

А.А. МУСАЕВ, И.А. БАРЛАСОВ
**ОЦЕНИВАНИЕ СОСТОЯНИЯ ФОНДОВЫХ
РЫНКОВ НА ОСНОВЕ МНОГОМЕРНОЙ
РЕГРЕССИИ НА СКОЛЬЗЯЩЕМ ОКНЕ
НАБЛЮДЕНИЯ**

Мусаев А.А., Барласов И.А. Оценивание состояния фондовых рынков на основе многомерной регрессии на скользящем окне наблюдения.

Аннотация. Рассмотрена задача восстановления индексов фондового рынка на основе алгоритмов многомерного регрессионного анализа. В качестве регрессоров используются валютные инструменты, используемые на электронном рынке «Forex». Параметры памяти, определяющие объем обучающей выборки, задаются размером скользящего окна наблюдения.

Ключевые слова: фондовый рынок, статистическое оценивание, многомерная регрессия.

Musaev A.A., Barlasov I.A. Stock market state estimation on the basis of multidimensional regression on a sliding observation window.

Abstract. The problem of stock market indexes restoration on the basis of algorithms multidimensional regression analysis is considered. As regressors currency tools are used in electronic market "Forex". The memory parameters, defining volume of training sample, are set by the size of a sliding observations window.

Keywords: stock market, statistical estimation, multidimensional regression analysis.

1. Введение. В предыдущих выпусках «Трудов СПИИРАН» [7] было показано, что между характеристическими параметрами фондового и валютного рынка существует неустойчивая корреляционная связь. Числовые характеристики этой связи не являются статическими, они изменяются во времени по величине и направлению. Более того, изменение коэффициентов корреляции носит характер реализации случайного процесса и существенно зависит от размера обучающей выборки, по которой они оцениваются. Учитывая перечисленные особенности задача регрессионного восстановления параметров одного рынка по наблюдениям состояния другого рынка не решается традиционными аналитическими средствами и требует изучения на основе численного анализа. Исследованию данного вопроса посвящена настоящая статья.

2. Ограничения регрессионного анализа. Прежде чем описать результаты практических исследований, кратко рассмотрим общее описание математической технологии регрессионного анализа и соответствующие ей ограничения, накладываемые на статистические свой-

ства исходных рядов наблюдений. Математический аппарат регрессионного анализа подробно описан в фундаментальных монографиях и справочных пособиях [1-4].

В основе технологии регрессионного анализа лежит ряд допущений и предположений, первым из которых является гипотеза о существовании функциональной зависимости $\mu = f(X_i, i=1, \dots, m)$ между условным средним генеральной совокупности μ и величинами некоторой переменной $X_i, i=1, \dots, m$. В исследуемом случае, μ - среднее значения фондового индекса, X_i - значения котировок валютных инструментов. Соответствующая кривая f образует график регрессии генеральной совокупности. Предположим, что для каждой совокупности валютных инструментов осуществляются выборочные наблюдения за фондовыми индексами $Y_i = Y_{i1}, Y_{i2}, \dots, Y_{ij}, \dots, Y_{im}$. Тогда выборочное среднее по каждой совокупности \bar{Y}_i может служить оценкой для μ_i . Задача регрессии в данном случае состоит в установлении формы связи $\bar{Y}_i = f(X_i)$ между индексом фондового рынка и независимыми переменными, в роли которых выступают валютные инструменты.

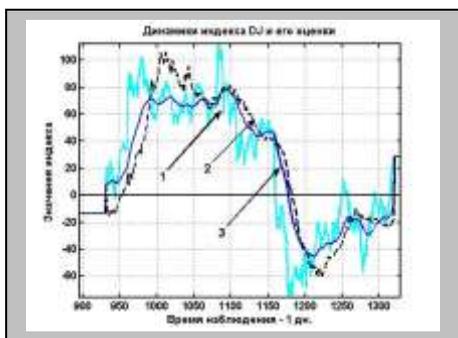


Рис. 1. Котировки индекса DJ $Y(t)$ (линия 1), их регрессионная модель $\mu(t)$ (линия 2) и условное среднее \bar{Y} (линия 3)

Соответствующий пример изменения индекса *Джу-Джонса* (DJ), его регрессионная модель и кривая изменения условного среднего приведены на рис. 1.

