

Модифицированная экспоненциальная скользящая средняя.

Станислав Булашев

Истинный механизм, который определяет динамику цен активов, вряд ли кому-нибудь доподлинно известен. Единственное, что можно сказать с уверенностью - в изменениях цен присутствует случайный фактор. А вот природа этой случайности может быть различной.

Согласно одной из возможных гипотез логарифмы изменения цен подчиняются нормальному распределению, однако это распределение нестационарно. То есть и математическое ожидание, и стандартное отклонение распределения могут меняться во времени. Вследствие этого, при обработке эмпирической выборки стандартными статистическими методами, предполагающими, что вся выборка получена из одной генеральной совокупности, получаем негауссовость выборки. Это может выражаться в тяжелых хвостах эмпирического распределения (вычисленный по выборке эксцесс превышает число 3, то есть эксцесс нормального распределения).

Согласно другой гипотезе логарифмы изменения цен изначально подчиняются распределению с эксцессом больше 3. В этой ситуации даже в случае стационарности самого распределения, полученная из этого распределения эмпирическая выборка может трактоваться как нестационарный во времени процесс. Дело в том, что оценкой математического ожидания случайной переменной x является среднее арифметическое по выборке:

$$\bar{X} = \frac{1}{N} \sum_{n=1}^N x_n$$

Среднее арифметическое случайных величин само является случайной величиной. Стандартное отклонение среднего арифметического зависит от стандартного отклонения случайной величины и объема выборки:

$$\sigma_{\bar{X}} = \sigma_x / \sqrt{N}$$

Таким образом, стандартное отклонение среднего значения меньше, чем стандартное отклонение самой случайной величины в \sqrt{N} раз, то есть точность оценки математического ожидания можно повысить путем увеличения объема выборки. Но это справедливо только для случайной величины с конечным математическим ожиданием и конечной дисперсией. Дело в том, что конечное математическое ожидание существует только у тех распределений, у которых при $x \rightarrow \pm\infty$ плотность вероятности падает как $1/|x|^{2+\delta}$ или круче, а конечная дисперсия только у тех распределений, у которых при $x \rightarrow \pm\infty$ плотность вероятности падает как $1/|x|^{3+\delta}$ или круче (δ - сколь угодно малое положительное число). Если же смоделировать ценовой график, используя в качестве логарифмов изменения цены случайную выборку, взятую из стационарного распределения с бесконечной дисперсией и/или бесконечным математическим ожиданием, и предложить эту выборку для анализа независимому наблюдателю, то у него может возникнуть иллюзия того, что он имеет дело с нестационарным во времени процессом.

Ну наконец нельзя исключить случай, когда не только параметры распределения, но и сам закон распределения приращений логарифмов цен нестационарен во времени, причем во временных рядах цен могут присутствовать участки, описываемые распределением с бесконечной дисперсией и/или бесконечным математическим ожиданием.

Таким образом, при отсутствии достоверной информации о механизме ценообразования, и связанных с этим трудностях прогнозирования рынка с приемлемой точностью и стабильностью, для извлечения спекулятивной прибыли необходимо просто использовать периодически возникающие участки направленного изменения цены, то есть тренды. При этом природа возникновения этих трендов в принципе не важна. Они могут быть вызваны как действиями крупных операторов рынка, так и образовываться спонтанно из колебаний цены, когда к зародившейся тенденции присоединяется все большее количество участников, то есть на какое то время возникает самоподдерживающаяся динамика изменения цены преимущественно в одном направлении.

Хочу сказать несколько слов о том, что в моем понимании включает в себя понятие "тренд". В этом вопросе может быть два подхода. С формальной точки зрения тренд - это модель процесса, которую можно выразить некоторой функцией. Неизвестные параметры этой модели определяются с помощью регрессионного анализа. При этом появляющиеся новые данные не должны приводить к отказу от выбранной модели. Вид функции может быть любым. В этом смысле и прямую, и экспоненту, и синусоиду допустимо назвать трендом. Однако такой подход годится для процессов физических, а не биржевых.

В связи с тем, что график цены невозможно описать подобного рода моделью, в биржевой торговле тренд можно рассматривать как участок ценового ряда с изменением цены преимущественно в одном направлении. Но так как на эту тенденцию накладывается шумовая составляющая, то в данном случае само понятие "тренд" невозможно отделить от результата его выявления. На одном и том же участке графика цены, в зависимости от используемого алгоритма, одна торговая система может показывать наличие тренда, а другая его отсутствие. Если согласиться с тем, что когда речь идет о динамике цен активов, понятие "тренд" не абсолютно, а относительно (по крайней мере в режиме реального времени, а не постфактум), то можно сказать, что тренд - это участок ценового ряда такой амплитуды, продолжительности и волатильности, которые позволят данному конкретному торговому алгоритму извлечь прибыль.

