



БИБЛИОТЕКА ЦЕНТРА ИССЛЕДОВАНИЙ ПЛАТЕЖНЫХ СИСТЕМ И РАСЧЕТОВ

Как предсказать курс доллара

Эффективные методы прогнозирования
с использованием Excel и Eviews

В.Г. Брюков



В.Г. БРЮКОВ

КАК ПРЕДСКАЗАТЬ КУРС ДОЛЛАРА

**ЭФФЕКТИВНЫЕ МЕТОДЫ
ПРОГНОЗИРОВАНИЯ
С ИСПОЛЬЗОВАНИЕМ
EXCEL И EIEWS**



Центр исследований
платежных систем и расчетов



МОСКВА
2011

УДК 338
ББК 65.054.3
Б87

Брюков В.Г.

Б87 Как предсказать курс доллара. Эффективные методы прогнозирования с использованием Excel и EViews / В.Г. Брюков. — М.: КНОРУС; ЦИПСИР, 2011. — 272 с.

ISBN 978-5-406-01441-7

Детально излагаются методики построения стационарных и нестационарных статистических моделей по прогнозированию курса доллара США с использованием программ EViews и Excel. Прогнозы по курсу доллара к рублю делаются с упреждением в один месяц, две и одну неделю, а по курсу евро к доллару — с упреждением в один день. Особый акцент сделан на составлении (с установленным инвестором уровнем надежности) прогнозов цен покупки и продажи валют для работы на валютном рынке на основе разработанных статистических моделей. Все методики с успехом применяются на практике.

Для валютных инвесторов, трейдеров и студентов, будущая профессия которых связана с работой в банке, финансовой компании или с операциями на финансовых и товарных рынках. Может быть полезна всем, кто интересуется валютным рынком, собирается зарабатывать или уже зарабатывает на этом рынке, хочет научиться делать прогнозы по курсам валют.

УДК 338
ББК 65.054.3

Брюков Владимир Георгиевич
**КАК ПРЕДСКАЗАТЬ КУРС ДОЛЛАРА.
ЭФФЕКТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ
С ИСПОЛЬЗОВАНИЕМ EXCEL И EIEWS**

Санитарно-эпидемиологическое заключение
№ 77.99.60.953.Д.006828.04.10 от 28.04.2010 г.

Изд. № 3685. Подписано в печать 18.02.2011. Формат 60×90/16.
Гарнитура «NewtonС». Печать офсетная.
Усл. печ. л. 17,0. Уч.-изд. л. 10,26. Тираж 1000 экз. Заказ № 1080.

ООО «Центр Исследований Платежных Систем и Расчетов».
E-mail: info@paysyscenter.ru Блог: <http://paysyscenter.livejournal.ru>
www.paysyscenter.ru

ООО «КноРус».

129085, Москва, проспект Мира, д. 105, стр. 1.
Тел.: (495) 741-46-28.

E-mail: office@knorus.ru <http://www.knorus.ru>

Отпечатано в полном соответствии с качеством
предоставленного издательством электронного оригинал-макета
в ГУП «Брянское областное полиграфическое объединение».
241019, г. Брянск, пр-т Ст. Димитрова, 40.

ISBN 978-5-406-01441-7

© ООО «Центр Исследований
Платежных Систем и Расчетов», 2011
© ООО «КноРус», 2011

Оглавление

Предисловие 5

Глава 1. Понятие о стационарном и нестационарном временном ряде, выявление нестационарности ряда графическим способом

- 1.1. Краткая характеристика стационарных
и нестационарных случайных процессов 9
- 1.2. Распознавание стационарности временного ряда
с помощью построения его графика 11

Глава 2. Метод наименьших квадратов и решение уравнения регрессии в Excel

- 2.1. Характеристика метода наименьших квадратов
и его применение при прогнозировании курса доллара 20
- 2.2. Решение уравнения регрессии в Excel с учетом
фактора времени. Интерпретация и оценка
значимости полученных параметров 22
- 2.3. Решение уравнений регрессии в Excel графическим способом 40

Глава 3. Уравнения авторегрессии и авторегрессии со скользящим средним

- 3.1. Специфика уравнений авторегрессии (AR) 50
- 3.2. Специфика уравнений авторегрессии
со скользящим средним (ARMA) 53
- 3.3. Коррелограмма и идентификация лаговых переменных
в уравнениях AR и ARMA 54
- 3.4. Решение в Excel уравнения авторегрессии
2-го порядка AR(2) 61
- 3.5. Решение в EViews уравнения авторегрессии
2-го порядка AR(2) 64
- 3.6. Интерпретация параметров уравнения
авторегрессии в EViews 69
- 3.7. Оценка точности решения уравнения
авторегрессии в EViews 76

Глава 4. Подбор адекватного уравнения авторегрессии и составление точечных и интервальных прогнозов по курсу доллара

- 4.1. Повышение статистической значимости коэффициентов
в уравнении авторегрессии 85
- 4.2. Оценка точности прогностической модели, проверка
остатков на автокорреляцию и стационарность 88
- 4.3. Описательная статистика и тестирование остатков
на нормальное распределение 95
- 4.4. Построение точечных и интервальных прогнозов 105
- 4.5. Проверка точности составленных интервальных прогнозов 113

Глава 5. Тестирование структурной нестационарности и построение нестационарной статистической модели с оптимизированным временным рядом

| | |
|--|-----|
| 5.1. Тестирование авторегрессионного процесса на стационарность путем нахождения обратных единичных корней | 120 |
| 5.2. Тестирование AR-структуры на стационарность с помощью функции импульсного ответа | 126 |
| 5.3. Влияние резких изменений курса доллара на смещение коэффициентов регрессии | 131 |
| 5.4. Стандартные и студентизированные остатки, влияние выбросов на точность уравнения регрессии | 135 |
| 5.5. Тесты Чоу на наличие структурной стабильности во временном ряде | 143 |
| 5.6. Структурные изменения в курсе доллара, произошедшие в августе—октябре 1998 г. | 156 |
| 5.7. Построение статистической модели с оптимальным диапазоном интервального прогноза | 166 |

Глава 6. Построение стационарной статистической модели

| | |
|---|-----|
| 6.1. Тестирование исходного и логарифмического временного ряда на стационарность | 174 |
| 6.2. Построение модели авторегрессии со скользящей средней и стационарной ARMA-структурой | 180 |
| 6.3. Тестирование модели авторегрессии со скользящей средней на автокорреляцию в остатках и проверка стационарности ее ARMA-структуры | 191 |
| 6.4. Оценка стабильности стационарной модели авторегрессии со скользящей средней | 199 |
| 6.5. Оценка точности стационарной модели ARMA | 203 |
| 6.6. Построение стационарной модели ARMA с оптимизированным временным рядом | 215 |

Глава 7. Использование статистических моделей в качестве инструмента торговой системы

| | |
|--|-----|
| 7.1. Использование в торговле модели для прогнозирования курса доллара к рублю с упреждением в один месяц | 229 |
| 7.2. Использование в торговле модели для прогнозирования курса доллара к рублю с упреждением в две недели | 242 |
| 7.3. Использование в торговле модели для прогнозирования курса доллара к рублю с упреждением в одну неделю | 251 |
| 7.4. Использование в торговле модели для прогнозирования курса евро к доллару с упреждением в один день | 258 |

| | |
|--------------------------|-----|
| Краткое заключение | 271 |
| Литература | 272 |

Каким будет курс американского доллара завтра, послезавтра, через неделю или через месяц? Сколько будут стоить в ближайшем будущем единая европейская валюта, японская йена, британский фунт, швейцарский франк, канадский, австралийский или новозеландский доллар и прочие валюты? Стоит ли сегодня (завтра, послезавтра или же через месяц) вкладывать российские рубли в доллары США, евро, фунт или в йену, либо, напротив, от иностранной валюты нужно поскорее избавляться? Каких рекомендуемых курсов покупки или продажи при этом следует придерживаться?

Мы живем в такое время, когда ответы на эти животрепещущие вопросы ежедневно приходится искать миллионам рядовых инвесторов и профессиональных трейдеров во всем мире, в том числе и в России. И это вполне понятно, поскольку от правильных ответов на эти вопросы зависит их будущее благосостояние.

По нашему мнению (и хочется надеяться, что читатель с ним согласится), тот, кто возьмет на себя труд внимательно прочитать, а еще лучше — протестировать эту книгу, сможет дать более компетентные ответы на эти поставленные вопросы.

Следовательно, книга предназначена для тех, кто интересуется валютным рынком, кто собирается зарабатывать или уже зарабатывает на этом рынке и кто хочет научиться делать прогнозы по курсам валют. Впрочем, эта книга будет полезна и интересна не только для валютных инвесторов и трейдеров, но и для студентов, будущая профессия которых так или иначе связана с работой в банке, финансовой компании или с операциями на финансовых и товарных рынках. Причем полученные из этой книги знания пригодятся не только для работы на валютном рынке, но и для анализа других товарных и финансовых рынков, поскольку методика прогнозирования курса валют принципиально не отличается от прогнозирования цен, например, на такие товары, как нефть или золото.

В книге детально излагается методика построения стационарных и нестационарных статистических моделей по прогнозированию курса доллара США с использованием программ EViews и Excel. При этом особенно большое внимание уделяется разработке моделей прогнозирования по курсу доллара США с упреждением в один месяц. После освоения этого материала в заключительной, седьмой главе даны статистические модели по прогнозированию курса доллара с упреждением в две недели и упреждением в одну неделю, а также по прогнозированию курса евро к доллару с упреждением в один день. В этой же заключительной главе рассказывается о методике составления на основе

разработанных статистических моделей рекомендуемых цен покупки и продажи валют. Причем эффективность этих цен рассматривается на конкретных примерах с учетом последних рыночных данных.

Очевидно, что помимо прогнозов по американской валюте и евро многие хотели бы научиться делать прогнозы и по курсам других валют, например по курсу йены, фунта стерлингов, австралийского доллара. И этот список интересных для инвестора валют можно расширить по меньшей мере еще на два десятка денежных единиц.

В связи с этим необходимо отметить: внимательно прочитав и освоив материал этой книги, читатель в дальнейшем сможет справиться с подобной задачей вполне самостоятельно, поскольку в основе составления прогнозов по курсам различных валют лежит одна и та же методика, основанная на использовании феноменальных технических возможностей таких мощных вычислительных программ, как EViews и Excel.

Материал книги, состоящей из семи глав, излагается следующим образом. Во-первых, в каждом параграфе главы сначала дается конкретная проблема, связанная с прогнозированием курсов валют. Во-вторых, предлагается определенный алгоритм действий с использованием EViews или Excel. В-третьих, даются некоторые математические подробности, разъясняющие суть каждого алгоритма действий. В-четвертых, на конкретном примере показывается, как нужно пользоваться алгоритмом действий для решения той или иной проблемы, связанной с прогнозированием курса доллара. И наконец, в-пятых, в конце каждой главы читателю задаются вопросы, помогающие закрепить пройденный материал.

После того как читатель придет к убеждению, что он может заниматься прогнозированием самостоятельно, ничто не мешает ему воспользоваться EViews или Excel, чтобы составить прогнозы по курсу, например, японской йены или британского фунта, а также относительно будущих цен на нефть, золото и прочие товары.

Вполне понятно, что в эпоху бурного роста электронно-вычислительной техники и всеобщей автоматизации любой желающий овладеть азами валютного прогнозирования может существенно сэкономить время на расчетах, если он сумеет овладеть компьютерными способами обработки статистических данных, поэтому необходимым условием для нашей совместной работы является умение читателя работать с компьютером на уровне рядового пользователя, а также наличие определенных навыков работы с программой Microsoft Excel. Желательно также иметь хотя бы самое элементарное представление об основах теории статистики.

В процессе работы над книгой читатель сможет расширить свои знания о функциональных возможностях программы Excel, а также научиться работать в эконометрической программе EViews. Вполне естественно, что в первую очередь мы будем обращать внимание на обучение тем функциям этих программ, которые потребуются для прогнозирования курса валют.

Начинающим пользователям, у которых еще недостаточно опыта в работе с Microsoft Excel, можно порекомендовать прочитать следующую книгу: *Макарова Н.И., Трофимец В.Я.* Статистика в Excel : учеб. пособие. М. : Финансы и статистика, 2003, а также другие учебные пособия (их довольно много), рассказывающие о возможностях этой программы. В свою очередь тем читателям, которые хотят научиться прогнозировать курсы валют с помощью эконометрической программы EViews, можно порекомендовать в качестве пособия следующие книги: *Молчанов И.Н., Герасимова И.А.* Компьютерный практикум по начальному курсу эконометрики (реализация на EViews) : практикум. Ростов-на-Дону, 2001; *Турунцева М.Ю.* Анализ временных рядов. М. : МИЭФ ГУ-ВШЭ, 2003.

В заключение следует сказать еще об одном важном условии, без которого работа с этой книгой будет не очень продуктивной: наличие у читателя желания овладеть приемами прогнозирования курса валют, которые автор постарался изложить максимально доступным языком.

И последнее, о чем хотелось бы сказать в этом предисловии: хочу выразить свою глубокую признательность порталам <http://bankir.ru> и <http://www.spekulant.ru>, сотрудничество с которыми и стало одной из причин того, что эта книга появилась на свет. Дело в том, что регулярное сотрудничество с этими уважаемыми порталами, на первом из которых автор этих строк с середины 2009 г. делает ежемесячные прогнозы, а на втором — с октября 2010 г. — еженедельные прогнозы по курсам ведущих мировых валют, явилось своего рода катализатором, способствовавшим написанию этой книги.

*Владимир Брюков,
независимый аналитик*

ПОНЯТИЕ О СТАЦИОНАРНОМ И НЕСТАЦИОНАРНОМ ВРЕМЕННОМ РЯДЕ, ВЫЯВЛЕНИЕ НЕСТАЦИОНАРНОСТИ РЯДА ГРАФИЧЕСКИМ СПОСОБОМ

1.1. Краткая характеристика стационарных и нестационарных случайных процессов

Внезапно изменяющиеся тренды на валютном рынке на первый взгляд носят настолько причудливый и непредсказуемый характер, что, по мнению многих инвесторов, делать какие-то прогнозы по поводу курса валют — дело абсолютно безнадежное. И действительно, если посмотреть, например, на динамику ежемесячного курса доллара США (как, впрочем, и на динамику других свободно конвертируемых валют), то этот временной ряд нельзя назвать стационарным. Чтобы понять, к каким последствиям — с точки зрения прогнозирования курса американской валюты — ведет этот факт, нам придется немного углубиться в теорию стационарных и нестационарных случайных процессов.

Как известно, в статистической литературе принято выделять три типа случайных процессов: строго стационарные, слабо стационарные и нестационарные процессы.

Случайный процесс, образующий временной ряд $X_1, X_2, X_3, \dots, X_t$ (буква X обозначает переменную, содержащую определенную рыночную информацию, например, по динамике курсов валют, а цифры 1, 2, 3, ..., t — моменты времени), называется *строго стационарным* (или, как еще говорят, стационарным в узком смысле), если совместное распределение вероятностей всех переменных $X_1, X_2, X_3, \dots, X_t$ точно такое же, как и для наблюдений $X_{1+T}, X_{2+T}, X_{3+T}, \dots, X_{t+T}$ (где $T = t_2 - t_1$ — временной

лаг). Иначе говоря, свойства строго стационарного временного ряда не меняются при изменении начала отсчета времени.

Однако в сфере экономики, в том числе в сфере финансовых и валютных рынков, строго стационарные процессы отсутствуют, а потому для нас гораздо больший интерес представляют так называемые слабые стационарные процессы, или стационарные процессы в широком смысле. Под *слабым стационарным процессом* понимается случайный процесс, у которого среднее и дисперсия — независимо от рассматриваемого периода времени — имеют постоянное значение, а автоковариация зависит только от длины лага между исследуемыми переменными.

Напомним читателям, что *среднее значение временного ряда* можно найти по следующей формуле:

$$X_{\text{ср}} = \frac{\sum (X_1 + X_2 + \dots + X_t)}{n}, \quad (1.1)$$

где n — количество членов во временном ряду.

Дисперсия (мера разброса случайной величины, например, отклонения курса доллара от его среднего значения, или, как еще говорят, от его математического ожидания) временного ряда представляет собой средний квадрат отклонений переменной (случайной величины) от ее среднего значения.

Соответственно дисперсия находится по следующей формуле:

$$D = \frac{\sum (X_t - X_{\text{ср}})^2}{n}. \quad (1.2)$$

В Excel дисперсию можно найти, пользуясь функцией ДИСПР (если исходные данные представляют собой генеральную совокупность) или функцией ДИСП (если данные представляют собой выборку).

Для оценки тесноты и направления связи между переменными одного временного ряда с определенным лагом используется автоковариация. В частности, *автоковариация* между значениями X_t и X_{t-T} , отделенными друг от друга интервалом в T единиц времени, называется автоковариацией с лагом (задержкой) T , которая находится по следующей формуле:

$$\text{Cov}(X_t, X_{t-T}) = \frac{\sum (X_t - X_{\text{ср}}) (X_{t-T} - X_{\text{ср}})}{n}. \quad (1.3)$$

Автоковариацию, согласно формуле (1.3), в Excel можно найти с помощью функции КОВАР, которая возвращает величину ковариации. Причем последнюю называют автоковариацией в том случае, когда ее ис-

пользуют для оценки тесноты и направления связи между переменными одного временного ряда с определенным лагом — например, с лагом минус один месяц. Доказано, что для независимых переменных X и Y ковариация всегда равна нулю, а для зависимых переменных она, как правило, отличается от нуля. Если лаг $T=0$, то автоковариация равна дисперсии.

Если временной ряд, характеризующий динамику, например, курса валют, является слабо стационарным, то это означает отсутствие: во-первых, тренда; во-вторых, строго периодических колебаний; в-третьих, систематических изменений дисперсии; в-четвертых, каких-либо иных систематических изменений во временном ряде. Таким образом, под *стационарным процессом* в слабом или в широком смысле понимается случайный процесс, для которого среднее и дисперсия — независимо от периода времени — имеют постоянное значение, а автоковариация зависит от длины лага между рассматриваемыми переменными.

Если временной ряд является нестационарным, то с точки зрения теории это предполагает, что он содержит не только случайную компоненту, но и тренд, а его средняя, дисперсия и автоковариация изменяются во времени. В связи с этим делать прогнозы по нестационарному временному ряду более затруднительно (особенно на длительный период или в период каких-либо резких изменений в его динамике), чем по стационарному ряду.

1.2. Распознавание стационарности временного ряда с помощью построения его графика

Существуют различные методы распознавания стационарности временного ряда, однако, пожалуй, самым простым из них является построение графика временного ряда с последующим визуальным определением наличия в нем тренда.

С этой целью построим график ежемесячных колебаний курса доллара к рублю за период с июня 1992 г. до апреля 2010 г. Читатели, которые не умеют строить диаграммы, могут ознакомиться с представленными ниже алгоритмами действий № 1 и 2.

Алгоритм действий № 1

Как строить диаграммы в Microsoft Excel

Шаг 1. Поиск данных, их загрузка и первичная обработка в Excel

Во-первых, нужно на сайте Банка России <http://www.cbr.ru/> взять необходимые данные по ежедневным курсам доллара за весь инте-

ресующий нас период. Во-вторых, после копирования рыночной статистики в файл Microsoft Excel все данные по курсу доллара к рублю с 1 июля 1992 г. по 1 января 1998 г. необходимо разделить на 1000, поскольку на сайте Банка России за этот период они приводятся в неденоминированном виде. В-третьих, чтобы из всего массива данных оставить только необходимые для нас данные, а именно курс доллара на конец месяца, необходимо их отфильтровать с помощью опций ДАННЫЕ/ДОПОЛНИТЕЛЬНО/РАСШИРЕННЫЙ ФИЛЬТР.

Шаг 2. Построение графика в Excel

Выделим с помощью мышки столбец с ежемесячными данными (на конец месяца) по курсу пары «рубль — доллар» за период с июня 1992 г. (на конец июня из-за отсутствия на сайте Банка России более ранних данных возьмем курс доллара на 1 июля 1992 г.) по апрель 2010 г. и столбец с соответствующими обозначениями месяцев. Далее выбираем в панели инструментов кнопку Вставка (в Excel 2007) либо кнопку Мастер диаграмм (в Excel 1997—2003), в которой выбираем опцию «График» (рис. 1.1).

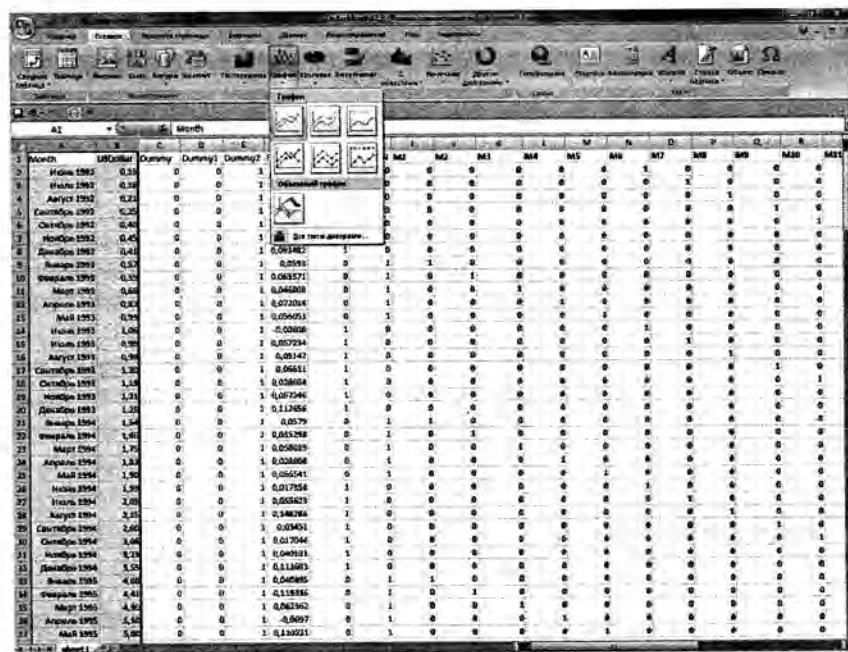


Рис. 1.1. Опция «График» — ВСТАВКА/МАСТЕР ДИАГРАММ

В результате у нас получился график (рис. 1.2), показывающий, что динамику колебаний ежемесячного курса доллара нельзя назвать стационарной. Судя по этому графику, можно прийти к выводу, что во временном ряде наблюдается тенденция к росту, а среднее значение курса доллара в разные периоды времени принимает различные значения. В частности, на графике хорошо видно, что во второй половине 1992 г. курс доллара хотя и систематически рос, но в целом был лишь немного выше нулевой отметки, в то время как к концу 1998 г. он превысил уровень 20 руб., а в 1999—2010 гг. курс американской валюты колебался в пределах от 24 до 35 руб.

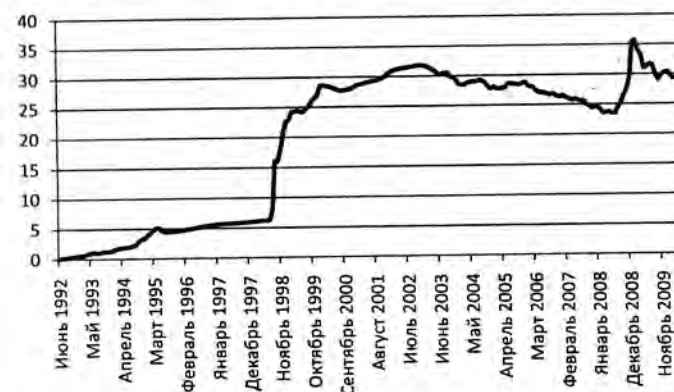


Рис. 1.2. Ежемесячный курс доллара США, руб.
Источник: данные Банка России.

Теперь построим аналогичный график в EViews. Однако прежде нам нужно научиться импортировать данные в эту программу из исходного экселевского файла. Умение выполнять эту процедуру потребуется для последующей работы в EViews. С этой целью следует ознакомиться с алгоритмом действий № 2.

Алгоритм действий № 2

Импорт данных и создание рабочего файла в EViews

Шаг 1. Подготовка данных в Excel для их последующего импорта в EViews

Прежде чем приступить к созданию диаграммы в EViews, нужно сначала импортировать в эту программу из Excel ежемесячные

данные о курсе доллара к рублю. При работе в более ранних версиях EViews импортируемые данные необходимо сохранять в формате Excel 5.0/95, поскольку при использовании других экселевских форматов в EViews появится сообщение об ошибке. Однако в последних версиях EViews можно загружать данные из экселевских файлов в любом формате, в том числе и из Excel 2007.

Импортируемые данные следует размещать в виде столбца в самой верхней строке экселевского листа слева. Например, заголовок первого столбца с данными должен быть помещен в ячейке B1, заголовок второго столбца с данными — в ячейке C1 и т.д., в то время как заголовок с соответствующими датами — в ячейке A1.

Заголовки столбцов следует обозначать латинскими буквами, поскольку англоязычная программа EViews не понимает кириллицу. В частности, столбец с ежемесячными данными по курсу доллара США мы решили обозначить как USDollar (поместили в ячейку B1), а заголовок (в ячейке A1) с названиями месяцев — Month. И последний важный момент: экселевский лист, на котором размещаются подготовленные к импорту в EViews данные, нужно также переименовать латинскими буквами. В нашем случае экселевский лист с импортируемыми данными назван sheet1.

Шаг 2. Создание рабочего файла в EViews

Чтобы создать рабочий файл, содержащий данные, с которыми мы собираемся работать, необходимо в главном меню EViews выбрать опции FILE/NEW/WORKFILE CREATE. В результате откроется следующее диалоговое мини-окно (рис. 1.3).

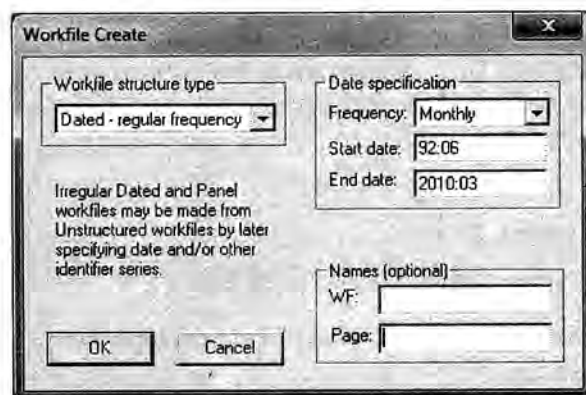


Рис. 1.3. Диалоговое окно Workfile Create

В этом диалоговом окне необходимо задать соответствующую информацию. Так, в мини-окне WORKFILE STRUCTURE TYPE (структура рабочего файла) мы задаем опцию DATED-REGULAR FREQUENCY (даты с определенной частотой). Соответственно в мини-окне FREQUENCY (частота данных) ставим опцию MONTHLY (ежемесячные данные), в START DATE (начальная дата) — 92:06 (июнь 1992 г.), в END DATE (конечная дата) — 2010:03 (март 2010 г.). В мини-окне END DATE дату года нужно обязательно давать четырехзначной, в то время как в START DATE она может быть двузначной. В результате у нас получится (рис. 1.4) неполный рабочий файл (WORKFILE): в нем будут отсутствовать данные, которые еще предстоит импортировать.

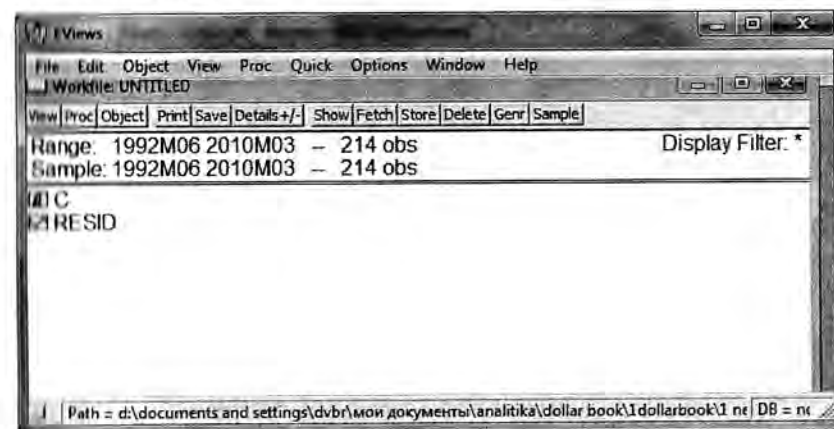


Рис. 1.4. Неполный рабочий файл Workfile

Шаг 3. Импорт данных в EViews

Перед импортом данных экселевский файл нужно обязательно закрыть, поскольку иначе появится сообщение об ошибке. При работе в последних версиях EViews в командной строке этой программы нужно воспользоваться опцией IMPORT/IMPORT FROM FILE. После открытия экселевского файла появится окошко EXCEL READ — Step 1 of 3 (рис. 1.5), в котором следует выбрать одну из двух опций: PREDEFINED RANGE (предопределенный диапазон) или CUSTOM RANGE (обычный диапазон). В случае выбора PREDEFINED RANGE в EViews загружаются уже выбранные программой данные, а если вы воспользуетесь опцией CUSTOM RANGE, то в этом случае можно самому выбрать требуемый диапазон данных, в том числе вне-

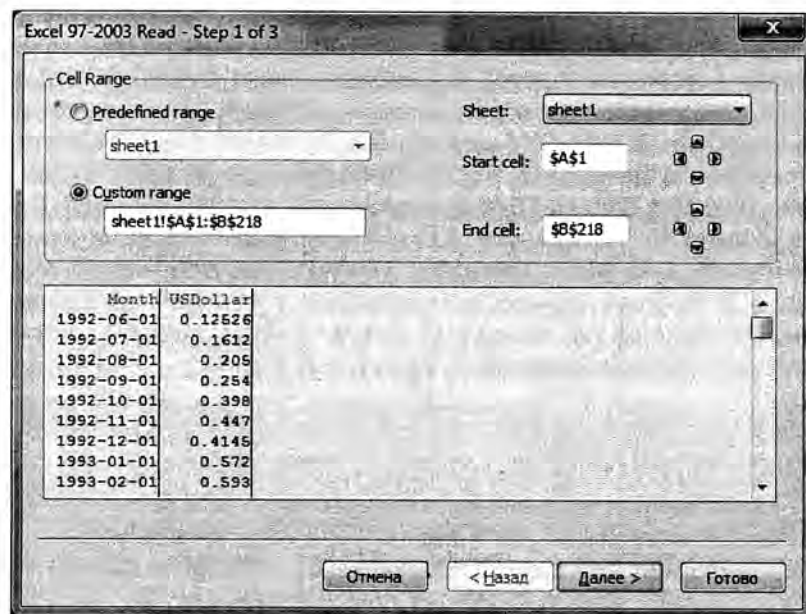


Рис. 1.5. Окно EXCEL READ — STEP 1 OF 3 (импорт листа в Excel)

сти необходимые правки в опции SHEET (лист), START CELL (начальная ячейка), END CELL (конечная ячейка).

Если вы работаете в более ранних версиях EViews, то при импорте данных в командной строке нужно выбрать опции FILE/IMPORT/READ TEXT-LOTUS-EXCEL. После этого появится новое диалоговое окно EXCEL SPREADSHEET IMPORT (импорт листа Excel). В открывшемся диалоговом окне (рис. 1.6) нужно отметить в мини-окне EXCEL5+ SHEET NAME название листа, которое у нас обозначено как sheet1. В мини-окне NAMES FOR SERIES OR NUMBER IF NAMED IN FILE (название для серии данных или номер серии данных, если у нее есть название в файле) поставим цифру 1, поскольку мы импортируем лишь одну серию данных, которую уже назвали USDollar. В остальных мини-окнах соответствующие опции в EViews устанавливаются по умолчанию. В частности, в мини-окне UPPER-LEFT DATA CELL (верхняя левая ячейка с данными) по умолчанию указывается ячейка B2.

Более подробно обо всех нюансах импорта данных из Excel в EViews можно прочитать, например, в книге М.Ю. Турунцевой «Анализ временных рядов»¹.

¹ Турунцева М.Ю. Анализ временных рядов. М.: МИЭФ ГУ-ВШЭ, 2003. С. 4—9.

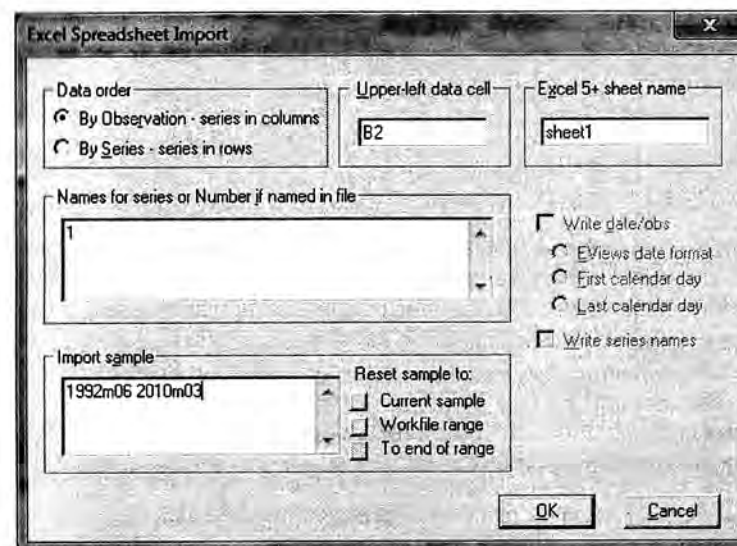


Рис. 1.6. Диалоговое окно EXCEL SPREADSHEET IMPORT (импорт листа Excel)

Поскольку мы уже создали рабочий файл в EViews, то построить график курса доллара, аналогичный экселевскому (см. рис. 1.2), не представляет особого труда. В рабочем файле (WORKFILE) EViews открываем файл USDollar, после чего используем для построения диаграммы в виде графика (LINE) опции VIEW/GRAPH/LINE (рис. 1.7).

В результате у нас получилась диаграмма (рис. 1.8), вполне аналогичная (если не считать различные типы форматирования, использованные при создании этих двух графиков) диаграмме на рис. 1.2, построенной в Excel. Чтобы сохранить полученную диаграмму в EViews на отдельном листе, следует нажать верхнюю кнопку FREEZE (окончательно принять).

Таким образом, построив соответствующие графики в EViews и Excel, мы выяснили, что временной ряд, характеризующий динамику ежемесячного курса доллара, является нестационарным, поскольку в нем наблюдается ярко выраженный тренд. Вместе с тем, как уже говорилось ранее, нестационарный временной ряд содержит не только тренд, но и случайную компоненту. Следовательно, чтобы сделать адекватный прогноз по курсу доллара, необходимо учесть как тренд, так и случайную компоненту, поскольку оба эти фактора существенно влияют на динамику валюты.

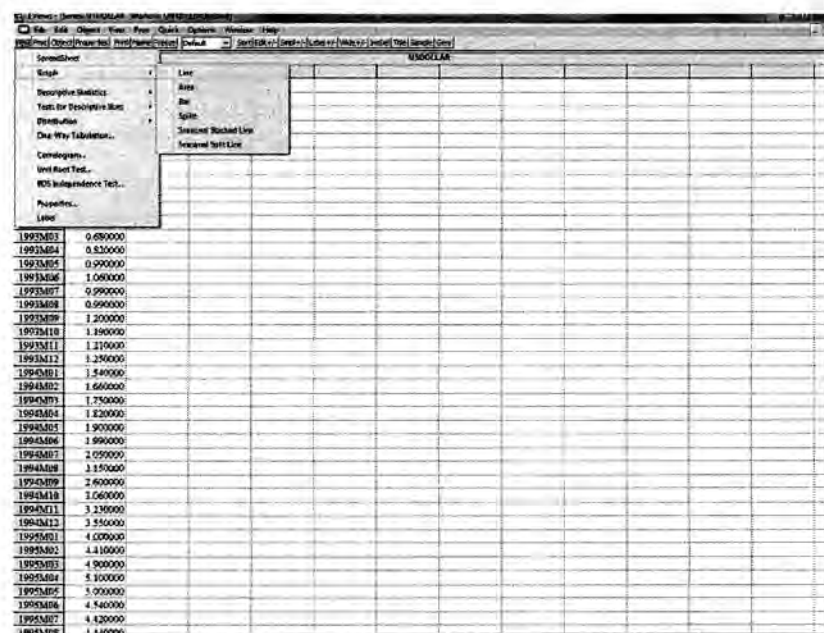


Рис. 1.7. Использование опции VIEW/GRAPH/LINE для построения в EViews диаграммы в виде графика (LINE)

Схематично наша дальнейшая работа, которой посвящены последующие главы этой книги, будет заключаться в следующем. Во-первых, нужно составить уравнение регрессии, с помощью которого можно будет делать прогнозы с необходимой точностью. Во-вторых, необходимо протестировать полученное уравнение регрессии (прогностическую модель) на его адекватность с точки зрения прогностических качеств. В-третьих, надо составить точечные прогнозы по курсу американской валюты, используя полученную математическую модель. В-четвертых, нужно удостовериться в приемлемой точности составленных точечных прогнозов. В-пятых, необходимо убедиться, что получившиеся в результате отклонения фактического курса доллара от его предсказанных (расчетных) значений представляют собой стационарный ряд. В-шестых, надо посмотреть, является ли распределение остатков нормальным, что позволит впоследствии составить интервальные прогнозы — с учетом диапазона отклонений точечных прогнозов от фактического курса доллара — с определенным уровнем надежности. В-седьмых, нужно проверить, соответствует ли точность

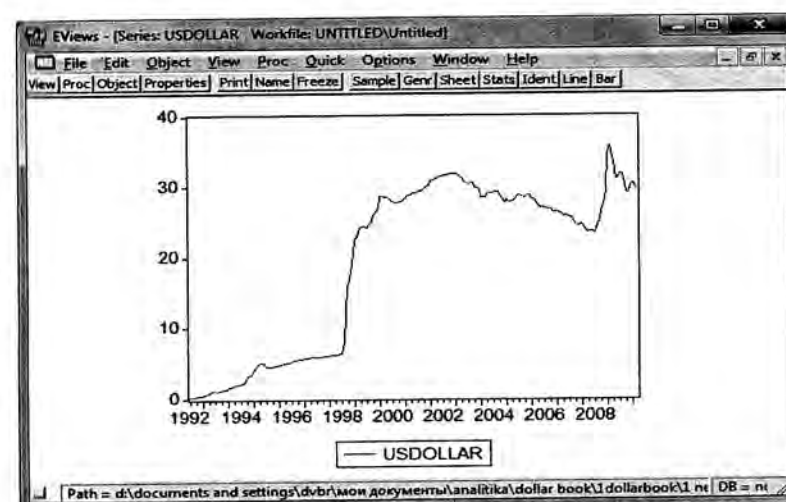


Рис. 1.8. График курса доллара, полученный в EViews
Источник: данные Банка России.

интервальных прогнозов заданному уровню надежности. В-восьмых, научиться применять полученную статистическую модель для составления рекомендуемых цен покупки и продажи валюты, используемых в качестве стоп-приказов при работе на валютном рынке. Выполнение всех этих процедур будет сопровождаться подробным рассказом о том, как их можно выполнить в Excel или EViews, что поможет читателям впоследствии самостоятельно решать эти задачи.

Контрольные вопросы и задания

1. Чем отличаются строго стационарные процессы от стационарных процессов в широком смысле?
2. Может ли стационарный процесс иметь тренд или какие-либо строго периодические колебания?
3. Чем нестационарный процесс отличается от стационарного? Может ли у нестационарного процесса быть тренд?
4. Если мы пришли к выводу о нестационарности временного ряда, что можно сказать об устойчивости его средней, дисперсии и автоковариации? Дайте определение средней, дисперсии и автоковариации.

ГЛАВА 2

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ И РЕШЕНИЕ УРАВНЕНИЯ РЕГРЕССИИ В EXCEL

2.1. Характеристика метода наименьших квадратов и его применение при прогнозировании курса доллара

Как выяснено в главе 1, динамика курса валют представляет собой временной ряд, имеющий не только тренд, но и случайную компоненту, поэтому в качестве метода оценки параметров прогностической модели, как правило, используется регрессионный анализ. Как известно, задачей регрессионного анализа является определение аналитического выражения (математической формулы), аппроксимирующего связь между зависимой переменной Y (ее называют также результативным признаком) и независимыми (их называют также факторными) переменными X_1, X_2, \dots, X_n . При этом форма связи результативного признака Y с факторами X_1, X_2, \dots, X_n либо с одним фактором X получила название *уравнения регрессии*. В качестве *метода аппроксимации* (приближения) в уравнении регрессии используется *метод наименьших квадратов* (МНК), который минимизирует сумму квадратов отклонений фактических значений Y от его предсказываемых значений, рассчитанных по определенной математической формуле. Причем решение уравнения регрессии относительно интересующих нас переменных y (курс доллара) и x (время или порядковый номер месяца), по сути, заключается в подборе прямой линии к совокупности пар данных, характеризующих динамику курса доллара и соответствующие порядковые номера месяцев. При этом линию, которая лучше всего подойдет к этим данным, выбирают так, чтобы сумма квадратов значений вертикальных отклонений зависимой переменной (фактического курса

доллара) от линии, рассчитанной по уравнению регрессии (предсказанный курс доллара), была минимальной.

Математические подробности оценки параметров уравнения регрессии методом наименьших квадратов

В самом общем виде формулу МНК можно представить следующим образом:

$$\sum (Y_i - Y_{\text{расч}})^2 \Rightarrow \min \sum e^2, \quad (2.1)$$

где Y_i и $Y_{\text{расч}}$ — фактические и расчетные значения зависимой (результативной) переменной для различных моментов времени;

$\min \sum e^2$ — минимальная сумма квадратов отклонений (остатков) фактических значений Y от его расчетных (предсказываемых) значений.

Поскольку $Y_{\text{расч}} = a + bX$,
где a — свободный член уравнения регрессии;
 b — коэффициент регрессии,

то уравнение (2.1) примет следующий вид:

$$\sum (Y_i - a - bX)^2 \Rightarrow \min \sum e^2. \quad (2.1.1)$$

Для отыскания параметров a и b , при которых функция $f(a, b)$ принимает минимальное значение, необходимо найти частные производные по каждому из параметров этой функции a и b и приравнять их к нулю. Если $\sum e^2$ обозначить через S , то в результате мы получим систему нормальных уравнений МНК для прямой:

$$\begin{aligned} \frac{dS}{da} &= -2 \sum Y + 2na + 2b \sum X = 0; \\ \frac{dS}{db} &= -2 \sum YX + 2a \sum X + 2b \sum X^2 = 0. \end{aligned} \quad (2.1.2)$$

Преобразовав систему уравнений (2.1.2), получим:

$$\begin{aligned} na + b \sum X &= \sum Y; \\ a \sum X + b \sum X^2 &= \sum XY. \end{aligned} \quad (2.1.3)$$

Решив систему уравнений (2.1.3) методом последовательного исключения переменных, найдем следующие оценки параметров:

$$a = \bar{Y} - b\bar{X}; \quad (2.1.4)$$

$$b = \frac{\bar{Y}\bar{X} - \bar{Y} \times \bar{X}}{(\bar{X}^2)_{\text{ср}} - \bar{X}^2}; \quad (2.1.5)$$

где \bar{Y} , \bar{X} , $(\bar{X}^2)_{\text{ср}}$ и $\bar{Y}\bar{X}$ — средние значения переменных X , Y , X^2 и их произведения XY .

С помощью оцененного таким образом уравнения регрессии можно предсказать, как в среднем изменится признак Y в результате роста факторов X_1, X_2, \dots, X_t (или одного фактора X).

В зависимости от того, какая математическая функция используется для прогнозирования результирующей переменной Y , различают линейную и нелинейную регрессию. При этом в основе *линейной регрессии* лежит уравнение линейного тренда, а в основе *нелинейной регрессии* — целое семейство уравнений нелинейных трендов (полиномиальный второй, третьей и прочих степеней, степенной, экспоненциальный и др.). В случае если результирующий признак Y зависит от одного фактора X , то такое уравнение регрессии называется *парным*, а если Y зависит от нескольких факторов X_1, X_2, \dots, X_t — то *уравнением множественной регрессии*.

Практически в любом учебнике по общей теории статистики и по эконометрике можно более подробно познакомиться со спецификой уравнений регрессии¹. Существуют формулы, по которым можно самостоятельно найти параметры как уравнения линейной регрессии, так и различных видов уравнений нелинейной регрессии. Однако с внедрением в широкую практику компьютеров и соответствующих компьютерных программ уже нет необходимости оценивать параметры уравнения регрессии вручную, тем более что это процесс довольно трудоемкий.

2.2. Решение уравнения регрессии в Excel с учетом фактора времени. Интерпретация и оценка значимости полученных параметров

Рассмотрим алгоритм решения уравнения регрессии с применением соответствующих вычислительных программ. При этом работу с уравнением регрессии в компьютерных программах можно разделить на три этапа.

На первом, подготовительном этапе необходимо определиться с набором факторов, которые необходимо включить в уравнение регрессии, а также с его аналитической формой, что в ряде случаев требует предварительной обработки данных. Например, в случае выбора степенного уравнения регрессии вместо исходных данных нужно взять их логарифмы.

Второй этап состоит из собственно решения уравнения регрессии и нахождения его параметров.

¹ См., например: Эконометрика : учебник / под ред. И.И. Елисеевой. 2-е изд., испр. и доп. М. : Финансы и статистика, 2006. С. 43—132.

На третьем этапе проводится оценка и тестирование общего качества уравнения регрессии, проверка статистической значимости каждого из коэффициентов регрессии, определяются их доверительные интервалы, а также принимается окончательное решение об адекватности или неадекватности полученного уравнения регрессии.

Как известно, одним из наиболее распространенных способов определения тренда в динамике курса валюты является построение его зависимости от фактора времени T . Так, если в качестве зависимой переменной Y мы возьмем ежемесячный курс доллара, а в качестве независимой переменной T — время (в данном случае порядковые номера месяцев начиная с июня 1992 г.), то у нас получится следующее уравнение парной линейной регрессии:

$$Y_{\text{расч}} = a + bT, \quad (2.2)$$

где a — свободный член уравнения регрессии;

b — линейный коэффициент регрессии, показывающий, как изменение величины независимой переменной (фактора) T в среднем способствует изменению зависимой переменной (результативного признака) Y ;

$Y_{\text{расч}}$ — расчетное значение результативного признака, вычисляемое по формуле 2.2.

Минимизируем сумму квадратов отклонений (остатков) $Y_{\text{факт}}$ от $Y_{\text{расч}}$, т.е. фактических значений курса доллара от его расчетных значений. В результате формулу МНК (2.1.1) для линейной регрессии можно представить в следующем виде:

$$\sum [Y_{\text{факт}} - (a + bT)^2] \Rightarrow \min \sum e^2. \quad (2.3)$$

Уравнение 2.3, в принципе, можно решить самостоятельно, если найти его параметры согласно формулам (2.1.4) и (2.1.5), но в целях ускорения этого процесса будем его решать с помощью Пакета анализа Excel. Кстати, желающие лучше усвоить суть МНК могут сначала самостоятельно в «ручном режиме» решить уравнение регрессии, а затем сверить свои результаты с теми, что мы получим в Excel.

Чтобы подготовить исходные данные к решению уравнения регрессии, разместим в Excel два столбца исходных данных. В первом столбце, который озаглавим Time, поместим порядковые номера месяцев, начиная с июня 1992 г. (с номером 1) и кончая апрелем 2010 г. (с номером 215). Во втором столбце, который озаглавим USDollar, поместим данные по курсу доллара на конец месяца, начиная с июня 1992 г. и заканчивая апрелем 2010 г.¹ Таким образом, столбец Time представляет

¹ Последние данные, имевшиеся у автора на тот момент, когда писались эти строки.

собой независимую переменную, которая в формуле (2.2) обозначена символом T , а столбец USDollar является зависимой переменной $Y_{\text{факт}}$. Далее переходим к решению уравнения регрессии в Пакете анализа Excel согласно алгоритму действий № 3.

Алгоритм действий № 3

Как решить уравнение регрессии в Excel

Шаг 1. Ввод в уравнение исходных данных

Сначала в Microsoft Excel 2007 в верхней панели инструментов выбирается опция ДАННЫЕ (в Microsoft Excel 1997—2003 нужно выбрать опцию СЕРВИС), потом в появившемся окне АНАЛИЗ ДАННЫХ — опция РЕГРЕССИЯ. После чего появляется новое окно РЕГРЕССИЯ (рис. 2.1), в котором в графе ВХОДНОЙ ИНТЕРВАЛ Y выделяем (с помощью мышки) столбец данных USDollar (ячейки $\$C\$1:\$C\216). Здесь же в графе ВХОДНОЙ ИНТЕРВАЛ X выделяем столбец данных Time (ячейки $\$B\$1:\$B\216), т.е. независимую переменную T из нашего уравнения регрессии (2.2).

Шаг 2. Дополнительные опции

Если бы мы хотели получить уравнение регрессии без свободного члена, который в формуле (2.2) обозначен символом a , то тогда нам следовало бы выбрать еще и опцию КОНСТАНТА-НОЛЬ. Но пока в использовании этой опции нет необходимости.

Опцию ОСТАТКИ следует выбирать тогда, когда есть необходимость, чтобы в выходных данных содержалась информация об отклонении расчетных Y от их фактических значений. При этом остатки находятся по формуле

$$\text{ОСТАТКИ} = Y_{\text{расч}} - Y_{\text{факт}} \quad (2.4)$$

где $Y_{\text{расч}}$ — расчетные;

$Y_{\text{факт}}$ — фактические значения результативного признака.

Опцию МЕТКИ применяют, чтобы переменные, включенные в уравнение регрессии, в ВЫВОДЕ ИТОГОВ были обозначены в виде заголовков соответствующих столбцов.

По умолчанию оценка в Excel параметров уравнения регрессии делается с 95%-ным уровнем надежности. Однако в случае необходимости в опции УРОВЕНЬ НАДЕЖНОСТИ можно поставить цифру 99, что означает задание для программы оценить коэффициенты

регрессии с 99%-ным уровнем надежности. В результате в ВЫВОДЕ ИТОГОВ мы получим данные, характеризующие как в целом уравнение регрессии, так и верхние и нижние интервальные оценки коэффициентов уравнения с 95%-ным и 99%-ным уровнями надежности. При 95%-ном уровне надежности существует риск, что в 5% случаях оценки коэффициентов уравнения регрессии могут оказаться статистически незначимыми, а при 99%-ном уровне надежности этот риск равен 1%.

Шаг 3. Вывод итогов

На заключительном этапе выбираем в параметрах вывода (окно РЕГРЕССИЯ) опцию ВЫХОДНОЙ ИНТЕРВАЛ, в которой указываем соответствующую ячейку Excel ($\$H\2), далее щелкаем по надписи ОК и получаем ВЫВОД ИТОГОВ (см. рис. 2.1, где можно увидеть все заданные нами параметры уравнения регрессии). В случае необходимости вывод итогов можно получить на отдельном листе (опция НОВЫЙ РАБОЧИЙ ЛИСТ) или в новой книге Excel (опция НОВАЯ РАБОЧАЯ КНИГА).

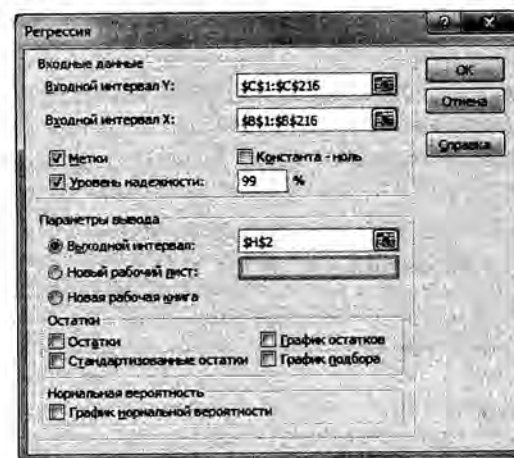


Рис. 2.1. Диалоговое окно РЕГРЕССИЯ для вывода итогов при решении уравнения регрессии

Результаты решения уравнения регрессии, которые в программе Excel выдаются в виде единой таблицы под заголовком ВЫВОД ИТОГОВ, у нас представлены в виде трех блоков (табл. 2.2—2.4). Так, в табл. 2.2 сгенерированы результаты по регрессионной статистике, в табл. 2.3 дается дисперсионный анализ, а в табл. 2.4 оценивается статистическая значимость коэффициентов регрессии.

Параметры, представленные в табл. 2.2, характеризуют уровень аппроксимации фактических данных, полученный с помощью уравнения регрессии. Так, параметр **МНОЖЕСТВЕННЫЙ R** обозначает коэффициент множественной корреляции R , который характеризует тесноту связи между результативным признаком Y и факторами переменных X_1, X_2, \dots, X_n . Этот коэффициент изменяется в пределах от 0 до 1, причем чем ближе к 1, тем теснее корреляционная связь между переменными, включенными в уравнение регрессии. Коэффициент множественной корреляции равен квадратному корню, извлеченному из коэффициента детерминации R^2 , который также приводится в регрессионной статистике. Коэффициент множественной корреляции R находят по формуле:

$$R = \sqrt{\frac{\sum (Y_{\text{расч}} - Y_{\text{ср}})^2}{\sum (Y_{\text{факт}} - Y_{\text{ср}})^2}}, \quad (2.5)$$

где $Y_{\text{факт}}$ — фактическое;

$Y_{\text{расч}}$ — расчетное (предсказанное по уравнению регрессии) значение результативного признака.

Зная величину коэффициента корреляции R , можно дать качественную оценку силы связи между зависимой и независимыми переменными, включенными в уравнение (2.5). С целью классификации силы связи обычно используют шкалу Чеддока (табл. 2.1).

Таблица 2.1

Шкала Чеддока для классификации силы связи

| Величина коэффициента множественной корреляции R | Оценка силы связи |
|--|-------------------|
| 0,1—0,3 | Слабая |
| 0,3—0,5 | Умеренная |
| 0,5—0,7 | Заметная |
| 0,7—0,9 | Высокая |
| 0,9—0,99 | Весьма высокая |

Если между переменными существует функциональная связь, то $R = 1$, а если корреляционная связь отсутствует, то $R = 0$. Поскольку в табл. 2.2 коэффициент множественной корреляции R равен 0,8456, то, согласно шкале Чеддока, связь между переменными, включенными в уравнение регрессии, можно считать высокой. Следует также заметить, что если коэффициент множественной корреляции меньше 0,7, то это означает, что величина коэффициента детерминации R^2 будет меньше 50%, а по-

тому регрессионные модели с таким коэффициентом детерминации не имеют большого практического значения.

Однако самым важным является другой параметр регрессионной статистики — **R-КВАДРАТ** (в табл. 2.2 он выделен шрифтом), обозначающий коэффициент детерминации R^2 . Коэффициент детерминации R^2 характеризует долю дисперсии результативного признака Y , объясняемую уравнением регрессии, в общей дисперсии результативного признака. Коэффициент детерминации R^2 находится по следующей формуле:

$$R^2 = \frac{D_{\text{объясн}}}{D_{\text{общ}}} = \frac{\sum (Y_{\text{расч}} - Y_{\text{ср}})^2}{\sum (Y_i - Y_{\text{ср}})^2}, \quad (2.6)$$

где Y_i — фактическое;

$Y_{\text{расч}}$ — расчетное (по уравнению регрессии) значение результативного признака (зависимой переменной);

$D_{\text{объясн}}$ — объясненная дисперсия;

$D_{\text{общ}}$ — общая дисперсия результативного признака Y .

Таблица 2.2

Регрессионная статистика

| ВЫВОД ИТОГОВ | |
|-------------------------------|----------------|
| Множественный R | 0,845642 |
| R-квадрат | 0,71511 |
| Нормированный R -квадрат | 0,713773 |
| Стандартная ошибка | 6,382421 |
| Наблюдения | 215 |

Коэффициент детерминации R^2 , как и коэффициент множественной корреляции R , изменяется в пределах от нуля до единицы. Если R^2 равен единице, то доля объясненной дисперсии составляет 100%, а следовательно, связь между зависимой переменной Y и независимыми переменными X_1, X_2, \dots, X_n носит функциональный характер. В том случае, когда R^2 равен нулю, какая-либо связь между переменными в этом уравнении регрессии отсутствует.

Величина коэффициента детерминации R^2 является одним из важнейших критериев при оценке качества уравнения регрессии. Так, при выборе из нескольких уравнений регрессии предпочтение (при прочих равных условиях) отдается тому, у которого коэффициент детерминации R^2 ближе к единице. И это вполне понятно: чем выше коэффи-

коэффициент детерминации уравнения регрессии, тем выше у него уровень аппроксимации и соответственно ниже доля необъясненной дисперсии. В нашем случае коэффициент детерминации $R^2 = 0,7151$, а потому можно сделать вывод, что в период с июня 1992 г. по апрель 2010 г. 71,51% ежемесячных колебаний курса доллара (зависимая переменная Y), согласно уравнению регрессии, объяснялись изменением порядкового номера месяца (независимая переменная T).

