



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

ПРЕДСКАЗАТЕЛЬНАЯ АНАЛИТИКА НА ОСНОВЕ ПОТОКОВ БОЛЬШИХ ДАННЫХ

Дмитриев Андрей Викторович
Факультет бизнес-информатики

Предсказательная аналитика – техника прогнозирования будущего поведения объектов и субъектов с целью принятия оптимальных решений с использованием статистических методов анализа данных, машинного обучения и интеллектуального анализа данных.

Charles N. (2007), Predictive Analytics White Paper, American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America.

Некоторые приложения предсказательной аналитики

Принятие решения по сделкам. Предсказательная аналитика использует паттерны, найденные в исторических и транзакционных данных, для анализа и прогноза рисков.

Системы поддержки принятия решений. В здравоохранении, методы предсказательной аналитики позволяют определить группы пациентов, подверженных риску развития определенных заболеваний.

Выявление мошенничества – кража личности, ложные страховые претензии, налоговые мошенничества, ...

Андеррайтинг – в банковском деле процедура оценки банком вероятности погашения или непогашения запрашиваемого кредита. Методы предсказательной аналитики используются для изучения платёжеспособности и кредитоспособности потенциального заёмщика.



Регрессионные методы

- линейная регрессия
- логистическая и множественная логистическая регрессия
- пробит-регрессия
- временные ряды
- классификации и регрессионные деревья
- анализ выживаемости
- многомерные адаптивные регрессионные сплайны

Методы машинного обучения

- искусственные нейронные сети
- многослойный перцептрон
- радиальные базисные функции
- метод опорных векторов
- наивный байесовский классификатор
- метод k ближайших соседей
- геопространственное предсказательное моделирование
- генетический алгоритм



Временные ряды

Временной ряд – динамический ряд последовательных значений, характеризующих изменение показателя во времени:

$$\{\vartheta_i\}_{i=1}^N = \{\vartheta_1, \vartheta_2, \dots, \vartheta_N\} \quad \vartheta_i = \vartheta(t_i), t_i = i\Delta t$$

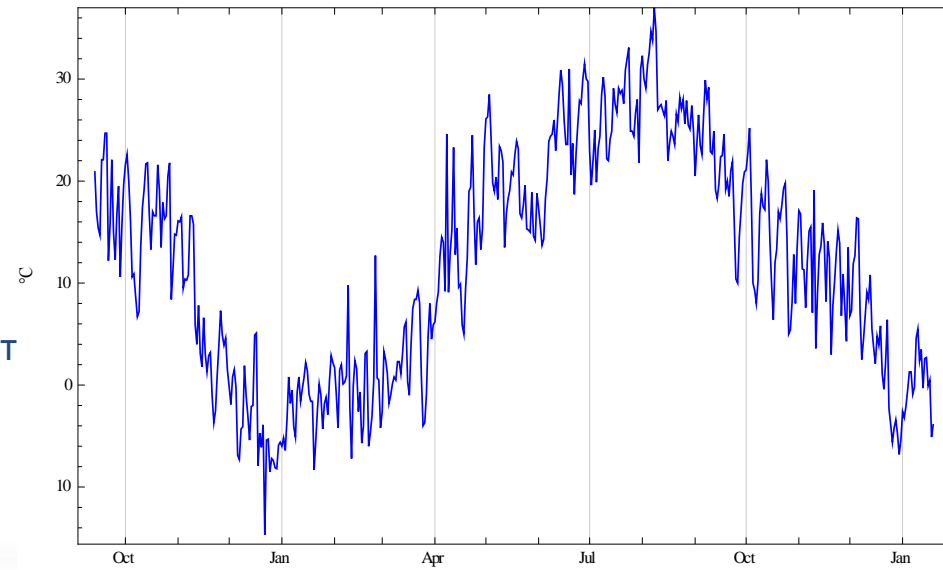
Классификация временных рядов

- **моментные ряды** (данные которых характеризуют величину явления по состоянию на определенные даты)
- **интервальные ряды** (характеризующие определенные периоды)
- **эволюторные процессы**, содержащие тренд
- **стационарные процессы**, не содержащие тренда

Методы анализа и прогнозирования временных рядов

- модель авторегрессии - скользящего среднего *ARMA*
- интегрированная модель авторегрессии - скользящего среднего *ARIMA*
- авторегрессия с условной дисперсией, зависящей от некоторой случайной величины *ARCH*
- обобщенная *ARCH*-модель *GARCH*

Daily maximum temperature , London , Ontario , 2000 [09]13 to 2002 [01]19



Задача реконструкции

по наблюдаемым значениям временного ряда оценить компоненты вектора параметров динамической системы

$$\vartheta_i = \vartheta(t_i) \Rightarrow \dot{x} = F(x, c)$$

Два постулата реконструкции

- в основе реального процесса лежит некоторая регулярная динамика
- агрегированный временной ряд содержит динамическую и стохастическую составляющие

Три задачи реконструкции

Прозрачный ящик. Структура (вид и число уравнений системы, явный вид F) модели исследуемого объекта полностью известны. Задача состоит только в оценке неизвестных параметров по наблюдаемым данным.

