



Анализ эффективности стратегий для торговли опционами на Московской Бирже с применением методов машинного обучения

О.Е. Кудрявцев¹, Х.М. Мамедзаде¹, В.В. Родоченко², А.А. Чивчян¹

¹Ростовский филиал Российской таможенной академии

²НПФ «ИнВайз Системс», Ростов-на-Дону

Аннотация: В статье исследуется эффективность стратегий, основанных на предсказании динамики подразумеваемой волатильности для торговли опционами российского финансового рынка. Мы предсказываем поведение волатильности на основе имеющихся исторических данных при помощи случайного леса, используя эффект кластеризации волатильности. На основе полученных результатов мы принимаем решение об открытии и закрытии торговых позиций. Результаты экспериментов показывают, что такой аппарат может быть применён как один из механизмов принятия торговых решений.

Ключевые слова: опцион, торговые стратегии, случайный лес, индекс РТС, срочный рынок.

Введение

Как в России, так и за рубежом в настоящее время наблюдается дефицит математического и программного обеспечения, которое позволяет эффективно управлять финансовыми рисками на срочном рынке. В качестве общего языка для описания процессов на большинстве мировых срочных рынков служит классическая диффузионная модель Блэка-Шоулса [1], дискретным аналогом которой является биномиальная модель [2]. Известно [3], что данная модель плохо соответствует реальности: в частности, недооценивает риски (эффект «тонких хвостов» распределения), не дает возможности моделировать скачки цен и не может объяснить наблюдаемые эффекты «улыбки волатильности» и «кластеризации волатильности». Широкий спектр более общих моделей, насколько известно авторам, в настоящее время на российском финансовом рынке активно не используется, но представляет интерес для участников как перспективное направление и активно исследуется [4].

Среди необходимых составляющих такого обеспечения можно назвать автоматизированные системы для принятия торговых решений. Стратегии



для российского срочного рынка, использующие опционы, нечасто становятся предметом исследования из-за ряда трудностей, связанных со сбором, обработкой и анализом исторических данных. Еще одна проблема состоит в особенностях поведения инструмента, связанных с изменениями центральных страйков и небольшой ликвидностью. Тем не менее, для некоторых других рынков существует достаточно обширная литература, в которой приводятся такие стратегии и соображения, лежащие в их основе. В качестве источников такого рода можно упомянуть, например книгу [5].

В рамках данного исследования мы используем явление, известное как «кластеризация волатильности» («volatility clustering»), в качестве основы для принятия решений по открытию или закрытию торговых позиций. Неформальное его определение состоит в том, что «волатильность», некоторым образом рассчитанная, хотя большую часть времени и лежит в некотором диапазоне (зависящем от конкретного актива), иногда принимает более высокие значения и «застревает» в них, не сразу возвращаясь к обычным. Более детальное описание кластеризации можно найти, например, в статье [6].

Для того, чтобы предсказать появление этого эффекта, мы используем случайный лес классификации [7]. В качестве альтернативного метода мы применяем случайный лес регрессии, чтобы получить прогнозные значения волатильности. Мы подготовили для обучения выборку с данными о состоянии торгов по всей доске опционов за предыдущий торговый период, а также данные о ходе торгов на базовый актив (цену закрытия торгового периода). Исходя из спрогнозированного значения или состояния рынка, мы принимали решение о покупке/продаже опциона put или call.

Несмотря на тот факт, что мы рассмотрели довольно короткую историю торгов, оказалось возможным как спрогнозировать изменение волатильности и появление кластеризации, так и использовать полученные



данные для построения стратегий, которые оказываются прибыльными при тестировании на исторических данных.

Исследования выполнялись с активным привлечением студентов студенческой научно-исследовательской лаборатории «Управление риском и прогнозирование» Ростовского филиала Российской таможенной академии (руководитель О.Е. Кудрявцев, д.ф.м.н., проф.).

Основные определения

Опцион – это договор о передаче права, но не обязательства, купли-продажи актива по определённой цене в будущем.

Объект такой купли-продажи называется «базовым активом». Опцион имеет фиксированное время экспирации, по наступлении которого прекращает действие и обесценивается. Фиксированная цена носит название «страйк».

Под «волатильностью» мы понимаем некоторый статистический финансовый показатель, характеризующий изменчивость цены. В работах [8, 9] приведён ряд методов, которые могут быть использованы для расчёта такого показателя.

