

## Вступление

В последнее время большой интерес статистиков — ученых и практиков — вызывают непараметрические методы исследований, которые, наряду с классическими параметрическими, значительно расширяют границы статистики.

Новым, еще не достигшим широкой известности, непараметрическим методом является квантильная регрессия.

Прогнозирование в целом и, в особенности, финансовых показателей сопряжено с рядом трудностей.

Колебания показателей финансового рынка могут быть вызваны не только объективными рыночными причинами, они часто связаны с политическими и психологическими факторами, тем более в условиях нестабильной российской экономики. Прогнозировать эти колебания нелегко, и хотя в большинстве случаев удастся определить будущую динамику рынка, но все же проигрывать исключить бывает невозможно.

Инвесторы, стараясь уменьшить свой риск, хотели бы знать пусть не точные значения, но хотя бы границы колебаний будущих значений.

Метод квантильной регрессии, рассматриваемый в данной работе, позволяет как раз получать не только точные прогнозы (случай медианной регрессии), но и предсказывать границы колебаний показателей.

К сожалению, в России этот метод почти не известен. Об этом свидетельствует тот факт, что в российской научной литературе квантильная регрессия до настоящего времени не рассматривалась. Кроме того, отсутствуют переводные публикации зарубежных ученых по данной тематике.

## Глава 1. Введение в теорию квантильной регрессии

Методы непараметрической статистики уже доказали свою широкую применимость, важность и актуальность, но к сожалению, не все из них описаны на русском языке и широко известны.

*Непараметрическими* называют такие методы, которые не предназначены специально для какого-нибудь параметрического семейства распределений и не используют его свойств. Наиболее распространенными непараметрическими методами являются ранговые и знаковые.

Любой непараметрический метод — это статистический метод с некоторыми желательными свойствами, сохраняющимися при относительно слабых допущениях о рассматриваемых генеральных совокупностях, из которых полу-

чены данные. Быстрое развитие непараметрических статистических методов можно отчасти отнести к следующему:

1. Непараметрические методы требуют немногих предположений относительно генеральной совокупности, из которой извлечены данные. В частности, эти методы, в отличие от своих предшественников, не требуют традиционных допущений о нормальности распределения генеральной совокупности.
2. Непараметрические методы часто (хотя не всегда) проще в применении, чем их конкуренты из нормальной теории.
3. Непараметрические приемы обычно понятны.
4. Непараметрические методы применимы в ситуациях, в которых методы классической теории не «работают». Например, для многих непараметрических методов требуются не действительные значения наблюдений, а только их ранги, или их знаки.
5. Хотя может показаться, что, применяя знаковые или ранговые процедуры, теряется много существенной информации, содержащейся в выборке, теоретические исследования показали, что это вовсе не так. Обычно непараметрические процедуры лишь ненамного менее эффективны, чем их конкуренты из классической теории, если выполняются все предположения данной теории, зато они оказываются более эффективны, если данные предположения не выполняются.
6. Итак, непараметрические методы расширяют область применения статистических методов по сравнению с классическими параметрическими методами, в которых считается, что распределение случайных ошибок следует определенному закону, известному с точностью до параметров.

К числу непараметрических методов относится и, еще малоизвестная в России, методика квантильной регрессии.

### **§1. Появление и развитие методика квантильной регрессии**

До настоящего времени справедлив афоризм, ставший знаменитым благодаря Пуанкаре (Poincaré) и цитируемый Крамером (Cramér), что, "каждый верит в [гауссов] закон ошибок, потому что экспериментаторы думают, что это — математическая теорема, математики, что это — экспериментальный факт" [9]. Эта "догма нормальности" в значительной степени относится к своего рода самообольщению. Ошибки, с которыми сталкиваются в эконометрических моделях, как известно, являются суммой большого количества маленьких и независимых элементарных ошибок, и, следовательно, принимая во внимание центральную предельную теорему, будут приблизительно нормально распределенными.

Потребность в альтернативе выборочному среднему (формула оценки наименьших квадратов в модели локализации<sup>1</sup>) стала очевидной в восемнадцатом столетии. В девятнадцатом веке стали широко использоваться, особенно в астрономических вычислениях, медиана и более сложные линейные комбинации порядковых статистик<sup>2</sup>.

В 1821 году Гаусс показал, что выборочное среднее обеспечивает "наиболее вероятную" оценку параметра локализации из случайной выборки с плотностью распределения, пропорциональной  $e^{-x^2/2\sigma^2}$ , но этот результат был скорее освобождением от иррациональности при использовании выборочного среднего, чем требованием эмпирической справедливости этого частного закона распределения ошибок. Обычно на практике, и это было отмечено рядом авторов, наблюдается распределение ошибок с более «длинными хвостами», чем у нормального распределения. В таких случаях было бы желательным выбрать формулы оценки, которые изменили бы выборочное среднее, помещая меньший вес на крайних наблюдаемых значениях.

Одновременно шло осознание потребности появления альтернативы формуле оценки методом наименьших квадратов для линейной модели. Аномальные наблюдения, или "выбросы", как их назвали в последствии, было сложно выделить в таких моделях. Многие знаменитые ученые (Гаусс, Лаплас и Лежандр, и другие) полагали, что, если некоторые наблюдения ненадежны, то минимизация абсолютных отклонений могла бы быть более предпочтительна, чем метод наименьших квадратов<sup>3</sup>. В 1818 году Лаплас доказал, что в простой модели двумерной регрессии без константы формула оценки по наименьшему абсолютному отклонению (last absolute distance, LAD) имеет меньшую асимптотическую дисперсию, чем формула оценки по методу наименьших квадратов (при условии, что ошибки в модели имеют дисперсию  $\sigma^2$ , и плотность в медиане,  $f(0)$ , удовлетворяет следующему условию:  $[2f(0)]^{-1} < \sigma$ . Этот

---

<sup>1</sup> В модели локализации рассматривается задача оценки вектора неизвестных параметров регрессии  $\beta$  для выборки независимых наблюдений за случайными величинами  $Y_1, Y_2, \dots, Y_T$ , распределенными согласно  $P(Y_t < y) = F(y - x_t\beta)$ , где  $x_t = 1, t=1, \dots, T$ . [5]

<sup>2</sup> Порядковыми статистиками называют вариационный ряд и его члены. Вариационный ряд — последовательность  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ , полученная в результате расположения в порядке не убывания исходной последовательности независимых, одинаково распределенных случайных величин  $X_{(1)}, \dots, X_{(n)}$ . вариационный ряд обычно используется в математической статистике как основа непараметрических методов. Вариационный ряд служит для построения функции эмпирического распределения  $f_n(x) = \mu_n(x)/n$ , где  $\mu_n(x)$  - число членов вариационного ряда, меньших  $x$ , которое является оценкой функции распределения  $f(x)$  случайных величин  $x_1, \dots, x_n$ . [1]

<sup>3</sup> Обычно первым формулам оценки, которые минимизируют сумму абсолютных отклонений, приписывается имя Босковича (Boscovitch).

результат положил начало исследованиям большей части теории статистики, основанной на обычных квантилях в модели локализации.

Мостелер (Mosteller) в 1946 году предложил использование ряда так называемой "неэффективной статистики", основанной на нескольких обычных квантилях, для замены «на скорую руку» ("quick and dirty") обычных формул оценки. Было обнаружено, что можно построить формулы оценки этого типа, которые были бы почти столь же эффективны как формулы оценки методом максимального правдоподобия для обычных параметрических моделей. Этот подход был далее развит Беннеттом (Bennett) для строго параметрических моделей, в то время как Гаствич (Gastwirth) и другие установили, что некоторые формулы оценки этого типа эффективны для широкого многообразия распределений. Например, взвешенное среднее значение квантилей  $1/3$ ,  $1/2$ , и  $2/3$  с весами  $0.3$ ,  $0.4$ ,  $0.3$  имеет асимптотическую эффективность почти восемьдесят процентов для гауссова, лапласова, логистического распределений и распределения Коши. Выборочное среднее же имеет асимптотическую эффективность, равную единице в Гауссовом случае, но меньше, чем половина для лапласового распределения и ноль для распределения Коши. Таким образом, хотя эта формула оценки «на скорую руку» может быть "неэффективной статистикой" для любой частной параметрической модели, фактически для предполагаемой "оптимальной" формулы оценки они предпочтительнее, чем выборочное среднее, если имеется некоторая погрешность относительно формы распределения, генерирующей выборку. Много работ было посвящено распространению этих результатов вне установленного числа квантилей к формулам оценки, которые являются линейными комбинациями порядковых статистик. Асимптотическая теория (L) формул оценки стала почти классической благодаря усилиям Бикеля (Bickel), Стиглера (Stigler) и других. Наиболее общая (L) формула оценки локализации — «урезанная» середина, которая является просто серединой выборки после удаления из нее частей самых больших и самых маленьких наблюдений.

Другая линия исследований, основанная на аналогах ранговых процедур в модели локализации, была расширена Jureckova, Jaeskel и другим к линейной модели. Бикель (Bickel) предложил третью линию решения, основанную на аналогах линейных комбинаций порядковых статистик, (L) оценок, из линейной модели. Его оценки — одношаговые итерации для последовательности наблюдений, основанные на некоторой предварительной оценке локализации, например, LAD-оценка.

В 1978 году была опубликована статья Роджера Коенкера (Koenker) и Гильберта Бассета-младшего (Bassett) «Квантильная регрессия» (REGRESSION QUANTILES). Представленная ими модель квантильной регрессии расширяет понятие обычных квантилей в модели локализации до более общего класса линейных моделей, в которых условные квантили имеют линейную форму.

Обобщая модель квантильной регрессии, Пауэлл (Powell) (1984 и 1986) представил модель цензурированной квантильной регрессии. Эта модель последовательно оценивает условные квантили, когда наблюдения за зависимой переменной цензурированы.

В 90-х годах вышли в свет еще ряд зарубежных статей по данной тематике.

В 1997 году опубликована книга М. В. Болдина, Г. И. Симоновой, Ю. Н. Тюрина “Знаковый статистический анализ линейных моделей”. Авторами рассмотрен знаковый непараметрический подход к статистическому анализу данных, который можно отнести к частному случаю квантильной регрессии.

Вышеуказанная работа является единственной публикацией российских ученых, затрагивающей данную тему.

## **§2. Преимущества метода квантильной регрессии**

При построении различных статистических процедур каждый раз исходят из некоторой совокупности условий. В основном эти условия касаются независимости наблюдений и их одинаковой распределенности, предположений о характере распределения элементов выборки.

На практике обсуждаемые условия — результат аппроксимации и неизбежной идеализации. Следовательно, в точном виде эти условия обычно не выполнены, и возникают опасения, связанные с правильностью выводов, сделанных с помощью той или иной выбранной статистической процедуры.

Перед применением разработанных методов необходимо выяснить, как велики должны быть отклонения от принятых условий, чтобы это заставило изменить сформулированные выводы.

С математической точки зрения эта задача очень близка к проблеме устойчивости. В англоязычной литературе для этого круга задач принят термин «робастность» (robustness).

Наиболее распространенными отклонениями от упомянутых выше условий являются следующие:

1. В серии наблюдений присутствует малая доля «выбросов», то есть наблюдений, вызванных каким-либо другим «мешающим» механизмом, отличным от исследуемой системы. Отделить эти наблюдения, как правило, невозможно. Вместо этого ищут процедуры, которые были бы мало чувствительны к таким «засорениям» выборки.
2. Распределение элементов выборки не может быть точно описано.
3. Элементы выборки не являются независимыми.

