

Московский государственный университет имени М. В. Ломоносова



Факультет Вычислительной Математики и Кибернетики
Кафедра Математических Методов Прогнозирования

КУРСОВАЯ РАБОТА

«Обзор и применение методов прогнозирования финансовых временных рядов»

Выполнил:

студент 3 курса 317 группы

Рыжков Александр Михайлович

Научный руководитель:

д.ф-м.н., доцент

Дьяконов Александр Геннадьевич

Москва, 2013

Содержание

1 Введение	3
2 Постановка задачи	3
3 Теоретические сведения	4
3.1 Комбинация скользящих средних	5
3.2 Полосы Боллинджера	7
3.3 Осцилляторы скользящих средних	9
4 Вычислительный эксперимент	11
5 Оценка вычислительной сложности алгоритма	12
6 Практическое применение	14
7 Заключение и выводы	15
Список литературы	15

1 Введение

В данной работе проводится анализ различных алгоритмов прогнозирования финансовых временных рядов и их комбинаций. В настоящее время эта область очень актуальна, поскольку может быть почти полностью автоматизирована и являться источником средних и больших доходов без ежеминутного вмешательства пользователя.

Для того, чтобы теория не казалась мифом или сильно оторванной от практики, в работе приводится вычислительный эксперимент, основанный на реальных данных, предложенных компанией Inventum, а также тестовый запуск комбинации написанных алгоритмов на временном ряде фьючерсов компании Газпром.

2 Постановка задачи

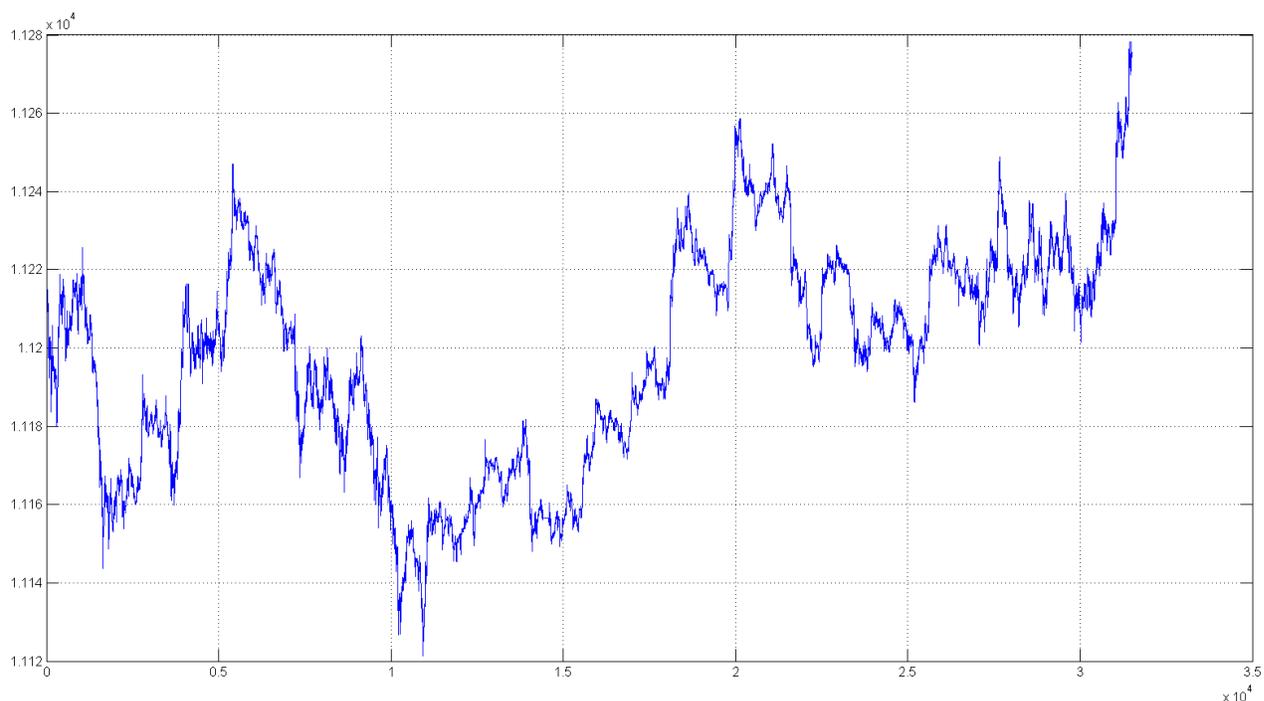


Рис. 1: Временной ряд финансового инструмента

Исходными данными при решении поставленной задачи является временной ряд 1 (или некоторый набор временных рядов в случае их тесной взаимосвязи - к этой категории относятся, например, ряды соотношений между ценами доллара и евро и количество сделок на обоих рядах в каждый момент времени). В общем случае в движениях графика временного ряда существуют несколько составляющих:

- Общая тенденция (тренд) - линия движения ряда на всем временном интервале (в данном случае 1 она возрастающая)