В рамках этого исследования в качестве «волатильности» мы используем подразумеваемую волатильность («implied volatility», далее IV) рассматриваемого опциона. Для того, чтобы получить IV, например, для опциона put (европейского либо маржируемого), необходимо при помощи некоторого численного метода решить для σ уравнение вида:

$$P(S, \sigma, T, K) = P_{\text{market}}, \quad (1)$$

где $P(S, \sigma, T, K)$ – цена опциона put в модели Блэка-Шоулса [1]; S – цена базового актива за рассматриваемый торговый период; σ – волатильность; T – срок действия опциона; K – цена исполнения опциона; P_{market} – рыночная цена опциона put.

Для различных страйков получаются разные значения подразумеваемой волатильности. Будучи нанесенными на график как функция страйка, они образуют так называемую «улыбку» или «ухмылку» волатильности, пример которой можно видеть на рис. 1.

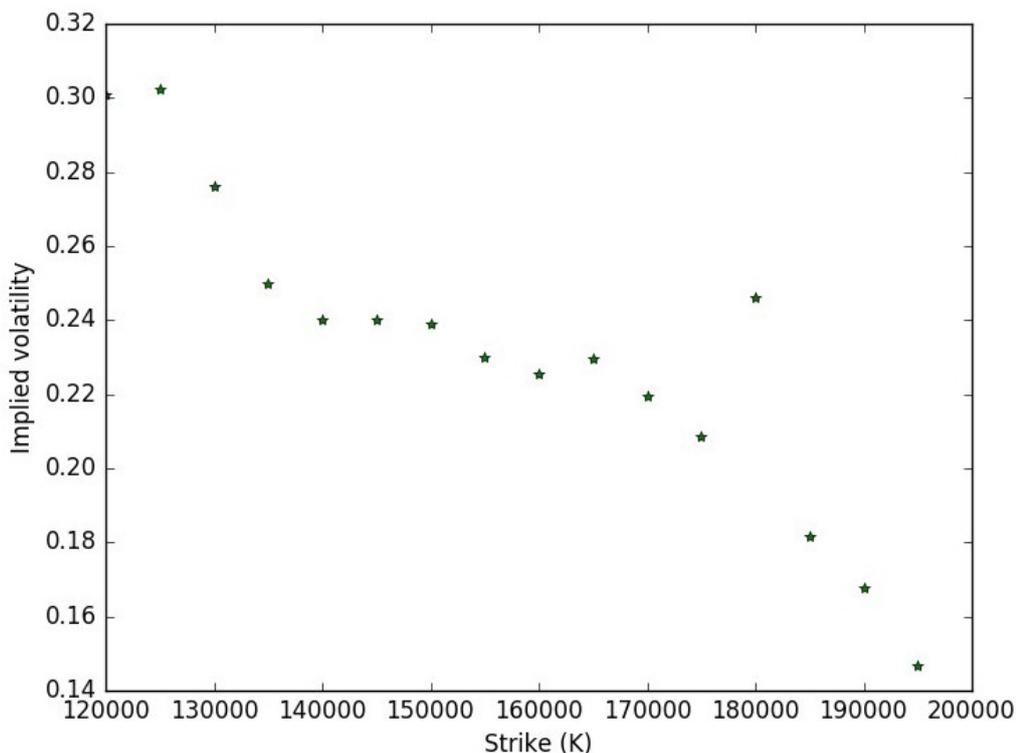


Рис. 1. – «Улыбка» для опциона «put» на фьючерс на индекс RTS

Для опционов call и put позиции, в которых оказывается участник торгов после осуществления таких сделок, различны. Мы придерживаемся соглашения, что «buy» означает уплату премии и приобретение опционного контракта, «short» означает подписание контракта (и постепенное приобретение премии). Соответственно, «sell» и «cover» используются в том смысле, что позволяют закрыть позицию таким образом, что между продавцом и подписчиком опциона более не остаётся обязательств.

Под «торговой стратегией» мы понимаем набор инструкций, регламентирующих правила открытия и закрытия длинных и коротких позиций в зависимости от имеющихся данных о ходе торгов и показателей, вычисляемых на их основе.



Обработка данных

Мы рассматриваем собранную историю сделок по всем опционам, базовым активом которых является фьючерс на индекс РТС, торгуемым на Московской Бирже за период с 31.05.2016 по 30.06.2016.