Другой параметр регрессионной статистики — **НОРМИРОВАННЫЙ R-КВАДРАТ**. Дело в том, что при добавлении в уравнение регрессии дополнительных факторов (независимых переменных) величина коэффициента детерминации R^2 соответственно растет. Поэтому чтобы сделать сравнения коэффициентов детерминации между уравнениями регрессии с разным числом факторов более сопоставимыми, используется *нормированный R^2* , величина которого корректируется в сторону уменьшения при добавлении в уравнение дополнительных факторов. В Пакете анализа Excel нормированный R^2 вычисляют по формуле:

$$R^2_{\text{норм}} = \frac{(n-1)R^2 - k}{n - k - 1}, \quad (2.7)$$

где n — количество наблюдений;

k — количество переменных в уравнении регрессии.

В нашем случае

$$R^2_{\text{норм}} = \frac{(215-1) \times 0,71511 - 1}{215 - 1 - 1} = 0,713773.$$

Еще один параметр регрессионной статистики — **СТАНДАРТНАЯ ОШИБКА**, или остаточное стандартное отклонение, которое можно найти по следующей формуле:

$$S_{\text{ост}} = \sqrt{\frac{\sum (Y_i - Y_{\text{расч}})^2}{n - (k + 1)}}, \quad (2.8)$$

где n — количество наблюдений;

k — количество переменных в уравнении регрессии.

НАБЛЮДЕНИЯ — этот параметр регрессионной статистики показывает число наблюдений n , равное 215 (т.е. числу месяцев с июня 1992 г. по апрель 2010 г., по которым у нас есть данные).

В таблице 2.3 дается дисперсионный анализ, т.е. анализ изменения результативного признака под воздействием включенных в уравнение регрессии факторов.

Таблица 2.3

Дисперсионный анализ

| | df | SS | MS | F | Значимость F |
|-----------|------|-----------|-----------|----------|----------------|
| Регрессия | 1 | 21 779,45 | 21 779,45 | 534,6578 | 5,4E-60 |
| Остаток | 213 | 8 676,619 | 40,7353 | | |
| Итого | 214 | 30 456,07 | | | |

Источник: расчеты автора.

При этом столбцы этой таблицы имеют следующую интерпретацию.

1. Столбец df (*degrees of freedom*) сообщает число степеней свободы.

Причем для строки РЕГРЕССИЯ число степеней свободы равно количеству факторов $k_{\text{факт}}$ включенных в уравнение регрессии. В нашем случае $df_{\text{регр}} = k = 1$.

Для строки ОСТАТОК число степеней свободы определяется числом наблюдений и количеством факторов, включенных в уравнении регрессии. При этом $df_{\text{ост}}$ находится по следующей формуле:

$$df_{\text{ост}} = n - (k + 1), \quad (2.9)$$

где n — число наблюдений;

k — количество факторов.

В нашем случае $df_{\text{ост}} = 215 - (1 + 1) = 213$.

Для строки ИТОГО число степеней свободы находится по следующей формуле:

$$df_{\text{итого}} = df_{\text{регр}} + df_{\text{ост}}. \quad (2.10)$$

В нашем случае $df_{\text{итого}} = 1 + 213 = 214$.

2. Столбец SS означает сумму квадратов отклонений.

Для строки РЕГРЕССИЯ этот столбец обозначает сумму квадратов отклонений рассчитанных (предсказанных) значений результативного признака от его среднего, рассчитанного по фактическим данным:

$$SS^2_{\text{регр}} = \sum (Y_{\text{расч}} - Y_{\text{ср}})^2. \quad (2.11)$$

Для строки ОСТАТОК столбец SS обозначает сумму квадратов отклонений фактических данных от их расчетных значений:

$$SS^2_{\text{ост}} = \sum (Y_i - Y_{\text{расч}})^2. \quad (2.12)$$

Для строки ИТОГО столбец SS обозначает сумму квадратов отклонений фактических данных от их среднего:

$$SS_{\text{итого}}^2 = \sum (Y_i - Y_{\text{ср}})^2. \quad (2.13)$$

$SS_{\text{итого}}^2$ можно также найти, сложив $SS_{\text{регр}}^2$ с $SS_{\text{ост}}^2$: $21\,779,45 + 8676,619 = 30\,456,07$.

3. Столбец MS означает дисперсию на одну степень свободы, которая находится по следующей формуле:

$$MS = \frac{SS}{df}. \quad (2.14)$$

Для строки РЕГРЕССИЯ — это факторная, или объясненная, дисперсия:

$$D_{\text{факт}} = MS_{\text{факт}} = 21\,779,45/1 = 21\,779,45.$$

Для строки ОСТАТОК — это остаточная дисперсия:

$$D_{\text{ост}} = MS_{\text{ост}} = 8676,619/213 = 40,7353.$$

4. В столбце F дается фактический F -критерий Фишера, который находится путем сопоставления факторной и остаточной дисперсии на одну степень свободы. При этом F -критерий Фишера рассчитывается по следующей формуле:

$$F_{\text{факт}} = \frac{MS(\text{Регрессия})}{MS(\text{Остаток})}. \quad (2.15)$$

Если нулевая гипотеза (об отсутствии связи между переменными, включенными в уравнение регрессии) справедлива, то факторная и остаточная дисперсия не отличаются друг от друга. Чтобы уравнение регрессии было признано значимым, требуется опровержение нулевой гипотезы, а для этого необходимо, чтобы факторная дисперсия превышала остаточную дисперсию в несколько раз. Статистиками разработаны соответствующие таблицы критических значений F -критерия при разных уровнях значимости нулевой гипотезы и различном числе степеней свободы. При этом следует иметь в виду, что табличное значение F -критерия — это максимальная величина отношения факторной дисперсии к остаточной дисперсии, которая может иметь место при случайном их расхождении для данного уровня вероятности наличия нулевой гипотезы. Если фактический (т.е. рассчитанный для этого уравнения регрессии) F -критерий больше его табличного значения, то нулевая гипотеза об отсутствии связи между результативным признаком и факторами отклоняется и делается вывод о существенности этой связи.

5. В столбце ЗНАЧИМОСТЬ F дается уровень значимости, который соответствует величине фактического F -критерия Фишера, вычисленного

для этого уравнения регрессии. В нашем случае значимость $F_{\text{факт}}$ практически равна нулю, т.е. $F_{\text{факт}}$ больше $F_{\text{табл}}$ (значения F -критерия Фишера при уровне значимости 0,05 или 5% можно найти в любом учебнике по статистике) при 1%-ном и 5%-ном уровне значимости. Отсюда можно сделать вывод о статистической значимости уравнения регрессии, поскольку связь между включенными в него факторами в этом случае доказана.

В тех случаях, когда значимость F бывает больше, например, 0,01, но меньше 0,05, то тогда делается вывод, что $F_{\text{факт}}$ меньше $F_{\text{табл}}$ при 1%-ном уровне значимости, но больше $F_{\text{табл}}$ при 5%-ном уровне значимости. Следовательно, в этой ситуации нулевая гипотеза об отсутствии связи между результативным признаком и факторами, включенными в уравнение регрессии, на 1%-ном уровне значимости не отклоняется, но отклоняется на 5%-ном уровне значимости. Таким образом, в этом случае каждый исследователь должен сам решить, считать ли 5%-ный уровень значимости F -критерия достаточным для того, чтобы сделать вывод о статистической значимости уравнения регрессии. При этом следует иметь в виду, что если значимость F -критерия выше 0,05, т.е. $F_{\text{факт}}$ меньше $F_{\text{табл}}$ при 5%-ном уровне значимости, то в этой ситуации уравнение регрессии, как правило, считается статистически незначимым.

В таблице 2.4 сгенерированы коэффициенты уравнения регрессии и оценки их статистической значимости.

Таблица 2.4

Коэффициенты уравнения регрессии и их статистическая значимость

| | Коэффициент | Стандартная ошибка | T-статистика | P-значение | Нижние 95% | Верхние 95% | Нижние 99,0% | Верхние 99,0% |
|---------------|-------------|--------------------|--------------|------------|------------|-------------|--------------|---------------|
| У пересечение | 1,995805 | 0,873601 | 2,284573 | 0,023323 | 0,273794 | 3,717815 | -0,27478 | 4,266387 |
| Time | 0,162166 | 0,007013 | 23,12267 | 5,4E-60 | 0,148342 | 0,175991 | 0,143938 | 0,180395 |

Источник: расчеты автора.

1. В столбце КОЭФФИЦИЕНТЫ представлены коэффициенты уравнения регрессии. На пересечении этого столбца со строкой У-ПЕРЕСЕЧЕНИЕ дан свободный член, который в формуле линейного уравнения регрессии (2.2) обозначен символом $a = 1,995805$.

Во второй строке этого столбца, обозначенной как TIME (независимая переменная — порядковый номер месяца), сгенерирован коэффициент уравнения регрессии, который в формуле (2.2) представлен символом $b = 0,162166$.

Таким образом, данные, представленные в столбце КОЭФФИЦИЕНТЫ, дают нам возможность составить путем подстановки соответ-

ствующих цифр в формулу (2.2) следующее уравнение линейной парной регрессии:

$$Y = 0,1622X + 1,9958,$$

где независимая переменная X означает порядковый номер месяца (июнь 1992 г. — 1, а апрель 2010 г. — 215);

зависимая переменная Y — ежемесячное значение курса доллара.

При этом экономическая интерпретация этого линейного уравнения следующая: в период с июня 1992 г. по апрель 2010 г. курс доллара к рублю ежемесячно рос со средней скоростью 16,22 коп. при исходном уровне временного ряда в размере 1 руб. 99,58 коп. В свою очередь геометрическая интерпретация этого линейного уравнения следующая: свободный член уравнения 1,9958 показывает точку пересечения линии тренда с осью Y , а коэффициент уравнения 0,1622 равен углу наклона линии тренда к оси X (см. рис. 2.5).

2. В столбце СТАНДАРТНАЯ ОШИБКА сгенерированы стандартные ошибки свободного члена и коэффициента регрессии, значения которых даны во втором столбце табл. 2.4. При этом стандартная ошибка свободного члена уравнения регрессии находится по следующей формуле:

$$m_0 = \sqrt{MS_{\text{ост}} \frac{\sum X^2}{n \sum (X - \bar{X})^2}}, \quad (2.16)$$

где $MS_{\text{ост}} = D_{\text{ост}}$ — остаточная дисперсия, приходящаяся на одну степень свободы.

Для нашего случая стандартная ошибка свободного члена уравнения регрессии равна

$$m_0 = \sqrt{40,7353 \times \frac{3\,335\,940}{215 \times 828\,180}} = 0,873601.$$

В свою очередь стандартная ошибка коэффициента регрессии оценивается по следующей формуле:

$$m_b = \sqrt{\frac{MS_{\text{ост}}}{\sum (X - \bar{X})^2}}. \quad (2.17)$$

Для нашего случая стандартная ошибка коэффициента регрессии имеет следующее значение:

$$m_b = \sqrt{\frac{40,7353}{828\,180}} = 0,007013.$$

3. В столбце t -СТАТИСТИКА даны расчетные значения t -критерия. При этом для свободного члена t -статистика вычисляется по формуле

$$t_{\text{ст}} = \frac{a}{m_a}, \quad (2.18)$$

где a — свободный член уравнения.

В нашем случае t -статистика находится следующим образом:

$$t_{\text{ст}} = \frac{1,995805}{0,873601} = 2,284573.$$

Для коэффициента регрессии t -статистика рассчитывается по формуле

$$t_{\text{ст}} = \frac{b}{m_b}, \quad (2.19)$$

где b — коэффициент регрессии.

Тогда t -статистика находится следующим образом:

$$t_{\text{ст}} = \frac{0,162166}{0,007013} = 23,12267.$$

4. В столбце P -ЗНАЧЕНИЕ сгенерированы уровни значимости, соответствующие значениям t -статистики.

В Excel P -значение находится с помощью следующей функции:

$$\text{СТБЮДРАСП} (X = t_{\text{ст}}; df = n - k - 1; \text{хвосты} = 2),$$

где в опции X дается t -статистика, для которой нужно вычислить двустороннее распределение;

в опции df — число степеней свободы;

в опции хвосты — цифра 2 для двустороннего распределения.

Для свободного члена уравнения эта функция приобретает следующий вид:

$$\text{СТБЮДРАСП} (2,284573; 215 - 1 - 1 = 213; 2) = 0,023323.$$

Следовательно, P -значение свободного члена уравнения показывает, что этот коэффициент значим лишь при 5%-ном уровне значимости, но не при 1%-ном уровне значимости.

Для коэффициента регрессии R -значение в Excel находится следующим образом¹:

$$\text{СТЮДРАСП} (23,12267; 215 - 1 - 1 = 213; 2) = 5,4E - 60 = 0,0.$$

Следовательно, R -значение коэффициента регрессии показывает, что этот коэффициент значим не только при 5%-ном уровне значимости, но и при 1%-ном уровне значимости.

5. Столбцы НИЖНИЕ 95% и ВЕРХНИЕ 95% показывают соответственно нижние и верхние интервалы значений коэффициентов при 95%-ном уровне значимости. Для расчета доверительных интервалов сначала устанавливается критическое значение t -критерия, которое в Excel находится с помощью функции

$$\text{СТЮДРАСПОБР} (\alpha = 0,05; df = n - k - 1);$$

где в опции α — величина риска, при котором коэффициент регрессии (или свободный член) может оказаться за рамками установленных доверительных интервалов;

в опции df — число степеней свободы.

Таким образом, для 95%-ного уровня надежности t -критерий = $\text{СТЮДРАСПОБР} (\alpha = 0,05; df = 215 - 1 - 1) = 1,9712$.

Далее для свободного члена уравнения находим:

1. Значение столбца НИЖНИЕ 95% = $\text{КОЭФФИЦИЕНТ} - \text{СТАНДАРТНАЯ ОШИБКА} \times t\text{-критерий} = 1,995805 - (0,873601 \times 1,9712) = 0,273794$.

2. Значение столбца ВЕРХНИЕ 95% = $\text{КОЭФФИЦИЕНТ} + \text{СТАНДАРТНАЯ ОШИБКА} \times t\text{-критерий} = 1,995805 + (0,873601 \times 1,9712) = 3,717815$.

Для коэффициента регрессии TIME находим:

1. Значение столбца НИЖНИЕ 95% = $\text{КОЭФФИЦИЕНТ} - \text{СТАНДАРТНАЯ ОШИБКА} \times t\text{-критерий} = 0,162166 - (0,007013 \times 1,9712) = 0,148342$.

2. Значение столбца ВЕРХНИЕ 95% = $\text{КОЭФФИЦИЕНТ} + \text{СТАНДАРТНАЯ ОШИБКА} \times t\text{-критерий} = 0,162166 + (0,007013 \times 1,9712) = 0,175991$.

¹ Здесь и далее необходимо учитывать, что при вычислении R -значения коэффициент регрессии компьютерная программа может использовать число в экспоненциальном формате, который обычно применяется для обозначения очень больших или очень малых чисел. Часть числа заменяется символами $E + n$ (для очень больших чисел) и $E - n$ (для очень малых чисел), где E (экспонента) означает умножение предшествующего числа на 10 в степени n . Если, например, число $5,4E - 60$ перевести в числовой формат, то оно будет равно 0,00...

6. Столбцы НИЖНИЕ 99% и ВЕРХНИЕ 99% показывают соответственно нижние и верхние интервалы значений коэффициентов при 99%-ном уровне значимости. При этом значения столбца НИЖНИЕ 99% и ВЕРХНИЕ 99% находятся аналогичным образом, как и значения столбцов НИЖНИЕ 95% и ВЕРХНИЕ 95%.

Единственное отличие — это расчет t -критерия для 99%-ного уровня надежности. При этом t -критерий = $\text{СТЮДРАСПОБР} (\alpha = 0,01; df = 215 - 1 - 1) = 3,3368$. Найденный t -критерий используют при нахождении 99% доверительных интервалов для свободного члена и коэффициента регрессии. Правда, со свободным членом уравнения у нас возникает довольно серьезная проблема. Дело в том, что при 99%-ном уровне надежности у свободного члена уравнения при переходе от столбца НИЖНИЕ 99% к столбцу ВЕРХНИЕ 99% происходит смена знака от минуса к плюсу. Вполне очевидно, что в практических расчетах столь неоднозначно изменяющийся свободный член уравнения (он может быть как положительным, так и отрицательным, а также равным нулю) невозможно использовать. Поэтому для 99%-ного уровня надежности свободный член уравнения считается статистически незначимым, в то время как для 95%-ного уровня надежности его можно считать статистически значимым, поскольку в последнем случае при переходе от столбца НИЖНИЕ 95% к столбцу ВЕРХНИЕ 95% не происходит смена знака от минуса к плюсу.

Суммируя сказанное, приведем краткий алгоритм принятия решения о статистической значимости уравнения регрессии на основе ВЫВОДА ИТОГОВ в Excel.

Алгоритм действий № 4

Оценка статистической значимости уравнения регрессии и его коэффициентов

Шаг 1. Принятие решения о значимости уравнения регрессии

1.1. Чем ближе R -квадрат к единице, тем лучше. Это дает важный критерий для выбора одного из нескольких уравнений регрессии.

1.2. Значимость F при 95%-ном уровне надежности должна быть меньше 0,05; при 99%-ном должна быть меньше 0,01.

Шаг 2. Принятие решения о значимости коэффициентов уравнения регрессии

2.1. R -значение должно быть меньше 0,05 при 95%-ном уровне надежности; при 99%-ном R -значение должно быть меньше 0,01.

Таблица 2.5
Вывод итогов и принятие решения о статистической значимости уравнения регрессии и значимости его коэффициентов

| ВЫВОД ИТОГОВ | | | SS | MS | F | Значимость F | Верхние 95% | Нижние 99,0% | Верхние 99,0% |
|--------------------------|----------------|-----|--------------------|--------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Регрессионная статистика | | | | | | | | | |
| Множественный R | 0,845642 | | | | | | | | |
| R-квадрат | 0,71511 | | | | | | | | |
| Нормированный R-квадрат | 0,713773 | | | | | | | | |
| Стандартная ошибка | 6,382421 | | | | | | | | |
| Наблюдения | 215 | | | | | | | | |
| Дисперсионный анализ | | df | | | | | | | |
| Регрессия | | 1 | 21779,45 | 21779,45 | 534,6578 | 5,4E -60 | | | |
| Остаток | | 213 | 8676,619 | 40,7353 | | | | | |
| Итого | | 214 | 30456,07 | | | | | | |
| | Коэффициент | | Стандартная ошибка | t-статистика | P-значение | Нижние 95% | Верхние 95% | Нижние 99,0% | Верхние 99,0% |
| У-пересечение | 1,995805 | | 0,873601 | 2,284573 | 0,023323 | 0,273794 | 3,717815 | -0,27478 | 4,266387 |
| Time | 0,162166 | | 0,007013 | 23,12267 | 5,4E -60 | 0,148342 | 0,175991 | 0,143938 | 0,180395 |

2.2. Коэффициенты регрессии и свободный член уравнения при переходе от столбцов НИЖНИЕ и ВЕРХНИЕ (при заданном уровне надежности) не должны менять свой знак. Если смена знака происходит, то коэффициенты регрессии и свободный член уравнения признаются статистически незначимыми.

Исходя из этого краткого алгоритма мы отметили жирным шрифтом в ВЫВОДЕ ИТОГОВ (табл. 2.5) именно те пункты, на которые следует обратить внимание. При этом те пункты, которые не являются статистически значимыми при определенном уровне надежности, мы не только выделили жирным шрифтом, но еще и подчеркнули.

Таким образом, взяв за основу данные из табл. 2.5 и действуя по алгоритму № 4, мы дадим ответы на все его четыре пункта.

1.1. Поскольку коэффициент детерминации R^2 для этого уравнения регрессии оказался равен 0,71511, то отсюда можно сделать вывод, что это уравнение в 71,51% случаях в состоянии объяснить ежемесячные колебания курса доллара.

1.2. Значимость F равна 5,4E -60 или нулю, а следовательно, уравнение регрессии статистически значимо как при 95%-ном уровне надежности, так и при 99%-ном уровне надежности.

2.1. P -значение для свободного члена уравнения равно 0,023323, а следовательно, этот коэффициент статистически значим лишь при 95%-ном уровне надежности, но незначим при 99%-ном уровне надежности, поскольку он больше 0,01. P -значение для коэффициента регрессии равно нулю, а следовательно, этот коэффициент статистически значим как при 95%-ном уровне надежности, так и при 99%-ном уровне надежности.

2.2. Свободный член (константа) уравнения при переходе от столбца НИЖНИЕ 99,0% к столбцу ВЕРХНИЕ 99,0% меняет знак с минуса на плюс, а потому статистически незначим при 99%-ном уровне надежности. При 95%-ном уровне надежности смены знаков не происходит, а потому свободный член уравнения при этом уровне надежности статистически значим. Коэффициент регрессии статистически значим как при 95%, так и при 99%-ном уровне надежности, поскольку и в том, и в другом случае смены знака у этого коэффициента не происходит. Следовательно, на основании табл. 2.5 можно сделать вывод, что в целом уравнение регрессии и все его коэффициенты статистически значимы при 95%-ном уровне надежности.

Как мы уже говорили ранее, уравнение регрессии в отличие от обычных уравнений, оценивающих функциональную, т.е. жестко детерминированную связь между переменными, дает прогноз зави-

симой переменной с учетом воздействия случайного фактора, поэтому фактические значения результативного признака практически всегда отличаются от его расчетных (теоретических) значений. При этом случайная компонента (остаток) находится следующим образом.

Сначала находится прогнозируемый курс доллара, например, на апрель 2010 г. С учетом того, что порядковый номер апреля 2010 г. равен 215 (июнь 1992 г. = 1), на этот месяц может быть предсказан следующий курс доллара:

$$Y_{\text{расч}} = 0,1622 \times 215 + 1,9958 = 36,8616;$$

$$E = Y_{\text{факт}} - Y_{\text{расч}} = -7,573.$$

Следовательно, прогноз, сделанный по уравнению регрессии, в апреле 2010 г. оказался выше фактического курса доллара на 7 руб. 57,3 коп. Вполне очевидно, что это слишком большая величина отклонения, чтобы исследуемое уравнение регрессии можно было бы использовать для прогноза валютного курса. В свою очередь чем ближе теоретические значения подходят к фактическим данным, тем лучше качество прогностической модели. Поскольку разница между фактическим и прогнозируемым значениями курса доллара ($Y_{\text{факт}} - Y_{\text{расч}}$) может быть величиной как положительной, так и отрицательной, то ошибку аппроксимации (подгонки модели к фактическим данным) следует определять как в абсолютных цифрах по модулю, так и в процентах по модулю.

При этом среднюю абсолютную ошибку по модулю находят по следующей формуле:

$$A_{\text{абс}} = \left| \frac{\sum (Y_{\text{факт}} - Y_{\text{расч}})}{n} \right|. \quad (2.20)$$

Для нашего уравнения регрессии средняя абсолютная ошибка по формуле (2.20) будет равна

$$A_{\text{абс}} = \left| \frac{1208,967}{215} \right| = 5,623.$$

Иначе говоря, прогноз по этой статистической модели в среднем по каждому наблюдению отклонялся от фактического значения курса доллара на 5 руб. 62,3 коп. по модулю.

Среднюю относительную ошибку по модулю в процентах вычисляют по следующей формуле:

$$A_{\text{отн}} = \left| \frac{\sum [(Y_{\text{факт}} - Y_{\text{расч}}) \times 100]}{n} \right|. \quad (2.21)$$

При этом средняя относительная ошибка по модулю в процентах имеет следующее значение:

$$A_{\text{отн}} = \left| \frac{8381,225}{215} \right| = 38,982.$$

Следовательно, прогноз по этой статистической модели в среднем по каждому наблюдению отклонялся от фактического значения курса доллара на 38,98%. В то время как о хорошем качестве уравнения регрессии можно говорить лишь в том случае, если средняя относительная ошибка по модулю составляет не более 5–7%¹.

Чтобы окончательно убедиться в непригодности для прогноза этого уравнения регрессии, построим табл. 2.6, в которой дадим прогнозы и фактический курс доллара за период с января 2009 г. по апрель 2010 г.

Таблица 2.6

Прогноз, фактический курс доллара и остатки с января 2009 г. по апрель 2010 г.

| Месяц | Фактический курс доллара | Прогноз | Остаток |
|------------------|--------------------------|---------|---------|
| Январь 2009 г. | 35,4146 | 34,4291 | 0,9855 |
| Февраль 2009 г. | 35,7205 | 34,5913 | 1,1292 |
| Март 2009 г. | 34,0134 | 34,7534 | –0,7400 |
| Апрель 2009 г. | 33,2491 | 34,9156 | –1,6665 |
| Май 2009 г. | 30,9843 | 35,0778 | –4,0935 |
| Июнь 2009 г. | 31,2904 | 35,2399 | –3,9495 |
| Июль 2009 г. | 31,7555 | 35,4021 | –3,6466 |
| Август 2009 г. | 31,5687 | 35,5643 | –3,9956 |
| Сентябрь 2009 г. | 30,0922 | 35,7264 | –5,6342 |
| Октябрь 2009 г. | 29,0488 | 35,8886 | –6,8398 |
| Ноябрь 2009 г. | 29,8179 | 36,0508 | –6,2329 |
| Декабрь 2009 г. | 30,2442 | 36,2129 | –5,9687 |
| Январь 2010 г. | 30,4312 | 36,3751 | –5,9439 |
| Февраль 2010 г. | 29,9484 | 36,5373 | –6,5889 |
| Март 2010 г. | 29,3638 | 36,6994 | –7,3356 |
| Апрель 2010 г. | 29,2886 | 36,8616 | –7,5730 |

Источник: расчеты автора по данным Банка России.

¹ См.: Эконометрика. С. 107.

Судя по табл. 2.6, с января 2009 г. по апрель 2010 г. отклонения от прогноза (остатки), сделанного по уравнению регрессии $Y_{расч} = 0,1622 \times 215 + 1,9958$, колебались в диапазоне от 98,5 коп. до 7 руб. 57,3 коп. что свидетельствует о невысокой точности этой прогностической модели. Более того, если построить график остатков по линейной прогностической модели, то легко обнаружить, что на нем имеется несколько локальных трендов (рис. 2.2). А это признак — как мы об этом уже говорили — нестационарности полученных остатков.

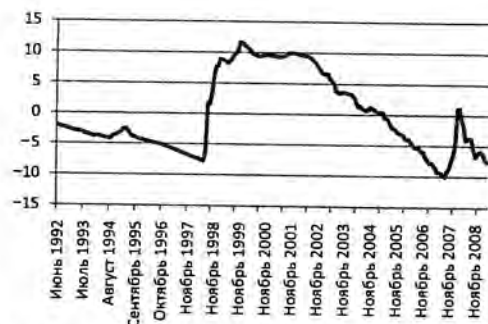


Рис. 2.2. Нестационарность остатков, полученных в линейной прогностической модели, руб.

Источник: расчеты автора по данным Банка России.

2.3. Решение уравнений регрессии в Excel графическим способом

Попробуем повысить точность нашего прогноза, используя алгоритм действий № 1 «Как строить диаграммы в Microsoft Excel». С этой целью обведем с помощью мышки столбец с ежемесячными данными (на конец месяца) по курсу пары «рубль — доллар» за период с июня 1992 г. по апрель 2010 г. и столбец с соответствующими обозначениями месяцев. Выбрав опцию ГРАФИК, строим соответствующую диаграмму, а затем щелкаем с помощью мышки по линии графика и выбираем в появившемся окне опцию ДОБАВИТЬ ЛИНИЮ ТРЕНДА (рис. 2.3).

Далее появляется диалоговое мини-окно ФОРМАТ ЛИНИИ ТРЕНДА, в котором мы можем выбрать соответствующие ПАРАМЕТРЫ ЛИНИИ ТРЕНДА (рис. 2.4), необходимые для построения

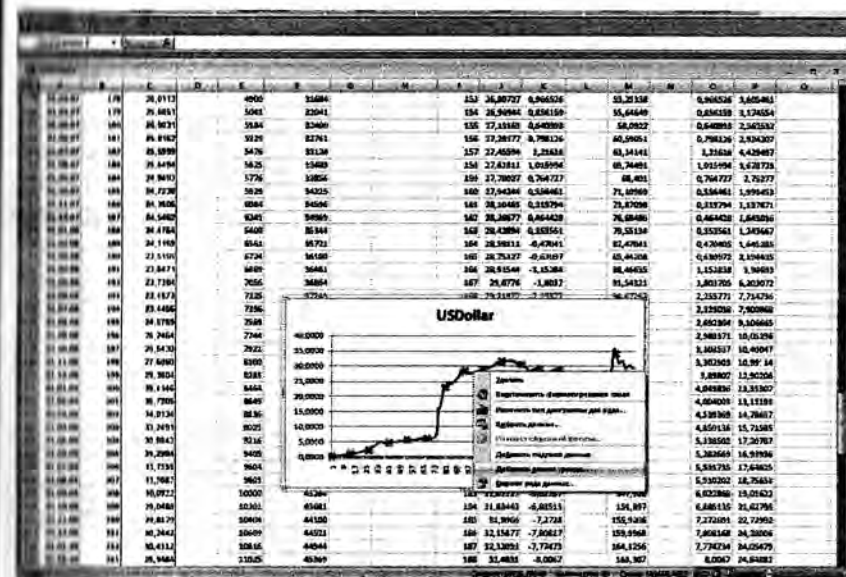


Рис. 2.3. Построение линии тренда на основе графика изменений курса доллара

прогностических моделей. При этом воспользуемся всеми имеющимися в Excel форматами тренда за одним-единственным исключением: из полиномиальных трендов возьмем тренды не выше третьей степени. В научной литературе обычно не рекомендуют использовать для аппроксимации фактических данных более сложные полиномы, поскольку они плохо поддаются интерпретации и, несмотря на высокий коэффициент детерминации (по включенной в статистическую модель базе данных), обладают низкой прогностической ценностью.

Сначала построим самый простой линейный тренд. С этой целью выберем в окне ФОРМАТ ЛИНИИ ТРЕНДА в опции ПАРАМЕТРЫ ЛИНИИ ТРЕНДА формат ЛИНЕЙНАЯ. При этом поставим галочку в опциях ПОКАЗЫВАТЬ УРАВНЕНИЕ НА ДИАГРАММЕ, ПОМЕСТИТЬ НА ДИАГРАММУ ВЕЛИЧИНУ ДОСТОВЕРНОСТИ АППРОКСИМАЦИИ (R^2). В результате получим диаграмму (рис. 2.5), показывающую линейный тренд, т.е. линейную зависимость роста курса доллара от времени (порядковый номер 1 — июнь 1992 г.).

¹ Знак R^2 используется в качестве обозначения степени числа, т.е. R^2 равно R^2 .

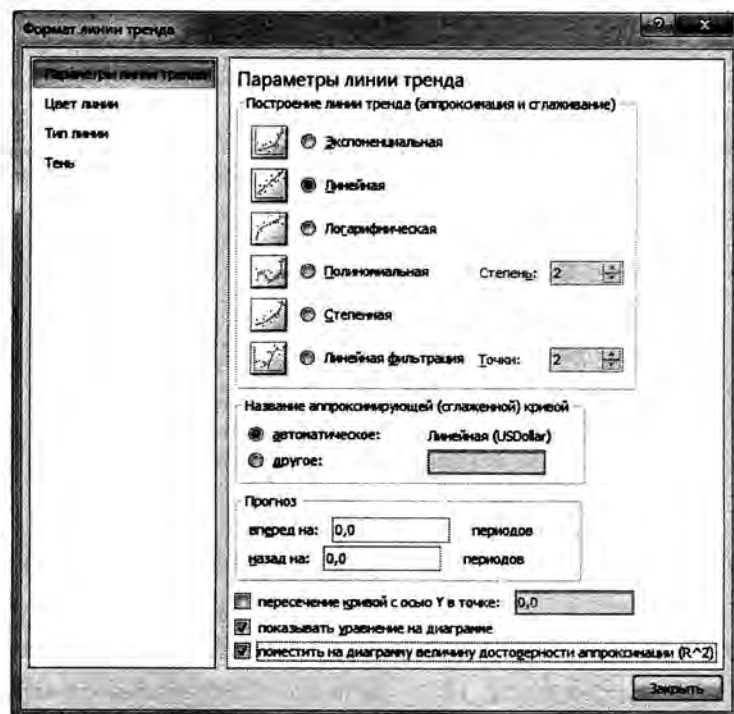


Рис. 2.4. Определение характера и параметров линии тренда

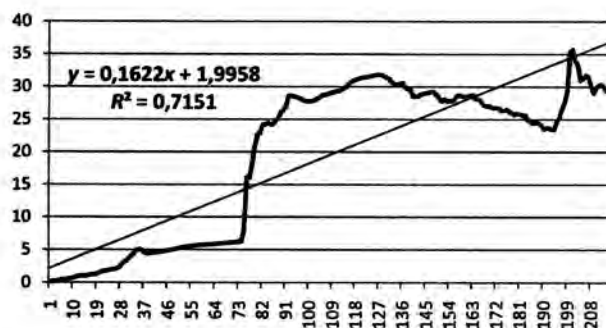


Рис. 2.5. Ежемесячный курс доллара и его линейный тренд: по оси абсцисс вместо названий месяцев даны их порядковые номера (июнь 1992 г. — 1, июль 1992 г. — 2, ..., апрель 2010 г. — 215)

Источник: расчеты автора по данным Банка России.

Поочередно задавая различные параметры тренда и сравнивая коэффициенты детерминации, составим табл. 2.7, в которой разместим по мере роста коэффициента детерминации прогностические модели с различным форматом тренда. Наиболее высокий коэффициент детерминации соответствует уравнению регрессии, полученному путем аппроксимации по степенному тренду. В этом случае R^2 оказался равен 0,919136, т.е. это уравнение регрессии объясняет 91,91% всех ежемесячных колебаний курса доллара. Соответственно доля случайной компоненты оказалась равна: $100\% - 91,91\% = 8,09\%$.

Таблица 2.7

Параметры тренда и величина коэффициента детерминации R^2

| Линия тренда | Формула | Коэффициент детерминации R^2 |
|----------------------------|---|--------------------------------|
| Экспоненциальная | $y = 1,9796007e^{0,0168591x}$ | 0,6734987 |
| Линейная | $y = 0,1622x + 1,9958$ | 0,7151 |
| Логарифмическая | $y = 10,408\ln(x) - 26,154$ | 0,6966 |
| Полиномиальная 2-й степени | $y = -0,0013x^2 + 0,433x - 7,7978$ | 0,8468 |
| Полиномиальная 3-й степени | $y = -0,0000073x^3 + 0,0011247x^2 + 0,2269865x - 4,0473597$ | 0,8602159 |
| Степенная | $y = 0,0443609x^{1,2807295}$ | 0,919136 |

Источник: расчеты автора.

Чтобы правильно интерпретировать уравнения регрессии, полученные графическим способом, необходимо иметь в виду, что в процессе построения тренда программа Excel автоматически задает в качестве зависимой переменной y ежемесячный курс доллара, а в качестве независимой x — порядковый номер месяца. Например, экономическая интерпретация уравнения регрессии со степенной функцией $y = 0,0443609x^{1,2807295}$ следующая: курс доллара в период с июня 1992 г. по апрель 2010 г. ежемесячно рос со средней скоростью 1,28% при исходном уровне 4,44 коп.¹

Как мы уже убедились, графический способ решения уравнения регрессии в программе Excel позволяет довольно существенно эконо-

¹ В книге все стоимостные выражения указываются в денонмированных единицах. В январе 1998 г. в России была проведена денонминация (уменьшение номинала) рубля, в результате которой его стоимость уменьшилась в 1000 раз. В июне 1992 г. доллар стоил 44,4 руб. и в дальнейшем продолжал быстрый рост. Однако для осуществления математических расчетов необходимо пользоваться едиными масштабами измерения стоимости, поэтому можно сказать, что в этот момент доллар стоил 4,44 коп. в копейках 1998 г., а к апрелю 2010 г. его цена превышала 30 руб.

мать время. Однако у этого способа есть и один весьма существенный недостаток, обусловленный тем, что при этом не проводится оценка статистической значимости как в целом уравнения регрессии, так и его коэффициентов.

Таким образом, графический способ решения уравнения регрессии целесообразно использовать на этапе предварительного отбора уравнений регрессии, имеющих наиболее высокий коэффициент детерминации. После отбора уравнения регрессии с высоким коэффициентом детерминации в Excel его нужно решить, используя в Пакете анализа опцию РЕГРЕССИЯ (см. алгоритм действий № 3). Однако решение уравнения регрессии, аппроксимирующего фактические данные степенным трендом, имеет определенную специфику. В отличие от линейного тренда уравнение регрессии решается не относительно имеющихся исходных данных, а по отношению к их логарифмам. Объясняется это тем, что уравнение регрессии со степенным трендом относится по оцениваемым параметрам к нелинейным моделям, но путем логарифмирования его можно привести к линейному виду.

В результате уравнение регрессии для степенного тренда (см. табл. 2.7) приобретет следующий вид:

$$\sum (\ln Y_{\text{факт}} - \ln Y_{\text{расч}})^2 \Rightarrow \min \sum e^2, \quad (2.22)$$

где $\ln Y_{\text{факт}}$ и $\ln Y_{\text{расч}}$ — натуральные логарифмы фактических и расчетных значений зависимой (результативной) переменной для различных моментов времени;

$\min \sum e^2$ — минимальная сумма квадратов отклонений фактических значений $\ln Y$ от его предсказываемых значений.

Следует иметь в виду, что приведение нелинейной функции к линейному виду с помощью логарифмирования используется очень часто, хотя это и приводит к некоторым коллизиям. Вот что пишут по этому поводу Е.М. Четыркин и И.Л. Калихман: «Однако такое преобразование приводит к тому, что оценка параметров базируется не на минимизации суммы квадратов отклонений, а на минимизации суммы квадратов отклонений в логарифмах. ...Следствием этого является некоторое смещение оценок параметров, получаемых обычным (линейным) МНК»¹.

Далее параметры этого уравнения регрессии находятся согласно формулам (2.1.4) и (2.1.5) либо решаются с помощью соответствующей компьютерной программы.

¹ Четыркин Е.М., Калихман И.Л. Вероятность и статистика. М.: Финансы и статистика, 1982. С. 255.

Поэтому прежде чем приступить к выполнению алгоритма действий № 3 «Как решить уравнение регрессии в Excel», нужно взять натуральные логарифмы (логарифмы, основанием которых служит число $e = 2,71828$) как от независимой переменной x — порядковый номер месяца, так и от зависимой переменной y — курс доллара. В Excel для этих целей можно воспользоваться функцией LN. Далее поступаем в полном соответствии с алгоритмом действий № 3, а данные, полученные после решения уравнения регрессии, занесем в табл. 2.8.

Согласно алгоритму действий № 4 «Оценка статистической значимости уравнения регрессии и его коэффициентов», проведем проверку статистической значимости этого уравнения регрессии. При этом выделим в табл. 2.8 все важнейшие пункты жирным шрифтом. В результате мы приходим к выводу, что у нас получились статистически значимыми уравнение регрессии и его коэффициенты как при 95%-ном, так и 99%-ном уровне надежности. Правда, поскольку уравнение регрессии мы решили относительно натуральных логарифмов, взятых от исходных данных, то в результате оно приобрело следующий вид:

$$\ln Y = -3,1154 + 1,28073 \ln X.$$

Согласно последнему уравнению регрессии, прогноз курса доллара рассчитывается на основе логарифмов, взятых от исходных данных. Например, прогноз относительно апреля 2010 г. вычисляется следующим образом:

$$\ln Y = -3,1154 + 1,28073 \times 5,370638 = 3,762939,$$

где $5,370638 = \ln(215)$ — натуральный логарифм от порядкового номера апреля 2010 г. — 215.

Отсюда находим (в Excel потенцирование натуральных логарифмов производится с помощью функции EXP), что прогноз курса доллара на апрель 2010 г. равен

$$Y = \text{EXP}(3,762939) = 43,07482.$$

После проведения соответствующих преобразований уравнение регрессии приобретет следующий вид:

$$Y = \text{EXP}(-3,1154 + 1,28073 \ln X) = 0,044361 \times X^{1,28073}.$$

С помощью последнего уравнения регрессии можно делать расчет прогнозов непосредственно от исходных данных, а не от их натуральных логарифмов. В результате можно получить следующий прогноз курса доллара на апрель 2010 г.:

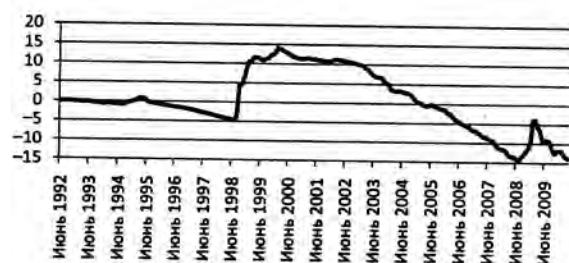


Рис. 2.6. Нестационарность остатков, полученных в степенной прогностической модели, руб.

Источник: расчеты автора.

5 руб. 92,4 коп. Следовательно, этот показатель для анализируемой модели оказался на 30 коп. больше, чем у линейной модели. В свою очередь средняя относительная ошибка по модулю в процентах (см. формулу (2.20)) для степенной модели оказалась равна 31,10%, т.е. на 7,78 процентных пункта ниже, чем у линейной модели. Более того, если построить график остатков по степенной прогностической модели (рис. 2.6), то легко обнаружить, что на нем наблюдается несколько локальных трендов. А это — как мы говорили ранее — наглядно свидетельствует о нестационарности остатков.

Отсюда можно сделать вывод, что не только линейная модель, но и степенная модель, в которой в качестве независимой переменной использовался фактор времени (порядковый номер месяца), оказались непригодны для прогнозирования курса доллара. Все это заставляет нас продолжить поиск адекватной прогностической модели.

Контрольные вопросы и задания

1. В чем суть метода наименьших квадратов, как этот метод можно представить графически?
2. Перечислите все пункты алгоритма действий, которые необходимо выполнить при решении уравнения регрессии в Microsoft Excel.
3. Какое уравнение регрессии (при прочих равных условиях) точнее: то, которое имеет коэффициент детерминации (R -квадрат) = 0,757, либо то, у которого R -квадрат равен 0,978? Объясните почему?

4. Можно ли признать статистически значимым уравнение регрессии в том случае, если значимость его F -критерия = 0,049? И если можно, то с каким уровнем надежности?
5. В каком случае делается вывод о статистической значимости коэффициентов уравнения регрессии? При каком P -значении коэффициенты уравнения регрессии являются статистически значимыми с 95%-ным и 99%-ным уровнем надежности?
6. Перечислите форматы трендов, используемых в Excel при решении уравнений регрессии графическим способом. Чем эти форматы трендов отличаются друг от друга?

ГЛАВА 3

УРАВНЕНИЯ АВТОРЕГРЕССИИ И АВТОРЕГРЕССИИ СО СКОЛЬЗЯЩИМ СРЕДНИМ

3.1. Специфика уравнений авторегрессии (AR)

Во главе 2 мы убедились, что использование фактора времени для прогноза курса доллара не дает достаточно точного результата, поэтому необходимо найти более подходящую независимую переменную (или переменные). Как известно, согласно теории эффективного рынка, наиболее полную информацию для прогноза будущей стоимости какого-либо финансового актива в момент времени t можно извлечь из его цены в момент времени $t - 1$. Причем точность прогноза уменьшается по мере того, как прогноз делается в моменты времени $t - 2$, $t - 3$ и т.д. Исходя из этого вполне очевидного постулата можно прийти к выводу, что наиболее полную информацию о курсе доллара на момент времени t содержит его курс на момент времени $t - 1$. Следовательно, наиболее точный прогноз курса американской валюты можно рассчитать на основе уравнения регрессии, включив в него в качестве независимой переменной курс доллара с лагом $t - 1$. Такого рода уравнения регрессии, в которых значения результативного признака прогнозируются на основе его предыдущих значений, в статистической литературе называют *уравнениями авторегрессии*.

Правда, в отличие от прогностической модели, в которой в качестве независимой переменной используется фактор времени, а потому горизонт для прогноза практически безграничен, прогноз по авторегрессионной модели имеет небольшой временной горизонт для прогноза, равный длине лага. В частности, модель авторегрессии с лагом в один месяц способна давать прогноз с упреждением в один месяц.

Помимо относительно небольшого временного горизонта для прогноза в процессе построения моделей авторегрессии возникает еще

одна серьезная проблема. Дело в том, что наличие лаговых значений зависимой переменной в правой части уравнения приводит к нарушению одной из важнейших предпосылок метода наименьших квадратов (МНК) — об отсутствии связи между зависимой (результативной) и независимой (факторной) переменными. Если перейти к языку формул, то теоретически эта проблема может быть изложена следующим образом:

$$Y_t = c + bY_{t-1} + e_t, \quad (3.1)$$

где c — свободный член (константа) уравнения;
 Y_t — зависимая (результативная) переменная;
 Y_{t-1} — независимая (факторная) переменная с лагом в один месяц;
 b — соответствующий коэффициент при Y_{t-1} ;
 e_t — отклонение прогноза от фактического курса доллара (остаток) в текущем месяце t .

Таким образом, из формулы (3.1) следует, что в уравнении авторегрессии может иметь место, во-первых, зависимость между e_t и e_{t-1} , т.е. может быть нарушена предпосылка МНК об отсутствии автокорреляции в остатках; во-вторых, может появиться зависимость между факторной переменной Y_{t-1} и остатками e_t , т.е. будет нарушена предпосылка МНК о *гомоскедастичности*¹ остатков.

Наличие автокорреляции в остатках означает определенную связь (корреляцию) между остатками текущих и предыдущих наблюдений. При наличии такой зависимости остатки могут содержать определенную тенденцию либо какие-то циклические колебания. В этом случае делается вывод, что отклонения от прогноза не могут иметь случайный характер. При наличии автокорреляции в остатках оценки коэффициентов уравнения регрессии нельзя назвать состоятельными и эффективными.

Гомоскедастичность остатков означает, что дисперсия остатков e_t не изменяется в зависимости от величины факторной переменной Y_{t-1} . Если это не так, то возникает *гетероскедастичность остатков*, что так же, как и в случае автокорреляции в остатках, влияет на состоятельность оценки коэффициентов уравнения регрессии.

Для справки заметим, что *состоятельными* называются такие оценки, чья точность повышается по мере роста объема выборки, объема данных, на основе которых строится уравнение регрессии. В свою оче-

¹ Гомоскедастичность, или гомогенность, дисперсии — состояние, при котором измерения вариативности колеблются внутри диапазона, ожидаемого при случайной вариативности; гетероскедастичность — состояние, при котором измерения вариативности являются большими, чем ожидаемые случайно.

редь *эффективными* называются такие оценки, которые имеют наименьшую дисперсию.

Несмотря на высказанные опасения, многие авторитетные специалисты полагают, что в случае больших выборок уравнения авторегрессии позволяют получать состоятельные и эффективные оценки. Вот как, например, оценивает авторегрессионные модели профессор статистики Стэнфордского университета Т. Андерсон: «Модель авторегрессии обладает рядом преимуществ по сравнению с моделью скользящего среднего и процессом авторегрессии с остатками в виде скользящего среднего, хотя последние в определенных случаях могут хорошо описывать образование наблюдаемых временных рядов. Оценки коэффициентов процесса авторегрессии легко вычисляются. Статистические процедуры для такого процесса, основывающиеся на теории больших выборок, легко выполняемы, поскольку они соответствуют обычной технике наименьших квадратов. Во многих случаях коэффициенты процесса авторегрессии допускают непосредственную интерпретацию, а линейные функции от запаздывающих переменных могут быть использованы для прогнозирования»¹.

Следует заметить, что в зависимости от того, сколько предыдущих значений временного ряда будет включено в уравнение авторегрессии в качестве лаговых (факторных) переменных, принято различать авторегрессионные процессы разного порядка. Так, в формуле (3.1) представлен авторегрессионный процесс 1-го порядка, который в англоязычной литературе обычно называется словосочетанием *Auto Regressive* и кратко обозначается как AR(1).

Например, в том случае, когда в авторегрессию 1-го порядка добавляются лаговые переменные Y_{t-2} и Y_{t-3} , его принято обозначать как AR(3), т.е. как авторегрессионный процесс 3-го порядка. При этом уравнение для AR(3) примет следующий вид:

$$Y_t = c + b_1 Y_{t-1} + b_2 Y_{t-2} + b_3 Y_{t-3} + e_t, \quad (3.2)$$

где Y_{t-1} , Y_{t-2} и Y_{t-3} — независимые (факторные) переменные с лагом в один, два и три месяца;

b_1 , b_2 и b_3 — соответствующие коэффициенты регрессии при лаговых переменных.

¹ Андерсон Т. Статистический анализ временных рядов / пер. под ред. Ю.К. Беляева. М.: Мир, 1976. С. 276.

3.2. Специфика уравнений авторегрессии со скользящим средним (ARMA)

Помимо авторегрессионных моделей нам необходимо также познакомиться и с моделями со скользящим средним в остатках, которые в англоязычной литературе обычно называются словосочетанием *Moving Average*. Полезность моделей со скользящим средним в остатках обусловлена тем, что для стационарного ряда предсказываемую переменную Y_t можно представить в виде линейной функции прошлых ошибок (отклонений прогнозов от их фактических значений). Следует иметь в виду, что термин «скользящая средняя» в данном случае не является синонимом скользящей средней, применяемой, например, для сезонного сглаживания уровней динамического ряда. При этом модель со скользящим средним в остатках 1-го порядка кратко обозначается как MA(1), а в виде формулы она приобретает следующий вид:

$$Y_t = c + e_t - \theta e_{t-1} + u_t, \quad (3.3)$$

где e_{t-1} — отклонение фактического курса доллара от прогноза в предыдущем периоде (в нашем случае с лагом в один месяц);

θ — соответствующий коэффициент при e_{t-1} ;

u_t — остатки, не коррелирующие с остатками предыдущего периода, или, как их еще называют в научной литературе, «белый шум», наличие которого свидетельствует о стационарности в остатках.

Объединение в одной модели авторегрессионного процесса AR и модели со скользящим средним в остатках MA приводит к созданию более экономичной модели с точки зрения количества используемых параметров. Эту объединенную модель в англоязычной литературе кратко называют ARMA. Эта аббревиатура произошла от словосочетания *Auto Regressive — Moving Average*, что в переводе означает «авторегрессионный процесс со скользящим средним в остатках».

Порядок в этой модели в буквенной форме принято обозначать как ARMA(p , q), где p — величина порядка авторегрессионного процесса, а q — величина порядка процесса со скользящим средним в остатках. Например, модель ARMA(2; 1) фактически представляет собой комбинацию модели AR(2) с моделью MA(1), т.е. в одной модели объединена авторегрессионная модель 2-го порядка с моделью со скользящим средним в остатках 1-го порядка. В результате модель ARMA(2; 1) приобретает следующий вид:

$$Y_t = c + b_1 Y_{t-1} + b_2 Y_{t-2} + e_t - \theta e_{t-1} + u_t, \quad (3.4)$$

Чтобы объединенная модель ARMA(2; 1) была более понятна, ее можно задать в виде двух уравнений. Так, для AR(2) формула будет иметь вид

$$Y_t = c + b_1 Y_{t-1} + b_2 Y_{t-2} + e_t, \quad (3.5)$$

в то время как уравнение для MA(1) можно представить в следующем виде:

$$e_t = \theta e_{t-1} + u_t. \quad (3.6)$$

Следовательно, формулу (3.4) модели ARMA(2; 1) можно получить путем вычитания из формулы (3.5) расчетного параметра θe_{t-1} из левой части уравнения (3.6).

3.3. Коррелограмма и идентификация лаговых переменных в уравнениях AR и ARMA

При практическом построении модели ARMA(p, q) наиболее трудным является определение параметров p и q , т.е. определение оптимального количества лагов. При этом инструментами для нахождения соответствующих лаговых переменных являются автокорреляционная функция и частная автокорреляционная функция.

Программа EViews позволяет довольно быстро найти оптимальные параметры p и q для модели ARMA, для этого используется коррелограмма зависимости между различными лагами временного ряда с ежемесячными курсами американского доллара к российскому рублю.

Алгоритм действий № 5

Как построить коррелограмму в EViews

Шаг 1. Выбор основных опций для построения коррелограммы

С этой целью загрузим в EViews ежемесячные данные по курсу доллара (столбец с данными обозначим как USDollar) в соответствии с алгоритмом действий № 2 «Импорт данных и создание рабочего файла в EViews», изложенным в главе 1.

Далее строим коррелограмму, тем более что в EViews сделать это довольно просто. С этой целью в Workfile (рабочем файле) этой программы открываем файл USDollar. После чего в файле USDollar нам необходимо выбрать опции VIEW/CORRELOGRAM, а в появившемся окне (рис. 3.1) CORRELOGRAM SPECIFICATION (спецификация

коррелограммы) оставить заданные по умолчанию опцию LEVEL (исходный уровень) и опцию LAGS TO INCLUDE (максимальная величина лага, включенного в коррелограмму). В результате у нас получится коррелограмма исходных уровней (фактических значений курса доллара) временного ряда USDollar с величиной лага от 1 до 36.

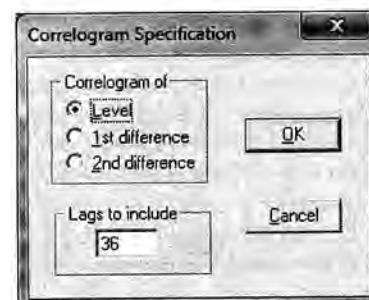


Рис. 3.1. Использование опций Level и Lags to include для построения коррелограммы

Шаг 2. Дополнительные возможности, которые можно использовать для построения коррелограммы

Если бы мы выбрали, например, опцию 1ST DIFFERENCE (разница исходных уровней 1-го порядка) или 2ND DIFFERENCE (разница исходных уровней 2-го порядка), тогда была бы построена коррелограмма не исходных уровней временного ряда, а соответственно их первых и вторых разностей. Например, исходный уровень для курса доллара по состоянию на апрель 2010 г. был равен 29,2886 руб. В то время как разница исходных уровней 1-го порядка на эту же дату оказалась равна -0,0752 руб. (т.е. по сравнению с прошлым месяцем курс доллара снизился на 7,52 коп.), а разница исходных уровней 2-го порядка составила 0,5094 руб. (т.е. падение курса доллара по сравнению с предыдущим месяцем уменьшилось на 50,94 коп.).

В полученной коррелограмме (см. табл. 3.1) можно увидеть, как меняются коэффициенты автокорреляции (*Autocorrelation*, или AC) и частной автокорреляции (*Partial Correlation*, или PAC) в зависимости от изменения величины лага. Корреляционную зависимость между последовательными уровнями временного ряда называют *автокорреляцией уровней ряда*. Так, коэффициент автокорреляции уровней первого порядка измеряет корреляционную зависимость между динамикой курса доллара временного ряда t и динамикой курса доллара временного

ряда $t-1$, т.е. в нашем случае измеряется коэффициент автокорреляции при лаге в один месяц. В свою очередь коэффициент автокорреляции уровней второго порядка измеряет зависимость между динамикой курса доллара временного ряда t и динамикой курса доллара временного ряда $t-2$, т.е. при лаге в два месяца. И так далее, вплоть до коэффициента автокорреляции уровней 36-го порядка, измеряющего зависимость между динамикой курса доллара временного ряда t и динамикой курса доллара временного ряда $t-36$, т.е. с лагом в 36 месяцев.

При этом коэффициент автокорреляции уровней k -го (т.е. 1-го, 2-го, ..., 36-го) порядка находится в EViews по следующей формуле:

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad (3.7)$$

где n — количество наблюдений во временном ряде;

k — величина лага;

Y_t — динамика курса доллара временного ряда t ;

Y_{t-k} — динамика курса доллара временного ряда $t-k$;

\bar{Y} — средняя для всей выборки.

Следует заметить, что коэффициент автокорреляции, рассчитываемый в EViews, несколько отличается от обычно вычисляемого коэффициента автокорреляции. Дело в том, что в EViews с целью упрощения вычислений в качестве \bar{Y} взята средняя для всей выборки, в то время как обычно для рядов Y_t и Y_{t-k} берутся свои средние.

