

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ
УНИВЕРСИТЕТ «СТАНКИН»**

**ИНСТИТУТ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ
РОССИЙСКОЙ АКАДЕМИИ НАУК**

***ФУНДАМЕНТАЛЬНЫЕ
ФИЗИКО-МАТЕМАТИЧЕСКИЕ
ПРОБЛЕМЫ
И
МОДЕЛИРОВАНИЕ
ТЕХНИКО-ТЕХНОЛОГИЧЕСКИХ
СИСТЕМ***

*Ежегодный сборник научных трудов
Выпуск 12*

*Материалы Международной научной конференции
«Моделирование нелинейных процессов и систем»
Том 2*

**Москва
Янус-К
2009**

НЕЛИНЕЙНЫЕ РОБАСТНЫЕ СХЕМЫ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ ПРОЦЕССОВ

В.В. Иванов¹, А.В. Крянев², Г.В. Лукин², Д.К. Удумян²

¹ Лаборатория информационных технологий, Объединенный институт ядерных исследований (Дубна)

² Московский инженерно-физический институт (государственный университет)

Выделение трендовой компоненты является одним из основных этапов анализа временного процесса (ряда), обычно предшествующего его прогнозированию, и позволяет изучать глобальное поведение исследуемого процесса, отвечающего анализируемому временному ряду [1,2]. Основной целью настоящей работы является применение эффективных нелинейных робастных методов выделения трендовой, аномальной и хаотической компонент анализируемого временного ряда и методов прогнозирования с возможностью учета априорной экспертной информации.

Одной из основных задач прикладного анализа является исследование динамики исследуемых процессов и, в частности, их прогнозирование. Трендовая компонента отражает изменения временного ряда, обусловленные влиянием различных факторов или причин. Для многих временных рядов прикладного характера причины, порождающие их трендовые компоненты, могут быть не достаточно ясными. Однако их совокупное влияние может быть устойчивым в течение относительно длительных промежутков времени. Последнее обстоятельство обеспечивает возможность прогноза временных рядов.

В то же время при анализе временных рядов довольно часто необходимо учитывать и главное уметь выделять аномальные составляющие, представляющие собой кратковременные выбросы большой амплитуды по сравнению с амплитудами обычных хаотических флуктуаций. Для временных процессов прикладного характера аномальным компонентам соответствуют кризисные состояния исследуемого объекта. Разделение исходного временного ряда на трендовую, аномальную и хаотическую компоненты облегчает задачу исследования временного ряда и его прогноза. Более того, традиционные (не робастные) методы выделения трендовой компоненты не обладают устойчивостью к кратковременным выбросам большой амплитуды, что приводит к нежелательному искажению трендовой компоненты, а также не

позволяют выделять аномальную составляющую, которая часто представляет самостоятельный интерес.

Таким образом, будем рассматривать представления исследуемого временного процесса $x(t)$ в виде трех слагаемых:

$$x(t) = y_{tr}(t) + y_{an}(t) + y_{ch}(t) \quad (1)$$

где $y_{tr}(t), y_{an}(t), y_{ch}(t)$ - соответственно, трендовая, аномальная и хаотическая компоненты.

Сформулируем основные требования к методам выделения трендовой составляющей $y_{tr}(t)$:

- * устойчивость к присутствию кратковременных выбросов большой амплитуды с возможностью выделения аномальной составляющей;
- * возможность использования нелинейной полиномиальной модели тренда с высокой степенью нелинейности;
- * наличие эффективных и устойчивых численных методов, реализующих предлагаемую схему выделения трендовой компоненты.

Проблема выделения аномальной компоненты решается с помощью робастной методики. В настоящей работе использована система ортогональных робастных полиномов, что позволило восстанавливать нелинейную полиномиальную модель трендовой компоненты любого порядка, не прибегая к операции обращения информационной матрицы. Тем самым устраняется возможная вычислительная неустойчивость рассматриваемой задачи [2,3].

Пусть известны n значений $x(t_i) = x_i, i = 1, \dots, n$ исследуемого временного процесса $x(t)$.

Представим трендовую компоненту в виде разложения

$$y_{tr}(t) = \sum_{k=0}^m u_k \cdot \Phi_k(t) \quad (2)$$

где u_k - искомые коэффициенты разложения, $\Phi_k(t)$ - ортогональные полиномы степени k .

