

АНАЛИЗ СТАБИЛЬНОСТИ ПРОГНОЗИРОВАНИЯ КОРОТКИХ ВРЕМЕННЫХ РЯДОВ ЦЕН НА ОСНОВЕ РАСПОЗНАВАНИЯ ПАТТЕРНОВ

В.Г. Саркисов¹, Т.В. Чихерева²

ФГБОУ ВПО «Самарский государственный технический университет»
443100, г. Самара, ул. Молодогвардейская, 244
E-mail: vigen.sarkisov@mail.ru

В качестве метода прогнозирования коротких временных рядов цен финансовых инструментов выбран метод аналогов. Проведен анализ изменения качества прогноза в зависимости от длины рассматриваемых паттернов. Исследована стабильность качества инвестиционных решений при оптимизации системы прогнозирования на различных временных отрезках. Рассмотрено несколько критериев оценки работы системы прогнозирования, учитывающих правильность прогноза.

Ключевые слова: временной ряд цен, паттерн, хэш-код, качество прогноза, стабильность, оптимизация.

Система прогнозирования изменений цен финансовых инструментов является неотъемлемой частью системы управления инвестиционным портфелем. Существенная проблема систем прогнозирования временных рядов состоит в нестабильности их работы.

Будучи однажды оптимизированной на исторических данных, система постепенно (а иногда и сразу) утрачивает свои прогнозирующие способности. Обычно это свойственно системам прогнозирования с большим количеством оптимизируемых параметров. Данное явление носит название *curve fitting* («подгонка под кривую»), или *overfitting* («переподгонка»), и состоит в том, что система с высочайшей точностью прямо или косвенно аппроксимирует используемый при оптимизации временной ряд исторических данных, а при небольшом изменении параметров ряда дает существенную погрешность прогноза.

В качестве решения подобной проблемы обычно рекомендуют следующее.

1. Производить периодическую оптимизацию параметров системы прогнозирования. При этом предполагается, что прогнозирующая способность системы будет ухудшаться постепенно, что верно далеко не всегда. Повторная оптимизация вовсе не гарантирует, что система не потеряет качество прогноза практически мгновенно.

2. Зафиксировать некоторые параметры системы и не производить их оптимизацию, снизив тем самым количество оптимизируемых параметров [1]. Слабым местом этого подхода является необходимость выбора фиксированных параметров, то есть их косвенная, скрытая оптимизация.

3. Вообще отказаться от оптимизации, выбрав из нескольких систем с фиксированными параметрами систему с наилучшим качеством прогноза на исторических данных. Этот способ, как и предыдущие, не исключает возникновения подгонки, лишь скрывая ее за выбором наиболее эффективной системы прогнозирования.

Очевидно, что ни один из этих подходов не решает проблему ухудшения качества прогноза при изменении характеристик временного ряда цен.

Рассматривая проблему в наиболее общем виде, приходится признать, что разрешить ее до конца принципиально невозможно. Это связано с тем, что всегда можно искусственно синтезировать временной ряд цен (возможно, не реализуемый на практике), который будет вынуждать рассматриваемую систему прогнозирования постоянно ошибаться, меняя на каждом следующем временном отрезке характеристики, важные именно для данной системы.

В настоящей работе акцент сделан на анализе стабильности системы прогнозирования, то есть ее способности выдавать приемлемый результат при различных характеристиках временных рядов.

¹ *Виген Геннадьевич Саркисов* (к.т.н.), доцент кафедры «Высшая математика и прикладная информатика».

² *Татьяна Владимировна Чихерева*, студентка направления «Менеджмент», профиль «Экономика и управление на предприятиях топливно-энергетического комплекса».

Метод тестирования системы прогнозирования на стабильность. Классический подход к выбору системы прогнозирования сводится к оптимизации параметров системы на одной части исторических данных и проверке работоспособности системы с оптимизированными параметрами на другой части. Лучшей при этом считают систему, показавшую наилучшее качество прогноза на второй части исторических данных. Такой подход может привести к выбору очень нестабильной системы, которая случайно показала высокую точность прогноза на второй части данных. В силу случайности выборки параметры, полученные в ходе оптимизации, также являются случайными. В итоге выбор системы сводится к косвенной оптимизации на некотором отрезке временного ряда.