- Сезонная тенденция (сезонность) - закономерность в движениях ряда, связанная с определенными жизненными ситуациями - праздниками, каникулами, отпусками и т. п.
- Местные тенденции - линия движения ряда на небольшом отрезке всего временного интервала (в данном случае 1 сначала идет возрастающая местная тенденция, потом убывающая, потом снова возрастающая и т. д.)
- Случайная составляющая - «шумовая» компонента временного ряда, создающая незначительные колебания графика ряда относительно своего ожидаемого, тенденционного значения

Задача состоит в том, чтобы спрогнозировать дальнейшее движение ряда для совершения сделок - покупок и продаж финансового инструмента - с целью получения максимальной итоговой прибыли. Единственное ограничение - это совпадение числа сделок-покупок с числом сделок-продаж в конце заранее заданного временного интервала (обычно в конце дня).

Поскольку количество сделок не ограничено, то их можно совершать хоть в каждый момент времени, предсказывая движение ряда всего в один последующий момент, но такой вариант оказывается невыгодным, поскольку за каждую сделку взимается комиссия в размере 0.000085% от цены финансового инструмента. На первый взгляд она кажется мизерной, но даже такое ее значение может сделать наше поведение на рынке убыточным в случае слишком частого заключения сделок. Однако мы, как и всегда, хотим зарабатывать чем больше, тем лучше. Поэтому необходимо найти «золотую середину» между числом сделок и полученной прибылью. Это и делают описанные далее алгоритмы.

3 Теоретические сведения

В этой части работы рассмотрим основные алгоритмы, используемые на практике для поддержания баланса между общим числом сделок и полученной прибылью.

Для того, чтобы оценивать ситуацию на рынке в некоторый момент времени t_0 , воспользуемся сглаживанием предыстории изменения цены финансового инструмента, называемым *Скользящим Средним (Moving Average)* и вычисляемым по следующей формуле:

$$MA = \frac{1}{n} \sum_{k=1}^n price(t_0 - k + 1), \quad (3.0.1)$$

где n - окрестность сглаживания

3.1 Комбинация скользящих средних

Наиболее простым и логичным с точки зрения интуитивного понимания ситуации является алгоритм, использующий для прогнозирования комбинации скользящих средних. Количество сглаживаний N , используемых этим алгоритмом, варьируется, но их количество часто берется равным $N = 2$ или $N = 3$. Для большей наглядности описания основной идеи алгоритма воспользуемся случаем $N = 2$.

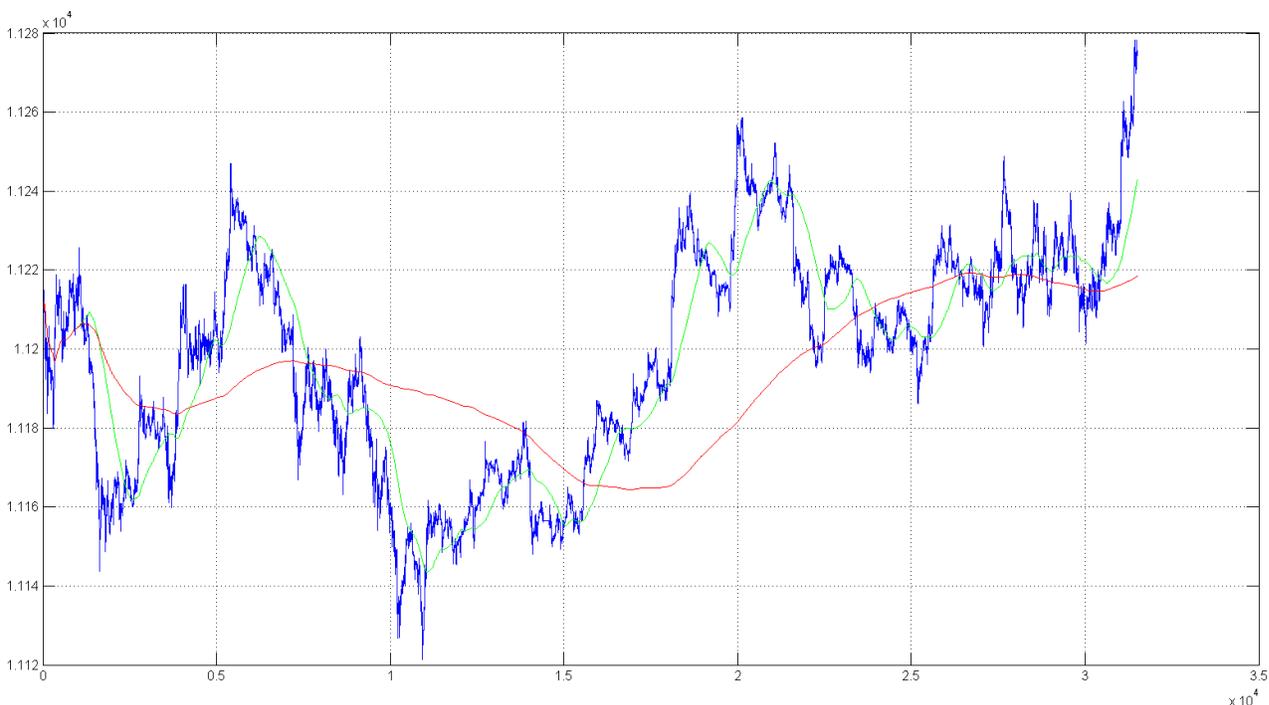


Рис. 2: Комбинация скользящих средних

Выберем два различных скользящих средних:

