наблюдения в структуре $Y_k = A\,e^{-\alpha k \Delta} + \varepsilon_k$

$$Y_k = \lambda_1 Y_{k-1} + \varepsilon_k - \lambda_1 \varepsilon_{k-1} \Rightarrow Y_k = \lambda_1 Y_{k-1} + \mu_k \Rightarrow$$

где $\mu_k = \varepsilon_k - \lambda\varepsilon_{k-1}$ - новая стохастическая компонента, обладающая автокорреляцией.

Казалось бы можно найти оценку параметра экспоненты по трем наблюдениям, но МНК на n наблюдениях из СЛАУ первого порядка обеспечит помехозащищенность оценки параметра экспоненты:

$$\Rightarrow \alpha^0 = -\frac{1}{\Delta}\ln\lambda_1^0 .$$

Затем конструируем второе СЛАУ (также первого порядка) с использованием α^0 для нахождения A^0 :

⇑

$$Y_k = A\,e^{-\alpha^o k \Delta} \Rightarrow A^o.$$

Рассмотрим теперь гармонику с аддитивной стохастической компонентой

$$S_k = A\,\mathrm{Cos}\,(\omega k\Delta + \phi)$$

$$S_k = \nu_1 S_{k-1} - S_{k-2}$$

$$\nu_1 = 2\cos\omega\Delta$$

- разностная схема 2 порядка, при $k \geq 2$ **.**

1 этап: модель авторегрессии для оценки частоты:

$$Y_\kappa = \nu_1 Y_{\kappa-1} - Y_{\kappa-2} + \xi_\kappa,$$

МНК, СЛАУ первого порядка при $k \geq 2$.

$$\omega^o = \frac{1}{\Delta} Arccos\frac{\nu_1^o}{2}$$

2 этап: оценка амплитуды и фазы:

$$A_1^\circ, A_2^\circ = \underset{A_1,A_2}{\mathrm{argmin}} \sum_{\kappa=0} \{Y_\kappa - A_1 \mathrm{Cos}\,\omega^\circ\kappa\Delta + A_2 \mathrm{Sin}\,\omega^\circ\kappa\Delta\}^2,$$

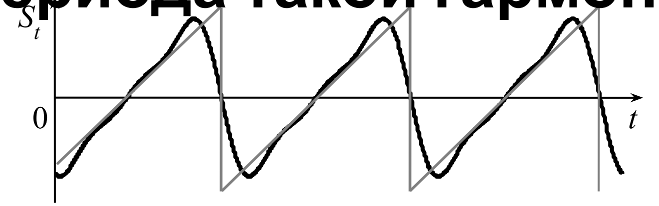
$$A_1 = A\cos\phi \qquad A_2 = A\sin\phi$$

МНК, СЛАУ

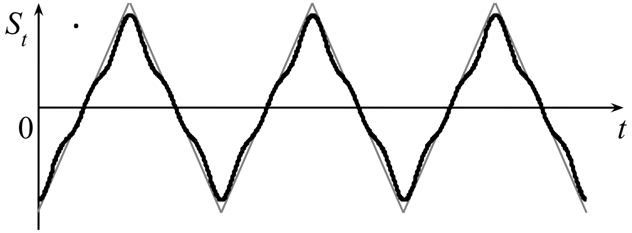
второго порядка

$$A^0 = ((A_1^0)^2 + (A_2^0)^2)^{1/2} \qquad \phi^0 = Arc\,tg(A_2^0 / A_1^0)$$

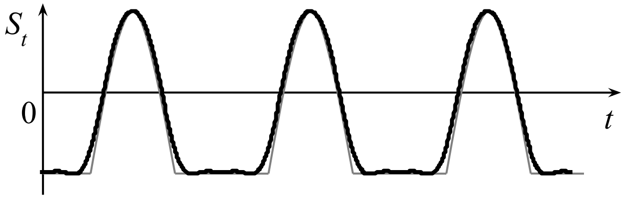
Напомню, что три члена разложения передают многообразие моделей колебательной компоненты при соблюдении условия Найквиста-Котельникова назначения интервала дискретизации по самой высокочастотной гармонике. Уникальное свойство метода – возможность оценки всех трех параметров всех гармоник даже на доле периода такой гармоники.



Колебания пилообразной формы,



Колебания треугольной формы,



Колебания куполообразной формы.

Автокорреляция стохастических компонент в моделях регрессии, методы их компенсации

Из анализа автокорреляции можно, как уже показано возможно определять параметры моделей. Например, для колебательной компоненты можно найти все три параметра на доле периода, что оказывается существенно быстрее и точнее, чем в других известных методах. Это позволяет работать на относительно коротких выборках траектории, осуществлять мониторинг эволюции. Однако при этом возникает и другая проблема, которую следует решать.

На примере стохастической компоненты (остатков) μ_k в модели экспоненты можно видеть, что она обладает автокорреляцией, нарушая тем самым одно из условий Гаусса-Маркова. Тоже явление будет наблюдаться и в других авторегрессиях.

Мощным, но теоретически сложным методом их компенсации является применение обобщенного метода наименьших квадратов (ОМНК), как и для компенсации гетероскедастичности.

Простым, но достаточно эффективным является метод прореживания выборки: использования наблюдений через одно, через два наблюдения и т.д., идентификация через них, причем, с учетом нового значения интервала дискретизации.

Увеличение требуемой выборки, происходящее при этом, все же не так велико, сохраняя возможность осуществлять мониторинг эволюции параметров анализируемых моделей трендов и колебательных компонент.

Преимущества и недостатки моделей трендов

Преимущества:

- возможность реализации и двух подходов к моделированию – в начале реализовать алгоритмический подход для сглаживания (текущее простое или взвешенное, или адаптивное), а затем по полученным данным – параметрический (например, по МНК), а по нему выполнить прогнозирование (среднесрочное и долгосрочное). Для краткосрочного прогнозирования необходимо идентифицировать и колебательные компоненты траекторий;

- минимум априорных предположений о процессах, протекающих в объекте исследований, дает большую возможность получения новых результатов, не ограниченных этими предположениями.

Недостатки:

- традиционные тренды применимы к только к обратимым (повторяющимся, не эволюционирующим) процессам (исключения - логисты, структурные сдвиги, эволюция амплитуд колебательных компонент, изменение структур траекторий);

- при получении новых наблюдений идентификацию надо осуществлять заново.

От последнего недостатка, сохраняя возможность прогнозирования, свободны модели авторегрессий наблюдений, которые и рассмотрим далее.

Модели Койка

- Модель (с геометрически распределенными лагами)

$$Y_k = \alpha + \beta_0 \sum_{i=0}^{\infty} \lambda^i X_{k-i} + \varepsilon_k$$

$$Y_k = \alpha(1-\lambda) + \beta_0 X_k + \lambda Y_{k-1} + \varepsilon_k - \lambda \varepsilon_{k-1}$$

- Модель адаптивных ожиданий по определяющему фактору

$$Y_k = \alpha + \beta X_k^* + \varepsilon_k,$$

$$X_k^* = \gamma X_k + (1-\gamma) X_{k-1}^*$$

$$Y_k = \alpha\gamma + \beta\gamma X_k + (1-\gamma) Y_{k-1} + \varepsilon_k - (1-\gamma)\varepsilon_{k-1}$$

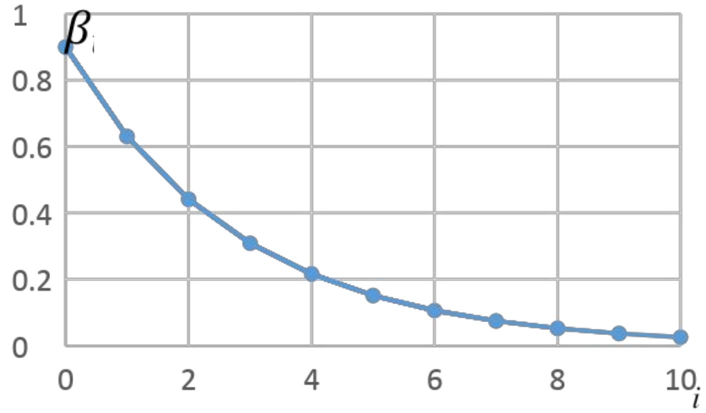
- Модель частичной корректировки по определяемому фактору

$$Y_k^* = \alpha + \beta X_k + \varepsilon_k,$$

$$Y_k = \lambda Y_k^* + (1-\lambda) Y_{k-1}$$

$$Y_k = \lambda\alpha + \beta\lambda X_k + (1-\lambda) Y_{k-1} + \lambda \varepsilon_k$$

Модель с геометрически распределенными лагами (1 метод)



Предполагается, что коэффициенты модели убывают в геометрической прогрессии: $\beta_i = \beta_0 \lambda^i, \quad 0 < \lambda < 1$.

Долгосрочный мультипликатор: $\boxed{\sum_{i=0}^{\infty} \beta_i = \frac{\beta_0}{1-\lambda}}$

$$Y_k = \alpha + \beta_0 X_k + \beta_1 X_{k-1} + \beta_2 X_{k-2} + \ldots + \varepsilon_k = \alpha + \beta_0 X_k + \beta_0 \lambda X_{k-1} + \beta_0 \lambda^2 X_{k-2} + \ldots + \varepsilon_k$$

$$\boxed{Y_k = \alpha + \beta_0 \sum_{i=0}^{\infty} \lambda^i X_{k-i} + \varepsilon_k}$$

Данная модель нелинейна по параметрам, для ее идентификации обычно используют простой перебор по сетке значений:

$$\boxed{\lambda^* = 0; 0,01; 0,02; \ldots 1} \qquad Z_k = \sum_{i=0}^{\infty} \left(\lambda^*\right)^i X_{k-i} \qquad Y_k = \alpha + \beta_0 Z_k + \varepsilon_k$$

Из всех наборов параметров выбирается тот, который даёт наибольший R^2.