В целях уменьшения вероятности подобной неприятной ситуации авторы предлагают увеличить количество отрезков времени и провести следующую процедуру оценки стабильности (аналогичная процедура для систем управления портфелем описана в [2]).

1. Весь доступный временной ряд исторических данных разбивается на равные промежутки. Далее в качестве примера будут рассмотрены 8 промежутков, по одному году каждый (2006–2013 гг.), для фьючерса на индекс РТС.

2. На каждом из промежутков система оптимизируется для достижения наилучшего качества прогноза. В результате получим 8 наборов (векторов) параметров.

3. Каждый из наборов параметров применяется на каждом из промежутков. Для каждого сочетания оценивается качество прогноза. Полученная матрица оценок позволяет количественно описать стабильность системы прогнозирования.

Сформулируем математическую постановку задачи максимизации стабильности качества прогноза. Пусть система имеет N оптимизируемых параметров, образующих вектор $k = (k_1, \dots, k_N)$, а временной ряд цены финансового инструмента разбит на T промежутков (τ_1, \dots, τ_T). Качество прогноза на каждом промежутке может быть оценено показателем качества q . Для удобства дальнейших рассуждений будем рассматривать q как функцию параметров и интервала времени $q(k, \tau_i)$, не вдаваясь пока в подробности формирования прогноза и способа оценки его точности.

Оптимальный набор параметров для t -го промежутка доставляет максимум показателю качества прогноза:

$$k^{*t} = \arg \max_k q(k, \tau_t) \quad (1)$$

При использовании на i -ом промежутке параметров, полученных при оптимизации на t -ом промежутке, показатель качества примет значение $q(k^{*t}, \tau_i)$. Проведя все описанные выше эксперименты, мы получим матрицу Q значений показателя q :

$$Q = \begin{pmatrix} q(k^{*1}, \tau_1) & q(k^{*1}, \tau_2) & \dots & q(k^{*1}, \tau_T) \\ q(k^{*2}, \tau_1) & q(k^{*2}, \tau_2) & \dots & q(k^{*2}, \tau_T) \\ \boxed{\times} & \boxed{\times} & \boxed{\times} & \boxed{\times} \\ q(k^{*T}, \tau_1) & q(k^{*T}, \tau_2) & \dots & q(k^{*T}, \tau_T) \end{pmatrix} \quad (2)$$

В главной диагонали матрицы Q располагаются значения показателя качества, полученные в ходе оптимизации. Сравнивая между собой значения в некотором столбце, можно определить, насколько влияет изменение настроек на работу системы на одном промежутке. Сравнивая значения в строке, можно оценить разброс качества на разных промежутках при одинаковых настройках.

Хорошая система прогнозирования должна демонстрировать приемлемое качество прогноза во всех случаях. Чем выше нижняя граница полученного множества значений показателя качества, тем лучше система. Также целесообразно рассматривать и другие характеристики положения – квантили низких порядков и среднее значение показателя качества, вычисленные без учета промежутка, на котором производилась оптимизация. Практический интерес представляют квантили порядков от 0 (нижняя граница) до 0,5 (медиана). Квантили более высоких порядков не представляют практического интереса, так как описывают лишь случаи наиболее точных прогнозов. В практической части данной работы будут использоваться квантили порядков 0,1, 0,2, 0,3 и 0,5 (медиана).

Оценки качества прогноза временного ряда цен. Конечной целью инвестирования на финансовом рынке (если исключить противозаконный вариант манипулирования рынком) является получение прибыли. Именно с точки зрения возможности получения прибыли и стоит рассматривать систему прогнозирования. Поэтому хорошая система прогнозирования должна давать инвестору возможность принять правильное инвестиционное решение. Учитывая, что множество возможных решений содержит лишь решения «купить», «продать» и «оставаться вне рынка», инвестору важно знать направление будущего изменения цены, но не размер этого изменения [3]. Таким образом, качество работы системы характеризуется процентом правильных прогнозов направления изменения цены.

Для более глубокого анализа интересно рассмотреть распределение потенциальных прибылей и убытков от применения системы прогнозирования. Подобное исследование предполагает наличие подробной информации о системе принятия решений о покупке и продаже финансового инструмента. В данной работе в демонстрационных целях используется простейшая система, предполагающая открытие позиции (покупку или продажу) в соответствии с направлением изменения цены, указанным системой прогнозирования, а закрытие позиции – по цене закрытия торгов в прогнозируемый день.