В простейшем и наиболее распространенном на практике случае используется линейная (относительно параметров) модель искомой зависимости вида $\mu_i = a_0 + a_1 X_i, i=1, \dots, n$, где a_0, a_1 - неизвестные константы. При этом для случайной выборки наблюдений, не имеющих смещения, математическое ожидание равно среднему, т.е. $E\{Y_{ij}\} = \mu_i, i=1, \dots, n$.

Предположим, что для распределения Y_{ij} выполняется условие

гомоскедастичности, т.е. $\sigma_i^2(Y_{ij}) = \sigma^2 = const$. Тогда $\sigma_i^2(\bar{Y}_i) = \sigma^2/m$.

В этом случае, для стационарного ряда наблюдений и для несмещенного среднего \bar{Y}_i ошибка оценивания $\varepsilon_i = \bar{Y}_i - \mu_i$ представляет собой случайную величину с параметрами $E\{\varepsilon_i\} = 0$, $\sigma^2\{\varepsilon_i\} = \sigma^2/m$. Тогда возмущенное уравнение регрессии имеет вид $\bar{Y}_i = a_0 + a_1 X_{1i} + \varepsilon_i$, $i = 1, \dots, n$. При решении практических задач обычно используются следующие, далеко не всегда соответствующие реальности, допущения:

1. Все X_i , $i = 1, \dots, m$ не являются случайными величинами и, следовательно, для $\forall X_i$ $cov\{\varepsilon_i, X_i\} = 0$.

2. Все Y_i , $i = 1, \dots, m$ - независимо распределенные случайные величины, $cov(Y_i, Y_j) = E\{\varepsilon_i \varepsilon_j\} = 0$ для $\forall i \neq j$.

Многомерная регрессия (*multiple regression*) представляет собой обобщение одномерной линейной регрессии на случай, когда имеется несколько взаимосвязанных предикатов X , определяющих структуру базовой модели. Схема множественной регрессии описывает зависимость выходного (прогнозируемого) параметра от значений входных параметров (предикатов), в роли которых выступают параметры управления и состояния. С точки зрения формирования оценки состояния фондового рынка, применение многомерной регрессии означает возможность последовательного оценивания влияния валютных инструментов на индексы фондовых рынков на различных торговых площадках.

В общем случае, многомерная линейная регрессионная модель имеет вид $\bar{Y}_i = a_0 + a_1 X_{1i} + \dots + a_m X_{mi} + \varepsilon_i$, $i = 1, \dots, n$. При $m = 1$ уравнение регрессии представляет собой гиперплоскость в $(m + 1)$ -ом пространстве. При этом используются те же допущения:

$$cov(Y_i, Y_j) = \begin{cases} 0, & i \neq j, \\ \sigma^2, & i = j. \end{cases} \quad E\{\varepsilon_i\} = 0, \quad cov(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, & i \neq j, \\ \sigma^2, & i = j. \end{cases}$$

$$E\{X_{ki} \varepsilon_i\} = 0, \quad \forall i = 1, \dots, n, \quad k = 1, \dots, m.$$

В дальнейшем для описания многомерной линейной регрессии будем использовать матричную нотацию:

$$a = (a_1, a_2, \dots, a_m)^T, \quad X = \begin{pmatrix} 1 & X_{11} & \dots & X_{m1} \\ 1 & X_{12} & \dots & X_{m2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{1n} & \dots & X_{mn} \end{pmatrix}, \quad \varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)^T.$$

Для получения искоемых оценок параметров регрессии \hat{a} , в соответствии с *методом наименьших квадратов* (МНК) [2, 4], будем минимизировать сумму квадратов ошибок

$$\sum_{i=1}^n \varepsilon^2 = \varepsilon^T \varepsilon = (Y - Xa)^T (Y - Xa).$$

Минимизация последнего выражения приводит к системе нормальных уравнений, решение которых, в свою очередь, позволяет определить хорошо известное соотношение для оценки параметров регрессии по МНК $\hat{a} = (X^T X)^{-1} X^T Y$.