Модель с геометрически распределенными лагами (2 метод - преобразование Койка)

$$Y_k = \alpha + \beta_0 X_k + \beta_0 \lambda X_{k-1} + \beta_0 \lambda^2 X_{k-2} + \ldots + \varepsilon_k$$

$$Y_{k-1} = \alpha + \beta_0 X_{k-1} + \beta_0 \lambda X_{k-2} + \beta_0 \lambda^2 X_{k-3} + \ldots + \varepsilon_{k-1}$$

$$Y_k - \lambda Y_{k-1} = \alpha + \beta_0 X_k + \beta_0 \lambda X_{k-1} + \beta_0 \lambda^2 X_{k-2} + \cdots + \varepsilon_k -$$

$$- \lambda\alpha - \beta_0 \lambda X_{k-1} - \beta_0 \lambda^2 X_{k-2} - \cdots - \lambda\varepsilon_{k-1} =$$

$$= \alpha - \lambda\alpha + \beta_0 X_k + \varepsilon_k - \lambda\varepsilon_{k-1}$$

Получим авторегрессионную модель Койка:

$$\boxed{Y_k = \alpha(1-\lambda) + \beta_0 X_k + \lambda Y_{k-1} + \nu_k, \quad \nu_k = \varepsilon_k - \lambda\varepsilon_{k-1}}$$

К данной модели может быть применен МНК. Однако для нее нарушаются условия Гаусса-Маркова и, в силу этого, полученные оценки параметров будут несостоятельными и смещенными (пояснить содержание этих недостатков).

Пример сравнения методик расчета параметров

Имеются данные о динамики цен на сырье X_k и цен на товар, производимый из этого сырья Y_k .

Число наблюдений n = 250.

По этим данным были найдены оценки параметров модели Койка:

МНК, примененным к модели Койка, дал:

$$\alpha^* = -0,38 \quad \beta_0^* = 4,78 \quad \lambda^* = 0,39$$

Через модель с геометрически распределенными лагами ($L = 20$) получили:

$$\alpha^* = -2,2 \quad \beta_0^* = 4,84 \quad \lambda^* = 0,39$$

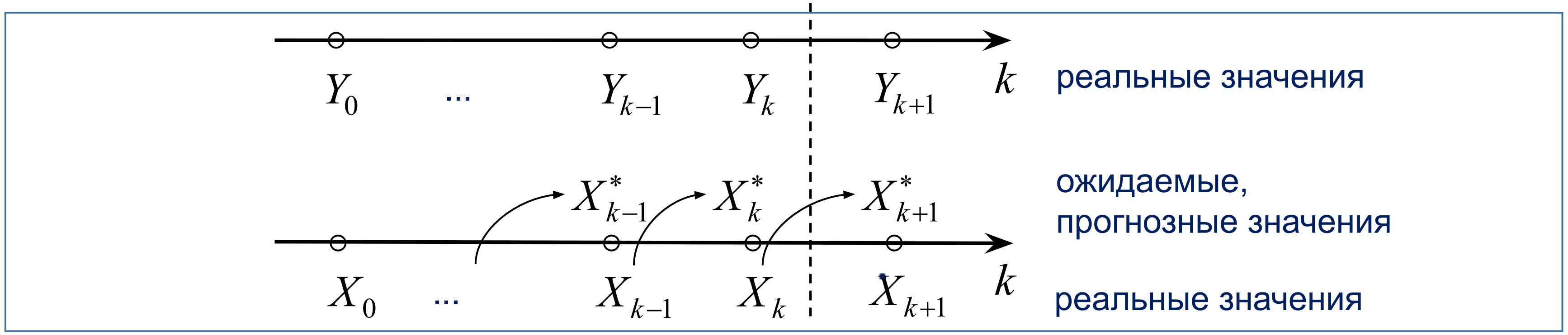
Модель адаптивных ожиданий и корректировки

- Модели используются для эмпирической верификации макроэкономических моделей, в которых учитываются ожидания экономических агентов относительно значений экономических показателей, включенных в модель в момент времени *t*.
- Механизмы этих моделей могут быть существенно различны и сложны, поэтому рассмотрим наиболее простые из них – адаптивных ожиданий определяющего фактора (определяющей переменной) и неполной корректировки определяемого фактора (определяемой переменной).
- Оценку параметров каждой из этих моделей можно проводить, используя модель авторегрессии.

Модель адаптивных ожиданий определяющего фактора

$$\boxed{Y_k = \alpha + \beta X_k + \varepsilon_k}$$ - традиционная модель авторегрессии.

(1) $Y_k = \alpha + \beta X^*_{k+1} + \varepsilon_k$ - определим ее через прогнозируемое ожидание X^*_{k+1} определяющего фактора.



(2) $$\underbrace{X^*_{k+1} - X^*_k}_{\text{коррекция прогноза на шаг вперед}} = \gamma \underbrace{\left(X_k - X^*_k\right)}_{\text{ошибка прогноза}}$$

$0 \le \gamma \le 1$ -коэффициент ожидания

Если γ стремится к 1, то реализуются ожидания агентов, если к нулю, то имеем устойчивость (сохранение) тенденции.

Из (2) получим (3) $X^*_{k+1} = \gamma X_k + (1-\gamma) X^*_k$

Из(1) и (3): (4) $$\boxed{\begin{aligned} &Y_k = \alpha + \beta X^*_{k+1} + \varepsilon_k, \\ &X^*_{k+1} = \gamma X_k + (1-\gamma) X^*_k \end{aligned}}$$

Преобразование модели адаптивных ожиданий использует подстановку (3) в модель авторегрессии (1), что дает:

$$\boxed{Y_k = \alpha + \beta\left(\gamma X_k + (1-\gamma) X_k^*\right) + \varepsilon_k,}$$

Уравнение авторегрессии (1) при новом аргументе примет вид:

$$Y_{k-1} = \alpha + \beta X_k^* + \varepsilon_{k-1},$$

$$Y_k - (1-\gamma) Y_{k-1} = \alpha + \beta\gamma X_k + (1-\gamma)\beta X_k^* + \varepsilon_k - (1-\gamma)\alpha - (1-\gamma)\beta X_k^* - (1-\gamma)\varepsilon_{k-1}$$

$$= \gamma\alpha + \beta\gamma X_k + \varepsilon_k - (1-\gamma)\varepsilon_{k-1}$$

Получим авторегрессионная модель Койка, в которой можно последовательно, используя МНК найти $\beta\gamma \Rightarrow \beta$, затем $1-\gamma \Rightarrow \gamma$ и, наконец, $\alpha\gamma \Rightarrow \alpha$.

(4) $$\boxed{Y_k = \alpha\gamma + \beta\gamma X_k + (1-\gamma) Y_{k-1} + \nu_k, \quad \nu_k = \varepsilon_k - (1-\gamma)\varepsilon_{k-1}}$$

Можно (в силу не выполнения условий Гаусса-Маркова) найти и другое решение:

Обратное преобразование Койка

(1) $$Y_k = \alpha\gamma + \beta\gamma X_k + (1-\gamma)Y_{k-1} + \varepsilon_k - (1-\gamma)\varepsilon_{k-1}$$

(2) $$Y_{k-1} = \alpha\gamma + \beta\gamma X_{k-1} + (1-\gamma)Y_{k-2} + \varepsilon_{k-1} - (1-\gamma)\varepsilon_{k-2}$$

Подставим (2) в (1):

$$Y_k = \alpha\gamma + \beta\gamma X_k + \varepsilon_k - (1-\gamma)\varepsilon_{k-1} + (1-\gamma)\alpha\gamma + \beta\gamma(1-\gamma)X_{k-1} + (1-\gamma)^2 Y_{k-2} +$$
$$+(1-\gamma)\varepsilon_{k-1} - (1-\gamma)^2\varepsilon_{k-2} = \alpha\gamma\big(1 + (1-\gamma)\big) + \beta\gamma(X_k + (1-\gamma)X_{k-1}) +$$
$$+(1-\gamma)^2 Y_{k-2} + \varepsilon_k - (1-\gamma)^2\varepsilon_{k-2}$$

$$Y_{k-2} = \alpha\gamma + \beta\gamma X_{k-2} + (1-\gamma)Y_{k-3} + \varepsilon_{k-2} - (1-\gamma)\varepsilon_{k-3}$$

$$Y_k = \alpha\gamma\left(1 + (1-\gamma) + (1-\gamma)^2\right) + \beta\gamma\left(X_k + (1-\gamma)X_{k-1} + (1-\gamma)^2 X_{k-2}\right) + (1-\gamma)^3 Y_{k-3} + \varepsilon_k - (1-\gamma)^3\varepsilon_{k-3}$$

…

$$Y_k = \alpha\gamma\sum_{i=0}^{L}(1-\gamma)^i + \beta\gamma\sum_{i=0}^{L}(1-\gamma)^i X_{k-i} + (1-\gamma)^L Y_{k-3} + \varepsilon_k - (1-\gamma)^L\varepsilon_{k-3}, \quad L \to \infty$$

Получим модель с геометрически распределенными лагами для которого вновь возможен перебор по сетке значений :

$$Y_k = \alpha + \beta\gamma\sum_{i=0}^{\infty}(1-\gamma)^i X_{k-i} + \varepsilon_k$$

Модель частичной корректировки

В данной модели, в отличие от метода адаптивных ожиданий рассматриваются корректировки определяемой переменной Y_k.