Рассматриваемая система прогнозирования. В качестве основы для построения системы прогноза была использована система, рассмотренная в [4]. Основным отличием является формирование не двоичного, а троичного $(-1, 0, +1)$ хеш-кода. Алгоритм работы исследуемой системы следующий.

1. Вычисляется относительное (в процентах) приращение цены инвестиционного инструмента за каждый день.

2. Если приращение меньше некоторого значения a , то данному дню присваивается код -1 , а если больше значения b , то $+1$ ($a \leq b$). В остальных случаях присваивается код 0 . В результате из последовательности значений относительного приращения цены формируется последовательность из $-1, 0$ и $+1$. Изменение значений a и b позволяет настраивать систему для получения оптимальных характеристик.

3. Для того дня, который будет прогнозироваться, формируется хеш-код – троичное число, соответствующее последовательности заданного числа кодов предыдущих дней. Далее будут рассмотрено применение хеш-кодов из 1, 2 и 3 кодов дней.

4. На исторических данных выбираются дни, имеющие такой же хеш-код, как у рассматриваемого дня. Строится распределение приращений цен за эти дни, которое и принимается в качестве прогноза распределения приращения цены на прогнозируемый день.

С увеличением количества дней в хеш-коде возрастает количество возможных значений хеш-кода. Для одного дня таких вариантов всего 3, для двух дней – $3^2=9$, а для трех дней уже $3^3=27$ (рис. 1).

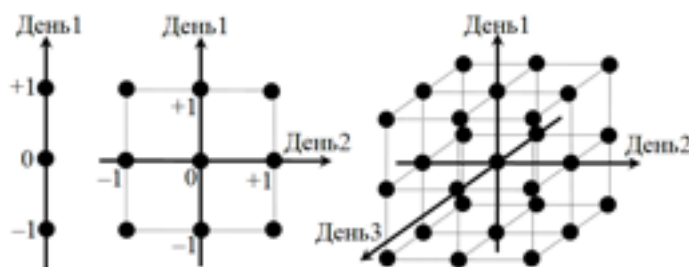


Рис. 1. Варианты хеш-кода (паттерны) при разном числе дней

Один год содержит примерно 250 рабочих дней. При формировании хеш-кода из трех дней на каждый вариант будет приходиться в среднем $250/27 \approx 9,3$ дня в год, а при хеш-коде из четырех дней – $250/81 \approx 3,1$, что делает невозможным получение корректных статистических оценок. Можно попытаться обойти подобную проблему, установив такие значения параметров, чтобы некоторые варианты хеш-кода возникали намного чаще остальных. В настоящей работе данный аспект отдельно не рассматривается, предпочтение отдано ограничению числа дней в хеш-коде.

Описание и результаты эксперимента. В качестве исторических данных был выбран временной ряд цен наиболее ликвидного инструмента российского финансового рынка – фьючерса на индекс Российской торговой системы [5]. Рассмотрен период времени с 1 января 2006 года по 31 декабря 2013 года (8 лет). Шаг дискретизации – 1 день. Весь временной ряд разбит на 8 интервалов по 1 году. График изменения цены фьючерса на индекс РТС в пунктах (1 пункт в соответствии со спецификацией контракта [6] равен 0,02 доллара США) представлен на рис. 2.



Рис. 2. График изменения цены фьючерса на индекс РТС (в пунктах)

На каждом из промежутков была проведена оптимизация параметров a и b , направленная на увеличение процента правильно предсказанных дней. На рис. 3 для иллюстрации приведено превышение количества правильно предсказанных дней над неправильно предсказанными для всего рассматриваемого периода 2006–2013 гг. при различных значениях a и b (всего 1992 дня).