Частной автокорреляционной функцией называют серию частных коэффициентов автокорреляции r_k , измеряющих связь между текущим лагом временного ряда Y_t и предыдущими лагами временного ряда Y_{t-1} , Y_{t-2} , ..., Y_{t-k+1} с устранением влияния других промежуточных временных лагов. Вполне естественно, что при нулевом лаге коэффициент частной корреляции $\rho_0 = 1$, а при лаге $k = 1$ $\rho_1 = r_1$, т.е. коэффициент частной корреляции равен коэффициенту автокорреляции.

Для лага k больше 1 EViews рекурсивно вычисляет частную автокорреляцию по следующей формуле:

$$\rho_k = \frac{r_k - \sum_{j=1}^{k-1} \rho_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \rho_{k-1,j} r_{k-j}}, \quad (3.8)$$

где r_k — коэффициент автокорреляции для лага k .

Этот алгоритм вычисления коэффициента частной корреляции, предложенный Боксом и Дженкинсом в 1976 г., представляет собой аппроксимацию. Чтобы найти его более точную оценку, следует решить следующее уравнение регрессии, с помощью которого мы найдем коэффициент частной корреляции ρ_k для лага k :

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_{k-1} Y_{t-(k-1)} + \rho_k Y_{t-k} + e_t, \quad (3.9)$$

где e_t — остатки.

Судя по полученной коррелограмме (см. табл. 3.1), уровень автокорреляции (AC) между исходными уровнями временного ряда USDollar постоянно убывает начиная с 1-го лага. В свою очередь уровень частной корреляции (PAC) резко снижается уже после 1-го лага, а после 2-го лага осциллирующим образом стремится к нулю (т.е. колеблется вокруг нуля).

В том случае, когда мы хотим построить модель авторегрессионного процесса $AR(p)$, для определения оптимального числа p мы должны использовать частную автокорреляционную функцию. При этом следует исходить из следующего критерия: оптимальное число p в уравнении авторегрессии должно быть меньше лага, в котором частная автокорреляционная функция начинает стремиться к нулю. Судя по коррелограмме, помещенной в табл. 3.1, коэффициент частной автокорреляции для лага один месяц (или лага 1-го порядка) равен 0,99, а для лага два месяца (или лага 2-го порядка) — -0,25. Однако для 3-го порядка коэффициент частной автокорреляции равен -0,014, причем начиная с этого лага величина этого коэффициента колеблется вокруг нулевого уровня. Следовательно, можно сделать вывод, что для прогнозирования курса доллара с помощью модели авторегрессии необходимо использовать модель $AR(2)$, которая примет следующий вид:

$$Y_t = c + b_1 Y_{t-1} + b_2 Y_{t-2} + e_t. \quad (3.10)$$

В свою очередь при идентификации модели $ARMA(p, q)$ в качестве лага p выбирается лаг, после которого начинает убывать частная автокорреляционная функция, а в качестве лага q — лаг, после которого начинает убывать автокорреляционная функция. Исходя из табл. 3.1 легко прийти к выводу, что коэффициент автокорреляции начинает убывать уже с лага 2-го порядка. Аналогичный вывод можно сделать и относительно коэффициента частной автокорреляции. Поэтому для прогнозирования курса доллара с помощью модели авторегрес-

сии со скользящими средними в остатках необходимо использовать модель ARMA(1, 1), которая примет следующий вид:

$$Y_t = c + b_1 Y_{t-1} + e_t - \theta e_{t-1} + u_t. \quad (3.11)$$

Два последних столбца в табл. 3.1 показывают соответственно Q -статистику Лjung — Бокса (Q -Stat) и ее значимость (Prob.) для каждого лага. Следует иметь в виду, что Q -статистика для лага k является тестовой статистикой при нулевой гипотезе об отсутствии автокорреляции между динамикой курса доллара временного ряда t и динамикой курса доллара временного ряда $t - k$.

Таблица 3.

Коррелограмма исходных уровней временного ряда USDollar
с величиной лага от 1 до 36

| Date: 05/13/10 Time: 10:47 | | | | | | |
|----------------------------|---------------------|----|--------------|---------------|--------|-------|
| Sample: 1992M06 2010M05 | | | | | | |
| Included observations: 215 | | | | | | |
| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob. |
| .***** | .***** | 1 | 0,990 | 0,990 | 213,53 | 0,000 |
| .***** | ** . | 2 | 0,974 | -0,250 | 421,45 | 0,000 |
| .***** | . | 3 | 0,958 | -0,014 | 623,25 | 0,000 |
| .***** | . | 4 | 0,940 | -0,024 | 818,72 | 0,000 |
| .***** | * . | 5 | 0,921 | -0,083 | 1007,3 | 0,000 |
| .***** | . | 6 | 0,901 | -0,015 | 1188,6 | 0,000 |
| .***** | . | 7 | 0,881 | -0,027 | 1362,6 | 0,000 |
| .***** | * . | 8 | 0,859 | -0,065 | 1528,9 | 0,000 |
| .***** | .* | 9 | 0,839 | 0,096 | 1688,2 | 0,000 |
| .***** | .* | 10 | 0,822 | 0,133 | 1841,8 | 0,000 |
| .***** | . | 11 | 0,806 | -0,000 | 1990,5 | 0,000 |
| .***** | . | 12 | 0,792 | 0,024 | 2134,6 | 0,000 |
| .***** | * . | 13 | 0,777 | -0,064 | 2274,1 | 0,000 |

Окончание

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob. |
|-----------------|---------------------|----|-------|--------|--------|-------|
| .***** | * . | 14 | 0,761 | -0,077 | 2408,7 | 0,000 |
| .***** | . | 15 | 0,745 | -0,043 | 2538,1 | 0,000 |
| .***** | . | 16 | 0,728 | -0,026 | 2662,2 | 0,000 |
| .***** | * . | 17 | 0,710 | -0,074 | 2780,9 | 0,000 |
| .***** | .* | 18 | 0,692 | 0,088 | 2894,4 | 0,000 |
| .***** | . | 19 | 0,675 | 0,022 | 3002,9 | 0,000 |
| .***** | . | 20 | 0,659 | 0,023 | 3106,7 | 0,000 |
| .***** | . | 21 | 0,642 | 0,025 | 3205,8 | 0,000 |
| .***** | . | 22 | 0,626 | -0,030 | 3300,5 | 0,000 |
| .***** | . | 23 | 0,610 | -0,033 | 3390,8 | 0,000 |
| .***** | . | 24 | 0,594 | -0,036 | 3476,9 | 0,000 |
| .**** | . | 25 | 0,577 | -0,055 | 3558,7 | 0,000 |
| .**** | . | 26 | 0,561 | -0,038 | 3636,3 | 0,000 |
| .**** | . | 27 | 0,544 | 0,013 | 3709,6 | 0,000 |
| .**** | . | 28 | 0,527 | 0,000 | 3778,8 | 0,000 |
| .**** | . | 29 | 0,510 | 0,026 | 3844,0 | 0,000 |
| .**** | . | 30 | 0,493 | 0,002 | 3905,2 | 0,000 |
| .**** | . | 31 | 0,476 | 0,014 | 3962,6 | 0,000 |
| .**** | . | 32 | 0,459 | -0,025 | 4016,3 | 0,000 |
| .*** | . | 33 | 0,442 | -0,026 | 4066,5 | 0,000 |
| .*** | * . | 34 | 0,425 | -0,057 | 4113,1 | 0,000 |
| .*** | . | 35 | 0,408 | -0,042 | 4156,3 | 0,000 |
| .*** | . | 36 | 0,391 | -0,021 | 4196,1 | 0,000 |

При этом Q -статистика Лjunga — Бокса для лага k -го порядка находится по следующей формуле:

$$Q = T(T+2) \sum_{k=1}^m \frac{r_k^2}{(T-k)}, \quad (3.12)$$

где T — число наблюдений;

r_k — автокорреляция k -го порядка;

m — число проверяемых лагов.

Например, для лага 1-го порядка формула (3.12) имеет следующее значение:

$$Q = 215 \times (215 + 2) \sum_{k=1}^1 \frac{0,990^2}{(215 - 1)} = 213,53.$$

При этом для проверки нулевой гипотезы об отсутствии автокорреляции величина Q сравнивается с величиной $\chi^2_{\text{крит}, \alpha, m}$. Причем последняя величина характеризует распределение χ^2 (хи-квадрат) с n степенями свободы, представляющее собой распределение суммы квадратов n независимых стандартных случайных величин. Следовательно, если $Q > \chi^2_{\text{крит}, \alpha, m}$, то отсюда делается вывод о неслучайном характере этого распределения, а следовательно, нулевая гипотеза опровергается.

В EViews значимость Q -статистики определяется автоматически. Однако при необходимости величину $\chi^2_{\text{крит}, \alpha, m}$ в Excel можно найти с помощью функции ХИ2ОБР, которая для r_1 , т.е. для автокорреляции с лагом 1-го порядка, находится следующим образом: ХИ2ОБР(0,05; 1) = 3,8414. Поскольку в этом случае Q -статистика = 213,53 > $\chi^2_{\text{крит}, \alpha, m} = 3,8414$ для лага 1-го порядка, то, следовательно, нулевая гипотеза об отсутствии автокорреляции отклоняется. При этом значимость Q -статистики для m в Excel находится с помощью функции ХИ2РАСП(213,53; 1) = 0,00.

Следует иметь в виду, что в том случае, когда в табл. 3.1 значимость (Prob.) Q -статистики будет больше 0,05, то нулевую гипотезу об отсутствии автокорреляции между уровнями ряда с лагом k -го порядка нельзя считать опровергнутой с 95%-ным уровнем надежности. Если значимость Q -статистики будет больше 0,01, но меньше 0,05, то нулевую гипотезу об отсутствии автокорреляции между уровнями ряда с лагом k -го порядка нельзя считать опровергнутой с 99%-ным уровнем надежности. Судя по коррелограмме исходных уровней временного ряда USDollar (см. табл. 3.1), значимость Q -статистики для всех 36 лагов равна нулю, поэтому нулевая гипотеза об отсутствии автокорреляции в остатках отклоняется для всех лагов.

3.4. Решение в Excel уравнения авторегрессии 2-го порядка AR(2)

После того как с помощью соответствующей коррелограммы (см. табл. 3.1) мы пришли к выводу, что для получения оптимального прогноза по курсу доллара следует построить модель авторегрессии 2-го порядка AR(2), следующим нашим шагом должно стать нахождение ее параметров. Правда, для этого развернутое уравнение авторегрессии AR(2), представленное в формуле (3.10), необходимо немного упростить. С этой целью из формулы следует убрать остатки, которые появятся только после решения этого уравнения. Кроме того, чтобы убрать у коэффициентов факторных переменных подстрочные индексы (цифры), обозначим их различными буквами. В результате формула (3.10) приобретет более удобный для решения вид:

$$Y_t = c + aY_{t-1} + bY_{t-2}, \quad (3.13)$$

где Y_{t-1} — курс доллара США с лагом в один месяц;

Y_{t-2} — курс доллара США с лагом в два месяца;

c — свободный член (константа).

Мы уже научились решать уравнения регрессии в Excel (см. алгоритм действий № 3 «Как решить уравнение регрессии в Excel»), поэтому, используя этот алгоритм, можно получить соответствующее уравнение авторегрессии, которое, как известно, является частным случаем уравнения регрессии и отличается от последнего лишь наличием лаговых факторных переменных. А для загрузки и первичной обработки данных по ежемесячному курсу доллара необходимо воспользоваться алгоритмом действий № 1 «Как строить диаграммы в Microsoft Excel» — Шаг 1 «Поиск данных, их загрузка и первичная обработка в Excel».

Далее создадим в Excel три столбца: во-первых, с зависимой переменной USDollar — ежемесячный курс доллара США; во-вторых, с двумя независимыми переменными USDollar(-1) — курс доллара США с лагом в один месяц и USDollar(-2) — курс доллара США с лагом в два месяца. При этом загруженная база данных по американской валюте охватывает период с июня 1992 г. по апрель 2010 г.

Далее, согласно алгоритму действий № 3 «Как решить уравнение регрессии в Excel», установим в появившемся окне РЕГРЕССИЯ следующие опции (рис. 3.2): ВХОДНОЙ ИНТЕРВАЛ Y (\$B\$1:\$B\$214); ВХОДНОЙ ИНТЕРВАЛ X (\$C\$1:\$D\$214); УРОВЕНЬ НАДЕЖНОСТИ (99); ВЫХОДНОЙ ИНТЕРВАЛ (\$L\$2).

Таблица 3.2

Вывод итогов в Excel для уравнения авторегрессии 2-го порядка AR(2)

| ВЫВОД ИТОГОВ | | | | | | | | | | | |
|--------------------------|----------|-------------|--------------------|--------------|------------|--------------|-------------|--|---------------|--------------|---------------|
| Регрессионная статистика | | | | | | | | | Верхние 99,0% | Нижние 99,0% | Верхние 99,0% |
| Множественный R | 0,99767 | | | | | | | | | | |
| R-квадрат | 0,995346 | | | | | | | | | | |
| Нормированный R-квадрат | 0,995301 | | | | | | | | | | |
| Стандартная ошибка | 0,811301 | | | | | | | | | | |
| Наблюдения | 213 | | | | | | | | | | |
| Дисперсионный анализ | | df | SS | MS | F | Значимость F | | | | | |
| Регрессия | | 2 | 29560,67 | 14780,34 | 22455,38 | 1,3E-245 | | | | | |
| Остаток | | 210 | 138,2239 | 0,658209 | | | | | | | |
| Итого | | 212 | 29698,9 | | | | | | | | |
| | | Коэффициент | Стандартная ошибка | t-статистика | P-значение | Нижние 95% | Верхние 95% | | | | |
| У-пересечение | | 0,225996 | 0,107792 | 2,096606 | 0,037226 | 0,013504 | 0,438489 | | | | |
| USDollar(-1) | | 1,297958 | 0,065686 | 19,76008 | 8,41E-50 | 1,16847 | 1,427446 | | | | |
| USDollar(-2) | | -0,30466 | 0,065384 | -4,65954 | 5,62E-06 | -0,43355 | -0,17577 | | | | |

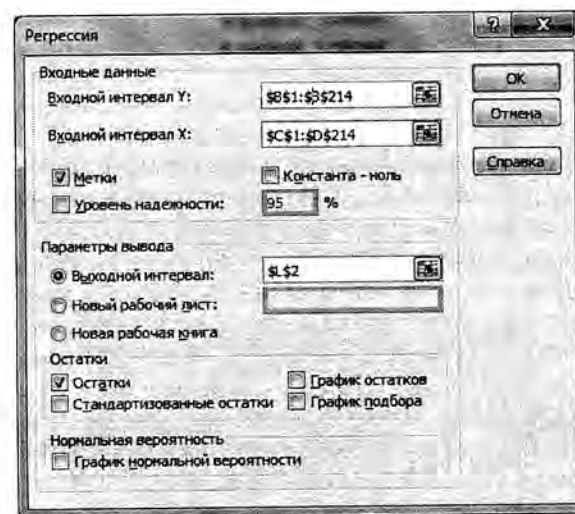


Рис. 3.2. Установка опций в окне РЕГРЕССИЯ

В результате решения в Excel уравнения авторегрессии AR(2) со свободным членом мы получим следующий **ВЫВОД ИТОГОВ**, представленный в виде табл. 3.2. Возьмем из этой таблицы значения коэффициентов (см. столбец «Коэффициенты») и, подставив их в формулу (3.13), получим следующее уравнение авторегрессии (с округлением):

$$\text{USDollar} = 0,2260 + 1,2980 \text{ USDollar}(-1) - 0,3047 \text{ USDollar}(-2),$$

где USDollar — зависимая переменная, курс доллара США;

USDollar(-1) — независимая переменная, курс доллара США с лагом в один месяц;

USDollar(-2) — независимая переменная, курс доллара США с лагом в два месяца;

0,2260 — свободный член (константа).

При этом экономическая интерпретация этого уравнения авторегрессии 2-го порядка следующая: во-первых, в период с июня 1992 г. по апрель 2010 г. при исходном уровне 0,2260 руб. рост на 1 руб. курса доллара в текущем месяце приводил к повышению прогнозируемого курса доллара в будущем месяце в среднем на 1,2980 руб.; во-вторых, одновременно с этим рост курса доллара в прошлом месяце приводил к снижению прогнозируемого курса доллара в будущем месяце в среднем на 0,3047 руб.

Действуя согласно алгоритму действий № 4 «Оценка статистической значимости уравнения регрессии и его коэффициентов», мы можем сделать следующие выводы.

1. Поскольку коэффициент детерминации R^2 для уравнения регрессии оказался равен 0,9977, то отсюда следует, что оно в 99,77% случаев в состоянии объяснить ежемесячные колебания курса доллара.
2. Значимость F равна 1,3Е – 245 или нулю, следовательно, уравнение регрессии статистически значимо как при 95%-ном уровне надежности, так и при 99%-ном уровне надежности.
3. P -значение для свободного члена (константы) равно 0,037226, следовательно, константа статистически значима лишь при 95%-ном уровне надежности, но незначима при 99%-ном уровне надежности, поскольку ее P -значение больше 0,01. P -значение для двух коэффициентов регрессии равно 0, следовательно, эти коэффициенты статистически значимы как при 95%-ном уровне надежности, так и при 99%-ном уровне надежности.

3.5. Решение в EViews уравнения авторегрессии 2-го порядка AR(2)

Уравнение авторегрессии 2-го порядка с константой можно решить не только в Excel, но и в EViews. Более того, решение этого уравнения регрессии в EViews имеет ряд преимуществ, обусловленных спецификой этой программы. Во-первых, в EViews можно быстрее оценить прогностическую точность полученной статистической модели; во-вторых, есть возможность протестировать полученные остатки на стационарность, наличие автокорреляции, а также провести ряд других важных тестов, о которых мы расскажем позднее. Тем читателям, которым еще не приходилось решать уравнения регрессии в EViews, советуем внимательно ознакомиться с алгоритмом действий № 6 «Как решить уравнение регрессии в EViews».

Алгоритм действий № 6

Как решить уравнение регрессии в EViews

Шаг 1. Импорт данных из Excel и создание рабочего файла в EViews

Для импорта ежемесячных данных по курсу доллара (на конец месяца) за период с июня 1992 г. по апрель 2010 г. из Excel в EViews

необходимо воспользоваться алгоритмом действий № 2 «Импорт данных и создание рабочего файла в EViews». При этом столбец с соответствующими данными по курсу доллара мы обозначили как USDollar.

Шаг 2. Выбор опций в EViews для решения уравнения регрессии

После импорта данных в Excel выбираем в командной строке EViews опции OBJECT/NEW OBJECT, а затем в появившемся окне (NEW OBJECT (НОВЫЙ ОБЪЕКТ)) выбираем опцию EQUATION (УРАВНЕНИЕ) — рис. 3.3).

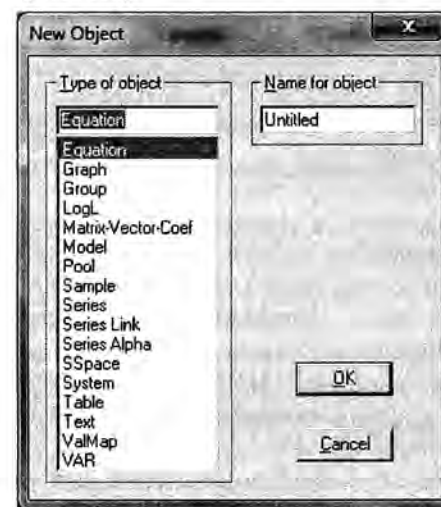


Рис. 3.3. Опция EQUATION в окне NEW OBJECT программы EViews

Далее в EViews появляется новое окно — EQUATION ESTIMATION (ОЦЕНКА УРАВНЕНИЯ), которое мы должны заполнить, как показано на рис. 3.4.

Следует иметь в виду, что в опции ESTIMATION SETTINGS (ПАРАМЕТРЫ ОЦЕНИВАЕМОЙ МОДЕЛИ) в мини-окне METHOD (МЕТОД РЕШЕНИЯ) по умолчанию появляется опция LS — LEAST SQUARES (NLC AND ARMA), название которой переводится как МЕТОД НАИМЕНЬШИХ КВАДРАТОВ (НЕЛИНЕЙНЫЙ МНК И ARMA). Поскольку это уравнение авторегрессии мы решаем с помощью метода наименьших квадратов, то эту опцию мы оставляем. Хотя при необходимости в EViews можно использовать несколько других методов решения уравнений, на которых мы сейчас не будем останавливаться.

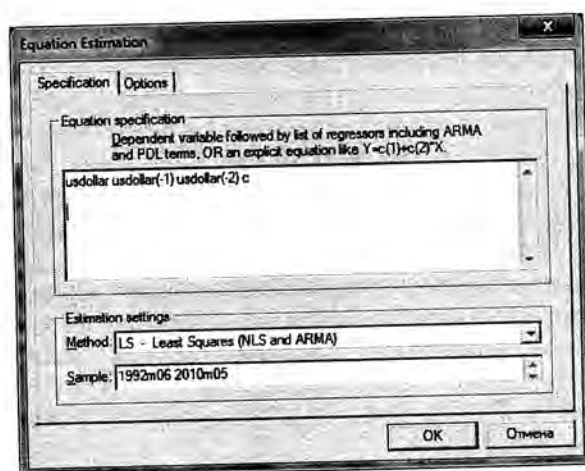


Рис. 3.4. Заполнение окна EQUATION ESTIMATION

Шаг 3. Выбор параметров оцениваемой статистической модели

В опции ESTIMATION SETTINGS (ПАРАМЕТРЫ ОЦЕНИВАЕМОЙ МОДЕЛИ) есть еще одно мини-окно — SAMPLE (ВЫБОРКА), в котором по умолчанию указывается либо общее количество наблюдений, либо период наблюдения. В данном случае в мини-окне SAMPLE появилась надпись: 1992M06 2010M05, что означает, что наша выборка содержит ежемесячные данные за период с июня 1992 г. по май 2010 г.

Особенно внимательным следует быть при заполнении мини-окна EQUATION SPECIFICATION (СПЕЦИФИКАЦИЯ УРАВНЕНИЯ), в котором нужно написать латиницей название зависимой переменной (ее в списке всегда пишут первой слева) и независимых переменных, а также — в случае необходимости — константу (свободный член уравнения), обозначаемую латинской буквой *c*.

В нашем случае мини-окно EQUATION SPECIFICATION заполняется следующим образом:

$$\text{USDollar} \text{ USDollar}(-1) \text{ USDollar}(-2) \text{ } c, \quad (3.14)$$

где USDollar — зависимая переменная, курс доллара США;

USDollar(-1) — независимая переменная, курс доллара США с лагом в один месяц;

USDollar(-2) — независимая переменная, курс доллара США с лагом в два месяца;

c — свободный член (константа).

Мини-окно EQUATION SPECIFICATION легко заполнить, если воспользоваться уравнением авторегрессии (3.13). При этом нужно сделать следующее: во-первых, убрать буквенные обозначения коэффициентов регрессии, но оставить константу *c*; во-вторых, вместо *Y*, поставить соответствующее название зависимой переменной — USDollar, а для факторных (независимых) переменных Y_{t-1} и Y_{t-2} в скобках еще и добавить соответствующую цифру лага со знаком минус.

Если вспомнить, что формула (3.14) фактически означает уравнение авторегрессии 2-го порядка со свободным членом, то мини-окно EQUATION SPECIFICATION можно заполнить другой, более краткой, но вполне равнозначной формулой:

$$\text{USDollar AR}(1) \text{ AR}(2) \text{ } c, \quad (3.15)$$

где USDollar — зависимая переменная;

AR(1) — авторегрессия 1-го порядка, или USDollar(-1);

AR(2) — авторегрессия 2-го порядка, или USDollar(-2).

Шаг 4. Вывод в EViews параметров уравнения авторегрессии

Итак, все опции, необходимые для решения уравнения авторегрессии, установлены. Далее щелкаем кнопку OK в окне EQUATION ESTIMATION. В результате чего получаем данные с параметрами уравнения авторегрессии, которые мы поместили в табл. 3.3. При этом не стоит удивляться тому, что после соответствующей корректировки количество наблюдений у нас сократилось с 215 до 213. Это обусловлено тем, что при создании факторных переменных с лагом в один и в два месяца мы потеряли два наблюдения. В результате теперь наша скорректированная выборка охватывает период не с июня 1992 г., а с августа 1992 г. по апрель 2010 г.

Чтобы нашему читателю было легче понять содержащиеся в табл. 3.3 англоязычные термины, они даются вместе с параллельным переводом в скобках. Если сравнить табл. 3.3 с выводом итогов, полученным после решения этого же уравнения авторегрессии в Excel (см. табл. 3.2), то можно прийти к выводу о тождественности большей части информации, имеющейся в обеих таблицах. Следует также заметить, что как в программе Excel, так и в EViews мы смогли получить коэффициенты уравнения регрессии с одинаковым уровнем точности.

Таблица 3.3

Вывод итогов в EViews и принятие решения о статистической значимости уравнения регрессии и значимости его коэффициентов

| Dependent Variable (зависимая переменная): USDOLLAR | | | | |
|---|------------------------------|---|-------------------------------|-----------------------|
| Method: Least Squares (метод наименьших квадратов) | | | | |
| Date: 05/21/10 Time: 11:18 (дата и время решения уравнения) | | | | |
| Sample (adjusted) (скорректированная выборка): 1992M08 2010M04 | | | | |
| Included observations: 213 after adjustments (включает 213 наблюдений после корректировки). | | | | |
| Variable (независимая переменная) | Coefficient (коэффициент) | Std. Error (стандартная ошибка) | t-Statistic (t-статистика) | Prob. (P-значение) |
| USDOLLAR(-1) | 1,297958 | 0,065686 | 19,76008 | 0,0000 |
| USDOLLAR(-2) | -0,304661 | 0,065384 | -4,659541 | 0,0000 |
| C | 0,225996 | 0,107792 | 2,096606 | 0,0372 |
| R-squared (коэффициент детерминации R^2) | 0,995346 | Mean dependent var (Среднее зависимой переменной) | | 19,69163 |
| Adjusted R-squared (скорректированный R^2) | 0,995301 | S.D. dependent var (Стандартное отклонение зависимой переменной) | | 11,83593 |
| S.E. of regression (стандартная ошибка регрессии) | 0,811301 | Akaike info criterion (Информационный критерий Акаика) | | 2,433629 |
| Sum squared resid (сумма остатков в квадрате) | 138,2239 | Schwarz criterion (Критерий Шварца) | | 2,480971 |
| Log likelihood (логарифм максимального правдоподобия) | -256,1815 | F-statistic (F-статистика) | | 22 455,38 |
| Durbin — Watson stat (критерий Дарбина — Уотсона) | 2,039140 | Prob. (F-statistic) (Значимость F-критерия) | | 0,00 |

3.6. Интерпретация параметров уравнения авторегрессии в EViews

Какой статистический смысл имеют те или иные параметры уравнения регрессии при выводе итогов в Excel, уже говорилось в главе 1 книги. Однако при выводе итогов в EViews мы получаем новую информацию о других важных параметрах уравнения регрессии, которых нет при выводе итогов в Excel. Чтобы обратить внимание читателя на эти дополнительные параметры, мы выделили их жирным шрифтом в табл. 3.3. Познакомимся со статистическим смыслом этих еще не изученных нами дополнительных параметров уравнения регрессии.

1. В таблице 3.3 среди пока неизвестных нам параметров уравнения регрессии можно назвать такой важный показатель, как **LOG LIKELIHOOD** (ЛОГАРИФМ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ), который используется в качестве критерия для отбора наиболее адекватных уравнений регрессии. Чем выше логарифм максимального правдоподобия, тем более адекватным считается уравнение регрессии. При этом логарифм максимального правдоподобия находится по следующей формуле:

$$LL = -\frac{T}{2} \left[1 + \log(2\pi) + \log \left(\frac{\sum e^2}{T} \right) \right], \quad (3.16)$$

где T — количество наблюдений;
 e — отклонение (остатки) прогноза от фактического курса доллара;
 π — число пи, равное 3,141593...

В нашем случае логарифм максимального правдоподобия имеет следующее значение:

$$LL = -\frac{213}{2} \left[1 + \log(2 \times 3,141593) + \log \left(\frac{138,2239}{213} \right) \right] = -256,181.$$

2. Следующим еще не изученным нами параметром уравнения регрессии является **DURBIN-WATSON STAT** (КРИТЕРИЙ ДАРБИНА — УОТСОНА), который является тестом на наличие автокорреляции в остатках. Как мы уже говорили, при наличии автокорреляции в остатках оценки коэффициентов уравнения регрессии нельзя называть состоятельными и эффективными. При этом критерий Дарбина — Уотсона находится следующим образом:

$$D = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}, \quad (3.17)$$

где n — количество наблюдений;

e_t — отклонение (остатки) прогноза от фактического курса доллара;

e_{t-1} — отклонение (остатки) прогноза от фактического курса доллара с лагом в один месяц.

В нашем случае критерий Дарбина — Уотсона имеет следующее значение:

$$D = \frac{281,8580}{138,2239} = 2,0391.$$

Правда, критерий Дарбина — Уотсона нельзя использовать для тестирования уравнений авторегрессии на наличие автокорреляции в остатках, поскольку в этом случае он теряет свою мощность. Это объясняется тем, что применение критерия Дарбина — Уотсона предполагает строгое соблюдение предпосылки о разделении переменных на зависимую (результативную) и независимую (факторную) переменную. В уравнениях авторегрессии, как известно, в правой части уравнения имеются лаговые значения результативной переменной, а следовательно, указанная предпосылка не соблюдается. В этом случае фактическое значение критерия Дарбина — Уотсона приблизительно равно 2 как при наличии, так и при отсутствии автокорреляции в остатках. Тем не менее в обычных уравнениях регрессии этот критерий весьма полезен для тестирования остатков на наличие автокорреляции.

3. Следующий параметр уравнения регрессии, на наш взгляд, не представляет каких-либо трудностей для его понимания — MEAN DEPENDENT VAR (СРЕДНЕЕ ЗНАЧЕНИЕ ЗАВИСИМОЙ ПЕРЕМЕННОЙ). При этом среднее значение зависимой переменной рассчитывается по довольно простой формуле

$$\bar{Y}_t = \frac{\sum Y_t}{n}, \quad (3.18)$$

где n — количество наблюдений;

Y_t — зависимая переменная, ежемесячный курс доллара.

В нашем случае среднее значение (вернее сказать, среднее хронологическое, поскольку мы берем период за 213 месяцев) зависимой переменной будет равно

$$\bar{Y}_t = \frac{4194,3169}{213} = 19,6916 \text{ руб. за один доллар США.}$$

4. Еще один показатель, характеризующий зависимую переменную данного уравнения регрессии — S.D. DEPENDENT VAR (СТАНДАРТНОЕ ОТКЛОНЕНИЕ ЗАВИСИМОЙ ПЕРЕМЕННОЙ). При этом стандартное отклонение зависимой переменной находится следующим образом:

$$S.D._{dep.} = \sqrt{\frac{\sum (Y_t - \bar{Y})^2}{n-1}}, \quad (3.19)$$

где \bar{Y} — среднее хронологическое значение зависимой переменной;

n — количество наблюдений;

Y_t — зависимая переменная, ежемесячный курс доллара.

В нашем случае стандартное отклонение зависимой переменной вычисляется достаточно легко:

$$S.D._{dep.} = \sqrt{\frac{29698,895}{213-1}} = \sqrt{140,089} = 11,83593.$$

5. Важными параметрами уравнения регрессии являются два информационных критерия — AKAIKE INFO CRITERION (ИНФОРМАЦИОННЫЙ КРИТЕРИЙ АКАЙКА) и SCHWARZ CRITERION (КРИТЕРИЙ ШВАРЦА). Оба этих информационных критерия можно использовать в качестве критериев для определения в уравнении регрессии оптимальной длины лага. При этом они основаны на принципе снижения остаточной суммы квадратов при добавлении значимого фактора. Так, информационный критерий Акайка находится по следующей формуле:

$$AIC = -2LL : T + 2k : T, \quad (3.20)$$

где LL — логарифм максимального правдоподобия;

T — количество наблюдений;

k — общее количество лагов в уравнении авторегрессии.

В нашем случае информационный критерий Акайка равен

$$AIC = -2 \times 256,1815 : 213 \times 2 \times 3 : 213 = 2,4336.$$

В свою очередь информационный критерий Шварца рассчитывается по формуле

$$SC = -2LL : T + (k \ln T) : T. \quad (3.21)$$

Относительно нашего уравнения регрессии информационный критерий Шварца имеет следующее значение:

$$SC = -2 \times 256,1815 : 213 + (3 \ln 213) : 213 = 2,4809.$$

Обычно оцениваемая статистическая модель лучше соответствует фактическим данным при более высоком порядке p и q в модели ARMA(p, q). Платой за это кажущееся повышение точности является вполне очевидная потеря в простоте статистической модели и в экономии включенных в него параметров, поэтому для достижения компромисса между точностью уравнения регрессии и экономией его параметров пользуются информационными критериями Акаика и Шварца.

При выборе из двух уравнений регрессии обычно предпочтение отдается той статистической модели, у которой меньше значения этих информационных критериев. Следует также заметить, что информационный критерий Шварца по сравнению с критерием Акаика позволяет отбирать уравнения регрессии с более экономичными параметрами.

Как мы уже говорили, в уравнениях авторегрессии при тестировании остатков на наличие автокорреляции критерий Дарбина — Уотсона теряет свою мощност, и в этих случаях приходится пользоваться иными критериями. Например, тем, кто работает в Excel, с этой целью проще воспользоваться критерием h Дарбина, или, как его еще называют, h -статистикой Дарбина. Его расчет выполняется по следующей формуле:

$$h = \left(1 - \frac{D}{2}\right) \sqrt{\frac{n}{1 - nV}}, \quad (3.22)$$

где D — критерий Дарбина — Уотсона;

n — количество наблюдений;

V — квадрат стандартной ошибки при лаговой факторной переменной Y_{t-1} .

Например, в нашем случае критерий h Дарбина имеет следующую величину:

$$h = \left(1 - \frac{2,03914}{2}\right) \sqrt{\frac{213}{1 - 213 \times 0,06569^2}} = -1,00368.$$

При увеличении объема выборки распределение h -статистики стремится к нормальному с нулевым математическим ожиданием и дисперсией, равной 1. Поэтому гипотеза об отсутствии автокорреляции в остатках отвергается, если фактическое значение h -статистики оказывается больше, чем критическое значение нормального распределе-

ния. Для проверки по критерию h Дарбина гипотезы о наличии автокорреляции в остатках проще воспользоваться следующим правилом.

1. Если $h > 1,96$, то нулевая гипотеза об отсутствии положительной автокорреляции в остатках отклоняется.
2. Если $h < -1,96$, то нулевая гипотеза об отсутствии отрицательной автокорреляции в остатках отклоняется.
3. Если $-1,96 < h < 1,96$, то нет основания отклонять нулевую гипотезу об отсутствии автокорреляции в остатках.

Поскольку критерий h Дарбина получился равным $-1,00368$, то у нас нет основания отклонять нулевую гипотезу об отсутствии автокорреляции в остатках.

Следует иметь в виду, что в использовании критерия h Дарбина есть определенная специфика. Во-первых, этот критерий нельзя применять, если произведение $nV \geq 1$. Во-вторых, h -статистику Дарбина можно использовать лишь для больших выборок ($n \geq 30$ наблюдений). В-третьих, критерий h Дарбина зависит только от V (квадрата стандартной ошибки) при лаговой факторной переменной Y_{t-1} и не зависит от числа лагов, используемых в уравнении авторегрессии.

В EViews для проверки статистических моделей на наличие автокорреляции в остатках целесообразно использовать LM -тест Бройша — Годфри (*Breusch — Godfrey Serial Correlation LM Test*), который в отличие от h -статистики Дарбина может быть применим не только для авторегрессии 1-го порядка, но и для авторегрессии более высоких порядков.

Суть этого теста заключается в построении уравнения регрессии остатков с заранее заданной величиной лага, решение которого позволяет сделать вывод о наличии или отсутствии автокорреляции в остатках:

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_m e_{t-m} + u, \quad (3.23)$$

где e — остатки;

m — заданная величина лага;

u — некоррелируемые остатки, т.е. «белый шум».

При этом выдвигается нулевая гипотеза, что $\rho_1 = \rho_2 = \dots = \rho_m = 0$, т.е. автокорреляция в остатках с различным лагом отсутствует. Вполне естественно, что альтернативной гипотезой в этом случае является гипотеза $\rho_1 \neq \rho_2 \neq \dots \neq \rho_m \neq 0$. По итогам решения уравнения регрессии 3.23 нулевая гипотеза либо принимается, либо отклоняется.

Поскольку LM -тест Бройша — Годфри проверяет остатки на автокорреляцию, то мы его проводим уже после того, как решили основное уравнение авторегрессии, а следовательно, нашли остатки, полученные на основе этой статистической модели.

Алгоритм действий № 7

Как выполняется LM-тест Бройша — Годфри в EViews

Шаг 1. Практическая реализация LM-теста Бройша — Годфри

В EViews реализация LM-теста Бройша — Годфри довольно проста. С этой целью необходимо в командной строке (1 Command) или в строке уравнение (3 EQUATION) выбрать следующие опции: View/Residual Tests/Serial Correlation LM Test... После чего появляется мини-окно LAG SPECIFICATION, в котором можно задать интересующую нас величину лага (рис. 3.5). В этом случае мы задаем величину лага, равную 2, что обусловлено структурой лаговых переменных, включенных в уравнение авторегрессии (см. формулу (3.14)). В общем виде величина задаваемого лага для модели ARMA (p, q) = max(p, q), которая в нашем случае приобретает вид: ARMA (2, 0) = max(2, 0) = 2.



Рис. 3.5. Мини-окно Lag Specification, в котором задается величина лага

Шаг 2. Интерпретация результатов тестирования

В результате мы получаем следующие данные по результатам проведения LM-теста Бройша — Годфри, которые заносим в табл. 3.4. EViews сообщает две тестовые статистики (см. две верхние строки в табл. 3.4, выделенные жирным шрифтом). При этом для оценки результатов тестирования в качестве основного используется критерий $Obs \times R\text{-squared}$ (Наблюдения $\times R^2$), который мы не только выделили жирным шрифтом, но и подчеркнули. Для нашего случая $Obs \times R\text{-squared} = 0,024005 \times 213 = 5,112998$. Правда, если мы попробуем сами провести это вычисление, то из-за округления R^2 у нас получится некоторое расхождение с цифрой, выданной EViews. При этом предполагается, что LM-тестовая статистика (критерий $Obs \times R\text{-squared}$) асимптотически распределена как χ^2 (хи-квадрат-распределение), о котором мы уже говорили выше. Поэтому значимость $Obs \times R\text{-squared}$ определяется с помощью табличного $\chi^2_{крит, \alpha, m}$.

Таблица 3.4

Результаты LM-теста Бройша — Годфри на выявление автокорреляции в остатках

автокорреляция в остатках

| | | | | |
|---|-----------------|-----------------------|-----------------|--------|
| Breusch — Godfrey Serial Correlation LM Test: | | | | |
| F-statistic | 2,557889 | Probability | 0,079903 | |
| Obs × R-squared | 5.112998 | Probability | 0.077576 | |
| Test Equation: | | | | |
| Dependent Variable: RESID | | | | |
| Method: Least Squares | | | | |
| Date: 05/23/10 Time: 12:14 | | | | |
| Presample missing value lagged residuals set to zero. | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 1,515562 | 0,678379 | 2,234093 | 0,0265 |
| USDOLLAR(-2) | -1,502682 | 0,672600 | -2,234138 | 0,0265 |
| C | -0,457845 | 0,231070 | -1,981414 | 0,0489 |
| RESID(-1) | -1,544221 | 0,685931 | -2,251276 | 0,0254 |
| RESID(-2) | -0,450418 | 0,218962 | -2,057061 | 0,0409 |
| R-squared | 0,024005 | Mean dependent var | -7,44E -16 | |
| Adjusted R-squared | 0,005236 | S.D. dependent var | 0,807465 | |
| S.E. of regression | 0,805348 | Akaike info criterion | 2,428111 | |
| Sum squared resid | 134,9059 | Schwarz criterion | 2,507014 | |
| Log likelihood | -253,5938 | F-statistic | 1,278944 | |
| Durbin — Watson stat | 1,998877 | Prob(F-statistic) | 0,279381 | |

В том случае, когда значимость (*Probability*) $Obs \times R\text{-squared}$ у нас оказывается меньше 0,05, нулевая гипотеза об отсутствии автокорреляции в остатках отклоняется. Если же $Obs \times R\text{-squared}$ больше 0,05, нулевую гипотезу об отсутствии автокорреляции в остатках нельзя отклонить. Поскольку в нашем случае значимость $Obs \times R\text{-squared} = 0,077576$, то, следовательно, нулевая гипотеза не отклоняется и можно сделать вывод об отсутствии автокорреляции в остатках.

В EViews приводится в качестве дополнительного F -критерий ($F\text{-statistic}$), который представляет собой тест на определение совокупной значимости всех лаговых остатков. В нашем случае F -критерий также подтверждает отсутствие автокорреляции в остатках.

Как мы уже убедились ранее, при построении уравнения авторегрессии у нас происходит уменьшение временного ряда данных, что ведет к пропуску в том числе и части лаговых остатков. Согласно предложению, выдвинутому в 1993 г. Давидсоном и Маккинном, в этом случае отсутствующие остатки следует приравнять к нулю. По их мнению, это дает лучшую статистику, чем в случае пропуска этих остатков. Однако, по мнению большинства исследователей, в этом случае распределение F-статистики становится не совсем точным. Тем не менее EViews дает F-критерий для справочных целей.

3.7. Оценка точности решения уравнения авторегрессии в EViews

Важным критерием оценки эффективности статистической модели является уровень точности, получаемый с помощью определенной статистической модели при прогнозе курса доллара. Его в EViews можно оценить с помощью алгоритма действий № 8.

Алгоритм действий № 8

Как оценить точность статистической модели в EViews

Шаг 1. Выбор необходимой опции

Чтобы оценить точность статистической модели, нужно в строке 3 EQUATION (уравнение) выбрать опцию FORECAST.

Шаг 2. Заполнение мини-окна FORECAST.

В результате откроется мини-окно FORECAST, которое следует заполнить таким образом (рис. 3.6).

По умолчанию в опции FORECAST NAME (название файла с прогнозом) задается название файла с точечным прогнозом путем прибавления к исходному файлу латинской буквы *f*. Например, если у нас исходный файл — USDollar, то название файла с прогнозом будет задано программой как USDollarf. В опции FORECAST SAMPLE (выборка для прогноза) по умолчанию задается исходная выборка данных для прогноза, которую при необходимости можно изменить. В опции METHOD (метод прогноза) нам следует выбрать STATIC FORECAST (статичный прогноз), т.е. мы оцениваем точность прогноза только на один следующий месяц. Если в опции

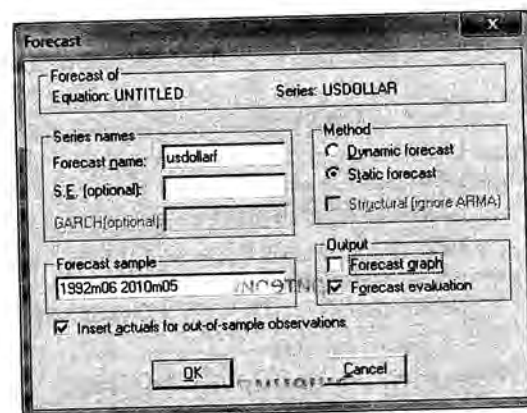


Рис. 3.6. Заполнение мини-окна Forecast

METHOD выбрать вариант DYNAMIC FORECAST (динамичный прогноз), то это увеличило бы временной горизонт для прогноза, но его точность существенно снизилась бы. Дело в том, что при динамическом прогнозировании предсказание на следующий месяц составляется так же, как и при статичном, но прогнозы на более длительные сроки составляются на основе расчетных, т.е. предсказанных, а не фактических значений независимой переменной.

В опции OUTPUT (вывод итогов) мы задали вариант FORECAST EVALUATION (оценка прогноза) и получили таблицу с оценкой точности прогноза этой статистической модели (см. табл. 3.6). При необходимости в последней опции можно задать еще и вариант FORECAST GRAPH (график прогнозов), после чего можно получить и график с прогнозами.

Шаг 3. Интерпретация параметров, характеризующих уровень точности статистической модели

Чтобы по табл. 3.6 вынести суждение о качестве статистической модели, сначала нужно ознакомиться с табл. 3.5. Причем в первую очередь следует обратить внимание на раздел этой таблицы «Идеальное значение параметра». Из него можно сделать вывод: чем ближе стремятся к нулю параметры, представленные в табл. 3.6, тем выше прогностическая ценность статистической модели. Единственным исключением из этого правила является параметр Covariance Proportion (доля ковариации, т.е. доля несистематической ошибки), для которого идеальным значением является единица.

Таблица 3.5
Критерии для оценки точности прогноза и их идеальные значения

| Англоязычный термин | Перевод | Диапазон изменений | Идеальное значение параметра |
|--------------------------------|---|--------------------|------------------------------|
| Root Mean Squared Error | Квадратный корень средней ошибки предсказания | От 0 до ∞ | $\Rightarrow 0$ |
| Mean Absolute Error | Средняя ошибка по модулю | От 0 до ∞ | $\Rightarrow 0$ |
| Mean Absolute Percentage Error | Средняя ошибка по модулю, % | От 0 до ∞ | $\Rightarrow 0$ |
| Theil Inequality Coefficient | Коэффициент неравенства Тейла | От 0 до 1 | $\Rightarrow 0$ |
| Bias Proportion | Доля систематической ошибки | | $\Rightarrow 0$ |
| Variance Proportion | Доля вариации, т.е. разницы между колебанием фактических и расчетных значений | От 0 до 1 | $\Rightarrow 0$ |
| Covariance Proportion | Доля ковариации, т.е. несистематической ошибки | От 0 до 1 | $\Rightarrow 1$ |

В алгоритме действий № 8 «Как оценить точность статистической модели в EViews» в самом общем виде уже говорилось об интерпретации параметров, характеризующих уровень точности статистической модели. Однако далее все желающие могут более подробно ознакомиться со спецификой параметров, содержащихся в табл. 3.6. «Оценка точности уравнения регрессии (статистической модели) с параметрами USDollar = 0,2260 + 1,2980 USDollar(-1) - 0,3047 USDollar(-2)».

Таблица 3.6
Оценка точности уравнения регрессии (статистической модели) с параметрами USDollar = 0,2260 + 1,2980 USDollar(-1) - 0,3047 USDollar(-2)

| |
|--|
| Forecast (прогноз): USDOLLARF |
| Actual (фактический курс): USDOLLAR |
| Forecast sample (выборка для прогноза): 1992M06 2010M05 |
| Adjusted sample (скорректированная выборка): 1992M08 2010M05 |
| Included observations (включено наблюдений): 213 |

Окончание

| | |
|---|----------|
| Root Mean Squared Error (квадратный корень средней ошибки предсказания) | 0,805567 |
| Mean Absolute Error (средняя ошибка по модулю) | 0,360739 |
| Mean Absolute Percentage Error (средняя ошибка по модулю, %) | 4,800196 |
| Theil Inequality Coefficient (коэффициент неравенства Тейла) | 0,017548 |
| Bias Proportion (доля систематической ошибки) | 0,000000 |
| Variance Proportion (доля вариации) | 0,001166 |
| Covariance Proportion (доля ковариации, т.е. несистематической ошибки) | 0,998834 |

В частности, *Root Mean Squared Error* (квадратный корень средней ошибки предсказания) представляет собой квадратный корень из суммы квадратов остатков (разницы между фактическим и предсказанным значением), деленной на общее количество наблюдений. Квадратный корень средней ошибки предсказания находят по следующей формуле:

$$\text{Root Mean Squared Error} = \sqrt{\frac{\sum (Y_t - \hat{Y}_t)^2}{n}} = \sqrt{\frac{\sum e^2}{n}}, \quad (3.24)$$

где Y_t — фактические значения курса доллара;

\hat{Y}_t — предсказанные значения курса доллара;

n — количество наблюдений;

e — ошибки (остатки) или разница между расчетным и фактическим значением курса доллара.

На основе имевшихся у нас данных квадратный корень средней ошибки предсказания по курсу доллара имеет следующее значение:

$$\text{Root Mean Squared Error} = \sqrt{\frac{138,2239}{213}} = \sqrt{0,648939} = 0,805567.$$

При этом следует иметь в виду, что величина квадратного корня средней ошибки предсказания всегда чуть больше стандартной ошибки, представленной, например, в «Выводе итогов в Excel для уравнения авторегрессии 2-го порядка AR(2)» (см. табл. 3.2). Это объясняется тем, что квадратный корень средней ошибки предсказания находится путем деления суммы квадратов остатков на общее количество наблюдений. В то время как стандартная ошибка находится путем деления суммы квадратов остатков на число степеней свободы. Так, в нашем случае

квадратный корень средней ошибки предсказания равен 0,805567 (при общем числе наблюдений, равном 213), а стандартная ошибка равна 0,811301 (при 210 степенях свободы). Причем число степеней свободы для нашей статистической модели находят следующим образом:

$$df = n - k - 1 = 213 - 2 - 1 = 210,$$

где k — количество факторных переменных в статистической модели.

Mean Absolute Error (средняя ошибка) по модулю представляет собой абсолютную (без учета знака) сумму остатков (ошибок), деленную на общее количество наблюдений. Поскольку при сложении сумма остатков стремится к нулю, для нахождения средней ошибки приходится использовать их модульные значения. Средняя ошибка по модулю вычисляется по формуле

$$\text{Mean Absolute Error} = \frac{\sum |Y_t - \hat{Y}_t|}{n} \quad (3.25)$$

Для нашей статистической модели средняя ошибка по модулю равна:

$$\text{Mean Absolute Error} = \frac{76,83736}{213} = 0,360739.$$

Mean Absolute Percentage Error (средняя ошибка по модулю, %) равна сумме относительных ошибок (остатков), деленной на общее количество наблюдений. Средняя ошибка по модулю находится следующим образом:

$$\text{Mean Absolute Percentage Error} = \frac{1}{n} \times \sum \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100. \quad (3.26)$$

В этом случае средняя ошибка по модулю имеет следующее значение:

$$\text{Mean Absolute Percentage Error} = \frac{1}{213} \times 1022,442 = 4,8000196.$$

Как мы уже говорили, по трем перечисленным выше параметрам — *Root Mean Squared Error*, *Mean Absolute Error* и *Mean Absolute Percentage Error* — выбор обычно делается в пользу той статистической модели, у которой значения этих параметров стремятся к нулю. Вполне очевидно, что чем меньше квадратный корень средней ошибки предсказания, средняя ошибка по модулю и средняя ошибка по модулю (%), тем выше прогностическая ценность модели. При этом следует иметь в виду, что все три перечисленных выше параметра имеют диапазон значений от нуля до бесконечности.

Несколько особняком стоят четыре остальных параметра, представленных в табл. 3.6. Из них главным является *Theil Inequality Coefficient* (коэффициент неравенства Тейла), в то время как три других можно назвать производными от первого. При этом значения этих четырех параметров изменяются в пределах от нуля до единицы.

Theil Inequality Coefficient служит для общей оценки качества прогностической модели. Как мы уже говорили, идеальным для статистической модели считается значение коэффициента Тейла, равное нулю. Таким образом, чем ближе этот коэффициент к нулю, тем ценнее предсказание.

Коэффициент неравенства Тейла находится по следующей формуле:

$$\begin{aligned} \text{Theil Inequality Coefficient} &= \frac{\sqrt{\sum (Y_t - \hat{Y}_t)^2 : n}}{\sqrt{\sum \hat{Y}_t^2 : n} + \sqrt{\sum Y_t^2 : n}} = \\ &= \frac{\text{Root Mean Squared Error}}{\sqrt{\sum \hat{Y}_t^2 : n} + \sqrt{\sum Y_t^2 : n}}. \end{aligned} \quad (3.27)$$

В нашем случае коэффициент Тейла имеет следующее значение:

$$\text{Theil Inequality Coefficient} = \frac{0,805567}{22,94652 + 22,96065} = 0,017548.$$

Квадратный корень средней ошибки предсказания может быть разложен на слагаемые по следующей формуле:

$$\text{Root Mean Squared Error} = ((\sum \hat{Y}_t : n) - \bar{Y})^2 + (S_y - S_{\hat{y}})^2 + 2(1-r)S_y S_{\hat{y}}, \quad (3.28)$$

где \hat{Y}_t — средняя величина фактических значений курса доллара;

S_y — стандартное отклонение предсказанных значений курса доллара;

$S_{\hat{y}}$ — стандартное отклонение фактических значений курса доллара;

r — коэффициент корреляции между фактическими и предсказанными значениями курса доллара.

Если мы найдем долю (относительно квадратного корня средней ошибки) каждого из трех слагаемых (см. формулу (3.30)), то в этом случае нам удастся вычислить еще три важных параметра, характеризующих качество прогноза, рассчитанного с помощью исследуемой статистической модели.

Так, *Bias Proportion* (доля систематической ошибки прогноза) служит своего рода индикатором, показывающим, насколько средняя величина прогнозируемого значения, например средняя величина прогноза

по курсу доллара, отклоняется от средней величины его фактического значения. Причем идеальной считается ситуация, когда доля систематической ошибки предсказания равна нулю. При этом доля систематической ошибки находится по следующей формуле:

$$\text{Bias Proportion} = \frac{[(\sum \hat{Y}_i : n) - \bar{Y}]^2}{\sum (Y_i - \hat{Y})^2 : n} \quad (3.29)$$

Чтобы найти долю систематической ошибки в предсказаниях, сделанных при помощи этой прогностической модели, нам пришлось проделать следующие вычисления:

$$\text{Bias Proportion} = \frac{(19,69163 - 19,69163)^2}{0,648939} = 0.$$

Индикатор *Variance Proportion* (доля вариации) показывает, насколько отличаются друг от друга вариации фактических и предсказываемых значений, например курса доллара. Чем меньше доля этой вариации, тем лучше, а в идеале она должна быть равна нулю. Доля вариации находится по следующей формуле:

$$\text{Variance Proportion} = \frac{(S_y - S_{\hat{y}})^2}{\sum (Y_i - \hat{Y})^2 : n} \quad (3.30)$$

где выражение $\sum (Y_i - \hat{Y})^2 : n$ по сути своей представляет дисперсию ошибки предсказания.

В Excel дисперсию и стандартное отклонение для генеральной совокупности данных можно вычислить с помощью соответствующих функций ДИСПР и СТАНДОТКЛОНП.

В нашей статистической модели доля вариации в предсказаниях оказалась равна:

$$\text{Variance Proportion} = \frac{(11,7806 - 11,80811)^2}{0,648939} = 0,001166.$$

Индикатор *Covariance Proportion* (доля ковариации) показывает долю несистематической ошибки в общей величине дисперсии ошибки предсказания. Поскольку этот индикатор показывает долю несистематической, остаточной ошибки в предсказаниях, которая присутствует во всех статистических моделях, то ее наличие не требует отказа от использования этого уравнения регрессии. Доля несистематической ошибки прогноза изменяется в диапазоне от 0 до 1. Причем в идеале этот показатель дол-

жен быть равен единице, чем он и отличается от всех остальных индикаторов, представленных в табл. 3.6.

В общем виде доля ковариации в предсказаниях находится следующим образом:

$$\text{Covariance Proportion} = \frac{2(1-r)S_y S_{\hat{y}}}{\sum (Y_i - \hat{Y})^2 : n} \quad (3.31)$$

где r — коэффициент корреляции между фактическими и предсказанными значениями курса доллара.

При этом по формуле (3.31) доля ковариаций в предсказаниях равна:

$$\text{Covariance Proportion} = \frac{2 \times (1 - 0,99767) \times 139,1066}{0,648939} = 0,998834.$$

Следует также иметь в виду, что доля систематической ошибки прогноза, доля вариации и доля ковариации в сумме равняются единице. В виде формулы это соотношение можно представить следующим образом:

$$\text{Bias Proportion} + \text{Variance Proportion} + \text{Covariance Proportion} = 1. \quad (3.32)$$

Следовательно, когда доля ковариации равна единице, это означает, что доля вариации и доля систематической ошибки в прогнозах равны нулю. В этом случае можно было бы сделать вывод об идеальном качестве полученных прогнозов, чего на практике, как правило, не бывает. Используя преобразованную формулу (3.32), можно быстрее найти долю ковариации, чем по формуле (3.31). В результате долю ковариации в наших прогнозах можно также вычислить более простым способом:

$$\text{Covariance Proportion} = 1 - (\text{Bias Proportion} + \text{Variance Proportion}) = 1 - 0 - 0,001166 = 0,998834.$$

В заключение остановимся на содержательной интерпретации индикаторов, представленных в табл. 3.6. Из этой таблицы можно сделать вывод, что квадратный корень средней ошибки предсказания по курсу доллара после округления оказался равным 0,8056 руб., или 80,56 коп., в то время как средняя ошибка по модулю — 0,3607 руб., или 36,07 коп. В свою очередь средняя ошибка предсказания по модулю равна 4,80%. Напомним, что ошибка аппроксимации в пределах 5—7% свидетельствует о хорошем соответствии статистической модели исходным данным.