При выполнении всех вышеуказанных предположений оценка по МНК является наилучшей по критерию минимума дисперсии и представляет собой линейную функцию от ошибок модели наблюдений v : $e = Mv$, где $M = I_n - X(X^T X)^{-1} X^T$ - симметричная идемпотентная матрица ($M^T M = M$) [2], I_n - единичная матрица порядка n . При

$$\text{этом } \varepsilon^T \varepsilon = \varepsilon^T M^T M \varepsilon = \varepsilon^T M \varepsilon = \sum_{i=1}^n M_{ii} v_i^2 + \sum_{i \neq i'} M_{ii'} v_i v_{i'}.$$

3. Особенности регрессионного анализа взаимных связей рыночных показателей. Для проведения численных исследований эффективности применения регрессионного анализа в задачах оценки состояния фондового рынка в настоящей работе будет использоваться скользящее окно наблюдения, позволяющее снизить влияние нестационарности на качество обработки данных.

Организация скользящего окна наблюдения и отвечающая ему схема структуризации данных представлена в статье [5]. При этом информационная платформа анализа формируется в виде двумерной таблицы, в левой части которой представлены данные наблюдений за регрессорами, в роли которых, в данном случае, выступают валютные инструменты рынка Forex. Заметим, что для задач прогнозирования в левую часть таблицы могут входить предшествующие параметры фондового рынка.

В правой части таблицы представлены параметры, описывающие состояние фондового рынка (например, индексы или котировки акций

активов, используемых при формировании инвестиционного портфеля). При этом для задач прогнозирования осуществляется сдвиг ретроспективных данных, соответствующих состоянию фондового рынка, относительно данных, описывающих состояние среды взаимодействия на число шагов $k(\tau)$, отвечающих интервалу планируемого прогноза τ . В задачах анализа состояния фондового рынка такой сдвиг делать не нужно.

Совокупность ретроспективных данных за промежуток времени $(k - L, k - 1)$, где k - отсчет времени, отвечающий текущему состоянию инвестиционной среды и текущему (для задач анализа) или сдвинутому на $k(\tau)$ шагов (для задач прогноза) состоянию фондового рынка, образует обучающую выборку наблюдений, используемую для построения математической модели инвестиционной ситуации.

Наличие скользящего окна наблюдения позволяет оперативно перестраивать модель как в части ее структуры, так и с позиции параметрической адаптации. Данный подход предоставляет возможность последовательно «отслеживать» динамику процесса, причем, в случае успешного восстановления тренда, оставшиеся невязки аппроксимации образуют процесс, весьма близкий к стационарному гауссовскому шуму.

4. Численные исследования. Рассмотрим задачу численного исследования регрессионного оценивания фондовых индексов по совокупности регрессоров, образованных группой валютных инструментов $mnR = \{EURUSD, EURJPY, EURCAD, USDJPY, GBPCHF, GBPJPY\}$.

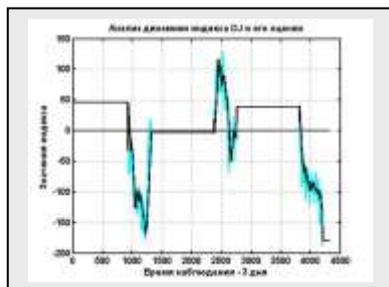


Рис. 2. Динамика изменения индекса DJ и его регрессионной оценки

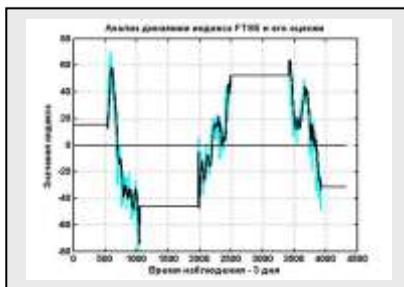


Рис. 3. Динамика изменения индекса FTSE и его регрессионной оценки

На рис. 2 и 3 представлены графики изменения фондовых индексов *FTSE* и *DJ* на интервале наблюдения 3 дня. Количество одноимен-

нутных шагов в окне обучения $n_{Window}=48$.