$$\boxed{Y_k^* = \alpha + \beta X_k + \varepsilon_k}$$

Частичная корректировка:

$$Y_k - Y_{k-1} = \lambda\left(Y_k^* - Y_{k-1}\right) \qquad Y_k = \lambda Y_k^* + (1-\lambda)Y_{k-1}$$

$$\boxed{\begin{gathered} Y_k^* = \alpha + \beta X_k + \varepsilon_k, \\ Y_k = \lambda Y_k^* + (1-\lambda)Y_{k-1} \end{gathered}}$$

Авторегрессионная модель Койка, для идентификации которой можно применить МНК: $Y_k = \lambda\alpha + \beta\lambda X_k + (1-\lambda)Y_{k-1} + \lambda\varepsilon_k$

Модель частичной корректировки

В данной модели рассматриваются корректировки определяемой переменной Y_k.

$$\boxed{Y_k^* = \alpha + \beta X_k + \varepsilon_k}$$

Частичная корректировка:

$$Y_k - Y_{k-1} = \lambda\left(Y_k^* - Y_{k-1}\right) \quad Y_k = \lambda Y_k^* + (1-\lambda)Y_{k-1}$$

$$\boxed{\begin{aligned} &Y_k^* = \alpha + \beta X_k + \varepsilon_k, \\ &Y_k = \lambda Y_k^* + (1-\lambda)Y_{k-1} \end{aligned}}$$

Вновь приходим к авторегрессионной модели Койка и известным способам её решения:

$$Y_k = \lambda\alpha + \beta\lambda X_k + (1-\lambda)Y_{k-1} + \lambda\varepsilon_k$$

Примеры применения моделей авторегрессии:

Для модели адаптивных ожиданий:

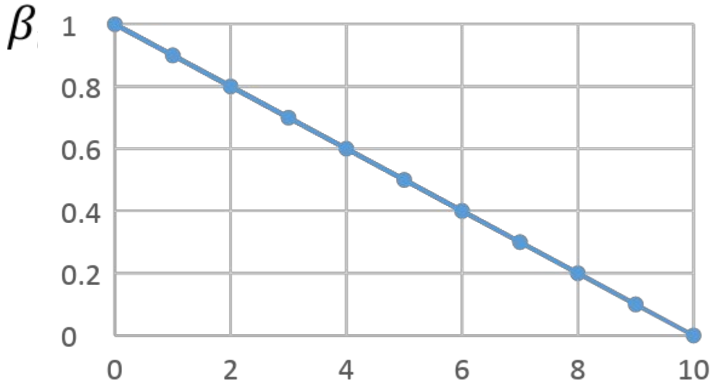
определяемая переменная - реальная заработанная плата, а определяющий фактор - ожидаемая величина уровня безработицы в условиях полной занятости.

Для модели частичной корректировки:

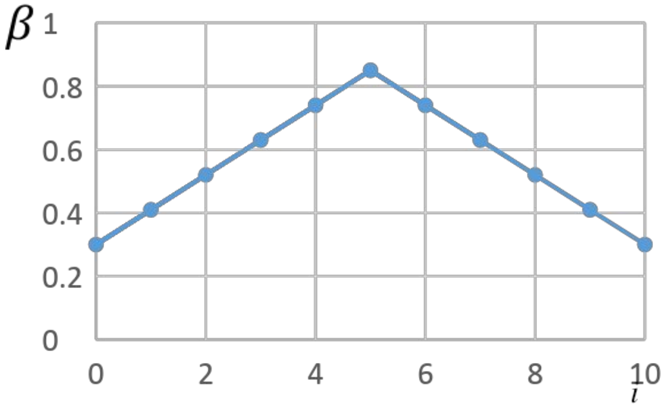
определяемый фактор - спрос на труд, а определяющий фактор – уровень занятости (при предпосылке в долгосрочной перспективе об определенном желательном уровне занятости).

Варианты структуры лага в авторегрессиях с распределенными лагами

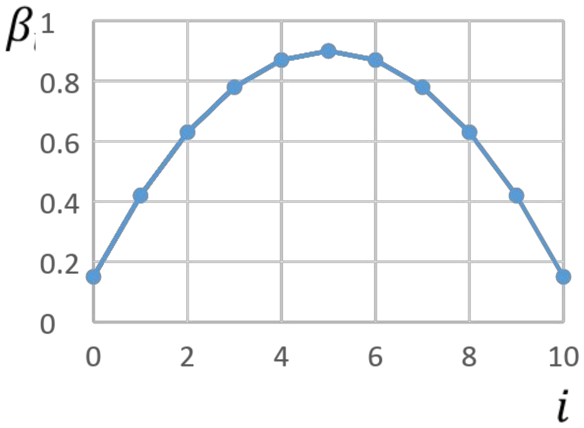
$$Y_k = \alpha + \beta_0 X_k + \beta_1 X_{k-1} + \cdots + \beta_L X_{k-L} + \varepsilon_k$$

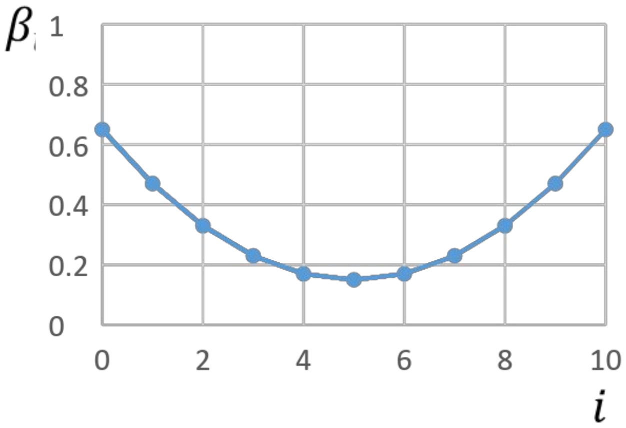


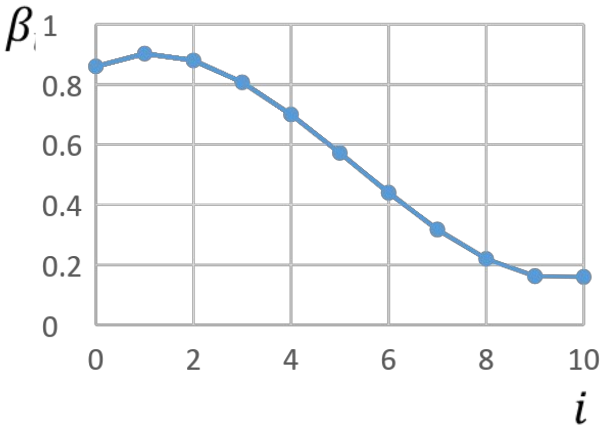
линейная структура лага



перевернутая V-образная структура лага

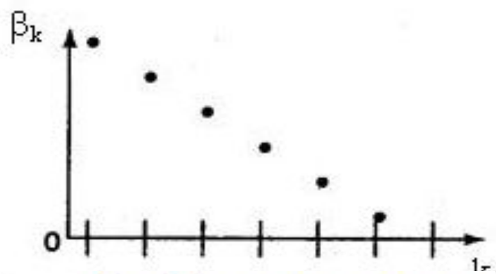




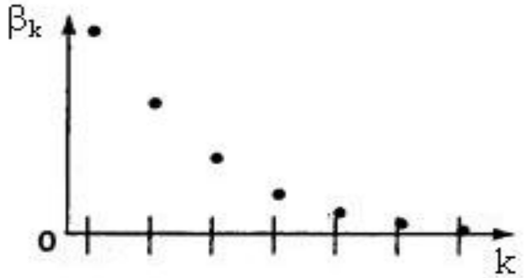


полиномиальная структура лага

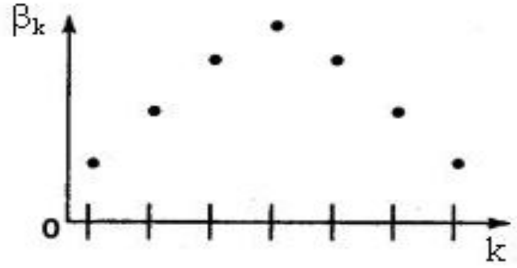
Метод Алмон для динамических моделей с распределенными размерами



Линейная структура лага

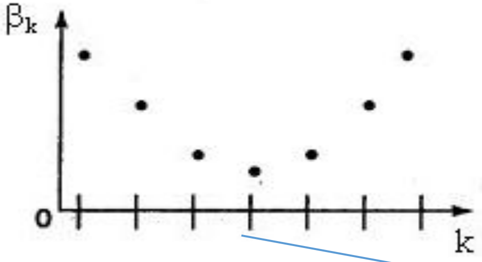


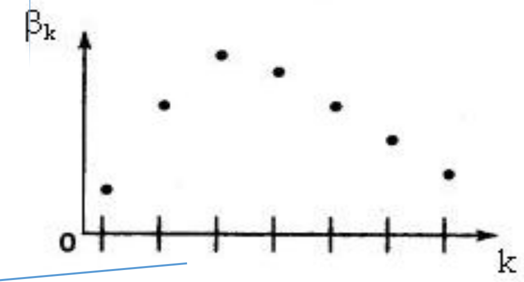
геометрическая структура лага (Модель Койка)