Коэффициент неравенства Тейла, фактически являющийся индексом, в этой таблице равен 0,0175, т.е. его значение довольно близко

подходит к нулю, что говорит о хорошем качестве предсказания. При этом доля систематической ошибки в предсказаниях равна 0 или 0%, в то время как доля вариации равна 0,0012, или 0,12%, а доля ковариации — 0,9988, или 99,88%.

Судя по табл. 3.6, с помощью двухфакторного уравнения регрессии со свободным членом нам удалось получить довольно качественную прогностическую модель. Тем не менее точность этой авторегрессионной модели можно повысить, причем довольно существенно.

Контрольные вопросы и задания

1. Какие уравнения называются уравнениями авторегрессии? Являются ли уравнения авторегрессии частным случаем уравнений регрессии? В чем преимущество использования в прогнозах лаговой переменной с точки зрения теории эффективного рынка?
2. Какая предпосылка метода наименьших квадратов (МНК) не соблюдается в уравнениях регрессии? В каких случаях с помощью уравнения авторегрессии можно получать состоятельные и эффективные оценки?
3. Что означают англоязычные аббревиатуры AR и ARMA? Чем отличается модель AR от модели ARMA? Какие переменные входят в модель ARMA(2; 1)?
4. Для чего необходима коррелограмма? В чем отличие автокорреляции от частной автокорреляционной функции? Что измеряет коэффициент автокорреляции уровней 1-го порядка?
5. Как производится идентификация моделей AR(p) и ARMA(p, q) с помощью коррелограммы? Как при этом используются автокорреляция и частная автокорреляция?
6. Почему критерий Дарбина — Уотсона нельзя использовать для тестирования уравнений авторегрессии на наличие автокорреляции в остатках? Какой тест на наличие автокорреляции в остатках в уравнениях авторегрессии используется в EViews? Какой лаг нужно задать в этом тесте при тестировании уравнения авторегрессии 2-го порядка?
7. Как находится квадратный корень средней ошибки предсказания? Почему для нахождения средней ошибки приходится использовать их модульные значения? Как находится средняя ошибка по модулю (%)? Для чего используется коэффициент неравенства Тейла? Какое значение коэффициента неравенства Тейла считается идеальным для статистической модели?

ГЛАВА 4

ПОДБОР АДЕКВАТНОГО УРАВНЕНИЯ АВТОРЕГРЕССИИ И СОСТАВЛЕНИЕ ТОЧЕЧНЫХ И ИНТЕРВАЛЬНЫХ ПРОГНОЗОВ ПО КУРСУ ДОЛЛАРА

4.1. Повышение статистической значимости коэффициентов в уравнении авторегрессии

Одним из способов повышения точности статистической модели является увеличение количества переменных, включаемых в уравнение регрессии. Однако в табл. 3.1 «Коррелограмма исходных уровней временного ряда USDollar с величиной лага от 1 до 36» хорошо видно, что коэффициент частной автокорреляции уже на лаге в три месяца становится близким к нулю. Отсюда следует вывод, что нет никакого смысла добавлять в уравнение авторегрессии 2-го порядка AR(2) со свободным членом факторную лаговую переменную с лагом в три месяца и более.

Вместе с тем вывод итогов как в Excel, так и в EViews для этого уравнения свидетельствует, что величина P -значений включенных в него коэффициентов далеко не одинакова (см. табл. 3.2 и 3.3). Так, P -значения для коэффициентов регрессии факторных переменных USDollar(-1) и USDollar(-2) практически равны нулю, что свидетельствует об их статистической значимости с 99%-ным уровнем надежности. А вот P -значение для коэффициента свободного члена (константы) этого уравнения регрессии равно 0,037226, что свидетельствует о его статистической значимости лишь с 95%-ным уровнем надежности (точнее сказать, с 96,28%-ным уровнем надежности: 100% — 3,72%).

Следовательно, чтобы повысить точность наших прогнозов, мы попробуем решить уравнение регрессии, исключив из формулы (3.14)

статистически менее значимый свободный член. С этой целью необходимо воспользоваться алгоритмом действий № 6 «Как решить уравнение регрессии в EViews» (см. главу 3), но при выборе параметров оцениваемой статистической модели (см. шаг 3 этого алгоритма) мини-окно EQUATION SPECIFICATION нужно заполнить следующим образом:

$$\text{USDollar USDollar}(-1) \text{ USDollar}(-2). \quad (4.1)$$

Фактически в буквенной форме формула (4.1) приобретет следующий вид:

$$\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2). \quad (4.2)$$

Причем, введя спецификацию (4.1) в EViews, мы тем самым даем программе задание оценить коэффициенты a и b из формулы (4.2). В результате EViews выдает итоги, которые заносятся в табл. 4.1. На основе данных этой таблицы мы получаем уравнение авторегрессии 2-го порядка AR(2) без константы со следующими параметрами:

$$\text{USDOLLAR} = 1,321092 \times \text{USDOLLAR}(-1) - 0,319415 \times \text{USDOLLAR}(-2), \quad (4.3)$$

где USDollar — зависимая переменная, курс доллара США;
USDollar(-1) — независимая переменная, курс доллара США с лагом в один месяц;
USDollar(-2) — независимая переменная, курс доллара США с лагом в два месяца.

Экономическая интерпретация этого уравнения авторегрессии 2-го порядка следующая: во-первых, в период с июня 1992 г. по апрель 2010 г. рост на 1 руб. курса доллара в текущем месяце приводил к повышению прогнозируемого курса доллара в будущем месяце в среднем на 1,3210 руб.; во-вторых, одновременно с этим рост курса доллара в прошлом месяце приводил к снижению прогнозируемого курса доллара в будущем месяце в среднем на 0,3194 руб.

Судя по табл. 4.1, все коэффициенты в этом уравнении имеют P -значения ($Prob.$) = 0, а следовательно, можно сделать вывод, что они значимы с 99%-ным уровнем надежности. Вполне очевидно, этого нам удалось добиться благодаря тому, что мы убрали из уравнения авторегрессии свободный член. Но как этот факт повлиял в целом на прогностические качества этой статистической модели?

Если посмотреть на коэффициент детерминации R^2 (R -squared), то видно, что после удаления константы он уменьшился весьма незначительно: с 99,53% (0,9953) до 99,52% (0,9952), или на 0,01 процентного пункта. Еще меньше снизился скорректированный коэффициент детерминации R^2 ($Adjusted R$ -squared). Вместе с тем в уравнении авторегрессии без свободного члена незначительно снизился логарифм максимального правдоподобия (его более высокое значение, как правило, свидетельствует о более высоком качестве прогноза) и одновременно с этим незначительно повысилась величина информационного критерия Акаика (его более низкое значение, как правило, свидетельствует о более высоком качестве прогноза). Однако плюсом для уравнения без константы стал тот факт, что информационный критерий Шварца, который сильнее «штрафует» включение в уравнение регрессии дополнительных факторов, у него оказался ниже (его более низкое значение, как правило, свидетельствует о более высоком качестве прогноза).

Таблица 4.1

Вывод итогов в EViews для уравнения авторегрессии 2-го порядка без свободного члена

| Dependent Variable: USDOLLAR | | | | |
|--|-------------|-----------------------|-------------|----------|
| Method: Least Squares | | | | |
| Date: 05/27/10 Time: 08:01 | | | | |
| Sample (adjusted): 1992M08 2010M04 | | | | |
| Included observations: 213 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 1,321092 | 0,065271 | 20,23999 | 0,0000 |
| USDOLLAR(-2) | -0,319415 | 0,065525 | -4,874679 | 0,0000 |
| R-squared | 0,995248 | Mean dependent var | | 19,69163 |
| Adjusted R-squared | 0,995226 | S.D. dependent var | | 11,83593 |
| S.E. of regression | 0,817803 | Akaike info criterion | | 2,444955 |
| Sum squared resid | 141,1173 | Schwarz criterion | | 2,476517 |
| Log likelihood | -258,3878 | Durbin — Watson stat | | 2,045083 |

Источник: расчеты автора по данным Банка России.

4.2. Оценка точности прогностической модели, проверка остатков на автокорреляцию и стационарность

Далее проверим уравнение AR(2) без константы на наличие автокорреляции в остатках с помощью *LM*-теста Бройша — Годфри, используя при этом алгоритм действий № 7. При этом в мини-окне LAG SPECIFICATION зададим величину лага, равную 2, поскольку мы тестируем уравнение авторегрессии 2-го порядка. Полученные результаты занесем в табл. 4.2. Поскольку значимость (*Probability*) главного критерия этого теста «Наблюдения $\times R^2$ » (*Obs \times R-squared*) равна 0,1069, то, следовательно, нулевая гипотеза об отсутствии автокорреляции в остатках не может быть отклонена с 95%-ным уровнем надежности (а точнее сказать, с 89,31%-ным уровнем надежности). Если сравнить последнюю цифру с аналогичными данными табл. 3.4, то об отсутствии автокорреляции в остатках в последнем случае можно говорить с большей уверенностью.

Таким образом, сравнение параметров, с одной стороны, уравнения AR(2) с константой (см. табл. 3.3), а с другой стороны, уравнения AR(2) без константы (см. табл. 4.1) не помогло нам сделать окончательный вывод в пользу одного из них. Аналогичный результат у нас получился и по итогам проведения *LM*-теста Бройша — Годфри на наличие автокорреляции в остатках. Поэтому мы решили оценить точность прогнозов, сделанных с помощью уравнения авторегрессии без константы, воспользовавшись алгоритмом действий № 8 «Как оценить точность статистической модели в EViews». В результате получилась табл. 4.3.

Если сравнить табл. 4.3 с табл. 3.6, то можно легко прийти к выводу, что шесть из восьми параметров, характеризующих точность прогнозов, свидетельствуют в пользу уравнения авторегрессии с константой. Правда, при этом разница между ними была весьма незначительной. Однако вполне естественно, что при прогнозировании курсов валют нас в первую очередь интересует точность предсказаний, которую можно получить с помощью той или иной статистической модели. И с этой точки зрения уравнение авторегрессии 2-го порядка без константы оказалось точнее. Так, средняя ошибка прогноза по модулю у последней модели оказалась (после округления) равна 33,50 коп., а средняя ошибка по модулю (%) — 2,78%, т.е. соответственно на 2,5 коп. и на 2,02 процентного пункта ниже, чем у уравнения авторегрессии с константой. Именно это обстоятельство и побудило нас сделать выбор в пользу уравнения авторегрессии без константы.

Чтобы сделать адекватный прогноз по курсу доллара, необходимо учесть как тренд, так и случайную компоненту, поскольку оба этих

Таблица 4.2

Результаты *LM*-теста Бройша — Годфри на выявление автокорреляции в остатках для уравнения авторегрессии 2-го порядка без свободного члена

в остатках для уравнения авторегрессии 2-го порядка без экзогенных переменных

| | | | | |
|---|-------------|-----------------------|-------------|----------|
| Breusch — Godfrey Serial Correlation LM Test: | | | | |
| F-statistic | 2,826594 | Probability | 0,061480 | |
| Obs*R-squared | 4,472558 | Probability | 0,106855 | |
| Test Equation: | | | | |
| Dependent Variable: RESID | | | | |
| Method: Least Squares | | | | |
| Date: 05/27/10 Time: 08:04 | | | | |
| Presample missing value lagged residuals set to zero. | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 1,502466 | 0,641949 | 2,340475 | 0,0202 |
| USDOLLAR(-2) | -1,506636 | 0,643725 | -2,340497 | 0,0202 |
| RESID(-1) | -1,535466 | 0,650172 | -2,361630 | 0,0191 |
| RESID(-2) | -0,460659 | 0,215474 | -2,137889 | 0,0337 |
| R-squared | 0,020998 | Mean dependent var | | 0,060106 |
| Adjusted R-squared | 0,006945 | S.D. dependent var | | 0,813645 |
| S.E. of regression | 0,810814 | Akaike info criterion | | 2,437045 |
| Sum squared resid | 137,4008 | Schwarz criterion | | 2,500168 |
| Log likelihood | -255,5453 | Durbin — Watson stat | | 1,999111 |

фактора существенно влияют на динамику валюты. Судя по табл. 4.3, нам удалось построить уравнение авторегрессии, с достаточно высокой степенью точности учитывающее тренд. Используя эту статистическую модель, можно делать точечные прогнозы, которые, правда, очень редко совпадают с фактическим курсом доллара. Объясняется это тем, что в ежемесячных колебаниях курса доллара достаточно большую роль играет не только тренд, но и случайная компонента. Судя по тому, что средняя ошибка по модулю равна 2,78%, вполне очевидно, что эту цифру можно считать своего рода среднестатистическим индикатором вклада случайной компоненты в динамику курса доллара.

Таблица 4.3

Оценка точности уравнения авторегрессии 2-го порядка без константы

| | |
|----------------------------------|----------|
| Forecast: USDOLLARF | |
| Actual: USDOLLAR | |
| Forecast sample: 1992M06 2010M05 | |
| Adjusted sample: 1992M08 2010M05 | |
| Included observations: 213 | |
| Root Mean Squared Error | 0,813955 |
| Mean Absolute Error | 0,334980 |
| Mean Absolute Percentage Error | 2,783718 |
| Theil Inequality Coefficient | 0,017731 |
| Bias Proportion | 0,005453 |
| Variance Proportion | 0,007783 |
| Covariance Proportion | 0,986764 |

Поскольку точечный прогноз по определению не в состоянии указать нам диапазон вероятного отклонения фактического курса доллара от его предсказываемого значения, то с этой целью приходится использовать так называемый *интервальный прогноз*. Суть интервального прогнозирования заключается в определении интервала значений, в который прогнозируемое значение попадет с определенной долей вероятности. Чем выше интервал прогноза (разница между максимальным и минимальным значениями прогноза), тем больше вероятность (ее еще называют уровнем надежности) его реализации.

Однако прежде чем перейти к составлению интервальных прогнозов, нам необходимо, во-первых, проверить полученные остатки на стационарность, во-вторых, посмотреть, является ли распределение остатков нормальным.

В EViews проверить остатки на стационарность достаточно просто, для этого нужно только точно следовать алгоритму действий № 9.

Алгоритм действий № 9

Как в EViews проверить остатки на стационарность

Шаг 1. Установка необходимых опций

С этой целью нужно выбрать строку 2 WORKFILE (рабочий файл), а затем открыть файл RESID (остатки), который появляется в рабо-

чем файле после того, как мы воспользовались опцией FORECAST (см. алгоритм действий № 8 «Как оценить точность статистической модели в EViews»).

Далее в файле RESID нам следует воспользоваться опцией UNIT ROOT TEST (тест на единичный корень), в результате чего появится (рис. 4.1) мини-окно UNIT ROOT TEST, в котором нам нужно выбрать следующие опции.

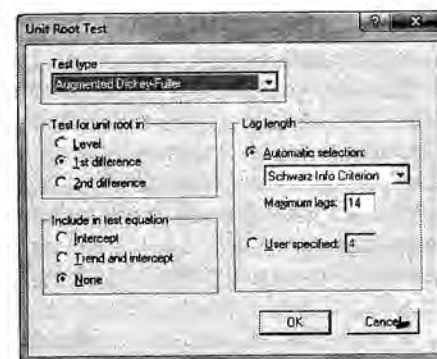


Рис. 4.1. Заполнение мини-окна UNIT ROOT TEST

Шаг 2. Заполнение мини-окна UNIT ROOT TEST

Параметр TEST TYPE (тип теста) установим на опции AUGMENTED DICKEY — FULLER (расширенный тест Дикки — Фуллера), поскольку этот тест чаще всего используется на практике, так как он учитывает возможную автокорреляцию в остатках. Параметр TEST FOR UNIT ROOT IN (тест на единичный корень для...) следует установить на опции 1ST DIFFERENCE (первых разностей), так как при исследовании остатков на стационарность не используются их исходные уровни. Параметр INCLUDE IN TEST EQUATION (включить в тестовое уравнение) установим на опции NONE (не включать тренд или тренд и константу), поскольку в остатках отсутствует тренд и свободный член уравнения (константа). Параметр LAG LENGTH (длина лага) установим на опции AUTOMATIC SELECTION (автоматический выбор), что позволит EViews самостоятельно выбрать длину лага. Вполне естественно, что при необходимости длину лага можно задать самому.

Шаг 3. Интерпретация результатов теста

Теория тестирования стационарности временных рядов изложена ниже. А чтобы просто сделать вывод о стационарности времен-

ного ряда на основе расширенного теста Дикки — Фуллера, нужно знать следующее. После того как ранее мы заполнили мини-окно Unit Root test и щелкнули кнопку ОК, в результате у нас получилась табл. 4.4 с итогами теста. При этом главное внимание нужно обратить на верхнюю строчку теста, выделенную жирным шрифтом: Augmented Dickey — Fuller test statistic (статистика расширенного теста Дикки — Фуллера). Поскольку статистика теста Дикки — Фуллера в этом случае равна 11,05764, а ее значимость (Prob.) равна 0,0000, то нулевая гипотеза о том, что $D(\text{RESID})$ имеет единичный корень, отвергается. Следовательно, мы можем принять альтернативную гипотезу о стационарности полученных остатков.

При этом в табл. 4.4 даются критические значения теста (Test critical values), на основе которых о стационарности остатков можно судить с различным уровнем надежности. Так, в том случае, когда статистика расширенного теста Дикки — Фуллера меньше -2,576127, то вывод о стационарности остатков можно сделать с 99%-ным уровнем надежности, а если меньше -1,942361, но больше -2,576127, то с 95%-ным уровнем надежности. Если интересующая нас статистика меньше -1,615684, но больше -1,942361, то уровень надежности вывода о стационарности остатков снижается до 90%.

В основе теории единичного корня лежит довольно простая формула, которая считается базовой для понимания стационарности в уравнениях авторегрессии:

$$Y_t = \rho Y_{t-1} + e_t, \quad (4.4)$$

где Y_t — результирующая зависимая переменная;

Y_{t-1} — независимая факторная переменная с лагом в один период (в нашем случае в один месяц);

ρ — коэффициент регрессии;

e_t — остатки.

Уравнение авторегрессии 1-го порядка считается стационарным в том случае, когда коэффициент регрессии $\rho < 1$. Соответственно если $\rho > 1$, то оно считается нестационарным, а следовательно, волатильность с течением времени может нарастать и стремиться к бесконечности. Следует заметить, что при необходимости в формулу (4.4) может быть добавлена константа либо константа и тренд, если, конечно, они будут статистически значимыми.

Проверка авторегрессионного процесса на стационарность проводится следующим образом. Согласно нулевой гипотезе, предполагается, что если $\rho = 1$, то временной ряд нестационарный, а в случае ее

опровержения принимается альтернативная гипотеза, утверждающая, что $\rho < 1$, а следовательно, ряд стационарный.

В ходе решения обычного уравнения регрессии рассчитывается t -статистика для коэффициента регрессии ρ , совпадающая с расчетными значениями статистики Дикки — Фуллера, которая потом сравнивается с критическими значениями статистики Дикки — Фуллера (обычно даются в таблице, но в EViews, естественно, мы их получим в готовом виде). Сравнение проводится по одностороннему критерию, но если бы альтернативная гипотеза состояла в утверждении, что $\rho \neq 1$, то тогда мы пользовались бы двусторонним критерием. Поскольку проверка гипотезы проводится по одностороннему критерию, то в этом случае, если расчетное значение t -статистики для коэффициента регрессии ρ будет меньше критического значения статистики Дикки — Фуллера (с поправкой на число наблюдений), нулевая гипотеза о том, что $\rho = 1$ отклоняется и принимается альтернативная гипотеза о том, что $\rho < 1$, а следовательно, временной ряд Y_t можно считать стационарным.

Стандартный тест Дикки — Фуллера проводится после вычитания Y_{t-1} из левой и правой частей уравнения (4.4). В результате мы получаем следующую формулу:

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + e_t. \quad (4.5)$$

Учитывая, что $dY_t = Y_t - Y_{t-1}$, а $\rho Y_{t-1} - Y_{t-1} = (\rho - 1)Y_{t-1}$, и приравняв $\alpha = (\rho - 1)$, получим новое уравнение:

$$dY_t = \alpha Y_{t-1} + e_t. \quad (4.6)$$

С учетом того, что при $\rho = 1$ параметр α становится равным нулю, то соответственно в случае принятия нулевой гипотезы $\alpha = 0$, а если принимается альтернативная гипотеза, то соответственно $|\alpha| < 1$, а следовательно, временной ряд считается стационарным.

Однако на практике большую популярность приобрел расширенный тест Дикки — Фуллера AUGMENTED DICKKEY — FULLER, так как он учитывает возможную автокорреляцию в остатках. При этом в правую часть уравнения (4.6) включаются дополнительные лаговые переменные Y . В результате это уравнение приобретает следующий вид:

$$dY_t = \alpha Y_{t-1} + \beta_1 dY_{t-1} + \beta_2 dY_{t-2} + \dots + \beta_p dY_{t-p} + u_t, \quad (4.7)$$

где β — коэффициенты при дополнительных лаговых переменных;

p — номер включенного дополнительного лага;

d — первые разности дополнительных лаговых переменных;

u — остатки без автокорреляции, или «белый шум».

В дальнейшем эти знания нам потребуются для проверки авторегрессионного процесса 2-го порядка (см. уравнение (4.1)) на стационарность, а пока применим эту теорию для проверки на стационарность остатков, полученных в результате решения этого уравнения. Заполнив в алгоритме № 9 мини-окно UNIT ROOT TEST и щелкнув кнопку ОК, мы фактически решили следующее уравнение регрессии:

$$d_2 e_t = \alpha d e_{t-1} + \beta_1 d_2 e_{t-1} + \beta_2 d_2 e_{t-2} + \dots + \beta_p d_2 e_{t-p} + u_t, \quad (4.8)$$

где e — остатки;

d и d_2 — первые и вторые разности дополнительных лаговых переменных.

В результате решения расширенного теста Дикки — Фуллера мы получили табл. 4.4 с итогами теста, свидетельствующими о стационарности остатков. О том, как мы пришли к этому выводу, подробно рассказано выше (см. алгоритм действий № 9 «Как проверить в EViews остатки на стационарность модели»).

Таблица 4.4

Итоги решения расширенного теста Дикки — Фуллера для остатков, полученных по статистической модели

$$\text{USDOLLAR} = 1,321092 \times \text{USDOLLAR}(-1) + 0,319415 \times \text{USDOLLAR}(-2)$$

| | | | | |
|---|-----------|--|-------------|--------|
| Null Hypothesis (нулевая гипотеза): D(RESID) has a unit root (D(RESID) имеет единичный корень) | | | | |
| Exogenous: None (константы и тренд отсутствуют) | | | | |
| Lag Length: 4 (Automatic based on SIC, MAXLAG = 14 (длина лага рассчитана автоматически на основе информационного критерия Шварца)) | | | | |
| | | | t-Statistic | Prob.* |
| Augmented Dickey — Fuller test statistic (статистика расширенного теста Дикки — Фуллера) | | | -11,05764 | 0,0000 |
| Test critical values (критические значения теста): | | | | |
| | 1% level | | -2,576127 | |
| | 5% level | | -1,942361 | |
| | 10% level | | -1,615684 | |
| *MacKinnon (1996) one-sided <i>p</i> -values (односторонние <i>p</i> -значения Маккин-нона). | | | | |

Окончание

Augmented Dickey — Fuller Test Equation (уравнение расширенного теста Дикки — Фуллера)

Dependent Variable: D(RESID,2)

Method: Least Squares

Date: 05/29/10 Time: 10:00

Sample (adjusted): 1993M02 2010M05

Included observations: 208 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| D(RESID(-1)) | -3,728305 | 0,337170 | -11,05764 | 0,0000 |
| D(RESID(-1),2) | 1,814453 | 0,296399 | 6,121645 | 0,0000 |
| D(RESID(-2),2) | 1,034685 | 0,230079 | 4,497082 | 0,0000 |
| D(RESID(-3),2) | 0,522649 | 0,147981 | 3,531877 | 0,0005 |
| D(RESID(-4),2) | 0,174322 | 0,069687 | 2,501487 | 0,0132 |
| R-squared | 0,824168 | Mean dependent var | | 0,004348 |
| Adjusted R-squared | 0,820703 | S.D. dependent var | | 2,056724 |
| S.E. of regression | 0,870888 | Akaike info criterion | | 2,585139 |
| Sum squared resid | 153,9647 | Schwarz criterion | | 2,665368 |
| Log likelihood | -263,8545 | Durbin — Watson stat | | 2,009083 |

Поскольку мы доказали, что остатки, полученные по модели авторегрессии 2-го порядка без константы, являются стационарными, то, следовательно, можно сделать вывод, что их распределение носит устойчивый характер.

4.3. Описательная статистика и тестирование остатков на нормальное распределение

Теперь нашей задачей является ответить на следующий важный вопрос: является ли распределение полученных остатков нормальным? При составлении интервальных прогнозов мы исходим из предпо-

ложения, что распределение остатков носит нормальный характер, поэтому теперь должны проверить, насколько это утверждение соответствует истине.

Алгоритм действий № 10

Как в EViews получить описательную статистику остатков

Шаг 1. Установка необходимых опций

Чтобы узнать характер распределения остатков, необходимо в рабочем файле открыть файл RESID, а затем выбрать опции VIEW (смотреть)/DESCRIPTIVE STATISTICS (описательная статистика)/STATS TABLE (таблица со статистикой). В результате мы получили табл. 4.5 с описательной статистикой для остатков.

Таблица 4.5

Описательная статистика для остатков, полученных по уравнению авторегрессии 2-го порядка без константы

| | RESID (остатки) |
|---------------------------------------|-----------------|
| Mean (среднее) | 0,060106 |
| Median (медиана) | 0,016478 |
| Maximum (максимум) | 7,613777 |
| Minimum (минимум) | -2,687712 |
| Std. Dev. (стандартное отклонение) | 0,813645 |
| Skewness (коэффициент асимметрии) | 4,939145 |
| Kurtosis (коэффициент эксцесса) | 45,83162 |
| Jarque — Bera (критерий Жарка — Бера) | 17147,64 |
| Probability (значимость) | 0,000000 |
| Sum (сумма) | 12,80254 |
| Sum Sq. Dev. (стандартное отклонение) | 140,3478 |
| Observations (наблюдения) | 213 |

Шаг 2. Интерпретация результатов теста

Во-первых, если Probability (значимость) больше 0,05, то гипотеза о нормальном распределении остатков подтверждается. Поскольку в нашем случае Probability = 0, то гипотеза о нормальном распределении остатков отклоняется. Во-вторых, если коэффициент асимметрии (Skewness) больше нуля, то в остатках наблюдается правосторонняя асимметрия, а если меньше нуля — левосторонняя асимметрия. Судя по табл. 4.5, в этом случае в распределении остатков наблюдается правосторонняя асимметрия. Если коэффициент эксцесса (Kurtosis) больше 3, то наблюдается островершинное распределение, а если меньше — плосковершинное распределение статистического ряда. В этом случае мы наблюдаем «островершинное» распределение остатков. С более подробной интерпретацией описательной статистики можно познакомиться ниже.

Дадим некоторые пояснения к табл. 4.5. Так, среднее (Mean) равно сумме всех остатков, деленной на количество наблюдений. В свою очередь медиана (Median) представляет собой величину, расположенную в середине нечетного ряда, ранжированного в порядке возрастания или убывания. В четном ряде медиана равна среднему значению двух соседних величин, расположенных в середине ряда. Соответственно максимум (Maximum) и минимум (Minimum) означают максимальное и минимальное значения временного ряда.

Стандартное отклонение является мерой дисперсии для временного ряда и находится по формуле для стандартного отклонения по выборке:

$$S = \sqrt{\frac{\sum (e_t - \bar{e})^2}{N - 1}}, \quad (4.9)$$

где e_t — остатки;

\bar{e} — среднее значение остатков;

N — количество наблюдений.

В нашем случае стандартное отклонение имеет следующее значение:

$$S = \sqrt{\frac{140,3478}{212}} = 0,813645.$$

Коэффициент асимметрии является своего рода индикатором, показывающим степень асимметричности распределения статистического ряда. Следует иметь в виду, что в случае полной симметрии, в том числе и при нормальном распределении, коэффициент асимметрии должен быть равен нулю. Если коэффициент асимметрии

меньше нуля, то говорят о левосторонней асимметрии, а если больше нуля, — то о правосторонней асимметрии. Коэффициент асимметрии для остатков в EViews рассчитан по следующей формуле:

$$A = \frac{1}{N} \sum \left(\frac{e_t - \bar{e}}{\hat{\sigma}} \right)^3, \quad (4.10)$$

где расчетное стандартное отклонение $\hat{\sigma}$ рассчитывается по формуле: $\hat{\sigma} = S \times \sqrt{(N-1)/N}$.

В этом случае коэффициент асимметрии имеет следующее значение:

$$A = \frac{1}{213} \times 1052,03793 = 4,939145.$$

Как мы видим, в этом случае коэффициент асимметрии равен 4,939145, следовательно, в распределении остатков наблюдается очень сильная правосторонняя асимметрия, т.е. имеет место преобладание положительных остатков над отрицательными. Отсюда можно сделать вывод, что в динамике курса доллара к рублю чаще наблюдались резкие (вполне очевидно, что незначительные плавные колебания курса легко поддаются прогнозированию) подъемы, чем аналогичные падения. С фундаментальной точки зрения этот факт объясняется многолетней политикой Банка России по поддержанию слабого курса рубля.

Коэффициент эксцесса можно назвать индикатором «крутизны» распределения статистического ряда. Коэффициент эксцесса для нормального распределения равен 3. В том случае, когда этот коэффициент больше 3, это является показателем «островершинного» распределения, а если меньше 3, это свидетельствует о «плосковершинном» распределении статистического ряда. Коэффициент эксцесса для остатков в EViews вычислен по следующей формуле:

$$S = \frac{1}{N} \sum \left(\frac{e_t - \bar{e}}{\hat{\sigma}} \right)^4, \quad (4.11)$$

где расчетное стандартное отклонение $\hat{\sigma}$ находится таким же образом, как и в формуле (4.10).

В нашем случае коэффициент эксцесса имеет следующее значение:

$$S = \frac{1}{213} \times 9762,135525 = 45,83162.$$

Поскольку коэффициент эксцесса равен 45,83162 (см. табл. 4.5), можно сделать вывод, что распределение остатков является «островер-

шинным». По сути это означает, что в этом распределении имеется ярко выраженное ядро плотности распределения, внутри которого диапазон колебаний величины остатков незначителен, и рассеянное «гало», где разброс колебаний величины остатков весьма значителен. С точки зрения предсказания курса доллара такой характер распределения позволяет задавать, например, при 80%-ном уровне надежности, не слишком широкие прогностические интервалы. Правда, если инвестор хочет иметь прогноз с более высоким 99%-ным уровнем надежности, то из-за рассеянного «гало» ширина этих интервалов начинает резко увеличиваться.

В EViews есть возможность посмотреть в графическом виде оценку ядра плотности распределения с помощью опций DISTRIBUTION/KERNEL DENSITY GRAPHS... (распределение/графики ядра плотности распределения). В появившемся мини-окне KERNEL DENSITY (ядро плотности распределения) по умолчанию устанавливается опция EPANECHNICOV, а всего их здесь семь и отличаются они друг от друга по используемому алгоритму сглаживания (рис. 4.2).

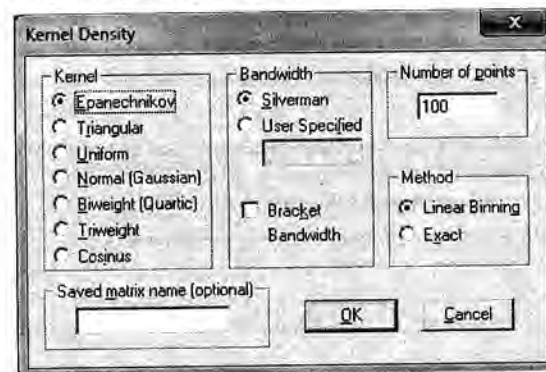


Рис. 4.2. Опция EPANECHNICOV в мини-окне KERNEL DENSITY

Дело в том, что в отличие от обычной гистограммы (столбчатая диаграмма, высота каждого прямоугольника которой пропорциональна частоте распределения в заданном интервале значений) график ядра плотности распределения создается с помощью сглаживания, в ходе которого различным наблюдениям присваиваются определенные веса. При этом соблюдается следующий принцип: чем дальше отдельное наблюдение от оцениваемой «точки», тем более легкий вес ему присваивается. В результате получается диаграмма, приведенная на рис. 4.3, на которой хорошо виден «островершинный» характер ядра плотности распределения остатков.

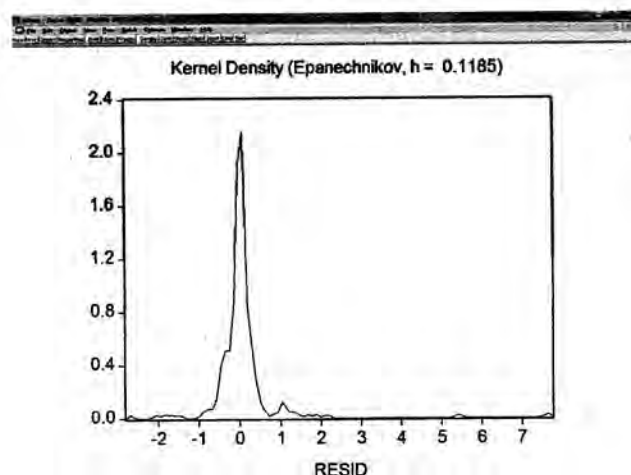


Рис. 4.3. «Острове́ршинный» характер ядра плотности распределения остатков (с правосторонней асимметрией), полученных после решения уравнения авторегрессии 2-го порядка без константы

Источник: расчеты автора.

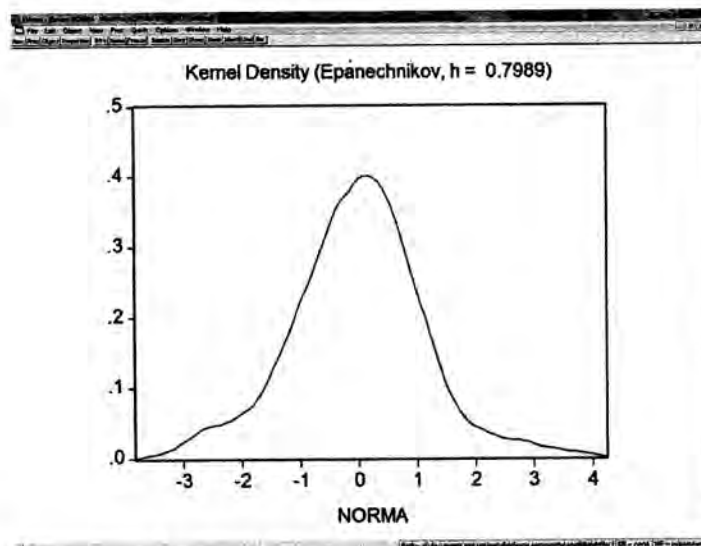


Рис. 4.4. Нормальное распределение с обычным ядром плотности распределения

Для большей наглядности ядро плотности распределения остатков можно сравнить с нормальным распределением, имеющим стандартное ядро плотности распределения (рис. 4.4). С этой целью мы получили в Excel нормальное распределение, используя опции АНАЛИЗ ДАННЫХ/ГЕНЕРАЦИЯ СЛУЧАЙНЫХ ЧИСЕЛ. Сравнив рис. 4.3 и 4.4, легко заметить, что у нормального распределения, во-первых, вершина гораздо более плоская; во-вторых, ядро плотности распределения значительно шире; в-третьих, рассеянное «гало» не столь широко разбросано по краям.

Продолжим анализ характера распределения остатков и с этой целью посмотрим оценку значимости критерия Жарка — Бера, представленную в табл. 4.5. При этом следует иметь в виду, что величина критерия Жарка — Бера служит для проверки нулевой гипотезы о нормальном распределении изучаемого статистического ряда. Тестовая статистика в этом случае измеряет разницу между нормальным распределением и коэффициентами асимметрии и эксцесса, вычисленными для данного статистического ряда. Критерий Жарка — Бера находится по следующей формуле:

$$Jarque - Bera = \frac{N - k}{6} \left[A^2 + \frac{(K - 3)^2}{4} \right], \quad (4.12)$$

где N — количество наблюдений;

A — коэффициент асимметрии;

K — коэффициент эксцесса;

k — количество параметров, использованных для создания данного временного ряда.

Численное значение теста Жарка — Бера получается следующее:

$$Jarque - Bera = \frac{213 - 0}{6} \left[4,939145^2 + \frac{(45,83162 - 3)^2}{4} \right] = 17147,64.$$

После этого значение теста Жарка — Бера сравнивают с распределением χ^2 (хи-квадрат) с двумя степенями свободы. В том случае, если критерий Жарка — Бера $> \chi^2_{крит}$, то делается вывод о неслучайном характере распределения, а следовательно, нулевая гипотеза о нормальном распределении опровергается. В нашем случае значение теста Жарка — Бера равно 17147,64, а следовательно, если сравнить с соответствующим табличным значением $\chi^2_{крит, 0,01; 2} = 9,21$, то рассчитанный нами критерий Жарка — Бера существенно выше последнего.

Впрочем, нам не обязательно заглядывать в таблицу. Чтобы вычислить значимость критерия Жарка — Бера в Excel, достаточно воспользоваться

функцией ХИ2РАСП (17147,64; 2) = 0. А в EViews значимость (*Probability*) критерия Жарка — Бера, равная нулю, выдается автоматически (см. табл. 4.4).

Поскольку при значимости критерия Жарка — Бера (*Probability*) < 0,05 нулевая гипотеза о нормальном распределении опровергается с 95% уровнем надежности, то, следовательно, в нашем случае мы вынуждены отвергнуть гипотезу о нормальном распределении остатков.

В EViews имеется и ряд других тестов, с помощью которых можно провести дополнительную проверку нулевой гипотезы о нормальном распределении. В частности, если в файле RESID воспользоваться опциями VIEW/DISTRIBUTION/EMPIRICAL DISTRIBUTION TESTS... (смотреть/распределение/тесты на проверку характера эмпирического распределения), то мы получим результаты проверки нулевой гипотезы о нормальном распределении остатков с помощью соответствующих тестов Лиллиефорса (*Lilliefors*), Крамера фон Мизеса (*Cramer von Mises*), Уотсона (*Watson*) и Андерсона — Дарлинга (*Anderson — Darling*). Результаты тестирования занесены в табл. 4.6. Поскольку значимость (*Probability*) критериев по всем четырем тестам равна нулю, то нулевая гипотеза о нормальном распределении остатков опровергается.

В связи с опровержением нулевой гипотезы возникает вопрос: можно ли в этой ситуации строить интервальные прогнозы по курсу доллара исходя из предположения о нормальном распределении остатков? Вот как на него отвечает известный американский профессор статистики Стэнфордского университета Т. Андерсон: «Приведенные процедуры проверки гипотез и построения доверительных областей были основаны на предположении о том, что наблюдения распределены нормально. Если предположение о нормальности не выполняется, то эти процедуры все же можно применять для больших выборок, используя асимптотическую теорию...

Значение приведенных теорем (доказывающих асимптотическую теорию. — Прим. авт.) состоит в том, что, опираясь на них, обычную теорию для нормального случая при больших объемах выборок можно использовать с достаточной точностью и в тех ситуациях, когда наблюдения не являются нормально распределенными»¹.

Можно ли применить асимптотическую теорию к распределению остатков, полученных с помощью статистической модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$? Поскольку малой выборкой принято называть выборку, имеющую до 30 степеней свободы, а в нашей выборке имеется 210 степеней свободы, то вполне естественно, что асимптотическую теорию в этом случае можно использовать.

¹ Андерсон Т. Статистический анализ временных рядов. С. 35, 39.

Таблица 4.6

Результаты проверки нулевой гипотезы о нормальном распределении остатков с помощью тестов Лиллиефорса, Крамера фон Мизеса, Уотсона и Андерсона — Дарлинга

| Empirical Distribution Test for RESID | | | | |
|--|-----------|---------------------|-------------|----------|
| Hypothesis: Normal | | | | |
| Date: 06/02/10 Time: 16:25 | | | | |
| Sample (adjusted): 1992M08 2010M04 | | | | |
| Included observations: 213 after adjustments | | | | |
| Method | Value | Adj. Value | Probability | |
| Lilliefors (D) | 0,266860 | NA | 0,0000 | |
| Cramer von Mises (W2) | 5,303318 | 5,315767 | 0,0000 | |
| Watson (U2) | 5,208463 | 5,220689 | 0,0000 | |
| Anderson — Darling (A2) | 29,67056 | 29,77651 | 0,0000 | |
| Method: Maximum Likelihood — d.f. corrected (Exact Solution) | | | | |
| Parameter | Value | Std. Error | z-Statistic | Prob. |
| MU | 0,060106 | 0,055750 | 1,078131 | 0,2810 |
| SIGMA | 0,813645 | 0,039514 | 20,59126 | 0,0000 |
| Log likelihood | -257,8066 | Mean dependent var. | | 0,060106 |
| No. of Coefficients | | S.D. dependent var. | | 0,813645 |

Почему столь важно строить интервальные прогнозы по курсу доллара исходя из предположения о нормальном распределении остатков? Дело в том, что нормальный закон распределения играет важнейшую роль в теории вероятностей. При этом главной особенностью этого закона является тот факт, что он является предельным законом, к которому — при определенных условиях — приближаются другие законы распределения. Предполагая, что остатки распределены согласно закону о нормальном распределении (т.е. их распределение определяется воздействием множества случайных причин), мы тем самым приписываем им следующие свойства, благоприятные для построения интервальных прогнозов.

Во-первых, график плотностей вероятностей нормального распределения (см. рис. 4.4) имеет колоколообразную форму, симметричную от-

носителем средней (математического ожидания) μ . При этом плотность вероятностей нормального распределения определяется по формуле

$$f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2s^2}}, \quad (4.13)$$

где s — стандартное отклонение.

Следовательно, плотность вероятностей нормального распределения полностью определяется двумя параметрами остатков — их средней величиной (математического ожидания) μ и стандартным отклонением s .

Во-вторых, график плотностей вероятностей нормального распределения показывает, что для нормально распределенных остатков вероятность отклонения от их средней величины (математического ожидания) μ быстро уменьшается с ростом этого отклонения.

В-третьих, если $\mu = 0$, а стандартное отклонение $s = 1$, то нормальное распределение с такими параметрами называется нормированным. При этом плотность вероятностей нормированного нормального распределения определяется по следующей формуле:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (4.14)$$

В-четвертых, если функцию нормированной плотности вероятностей распределения $f(x)$ перевести в проценты (при 1 = 100%), а затем построить график плотности вероятностей нормированного нормального распределения, то мы получим диаграмму, изображенную на рис. 4.5.

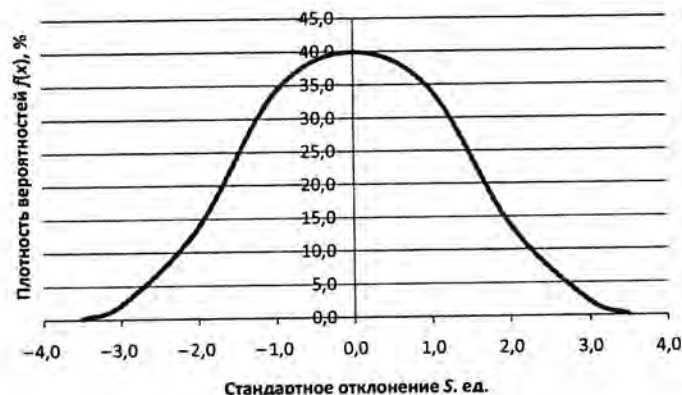


Рис. 4.5. Нормированная плотность вероятностей нормального распределения

Исходя из этого рисунка можно прийти к выводу: если мы будем суммировать (интегрировать) вероятность попадания остатка в область интервального прогноза при стандартном отклонении s , то выяснится, что в этом случае в область интервального прогноза попадет 68,17% всех нормально распределенных остатков. Соответственно при стандартном отклонении $s = \pm 2$ в область интервального прогноза попадет 95,45% всех нормально распределенных остатков, а при стандартном отклонении $s = \pm 3$ в область интервального прогноза попадет 99,73% всех нормально распределенных остатков. Заметим также, что чаще всего интервальные прогнозы строят исходя из 95%-ного уровня надежности при нормированном стандартном отклонении, равном 1,96; либо при 99%-ном уровне надежности при нормированном стандартном отклонении, равном 2,58; либо при 99,9%-ном уровне надежности при нормированном стандартном отклонении 3,29.

4.4. Построение точечных и интервальных прогнозов

Таким образом, предположение о нормальном распределении остатков позволяет нам строить интервальные прогнозы исходя из определенных доверительных интервалов, точность которых можно проверить. Именно этим мы сейчас и займемся.

Однако прежде чем перейти к составлению интервальных прогнозов, нам необходимо исходя из уже решенного уравнения регрессии (4.3) составить точечный прогноз на май 2010 г. Последняя дата выбрана не случайно, так как это позволяет нам смоделировать ситуацию реального прогноза. Дело в том, что в этом случае в качестве базы данных мы использовали информацию по ежемесячному курсу доллара за период с июня 1992 г. по апрель 2010 г., а точечный прогноз составили на май 2010 г. Именно по такой схеме обычно в реальной жизни и делаются прогнозы.

Алгоритм действий № 11

Как в EViews построить точечный прогноз

С помощью опции FORECAST находим точечный прогноз на будущий месяц. В EViews точечные прогнозы выдаются автоматически: сразу же после использования опции FORECAST (см. алгоритм действий № 8 «Как оценить точность статистической модели в EViews») у нас в рабочем файле появляется файл USDollar.f. Поскольку в ка-

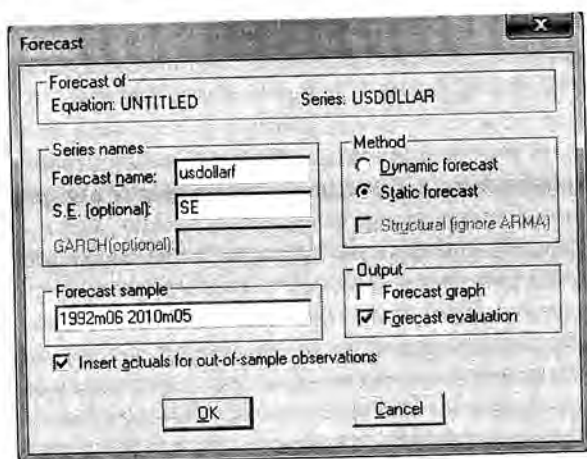


Рис. 4.6. Построение точечного прогноза в мини-окне FORECAST на май 2010 г.

честве конечной даты для прогноза нас интересует май 2010 г., то при этом нужно проследить, чтобы в опции FORECAST SAMPLE (выборка для прогноза) последней датой был именно этот месяц, который в EViews обозначается как 2010m05 (рис. 4.6). Таким образом, открыв файл USDollarf, мы увидим предсказанные значения курса доллара с августа 1992 г. (прогноз на июнь и июль 1992 г. не строится из-за потери двух наблюдений, обусловленной созданием двух факторных лаговых переменных) по май 2010 г. В результате мы выяснили, что точечный прогноз курса доллара на конец мая 2010 г. у нас равен 29 руб. 31,37 коп. Этот прогноз на май 2010 г. (как, впрочем, и точечные прогнозы на другие месяцы) был рассчитан путем подстановки в формулу (4.3) соответствующих значений переменных:

$$\begin{aligned} \text{USDOLLAR} &= 1,321092 \times \text{USDOLLAR}(-1) - 0,319415 \times \\ &\times \text{USDOLLAR}(-2) \times \text{USDOLLAR} = 1,321092 \times 29,28860 - \\ &- 0,319415 \times 29,36380 = 29,31370. \end{aligned}$$

Алгоритм действий № 12

Как в EViews построить интервальные прогнозы

Шаг 1. Как найти средние ошибки прогнозируемого курса доллара

Чтобы одновременно с точечным прогнозом вычислить величину интервального прогноза, нам следует в мини-окне FORECAST (см. шаг 2

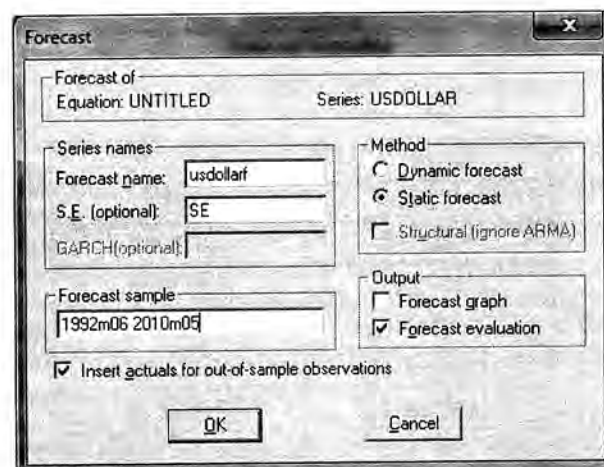


Рис. 4.7. Использование дополнительной опции S.E. в мини-окне FORECAST

алгоритма действий № 8 — заполнение мини-окна FORECAST) использовать дополнительную опцию S.E. (optional). Например, написать в этой опции аббревиатуру SE в качестве названия для нового файла, в который мы собираемся поместить средние ошибки прогнозируемого индивидуального значения курса доллара (рис. 4.7). В отличие от широко используемого (из-за простоты в расчетах) стандартного отклонения средняя ошибка прогнозируемого индивидуального значения включает не только стандартную ошибку, но и случайную ошибку (см. далее — «Математические подробности, связанные с расчетом интервальных прогнозов»), а потому делает интервал прогноза более надежным. Причем разница между стандартным отклонением и средней ошибкой прогнозируемого индивидуального значения нарастает из-за резких колебаний факторной переменной (предыдущего значения курса доллара), в то время как в обычные периоды она незначительна.

В результате появится файл SE, открыв который можно увидеть значения средних ошибок прогнозируемого индивидуального значения курса доллара. Они показаны в табл. 4.7, правда, в целях экономии приведены данные только за 1992 и 2010 гг., в то время как в EViews они приводятся полностью.

Шаг 2. Вычисление интервальных прогнозов

После нахождения средней ошибки прогнозируемого индивидуального значения курса доллара на май 2010 г. следующей задачей

Таблица 4.7

Средние ошибки прогнозируемого индивидуального значения курса доллара

| | |
|--------------------------------|----------|
| Last updated: 05/31/10 — 12:28 | |
| 1992M06 | NA |
| 1992M07 | NA |
| 1992M08 | 0,817807 |
| 1992M09 | 0,817808 |
| 1992M10 | 0,817810 |
| 1992M11 | 0,817857 |
| 1992M12 | 0,817810 |
| ... | ... |
| 2010M01 | 0,821369 |
| 2010M02 | 0,821200 |
| 2010M03 | 0,822092 |
| 2010M04 | 0,822312 |
| 2010M05 | 0,821060 |

является определение доверительного интервала прогноза, величина которого зависит от заданного уровня надежности. При этом границы доверительного интервала для 95%-ного уровня надежности (в случае необходимости и для иных уровней надежности) и для степеней свободы $n = 214 - k - 1 = 214 - 2 - 1 = 211$ (где 214 — количество наблюдений во временном ряде, а k — количество факторных переменных) в Excel находятся с помощью функции СТЬЮДРАСПОБР, которая возвращает двустороннее значение t -критерия Стьюдента как функцию вероятности и числа степеней свободы. Двустороннее значение t -критерия в этом случае используется, поскольку необходимо определить двусторонний интервал (Нижняя граница интервального прогноза \leq Фактический курс доллара \leq Верхняя граница интервального прогноза), в котором прогнозируемый курс доллара должен оказаться с 95%-ным уровнем надежности. (Для справки заметим, что односторонний t -критерий используется в том случае, если мы поставим перед собой задачу найти такое значение Y , которое с определенным уровнем надежности будет больше X ; либо, напротив, прямо противоположную задачу — когда Y будет меньше X . Односторонний критерий менее консервативен, поскольку когда

мы вычисляем (с определенным уровнем надежности) вероятность того, что значение $Y > X$, то в этом случае вероятность того, что значения $Y < X$, в расчетах не учитывается. Это же происходит и при расчетах вероятности того, что значения $Y < X$.)

Таким образом, в нашем случае t -значение находится следующим образом:

$$\text{СТЮДРАСПОБР}(1 - 0,95 = 0,05; 211) = 1,9713.$$

(При определении границы доверительного интервала, например с 99%-ным уровнем надежности t -значение имеет следующую величину: $\text{СТЮДРАСПОБР}(1 - 0,99 = 0,01; 211) = 2,5993$. Вполне очевидно, что таким образом можно найти двусторонние t -значения для любого заданного уровня надежности.)

После того как мы нашли t -значение для 95%-ного уровня надежности, появляется возможность составить интервальный прогноз на конец мая 2010 г., т.е. вычислить как нижнюю, так и верхнюю границу прогноза курса доллара на эту дату:

$$\hat{Y}_{\min} = \hat{Y}_{\text{май 2010}} - t \times SE_{\text{ind}} = 29,3137 - 1,9713 \times 0,82106 = 27,6952; \quad (4.15)$$

$$\hat{Y}_{\max} = \hat{Y}_{\text{май 2010}} + t \times SE_{\text{ind}} = 29,3137 + 1,9713 \times 0,82106 = 30,9322, \quad (4.16)$$

где Y_{\min} — нижняя граница интервального прогноза;

Y_{\max} — верхняя граница интервального прогноза;

$\hat{Y}_{\text{май 2010}}$ — точечный прогноз на конец мая 2010 г.;

t — значение двустороннего t -критерия для 95%-ного уровня надежности;

SE_{ind} — средняя ошибка прогнозируемого индивидуального значения курса доллара на май 2010 г.

Математические подробности, связанные с расчетом интервальных прогнозов

Тем, кому интересно знать, как мы получили табл. 4.7, дадим необходимое пояснение. EViews вычисляет среднюю ошибку индивидуального значения курса доллара следующим образом:

$$SE_{\text{ind}} = S \times \sqrt{(1 + X_t^T (X^T X)^{-1} X_t)}, \quad (4.17)$$

где X — матрица исходных значений факторных переменных по всему временному ряду;
 X^T — транспонированная матрица исходных значений факторных переменных по всему временному ряду;

X_t — матрица-столбец значений факторных переменных для момента времени t ;

X_t^T — транспонированная матрица-столбец значений факторных переменных для момента времени t ;

S — стандартное отклонение уравнения регрессии.

При этом стандартное отклонение уравнения регрессии находим по формуле

$$S = \sqrt{\frac{\sum e_i^2}{n-2}}, \quad (4.18)$$

где e — остатки (или отклонения прогноза от фактического значения курса доллара);

n — количество наблюдений.

Для справки заметим, что в Excel умножение матриц производится с помощью функции МУМНОЖ, а обратная матрица $(X^T X)^{-1}$ находится с помощью функции МОБР.

Для нашего случая X -матрицу исходных факторных значений по всему временному ряду в EViews можно найти, воспользовавшись опциями EQUATION/PROC/MAKE REGRESSOR GROUP (уравнение/выполнить/создать группу регрессоров).

В результате этого мы получим три столбца с результативной переменной USDOLLAR и факторными переменными USDOLLAR(-1) и USDOLLAR(-2) за весь период с июня 1992 г. по апрель 2010 г. Убрав столбец с результативной переменной, мы тем самым получили X -матрицу исходных факторных значений по всему временному ряду (табл. 4.8). Правда, в целях экономии места в этой таблице представлены данные лишь за 1992 и 2010 гг. Причем крайний правый столбец с датировкой наблюдений здесь нами дан только для справки, а в X -матрицу исходных факторных значений входят только два столбца, выделенные жирным шрифтом, с двумя факторными переменными. Следует также заметить, что если бы уравнение регрессии было с константой, то в табл. 4.8 нам пришлось бы поместить дополнительный столбец с единичным вектором.