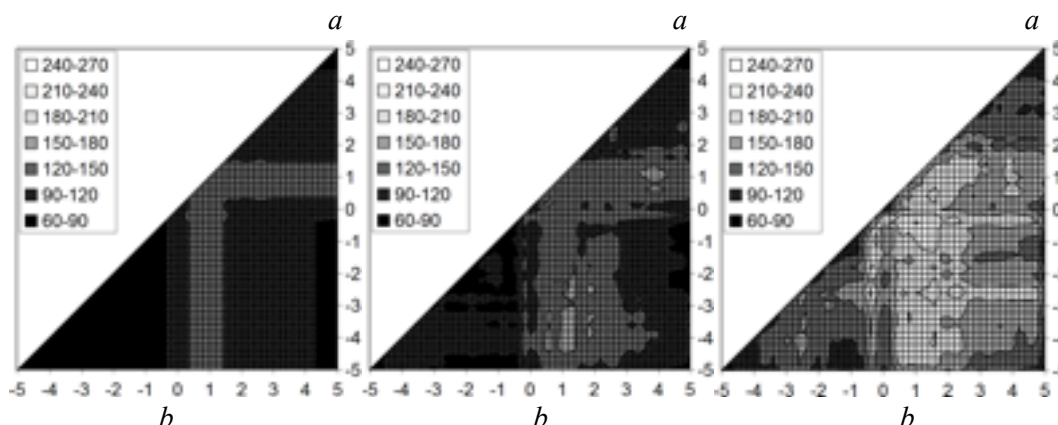


Рис. 3. Превышение правильно предсказанных дней над неправильно предсказанными при различных значениях параметров на периоде 2006–2013 гг. для хеш-кодов из 1, 2 и 3 дней

Очевидно, что увеличение количества дней, по которым строится хеш-код, дает больше возможностей для оптимизации, одновременно усиливая эффект подгонки под кривую. Результаты оптимизации даны в табл. 1.

Таблица 1

Оптимальные значения параметров системы прогнозирования направления

Дней в хеш-коде	Параметр	2006	2007	2008	2009	2010	2011	2012	2013
1	a , %	+1,8	+0,2	-4,2	+2,6	-3,0	+0,6	-2,0	-2,0
	b , %	+3,2	+2,0	-3,2	+2,6	+3,6	+2,6	-0,2	+0,8

2	$a, \%$	-2,0	-0,2	-2,4	-0,6	+0,2	-0,2	-1,0	-1,8
	$b, \%$	+2,2	+1,4	-0,8	+0,6	+2,0	+0,2	+1,4	+0,4
3	$a, \%$	-0,6	-1,0	-1,6	-1,4	-1,4	-1,6	-0,4	-0,4
	$b, \%$	+1,2	+1,0	-0,4	+0,2	+1,2	+1,4	+1,0	+0,8

Аналогичные расчеты проведены при использовании в качестве критерия доходности системы с открытием позиции (покупкой или продажей) в соответствии с прогнозом и закрытием позиции в конце дня (рис. 4, табл. 2).

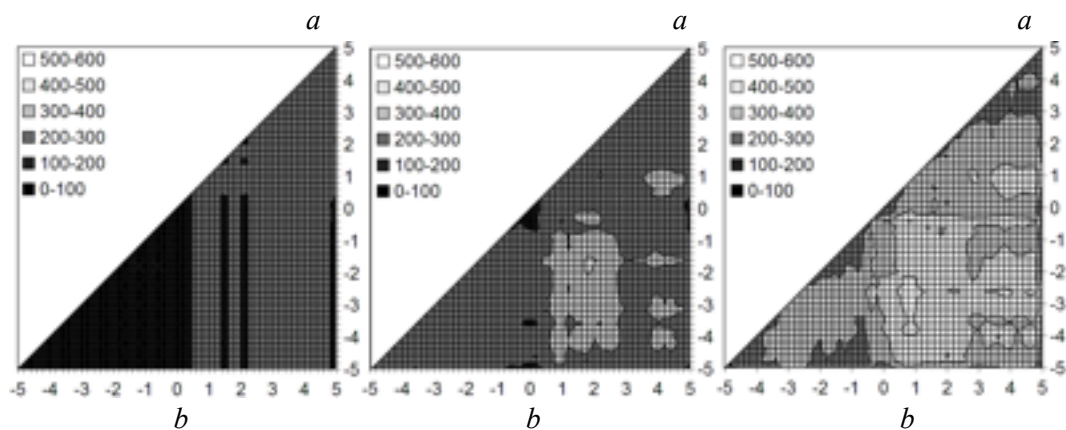


Рис. 4. Доходности, %, системы при различных значениях параметров на периоде 2006–2013 гг. для хеш-кодов из 1, 2 и 3 дней

