

АВТОМАТИЗИРОВАННЫЙ АНАЛИЗ ВЛИЯНИЯ СТАТИСТИЧЕСКОЙ НЕОПРЕДЕЛЕННОСТИ НА ТОЧНОСТНЫЕ ХАРАКТЕРИСТИКИ ОЦЕНОК ПАРАМЕТРОВ ТОРГОВОЙ СИТУАЦИИ

МУСАЕВ А.А.

УДК 006.72

Мусаев А.А. Автоматизированный анализ влияния статистической неопределенности на точностные характеристики оценок параметров инвестиционной ситуации

Аннотация. В статье представлены результаты численных исследований качества применения традиционных вариантов статистической обработки данных к задачам оценивания и прогнозирования состояния рынков капитала.

Ключевые слова: статистическое оценивание, прогнозирование, метод наименьших квадратов.

Musaev A.A. The numerical analysis of influence statistical uncertainty on accuracy of trade situation parameters estimation.

Abstract. In article are presented results of numerical researches of application of traditional variants statistical data processing to the capital markets estimation and forecasting accuracy.

Keywords: statistical estimation, forecasting, method of the least squares.

1. Введение. Статистические характеристики реальных процессов, протекающих на рынках капитала, существенно отличаются от типовых математических моделей, используемых в задачах прогнозирования и поддержки принятия решений [1, 2]. Возникает вопрос: насколько существенно влияют подобного рода отклонения на эффективность формируемых инвестиционных решений? Очевидно, что эффективность решений, с точки зрения технического анализа динамики капитала, будет в свою очередь, определяться точностью оценки текущей ситуации и достоверностью прогноза ее развития на интересующий инвестора период. Таким образом, в качестве метрики влияния вариаций вероятностных характеристик реальных экономических процессов на качество формируемых решений естественно использовать изменение точности оценок, используемых при идентификации и прогнозе состояния характеристических показателей рынков капитала.

В свою очередь, в качестве оценки точности оценивания в практике статистического анализа обычно используются оценка среднеквадратического отклонения (СКО) или, в более общем случае, полный квадрат ошибки d^2 (ПКО), включающий в себя оценку дисперсии σ^2 и квадрата смещения оценки b^2 . Численная оценка ПКО для за-

данного объема выборки наблюдений n осуществляется исходя из известного соотношения

$$d^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - x)^2, \quad (1)$$

где x — истинное значение оцениваемой величины.

Следует заметить, что проблема борьбы с априорной статистической неопределенностью не нова и ее многочисленные частные решения представлены в публикациях [3, 4]. Однако в большинстве случаев данные решения основаны на тех или иных смягчающих допущениях, позволяющих получить вполне приемлемые для практики частные решения. В то же время практика обработки данных, полученных путем мониторинга состояния рынков капитала, в силу безынерционности наблюдаемых процессов и большого числа факторов влияния со стороны среды взаимодействия, столкнулась с неопределенностью самого общего вида. Иными словами, протекающие процессы по своей природе являются существенно неэргодичными и нестационарными и не допускают каких-либо комфортных для аналитиков предположений относительно природы данных.

В связи с этим для анализа влияния указанной неопределенности следует рассматривать множество моделей данных, имитирующих различные типы возможных вариаций статистических характеристик.

2. Основные статистические модели. Наиболее типичной моделью для стохастических информационных процессов является аддитивная смесь детерминированной компоненты $x(t)$, отражающей базовые тенденции рынка, со стационарным случайным процессом $v(t)$, ответственным за локальные несистемные флуктуации, т. е.

$$y(t) = x(t) + v(t).$$

При этом детерминированная компонента, описывающая динамику среднего, как правило, достаточно успешно аппроксимируется полиномиальной моделью вида

$$x(t) = \sum_{i=0}^n a_i t^i.$$

Действительно, в соответствии с теоремой Вейерштрассе об аппроксимации [5], любой аperiодический процесс может быть с достаточной точностью аппроксимирован конечным полиномом.

Заметим, что в случае использования простейшей модели с одним предиктором оценки по методу наименьших квадратов сведутся к простейшим выражениям вида [6]:

$$\hat{a}_1 = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}, \quad \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x},$$

где $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$, $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ — выборочные средние.

При необходимости учета периодических колебаний процесс может быть дополнен тригонометрическим рядом Фурье.