Получается, что тренд (в режиме реального времени) - это просто такая тенденция, которую мы смогли идентифицировать. Таким образом надо различать тренды "реальные" (которые удастся использовать в торговле), и тренды "идеальные" (которые все мы хорошо видим на графике цены постфактум). Отличия у этих трендов следующие:

- амплитуда реального тренда всегда меньше амплитуды соответствующего идеального тренда,

- моменты начала и окончания реального тренда всегда сдвинуты вправо по оси времени относительно начала и окончания соответствующего идеального тренда.

Если исходить из того, что нельзя дать достоверный прогноз касательно возможной амплитуды и продолжительности тренда, то в случае открытия позиции мы не в состоянии предвидеть свою прибыль, однако вполне можем ограничить свой убыток заранее заданной величиной. То есть при трендследящем методе торговли представляется разумной следующая тактика:

- до момента открытия позиции задается предельная сумма, которая может быть потеряна в ходе планируемой сделки,
- до момента открытия позиции определяется начальное значение и алгоритм изменения цены выхода, причем в случае открытия длинной позиции цена выхода должна быть неубывающей функцией времени (или по крайней мере не опускаться ниже начального значения цены выхода), а в случае открытия короткой позиции цена выхода должна быть невозрастающей функцией времени (или по крайней мере не подниматься выше начального значения цены выхода),
- объем открываемой позиции должен соответствовать предельной сумме потерь и начальной цене выхода (с учетом комиссий и проскальзываний в два конца).

При таком подходе задача определения момента открытия и закрытия позиции идеологически проста - с минимально возможной задержкой идентифицировать начало и окончание тренда. Однако эта задача сложна практически, так как возникающие на рынке тенденции зашумлены, вследствие чего начало и окончание идеального и реального трендов сдвинуты во времени. Одним из вариантов решения этой проблемы может быть торговая система, основанная на сглаживании рядов цен активов. Целью сглаживания является фильтрация шумовых колебаний с целью выявления устойчивых тенденций движения. Предположим, что выявляемые с помощью сглаживания тенденции движения цены являются "достаточно гладкими" линиями во времени. Тогда в идеальной скользящей средней (moving average или МА) максимально возможное сглаживание шумовых колебаний должно сочетаться с минимальным искажением самих тенденций. Если же перейти от идеальных к реальным МА, то при сглаживании необходимо добиться разумного компромиса между мерой гладкости МА и мерой близости этой МА к исходному ценовому ряду. Разумеется "достаточная близость" и при этом "достаточная гладкость" могут быть разными в зависимости от амплитуды и продолжительности выявляемых трендов.

* * *

До того, как перейти к собственно используемому мной алгоритму сглаживания временных рядов и вариантам построения на его основе торговых систем, для сохранения логической стройности изложения необходимо предварительно кратко рассмотреть одну из всем хорошо известных скользящих средних. Вернемся к упомянутым выше мере гладкости скользящей средней и мере близости скользящей средней к исходному ценовому ряду. Введем обозначения:

- $\{x_t\}$ - ряд цены (или логарифма цены) актива,
- $\{y_t\}$ - скользящая средняя ряда $\{x_t\}$.

В качестве меры близости рядов $\{x_t\}$ и $\{y_t\}$ в момент времени t выберем функцию $(y_t - x_t)^2$, а в качестве меры гладкости ряда $\{y_t\}$ функцию $(y_t - y_{t-1})^2$. Если взять взвешенную сумму этих функций, то есть

$$\Delta S_t = w \cdot (y_t - x_t)^2 + (1 - w) \cdot (y_t - y_{t-1})^2$$

то для расчета текущей величины y_t , которая минимизирует функцию ΔS_t , надо решить следующее уравнение:

$$\frac{\partial \Delta S_t}{\partial y_t} = 0$$

После несложных преобразований получим, что $y_t = y_{t-1} + w \cdot (x_t - y_{t-1})$. Это формула для вычисления всем хорошо известной экспоненциальной скользящей средней с фактором сглаживания w (которую будем обозначать как $EMA_t^{[w]}$).

Одним из важнейших показателей скользящей средней является время ее запаздывания относительно исходного сигнала. Для вычисления запаздывания формулу скользящей средней нужно записать в явном виде, то есть выразить ее только через значения ряда $\{x_t\}$. Для этого подставим в формулу

$$y_t = y_{t-1} + w \cdot (x_t - y_{t-1})$$

на 1-м шаге $y_{t-1} = y_{t-2} + w \cdot (x_{t-1} - y_{t-2})$,

на 2-м шаге $y_{t-2} = y_{t-3} + w \cdot (x_{t-2} - y_{t-3})$, и т.д.