- с окрестностью n_1 - короткое скользящее среднее
- с окрестностью n_2 ($n_2 > n_1$) - длинное скользящее среднее

Заметим следующие особенности предложенного графика 2 и обоснуем их:

1. Графики используемых скользящих средних пересекаются, причем происходит это через некоторый промежуток времени после того, как временной ряд цены финансового инструмента изменяет направление своего движения с возрастания на убывание и наоборот. Это связано с тем, что зеленое сглаживание больше похоже на исходный график временного ряда, чем красное, ведь оно имеет меньшую окрестность
2. При изменении направления движения временного ряда с возрастания на убывание, через сравнительно небольшой промежуток времени зеленое сглаживание пробивает красное **сверху вниз** (такая ситуация изображена на графике 3), что говорит о

наличии **убывающей средней тенденции** на рынке. В такой момент лучше всего совершить продажу финансового инструмента, дабы потом купить его по более низкой цене, тем самым получив некоторую прибыль

3. При изменении направления движения временного ряда с убывания на возрастание, также как и в предыдущем случае через сравнительно небольшой промежуток времени зеленое сглаживание пробивает красное **снизу вверх** (такая ситуация изображена на графике 4), что говорит о наличии **возрастающей средней тенденции** на рынке. В такой момент лучше всего совершить покупку финансового инструмента, чтобы потом продать его после возрастания цены до точки локального максимума



Рис. 3: Убывающая средняя тенденция

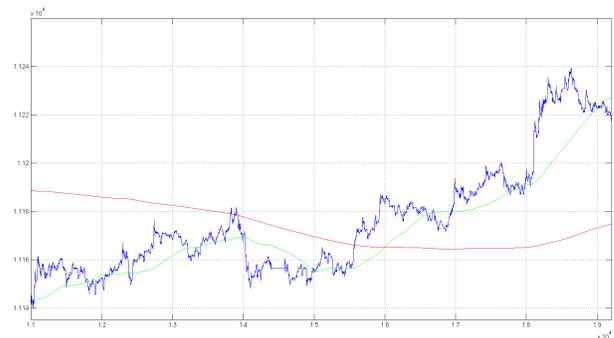


Рис. 4: Возрастающая средняя тенденция

Также следует отметить особенность, которая позволяет использовать весь график временного ряда, а не только некоторые его отрезки. Рассмотрим ситуацию с графиком 2: первым делом у нас будет совершена покупка (зеленое сглаживание отрывается от красного вверх). Далее должна быть совершена продажа (зеленое сглаживание пробивает красное сверху вниз), но здесь мы прибегнем к хитрости - сделаем не одну продажу, а сразу две! Это позволит нам не только закрыть позицию покупки с предыдущего момента времени, но и сразу же открыть новую позицию продажи, которую мы успешно закроем, когда в следующий раз решим совершить покупку. Продолжая делать по две покупки или две продажи в зависимости от ситуации, придем к ситуации, когда день уже скоро завершится, а нам нужно совершить продажу - теперь вместо двух мы совершим только одну продажу, тем самым сравнивая число покупок и продаж за день. Такая схема условно называется «1 – 2 – ... – 2 – 1». Использование этой схемы повышает прибыль в среднем в 1.5 – 2 раза, что достаточно существенно и не может быть сброшено со счетов.

Таким образом, делая сделки указанных выше типов в соответствующие моменты времени (в точках пересечения скользящих средних) можно получить достаточно высокую прибыль, что говорит о реальной возможности применения этого метода в практических целях.

3.2 Полосы Боллинджера

Наиболее прибыльные варианты алгоритма «Комбинация скользящих средних» получаются при выборе достаточно большой окрестности длинного скользящего среднего, которое делает основной акцент на общую тенденцию временного ряда (это можно видеть на графике 2). Таким образом, за весь промежуток времени совершается сравнительно небольшое число сделок, что не позволяет использовать более мелкие движения ряда - из этого и возникает идея придумать алгоритм, который способен работать на небольших временных отрезках.