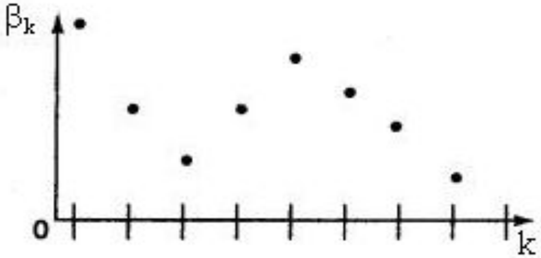
перевернутая V-образная

Имеют тенденции убывать во времени





Полиномиальная 1 степени



полиномиальная 3-ей степени

Не имеют тенденции убывать во времени

полиномиальные

Основная трудность в выявлении структуры лага в том, как получены значения β_k

Распределенные лаги Ш. Алмон

Предполагается, что коэффициенты изменяются по полиномиальному закону:

$$\beta_i = c_0 + c_1 i + c_2 i^2 + \dots + c_m i^m, \qquad i = \overline{1,L}$$

m - выбранный порядок полинома (на практике обычно 2 или 3-ий)

$$\beta_0 = c_0$$

$$\beta_1 = c_0 + c_1 + c_2 + \dots + c_m$$

$$\beta_2 = c_0 + 2c_1 + 4c_2 + \dots + 2^m c_m$$

$$\dots$$

$$\beta_L = c_0 + c_1 L + c_2 L^2 + \dots + c_m L^m$$

Подставим значения коэффициентов в модель:

$$Y_k = \alpha + c_0 X_k + (c_0 + c_1 + \dots + c_m)X_{k-1} + (c_0 + 2c_1 + \dots + 2^m c_m)X_{k-2} +$$

$$+ \dots + (c_0 + Lc_1 + \dots + L^m c_m)X_{k-L} + \varepsilon_k =$$

$$\alpha + c_0(X_k + X_{k-1} + \dots + X_{k-L}) + c_1(X_{k-1} + 2X_{k-2} + \dots + LX_{k-L}) +$$

$$+ \dots + c_m(X_{k-1} + 2^m X_{k-1} + \dots + L^m X_{k-L}) + \varepsilon_k$$

Характеристики метода Алмон для динамических моделей с распределенными размерами

Основные недостатки метода Алмон:

-величины лага L должна быть известна заранее. При ее определении лучше исходить из минимально возможного лага, чем ограничивается лагами небольшой длинны. Выбор меньшего лага, чем его реальное значение приведет к тому, что в модели регрессии их будет учтен фактор, оказывающий значительное влияние на результат, т.е. к неверной спецификации (структурной идентификации) модели;

-будут соблюдаться условия Гаусса – Маркова: полученные оценки окажутся не эффективными и смещенными;

-при выборе большого лага в сравнении с реальными – оценки будут не смещенными, эффективность будет хуже;

-необходимо установить степень полинома «m». Обычно ограничиваются рассматривая полином 2-ой и 3-ей степени.

Достоинства:

- универсальность (разнообразность структур лага);
- можно строить модели с распределенным лагом любой длинны;

Задачи формирования инструментария моделирования и прогнозирования

- сравнить современные аналитические и численные методы решения ДУ: Левенберга - Марквардта, Фишера, Готеллинга, Родса, Нейра, трех сумм и др.;
- использовать обобщенные параметрические модели авторегрессии-скользящего среднего;
- реализовать и оптимизировать и генетическое моделирование, нейросети, инструментальную среду *R*;
- использовать методологию оценки точности моделирования и прогнозирования методом Монте-Карло (до сотен тысяч выборок) для оценки области применения моделей, точностных характеристик параметров моделей, доверительных интервалов нелинейных моделей;
- предполагать мультимодельный мониторинг эволюции траекторий, итерационную параметрическую декомпозиция тренд-колебательных рядов;
- применять «бутстреп» для размножения относительно коротких выборок.

Актуальные области моделирования и прогнозирования моделями, рассматриваемыми в рамках курса

- динамика добычи невосполняемых ресурсов: нефти, газа, угля, золота (во всем мире, в отдельных странах мира, в регионах отдельных стран, на отдельных месторождениях многих стран с рекомендациями по частотам применения тех или иных моделей, анализом эффективности управленческих и технологических инноваций, мониторингом эволюции многокомпонентных моделей динамики траекторий, оценками точности прогнозирования;
- анализ динамики потребления товаров на макрорынках: ИТ-технологий, отдельных гаджетов, динамики энергопотребления;
- ценовая динамика первичного и вторичного жилья, объемов валового сбора зерна, цен на бензин в различных странах, статистики поисковых запросов на товары и т.д.;
- социальная-экономическая динамика: безработица, прогноз динамики населения, миграции, заболеваний, маркетинговые задачи, ВВП , ВРП и т. д.

Компьютерно-интенсивные методы моделирования (рандомизация, бутстреп и методы Монте-Карло)

- Семейство процедур Монте-Карло (метода статистических испытаний) – многократная генерация случайных выборок и статистический анализ откликов (параметров и критериев точности моделирования и прогнозирования) объекта анализа. Количество итераций: увеличение точности в 10 раз требует увеличения количества генераций примерно в 100 раз.
- Первоначально процедуры предлагались в математике (с 1777г. Бюффон) для расчета интегралов (особенно кратных).
- Эконометрика – оценка области применения тех или иных моделей экономической динамики (пространственной и временной). Назначаются диапазоны изменения параметров моделей (их динамические диапазоны) и сравниваются отношения дисперсий модели и назначаемой (актуальной для практики) стохастической компоненты помехи (например, от нуля – до 30%).
- Проблема генерации стохастических выборок с заданными свойствами: вид распределения, обычно, нормальный или равномерный. Рассматриваются структуры взаимодействия с детерминированной компонентой: аддитивный или мультипликативный, а центрированность и нормированность выборок формировались для задания нужного соотношения дисперсий.

Идея бутстрепа (бутстрапа) по Б. Эфрону (1979г.)

Приближенную оценку статистик стохастической компоненты, доверительных интервалов параметров и регрессий на относительно коротких выборках при отсутствии эволюции параметров и структур моделей позволяет достичь формируемый на основе метода Монте-Карло бутстреп. Он многократно генерирует повторные выборки из исходной, создавая тем самым альтернативу асимптотическим приближениям, позволяя обходиться без сложных аналитических выводов и мифических предположений о нормальности стохастической компоненты. Например, для выборки из n наблюдений последовательными итерациями с помощью датчика случайных чисел, равномерно распределенных на интервале $[1,n]$, «вытягивается» произвольный элемент X_k , который «возвращается» в исходную выборку, заменяя какое-то реальное наблюдение, формируя тем самым «псевдовыборку», которая «удлиняет» при суммировании с ней имеющуюся исходную выборку.

Многократное повторение этих действий позволяет как бы существенно «удлинить» исходную выборку и с большей точностью получить моментные статистики и доверительные интервалы (аналог эргодической теоремы в теории случайных процессов).

- На основе только исходной выборки, мы всегда можем получить бутстреп-модели генеральных распределений и для нулевой и для альтернативной гипотезы, и после этого вычислить ошибки первого и второго рода для любого выбранного нами порогового значения.

Имеется много вариантов формирования псевдовыборок выборок на основе этой идеи , например, формируя их не из наблюдений траектории, а из невязок – разностей между модельными и реальными значениями.

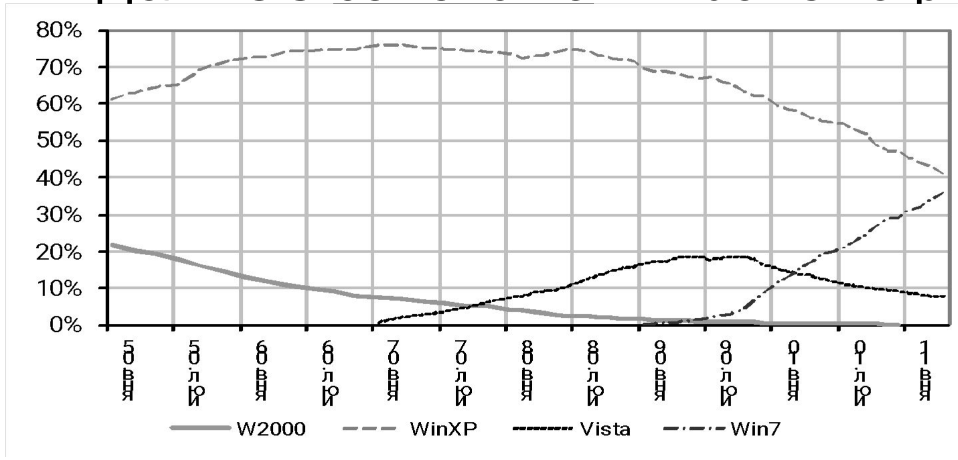
Этимология многих эконометрических терминов достаточно необычна: бутстреп - «шнурки от ботинок скаута» или «как бы из Мюнхгаузена», «метод складного ножа», «метод отжига», «метод гусеницы», что может пояснить приход «аспиранта к математику-профессору» - недостаток его воображения ведет скорее к поэзии, не позволяя заниматься математикой.

