УДК 681.5.01

В.А. АНТОНОВ, АЛЬ-ГУЛИ АБЕД, Б.В. ШАМША

ПОСТРОЕНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ В ДИЛИНГОВЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ ПО ФУНДАМЕНТАЛЬНЫМ ФАКТОРАМ

Рассматриваются вопросы использования фундаментального анализа для построения математических моделей долгосрочного и среднесрочного прогнозирования курса валют в дилинговых информационных системах по макроэкономическим факторам. Исследуется целесообразность использования регрессионного анализа, робастных оценок и нейронных сетей для построения моделей изменения курса валют по фундаментальным факторам.

1. Введение

Построение математических моделей по фундаментальным факторам позволяет формировать и предсказывать тренды в динамике цен, проводить анализ и прогноз фундаментальных факторов, а также оценивать их влияние на трендовую динамику цен.

Стратегические инвесторы, осуществляющие долгосрочное инвестирование, основное внимание в своей работе уделяют именно фундаментальному анализу. Математические модели здесь могут сыграть решающую роль. Более того, по степени влияния фундаментальных факторов на курс национальной валюты можно определить состояние национальной экономики страны и оценить объем денежных средств, учетную и рыночную ставки. Математические модели могут использоваться органами государственного управления для планирования финансово-экономической политики.

2. Постановка задачи анализа применимости методов математического моделирования в дилинговых информационных системах

Построение математических моделей прогноза по фундаментальным факторам представляет собой достаточно сложную задачу. Это связано с тем, что уровень валютного курса формируется под воздействием множества факторов в предыдущие моменты времени. В этом случае в наборе экспериментальных данных присутствует автокорреляция, что существенно усложняет процесс интерпретации результатов статистического анализа. Более того, для различных значений лагов имеет место цикличность данных.

Зависимость между фундаментальными факторами и курсом валют является нелинейной. Известно, что использование линейных математических моделей не всегда приводит к удовлетворительным результатам. Поэтому исследование применимости того или иного метода построения моделей прогноза для такого класса случайных последовательностей представляет интерес и является предметом внимания многих ученых, занимающихся вопросами разработки аналитических методов прогнозирования курсов валют в дилинговых системах. В предлагаемой работе рассматриваются некоторые аспекты применения регрессионного анализа (РА), робастных оценок и нейронных сетей для прогнозирования курсов валют по фундаментальным факторам.

Валютный курс является отражением международных экономических отношений со страной валюты-измерителя. Влияние некоторых фундаментальных факторов на курс валюты в разных странах имеет различную качественную оценку. Так, изменение платежного баланса в Германии и Франции несущественно сказывается на курсе валюты. Определенную специфику имеют и фундаментальные факторы Японии, США. В этой связи представляет интерес оценка влияния фундаментальных факторов на курс валюты в Украине по математическим моделям. Однако данные, полученные на международном валютном рынке (FOREX), имеют специфические характеристики, которые противоречат многим предпосылкам и предположениям использования различных методов построения моделей. В связи с этим представляет интерес определение оценки степени применимости некоторых методов построения модели прогнозирования по динамическим рядам. В статье приведены результаты исследования применимости РА, робастных оценок и нейронных сетей для построения моделей прогнозирования детерминированных составляющих временного ряда курса валют по фундаментальным факторам.

3. Построение математической модели прогноза курса валют с использованием регрессионного анализа

Исходные данные для построения моделей прогнозирования курса валют в Украине по фундаментальным факторам были получены из INTERNET. Данные за период с января 1996 по декабрь 2001 года приведены на www.bank.gov.ua с лагом, равным месяцу, кварталу, и будут использованы для синтеза моделей среднесрочного и долгосрочного прогнозирования валютного курса.

В качестве зависимой переменной определим курс валюты (Y).

Независимые переменные представим в виде:

x_1 - валовой внутренний продукт;

x_2 - денежные доходы населения;

x_3 - денежные затраты;

x_4 - индекс цен производителей;

x_5 - индекс цен потребителей;

x_6 - размер дефицита;

x_7 - процентные ставки банков;

x_8 - депозиты;

x_9 - объем денежной массы вне банков;

x_{10} - объем денежной массы вне банков, а также на текущих и расчетных счетах в национальной валюте;

x_{11} - объем денежной массы вне банков, на текущих и расчетных счетах в национальной валюте, срочные вклады в национальной и зарубежной валютах.