Первоначальными, пробными реализациями этой идеи является алгоритм «конверт». Конверт - это две извилистые «параллельные линии» выше и ниже некоторого скользящего среднего, образующие полосу, которая охватывает большую часть ценовых колебаний временного ряда. Для того, чтобы рассчитать полосу, необходимо на 2% (этот коэффициент может быть изменен в зависимости от диапазона значений цены финансового инструмента) увеличить и уменьшить значения короткого скользящего среднего:

- Верхняя граница = $MA * (1 + 0.02) = MA * 1.02$
- Нижняя граница = $MA * (1 - 0.02) = MA * 0.98$

Сигнал к продаже подается, когда цена прорывает верхнюю границу. Когда цена пробивает нижнюю границу, подается сигнал к покупке. Поскольку сигналы конверта очень недолговечны и часто направлены против основной тенденции рынка, решающее значение при использовании алгоритма «конверт» имеет скорость совершения сделок.

Теперь рассмотрим более сложный, но в тоже время более качественный и долговечный алгоритм «полосы Боллинджера», предложенный Джоном Боллинджером. Для того, чтобы начать его описание введем определение волатильности финансового инструмента:

Определение 1. *Волатильность финансового инструмента - статистический финансовый показатель, характеризующий изменчивость цены и позволяющий определить риск приобретения финансового инструмента*

Алгоритм «полосы Боллинджера» комбинирует в себе алгоритм «конверт», рассмотренный ранее, с волатильностью финансового инструмента. Данный инструмент был разработан для того, чтобы определять, являются ли цены относительно высокими или низкими. В качестве меры волатильности в этом алгоритме используется стандартное отклонение из-за его чувствительности к экстремальным значениям. Границы полос обычно берутся на 2 (этот коэффициент также может быть изменен в зависимости от диапазона значений цены финансового инструмента, как и в алгоритме «конверт») стандартных отклонения выше и ниже короткого скользящего среднего.

Полосы Боллинджера во многом напоминают то сужающийся, то расширяющийся конверт 5 и рассчитываются по следующим формулам:

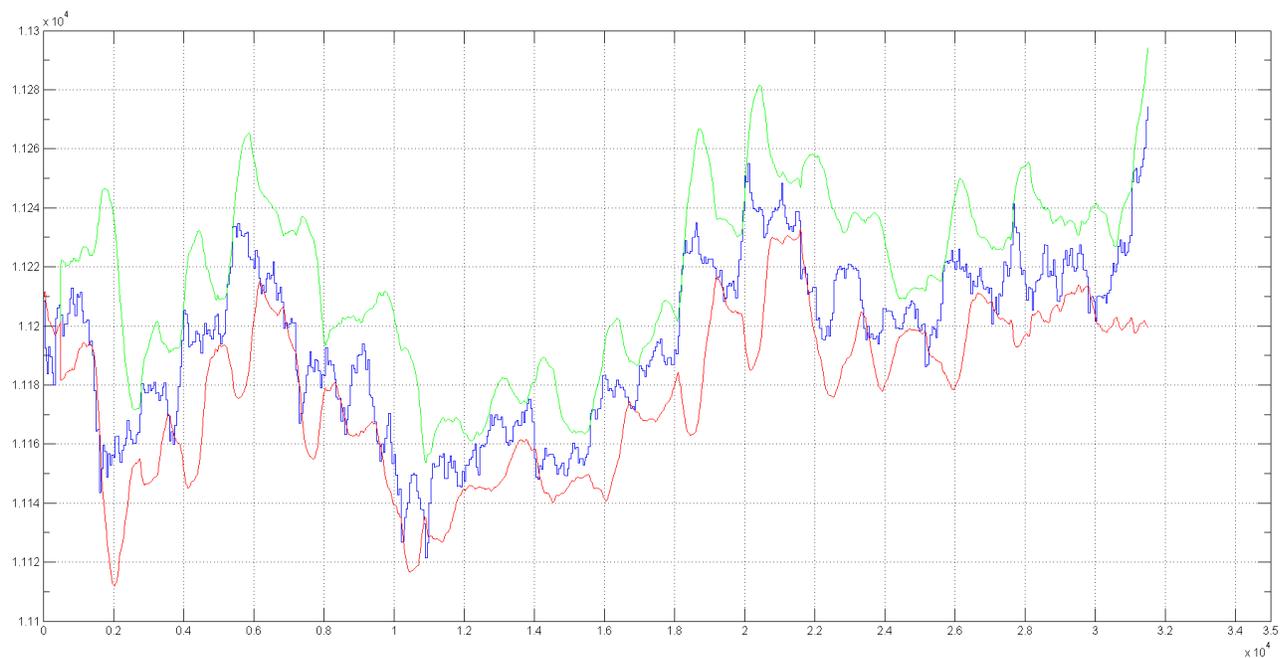


Рис. 5: Полосы Боллинджера

$$d = \frac{((P_1 - MA_1)^2 + (P_2 - MA_2)^2 + \dots + (P_n - MA_n)^2)}{n} \quad (3.2.1)$$

$$\sigma = \sqrt{d} \quad (3.2.2)$$

$$LB = MA - 2\sigma, \quad (3.2.3)$$

$$UB = MA + 2\sigma, \quad (3.2.4)$$

где:

- d - отклонение от скользящего среднего
- P_i - значение цены в нужной точке
- MA_i - значение скользящего среднего в нужной точке
- n - число слагаемых
- σ - стандартное отклонение
- $LB(low\ bound)$ - нижняя граница полосы
- $UB(up\ bound)$ - верхняя граница полосы

Если полоса сильно сузилась, это означает, что волатильность финансового инструмента низка и что в ближайшем будущем полоса резко расширится. Однако малая волатильность не указывает направления будущего изменения цены.