Таблица 4.8

X -матрица исходных факторных значений по всему временному ряду для уравнения регрессии 2-го порядка без константы

| Наблюдения | USDOLLAR(-1) | USDOLLAR(-2) |
|------------|---------------------|---------------------|
| 1992M08 | 0,1612 | 0,12526 |
| 1992M09 | 0,205 | 0,1612 |
| 1992M10 | 0,254 | 0,205 |
| 1992M11 | 0,398 | 0,254 |
| 1992M12 | 0,447 | 0,398 |
| ... | ... | ... |
| 2010M01 | 30,2442 | 29,8179 |
| 2010M02 | 30,4312 | 30,2442 |
| 2010M03 | 29,9484 | 30,4312 |
| 2010M04 | 29,3638 | 29,9484 |
| 2010M05 | 29,2886 | 29,3638 |

В свою очередь X^T — транспонированная матрица исходных факторных значений по всему временному ряду. По определению она представляет собой матрицу, столбцами которой являются строки X -матрицы исходных факторных значений по всему временному ряду. Кстати, в Excel транспонированную матрицу X^T можно получить следующим образом: обвести значения исходной матрицы X , скопировать их, а затем, вставляя данные, выбрать в мини-окне СПЕЦИАЛЬНАЯ ВСТАВКА опцию ТРАНСПОНИРОВАТЬ (рис. 4.8).

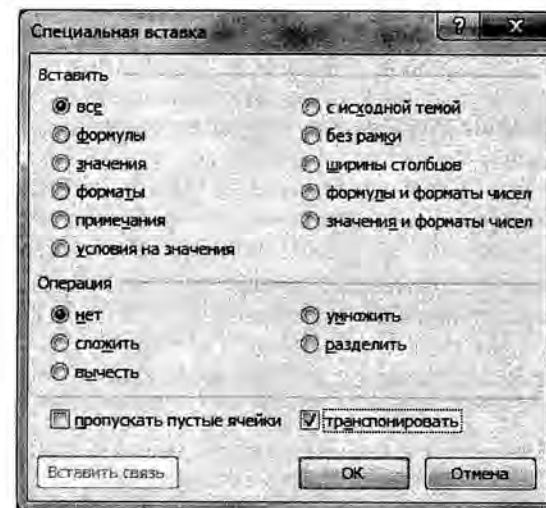


Рис. 4.8. Получение X^T -матрицы при помощи мини-окна СПЕЦИАЛЬНАЯ ВСТАВКА и опции ТРАНСПОНИРОВАТЬ

Следует также иметь в виду, что при расчете средней ошибки индивидуального значения курса доллара на май 2010 г. X_t -матрица-столбец факторных значений для момента времени t приобретает следующий вид:

29,2886

29,3638.

X_t^T -матрица-строка факторных значений для момента времени t в этом случае выглядит таким образом:

29,2886

29,3638.

Следовательно, при расчете средней ошибки индивидуального значения курса доллара на каждый месяц как матрица-столбец X_t , так и матрица-строка X_t^T приобретают разные значения. И еще один важный момент: если у нас было бы уравнение регрессии со свободным членом, то в матрицу-столбец и в матрицу-строку следовало бы добавить по единице.

В частности, средняя ошибка прогнозируемого индивидуального значения курса доллара на май 2010 г. по формуле (4.17) вычисляется следующим образом:

$$SE_{ind} = 0,8178 \times 1,003983 = 0,82106.$$

В зависимости от того, какой уровень волатильности курса доллара инвестор ожидает в будущем месяце, он может составить интервальные прогнозы с разными уровнями надежности. Так, если на рынке ожидают очень высокую волатильность в динамике курса американской валюты, то в этом случае интервальный прогноз целесообразно составить, например, с 99%-ным уровнем надежности, а если относительно низкую, то требуемый уровень надежности можно уменьшить до 95%.

При этом следует иметь в виду: чем выше уровень надежности, тем больше t -значение и шире интервал прогноза, а следовательно, ниже точность прогноза. Как в зависимости от уровня надежности меняется диапазон интервального прогноза, можно посмотреть в табл. 4.9, в которой представлены интервальные прогнозы на май 2010 г., рассчитанные с разными уровнями надежности. Например, при 10%-ном уровне надежности диапазон интервального прогноза составил лишь 20,66 коп., в то время как при 99,9%-ном уровне надежности диапазон интервального прогноза вырос до 5 руб. 48,05 коп.

Таблица 4.9

**Интервальные прогнозы по курсу доллара на май 2010 г.
с разными уровнями надежности**

| Уровень надежности, % | t -значение | Нижняя граница интервального прогноза, руб. | Верхняя граница интервального прогноза, руб. | Диапазон интервального прогноза, руб. |
|-----------------------|---------------|---|--|---------------------------------------|
| 99,9 | 3,3375 | 26,5734 | 32,0540 | 5,4805 |
| 99,0 | 2,5994 | 27,1794 | 31,4480 | 4,2686 |
| 95,0 | 1,9713 | 27,6951 | 30,9323 | 3,2372 |
| 90,0 | 1,6521 | 27,9572 | 30,6702 | 2,7130 |
| 80,0 | 1,2856 | 28,2581 | 30,3693 | 2,1111 |
| 70,0 | 1,0390 | 28,4606 | 30,1668 | 1,7062 |
| 60,0 | 0,8433 | 28,6213 | 30,0061 | 1,3849 |
| 50,0 | 0,6757 | 28,7589 | 29,8685 | 1,1095 |
| 40,0 | 0,5252 | 28,8825 | 29,7449 | 0,8624 |
| 30,0 | 0,3858 | 28,9969 | 29,6305 | 0,6336 |
| 20,0 | 0,2537 | 29,1054 | 29,5220 | 0,4166 |
| 10,0 | 0,1258 | 29,2104 | 29,4170 | 0,2066 |

Интуитивно нетрудно понять: чем шире диапазон интервального прогноза, тем выше вероятность реализации прогноза, а следовательно, выше и его надежность. Так, фактический курс доллара к рублю на конец мая 2010 г. составил 30 руб. 49,56 коп., т.е. оказался на 1 руб. 18,19 коп. выше точечного прогноза. Несмотря на довольно значительное отклонение, мы все же не ошиблись с интервальным прогнозом, однако только в том случае, когда составили его с 90%-ным и выше уровнем надежности, в то время как при более низком уровне надежности он оказался бы неточным.

4.5. Проверка точности составленных интервальных прогнозов

И еще один важный момент: мы уже ранее говорили, что интервальные прогнозы составляются исходя из предположения о нормальном распределении остатков, однако в действительности их распределение нельзя назвать нормальным. В связи с этим возникает вопрос: насколько, например, 95%-ный уровень надежности соответствует фактическому попаданию курса доллара в интервал прогноза?

Чтобы ответить на этот важный вопрос, нам необходимо сделать следующее. Во-первых, решить в EViews уравнение авторегрессии 2-го порядка без константы (см. алгоритм действий № 6 «Как решить уравнение регрессии в EViews»). Во-вторых, найти точечные прогнозы с помощью опции FORECAST (см. алгоритм действий № 8 «Как оценить точность статистической модели в EViews»), в которой надо обязательно заполнить дополнительную опцию S.E. (optional) аббревиатурой SE. Таким образом, в файле SE у нас появятся средние ошибки прогнозируемого индивидуального значения курса доллара (см. табл. 4.7). В-третьих, чтобы вывести на экран точечные прогнозы, нам необходимо выбрать опции EQUATION/VIEW/ACTUAL, FITTED, RESIDUAL /ACTUAL, FITTED, RESIDUAL TABLE (уравнение/вид/фактические, расчетные значения, остатки/таблица фактических, расчетных значений и остатков). Так мы получим данные по фактическим и предсказанным значениям курса доллара и по величине остатков за весь период с августа 1992 г. по май 2010 г., т.е. за 214 месяцев (табл. 4.10). В результате у нас получилась табл. 4.11, в которую мы в целях экономии места занесем фактические и предсказанные значения по курсу доллара лишь за два небольших периода — с августа по декабрь 1992 г. и с июля 2008 г. по май 2010 г.

Таблица 4.10

Доля точных интервальных прогнозов при разных уровнях надежности
в период с августа 1992 г. по май 2010 г. (214 наблюдений)

| t-значение для двусто- роннего критерия | Уровень надеж- ности, % | Количество точных прогнозов | Доля точных интервальных прогнозов, % | Разница между долей точных интервальных прогнозов и уровнем надежности, проц. пункт |
|--|----------------------------------|-----------------------------------|---|---|
| 3,3375 | 99,9 | 212 | 99,1 | -0,8 |
| 2,5994 | 99,0 | 211 | 98,6 | -0,4 |
| 1,9713 | 95,0 | 206 | 96,3 | 1,3 |
| 1,6521 | 90,0 | 203 | 94,9 | 4,9 |
| 1,2856 | 80,0 | 199 | 93,0 | 13,0 |
| 1,0390 | 70,0 | 194 | 90,7 | 20,7 |
| 0,8433 | 60,0 | 190 | 88,8 | 28,8 |
| 0,6757 | 50,0 | 187 | 87,4 | 37,4 |
| 0,5252 | 40,0 | 177 | 82,7 | 42,7 |
| 0,3858 | 30,0 | 157 | 73,4 | 43,4 |
| 0,2537 | 20,0 | 130 | 60,7 | 40,7 |
| 0,1258 | 10,0 | 98 | 45,8 | 35,8 |

Источник: расчеты автора.

В-четвертых, используя данные в файле SE по средним ошибкам прогнозируемого индивидуального значения курса доллара составим с 95%-ным уровнем надежности интервальные прогнозы за весь период с августа 1992 г. по май 2010 г. При этом расчеты нижней и верхней границы интервальных прогнозов будем делать согласно формулам (4.15) и (4.16). Таким образом, табл. 4.11 будет дополнена еще и интервальными прогнозами по включенным в нее наблюдениям.

После проведения соответствующих подсчетов удалось выяснить, что при 95%-ном уровне надежности из 214 составленных нами интервальных прогнозов в 206 случаях фактический курс доллара оказался в рамках интервального прогноза, т.е. был точным. Следовательно, при 95%-ном уровне надежности фактическая вероятность точного интервального прогноза достигла 96,3%, т.е. получилась на 1,3 процентного пункта выше заданного уровня надежности.

А теперь посмотрим, насколько удачными были наши интервальные прогнозы при других уровнях надежности. С этой целью мы должны: во-первых, найти t-значения для каждого уровня надежности; во-вторых, для каждого уровня надежности вычислить по формулам (4.15)

и (4.16) верхние и нижние границы интервального прогноза, используя при этом данные из файла SE по средним ошибкам прогнозируемых индивидуальных значений курса доллара; и, в-третьих, подсчитать для каждого уровня надежности количество попаданий фактического курса доллара в диапазон интервального прогноза.

Таблица 4.11

Интервальные и точечные прогнозы по курсу доллара с августа
по декабрь 1992 г. и с июля 2008 г. по май 2010 г.

| Месяц | Факт* | Прогноз | Нижняя граница интерваль- ного прогноза | Верхняя граница интерваль- ного прогноза | Точность интерваль- ного прогноза |
|---------|---------|---------|---|--|---|
| 1992M08 | 0,2050 | 0,1730 | -1,4392 | 1,7851 | В интервале прогноза |
| 1992M09 | 0,2540 | 0,2193 | -1,3928 | 1,8315 | В интервале прогноза |
| 1992M10 | 0,3980 | 0,2701 | -1,3421 | 1,8822 | В интервале прогноза |
| 1992M11 | 0,4470 | 0,4447 | -1,1676 | 2,0569 | В интервале прогноза |
| 1992M12 | 0,4145 | 0,4634 | -1,1488 | 2,0756 | В интервале прогноза |
| ... | ... | ... | ... | ... | ... |
| 2008M07 | 23,4456 | 23,4069 | 21,7900 | 25,0238 | В интервале прогноза |
| 2008M08 | 24,5769 | 23,4812 | 21,8650 | 25,0974 | В интервале прогноза |
| 2008M09 | 25,2464 | 24,9795 | 23,3576 | 26,6014 | В интервале прогноза |
| 2008M10 | 26,5430 | 25,5026 | 23,8843 | 27,1209 | В интервале прогноза |
| 2008M11 | 27,6060 | 27,0017 | 25,3773 | 28,6261 | В интервале прогноза |
| 2008M12 | 29,3804 | 27,9918 | 26,3696 | 29,6140 | В интервале прогноза |
| 2009M01 | 35,4146 | 29,9964 | 28,3642 | 31,6286 | Выше интервала прогноза |
| 2009M02 | 35,7205 | 37,4014 | 35,6115 | 39,1913 | В интервале прогноза |

Окончание

| Месяц | Факт* | Прогноз | Нижняя граница интервального прогноза | Верхняя граница интервального прогноза | Точность интервального прогноза |
|---------|---------|---------|---------------------------------------|--|---------------------------------|
| 2009M03 | 34,0134 | 35,8781 | 34,2566 | 37,4996 | Ниже интервала прогноза |
| 2009M04 | 33,2491 | 33,5252 | 31,8869 | 35,1635 | В интервале прогноза |
| 2009M05 | 30,9843 | 33,0607 | 31,4362 | 34,6852 | Ниже интервала прогноза |
| 2009M06 | 31,2904 | 30,3129 | 28,6642 | 31,9616 | В интервале прогноза |
| 2009M07 | 31,7555 | 31,4407 | 29,8213 | 33,0601 | В интервале прогноза |
| 2009M08 | 31,5687 | 31,9573 | 30,3373 | 33,5773 | В интервале прогноза |
| 2009M09 | 30,0922 | 31,5620 | 29,9421 | 33,1819 | В интервале прогноза |
| 2009M10 | 29,0488 | 29,6711 | 28,0390 | 31,3032 | В интервале прогноза |
| 2009M11 | 29,8179 | 28,7642 | 27,1388 | 30,3896 | В интервале прогноза |
| 2009M12 | 30,2442 | 30,1136 | 28,4930 | 31,7342 | В интервале прогноза |
| 2010M01 | 30,4312 | 30,4311 | 28,8119 | 32,0503 | В интервале прогноза |
| 2010M02 | 29,9484 | 30,5420 | 28,9231 | 32,1609 | В интервале прогноза |
| 2010M03 | 29,3638 | 29,8444 | 28,2238 | 31,4650 | В интервале прогноза |
| 2010M04 | 29,2886 | 29,2263 | 27,6053 | 30,8473 | В интервале прогноза |
| 2010M05 | 30,4956 | 29,3137 | 28,8770 | 32,1142 | В интервале прогноза |

* Данные за 1992 г. с учетом деноминации рубля, проведенной в 1998 г.

Источник: расчеты автора.

После проведения этих вычислений получилась табл. 4.10, в которой заданные уровни надежности сопоставляются с фактической долей точных прогнозов. Судя по этой таблице, доля точных прогнозов оказалась выше заданного уровня надежности при 95%-ном и более низких уровнях надежности. Причем эта положительная разница растет при снижении уровня надежности, достигая своего максимума при 30%-ном уровне надежности, когда она равна 43,4 процентного пункта.

Однако поскольку в этом случае мы оцениваем долю точных интервальных прогнозов, полученных по уравнению авторегрессии за 18-летний период, то это приводит к определенным сложностям. В этом легко убедиться, если посмотреть на формулу (4.17): $SE_{ind} = S\sqrt{1 + X_i^T(X^T X)^{-1}X_i}$, по которой вычисляются средние ошибки прогнозируемого индивидуального значения курса доллара. Судя по этой формуле, S (стандартное отклонение) является постоянным множителем при расчетах для всех наблюдений, в то время как выражение $X_i^T(X^T X)^{-1}X_i$ способствует увеличению средней ошибки для каждого конкретного наблюдения. Причем увеличение средней ошибки прогнозируемого индивидуального значения зависит только от X_i (матрица-столбец значений факторных переменных для момента времени t) и X_i^T (матрица-строка значений факторных переменных для момента времени t), поскольку X и X^T представляют собой соответственно обычную и транспонированную матрицы исходных значений факторных переменных по всему временному ряду.

Отсюда можно сделать вывод, что средняя ошибка прогнозируемого индивидуального значения курса доллара по своей сути является суммой стандартного отклонения и индивидуального значения средней ошибки. При этом рост последнего слагаемого зависит в нашем случае от значений переменных USDollar(-1) USDollar(-2) в момент прогнозирования, а если сказать точнее, то от их разницы, что фактически означает рост волатильности на валютном рынке. Именно поэтому максимального своего значения, равного 0,975232 (см. файл SE), средняя ошибка прогнозируемого индивидуального значения курса доллара достигла при прогнозировании на октябрь 1998 г., поскольку по итогам предыдущего месяца USDollar(-1) был равен 16,0645 руб., а USDollar(-2) — 7,9050 руб., т.е. разница между ними составила более 8 руб. Для справки также заметим, что в сентябре 1998 г. разница между предсказанным и фактическим курсом доллара составила 7 руб. 61,37 коп., т.е. также была наивысшей за все время наблюдений.

В целом же по всему исследуемому временному ряду средняя ошибка прогнозируемого индивидуального значения курса доллара не слишком заметно отличается от стандартного отклонения, которое равно 0,817803.

Так, если в августе 1992 г. величина средней ошибки составила 0,817807, то в мае 2010 г. она была чуть больше и равнялась 0,821060 (выше первой цифры лишь в 1,004 раза). Таким образом, в течение 18 лет средняя ошибка прогнозируемого индивидуального значения курса доллара в основном лишь незначительно отступала от стандартного отклонения, в то время как фактический курс доллара вырос с 0,1253 руб.¹ в июне 1992 г. до 30,4956 руб. в мае 2010 г., т.е. увеличился в 243,5 раза!

С учетом методики расчета нижней и верхних границ интервального прогноза (см. формулы (4.15) и (4.16)) становится вполне понятным, почему у нас в результате возникла проблема избыточной ширины интервального прогноза. Так, при прогнозе на август 1992 г. общий диапазон интервального прогноза (верхняя граница интервального прогноза минус нижняя граница интервального прогноза) при 95%-ном уровне надежности составил 3 руб. 22,43 коп. (см. табл. 4.11), в то время как фактическое значение курса доллара было равно лишь 20,5 коп. В свою очередь при прогнозе на май 2010 г. общий диапазон интервального прогноза был равен 3 руб. 23,72 коп., а фактический курс доллара составил 30 руб. 49,56 коп. Легко подсчитать, что в августе 1992 г. диапазон интервального прогноза в 15,7 раза превышал фактический курс доллара, в то время как в мае 2010 г. его доля в стоимости курса американской валюты составила вполне приемлемые 10,62%.

Почему для начальных наблюдений временного ряда у нас получился столь широкий диапазон интервального прогноза? Как построить статистическую модель с приемлемым диапазоном интервального прогноза? Стоит ли при этом исключать из расчетной базы данных часть наблюдений? И если исключать часть наблюдений все-таки необходимо, то как определить оптимальную выборку данных, которая необходима нам для составления предсказаний с оптимальным диапазоном интервального прогноза?

Чтобы ответить на эти вопросы, необходимо сделать следующее. Во-первых, познакомиться с такими понятиями, как устойчивость прогнозностической модели к внешним шокам; во-вторых, узнать, какого рода бывают изменения структурной стабильности временного ряда; в-третьих, научиться проводить тесты Чоу на структурную стабильность и на точность прогноза; в-четвертых, освоить методику проведения теста Д. Гуйарати по определению характера структурного сдвига; и, в-пятых, на основе результатов последнего теста научиться выделять выборку данных, необходимую для получения предсказаний с оптимальным диапазоном интервального прогноза.

¹ В этом и следующем абзацах стоимость доллара в рублях приводится с учетом деноминации в 1998 г.

Контрольные вопросы и задания

1. При каком уровне надежности статистически значим свободный член уравнения авторегрессии, если его p -значение равно 0,037226? Стоит ли его включать в уравнение, если мы хотим составить уравнение регрессии с 99%-ным уровнем надежности?
2. С помощью какого алгоритма действий уравнения авторегрессии проверяются на автокорреляцию в остатках? При использовании LM -теста Бройша — Годфри какой лаг следует установить в мини-окне LAG SPECIFICATION при тестировании уравнений авторегрессии 1-го AR(1), 2-го AR(2) и 3-го порядков AR(3)? В каком случае LM -тест Бройша — Годфри свидетельствует об отсутствии автокорреляции в остатках?
3. С помощью какого алгоритма действий проверяются остатки на стационарность? Используются ли при тестировании остатков на стационарность их исходные уровни или первые разности? В каком случае результаты расширенного теста Дикки — Фуллера показывают стационарность остатков?
4. С помощью какого алгоритма действий можно получить описательную статистику? Назовите тест, с помощью которого остатки определяются на нормальное распределение? Как интерпретируются результаты этого теста? В каком случае можно говорить о левосторонней или правосторонней асимметрии в остатках, их «островершинном» или «плосковершинном» распределении?
5. Каким образом в EViews можно рассчитать точечный прогноз? Можно ли строить интервальные прогнозы исходя из их нормального распределения, если тестирование показало, что их распределение нельзя считать нормальным? Если — да, то в каком случае это можно делать?
6. Внимательно изучите табл. 4.10, а затем ответьте на следующие вопросы. Назовите уровень надежности, при котором доля точных интервальных прогнозов в большей степени соответствует заданному уровню надежности. При каком уровне надежности разница между фактическим и заданным уровнем надежности достигает своего максимума? Какую долю точных интервальных прогнозов можно получить, снизив заданный уровень надежности до 90%?
7. Почему в полученной статистической модели возникла проблема избыточной ширины интервального прогноза? Подтвердите наличие этой проблемы конкретными цифрами. Как избыточный интервальный прогноз отражается на качестве прогнозирования?

ГЛАВА 5

ТЕСТИРОВАНИЕ СТРУКТУРНОЙ НЕСТАБИЛЬНОСТИ И ПОСТРОЕНИЕ НЕСТАЦИОНАРНОЙ СТАТИСТИЧЕСКОЙ МОДЕЛИ С ОПТИМИЗИРОВАННЫМ ВРЕМЕННЫМ РЯДОМ

5.1. Тестирование авторегрессионного процесса на стационарность путем нахождения обратных единичных корней

В главе 4 мы убедились, что с помощью уравнения авторегрессии $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$ можно строить точные интервальные прогнозы с 95%-ным уровнем надежности. Во всяком случае, прогноз по этой статистической модели на май 2010 г. показал, что доля точных интервальных прогнозов очень близка к заданному 95%-ному уровню надежности, рассчитанному на основе нормального распределения. И это несмотря на то, что сами остатки, полученные в результате решения уравнения регрессии, нельзя назвать нормально распределенными. Правда, при этом для части наблюдений у нас получились слишком широкие интервальные прогнозы. Как далее выяснится, решить эту проблему можно с помощью тестирования произошедших во временном ряде структурных изменений.

Однако сначала давайте посмотрим, насколько устойчива полученная прогностическая модель к внезапному росту волатильности на валютном рынке? Чтобы убедиться в устойчивости этой прогностической модели, необходимо проверить авторегрессионный процесс (AR-структуру этой модели) на стационарность. В EViews провести эту проверку достаточно несложно. При этом следует иметь в виду, что в ходе решения уравнения регрессии (см. алгоритм действий № 6

«Как решить уравнение регрессии в EViews») диалоговое мини-окно EQUATION SPECIFICATION заполняется иначе, а именно вместо записи $USDollar\ USDollar(-1)\ USDollar(-2)$ в него надо вставить формулу

$$USDollar\ AR(1)\ AR(2), \quad (5.1)$$

где $AR(1)$ — переменная с лагом в один месяц;

$AR(2)$ — переменная с лагом в два месяца.

Формула (5.1) по своей математической сути аналогична формуле $USDollar\ USDollar(-1)\ USDollar(-2)$, однако ввод в EViews уравнения по этой формуле дает возможность оценить авторегрессионный процесс на стационарность. Естественно, что при выводе итогов мы получим данные, практически аналогичные тем, которые уже содержатся в табл. 4.1. Одно из незначительных отличий заключается в том, что при выводе итогов ранее использовавшиеся обозначения переменных в виде $USDOLLAR(-1)$ и $USDOLLAR(-2)$ будут заменены соответственно на $AR(1)$ и $AR(2)$. Но самое главное заключается в том, что помимо уже известной нам информации в выводе итогов внизу появятся две дополнительные строки, в которых содержится оценка ARMA-структуры этого уравнения на стационарность (табл. 5.1).

Таблица 5.1

Оценка AR-структуры на стационарность для уравнения $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$

| | | |
|---|-------|-------|
| Inverted AR Roots (обратные AR единичные корни) | | |
| | 1,002 | 0,319 |
| Estimated AR process is nonstationary (оцененный процесс авторегрессии нестационарный) | | |

Судя по информации в этой таблице, AR-структура этого уравнения оказалась нестационарной, поскольку один из обратных единичных корней оказался больше единицы (подробнее об этом чуть позже). А из нестационарности AR-процесса вытекает вывод, что коэффициенты уравнения авторегрессии будут неустойчивыми. Таким образом, несмотря на довольно неплохие прогностические качества этой статистической модели, ее параметры нельзя назвать достаточно надежными к воздействию внешних «шоков», т.е. к случаям внезапного и резкого повышения курса доллара.

Чтобы точнее оценить степень устойчивости этой прогностической модели, продолжим проверку ее авторегрессионной структуры, тем

более что EViews позволяет сделать это с минимальными затратами времени.

Алгоритм действий № 13

Тестирование на стационарность AR-структуры уравнения
 $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$
 путем нахождения корней характеристического уравнения

Шаг 1. Нахождение корней характеристического уравнения

С этой целью в меню оцененного уравнения регрессии следует воспользоваться следующими опциями: VIEW/ARMA STRUCTURE (посмотреть/структуру модели ARMA). В результате чего на экране появится диалоговое мини-окно ARMA DIAGNOSTIC VIEWS (посмотреть диагностику модели ARMA).

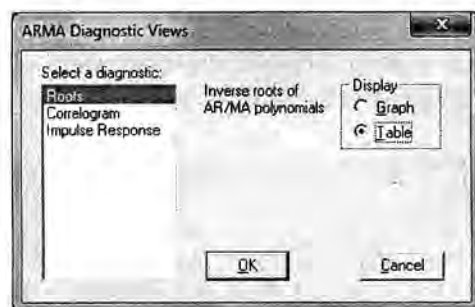


Рис. 5.1. Диалоговое мини-окно ARMA DIAGNOSTIC VIEWS для получения корней характеристического уравнения

Если в этом окне (рис. 5.1) выбрать опции ROOTS (корни) и TABLE (таблица), то в результате у нас получатся обратные корни характеристического уравнения в виде табл. 5.2. Судя по таблице, один из корней (по модулю) этого характеристического уравнения оказался больше единицы.

Шаг 2. Интерпретация корней характеристического уравнения

Чуть ниже мы остановимся подробнее на специфике корней, получаемых в результате решения характеристического уравнения. А сейчас отметим их самое важное для нас свойство: в том случае, когда абсолютные значения (по модулю) всех обратных корней этого уравнения меньше единицы, т.е. лежат внутри единичного круга,

то этот авторегрессионный процесс можно считать стационарным, а следовательно, обладающим устойчивыми вероятностными характеристиками. Если же хотя бы один из обратных корней характеристического уравнения больше единицы, т.е. лежит за пределами единичного круга, то тогда авторегрессионный процесс является нестационарным.

Таблица 5.2

Оценка авторегрессионного процесса на стационарность
 для уравнения
 $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$

| Inverse Roots of AR/MA Polynomial(s) (обратные единичные корни AR/MA полинома) | | |
|--|------------------------|-----------------|
| Specification: USDOLLAR AR(1) AR(2) | | |
| Date: 06/13/10 Time: 09:56 | | |
| Sample: 1992M06 2010M05 | | |
| Included observations: 213 | | |
| AR Root(s) (AR корни) | Modulus (по модулю) | Cycle (цикл) |
| 1,002461 | 1,002461 | |
| 0,318631 | 0,318631 | |
| Warning: At least one root outside the unit circle. | | |
| ARMA model is not stationary. (Предупреждение: по меньшей мере один корень лежит вне единичного круга. Модель ARMA — нестационарна.) | | |

Шаг 3. Построение графика корней характеристического уравнения

Если в мини-окне ARMA DIAGNOSTIC VIEWS выбрать опции ROOTS (корни) и GRAPH (график), то в этом случае мы получим обратные единичные корни характеристического уравнения в наглядном, графическом виде. Судя по рис. 5.2, один из корней находится внутри единичного круга, в то время как другой корень, хотя и расположен довольно близко к этому кругу, но все-таки лежит за его пределами. При этом следует иметь в виду, что горизонтальная ось на этом графике показывает фактические значения полученных обратных корней характеристического уравнения, в то время как вертикальная ось — их воображаемые значения.

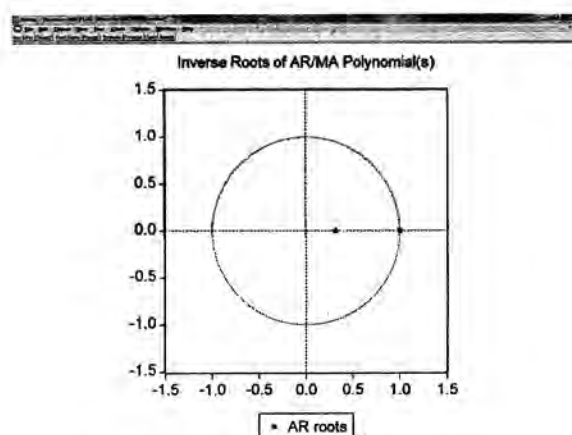


Рис. 5.2. Графическое изображение обратных единичных корней характеристического уравнения

Теперь остановимся несколько подробнее на процедуре получения обратных единичных корней, с помощью которой в EViews доказывалась стационарность AR-процессов. В главе 4 уже говорилось, что в основе теории единичного корня лежит довольно простая формула (4.4), которая считается базовой для понимания стационарности в уравнениях авторегрессии:

$$Y_t = \rho Y_{t-1} + e_t,$$

где Y_t — результативная зависимая переменная;

Y_{t-1} — независимая факторная переменная с лагом в один период (в нашем случае в один месяц);

ρ — коэффициент регрессии;

e_t — остатки.

При этом уравнение авторегрессии 1-го порядка считается стационарным в том случае, когда коэффициент регрессии $\rho < 1$. Соответственно, если $\rho > 1$, то оно считается нестационарным, а следовательно, волатильность в процессе авторегрессии с течением времени может нарастать и стремиться к бесконечности.

Применительно к авторегрессионным процессам, содержащим большое количество лаговых переменных, наличие стационарности предполагает следующее. AR-процессы считаются стационарными в том случае, если в уравнении (5.2) коэффициенты a_1, a_2, \dots, a_p образуют сходящийся ряд и все корни характеристического уравнения $1 - a_1z - a_2z^2 - \dots - a_pz^p = 0$ (вещественные и комплексные) должны лежать вне

единичного круга (см. рис. 5.2), их абсолютное значение (по модулю) должно быть больше единицы.

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} + e_t, \quad (5.2)$$

где $Y_t, Y_{t-1}, \dots, Y_{t-p}$ — переменные с различным лагом;

a — коэффициенты;

e — остатки.

Например, для решенного нами уравнения авторегрессии USDOLLAR = 1,321092 × USDOLLAR(-1) - 0,319415 × USDOLLAR(-2) (см. формулу (4.3)) характеристическое уравнение приобретает следующий вид:

$$1 - 1,321092 Y_{t-1} + 0,319415 Y_{t-2} = 0. \quad (5.3)$$

Корни в этом уравнении находятся с помощью известной со школьной скамьи формулы по нахождению корней в многочлене второй степени:

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{1,321092 \pm \sqrt{1,321092^2 - 4 \times 1 \times 0,319415}}{2 \times 0,319415}.$$

Отсюда следует, что первый единичный корень $x_1 = 3,138429$, а второй $x_2 = 0,997545$. Таким образом, один из этих двух корней характеристического уравнения лежит внутри единичного круга, а потому этот авторегрессионный процесс нельзя назвать стационарным. Однако мы уже говорили, что в EViews находятся не просто единичные корни, а именно ОБРАТНЫЕ единичные корни, которые мы получаем в выводе итогов (см. табл. 5.1) после небольших дополнительных вычислений. При этом первый и второй обратные единичные корни находятся из обычных единичных корней, полученных из уравнения (5.3), следующим образом: $x_1 = 1 : 3,138429 = 0,318631$, а второй $x_2 = 1 : 0,997545 = 1,002461$.

По сути, тот факт, что вместо единичных корней мы находим обратные единичные корни, ничего не меняет, однако — и это вполне понятно — при этом требования к тестированию стационарности AR-процесса формулируются противоположным образом. В этом случае авторегрессионный процесс считается стационарным тогда и только тогда, когда абсолютные значения (по модулю) всех обратных корней его характеристического уравнения лежат в пределах единичного круга. Поскольку один из обратных корней больше единицы, то, следовательно, AR-процесс, описанный формулой $1,321092 \times \text{USDOLLAR}(-1) - 0,319415 \times \text{USDOLLAR}(-2)$, нельзя считать стационарным.

5.2. Тестирование AR-структуры на стационарность с помощью функции импульсного ответа

Теперь остановимся еще на одном важном инструменте, который дает EViews для оценки устойчивости статистических моделей к внешним шокам (в нашем случае под ними подразумеваются резкие скачки курса доллара). Это тестирование AR-структуры авторегрессионного процесса на импульсный ответ (IMPULSE RESPONSE). При этом у нас появляется возможность получить также и оценку инновационной неопределенности, возникающей в этом авторегрессионном процессе в результате воздействия внешнего шока.

Алгоритм № 14

Тестирование на импульсный ответ AR-структуры нестационарного процесса, описываемого уравнением $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$

Шаг 1. Построение функции импульсного и накопленного импульсного ответа

Чтобы провести исследование AR-структуры на импульсный ответ (IMPULSE RESPONSE), нам необходимо в диалоговом мини-окне ARMA DIAGNOSTIC VIEWS (посмотреть диагностику ARMA) выбрать опции IMPULSE RESPONSE и TABLE, а также определить предполагаемую величину импульса (внешнего шока или величину отклонения курса доллара) (рис. 5.3). При этом по умолчанию используется опция ONE STANDARD DEVIATION (одно стандартное отклонение), а длительность проводимого теста ограничивается 24 периодами (см. опцию PERIODS). Однако при необходимости количество тестируемых периодов можно изменить, как, впрочем, изменить и заданную величину импульса. В последнем случае надо в опции IMPULSE выбрать вариант USER SPECIFIED (по усмотрению пользователя) и самому установить требуемую величину начального импульса (внешнего шока).

Шаг 2. Интерпретация функции импульсного и накопленного импульсного ответа

В опции IMPULSE мы решили выбрать вариант по умолчанию — ONE STANDARD DEVIATION (одно стандартное отклонение). В случае выбора этой опции исходная величина внешнего шока приравнивается к стандартной ошибке коэффициента регрессии факторной лаго-

вой переменной, т.е. к стандартной ошибке коэффициента регрессии $USDOLLAR(-1) = 0,06527$. При этом предполагается, что на первом шаге, когда внешний шок отсутствует, величина импульсного ответа равна стандартной ошибке уравнения регрессии $-0,817803$ (см. табл. 4.1).

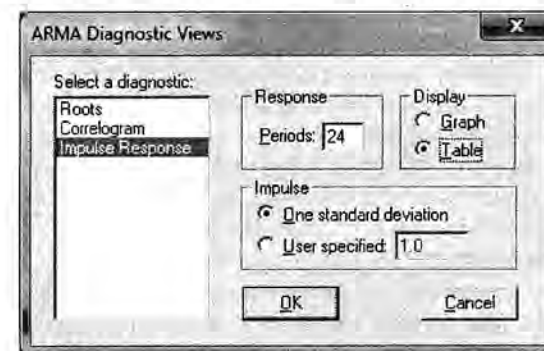


Рис. 5.3. Выбор опций IMPULSE RESPONSE и TABLE в диалоговом мини-окне ARMA DIAGNOSTIC VIEWS

В результате получим табл. 5.3, в которой содержится информация, характеризующая величину стандартной ошибки импульсного ответа AR-структуры на рост инновационной неопределенности, значения которой помещены в двух разделах Std. Err. Если бы вместо опции ONE STANDARD DEVIATION была бы выбрана опция USER SPECIFIED (по усмотрению пользователя), то тогда вместо величины инновационной неопределенности мы получили бы в разделах Std. Err. величину внешнего шока в виде стандартных ошибок факторной переменной, нарастающей по мере повышения инновационной неопределенности.

Таким образом, величина стандартной ошибки функции импульсного ответа позволяет оценить, как модель ARMA реагирует на единовременное шоковое воздействие (однократное резкое изменение курса доллара). В таблице 5.3 также приводится величина стандартной ошибки функции накопленного импульсного ответа. Последнюю функцию можно интерпретировать и как ответ на очередной текущий импульс (шоковое воздействие), но при условии, что аналогичные шоковые воздействия происходят непрерывно в течение всего исследуемого времени, начиная с 1-го периода.

Важным свойством стационарных моделей является то обстоятельство, что у них уровень инновационной неопределенности, как и величина ответа на импульс, асимптотически — по мере нара-

тания выборки (количества периодов) — стремятся к нулю. Это свидетельствует об устойчивости стационарных процессов к единовременным шокным воздействиям. Причем в случае анализа стационарного AR-процесса EViews дает внизу вывода итогов асимптотическую оценку как величины импульсного ответа, так и уровня инновационной неопределенности с указанием, что они равны нулю. Однако в табл. 5.3 этих оценок нет, поскольку исследуется нестационарный процесс, в котором обе эти величины постоянно нарастают, не имея при этом четко ограниченных пределов.

Как легко увидеть в табл. 5.3, в 1-м периоде величина инновационной неопределенности незначительна, а ошибка импульсного ответа равна стандартной ошибке уравнения регрессии. Во 2-м периоде инновационная неопределенность возрастает в силу воздействия внешнего шока, величина которого приравнивается к стандартной ошибке коэффициента регрессии независимой лаговой переменной USDOLLAR(-1). Далее в последующие периоды величина инновационной неопределенности (см. оба столбца Std. Err.) нарастает как в функции импульсного, так еще больше и в функции накопленного импульсного ответа. Так, уровень инновационной неопределенности в функции импульсного ответа увеличивается с 0,03962 в 1-м периоде до 0,16794 в 25-м периоде; соответственно в функции накопленного импульсного ответа эти цифры выросли с 0,03962 в 1-м периоде до 3,28261 в 25-м периоде.

Если при тестировании AR-структуры нестационарного процесса увеличить количество исследуемых периодов, то в результате уровень инновационной неопределенности и величина стандартной ошибки импульсного ответа в нестационарной модели еще больше вырастут к концу последнего периода.

Так, если в диалоговом мини-окне ARMADIAGNOSTIC VIEWS в опции PERIODS установить 50 периодов вместо используемых по умолчанию 24, то уровень инновационной неопределенности в функции импульсного ответа будет равен 0,27308, а в функции накопленного импульсного ответа — 8,65349. Соответственно при 100 периодах эти цифры в последнем периоде будут существенно выше и равны 0,56117 и 28,4379.

Шаг 3. Построение графика функций импульсного и накопленного импульсного ответа

При желании функции импульсного и накопленного импульсного ответа можно получить и в графическом виде. С этой целью в диалоговом мини-окне ARMA DIAGNOSTIC VIEWS необходимо выбрать опции IMPULSE RESPONSE и GRAPH. Если мы при этом оставим те же опции,

Таблица 5.3

Импульсный ответ AR-структуры и оценка инновационной неопределенности для уравнения авторегрессии USDOLLAR = a × USDOLLAR(-1) + b × USDOLLAR(-2)

| ARMA Impulse Response | | | | |
|--|---------------------|-----------|---------------------------------------|-----------|
| Specification: USDOLLAR AR(1) AR(2) | | | | |
| Date: 06/13/10 Time: 12:03 | | | | |
| Sample: 1992M06 2010M05 | | | | |
| Included observations: 213 | | | | |
| Warning: estimated ARMA model is NOT stationary(предупреждение: оцененная модель ARMA является нестационарной) | | | | |
| Period (период) | Response (ответ) | Std. Err. | Accumulated (накопленный ответ) | Std. Err. |
| 1 | 0,817803 | (0,03962) | 0,817803 | (0,03962) |
| 2 | 1,080394 | (0,07476) | 1,898197 | (0,10634) |
| 3 | 1,166081 | (0,10416) | 3,064278 | (0,20467) |
| 4 | 1,195407 | (0,11907) | 4,259685 | (0,32025) |
| 5 | 1,206779 | (0,12579) | 5,466463 | (0,44383) |
| 6 | 1,212435 | (0,12889) | 6,678898 | (0,57115) |
| 7 | 1,216275 | (0,13060) | 7,895173 | (0,70048) |
| 8 | 1,219541 | (0,13186) | 9,114715 | (0,83115) |
| 9 | 1,222630 | (0,13304) | 10,33735 | (0,96295) |
| 10 | 1,225667 | (0,13429) | 11,56301 | (1,09584) |
| 11 | 1,228693 | (0,13565) | 12,79171 | (1,22989) |
| 12 | 1,231720 | (0,13714) | 14,02343 | (1,36518) |
| 13 | 1,234753 | (0,13877) | 15,25818 | (1,50179) |
| 14 | 1,237792 | (0,14053) | 16,49597 | (1,63982) |
| 15 | 1,240839 | (0,14243) | 17,73681 | (1,77939) |
| 16 | 1,243894 | (0,14445) | 18,98070 | (1,92059) |

Окончание

| Period (период) | Response (ответ) | Std. Err. | Accumulated (накопленный ответ) | Std. Err. |
|--------------------|---------------------|-----------|---------------------------------------|-----------|
| 17 | 1,246955 | (0,14660) | 20,22766 | (2,06354) |
| 18 | 1,250025 | (0,14887) | 21,47768 | (2,20835) |
| 19 | 1,253101 | (0,15126) | 22,73078 | (2,35510) |
| 20 | 1,256186 | (0,15377) | 23,98697 | (2,50393) |
| 21 | 1,259278 | (0,15640) | 25,24625 | (2,65492) |
| 22 | 1,262378 | (0,15913) | 26,50863 | (2,80818) |
| 23 | 1,265485 | (0,16196) | 27,77411 | (2,96381) |
| 24 | 1,268600 | (0,16490) | 29,04271 | (3,12192) |
| 25 | 1,271722 | (0,16794) | 30,31443 | (3,28261) |

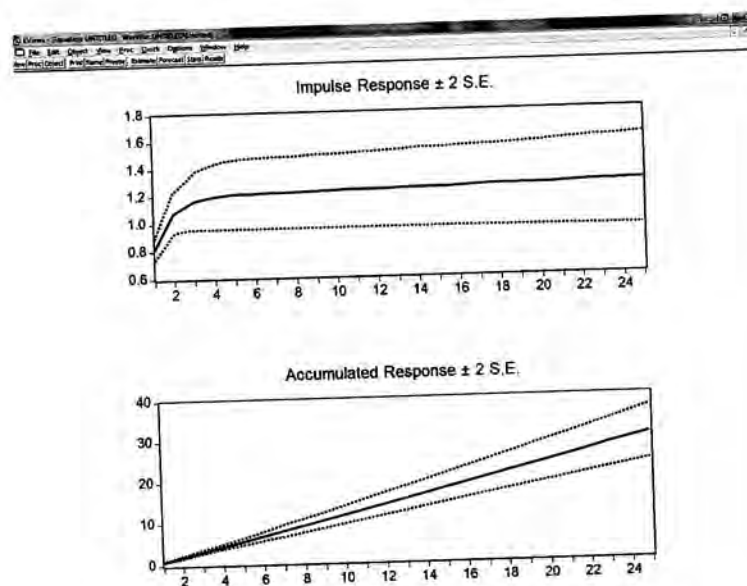


Рис. 5.4. Построение графика функций импульсного и накопленного импульсного ответа при помощи диалогового мини-окна ARMA DIAGNOSTIC VIEWS

что установили при выводе данных, размещенных в табл. 5.3, то тогда получим следующие графики роста величины импульсного и накопленного импульсного ответа (рис. 5.4). Помимо удобств, связанных с наглядностью, эти графики позволяют также увидеть нижние и верхние границы интервалов, в рамках которых рассчитанная в EViews величина импульсного и накопленного импульсного ответа может колебаться. В качестве доверительного интервала для оценки величины этих функций берется диапазон в размере двух стандартных отклонений.

5.3. Влияние резких изменений курса доллара на смещение коэффициентов регрессии

Продолжим далее наш анализ устойчивости к воздействию внешних шоков нестационарного AR-процесса, описанного уравнением $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$. С этой целью составим рейтинг наиболее резких изменений в курсе доллара, зафиксированных за период с августа 1998 г. по апрель 2010 г. При этом в качестве критерия для отбора будем использовать величину изменения курса доллара за один месяц в процентах по модулю. В результате получилась табл. 5.4. Из нее следует, что три самых крупных колебания курса доллара наблюдались в сентябре и августе 1998 г., а также в январе 2009 г. Кроме того, из этой таблицы (см. раздел «Скачок курса доллара по сравнению с предыдущим месяцем, руб.») можно сделать вывод, что резкие скачки доллара по преимуществу были положительными. Так, из 20 наблюдений, включенных в этот рейтинг, в 15 случаях рубль резко укреплялся и лишь в пяти случаях резко падал. Причем в шестерку самых волатильных месяцев вошли только те месяцы, когда был зафиксирован резкий рост, а не падение курса доллара. С фундаментальной точки зрения это объясняется многолетней политикой Банка России по поддержанию слабого курса рубля, а с точки зрения статистического анализа этот факт можно подтвердить с помощью описательной статистики (см. табл. 4.4), показывающей значительную правостороннюю асимметрию в остатках.

Таблица 5.4, в которую включена топ-двадцатка самых волатильных (с августа 1998 г.) месяцев, понадобится для того, чтобы оценить надежность нашей прогностической модели. Вполне очевидно, что слишком сильные колебания курса доллара довольно существенно сказывались на качестве прогноза. Своего рода рекорд по неточности предсказания можно было бы установить в конце сентября 1998 г. при прогнозировании

вании курса доллара на октябрь 1998 г. на основе данных за период с июня 1992 г. по сентябрь 1998 г. Проверим это утверждение. Однако прежде нам надо научиться оперативно изменять выборку данных в EViews, поскольку каждый раз импортировать новые данные нерационально в смысле затрат времени (см. алгоритм действий № 15).

Таблица 5.4

Рейтинг наблюдений по величине скачка курса доллара

| Место наблюдения по скачку доллара, % (по модулю) | Месяц | Курс доллара, руб. | Скачок курса доллара по сравнению с предыдущим месяцем, руб. | Скачок курса доллара, % (по модулю) |
|---|------------------|--------------------|--|-------------------------------------|
| 1 | Сентябрь 1998 г. | 16,06 | 8,16 | 103,2 |
| 2 | Август 1998 г. | 7,91 | 1,67 | 26,7 |
| 3 | Январь 2009 г. | 35,41 | 6,03 | 20,5 |
| 4 | Декабрь 1998 г. | 20,65 | 2,77 | 15,5 |
| 5 | Ноябрь 1998 г. | 17,88 | 1,87 | 11,7 |
| 6 | Январь 1999 г. | 22,60 | 1,95 | 9,4 |
| 7 | Май 2009 г. | 30,98 | -2,26 | 6,8 |
| 8 | Декабрь 2008 г. | 29,38 | 1,77 | 6,4 |
| 9 | Март 1999 г. | 24,18 | 1,32 | 5,8 |
| 10 | Январь 2000 г. | 28,55 | 1,55 | 5,7 |
| 11 | Октябрь 2008 г. | 26,54 | 1,30 | 5,1 |
| 12 | Август 2008 г. | 24,58 | 1,13 | 4,8 |
| 13 | Март 2009 г. | 34,01 | -1,71 | 4,8 |
| 14 | Сентябрь 2009 г. | 30,09 | -1,48 | 4,7 |
| 15 | Ноябрь 2008 г. | 27,61 | 1,06 | 4,0 |
| 16 | Октябрь 1999 г. | 26,05 | 0,97 | 3,9 |
| 17 | Октябрь 2009 г. | 29,05 | -1,04 | 3,5 |
| 18 | Январь 2004 г. | 28,49 | -0,96 | 3,3 |
| 19 | Сентябрь 2007 г. | 24,95 | -0,70 | 2,7 |
| 20 | Сентябрь 2008 г. | 25,25 | 0,67 | 2,7 |

Источник: расчеты автора по данным Банка России.

Алгоритм действий № 15

Как в EViews можно быстро изменить выборку данных

Чтобы смоделировать ситуацию реального прогноза на октябрь 1998 г., необходимо оставить во временном ряде данные лишь за период с июня 1992 г. по сентябрь 1998 г. С этой целью надо в верхней строке рабочего файла выбрать опции QUICK /SAMPLE (быстро/выборка), после чего появится диалоговое мини-окно SAMPLE, в котором мы должны ввести текст: @first 1998m08. Введенный в диалоговое мини-окно текст означает, что в выборке должны остаться данные с первого наблюдения по август 1998 г. Для справки заметим, что если бы мы захотели ограничить выборку снизу (например, с сентября 1998 г. до последнего наблюдения), то тогда команда в диалоговом мини-окне SAMPLE выглядела бы так: 1998m09@last. А если бы нам потребовалось вновь использовать всю выборку, то в этом случае в диалоговое мини-окно надо было бы ввести команду @all.

После сокращения выборки (период уменьшили до 74 наблюдений — с июня 1992 г. по сентябрь 1998 г.) займемся решением уравнения регрессии (см. алгоритм действий № 6 «Как решить уравнение регрессии в EViews»). А затем делаем прогноз и соответственно сразу же находим остатки (см. алгоритм действий № 8 «Как оценить точность статистической модели в EViews»). Вывод данных по уравнению регрессии у нас представлен в табл. 5.5.

В результате, согласно точечному прогнозу, составленному по этому уравнению регрессии, американский доллар в октябре 1998 г. должен был бы стоить 43 руб. 02 коп.! На самом деле за всю историю наблюдений за валютным рынком стоимость доллара никогда не достигала таких значений, а фактический его курс в конце октября 1998 г. равнялся 16 руб. 1 коп. Таким образом, разница (или остаток) составила 27 руб. и 1 коп.!!! Впрочем, столь неудачный прогноз — все-таки единственное исключение из всего временного ряда. Для справки заметим, что аналогичный прогноз, сделанный на ноябрь 1998 г., а тем более все последующие прогнозы уже не столь существенно отклонялись от фактического курса доллара. Так, согласно прогнозу на ноябрь 1998 г., рассчитанному на основе рыночных данных по курсу доллара за период с июня 1992 г. по октябрь 1998 г., американский доллар должен был стоить 16 руб. 75 коп. Однако фактическая стоимость доллара оказалась равна 17 руб. 88 коп., т.е. разница (остаток) составила 1 руб. 13 коп.

Таблица 5.5

Вывод данных по уравнению регрессии
 $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$,
 на основе которого был составлен прогноз на октябрь 1998 г.

| Dependent Variable: USDOLLAR | | | | |
|---|-------------|-----------------------|-------------|----------|
| Method: Least Squares | | | | |
| Date: 06/24/10 Time: 19:24 | | | | |
| Sample (adjusted): 1992M08 1998M09 | | | | |
| Included observations: 74 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 4,316452 | 0,279236 | 15,45809 | 0,0000 |
| USDOLLAR(-2) | -3,329701 | 0,285766 | -11,65185 | 0,0000 |
| R-squared | 0,952418 | Mean dependent var | | 3,973047 |
| Adjusted R-squared | 0,951757 | S.D. dependent var | | 2,544697 |
| S.E. of regression | 0,558925 | Akaike info criterion | | 1,701052 |
| Sum squared resid | 22,49260 | Schwarz criterion | | 1,763325 |
| Log likelihood | -60,93894 | Durbin — Watson stat | | 0,747672 |

Источник: расчеты автора по данным Банка России.

В чем же причины столь серьезной ошибки в прогнозе на октябрь 1998 г., сделанном с помощью уравнения регрессии $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$? Причина вполне очевидна — беспрецедентный 103,2%-ный скачок курса доллара в сентябре 1998 г. по сравнению с предыдущим месяцем. В результате этого колоссального внешнего шока оба коэффициента b и c в уравнении регрессии резко сместились в сторону предсказания необычно высокого роста, что характерно для статистических моделей с нестационарной ARMA-структурой. Смещение коэффициентов регрессии наглядно показано в табл. 5.6: если в уравнении регрессии, на основе которого был составлен прогноз на сентябрь 1998 г., первый коэффициент регрессии b был равен 1,6309, а второй коэффициент $c = -0,6185$, то уже в следующем месяце эти коэффициенты выросли соответственно до 4,3165 и $-3,3297$. Именно это значительное смещение в коэффициентах регрессии и привело к серьезному искажению в прогнозе на октябрь 1998 г. В пользу этого говорит и тот факт,

что если бы мы воспользовались при прогнозе на октябрь 1998 г. коэффициентами регрессии, на основе которых делался прогноз на сентябрь 1998 г., то ошибка в прогнозе у нас получилась бы меньше. Наш прогноз равнялся бы 21 руб. 31 коп., т.е. разница с фактическим курсом доллара, равным 16 руб. 1 коп., составила бы 5 руб. 30 коп.

5.4. Стандартные и студентизированные остатки, влияние выбросов на точность уравнения регрессии

Очевидно, что любое резкое повышение курса доллара приводит к возникновению так называемых выбросов, т.е. необычно больших остатков (разницы между фактическим и прогнозным курсом доллара), отрицательно влияющих на точность последующих прогнозов. Их определение можно взять в соответствующей литературе: «Выбросом

Таблица 5.6

Смещение коэффициентов в уравнении регрессии
 $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$
 после скачка курса доллара в августе 1998 г.

| Параметр уравнения | Прогноз на август 1998 г. | Прогноз на сентябрь 1998 г. | Прогноз на октябрь 1998 г. | Прогноз на ноябрь 1998 г. | Прогноз на декабрь 1998 г. | Прогноз на июль 2000 г. | Прогноз на май 2010 г. |
|--|---------------------------|-----------------------------|----------------------------|---------------------------|----------------------------|-------------------------|------------------------|
| Коэффициент b | 1,6812 | 1,6309 | 4,3165 | 1,1432 | 1,1268 | 1,235566 | 1,3211 |
| Коэффициент c | -0,6774 | -0,6185 | -3,3297 | -0,0964 | -0,0681 | -0,217671 | -0,3194 |
| Значимость коэффициента b | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 |
| Значимость коэффициента c | 0,0000 | 0,0008 | 0,0000 | 0,4539 | 0,5892 | 0,0408 | 0,0000 |
| Коэффициент детерминации уравнения R^2 | 0,9971 | 0,9895 | 0,9524 | 0,8936 | 0,9170 | 0,989746 | 0,9952 |

Источник: расчеты автора.

среди остатков называется остаток, который по абсолютной величине значительно превосходит остальные и отличается от среднего по остаткам на три, четыре или даже более стандартных отклонений¹.

Следовательно, решая уравнение регрессии, надо всегда проверять величину полученных остатков, а также оценивать риск, связанный с влиянием выбросов на смещение коэффициентов в уравнении регрессии. Очевидно, что в случае выявления во временном ряде наблюдения, имеющего остаток, равный трем и более стандартным отклонениям, нужно всегда проводить тщательный анализ с точки зрения его влияния на точность составленного прогноза. Впрочем, во многих статистических программах критическими считаются и те случаи, когда величина остатка больше двух стандартных отклонений.

Теперь посмотрим, какие результаты дал бы анализ остатков по прогнозу на октябрь 1998 г., сделанному на основе данных по курсу доллара с июня 1992 г. по сентябрь 1998 г.

Алгоритм действий № 16

Как в Excel можно найти стандартные остатки

В Excel анализ остатков на наличие в них выбросов проводится следующим образом. Решая уравнение регрессии, нужно действовать согласно алгоритму № 3 «Как решить уравнение регрессии в Excel». После того как на экране появится диалоговое мини-окно РЕГРЕССИЯ, в параметре ОСТАТКИ нужно установить две опции: ОСТАТКИ и СТАНДАРТИЗИРОВАННЫЕ ОСТАТКИ (рис. 5.5).

