

Сравнительный анализ различных алгоритмов скользящего усреднения с использованием критерия минимального среднего квадрата ошибок.

Тихонова О.А.

Донецкий национальный технический университет

oks_mail777@mail.ru

Abstract

The comparative analysis of classical algorithms of sliding averaging and new algorithms by criterion of the minimal average square of mistakes is made.

Введение

Мониторинг технической или экономической системы, решает задачи выделения тенденции (тренда) развития каждого из совокупности контролируемых параметров и прогнозирует моменты их завершений. В работе используется следующий вид временного ряда:

$$z(t) = x(t) + x'(t), \quad (1)$$

где $z(t)$ – временной ряд;

$x(t)$ – некоторая медленно меняющаяся функция, называемая трендом (систематическая составляющая).

$x'(t)$ – случайная (несистематическая, нерегулярная) компонента, как правило представляющая собой стационарный случайный процесс с нулевым средним.

Дальнейшая задача распадается на две независимых:

- выделение $x(t)$ из наблюдаемого $z(t)$;
- изучение (измерение, анализ, оценка) параметров $x(t)$ с целью принятия определенных технических или управленческих решений.

Целью данной работы является сравнительный анализ различных методов выделения трендов временных рядов (на пример, ценового графика) по критерию минимума среднего квадрата ошибок и выявления основных причин, влияющих на его величину.

Для выявления тенденций самое широкое применение получило скользящее усреднение. График, построенный по экспериментальным данным (на пример, цена закрытия), представляет собой ломанную кривую. Скользящие средние сглаживают цифровой ряд и облегчают определение тренда.

Обзор алгоритмов скользящего среднего

Простое скользящее среднее (Moving Average - MA) формируется путем вычисления среднего показателя за указанное число периодов.

$$MA = \frac{\sum_{i=1}^m z_i(t)}{n} \quad (2)$$

где $z(t)$ – цифровой временной ряд;

n – величина временного окна усреднения;

Усредняя временной ряд, простое скользящее среднее всегда следует с определенным лагом за главной тенденцией рынка, фильтруя мелкие колебания. Чем меньше параметр скользящего среднего, тем оно быстрее определяет новую тенденцию, но и одновременно делает больше ложных колебаний, и наоборот чем больше параметр, тем медленнее определяется новый тренд, но поступает меньше ложных колебаний (уменьшение колеблемости или сглаживание). На рисунке 1 представлены графики произвольного цифрового ряда и простого скользящего среднего.

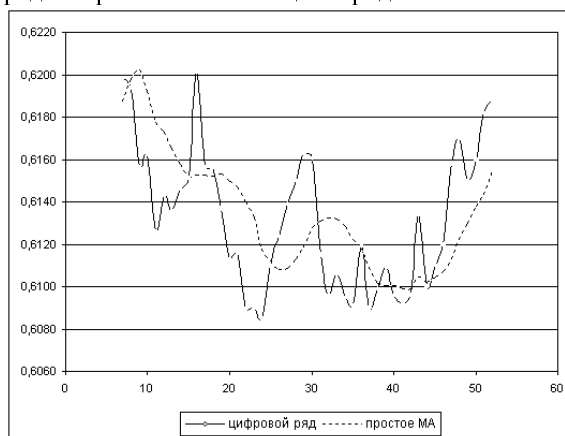


Рисунок 1 – Простое скользящее среднее.

Экспоненциальное скользящее среднее (Exponential Moving Average - EMA) рассчитывается по формуле:

$$EMA_i = EMA_{i-1} + \alpha \cdot (z_i(t) - EMA_{i-1}) \quad (3)$$

где

i – текущий момент времени,

$i - 1$ – предыдущий момент времени,

α – коэффициент экспоненциального

усреднения: $\alpha = 2 / (n + 1)$,

n – величина временного окна усреднения;

Экспоненциальная скользящая средняя уменьшает лаг, придавая больший вес последним данным цифрового ряда по сравнению с более дальними данными цифрового ряда. Этот позволяет быстрее реагировать на текущие изменения данных (на пример, цен) по сравнению с простой скользящей средней. Вес, придаваемый последним данным, зависит от периода скользящей средней. Чем короче период ЕМА, тем больший вес будет придаваться последним данным. На рисунке 2 представлены графики произвольного цифрового ряда и экспоненциального скользящего среднего.