Для расчётов мы используем пакет библиотек Anaconda сборки 4.2.0 для языка Python версии 3.5. Графики нарисованы в пакете Matplotlib, обработка данных выполнена при помощи библиотеки Pandas. Реализации методов машинного обучения, и расчёта метрик их производительности взяты из библиотеки Scikit-learn.

Для опционов мы располагаем собранной историей совершённых сделок, а также историей торговых приказов («order log»), состоящей из цен «bid» – лучшего заказа на покупку на бирже в данный момент и «ask» – лучшего предложения на продажу. Общее число зарегистрированных сделок за период исследования, использованных для анализа, составило 1 334 300.

Выбор инструмента обусловлен сравнительно высокой ликвидностью, которая дает возможность получить исторические данные, представляющие ценность для анализа, более чем для одного страйка.

В качестве минимального периода времени для расчётов положим один торговый час. В качестве «цены» базового актива для торгового периода используем цену закрытия торгов за этот период.

Обрабатываем историю и «order log» по следующему алгоритму:

1. Определяем «центральный страйк», также называемый «страйк на деньгах» (ATM – at the money), который соответствует страйку с минимальной разницей между ценой put и call. Чтобы определить на этом этапе разницу между put и call, мы используем цены последних сделок. В случае равенства разниц выбираем максимальный страйк.

2. Рассматриваем по два «соседних» страйка с каждой стороны от центрального.



3. Для каждого из таковых формируем цену опциона за торговый период как среднее из цен сделок за период.

4. Если сделок по страйку не произошло, но за период было размещено более определённого числа o_1 приказов «bid» и более o_1 приказов «ask», при этом по крайней мере для двух из них выполнено условие $\frac{\text{ask}}{\text{bid}} \leq c$, то есть разрыв между «ценой спроса» и «ценой предложения» невелик, то полагаем цену равной $\frac{\text{ask} - \text{bid}}{2}$. Значения констант мы эмпирически выбрали следующими: $o_1 = 5$, $c = 1.5$.

5. В случае, если цена по данному страйку отсутствует в данный период, но присутствует в последующий и предыдущий периоды - полагаем цену равной среднему между имеющимися данными

Такая схема позволяет дистанцироваться от эффектов, возникающих на дальних страйках, где могут происходить сделки, закономерности появления которых, по-видимому, существенно отличаются от таковых для центральных страйков. Наличие таких «выбросов» существенно и негативно влияет на точность прогнозирования и создаёт препятствие для трактовки полученных результатов.

С другой стороны, алгоритм позволяет не игнорировать данные, созданные ожиданиями участников рынка, даже в том случае, если сделки не были заключены.

Подготовив таким образом данные, мы получаем возможность для каждого из оставшихся страйков вычислить показатель волатильности, решая уравнение (1). Полученный показатель волатильности мы используем далее в качестве обучающего признака (для предыдущих периодов) и целевого значения (для текущего периода) алгоритмов машинного обучения.

Для того, чтобы уменьшить шанс получить переобученные модели и ограничить влияние исходной выборки на результаты исследования, мы



использовали технику, которая известна под названием «кросс-валидации» и детально описана, например, в [10]. Она состоит в том, что имеющаяся в наличии выборка была разделена на три части: собственно обучающую, тестовую и итоговую. Первая используется в качестве набора признаков для обучения случайного леса и подбора параметров, вторая – для оценки переобученности модели и анализа поведения на отличающихся данных. Результаты применения торговых стратегий проверяются на третьей, итоговой выборке. «Собственно обучающая» и «тестовая» выборки были взяты в соотношении 70% и 30%.

Для того, чтобы учесть эффект кластеризации, мы выделяем участки «относительно большой» и «относительно малой» волатильности. Для этого мы вычисляем среднюю волатильность (отдельно для опционов put и call) для каждого из страйков, и подбираем порог c_d так, чтобы при заданном c_p (например, $c_p = 0.51$) выполнялись соотношения $\left(\frac{\sigma_t}{\bar{\sigma}}\right) > c_d$ и $P\left(\frac{\sigma_{t+1}}{\bar{\sigma}} \middle| \frac{\sigma_t}{\bar{\sigma}}\right) > c_p$.

