

ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ ВЫСОКОЧАСТОТНОЙ БИРЖЕВОЙ ТОРГОВЛИ

Гасанов И.И., Ерешко Ант.Ф.

ВЦ РАН ФИЦ ИУ РАН, Россия, Москва, ул. Вавилова д.40

gasanov48@yandex.ru, asprs@yandex.ru

Аннотация: Описывается программный комплекс, предназначенный для автоматизированной торговли на фондовом и срочном рынках. Разработка относится к классу высокочастотных роботов. Робот эффективно использовался для фьючерсной торговли по индексу РТС и на протяжении нескольких лет завоёвывал призовые места на проводимых биржей конкурсах инвесторов.

Ключевые слова: биржа, робот, программный комплекс.

Введение

Алгоритмический трейдинг уверенно обосновался на всех мировых биржах и, как подтверждает статистика, используется для большинства торгуемых инструментов [1]. Значительная часть сделок на фондовых биржах осуществляется автоматическими системами в режиме высокочастотной торговли (HFT-алгоритмы). В [2] отмечается: в 2007 г. объем биржевых сделок на Чикагской товарной бирже (СМЕ) достиг 2,2 миллиарда контрактов стоимостью \$1,1 квадрильона; три четверти из них были заключены посредством электронных торгов.

Официальным началом использования алгоритмов является 1998 год, когда американский регулятор SEC (U.S. Securities and Exchange Commission) в США разрешил применение электронных площадок. Высокая эффективность алгоритмического трейдинга привлекает внимание профессиональных трейдеров, и они предпринимают значительные усилия для достижения ещё большей эффективности.

Системы автоматической торговли, в том числе высокочастотной, активно используются в решении проблем хеджирования, маркет-мейкинга, арбитража, управления клиентскими позициями. Существенная доля торговых операций, совершаемых автоматизированными системами, производится со спекулятивными целями.

В настоящей работе рассматриваются типичные проблемы, с которыми сталкиваются разработчики моделей и программ, такие как: подключение к торгам, актуальность биржевой информации, способы тестирования и нахождения оптимальных настроек.

1 Организация биржевой торговли

Задачей проекта было создание программного комплекса для проведения алгоритмической торговли на биржевых площадках. На мировых биржах торгуются многочисленные инструменты (товары). В общих чертах торговля разными инструментами на биржах организована, примерно, по одним и тем же принципам.

В изложении будем рассматривать срочный рынок и какой-то один инструмент. Вся торговля проходит в компьютеризированном, электронном виде. Сделки производятся согласно ордерам, поступающим на биржу от участников торгов по цифровым каналам связи.

Каждый ордер – это заявка на операцию с каким-то одним из торгуемых инструментов. Используются также ордера на снятие выставленных заявок.

Существует несколько типов заявок.

Мы здесь будем рассматривать только, так называемые, лимитные заявки. Лимитная заявка характеризуется множеством параметров, из которых для нас важны следующие:

- Направление заявки, т.е. какая это заявка – на покупку товара или продажу.
- Цена заявки, т.е. та наименее выгодная цена, на которую согласен участник.
- Объем заявки в лотах (контрактах).
- Время авторизации, т.е. момент, когда заявка была обработана биржей.

Цены заявок не произвольны, а имеют минимальный шаг дискретизации h . Также дискретизированы и объёмы заявок, которые измеряются в лотах.

Поступающие от участников торгов заявки встают в очередь на обработку. Время от поступления заявки до её обработки биржей очень невелико, но всё же существенно для участника. Это время не фиксировано и зависит от интенсивности потока поступающих заявок. Оно уменьшается с развитием используемой вычислительной техники.

В момент обработки новой заявки биржа пытается свести её с противоположной по направлению ранее авторизованной заявкой. Под сведением понимается осуществление сделки между инвесторами, подавшими сводимые заявки. Пусть, для примера, обрабатывается заявка на покупку B с ценой C^B и объёмом V^B . Биржа ищет среди уже обработанных, но не реализованных заявок на продажу такие, у которых цена $C^A \leq C^B$. Если такие заявки имеются, то оформляется сделка.