Для моделирования случайной динамической компоненты, как правило, используется тот или иной стационарный процесс. Разумеется, он не в полной мере отвечает структуре реальных флуктуаций на рынках капитала, однако позволяет использовать широкий арсенал методов и алгоритмов обработки данных, разработанных в прикладной статистике. Очевидно, что в этой ситуации возникает вопрос о том, насколько сохраняются оптимальные свойства традиционно используемых схем обработки. Более точно вопрос состоит в оценке снижения эффективности указанных алгоритмов оценивания, прогнозирования и принятия решений в условиях вариации статистической структуры нестационарных реальных данных относительно используемой вероятностной модели. На этот вопрос ответить аналитически крайне сложно в силу бесконечного числа возможных модификаций нестационарных процессов. В связи с этим рассмотрим данный вопрос численно, опираясь на некоторый конечный набор типичных отклонений от традиционной модели, выявленный в работах [1, 2].

3. Численные исследования статистической устойчивости. В качестве первого примера рассмотрим случай, когда случайные вариации v_k , $k = 1, \dots, N$, образуют ряд автокоррелированных данных. Простейшей моделью таких процессов может служить модель временных рядов типа авторегрессии первого порядка AR(1) вида

$$v_k = \alpha v_{k-1} + \xi_k, \quad k = 1, \dots, N,$$

где ξ_k — традиционный стационарный гауссовский процесс с распределением $N\{0, 1\}$.

На рис. 1 приведены графики, образованные аддитивной смесью линейного полинома с шумовой компонентой указанного типа AR(1) для разных значений α .

Из приведенных графиков можно видеть, что увеличение коэффициента α в модели AR делает случайный процесс более структури-

зированным, формируя в нем существенные колебательные непериодические процессы. При визуальном сравнении с процессами, наблюдаемыми на рынках капитала, визуально видно, что AR-модель (даже без системной компоненты) значительно лучше соответствует их реальной динамике, чем гауссовская.

Для оценки влияния отклонений, связанных с наличием автокорреляции, оценим параметры систематической компоненты, т. е. коэффициентов a_0 , a_1 , для указанной выше модели с различными значениями α для 1000 реализаций описанного процесса, каждая из которых содержит, как на приведенных выше графиках, по 100 временных наблюдений. Качество восстановления искомым коэффициентов оценим на основе усредненной (по числу реализаций) относительной ошибки. На рис. 2 приведены графики изменения значений данной ошибки оценивания для коэффициентов полиномиальной модели a_0 , a_1 при изменении коэффициента авторегрессии от 0 до 1. Можно видеть, что при значении $\alpha = 0.7$ средняя относительная ошибка оценивания превышает 10 %, а дальнейший рост автокорреляции снижает точность восстановления коэффициентов модели до 70–80 %.

Очевидно, что столь существенные погрешности в идентификации систематической составляющей модели непосредственно отразятся и на качестве прогноза состояния рынка.

В качестве другого примера влияния автокорреляции рассмотрим задачу одношагового прогнозирования состояния процесса на основе линейной регрессионной модели, параметры которой оцениваются по выборке на скользящем окне наблюдения.

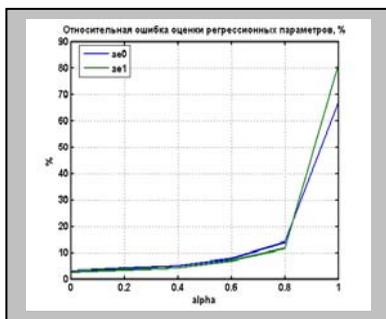


Рис. 2. Изменение относительной ошибки оценки параметров регрессионной модели в зависимости от величины коэффициента авторегрессии.

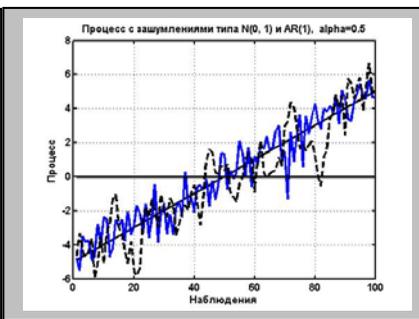


Рис. 3. Линейный процесс со случайными составляющими $N(0, 1)$ (сплошная линия) и $AR(1)$ с $\alpha = 0.5$ (пунктир).

На рис. 3. представлен линейный процесс (тренд) с двумя вариантами случайной составляющей. Первый вариант (сплошная линия) представляет собой традиционный стационарный гауссовский процесс с распределением $N(0, 1)$, второй вариант — процесс с автокорреляцией, описываемый моделью авторегрессии первого порядка $AR(1)$ с коэффициентом $\alpha = 0.5$. Визуально данные процессы вполне различимы, процесс $AR(1)$ имеет заметно более низкий уровень вариабельности.