В итоге получим, что

$$y_t = w \cdot \sum_n (1 - w)^n \cdot x_{t-n}$$

Традиционно скользящую среднюю ставят в соответствие с последней точкой интервала расчета, то есть с моментом времени t , хотя это некорректно. Ее значение нужно ставить в соответствие с моментом времени, определяемом по формуле

$$\bar{t} = w \cdot \sum_n (1 - w)^n \cdot (t - n)$$

Раскрыв сумму получим при достаточно больших t

$$\bar{t} = t - \frac{(1 - w)}{w} + \frac{(1 - w)^{t+1}}{w} \approx t - \frac{(1 - w)}{w}$$

Таким образом время запаздывания $EMA_t^{[w]}$ относительно исходного сигнала равно

$$\tau_{EMA} = \frac{1 - w}{w}$$

Соответственно фактор сглаживания выражается через время запаздывания как

$$w = \frac{1}{\tau_{EMA} + 1}$$

Из формулы для расчета $EMA_t^{[w]}$ следует, что ее 1-я производная меняет свой знак при пересечении линией цены линии скользящей средней (первая производная обращается в нуль при $x_t = y_{t-1}$). Конечно если графики цен представлены в дискретном виде (бары), то корректнее вместо термина "производная" употреблять термин "разность", но здесь и далее будет говориться о производных, помня об этой условности.

Указанное свойство ЕМА является логическим обоснованием торговых алгоритмов, в которых сигналами к покупкам/продажам являются моменты пересечений цены и средней. Однако, из этого же свойства следует, что амплитуда ЕМА всегда меньше, чем амплитуда цены. То есть локальный максимум цены всегда раньше по времени и выше соответствующего максимума ЕМА, а локальный минимум цены всегда раньше и ниже минимума ЕМА. В связи с этим возникла идея модифицировать алгоритм расчета ЕМА таким образом, чтобы при одном и том же времени запаздывания, увеличить амплитуду и гладкость скользящей средней.

* * *

Итак, переходим к используемому мной в торговой системе алгоритму сглаживания временных рядов. Сохраняем в качестве меры близости рядов $\{x_t\}$ и $\{y_t\}$ в момент времени t функцию $(y_t - x_t)^2$. А вот в качестве меры гладкости ряда $\{y_t\}$ выберем комбинацию функций $(y_t - y_{t-1})^2$ и $(y_t - 2 \cdot y_{t-1} + y_{t-2})^2$, то есть комбинацию 1-й и 2-й производной ряда $\{y_t\}$. Тогда взвешенная сумма всех этих функции примет вид

$$\Delta S_t = w \cdot (y_t - x_t)^2 + w_1 \cdot (y_t - y_{t-1})^2 + w_2 \cdot (y_t - 2 \cdot y_{t-1} + y_{t-2})^2$$

$$w + w_1 + w_2 = 1$$

Решением уравнения

$$\frac{\partial \Delta S_t}{\partial y_t} = 0$$

будет $y_t = y_{t-1} + w \cdot (x_t - y_{t-1}) + w_2 \cdot (y_{t-1} - y_{t-2})$.

Назовем эту функцию модифицированной экспоненциальной скользящей средней и будем обозначать ее как $MEMA_t^{[w, w_2]}$. Заметим, что $EMA_t^{[w]}$ является частным случаем функции $MEMA_t^{[w, w_2]}$, то есть $EMA_t^{[w]}$ равна $MEMA_t^{[w, w_2]}$ с коэффициентом $w_2 = 0$. В отличии от $EMA_t^{[w]}$, для того, чтобы первая производная функции $MEMA_t^{[w, w_2]}$ стала равной нулю, необходимо, чтобы

$$x_t = y_{t-1} - \frac{w_2}{w} \cdot (y_{t-1} - y_{t-2})$$

То есть $MEMA_t^{[w, w_2]}$ обладает большей инерцией по сравнению с $EMA_t^{[w]}$, и "малые" пересечения линией цены x_t линии y_t , не меняют знак первой производной скользящей средней.

Для вычисления времени запаздывания $MEMA_t^{[w, w_2]}$ относительно цены необходимо записать ее в явном виде. Для этого подставим в формулу

$$y_t = y_{t-1} + w \cdot (x_t - y_{t-1}) + w_2 \cdot (y_{t-1} - y_{t-2})$$

на 1-м шаге $y_{t-1} = y_{t-2} + w \cdot (x_{t-1} - y_{t-2}) + w_2 \cdot (y_{t-2} - y_{t-3})$,

на 2-м шаге $y_{t-2} = y_{t-3} + w \cdot (x_{t-2} - y_{t-3}) + w_2 \cdot (y_{t-3} - y_{t-4})$, и т.д.