Рассматриваемая заявка на покупку будет сведена в первую очередь с уже имеющейся заявкой на продажу A с минимальной ценой C^A . Если таких заявок несколько, то сначала с той из них, которая была авторизована раньше. Если операция не исчерпывает объём обрабатываемой заявки полностью, т.е. если объём заявки A , $V^A < V^B$, то несведённый остаток заявки B , $V^B - V^A$ биржа пытается свести со следующей по приоритету имеющейся заявкой на продажу A' . Это будет заявка с той же ценой C^A , но авторизованная после заявки A , если такая имеется. Если такой заявки нет, то для сведения берётся самая ранняя заявка с ценой $C^{A'} = C^A + h$ и т.д. При сведении двух заявок сделка всегда производится по цене той заявки, которая была авторизована раньше.

Если лимитная заявка не может быть сведена при авторизации полностью, по причине отсутствия контрагентов, то она с оставшимся объёмом присоединяется к множеству других несведённых заявок. После этого она находится в ожидании сведения с вновь поступающими заявками или снятия создавшим её инвестором. Отметим, что такая заявка в дальнейшем может быть сведена не только с лимитными заявками, но и с противоположными заявками других типов.

Множество лимитных заявок, находящихся в ожидании сведения разделено на два структурированных подмножества, это т.н. Книги заявок (Order Books), которые также принято называть Стаканами.

Одна Книга заявок, стакан Бид – это заявки на покупку; другая Книга заявок, стакан Аск – это заявки на продажу. Каждый из стаканов рассекается на уровни B_0, B_1, \dots, B_n для Бид и A_0, A_1, \dots, A_n для Аск. Уровень – это множество заявок с одинаковой ценой. Цены уровней будем обозначать теми же буквами A и B . Лучшие предложения – это цены A_0 и B_0 . Отрезок $A_0 - B_0$ будем называть спредом. Чем больше индекс у уровня, тем дальше он отстоит от спреда, т.е. $B_n \leq \dots \leq B_1 \leq B_0$ и $A_0 \leq A_1 \leq \dots \leq A_n$.

Понятно, что минимальное расстояние между уровнями равно шагу цены h . Но уровень, на котором нет заявок, игнорируется, поэтому между двумя соседними уровнями может быть разница в цене в несколько минимальных шагов.

Помимо цены, уровень характеризуется количеством и общим объёмом составляющих его заявок. Биржа регулярно и с высокой частотой передаёт участникам торгов информацию о произведённых сделках и текущем состоянии стаканов, указывая параметры каждого уровня. Делается это либо через заявленные промежутки времени, либо при авторизации каждой очередной заявки.

Участников торгов можно разбить на несколько типов, в зависимости от целей торговли.

Это трейдеры, которые выполняют поручения других инвесторов. Даже если трейдеры не инвестируют в рынок сами, они заинтересованы в выгодных ценах своих сделок, т.к. получают от этого дополнительную прибыль.

Это покупатели биржевого товара с целью использования его в своей экономической деятельности. Например, нефти, зерна, валюты. Как правило, дополнительный доход, связанный с сиюминутным колебанием цен при биржевых операциях у такого участника не сопоставим с денежными потоками от его долговременных операций и не является предметом его интереса.

То же относится и к участникам, использующим торговлю фьючерсами и опционами для хеджирования долговременных финансовых операций.

Наконец, значительную долю биржевых агентов составляют игроки, пытающиеся заработать на биржевых спекуляциях. Они с разной частотой продают и покупают биржевые товары, стремясь извлечь выгоду из разницы между ценой покупки и ценой продажи. Несмотря на то, что их операции

никак, не влияют на экономическое бытие товара вне стен биржи, они всё же играют определённую позитивную роль в экономике, обеспечивая, как бы, смазку для механизма биржи – обеспечивая бирже ликвидность. Кроме того, спекулянты выгодны бирже тем, что, приносят ей значительный доход на комиссиях от операций. Если спекулятивная фракция биржевых агентов в совокупности получает прибыль, то происходит это за счёт участников других типов, которые, как бы, платят спекулянтам за дополнительную ликвидность операций. Именно разработка и реализация алгоритмов для проведения спекулятивных биржевых операций были целью описываемого проекта.

