

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Математико-механический факультет

Кафедра системного программирования

Разработка и реализация модуля прогнозирования
волатильности с использованием рандомизированных
алгоритмов

Дипломная работа студента 545 группы

Федяшова Виктора Алексеевича

Научный руководитель / подпись /	д.ф.-м.н., профессор Граничин О.Н.
Рецензент / подпись /	к.ф.-м.н. Шалымов Д.С.
“Допустить к защите” заведующий кафедрой, / подпись /	д.ф.-м.н., проф. Терехов А.Н.

Санкт-Петербург
2010

SAINT PETERSBURG STATE UNIVERSITY
Mathematics & Mechanics Faculty

Software Engineering Chair

Development of a volatility forecasting tool using SPSA

by

Victor, Fedyashov

Master's thesis

Supervisor	Professor Dr. O. N. Granichin
Reviewer	Dr. D. S. Shalymov
“Approved by”	Professor A. N. Terekhov
Head of Department		

Saint Petersburg
2010

Оглавление:

ОГЛАВЛЕНИЕ:	3
ВВЕДЕНИЕ	4
ГЛАВА I: ТЕОРЕТИЧЕСКИЕ ОСНОВЫ	9
1.1. GARCH-МОДЕЛЬ	9
1.2. QMLE (QUASI-MAXIMUM LIKELIHOOD ESTIMATOR)	10
1.3. ОБЗОР СУЩЕСТВУЮЩИХ ПОДХОДОВ	12
ГЛАВА II: ПОСТАНОВКА ЗАДАЧИ И ОБЩИЕ ПОЛОЖЕНИЯ	15
2.1. ПОСТАНОВКА ЗАДАЧИ	15
2.2. НОВАЯ ПАРАМЕТРИЗАЦИЯ	15
2.3. ДОПОЛНИТЕЛЬНЫЙ ПАРАМЕТР	16
2.3.1. РАССТОЯНИЕ КУЛЬБАКА-ЛЕЙБЛЕРА	17
2.3.2. ПАРАМЕТРИЧЕСКОЕ СЕМЕЙСТВО	17
2.4. СВЯЗЬ С MLE И QMLE	18
ГЛАВА III: ОСНОВНОЙ РЕЗУЛЬТАТ	20
3.1. ИДЕНТИФИЦИРУЕМОСТЬ ПАРАМЕТРА ШТРАФА	20
3.2. ИДЕНТИФИЦИРУЕМОСТЬ ИСТИННОГО ПАРАМЕТРА	21
3.3. СОСТОЯТЕЛЬНОСТЬ И АСИМПТОТИЧЕСКИЕ СВОЙСТВА	21
3.4. РАНДОМИЗИРОВАННЫЕ АЛГОРИТМЫ	23
ГЛАВА IV: РЕАЛИЗАЦИЯ	25
4.1. ОСНОВНОЙ ДВУХСТУПЕНЧАТЫЙ АЛГОРИТМ	25
4.2. АДАПТИВНАЯ ПРОЦЕДУРА ВЫБОРА ПАРАМЕТРОВ ФУНКЦИИ ПРАВДОПОДОБИЯ	25
4.3. МОДЕЛИРОВАНИЕ	26
ЗАКЛЮЧЕНИЕ	28
СПИСОК ЛИТЕРАТУРЫ:	29

Введение

В последнее время, в связи с усложнением механизмов, лежащих в основе финансовых рынков и институтов, для того, чтобы принять правильное, взвешенное решение и выработать грамотную стратегию поведения, требуется учитывать все большее количество факторов. Наибольший интерес с научной точки зрения представляет изучение неопределенности рыночного процесса. Ключевым параметром, который численно ее характеризует, является волатильность.

Волатильность уже давно стала темой чрезвычайной важности для всех, кто связан с финансовыми рынками, даже в качестве наблюдателя. Для многих представителей неискушенной публики этот термин – это просто синоним слова риск. То есть высокая волатильность считается симптомом нарушения работы рыночной структуры. Для них волатильность значит, что финансовые активы оцениваются не вполне справедливо, а рынок капитала функционирует не так хорошо, как должен. Для тех же, кто имеет дело с производными ценными бумагами, понимание волатильности, умение аккуратно ее прогнозировать и управлять степенью защищенности их инвестиционных портфелей от ее эффектов является критически важным.

В рамках дипломной работы подробный обзор областей использования волатильности не будет представлен. Однако стоит упомянуть о некоторых из них здесь, чтобы показать актуальность дальнейших исследований не только с теоретической, но и с практической точки зрения. Одним из наиболее широко известных применений является оценка VaR (value at risk). Этот показатель на сегодняшний день является ключевым при работе с рисками, и отражает стоимостное выражение потерь, которые не будут превышены за определенный период времени с некоторой вероятностью (напр. [11]).

Также современная теория опционов, начиная с модели Блэка-Шоулза [1973], отводит волатильности центральную роль в определении

справедливой цены опциона или другой производной ценной бумаги с его свойствами. Несмотря на то, что она является лишь одним из минимум пяти параметров, которые учитываются при оценивании, ее важность увеличивается тем фактом, что только она не может быть получена напрямую из наблюдений. На современном этапе развития вопроса существует масса моделей, описывающих волатильность (в том числе некоторые модификации GARCH-модели, которая будет использоваться нами в дальнейшем, и которая показывает более высокую эффективность, чем стандартные модели Блэка-Шоулза и стохастической волатильности [3]), а также достаточно много способов их применения для прогнозирования. И даже несмотря на это оно остается во многом скорее искусством, нежели наукой, в особенности для трейдеров, связанных с производными ценными бумагами.

Для перехода к дальнейшим рассуждениям уже требуется формальное определение волатильности. Начиная с 50-х годов XX века как теоретические исследования, так и эмпирические данные говорили в пользу теории эффективного рынка или, что по сути то же самое, модели случайного блуждания. В рамках этой теории, движение цен любого финансового актива может быть записано в следующей форме:

$$r_t = \mu_t + \varepsilon_t v_t$$

где $\mu_t \stackrel{def}{=} E(r_t | I_{t-1})$, а ε_t – белый шум. Тогда:

$$VAR(r_t | I_{t-1}) = E((\varepsilon_t v_t)^2 | I_{t-1}) = v_t^2,$$

где $r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$ – доходность, I_{t-1} – история до предыдущего периода

включительно, а v_t – волатильность. Таким образом, формально говоря, волатильность – это условное по предыдущей истории стандартное отклонение процесса доходности. Таким образом, в рамках любой параметрической модели, описывающей волатильность, общий метод прогнозирования можно сформулировать так: $\hat{v}_t = f(\hat{\theta}, \hat{I}_{t-1})$, где $\hat{\theta}$ –