Оптимальные значения параметров системы максимизации доходности

Дней в хеш-коде	Параметр	2006	2007	2008	2009	2010	2011	2012	2013
1	a, %	+2,6	+1,0	-3,2	+2,4	-0,4	+0,4	-2,0	-1
	b, %	+3,4	+2,2	-3,2	+2,6	+0,2	+0,4	-0,2	0
2	a, %	-0,2	+1,0	-3,2	-1,0	+0,6	-2,4	-0,8	-1,2
	b, %	+2,2	+2,6	-0,8	+0,2	+1,8	+0,4	+1,4	+0,4
3	a, %	-0,6	-0,8	-1,6	-0,8	-1,4	-1,6	-0,8	-0,4
	b, %	+1,4	+1,0	+0,2	+0,2	-0,2	+3,6	+1,2	+0,8

После оптимизации были вычислены средние значения и квантили процента верных прогнозов при всевозможных сочетаниях оптимизированных параметров (из табл. 1) и временных интервалов. В качестве примера приведены матрицы превышения правильно предсказанных дней над неправильно предсказанными (3) и доходности (4) для хеш-кода по 3 дням.

$$Q_{\text{прогн}} = \begin{pmatrix} 89 & -11 & -1 & 17 & -4 & 9 & -43 & 26 \\ -27 & 73 & -31 & 19 & 24 & 5 & 45 & -18 \\ -9 & -1 & 87 & 4 & 3 & 7 & -11 & 26 \\ -1 & -7 & -13 & 90 & 23 & 17 & 9 & -16 \\ 7 & 9 & 1 & 19 & 86 & -11 & 1 & 22 \\ 7 & -9 & -21 & 37 & -18 & 73 & -1 & 8 \\ -49 & 11 & 1 & -9 & 8 & -7 & 93 & -34 \\ 17 & 31 & -29 & 19 & 6 & -5 & -19 & 80 \end{pmatrix}, \quad (3)$$

При расчетах доходностей (4) в целях упрощения вычислений использовались простые проценты. Это привело к деформации шкалы доходностей, переоценке убытков (например к появлению нереально низкой доходности -112 % годовых) и недооценке прибылей. При практическом применении предлагаемого метода для построения систем управления реальными инвестиционными портфелями в расчетах необходимо использовать сложные проценты.

$$Q_{\text{доход}} = \begin{pmatrix} 185 & 50 & -26 & 41 & 10 & -16 & -22 & 3 \\ -22 & 124 & -112 & 62 & 27 & 42 & 53 & -1 \\ -74 & -30 & 245 & -98 & 8 & -7 & -50 & 27 \\ 25 & 21 & -70 & 272 & -1 & 20 & 29 & -37 \\ 61 & -12 & -37 & 33 & 129 & -42 & -10 & -16 \\ -32 & 5 & 43 & 19 & -27 & 182 & 1 & 5 \\ -51 & 22 & -82 & 39 & 5 & 73 & 151 & -1 \\ 30 & 42 & -3 & -5 & -5 & -29 & -11 & 97 \end{pmatrix}. \quad (4)$$

В матрицах (3) и (4) в каждой строке максимальное значение достигается на главной диагонали. Это явление связано с тем, что значения главной диагонали получены при параметрах, оптимальных для данного промежутка времени. Значения, полученные с теми же параметрами на других промежутках времени, заметно хуже.

В табл. 3 приведены показатели, найденные по (3) и (4) с учетом значений на главных диагоналях и характеризующие стабильность исследованных систем.

Таблица 3

Статистические характеристики стабильности систем

Хеш-код	Прогнозирование направления			Доходность		
	1 день	2 дня	3 дня	1 день	2 дня	3 дня
Квантиль 0,1	-13	-14	-20	-29	-40	-48
Квантиль 0,2	-5	-6	-11	-18	-14	-28
Квантиль 0,3	-2	-3	-7	-7	-9	-13
Медиана	+8	+5	+6	+8	+14	+5
Среднее значение	+8	+9	+11	+16	+18	+20
СКО	+18	+22	+33	50	51	72

Более качественной системе соответствуют более высокие значения (выделены жирным шрифтом в табл. 3) всех показателей, кроме СКО, характеризующего разброс значений. По стабильности прогнозирования направления доминирующей является система с хеш-кодом, построенным по 1 дню. Система с трехдневным хеш-кодом имеет преимущество только по среднему значению.

Исключение из расчетов элементов главных диагоналей Q привело к следующему изменению значений показателей (табл. 4).