В итоге получим, что

$$y_t = w \cdot \sum_n q_n \cdot x_{t-n}$$

где

$$q_n = (1 - w + w_2)^n \cdot \sum_{k=0}^{n/2} \frac{(n-k)!}{(n-2k)!k!} \cdot \left(\frac{-w_2}{(1-w+w_2)^2} \right)^k$$

Последняя сумма является табличной и ее значение можно взять из справочника по рядам и частичным суммам. Проведя несложные преобразования, приходим к следующей формуле:

$$q_n = \frac{a^{n+1} - b^{n+1}}{a - b}$$

где

$$a = \frac{(1 - w + w_2) + \sqrt{(1 - w + w_2)^2 - 4w_2}}{2}$$

$$b = \frac{(1 - w + w_2) - \sqrt{(1 - w + w_2)^2 - 4w_2}}{2}$$

Итак

$$y_t = w \cdot \sum_n q_n \cdot x_{t-n} = w \cdot \sum_n \frac{a^{n+1} - b^{n+1}}{a - b} \cdot x_{t-n} = \frac{w}{a - b} \cdot \left(a \cdot \sum_n a^n \cdot x_{t-n} - b \cdot \sum_n b^n \cdot x_{t-n} \right)$$

Отсюда следует, что

$$y_t = \frac{w}{a - b} \cdot \left(\frac{a}{1 - a} \cdot EMA_t^{[1-a]} - \frac{b}{1 - b} \cdot EMA_t^{[1-b]} \right)$$

Получилось, что $MEMA_t^{[w, w_2]}$ является комбинацией двух экспоненциальных скользящих средних, взятых с соответствующими весами. Поэтому время запаздывания $MEMA_t^{[w, w_2]}$ относительно цены можно найти из времен запаздывания двух ЕМА, взятых с теми же самыми весами. После проведения необходимых вычислений в итоге получается, что время запаздывания $MEMA_t^{[w, w_2]}$ относительно исходного сигнала равно

$$\tau_{MEMA} = \frac{1 - w - w_2}{w}$$

Конечно работать со скользящей средней, имеющей 2 параметра, не всегда удобно, в частности - время запаздывания не определяет однозначно оба коэффициента w и w_2 . Однако это компенсируется потенциально более точной аппроксимацией исходного ряда цен. В принципе от одного из параметров можно избавиться, задав

например w_2 как функцию w . В этом случае надо учитывать, что $w + w_1 + w_2 = 1$, то есть $w_2 \leq 1 - w$.

* * *

Проиллюстрируем различие между ЕМА и МЕМА на примере. В качестве сглаживаемого ряда $\{x_t\}$ возьмем натуральный логарифм рублевых цен закрытия акций РАО ЕЭС на интервале с июня 1997г. по март 2007г. (недельные бары). По оси абсцисс будет указан порядковый номер бара. Время запаздывания сглаженного сигнала относительно исходного выберем равным 4 барам. Ниже приведены 4 рисунка, иллюстрирующие различия между ЕМА и МЕМА.

На рисунке 1 изображен исходный ряд и сглаженные линии ЕМА и МЕМА. Очевидно, что при выбранных значениях w и w_2 , МЕМА имеет большую амплитуду, по сравнению с ЕМА, при одном и том же времени запаздывания.

На рисунке 2 сглаженные линии сдвинуты в прошлое на время запаздывания, то есть фактически ЕМА и МЕМА показывают "идеальный" прогноз цены (прогноз с учетом будущих цен).

Рисунки 3 и 4, на которых изображены 1-я и 2-я производная сглаженных линий, показывают существенно меньшую волатильность производных МЕМА, то есть ее большую гладкость, по сравнению с ЕМА.