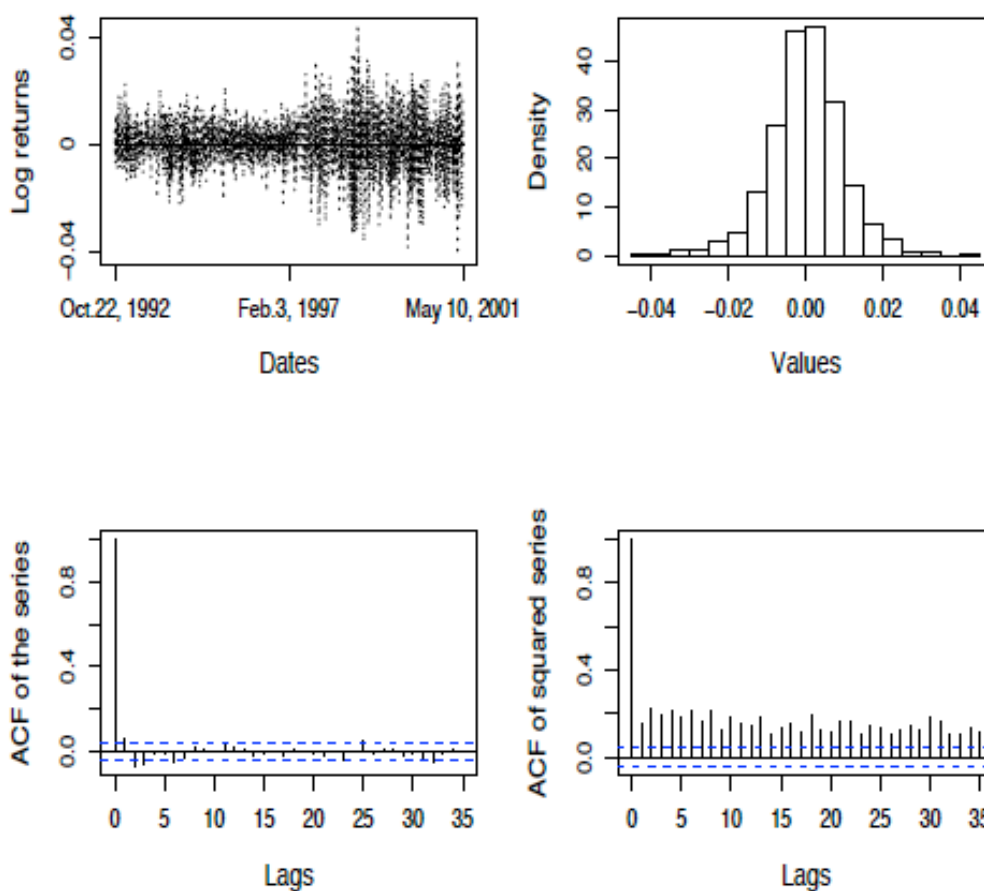
результат оценки параметров, а в качестве \hat{I}_{t-1} подставлены значения, взятые из наблюдений.

Однако, как уже было упомянуто, в современных математических финансах существует огромное количество различных моделей описания волатильности. Для того, чтобы понять, какую же из них выбрать в качестве основы при прогнозировании, нужно взглянуть на свойства временных рядов финансовой природы, многие из которых хорошо известны на сегодняшний день. Вот основные:

- Средняя доходность ряда, то есть суть эмпирическое математическое ожидание доходности, достаточно близка к нулю.
- Предельное распределение почти симметрично, с возможным небольшим перекосом, с пиком в районе нуля и почти всегда тяжелыми хвостами.
- Автокорреляция самого ряда незначительно мала, в то время как автокорреляция квадратов или абсолютных значений ряда может быть существенна для большого числа шагов.
- Кластеризация волатильности, т.е. часто можно наблюдать, что большие изменения (вариации) ряда следуют за большими, а малые – за малыми.

На диаграммах ниже можно увидеть практически все эти свойства. Верхний левый график – это типичный временной ряд (логарифм дневной доходности индекса Financial Times Stock Exchange 100), правее – гистограмма ряда, которая показывает, что (эмпирическая) плотность почти симметрична. Ниже представлены автокорреляции соответственно ряда (незначительна по крайней мере для 35 шагов) и квадратов ряда (значительна по крайней мере для 35 шагов). Свидетельство тяжелых хвостов значительно проще увидеть на примере отдельных акций, а не индексов. Так например, средний куртозис доходностей по 367 индивидуальным компаниям, включенным в индекс Standard&Poor 500 (далее S&P500) с 1990 по 2001 годы составляет примерно 10.6, причем некоторые отдельные показатели достигали 165, в то время как

куртозис нормального распределения составляет 3. Таким образом, эволюция математических моделей представления волатильности происходила по мере включения в них тех или иных свойств.



Свойства временных рядов, такие как (3) – (4), не объяснялись в достаточной мере традиционными эконометрическими моделями, до разработки модели ARCH (autoregressive conditional heteroscedasticity, [Engle, 1982]), а затем и самого известного ее обобщения – GARCH (generalized ARCH, [Bollerslev, 1986]). Сразу же после своего появления они завоевали огромную популярность, до сих пор оставаясь ведущими моделями описания ошибок финансовых временных рядов. Дипломная работа посвящена созданию модуля прогнозирования волатильности на базе модели GARCH. Также в работе подробно описано математическое обеспечение алгоритмов, которые лежат в основе процесса, сформулированы и частично доказаны основные теоретические результаты, а также представлен способ

оптимизации вычислений с помощью рандомизированных алгоритмов стохастической аппроксимации [1].

Структура работы следующая: в Главе I приведено описание модели и общих теоретических понятий, а также произведен обзор существующих подходов, Глава II посвящена формальной постановке задачи и основным положениям, в Главе III сформулирован основной теоретический результат, в Главе IV описана реализация, в Заключение подведены итоги.

Глава I: Теоретические основы

Глава построена следующим образом: сначала будет дано определение GARCH-модели, которая лежит в основе работы модуля прогнозирования. Затем будут кратко описаны теоретические аспекты, которые необходимы для понимания дальнейшего материала. В последнем пункте приведен обзор существующих подходов к оценке параметров этой модели, что является ключевым шагом при прогнозировании.