Серый ящик. О структуре модели известно почти все за исключением некоторых компонент функции F .

Черный ящик. О структуре модели заранее ничего не известно.

Основные подходы к решению задач реконструкции

- метод наименьших квадратов
- полный МНК при больших шумах
- метод максимального правдоподобия

Нелинейная модель динамики агрегированных рыночных цен спроса и предложения

$$\begin{cases} \dot{x}_1 = -\sigma(x_1 - x_2) \\ \dot{x}_2 = \rho x_1 - x_2 - x_1 x_3 \\ \dot{x}_3 = -\beta x_3 + x_1 x_2 \end{cases}$$

Переменные

$x_1 = [X_1(t) - P_0]$ - вариация агрегированных цен спроса

$x_2 = [X_2(t) - P_0]$ - вариация агрегированных цен предложения

$x_3 = E_{12}(t)$ - эластичность x_1 по x_2

Параметры

$\sigma = \frac{\tau_1}{\tau_2}$ - отношение времен релаксации

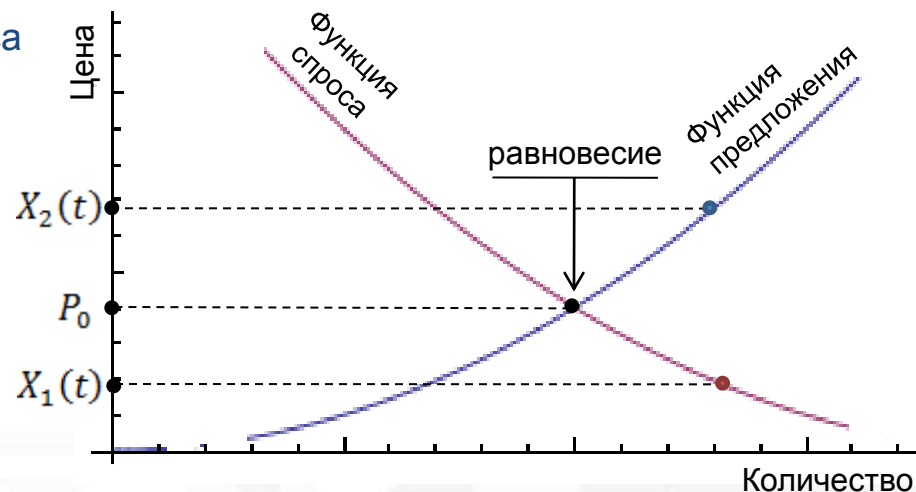
$\beta = \frac{\tau_1}{\tau_E}$ - отношение времен релаксации

τ_1 - характерное время релаксации $X_1 \rightarrow P_0$

τ_2 - характерное время релаксации $X_2 \rightarrow P_0$

τ_E - характерное время релаксации $E_{12} \rightarrow E_{12}^{(0)}$

$\rho = \frac{E_{12}^{(0)}}{\lambda}$ - масштабированная эластичность в состоянии равновесия системы



См. <http://www.hse.ru/org/persons/98405261>

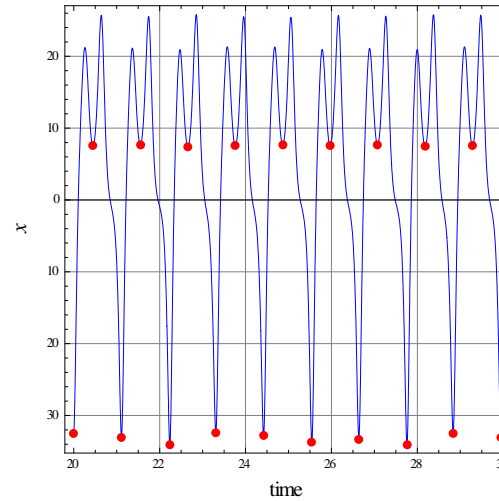


Предхаотические состояния в динамической системе цен спроса и предложения

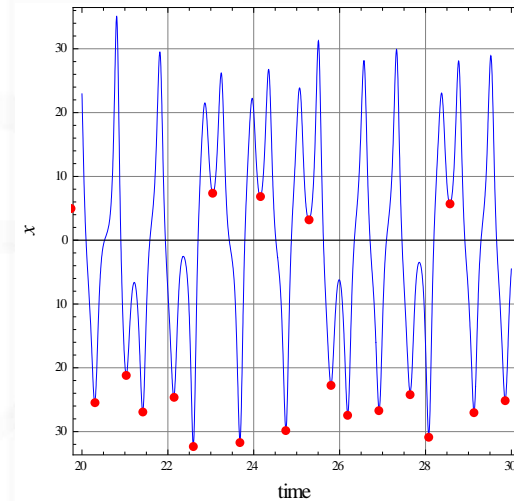
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Переход к хаосу через удвоение периода