Для сравнительной оценки вариантов использовалась эргодичность обоих процессов, т. е. оценка качества прогноза осуществлялась по одной выборке (одному временному ряду) длиной в 1000 отсчетов.

Оценка параметров модели в обоих случаях осуществлялась по методу наименьших квадратов (МНК) на скользящем окне наблюдений размер в 50 отсчетов.

В качестве прогностической схемы использовалась простейшая линейная экстраполяция на один отсчет.

Оценка качества прогноза на каждом шаге осуществлялась путем усреднения абсолютных значений ошибок прогноза на каждом шаге для набора значений коэффициента авторегрессии $\alpha = 0 : 0.25 : 1$. При этом первое значение набора соответствует независимому гауссовскому процессу. Соответствующие результаты качества прогноза представлены на рис. 4, из которого видно, что точность прогнозирования нелинейно зависит от степени его автокоррелированности. Наиболее

существенное снижение точности наблюдается при значениях коэффициента автокорреляции, превышающих 0.4.

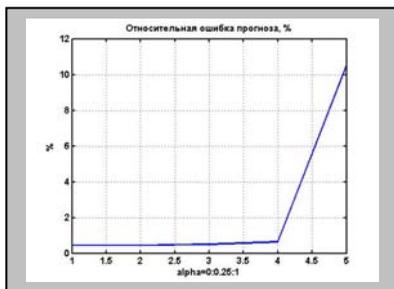


Рис. 4. Оценка относительной погрешности линейного прогноза в зависимости от величины α .

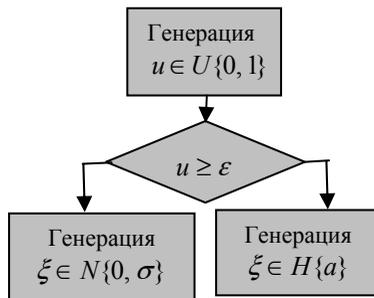


Рис. 5. Структурная схема программы генератора случайных данных, описываемых моделью Тьюки.

Разумеется, точность прогноза для AR-процесса можно существенно повысить, если использовать адекватный алгоритм обработки. Соответствующая оптимальная вычислительная схема последовательной обработки для конкретного варианта AR или ARMA (модель авторегрессии — скользящего среднего) процесса приведена, например, в работе [7]. Проблема состоит в том, что реальный процесс не является стационарным. Это означает, что параметры AR-процесса меняются во времени, а адаптивная подстройка к этим изменениям практически неприемлема в силу безынерционного характера информационных процессов, отражающих динамику рынков капитала.

В качестве третьего примера рассмотрим влияние больших отклонений на точность прогнозирования. В технической и математической литературе такие отклонения называются выбросами или аномальными наблюдениями. Однако в отношении рыночной динамики данные термины не вполне корректны. Внезапное или скачкообразное изменение тех или иных котировок не является аномальной ошибкой наблюдения, это реальное изменение стоимости актива, которое даже при одномоментной скачкообразной реализации может вызвать крайне серьезные последствия для инвестора или спекулянта. Поэтому типовые схемы обработки аномалий, такие как их отбраковка или винзорирование, могут оказаться непригодными. Однако оценка их влияния на среднестатистические характеристики качества прогноза могут ока-

заться полезными с точки зрения формирования соответствующих вычислительных схем прогнозирования и принятия решений.

С этой целью рассмотрим две схемы формирования стохастической компоненты модели изменения состояния котировок.

Первая схема, используемая как базовая платформа, представляет собой описанный выше стационарный гауссовский процесс с параметрами $N(0, 1)$. В качестве второй схемы, имитирующей последовательность наблюдений с повышенным уровнем скачкообразных изменений состояния (будем по-прежнему называть их аномалиями), используем модель данных с засоренным (contaminated) распределением, предложенную Хьюбером [8]:

$$CNN(\varepsilon, \sigma^2) = (1 - \varepsilon)N(0, 1) + \varepsilon N(0, \sigma^2), \quad (2)$$

где $\varepsilon \in [0, 1]$ — уровень засорения; $N(0, \sigma^2)$ — засоряющая компонента с уровнем дисперсии существенно большим, чем у основной компоненты (в данном случае $N(0, 1)$).