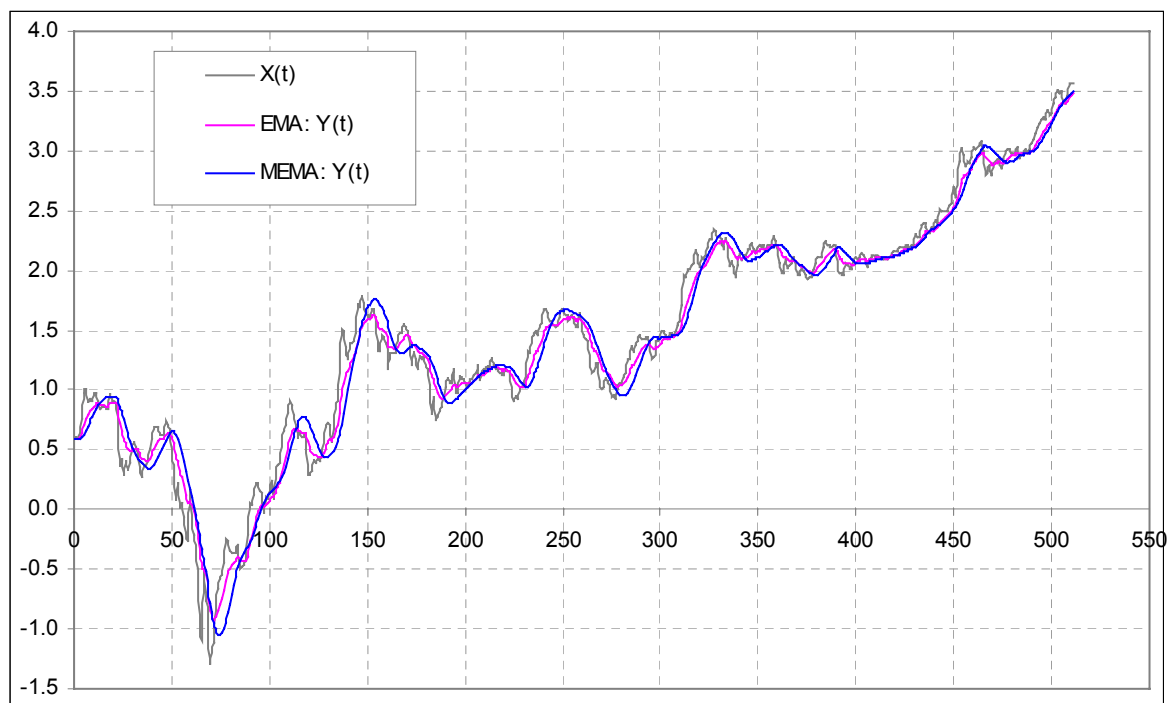


Рис.1. Исходный ряд и сглаженные линии ЕМА и МЕМА.

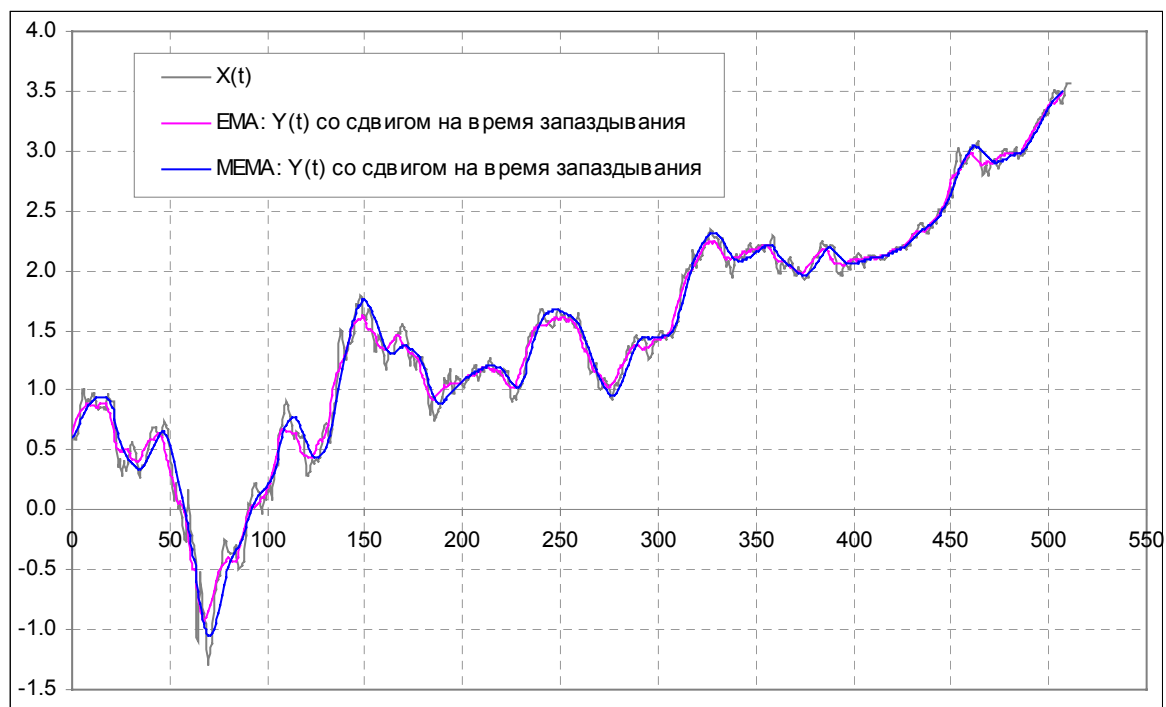


Рис.2. Исходный ряд и сглаженные линии ЕМА и МЕМА, сдвинутые в прошлое на время запаздывания.