Систему ортонормированных полиномов можно построить, используя, например, рекуррентную схему Форсайта [2,3].

Традиционные схемы метода наименьших квадратов не позволяют выделить все три компоненты представления (1) и не удовлетворяют сформулированным выше требованиям.

В отличие от традиционных схем метода наименьших квадратов будем искать коэффициенты разложения $u_k, k = 0, \dots, m$ как решение следующей экстремальной задачи

$$\vec{u} = \text{Arg} \min_{\vec{u}} \sum_{i=1}^n \rho \left(\frac{x(t_i) - y_{tr}(t_i)}{\sigma_i} \right) \quad (3)$$

где σ_i^2 - дисперсии хаотической компоненты, $i=1, \dots, n$, а $\rho(s)$ - функция Хьюбера [3,4].

Нами разработана устойчивая схема численного решения задачи (3), позволяющая выделить все три компоненты представления (1) и удовлетворяющая сформулированным выше необходимым требованиям.

Другая разработанная нами робастная схема выделения трех компонент представления (1) используют кубические сплайны и основана на минимизации функции:

$$J_{\alpha j}(S_{\alpha j}(t)) = \sum_{i=1}^n \rho\left(\frac{y_{j,i} - S_{j,i}}{\sigma_j}\right) + \alpha \sum_{i=2}^{n-1} (S_{j,i+1} - 2S_{j,i} + S_{j,i-1})^2, \quad (4)$$

где $\rho(s)$ - робастная функция Тьюки, $S_{j,i} = S_j(t_i) = y_{jdet}(t_i)$ - значения робастного кубического сплайна, $\alpha > 0$ - сглаживающий параметр (параметр регуляризации) [5].

Для нахождения искомых значений робастных кубических сплайнов S_{ji} , определяемых экстремальной задачей (4), нами разработан итерационный процесс относительно векторов вторых производных [6].

На рисунках 1,2 представлены результаты выделения трендовых компонент из исходных временных рядов с помощью изложенных выше схем, использующих робастную функцию Тьюки. Из рисунков 1,2 видно, что разработанные нами робастные методы позволяют эффективно выделять трендовые компоненты исследуемых временных рядов.

Часто в распоряжении аналитика наряду со статистическими (историческими) данными имеется дополнительная информация, носящая, как правило, характер априорных экспертных оценок. Одной из актуальных задач анализа является разработка методов, позволяющих прогнозировать исследуемый временной процесс с учетом априорных экспертных оценок, что в общем случае, позволит улучшить точность прогноза. Разработанные нами схемы прогнозирования используют априорные экспертные оценки, задаваемые в виде совокупности пар

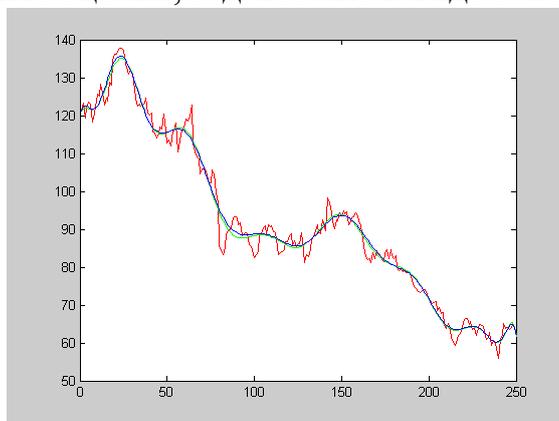


Рис.1. Выделенные тренды разного уровня сглаживания (синяя и зеленая линии). Красная линия соответствует исходному временному ряду

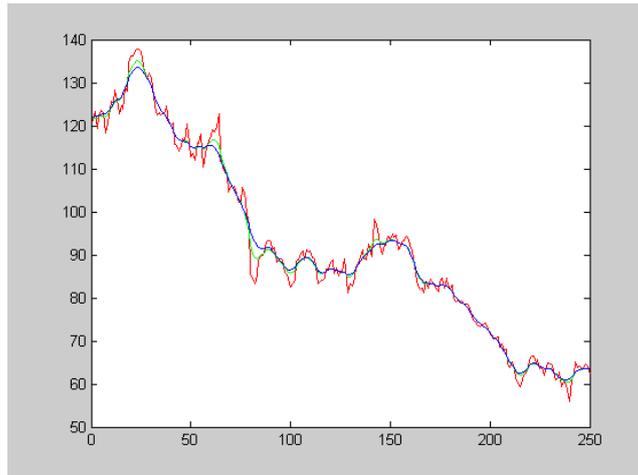


Рис.2. Выделенные тренды разного уровня сглаживания (синяя и зеленая линии). Красная линия соответствует исходному временному ряду