Здесь $\bar{\sigma}$ – выборочное среднее волатильности по выбранному опциону, рассчитанное только на «собственно обучающей» выборке, индекс t обозначает предыдущий торговый период.

Первое соотношение отражает тот факт, что волатильность должна превышать среднюю для выбранного опциона. Второе – что вероятность события «в текущий торговый период волатильность останется выше средней» должна быть также выше определённого порога. Цель подбора – получить достаточно кластеров, чтобы работать с ними, в то же время минимизируя влияние небольших всплесков.

В качестве набора признаков для обучения мы используем доступные данные за один (предыдущий) торговый период. Доступные данные включают всю известную информацию об опционном контракте (включая текущую волатильность и время до даты истечения). Для того, чтобы учесть



кластеризацию при обучении, мы добавили также счётчик дней, в течение которых сохраняется «высокая волатильность».

На подготовленной «собственно обучающей выборке» мы обучаем два случайных леса – лес регрессии и лес классификации. Первый используется для прогнозирования значения волатильности. Второй – разделяет выборку на два класса – класс «высокой волатильности» (с ненулевым значением счётчика) и класс «низкой волатильности» – и предсказывает, в каком классе будет лежать следующий торговый период.

Описание стратегий

Мы составили две стратегии, каждая из которых была применена как к опционам call, так и опционам put. Их суть состоит в том, что перед началом каждого из торговых дней мы используем, не переобучая, случайный лес регрессии либо классификации для прогнозирования волатильности на текущий период.

Для того, чтобы учесть специфику опционных контрактов, мы закрываем все контракты до того момента, как наступит срок истечения. При изменении центральных страйков, если остаются открытые позиции, они закрываются по последней доступной цене.

Подход к реализации стратегии, основанной на предсказаниях леса регрессии (стратегия «регрессия и волатильность»), сводится к набору правил, описанных в таблице №1:

Таблица № 1

Описание правил стратегии «Регрессия и волатильность»

№ п/п	Действие	Условие	Предсказание
1	Открыть длинную позицию («buy»)	Нет открытых позиций	Волатильность σ_t на следующий период поднимется выше порога $\sigma_t > c_d \bar{\sigma}$
2	Закреть длинную позицию	Открыта длинная	Волатильность σ_t на следующий период опустится



	(«sell»)	позиция	ниже порога $\sigma_t < c_d \bar{\sigma}$
3	Открыть короткую позицию («short»)	Нет открытых позиций	Волатильность σ_t на следующий период опустится ниже порога $\sigma_t < c_d \bar{\sigma}$
4	Закрыть короткую позицию («cover»)	Открыта короткая позиция	Волатильность σ_t на следующий период поднимется выше порога $\sigma_t > c_d \bar{\sigma}$

Подход к реализации стратегии, основанной на предсказаниях леса регрессии (стратегия «классификация и волатильность») аналогичен. Отличие состоит в том, что для колонки «предсказание» мы используем прогноз относительно того, в каком кластере мы окажемся в следующем торговом периоде.

Результаты применения стратегий

Мы сравнили результаты стратегий с теми, которые можно получить в результате применения наивной стратегии «купи и держи». При подсчёте прибыли/убытков по стратегии мы предполагали, что в наличии всегда имеется необходимая сумма для того, чтобы продолжать держать позицию открытой в том случае, если она убыточна.

При расчёте динамики портфеля мы делали предположение, что используется одна и только одна из стратегий, но торговля ведётся одновременно по всем доступным центральным страйкам всеми доступными активами. Краткое изложение результатов приведено в таблице №2, показатели приведены в процентах от стоимости портфеля, состоящего из одного фьючерса.

Таблица № 2

Результаты применения торговых стратегий

№ п/п	Торговая стратегия	Общая прибыль, %	Максимальная просадка, %	Максимальная прибыль, %
1	«Купи и держи»	1,3	6,0	4,0
2	«Классификация и	21,2	13,3	37,0



	волатильность»			
3	«Регрессия и волатильность»	11,3	7,8	19,7

Заключение

На рассматриваемой истории торгов оказалось возможным как спрогнозировать изменение волатильности и появление кластеризации (точность классификации составила 61%), так и использовать полученные данные для построения стратегий, которые принесли прибыль при тестировании.