С учетом сохранения участков, отвечающих часам работы фондовых бирж, результирующее время наблюдения существенно (примерно на треть) уменьшится.

Из приведенных графиков можно видеть, что, несмотря на относительно слабые корреляционные связи между валютными инструментами и фондовыми индексами, схема со скользящим окном наблюдения, в принципе, позволяет воспроизводить динамику изменения состояния фондовых рынков.

В более общем случае можно использовать схему структурной адаптации данной процедуры, основанную на последовательном корреляционном анализе на скользящем окне наблюдения и последовательном выборе квазиоптимального набора регрессоров из совокупности используемых валютных инструментов. Соответствующая блок-схема программы оценивания и прогнозирования на основе модели множественной регрессии приведена на рис. 4.

Два верхних функциональных блока в основном цикле программы предназначены для анализа корреляционных связей индексов фондового рынка и валютных инструментов, используемых в качестве регрессоров. Однако, как видно из динамических свойств коэффициентов корреляции, приведенных в [7], применение подобных схем самоорганизации регрессионной модели связано с существенным сложностями.

В частности, в рамках спекулятивных краткосрочных стратегий, ориентированных на работу в течение одного дня работы биржи, оценка коэффициента корреляции осуществляется на коротком интервале времени и не обладает статистической устойчивостью. Для долгосрочных стратегий эти же оценки оказываются излишне инерционными и не отражают реальные связи в переходных процессах.

Рассмотрим вопрос о качестве восстановления значений фондовых индексов по валютным инструментам. С этой целью изучим в качестве примера изменение индекса DJ и его регрессионных оценок в течение одной торговой сессии. Соответствующие графики для скользящих окон размером, соответственно, в 48 (сплошная черная линия) и 96 минут (пунктирная линия) наблюдения приведены на рис. 5. Из представленных графиков видно, что относительно короткие окна наблюдений, используемых для формирования оценок по МНК, позволяют, даже при небольших значениях корреляций между регрессорами и оцениваемыми величинами, успешно восстанавливать искомые оценки.