Среди прикладных методов наиболее распространенным является метод наименьших квадратов, позволяющий получить глубокие статистические результаты при предположении, что случайные ошибки распределены по нормальному (гауссовскому) закону. Но, несмотря на все теоретическое богатство гауссовской модели, положенные в ее основу математические предположения на практике могут не выполняться. В этом случае полученные выводы рискуют оказаться ошибочными. Методов, позволяющих надежно проверить, следуют ли ошибки нормальному распределению или нет, в настоящее время не существует. Кроме того, часто вообще не удается указать какого-либо параметрического семейства, в рамках которого находится конкретное распре-

деление ошибок. Поэтому на практике нужна статистическая теория, ориентированная не на какое-либо параметрическое семейство, а применимая к более широкому классу ошибок, то есть непараметрическая. При этом по кругу решаемых задач и по своим возможностям данный метод не должен уступать гауссовскому, то есть он тоже должен предоставлять целую систему статистической обработки, включая точечное и доверительное оценивание и проверку гипотез.

Предложенный метод квантильной регрессии относится к робастным методам, так как является устойчивым к отклонениям от предположений классических моделей.

Одним из важных свойств квантильной регрессии является то, что данная модель устойчива к «выбросам», которые часто встречаются на практике, в особенности в случае показателей финансового рынка. Поэтому при прогнозе финансовых показателей использование модели квантильной регрессии более предпочтительно, чем использование классических моделей.

Кроме того, на практике в экономике часто некоторые значения зависимой величины могут быть цензурированы. В этом случае оценивание классическими параметрическими методами затруднено. При использовании же модели квантильной регрессии не нужно знать истинную величину зависимой переменной<sup>4</sup>.

Общеупотребительные тесты, основанные на оценках методами наименьших квадратов и абсолютных модулей, свободны от распределения лишь асимптотически.

Для функций распределения с «тяжелыми хвостами» асимптотическая относительная эффективность квантильных оценок может быть больше единицы и сколь угодно большой.

В сущности, единственная серьезная претензия к квантильным процедурам — невысокая асимптотическая относительная эффективность для функций распределений со стандартными, типа гауссовских, «хвостами». Хорошей альтернативой квантильным, и в частности знаковым, методам служат ранговые процедуры. Впрочем, у квантильных процедур перед ранговыми есть серьезное преимущество: хотя ранговые тесты также свободны от вида распределения при конечных  $n$ , но ранговые статистики требуют одинаковой распределенности величин  $u_i$ .

---

<sup>4</sup> В случаях, когда некоторые из наблюдений, которые являются цензурированными справа (right-censored), имеют отрицательные разности, или некоторые из наблюдений, которые цензурированы слева (left-censored), имеют положительные разности, оценка по уравнению квантильной регрессии смещена. В этом случае нужно прибегнуть к модели цензурированной квантильной регрессии.

Основным преимуществом метода является не только то, что он дает хорошие оценки, но и то, что он позволяет делать выводы о размахе колебаний показателей. Для финансового рынка это является наиболее актуальным.

## Глава 2. Теория квантильной регрессии

В данной главе изложена первоначальная модель квантильной регрессии, представленная Кенкером и Бассетом в 1978 году; ее преобразование в задачу линейного программирования; основные формулы оценки асимптотической матрицы ковариации; а так же приведены тесты на гомоскедастичность и симметрию.

### § 1. Общая модель квантильной регрессии

Пусть  $(y_i, x_i)$ ,  $i = 1, \dots, n$  — некоторые переменные, где  $x_i$  —  $K \times 1$ -вектор независимых переменных в уравнении регрессии. Допускается, что

$$\Pr(y_i \leq \tau | x_i) = F_{u_\theta}(\tau - x_i' \beta_\theta | x_i), \quad i = 1, \dots, n.$$

Это соотношение — в другой формулировке — может быть переписано как

$$y_i = x_i' \beta_\theta + u_{\theta i}, \quad \text{Quant}_\theta(y_i | x_i) = x_i' \beta_\theta, \quad (1)$$

Где  $\text{Quant}_\theta(y_i | x_i)$  обозначает условную квантиль  $y_i$  на векторе регрессора  $x_i$ . Нужно отметить, что здесь предполагается, что и  $x_i$ , и  $y_i$  наблюдаются без ошибки, и что уравнение (1) правильно определено. Такие проблемы, как ошибки измерения и пропущенные переменные здесь не рассматриваются. Если (1) определено некорректно (то есть, не линейно), тогда модель можно рассматривать как лучший линейный прогноз для условной квантили.

Если функция  $F_{u_\theta}(\cdot)$  известна, то для оценки  $\beta_\theta$  могут использоваться различные подходы. Однако в данной модели распределение остаточного члена  $u_{\theta i}$  остается неопределенным. Как показывает соотношение (1), единственным предположением является лишь то, что  $u_{\theta i}$  удовлетворяет ограничению на квантиль:

$$\text{Quant}_\theta(u_{\theta i} | x_i) = 0.$$

Вообще,  $\theta$ -ая простая квантиль ( $0 < \theta < 1$ ) переменной  $y$ , скажем  $\hat{\mu}_\theta$ , является решением следующей задачи:

$$\min_b \left\{ \sum_{i: y_i \geq b} \theta |y_i - b| + \sum_{i: y_i < b} (1 - \theta) |y_i - b| \right\}.$$

В частности, медиана ( $\theta=1/2$ ) находится минимизацией суммы модулей:

$$\min_b \sum |y_i - b|.$$

Аналог линейной модели для  $\theta$ -ого квантиля определяется таким же образом. То есть  $\hat{\beta}_\theta$ , оценка для  $\beta_\theta$  в соотношении (1) — называемом  $\theta$ -ой квантильной регрессией — решает задачу

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i: y_i \geq x_i' \beta} \theta |y_i - x_i' \beta| + \sum_{i: y_i < x_i' \beta} (1 - \theta) |y_i - x_i' \beta| \right\} = \min_{\beta} \frac{1}{n} \sum_{i=1}^n \rho_\theta(u_{\theta i}), \quad (2)$$

где  $\rho_\theta(\lambda) = (\theta - I(\lambda < 0))\lambda$  — контрольная функция (check function), а  $I(\cdot)$  — обычная индикаторная функция.  $\theta$ -квантильная регрессия в (2) может быть переписана как

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^n (\theta - 1/2 + 1/2 \operatorname{sgn}(y_i - x_i' b))(y_i - x_i' b) \quad (3)$$

$$\text{где } \operatorname{sgn}(z) = \begin{cases} 1, & z > 0 \\ -1, & z < 0 \end{cases}$$

При  $\theta=1/2$  получаем классическую задачу наименьших расстояний (LAD):

$$\min_{\beta} \frac{1}{n} \sum_i \frac{1}{2} |y_i - x_i' \beta| = \min_{\beta} \frac{1}{n} \sum_i \rho_{1/2}(u_{(1/2)i}) \text{ или}$$

$$\min_{\beta} \frac{1}{2n} \sum_{i=1}^n (\operatorname{sgn}(y_i - x_i' b))(y_i - x_i' b)$$

$K \times 1$  вектор условий первого порядка для данной задачи (3) задается уравнением<sup>5</sup>

$$\frac{1}{n} \sum_{i=1}^n (\theta - 1/2 + 1/2 \operatorname{sgn}(y_i - x_i' \hat{\beta})) x_i = 0. \quad (4)$$

Фактически, можно показать, что условия первого порядка, как они записаны в (4), можно рассматривать как условия для моментов в обобщенном методе моментов (ОММ). Определим функцию момента как

$$\psi(x_i, y_i, \beta) = (\theta - 1/2 + 1/2 \operatorname{sgn}(y_i - x_i' \beta)) x_i. \quad (5)$$

---

<sup>5</sup> Вообще эти условия первого порядка выполняются только приближенно, так как только  $\operatorname{sgn}(0)=0$ . Однако, при  $n \rightarrow \infty$  левая часть этого уравнения сходится к 0, или выражаясь более точно, левая сторона уравнения является . В задаче линейного программирования, представленной ниже, в оптимальном решении

выполнено точное равенство  $\sum_{i=1}^n \varphi_i x_i = 0$

По определению квантиля  $E(\text{sgn}(y_i - x_i'\beta) | x_i) = 1 - 2\theta$ , следовательно, должно быть выполнено условие  $E[\psi(x_i, y_i, \beta_\theta)] = 0$ . Уравнение (4) является выборочным аналогом этого условия<sup>6</sup>. Таким образом, состоятельность и асимптотическую нормальность оценки  $\hat{\beta}_\theta$  для  $\beta_\theta$  можно установить в рамках идеологии обобщенного метода моментов. Для некоторых условий регулярности можно показать что

$$\sqrt{n}(\hat{\beta}_\theta - \beta_\theta) \longrightarrow N(0, \Lambda_\theta),$$

где

$$\Lambda_\theta = \theta(1 - \theta) (E[f_{u_\theta}(0|x_i)x_i x_i'])^{-1} E[x_i x_i'] (E[f_{u_\theta}(0|x_i)x_i x_i'])^{-1}. \quad (6)$$

Нужно отметить, что здесь

$$E[f_{u_\theta}(0|x_i)x_i x_i'] = \partial E[\psi(x_i, y_i, \beta_\theta)] / \partial \beta_\theta',$$

и

$$\theta(1 - \theta) E[x_i x_i'] = E[\psi(x_i, y_i, \beta_\theta) \psi(x_i, y_i, \beta_\theta)']$$

Если  $f_{u_\theta}(0|x) = f_{u_\theta}(0)$  с вероятностью 1 (то есть, плотность остаточного члена  $u_\theta$  оцененного в 0 не зависит от  $x$ ), то  $\Lambda_\theta$  в (6) упрощается до

$$\Lambda_\theta = \frac{\theta(1 - \theta)}{f_{u_\theta}^2(0)} (E[x_i x_i'])^{-1} \quad (7)$$

Как отмечалось выше,  $\theta$ -ая условная квантиль данного  $y$  условная по  $x$  дается  $\text{Quant}_\theta(y_i|x_i) = x_i'\beta_\theta$ , оценка которой  $\hat{\text{Quant}}_\theta(y_i|x_i) = x_i'\hat{\beta}_\theta$ .  $\theta$  лежит в интервале от 0 до 1. На практике, при условии, что любой набор данных содержит только конечное число наблюдений, только конечное число  $\theta$  квантильных оценок будет численно различно, хотя это количество может быть большим. Нужно отметить также, что различные оценки квантильной регрессии скоррелированы.

---

<sup>6</sup> Так же как для математического ожидания переменной  $E(x_i) = \mu$ , выборочным аналогом является  $\frac{1}{n} \sum_i x_i = \mu$ , то есть среднее является оценкой метода моментов для математического ожидания. Оценки МНК также являются оценками метода моментов. Из условия  $E(\varepsilon_i|x_i) = 0$  имеем  $E(\varepsilon_i x_i) = 0$ , выборочный аналог представляет собой следующее  $\frac{1}{n} \sum_i \varepsilon_i x_i = 0$ , а это условия первого порядка для МНК.

## Эффективная оценка

Формула оценки квантильной регрессии, описанная выше, не является эффективной оценкой для  $\beta_\theta$ . Эффективная оценка может быть получена из решения уравнения

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^n f_{u_\theta}(0|x) (\theta - 1/2 + 1/2 \operatorname{sgn}(y_i - x_i' \beta)) (y_i - x_i' \beta)$$

Однако при использовании данной процедуры оценивания требуется некоторая оценка для неизвестной плотности .