1.1. GARCH-модель

Одной из главных особенностей финансовых временных рядов является кластеризация волатильности. Эта характеристика финансовой информации предполагает, что волатильности автокоррелированы и зависят от времени. Именно попытка создать модель, которая бы объясняла этот феномен, и привела в 1982 году к идее описания динамики волатильности через взвешенное среднее ошибок прошлых периодов [5]. В обозначениях, введенных выше, можно определить поведение процесса следующим образом:

$$\begin{aligned}r_t &= \mu_t + \varepsilon_t, \\ \varepsilon_t &= \eta_t v_t, \\ v_t^2 &= \sigma + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2,\end{aligned}$$

где $\forall i \alpha_i \geq 0$, а $\{\eta_t\}$ – процесс инноваций, соответствующий белому шуму. Такая модель получила название ARCH. Естественное и более эффективное (в среднем требующее обработки значительно меньшей исторической информации) расширение, предложенное несколько лет спустя, предполагало зависимость условной вариации v_t от предыдущих условных вариаций, т. е. $\{v_k; k < t\}$. Усовершенствованная модель и получила название GARCH, и в общем виде уравнение для квадрата волатильности может быть записано следующим образом:

$$v_t^2 = \sigma + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j v_{t-j}^2,$$

где все коэффициенты также неотрицательны. Модель GARCH обладает значительно более гибкой структурой параметров, нежели ARCH. Так например, из эмпирических данных следует, что в то время как для модели ARCH требуется достаточно длинный лаг (большое q), уже GARCH(1,1) обычно хорошо описывает большое большинство временных рядов. Таким образом, простота и интуитивная понятность подхода сделали эту модель хорошей стартовой точкой во многих финансовых приложениях.

Прогнозирование с помощью модели GARCH вписывается в общую схему, представленную ранее. Основная задача – определить вектор параметров $\theta = [\sigma, \alpha, \beta]$, используя имеющиеся наблюдения. Затем для получения прогноза достаточно подставить его в уравнение волатильности вместе с историческими значениями.

1.2. QMLE (quasi-maximum likelihood estimator)

Перед переходом к дальнейшим рассуждениям, введем несколько ключевых теоретических понятий. Существует много различных вариантов оценки параметров модели, но в большинстве работ используется популярный в статистике метод максимального правдоподобия. На интуитивном уровне можно сказать, что метод выбирает такие значения параметров модели, при которых вероятность появления имеющийся выборки больше, чем при любом другом. Предположим, что у нас имеется набор наблюдений $\{y_1, \dots, y_n\}$, и вероятностные распределения зависят от вектора параметров θ . Допустим, также, что смешанная плотность (которая и называется функцией правдоподобия) это

$$L_n(\theta) = L(y_1, \dots, y_n | \theta).$$

Оценка максимального правдоподобия (MLE) определяется как любая оценка $\hat{\theta}_n$, которая максимизирует функцию правдоподобия внутри пространства параметров Θ , т.е.

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} L_n(\theta).$$

Существует несколько основных критериев качества оценок. Самый интуитивно понятный из них – это несмещенность. *Несмещенная оценка* – это точечная оценка, математическое ожидание которой равно оцениваемому параметру, т. е. $E(\hat{\theta}_n) = \theta_0$. Также важными критериями являются состоятельность и асимптотическая нормальность [1]. Они касаются асимптотического поведения оценок, иначе говоря когда выборка становится все больше. *Состоятельная оценка* – это точечная оценка, которая сходится по вероятности (слабая состоятельность) или почти наверное (сильная состоятельность) к оцениваемому параметру. *Асимптотическая нормальность* не только гарантирует состоятельность, но также дает более ясное представление по поводу скорости (относительно количества наблюдений), с которой оценка сходится к истинному параметру. Возможно также построение доверительного интервала. Этот критерий обычно записывается в следующей форме:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{D} N(0, \Sigma).$$

Оценка параметров GARCH модели обычно производится с помощью максимального правдоподобия. Однако, получение такой функции – задача далеко не простая, потому как для этого необходимо знать истинное распределение инноваций (в принятой в 1.1 системе обозначений – η_t) . Более того, даже в случае, если оно заранее известно, распределение v_t^2 по-прежнему не определено. Обычной практикой при оценке параметров является предположение об условной нормальности ε_t , иными словами о гауссовом распределении инноваций. В таком случае можно можно в явном виде выписать функцию максимального правдоподобия с точностью до константы (для удобства вместо нее используется ее минус логарифм [1]):

$$L_n(\theta) = L(r_1, \dots, r_n | \theta) = -\frac{1}{2} \sum_{t=1}^n \left(\frac{(r_t - \mu_t)^2}{v_t^2} + 2 \ln v_t \right).$$

Ввиду того, что описанная функция использует функцию Гаусса вместо реальной плотности распределения инноваций, она получила название функции квази максимального правдоподобия.

1.3. Обзор существующих подходов

В ранней литературе по изучению ARCH/GARCH моделей главным инструментом был описанный выше метод максимального правдоподобия с предположениями об условной нормальности инноваций. Тем не менее, многочисленные экспериментальные данные свидетельствовали в пользу того, что распределение инноваций обладает тяжелыми хвостами и асимметрично, что противоречило такой гипотезе (напр. [5]). В связи с этим стали появляться оценки максимального правдоподобия с использованием распределения Стьюдента (Student's t) или обобщенного нормального (Generalized Gaussian) (напр. [4], [10], [12]). Однако эти методы могут давать несостоятельные оценки, если истинное распределение инноваций неправильно выбрано. Напротив, оценки Гауссова метода максимального правдоподобия, определенные выше (далее просто QMLE) могут быть состоятельными и асимптотически нормальными (в случае если гарантирована конечность четвертого момента инноваций), даже если распределение далеко от Гауссова ([3], [11]).

Тем не менее, большая устойчивость влечет за собой большие потери в эффективности. Теоретически, отклонение истинной плотности инноваций от Гауссова распределения может существенно увеличить дисперсию оценок, которые, таким образом, оказываются достаточно далеко от границы Рао-Крамера. В 1991 году была предложена полупараметрическая процедура, которая позволяла получить выигрыш в эффективности над QMLE порядка 50%, основываясь на симуляциях Монте-Карло [8]. Тем не менее авторы признались, что не удалось получить максимально возможный прирост эффективности. Также была предложена адаптивная двухступенчатая процедура, основанная на измененной параметризации модели GARCH(1,1) с

неизвестной, но симметричной ошибкой [7]. После сравнения ее производительности с QMLE и MLE (т. е. процедуры со знанием истинной плотности распределения ошибок) оказалось, что все усилия напрасны в случае, если инновации не обладают конечным четвертым моментом [6]. Было также показано, что асимптотически QMLE сходится к устойчивому распределению (stable distribution), а не к нормальному.