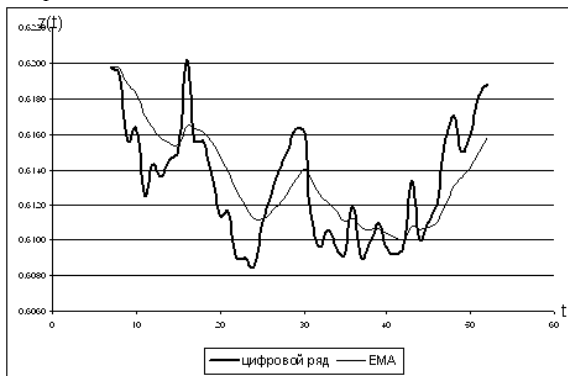


Рисунок 2 – Экспоненциальное скользящее среднее.

При использовании со скользящих средних нужно учитывать следующую особенность: с увеличением величины n уменьшается чувствительность (теряются элементы усредненного временного ряда). С уменьшением n увеличивается колеблемость выделяемых трендов, а чувствительность наоборот возрастает.

Недостатки скользящих средних:

1. Запаздывание отклика относительно элементов временного ряда на величину $(n-1)/2$, где n – величина временного окна усреднения; На рисунке 3 изображено отставание по времени начало тенденции скользящего среднего от начала тенденции цифрового ряда.

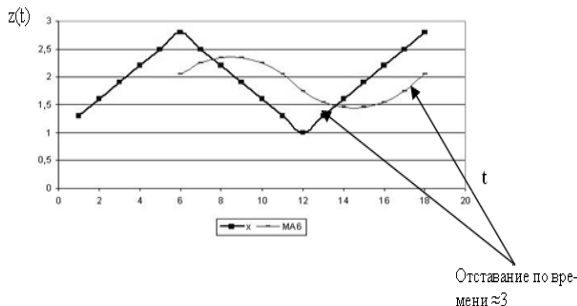


Рисунок 3 – Запаздывание отклика относительно элементов временного ряда

2. Скользящие средние при усреднении нелинейных трендов выделяют не истинные тренды, а их линейаризованные модели (при этом возникают определенные смещения). [3].

3. На рисунке 4 представлен фрагмент графика цифрового не линейного ряда $x(t)$, нелинейное уравнение авторегрессии, которого $y(t) = -0,0354x(t)^2 + 0,6975x(t) - 0,4327$ (штриховая линия). Увеличенный фрагмент данного рисунка наглядно показывает три различных смещения от истинного ряда $x(t)$: первое смещение $y(t)$, что составляет 1,1%; второе $MA(n=2) - 5,1\%$; третье $MA(n=4) - 19,9\%$

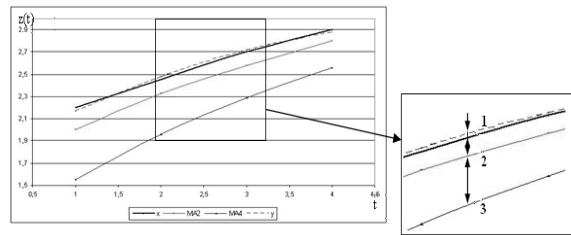


Рисунок 4 – Пример смещений от ряда $x(t)$.

4. Высокая колеблемость трендов, которая слабо уменьшается с ростом величины n [2], [4].

5. Наличие линейных частотных искажений при выделении трендов – форма выделяемых трендов отличается от реального тренда временного числового ряда, из-за нелинейности амплитудно-частотной характеристики (АЧХ).

Фильтры низких частот

Рассматривая скользящие средние с точки зрения цифровой фильтрации, нужно отметить, что скользящие средние являются фильтрами низких частот (ФНЧ). При применении такого фильтра происходят нежелательные изменения соотношений между амплитудами частотных составляющих входного и выходного цифрового ряда, пример которых можно увидеть на рисунке 5.

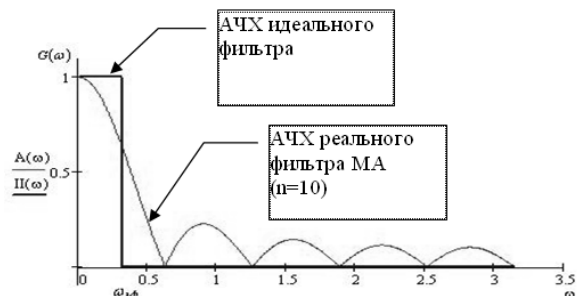


Рисунок 5 – Графики АЧХ реального и идеального фильтров.