Нужно отметить, что оценки квантильной регрессии имеют несколько важных свойств, которые облегчают процедуру вычисления. Обозначим набор возможных решений задачи (3), как  $B(\theta, y, X)$ . Затем для каждого  $\hat{\beta}_\theta \equiv \hat{\beta}(\theta, y, X) \in B(\theta, y, X)$  мы имеем<sup>7</sup>:

$$\hat{\beta}(\theta, \lambda y, X) = \lambda \hat{\beta}(\theta, y, X), \quad \lambda \in [0, \infty), \quad (8)$$

$$\hat{\beta}(1 - \theta, \lambda y, X) = \lambda \hat{\beta}(\theta, y, X), \quad \lambda \in (-\infty, 0] \quad (9)$$

$$\hat{\beta}(\theta, y + X\gamma, X) = \hat{\beta}(\theta, y, X) + \gamma, \quad \gamma \in \mathfrak{R}^k \quad (10)$$

$$\hat{\beta}(\theta, y, XA) = A^{-1} \hat{\beta}(\theta, y, X), \quad A_{k \times k} \text{ неединичная} \quad (11)$$

## §2. Представление квантильной регрессии в виде задачи линейного программирования

Задача (2), т.е.

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i: y_i \geq x_i' \beta} \theta |y_i - x_i' \beta| + \sum_{i: y_i < x_i' \beta} (1 - \theta) |y_i - x_i' \beta| \right\} = \min_{\beta} \frac{1}{n} \sum_{i=1}^n \rho_\theta(u_{\theta i}),$$

может рассматриваться как задача линейного программирования (ЛП). Кроме того, двойственная задача ЛП — приблизительно то же, что и условия первого порядка (4). Пару двойственных задач линейного программирования можно представить следующим образом:

$$\begin{array}{l|l} \theta \cdot \mathbf{1}' u^+ + (1 - \theta) \mathbf{1}' u^- \rightarrow \min & y' \varphi \rightarrow \max \\ X\beta + u^+ - u^- = y & X' \varphi = 0 \\ u^+ \geq 0 & \varphi \leq \theta \cdot \mathbf{1} \\ u^- \geq 0 & -\varphi \leq (1 - \theta) \cdot \mathbf{1} \end{array}$$

При этом

<sup>7</sup> (Koenker и Bassett, 1978. Theorem 3.2)

$$\varphi_i \approx \theta - \frac{1}{2} + \frac{1}{2} \operatorname{sgn}(y_i - x_i \beta)$$

$$u_i^+ = \begin{cases} y_i - x_i \beta, & y_i \geq x_i \beta \\ 0, & \text{иначе} \end{cases}$$

$$u_i^- = \begin{cases} y_i - x_i \beta, & y_i \leq x_i \beta \\ 0 & \text{иначе} \end{cases}$$

Из теоремы двойственности следует существование решения задачи (2), так как обе задачи имеют допустимое решение если матрица  $X$  имеет полный ранг. Из теории линейного программирования можно вывести и некоторые другие полезные свойства задачи (2).

Нужно отметить, что число итераций симплекс-метода при использовании соотношений (8)-(11) существенно сокращается. Если, кроме этого, известно заранее хорошее начальное значение  $\hat{\beta}_\theta$  (обозначим его как  $\hat{\beta}_\theta^0$ ), это может также сократить вычислительное время, так как это сразу помещает большинство наблюдений в «правильную» сторону гиперплоскости. Допустим  $y_R = y - X\hat{\beta}_\theta^0$  и допустим также, что  $\hat{\beta}_\theta^R$  оценка квантильной регрессии  $y_R$  от  $X$ , затем, используя свойство (10),  $\hat{\beta}_\theta = \hat{\beta}_\theta^R + \hat{\beta}_\theta^0$ . Получение  $\hat{\beta}_\theta^R$  и  $\hat{\beta}_\theta^0$  может быть намного быстрее, чем получение непосредственно  $\hat{\beta}_\theta$ .

Одно из возможных начальных значений — скорректированная оценка наименьших квадратов, где константа заменена  $n_\theta$ -ой порядковой статистикой остатков  $\hat{u}_1, \dots, \hat{u}_n$ . Альтернативное начальное значение может быть получено из квантильной регрессии, основанной только на маленькой части выборки; это может быть особенно полезно, когда набор данных очень большой.

Представление модели квантильной регрессии как задачи линейного программирования имеет несколько важных следствий, как с теоретической, так и с практической точки зрения. Во-первых, гарантируется, что оценка уравнения квантильной регрессии будет получена за конечное число итераций симплекс-метода<sup>8</sup>. Во-вторых, в отличие от классического случая регрессии (оценки параметров получены с помощью МНК), оценка вектора параметров устойчива к выбросам. То есть, если  $y_i - x_i' \hat{\beta}_\theta > 0$ , то  $y_i$  можно увеличивать до  $+\infty$ , или, если  $y_i - x_i' \hat{\beta}_\theta < 0$ , то  $y_i$  может быть уменьшен до  $-\infty$  без изменения решения  $\hat{\beta}_\theta$ . Другими словами, имеет значение не величина зависимой переменной, а только лишь то, на какой стороне оцениваемой гиперплоскости находится наблюдение.

---

<sup>8</sup> Обычно, число требуемых итераций относительно мало для эффективного ЛП алгоритма

### §3. Асимптотическая оценка матрицы ковариации.

Уравнения (6) и (7) дают формулы для асимптотической матрицы ковариаций для  $\hat{\beta}_\theta$  согласно двум альтернативным предположениям относительно  $f_{u_\theta}(0|x)$ . Проблемы в оценке матрицы ковариаций в (6) возникают в основном относительно  $f_{u_\theta}(0|x)$ , или в качестве альтернативы  $E[f_{u_\theta}(0|x)xx']$ . Следовательно, оценка  $\Lambda_\theta$  зависит от субъективного решения исследователя, и не существует решительно "лучшего" пути осуществления оценивания. Каждый подход имеет некоторые преимущества и недостатки, и каждый влечет за собой некоторую степень произвольности. Одной из формул оценки для  $\Lambda_\theta$  является оценка, основанная на порядковых статистиках.

#### Оценка на основе порядковых статистик

Эта формула оценки применима, когда  $f_{u_\theta}(0|x) = f_{u_\theta}(0)$ , то есть когда выполнено предположение независимости<sup>9</sup>. Учитывая это предположение, асимптотическую матрицу ковариаций можно упростить до

$$\Lambda_\theta = \sigma_\theta^2 (E[xx'])^{-1}, \quad \text{где} \quad \sigma_\theta^2 = \frac{\theta(1-\theta)}{f_{u_\theta}^2(0)}$$

Второй член — асимптотической дисперсии — легко оценить с помощью  $\widehat{E}(xx') = \frac{1}{n} \sum_{i=1}^n x_i x_i'$ . Первый член,  $\sigma_\theta^2$ , может быть получен из доверительного интервала, построенного для  $n\theta$ -ой порядковой статистики  $\hat{u}_{\theta_1}, \dots, \hat{u}_{\theta_n}$ . Вообще, точный доверительный интервал может быть вычислен для  $\theta$ -ой квантили случайной величины  $Y \sim F_Y(\cdot)$ . В частности,

$$P(y_{(j)} \leq \xi_\theta \leq y_{(k)}) = P(y_{(j)} \leq \xi_\theta) - P(y_{(k)} < \xi_\theta) \quad (12)$$

где  $y_{(j)}$  и  $y_{(k)}$  —  $j$ -ая и  $k$ -ая порядковые статистики  $y_1, \dots, y_n$ , соответственно. Отмечая это

$$P(y_{(j)} \leq \xi_\theta) = P\left(j \text{ или большее наблюдение} \leq \xi_\theta\right) = \sum_{i=j}^n \binom{n}{i} \theta^i (1-\theta)^{n-i}. \quad (13)$$

Аналогично,

---

<sup>9</sup> Эти формулы оценки были предложены Гэри Чемберлином (1994). Подобные методы вычисления, используя точный доверительный интервал для порядковой статистики, могут быть найдены в Эфрона (Efron) (1982) и Хьюбера (Huber) (1981).

$$P(y_{(k)} < \xi_\theta) = \sum_{i=k}^n \binom{n}{i} \theta^i (1-\theta)^{n-i} \quad (14)$$

Подставляя (13) и (14) в (12), получаем:

$$P(y_{(j)} \leq \xi_\theta \leq y_{(k)}) = \sum_{i=j}^{k-1} \binom{n}{i} \theta^i (1-\theta)^{n-i} \quad (15)$$

Построение симметричный доверительный интервал уровня  $1 - \alpha$  для  $\xi_\theta$  прямо следует. Допустим  $j = [n\theta - l]$ ,  $k = [n\theta + l]$ , и  $X \sim B(n, \theta)$ . Затем,

$$\begin{aligned} P(y_{[n\theta-l]} \leq \xi_\theta < y_{[n\theta+l]}) &= P([n\theta - l] \leq X < [n\theta + l]) \\ &\approx P\left(\left|\frac{X - n\theta}{\sqrt{n\theta(1-\theta)}}\right| \leq \frac{l}{\sqrt{n\theta(1-\theta)}}\right) \end{aligned} \quad (16)$$

Так как  $(X - n\theta)/\sqrt{n\theta(1-\theta)} \rightarrow N(0,1)$ , то приравнивая вероятность в (16) к  $1 - \alpha$ , получаем  $l = Z_{1-\alpha/2} \sqrt{n\theta(1-\theta)}$ . Сопоставление длины точного доверительного интервала в (15) с асимптотическим нормальным доверительным интервалом дает оценку для  $\sigma_\theta^2$ :

$$\hat{\sigma}_\theta^2 = \frac{n(y_{[n\theta+l]} - y_{[n\theta-l]})^2}{4Z_{1-\alpha/2}^2}. \quad (17)$$

По существу, то, что мы неявно получили — это формула оценки для  $f_{u_\theta}(0)$ , заданная  $\hat{f}_{u_\theta}(0) = 2Z_{1-\alpha/2} \sqrt{\theta(1-\theta)} / \sqrt{n} (Y_{[n\theta+l]} - Y_{[n\theta-l]})$ . Эта оценка может использоваться при построении оценки для матрицы ковариаций последовательности квантильных регрессий, как обсуждено ниже.

#### **§4. Последовательности квантильных регрессий**

До этого обсуждалась оценка отдельной квантильной регрессии для определенного значения  $\theta$ . Интересно оценить отдельные квантильные регрессии в различных точках условного распределения зависимой переменной. Так как эти квантильной регрессии будут оценены, используя одинаковые данные с различными весами, они должны быть взаимосвязаны. Ниже обсуждается оценка конечной последовательности квантильных регрессий и дается выражение для асимптотического распределения. Рассмотрим модель (2) (нижний индекс для простоты опускается) для  $p$  альтернатив  $\theta$ ,

$$y = x' \beta_{\theta_j} + u_{\theta_j} \quad \text{и} \quad \text{Quant}_{\theta_j}(u_{\theta_j} | x) = 0, \quad j = 1, \dots, p$$

Без потери общности принимают что  $0 < \theta_1 < \theta_2 < \dots < \theta_p < 1$ . Допустим

$$\begin{aligned} \psi(x, y, \beta_1, \dots, \beta_p)' &= \left( \psi_1(x, y, \beta_1)', \dots, \psi_p(x, y, \beta_p)' \right) \quad \text{где} \\ \psi_j(x, y, \beta) &= (\theta_j - 1/2 + 1/2 \operatorname{sgn}(y - x'\beta))x, \quad j = 1, \dots, p \end{aligned}$$

Определим  $p$  функций момента от  $\theta_1$ -ой до  $\theta_p$ -ой квантильной регрессии, соответственно. Допустим, что  $\beta' = (\beta_1', \dots, \beta_p')$  и  $\beta'_\theta = (\beta'_{\theta_1}, \dots, \beta'_{\theta_p})$  — истинные значения.