В более общем случае, в соответствии с моделью Тьюки [9]:

$$CNH(a) = (1 - \varepsilon)N(0, 1) + \varepsilon H(a),$$

где засоряющая компонента $H(a)$ может иметь произвольное распределение, имитирующее «утяжеление» хвостов распределений, асимметрию, эксцессы и другие отклонения формы от базовой модели.

Структурная схема программы генератора случайных данных, описываемых моделью Тьюки, приведена на рис. 5. Здесь u — случайная величина из генеральной совокупности, подчиненной равномерному распределению на интервале $[0, 1]$. В случае, когда u превышает значение $\varepsilon \in [0, 1]$, генерируется случайная величина с распределением $N(0, \sigma^2)$ (базовая компонента). В противном случае с небольшой вероятностью $P = \varepsilon$ генерируется засоряющая компонента из совокупности с распределением $\xi \in H\{a\}$, где a — векторный параметр числовых характеристик H .

На рис. 6 приведен график линейного процесса со стохастическими аддитивными компонентами гауссовского типа без засорения и с 10%-м засорением аномальными флуктуациями с утроенной дисперсией.

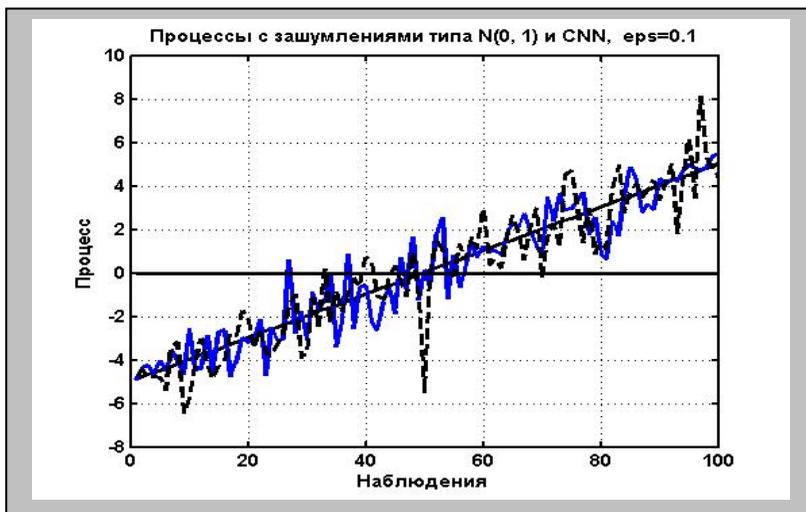


Рис. 6. Линейный процесс со случайными составляющими типа $N(0, 1)$ (сплошная линия) и CNN с $\varepsilon = 0.1$, $\sigma = 3$ (пунктир).

Рассмотрим решение поставленной выше задачи сравнительного анализа результатов оценивания параметров линейного регрессионного процесса для этих двух случаев. Степень засорения ε будем изменять в пределах 0–0.9 с шагом 0.1. Линейный процесс с теми же параметрами $a_0 = -5$, $a_1 = 0.1$ оценим на интервале из 100 наблюдений. Предположение эргодичности не используем, усредненную оценку для каждого ε осуществляем по 100 сериям наблюдений. Качество идентификации процесса оцениваем относительной ошибкой в процентах.

На рис. 7 и 8 приведены значения усредненных относительных ошибок оценивания для различных ε соответственно для параметров a_0 и a_1 в случаях использования моделей традиционного гауссовского процесса и засоренного гауссовского процесса (2). Из приведенных графиков видно, что изменение уровня засорения от 0 до 0.9 приводит к росту усредненной относительной ошибки оценивания параметра положения от 4 до 10 %, а параметра скорости — от 4 до 8,5 %. Достаточно очевидно, что такие ошибки в оценке параметров, особенно ошибки в оценивании скорости, существенно снижают точность экстраполяционного прогноза.

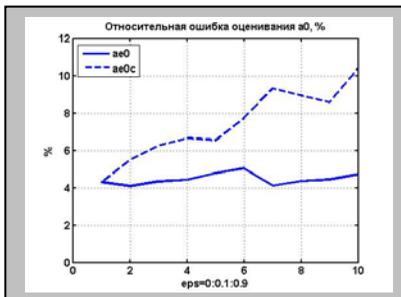


Рис. 7. Изменение усредненной относительной ошибки оценивания параметра положения в зависимости от уровня засорения.