Примеры применения методов моделирования и прогнозирования в экономической практике

Примеры приложений дробно-рациональных моделей на примерах ЖЦ

IT-технологий

Доли ОС семейства Windows на рынке



$$Y_k = T_k + \varepsilon_k$$

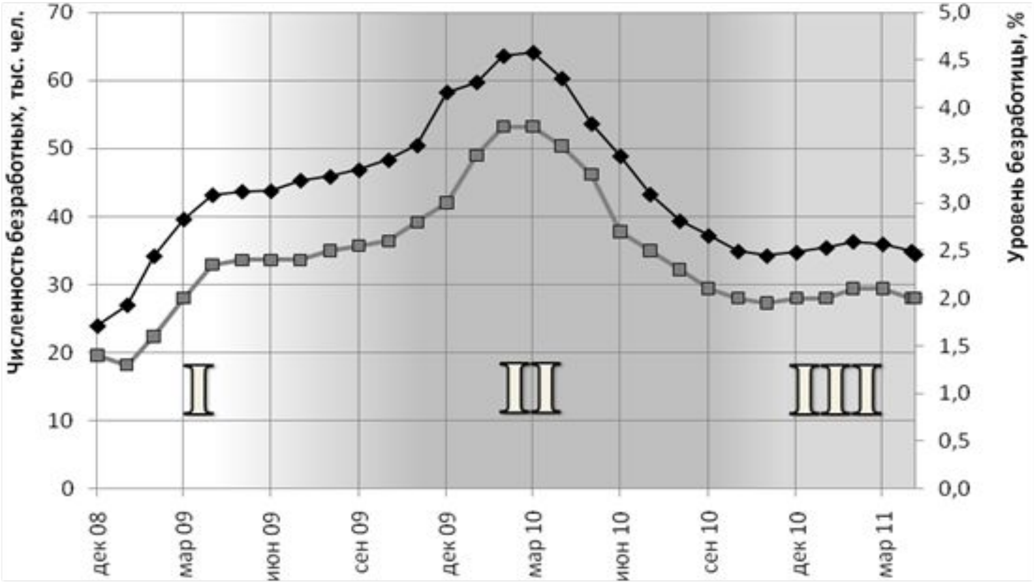
$$T_k = \frac{\alpha A(k\Delta - B) + C}{1 + A(k\Delta - B)^2}$$

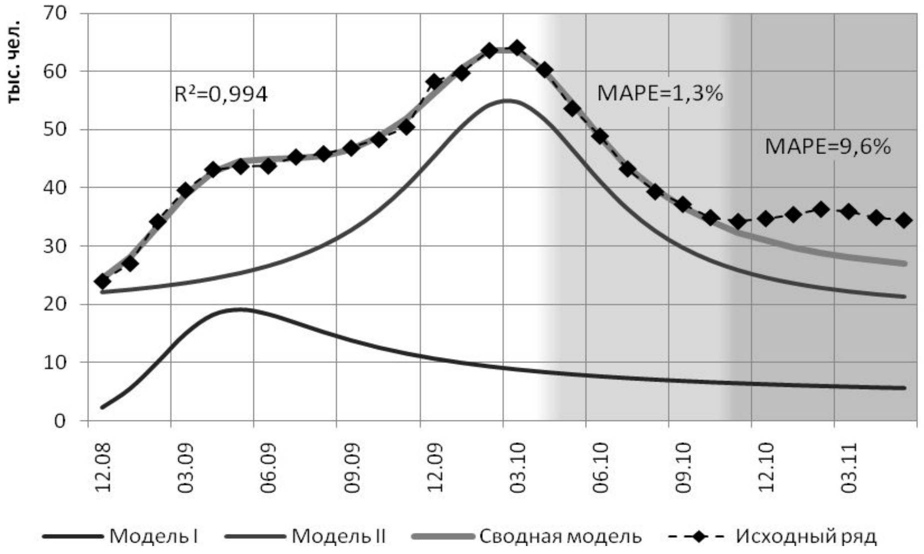
Моделирование жизненного цикла ОС Windows XP

Моделирование жизненного цикла ОС Windows 2000 $(R^2 = 0{,}999,\ K_{T2} = 0{,}061)$.

Примеры применения сумм дробно-рациональных моделей для безработицы в Самарской области

$$Y_t^* = Y_t^I + Y_t^{II} + Y_t^{III}, \ R^2 = 0{,}995.$$





$R^2 = 0{,}994$ $\qquad$ $MAPE = 1{,}25\%$

Результаты моделирования численности безработных в Самарской области

$$Y_k^* = Y_k^I + Y_k^{II} = \frac{1463 + 2328k\Delta + 199(k\Delta)^2}{1 - 0{,}27k\Delta + 0{,}053(k\Delta)^2} + \frac{22250 + 2339k\Delta + 76(k\Delta)^2}{1 - 0{,}12k\Delta + 0{,}004(k\Delta)^2} + \varepsilon_k,$$

ЖЦП Electronic Arts (EA), разработчик компьютерных видеоигр

$$Y_k = \frac{A_0}{1 + A_1 e^{-\alpha k\Delta}} + C_0 + C_1 k\Delta + \varepsilon_k$$

	NFS Pro Street	NFS Undercover
Модель	$Y_k = \frac{1,04}{1+83437\,e^{-1,6k\Delta}} - 0,06 + {} +0,08k\Delta + \varepsilon_k$	$Y_k = \frac{1,3}{1+14534\,e^{-1,9k\Delta}} - 0,11 + {} +0,13k\Delta + \varepsilon_k$
R^2	0,98	0,99
k_{T2}, %	2,75	1,72

ЖЦП Electronic Arts (EA) - разработчик компьютерных видеоигр

Результаты моделирования и прогнозирования рынков сотовой связи стран мира

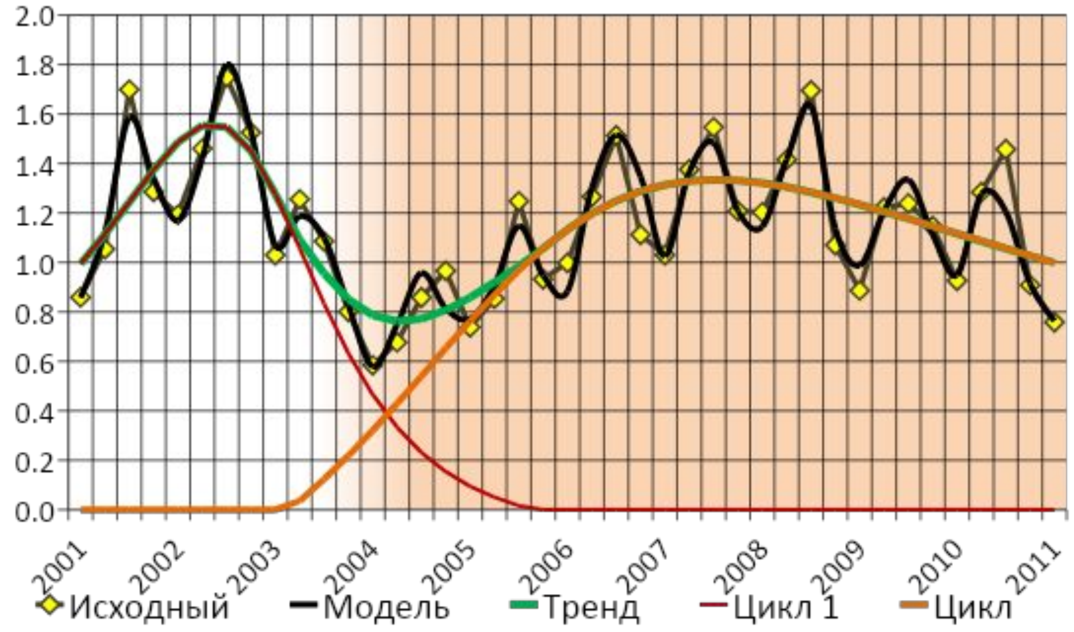
Страна	Модель	Емкость рынка, %	Точка перегиба		R^2_{adj}	k_{T2}
			Абсцисса	Ордината		
Япония	Шарифа	139,57	1995-1996гг.	16,52	0,96	3,66%
Дания	Гомпертца	136,66	1999-2000гг.	52,35	0,98	1,10%
США	Ричардса	113,16	2002-2003гг.	51,19	0,97	1,59%
Россия	Скиадаса	192,29	2004-2005гг.	55,13	0,98	0,40%
Белоруссия	Скиадаса	138,02	2004-2005гг.	34,22	0,98	1,17%
Украина	Ричардса	120,92	2005-2006гг.	77,14	0,97	1,01%

Динамика развития сотовой связи имеет логистический характер. Уровень проникновения сотовой связи в развитых странах в 2010 г. зафиксирован более 100%, темпы роста не превышают 1,5% в год.

В развивающихся странах до сих пор наблюдаются высокие темпы роста числа новых абонентов (до 20% в год), а признаки замедления роста отсутствуют.

Моделирование циклов продаж товара одного из самарских производителей: мультимодельность (повторный цикл)