Под некоторыми условиями регулярности понимаются следующие:

$$E[\psi(x, y, \beta_{\theta_1}, \dots, \beta_{\theta_p})] = 0.$$

Из принципа оценивания по аналогии (Manski, 1988) формула оценки  $\hat{\beta}_\theta$  для  $\beta_\theta$  получается как решение

$$\frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, \hat{\beta}_{\theta_1}, \dots, \hat{\beta}_{\theta_p}) = 0.$$

Однако данная формула оценки не является решением для всех  $\hat{\theta}_j$  ( $j = 1, \dots, p$ ) одновременно. Фактически оценка составляет решение задачи (3) для каждой квантили отдельно, так как нет никаких взаимных ограничений, наложенных на  $\hat{\beta}_{\theta_1}, \dots, \hat{\beta}_{\theta_p}$ .

Условия регулярности  $\hat{\beta}_\theta$  имеют нормальное асимптотическое распределение,

$$\sqrt{n}(\hat{\beta}_\theta - \beta_\theta) \longrightarrow N(0, \Lambda_\theta),$$

где  $\Lambda_\theta = \{\Lambda_{\theta_{jk}}\}_{j,k=1,\dots,p}$  и

$$\Lambda_{\theta_{jk}} = (\min\{\theta_j, \theta_k\} - \theta_j \theta_k) \left( E[f_{u_{\theta_j}}(0|x)xx'] \right)^{-1} E[xx'] \left( E[f_{u_{\theta_k}}(0|x)xx'] \right)^{-1} \quad (24)$$

Снова заметим, что если  $f_{u_{\theta_j}}(0|x) = f_{u_{\theta_j}}(0)$  ( $j=1, \dots, p$ ), то (24) упрощаем до

$$\Lambda_\theta = \Omega_\theta \otimes (E[xx'])^{-1}, \quad \text{где}^{10} \quad (25)$$

$$\Omega_\theta = \left\{ \Omega_{\theta_{jk}} \right\}_{j,k=1,\dots,p} \quad \text{и} \quad \Omega_{\theta_{jk}} = \frac{\min\{\theta_j, \theta_k\} - \theta_j \theta_k}{f_{u_{\theta_j}}(0) f_{u_{\theta_k}}(0)}$$

---

<sup>10</sup> [5], Теорема 4.2

Заметим, что оцененные условные квантили, условные на  $x$ , даются  $x'\hat{\beta}_{\theta_1}, \dots, x'\hat{\beta}_{\theta_p}$ . Так как оценки  $\hat{\beta}_{\theta_j}$  ( $j = 1, \dots, p$ ) для  $p$  квантилей получены независимо в каждой квантили, то нет необходимости в том, чтобы  $x'\hat{\beta}_{\theta_j} > x'\hat{\beta}_{\theta_k}$  если  $\theta_j > \theta_k$ . Фактически, можно всегда найти вектор  $x_0$  так чтобы  $x'_0\hat{\beta}_{\theta_j} > x'_0\hat{\beta}_{\theta_k}$ , т.е. условные квантили расходятся (не совпадают).

### **§5. Тесты на гомоскедастичность и симметрию.**

В предыдущих разделах были представлены асимптотические свойства квантильной регрессии. В данном разделе рассматриваются некоторые практические вопросы, касающиеся этих оценок. В частности, рассматриваются тесты, которые могут быть сделаны при наличии приближенных квантильных оценок. Заметьте, что если  $f_{u_\theta}(0|x) = f_{u_\theta}(0)$ , то любые два вектора параметров квантили,  $\hat{\beta}_{\theta_1}$  и  $\hat{\beta}_{\theta_2}$ , должны отличаться только своей длиной, но не угловыми коэффициентами. Если распределение  $u_\theta$  было симметричным, то совокупность векторов коэффициентов подразумевает несколько иную (альтернативную) структуру. Данный раздел исследует тесты на равенство угловых коэффициентов и на симметрию, используя метод минимального расстояния (minimum distance, MD, то же что LAD).

#### **1 Тест на гомоскедастичность**

Главное различие между методикой, используемой здесь и той, которой придерживаются Koenker и Bassett (1982) — то, что матрица ковариаций для тестовой статистики, используемой здесь, действительна при выполнении нелокальных альтернативных гипотез, в то время как ковариационная матрица, используемая Koenker и Bassett, действительна только при выполнении локальных альтернативных гипотез<sup>11</sup>.

В методе LAD, который используется в данном разделе, сначала оцениваются угловые коэффициенты при ограничениях, следующих из гомоскедастичности. То есть ограниченный вектор коэффициентов минимизируется, относительно  $\beta^R$ ,

$$Q(\beta^R) = (\hat{\beta}_\theta - R\beta^R)' A^{-1} (\hat{\beta}_\theta - R\beta^R), \quad (26)$$

Где  $A$  матрица весов с  $A \rightarrow \psi$  положительно определенной матрицей и где  $\hat{\beta}_\theta = (\hat{\beta}_{\theta_1}, \dots, \hat{\beta}_{\theta_p})$  — неограниченный вектор  $p$  оценок квантильной регрессии<sup>12</sup>.

---

<sup>11</sup> Koenker и Bassett рассмотрели частную модель мультипликативной гомоскедастичности, которая подразумевает некоторую структуру для векторов параметров квантили.

<sup>12</sup> Отмечу, что если  $\psi = \Lambda_\theta$  (определено в (25)), то результирующая оценка для

Замечу, что  $\beta_{\theta}^R = (\beta_{\theta_1}, \dots, \beta_{\theta_p}, \beta_2, \dots, \beta_k)$  является  $(p + K - 1) \times 1$  вектор ограниченных параметров. Матрица ограничений  $R$  задается следующим образом:

$$R' = (R_1, \dots, R_p), \quad \text{где} \quad R_j = \begin{pmatrix} e_j & 0_m \\ 0_v & I_{K-1} \end{pmatrix}$$

$e$  является  $p \times 1$  вектором нулей кроме 1 на  $j$ -ом месте,  $0_v$  — это  $1 \times (K-1)$  вектор нулей,  $0_m$  —  $(K-1) \times p$  матрица нулей, и  $I_{K-1}$  — единичная матрица порядка  $K-1$ . Длины альтернативных квантильных регрессий  $(\beta_{\theta_1}, \dots, \beta_{\theta_p})$  не должны быть равны.

Асимптотическое распределение оптимальной LAD оценки,  $\hat{\beta}_{\theta}^R$ , задается

$$\sqrt{n}(\hat{\beta}_{\theta}^R - \beta_{\theta}^R) \longrightarrow N(0, \Lambda_{\theta}^R) \quad \text{где} \quad \Lambda_{\theta}^R = (R' \Lambda_{\theta}^{-1} R)^{-1}.$$

Метод LAD позволяет напрямую получить тестовую статистику. Для  $\psi = \Lambda_{\theta}$ , определенной в (25), имеем, при выполнении основной гипотезы (равенства угловых коэффициентов):

$$n(\hat{\beta}_{\theta} - R\hat{\beta}_{\theta}^R)' A^{-1}(\hat{\beta}_{\theta} - R\hat{\beta}_{\theta}^R) \longrightarrow \chi_{(pK-p-K+1)}^2. \quad (27)$$

## 2 Тест на симметрию

Допустим, что  $\hat{\beta}_{\theta}$  и  $\hat{\beta}_{1-\theta}$  — оцененные векторы параметра для  $\theta$ -ой и  $(1-\theta)$ -ой квантильной регрессии, соответственно. **Симметрия** подразумевает далее, что  $\beta_{0,5} = 0,5(\beta_{\theta} + \beta_{(1-\theta)})$ , где  $\beta_{0,5}$  — медианный ( $\theta = 1/2$ ) вектор параметра.

Допустим, что  $\beta_{\theta_1}, \dots, \beta_{\theta_p}$  — векторы параметра, связанные с  $\theta_1, \dots, \theta_p$  квантилями, соответственно, и допустим также, что  $p$  — нечетное число. Кроме того, чтобы проверить на симметрию нам нужно также  $\theta_{p-j} = 1 - \theta_{j+1}$  для  $j=0, \dots, (p-1)/2-1$ , и средняя квантиль  $\beta_{\theta_{1+(p-1)/2}}$  для  $\theta_{1+(p-1)/2} = 0,5$ . Из предположения симметрии следует, что  $\beta_{\theta_{p-j}} = 2\beta_{0,5} - \beta_{\theta_{j+1}}$ .

Допустим, что  $\hat{\beta}_{\theta}^U = (\hat{\beta}'_{\theta_1}, \dots, \hat{\beta}'_{\theta_p})$  — сложенный вектор  $p$  неограниченных оценок квантили и  $\beta_{\theta}^R = (\beta'_1, \dots, \beta'_{1+(p-1)/2})$  — ограниченный параметр. Затем оценка для  $\beta_{\theta}^R$  получена из решения

---

$\beta_{\theta}^R$  - оптимальная LAD оценка, чья асимптотическая матрица ковариаций приведена ниже. Если  $\psi \neq \Lambda_{\theta}$ , то асимптотическая матрица ковариаций для LAD оценки —  $\Lambda_{\theta}^R = (R' \psi^{-1} R)^{-1} R' \psi^{-1} \Lambda_{\theta} \psi^{-1} R (R' \psi^{-1} R)^{-1}$ .

$$\min_b (\hat{\beta}_0^U - Rb)' A^{-1} (\hat{\beta}_0^U - Rb), \quad (28)$$

где матрица ограничений  $R$  дается  $R = (R_1', R_2')'$ ,  $R_1$  является единичной матрицей размерности  $(1+(p-1)/2)K$ , и

$$R_2 = \begin{pmatrix} -I_K & 0_K & 0_K & \cdot & \cdot & \cdot & 0_K & 2I_K \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0_K & 0_K & 0_K & \cdot & \cdot & \cdot & -I_K & 2I_K \end{pmatrix}$$

является  $K(p-1)/2 \times K(1+(p-1)/2)$  матрица.

Асимптотическое распределение испытательной статистики, при основной гипотезе, является подобной данному в (27). То есть для  $A \xrightarrow{p} \Lambda_\theta$  при  $n \rightarrow \infty$ , мы имеем

$$n (\hat{\beta}_0^U - R\hat{\beta}_0^R)' A^{-1} (\hat{\beta}_0^U - R\hat{\beta}_0^R) \longrightarrow \chi^2_{((p-1)K/2)}, \quad (29)$$

где  $\hat{\beta}_0^R$  обозначает оценку LAD.