На практике причиной того, что Гауссов QMLE недостаточно эффективен, является тот факт, что распределение финансовой информации обладает тяжелыми хвостами. Таким образом, предположение об условной нормальности не выполняется. Например можно посмотреть исследование о сравнении куртозиса (kurtosis, т.е. $E(Y_t^4) / E^2(Y_t^2)$) после применения QMLE (~ 4.6) к месячной информации индекса S&P500 с обычным Гауссовым ($= 3$) [Bollerslev & Wooldbridge, 1992]. Таким образом, интуитивно понятна идея разработки QMLE, который бы базировался на не-Гауссовых функциях правдоподобия. Потери эффективности могут быть существенно уменьшены за счет использования распределений с тяжелыми хвостами.

Несмотря на все вышесказанное, в литературе довольно редко можно встретить упоминание об оценке с помощью не-Гауссовых методов. Частично это связано с тем фактом, что Гауссов QMLE устойчив относительно неверной спецификации распределения ошибок, в то время как не-Гауссов QMLE в общем случае не выдает состоятельные оценки, если истинное распределение ошибок отклоняется от используемого в функции правдоподобия. Более того, эта несостоятельность не может быть исправлена даже при добавлении нового параметра формы, который позволял бы выбрать нужное распределение из семейства (напр. t-distribution), кроме случая, когда истинное распределение входит в это семейство [14]. Действенный способ, который может влиять на эффективность не-Гауссовых QMLE, – это манипуляция с изначальными предположениями о модели. Так например, было показано, что с другим условием на момент истинных инноваций, которое заменило бы стандартное $E(\eta^2) = 1$, соответствующие

оценки могут стать состоятельными и асимптотически нормальными [3]. Однако понятно, что такое условие является необходимым для возможности естественной интерпретации v_t как условного стандартного отклонения, т.е. волатильности. Более того, совершенно очевидно, что условия на моменты – это часть описания модели, и должны даваться до и независимо от выбора правдоподобия. Поэтому их изменение не устранил проблем с устойчивостью не-Гауссовых оценок, а всего лишь сделает состоятельной некоторую комбинацию изначальных условий и правдоподобия, которую невозможно определить, не зная истинную структуру инноваций.

Для повышения качества прогнозов волатильности, в дипломной работе разработан способ оценивать коэффициенты GARCH-модели с помощью не-Гауссова QMLE эффективно, независимо от изначальных предположений, и устойчиво относительно распределения инноваций. Помимо очевидной теоретической пользы, разработанный метод может быть достаточно легко внедрен в существующие программные модули по оценке GARCH параметров. Как правило они уже позволяют вручную выбирать функцию правдоподобия, основанную на распределении с тяжелыми хвостами [в основном на распределении Стьюдента или на обобщенном Гауссовом]. Таким образом метод, который будет описан, не требует никаких изменений в интерфейсе, но позволяет получать состоятельные оценки независимо от попадания/непопадания в правильное семейство. При реализации модуля использовалась адаптивная процедура выбора параметров функции правдоподобия, описанная в Главе IV.

Глава II: Постановка задачи и общие положения

Все дальнейшее рассуждение, можно логически разбить на два глобальных направления: с одной стороны нас будет интересовать теоретический аспект, а именно построение эффективного метода, который давал бы состоятельные и асимптотически нормальные оценки. С другой стороны, вторым ключевым моментом является выбор эффективного алгоритма его реализации. Ввиду того, что основная процедура, которая повторяется не один раз – это минимизация некоторой многомерной функции итеративными методами, мы будем использовать рандомизированные алгоритмы, которые не только позволяют экономить вычислительные мощности, но и дают возможность применения теоретических результатов при менее идеальных условиях, а именно в случае появления помех измерения.

2.1. Постановка задачи

Таким образом, в рамках дипломной работы ставились следующие задачи:

- Разработка нового алгоритма прогнозирования волатильности на основе модели GARCH, который работал бы эффективно независимо от специфики конкретного временного ряда.
- Математическое обоснование его корректности и состоятельности, а также сравнительный анализ с имеющимися подходами.
- Изучение искусственных симуляций алгоритма с использованием различных распределений.
- Предложение эффективной реализации основной процедуры с использованием рандомизированных алгоритмов.
- Реализация модуля прогнозирования волатильности и его интеграция в уже существующую оболочку.

В остальных разделах главы II основное внимание будет уделено базовым теоретическим идеям, которые лежат в основе алгоритма.

2.2. Новая параметризация

В рамках дипломной работы мы будем придерживаться частного случая параметризации, которая разделяет гетероскедастические параметры и параметр масштаба волатильности (впервые [7]). Изначально она выглядела следующим образом:

$$y_t = v_t \xi_t = \mu v_t + \sigma v_t \varepsilon_t,$$

где μ – вещественное число, ε_t – процесс инноваций, а v_t – GARCH-процесс с единичным свободным членом. Так как ключевой интерес для нас представляет измерение непосредственно волатильности, в дальнейшем мы будем полагать коэффициент μ равным нулю, так как он отвечает за среднее процесса. Таким образом, параметрическая форма, на которой мы будем основываться, принимает следующий вид:

$$\begin{cases} \varepsilon_t = \sigma v_t \eta_t \\ v_t^2 = 1 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j v_{t-j}^2 \end{cases},$$

где параметры модели – это вектор $\theta = \{\sigma, [\gamma]\}$, где первая компонента – это параметр масштаба, а $\gamma = \{[\alpha], [\beta]\}$, то есть вектор авторегрессивных параметров. Истинный параметр θ_0 находится внутри Θ , компактного подмножества \mathbb{R}_+^{1+p+q} . Что касается процесса инноваций, то мы сохраним исходные условия на первый и второй моменты, но откажемся от условной нормальности. Таким образом, плотность распределения неизвестна (в дальнейшем $g(\cdot)$). В нашей новой параметризации QMLE с использованием четко разделены параметры, которые всегда находятся асимптотически нормально не-Гауссовым QMLE, и параметр σ , который стандартный метод определить не в состоянии [14]. Об идее нового способа борьбы с несостоятельностью оценок σ будет рассказано в следующем пункте.

2.3. Дополнительный параметр

В этом разделе будет рассказано про введение дополнительного параметра, который будет выбираться специальным образом, чтобы компенсировать неправильную спецификацию распределения инноваций. Как было сказано выше, в большинстве современных GARCH-пакетов существует опция выбора семейства функций правдоподобия. Пользователь делает выбор, основываясь на своих представлениях об истинном распределении инноваций. Задача состоит в том, чтобы в случае, когда семейство выбрано неправильно, алгоритм тем не менее давал бы состоятельные оценки, тем самым совмещая устойчивость с использованием правдоподобий с тяжелыми хвостами.