Мы видим несколько возможных путей для улучшения, которые планируется положить в основу дальнейших исследований:

- Более точное прогнозирование, если оно достижимо при использовании большего количества торговых данных либо более длинной истории, может дать прирост прибыльности таких стратегий и уменьшение просадки.

- Интересным представляется использование комплексных позиций по опционам (бабочка, кондор, стреддл, стренгл) в контексте работы с волатильностью.

- В качестве дополнительного источника информации о волатильности можно использовать представленный в статье [8] индекс волатильности.

Стоит заметить, что при определении прибыльности стратегии мы пренебрегли для простоты рядом важных для практической торговли факторов. Среди них – комиссии брокера, ограничения, касающиеся возможностей заключения контрактов, «проскальзывания» и многие другие.

Тем не менее, предложенный аппарат может быть применён как один из механизмов принятия торговых решений в составе комплексной системы анализа торговых рисков.



Исследование выполнено при финансовой поддержке РФФИ, проект «Математические методы анализа и управления рисками российского срочного рынка», №15-32-01390.

Литература

1. Black F., Scholes M. The Pricing of Options and Corporate Liabilities // Journal of Political Economy. 1973. №3(81). pp. 637–654.

2. Богачева М.Н., Прянишникова Л.И. Оценка справедливой цены опциона для обобщенной модели Кокса-Росса-Рубинштейна в случае m состояний // Инженерный вестник Дона, 2013, №4 URL: ivdon.ru/ru/magazine/archive/n4y2013/2114.

3. Гречко А.С., Кудрявцев О.Е., Родоченко В.В. Адекватное моделирование российского срочного рынка // Наука и образование: хозяйство и экономика; предпринимательство; право и управление. 2015. №6. С. 59 – 64.

4. Кудрявцев О.Е. Вычисление цен барьерных и американских опционов в моделях Леви // Обозрение прикладной и промышленной математики. 2010. №2. С. 210–220.

5. Augen J. The volatility edge in options trading: new technical strategies for investing in unstable markets. Upper Saddle River, NJ.: FT Press, 2008. 301 p.

6. Automated Options Trading Using Machine Learning // ResearchGate URL: researchgate.net/publication/265944840_Automated_Options_Trading_Using_Machine_Learning (date: 11.03.2017).

7. Breiman L. Random Forests // Machine Learning. 2001. №45. pp. 5-32.

8. Гречко А.С., Кудрявцев О.Е. Методы анализа волатильности российского финансового рынка для широкого класса моделей // Инженерный вестник Дона, 2016, №4 URL: ivdon.ru/ru/magazine/archive/n4y2016/3924.



9. Гречко А.С., Кудрявцев О.Е. Особенности построения Российского индекса волатильности, учитывающего возможные скачки // Теория вероятностей и ее применения. 2016. №3. С. 608.

10. Raschka S. Python machine learning. Birmingham, UK: Packt Publishing, 2015. 454 p.

References

1. Black F., Scholes M. The Pricing of Options and Corporate Liabilities. Journal of Political Economy. 1973. №3(81). pp. 637–654.

2. Bogacheva M.N., Pryanishnikova L.I. Inzhenernyj vestnik Dona (Rus), 2013, №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2114.

3. Grechko A.S., Kudryavtsev O.E., Rodochenko V.V. Nauka i obrazovanie: khozyaystvo i ekonomika; predprinimatel'stvo; pravo i upravlenie. 2015. № 6. pp. 59-64.

4. Kudryavtsev O.E. Obozrenie prikladnoj i promyshlennoj matematiki. 2010. № 2. pp. 210–220.

5. Augen J. The volatility edge in options trading: new technical strategies for investing in unstable markets. Upper Saddle River, NJ.: FT Press, 2008. 301 p.

6. Automated Options Trading Using Machine Learning. ResearchGate URL: researchgate.net/publication/265944840_Automated_Options_Trading_Using_Machine_Learning (date: 11.03.2017).

7. Breiman L. Random Forests. Machine Learning. 2001. №45. pp. 5-32.

8. Grechko A.S., Kudryavtsev O.E. Inzhenernyj vestnik Dona (Rus), 2016, №4. URL: ivdon.ru/ru/magazine/archive/n4y2016/3924.

9. Grechko A.S., Kudryavtsev O.E. Teorija verojatnostej i ee primenenie. 2016. №3. p. 608.

10. Raschka S. Python machine learning. Birmingham, UK: Packt Publishing, 2015. 454 p.