Относительно методики, описанной выше, необходимо заметить, что, во-первых, результатом проверки гипотез является или отклонением, или не отклонение нулевой гипотезы — мы не получаем никакую меру мощности против альтернативных гипотез. Во-вторых, заметим, что в случае испытания однородности, если основная гипотеза не отвергается, то LAD структура обеспечивает оптимальный способ объединения различных угловых коэффициентов. В заключение, замечу, что необходимо выбрать матрицу весов в (26) или (28). Эта матрица может быть любой матрицей (включая неслучайную матрицу), чей предел является не случайной положительной определенной матрицей  $\Psi$ . Для  $A$  в (27) или (29), можно использовать асимптотическую оценку ковариации, описанную выше. Однако в то время как для гетероскедастичности проверяют все допустимые матрицы ковариации, для испытания симметрии, могут использоваться только оценки для общей формулы в (6).

### 3. Тесты, использующее структуру GMM

Альтернативная методика проверки гипотез может применяться, используя обобщенный метод моментов, предложенный Хансеном в 1982 (метод GMM). Функция момента определенная в (23) —  $pK \times 1$ - вектор. При нулевой гипотезе как для симметрии, так и для гетероскедастичности (форм, описанных выше) оценивается меньшее количество параметров, скажем  $q$ , а не  $pK$ . Структура GMM обеспечивает в формуле оценки для  $\beta_\theta^R$ , называемой  $\hat{\beta}_\theta^R$ , которая получена из решение задачи:

$$\min_b \left( \frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, b) \right)' A^{-1} \left( \frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, b) \right) \quad (30)$$

Эффективная оценка может быть получена, если выбрать так, чтобы

$$A \xrightarrow{p} E \left[ \psi(x, y, b) \psi(x, y, b)' \right] \quad \text{при } n \rightarrow \infty,$$

где  $\psi(x, y, b)$  определена в (5). Эта структура обеспечивает нас непосредственной методикой испытаний. Под нулевой гипотезой (или гетероскедастичности, или симметрии) мы понимаем:

$$n \left( \frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, \hat{\beta}_0^R) \right)' A^{-1} \left( \frac{1}{n} \sum_{i=1}^n \psi(x_i, y_i, \hat{\beta}_0^R) \right) \longrightarrow \chi^2(pK - q) \quad \text{при } n \rightarrow \infty$$

Недостаток этого подхода — это то, что решение задачи (30) может быть очень трудно, так как она не представима задачей линейного программирования. Заметим, что из-за линейности условных квантилей оценки, полученные при решении (30), с оптимальным  $A$ , являются (асимптотически) эквивалентными к тем, которые получены при оптимальном LAD, обсужденным выше.

### **Глава 3. Практическое применение метода квантильной регрессии**

Целью данной главы является построение различными методами прогнозов индекса РТС и сравнение полученных результатов.

Для построения непосредственно самого прогноза, необходимо вначале построить модель, описывающую поведение исходных данных во времени. Параграфы 2 и 3 посвящены построению, соответственно, последовательности квантильных регрессий и классической модели временных рядов на основе имеющихся данных.

Данными для расчетов являются валютные значения индекса РТС, описанию которых посвящен параграф 1.

Для того чтобы исключить возможные зависимости между значениями переменной, произведен переход от их абсолютных значений к натуральным логарифмам. Весь анализ ведется уже на их основе.

В последнем параграфе главы приведены некоторые экономические сведения, касающиеся изменения индекса РТС.

#### **§1. Описание данных**

Для анализа взят индекс РТС — официальный индикатор Российской торговой системы. Его расчет ведется с 1 сентября 1995 года ежедневно. В разработке методики расчета индекса РТС, основанной на информации о сделках по наиболее ликвидным акциям, принимают участие ведущие аналитики

компаний — профессиональных участников рынка ценных бумаг. С января 1998 года Индекс РТС рассчитывается каждые полчаса торговой сессии, начиная с 12:00 и заканчивая в 18:10 (в 18:00 индекс не рассчитывается). Значение индекса на 18:10 является значением закрытия. Соответственно значение индекса на 12:00 — это значение открытия.

Индекс рассчитывается в двух значениях — валютном и рублевом. Рублевые значения являются вспомогательными и определяются на основе валютных значений. Индекс (валютное значение) на расчетное время ( $I_n$ ) рассчитывается как отношение суммарной рыночной капитализации акций ( $MC_n$ ), включенных в список для расчета индекса, к суммарной рыночной капитализации этих же акций на начальную дату ( $MC_1$ ), умноженное на значение индекса на начальную дату ( $I_1$ ):

$$I_n = I_1 \cdot \frac{MC_n}{MC_1}$$

$MC_n$  — сумма рыночных капитализаций акций на текущее время в долларах США:

$$MC_{ti} = \sum_{i=1}^N P_{ti} * Q_i$$

где  $Q_i$  — количество акций соответствующего наименования, выпущенных эмитентом на текущую дату,

$P_{ti}$  — цена  $i$ -той акции в долларах США на расчетное время  $t$ ,

$N$  — число наименований акций в списке, по которому рассчитывается индекс.

Рублевое значение индекса РТС ( $I_m$ ) определяется как произведение валютного значения индекса на коэффициент, рассчитанный как отношение текущего значения курса рубля ( $K_n$ ) к начальному значению ( $K_1$ ):

$$I_m = I_n * \frac{K_n}{K_1}$$

Начальное значение индекса: 100 на 1 сентября 1995 года,  $K_1 = 4.447$

С 5 января 2000 года используются следующие начальные параметры:

$$I_1 = 175,2623$$

$$MC_1 = \$32\,648\,637\,998,5513$$

Цена  $i$ -той акции на момент времени  $T$  рассчитывается одним из следующих способов:

1. Если за расчетный период времени в торговой системе было заключено  $n$  сделок с  $i$ -той акцией, то

$$P_i = \frac{\sum_{k=1}^n (P_{ik} \times Q_{ik})}{\sum_{k=1}^n Q_{ik}}$$

$P_{ik}$  — цена  $k$ -той сделки по  $i$ -той акции,

$Q_{ik}$  — объем  $k$ -той сделки по  $i$ -той акции.

2. Если за расчетный период времени в торговой системе не было заключено ни одной сделки с  $i$ -той акцией, но за последние десять торговых дней сделки заключались и рассчитывалась цена согласно п.1, то в качестве цены  $i$ -той акции используется последнее рассчитанное значение (на момент времени  $t-1$ ):

$$P_{it} = P_{it-1}$$

3. Если в торговой системе не было заключено ни одной сделки в течение последних десяти торговых дней, то в качестве цены используется цена лучшего предложения на покупку:

$$P_i = P_{bid_i}$$

4. Если в торговой системе не было заключено ни одной сделки в течение последних десяти торговых дней, а также нет и котировок на покупку, то в качестве цены используется последняя зафиксированная цена лучшего предложения на покупку:

$$P_i = LastP_{bid_i}$$

Список акций для расчета индекса (Приложение 1) состоит из акций, входящих в котировочные листы первого и второго уровней, а также акций, отобранных Информационным комитетом на основе экспертной оценки.

Список для расчета индексов РТС может пересматриваться не чаще, чем один раз в три месяца. При принятии решения о составе списка исследуются характеристики акций за три календарных месяца. В список для расчета индексов включаются акции, входящие на конец исследуемого периода в котировочные листы первого и второго уровней и акции, отобранные на основе экспертной оценки. Изменения в списке вступают в силу через месяц после окончания исследуемого периода.

При изменении списка акций на  $n$ -дату индекс рассчитывается с использованием капитализации ( $MC_n^*$ ) по новому списку. Для предотвращения скачка, обусловленного расчетами по новому списку, производится расчет капитализации ( $MC_{n-1}^*$ ) по новому списку на  $(n-1)$ -дату. Значение  $MC_{n-1}^*$  используется в дальнейшем в качестве капитализации начального дня ( $MC_1$ ), а в качестве

значения индекса на начальную дату ( $I_1^*$ ) — значение индекса предыдущей даты ( $I_{n-1}$ ).

$$I_n = I_1^* \cdot \frac{MC_n^*}{MC_1^*}$$

При прогнозе рассматриваются валютные значения индекса на момент закрытия (Приложение 2).

## **§2. Расчеты с помощью квантильной регрессии**

В данном параграфе производится построение пяти квантильных регрессий при следующих уровнях квантили:  $\theta_1=0.1$ ,  $\theta_2=0.25$ ,  $\theta_3=0.5$ ,  $\theta_4=0.75$ ,  $\theta_5=0.9$ . Во всех случаях расчет ведется на основе одних и тех же данных.

Все результаты расчетов данного параграфа будут представлены в графиках и таблицах, часть из которых дана в приложении.

Целью данной главы является построение прогнозов индекса РТС различными методами и сравнение полученных результатов.

Для построения непосредственно самого прогноза, вначале необходимо построить модель, описывающую поведение во времени исходных данных.

Предполагается, что данные следуют процессу авторегрессии AR(q). Порядок q определяется так, чтобы процесс наиболее точно описывал изменения данных, т.е. при значимых коэффициентах, сумма квадратов остатков и целевая функция были минимальны,  $R^1$  (аналог  $R^2$  в классических моделях) был как можно больше.

Моделирование происходит с помощью статистического пакета *Matrixer*.

Случай 1.1. Важным частным случаем модели квантильной регрессии является медианная регрессия ( $\theta=1/2$ ) или знаковый статистический метод. Рассмотрим его первым. Знаковый метод исходит из предположения, что случайные ошибки могут быть положительными или отрицательными с равными вероятностями  $P\{\varepsilon_i > 0\} = P\{\varepsilon_i < 0\} = 1/2$ . Следовательно, линия регрессии будет проходить так, чтобы с каждой стороны от нее было равное количество наблюдений.

Минимизируемая функция медианной квантильной регрессии имеет следующий вид<sup>13</sup>:

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{1}{2} \operatorname{sgn}(\varepsilon_i) \right) (\varepsilon_i)$$

$$\varepsilon_i = u_i - \beta_0 - \beta_1 u_{i-1} - \dots - \beta_q u_{i-q}$$

---

<sup>13</sup> Здесь и далее, если не оговорено иное,  $n=1127-q$

Также как и раньше,

$$\text{sgn}(z) = \begin{cases} 1, & z > 0 \\ -1, & z < 0 \end{cases}$$

В результате моделирования получился процесс AR(2)<sup>14</sup> с константой равной нулю:

$$u_i = 1.24u_{i-1} - 0.24u_{i-2}$$

Как видно из представленной ниже таблицы все коэффициенты при переменных значимы<sup>15</sup>.

Переменная	Коэффициент	Станд. Ошибка	t-статистика	Знач.
1 U[-1]	1.239501	0.04325296	28.657027296	[0.0000]
2 U[2]	0.239435	0.04324609	5.536576568	[0.0000]

Гипотеза о нормальности распределения ошибок отвергается (см. приложение 3.1). Это говорит о том, что применение в этом случае квантильной регрессии предпочтительнее классических методов, в которых предполагается, что распределение ошибок следует нормальному закону.

---

<sup>14</sup> Количество переменных в уравнении равно 2, т.к., хотя включение дополнительных переменных позволило бы увеличить  $R^1$  (аналог  $R^2$ ), это увеличение не было бы вызвано ростом объясненной дисперсии из-за того, что последующие лаговые переменные сильно скоррелированы.

<sup>15</sup> Для проверки нулевой гипотезы  $H_0: \beta_i=0$  применяется t-критерий. Гипотеза  $H_0$  отвергается, если значимость ошибки, с которой можно отвергнуть гипотезу, меньше 5%.