2.3.1. Расстояние Кульбака-Лейблера

Идея введения нового параметра заключается в том, что вместо того, чтобы напрямую работать с выбранной функцией правдоподобия (далее ее плотность будем обозначать $f(\cdot)$) мы будем выбирать из параметрического семейства (которое индуцируется f) элемент, плотность которого наиболее близка к истинной. Для этого нам понадобится мера «близости» распределений. Мы будем использовать понятие дистанции Кульбака-Лейблера, которое определяется следующим образом:

$$D_{KL}(g, f) = \int_{-\infty}^{+\infty} g(x) \ln\left(\frac{g(x)}{f(x)}\right) dx,$$

в случае, если заданы два непрерывных распределения с соответствующими плотностями g и f . В рамках этой работы мы не будем останавливаться на обосновании выбора именно этого метода для определения меры различия между распределениями, подробно о его предпосылках и преимуществах можно посмотреть у [Kullback & Leibler, 1951].

2.3.2. Параметрическое семейство

Введем в рассмотрение параметрическое семейство $\left\{ \lambda : \frac{1}{\lambda} f\left(\frac{1}{\lambda}\right) \right\}$, зависящее от $\lambda > 0$ для любой заданной плотности f . Параметр λ будем считать показателем штрафа за неверную спецификацию распределения. Непосредственно в процедуре оценки параметров будет использоваться специальное квази-правдоподобие, с λ_f , минимизирующим невязку в смысле, описанном в предыдущем пункте (именно такой вид параметрического семейства был впервые предложен Уайтом в 1982 году независимо от моделей ARCH/GARCH, которые на тот момент еще не были изобретены [18]). Иначе говоря:

$$\begin{aligned}
D_{KL}(g, f) &= \int_{-\infty}^{+\infty} g(x) \ln \left(\frac{g(x)}{\frac{1}{\lambda} f\left(\frac{x}{\lambda}\right)} \right) dx = \int_{-\infty}^{+\infty} g(x) \ln(g(x)) dx - \\
&- \int_{-\infty}^{+\infty} \left(\ln \lambda - \ln f\left(\frac{x}{\lambda}\right) \right) g(x) dx, \\
\lambda_f &= \arg \min_{\lambda > 0} D_{KL}(g, f) = \arg \max_{\lambda > 0} \int_{-\infty}^{+\infty} \left(\ln f\left(\frac{x}{\lambda}\right) - \ln \lambda \right) dx \\
&= \arg \max_{\lambda > 0} E_g \left\{ -\ln \lambda + \ln f\left(\frac{\eta}{\lambda}\right) \right\}.
\end{aligned}$$

Таким образом, на первом этапе находится оптимальный параметр штрафа, после чего можно переходить к основной процедуре, в которой он участвует следующим образом:

$$L_T(\theta) = \frac{1}{T} \sum_{t=1}^T l_t(\theta) = \frac{1}{T} \sum_{t=1}^T \left(-\ln(\sigma \sqrt{h_t}) + \ln f\left(\frac{\varepsilon_t}{\lambda_f \sigma \sqrt{h_t}}\right) \right) \quad (1)$$

для удобства будем считать, что следующие условия всегда выполняются: $f \in C^2$, а также $\forall \lambda > 0$ имеем $\sup_{\theta \in \Theta} E |l_t(\theta)| < \infty$, $E \sup_{\theta \in V} |\nabla l_t(\theta)| < \infty$, а также $E \sup_{\theta \in V} |\nabla^2 l_t(\theta)| < \infty$, где V – некоторая окрестность истинного параметра θ_0 .

2.4. Связь с MLE и QMLE

Первый факт, на который стоит обратить внимание – это то, что первый шаг нашего алгоритма, строго говоря, невозможно выполнить, не зная реальную плотность распределения инноваций. Это связано с тем, что параметр штрафа максимизирует ожидание, которое берется по истинной плотности. В следующей главе будет представлен способ преодолеть эту трудность, но сначала требуется доказать следующее утверждение:

Утверждение: Если $f \propto \exp(-x^2 / 2)$ или $f = g$, то $\lambda_f = 1$

Доказательство: Рассмотрим последовательно оба эти случая и докажем про каждый из них в отдельности:

1) Предположим сначала, что $f \propto \exp(-x^2 / 2)$. Тогда:

$$E_g \left\{ -\ln \lambda + \ln f \left(\frac{\eta}{\lambda} \right) \right\} = -\ln \lambda - \frac{1}{2\lambda^2} E_g \{ \eta^2 \} =$$

$$= -\ln \lambda - \frac{1}{2\lambda^2}.$$

Получаем что и требовалось простым дифференцированием по λ и приравниванием результата нулю.

2) Теперь предположим, что $f = g$, и введем следующее обозначение

$$f' = \frac{1}{\lambda} f \left(\frac{\cdot}{\lambda} \right), \text{ тогда:}$$

$$-D_{KL}(f, f') = E \ln \left(\frac{\frac{1}{\lambda} f \left(\frac{\eta}{\lambda} \right)}{f(\eta)} \right) \leq 2E \left(\sqrt{\frac{\frac{1}{\lambda} f \left(\frac{\eta}{\lambda} \right)}{f(\eta)}} - 1 \right) =$$

$$= 2 \int_{-\infty}^{+\infty} \sqrt{\frac{1}{\lambda} f \left(\frac{x}{\lambda} \right)} f(x) dx - 2 \leq - \int_{-\infty}^{+\infty} \left(\sqrt{\frac{1}{\lambda} f \left(\frac{x}{\lambda} \right)} - \sqrt{f(x)} \right)^2 dx \leq$$

$$\leq 0.$$

Причем равенство сохраняется только при $\lambda = 1$. Что и требовалось доказать.

Из доказанного утверждения становится понятно, почему на этот параметр долгое время никто не обращал внимания. Это связано с тем, что, как было ранее упомянуто, в большинстве случаев используется Gaussian QMLE, а в нем значение штрафа равно единице. В случае же произвольной функции правдоподобия, его отклонение от единичного значения как раз и является численной мерой «ошибки».

Глава III: Основной результат

Глава носит сугубо теоретический характер: в ней делаются определенные предположения на выбранную плотность распределения, и в случае их удовлетворения формулируются результаты касательно состоятельности и асимптотической нормальности.

3.1. Идентифицируемость параметра штрафа

Текущий раздел состоит из формулировки достаточных условий для возможности идентификации параметра λ_f .

Утверждение: введем $Q(\lambda) = -\ln \lambda + E_g(\ln f\left(\frac{\eta}{\lambda}\right))$, $h(x) = x \frac{f'(x)}{f(x)}$, а $f \in C^2$.