$99 < \rho < 100.3$

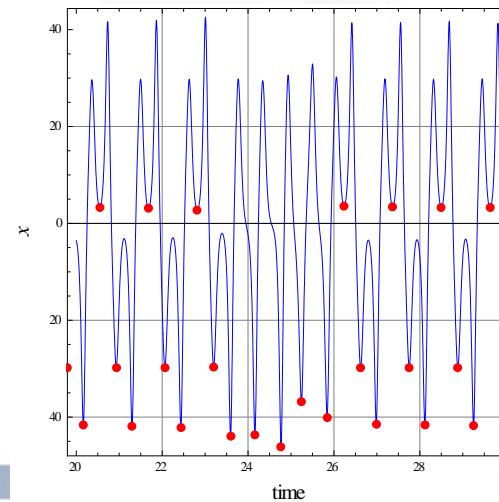


$\rho = 100.4$

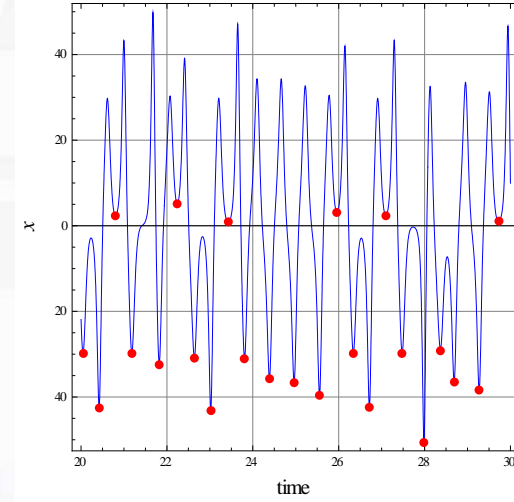


Переход к хаосу через перемежаемость

$166.1 < \rho < 166.45$



$\rho = 166.5$





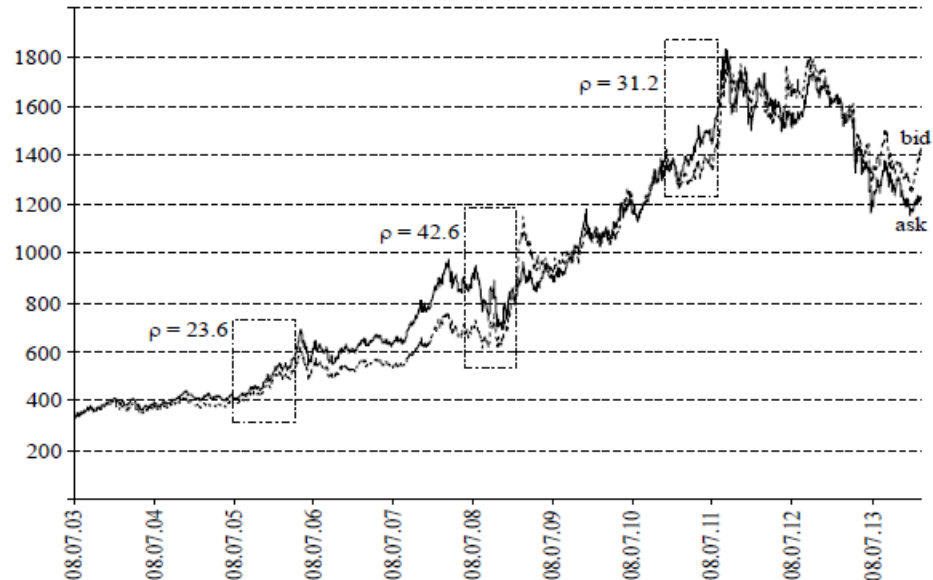
Реконструкция динамической системы агрегированных цен спроса и предложения по наблюдаемым временным рядам

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Наблюдаемые агрегированные данные для реконструкции системы цены спроса и предложения на

- драгоценные металлы
- обменный курс EUR/USD

- Установлены**
- интервалы значительной стохастичности, предшествующие кризисным состояниям на рынке драгоценных металлов
 - интервалы значительной стохастичности, предшествующие кризисным состояниям на рынке иностранной валюты



Технология реализации модели

- наблюдаемые данные брали из Bloomberg Terminal
- разработали прототип программы на основе Open API Bloomberg с использованием кода C++
- вход – потоки данных цен спроса и предложения в режиме реального времени
- выход – индикация предкризисного режима на рынке



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Спасибо за внимание!

101000, Россия, Москва, Мясницкая ул., д. 20

Тел.: (495) 621-7983, факс: (495) 628-7931

www.hse.ru