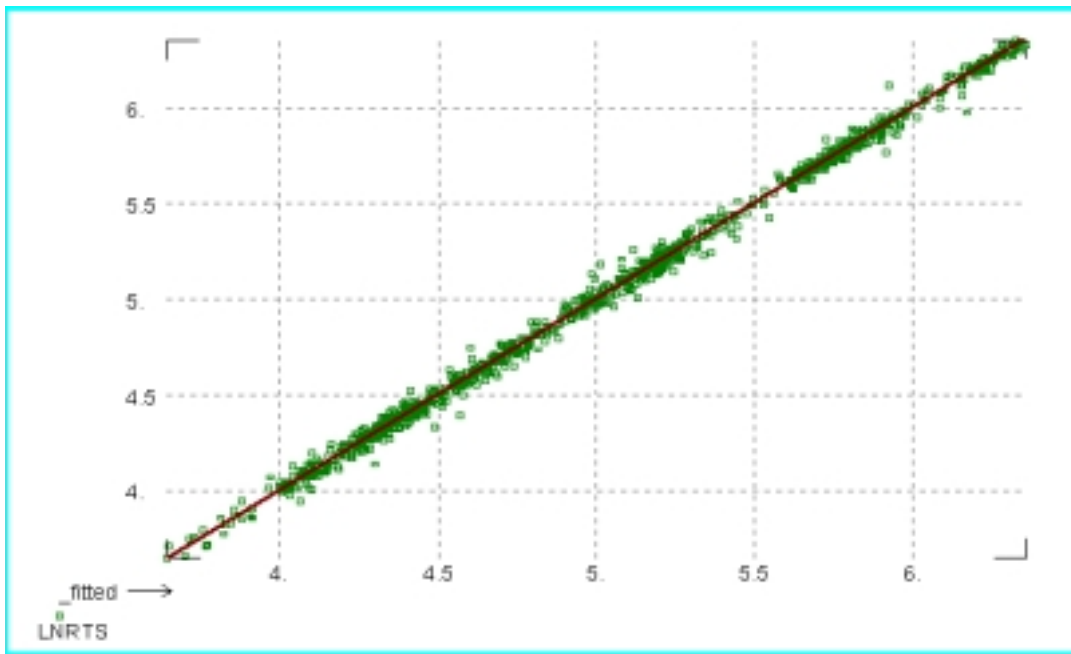


График 1. Диаграмма фактических и расчетных значений ( $\theta=0.5$ ).

На графике 1. представлена диаграмма фактических значений и результатов, полученных построением медианной квантильной регрессии.

Аналогично рассмотрим другие случаи.

Случай 1.2. Построение квантильной регрессии при  $\theta = 0,1$ .

Целевая функция имеет вид:

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i:\varepsilon_i \geq 0} \frac{1}{10} |\varepsilon_i| + \sum_{i:\varepsilon_i < 0} \frac{9}{10} |\varepsilon_i| \right\}$$

или

$$\min_b \frac{1}{n} \sum_{i=1}^n \left( -\frac{2}{5} + \frac{1}{2} \operatorname{sgn}(\varepsilon_i) \right) (\varepsilon_i)$$

$$\text{где } \varepsilon_i = u_i - \beta_0 - \beta_1 u_{i-1} - \dots - \beta_q u_{i-q}$$

По определению квантили, вероятность того, что ошибки меньше нуля будет равна 10%. Следовательно, линия регрессии будет проходить ниже линии медианной регрессии, и большее количество наблюдений будет находиться слева от нее (см. график 2).

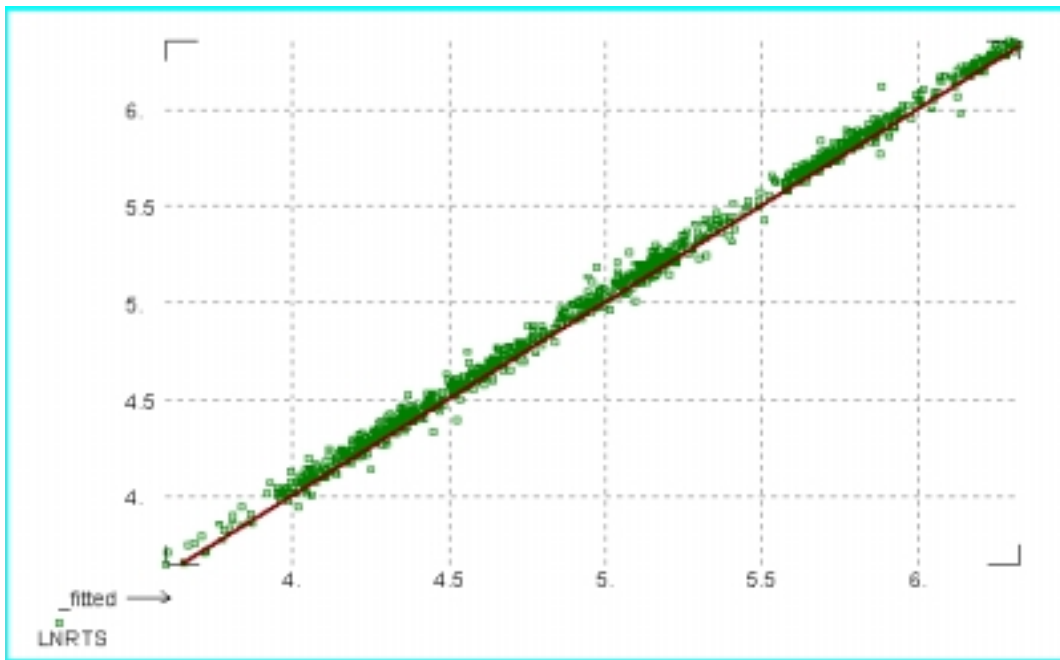


График 2. Диаграмма фактических и расчетных значений ( $\theta=0.1$ ).

В результате моделирования получился процесс AR(2), описывающийся функцией:

$$u_i = -0.056 + 1.27u_{i-1} - 0.267u_{i-2}$$

Все коэффициенты, в соответствии с критерием Стьюдента, с вероятностью ошибки близкой к нулю являются статистически значимыми. Распределение ошибок в данном случае также как и в случае 1.1. не соответствует гауссовскому. (см. приложение 3.П.), т.е. в данном случае, несмотря на нарушение одной из основных гипотез, метод квантильной регрессии дает хороший результат.

### Случай 1.3. Квантильная регрессия при $\theta=0,25$

Целевая функция квантильной регрессии при  $\theta=0,25$  имеет вид:

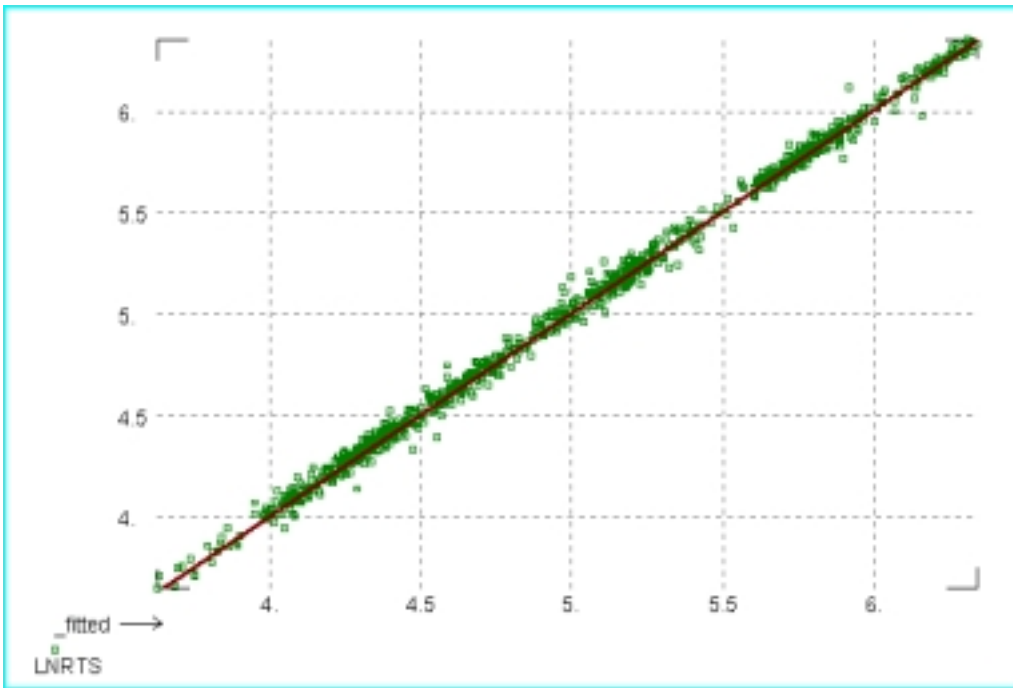
$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i:\varepsilon_i \geq 0} \frac{1}{4} |\varepsilon_i| + \sum_{i:\varepsilon_i < 0} \frac{3}{4} |\varepsilon_i| \right\}$$

или

$$\min_b \frac{1}{n} \sum_{i=1}^n \left( -\frac{1}{4} + \frac{1}{2} \operatorname{sgn}(\varepsilon_i) \right) (\varepsilon_i)$$

$$\text{где } \varepsilon_i = u_i - \beta_0 - \beta_1 u_{i-1} - \dots - \beta_q u_{i-q}$$

График 3. Диаграмма фактических и расчетных значений ( $\theta=0.25$ ).



Так как  $\theta=0.25$ , то вероятность того, что наблюдение лежит справа от линии регрессии, равна 25% (график 3). Следовательно, линия квантильной регрессии при  $\theta=0.25$  будет проходить между линиями регрессии при  $\theta=0.1$  и  $\theta=0.5$ .

Процесс, описывающий поведение переменной во времени, имеет вид:

$$u_i = -0.021 + 1.195u_{i-1} - 0.194u_{i-2}$$

Гипотеза о равенстве нулю коэффициентов отклоняются с вероятностью ошибки близкой 0 (вероятность того, что константа в уравнении не равна нулю 96,2%). Гипотеза о нормальности распределения ошибок, как и в предыдущих случаях, отвергается. Это также как и раньше свидетельствует о том, что применение классических методов было бы затруднено. Оценки параметров и значения основных статистик приведены в приложении 3.III

Случай 1.4.  $\theta = 0,75$  (график 4, приложение 3.IV).

Данный случай противоположен предыдущему. Вероятность того, что ошибки положительны, равна 25% .