Нельзя не отметить определенное преимущество МА – линейность ФЧХ и следовательно постоянную групповую задержку (т.е. все частотные составляющие сигнала задерживаются на оди-

наковое время), примеры которых изображены на рисунке 6.

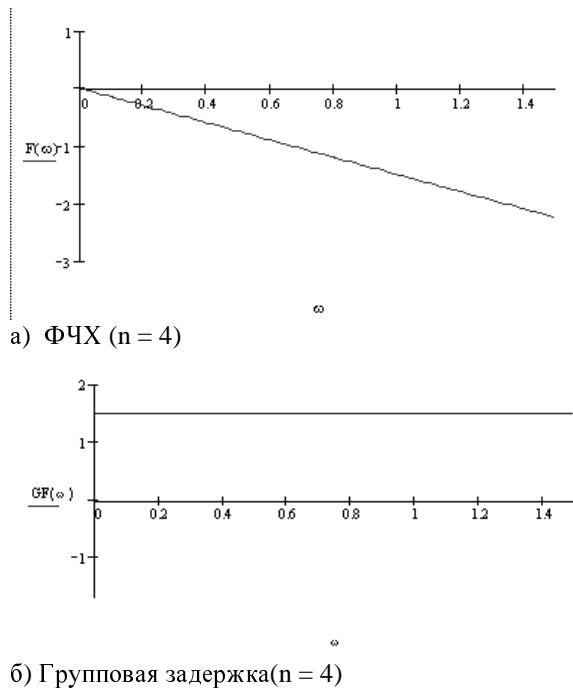


Рисунок 6

В отличие от МА групповая задержка ЕМА не является постоянной, т.к. ФЧХ имеет не линейную форму. Примеры графиков представлены на рисунке 7.

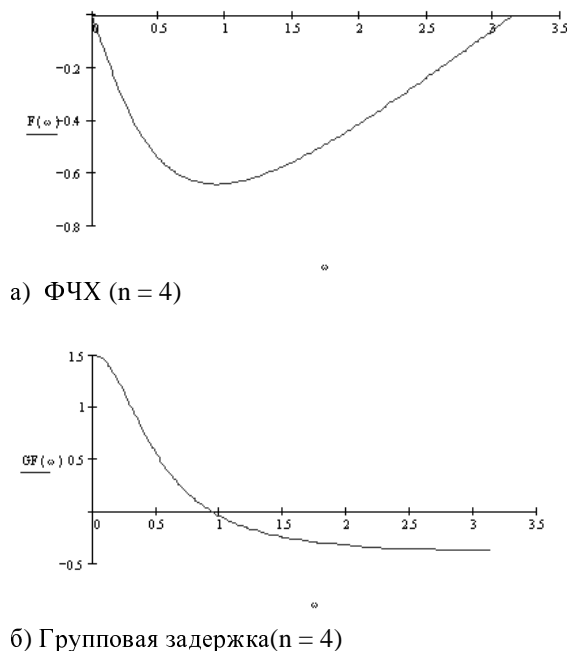


Рисунок 7

Исследования, простого и экспоненциального скользящих средних были подробно рассмотрены в работе [2] и доказали неэффективность применения данных скользящих средних.

Для устранения, выше перечисленных недостатков, были предложены новые методы решения:

новый вид скользящих средних – синтетические скользящие средние [1];

метод скользящей авторегрессии, адаптивной к типу уравнения выделяемого тренда.

Синтетическое скользящее среднее

Алгоритм синтетического скользящего среднего (рисунок 8) базируется на классическом алгоритме экспоненциального скользящего среднего, с многократным усреднением «назад-вперед», при условии постоянной величины α : $\alpha=2/(n'+1)$, где $n' = 2$, что обеспечивает минимальные искажения спектральных составляющих, характеризующих выделяемый тренд, минимизирует линейных искажений и смещения при выделении тренда не линейной формы. ФЧХ алгоритма определяется выражением [5]:

$$\tau_{\phi}(\omega_n) = \frac{-\phi(\omega_n)}{\omega_n} \quad (6)$$

где $\phi(\omega_n)$ – фазочастотная характеристика алгоритма усреднения;

ω_n – нормированная частота.