Предположим также, что инновации обладают конечным моментом степени p . Если дополнительно:

1. $h(x) \leq 0$.
2. $xh'(x) \leq 0$ и равенство верно только при $x = 0$.
3. $|h(x)| \leq C|x|^p$, а также $|xh'(x)| \leq C|x|^p$ для каких-то констант $C > 0$ и $p \geq 0$.
4. $\limsup_{x \rightarrow \infty} h(x) < -1$.

Тогда у $Q(\lambda)$ существует единственная точка максимума на положительной полуоси.

Доказательство: Сначала продифференцируем $Q(\lambda)$:

$$Q'(x) = -\frac{1}{x} E\left(1 + h\left(\frac{\eta}{x}\right)\right), \text{ далее введем } H(x) = E\left(1 + h\left(\frac{\eta}{x}\right)\right).$$

$$Q''(x) = \frac{1}{x^2} H(x) - \frac{1}{x} H'(x), \text{ где } H'(x) = -\frac{1}{x^2} E(\eta h'\left(\frac{\eta}{x}\right)) > 0,$$

$$\lim_{x \rightarrow \infty} |H(x) - 1| = \lim_{x \rightarrow \infty} \left| E\left(h\left(\frac{\eta}{x}\right)\right) \right| \leq \lim_{x \rightarrow \infty} \frac{E|\eta|^p}{x^p} \rightarrow 0, \text{ с другой стороны}$$

$$\limsup_{x \rightarrow 0+} H(x) = \limsup_{x \rightarrow 0+} E\left(1 + h\left(\frac{\eta}{x}\right)\right) \leq 1 + E\left(\limsup_{x \rightarrow 0+} h\left(\frac{\eta}{x}\right)\right) < 0.$$

Из вышесказанного следует, что существует единственная точка λ_f , такая что $H(\lambda_f) = 0 \Rightarrow Q'(\lambda_f) = 0$, и это точка максимума. Что и требовалось доказать.

3.2. Идентифицируемость истинного параметра

Идентифицируемость параметра играет ключевую роль при разговоре о состоятельности оценок. Она требует, чтобы ожидание функции правдоподобия, то есть суть $\bar{L}_T(\theta) = E(L_T(\theta))$, имело бы единственный максимум в истинном значении параметра θ_0 . Для того, чтобы показать, что вектор-параметр определяем, сделаем следующие предположения:

Предположение:

1. для идентификации авторегрессионных параметров потребуем:

$$v_t(\gamma)/v_t(\gamma_0) \neq \text{const.}$$
2. параметр штрафа определяем (см. предыдущий пункт)

Утверждение: При сделанном предположении параметр идентифицируем.

Доказательство:

$$\begin{aligned} E(l_t(\theta)) &= Q\left(\frac{\lambda_f \sigma v_t(\gamma)}{\sigma_0 v_t(\gamma_0)}\right) - \ln \sigma_0 v_t(\gamma_0) + \ln \lambda_f \leq Q(\lambda_f) - \ln \sigma_0 v_t(\gamma_0) + \ln \lambda_f = \\ &= E(l_t(\theta_0)). \end{aligned}$$

Просуммировав по t получим, что и требовалось доказать.

3.3. Состоятельность и асимптотические свойства

Прежде чем перейти к описанию двухступенчатой процедуры, сформулируем две ключевые теоремы, которые показывают асимптотическое поведение оценок при условии, что λ_f нам известно.

Теорема: При сделанных выше предположениях, в случае, если λ_f известно, оценки, полученные максимизацией (1) обладают свойством слабой состоятельности.

Доказательство: в рамках дипломной работы представлено не будет. Аналогичное доказательство можно посмотреть у Elie & Jeantheau, 1995.

Предположение:

$$1. \quad 0 < E \left(h^2 \left(\frac{\eta}{\lambda_f} \right) \right) < \infty, \quad 0 < E \left| \eta h' \left(\frac{\eta}{\lambda_f} \right) \right| < \infty.$$

$$2. \quad M = E(k_0 k_0^T) < \infty, \quad \text{где } k_0 = \left(\frac{1}{\sigma}, \left[\frac{1}{v_i} \frac{\partial v_i}{\partial \gamma} \right] \right) \Bigg|_{\theta=\theta_0}.$$

Теорема: При всех сделанных предположениях оценки, получающиеся максимизацией (1) асимптотически нормальны, то есть

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow N(0, \Sigma), \quad \text{где } \Sigma = M^{-1} \frac{E h_1^2}{E^2(h_2)}, \quad \text{где } h_1 = 1 + h \left(\frac{\eta}{\lambda_f} \right), \quad \text{а } h_2 = \frac{\eta}{\lambda_f} h' \left(\frac{\eta}{\lambda_f} \right).$$

Доказательство: также не будет представлено ввиду громоздкости. Идеи доказательства можно посмотреть в [12].

С использованием последней теоремы можно провести сравнительный анализ эффективности различных процедур оценки, в том числе Гауссова QMLE и нашего метода. Рассуждения достаточно громоздки, поэтому в рамках дипломной работы приведены не будут. Краткая идея такова: ввиду того, что данные финансовой природы обладают тяжелыми хвостами, в случае, когда выбранное распределение «тяжелее» нормального, эффективность нашей процедуры выше.

Стоит отметить, что условия на момент, который описаны в предположении, зависят только от хвостов истинной плотности инноваций и от выбранной нами функции правдоподобия. Таким образом, видно преимущество не-Гауссова метода: он может не требовать дополнительных условий, кроме налагаемых моделью (например, для распределения Стьюдента). Также видна еще одна трактовка параметра λ_f : он корректирует оценки простого не-Гауссова метода, представляя $\sigma_0 \lambda_f$ вместо σ_0 . Таким образом, метод без поправки на штраф может выдавать состоятельные

результаты только в случае, когда параметр штрафа строго равен единице. Тот факт, что это неверно для произвольного распределения, сделал стандартный QMLE столь популярным, так как в нем это условие выполняется (см. утверждение в Главе II). Отсюда возникает основная идея метода: встроить поправку на параметр штрафа в не-Гауссовый QMLE, что поможет потенциально достичь лучшей эффективности. Однако напрямую это сделать не удастся, потому как у нас нет никакого знания про истинную структуру инноваций. Способ преодоления этой проблемы, а также результаты моделирования, будут представлены в следующей главе.