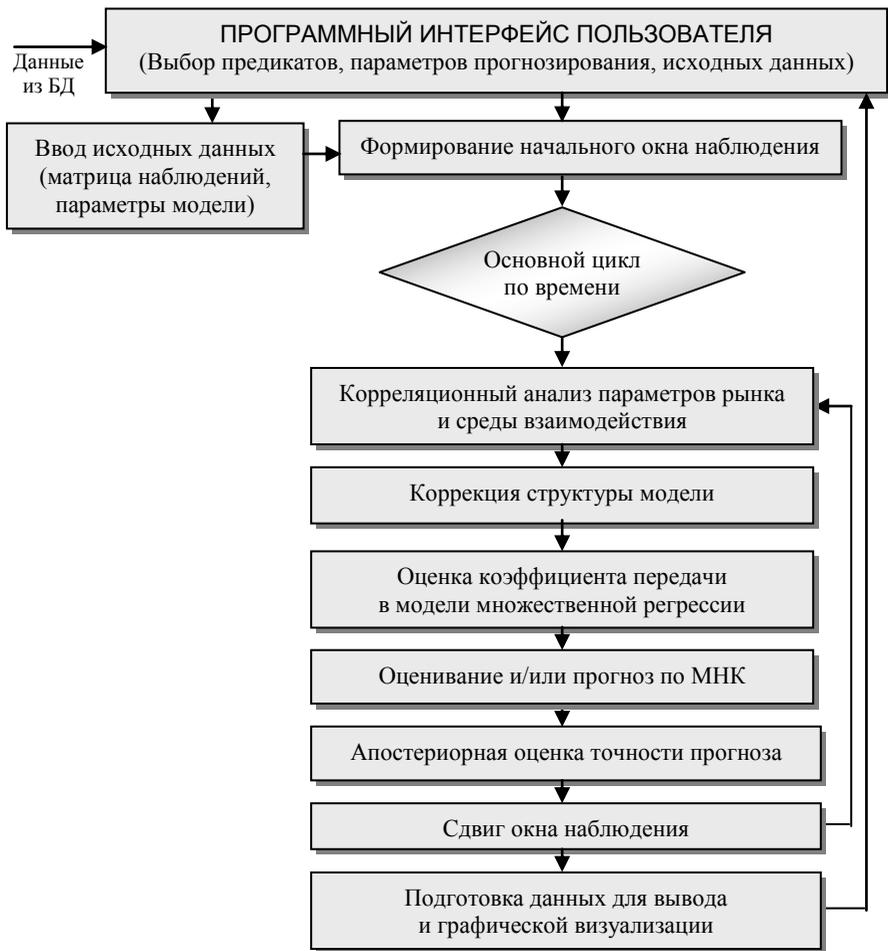


Рис. 4. Блок-схема программы оценивания и прогнозирования на основе модели множественной регрессии

Следует заметить, что рост размера скользящего окна, на котором формируется коэффициент передачи по МНК, заметно повышает инерционность оценки. Особенно заметен эффект запаздывания на

высоко динамичных участках траектории с ярко выраженным трендом. Отсюда возникает гипотеза о возможности применения разности между двумя синхронными сглаженными регрессионными оценками с различными скользящими окнами для автоматизированной идентификации участков траектории с различной динамикой.

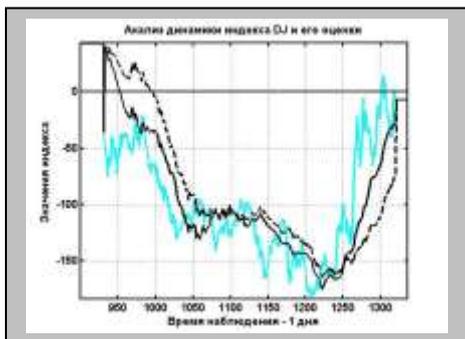


Рис. 5. Изменения индекса DJ и его регрессионных оценок со скользящими окнами в 48 и 96 минут в течение торговой сессии.

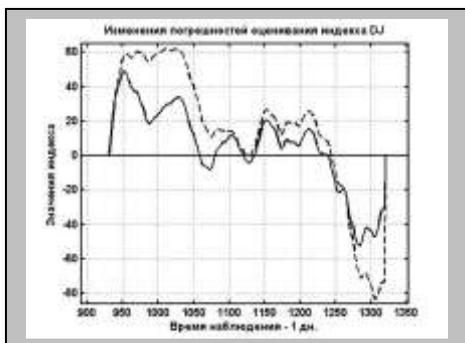


Рис. 6. Сглаженные изменения погрешности оценивания индекса DJ со скользящими окнами в 48 и 96 минут в течение торговой сессии.

онные связи между отдельными показателями фондового и валютного рынков схема регрессионного оценивания на скользящем окне наблюдения позволяет успешно восстанавливать значения фондовых индексов по значениям набора валютных инструментов;

Не менее важным вопросом является исследование динамики изменения погрешности регрессионной оценки и ее взаимосвязи с изменениями самих фондовых индексов.

В качестве примера на рис. 6 представлены графики сглаженных изменений погрешностей оценивания индекса DJ для тех же скользящих окон в течение выбранной суточной торговой сессии. Погрешности оценивания представляли собой разницы между реальными котировками индекса и его регрессионными оценками. В качестве оператора сглаживания использовался экспоненциальный фильтр с коэффициентом передачи $\alpha = 0.05$.