$$Y_t = \left(T_t^I + T_t^{II}\right)\left(1 + S_t\right) + \varepsilon_t$$



2

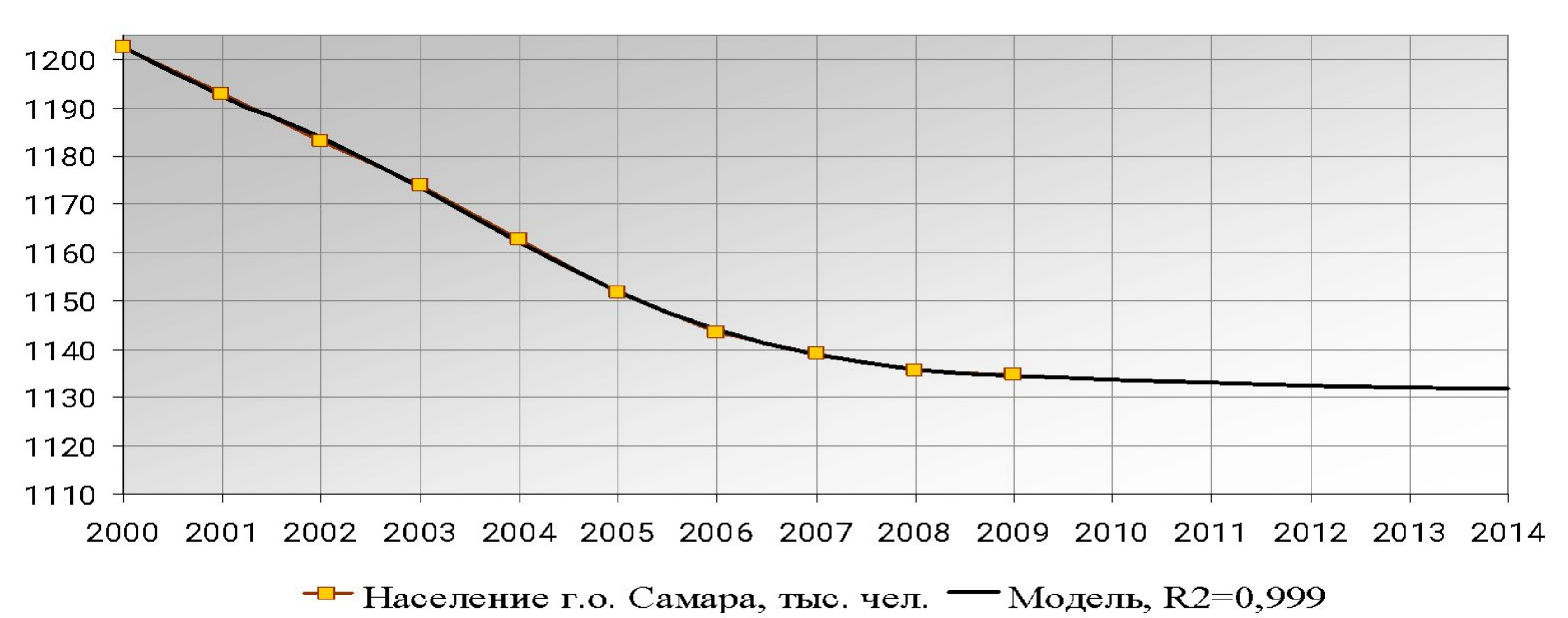
$$T_t^{II*} = \frac{-37,9 + 9,2t}{1 - 0,019t + 3,65 \cdot 10^{-4} t^2}$$

$$T_t^{I*} = \frac{326,8 - 5,86t}{1 - 0,054t + 1,24 \cdot 10^{-3} t^2}$$

$R^2 = 0,790$; $MAPE = 12,1\%$

Численность населения г.о. Самара

$$Y_k = 1130{,}5 - 129{,}7e^{-0{,}327k\Delta} + 37e^{-0{,}348k\Delta}\sin(0{,}665k\Delta - 1{,}628) + \varepsilon_k$$



Ошибка прогноза	Глубина прогноза, лет				
	1	2	3	4	5
чел.	450	389	363	427	392
%	0,04	0,034	0,032	0,037	0,034

Мониторинг цен на бензин

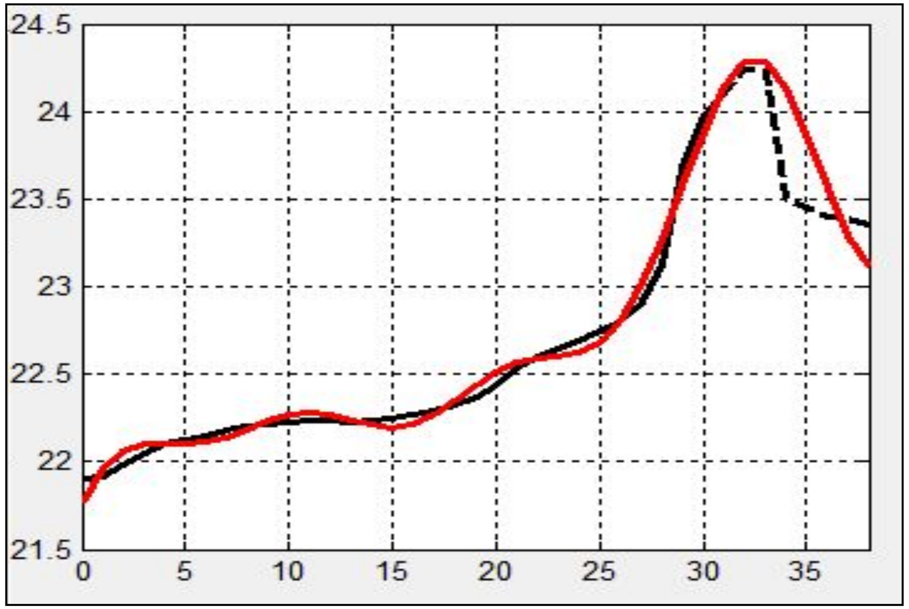
Полиномиальная модель с аддитивно-мультипликативными коле комп

$$Y_k = T_k(1 + S_k^M) + S_k^A$$

$$T_k = \sum_{i=0}^{N_T} \left[D_{i+1} \cdot (k\Delta)^{N_T - i}\right]$$

$$S_k^M = \sum_{q=1}^{N_M} \left[A_q \cdot \sin(\omega_q \cdot k\Delta) + B_q \cdot \cos(\omega_q \cdot k\Delta)\right],$$

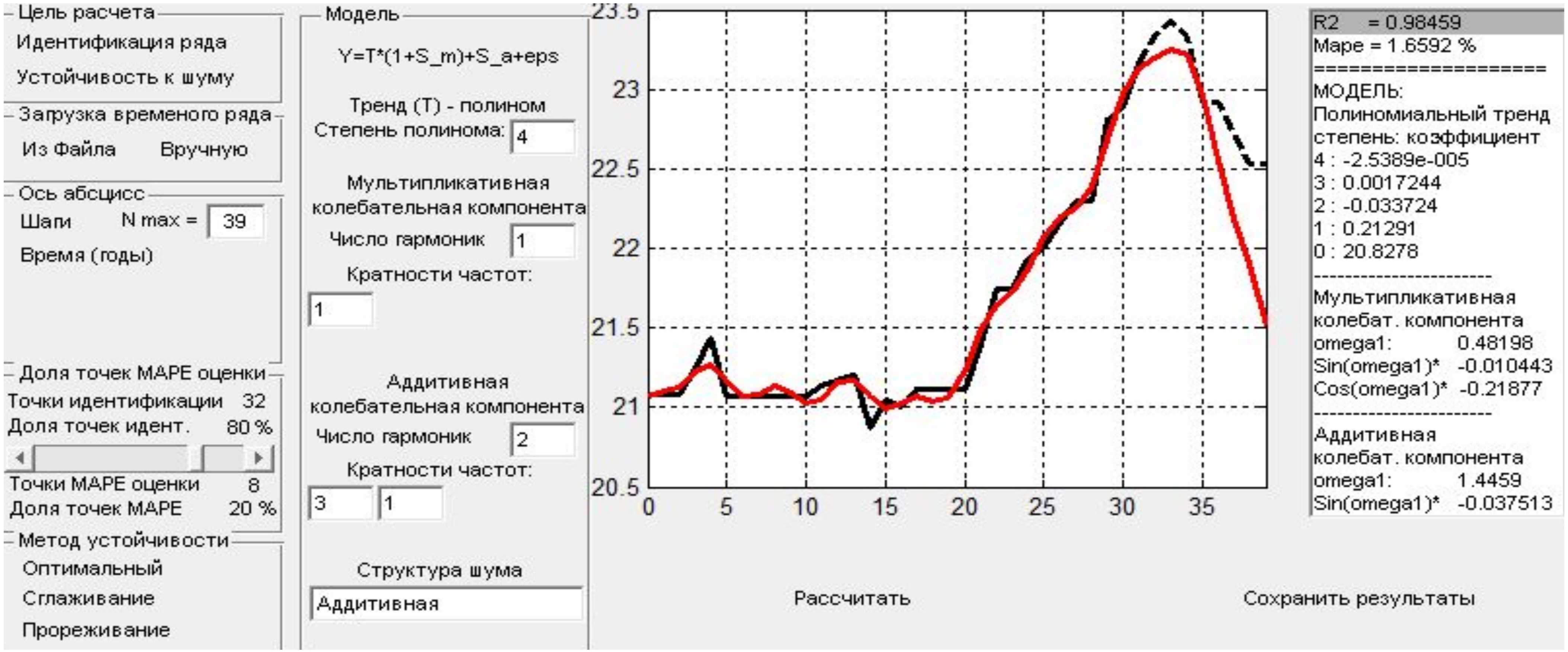
$$S_k^A = \sum_{r=1}^{N_A} \left[E_r \cdot \sin(\Omega_r \cdot k\Delta) + F_r \cdot \cos(\Omega_r \cdot k\Delta)\right].$$



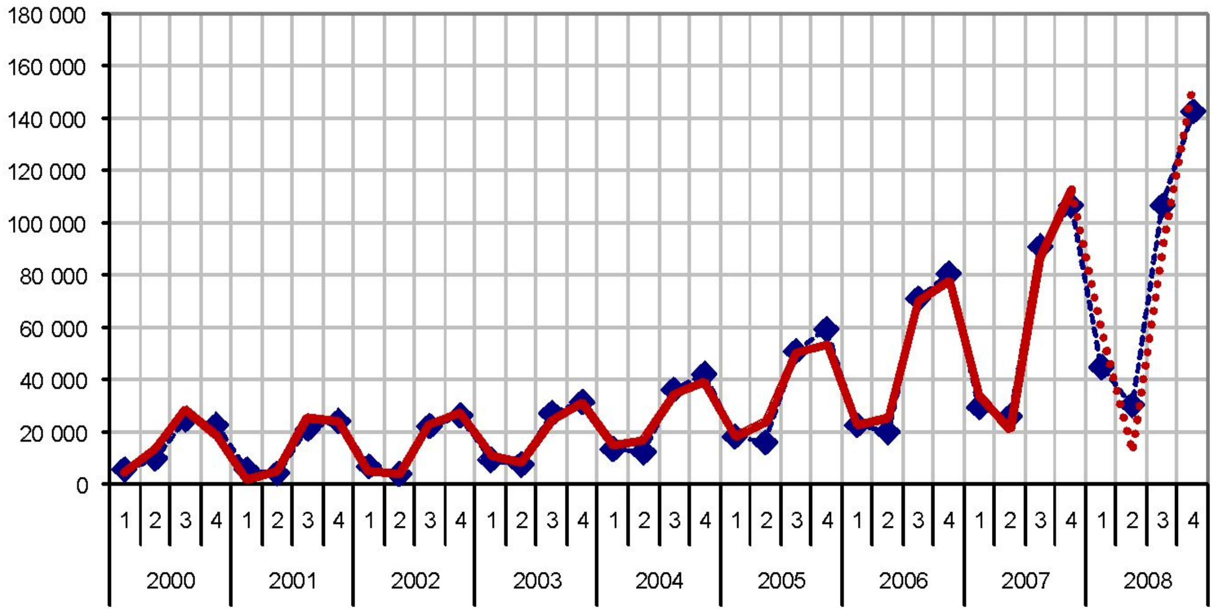
Динамика изменения цен на бензин на территории РФ, руб.