3.4. Рандомизированные алгоритмы

Во всех современных GARCH-пакетах, которые используют метод максимального правдоподобия (исключения встречаются крайне редко, из теоретически обоснованных методов можно отметить [17], где используется метод наименьших квадратов), для решения задачи максимизации используются градиентные методы, а именно: метод внешнего произведения градиентов и обобщенный метод моментов. В обоих случаях необходимым шагом для эффективной оценки является обращение матриц размерности $1 + p + q$ в нашей исходной нотации. Также на каждой итерации происходит дифференцирование. Это весьма дорогостоящие операции, и потому в нашей реализации использовались рандомизированные методы стохастической аппроксимации. Несмотря на то, что формально подобная процедура предполагается для другого класса задач, а именно когда мы не можем точно вычислить ни значения функции, ни градиент, тем не менее, по вышеупомянутым причинам, целесообразно использовать алгоритм одновременного пробного возмущения (обобщенная процедура Кифера-Вольфовица). В общем случае ее можно записать следующим образом:

$$\begin{aligned}\tilde{\theta}_n &= \tilde{\theta}_{n-1} - \alpha_n \Delta_n \frac{L_n^+ - L_n^-}{2\beta_n}, \\ L_n^\pm &= L(w_n, \tilde{\theta}_{n-1} \pm \beta_n \Delta_n),\end{aligned}$$

где $\{\alpha_n\}$, $\{\beta_n\}$ – числовые последовательности, удовлетворяющие следующим свойствам:

$$\sum_n \alpha_n = \infty, \beta_n \rightarrow 0, \sum_n \frac{\alpha_n^2}{\beta_n^2} < \infty, \sum_n \alpha_n^2 \beta_n^2 < \infty.$$

таким образом, на каждой итерации значение максимизируемого функционала измеряется только дважды. Подробно в [1]. В нашей реализации последовательности $\{\alpha_n\}$ и $\{\beta_n\}$ были взяты, соответственно, $1/n$ и $1/\sqrt{n}$. w_n - возможная помеха.

Стоит также отметить, что процедура, оптимизированная вышеописанным способом, является устойчивой к всевозможным помехам, как системным, так и элементарным погрешностям вычислений. Например, в рамках этой работы все внимание уделялось временной серии ошибок, и процесс вычисления условного математического ожидания доходности не был упомянут. Однако, в реальных условиях может использоваться не самый эффективный метод. Тогда в общем случае, используя принятую нотацию, можно записать, что на входе мы имеем ряд:

$$y_t = \varepsilon_t + w_t = \sigma \eta_t v_t + w_t,$$

где w_t является помехой, которая появляется при измерении ожидаемой величины процесса. В таком случае стандартные градиентные методы не работают, в то время как рандомизированная процедура дает состоятельные оценки, как и в случае без помехи. На каждом из трех этапов алгоритма, при максимизации того или иного функционала, можно считать, что данные не идеальны. В случае использования нашей оптимизации результат тем не менее корректен [1].

При моделировании были рассмотрены случаи с искусственным введением подобной погрешности, причем ненулевого среднего. Наш алгоритм, оптимизированный рандомизированными методами, показал стабильную работу даже в такой ситуации.

Глава IV: Реализация

4.1. Основной двухступенчатый алгоритм

В принципе, для того, чтобы измерить λ_f , необходимо знать выборку истинных инноваций, чтобы иметь возможность посчитать статистическое среднее. Ввиду того, что такой выборки в нашем распоряжении нет, мы заменим ее на специальный набор $\{\tilde{\eta}_t\}$. Для того, чтобы получить его, на первом шаге мы применим стандартный QMLE, в котором параметр штрафа, согласно ранее доказанному, равен единице. Таким образом, вся процедура выглядит следующим образом:

$$\hat{\lambda}_f = \arg \max_{\lambda} \frac{1}{T} \sum_{t=1}^T l_2(\varepsilon_t, \hat{\theta}_T, \lambda) = \arg \max_{\lambda} \frac{1}{T} \sum_{t=1}^T \left(-\ln \lambda + \ln f\left(\frac{\tilde{\eta}_t}{\lambda}\right) \right), \text{ где}$$
$$\tilde{\theta}_T = \arg \max_{\theta} \frac{1}{T} \sum_{t=1}^T \left(-\ln(\sigma v_t) - \frac{\varepsilon_t^2}{2\sigma^2 v_t^2} \right), \text{ а } \tilde{\eta}_t = \varepsilon_t / (\tilde{\sigma} v_t(\tilde{\gamma})).$$

На этом заканчивается первая ступень алгоритма. В результате мы получаем искомый параметр штрафа с какой-то приемлемой погрешностью. После этого надо, как было написано выше, применить не-Гауссов QMLE с поправкой на этот параметр:

$$\hat{\theta}_T = \arg \max_{\theta} \frac{1}{T} \sum_{t=1}^T \left(-\ln(\sigma v_t) + \ln f\left(\frac{\varepsilon_t}{\hat{\lambda}_f \sigma v_t}\right) \right).$$

Таким образом, теоретическая часть процедуры описана полностью. Единственное, что остается сделать – это выбрать метод реализации.

4.2. Адаптивная процедура выбора параметров функции правдоподобия

Из всего вышесказанного следует, что существует два глобальных преимущества выбора правдоподобия с тяжелыми хвостами над стандартным Гауссовым. Во-первых, в рамках принятой нами параметризации, асимптотическая нормальность оценок γ больше не зависит от условий на

четвертый момент инноваций. Вместо этого условие накладывается на $E(h_1^2) / E^2(h_2)$, что значительно слабее и может быть удовлетворено многими распределениями. Во-вторых, даже в случае конечного четвертого момента, на данных финансовой природы наша процедура обладает значительно меньшей дисперсией ошибок.

Следующий логичный шаг – это адаптивный относительно входного временного ряда выбор правдоподобия. Выбор будет совершаться таким образом, чтобы уменьшить потери в эффективности, то есть минимизировать $E(h_1^2) / E^2(h_2)$. В рамках этой работы в качестве семейств, по которым будет производиться минимизация, предлагаются два стандартных семейства с тяжелыми хвостами: Student's t со степенью свободы $\nu > 2$, и Generalized Gaussian с $\beta \leq 1$. таким образом, формально говоря, для истинного распределения и кандидата определим:

$$F(f, g) = \frac{E_g h_1^2}{E_g^2(h_2)}, \text{ где как и раньше } h_1 = 1 + h \left(\frac{\eta}{\lambda_f} \right), \text{ а } h_2 = \frac{\eta}{\lambda_f} h' \left(\frac{\eta}{\lambda_f} \right).$$

Тогда оптимальное правдоподобие выбирается следующим образом:

$$f^* = \arg \min_{\nu, \beta} \{ \{F(f_\nu^t, \bar{g})\}_{\nu > 2}, \{F(f_\beta^{gg}, \bar{g})\}_{\beta \leq 1} \}, \text{ где } \bar{g} - \text{ эмпирическое}$$

распределение, полученное на первой ступени.