Вследствие ошибок измерения, которые присутствуют в регистрационном эксперименте, и влияния различных неучтенных факторов зависимая переменная является случайным процессом с определенными статистическими характеристиками.

Любая модель отражает только некоторые характерные черты объекта и никогда не бывает его точной копией. Следовательно, нет оснований говорить об «истинной» модели в полном смысле слова. Обычно под «истинным» значением понимают условное математическое ожидание зависимой переменной при заданных значениях независимых.

Структура модели может быть представлена различными функциями, но в рассматриваемом случае – обязательно линейными относительно коэффициентов. В модели, наряду со стандартными данными, необходимо использовать квадраты независимых переменных и различные варианты их произведений. Это связано с априорной информацией о нелинейном влиянии фундаментальных факторов на курс валюты, а также с результатом анализа диаграмм рассеивания, который показал целесообразность преобразования данных.

Как известно, для корректного использования РА необходимо соблюдение определенных предпосылок и предположений, которые достаточно полно изложены в работах [1,2].

При решении задачи построения модели курса валют по фундаментальным факторам перечисленные допущения классического РА представляют собой априори заданные требования. В связи с этим при использовании РА для обработки экспериментальной информации необходимо решить ряд статистических проблем:

– провести статистический анализ экспериментальных данных;

– задать структуру модели и получить наилучшие точечные и интервальные оценки ее параметров;

– провести интерпретацию модели, проверив гипотезы относительно ее параметров, оценив адекватность модели и проверив предположения, на которых основан РА, используя анализ остатков.

На первом этапе построения математической модели проверим гипотезы о случайном характере зависимых и независимых векторов при помощи критерия серий [3].

Суть его состоит в определении последовательности наблюденных значений, перед которыми и после которых расположены значения другой категории. Рассматривается гипотеза об отсутствии тренда, т.е. последовательность N наблюдений содержит только независимые значения одной и той же случайной величины. Гипотеза проверяется при любом заданном уровне значимости α путем сопоставления наблюденного числа серий с граничными значениями $r_{n;1-\alpha/2}$ и $r_{n;\alpha/2}$, где $n=N/2$. Если наблюденное число серий выходит за границы этого интервала, гипотезу следует отвергнуть при уровне значимости α. В [3] для различных уровней α приведены значения $r_{n;1-\alpha/2}$ и $r_{n;\alpha/2}$.

Определим, являются ли независимые переменные x_1-x_{11} и зависимая переменная Y случайными. Выполним проверку при уровне значимости $\alpha=0{,}05$. Анализ показал, что переменные x_1, x_3, x_6, x_{11} и зависимая переменная Y имеют тренд. Переменные x_4, x_5 являются случайными.

На следующем этапе проверим гипотезу о том, что наблюдения распределены по нормальному закону.

Напомним, что в РА независимые переменные рассматриваются как неслучайные величины, независимо от их истинного закона распределения. Проверка нормальности закона распределения Y необходима лишь для проверки значимости уравнения регрессии и его параметров, а также для интервального оценивания параметров. Для получения точечных оценок параметров модели этого условия не требуется. Заметим, что проверку можно осуществить на основании анализа графических и аналитических процедур. Для выяснения вопроса о виде закона распределения используют представления эмпирического распределения в виде гистограмм или графика функции распределения, по которым предварительно оценивают тип распределения. Выбор типа теоретического распределения обычно осуществляется на основе глазомерного анализа. Как правило, о виде распределения судят по выборочным точечным оценкам