Если же полоса сильно расширилась, то волатильность слишком высока и следует ожидать ее понижения. При таком развитии событий первоначальное направление тенденции известно, поэтому и направление разворота очевидно. Нужно только уловить момент, когда этот самый разворот произойдет.

Еще один сигнал полоса Боллинджера подает, когда образуется две следующие друг за другом вершины временного ряда, первая из которых лежит вне полосы, а вторая — внутри нее. Это — сигнал к продаже (зеленые маркеры графика 6). И наоборот: две следующие друг за другом вершины временного ряда, первая из которых лежит под нижней границей полосы, а вторая — над ней, подают сигнал к покупке (красные маркеры графика 6).

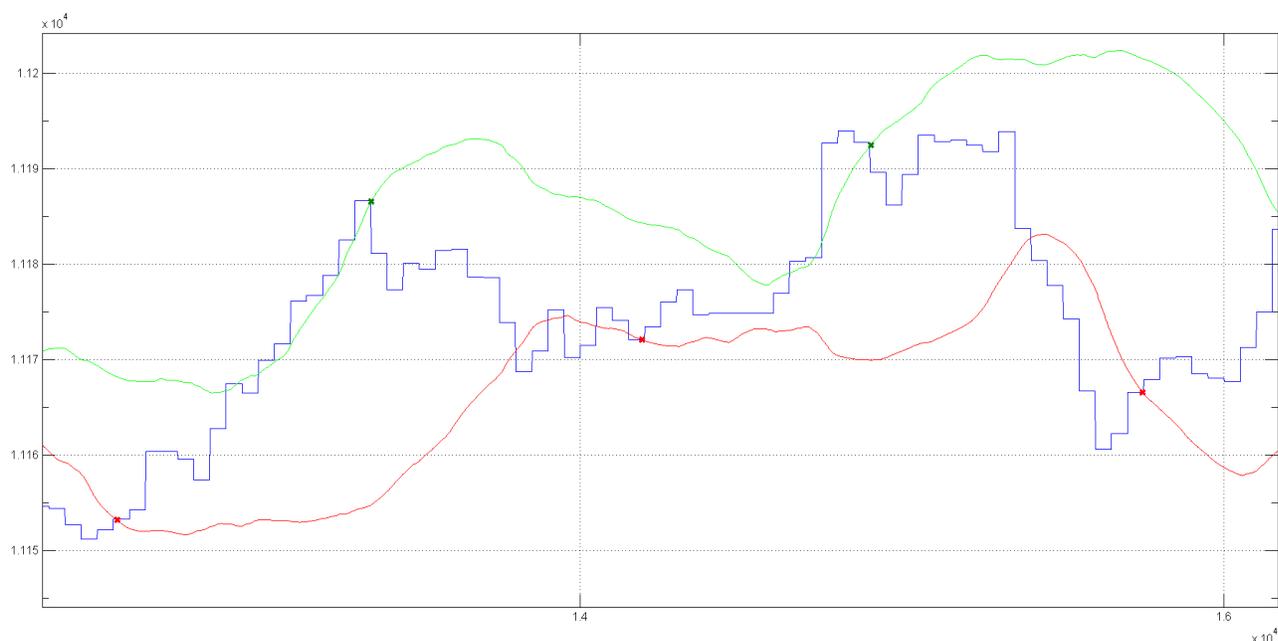


Рис. 6: Сигналы полосы Боллинджера

3.3 Осцилляторы скользящих средних

Для начала дадим определение осциллятора:

Определение 2. *Осциллятор — это индикатор технического анализа, который характеризует состояние перекупленности (overbought) или перепроданности (oversold) рынка.*

Осциллятор, как правило, эффективен при стационарном состоянии рынка, когда цена двигается в пределах сравнительно узкого «рыночного коридора». Существует достаточно большое количество осцилляторов, но мы остановимся на наиболее успешном и популярном, который называется *MACD (Moving Average Convergence/Divergence, Схождение/Расхождение Скользящих Средних)*. MACD представляет собой разность двух скользящих средних, опять же короткого с окрестностью сглаживания n_1 и длинного с окрестностью n_2 ($n_2 > n_1$), которая вычерчивается в открытой шкале около нулевой линии. Точки на нулевой линии соответствуют моментам, когда значения обеих скользящих средних совпадают. График MACD обычно изображается в виде гистограммы (нижняя часть 7) с отмеченными на ней сигнальными линиями - зеленой (положительной) и красной (отрицательной).