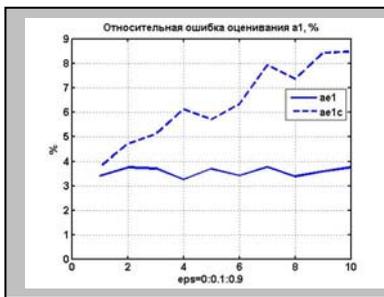


Рис. 8. Изменение усредненной относительной ошибки оценивания параметра скорости в зависимости от уровня засорения.

4. Заключение. Приведенные исследования далеко не исчерпывают множество вариантов структур стохастической компоненты наблюдений за состоянием рынка. В частности, возможны варианты, связанные с наличием различных форм колебательных составляющих, гетерогенности и гетероскедастичности, определяемой изменяющимися во времени характеристиками рассеяния, различными формами асимметрии распределений и т. п. При этом параметры процессов непрерывно изменяются во времени, т. е. система является сугубо нестационарной. Данный вывод подтверждает результаты численных экспериментов, описанных в работах [1, 2].

Приведенные в статье материалы позволяют подтвердить гипотезу о том, что применение традиционных схем обработки данных (метод максимального правдоподобия, МНК, экстраполяционный прогноз и т. п.) в этих ситуациях приводят к существенному снижению точности прогнозирования, и, как следствие, к снижению эффективности управляющих решений. В то же время применение адаптивных алгоритмов оценивания, ориентированных на слежение за текущими характеристиками процесса и подстройку соответствующих процедур прогнозирования, оказывается неприемлемым в силу безынерционности информационных процессов, описывающих динамику рынков капитала. Отсюда непосредственно вытекает вывод о целесообразности применения в подобных задачах робастифицированных алгоритмов обработки данных, обладающих статистической устойчивостью к перечисленным выше нестационарным вариациям вероятностных характеристик контролируемых рыночных процессов. Результаты построе-

ния и применения подобных вычислительных схем предполагается представить в последующих выпусках сборников СПИИРАН.

Литература:

1. *Мусаев А.А., Барласов И.А.* Анализ динамических характеристик состояния рынка ценных бумаг // Тр. СПИИРАН. 2008. Вып. 6. С. 150–160.
2. *Мусаев А.А., Барласов И.А.* Моделирование хаотических процессов на рынках капитала // Тр. СПИИРАН. 2008. Вып. 7. С. 255–264.
3. *Ершов А.А.* Стабильные методы оценки параметров // Автоматики и телемеханика. 1978. № 8. С. 66–100.
4. *Стогов Г.В., Макианов А.В., Мусаев А.А.* Устойчивые методы обработки результатов измерений // Зарубежная радиоэлектроника. 1982. № 9. С. 3–46.
5. *Lang S.* Math Talks for Undergraduates. Springer, 1999. 150 p.
6. *Болч Б., Хуань К.* Многомерные статистические методы для экономики / Под ред. С. А. Айвазяна. М.: Статистика, 1979. 317 с.
7. *Современные методы идентификации систем* / Под ред. П. Эйкхоффа. М.: Мир, 1983. 462 с.
8. *Хьюбер П.* Робастность в статистике / Под ред. И.Г. Журбенко. М.: Мир, 1984. 303 с.
9. *Тьюки Дж.* Анализ результатов наблюдений / Пер. с англ. В. Ф. Писаренко. М.: Мир, 1981. 693 с.

Мусаев Александр Азерович — д-р техн. наук, профессор; ведущий научный сотрудник научно-исследовательской группы информационных технологий в образовании Учреждения Российской академии наук Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН), научный консультант ОАО Специализированная инжиниринговая компания «Севзапмонтажавтоматика». Область научных интересов: анализ данных, управление и прогнозирование в сложных динамических системах, хаотические системы. Число научных публикаций — 180. amusaev@szma.com, www.szma.com; СПИИРАН, 14-я линия В.О., д.39, Санкт-Петербург, 199178, РФ; р.т. +7(812)350-5885, факс +7 (812)350-1113.

Musaev Alexander Azerovich — Dr.Sc. in Appl. Math., professor; leading researcher, Education Information Technology Group, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), expert, public corporation Specialized engineering company «Sevzapmontageautomatica». Research interests: data analysis, complicated dynamic systems prognosis and control, chaos systems. The number of publications — 182. amusaev@szma.com, www.szma.com; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)350-5885, fax +7(812)350-1113.

Рекомендовано СПИИРАН, директор Р.М. Юсупов, чл.-корр. РАН.
Статья поступила в редакцию 16.11.2009.