$$\{x_{\text{exp}}(t_i), \sigma_i^2, i = n + 1, \dots, n + L\} \quad (5)$$

где $x_{\text{exp}}(t_i)$ - наиболее вероятная экспертная оценка значения исследуемого временного ряда на будущую дату t_i , σ_i^2 - дисперсия погрешности экспертной оценки, L - горизонт прогнозирования.

Схема прогнозирования основана на выделении из временного ряда

$$x(t_{n-1}), x(t_{n-l+1}), \dots, x(t_n), x_{\text{exp}}(t_{n+1}), \dots, x(t_{n+L}) \quad (6)$$

детерминированной трендовой компоненты с помощью робастных ортогональных полиномов. В качестве прогнозируемых значений $x_{\text{for}}(t_{n+i}), i = 1, \dots, L$ берутся соответствующие значения $x_{\text{tr}}(t_{n+i})$ трендовой компоненты $x_{\text{tr}}(t)$, выделенной с помощью описанных выше схем.

Для выделения трендовых компонент и прогнозирования временных процессов можно использовать также сингулярно-спектральный анализ, основанный на представлении тренда в виде разложения по главным компонентам [2,6,7].

На рисунке 3 представлены результаты прогнозирования на 2 года вперед месячных значений характеристики макроэкономики Российской Федерации. Красным цветом выделены прогнозные значения, полученные с помощью сингулярно-спектрального анализа.

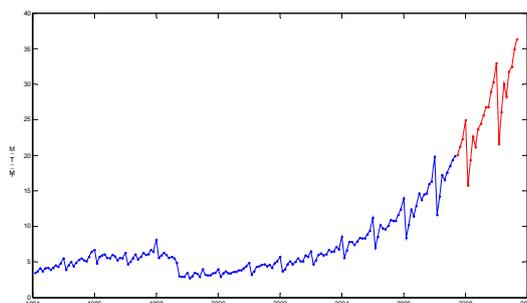


Рис.3. Суммарный импорт в период 1.01.1994-31.08.2007

Из рисунка 3 видно, что наряду с трендовой компонентой исследуемый макроэкономический показатель имеет резко выраженные сезонные колебания, которые с помощью разработанных нами схем выделяются и прогнозируются с высокой точностью. Ошибка месячного прогноза на 2 года вперед для экономического показателя, представленного на рис.3 составляет менее 1%. Как показали многократные применения разработанных нами методов, эти методы позволяют в большинстве случаев прогнозировать исследуемые временные процессы с приемлемой точностью.

Выводы

В работе новые робастные схемы выделения трендовых, аномальных и хаотических компонент и прогнозирования применяются для анализа и прогнозирования динамики процессов. Используемые методы обладают вычислительной устойчивостью и устойчивостью к большим по амплитуде кратковременным выбросам в исследуемом временном процессе. С их помощью выделяются аномальные компоненты, в том числе не просматриваемые явно, как по их положению во времени, так и по амплитуде. Это предоставляет возможность для более глубокого изучения аномальных и кризисных явлений исследуемых процессов. Разработанные нами схемы прогнозирования позволяют учесть дополнительную априорную экспертную информацию и, тем самым, получить более надежные прогнозные значения для исследуемых временных процессов.

Литература

1. Haykin S. Adaptive Filter Theory. Prentice-Hall, Englewood Cliffs. 1991.
2. Крянев А.В., Лукин Г.В. Математические методы обработки неопределенных данных. М.: Физматлит, 2006.
3. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980.
4. Хьюбер П. Робастность в статистике. М.: Мир, 1984.
5. Arsenin V.Ya., Kryanev A.V. Generalized maximum likelihood method and its application for solving ill-posed problems // Ill-Posed Problems in Natural Sciences. A. Tikhonov (Ed.). Netherland, VSP-BV. 1992.
6. Antoniou I., Akritas P., Burak D.A., Ivanov V.V., Kryanev A.V., Lukin G.V. Robust methods for stock market data analysis // Physica A 336, 2004.
7. Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure. SSA and Related Techniques. Champan & Hall/CRS, 2001.