Таблица 4

Статистические характеристики стабильности систем без учета элементов главных диагоналей

Хеш-код	Прогнозирование направления			Доходность		
	1 день	2 дня	3 дня	1 день	2 дня	3 дня
Квантиль 0,1	-13	-16	-24	-38	-43	-50
Квантиль 0,2	-8	-7	-13	-22	-19	-30
Квантиль 0,3	-3	-3	-9	-11	-10	-19
Медиана	+3	+3	+1	+4	+8	-1
Среднее значение	+6	+3	+1	+6	+6	-2
СКО	17	15	19	42	36	40
Среднее значение элементов главной диагонали Q	26	54	84	87	104	173

Система с трехдневным хеш-кодом не является доминирующей ни по одному из показателей (см. табл. 4), несмотря на наилучшие значения на главной диагонали. Наилучшей по доходности системой оказалась система с двухдневным хеш-кодом, а по прогнозированию направления – с однодневным.

Выводы:

1. Использование на одном временном промежутке параметров, оптимизированных на другом промежутке, часто дает неприемлемые результаты управления портфелем. Это говорит об опасности непосредственного применения периодической оптимизации параметров системы без учета результатов анализа стабильности.

2. Система, имеющая большие возможности подгонки под кривую (максимальные значения главной диагонали матриц Q), оказалась в итоге самой нестабильной по всем показателям. Этот результат подтверждает (хотя и для частного случая) общепринятый тезис о необходимости сокращения возможностей подгонки под кривую.

3. Предлагаемый механизм оценки стабильности работы системы позволяет количественно описать потерю качества работы системы на временных отрезках, на которых не производилась оптимизация параметров.

4. Оценка стабильности может рассматриваться не только в качестве оценки самой системы прогнозирования, но и в качестве оценки подхода к оптимизации системы и оценки соответствующего критерия оптимальности.

Проведенное исследование показало целесообразность применения предлагаемого метода оценки стабильности в процессе построения и выбора систем прогнозирования и управления инвестиционным портфелем. Также представляется интересной возможность использования предлагаемого метода при оценке стабильности качества прогнозирования в социально-экономических и технических системах.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Пардо Р. Разработка, тестирование, оптимизация торговых систем для биржевого трейдера. – М.: Минакс, 2002. – 224 с.

2. Саркисов В.Г., Саркисов Г.А. Метод оценки и выбора алгоритма принятия инвестиционных решений на основе его параметрической оптимизации // Вестник Самарского государственного технического университета. Сер. Технические науки. – 2009. – № 2(24). – С. 39-46.

3. Саркисов В.Г., Чихерева Т.В. Оптимизация системы прогнозирования на основе распознавания паттернов цен финансовых инструментов // Вестник Самарского государственного технического университета. Сер. Экономические науки. – 2014. – № 1(11). – С. 139-144.

4. Спири А.Г., Киселева Н.Е. Процедура хеширования в задачах структурного анализа временных рядов на финансовых рынках // XII Всероссийское совещание по проблемам управления ВСПУ-2014. Москва, 16-19 июня 2014. – М.: Институт проблем управления им. В.А. Трапезникова, 2014. – С. 8130-8138.

5. Фьючерсы ФОРТС RTS: Электронный ресурс. – Режим доступа: <http://www.finam.ru/analysis/profile0442F>.

6. Файловая библиотека Московской биржи. Фьючерсный контракт на Индекс РТС : Электронный ресурс. – Режим доступа: <http://fs.moex.com/files/3244>.

Поступила в редакцию 20/VII/2014;

в окончательном варианте – 25/VII/2014

UDC 007, 336.76

THE ANALYSIS OF STABILITY OF PRICES SHORT TIME SERIES FORECASTING BASED ON PATTERN RECOGNITION

V.G. Sarkisov, T.V. Chikhereva

Samara State Technical University

244, Molodogvardeyskaya st., Samara, 443100

The method of analogs is applied to forecast short time series of financial instruments prices. The dependence of the forecast quality on length of considered patterns is analyzed. Stability of investment decisions quality is investigated by optimization on various time ranges. Some criteria of the forecasting system quality taking into account the forecast accuracy are considered.

Keywords: prices time series, pattern, hash code, quality of the forecast, stability, optimization.

Original article submitted 20/VII/2014;

revision submitted – 25/VII/2014.

Vigen G. Sarkisov (Ph.D.), Chair of mathematics and applied informatics.

Tatyana V. Chikhereva, Student direction «Management» profile «Economics and Management of Fuel and Energy Complex».