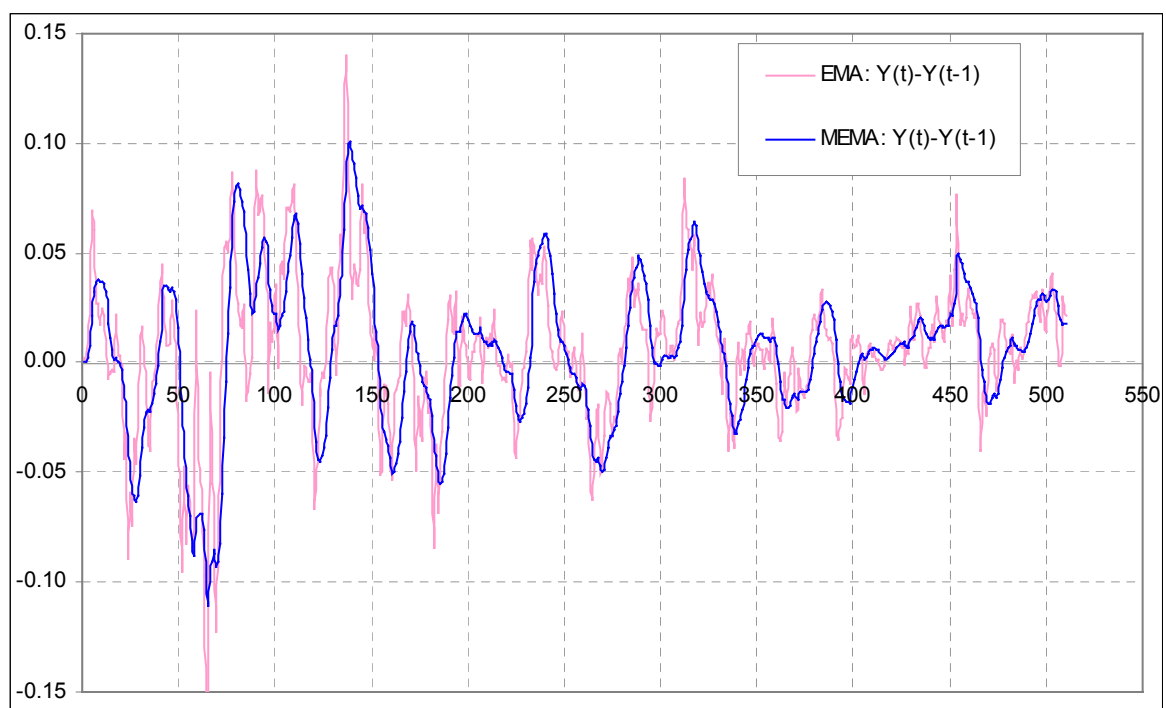


Рис.3. Первая производная ЕМА и МЕМА.

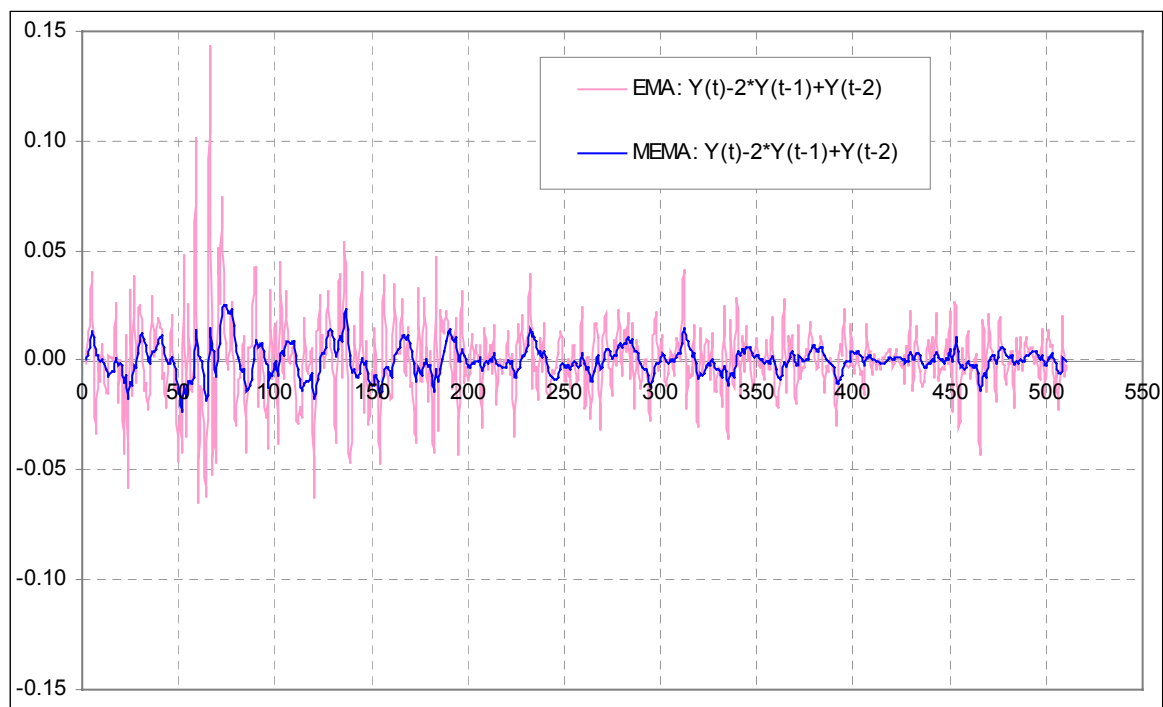


Рис.4. Вторая производная ЕМА и МЕМА.