2 Концепция проекта

Будем рассматривать некоторый актив, торгуемый на одной из биржевых площадок. Эта торговля представляет собой развивающийся во времени многомерный процесс генерации новых ордеров инвесторами, авторизации, сведения и снятия заявок. Частью этого процесса является динамика биржевых цен, которая понимается нами как динамика стаканов и которая полностью определяется потоком поступающих от участников заявок разных типов, их последовательной авторизацией биржей и сведением с другими заявками, уже находящимися в стаканах. В более узком смысле, динамику цен можно ассоциировать с изменением лучших предложений покупки и продажи актива, т.е. B_0 и A_0 .

В каждый момент времени биржа обрабатывает не более одной новой заявки. Моменты авторизации – это некоторые точки на временной оси $t_0, t_1, \dots, t_n, \dots$, где через t_0 обозначен момент начала торгов, а через t_n – текущий момент. Промежутки между очередными авторизациями варьируются и зависят как от производительности биржевого сервера, так и от интенсивности торговли. Это время составляет от долей миллисекунды до нескольких секунд.

Весь поток авторизуемых заявок в самом общем виде можно выразить формулой

$$\Phi = o_0, o_1, \dots, o_n, \dots,$$

где o – вектор параметров очередной заявки.

Динамика стаканов, как вектор-функции от потока заявок, в общем виде запишется формулами

$$\tilde{B}_n(o_0, o_1, \dots, o_n) \text{ и } \tilde{A}_n(o_0, o_1, \dots, o_n)$$

Здесь \tilde{B}_n и \tilde{A}_n – текущие стаканы Бид и Аск. Это текущие "слепки" множества лимитных заявок, которые были ранее авторизованы, но ещё не исполнены целиком и не сняты их авторами.

Поток Φ зависит от множества различных факторов. Концептуально этот поток можно выразить формулой $\Phi(R)$, где определяющие его факторы обозначены через $R \in \Theta$. Часть из них – это показатели, доступные для наблюдения инвесторами. Обозначим их через P . Это, например, уже авторизованные к моменту наблюдения заявки, текущий и предшествующие стаканы, биржевые сводки по всему множеству инструментов.

Словесно цель проекта можно выразить так: хотелось бы создать программу-робот, управляющей ордерами инвестора посредством функций от множества показателей P , которая за счёт сделок купли-продажи актива приносила бы инвестору совокупный доход.

Совершая сделки купли-продажи, инвестор изменяет состав корзины своих контрактов и объём оборотных средств. В корзине могут быть лоты на покупку или на продажу актива. Если, предположим, к моменту t_n корзина состоит из m купленных лотов актива и в этот момент приобретает ещё μ лотов по цене C_{t_n} , то в корзине оказывается $m + \mu$ купленных лотов, а оборотные средства уменьшаются на величину $\mu \cdot (C_{t_n} + \nu)$, где ν – комиссия биржи. Если μ лотов продаётся, то денежные средства инвестора возрастают на величину $\mu \cdot (C_{t_n} - \nu)$. При этом, если $\mu \leq m$, то в корзине остаются $m - \mu$ купленных лотов, но если $\mu > m$, то тогда корзина будет содержать $\mu - m$ проданных лотов. Количество лотов в корзине именуется позицией. Позиция

определяется как число со знаком: если лоты в корзине – это купленный актив, то знак положительный, если проданный – отрицательный.