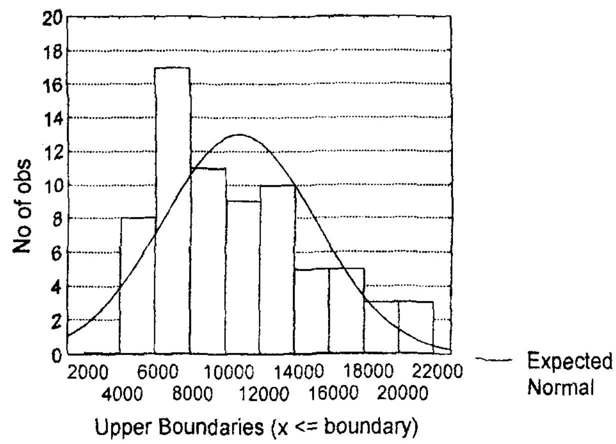

Рис. 1. Гистограмма переменной x_1

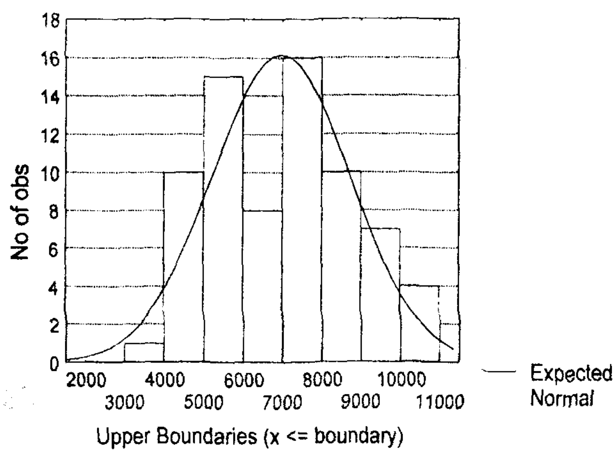

Рис. 2. Гистограмма переменной x_3

числовых характеристик - стандартизованные асимметрия и эксцесс (деленные на стандартное отклонение), которые характеризуют близость распределения к нормальному. Если эти коэффициенты находятся в диапазоне [-2, 2], то можно ожидать, что распределение близко к нормальному. Если плотность распределения симметрична, то коэффициент асимметрии равен нулю. Для нормального распределения коэффициент эксцесса равен нулю. Если распределение сконцентрировано вокруг среднего больше, чем нормальное, *то он меньше нуля.*

На рис. 1-4 приведены гистограммы переменных x_1, x_3, x_6, x_{11}.

Визуально для переменных x_1, x_3, x_6, x_{11} можно принять гипотезу о нормальном законе распределения. Гистограммы для остальных переменных показывают, что принять гипотезу о нормальном законе распределения нельзя.

Дальнейшую проверку гипотезы проведем по графикам нормальности. На рис. 5-8 приведены графики расположения данных переменных x_1, x_3, x_6, x_{11} на прямой, соответствующей идеальному нормальному закону распределения.

Гипотеза о равенстве эмпирического закона распределения нормальному принимается в случае, если исходные данные лежат на прямой идеального закона распределения. Заметим, что S-образное расположение данных на графиках свидетельствует о необходимости каким-то образом преобразовать переменные, например, логарифмирование часто используется для того, чтобы выровнять край распределения.

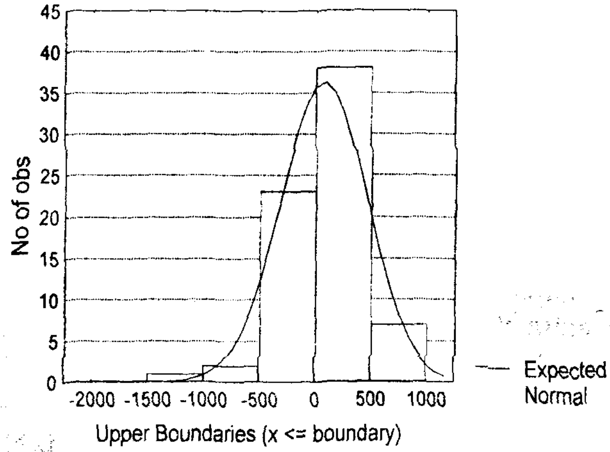

Рис. 3. Гистограмма переменной x_6

На этих графиках можно визуально обнаружить выбросы. Анализ независимых переменных показывает, что только для переменных x_3, x_6 можно принять гипотезу о нормальном законе распределения.