Рассмотрим варианты построения торговых алгоритмов на основе МЕМА.

1 вариант

Открытие/закрытие позиций при изменении знака 1-й производной сглаженной линии. Напомним, что цена, меняющая на текущем баре знак 1-й производной МЕМА, может быть найдена по формуле

$$x_t = y_{t-1} - \frac{w_2}{w} \cdot (y_{t-1} - y_{t-2})$$

Данный вариант очевиден и не требует более детального рассмотрения.

2 вариант

Рассмотрим ошибки аппроксимации ценового ряда $e_t = x_t - y_t$, где $y_t = MEMA^{[w, w_2]}(x_t)$.

Сгладим модуль этих ошибок тем же алгоритмом и с теми же параметрами сглаживания, то есть найдем $E_t = MEMA^{[w, w_2]}(|e_t|)$. Далее построим канал, верхняя и нижняя граница которого вычисляются как

$$YMax_t = y_t + E_t$$

$$YMin_t = y_t - E_t$$

Построенный таким образом канал обладает следующим свойством:

- после пробоя ценой снизу вверх верхней границы канала, его нижняя граница является как правило неубывающей функцией времени (*up-trend*),
- после пробоя ценой сверху вниз нижней границы канала, его верхняя граница является как правило невозрастающей функцией времени (*down-trend*).

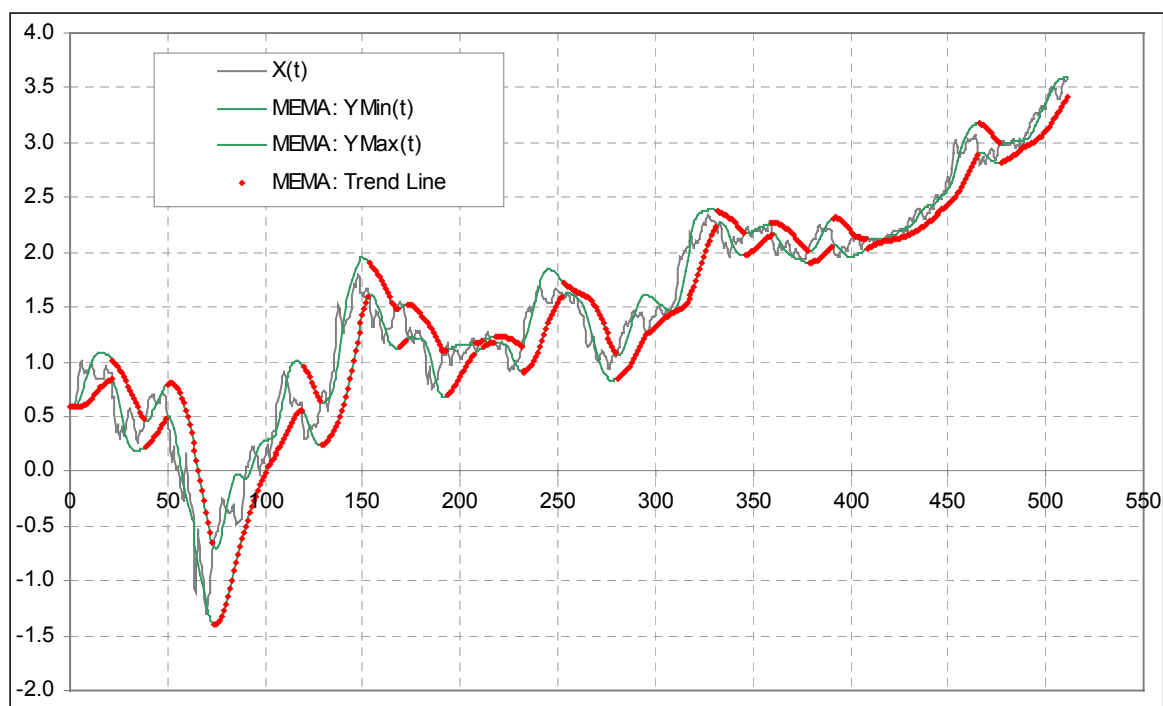


Рис.5. Канал на основе МЕМА и линия текущего тренда.