РЕФЕРАТ

Мусаев А.А. Автоматизированный анализ влияния статистической неопределенности на точностные характеристики оценок параметров инвестиционной ситуации

Практика электронных торгов на рынках капитала показала, что статистические характеристики реальных процессов изменения котировок существенно отличаются от типовых математических моделей, используемых в задачах прогнозирования и поддержки принятия решений.

Возникает задача исследования влияния подобного рода отклонений на эффективность формируемых торговых решений. Очевидно, что эффективность решений, с точки зрения технического анализа динамики капитала, будет в свою очередь определяться точностью оценки текущей ситуации и достоверностью прогноза ее развития. В качестве метрики влияния вариаций статистических характеристик реальных экономических процессов на качество формируемых решений естественно использовать изменение точности оценок, используемых при идентификации и прогнозе состояния характеристических показателей рынков капитала.

Известные способы борьбы с априорной статистической неопределенностью в большинстве случаев основаны на тех или иных «смягчающих» допущениях, позволяющих получить приемлемые решения для некоторых частных задач. В то же время практика обработки данных, полученных в мониторинге состояния рынков капитала, в силу безынерционности наблюдаемых процессов и большого числа факторов влияния со стороны среды взаимодействия, столкнулась с неопределенностью самого общего вида. Иными словами, протекающие процессы по своей природе являются существенно неэргодичными и нестационарными и не допускают каких-либо комфортных для аналитиков предположений относительно природы данных. В связи с этим для анализа влияния указанной неопределенности следует рассматривать множество моделей данных, имитирующих различные типы возможных вариаций статистических характеристик.

Приведенные в статье материалы позволяют подтвердить гипотезу о том, что применение традиционных схем обработки данных в этих ситуациях приводят к существенному снижению точности прогнозирования, и, как следствие, к снижению эффективности управляющих решений. В то же время применение адаптивного подхода, ориентированного на слежение за текущими характеристиками процесса и подстройку соответствующих алгоритмов прогнозирования, оказывается недопустимым в силу безынерционности информационных процессов, описывающих динамику рынков капитала. Отсюда непосредственно вытекает вывод о целесообразности применения в подобных задачах робастифицированных алгоритмов обработки данных, обладающих статистической устойчивостью к перечисленным выше нестационарным вариациям вероятностных характеристик торговых процессов.

SUMMARY

***Musaev A.A.* The numerical analysis of influence statistical uncertainty on trade situation parameters accuracy estimation.**

Practice of the electronic auctions in the capital markets has shown that statistical characteristics of real processes of quotations evolution significantly differ from the typical mathematical models used in tasks of forecasting and support of decision-making.

There is a research problem of statistical deviations influence on efficiency of formed trading decisions. It is obvious that decisions efficiency, from the capital dynamics technical analysis point of view, will be, in turn, defined by accuracy of current situation estimation and its development forecast reliability.

As the metrics of influence of real economic processes statistical characteristics variations on formed decisions quality it is natural to use change of estimations accuracy. This estimation are used at identification and the forecast of characteristic indicators of the capital markets. Usually as an estimation of accuracy of estimation in practice of the statistical analysis values standard deviations or, in the more general case, a full square of an error are used.

Known ways of struggle against aprioristic statistical uncertainty are in most cases based on some "softening" assumptions, allowing to receive comprehensible decisions for some private problems.

At the same time, practice of the data processing, received in the course of the capital markets monitoring, in view of inertialless observable processes and a great number of influence factors from outside interaction environments, has faced uncertainty of the general view.

Differently proceeding processes, by the nature, are essentially non-ergodic and non-stationary and do not suppose any comfortable for analysts assumptions concerning the nature of the data.

Thereupon for the analysis of the specified uncertainty influence it is necessary to consider set of the data simulating models with various types of statistical characteristics possible variations.

The materials resulted in article allow to confirm a hypothesis that application of data processing traditional schemes (least squares method, maximum likelihood, extrapolation forecast, etc.) in these situations lead to essential decrease in accuracy of forecasting, and, as a result, to decrease in operating decisions efficiency.

At the same time application the adaptive approach focused on current characteristics of process tracking and fine tuning of corresponding prognosis algorithms, is inadmissible on account of inertia-free of information processes describing capital markets dynamics. From here there is immediate conclusion about expediency of application in similar tasks data processing robust algorithms, possessing statistical stability to the listed above follows non-stationary variations probability characteristics of trading processes.