Используя в Excel опции ОСТАТКИ и СТАНДАРТИЗИРОВАННЫЕ ОСТАТКИ, мы получили после решения уравнения регрессии следующие данные по остаткам, часть из которых (за 1997—1998 гг.) поместили в табл. 5.7. Нужно сразу заметить, что в Excel *стандартные* (в литературе чаще используется последнее определение, хотя иногда встречается и термин «стандартизированные») остатки находятся путем деления их фактических, а не абсолютных значений. Поэтому в нашей таблице есть наблюдения как с положительными (в том случае, когда прогноз оказался ниже фактического курса доллара), так и с отрицательными (когда прогноз оказался выше фактического курса доллара) стандартными остатками.

¹ Дрейнер Н., Смит Г. Прикладной регрессионный анализ : пер. с англ. Кн. 1. М. : Финансы и статистика, 1986. С. 199.

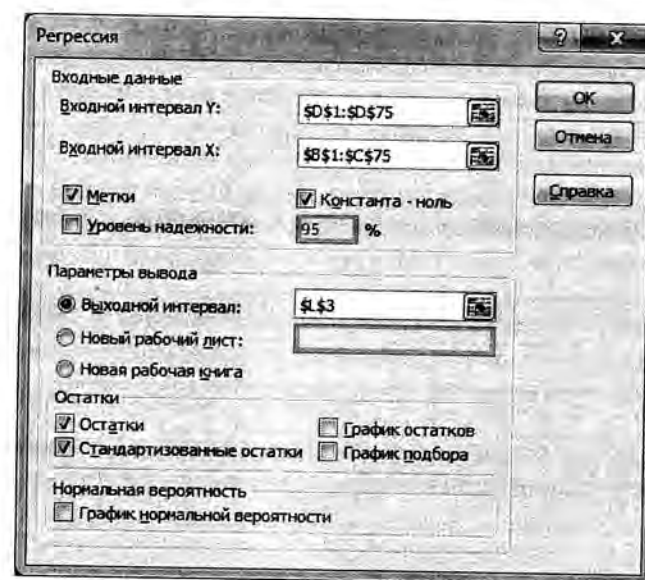


Рис. 5.5. Установление опций ОСТАТКИ и СТАНДАРТИЗИРОВАННЫЕ ОСТАТКИ в диалоговом мини-окне РЕГРЕССИЯ

Анализируя по табл. 5.7 динамику стандартных остатков, легко заметить, что своего максимума они достигли в августе и сентябре 1998 г. Чтобы обратить внимание читателей к этим двум наблюдениям, их выделили жирным шрифтом. При этом можно увидеть, что в то время как стандартные остатки в августе 1998 г. оказались равны 2,931979 и были меньше трех стандартных отклонений, то в сентябре 1998 г. их величина составила 4,922042, приблизившись тем самым к пяти стандартным отклонениям. Таким образом, с уверенностью можно констатировать, что сентябрьский остаток представляет собой выброс — это весьма осложняет получение (по применяемой статистической модели) точного прогноза на октябрь 1998 г. Впрочем, выбросом можно считать и остатки, полученные в августе 1998 г., если снизить уровень для выбросов до двух стандартных отклонений.

Помимо стандартных остатков для анализа выбросов используются также студентизированные остатки, которые представляют собой частное от деления обычного остатка на оценку его стандартного отклонения. Хотя теоретически все случайные ошибки, полученные после решения уравнения регрессии, считаются независимыми и имеющими одну и ту же дисперсию, однако в действительности конкретные

Таблица 5.7

Стандартные остатки, полученные при решении уравнения регрессии в Excel на основе данных за период с июня 1992 г. по сентябрь 1998 г.

| Месяцы | Наблю- дение | Предсказанное USDollar | Остатки | Стандартные остатки |
|------------------|-----------------|---------------------------|----------|------------------------|
| Январь 1997 г. | 54 | 5,649491 | -0,02049 | -0,03717 |
| Февраль 1997 г. | 55 | 5,784171 | -0,10817 | -0,1962 |
| Март 1997 г. | 56 | 5,757295 | -0,03129 | -0,05676 |
| Апрель 1997 г. | 57 | 5,816621 | -0,05462 | -0,09907 |
| Май 1997 г. | 58 | 5,805528 | -0,03253 | -0,059 |
| Июнь 1997 г. | 59 | 5,73314 | 0,04886 | 0,088623 |
| Июль 1997 г. | 60 | 5,735361 | 0,062639 | 0,113615 |
| Август 1997 г. | 61 | 5,774457 | 0,055543 | 0,100745 |
| Сентябрь 1997 г. | 62 | 5,859309 | 0,000691 | 0,001254 |
| Октябрь 1997 г. | 63 | 5,882252 | 0,004748 | 0,008612 |
| Ноябрь 1997 г. | 64 | 5,898905 | 0,020095 | 0,036449 |
| Декабрь 1997 г. | 65 | 5,94713 | 0,01287 | 0,023345 |
| Январь 1998 г. | 66 | 6,017554 | 0,008446 | 0,01532 |
| Февраль 1998 г. | 67 | 6,165922 | -0,09392 | -0,17036 |
| Март 1998 г. | 68 | 6,144718 | -0,03872 | -0,07023 |
| Апрель 1998 г. | 69 | 6,138311 | -0,00531 | -0,00963 |
| Май 1998 г. | 70 | 6,141646 | 0,022354 | 0,040547 |
| Июнь 1998 г. | 71 | 6,185554 | 0,012446 | 0,022575 |
| Июль 1998 г. | 72 | 6,229092 | 0,008908 | 0,016157 |
| Август 1998 г. | 73 | 6,288541 | 1,616459 | 2,931979 |
| Сентябрь 1998 г. | 74 | 13,35088 | 2,713622 | 4,922042 |

Источник: расчеты автора.

Примечания:

1. В таблице помещены данные за 1997–1998 гг.
2. Графа «Месяц» в Excel не выводится, но дана автором для большей наглядности представленной информации.

остатки в силу своего различного положения во временном ряду отнюдь не независимы и, следовательно, не имеют одинаковых дисперсий¹. Поэтому чтобы учесть эту разницу в дисперсии остатков, их необходимо студентизировать, т.е. оценить с учетом их положения в выборке. Формулу по расчету студентизированных остатков мы дадим далее, а сейчас приведем алгоритм действий, с помощью которого можно быстро получить студентизированные остатки. Правда, в Excel такая возможность отсутствует, но в последних версиях EViews эту процедуру можно реализовать с помощью статистики влияния остатков (INFLUENCE STATISTICS). Чтобы уяснить, как это делается, надо прочитать алгоритм действий № 17.

Алгоритм действий № 17

Диагностика в EViews влияния студентизированных остатков на уравнение регрессии для прогностической модели

$$USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$$

Шаг 1. Как получить студентизированные остатки

После решения уравнения регрессии (на основе рыночных данных по курсу доллара за период с июня 1992 г. по сентябрь 1998 г.) в строке EQUATION выбираем опции VIEW/STABILITY DIAGNOSTICS/INFLUENCE STATISTICS. В результате на экране появляется диалоговое мини-окно INFLUENCE STATISTICS, которое нужно соответствующим образом заполнить, чтобы провести диагностику остатков (рис. 5.6). Чтобы получить как графический, так и табличный вариант по статистике влияния остатков в параметре OUTPUT TYPE (тип выходной статистики), следует установить опции GRAPH (график) и TABLE (таблица). Далее в параметре OUTPUT STATISTICS (выходная статистика) ставим галочку у опции RSTUDENT (студентизированные остатки) и рядом пишем RS — название файла, который будет помещен в рабочий файл.

Шаг 2. Интерпретация влияния студентизированных остатков на точность прогноза

В результате шага 1 получены диаграмма (она приведена на рис. 5.7) и табл. 5.8. Интерпретация диаграммы довольно проста, поскольку на ней представлен график значений студентизированных остатков,

¹ Дрейнер Н., Смит Г. Прикладной регрессионный анализ. С. 190.

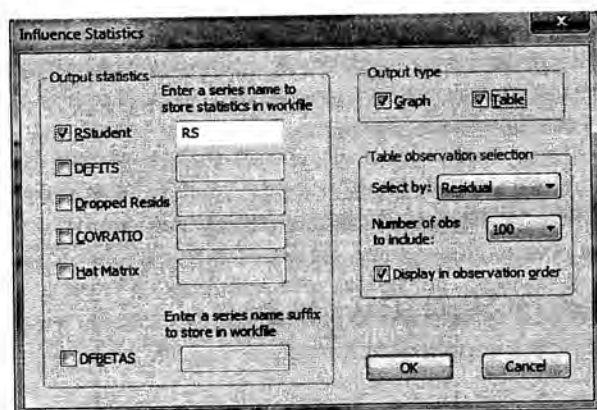


Рис. 5.6. Диалоговое мини-окно INFLUENCE STATISTICS для проведения диагностики остатков

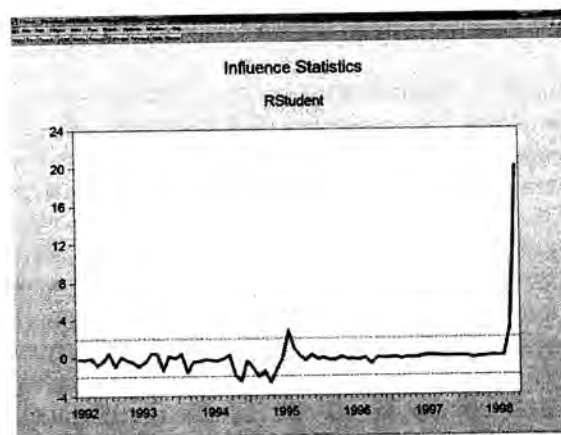


Рис. 5.7. График значений студентизированных остатков
Источник: расчеты автора.

который с обеих сторон выделен пунктирной линией и обозначает область допустимых значений, равных ± 2 . Когда студентизированные остатки выходят за пределы этой пунктирной линии, в этом наблюдении их можно считать выбросами. Легко заметить, что особенно велик студентизированный остаток, полученный в сентябре 1998 г.

В таблице 5.8 приведена часть полученных с помощью EViews значений студентизированных остатков (за период с января 1997 г. по сентябрь 1998 г.). При этом студентизированные остатки, которые

Таблица 5.8

Студентизированные остатки, выбросы и их влияние на точность прогнозирования, полученные при решении уравнении регрессии в EViews на основе данных за период с июня 1992 г. по сентябрь 1998 г.

| Influence Statistics | | |
|----------------------------|-----------|-----------------|
| Date: 06/27/10 Time: 10:35 | | |
| Sample: 1992M06 1998M09 | | |
| Included observations: 74 | | |
| Obs. | Resid. | RStudent |
| 1997M01 | -0,020491 | -0,036856 |
| 1997M02 | -0,108171 | -0,194596 |
| 1997M03 | -0,031295 | -0,056323 |
| 1997M04 | -0,054621 | -0,098327 |
| 1997M05 | -0,032528 | -0,058585 |
| 1997M06 | 0,048860 | 0,088075 |
| 1997M07 | 0,062639 | 0,112930 |
| 1997M08 | 0,055543 | 0,100119 |
| 1997M09 | 0,000691 | 0,001246 |
| 1997M10 | 0,004748 | 0,008557 |
| 1997M11 | 0,020095 | 0,036224 |
| 1997M12 | 0,012870 | 0,023201 |
| 1998M01 | 0,008446 | 0,015225 |
| 1998M02 | -0,093922 | -0,169284 |
| 1998M03 | -0,038718 | -0,069825 |
| 1998M04 | -0,005311 | -0,009583 |
| 1998M05 | 0,022354 | 0,040349 |
| 1998M06 | 0,012446 | 0,022466 |
| 1998M07 | 0,008908 | 0,016080 |
| 1998M08 | 1,616459 | <u>3,110573</u> |
| 1998M09 | 2,713622 | <u>19,91993</u> |

Источник: расчеты автора.

Примечание: в таблице помещены данные за 1997–1998 гг.

считаются выбросами (их величина больше или меньше 2), при выводе итогов обозначаются EViews красным шрифтом (в таблице они подчеркнуты). При этом область допустимых значений определяется с помощью уже известной нам t -статистики. В частности, выбросами считаются остатки, которые получены не только в сентябре, но и в августе 1998 г. Если сравнить стандартные остатки из табл. 5.7 со студентизированными остатками, то легко заметить, что значения последних — за счет выросшей дисперсии между наблюдениями — наиболее сильно отличаются от значений первых для августа и сентября 1998 г.

Некоторые математические подробности по расчету студентизированных остатков в EViews

Теоретически все случайные ошибки предполагаются независимыми и имеющими одну и ту же дисперсию σ^2 , однако в действительности конкретные остатки отнюдь не независимы и, следовательно, не имеют одинаковых дисперсий. В действительности дисперсия остатков зависит не только от величины σ^2 , но и от h_i — i -го диагонального элемента матрицы вида $X_i(X'X)^{-1}X_i$, с которой мы уже познакомились в главе 3.

Студентизированные остатки в EViews рассчитываются по формуле

$$\bar{e}_i = \frac{e_i}{s(i)\sqrt{1-h_i}}, \quad (5.4)$$

где e_i — остаток для конкретного наблюдения, полученный по уравнению регрессии, построенному с учетом всех наблюдений временного ряда;

$s(i)$ — стандартное отклонение остатков, полученное по уравнению регрессии, построенному по тому же временному ряду без учета наблюдения i ;

h_i — i -ный диагональный элемент матрицы вида $X_i(X'X)^{-1}X_i$.

При необходимости i -ный диагональный элемент матрицы $X_i(X'X)^{-1}X_i$ можно найти для каждого наблюдения, если в диалоговом мини-окне INFLUENCE STATISTICS установить опцию HAT MATRIX (т.е. матрица $X_i(X'X)^{-1}X_i$).

Например, величина студентизированного остатка для сентября 1998 г. равна

$$\bar{e}_{\text{сен. 98 г.}} = \frac{2,7136}{0,2192\sqrt{1-0,6140}} = 19,9199.$$

Распределение студентизированных остатков подчиняется t -статистике, получаемой в результате подстановки фиктивной переменной в первоначальное уравнение регрессии. Причем фиктивная переменная для интересующего нас наблюдения i равна 1, а для всех остальных наблюдений она равна 0. Таким образом, студентизированный остаток можно интерпретировать как тест на значимость остатка определенного наблюдения с точки зрения его влияния на уравнение регрессии.

Следует заметить, что если у кого-то из читателей нет последней версии EViews или иных программ, умеющих рассчитывать студентизированные остатки, то в принципе для обнаружения выбросов вполне возможно пользоваться стандартными остатками. Во всяком случае, как утверждают Н. Дрейпер и Г. Смит, в подавляющем большинстве случаев, хотя и не во всех, для обнаружения выбросов вполне достаточно пользоваться графиками обычных и стандартных остатков¹.

Чтобы убедиться в справедливости этих слов, мы провели небольшой эксперимент. С этой целью уравнение регрессии USDOLLAR = $a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2)$ решено на основе данных за период с июня 1992 г. по апрель 2010 г., а затем рассмотрены полученные остатки (табл. 5.9). В том случае, когда студентизированные остатки диагностируют выбросы, стандартные остатки также их выявляют (если к выбросам отнести остатки, имеющие два стандартных отклонения). Правда, поскольку студентизированные остатки учитывают не только стандартное отклонение, но и дисперсию между различными наблюдениями (формула (5.6)), то величина студентизированных остатков всегда выше. Причем особенно заметна эта разница относительно сентября 1998 г. и января 2009 г., т.е. когда на валютном рынке наблюдалась максимальная волатильность, обусловленная в первом случае августовским дефолтом 1998 г., а во втором случае — глобальным финансовым кризисом 2008–2009 гг.

5.5. Тесты Чоу на наличие структурной стабильности во временном ряде

Диагностика выбросов в остатках является не единственным инструментом для выявления проблем, мешающих повышению точности прогностических моделей. В этом смысле, пожалуй, еще большее значение имеет тест Грегори Чоу на наличие структурной стабильности временного ряда. Поэтому следующим нашим шагом будет оценка на основе этого теста стабильности временного ряда за период с июня 1992 г. по апрель 2010 г. С методикой проведения этого теста можно познакомиться в алгоритме действий № 18.

¹ Дрейпер Н., Смит Г. Прикладной регрессионный анализ. С. 190.

Таблица 5.9

Сравнение стандартных и студентизированные остатки, полученных при решении уравнения регрессии

$$\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2)$$

на основе данных за период с июня 1992 г. по апрель 2010 г.

| Наблюдения | Остатки | Стандартные остатки | Студентизированные остатки |
|------------|---------|------------------------|-------------------------------|
| 1998M08 | 1,6438 | <u>2,0195</u> | <u>2,0250</u> |
| 1998M09 | 7,6138 | <u>9,3541</u> | <u>12,2852</u> |
| 1998M10 | -2,6877 | <u>-3,3020</u> | <u>-4,5175</u> |
| 1998M11 | 1,8606 | <u>2,2858</u> | <u>2,3009</u> |
| 1998M12 | 2,1427 | <u>2,6325</u> | <u>2,6902</u> |
| 2008M12 | 1,3886 | 1,7059 | 1,7165 |
| 2009M01 | 5,4182 | <u>6,6566</u> | 7,5472 |
| 2009M02 | -1,6809 | <u>-2,0651</u> | <u>-2,3719</u> |
| 2009M03 | -1,8647 | <u>-2,2909</u> | <u>-2,3171</u> |
| 2009M04 | -0,2761 | -0,3392 | -0,3425 |
| 2009M05 | -2,0764 | <u>-2,5510</u> | <u>-2,5932</u> |
| 2009M06 | 0,9775 | 1,2010 | 1,2251 |
| 2009M07 | 0,3148 | 0,3868 | 0,3859 |
| 2009M08 | -0,3886 | -0,4774 | -0,4766 |
| 2009M09 | -1,4698 | -1,8057 | -1,8158 |
| 2009M10 | -0,6223 | -0,7645 | -0,7698 |
| 2009M11 | 1,0537 | 1,2945 | 1,3013 |
| 2009M12 | 0,1306 | 0,1605 | 0,1602 |
| 2010M02 | -0,5936 | -0,7292 | -0,7280 |
| 2010M03 | -0,4806 | -0,5905 | -0,5899 |
| 2010M04 | 0,0623 | 0,0765 | 0,0764 |

Источник: расчеты автора.

Алгоритм действий № 18

Методика проведения теста Чоу на наличие структурной стабильности во временном ряде для прогностической модели $\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2)$

Шаг 1. Основные идеи, на которых строится тест Чоу на наличие структурной стабильности

Тест Чоу на диагностирование структурной стабильности проводится следующим образом. Сначала берется временной ряд (на-

пример, данные по ежемесячному курсу доллара за период с июня 1992 г. по апрель 2010 г.), относительно которого выдвигается нулевая гипотеза о его структурной стабильности. Потом этот временной ряд делится на два периода наблюдений, граница между которыми проводится в момент времени t , т.е. в момент предполагаемых структурных изменений. (При необходимости EViews позволяет проводить тест на наличие во временном ряде структурных изменений не только в какой-то один момент времени t , но и сразу для нескольких моментов, деля выборку на несколько соответствующих периодов.)

Проверка нулевой гипотезы идет путем сравнения разницы между суммой квадратов остатков, которую мы получаем, построив уравнение регрессии для единого временного ряда, и суммой квадратов остатков, получаемой при построении уравнения регрессии отдельно для каждого периода этого ряда. При этом в соответствии с методикой, предложенной Г. Чоу, определяется фактическое значение F -критерия и LR -статистики (*log likelihood ratio statistic* — соотношение статистики логарифмов правдоподобия). Если уровни значимости F -критерия и LR -статистики оказываются меньше 0,05, то тогда нулевая гипотеза о структурной стабильности временного ряда отвергается, а следовательно, влияние структурных изменений признается существенным.

Шаг 2. Проведение в EViews теста Чоу на наличие структурной стабильности

В соответствии с данными табл. 5.4 вполне логично предположить, что самые значительные структурные изменения в исследуемом временном ряде могли произойти после самого крупного скачка курса доллара, имевшего место в сентябре 1998 г. Поэтому мы решили выделить в нашем временном ряде период с октября 1998 г. по апрель 2010 г. При этом для корректного проведения теста необходимо, чтобы количество наблюдений в каждом из выделенных периодов временного ряда было по меньшей мере равно количеству параметров в оцененной нами статистической модели. Впрочем, это требование соблюдено, поскольку в анализируемой статистической модели всего лишь два параметра, а в самом малом выделенном периоде временного ряда имеется 74 наблюдения.

Чтобы в EViews провести тест Чоу на наличие структурной стабильности, в меню оцененного уравнения регрессии необходимо воспользоваться опциями VIEW/STABILITY TESTS/CHOW BREAKPOINT TEST... (смотреть/тесты на стабильность/тест Чоу на структурные изменения). В результате открывается диалоговое мини-окно CHOW TESTS (тесты

Чоу), в котором нужно указать конкретное наблюдение, когда произошло предполагаемое структурное изменение во временном ряде. В этом случае в мини-окно введено обозначение — 98m10, т.е. указан октябрь 1998 г. (рис. 5.8). Следовательно, можно посмотреть, произошли ли структурные изменения в октябре 1998 г.

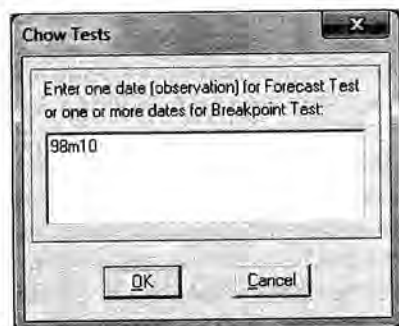


Рис. 5.8. Проверка на структурные изменения в диалоговом мини-окне CHOW TESTS

После того как мы щелкнули кнопку ОК, в мини-окне CHOW TESTS появился вывод данных по результатам тестирования, которые приведены в табл. 5.10. Поскольку уровни значимости (*Probability*) как *F*-критерия (*F-statistic*), так и *LR*-статистики (*Log likelihood ratio* — соотношения логарифмов правдоподобия) у нас оказались равны нулю, т.е. получились меньше критического значения, равного 0,05, следовательно, нулевая гипотеза о наличии структурной стабильности во временном ряде в октябре 1998 г. отвергается.

Таблица 5.10

Вывод данных по результатам теста Чоу на наличие структурной стабильности в октябре 1998 г. во временном ряде для модели
USDOLLAR = a × USDOLLAR(−1) + b × USDOLLAR(−2)

| Chow Breakpoint Test: 1998M10 | | | |
|-------------------------------|----------|-------------|----------|
| F-statistic | 42,11122 | Probability | 0,000000 |
| Log likelihood ratio | 72,12123 | Probability | 0,000000 |

Некоторые математические подробности для теста Чоу на наличие структурной стабильности во временном ряде

После того как была выдвинута нулевая гипотеза о структурной стабильности временного ряда, далее нам приходится решать несколько уравнений регрессии $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$ как от-

носительно единого временного ряда, так и относительно каждого выделенного периода наблюдений. Напомним, что в этом случае мы предположили, что структурная нестабильность возникла в октябре 1998 г., а потому временной ряд нами разделен на два периода: с июня 1992 г. по сентябрь 1998 г. и с октября 1998 г. по апрель 2010 г. Таким образом, мы находим сумму квадратов остатков, полученных как по единому уравнению регрессии для всего временного ряда, так и по остальным уравнениям регрессии (назовем их совокупность объединенной кусочно-линейной прогностической моделью) для каждого выделенного периода наблюдений.

Далее складываем суммы квадратов остатков, полученных в объединенной кусочно-линейной прогностической модели, по формуле

$$SS_{\text{ост}}^{\text{кл}} = SS_{\text{ост}}^1 + SS_{\text{ост}}^2, \quad (5.5)$$

где $SS_{\text{ост}}^{\text{кл}}$ — сумма квадратов остатков, полученных в объединенной кусочно-линейной модели;

$SS_{\text{ост}}^1$ и $SS_{\text{ост}}^2$ — сумма квадратов остатков, полученных по уравнениям регрессии, решенных относительно каждого выделенного периода временного ряда.

После чего находим фактическое значение *F*-критерия по формуле

$$F_{\text{факт}} = \frac{(SS_{\text{ост}}^{\text{кл}} - SS_{\text{ост}}^1) : k}{SS_{\text{ост}}^1 : (n - 2k)}, \quad (5.6)$$

где $SS_{\text{ост}}^{\text{кл}}$ — сумма квадратов остатков, полученных по единому уравнению регрессии для всего временного ряда;

n — количество наблюдений во всем временном ряде;

k — количество параметров в уравнении.

$$F_{\text{факт}} = \frac{141,117 - 100,584 : 2}{100,584 : 209} = 42,111.$$

Затем в Excel с помощью функции ФРАСП находим значимость фактического *F*-критерия:

ФРАСП($F_{\text{факт}}$; числитель степеней свободы; знаменатель степеней свободы) = ФРАСП(42,111; 2; 209) = 0.

Таким образом, поскольку значимость фактического *F*-критерия равна нулю, это позволяет нам отвергнуть нулевую гипотезу о структурной стабильности временного ряда.

LR-статистика в этом тесте рассчитывается путем сравнения ограниченного и неограниченного максимума функции логарифма правдоподобия. Причем *LR*-статистика — при нулевой гипотезе об отсутствии структурных изменений — имеет асимптотическое χ^2 (хи-квадрат) распределение со степенями свободы, равными

$$(m - 1)k,$$

где m — число периодов во временном ряде;

k — количество параметров в уравнении регрессии.

При значимости LR-статистики меньше 0,05 нулевая гипотеза о структурной стабильности отвергается.

Поскольку мы уже научились проводить тест Чоу на структурную стабильность временного ряда, то продолжим наше исследование уровня надежности модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$, используя при этом рыночные данные за период с июня 1992 г. по апрель 2010 г. Теперь нашей задачей будет последовательное тестирование структурной стабильности временного ряда после резких скачков курса доллара, вошедших в топ-двадчатку самых волатильных месяцев (см. табл. 5.4).

Следует иметь в виду, что применение теста Чоу, как правило, предполагает соблюдение предпосылок о нормальном распределении остатков и независимости их распределения. К сожалению, в этом случае, поскольку мы имеем дело с уравнением авторегрессии, эти предпосылки не выполняются. Тем не менее тест и в этой ситуации показал себя достаточно чувствительным к структурным изменениям.

По результатам тестирования мы составили табл. 5.11, в которой все наблюдения размещены в календарной последовательности. При этом каждый месяц резкого скачка курса доллара взят в качестве первого наблюдения, включенного в период после предполагаемых структурных изменений во временном ряде, поскольку в этом случае тест становится наиболее чувствительным к изменению стабильности.

И еще один важный момент в связи с тем, что поскольку в период с января 2000 г. по январь 2004 г. ни одно из наблюдений не вошло в топ-двадчатку самых волатильных, а информация по этому периоду нам также важна, то мы заполнили этот период, протестировав менее волатильные месяцы. Их мы не стали выделять жирным шрифтом, чтобы отличить от остальных, гораздо более волатильных наблюдений.

Судя по табл. 5.11, начиная с октября 1998 г. и по июнь 2000 г. в исследуемом временном ряде наблюдаются структурные изменения. Однако с конца 1998 г. и до середины 2000 г. волатильность на валютном рынке стала понемногу затухать, а к июлю 2000 г. тест Чоу вновь стал уверенно показывать наличие структурной стабильности. Впрочем, за одним-единственным исключением: после резкого роста доллара в январе 2009 г. на 18,7% — в ходе так называемой плавной девальвации рубля — тест Чоу вновь выявил структурную нестабильность временного ряда, которая, впрочем, восстановилась уже в следующем месяце.

Вызывает большие сомнения, что 26,8%-ное повышение курса доллара в августе 1998 г. не привело к структурным изменениям во временном ряде. Тем не менее уровни значимости F -критерия и LR -статистики, по-

Таблица 5.11

Влияние скачка доллара на структурную стабильность временного ряда по результатам теста Чоу для модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$

| Месяц, тестируемый на структурные изменения | Скачок курса доллара в этом месяце, % (по модулю) | Период до структурных изменений | Период после структурных изменений | Значимость F -критерия | Значимость LR -статистики |
|---|---|---------------------------------|------------------------------------|--------------------------|-----------------------------|
| Август 1998 г. | 26,7 | Август 1992—Июль 1998 | Август 1998—Апрель 2010 | 0,714 | 0,710 |
| Сентябрь 1998 г. | 103,2 | Август 1992—Август 1998 | Сентябрь 1998—Апрель 2010 | 0 | 0 |
| Ноябрь 1998 г. | 11,7 | Август 1992—Октябрь 1998 | Ноябрь 1998—Апрель 2010 | 0,001 | 0,001 |
| Декабрь 1998 г. | 15,5 | Август 1992—Ноябрь 1998 | Декабрь 1998—Апрель 2010 | 0,001 | 0,001 |
| Январь 1999 г. | 9,4 | Август 1992—Декабрь 1998 | Январь 1999—Апрель 2010 | 0,001 | 0,001 |
| Март 1999 г. | 5,8 | Август 1992—Февраль 1999 | Март 1999—Апрель 2010 | 0,001 | 0,001 |
| Октябрь 1999 г. | 3,9 | Август 1992—Сентябрь 1999 | Октябрь 1999—Апрель 2010 | 0,018 | 0,017 |
| Январь 2000 г. | 5,7 | Август 1992—Декабрь 1999 | Январь 2000—Апрель 2010 | 0,007 | 0,006 |
| Май 2000 г. | 0,6 | Август 1992—Апрель 2000 | Май 2000—Апрель 2010 | 0,048 | 0,045 |
| Июнь 2000 г. | 1 | Август 1992—Май 2000 | Июнь 2000—Апрель 2010 | 0,067 | 0,064 |
| Декабрь 2000 г. | 1,1 | Август 1992—Ноябрь 2000 | Декабрь 2000—Апрель 2010 | 0,138 | 0,132 |
| Декабрь 2001 г. | 0,8 | Август 1992—Ноябрь 2001 | Декабрь 2001—Апрель 2010 | 0,174 | 0,168 |
| Декабрь 2002 г. | 1 | Август 1992—Ноябрь 2002 | Декабрь 2002—Апрель 2010 | 0,208 | 0,202 |
| Январь 2004 г. | 3,3 | Август 1992—Декабрь 2003 | Январь 2004—Апрель 2010 | 0,640 | 0,635 |
| Сентябрь 2007 г. | 2,7 | Август 1992—Август 2007 | Сентябрь 2007—Апрель 2010 | 0,911 | 0,909 |
| Август 2008 г. | 4,8 | Август 1992—Июль 2008 | Август 2008—Апрель 2010 | 0,907 | 0,905 |

Окончание

| Месяц, тестируемый на структурные изменения | Скачок курса доллара в этом месяце, % (по модулю) | Период до структурных изменений | Период после структурных изменений | Значимость F-критерия | Значимость LR-статистики |
|---|---|---------------------------------|------------------------------------|-----------------------|--------------------------|
| Сентябрь 2008 г. | 2,7 | Август 1992–Август 2008 | Сентябрь 2008–Апрель 2010 | 0,929 | 0,927 |
| Октябрь 2008 г. | 5,1 | Август 1992–Сентябрь 2008 | Октябрь 2008–Апрель 2010 | 0,933 | 0,932 |
| Ноябрь 2008 г. | 4,0 | Август 1992–Октябрь 2008 | Ноябрь 2008–Апрель 2010 | 0,91 | 0,908 |
| Декабрь 2008 г. | 6,4 | Август 1992–Ноябрь 2008 | Декабрь 2008–Апрель 2010 | 0,705 | 0,701 |
| Январь 2009 г. | 20,5 | Август 1992–Декабрь 2008 | Январь 2009–Апрель 2010 | 0,004 | 0,003 |
| Март 2009 г. | 4,8 | Август 1992–Февраль 2009 | Март 2009–Апрель 2010 | 0,344 | 0,337 |
| Май 2009 г. | 6,8 | Август 1992–Апрель 2009 | Май 2009–Апрель 2010 | 0,552 | 0,546 |
| Сентябрь 2009 г. | 4,7 | Август 1992–Август 2009 | Сентябрь 2009–Апрель 2010 | 0,969 | 0,968 |
| Октябрь 2009 г. | 3,5 | Август 1992–Сентябрь 2009 | Октябрь 2009–Апрель 2010 | 0,773 | 0,770 |

Источник: расчеты автора.

лученные в ходе тестирования, оказались в августе 1998 г. существенно выше 0,05. Это объясняется тем, что в табл. 5.6 представлены результаты тестирования для статистической модели, построенной на базе данных за период с июня 1992 г. по апрель 2010 г. При такой базе данных прирост курса доллара в августе 1998 г. на 1,67 руб. не выглядит чем-то экстраординарным, хотя сразу после августовского дефолта столь значительный взлет американской валюты буквально шокировал участников рынка.

Посмотрим, что произойдет, если мы возьмем в качестве базы данных период с июня 1992 г. по август 1998 г., т.е. фактически смоделируем ситуацию реального прогнозирования. С этой целью представим, что мы делаем прогноз на сентябрь 1998 г. в августе 1998 г., а потому более поздней информацией по курсу доллара не обладаем. Тем более что мы уже умеем быстро менять нашу базу данных (см. алгоритм действий № 15 «Как в EViews можно быстро изменить выборку данных»).

После того как мы в очередной раз изменили выборку, у нас появилась возможность построить прогноз на сентябрь 1998 г. на основе рыночных данных (по курсу доллара на конец каждого месяца) за период с июня 1992 г. по август 1998 г. С этой целью мы сначала решаем уравнение регрессии (см. алгоритм действий № 6 «Как решить уравнение регрессии в EViews»), а потом делаем прогноз и соответственно сразу же находим остатки (см. алгоритм действий № 8 «Как оценить точность статистической модели в EViews»). Вывод данных по этому уравнению регрессии представлен в табл. 5.12.

Таблица 5.12

Вывод данных по уравнению регрессии
 $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$,
 на основе которого был составлен прогноз на сентябрь 1998 г.

| Dependent Variable: USDOLLAR | | | | |
|---|-------------|-----------------------|-------------|-----------|
| Method: Least Squares | | | | |
| Date: 06/29/10 Time: 08:31 | | | | |
| Sample (adjusted): 1992M08 1998M08 | | | | |
| Included observations: 73 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 1,630888 | 0,173714 | 9,388327 | 0,0000 |
| USDOLLAR(-2) | -0,618537 | 0,176331 | -3,507809 | 0,0008 |
| R-squared | 0,989480 | Mean dependent var | | 3,807411 |
| Adjusted R-squared | 0,989332 | S.D. dependent var | | 2,122971 |
| S.E. of regression | 0,219274 | Akaike info criterion | | -0,169970 |
| Sum squared resid | 3,413772 | Schwarz criterion | | -0,107218 |
| Log likelihood | 8,203908 | Durbin — Watson stat | | 1,225332 |

Источник: расчеты автора по данным Банка России.

Согласно прогнозу, составленному по этому уравнению регрессии, курс американского доллара к концу сентября 1998 г. должен был вырасти до 9 руб. 3 коп. Однако в действительности стоимость американской валюты к тому времени взлетела до 16 руб. 6 коп., т.е. оказалась выше прогнозируемого значения на 7 руб. 3 коп. Теперь посмотрим, сможем ли мы диагностировать структурные изменения в августе 1998 г., не обращаясь при этом к более поздней рыночной информации.

С этой целью лучше воспользоваться другим тестом Чоу — тестом на точность прогноза. Дело в том, что тест на структурную стабильность требует, чтобы количество наблюдений в каждом из выделенных периодов временного ряда было равно количеству параметров в оцененной статистической модели, которых у нас два (по числу переменных). Таким образом, для проведения теста Чоу на стабильность нам пришлось бы выделять в отдельный период два месяца — июль и август 1998 г., хотя в первом месяце, как известно, курс доллара еще колебался в рамках установленного российским правительством коридора. Тест Чоу на точность прогноза лишен этого недостатка, поэтому нам нужно уметь им пользоваться. Для этого используется алгоритм действий № 19.

Алгоритм действий № 19

Методика проведения теста Чоу на точность прогноза для прогностической модели

$$\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2)$$

Шаг 1. Основные идеи, на которых построен тест Чоу на точность прогноза

Тест Чоу на точность прогноза оценивает две статистические модели: одну, построенную при помощи всей выборки данных, и другую, построенную на неполной части выборки, в которую не включаются прогнозируемые наблюдения. При этом выдвигается нулевая гипотеза о структурной стабильности во временном ряде. Однако в случае выявления существенной разницы между двумя моделями нулевая гипотеза отвергается и делается вывод о значимости структурных изменений, произошедших в момент перехода от неполной выборки к прогнозируемым наблюдениям.

Для оценки результатов теста EViews сообщает две статистики: величины F -критерия и LR -статистики. Расчет F -критерия основан на сравнении суммы квадратов остатков, полученных для моделей, основанных на всей и неполной выборках (см. формулу (5.6)), а LR -статистики — на сравнении соотношения ограниченного и неограниченного максимума логарифма правдоподобия. При этом в случае если уровень значимости F -критерия и LR -статистики меньше 0,05, нулевая гипотеза отвергается. Некоторые математические подробности по этому тесту приводятся ниже.

Шаг 2. Проведение в EViews теста Чоу на точность прогноза

Чтобы в EViews получить результаты теста Чоу на точность прогноза, необходимо выбрать следующие опции: VIEW/STABILITY TESTS/CHOW FORECAST TEST... (посмотреть/тесты на стабильность/тест Чоу на точность прогноза). После чего в появившемся диалоговом мини-окне CHOW TESTS мы указываем прогнозируемое наблюдение — 98m08, т.е. август 1998 г. (рис. 5.9). Таким образом, все остальные наблюдения у нас попадут в неполную выборку, охватывающую период с июня 1992 г. по июль 1998 г.

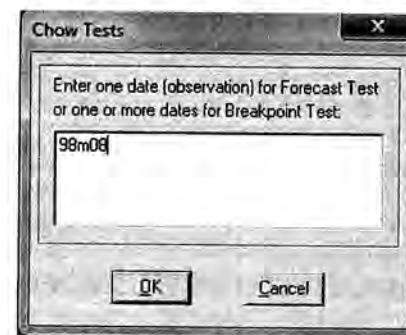


Рис. 5.9. Формирование выборки с июня 1992 г. по июль 1998 г. в диалоговом мини-окне CHOW TESTS

Если в диалоговом мини-окне CHOW TESTS мы щелкнем кнопку OK, то получим готовый вывод данных с результатами теста Чоу на точность прогноза. Эти данные поместим в табл. 5.13, из которой следует, что уровень значимости как F -критерия, так LR -статистики у нас оказался равен нулю. Следовательно, нулевая гипотеза о структурной стабильности во временном ряде отвергается и делается вывод о значимости структурных изменений во временном ряде, произошедших в августе 1998 г. Таким образом, вывод о наличии структурных изменений зависит не только от этих изменений, но и от объема взятой нами выборки.

Таблица 5.13

Вывод данных по результатам теста Чоу на точность прогноза относительно августа 1998 г. для модели $\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2)$

| Chow Forecast Test: Forecast from 1998M08 to 1998M08 | | | |
|--|----------|-------------|----------|
| F-statistic | 200,2762 | Probability | 0,000000 |
| Log likelihood ratio | 98,61930 | Probability | 0,000000 |

Окончание

| Test Equation: | | | | |
|------------------------------|-------------|-----------------------|-------------|-----------|
| Dependent Variable: USDOLLAR | | | | |
| Method: Least Squares | | | | |
| Date: 06/29/10 Time: 10:15 | | | | |
| Sample: 1992M08 1998M07 | | | | |
| Included observations: 72 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 1,681169 | 0,089106 | 18,86708 | 0,0000 |
| USDOLLAR(-2) | -0,677362 | 0,090472 | -7,486989 | 0,0000 |
| R-squared | 0,997125 | Mean dependent var | | 3,750500 |
| Adjusted R-squared | 0,997083 | S.D. dependent var | | 2,081037 |
| S.E. of regression | 0,112386 | Akaike info criterion | | -1,506365 |
| Sum squared resid | 0,884147 | Schwarz criterion | | -1,443124 |
| Log likelihood | 56,22914 | Durbin — Watson stat | | 1,835158 |

Некоторые математические подробности для теста Чоу на точность прогноза

Расчет F -критерия для теста Чоу на точность прогноза построен на сравнении суммы квадратов остатков, полученных для двух моделей, основанных соответственно на всей и неполной выборках.

При этом вычисления делаются по следующей формуле:

$$F_{\text{факт}} = \frac{(SS_1 - SS_2) : T_2}{SS_2 : (T_1 - k)}, \quad (5.7)$$

где SS_1 — сумма квадратов остатков, полученных по уравнению регрессии, построенному на всей выборке;

SS_2 — сумма квадратов остатков, полученных по уравнению регрессии, построенному на неполной выборке;

T_1 — количество наблюдений в неполной выборке;

T_2 — количество прогнозируемых наблюдений, т.е. наблюдений, не вошедших в полную выборку;

k — количество параметров в уравнении регрессии.

Таким образом, в нашем случае фактический F -критерий в тесте Чоу на точность прогноза относительно прогнозируемого наблюдения — августа 1998 г. будет иметь следующее значение:

$$F_{\text{факт}} = \frac{(3,4137 - 0,8841) : 1}{0,8841 : (72 - 2)} = 200,28.$$

Далее находим уровень значимости $F_{\text{факт}}$ с помощью функции в Excel $\text{FPASP}(200,28; 1; 70) = 0$. Поскольку уровень значимости $F_{\text{факт}}$ равен нулю, то, следовательно, нулевая гипотеза отвергается.

Как мы уже говорили ранее, LR -статистика этого теста основана на сравнении соотношения ограниченного и неограниченного максимума логарифма правдоподобия. Причем как ограниченный, так и неограниченный логарифм правдоподобия находятся путем оценки всей выборки наблюдений. Однако при расчете ограниченного логарифма правдоподобия используется первоначальный набор независимых переменных, в то время как для нахождения неограниченного логарифма правдоподобия в первоначальный набор регрессоров добавляют еще фиктивную переменную, которая равна единице — для прогнозируемых наблюдений выборки и равна нулю — для остальных наблюдений выборки. Следовательно, в нашем случае фиктивная переменная равна единице лишь для августа 1998 г.

Следует иметь в виду, что при нулевой гипотезе об отсутствии структурных изменений LR -статистика имеет асимптотическое χ^2 (хи-квадрат) распределение со степенями свободы, равными количеству прогнозируемых наблюдений. В том случае, если уровень значимости LR -статистики оказывается меньше 0,05, нулевая гипотеза о структурной стабильности отвергается.

Таким образом, тесты Чоу на структурную стабильность и на точность прогноза помогают анализировать устойчивость временного ряда. При этом тест на структурную стабильность, на наш взгляд, лучше подходит для ретроспективного анализа устойчивости статистической модели за весь период наблюдений, а тест на точность прогноза — для анализа ее стабильности относительно последнего наблюдения.

Причем в том случае, когда тест на точность прогноза свидетельствует о структурной неустойчивости, возникшей в модели в результате резкого изменения курса доллара в последнем наблюдении, то для устранения смещения в коэффициентах регрессии (и (или) величины константы) в уравнение можно ввести фиктивную переменную. Приравняем к единице фиктивную переменную для последнего наблюдения, а все остальные наблюдения приравняем к нулю, и тем самым прогностической моделью будет аппроксимирован последний рост без изменения коэффициентов регрессии и константы (свободного члена) уравнения. Еще более надежным способом получения точного прогноза в ситуации, когда тест Чоу на точность прогноза показал структурную неустойчивость, является отказ от уравнения авторегрессии с нестационарной ARMA-структурой и переход к уравнению ав-

торегрессии со стационарной ARMA-структурой, поскольку внешние шоки в гораздо меньшей степени влияют на коэффициенты регрессии и константу последнего уравнения. О том, как построить прогностическую модель со стационарной ARMA-структурой, мы будем говорить в главе 6.

5.6. Структурные изменения в курсе доллара, произошедшие в августе—октябре 1998 г.

Пока остановимся на тестировании характера структурных изменений во временном нестационарном ряде, поскольку по форме они могут быть различными. Вполне очевидно, что в том случае, когда тестирование показывает нестабильность временного ряда, тогда перед нами стоит задача выявить характер произошедших структурных изменений. В общем виде этот анализ проводится следующим образом. Например, предположим, что в момент времени $t = 5$ в динамике временного ряда произошли кардинальные изменения. Чтобы понять характер этих изменений, нужно сравнить параметры следующего уравнения регрессии:

$$Y = a_1 + b_1 \times Y(-1) \text{ в момент времени } t \leq 5;$$

$$Y = a_2 + b_2 \times Y(-1) \text{ в момент времени } t > 5,$$

где $Y(-1)$ — независимая переменная с лагом в один месяц;
 a — свободный член уравнения регрессии;
 b — коэффициент регрессии уравнения регрессии.

Если, например, после момента времени $t = 5$ в уравнении регрессии (5.8) статистически значимо изменился свободный член уравнения, т.е. если мы пришли к выводу, что $a_1 \neq a_2$, это свидетельствует о произошедшем структурном изменении в виде сдвига. Геометрически это означает, что графики стабильного тренда и тренда со сдвигом продолжают оставаться параллельными друг другу (рис. 5.10), в то время как изменение в начальном уровне тренда со сдвигом произошло одновременно в момент времени $t = 5$ при неизменном среднем темпе прироста в обоих трендах за весь период времени t .

Если, например, после момента времени $t = 5$ в уравнении (5.8) статистически значимо изменился коэффициент регрессии, т.е. если мы пришли к выводу, что $b_1 \neq b_2$, это свидетельствует о произошедшем структурном изменении в виде изменения наклона. Геометрически это означает, что графики стабильного тренда и тренда с изменением на-



Рис. 5.10. Сравнение графиков стабильного тренда и тренда со сдвигом

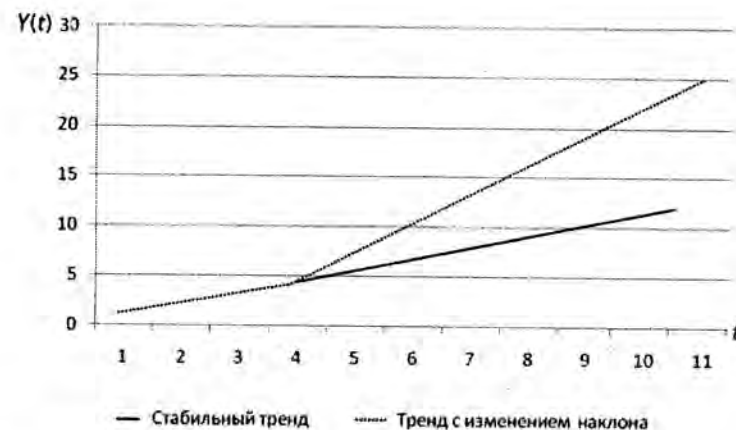


Рис. 5.11. Графики стабильного тренда и тренда с изменением наклона

клона становятся непараллельными друг другу, пересекаясь в момент времени $t = 5$ (рис. 5.11). При этом изменения в динамике обоих трендов обусловлены возникшей у них существенной разницей в среднем темпе прироста.

Если после момента времени $t = 5$ в уравнении регрессии (5.8) статистически значимо изменились как свободный член уравнения ($a_1 \neq a_2$), так и коэффициент регрессии ($b_1 \neq b_2$), это свидетельствует о произошедшем структурном изменении в виде одновременного сдвига и изменения наклона. В этой ситуации можно говорить о том,

что изменение в начальном уровне «тренда со сдвигом и изменением наклона» произошло одновременно в момент времени $t = 5$, что совпало и с возникшей в этот момент существенной разницей в среднем темпе прироста между обоими трендами. Поэтому вполне понятно, что с геометрической точки зрения график тренда со сдвигом и изменением наклона представляет собой сочетание тренда с изменением наклона и тренда со сдвигом. А потому график тренда со сдвигом и изменением наклона не параллелен стабильному тренду и резко отклоняется от последнего в момент времени, равный 5 (рис. 5.12).

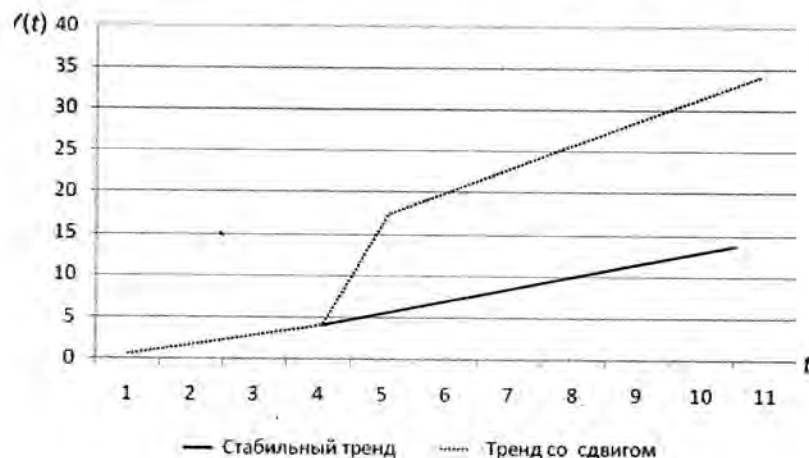


Рис. 5.12. Графики стабильного тренда и тренда со сдвигом и изменением наклона

После краткой общей характеристики различных видов структурных изменений нужно применить эти знания к исследованию нашей статистической модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$. Поэтому предположим, что в августе 1998 г. в динамике курса доллара произошли структурные изменения, характер которых нам следует определить. Чтобы справиться с поставленной задачей, необходимо воспользоваться методом, предложенным американским экономистом Д. Гуйарати¹.

¹ Gujarati D.N. Basic Econometrics. Third Ed. Mc. Graw-Hill. Inc., 1995. P. 509–513.

Алгоритм действий № 20

Методика проведения теста Д. Гуйарати по определению характера структурного сдвига (на примере прогностической модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$)

Шаг 1. Основная идея, на которой построен тест Д. Гуйарати

В основе метода Д. Гуйарати лежит достаточно простая и вполне понятная идея: поскольку основной задачей уравнения регрессии является аппроксимация динамики временного ряда, то, разделив этот ряд с помощью фиктивной переменной на два периода — до и после структурного изменения, можно выяснить характер произошедшего структурного изменения. При этом фиктивная переменная для наблюдений, расположенных до момента предполагаемого структурного изменения, у нас приравнивается к нулю, а на остальном участке временного ряда приравнивается к единице. Следует также заметить, что структурные изменения в виде сдвига диагностируются с помощью обычной фиктивной переменной (назовем ее *фиктивной переменной сдвига*), а изменение в виде наклона — с помощью еще одной переменной, представляющей собой произведение фиктивной переменной и независимой переменной (назовем ее *фиктивной переменной наклона*).

Перед тестированием выдвигается нулевая гипотеза об отсутствии в динамике временного ряда структурных изменений в виде сдвига и в виде наклона. Но если после решения уравнения регрессии фиктивные переменные сдвига и наклона окажутся статистически значимыми, то нулевая гипотеза будет считаться опровергнутой и будет принята альтернативная гипотеза.

Шаг 2. Проведение теста Д. Гуйарати

Поскольку мы хотим узнать характер структурных изменений, произошедших в августе 1998 г. во временном ряде, охватывающем период с августа 1992 г. по апрель 2010 г., то, следовательно, фиктивная переменная DUMMY до июля 1998 г. (включительно) будет приравнена к нулю, а для последующих наблюдений — к единице. Соответственно структурные изменения в виде сдвига будут выявлены в том случае, если фиктивная переменная DUMMY окажется статистически значимой. Кроме того, в уравнение регрессии $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$ будут введены не только фиктивная переменная сдвига DUMMY, но и новые переменные

DUMMY \times USDOLLAR(-1) и DUMMY \times USDOLLAR(-2), которые в случае их статистической значимости помогут нам выявить изменения в наклоне соответственно коэффициентов a , b и c . Таким образом, процедура тестирования по сути будет представлять собой решение обычного уравнения регрессии (см. алгоритм действий № 6 «Как решить уравнение регрессии в EViews»). При этом в диалоговое мини-окно EQUATION ESTIMATION следует ввести соответствующую формулу (рис. 5.13): USDOLLAR USDOLLAR(-1) USDOLLAR(-2) DUMMY DUMMY \times USDOLLAR(-1) DUMMY \times USDOLLAR(-2).

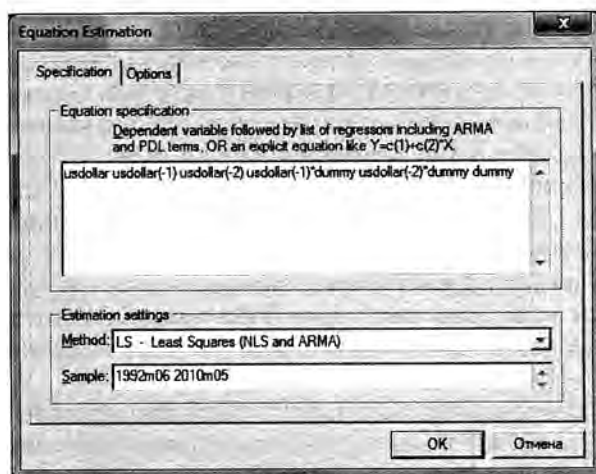


Рис. 5.13. Ввод формулы в диалоговое мини-окно EQUATION ESTIMATION для проведения теста Д. Гуйарати

Шаг 3. Интерпретация теста Д. Гуйарати

В результате решения нового уравнения регрессии мы получили следующий вывод данных (табл. 5.14). При этом коэффициент фиктивной переменной сдвига DUMMY у нас получился статистически значимым (Prob. = 0). Однако поскольку уровень значимости (*Probability*) у коэффициентов таких переменных, как USDOLLAR(-2) и фиктивных переменных наклона USDOLLAR(-1) \times DUMMY и USDOLLAR(-2) \times DUMMY, оказался больше 0,05, то, следовательно, их нельзя признать статистически значимыми, а потому состав переменных, включенных в это уравнение регрессии, нужно пересмотреть.

Поэтому мы решили избавиться от статистически незначимых фиктивных переменных наклона, введя вместо них новые. Исходя

Таблица 5.14
Вывод данных по результатам проведения теста Д. Гуйарати на определение характера структурных изменений в августе 1998 г.

| Dependent Variable: USDOLLAR | | | | |
|--|-------------|-----------------------|-------------|--------|
| Method: Least Squares | | | | |
| Date: 07/01/10 Time: 18:31 | | | | |
| Sample (adjusted): 1992M08 2010M04 | | | | |
| Included observations: 213 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 1,681169 | 0,567074 | 2,964638 | 0,0034 |
| USDOLLAR(-2) | -0,677362 | 0,575767 | -1,176452 | 0,2408 |
| USDOLLAR(-1) \times DUMMY | -0,629353 | 0,570886 | -1,102414 | 0,2716 |
| USDOLLAR(-2) \times DUMMY | 0,499182 | 0,578884 | 0,862318 | 0,3895 |
| DUMMY | 3,630466 | 0,442494 | 8,204555 | 0,0000 |
| R-squared | 0,996417 | Mean dependent var | 19,69163 | |
| Adjusted R-squared | 0,996348 | S.D. dependent var | 11,83593 | |
| S.E. of regression | 0,715231 | Akaike info criterion | 2,190772 | |
| Sum squared resid | 106,4035 | Schwarz criterion | 2,269676 | |
| Log likelihood | -228,3172 | Durbin — Watson stat | 2,066341 | |

Источник: расчеты автора.

из этого предполагаем, что изменение наклона в динамике курса доллара происходило в два этапа. Согласно нашему предположению, в сентябре 1998 г. структурное изменение в виде первого изменения наклона произошло за счет изменения коэффициента регрессии переменной USDOLLAR(-2), а в октябре 1998 г. имело место второе изменение наклона — за счет изменения коэффициента регрессии в переменной USDOLLAR(-1). (Попутно заметим, что сначала мы проверили предположение об изменении наклона относительно обеих переменных в сентябре 1998 г., но оно не подтвердилось, так как не все коэффициенты в уравнении регрессии оказались статистически незначимыми.)

Чтобы проверить наше последнее предположение, нам пришлось создать две дополнительные фиктивные переменные: DUMMY09

и DUMMY10. При этом DUMMY09 принимает нулевые значения с августа 1992 г. и до августа 1998 г. (включительно), а область нулевых значений — для DUMMY10 с августа 1992 г. до сентября 1998 г. (включительно). Во всех последующих наблюдениях вплоть до апреля 2010 г. эти фиктивные переменные равны единице. Две дополнительные фиктивные переменные потребовались нам для создания новых переменных наклона $USDOLLAR(-1) \times DUMMY10$ и $USDOLLAR(-2) \times DUMMY09$. После чего в диалоговое мини-окно EQUATION ESTIMATION была введена формула в следующем виде: $USDOLLAR \text{ } USDOLLAR(-1) \text{ } USDOLLAR(-2) \text{ } DUMMY \text{ } DUMMY10 \times \text{ } USDOLLAR(-1) \text{ } DUMMY09 \times \text{ } USDOLLAR(-2)$.