Доход по операциям инвестора в течение текущей сессии к моменту t_n – это сумма двух величин: денежный доход (траты) инвестора от начала сессии до момента t_n и стоимость лотов в корзине (стоимость позиции). Пусть, к примеру, доход инвестора к моменту t_n равен I_{t_n} , его позиция составляет m купленных лотов и на шаге t_n покупается ещё μ лотов по цене C_{t_n} . Тогда текущий доход инвестора к моменту $n+1$ составит величину

$$I_{t_{n+1}} = I_{t_n} - \mu \cdot v + (m + \mu) \cdot (C_{t_{n+1}} - C_{t_n})$$

Рассматривается торговая стратегия, при которой инвестор совершает сделки купли и продажи в течение сессии, но закрывает (обнуляет) позицию к моменту её окончания. Доход инвестора за сессию I_T составит сумму из дохода к моменту последней сделки и цены этой последней сделки.

Формально цель проекта можно записать так. Для торгового робота формируются управления $U_t(P|_{t-\tau}^t, q_t, Q_t)$ следующего вида.

Обозначим как $P|_{t-\tau}^t$ показатели P на отрезке времени $[t-\tau, t]$.

Через q_t будем обозначать позицию инвестора в момент t , а через Q_t – множество его заявок, находящихся в момент t в стаканах.

Пусть задана, вообще говоря, многомерная функция $\varphi_t(P|_{t-\tau}^t)$, которую будем называть признаком, а также, $\Omega_t^+, \Omega_t^-, \Omega_t^0$ некоторые области значений этого признака.

Если в момент t признак $\varphi_t(P|_{t-\tau}^t) \in \Omega_t^+$, то выставляется ордер на покупку актива. Назовём это сигналом к покупке.

Объём ордера V_t зависит от текущей позиции инвестора q_t и заявок инвестора Q_t , уже находящихся в стаканах, т.е. $V_t = V_t(q_t, Q_t)$. Это может быть и нулевой объём, если достигнут заданный инвестором лимит позиции. В этом случае заявка не формируется.

Цена заявки $C_t(P|_{t-\tau}^t)$ также может зависеть от наблюдаемых факторов.

Если $\varphi_t(P|_{t-\tau}^t) \in \Omega_t^-$, то это сигнал к продаже. В этом случае аналогичным образом формируется ордер на продажу.

Если $\varphi_t(P|_{t-\tau}^t) \in \Omega_t^0$, то снимается заявка инвестора, уже авторизованная и находящаяся в стакане.

Таким образом, управление $U_t(P|_{t-\tau}^t, q_t, Q_t)$ – это ордер, формируемый в момент t с атрибутами объёма V_t и цены C_t .

Если на некотором временном промежутке $[t^0, T]$ инвестор задаёт стратегию управлений $S = \{U_t(P|_{t-\tau}^t, q_t, Q_t)\}_{t^0}^T$, то она однозначно определяет его ордера на потоке заявок $\Phi|_{t^0}^T$, траекторию его позиции $q|_{t^0}^T$, торговые сделки с его заявками и доход $I_t(S)$. Инвестор хочет, чтобы доход I_T был положительным и как можно большим. Мы понимаем поток Φ как нестационарный вероятностный процесс. Такой подход позволяет записать цель инвестора как максимизацию математического ожидания

$$M(I_T(S)) \xrightarrow{S} \max.$$

Следует заметить, что приведённая выше запись управлений не является максимально общей. В то же время, в практической работе использовался ещё более узкий класс управлений. Признаки φ были

скалярными функциями от некоторого подмножества показателей P , причём использовались только данные о ценах и заявках на той биржевой площадке, для которой и разрабатывалась торговая стратегия. Области Ω – задавались барьерными величинами β_i и $-\beta_i$. Сигналом к операции служил переход значения признака φ_i за барьер. То есть:

- если для $t' \in [t - \delta, t)$: $\varphi_{t'} < \beta_i$, а $\varphi_t \geq \beta_i$, то это служило сигналом для формирования заявки на покупку актива.