4.3. Моделирование

Ввиду того, что волатильность является скрытым параметром, то есть не может наблюдаться напрямую, для демонстрации качества работы нашего модуля результаты прогнозов бессмысленны, так как даже зная реальные ошибки будущих периодов получить волатильность в явном виде – задача невозможная. Вместо этого мы приведем таблицу сравнительного качества оценки параметров GARCH(1,1) модели. Они являются ключевым элементом для прогнозирования, и потому именно на этом этапе недопустимы потери в эффективности. Для того, чтобы показать преимущество наиболее ярко, будем искусственно моделировать процесс ошибок, где инновации – это

распределения со сверхтяжелыми хвостами, в приведенном случае устойчивые распределения с $\alpha \in [1.1, 1.9]$ (α - stable. Распределения трансформированы таким образом, что они не имеют четвертого момента, но второй, как и требуется в изначальных предположениях о модели, равен единице). Волатильность же будем строить с помощью GARCH(1,1) процесса с использованием истинных параметров $(\sigma_0, \alpha_1, \beta_1) = (0.5, 0.35, 0.3)$. В рамках теста было проведено 2500 симуляций с $T = 3000$. Из-за того, что плотности распределения не могут быть выписаны в явном виде, оценки истинного метода максимального правдоподобия труднополучимы. Поэтому мы приведем сравнение с Гауссовым QMLE, который используется в большинстве случаев.

Распределение	Отношение дисперсий			Отношение MSE		
	σ_0	α_1	β_1	σ_0	α_1	β_1
$\alpha = 1.9$	1.266	1.446	1.205	1.285	1.470	1.215
$\alpha = 1.7$	2.502	5.072	2.175	2.551	5.301	2.177
$\alpha = 1.5$	5.381	148.9	4.000	5.600	154.2	3.954
$\alpha = 1.3$	9.774	499.1	6.911	10.11	524.54	6.868
$\alpha = 1.1$	16.078	1313	10.20	16.94	1445	10.00

В соответствующих столбцах стоят отношения дисперсий и среднеквадратичных ошибок (MSE) Гауссова QMLE и нашего метода. Из таблицы ясно видно, что QMLE становится неэффективным по мере уменьшения α , то есть с «утолщением» хвостов. Он выдает слишком большие измерения, демонстрируя яркий сдвиг вверх, особенно для α_1 . В то же время наша процедура отклоняется от истинных значений параметров незначительно, и чем тяжелее хвосты – тем меньше.

Таким образом, описанный пример является яркой демонстрацией преимущества нашего метода, который доказывает работоспособность даже в случаях, на которые не откалиброваны стандартные алгоритмы.

Заключение

В ходе дипломной работы была подробно исследована область прогнозирования волатильности, в частности с помощью GARCH-моделей. Для этого были рассмотрены различные современные методы оценки ее параметров. Была предложена и математически обоснована новая процедура, которая позволяет добиться большей эффективности, нежели стандартные алгоритмы, с помощью более скрупулезного анализа входных данных. Был также разработан метод адаптивного выбора начальных параметров. В проведенных искусственных симуляциях новый алгоритм, в основном, демонстрировал результаты, практически не отличающиеся от случая, когда все скрытые параметры модели нам известны.

Для снижения трудоемкости вычислений было решено использовать оптимизацию с помощью рандомизированных алгоритмов стохастической аппроксимации, так как на различных этапах метод сводится к максимизации выпуклых функционалов. В случае большой размерности задачи, то есть когда модель зависит от многих параметров, подход с использованием SPSA показывает значительно лучшее время работы, нежели стандартные градиентные методы.

Также были проведены специальные тесты с введением дополнительной помехи, которая может трактоваться как неидеальность системы. Разработанный алгоритм, оптимизированный рандомизированными методами, показал стабильную работу даже в такой ситуации.

На основе придуманной процедуры был успешно реализован модуль по прогнозированию волатильности, который был внедрен в более общую финансово-аналитическую систему в финансовом отделе компании «Теплоэнерго» (акт о внедрении прилагается).

Список литературы:

1. Граничин, О.Н. (2003). Введение в методы стохастической оптимизации и оценивания. *Издательство СПбГУ*.
2. Andersen, T.G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Econometric Review*.
3. Barone-Adesi, G. and Engle, R. and Mancini, L. (2007) A GARCH Option Pricing Model in Incomplete Markets.
4. Berkes, I. and Horvath, L. (2004). The efficiency of the estimators of the parameters in garch processes. *The Annals of Statistics*, 32 633-655.
5. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedacity. *Journal of Econometrics*, 31 307-327.
6. Diebold, F. (1988) Empirical modeling of exchange rate dynamics. *Springer*.
7. Drost, F. and Klaassen, C. (1997). Efficient estimation in semiparametric garch models. *Journal of Econometrics*, 81 193-221.
8. Drost, F. and Klaassen, C. (1997). Adaptive estimation in time-series models. *The Annals of Statistics*, 25 786-817.
9. Engle, R.F. and Gonzalez-Rivera, G. (1991). Semiparametric arch models, *Journal of Business and Economic Statistics*, 9 345-359.
10. Engle, R.F. and Gonzalez-Rivera, G. (1999). Efficiency comparisons of maximum-likelihood-based estimators in garch models. *Journal of Econometrics*, 93 93-111.
11. Engle, R. and Focardi, S. and Fabozzi, F. (2006). ARCH/GARCH models in applied econometrics. *Applied Econometrics*.
12. Fan, J. and Wang, M. and Yao, Q. (2008). Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of Royal Statistical Society*, 38 1-15.
13. Hall, P. and Yao, Q. (2003). Inference in arch and garch models with heavy-tailed errors. *Econometrica*, 71 285-317.

- 14.Hsieh, D. A. (1989). Modeling heteroscedacity in daily foreign-exchange rates. *Journal of Business and Economic Statistics*, 7 307-317.
- 15.Moskoni, R. (2008) Assesing GARCH model's predictive ability from trader's point of view, *Financial Mathematics*, 23 45-58.
- 16.Newey, W.K. and Steigerwald, D.G. (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica*, 65 587-599.
- 17.Preminger, A. and Storti, G. (2009). A GARCH(1,1) Estimator With no Moment Conditions on the Error Term.
- 18.White, A. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*.
- 19.Yingfu, X. (2009). Maximum likelihood estimation and forecasting for GARCH, markov switching, and locally stationary wavelet processes. *PhD thesis*.