На рисунке 5 изображен канал на основе МЕМА и линии повышательных и понижительных трендов. Как правило сигналы об открытии/закрытии позиций по варианту 2 поступают ранее соответствующих сигналов по варианту 1, или одновременно с ними. Объем открываемой позиции зависит от предельного риска на сделку, то есть в данном случае от текущей ширины канала. Например, длинная позиция на 74 баре открывается вообще говоря меньшим объемом, чем длинная позиция на 281 баре.

* * *

Описанный выше торговый алгоритм на основе МЕМА обладает всеми достоинствами и недостатками любой трендследающей системы - он вполне приемлемо работает на трендовых участках цен, но теряет деньги на флэтах. Дальнейшие его модификации могут быть связаны с введением переменных факторов сглаживания. При этом на трендах сглаженная линия должна следовать за ценой, а ширина канала, выраженная в ошибках аппроксимации, быть минимальной. На флэтах сглаженная кривая должна вырождаться в горизонтальную линию, а ширина канала увеличиваться для уменьшения количества ложных сигналов.

Обратимся вновь к аналогии с ЕМА. Рекуррентная формула для ее вычисления при переменном факторе сглаживания принимает вид $y_t = y_{t-1} + w_t \cdot (x_t - y_{t-1})$, где w_t принимает значения от 0 на флэте до некоторого w_{\max} на тренде, причем w_t тем больше, чем более явно текущую динамику можно трактовать как "тренд".

В случае же МЕМА с переменными факторами сглаживания рекуррентная формула запишется как $y_t = y_{t-1} + w_t \cdot (x_t - y_{t-1}) + w_{(2)t} \cdot (y_{t-1} - y_{t-2})$, где

на флоте	$w_t \rightarrow 0$	на тренде	$w_t \rightarrow w_{\max}$
	$w_{(2)t} \rightarrow 0$		$w_{(2)t} \rightarrow w_{(2)\max}$

На рисунке 6 приведен пример подобного рода сглаженной линии.

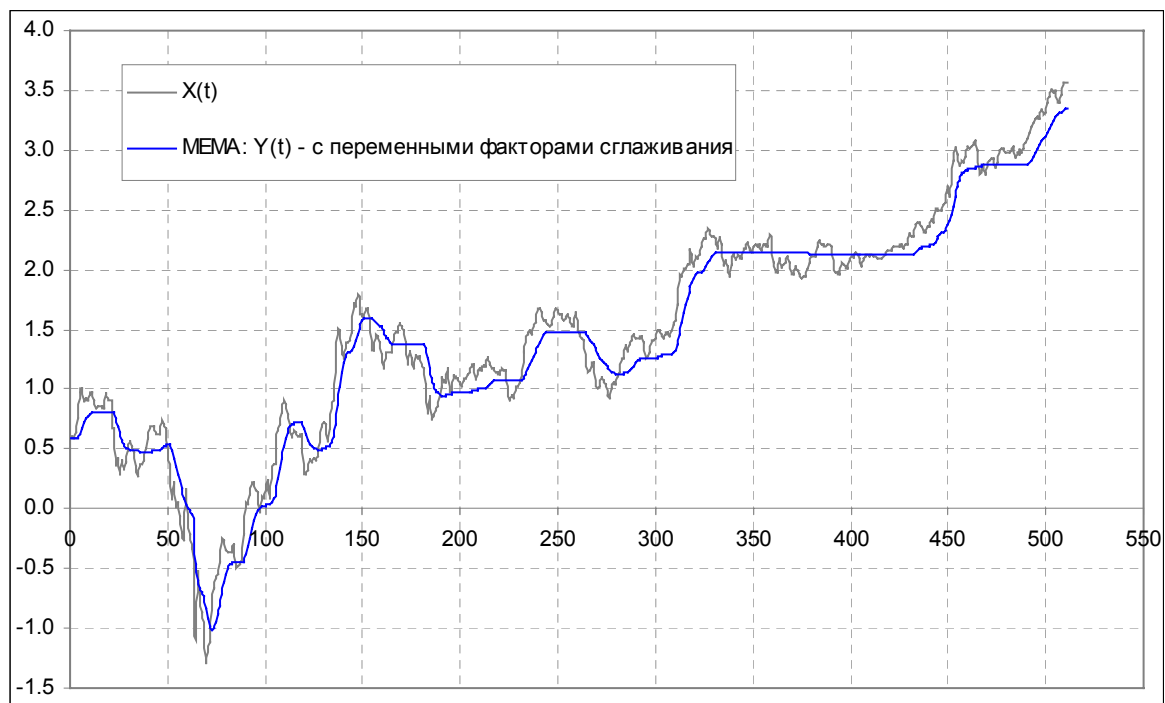


Рис.6. МЕМА с переменными факторами сглаживания.