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i:\varepsilon_i \geq 0} \frac{3}{4} |\varepsilon_i| + \sum_{i:\varepsilon_i < 0} \frac{1}{4} |\varepsilon_i| \right\}$$

или

$$\min_b \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{4} + \frac{1}{2} \operatorname{sgn}(\varepsilon_i) \right) |\varepsilon_i|$$

$$\text{z}\partial e \quad \varepsilon_i = u_i - \beta_0 - \beta_1 u_{i-1} - \dots - \beta_q u_{i-q}$$

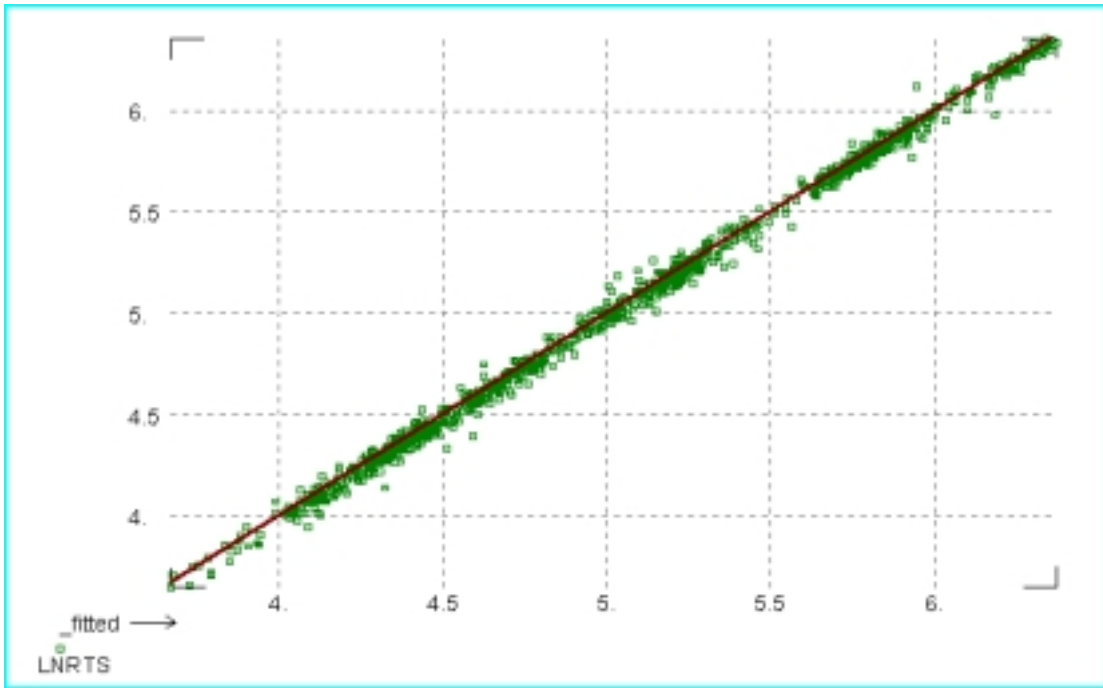


График 4. Диаграмма фактических и расчетных значений ( $\theta=0.75$ )

После получения статистически значимых коэффициентов, линия квантильной регрессии описывается следующим уравнением:

$$u_i = 0.04 + 1.215u_{i-1} - 0.219u_{i-2}$$

Также как и раньше квантильная регрессия демонстрирует устойчивость к отклонению от одного из основных модельных предположений.

Случай 1.5.  $\theta = 0,9$  (график 5, приложение 3.V).

Это случай противоположен случаю 1.2. Здесь, в отличие от случая квантильной регрессии при  $\theta=0.1$ , вероятность отрицательных значений ошибок больше и равна 0.9 (график 5).

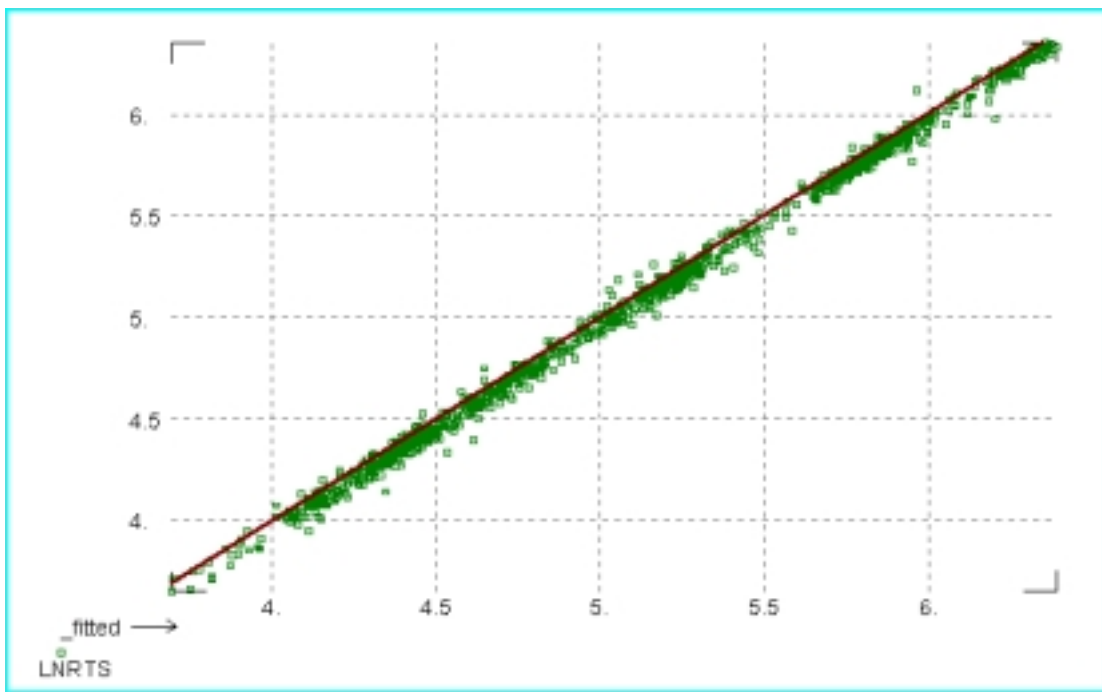


График 5. Диаграмма фактических и расчетных значений ( $\theta=0.9$ )

Целевой функцией является следующая:

$$\min_b \frac{1}{n} \sum_{i=1}^n \left( -\frac{2}{5} + \frac{1}{2} \operatorname{sgn}(\varepsilon_i) \right) |\varepsilon_i|$$

$$\text{где } \varepsilon_i = u_i - b_0 - b_1 u_{i-1} - \dots - b_q u_{i-q}$$

Результатом построения такой квантильной регрессии явилось уравнение:

$$u_i = 0.069 + 1.214u_{i-1} - 0.22u_{i-2}$$

Здесь, как и во всех предыдущих случаях коэффициенты являются статистически значимыми. Требование классических методов о нормальности распределения ошибок не выполнено. Следовательно, применение к любому из описанных выше случаев классических методов было бы затруднено.

### Интерпретация коэффициентов.

Таблица 1. Коэффициенты квантильных регрессий

Коэф <sup>16</sup> .	кван- тиль0.5	кван- тиль0.1	квантиль0.9	Квантиль 0.75	кван- тиль0.25
----------------------	------------------	------------------	-------------	------------------	-------------------

<sup>16</sup>  $\alpha$  — константа,  $\beta_1$  — коэффициент при  $u_{i-1}$ ,  $\beta_2$  — коэффициент при  $u_{i-2}$ .

$\beta_0$		-0,0558472	0,0693064 7	0,040225407	-0,02127
$\beta_1$	1,23950126	1,26958185	1,2143108 6	1,215117555	1,194515921
$\beta_2$	-0,2394353	-0,2666553	-0,2198365	- 0,219334951	-0,19367881

Как видно из таблицы 1, соответствующие коэффициенты в различных квантильных регрессиях различны.

Текущие значения ( $u_i$ ) переменной во всех квантильных регрессиях прямо пропорциональны значениям переменной с лагом, равным 1 ( $u_{i-1}$ ), и обратно пропорциональны значениям переменной с лагом — 2 ( $u_{i-2}$ ).

Различие угловых коэффициентов говорит о том, что в квантильных регрессиях с различными уровнями квантиля, влияние лаговых переменных на текущее значение несколько различно.

### **§3. Расчеты на основе классических методов**

Для сравнения строится модель классическим методом. Из временного ряда выделяются три составляющие: тренд, циклические компоненты и несистематическая или случайная компонента.

Пусть  $x_t$  — временной ряд,  $f(t) = x(t)$  — тенденция или тренд,  $z_t$  — циклическая составляющая,  $\varepsilon_t$  — остаток (случайная последовательность, подчиняющаяся некоторому вероятностному закону).

Нестационарный временной ряд можно представить как:  $x_t = x(t) + z_t + \varepsilon_t$

Для каждой из этих составляющих можно построить модели, описывающие их поведение.

В данном случае, функция тренда представляет собой полином третьей степени

$$x(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3,$$

параметры которого находятся с помощью обычного метода наименьших квадратов.

После исключения из нестационарного ряда тренда, должен получиться стационарный ряд, состоящий из систематической (или циклической) компоненты и белого шума.

В общем виде уравнение циклической составляющей можно представить в виде ряда Фурье:

$$z_t = \sum_{j=0}^{\lfloor T/2 \rfloor} \left( \alpha_j \cos \frac{2\pi jt}{T} + \beta_j \sin \frac{2\pi jt}{T} \right).$$

В данном случае выделяется одна циклическая компонента на частоте 1,02.

Остатки описываются процессом AR(2).

Все рассчитанные коэффициенты и некоторые статистики приведены в приложении 5.

Интересно отметить то, что гипотеза о нормальности распределения остатков отклоняется, следовательно, нарушено одно из основных предположений классического метода.

### §3. Построение прогнозов

Используя полученные квантильные регрессии и классическую модель временного ряда, можно по каждой модели спрогнозировать будущие значения. Прогноз делается на 10 пунктов вперед.

Для более точного прогноза необходимо, чтобы дисперсия ошибки была минимальна.

Находясь в начальной точке отчета  $t$ , производится прогноз величины  $z_{t+L}$ , который обозначается как  $\hat{z}_t(L)$ .

$z_{t+L} = \hat{z}_t(L) + e_t(L)$ , где  $e_t(L)$  – ошибка прогноза.

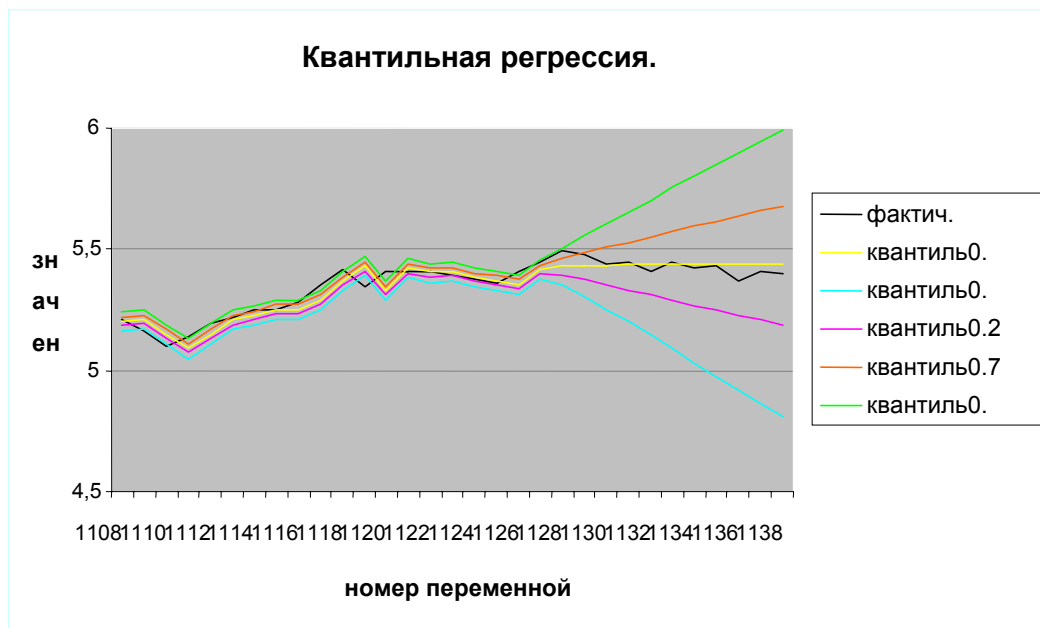


График 6

На графике 6 представлены последние 20 расчетных и 10 прогнозных значений, а также фактические значения за этот период<sup>17</sup>. Значения, полученные на основе медианной квантильной регрессии, могут служить в качестве точных прогнозных значений, а значения других квантилей — границами возможных колебаний показателя.