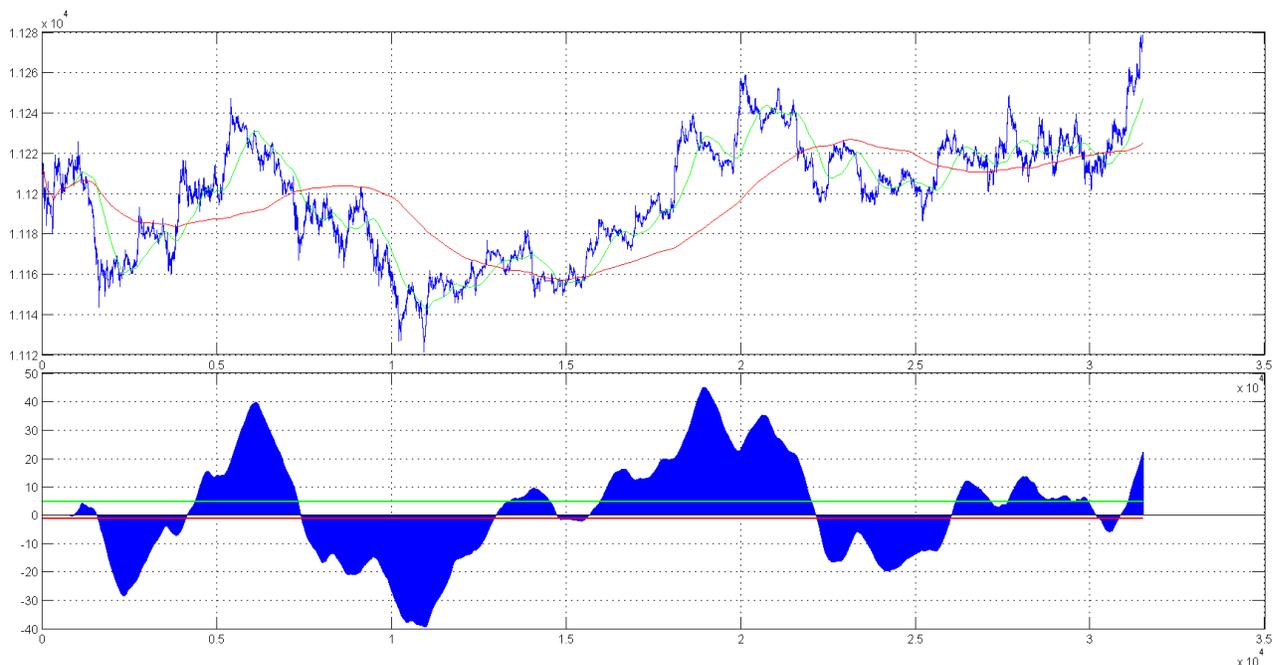


Рис. 7: Осциллятор MACD

Осциллятор MACD подает следующие торговые сигналы:

- К покупке:
 - MACD поднимается выше нулевой линии
 - MACD поднимается выше положительной (зеленой) сигнальной линии
- К продаже:
 - MACD опускается ниже нулевой линии
 - MACD опускается ниже отрицательной (красной) сигнальной линии

4 Вычислительный эксперимент

Теперь, изучив важные основные определения и методы, самое время перейти от теории к практике.

После победы в соревновании *Inventum Data Mining Contest* [4], проводимом нашей кафедрой в декабре 2012 года, я продолжил сотрудничество с компанией *Inventum*, занявшись разработкой алгоритма, позволяющего делать сделки на бирже в нужные моменты, принося тем самым прибыль.

Для разработки всех алгоритмов использовался пакет прикладных программ для решения задач технических вычислений MATLAB [5], позволивший свести к минимуму временные затраты на работу с матрицами, а также сильно упростивший зрительное восприятие получающейся в процессе работы алгоритмов картины за счет построения цветных графиков с возможностью использования маркеров. Для систематизации полученных результатов и возможности использования прошлых версий кода используемых алгоритмов, применялась система контроля версий TortoiseSVN [6].

Перейдем непосредственно к постановке вычислительного эксперимента. Изначально у меня было 48 временных рядов, каждый из которых соответствовал одному рабочему торговому дню на бирже. Отдельно взятый ряд представлял собой зависимость цены от времени (в формате часы:минуты:секунды:миллисекунды), а значения цены были известны только в моменты совершения сделок (то есть иногда практически каждую миллисекунду, иногда за пару секунд было всего 1-2 значения), поэтому все ряды имели различную длину.

Первым делом необходимо было все ряды привести к единому формату, и для этого был использован такой прием, как **дискретизация**. Весь интервал рабочего дня от 10:00 до 18:45 был поделен на секундные интервалы, и в каждом из этих интервалов за значениями цены взято среднее арифметическое значение цен сделок за эту секунду. Если в данном интервале не совершено ни одной сделки, то интервал следует заполнить предыдущим значением цены, поскольку в реальной ситуации нет возможности заглянуть в будущее. Таким образом, все полученные ряды имеют длину в $(18 - 10) * 3600 + 45 * 60 + 1 = 31501$ точку.