- если для $t' \in [t - \delta, t)$: $\varphi_{t'} > \beta_i$, а $\varphi_t \leq \beta_i$, то это служило сигналом для формирования заявки на покупку актива.

Формировался или нет ордер зависело от текущей позиции робота, т.е. от того, не приведёт ли его исполнение к превышению определённого инвестором максимума корзины.

Цены формируемых роботом заявок C_i имели фиксированный сдвиг относительно $B0_i$ для покупки и $A0_i$ для продажи.

Кроме того, чтобы подробности не уводили от сути проблем, здесь опущены некоторые детали. Так, например, в нашем случае торговля не предполагает одновременного наличия в стаканах нескольких созданных роботом лимитных заявок. И это регулируется дополнительными условиями на выставление и снятие заявок.

Следует иметь в виду, что поток Φ формируется огромным числом факторов из множества R . Какие-то из них не известны инвестору по значениям, про какие-то он вообще не знает об их существовании. В них входят не только внутренние биржевые факторы, но и события мировой экономики и политики. Значительная их часть не выражается числом, носит разовый, уникальный характер. Поэтому, трудно надеяться на долгосрочный прогноз не только стаканов \tilde{B}_n и \tilde{A}_n , но и каких-то отдалённых характеристик для их распределений, типа математических ожиданий и дисперсий. Эти соображения послужили поводом для поиска решений в классе высокочастотных роботов. В этом случае горизонт прогноза будет достаточно малым, что позволяет надеяться на его достоверность.

Разработанный алгоритм функционирует в течение отдельной торговой сессии, открывая позицию в её начале и закрывая в конце. Небольшой ряд последовательных сессий – это промежуток времени, о котором можно предполагать, что факторы R на его протяжении остаются, примерно, одинаковыми. Будем ориентироваться не на весь спектр событий из множества Θ , а на некоторое его подмножество Θ' , отбрасывая возможные катастрофические сценарии. Они случаются, но далеко не каждую сессию. От робота требуется, чтобы он был эффективен в обычных условиях и защищён от больших потерь при форсмажоре. (От серьёзных потерь защищает залоговая сумма). Если в течение сессии не происходит событий из множества $\Theta \setminus \Theta'$, то будем её понимать как стандартную. В черед таких сессий при большом числе операций, производимых роботом, можно говорить о регулярности и статистической значимости испытаний (операций). Можно предполагать, что факторы R , определяющие процесс, за несколько последовательных стандартных сессий не успевают поменяться слишком существенно, и рассматриваемый процесс в этом промежутке времени близок к периодическому.

Предположим, в момент t поступает сигнал на покупку $\alpha^+(P|_{t-\tau}^t)$. Пусть t' – первый после t момент, когда реализуется сигнал $\alpha^-(P|_{t'-\tau}^{t'})$ на продажу актива. Сигналу $\alpha^+(P|_{t-\tau}^t)$ соответствует вероятностное распределение цен в момент t' . Обозначим через $\theta_i^B(R_t)$ и $\theta_i^A(R_t)$ распределения величин $B0_{t'} - B0_t$ и $A0_{t'} - A0_t$ на момент первого обратного сигнала t' . Полагаем, что параметры R на соседних стандартных сессиях остаются, примерно, одинаковыми. Тогда можно надеяться, что и математические ожидания $M\theta_i^B$ и $M\theta_i^A$, приблизительно, одинаковы для сигналов, если эти

сигналы относятся к соседним сессиям, но отстоят от начала сессий, примерно, на одинаковом расстоянии.

Для целей торговли при большой частоте операций достаточно использовать только эти оценки. Если средний ожидаемый рост цен при сигнале о покупке будет достаточно большим, то есть основания для заявки на покупку актива. То же для продажи, но с обратным знаком.