Наиболее достоверный результат проверки гипотезы о нормальном распределении зависимых и независимых переменных можно получить с помощью существующих критериев согласия эмпирического и теоретического распределения: Пирсона, Крамера-Мизеса, Колмогорова - Смирнова, Шапиро-Уилкса [3].

Для исследуемых переменных значения основных статистик и результаты *исследования* нормальности распределения представлены в таблице, где приняты следующие обозначения: m – математическое ожидание; s – стандартное отклонение; d – критерий Колмогорова-Смирнова; L – критерий Lilliefors; W – критерий Шапиро-Уилкса.

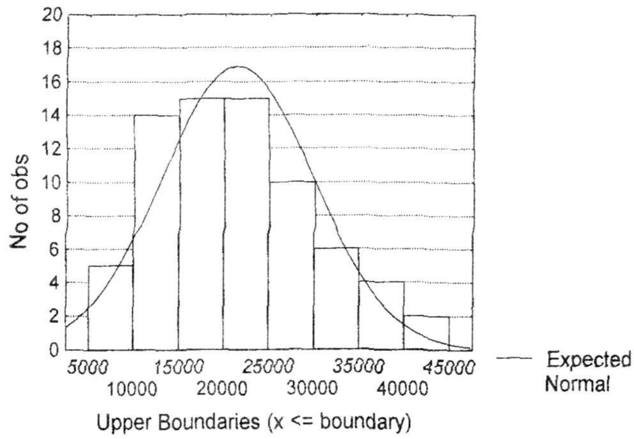

Рис. 4. Гистограмма переменной x_{11}

Согласно таблице результатов, *принимаем гипотезу о* нормальном распределении только для переменных x_1, x_7, x_8, x_9.

Анализ диаграмм рассеивания и графиков нормальности привел к выводу, что в распределениях имеются «хвосты», графики нормальности имеют S-образную форму и выборки исходных данных имеют различный масштаб. Эти факты свидетельствуют о необходимости преобразования исходных данных – следует произвести логарифмирование (переменные $x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9, x_{10}, x_{11}$) и масштабирование ($x_6$).

В связи с большим числом фундаментальных факторов и ограниченной длиной выборки будем применять пошаговый алгоритм оценки параметров уравнения регрессии.

В результате расчетов *были получены оценки* параметров модели. В соответствии с t-критерием значимыми оказались 6 независимых переменных (значение $p<0{,}05$). Значение F-критерия полученной модели оказалось равно 249,91, а в значении вероятности p ненулевые цифры появляются только в шестом знаке после запятой. Коэффициент детерминации и его квадрат равны 0,9825 и 0,9652, соответственно. Масштаб зависимой переменной оказался равным 31,055. Эти результаты свидетельствуют об адекватности полученной модели, приведенной ниже:

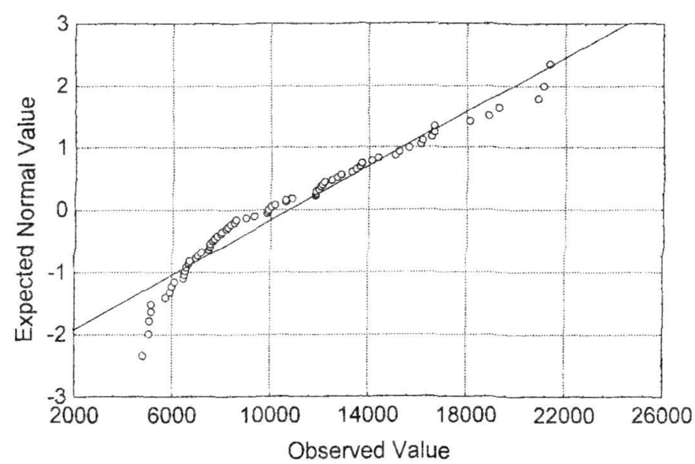

Рис. 5. График нормальности переменной x_1

$$Y = -3211{,}14 + 0{,}071\,Ln(x_5) + 0{,}276\,Ln(x_7) - 0{,}21\,Ln(x_8) - \\ - 0{,}81\,Ln(x_9) - 2{,}3\,Ln(x_{10}) + 4{,}08\,Ln(x_{11}). \tag{1}$$

Следующим шагом при оценке адекватности является анализ остатков. Оценка постоянства математического ожидания и дисперсии дала положительный результат. Необходимо оценить соответствие плотности распределения остатков модели нормальному закону распределения. График нормальности остатков регрессионной модели приведен на рис. 9. Визуальный анализ графика нормальности остатков и диаграммы рассеивания позволил выявить наличие «тяжелых хвостов». Этот факт ставит под сомнение соответствие закона распределения остатков модели нормальному закону и, следовательно, адекватность модели в целом.