Отдельно фактические и расчетные значения, полученные знаковым методом показаны на графике 7.

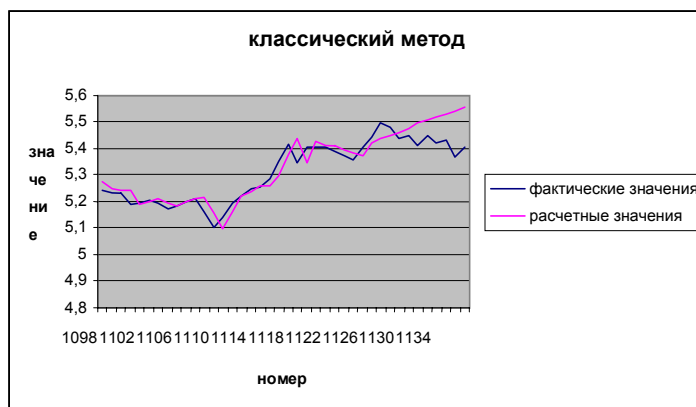
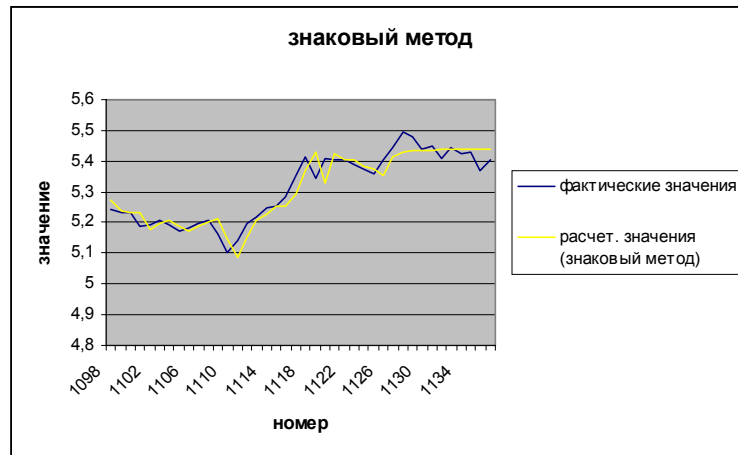


График 8. Построение прогноза на основе классического метода

Из графиков видно, что, хотя построенные квантильные регрессии и традиционная модель хорошо описывают происходящий процесс, прогнозные значения (начиная с 1128 номера), полученные медианной квантильной регрессией, более точны.

<sup>17</sup> Полностью график представлен в приложении 4

В таблицах 1 и 2 приведены ошибки при прогнозе знаковым и классическим методами, соответственно.

Таблица 1. Знаковый метод

номер	фактич.	прогноз	ошибки
1128	243,92	228,21992	15,70008
1129	239,86	229,10403	10,75597
1130	229,88	229,39847	0,481526
1131	232,45	229,55135	2,898645
1132	223,92	229,67035	-5,75035
1133	231,88	229,78128	2,098723
1134	226,22	229,89032	-3,67032
1135	228,17	229,99895	-1,82895
1136	214,59	230,10753	-15,5175
1137	222,43	230,21614	-7,78614

Таблица 2

номер	фактич.	прогноз.	ошибки
1128	243,92	230,4504	13,46958
1129	239,86	232,7382	7,121839
1130	229,88	235,5695	-5,68948
1131	232,45	239,4258	-6,97577
1132	223,92	243,6409	-19,7209
1133	231,88	247,1894	-15,3094
1134	226,22	249,7104	-23,4904
1135	228,17	251,8896	-23,7196
1136	214,59	254,8143	-40,2243
1137	222,43	258,9401	-36,5101

Дисперсия ошибки при применении знакового метода равна 72,6; при классическом методе — 272. Видно, что дисперсия ошибки в классическом методе почти в 4 раза больше. Следовательно, прогноз, полученный знаковым методом лучше.

#### §4. Некоторые экономические факты

События на мировом фондовом рынке, происшедшие в октябре 1997 года — дестабилизация ситуации в странах Юго-Восточной Азии и цепной кризис мировых финансовых рынков — не могли не отразиться на развивающемся российском финансовом рынке. 28 октября кризисные события затронули Россию, индекс РТС упал за сутки почти на 20% (см. график 9). После этих событий общая тенденция к снижению индекса сохранялась вплоть до ноября 1998 года<sup>18</sup>, после чего, благодаря мерам, принятым правительством Е. Примакова, а так же окончанию общемирового кризиса, он начал медленно расти. Рост индекса продолжается до сих пор.

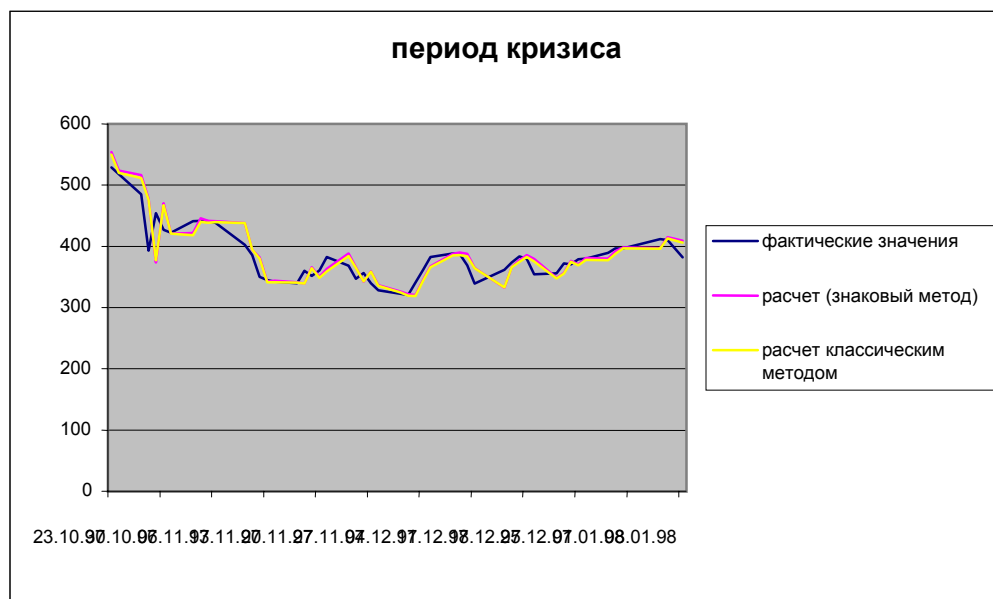


График 9

В связи с кризисами и другими событиями, резко менявшимися значения индекса, расчетные значения отличаются от фактических, этим объясняются в данные периоды большие остатки в моделях (см. график 10).

Хотелось бы отметить также, что на рост котировок в этом году будут оказывать сильное влияние три основных фактора.

Первый фактор, конечно, «президентский»: избрание Владимира Путина главой государства инвесторы, особенно иностранные, наверняка оценят как сигнал к увеличению масштабов операций.

<sup>18</sup> индекс начал расти с 6.10.98

Второй фактор носит более глобальный характер. В 2000 году прогнозируется новый виток внимания к развивающимся рынкам. По оценкам ряда крупных мировых инвестбанков, восстановление доверия должно перерасти в бум (что соответствует доходности инвестиций в местные акции примерно в 300% годовых). Естественно, эта волна должна увеличить и котировки российских акций.

Третий фактор — цены на нефть. Россия, к сожалению, все еще остается сырьевым придатком Запада. А это значит, что валютная часть дохода бюджета будет формироваться за счет экспорта газа и нефти. И если цены на газ меняются относительно медленно, то нефть дорожает. Следовательно, валютные доходы государства будут расти, а это должно вызвать и рост внутреннего потребления. Таким образом, ожидание улучшения платежного баланса государства и доходов населения вызовет рост активности инвесторов, как российских, так и иностранных.

Но существуют и негативные факторы, которые могут сильно повлиять на российский рынок акций. Один из таких факторов: риск обвального падения индекса Доу-Джонса. Крах американского фондового рынка может вызвать самые непредсказуемые последствия.

## **Заключение**

Наиболее важные моменты моделей квантильной регрессии могут быть представлены следующим образом:

- (a) модели могут использоваться для того, чтобы характеризовать полное условное распределение данной зависимой переменной по набору независимых переменных в уравнении регрессии;
- (b) модель квантильной регрессии может быть представлена как задача линейного программирования (ЛП), что упрощает оценку;
- (c) как и минимизируемая функция LAD (формула оценки по наименьшему абсолютному отклонению), целевая функция квантильной регрессии — взвешенная сумма абсолютных отклонений, которая дает хорошую меру локализации, так как оцениваемый вектор коэффициентов не чувствителен к посторонним наблюдениям по зависимой переменной;
- (d) когда ошибки не подчинены нормальному закону распределения, формулы оценки квантильной регрессии могут быть более эффективны, чем МНК-оценки;
- (e) возможные различные решения при различных квантилях могут интерпретироваться как различия в реакции зависимой переменной к изменениям независимых переменных (регрессоров) уравнения регрессии в различных точках условного распределения зависимой переменной;
- (f) L-оценки, основанные на линейной комбинации квантильных формул оценки, вообще более эффективны, чем МНК-оценки.

Проведенные в дипломной работе расчеты позволяют с большой долей уверенности сказать — метод квантильной регрессии предпочтительного классического метода. С помощью медианной квантильной регрессии можно, так же как и при использовании классических методов, строить прогнозные значения. Результаты прогнозов в этом случае значительно точнее.

Использование модели при других уровнях квантили позволяет получать прогноз с заранее определенной вероятностью (границы риска). Это свойство квантильной регрессии должно быть особенно привлекательно для инвесторов, так как риск неудачи может быть значительно снижен.

Вне сомнения, метод квантильной регрессии найдет широкое применение в практике статистических исследований и прогнозов. Возможности, которые он открывает — тому свидетельство.

### **Использованная литература**

1. Математический энциклопедический словарь/ под ред. Прохорова М.,1988
2. Болдин М.Б., Симонова Г.И., Тюрин Ю.Н. Знаковый статистический анализ линейных моделей. М.,1997
3. Рунион Р. Справочник по непараметрической статистике. М.,1982
4. Холлендер М., Вульф Д.А. Непараметрические методы статистики. М.,1983
5. Koenker and G. Bassett, Jr. "Regression Quantiles," *Econometrica*, Vol.46 No1 (January, 1978)
6. W. Muller. "Quantile Regression". Institut fur Angewandte Mathematik, Universitat Heidelberg, Germany, Preprints Nr 624, April 1991.
7. Buchinsky. "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research." Brown University and NBER.1997
8. Robert F. Engle, Simone Manganelli. "CAViaR: Conditional Value at Risk by Quantile Regression." Working Paper 7341, National Bureau of Economic Research, September 1999
9. Cramer, *Mathematical Method of Statistics*. 1946
10. Peter C. B. Phillips, James W. McFarland. "Forward Exchange Market Unbiasedness: the Case of the Australian Dollar Since". 1996
11. <http://www.fe.msk.ru/infomarket/rinacoplus>
12. <http://www.paufor.ru/scripts/rts/index.idc?>