Расчёт был на то, что существуют и удастся отыскать такие сигналы, которые обладают достаточно большой инерцией относительно изменения параметров R в области Θ' . Так, что величины $M\theta_t^B$ и $M\theta_t^A$ не слишком сильно меняются в течение значимого периода времени в несколько сессий и при этом достаточно велики, чтобы доход от операций компенсировал затраты на их обслуживание (комиссионные, оборудование). Так как со временем изменения в факторах R нарастают, то и параметры сигналов U_t следует корректировать, адаптировать к этим изменениям.

Приведённые рассуждения носят неформальный, качественный характер. Но они задают направление для развития проекта. Если надежды оправданы, то существуют такие сигналы, которые при удачных настройках их параметров дают средний положительный доход на протяжении какого-то относительно длительного промежутка времени.

Тогда ряды данных о прошедших торгах можно использовать для поиска и тестирования таких сигналов, а затем использовать на протяжении некоторого времени в практической работе.

Сигналы определяются функциями от показателей P . Оптимальные значения параметров этих функций зависят от текущего состояния факторов R . Для настройки этих параметров на очередную сессию можно использовать некоторое множество ей предшествующих.

Таким образом мы приходим к поиску и настройке управлений на рядах наблюдений, и **центральное место в проекте занимает разработка программного имитационного комплекса**. Он предназначен для того, чтобы для тестируемой стратегии управления с приемлемой точностью рассчитывать динамику позиции и доходов так, как если бы она использовалась роботом на множестве уже прошедших биржевых сессий.

3 Краткий обзор аппаратно-программного комплекса

Комплекс составлен из двух крупных взаимодополняющих частей.

Одна часть системы – это собственно торговый робот. Это программа, инсталлированная на сервере пользователя в дата-центре и коммутированная по каналам связи с сервером биржи.

У торгового робота следующие функции.

1. Робот получает с биржевого сервера и обрабатывает обширную информацию о текущих торгах. Это поток авторизованных заявок Φ и данные об исполненных сделках. Все данные сохраняются на диске пользовательского компьютера и образуют "сырой" лог сессии.

Кроме того, по этим данным робот в режиме реального времени формирует рабочий лог текущей сессии. Он содержит последовательные слепки стаканов, сделанные с некоторым временным шагом. Данные каждого временного шага составляют отдельную строку записи в рабочем логе.

Минимальный шаг для очередного слепка стаканов η задаётся пользователем и составляет несколько миллисекунд. Однако в реальности он может быть и большим. Это минимум между величиной η и двумя последовательными авторизациями биржевых заявок.

Рабочие логи также содержат информацию о собственных операциях робота, о его заявках поставленных в очередь на авторизацию, об их авторизации и сведении, о текущей позиции и текущем доходе робота.

2. В ходе торгов робот, согласно встроенному алгоритму, используя данные за несколько последних шагов логирования, формирует сигналы. Если сигнал указывает на операцию, то робот формирует соответствующую заявку (ордер) и пересылает её бирже.

Следует подчеркнуть, что модуль робота, отвечающий за алгоритм управления имеет широкие возможности настройки. Это позволяет использовать самые разные алгоритмы управления на базе параметров, которые включены в рабочие логи.

Другая часть системы – это имитационный комплекс. Основным его инструментом является виртуальный экземпляр робота. Виртуальному роботу на вход подаются сформированные ранее реальным роботом рабочие логи биржевых торгов.

Модуль виртуального робота, обслуживающий алгоритм управления, настраивается на какой-то, сконструированный пользователем и предназначенный для тестирования торговый алгоритм.

Робот, согласно этому алгоритму, идёт от шага к шагу логированных данных и осуществляет виртуальную торговлю, формируя виртуальные ордера, а также, рассчитывает требуемые показатели, в частности, текущие и итоговые виртуальные доходы.

Моделирование исполнимости заявок и коррекция доходов. При реальных торгах биржа сводит заявки, отслеживает позиции участников, их текущие доходы, текущие объёмы залоговой суммы и регулярно информирует каждого из участников об этих параметрах.

При виртуальных торгах робот, согласно его торговому алгоритму, добавляет к реальным заявкам свои виртуальные заявки. Так как они не участвуют в реальных торгах, то их исполнение, также виртуальное, должно моделироваться внутри имитационного комплекса.