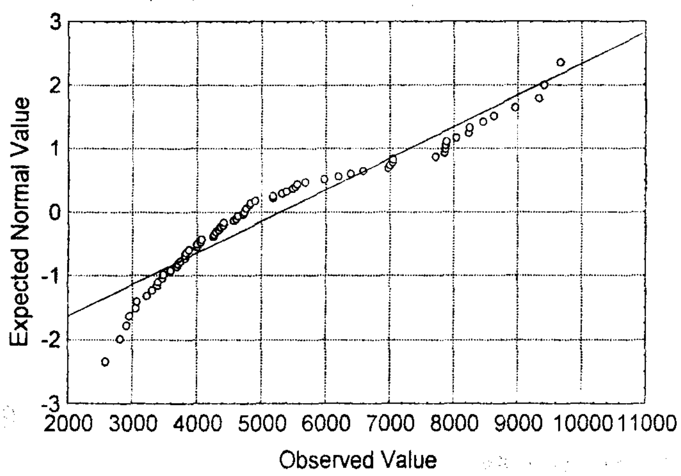

Рис. 6. График нормальности переменной x_3

В связи с этим целесообразно перейти к робастным методам оценки параметров регрессионной модели. Этот класс методов позволяет исключить влияние выбросов в исходных данных, которые обуславливают «тяжелые хвосты» в распределении остатков.

4. Робастные методы оценивания параметров регрессии

В качестве робастного будем использовать метод наименьшей медианы квадратов (LMS), предложенный Рауссеу [4,5]. Его суть состоит в последовательном переборе подмножеств исходных данных и определении такого подмножества, которое обеспечит наименьшее значение масштаба зависимой переменной. Далее на основании этого подмножества и строится регрессионная модель. Благодаря тому, что модель строится по подмножеству значений, данный метод позволяет исключить из набора исходных данных выбросы, что обеспечивает получение более качественного результата по сравнению с классическим РА. В общем случае этот метод способен выявить до 50 % ошибочных данных, что и обусловило его применение в нашем случае.

В результате применения указанного метода была получена модель следующего вида:

$$\begin{aligned} Y = &-751{,}21 + 8{,}1\mathrm{Ln}(x_1) - 37{,}4\mathrm{Ln}(x_2) - 94{,}02\mathrm{Ln}(x_3) + 136{,}51\mathrm{Ln}(x_4) - \\ &- 412{,}64\mathrm{Ln}(x_5) + 3{,}33(x_6/100) + 72\mathrm{Ln}(x_7) - 62{,}61\mathrm{Ln}(x_8) - \\ &- 194{,}06\mathrm{Ln}(x_9) - 1364\mathrm{Ln}(x_{10}) + 1829{,}57\mathrm{Ln}(x_{11}). \end{aligned} \tag{2}$$

Коэффициент детерминации при этом составил 0,96476, а масштаб зависимой переменной 29,234. Значение масштаба меньше, чем в случае классического РА, что свидетельствует о получении более точного результата аппроксимации по сравнению с моделью (1).