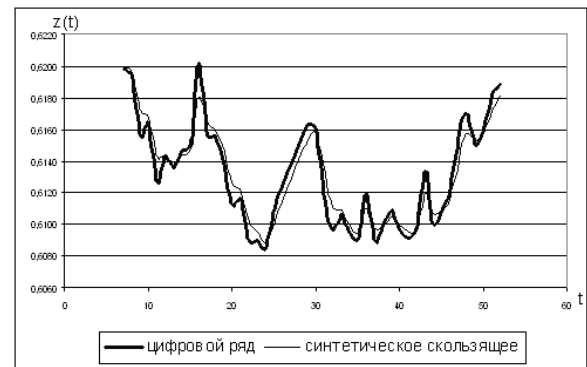


Рисунок 8 – Синтетическое скользящее среднее.

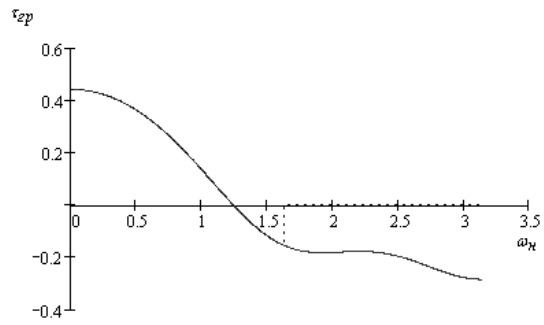
Запаздывание, выделяемого тренда относительно его реального положения, характеризуется групповой задержкой:

$$\tau_{gp}(\omega_n) = -\frac{d\phi(\bar{\omega}_n)}{d\bar{\omega}_n} \quad (7)$$

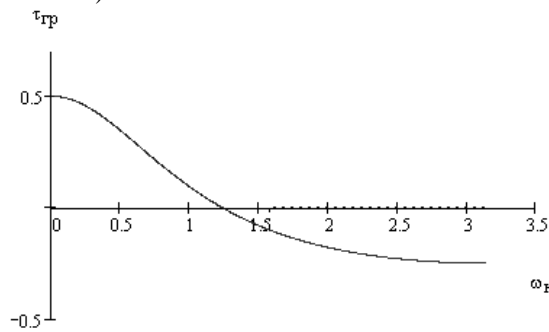
где $\bar{\omega}_n$ – среднее значение нормированной частоты;

$\frac{d}{d\bar{\omega}_n}$ – первая производная ФЧХ алгоритма.

Необходимо отметить, что с ростом n, групповая задержка становится постоянной и равна 0,5. На рисунке 9 представлены графики групповой задержки для различных окон скользящего.



а) n=4



б) n=24

Рисунок 9 – Групповая задержка синтетического скользящего среднего

Эффективность и преимущество данного алгоритма синтетического скользящего среднего приведены в работе [5].

Недостаток \ Скользящее среднее		Скользящие среднее		
		MA	EMA	SMA
1) Смещения относительно линейных трендов	1	*	*	-
	2	**	**	*
	3	**	**	*
2) Смещения относительно не линейных трендов	1	*	*	-
	2	**	**	*
	3	***	***	*
3) Линейные частотные искажения	1	*	*	*
	2	**	**	*
	3	***	***	*
4) Временная задержка	1	*	*	-
	2	**	**	*
	3	***	***	*
5) Низкая чувствительность	1	*	*	-
	2	**	**	*
	3	***	***	*
6) Зависимость групповой задержки от частоты (0..ωнφ)	1	-	***	*
	2	-	***	*
	3	-	***	*

- - отсутствует; * - слабо выражен; ** - средне выражен; *** - сильно выражен;
 1, 2, 3 – различные диапазоны окон скользяния: 2 ≤ n ≤ 10; 11 ≤ n ≤ 20; n больше 20.

Метод авторегрессии.

Сущность метода скользящей авторегрессии, заключается в следующем: на скользящем интервале усреднения по данным цифрового ряда методом наименьших квадратов вычисляют неизвестные параметры счетного множества уравнений авторегрессии различных типов. Для каждого из N уравнений авторегрессии вычисляется остаточная дисперсия:

$$\sigma_{ост,j}^2 = \frac{\sum_{t=1}^m (U_t - \bar{y}_j)^2}{m - k - 1} \quad (8)$$

где k – число неизвестных параметров j-го регрессионного уравнения.

Выбирается то j-ое уравнение остаточная дисперсия которого имеет наименьшее значение [3]. Более подробную информацию можно получить в работах [3] и [5].

Для сравнения классических алгоритмов скользящего усреднения, нового синтетического и метода скользящей авторегрессии рассмотрим критерий минимума среднего квадрата ошибок.