Хотя сами процедуры сведения заявок известны участникам, точно рассчитать их исполнение на том или ином шаге модели не представляется возможным. Дело в том, что от формирования заявки роботом, до её авторизации проходит некоторое время, и это время, вообще говоря, плавающее. Оно зависит от интенсивности торгов.

Кроме того, не удаётся точно синхронизировать астрономическое время на биржевом и клиентском серверах, а значит, невозможно точно определить, когда заявка робота попадёт в очередь на авторизацию. По техническим соображениям, чтобы они не были слишком "тяжёлыми", логи подаваемые на вход виртуальному роботу, не содержат каждую авторизуемую заявку по отдельности.

Рабочие логи проходят предварительную обработку, в процессе которой рассчитываются интегральные данные об объёмах заявок за период между моментами логирования. Поэтому исполнение виртуальных заявок робота приходится вычислять приближённо. Для этого использовалась следующая эвристическая процедура.

Делается посылка, что задержка от формирования заявки до её авторизации составляет одинаковое время в процентах от длительности промежутка между шагами логирования (строками лога). Обозначим его τ . Рассмотрим произвольную строку лога n . Ей соответствует время логирования t_n . Допустим в момент t_n робот принимает решение о формировании нового ордера. Условно считаем, что новая заявка формируется в этот же момент t_n . Тогда авторизация заявки происходит в момент $t_n + \tau \cdot (t_{n+1} - t_n)$.

Также предполагается, что объёмы авторизуемых заявок каждого вида и цены в промежутке времени $[t_n, t_{n+1}]$ изменяются линейно.

Этого достаточно, чтобы оценить объём сведения (исполнения) заявки робота при её авторизации. В том же предположении о линейности поступления заявок рассчитывается исполнение заявки робота на последующих шагах пребывания её в стакане и на шаге принудительного снятия.

Необходимо настроить параметр τ . Для этого используются логи таких сессий, на которых робот участвовал в реальной торговле. Брался массив из, примерно, 40 сессий – их число ограничивалось ресурсами компьютерной техники.

В реальной торговле каждая биржевая сессия была разбита на временные периоды с разными настройками параметров сигнала. Поэтому общее число тестовых периодов составляло около полутысячи.

Реальные доходы робота в каждый из этих периодов были известны. Требовалось подобрать такое значение задержки τ , которое минимизировало бы отклонение виртуальных доходов от реальных. В силу грубости метода, такое отклонение по периодам неизбежно, иногда значительное. Но важно сгладить интегральные величины.

Тестируя значения τ мы вводили поправочный коэффициент к виртуальным доходам так, чтобы совпадали суммарные реальные и виртуальные доходы по всему тестовому массиву. Поправочный коэффициент k рассчитывался по следующей формуле

$$k = \left(\sum_{j \in J, n \in N} I_{j,n}^{real} - \sum_{j \in J, n \in N} I_{j,n}^{vir} \right) / \sum_{j \in J, n \in N} D_{j,n}^{vir}$$

Здесь, j – номер торгового периода внутри сессии, n – номер сессии тестового полигона, $I_{j,n}^{real}$ – реальный доход, $I_{j,n}^{vir}$ – виртуальный доход до поправки, $D_{j,n}^{vir}$ – количество виртуальных сделок за период.

Затем выбиралось такое значение τ , которое минимизировало сумму модулей отклонений по сессиям между реальными и виртуальными доходами.

$$\sum_{j \in J, n \in N} \left| I_{j,n}^{real} - I_{j,n}^{vir} - k \cdot D_{j,n}^{vir} \right| \rightarrow \min.$$

Таким образом поправка доходов за торговый период зависела от интенсивности в нём торгов.