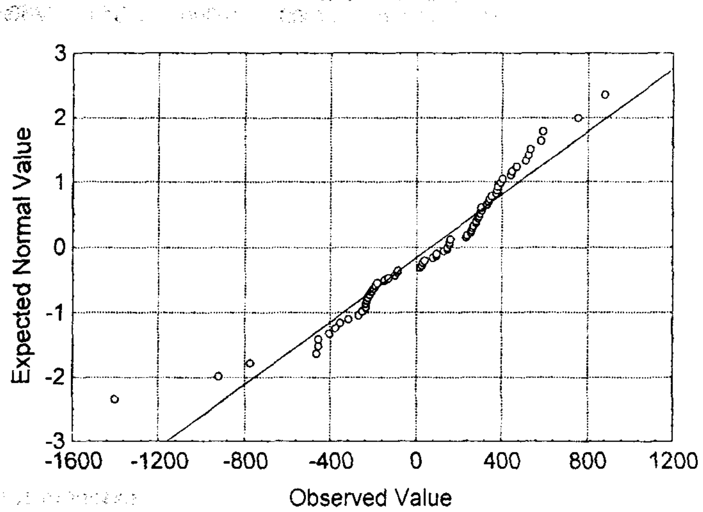

Рис. 7. График нормальности переменной x_6

Исследования робастных методов в [6,7] показали, что метод LMS не обеспечивает получение эффективных оценок параметров. В этой связи воспользуемся иным робастным методом - методом взвешенных наименьших квадратов (RLS) [4,6]. Он базируется на результатах метода LMS. Метод RLS использует информацию о наличии выбросов в исходных данных, полученную с помощью метода LMS, и учитывает их со сниженными весами при построении модели с использованием классического РА. Таким образом, достигается максимальная эффективность результатов при сохранении свойств робастности. С использованием метода RLS была получена следующая модель регрессии:

$$\begin{aligned} Y = & -2207{,}95 + 28{,}51\mathrm{Ln}(x_1) + 47{,}73\mathrm{Ln}(x_2) - 45{,}98\mathrm{Ln}(x_3) - 252{,}66\mathrm{Ln}(x_4) + \\ & + 282{,}93\mathrm{Ln}(x_5) - 0{,}89(x_6/100) + 68{,}88\mathrm{Ln}(x_7) - 61{,}54\mathrm{Ln}(x_8) - \\ & - 95{,}03\mathrm{Ln}(x_9) - 1329\mathrm{Ln}(x_{10}) + 1574{,}8\mathrm{Ln}(x_{11}). \end{aligned} \quad (3)$$

Коэффициент детерминации при этом составил 0,98028, а масштаб зависимой переменной 24,522. Значение масштаба меньше, чем в случае классического РА модели, полученной с использованием LMS, что свидетельствует о получении более точного результата аппроксимации по сравнению с моделями (1) и (2).

Однако даже последняя полученная модель, обеспечивающая наименьший масштаб значений зависимой переменной, не обеспечивает необходимую точность аппроксимации курса национальной валюты. Проведенные исследования позволят лишь судить о влиянии того или иного фундаментального фактора на состояние курса валюты, но непригодны для получения точных результатов, например, при прогнозировании.

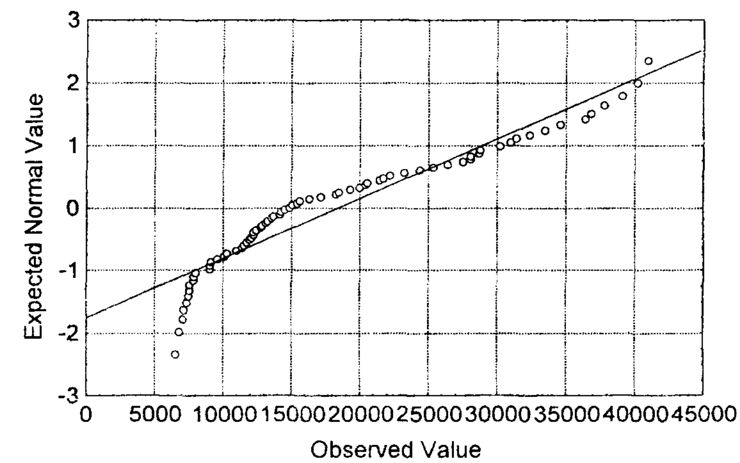

Рис. 8. График нормальности переменной x_{11}

В связи с этим необходимо обратиться к математическому аппарату, который в состоянии обеспечить приемлемое качество аппроксимации, вследствие чего позволит получить наиболее точную модель, описывающую зависимость курса валюты от исследуемых фундаментальных факторов.