В результате мы получили вывод данных для уравнения регрессии с новыми фиктивными переменными наклона. Судя по табл. 5.15, все коэффициенты в этом уравнении регрессии оказались статистически значимыми, поскольку уровень их значимости оказался равен нулю. Следовательно, нулевую гипотезу об отсутствии в динамике курса доллара в августе 1998 г. структурных изменений сдвига и отсутствий в сентябре и октябре 1998 г. структурных изменений наклона можно считать опровергнутой с 1%-ным уровнем значимости. Это и позволяет нам взять за основу альтернативную гипотезу, что эти структурные изменения имели существенное значение.

Таблица 5.15

Вывод данных по результатам проведения теста Д. Гуйарати на определение характера структурных изменений в августе—октябре 1998 г.

| Dependent Variable: USDOLLAR | | | | |
|--|-------------|------------|-------------|--------|
| Method: Least Squares | | | | |
| Date: 07/02/10 Time: 10:43 | | | | |
| Sample (adjusted): 1992M08 2010M04 | | | | |
| Included observations: 213 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 1,927281 | 0,098494 | 19,56754 | 0,0000 |
| USDOLLAR(-2) | -0,928236 | 0,100336 | -9,251286 | 0,0000 |
| USDOLLAR(-1)*DUMMY10 | -0,817790 | 0,080535 | -10,15447 | 0,0000 |
| USDOLLAR(-2)*DUMMY09 | 0,754432 | 0,088115 | 8,561899 | 0,0000 |
| DUMMY | 1,865490 | 0,368679 | 5,059936 | 0,0000 |

Окончание

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0,997541 | Mean dependent var | 19,69163 |
| Adjusted R-squared | 0,997493 | S.D. dependent var | 11,83593 |
| S.E. of regression | 0,592593 | Akaike info criterion | 1,814577 |
| Sum squared resid | 73,04267 | Schwarz criterion | 1,893481 |
| Log likelihood | -188,2525 | Durbin — Watson stat | 1,638073 |

Источник: расчеты автора.

Из таблицы 5.15 можно сделать вывод, что до августа 1998 г. динамика курса доллара описывалась трендом, имеющим следующую формулу:

$$USDOLLAR = 1,927 \times USDOLLAR(-1) - 0,928 \times USDOLLAR(-2) \quad (5.9.1)$$

При этом интерпретация уравнения (5.9.1) была следующая: во-первых, рост на 1 руб. курса доллара в текущем месяце способствовал повышению курса доллара в будущем месяце в среднем на 1,927 руб.; во-вторых, рост на 1 руб. курса доллара в прошлом месяце способствовал снижению курса доллара в будущем месяце в среднем на -0,928 руб.

В августе 1998 г., когда произошел структурный сдвиг, формула тренда приобрела иной вид:

$$USDOLLAR = 1,927 \times USDOLLAR(-1) - 0,928 \times USDOLLAR(-2) + 1,865 \times DUMMY \quad (5.9.2)$$

Интерпретация формулы (5.9.2) аналогична предыдущей, за исключением того, что появление в уравнении фиктивной переменной DUMMY свидетельствует о единовременном повышении курса доллара в августе 1998 г. на 1,865 руб.

В сентябре 1998 г., когда имело место первое структурное изменение наклона, формула тренда снова изменилась:

$$USDOLLAR = 1,927 \times USDOLLAR(-1) - 0,928 \times USDOLLAR(-2) + 0,754 \times USDOLLAR(-2) \times DUMMY09 + 1,865 \times DUMMY \quad (5.9.3)$$

Формула (5.9.3) отличается от уравнения (5.9.2) тем, что у нее появилась фиктивная переменная наклона $USDOLLAR(-2) \times DUMMY09$, показывающая увеличение коэффициента регрессии у переменной

USDOLLAR(-2) на 0,754. Следовательно, с этого месяца вклад переменной USDOLLAR(-2) в динамику курса доллара с учетом структурного наклона, стал следующим: рост на 1 руб. курса доллара в прошлом месяце способствовал снижению курса доллара в будущем месяце в среднем на -0,174 руб.

В октябре 1998 г. после второго структурного изменения наклона наша формула приобрела следующий вид:

$$\begin{aligned} \text{USDOLLAR} = & 1,927 \times \text{USDOLLAR}(-1) - 0,818 \times \\ & \times \text{USDOLLAR}(-1) \times \text{DUMMY10} - 0,928 \text{ USDOLLAR}(-2) + \\ & + 0,754 \times \text{USDOLLAR}(-2) \times \text{DUMMY09} + 1,865 \times \text{DUMMY} \end{aligned} \quad (5.9.4)$$

Формула (5.9.4) отличается от предыдущей лишь тем, что у нее появилась фиктивная переменная наклона USDOLLAR(-1) × DUMMY10, которая показывает уменьшение коэффициента регрессии у переменной USDOLLAR(-1) на -0,818. Следовательно, с октября 1998 г. вклад переменной USDOLLAR(-2) в динамику курса доллара изменился таким образом: рост на 1 руб. курса доллара в текущем месяце способствовал повышению курса доллара в будущем месяце в среднем на 1,109 руб.

Следует иметь в виду, что формула (5.9.4) отличается, например, от формулы (3.4), описывающей аналогичный период, поскольку у этих формул разное число оцениваемых параметров.

На диаграмме, представленной на рис. 5.14, изображен график, наглядно показывающий характер структурных изменений в дина-

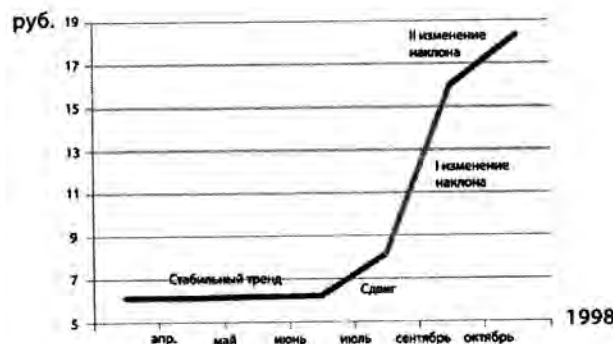


Рис. 5.14. Структурные изменения в динамике курса доллара в апреле – октябре 1998 г.

Источник: расчеты автора.

мике курса доллара, произошедших во время кризиса 1998 г. В частности, здесь можно увидеть как структурный сдвиг, произошедший в августе 1998 г., так и первое, и второе структурные изменения наклона, имевшие место в сентябре и октябре 1998 г.

Математические подробности по тесту Д. Гуйарати

Американский экономист Д. Гуйарати для оценки структурных изменений в динамике тренда, происходящих в момент времени t^* , предложил оценивать параметры следующего уравнения регрессии с фиктивными переменными:

$$Y_t = a + b \times Z_t + c \times t + d \times (Z_t \times t) + e, \quad (5.10)$$

где Y_t — зависимая переменная;

t — время;

a, b, c, d — параметры уравнения регрессии;

e — ошибка (остатки);

Z_t — фиктивная переменная, которая при $t < t^*$ равна нулю, а при $t \geq t^*$ равна единице.

Следовательно для момента времени $t < t^*$ мы получим следующее уравнение регрессии:

$$Z_t = 0 \Rightarrow Y_t = a + b \times 0 + c \times t + d \times (0 \times t) + e \Rightarrow Y_t = a + c \times t + e. \quad (5.10.1)$$

Соответственно для момента времени $t \geq t^*$ уравнение примет такой вид:

$$\begin{aligned} Z_t = 1 \Rightarrow Y_t &= a + b \times 1 + c \times t + d \times (1 \times t) + e \Rightarrow \\ &\Rightarrow Y_t = (a + b) + (c + d) \times t + e. \end{aligned} \quad (5.10.2)$$

Сравнив уравнение (5.10.1) с уравнением (5.10.2), нетрудно прийти к выводу, что при $Z_t = 1$ свободный член уравнения $a_2 = (a + b)$, а коэффициент регрессии — $c_2 \times t = (c + d) \times t$. Соответственно при $Z_t = 0$ свободный член уравнения $a_1 = a$, а коэффициент регрессии — $c_1 \times t = c \times t$. Таким образом, параметр b можно рассматривать как разницу между a_1 и a_2 , т.е. между свободными членами уравнений (5.10.1) и (5.10.2). В свою очередь параметр d следует рассматривать как разницу между c_1 и c_2 , т.е. между коэффициентами регрессии уравнений (5.10.1) и (5.10.2). Следовательно, параметр b оценивает структурный сдвиг, а параметр d — структурное изменение наклона в уравнении регрессии, произошедшее в момент времени t^* . Оценку параметров b и d можно провести, решив уравнение регрессии, а затем оценив их значимость при помощи t -критерия Стьюдента.

Подробности по этой теме можно узнать, ознакомившись с соответствующей литературой¹.

Таким образом, с помощью метода, предложенного американским экономистом Д. Гуйарати, нам удалось выяснить, что во временном

¹ См., например: Эконометрика. С. 327–334.

ряде по ежемесячному курсу доллара, охватывающем период с июня 1992 г. по апрель 2010 г., после августовского дефолта 1998 г. произошли следующие структурные изменения: во-первых, в августе 1998 г. произошел структурный сдвиг; во-вторых, в сентябре 1998 г. имело место первое структурное изменение наклона (изменился коэффициент регрессии факторной переменной с лагом в два месяца); в-третьих, в октябре 1998 г. имело место второе структурное изменение наклона (изменился коэффициент факторной переменной с лагом один месяц).

5.7. Построение статистической модели с оптимальным диапазоном интервального прогноза

А теперь посмотрим, подтвердит ли эти структурные изменения во временном ряде тест Чоу на точность прогноза. С этой целью проведем тестирование модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$, построенной на основе данных за период с июня 1992 г. по апрель 2010 г. При этом проверять на наличие структурной стабильности будем такие месяцы, как июль — декабрь 1998 г. и январь 1999 г., поскольку тест Д. Гуйарати показал наличие структурных изменений в августе — октябре 1998 г., но для большей надежности мы решили несколько расширить этот временной диапазон.

В результате получилась табл. 5.16, согласно которой уровень значимости F -критерия и LR -статистики для июля — октября 1998 г. оказался равен нулю, что опровергает нулевую гипотезу о структурной стабильности временного ряда относительно тестируемых наблюдений. Кстати, на основе этого теста аналогичный вывод можно сделать и для всех наблюдений с августа 1992 г. по октябрь 1998 г. включительно; в то время как уровень значимости F -критерия и LR -статистики по итогам проведения теста Чоу на точность прогноза относительно ноября 1998 г. и остальных, более поздних наблюдений (за исключением января — февраля 2009 г., но этот факт мы рассматриваем как временное явление) будет выше 0,05. Отсюда можно сделать вывод, что нулевая гипотеза о наличии структурной стабильности в этой части временного ряда подтверждается, что совпадает с аналогичными результатами, полученными с помощью метода Д. Гуйарати.

Чтобы построить статистическую модель с приемлемым диапазоном интервального прогноза, попробуем — с учетом итогов теста Чоу на точность прогноза — исключить из расчетной базы данных период с июля 1992 г. по октябрь 1998 г. (включительно). После чего на основе рыночных данных с ноября 1998 г. по апрель 2010 г. с помощью моде-

Таблица 5.16
Влияние скачка доллара на структурную стабильность временного ряда по результатам теста Чоу на точность прогноза для модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$

| Месяц, тестируемый на структурные изменения | Скачок курса доллара в этом месяце, % (по модулю) | Период до структурных изменений | Период после структурных изменений | Значимость F -критерия | Значимость LR -статистики |
|---|---|---------------------------------|------------------------------------|--------------------------|-----------------------------|
| Июль 1998 г. | 26,7 | Август 1992—Июнь 1998 | Июль 1998—Апрель 2010 | 0 | 0 |
| Август 1998 г. | 26,7 | Август 1992—Июль 1998 | Август 1998—Апрель 2010 | 0 | 0 |
| Сентябрь 1998 г. | 103,2 | Август 1992—Август 1998 | Сентябрь 1998—Апрель 2010 | 0 | 0 |
| Октябрь 1998 г. | 103,2 | Август 1992—Сентябрь 1998 | Октябрь 1998—Апрель 2010 | 0 | 0 |
| Ноябрь 1998 г. | 11,7 | Август 1992—Октябрь 1998 | Ноябрь 1998—Апрель 2010 | 0,993316 | 0,068146 |
| Декабрь 1998 г. | 15,5 | Август 1992—Ноябрь 1998 | Декабрь 1998—Апрель 2010 | 0,994214 | 0,086182 |
| Январь 1999 г. | 15,5 | Август 1992—Декабрь 1998 | Январь 1999—Апрель 2010 | 0,994214 | 0,086182 |

Источник: расчеты автора.

ли $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$ построим новое уравнение регрессии, вывод итогов которого представлен в табл. 5.17. Как нетрудно заметить, все переменные, включенные в это уравнение регрессии, оказались статистически значимыми ($Prob. = 0$). Далее эту модель будем называть нестационарной моделью с оптимизированным временным рядом, чтобы отличить ее от модели с полным временным рядом.

В результате точечный прогноз по курсу доллара на май 2010 г. оказался равен 29,287 руб. Таким образом, последняя цифра лишь немного отклонилась от 29,3137 руб., т.е. от точечного прогноза на май 2010 г., составленного по статистической модели с полным временным рядом (на основе данных с июня 1992 г. по апрель 2010 г.).

Далее составим точечные и интервальные прогнозы как для всех предыдущих наблюдений, включенных в базу данных, начиная с ноя-

Таблица 5.17

Вывод итогов по решению уравнения регрессии
 $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$
 на основе данных за период с июня 1992 г. по апрель 2010 г.

| Dependent Variable: USDOLLAR | | | | |
|------------------------------|-------------|-----------------------|-------------|----------|
| Method: Least Squares | | | | |
| Date: 08/24/10 Time: 16:30 | | | | |
| Sample: 1998M11 2010M04 | | | | |
| Included observations: 138 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 1,420453 | 0,077895 | 18,23557 | 0,0000 |
| USDOLLAR(-2) | -0,419429 | 0,078110 | -5,369739 | 0,0000 |
| R-squared | 0,933017 | Mean dependent var | | 28,14711 |
| Adjusted R-squared | 0,932524 | S.D. dependent var | | 2,785302 |
| S.E. of regression | 0,723514 | Akaike info criterion | | 2,204992 |
| Sum squared resid | 71,19220 | Schwarz criterion | | 2,247416 |
| Log likelihood | -150,1445 | Durbin — Watson stat | | 1,888462 |

Источник: расчеты автора.

бря 1998 г., так и на май 2010 г. (это наблюдение в базу данных не вошло). При этом будем пользоваться алгоритмом действий № 11 «Как в EViews построить точечный прогноз» и алгоритмом действий № 12 «Как в EViews построить интервальные прогнозы».

В результате появилась возможность сопоставить заданные уровни надежности с фактической долей точных интервальных прогнозов. После проведения соответствующих подсчетов получилась табл. 5.18. Судя по этой таблице, доля точных прогнозов у предложенной модели оказалась незначительно ниже заданного уровня при 99,9%-ном и 99%-ном уровнях надежности. В то время как при 95%-ном уровне надежности и ниже доля точных интервальных прогнозов становится на 0,7 процентного пункта выше заданного уровня. По мере снижения заданного уровня надежности эта положительная разница растет, достигая максимума при 40%-ном уровне надежности, когда она равна 31,2 процентного пункта.

Сравнив табл. 5.18 и табл. 4.8, можно сделать следующие выводы. Во-первых, фактическая точность интервальных прогнозов, составленных по нестационарной модели с оптимизированным временным

Таблица 5.18

Доля точных интервальных прогнозов, составленных по модели
 $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$, при различном уровне
 надежности в период с ноября 1998 г. по май 2010 г. (142 наблюдения)

| t-значе- ние | Уровень на- дежности, % | Количество точных про- гнозов | Доля точ- ных интер- вальных прогнозов, % | Разница между долей точных интервальных прогнозов и уровнем надежности, проц. пунктов |
|-----------------|----------------------------|-------------------------------------|---|---|
| 3,3641 | 99,9 | 138 | 99,3 | -0,6 |
| 2,6127 | 99 | 135 | 97,1 | -1,9 |
| 1,9777 | 95 | 133 | 95,7 | 0,7 |
| 1,6562 | 90 | 129 | 92,8 | 2,8 |
| 1,2879 | 80 | 124 | 89,2 | 9,2 |
| 1,0404 | 70 | 121 | 87,1 | 17,1 |
| 0,8443 | 60 | 118 | 84,9 | 24,9 |
| 0,6763 | 50 | 112 | 80,6 | 30,6 |
| 0,5256 | 40 | 99 | 71,2 | 31,2 |
| 0,3861 | 30 | 80 | 57,6 | 27,6 |
| 0,2538 | 20 | 65 | 46,8 | 26,8 |
| 0,1259 | 10 | 40 | 28,8 | 18,8 |

Источник: расчеты автора.

рядом, как и точность интервальных прогнозов, составленных по аналогичной модели с полным временным рядом, при 95%-ном уровне надежности оказалась выше заданного уровня. Во-вторых, фактическая точность интервальных прогнозов, рассчитанных по модели с полным временным рядом, в среднем немного выше, чем у модели с оптимизированным временным рядом, хотя при 95%-ном уровне надежности эта разница и незначительна.

Однако последняя модель опережает первую по таким важным параметрам, как средний диапазон интервального прогноза (в рублях); средний диапазон интервального прогноза (в процентах от среднего фактического курса); и индекс оптимальности интервальных прогнозов (см. табл. 6.24).

Стоит также отметить, что благодаря оптимизации временного ряда нам удалось получить временной ряд с приемлемым диапазоном интервального прогноза. В этом можно убедиться, если познакомиться с табл. 5.19, в которой в целях экономии места помещена лишь часть интервальных прогнозов.

Так, в ноябре 1998 г. общий диапазон интервального прогноза (верхняя граница интервального прогноза минус нижняя граница интервального прогноза) при 95%-ном уровне надежности составил 2 руб. 86,53 коп. при фактическом курсе доллара, равном 17 руб. 88 коп. В свою очередь при прогнозе на май 2010 г. общий диапазон интервального прогноза был равен 2 руб. 88,07 коп., а фактический курс доллара составил 30 руб. 49,56 коп. Следовательно, в ноябре 1998 г. общий диапазон интервального прогноза составлял 16,03% от фактической стоимости доллара, в то время как в мае 2010 г. эта цифра равнялась 9,42%.

Нетрудно также заметить, что за счет уменьшения стандартного отклонения (в структурно стабильном временном ряде, естественно, наблюдается более низкий уровень волатильности) ширина диапазона интервального прогноза в табл. 5.19 оказалась несколько меньше, чем в табл. 4.9.

Таблица 5.19

**Интервальные и точечные прогнозы по курсу доллара за период
ноябрь—декабрь 1998 г. и за период с января 2008 г. по май 2010 г.**

| Месяц | Факт | Прогноз | Нижняя граница интервального прогноза | Верхняя граница интервального прогноза | Точность интервального прогноза |
|---------|---------|---------|---------------------------------------|--|---------------------------------|
| 1998M11 | 17,88 | 16,0035 | 14,5709 | 17,4361 | Выше интервала прогноза |
| 1998M12 | 20,65 | 18,6826 | 17,2225 | 20,1427 | Выше интервала прогноза |
| ... | ... | ... | ... | ... | ... |
| 2008M01 | 24,4764 | 24,6534 | 23,2185 | 26,0883 | В интервале прогноза |
| 2008M02 | 24,1159 | 24,4722 | 23,0372 | 25,9072 | В интервале прогноза |
| 2008M03 | 23,5156 | 23,9894 | 22,5531 | 25,4257 | В интервале прогноза |
| 2008M04 | 23,6471 | 23,2879 | 21,8496 | 24,7262 | В интервале прогноза |
| 2008M05 | 23,7384 | 23,7265 | 22,2919 | 25,1611 | В интервале прогноза |
| 2008M06 | 23,4573 | 23,8010 | 22,3664 | 25,2356 | В интервале прогноза |
| 2008M07 | 23,4456 | 23,3634 | 21,9279 | 24,7989 | В интервале прогноза |

Продолжение

| Месяц | Факт | Прогноз | Нижняя граница интервального прогноза | Верхняя граница интервального прогноза | Точность интервального прогноза |
|---------|---------|---------|---------------------------------------|--|---------------------------------|
| 2008M08 | 24,5769 | 23,4647 | 22,0302 | 24,8992 | В интервале прогноза |
| 2008M09 | 25,2464 | 25,0766 | 23,6326 | 26,5206 | В интервале прогноза |
| 2008M10 | 26,543 | 25,5530 | 24,1151 | 26,9909 | В интервале прогноза |
| 2008M11 | 27,606 | 27,1140 | 25,6663 | 28,5617 | В интервале прогноза |
| 2008M12 | 29,3804 | 28,0801 | 26,6364 | 29,5238 | В интервале прогноза |
| 2009M01 | 35,4146 | 30,1547 | 28,6949 | 31,6145 | Выше интервала прогноза |
| 2009M02 | 35,7205 | 37,9818 | 36,2765 | 39,6871 | Ниже интервала прогноза |
| 2009M03 | 34,0134 | 35,8854 | 34,4459 | 37,3249 | Ниже интервала прогноза |
| 2009M04 | 33,2491 | 33,3322 | 31,8667 | 34,7977 | В интервале прогноза |
| 2009M05 | 30,9843 | 32,9626 | 31,5183 | 34,4069 | Ниже интервала прогноза |
| 2009M06 | 31,2904 | 30,0661 | 28,5834 | 31,5488 | В интервале прогноза |
| 2009M07 | 31,7555 | 31,4508 | 30,0132 | 32,8884 | В интервале прогноза |
| 2009M08 | 31,5687 | 31,9831 | 30,5446 | 33,4216 | В интервале прогноза |
| 2009M09 | 30,0922 | 31,5227 | 30,0847 | 32,9607 | В интервале прогноза |
| 2009M10 | 29,0488 | 29,5037 | 28,0466 | 30,9608 | В интервале прогноза |
| 2009M11 | 29,8179 | 28,6409 | 27,1939 | 30,0879 | В интервале прогноза |
| 2009M12 | 30,2442 | 30,1710 | 28,7305 | 31,6115 | В интервале прогноза |

Окончание

| Месяц | Факт | Прогноз | Нижняя граница интервального прогноза | Верхняя граница интервального прогноза | Точность интервального прогноза |
|---------|---------|---------|---------------------------------------|--|---------------------------------|
| 2010M01 | 30,4312 | 30,4540 | 29,0163 | 31,8917 | В интервале прогноза |
| 2010M02 | 29,9484 | 30,5408 | 29,1038 | 31,9778 | В интервале прогноза |
| 2010M03 | 29,3638 | 29,7766 | 28,3371 | 31,2161 | В интервале прогноза |
| 2010M04 | 29,2886 | 29,1487 | 27,7084 | 30,5890 | В интервале прогноза |
| 2010M05 | 30,4956 | 29,2870 | 27,8503 | 30,7237 | В интервале прогноза |

Источник: расчеты автора.

Контрольные вопросы и задания

1. Какую формулу нужно ввести в мини-окно EQUATION SPECIFICATION, чтобы проверить AR-структуру статистической модели на стационарность? Как находятся обратные единичные корни? При каком значении обратных единичных корней авторегрессионный процесс считается нестационарным?
2. Чем различаются функция импульсного ответа и функция накопленного импульсного ответа? Как изменяется инновационная неопределенность и импульсный ответ в нестационарной и в стационарной AR-моделях?
3. Почему в шестерку самых волатильных месяцев вошли только те месяцы, когда был зафиксирован резкий рост, а не падение курса доллара? Как этот факт можно объяснить с фундаментальной точки зрения? Как это влияет на распределение остатков?
4. Чем можно объяснить большую неточность прогноза, составленного на октябрь 1998 г.? Какие остатки считаются выбросами, по мнению Н. Дрейпера и Г. Смита? Как величина выброса определяется во многих статистических программах? Что такое студентизированные остатки и чем они отличаются от стандартных остатков? Какие остатки точнее учитывают волатильность во временном ряде?

5. В чем сходство и различие тестов Чоу на структурную стабильность и на точность прогноза? Какой из этих тестов лучше подходит для анализа стабильности статистической модели относительно последнего наблюдения? Как можно изменить статистическую модель в том случае, когда тест на точность прогноза свидетельствует о структурной неустойчивости, возникшей в модели в результате резкого изменения курса доллара в последнем наблюдении?
6. Чем отличается структурное изменение в виде тренда со сдвигом от структурного изменения в виде тренда с наклоном? С помощью какого теста выявляются такого рода изменения во временном ряде? Какие структурные изменения в динамике курса доллара к рублю были выявлены в августе, сентябре и октябре 1998 г.?
7. Почему из расчетной базы данных, на которых строилась статистическая модель $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$, была исключена часть наблюдений? Удалось ли в результате получить статистическую модель с оптимальным диапазоном интервального прогноза? Вывод свой обоснуйте.

ГЛАВА 6

ПОСТРОЕНИЕ СТАЦИОНАРНОЙ СТАТИСТИЧЕСКОЙ МОДЕЛИ

6.1. Тестирование исходного и логарифмического временного ряда на стационарность

В главе 5 с помощью анализа остатков на выбросы, тестов Чоу на стабильность и точность прогноза, а также метода Гуйарати по определению характера структурных изменений была выявлена нестабильность параметров уравнения регрессии $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$. Причем эта проблема особенно обостряется во время резких колебаний курса доллара (в первую очередь в периоды кризисов 1998 г. и 2008–2009 гг.). Мы также выяснили, что нестабильность параметров в этом уравнении регрессии обусловлена не только высокой волатильностью на рынке, но и его нестационарной AR-структурой. Об этом, в частности, свидетельствует тестирование этого уравнения регрессии на импульсный ответ (см. алгоритм № 14).

В связи с этим нам предстоит задача по созданию уравнения регрессии, обладающего стационарной AR- или ARMA-структурой. Напомним нашим читателям, что отличие первой от второй заключается в том, что первое уравнение представляет уравнение авторегрессии, а второе — уравнение авторегрессии со скользящей средней.

Вот что пишет о специфике стационарных временных рядов профессор статистики Стэнфордского университета Т. Андерсон: «Предполагается, что случайные составляющие имеют в каждый момент времени одинаковые дисперсии и некоррелированы. Они могут представлять собой ошибки наблюдения или нерегулярности иного рода. Предположения о равенстве дисперсий и отсутствии корреляции являются определенным приближением к действительному положению вещей. ...Иногда наблюдения лучше соответствуют условиям равенства дисперсий и аддитивности ошибки, если преобразовать масштаб

измерений изучаемой величины. Например, в ряде экономических исследований производится анализ не самих цен, а их логарифмов...»¹

Попробуем получить стационарные ряды, взяв логарифмы от исходного уровня временного ряда, содержащего данные по курсу доллара за период с июня 1992 г. по июнь 2010 г. Однако сначала убедимся, что исходный временной ряд, содержащий данные по ежемесячному курсу доллара за период с июня 1992 г. по июнь 2010 г., действительно нестационарен, и с этой целью воспользуемся указаниями алгоритма действий № 21.

Алгоритм действий № 21

Как провести тест на стационарность исходного уровня временного ряда

Мы уже проверяли остатки на стационарность (см. алгоритм действий № 9 «Как проверить в EViews остатки на стационарность модели»). В этом случае будем действовать аналогичным образом, однако вместо файла RESID откроем файл USDOLLAR, после чего воспользуемся опциями VIEW/UNIT ROOT TEST (посмотреть/тест на единичный корень), в результате чего появится диалоговое мини-окно UNIT ROOT TEST (рис. 6.1). Его мы заполним следующим образом. Параметр TEST TYPE (тип теста) установим на опции

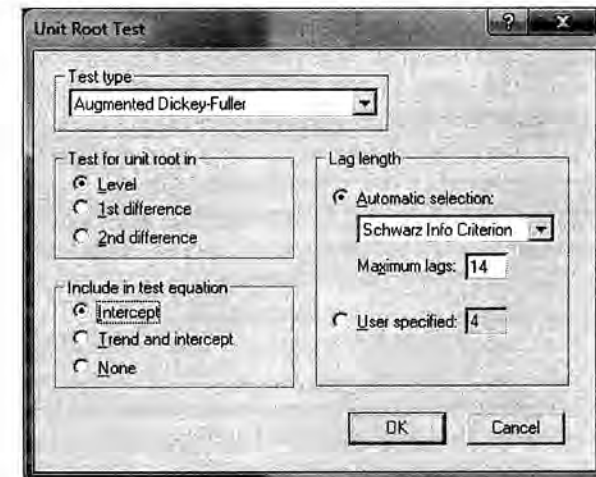


Рис. 6.1. Заполнение диалогового мини-окна UNIT ROOT TEST для получения результатов решения теста на стационарность

¹ Андерсон Т. Статистический анализ временных рядов. С. 43.

AUGMENTED DICKEY-FULLER (расширенный тест Дикки — Фуллера), а TEST FOR UNIT ROOT IN (тест на единичный корень для...) следует установить на опции LEVEL (исходный уровень ряда), так как мы проводим исследование исходного уровня временного ряда на стационарность. Параметр INCLUDE IN TEST EQUATION (включить в тестовое уравнение) установим на опции INTERCEPT (включить константу), поскольку мы предполагаем, что в исследуемом временном ряде может присутствовать свободный член уравнения (константа). Параметр LAG LENGTH (длина лага) установим на опции AUTOMATIC SELECTION (автоматический выбор), что позволит EViews самостоятельно выбрать длину лага. Вполне естественно, что при необходимости длину лага можно задать самому.

После щелчка мышкой кнопки OK в диалоговом мини-окне Unit Root test получим табл. 6.1 с результатами решения теста на стационарность. Однако полученный уровень значимости (*Prob.**) одностороннего *t*-критерия (*t-Statistic*), который равен 0,6166, свидетельствует, что нулевая гипотеза о наличии единичного корня не опровергается, а следовательно, исходный временной ряд нестационарен.

Таблица 6.1

Результаты проведения теста на стационарность исходного уровня временного ряда «Курс доллара за период с июня 1992 г. по июнь 2010 г.»

| | | | | |
|--|-----------|--|-------------|--------|
| Null Hypothesis: USDOLLAR has a unit root (нулевая гипотеза USDOLLAR имеет единичный корень) | | | | |
| Exogenous: Constant | | | | |
| Lag Length: 1 (Automatic based on SIC, MAXLAG = 14) | | | | |
| | | | t-Statistic | Prob.* |
| Augmented Dickey — Fuller test statistic | | | -1,328121 | 0,6166 |
| Test critical values: | 1% level | | -3,460739 | |
| | 5% level | | -2,874804 | |
| | 10% level | | -2,573917 | |
| *MacKinnon (1996) one-sided p-values. | | | | |
| Augmented Dickey — Fuller Test Equation | | | | |
| Dependent Variable: D(USDOLLAR) | | | | |
| Method: Least Squares | | | | |
| Date: 07/19/10 Time: 07:15 | | | | |

Окончание

| Sample (adjusted): 1992M08 2010M06 | | | | |
|--|-------------|-----------------------|-------------|----------|
| Included observations: 215 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | -0,006207 | 0,004674 | -1,328121 | 0,1856 |
| D(USDOLLAR(-1)) | 0,305284 | 0,065173 | 4,684193 | 0,0000 |
| C | 0,223195 | 0,107823 | 2,070020 | 0,0397 |
| R-squared | 0,101765 | Mean dependent var | | 0,144345 |
| Adjusted R-squared | 0,093291 | S.D. dependent var | | 0,852635 |
| S.E. of regression | 0,811889 | Akaike info criterion | | 2,434950 |
| Sum squared resid | 139,7428 | Schwarz criterion | | 2,481982 |
| Log likelihood | -258,7571 | F-statistic | | 12,00924 |
| Durbin — Watson stat | 2,034030 | Prob (F-statistic) | | 0,000011 |

Источник: расчеты автора.

Напомним, что альтернативная гипотеза об отсутствии единичного корня и стационарности исходного временного ряда может быть принята лишь при уровне значимости менее 0,05. В принципе, можно попробовать получить стационарный ряд, включив в тестовое уравнение (INCLUDE IN TEST EQUATION) вместо опции константа (INTERCEPT) другую опцию TREND AND INTERCEPT (тренд и константа) (см. рис. 6.1). Однако в результате у нас получился бы еще более высокий уровень значимости *t*-критерия = 0,9033, который с еще большим уровнем надежности подтвердил бы нулевую гипотезу о наличии единичного корня и нестационарности временного ряда.

Чтобы получить стационарный ряд, попробуем взять логарифмы от исходного уровня временного ряда. С этой целью нужно открыть файл USDOLLAR и воспользоваться опциями PROC/GENERATE BY EQUATION (выполнить/создать с помощью уравнения), после чего на экране появится диалоговое мини-окно GENERATE SERIES BY EQUATION (создать временной ряд по уравнению) (рис. 6.2), которое мы должны заполнить таким образом: USDOLLAR1 = log(USDOLLAR). В результате у нас появится новый логарифмический временной ряд, который поместим в файле USDOLLAR1.

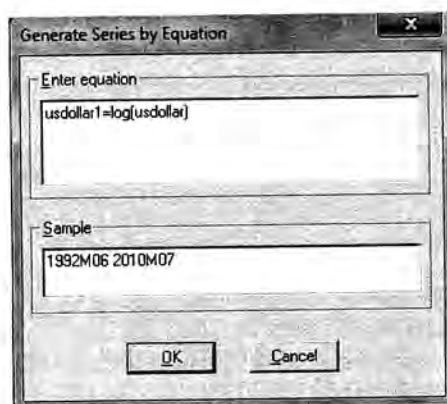


Рис. 6.2. Заполнение диалогового мини-окна GENERATE SERIES BY EQUATION

Следующей задачей будет тестирование логарифмического временного ряда на стационарность. С этой целью откроем файл USDOLLAR1 и воспользуемся опциями VIEW/UNIT ROOT TEST (посмотреть/тест на единичный корень). Далее будем действовать таким же образом, как и в алгоритме действий № 21. При этом параметр INCLUDE IN TEST EQUATION (включить в тестовое уравнение) установим на опции INTERCEPT (включить константу). В результате диалоговое мини-окно UNIT ROOT TEST приобретет следующий вид (рис. 6.3).

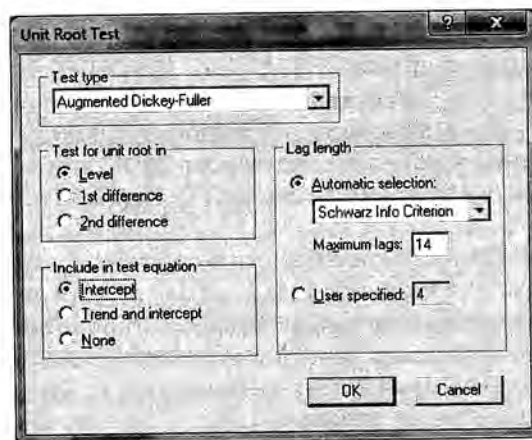


Рис. 6.3. Заполнение диалогового мини-окна UNIT ROOT TEST для получения логарифмического временного ряда

Нажав на кнопку ОК, получим следующий вывод итогов по результатам расширенного теста Дикки — Фуллера (табл. 6.2). В результате удастся получить уровень значимости (*Prob. **) одностороннего *t*-критерия (*t*-Statistic), равный нулю. Таким образом, нулевая гипотеза о наличии единичного корня и нестационарности логарифмического временного ряда опровергается и принимается альтернативная гипотеза о его стационарности.

Таблица 6.2

Результаты проведения теста на стационарность логарифмического временного ряда, полученного от исходного временно го ряда «Курс доллара за период с июня 1992 г. года по июнь 2010 г.»

Null Hypothesis: USDOLLAR1 has a unit root

Exogenous: Constant

Lag Length: 1 (Automatic based on SIC, MAXLAG = 14)

| | | | |
|--|-----------|-------------|--------|
| | | t-Statistic | Prob.* |
| Augmented Dickey—Fuller test statistic | | −5,549038 | 0,0000 |
| Test critical values: | 1% level | −3,460739 | |
| | 5% level | −2,874804 | |
| | 10% level | −2,573917 | |

*MacKinnon (1996) one-sided p-values.

Augmented Dickey — Fuller Test Equation

Dependent Variable: D(USDOLLAR1)

Method: Least Squares

Date: 07/19/10 Time: 08:45

Sample (adjusted): 1992M08 2010M06

Included observations: 215 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------------------|-------------|-----------------------|-------------|-----------|
| USDOLLAR1(−1) | −0,022988 | 0,004143 | −5,549038 | 0,0000 |
| D(USDOLLAR1(−1)) | 0,234086 | 0,064451 | 3,631981 | 0,0004 |
| C | 0,076651 | 0,012205 | 6,280174 | 0,0000 |
| R-squared | 0,267763 | Mean dependent var | | 0,024490 |
| Adjusted R-squared | 0,260855 | S.D. dependent var | | 0,078198 |
| S.E. of regression | 0,067229 | Akaike info criterion | | −2,547562 |
| Sum squared resid | 0,958191 | Schwarz criterion | | −2,500530 |
| Log likelihood | 276,8630 | F-statistic | | 38,76180 |
| Durbin — Watson stat | 2,013422 | Prob (F-statistic) | | 0,0000 |

6.2. Построение модели авторегрессии со скользящей средней и стационарной ARMA-структурой

Нам удалось выяснить, что созданный логарифмический временной ряд стационарен. Однако нужно еще построить уравнение авторегрессии со стационарной ARMA-структурой, что очень важно с точки зрения получения устойчивых (к воздействию внешних шоков) коэффициентов регрессии и получения надежных прогнозов. Этой проблемой мы уже занимались (см. алгоритм действий № 13 «Тестирование стационарности авторегрессионного процесса, описываемого уравнением $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$, путем нахождения корней характеристического уравнения»), но тогда нам не удалось получить уравнение авторегрессии со стационарной ARMA-структурой.

Мы уже довольно много времени уделили построению нестационарной прогностической модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$. Поскольку при этом нам приходилось учиться, то все процедуры, необходимые для построения этой модели, вводились не сразу, а постепенно, чтобы облегчить их усвоение. Теперь перед нами стоит задача построить стационарную прогностическую модель. При этом мы будем пользоваться теми же процедурами, которые использовались при создании прогностической модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$. Чтобы не останавливаться на уже пройденном, но вместе с тем более четко структурировать полученные ранее знания, перечислим основные статистические процедуры, которые необходимо использовать при построении любой авторегрессионной (AR) или авторегрессионной со скользящей средней (ARMA) прогностической модели (алгоритм действий № 22).

Алгоритм действий № 22

Перечень действий, необходимых для построения статистической модели, представляющей собой уравнения авторегрессии (AR) или уравнения авторегрессии со скользящей средней (ARMA)

1. Построение коррелограммы в EViews с целью определения параметров p и q в модели $ARMA(p, q)$. Коррелограмма поможет нам определить лаговые переменные в уравнении авторегрессии (см. алгоритм действий № 5 «Как построить коррелограмму в EViews»).
2. Решение уравнения регрессии и проверка значимости всех его параметров. Этот вопрос можно считать самым важным из всего на-

шего перечня (см. алгоритм действий № 3 «Как решить уравнение регрессии в Excel», алгоритм действий № 4 «Оценка статистической значимости уравнения регрессии и его коэффициентов» и алгоритм действий № 6 «Как решить уравнение регрессии в EViews»).

3. Тестирование AR- или ARMA-структуры уравнения на стационарность (см. алгоритм действий № 13 «Тестирование на стационарность AR-структуры уравнения $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$ путем нахождения корней характеристического уравнения»). Тестирование исходного (а при необходимости и логарифмического) временного ряда на стационарность (см. алгоритм действий № 21 «Как провести тест на стационарность исходного уровня временного ряда»).

4. Тестирование AR- или ARMA-структуры уравнения на импульсный ответ (см. алгоритм действий № 14 «Тестирование на импульсный ответ AR-структуры нестационарного процесса, описываемого уравнением $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$ »).

5. Проверка остатков, полученных в результате решения уравнения регрессии, на наличие в них автокорреляции (см. алгоритм действий № 7 «Как выполняется LM-тест Бройша — Годфри в EViews»).

6. Проверка остатков на стационарность (см. алгоритм действий № 9 «Как в EViews проверить остатки на стационарность»).

7. Проведение анализа стандартных или студентизированных остатков на наличие выбросов (в первую очередь тех, которые влияют на текущий прогноз), теста Чоу на точность прогноза (см. алгоритм действий № 17 «Диагностика в EViews влияния студентизированных остатков на уравнение регрессии для прогностической модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$ » и алгоритм действий № 19 «Методика проведения теста Чоу на точность прогноза для прогностической модели $USDOLLAR = a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$ »).

8. Если мы получили стационарные остатки и автокорреляция в остатках не выявлена (в противном случае над уравнением регрессии придется снова поработать), то следующим нашим шагом будет оценка точности полученного уравнения регрессии (см. алгоритм действий № 8 «Как оценить точность статистической модели в EViews»).

9. Если точность прогностической модели нас устроила (см. алгоритм действий № 11 «Как в EViews построить точечный прогноз»), то в этом случае мы проверяем остатки на нормальное распределение. А затем строим интервальные прогнозы (см. алгоритм действий № 12 «Как в EViews построить интервальные прогнозы»), проверяя

уровень их надежности на соответствие нормальному распределению, на основе которого строятся доверительные интервалы.

10. Чтобы использовать прогнозы стационарной статистической модели в качестве инструмента торговой системы, необходимо на основе этой модели составить прогнозы по рекомендуемым курсам покупки и продажи валюты с односторонним ограничением риска на уровне 60—90% надежности. При этом прогнозы по рекомендуемым курсам покупки и продажи валюты могут использоваться в качестве стоп-заявок (подробнее об этом читатель узнает в главе 7).

Таким образом, создавая уравнение авторегрессии со стационарной ARMA-структурой на основе логарифмического временного ряда, необходимо выполнить все действия, которые перечислены в алгоритме действий № 22. При необходимости этот перечень каждый исследователь может расширить, если сочтет необходимым исходя из тех или иных соображений.

Однако далее при составлении стационарной статистической модели мы остановимся лишь на наиболее важных моментах этой работы. Первым делом нам необходимо найти с помощью EViews коррелограмму логарифмического временного ряда. С этой целью нужно открыть ранее созданный файл USDOLLAR1 с логарифмическим временным рядом и воспользоваться алгоритмом действий № 5 «Как построить коррелограмму в EViews». В результате у нас получилась табл. 6.3 с коррелограммой логарифмического временного ряда, полученного от исходного временного ряда «Курс доллара за период с июня 1992 г. по июнь 2010 г.». В полученной коррелограмме можно увидеть, как меняются коэффициенты автокорреляции (*Autocorrelation*, или AC) и частной автокорреляции (*Partial Correlation*, или PAC) в зависимости от изменения величины лага.

Таблица 6.3

Коррелограмма логарифмического временного ряда, полученного от исходного временного ряда «Курс доллара за период с июня 1992 г. по июнь 2010 г.»

| Date: 08/03/10 Time: 10:41 | | | | | | |
|----------------------------|---------------------|---|--------------|--------------|--------|-------|
| Sample: 1992M06 2010M06 | | | | | | |
| Included observations: 217 | | | | | | |
| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
| ***** | ***** | 1 | 0,967 | 0,967 | 205,58 | 0,000 |
| ***** | . | 2 | 0,935 | 0,005 | 398,71 | 0,000 |

Продолжение

| | | | | | | |
|-------|---|----|-------|--------|--------|-------|
| ***** | . | 3 | 0,902 | −0,022 | 579,58 | 0,000 |
| ***** | . | 4 | 0,871 | −0,001 | 748,96 | 0,000 |
| ***** | . | 5 | 0,845 | 0,058 | 908,99 | 0,000 |
| ***** | . | 6 | 0,819 | −0,001 | 1060,2 | 0,000 |
| ***** | . | 7 | 0,791 | −0,050 | 1201,8 | 0,000 |
| ***** | . | 8 | 0,767 | 0,038 | 1335,5 | 0,000 |
| ***** | . | 9 | 0,741 | −0,019 | 1461,1 | 0,000 |
| ***** | . | 10 | 0,717 | 0,002 | 1579,1 | 0,000 |
| ***** | . | 11 | 0,696 | 0,030 | 1690,8 | 0,000 |
| ***** | . | 12 | 0,677 | 0,030 | 1797,0 | 0,000 |
| ***** | . | 13 | 0,658 | −0,002 | 1898,0 | 0,000 |
| ***** | . | 14 | 0,639 | −0,033 | 1993,4 | 0,000 |
| ***** | . | 15 | 0,618 | −0,011 | 2083,3 | 0,000 |
| ***** | . | 16 | 0,600 | 0,021 | 2168,3 | 0,000 |
| **** | . | 17 | 0,580 | −0,026 | 2248,2 | 0,000 |
| **** | . | 18 | 0,560 | −0,022 | 2323,1 | 0,000 |
| **** | . | 19 | 0,539 | −0,018 | 2392,9 | 0,000 |
| **** | . | 20 | 0,521 | 0,028 | 2458,4 | 0,000 |
| **** | . | 21 | 0,504 | 0,001 | 2519,9 | 0,000 |
| **** | . | 22 | 0,486 | −0,007 | 2577,5 | 0,000 |
| **** | . | 23 | 0,469 | −0,004 | 2631,5 | 0,000 |
| *** | . | 24 | 0,452 | −0,010 | 2681,8 | 0,000 |
| *** | . | 25 | 0,435 | −0,011 | 2728,6 | 0,000 |
| *** | . | 26 | 0,418 | −0,013 | 2772,0 | 0,000 |

Окончание

| | | | | | | |
|-------|-----|----|-------|--------|--------|-------|
| . *** | . . | 27 | 0,400 | -0,007 | 2812,2 | 0,000 |
| . *** | . . | 28 | 0,385 | 0,015 | 2849,4 | 0,000 |
| . *** | . . | 29 | 0,371 | 0,012 | 2884,3 | 0,000 |
| . *** | . . | 30 | 0,358 | -0,001 | 2916,8 | 0,000 |
| . *** | . . | 31 | 0,345 | 0,008 | 2947,3 | 0,000 |
| . *** | . . | 32 | 0,334 | 0,017 | 2975,9 | 0,000 |
| . ** | . . | 33 | 0,324 | 0,012 | 3002,9 | 0,000 |
| . ** | . . | 34 | 0,315 | 0,011 | 3028,7 | 0,000 |
| . ** | . . | 35 | 0,306 | 0,005 | 3053,2 | 0,000 |
| . ** | . . | 36 | 0,297 | -0,011 | 3076,4 | 0,000 |

Судя по табл. 6.3, уровень автокорреляции (AC) между исходными уровнями временного ряда USDollar1 постоянно убывает, начиная с первого лага. В свою очередь уровень частной корреляции (PAC) резко снижается уже после первого лага, а после второго лага осциллирующим образом стремится к нулю (фактически колеблется вокруг нуля). Если мы хотим построить модель авторегрессионного процесса $AR(p)$, то для определения оптимального числа p мы должны использовать частную автокорреляционную функцию. При этом следует исходить из следующего критерия: оптимальное число p в уравнении авторегрессии должно быть меньше лага, в котором частная автокорреляционная функция начинает стремиться к нулю. Судя по коррелограмме, помещенной в табл. 6.3, коэффициент частной автокорреляции для лага в один месяц (или лага 1-го порядка) равен 0,967, а для лага в два месяца (или лага 2-го порядка) = 0,005. Причем начиная с этого лага величина коэффициента колеблется вокруг нулевого уровня. Следовательно, можно сделать вывод, что для прогнозирования курса доллара с помощью модели авторегрессии необходимо использовать модель $AR(1)$, которая примет следующий вид:

$$Y_t = c + b_1 Y_{t-1} + e_t. \quad (6.1)$$

В свою очередь при идентификации авторегрессионной модели со скользящей средней модели $ARMA(p, q)$ в качестве p выбирается лаг,

после которого начинает убывать частная автокорреляционная функция, а в качестве q выбирается лаг, после которого начинает убывать автокорреляционная функция. Исходя из табл. 6.3 можно легко прийти к выводу, что коэффициент автокорреляции начинает убывать уже с лага 2-го порядка. Аналогичный вывод можно сделать и относительно коэффициента частной автокорреляции. Поэтому для прогнозирования курса доллара с помощью модели авторегрессии со скользящим средним в остатках можно использовать модель $ARMA(1, 1)$, которая примет следующий вид:

$$Y_t = c + b Y_{t-1} + e_t - \theta e_{t-1} + u_t. \quad (6.2)$$

Однако попробуем обойтись меньшим числом параметров, а потому сначала будем использовать авторегрессионную модель $AR(1)$ согласно формуле (6.1). После того как в эту формулу вместо Y будет вставлено логарифмированное значение курса USDollar, оно приобретет следующий вид:

$$\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + e. \quad (6.3)$$

Однако для ввода в EViews следует использовать эту формулу в понятном для программы виде:

$$\log(\text{USDollar}) \log(\text{USDollar}(-1)) c. \quad (6.4)$$

При этом целесообразно использовать для создания переменных файл USDollar, т.е. файл с исходным временным рядом (с данными за период с июня 1992 г. по июнь 2010 г.), поскольку в этом случае мы сможем получить в EViews прогноз в исходном, а не в логарифмическом виде (рис. 6.4).

Таким образом, мы получили следующий вывод данных по итогам решения уравнения авторегрессии $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1))$ (табл. 6.4). При этом нетрудно увидеть, что все коэффициенты в этом уравнении регрессии оказались статистически значимыми (поскольку *Prob.*, т.е. уровень их значимости, оказался равен нулю). После замены букв найденными коэффициентами это уравнение в логарифмическом виде приобретет следующий вид:

$$\log(\text{USDollar}) = 0,103059 + 0,969092 \times \log(\text{USDollar}(-1)). \quad (6.5)$$

Однако интерпретация формулы (6.5) не столь очевидна, поскольку она относится к логарифмическому ряду, поэтому с помощью потенцирования этой формулы можно перейти от логарифмов к исходному временному ряду. Поскольку в EViews при логарифмировании

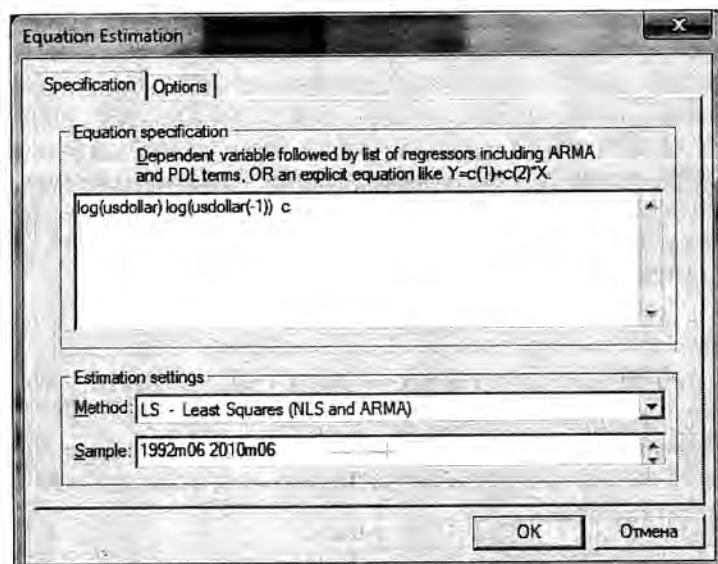


Рис. 6.4. Ввод формулы в диалоговое мини-окно EQUATION ESTIMATION/SPECIFICATION с целью получения параметров уравнения в числовом виде

исходного временного ряда используются натуральные логарифмы, в основании которых лежит $e = 2,718281\dots$, формулу (6.5) можно преобразовать следующим образом:

$$\text{EXP}(\log(\text{USDollar})) = \text{EXP}(0,103059) + (\text{EXP } 0,969092 \times \log(\text{USDollar}(-1))). \quad (6.6)$$

В частности, новое значение свободного члена (константы) в формуле (6.6) легко найти в Excel с помощью функции $\text{EXP}(0,103059) = 1,1085568$. В результате исходная линейная функция с константой 6.5, решенная относительно логарифмического временного ряда, станет степенной функцией с константой, которую можно применять относительно исходного временного ряда:

$$\text{USDollar} = 1,1085568 \times \text{USDollar}(-1)^{0,969092}. \quad (6.7)$$

При этом интерпретация формулы (6.7) будет следующей: рост на 1% курса доллара в текущем месяце в среднем способствовал повышению курса доллара в прогнозируемом месяце на 0,969% при исходном уровне курса доллара, равном 1,1086 рублей.

Таблица 6.4

Вывод данных по итогам решения уравнения авторегрессии
 $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1))$

| Dependent Variable: LOG(USDOLLAR) | | | | |
|--|-------------|-----------------------|-------------|-----------|
| Method: Least Squares | | | | |
| Date: 08/03/10 Time: 11:45 | | | | |
| Sample (adjusted): 1992M07 2010M06 | | | | |
| Included observations: 216 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| LOG(USDOLLAR(-1)) | 0,969092 | 0,003698 | 262,0245 | 0,0000 |
| C | 0,103059 | 0,010403 | 9,906766 | 0,0000 |
| R-squared | 0,996893 | Mean dependent var | | 2,533457 |
| Adjusted R-squared | 0,996878 | S.D. dependent var | | 1,238987 |
| S.E. of regression | 0,069226 | Akaike info criterion | | -2,493669 |
| Sum squared resid | 1,025534 | Schwarz criterion | | -2,462416 |
| Log likelihood | 271,3163 | F-statistic | | 68656,84 |
| Durbin - Watson stat. | 1,508627 | Prob (F-statistic) | | 0,0000 |

Проверим полученную статистическую модель на наличие автокорреляции в остатках. Для проверки будем использовать алгоритм действий № 7 «Как выполняется LM-тест Бройша — Годфри в EViews». Для выполнения теста в диалоговом мини-окне LAG SPECIFICATION (лаговая спецификация) нужно установить 1, поскольку в нашем уравнении авторегрессии есть только одна факторная переменная с одним лагом (см. формулу (6.5)). Поскольку величина задаваемого лага определяется для модели $\text{ARMA}(p, q) = \max(p, q)$, то в этом случае она приобретает следующий вид: $\text{ARMA}(1, 0) = \max(1, 0) = 1$.

По результатам проведения этого теста получилась табл. 6.5, из которой следует, что наблюдается автокорреляция в остатках. Об этом свидетельствует нулевой уровень значимости как основного критерия теста $\text{Obs} \times R\text{-squared}$ (Наблюдения $\times R^2$), так и дополнительного — $F\text{-statistic}$ (F -критерия).

Таблица 6.5

Вывод данных по итогам тестирования уравнения авторегрессии
 $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1))$
 на наличие автокорреляции в остатках

| | | | | |
|---|-------------|-----------------------|-------------|-----------|
| Breusch — Godfrey Serial Correlation LM Test: | | | | |
| F-statistic | 13,28136 | Probability | 0,000337 | |
| Obs × R-squared | 12,67790 | Probability | 0,000370 | |
| Test Equation: | | | | |
| Dependent Variable: RESID | | | | |
| Method: Least Squares | | | | |
| Date: 08/03/10 Time: 12:40 | | | | |
| Presample missing value lagged residuals set to zero. | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| LOG(USDOLLAR(-1)) | -0,000694 | 0,003602 | -0,192707 | 0,8474 |
| C | 0,001769 | 0,010128 | 0,174667 | 0,8615 |
| RESID(-1) | 0,242684 | 0,066592 | 3,644360 | 0,0003 |
| R-squared | 0,058694 | Mean dependent var | | 1,31E -16 |
| Adjusted R-squared | 0,049855 | S.D. dependent var | | 0,069065 |
| S.E. of regression | 0,067321 | Akaike info criterion | | -2,544897 |
| Sum squared resid | 0,965342 | Schwarz criterion | | -2,498018 |
| Log likelihood | 277,8489 | F-statistic | | 6,640678 |
| Durbin — Watson stat | 1,999108 | Prob (F-statistic) | | 0,001593 |

Чтобы устранить автокорреляцию в остатках, необходимо изменить спецификацию статистической модели. С этой целью попробуем использовать для прогнозирования курса доллара модель авторегрессии со скользящим средним в остатках ARMA(1, 1) (см. формулу (6.2)). Однако ввести в EViews эту формулу необходимо в следующем виде:

$$\log(\text{USDollar}) = c + \log(\text{USDollar}(-1)) \text{ MA}(1), \quad (6.8)$$

где c — константа;
 $\log(\text{USDollar}(-1))$ — логарифм от переменной с лагом в один месяц;
 $\text{MA}(1)$ — представляет собой скользящую среднюю первого порядка.