$R^2 = 0{,}98$, $MAPE = 1\%$

Динамики цен на бензин на территории Самарской области (руб./л)



Инвестиции в основной капитал Самарской области



$$R^2 = 0,98, \qquad MAPE = 19\%$$

В структуре модели инвестиций присутствуют три колебательные компоненты, пропорциональные тренду, и две колебательные аддитивные компоненты.

Кроме того, можно сделать вывод и о том, что горизонт прогноза взят излишне большой, не обеспечивая высокую точность в связи с эволюцией амплитуды колебательной компоненты (ее надо изменить!).

$$\boldsymbol{Y_k} = \left(\boldsymbol{D_1(k\Delta)^2} + \boldsymbol{D_2 k\Delta} + \boldsymbol{D_3}\right)\cdot$$

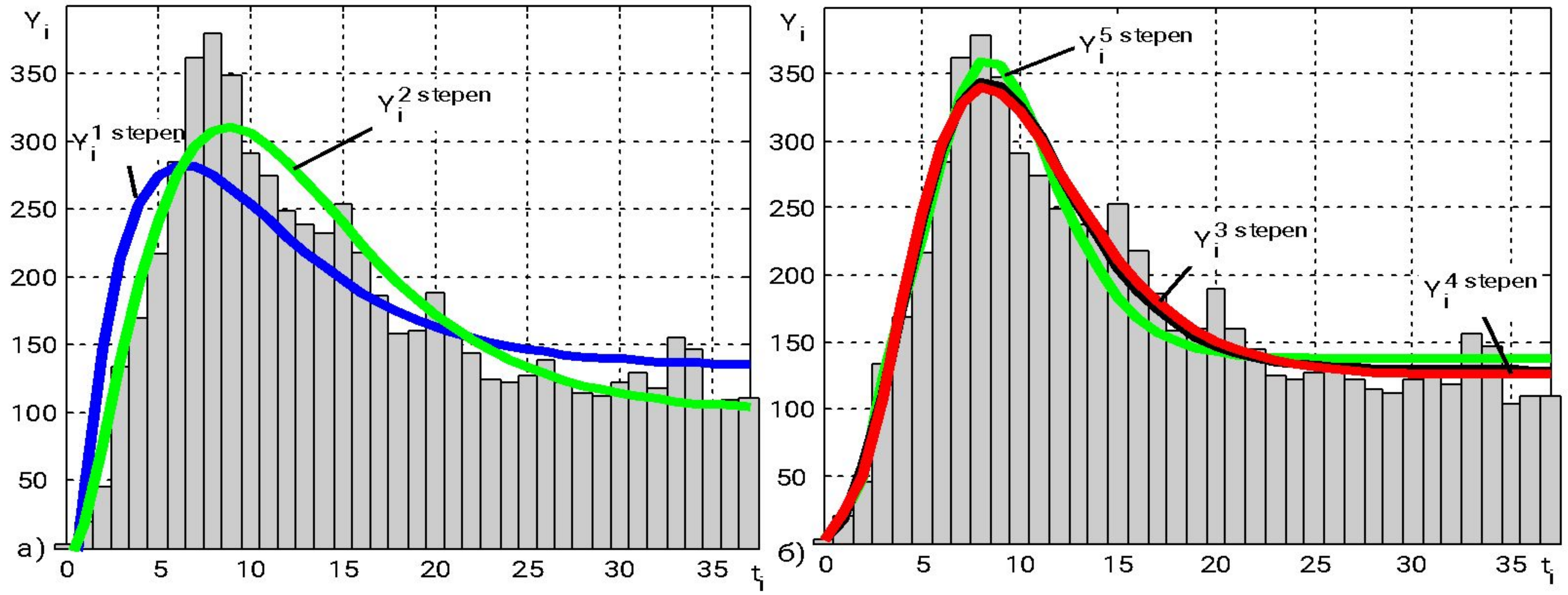
$$\cdot\left(1 + A_1\sin(\omega k\Delta) + B_1\cos(\omega k\Delta) + A_2\sin(2\omega k\Delta) + B_2\cos(2\omega\right.$$

$$+ E_1\sin(\omega k\Delta) + F_1\cos(\omega k\Delta) + \boldsymbol{\varepsilon_k}.$$

Модели добычи нефти

Фильтрационные модели разработки месторождений	Феноменологические модели описания истории добычи нефти и газа
+ **Новые месторождения** Опора на полевые исследования пласта	+ **Любой масштаб объекта** Месторождение, регион добычи, государство
– **Ограничение масштаба** Невозможно построить модель добычи нефти в регионе или стране	+ **Учет всех факторов добычи** Все факторы отражаются на объеме добычи
– **Отсутствие учета человеческого и экономического факторов**	– **Желательна история добычи** (лишь рекомендации по частоте используемых моделей)

Годовая добыча нефти месторождения Самарской области функцией Рамсея



$$Y_i^{3stepen} = (20,13168t_i^3 - 53,633t_i^2 - 1,033t_i - 123,093)e^{-0,43127t_i} + 125,643 +$$

$$R^2 = 0,96$$

Колоколообразные модели (импульсные логисты) тренда добычи нефти (и других невосполняемых ресурсов)

Эмпирические (феноменологические) модели добычи:

- **Коши-Капицы**

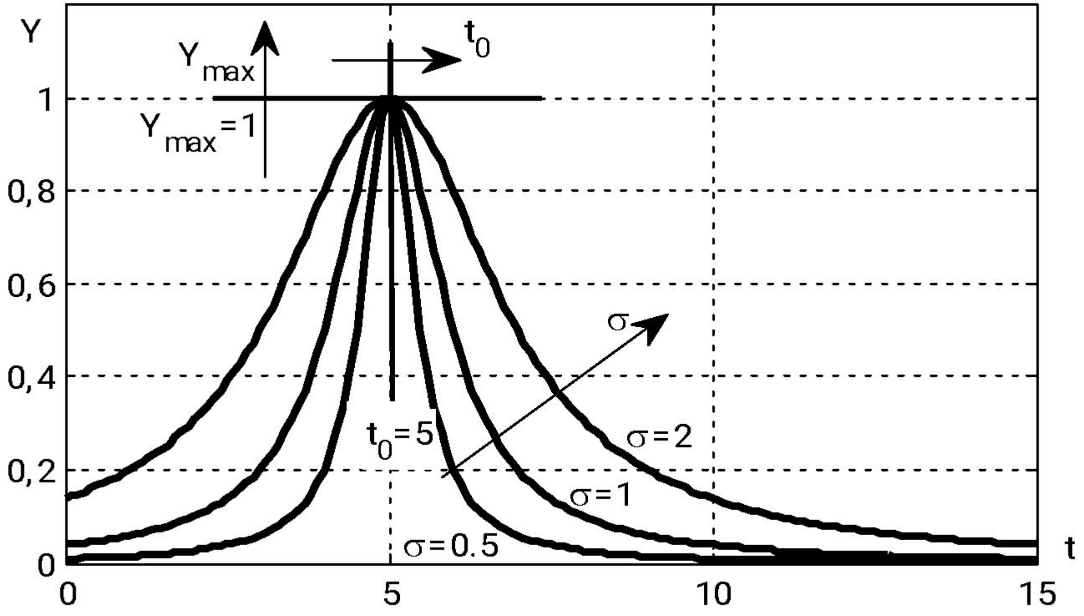
$$Y(t) = \frac{Y_{\max} \cdot \sigma^2}{(t - t_0)^2 + \sigma^2}$$

- **Гаусса**

$$Y(t) = Y_{\max} e^{-(t-t_0)^2/\sigma^2}$$

- **Хабберта**

$$Y(t) = \frac{2Y_{\max}}{1 + \operatorname{ch}\left[\sigma\left(t - t_0\right)\right]}$$



- **Логнормальная модель**

$$Y(t) = Y_{\max} \frac{t_0}{t} e^{-\frac{1}{2}\frac{\ln\frac{t_0}{t}\left(t_0^2 \cdot \ln\frac{t_0}{t} + 2\sigma^2\right)}{\sigma^2}}$$