Заключение. Полученные в подразделе материалы позволяют сделать следующие выводы:

1. Несмотря на относительно слабые корреляци-

2. Качество восстановления и прогноза индикаторов фондового рынка следует исследовать на основе изучения не отдельной функции ошибки регрессионной оценки, а их совокупности, образованной параметрическими и структурными вариациями исходных посылок;

3. Визуальный анализ представленных графиков дает возможность выдвинуть ряд предположений о возможных путях повышения эффективности традиционной схемы регрессионного анализа в условиях нестационарных возмущений:

гипотеза о возможности применения разности между погрешностями оценивания, формируемыми для различных окон наблюдения и различными степенями сглаживания в качестве индикаторов трендов динамики фондовых инструментов;

гипотеза о целесообразности использования адаптивных схем оценивания путем изменения размера скользящего окна наблюдения и перечня регрессоров на основе тестов на предшествующем окне обучения;

гипотеза о возможности применения технологии робастификации формируемых оценок на основе методов, рассмотренных, например, в [6].

Литература

1. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрии: Учебник для вузов / М.: ЮНИТИ, 1998. 1022с.
2. Болч Б., Хуань К. Многомерные статистические методы для экономики / Пер. с англ. под ред. С. А. Айвазяна М.: Статистика, 1979. 317с.
3. Демиденко Е. З. Линейная и нелинейная регрессии / М.: Финансы и статистика, 1981. 302с.
4. Линник Ю.В. Метод наименьших квадратов и основы теории обработки наблюдений / М.: Физматгиз, 1958. 349с.
5. Мусаев А.А., Барласов И.А. Анализ динамических характеристик состояния рынка ценных бумаг // Труды СПИИРАН. Вып. 6. 2008. С. 150–160.
6. Мусаев А.А. Методы построения робастифицированных систем анализа торговых ситуаций // Труды СПИИРАН. Вып. 14. 2010. С. 187–215.
7. Мусаев А.А. Корреляционный анализ процессов изменения состояния фондовых и валютных рынков // Труды СПИИРАН. Вып. 18. 2011. С. 5–18.

Мусаев Александр Азерович — д.т.н., профессор; ведущий научный сотрудник научно-исследовательской группы информационных технологий в образовании Учреждения Российской академии наук Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН), научный консультант ОАО Специализированная инженеринговая компания «Севзапмонтажавтоматика». Область научных интересов: анализ данных, управление и прогнозирование в сложных динамических системах, стохастические хаотические системы. Число научных публикаций — 176. amusaev@szma.com, www.szma.com; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)350-5885, факс +7 (812)350-1113.

Musaev Alexander Azerovich — Dr. 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, stochastic chaos systems. The number of publications — 176. amusaev@szma.com, www.szma.com; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)350-5885, fax +7(812)350-1113.

Барласов Илья Антонович — менеджер группы корпоративного бизнеса ЗАО «Райффайзенбанк». Область научных интересов: анализ данных, управление и прогнозирование в экономике и финансах. Число научных публикаций — 4. barlasov@yandex.ru, www.raiffeisen.ru; Райффайзенбанк, отд. «Сенная площадь», ул. Ефимова, д. 4а, г. Санкт-Петербург, 198150, РФ; р.т. +7(812)334-4343, факс +7 (812)343-4145.

Barlasov Ilya Antonovich — corporative business group manager of closed corporation «Raiffeisenbank». Research interests: data analysis, prognosis and management in economics and finance. The number of publications — 4. barlasov@yandex.ru, www.raiffeisen.ru; Raiffeisenbank, dep. «Hay square», 4a, Efimova street, St. Petersburg, 198150, Russia; office phone +7(812)334-4343, fax +7(812)334-4145.

Рекомендовано ИГИТО СПИИРАН, рук. ктн, доц. А.В. Тишков.
Статья поступила в редакцию 08.11.2011.