5. *Аппарат искусственных нейронных сетей*

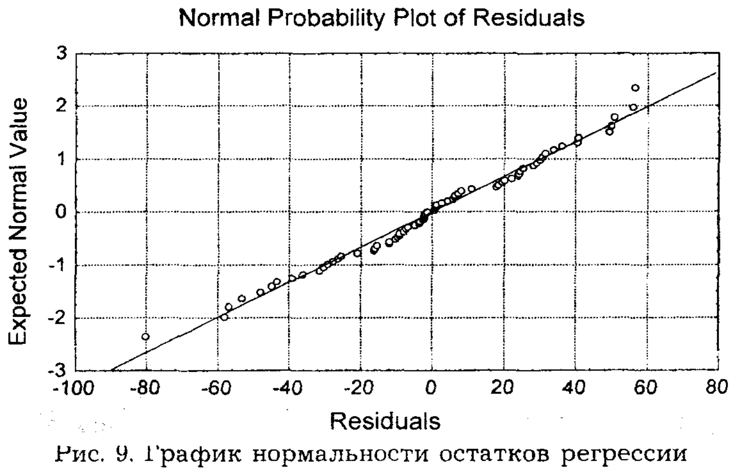

Рис. 9. График нормальности остатков регрессии

На сегодняшний день наиболее универсальным и простым в использовании является аппарат искусственных нейронных сетей (ИНС), который позволяет с высокой точностью аппроксимировать нелинейные функции [8]. Используем аппарат ИНС для построения модели зависимости курса валют от ряда фундаментальных факторов.

Используем нейронную сеть прямого распространения с одним скрытым слоем. В скрытом слое возьмем 4 нейрона с сигмодной функцией в качестве функции активации. В выходном слое используем линейную функцию активации, что

Результаты проверки гипотезы нормальности распределения исходных данных по критериям согласия Колмогорова – Смирнова, Lilliefors, Шапиро-Уилкса

Переменная	m	s	d	L	W
x_1	10790,18	4357	d=,12951, p<,20	p<,01	W=,91944, p<,0001
x_2	5502,831	2070	d=,16640, p<,05	p<,01	W=,89461, p<,0000
x_3	5301,746	1898	d=,16043, p<,10	p<,01	W=,89881, p<,0000
x_4	101,231	1,787	d=,23697, p<,01	p<,01	W=,63432, p<,0000
x_5	101,5239	1,88	d=,18114, p<,05	p<,01	W=,85337, p<,0000
x_6	69,77439	390,44	d=,11185, p>,20	p<,05	W=,94961, p<,0151
x_7	52,31268	17,617	d=,10873, p>,20	p<,05	W=,87268, p<,0000
x_8	20,05211	8,7417	d=,12311, p>,20	p<,01	W=,87998, p<,0000
x_9	7931,197	3802,9	d=,15746, p<,10	p<,01	W=,92490, p<,0003
x_{10}	12268,54	6195,1	d=,17206, p<,05	p<,01	W=,88856, p<,0000
x_{11}	18433,24	9829,8	d=,16150, p<,05	p<,01	W=,88613, p<,0000
Y	348,8841	158,02	d=,24645, p<,01	p<,01	W=,75983, p<,0000

значительно упростит аналитическое выражение полученной модели. В качестве алгоритма обучения применим алгоритм обратного распространения.

Для обеспечения наиболее точной аппроксимации перед обучением исходные данные (обучающая выборка) были нормированы таким образом, чтобы их значения лежали в диапазоне [0; 1].

В результате обучения нейронная сеть аппроксимирует исследуемую выборку со средней ошибкой 0,00028 и максимальной ошибкой 0,0017, что составляет 0,1 и 0,63 копейки. Это более чем удовлетворительный результат, с точки зрения его точности. Точность аппроксимации на два порядка превышает точность моделей (1)-(3). С учетом топологии ИНС и полученных в результате обучения весов модель в аналитическом виде будет иметь следующий вид:

$$\begin{aligned}
Y &= -1,65-0,81/(1+e^{-(\alpha)})+2,03/(1+e^{-(\beta)})-2,03/(1+e^{-(\gamma)})-0,02/(1+e^{-(\delta)});\\
\alpha &= 0,54+0,64x_1-1,38x_2-0,98x_3+0,33x_4-0,62x_5+5,68x_6+0,18x_7-1,85x_8-1,89x_9+0,36x_{10}-6,82x_{11};\\
\beta &= -5,47-5,04x_1+9,31x_2-0,4x_3+0,68x_4+1,36x_5+5,18x_6+3,59x_7-6,74x_8+5,38x_9+14,75x_{10}-5,44x_{11};\\
\gamma &= -0,64-5,72x_1+3,02x_2+4,57x_3+1,39x_4+0,36x_5+1,03x_6-1,25x_7-4,39x_8-9x_9+6,49x_{10}+0,16x_{11};\\
\delta &= -2,01+4,65x_1-0,22x_2-5,65x_3-1,23x_4+4,26x_5+1,61x_6+1,87x_7-1,64x_8+12,38x_9+1,69x_{10}-4,79x_{11}.
\end{aligned} \qquad (4)$$

6. Выводы

Приведенные результаты исследований позволяют сделать следующие выводы относительно применимости тех или иных методов для построения математических моделей зависимости курса валют от ряда фундаментальных факторов в информационных дилинговых системах. Применение регрессионного анализа позволило получить модель (1), которую можно использовать для качественной оценки влияния тех или иных фундаментальных факторов на динамику изменения курса валют. Точностные характеристики модели не состоятельны. Преодолеть наличие выбросов в исходных данных позволяют робастные методы – модели (2) и (3) соответственно. Их точность выше, чем для модели, полученной с использованием РА, но не существенно. Для практического применения пригодна математическая модель (4), полученная с использованием аппарата ИНС –

точность аппроксимации выше на два порядка по сравнению со статистическими моделями. Однако по виду модели (4) достаточно сложно судить о влиянии каждого фундаментального фактора на динамику изменения курса валют, что ограничивает ее применение для качественной оценки ситуации.

Список литературы: 1. *Вучков И., Бояджиева Л., Солаков Е.* Прикладной линейный регрессионный анализ. М.: Финансы и статистика, 1987. 239 с. 2. *Демиденко Е.З.* Линейная и нелинейная регрессии. М.: Финансы и статистика, 1981. 302 с. 3. *Тюрин Ю.Н., Макаров А.А.* Статистический анализ данных на компьютере/Под ред. В.Э. Фигурнова. М.: ИНФРА-М, 1998. 528 с. 4. *Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В.* Робастность в статистике, подход на основе функции влияния. М.: Мир, 1989. 511с. 5. *Rousseeuw P.J., Hubert M.* Recent developments in PROGRESS. // J1 -Statistical Procedures and Related Topics. 1997. Vol. 31. P.201-214. 6. *Шамша Б.В., Антонов В.А., Васин И.С.* Сравнение эффективности методов робастного оценивания на основе исходной статистической информации // Информационно-управляющие системы на железнодорожном транспорте. 1999. № 5. С.68-72. 7. *Meintanis S.G., Donatos G.S.* A comparative study of some robust method for coefficient-estimation in linear regression. Computational Statistics & Data Analysis 1997. Vol. 23. P. 525-540. 8. *Бэстен Д.-Э., Ван ден Берг В.-М., Вуд Д.* Нейронные сети и финансовые рынки: принятие решений в торговых операциях. М.: ТВП, 1997. 236 с

Поступила в редколлегию 25.06.2003

Антонов Владислав Александрович, канд. техн. наук, ст. преп. кафедры ИУС ХНУРЭ. Научные интересы: теоретические и методологические основы разработки ИУС, математическое обеспечение ИУС, компьютерные сети. Адрес: Украина, 6116, Харьков, пр. Леннина, 14, тел. 70-21-451.

Аль-Гули Абед, аспирант каф. ИУС ХНУРЭ. Научные интересы: математические модели дилинговых информационных систем. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-451.

Шамша Борис Владимирович, канд. техн. наук,. доцент, профессор кафедры ИУС ХНУРЭ. Научные интересы: обработка данных и управление. Адрес: Украина, 61018, Харьков, ул. Космонавтов, 5, кв. 32, тел. 33-27-78.