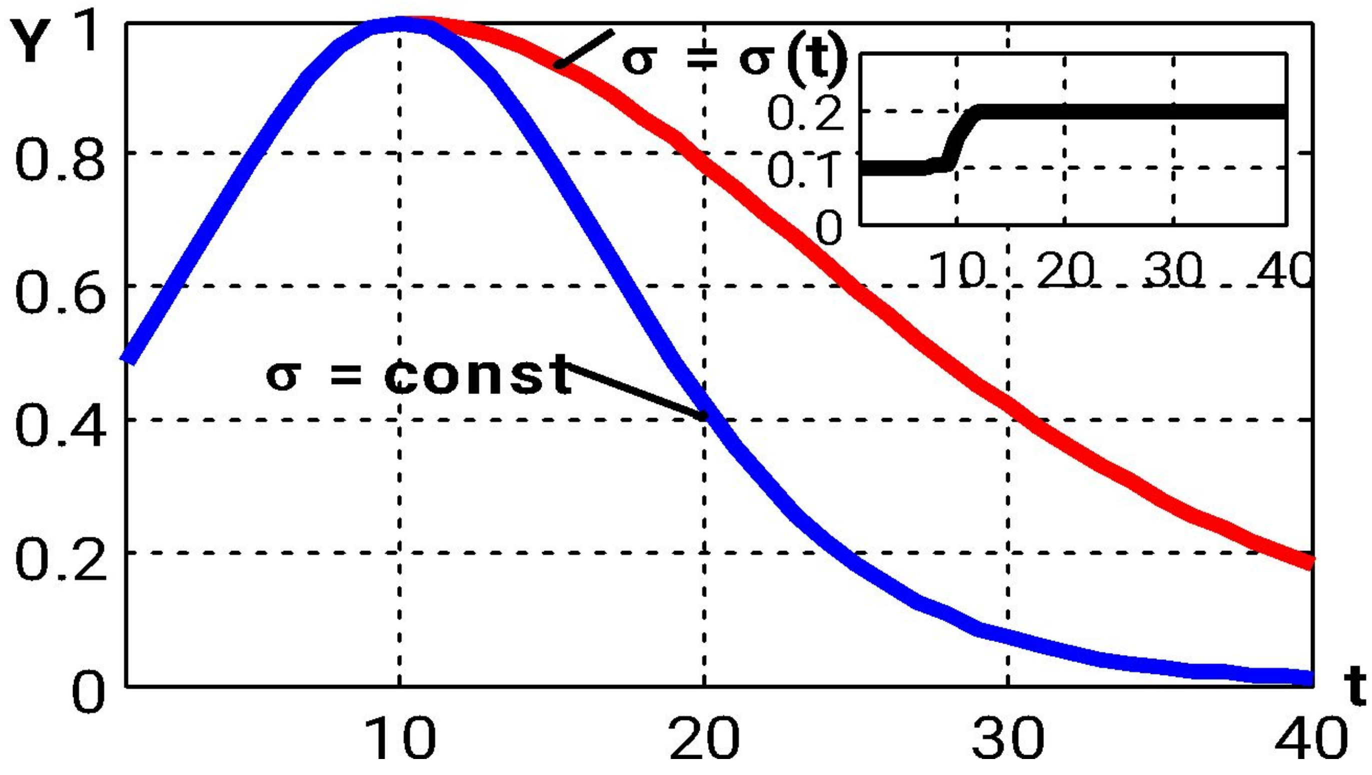
- **Хаммонда-Маккея**

$$Y(t) = \frac{Y_{\max}}{Norm} t^{\sigma(t)\cdot t_0} e^{-\sigma(t)\cdot t}$$

$$Norm = t_0^{\sigma(t)\cdot t_0} e^{-\sigma(t)\cdot t_0}$$

Задание асимметрии путем введения функции σ = σ(t)

Брандт: исследовано 67 месторождений, уровень падения в среднем на 5 % ниже уровня роста – кривая добычи нефти ассиметрична



- **Ферхюльста** $$\sigma = \sigma_1 + \frac{\sigma_2 - \sigma_1}{1 + e^{-\frac{t - t_0}{\sigma_T}}}$$

- **Ричардса** $$\sigma = \sigma_1 + \frac{\sigma_2 - \sigma_1}{\left[1 + e^{-\frac{t - t_0}{\sigma_T}}\right]^{1/\sigma_{T1}}}$$

- **Гомперца** $$\sigma = \sigma_1 + (\sigma_2 - \sigma_1) e^{-0.7 e^{-\frac{t - t_0}{\sigma_T}}}$$

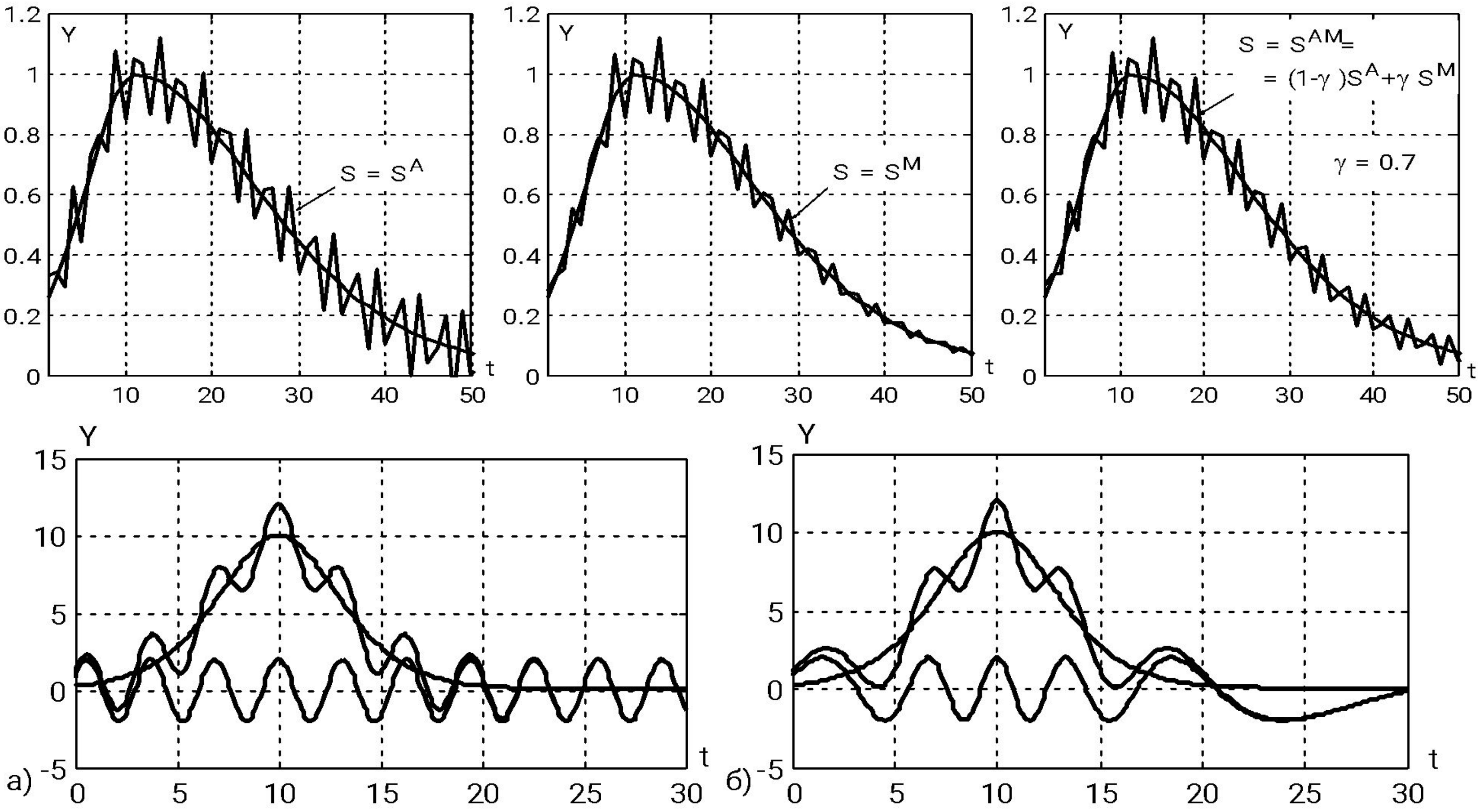
- **Рамсея** $$\sigma = \begin{cases} \sigma_1 + (\sigma_2 - \sigma_1)\left(1 + \left[1 + \dfrac{t - t_0^*}{\sigma_T}\right] e^{-\frac{t - t_0^*}{\sigma_T}}\right), t \geq t_0^* \\ \sigma_1, \qquad t < t_0^*, \qquad\qquad t_0^* = t_0 - 1,678\sigma_T \end{cases}$$

Выбор предпочтительной феноменологической модели для анализа добычи нефти и газа на отдельных месторождениях

Для выбора лучших моделей проведен представительный анализ данных добычи на 320 месторождений: 140 нефтяных месторождений ОАО НК Роснефть, 90 нефтяных и 90 газовых месторождений штата Техас.

Модель Коши показала себя лучшей при моделировании и прогнозировании добычи нефти на отдельных месторождениях в России и США в 74 – 79% случаев и не уступает лучшей модели более чем 0,9 от величины критерия в 88 – 93% случаев, что позволяет уверенно рекомендовать ее для описания добычи нефти и газа на отдельных месторождениях, как наиболее вероятную.

Учет колебательной компоненты



На сегодняшний день в мире насчитывается больше 1000 статистических пакетов. Статистические пакеты делятся на несколько категорий.

Во-первых, коммерческие (платные) и находящиеся в свободном доступе (бесплатные).

Во-вторых, универсальные и специализированные.

В-третьих, интерактивные и предлагающие пакетную обработку.

В-четвертых, закрытые и настраиваемые.

Наиболее известны зарубежные статистические пакеты: SAS, SPSS, STATISTICA, R, S-plus и т.п.

Прежде чем получить результат, определите, что Вы с ним будете делать...

Спасибо за внимание!