РЕФЕРАТ

Мусаев А.А., Барласов И.А. **Оценивание состояния фондовых рынков на основе многомерной регрессии на скользящем окне наблюдения.**

Котировки рыночных активов представляют собой результат групповой ментальной оценки рыночной ситуации и образуют безынерционный стохастический процесс с элементами хаотической динамики. Указанная специфика динамики котировок существенно ограничивает возможность построения эффективного прогноза развития рыночной ситуации. Отсюда возникает необходимость поиска новых нетрадиционных вычислительных схем, основанных на комплексном использовании взаимных связей между процессами, протекающими на различных рынках капитала. В частности, в настоящей статье рассмотрена возможность оценивания состояния фондового рынка по наблюдениям за динамикой изменения финансовых инструментов на валютном рынке Fogex.

Предыдущие исследования данного вопроса показали наличие неустойчивых корреляционных связей между отдельными показателями фондового и валютного рынков. Тем не менее, схема регрессионного оценивания на скользящем окне наблюдения позволяет успешно восстанавливать значения фондовых индексов по значениям набора валютных инструментов. При этом качество восстановления и прогноза индикаторов фондового рынка существенно зависит от размера окна наблюдения и совокупности регрессоров, используемых процедурой оценивания.

Небольшой размер окна наблюдения приводит к сильным безынерционным колебаниям оценки, в результате которых соответствующие управляющие решения будут содержать большое число статистических ошибок второго рода. Увеличение размера окна повышает устойчивость оценивания, но вместе с этим увеличивается инерционность формируемых оценок и, как следствие, число статистических ошибок первого рода. При этом традиционные схемы параметрической адаптации размера скользящего окна, как и другие варианты построения адаптивных решений, не приводят к положительному результату в силу хаотической структуры наблюдаемых процессов.

Интересным выводом из представленных материалов является гипотеза о возможности применения разности между погрешностями оценивания, формируемыми для различных окон наблюдения в качестве индикаторов изменения трендов динамики фондовых инструментов.

Наиболее эффективным средством развития данной задачи, по-видимому, является возможность применения технологии робастного оценивания, обладающей повышенной устойчивостью к вариациям вероятностных характеристик исходных данных относительно опорной вероятностной модели, используемой для описания наблюдаемых процессов. Основная проблема состоит в поиске компромисса, обеспечивающего относительно устойчивое решение при сохранении чувствительности оценок к значимым изменениям трендов.

SUMMARY

Barlasov I.A., Musaev A.A. **Stock market state estimation on the basis of multidimensional regression on a sliding observation window.**

Quotations of market assets are represented as result of a group mental estimations of a market situation and form uninertia stochastic process with elements of chaotic dynamics. This specificity of quotations dynamics essentially limits possibility of a market situation development effective forecast construction. That's why there is a necessity of search of new nonconventional computing schemes based on complex use of an interconnection between processes proceeding at the various markets of the capital. In particular, in the present paper possibility of estimation of a stock market state supervision based on change of dynamics of financial tools in currency market Forex is considered.

The previous researches of this problem have shown presence of unstable correlation between separate indicators of the share and currency markets. Nevertheless, the regression estimations scheme on sliding watch window allows to restore successfully values of share indexes on values of a currency tools set. Thus restoration and the forecast of indicators of stock market quality essentially depend on the size of a watch window and set of the regressors used by procedure of estimation.

The small size of a watch window leads to strong uninertia fluctuations of estimation which result that corresponding operating decisions will contain a great number of statistical errors of the second sort. The increase in the size of a window raises stability of estimation, but together with it the lag effect of formed estimations and, as consequence, number of statistical errors of the first sort increase. Thus traditional schemes of parametrical adaptation of a sliding window size, as well as other variants of construction of adaptive decisions, don't lead to positive result owing to chaotic structure of observable processes.

Interesting conclusion from the presented materials is the hypothesis about possibility of application of a difference between errors of the estimation, formed for various watch windows as indicators of change of trends of dynamics of share tools.

The most effective mean of the given problem development, apparently, is a possibility of application of robust estimations technology, possessing raised stability to variations of likelihood characteristics of the initial data concerning the basic likelihood model used for the description of observable processes. The basic problem consists in search of the compromise providing rather steady decision for preservation of sensitivity of estimations of significant trends changes.