Выходной сигнал, исходя из линейности и стационарности рассматриваемой системы, должен представлять собой линейную комбинацию импульсных характеристик:

$$y(k) = \sum_{m \in S} x(m) \cdot h(k - m) \quad (9)$$

где y – выходной сигнал;
 x – входной сигнал;
 h – отсчеты импульсной характеристики;

Выбор оптимального алгоритма

Наиболее распространенным критерием оптимальности, применяемым для оценки качества обработки, является критерий минимума среднего квадрата ошибок. Применительно к фильтрации запишем его выражение в виде:

$$E \left\{ \left[y(k) - \sum_{m \in S} x(m) \cdot h(k - m) \right]^2 \right\} \rightarrow \min_h \quad (10)$$

где E – средний квадрат ошибки.

Выбор оптимального, в среднеквадратичном смысле, фильтра заключается в том, чтобы средний квадрат ошибки $\epsilon(t) = y(t) - x(t)$, выражающей разнице между входным сигналом и откликом, формируемой фильтром, был минимальным. Средний квадрат ошибки является статистической характеристикой и на практике характеризует смещение, получаемое между реальным трендом и выделяемым трендом.

Применим данный критерий к рассматриваемым алгоритмам скользящего усреднения. В качестве временного цифрового ряда, для всех

алгоритмов, будем использовать один и тот же фрагмент ценового графика рынка валютной биржи. Для обеспечения одинаковых условий и из-за недостатка алгоритма синтетических скользящих (работает только с временными окнами усреднения кратными 4) выбираем $n = 4, 8, 12, \dots, 48$. Получив таблицу соответствий:

$$E(n) \rightarrow \sum (C_i - MA_i(n))^2$$

построим график для каждого алгоритма усреднения (рисунок 10).

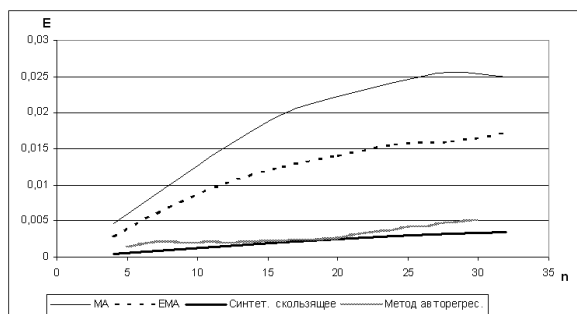


Рисунок 10 – Значение среднеквадратической ошибки для различных алгоритмов усреднения.

Выводы.

Используя полученные результаты можно сделать следующие выводы:

1. Используя классические алгоритмы MA и EMA пользователь вынужден находить компромисс при подборе величины окна усреднения. При $2 \leq n \leq 10$ алгоритм скользящего среднего обладает минимальными недостатками, но при этом высока колеблемость выделяемых трендов. При $n > 10$ пользователь получает сглаженный график, где проявляется наличие тенденции (если она есть), но растет величина задержки, уменьшается чувствительность и увеличивается смещение относительно наклонного тренда.

2. Метод скользящей авторегрессии реализует алгоритм нелинейной фильтрации. Алгоритм данного метода показывает его преимущества перед методами MA и EMA. Однако этот метод реализует на практике трудно вычисляемые характеристики. Величина временного окна должна быть много больше величины переменной подобранного регрессионного уравнения. При малых n ($n < 10$) возникают трудности правильного подбора регрессионного уравнения. Оптимальные условия для применения данного метода при $10 \leq n \leq 20$.

3. Синтетическое скользящее среднее является методом линейной фильтрации. Критерий среднего квадрата ошибки данного метода является наименьшим, и слабо растет с увеличением n . Применение синтетического скользящего среднего минимизирует линеаризацию нелинейных трендов и их смещение, что доказывает уменьшение эффекта линейных частотных искажений.

Простота вычислений делает его удобным для построения на его основе новых систем для принятия торговых или технических решений.

Литература:

1. Смирнов А., Тихонова О. Секрет совершенства индикаторов Марка Джурика раскрыт? // Валютный спекулянт. 2006, №1, с.32-35
2. Смирнов А., Михайлов С. Выбор типа скользящих средних // Валютный спекулянт. 2003, №7, с.50-55
3. Смирнов А., Гизатулин А. Новый метод сглаживания ценовых графиков // Валютный спекулянт. 2002, №12, с.38-40
4. Сиргиенко А.Б. Цифровая обработка сигналов // Питер. 2006, -750с.