В дальнейшем для вычисления различных скользящих средних был создан большой массив сглаживаний (размером чуть больше, чем 1.7 Гб), в котором хранились сглаженные копии всех 48 рядов с окрестностями от 50 до 1000 точек с 50-точечным интервалом. Из них впоследствии были выбраны лишь некоторые, дающие наилучший результат при использовании выше описанных алгоритмов, а именно для алгоритма «Комбинация скользящих средних» понадобились сглаживания по 1000 и 8500 точкам, а для MACD - сглаживания по 750 и 4950 точкам.

На подбор всех параметров алгоритма, состоящего из трех независимых частей:

- Алгоритм «Комбинация скользящих средних»,
- Алгоритм «MACD»,
- Алгоритм «Полосы Боллинджера»,

ушло достаточно большое количество времени (около 3.5 часов), но и результат стоил того - на всех 48 рядах было сделано 3196 сделок (покупок и продаж вместе взятых) и заработано 8856.750 у.е. с учетом комиссии в размере 0.000085% от цены финансового инструмента. Получается, что мы совершили в среднем по 66 - 68 сделок на каждом графике за один рабочий день. Приведем график временного ряда, на котором отображены все сделки, совершенные за этот рабочий день 8.

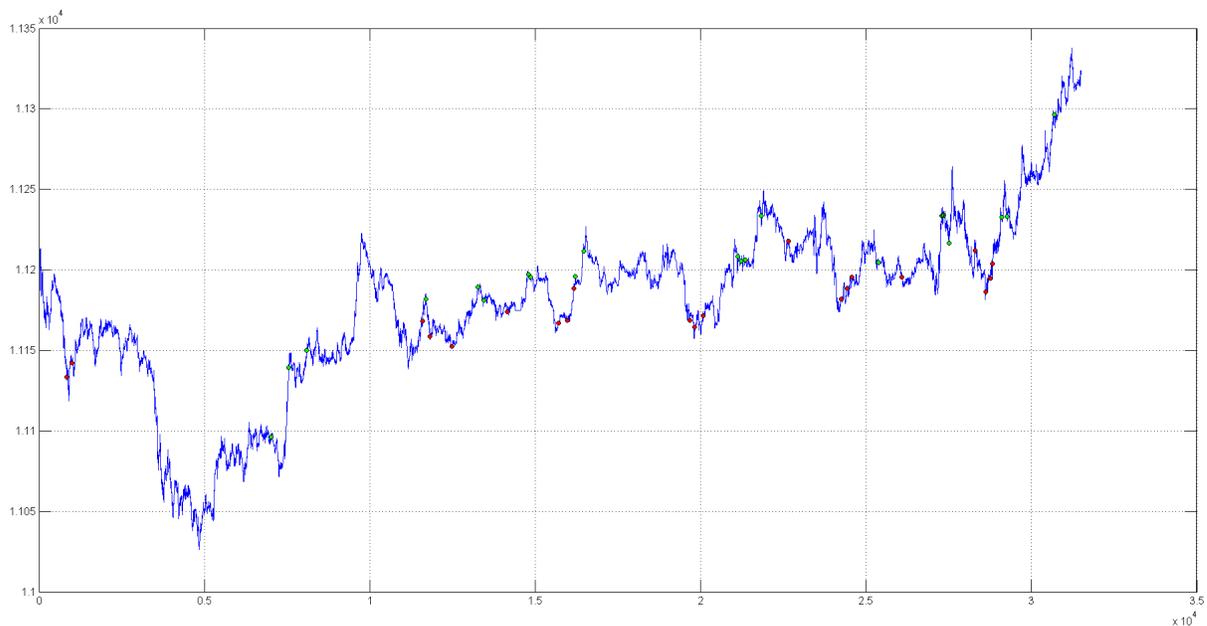


Рис. 8: Вычислительный эксперимент

На нем зелеными маркерами отмечены продажи (они увеличивают текущий баланс), а красными - покупки (они забирают деньги с баланса на покупку единиц финансового актива). Если посчитать количество покупок и продаж, то окажется, что они равны - этого мы и добивались. Даже если учитывать модель «1 – 2 – ... – 2 – 1», то количества покупок и продаж все также остаются равными, поскольку первой идет покупка, а последней - продажа.

5 Оценка вычислительной сложности алгоритма

Составим общую оценку вычислительной сложности алгоритма из оценок для отдельных частей (с учетом того факта, что $n \approx 30000$):