Доход за период при уже выбранном значении τ в процессе тестирования и настройки параметров сигнала вычислялся по формуле

$$I_{j,n}^{vir} = I_{j,n}^{real} + k \cdot D_{j,n}^{vir}.$$

Настройка параметров сигналов. Регулярная корректировка этих параметров, необходима, т.к. требуется их адаптация к изменяющимся факторам R .

Используемые в проекте сигналы определяются как функции достаточно большого числа различных параметров. И это создаёт проблему. Чем больше настраиваемых параметров вероятностного процесса, тем большей должна быть используемая для их оптимизации выборка. Иначе исследователь попадает в ловушку, приспособливая параметры не столько к вероятностному процессу, сколько к специфическим особенностям некоторого ограниченного множества его реализаций.

Но массив логов, пригодных для коррекции параметров, в случае большого количества параметров, будет неизбежно недостаточным. Его ограничивают, по меньшей мере, два обстоятельства.

Ресурсы, задействованной в проекте компьютерной техники не безграничны, а расчёты достаточно трудоёмки.

Ещё более существенно то, что изменения в факторах R не позволяют даже условно, приближённо описывать динамику рынка одним распределением в течение длительного отрезка из торговых сессий.

Оптимальные параметры в начале и конце периода настройки могут оказаться слишком далёкими. Необходим компромисс. Не сразу, но в результате длительных экспериментов мы пришли к следующей тактике коррекции параметров.

Большинство варьируемых параметров после многочисленных вычислительных экспериментов было фиксировано или подвергалось ревизии крайне редко. Это такие параметры, которые, как показали эксперименты, или в широкой области значений мало влияют на эффективность стратегии, или, напротив, давали однозначно лучший результат при каком-то из значений. К ним относятся глубина учитываемых уровней цен в стаканах, глубина временного отрезка данных – число шагов лога, предшествующих текущему моменту расчёта сигнала. К ним же относится и цена формируемой заявки. Она, как правило, была фиксирована на уровне $BO_t + h$ для покупки и $AO_t - h$ для продажи. Здесь h – биржевой шаг цены.

Значения барьера β_t корректировались каждую неделю. Для этого использовался, примерно, двухмесячный массив логов.

Чтобы снизить влияние подстройки параметра под специфику полигона данных, применялся следующий эвристический приём. Ретроспективно вычислялось оптимальное значение β на предшествующей расчёту неделе. Данное значение усреднялось с такими же оптимальными β , за предыдущие 9 недель. Это и было то значение барьера, которое использовалось на реальных торгах следующей недели.

Заключение

Описанный выше подход – это один из вариантов поиска эффективной стратегии управления в условиях сложной, многофакторной неопределённости методом оптимизации на исторических рядах данных. Здесь в качестве таковых использовались логи торговых сессий.

Подход не гарантирует успех, но в описываемом здесь случае удалось построить и поддерживать торговую стратегию, которая на протяжении нескольких лет завоёвывала призовые места на проводимых биржей конкурсах инвесторов.

Литература

1. *Marco Avellaneda*. Algorithmic and High-frequency trading: an overview – New York University & Finance Concepts LLC, 2011, 58 с.
<https://www.math.nyu.edu/faculty/avellane/QuantCongressUSA2011AlgoTradingLAST.pdf>
2. https://www.cmegroup.com/trading/agricultural/files/AC216_GrainOilseed__Hedging__Guide__Russian_SR.pdf. CME Group, 2010, 76 с.
3. <https://investingnotes.trade/algotrejding-rukovodstvo.html>
4. *Gasnov I.I., Ereshko Ant.F., Vakhranov A.* Methods of Monte-Carlo in the problems of control at the ranks of historical data. IEEE Xplore Digital Library. Tenth International Conference Management of Large-Scale System Development (MLSD), Moscow, Russia, 2017.
5. *Гасанов И.И., Ерешко Ант.Ф., Вахранев А.В.* Имитационные эксперименты с моделью книги заявок. Материалы десятой международной конференции «Управление развитием крупномасштабных систем MLSD'2017 (2-4 октября 2017 г., Москва, Россия), Том I, с. 193-195.