1. Дискретизация данных занимает время $\underline{O}(kn)$, где $k \in [0; 1000]$ — число сделок за 1 секунду. Однако все использованные временные ряды имеют длину не более 210000 тиков, поэтому окончательный результат для первой части алгоритма выглядит так: $\underline{O}\left(\frac{210000}{30000}n\right) = \underline{O}(7n)$
2. Вычисление скользящих средних (для корректной работы алгоритма их необходимо 4 штуки) занимает время $\underline{O}((1000 + 750 + 8500 + 4950)n) = \underline{O}(14750n) \approx \underline{O}\left(\frac{n^2}{2}\right)$
3. Первая часть основного алгоритма (алгоритм «Комбинация скользящих средних») затрачивает $\underline{O}(3n)$ на поточечное сравнение скользящих средних и нахождение подходящих моментов времени для сделок, а также $\underline{O}(C_1)$, $C_1 = const$ для расстановки сделок в «массиве сделок» с учетом комиссий и модели «1 – 2 – ... – 2 – 1»
4. Вторая часть основного алгоритма (алгоритм «MACD») затрачивает $\underline{O}(5n)$ на вычисление поточечных разностей скользящих средних и нахождение нужных моментов времени для сделок (подъемов и спусков), а также $\underline{O}(C_2)$, $C_2 = const$ для расстановки сделок в «массиве сделок» с учетом комиссий и модели «1 – 2 – ... – 2 – 1»
5. Третья часть основного алгоритма (алгоритм «Полосы Боллинджера») затрачивает гораздо больше, чем предыдущие основные составляющие:
 - Более крупная дискретизация затрачивает $\underline{O}(50n)$
 - Вычисление границ полосы Боллинджера занимает $\underline{O}(2Nn) = \underline{O}(2 * 500n) \approx \underline{O}\left(\frac{n^2}{25}\right)$, где N - окрестность вычисления границ
 - Вычисление пересечений цены с границами полосы Боллинджера требует $\underline{O}(2n)$ времени
 - Проход по точкам пересечения выполняется за $\underline{O}\left(\frac{n}{2}\right)$
 - Используется $\underline{O}(C_3)$, $C_3 = const$ для расстановки сделок в «массиве сделок» с учетом комиссий и модели «1 – 2 – ... – 2 – 1»

Учтем также, что необходимо потратить $\underline{O}(2n)$, чтобы сложить все усилия алгоритма и получить конечный результат. Таким образом, общая вычислительная сложность алгоритма равна

$$\underline{O}\left(7n + \frac{n^2}{2} + 3n + C_1 + 5n + C_2 + 50n + \frac{n^2}{25} + 2n + 2n + \frac{n}{2} + C_3\right) = \underline{O}\left(\frac{26n^2}{50} + \frac{139n}{2} + C\right),$$

$$C = C_1 + C_2 + C_3 = const$$

6 Практическое применение

Описанный в части 4 основной алгоритм был также использован на реальных данных фьючерсов компании «Газпром» за март-апрель 2013 года (пример одного дня с отображением линий Боллинджера можно увидеть на графике 9), которые были получены с сайта [7].



Рис. 9: Фьючерсы компании «Газпром»

Работа с такими данными немного отличалась от той, что была описана в вычислительном эксперименте в силу различной структуры исходных данных. Так, например, сделки в новых данных производились с 10:00 до 23:50, а не до 18:45 как было раньше. Однако это нововведение не сильно отразилось на алгоритме и он все также остался работоспособен.

После настройки всех необходимых параметров по данным за март, на апрельских данных был получен результат в 5458.634 у.е, делающий порядка 2000 сделок, что как раз примерно соответствует $\frac{2}{3}$ от заработка за 48 рядов, рассмотренных в вычислительном эксперименте. Этот факт говорит о том, что комбинация предложенных мною алгоритмов с правильно подобранными параметрами способна приносить **реальную положительную прибыль** на биржевых торгах.

7 Заключение и выводы

В заключении данной работы хотелось бы отметить, что прогнозирование временных рядов в ближайшее время станет наиболее популярной областью для исследований и экспериментов, поскольку оно может приносить (и даже уже сейчас приносит многим) высокую материальную прибыль. Именно поэтому будет появляться все больше и больше качественных и быстрых алгоритмов прогнозирования (даже в интерактивном режиме с постоянным получением текущих, актуальных данных о котировках финансовых инструментов), что предоставит людям, желающим начать исследования в этой сфере, большой интеллектуальный, информационный и алгоритмический фундамент для последующей работы и развития.

Список литературы

- [1] Лука К. Применение технического анализа на мировом валютном рынке Forex — М., Евро, 2003
- [2] Афанасьев В.Н., Юзбашев М. М. Анализ временных рядов и прогнозирование — М., Финансы и статистика, Инфра-М, 2010
- [3] Орлов Ю. Н., Осминин К. П. Нестационарные временные ряды. Методы прогнозирования с примерами анализа финансовых и сырьевых рынков — М., Либроком, 2011
- [4] http://www.machinelearning.ru/wiki/index.php?title=%D0%A1%D0%BE%D1%80%D0%B5%D0%B2%D0%BD%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_Inventum_Data_Mining_Contest
- [5] <http://www.mathworks.com>
- [6] <http://tortoisesvn.net>
- [7] www.finam.ru
- [8] <http://en.wikibooks.org/wiki/LaTeX>
- [9] Воронцов К.В. \LaTeX 2 ϵ в примерах
- [10] Львовский С. М. Набор и вёрстка в пакете \LaTeX . — М., Космосинформ, 1994.