При этом следует иметь в виду одну тонкость: в опциях диалогового мини-окна EQUATION ESTIMATION (оценка уравнения) по умолчанию устанавливается параметр BACKCAST MA TERMS (рекурсивное вычисление предыдущей ошибки в МА-процессе). В этом случае значение предыдущей ошибки (отклонения прогноза от фактического курса доллара в прошлом месяце) для точечного прогноза по первому наблюдению находится рекурсивным методом (см. ввод этой опции на рис. 6.5).

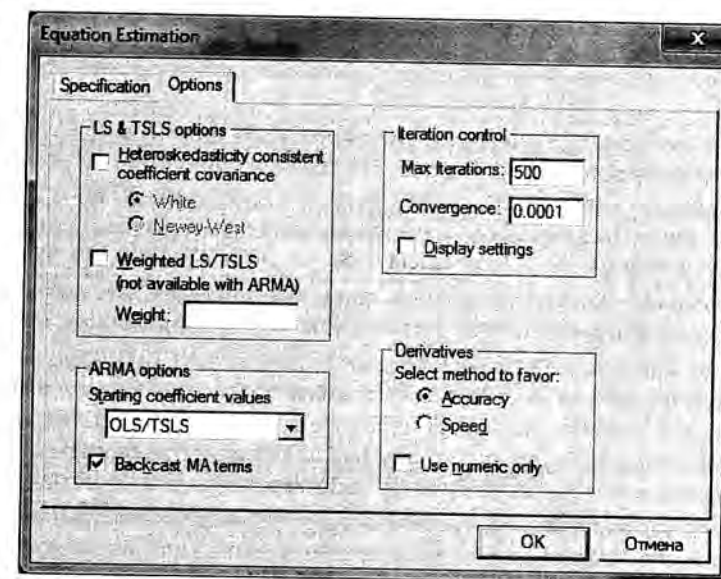


Рис. 6.5. Установление опции BACKCAST MA TERMS в диалоговом мини-окне EQUATION ESTIMATION

Если в параметре BACKCAST MA TERMS убрать «галочку», значение прошлой ошибки в точечном прогнозе для первого наблюдения приравняется нулю, поскольку оно на тот момент отсутствует. В свою очередь в случае применения в статистической модели скользящей средней второго порядка прошлые ошибки при прогнозировании не только первого, но и второго наблюдений либо приравняются нулю, либо вычисляются рекурсивным методом.

В результате ввода в диалоговое мини-окно EQUATION ESTIMATION формулы (6.7) получается следующий вывод данных (табл. 6.6). Причем все полученные коэффициенты в этом уравнении у нас полу-

чилились статистически значимыми (все выделенные жирным шрифтом значения *Prob.* меньше 0,05). Кстати, в таблице появляется информация о рекурсивном вычислении предыдущей ошибки относительно июня 1992 г. — Backcast: 1992M06. После замены буквенных обозначений вычисленными коэффициентами получилось следующее уравнение регрессии со скользящей средней:

$$\log(\text{USDollar}) = 0,105219 + 0,968257 \times \log(\text{USDollar}(-1)) + 0,253616 e_{t-1}, \quad (6.9)$$

где e_{t-1} — прошлая ошибка (отклонение фактического курса доллара от его прогноза), которая в уравнении (6.8) представлена скользящей средней MA(1).

К сожалению, в отличие от формулы (6.5) формулу (6.9) нельзя путем потенцирования привести к исходному временному ряду, что обусловлено применением в этой статистической модели скользящей средней, рассчитанной применительно к остаткам, полученным от логарифмического ряда. Поэтому интерпретация формулы (6.9) будет достаточно затруднительной, поскольку мы вынуждены ее дать относительно логарифмического, а не исходного временного ряда.

Тем не менее все-таки эту интерпретацию нужно представить, чтобы смысл уравнения (6.9) был для читателя более понятен. Во-первых, рост на одну единицу логарифмического значения курса доллара в текущем месяце в среднем способствовал повышению логарифмического значения курса доллара в прогнозируемом месяце на 0,968 ед. (при исходном уровне логарифмического значения курса доллара, равном 0,105 ед.). Во-вторых, рост на одну единицу отклонения логарифмического значения фактического курса доллара от его прогноза способствовал повышению логарифмического значения курса доллара в прогнозируемом месяце в среднем на 0,254 ед.

Таблица 6.6

Вывод итогов решения уравнения авторегрессии со скользящей средней
 $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$

Dependent Variable: LOG(USDOLLAR)
 Method: Least Squares
 Date: 08/04/10 Time: 08:22
 Sample (adjusted): 1992M07 2010M06
 Included observations: 216 after adjustments

Окончание

| | | | | |
|--|-------------|-----------------------|-------------|---------------|
| Convergence achieved after 5 iterations | | | | |
| Backcast (рекурсивное вычисление предыдущей ошибки): 1992M06 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 0,105219 | 0,012630 | 8,330752 | 0,0000 |
| LOG(USDOLLAR(-1)) | 0,968257 | 0,004491 | 215,6184 | 0,0000 |
| MA(1) | 0,253616 | 0,066335 | 3,823243 | 0,0002 |
| R-squared | 0,997077 | Mean dependent var | | 2,533457 |
| Adjusted R-squared | 0,997049 | S.D. dependent var | | 1,238987 |
| S.E. of regression | 0,067302 | Akaike info criterion | | 2,545470 |
| Sum squared resid | 0,964789 | Schwarz criterion | | -2,498591 |
| Log likelihood | 277,9107 | F-statistic | | 36326,02 |
| Durbin – Watson stat | 2,000391 | Prob(F-statistic) | | 0,00000 |
| Inverted MA Roots | -0,25 | | | |

6.3. Тестирование модели авторегрессии со скользящей средней на автокорреляцию в остатках и проверка стационарности ее ARMA-структуры

Теперь посмотрим, есть ли автокорреляция в остатках у полученной статистической модели, а потому вновь проведем тестирование с помощью *LM*-теста Бройша — Годфри. Причем при выполнении теста в диалоговом мини-окне LAG SPECIFICATION (лаговая спецификация) нужно, как и в предыдущем случае, установить 1, поскольку в нашем уравнении авторегрессии со скользящим средним ARMA(1, 1) как факторная переменная, так и скользящая средняя имеют один лаг (см. формулу (6.8)).

По результатам проведения этого теста у нас получилась табл. 6.7, данные которой уверенно свидетельствуют об отсутствии автокорреляции в остатках. Такой вывод можно сделать исходя из того, что уровень значимости как основного критерия теста *Obs × R-squared* (Наблюдения × R^2), так и дополнительного — *F-statistic* (*F*-критерия) существенно выше 0,05.

Таблица 6.7

Вывод данных по итогам тестирования уравнения авторегрессии
со скользящей средней

$$\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$$

на наличие автокорреляции в остатках

| | | | | |
|---|-------------|-----------------------|-------------|-----------|
| Breusch — Godfrey Serial Correlation LM Test: | | | | |
| F-statistic | 0,038296 | Probability | 0,845037 | |
| Obs*R-squared | 0,038875 | Probability | 0,843696 | |
| Test Equation: | | | | |
| Dependent Variable: RESID | | | | |
| Method: Least Squares | | | | |
| Date: 08/04/10 Time: 11:45 | | | | |
| Presample missing value lagged residuals set to zero. | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -0,000204 | 0,012702 | -0,016068 | 0,9872 |
| LOG(USDOLLAR(-1)) | 7,69E-05 | 0,004518 | 0,017020 | 0,9864 |
| MA(1) | 0,049512 | 0,261595 | 0,189270 | 0,8501 |
| RESID(-1) | -0,053008 | 0,270873 | -0,195694 | 0,8450 |
| R-squared | 0,000180 | Mean dependent var | | -5,31E-05 |
| Adjusted R-squared | -0,013968 | S.D. dependent var | | 0,066988 |
| S.E. of regression | 0,067454 | Akaike info criterion | | -2,536391 |
| Sum squared resid | 0,964614 | Schwarz criterion | | -2,473886 |
| Log likelihood | 277,9302 | F-statistic | | 0,012721 |
| Durbin — Watson stat. | 1,989783 | Prob(F-statistic) | | 0,998033 |

Теперь протестируем ARMA-структуру этого уравнения на стационарность, воспользовавшись при этом алгоритмом действий № 13. В результате у нас получится табл. 6.8, свидетельствующая, что ARMA-структура этой статистической модели получилась стационарной, поскольку все обратные корни в этом уравнении лежат внутри единичного круга. Этот вывод можно найти в нижней части этой таблицы.

Таблица 6.8

Оценка авторегрессионного процесса на стационарность для уравнения
 $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$

| Inverse Roots of AR/MA Polynomial(s) | | |
|--|----------|-------|
| Specification: LOG(USDOLLAR) C AR(1) MA(1) | | |
| Date: 08/04/10 Time: 12:46 | | |
| Sample: 1992M06 2010M06 | | |
| Included observations: 216 | | |
| AR Root(s) | Modulus | Cycle |
| 0,968257 | 0,968257 | |
| No root lies outside the unit circle (ни один обратный корень не лежит вне единичного круга) | | |
| ARMA model is stationary (модель ARMA стационарная) | | |
| MA Root(s) | Modulus | Cycle |
| -0,253617 | 0,253617 | |
| No root lies outside the unit circle (ни один обратный корень не лежит вне единичного круга) | | |
| ARMA model is invertible (модель ARMA обратимая) | | |

Далее посмотрим, как стационарная ARMA-структура уравнения $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ влияет на надежность полученных с ее помощью прогнозов, поэтому, воспользовавшись алгоритмом действий № 14, протестируем эту статистическую модель на импульсный ответ.

При этом в опции IMPULSE (импульс) мы выбрали вариант по умолчанию — ONE STANDARD DEVIATION (одно стандартное отклонение), т.е. поступили так же, как и в главе 4 при анализе импульсного ответа для уравнения $\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2)$.

В результате получим табл. 6.9, в которой содержится информация, характеризующая специфику импульсного и накопленного импульсного ответа этой ARMA-модели. Поскольку мы выбрали величину импульса в размере одного стандартного отклонения, то EViews в этом случае выдает нам информацию об уровне инновационной неопределенности, полученной после оценки размера стандартной

ошибки импульсного ответа. Важным свойством стационарных моделей является то обстоятельство, что у них как уровень инновационной неопределенности, так и величина стандартного отклонения импульсного ответа — по мере увеличения количества тестируемых периодов — стремятся к нулю. Судя по табл. 6.9, уровень инновационной неопределенности и величина ответа на импульс асимптотически у стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ действительно стремятся к нулю. При этом в нижней части раздела *Response* и крайнего правого раздела *Std. Err.* дается асимптотическая оценка того, что эти параметры равны нулю.

Таблица 6.9

Импульсный ответ ARMA-структуры и оценка инновационной неопределенности для уравнения $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$

| ARMA Impulse Response | | | | |
|--|----------|-----------|-------------|-----------|
| Specification: LOG(USDOLLAR) AR(1) MA(1) C | | | | |
| Date: 09/15/10 Time: 11:05 | | | | |
| Sample: 1992M06 2010M06 | | | | |
| Included observations: 216 | | | | |
| Period | Response | Std. Err. | Accumulated | Std. Err. |
| 1 | 0,067302 | (0,00324) | 0,067302 | (0,00324) |
| 2 | 0,082234 | (0,00596) | 0,149536 | (0,00846) |
| 3 | 0,079624 | (0,00579) | 0,229160 | (0,01410) |
| 4 | 0,077096 | (0,00565) | 0,306256 | (0,01967) |
| 5 | 0,074649 | (0,00553) | 0,380906 | (0,02512) |
| ... | ... | ... | ... | ... |
| 197 | 0,000152 | (0,00014) | 2,653304 | (0,41080) |
| 198 | 0,000148 | (0,00013) | 2,653451 | (0,41093) |
| 199 | 0,000143 | (0,00013) | 2,653594 | (0,41105) |
| 200 | 0,000138 | (0,00013) | 2,653733 | (0,41117) |
| 201 | 0,000134 | (0,00012) | 2,653867 | (0,41128) |
| LR | 0,000000 | (0,00000) | 2,657955 | (0,41530) |

Источник: расчеты автора.

Кроме того, в табл. 6.9 хорошо видно, что по мере увеличения количества исследуемых периодов величина стандартного отклонения у накопленного импульсного ответа (см. раздел таблицы *Accumulated*) и уровень накопленной инновационной неопределенности (см. раздел в центре таблицы — *Std. Err.*) стремятся к определенному асимптотическому пределу, значение которого приводится внизу. Следует заметить, что у статистической модели с нестационарной ARMA-структурой указанный предел отсутствует.

В целях экономии места в табл. 6.9 приведена лишь часть данных. Однако эта информация в наглядном виде представлена на рис. 6.6, который полностью подтверждает наши выводы. Кроме того, на рисунке точечными линиями с двух сторон обозначены доверительные интервалы, показывающие возможную погрешность в оценке величины импульсного и накопленного импульсного ответов.

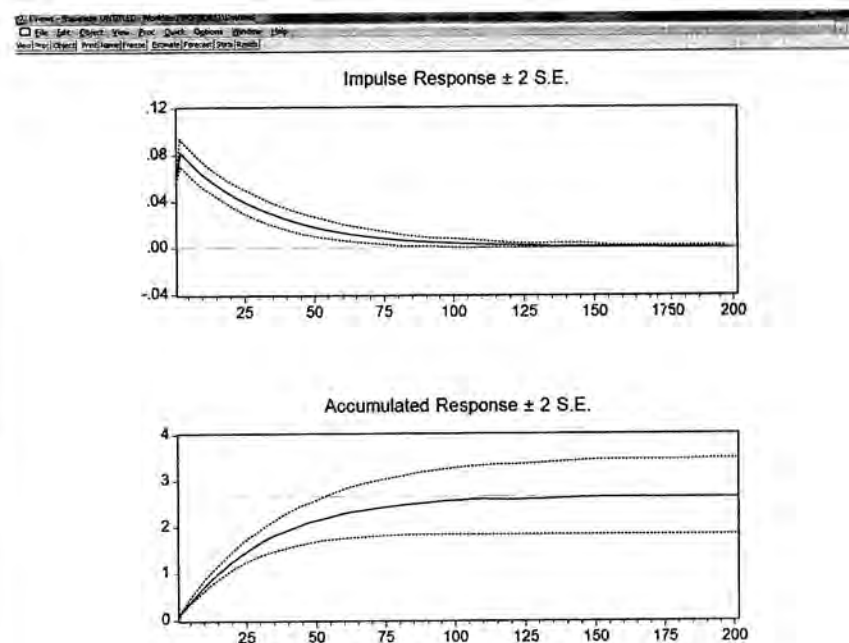


Рис. 6.6. Графическое изображение импульсного и накопленного импульсного ответов, стремящихся к асимптотическому пределу в рамках доверительных интервалов

Алгоритм действий № 23

Как сравнить коррелограмму остатков стационарной модели с ее теоретическим аналогом

Насколько хорошо построена стационарная модель, можно судить по оценке соответствия фактических значений коррелограммы остатков их теоретическим значениям. С этой целью воспользуемся опциями VIEW/ARMA STRUCTURE (посмотреть/структура модели ARMA). В результате на экране появится диалоговое мини-окно ARMA DIAGNOSTIC VIEWS (посмотреть диагностику модели ARMA), в которой нужно выбрать параметр CORRELOGRAM (рис. 6.7). Причем если нам нужна коррелограмма в табличной форме, то в опции DISPLAY мы выбираем надпись TABLE, а если в виде графика, то следует выбрать надпись GRAPH. При этом по умолчанию составляется коррелограмма для 24 лагов, но при необходимости пользователь может выбрать и иное количество лагов.

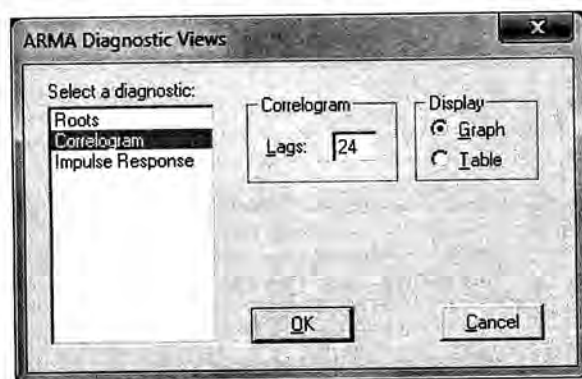


Рис. 6.7. Установление опции CORRELOGRAM и ее вида в диалоговом мини-окне ARMA DIAGNOSTIC VIEWS

В таблице 6.10 представлены как фактические, так и теоретические значения коррелограммы остатков, полученных после решения уравнения регрессии $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$. В таблице представлены значения автокорреляционной и частной автокорреляционной функций (т.е. автокорреляция между двумя лагами без учета влияния других промежуточных временных лагов). Как вычисляются коэффициенты автокорреляции и частной автокорреляции, можно уточнить в формулах (3.7—3.9).

Таблица 6.10

Фактическая и теоретическая коррелограмма остатков для модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$

| Actual and ARMA Model Correlogram | | | | | | | |
|--|-------------------------------------|--------------------------------------|------------|----|---|--------------------------------------|------------|
| Specification: LOG(USDOLLAR) AR(1) MA(1) C | | | | | | | |
| Date: 09/22/10 Time: 11:37 | | | | | | | |
| Sample: 1998M10 2010M07 | | | | | | | |
| Included observations: 142 | | | | | | | |
| | Autocorrelation (автокорреляция) | | | | Partial Autocorrelation (частная автокорреляция) | | |
| | Actual (фактические значения) | Model (теоретические значения) | Difference | | Actual (фактические значения) | Model (теоретические значения) | Difference |
| 0 | 1,000 | 1,000 | 0,000 | 0 | 1,000 | 1,000 | 0,000 |
| 1 | 0,877 | 0,934 | -0,057 | 1 | 0,877 | 0,934 | -0,057 |
| 2 | 0,753 | 0,819 | -0,066 | 2 | -0,069 | -0,416 | 0,347 |
| 3 | 0,662 | 0,718 | -0,056 | 3 | 0,073 | 0,224 | -0,151 |
| 4 | 0,592 | 0,630 | -0,037 | 4 | 0,037 | -0,127 | 0,164 |
| 5 | 0,524 | 0,552 | -0,028 | 5 | -0,025 | 0,073 | -0,098 |
| 6 | 0,466 | 0,484 | -0,019 | 6 | 0,017 | -0,042 | 0,059 |
| 7 | 0,407 | 0,425 | -0,018 | 7 | -0,034 | 0,024 | -0,058 |
| 8 | 0,355 | 0,372 | -0,017 | 8 | 0,005 | -0,014 | 0,019 |
| 9 | 0,308 | 0,326 | -0,018 | 9 | -0,013 | 0,008 | -0,021 |
| 10 | 0,264 | 0,286 | -0,022 | 10 | -0,018 | -0,005 | -0,013 |
| 11 | 0,218 | 0,251 | -0,033 | 11 | -0,029 | 0,003 | -0,031 |
| 12 | 0,169 | 0,220 | -0,051 | 12 | -0,049 | -0,002 | -0,047 |
| 13 | 0,128 | 0,193 | -0,065 | 13 | -0,002 | 0,001 | -0,003 |

Окончание

| | Autocorrelation (автокорреляция) | | | | Partial Autocorrelation (частная автокорреляция) | | |
|----|---|--|-----------------|----|---|--|-----------------|
| | Actual (фак- тические значе- ния) | Model (теоре- тические значе- ния) | Diffe- rence | | Actual (фак- тические значе- ния) | Model (теоре- тические значе- ния) | Diffe- rence |
| 14 | 0,088 | 0,169 | -0,081 | 14 | -0,033 | -0,001 | -0,033 |
| 15 | 0,056 | 0,148 | -0,092 | 15 | 0,000 | 0,000 | -0,000 |
| 16 | 0,040 | 0,130 | -0,090 | 16 | 0,041 | -0,000 | 0,042 |
| 17 | 0,023 | 0,114 | -0,091 | 17 | -0,025 | 0,000 | -0,025 |
| 18 | 0,000 | 0,100 | -0,100 | 18 | -0,026 | -0,000 | -0,026 |
| 19 | -0,021 | 0,088 | -0,108 | 19 | -0,010 | 0,000 | -0,010 |
| 20 | -0,035 | 0,077 | -0,112 | 20 | -0,001 | -0,000 | -0,001 |
| 21 | -0,051 | 0,067 | -0,118 | 21 | -0,021 | 0,000 | -0,021 |
| 22 | -0,066 | 0,059 | -0,126 | 22 | -0,017 | -0,000 | -0,017 |
| 23 | -0,080 | 0,052 | -0,132 | 23 | -0,009 | 0,000 | -0,009 |
| 24 | -0,092 | 0,045 | -0,137 | 24 | -0,016 | -0,000 | -0,016 |

Важной особенностью коррелограммы остатков, полученных по стационарным моделям, является то, что с увеличением величины лага значения автокорреляционной функции медленно, но с завидным постоянством убывают к нулю, в то время как частная автокорреляционная функция начинает колебаться около нуля уже со второго лага, при этом то немного вырастая, то убывая.

Стационарная модель считается хорошо построенной, если фактические значения коррелограммы окажутся близкими к ее теоретическим значениям. Как видим, в этом случае у нас это получилось.

Близость между фактическими и теоретическими значениями коррелограммы наглядно представлена на рис. 6.8. При этом теоретические значения коррелограммы с целью большей наглядности обозначены на рисунке горизонтальной линией, а фактические значения вертикальными линиями.

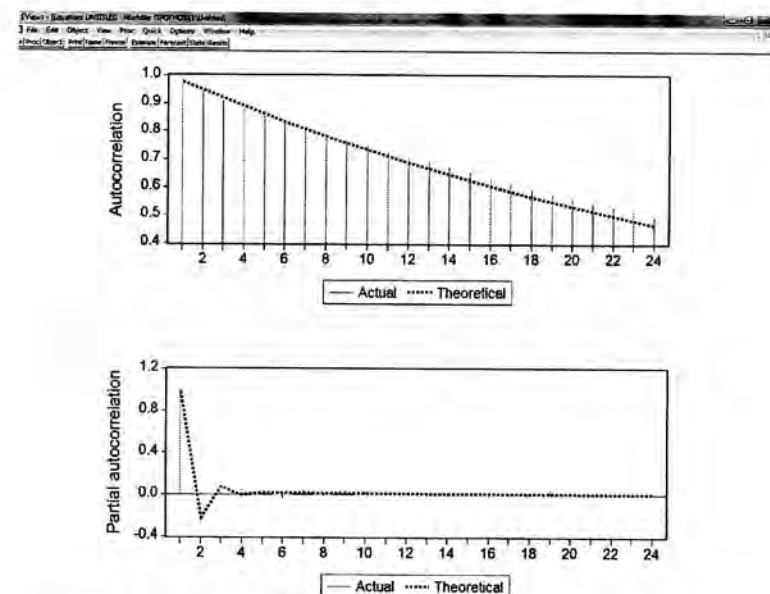


Рис. 6.8. Сравнение фактических и теоретических значений коррелограммы модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$

6.4. Оценка стабильности стационарной модели авторегрессии со скользящей средней

На основе данных за период с июня 1992 г. по июнь 2010 г. необходимо с помощью модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ составить точечный и интервальный прогнозы по курсу доллара на июль 2010 г. Однако прежде проведем анализ стандартных и студентизированных остатков, полученных в этой модели, на предмет наличия выбросов, причем особое внимание будем обращать на наличие выбросов в последних наблюдениях, которые в большей степени могут повлиять на точность текущего прогнозирования. Для расчета стандартных и студентизированных остатков следует воспользоваться алгоритмами действий № 16 и 17.

В результате у нас получилась табл. 6.11, а также диаграмма студентизированных остатков на рис. 6.9. Если эту таблицу сравнить с табл. 5.9, то выяснится следующее важное обстоятельство. В статистической модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ из 11 выбросов, выявленных с помощью стандартных и студентизированных остатков, шесть

Таблица 6.11

Выбросы в стандартных и студентизированных остатках, полученных после решения уравнения $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, на основе данных за период с июня 1992 г. по июнь 2010 г.

| Наблюдения | Остатки | Стандартные остатки | Студентизиро- ванные остатки |
|------------|----------|------------------------|---------------------------------|
| 1992M10 | 0,289192 | 4.29693 | 4.607597 |
| 1992M11 | -0,0917 | -1,36253 | -1,45387 |
| 1992M12 | -0,18301 | -2.71919 | -2.84911 |
| 1993M01 | 0,235305 | 3.496256 | 3.701469 |
| 1993M02 | -0,14657 | -2.17784 | -2.32608 |
| 1993M07 | -0,15765 | -2.34246 | -2.39946 |
| 1993M08 | -0,07013 | -1,042 | -1,0638 |
| 1993M09 | 0,110355 | 1,639699 | 1,666064 |
| 1993M10 | -0,13996 | -2.0796 | -2.13392 |
| 1994M09 | 0,112262 | 1,668034 | 1,68639 |
| 1998M06 | -0,03342 | -0,49654 | -0,49729 |
| 1998M07 | -0,03241 | -0,48149 | -0,4822 |
| 1998M08 | 0,197945 | 2.941146 | 3.005852 |
| 1998M09 | 0,619323 | 9.202149 | 12.30732 |
| 1998M10 | -0,17755 | -2.63811 | -3.26299 |
| 1998M11 | 0,138309 | 2.05505 | 2.192255 |
| 1998M12 | 0,095272 | 1,415589 | 1,457417 |
| 1999M01 | 0,056961 | 0,846349 | 0,848719 |
| 2008M08 | 0,042431 | 0,630457 | 0,631409 |
| 2008M09 | 0,01253 | 0,186176 | 0,186467 |
| 2008M10 | 0,044173 | 0,65634 | 0,657475 |
| 2008M11 | 0,026922 | 0,400018 | 0,400846 |
| 2008M12 | 0,055571 | 0,825696 | 0,827868 |
| 2009M01 | 0,174784 | 2.59701 | 2.64555 |
| 2009M02 | -0,02772 | -0,41182 | -0,41789 |
| 2009M03 | -0,03366 | -0,50009 | -0,50243 |
| 2009M04 | -0,00746 | -0,11086 | -0,11105 |
| 2009M05 | -0,06265 | -0,93083 | -0,93404 |

Окончание

| Наблюдения | Остатки | Стандартные остатки | Студентизиро- ванные остатки |
|------------|----------|------------------------|---------------------------------|
| 2009M06 | 0,029488 | 0,438144 | 0,439729 |
| 2009M07 | 0,011357 | 0,168747 | 0,169116 |
| 2009M08 | -0,00423 | -0,06287 | -0,06295 |
| 2009M09 | -0,04247 | -0,63096 | -0,63235 |
| 2009M10 | -0,02168 | -0,3221 | -0,32282 |
| 2009M11 | 0,033351 | 0,495542 | 0,496372 |
| 2009M12 | 0,008288 | 0,123146 | 0,123358 |
| 2010M01 | 0,007063 | 0,104945 | 0,105072 |
| 2010M02 | -0,01459 | -0,21674 | -0,21701 |

выбросов приходится на период 1992–1993 гг., т.е. имели место в период самых первых наблюдений. В свою очередь остальные четыре выброса произошли с августа по ноябрь 1998 г., в период после дефолта. В то же время в период глобального финансового кризиса в остатках этой модели обнаруживается лишь один выброс, относящийся к январю 2009 г.

Для справки заметим, что в остатках, полученных после решения уравнения регрессии $\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2)$, имели место девять выбросов. Причем до августа 1998 г. в этой модели выбросы не выявлены, но зато было пять выбросов после августовского дефолта – с августа по декабрь 1998 г. и четыре выброса в период глобального финансового кризиса – в январе, феврале, марте и мае 2009 г. Таким образом, в последние годы стационарная модель $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ демонстрирует гораздо большую стабильность, чем нестационарная модель $\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2)$.

На рисунке 6.9 приведена диаграмма, из которой хорошо видно, что, за исключением одного уже упомянутого случая, выбросы в стационарной модели после 1998 г. уже не наблюдались.

В главе 4 уже говорилось, что тест Чоу на точность прогноза хорошо подходит для анализа стабильности статистической модели относительно последнего наблюдения. Поэтому мы воспользовались этим тестом, чтобы еще раз убедиться в стабильности модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ относительно июня 2010 г. (см. алгоритм действий № 19). В результате у нас получилась табл. 6.12. Судя по уровню значимости F -критерия (F -statistic) и логарифма

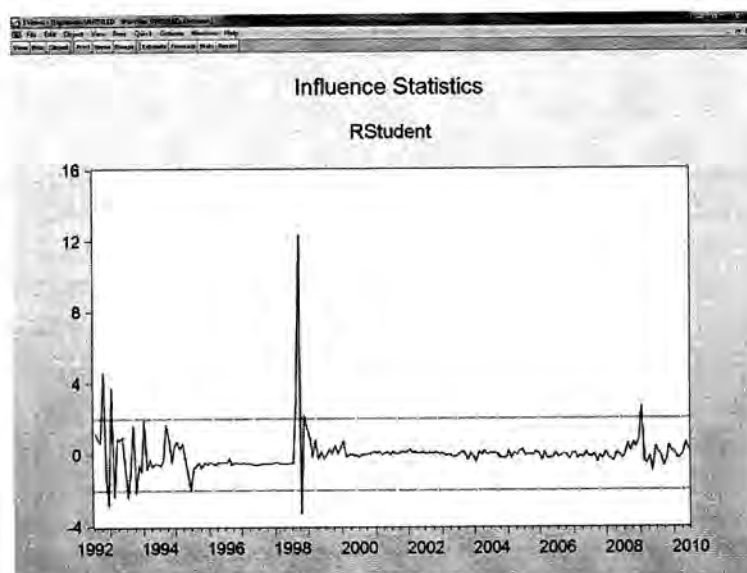


Рис. 6.9. Выбросы в студентизированных остатках, полученных после решения уравнения $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, на основе данных за период с июня 1992 г. по июнь 2010 г.

Источник: расчеты автора

рифма правдоподобия (*Log likelihood ratio*), можно сделать вывод, что нулевая гипотеза о структурной стабильности статистической модели относительно последнего наблюдения подтверждается с большим уровнем надежности. Отметим еще раз, что нулевая гипотеза может быть отвергнута, если уровень значимости (*Probability*) *F*-критерия и логарифма правдоподобия будет ниже 0,05.

Таблица 6.12

Вывод данных по результатам теста Чоу на точность прогноза относительно июня 2010 г. для модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$

| | | | |
|--|----------|-------------|----------|
| Chow Forecast Test: Forecast from 2010M06 to 2010M06 | | | |
| F-statistic | 0,052525 | Probability | 0,818948 |
| Log likelihood ratio | 0,053523 | Probability | 0,817042 |
| Test Equation: | | | |
| Dependent Variable: LOG(USDOLLAR) | | | |
| Method: Least Squares | | | |

Окончание

Date: 09/04/10 Time: 09:08
 Sample: 1992M07 2010M05
 Included observations: 215
 Convergence achieved after 5 iterations
 Backcast: 1992M06

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------------------|-------------|-----------------------|-------------|---------------|
| LOG(USDOLLAR(-1)) | 0,968209 | 0,004504 | 214,9855 | 0,0000 |
| C | 0,105251 | 0,012653 | 8,318197 | 0,0000 |
| MA(1) | 0,252999 | 0,066554 | 3,801393 | 0,0002 |
| R-squared | 0,997070 | Mean dependent var | | 2,529239 |
| Adjusted R-squared | 0,997043 | S.D. dependent var | | 1,240323 |
| S.E. of regression | 0,067452 | Akaike info criterion | | -2,540948 |
| Sum squared resid | 0,964550 | Schwarz criterion | | -2,493916 |
| Log likelihood | 276,1519 | F-statistic | | 36073,67 |
| Durbin — Watson stat | 1,998846 | Prob (F-statistic) | | 0,000000 |
| Inverted MA Roots | -0,25 | | | |

6.5. Оценка точности стационарной модели ARMA

Поскольку мы уже убедились в относительной стабильности стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, то теперь можем сделать точечный прогноз на июль 2010 г. на основе данных за период с июня 1992 г. по июнь 2010 г. (см. алгоритм действий № 11 «Как в EViews построить точечный прогноз»). При этом следует иметь в виду, что составление прогнозов по логарифмическому временному ряду имеет некоторую специфику. По умолчанию диалоговое мини-окно FORECAST (прогноз) при работе с логарифмическим рядом в опции SERIES TO FORECAST (ряд для прогноза) указывает на файл с данными для исходного временного ряда USDOLLAR (рис. 6.10). В этом случае прогнозы будут даваться не в логарифмическом, а в исходном виде, т.е. в том виде, который обычно необходим для прогноза по валютному рынку. Однако при необходимости пользователь может

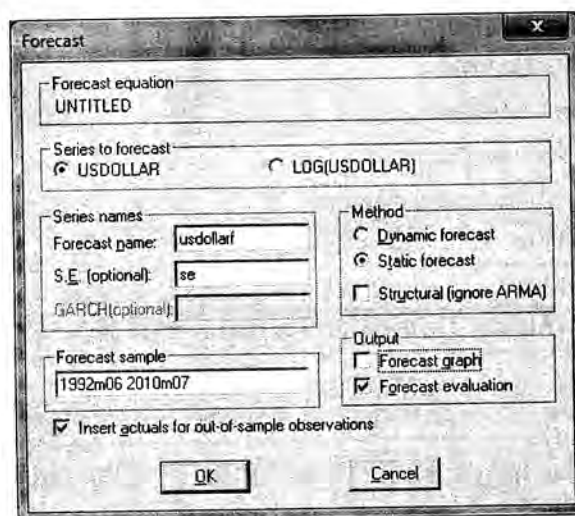


Рис. 6.10. Диалоговое мини-окно FORECAST по умолчанию указывает на файл USDOLLAR с данными для исходного временного ряда

самостоятельно поставить «галочку» у файла LOG(USDOLLAR) и получить прогнозы в логарифмическом виде.

В результате мы получили табл. 6.13, в которой наряду с оценкой точности стационарной прогностической модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ поместили и оценку точности нестационарной статистической модели $\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + a \times \text{USDOLLAR}(-2)$ за период с июня 1992 г. по июнь 2010 г.

Таблица 6.13

Оценка точности стационарной и нестационарной статистических моделей, построенных на основе данных за период с июня 1992 г. по июнь 2010 г.

| Параметр | Точность стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ | Точность нестационарной модели $\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2)$ | Разница |
|---|--|--|---------|
| Root Mean Squared Error (квадратный корень средней ошибки предсказания) | 0,8142 | 0,8172 | -0,0030 |

Окончание

| Параметр | Точность стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ | Точность нестационарной модели $\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + b \times \text{USDOLLAR}(-2)$ | Разница |
|--|--|--|---------|
| Mean Absolute Error (средняя ошибка по модулю) | 0,3679 | 0,3434 | 0,0245 |
| Mean Absolute Percentage Error (средняя ошибка по модулю, %) | 3,2386 | 2,7806 | 0,4580 |
| Theil Inequality Coefficient (коэффициент неравенства Тейла) | 0,0177 | 0,0177 | 0,0000 |
| Bias Proportion (доля систематической ошибки) | 0,0046 | 0,0041 | 0,0005 |
| Variance Proportion (доля вариации) | 0,0032 | 0,0086 | -0,0055 |
| Covariance Proportion (доля ковариации, т.е. несистематической ошибки) | 0,9922 | 0,9873 | 0,0050 |

Источник: расчеты автора.

О содержательной интерпретации параметров, представленных в табл. 6.13, мы уже говорили (см. алгоритм действий № 8 «Как оценить точность статистической модели в EViews»).

Нетрудно заметить, что хотя в целом по уровню точности обе модели имеют довольно близкие оценки, тем не менее стационарная модель по ряду показателей уступает нестационарной модели. Так, довольно существенным кажется отклонение по величине средней ошибки по модулю (Mean Absolute Error) и по величине средней ошибки по модулю

в процентах (*Mean Absolute Percentage Error*). Например, в целом за весь период средняя ошибка по модулю для стационарной модели оказалась на 2,45 процентного пункта выше, чем у нестационарной, а по величине средней ошибки по модулю в процентах — почти на 0,46 пункта.

Однако если посмотреть, как изменялась точность обеих статистических моделей в различные периоды времени, то начиная с 1999 г. стационарная модель дает более точные прогнозы. В частности, в период с января 1999 г. по июнь 2010 г. средняя точность стационарной модели оказалась выше точности нестационарной модели на 0,2 коп. по модулю (см. цифры, выделенные жирным шрифтом в табл. 6.14). А с января 2009 г. по июнь 2010 г. эта разница составила уже 8,7 коп.

Таблица 6.14

Средняя точность прогноза (в руб. по модулю) стационарной и нестационарной статистических моделей за различные периоды времени

| Период | Средняя точность стационарной модели (в руб. по модулю) | Средняя точность нестационарной модели (в руб. по модулю) | Разница |
|------------------------|---|---|----------------|
| Весь период | 0,3695 | 0,3434 | 0,0262 |
| 1999 г. — июнь 2010 г. | 0,3775 | 0,3796 | -0,0021 |
| 2005 г. — июнь 2010 г. | 0,5045 | 0,5311 | -0,0266 |
| 2006 г. — июнь 2010 г. | 0,5693 | 1,9443 | -1,375 |
| 2007 г. — июнь 2010 г. | 0,6613 | 2,4127 | -1,7515 |
| 2008 г. — июнь 2010 г. | 0,8125 | 0,8484 | -0,036 |
| 2009 г. — июнь 2010 г. | 0,9752 | 1,0622 | -0,087 |

Источник: расчеты автора.

Естественно, что и по величине средней точности прогноза (в % по модулю) стационарная модель с января 1999 г. также дает более точные прогнозы. В частности, в период с января 1999 г. по июнь 2010 г. средняя точность стационарной модели (в % по модулю) оказалась выше точности нестационарной модели на 0,08 процентного пункта (см. цифры, выделенные жирным шрифтом в табл. 6.15). В свою очередь с января 2009 г. по июнь 2010 г. эта разница составила уже более 0,26 процентного пункта. С учетом этого можно сделать вывод, что точность стационарной статистической модели за последние 11,5 лет оказалась выше, чем у нестационарной модели.

Таблица 6.15

Средняя точность прогноза (в % по модулю) стационарной и нестационарной статистических моделей за различные периоды времени

| Период | Средняя точность стационарной модели (в % по модулю) | Средняя точность нестационарной модели (в % по модулю) | Разница |
|------------------------|--|--|----------------|
| Весь период | 3,2386 | 2,7806 | 0,4580 |
| 1999 г. — июнь 2010 г. | 1,3133 | 1,3166 | -0,0033 |
| 2005 г. — июнь 2010 г. | 1,7177 | 1,7983 | -0,0807 |
| 2006 г. — июнь 2010 г. | 1,9321 | 2,0093 | -0,0773 |
| 2007 г. — июнь 2010 г. | 2,2260 | 2,2983 | -0,0723 |
| 2008 г. — июнь 2010 г. | 2,6742 | 2,7717 | -0,0975 |
| 2009 г. — июнь 2010 г. | 3,0167 | 3,2789 | -0,2622 |

Источник: расчеты автора.

Воспользовавшись диалоговым мини-окном FORECAST, мы получили не только оценку точности прогноза для стационарной статистической модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, но и файл с точечными прогнозами USDOLLARF за период с июля 1992 г. по июль 2010 г. Открыв этот файл, мы выяснили, что точечный прогноз на июль 2010 г. оказался равен 31 руб. 19 коп., однако фактический курс доллара в июле 2010 г. был равен 30 руб. 19 коп. Следовательно, разница составила 1 руб. Посмотрим, попал ли фактический курс доллара в диапазон интервального прогноза?

Однако, прежде чем это сделать, проверим остатки, полученные по модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, на нормальное распределение и на стационарность (см. алгоритм действий № 9).

В первом случае откроем файл RESID и выберем опции VIEW (смотреть)/DESCRIPTIVE STATISTICS (описательная статистика)/STATS TABLE (таблица со статистикой). При этом следует иметь в виду, что проверку на нормальное распределение остатков целесообразно проводить относительно логарифмических остатков, поскольку наша статистическая модель построена на логарифмическом временном ряде. Логарифмические остатки нетрудно найти, если при составлении прогнозов в диалоговом мини-окне FORECAST (прогноз) поставим «галочку» у файла LOG(USDOLLAR) (см. рис. 6.6). В результате мы получили табл. 6.16.

Таблица 6.16

Описательная статистика для остатков, полученных по модели

$$\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$$

| | RESID |
|---------------|-----------|
| Mean | -0,000204 |
| Median | -0,002729 |
| Maximum | 0,619323 |
| Minimum | -0,183007 |
| Std. Dev. | 0,066870 |
| Skewness | 4,210730 |
| Kurtosis | 38,42839 |
| Jarque — Bera | 11990,08 |
| Probability | 0,000000 |
| Sum | -0,044252 |
| Sum Sq. Dev. | 0,965854 |
| Observations | 217 |

Судя по тому, что коэффициент асимметрии (*Skewness*) в табл. 6.16 положителен, можно прийти к выводу, что в распределении остатков, полученных по стационарной модели, наблюдается положительная асимметрия. Отсюда можно сделать вывод, что в динамике курса доллара к рублю чаще наблюдались резкие (вполне очевидно, что незначительные плавные колебания курса легко поддаются прогнозированию) подъемы, чем аналогичные падения. В свою очередь величина коэффициента эксцесса (*Kurtosis*) существенно выше 3, что свидетельствует об «островершинном» распределении остатков. По сути, это означает, что в этом распределении имеется ярко выраженное ядро плотности распределения, внутри которого диапазон колебаний величины остатков незначителен, и рассеянное «гало», где разброс колебаний величины остатков весьма значителен. Поскольку величина тестовой статистики Жарка — Бера составила 11990,08, а уровень ее значимости (*Probability*) оказался равен нулю, то, следовательно, мы вынуждены отвергнуть гипотезу о нормальном распределении остатков. Поскольку, как мы уже говорили нашим читателям, при уровне значимости критерия Жарка — Бера (*Probability*) меньше 0,05 нулевая гипотеза о нормальном распределении отклоняется.

В EViews есть возможность посмотреть в графическом виде оценку ядра плотности распределения с помощью опций **DISTRIBUTION/KERNEL DENSITY GRAPHS...** (распределение/графики ядра плотности распределения). В появившемся мини-окне **KERNEL DENSITY** (ядро плотности распределения) установим опцию **EPANECHNIKOV**. В результате получим рис. 6.11, наглядно показывающий «островершинный» характер распределения остатков с правосторонней асимметрией.

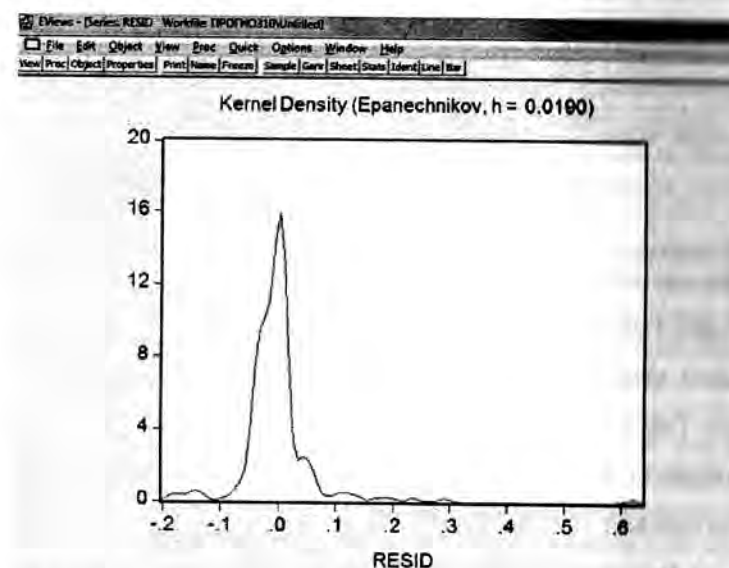


Рис. 6.11. «Островершинный» характер ядра плотности распределения остатков (с правосторонней асимметрией), полученных по модели $\log(\text{USDollar}) = c + a \times \log[\text{USDollar}(-1)] + \text{MA}(1)$

Для проверки остатков на стационарность воспользуемся расширенным тестом Дикки — Фуллера. Следует заметить, что проверку остатков на стационарность также целесообразно проводить относительно логарифмических остатков. После проведения тестирования мы убедились, что получили стационарные остатки (табл. 6.17). Поскольку статистика теста Дикки — Фуллера в этом случае составила -15,61466, а ее значимость (*Probability*) равна 0,0000, то нулевая гипотеза о том, что $D(\text{RESID})$ имеет единичный корень, отвергается. Следовательно, мы можем принять альтернативную гипотезу о стационарности полученных остатков.

Таблица 6.17

Итоги решения расширенного теста Дикки — Фуллера для остатков,
полученных по статистической модели
 $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$

| | | | | |
|---|-------------|-----------------------|-------------|-----------|
| Null Hypothesis: D(RESID) has a unit root | | | | |
| Exogenous: None | | | | |
| Lag Length: 2 (Automatic based on SIC, MAXLAG=14) | | | | |
| | | | t-Statistic | Prob.* |
| Augmented Dickey — Fuller test statistic | | | -15,61466 | 0,0000 |
| Test critical values: | 1% level | | -2,575864 | |
| | 5% level | | -1,942324 | |
| | 10% level | | -1,615707 | |
| *MacKinnon (1996) one-sided p-values. | | | | |
| Augmented Dickey — Fuller Test Equation | | | | |
| Dependent Variable: D(RESID,2) | | | | |
| Method: Least Squares | | | | |
| Date: 08/05/10 Time: 11:35 | | | | |
| Sample (adjusted): 1992M11 2010M07 | | | | |
| Included observations: 213 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| D(RESID(-1)) | -2,736597 | 0,175258 | -15,61466 | 0,0000 |
| D(RESID(-1),2) | 0,881600 | 0,128023 | 6,886244 | 0,0000 |
| D(RESID(-2),2) | 0,255218 | 0,065183 | 3,915388 | 0,0001 |
| R-squared | 0,824863 | Mean dependent var | | -0,001376 |
| Adjusted R-squared | 0,823195 | S.D. dependent var | | 0,165325 |
| S.E. of regression | 0,069516 | Akaike info criterion | | -2,480528 |
| Sum squared resid | 1,014827 | Schwarz criterion | | -2,433186 |
| Log likelihood | 267,1762 | Durbin — Watson stat. | | 1,957093 |

Несмотря на то что остатки, полученные по стационарной модели, нельзя считать нормально распределенными, мы тем не менее уже знаем, что при больших выборках можно строить интервальные прогнозы исходя из их нормального распределения. Поэтому нашим следующим шагом будет расчет интервальных прогнозов не только на июль 2010 г. (курс доллара по этому месяцу не включен в базу данных), но и для всех наблюдений, на основе которых составлена статистическая модель $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$. Это поможет нам проверить соответствие составленных интервальных прогнозов нормальному распределению, поскольку уровень надежности для интервальных прогнозов рассчитывался исходя из предположения о нормальном распределении остатков. Полупно заметим, что интервальные прогнозы будут построены начиная с июля 1992 г., поскольку первое наблюдение во временном ряде нам потребовалось для создания факторной переменной $\log(\text{USDollar}(-1))$.

После того как на основе алгоритма действий № 12 будут составлены соответствующие интервальные прогнозы, у нас появится возможность сопоставить заданные уровни надежности с фактической долей точных интервальных прогнозов. Судя по табл. 6.18, доля точных прогнозов оказалась незначительно ниже заданного уровня надежности при 99,9%-ном и 99%-ном уровнях надежности и практически ему равна при 95%-ном (отклонение на -0,1 процентного пункта). Однако при 90%-ном уровне надежности и более низких уровнях доля фактических прогнозов становится выше заданного уровня. Причем эта положительная разница растет при снижении уровня надежности, достигая своего максимума при 40%-ном уровне надежности, когда она равна 36 процентным пунктам.

Таблица 6.18

Доля точных интервальных прогнозов при разных уровнях надежности
в период с июля 1992 г. по июль 2010 г. (217 наблюдений)

| t-значение | Уровень надежности, % | Количество точных прогнозов | Доля точных интервальных прогнозов, % | Разница между долей точных интервальных прогнозов и уровнем надежности, проц. пункт |
|------------|-----------------------|-----------------------------|---------------------------------------|---|
| 3,3375 | 99,9 | 214 | 98,6 | -1,3 |
| 2,5994 | 99 | 212 | 97,7 | -1,3 |
| 1,9713 | 95 | 206 | 94,9 | -0,1 |
| 1,6521 | 90 | 202 | 93,1 | 3,1 |

Окончание

| t-значение | Уровень надежности, % | Количество точных прогнозов | Доля точных интервальных прогнозов, % | Разница между долей точных интервальных прогнозов и уровнем надежности, проц. пункт |
|------------|-----------------------|-----------------------------|---------------------------------------|---|
| 1,2856 | 80 | 201 | 92,6 | 12,6 |
| 1,039 | 70 | 199 | 91,7 | 21,7 |
| 0,8433 | 60 | 189 | 87,1 | 27,1 |
| 0,6757 | 50 | 181 | 83,4 | 33,4 |
| 0,5252 | 40 | 165 | 76,0 | 36,0 |
| 0,3858 | 30 | 125 | 57,6 | 27,6 |
| 0,2537 | 20 | 112 | 51,6 | 31,6 |
| 0,1258 | 10 | 75 | 34,6 | 24,6 |

Однако если сравнить табл. 6.18, характеризующую точность интервальных прогнозов для стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, с табл. 4.8, которая характеризует их точность для нестационарной модели $\text{USDOLLAR} = a \times \text{USDOLLAR}(-1) + a \times \text{USDOLLAR}(-2)$, то выяснится, что последняя модель с этой точки зрения несколько точнее.

Еще более серьезным минусом интервальных прогнозов, составленных по стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ за весь период наблюдений (т.е. на основе данных с июля 1992 г. по июнь 2010 г.), являются слишком широкие интервалы прогнозов для большей части временного ряда, начиная с октября 1999 г. Так, при прогнозе на июль 1992 г. общий диапазон интервального прогноза (верхняя граница интервального прогноза минус нижняя граница интервального прогноза) при 95%-ном уровне надежности составил лишь 4 коп. (табл. 6.19), в то время как фактическое значение курса доллара было равно лишь 16,12 коп. В свою очередь при прогнозе на июль 2010 г. общий диапазон интервального прогноза был равен 8 руб. 32,27 коп., а фактический курс доллара составил 30 руб. 18,69 коп. При этом в июле 1992 г. диапазон интервального прогноза составлял 25,77% от фактического курса доллара, а в июле 2010 г. его доля в стоимости курса американской валюты равнялась 27,57%, в то время как для нестационарной модели доля интервального прогноза для последнего наблюдения равнялась 10,62%.

Таблица 6.19

Интервальные и точечные прогнозы по курсу доллара за период с июля по декабрь 1992 г. и за период с января по май 2010 г.

| Месяц | Факт | Прогноз | Нижняя граница интервального прогноза | Верхняя граница интервального прогноза | Точность интервального прогноза |
|---------|---------|---------|---------------------------------------|--|---------------------------------|
| 1992M07 | 0,1612 | 0,1492 | 0,1284 | 0,1700 | В интервале прогноза |
| 1992M08 | 0,2050 | 0,1935 | 0,1667 | 0,2203 | В интервале прогноза |
| 1992M09 | 0,2540 | 0,2430 | 0,2095 | 0,2766 | В интервале прогноза |
| 1992M10 | 0,3980 | 0,2981 | 0,2571 | 0,3390 | Выше интервала прогноза |
| 1992M11 | 0,4470 | 0,4899 | 0,4230 | 0,5568 | В интервале прогноза |
| 1992M12 | 0,4145 | 0,4977 | 0,4299 | 0,5656 | Выше интервала прогноза |
| ... | ... | ... | ... | ... | ... |
| 2008M01 | 24,4764 | 24,6888 | 21,3980 | 27,9797 | В интервале прогноза |
| 2008M02 | 24,1159 | 24,5137 | 21,2463 | 27,7812 | В интервале прогноза |
| 2008M03 | 23,5156 | 24,1168 | 20,9024 | 27,3312 | В интервале прогноза |
| 2008M04 | 23,6471 | 23,4823 | 20,3527 | 26,6119 | В интервале прогноза |
| 2008M05 | 23,7384 | 23,8033 | 20,6308 | 26,9757 | В интервале прогноза |
| 2008M06 | 23,4573 | 23,8334 | 20,6569 | 27,0099 | В интервале прогноза |
| 2008M07 | 23,4456 | 23,4815 | 20,3520 | 26,6110 | В интервале прогноза |
| 2008M08 | 24,5769 | 23,5559 | 20,4165 | 26,6953 | В интервале прогноза |
| 2008M09 | 25,2464 | 24,9320 | 21,6088 | 28,2553 | В интервале прогноза |
| 2008M10 | 26,5430 | 25,3960 | 22,0106 | 28,7814 | В интервале прогноза |

Продолжение

| Месяц | Факт | Прогноз | Нижняя граница интервального прогноза | Верхняя граница интервального прогноза | Точность интервального прогноза |
|----------------|----------------|----------------|---------------------------------------|--|---------------------------------|
| 2008M11 | 27,6060 | 26,8727 | 23,2899 | 30,4555 | В интервале прогноза |
| 2008M12 | 29,3804 | 27,7922 | 24,0863 | 31,4982 | В интервале прогноза |
| 2009M01 | 35,4146 | 29,7355 | 25,7695 | 33,7014 | Выше интервала прогноза |
| 2009M02 | 35,7205 | 36,7244 | 31,8224 | 41,6264 | В интервале прогноза |
| 2009M03 | 34,0134 | 35,1777 | 30,4819 | 39,8734 | В интервале прогноза |
| 2009M04 | 33,2491 | 33,4981 | 29,0276 | 37,9687 | В интервале прогноза |
| 2009M05 | 30,9843 | 32,9875 | 28,5855 | 37,3894 | В интервале прогноза |
| 2009M06 | 31,2904 | 30,3812 | 26,3282 | 34,4341 | В интервале прогноза |
| 2009M07 | 31,7555 | 31,3969 | 27,2083 | 35,5855 | В интервале прогноза |
| 2009M08 | 31,5687 | 31,7026 | 27,4729 | 35,9322 | В интервале прогноза |
| 2009M09 | 30,0922 | 31,3976 | 27,2087 | 35,5865 | В интервале прогноза |
| 2009M10 | 29,0488 | 29,6854 | 25,7257 | 33,6450 | В интервале прогноза |
| 2009M11 | 29,8179 | 28,8399 | 24,9935 | 32,6862 | В интервале прогноза |
| 2009M12 | 30,2442 | 29,9946 | 25,9938 | 33,9953 | В интервале прогноза |
| 2010M01 | 30,4312 | 30,2170 | 26,1864 | 34,2477 | В интервале прогноза |
| 2010M02 | 29,9484 | 30,3885 | 26,3348 | 34,4421 | В интервале прогноза |
| 2010M03 | 29,3638 | 29,7577 | 25,7884 | 33,7269 | В интервале прогноза |
| 2010M04 | 29,2886 | 29,2044 | 25,3093 | 33,0995 | В интервале прогноза |

Окончание

| Месяц | Факт | Прогноз | Нижняя граница интервального прогноза | Верхняя граница интервального прогноза | Точность интервального прогноза |
|---------|---------|---------|---------------------------------------|--|---------------------------------|
| 2010M05 | 30,4956 | 29,2520 | 25,3505 | 33,1534 | В интервале прогноза |
| 2010M06 | 31,1954 | 30,7189 | 26,6211 | 34,8166 | В интервале прогноза |
| 2010M07 | 30,1869 | 31,1929 | 27,0315 | 35,3542 | В интервале прогноза |

Источник: расчеты автора.

6.6. Построение стационарной модели ARMA с оптимизированным временным рядом

Чем же объясняется слишком широкий диапазон интервальных прогнозов для большей части наблюдений, полученных по модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1))$? Ведь мы уже знаем, что если временной ряд является слабо стационарным, то это означает отсутствие, во-первых, тренда; во-вторых, строго периодических колебаний; в-третьих, систематических изменений дисперсии; в-четвертых, каких-либо иных систематических изменений во временном ряде (см. главу 1). Если систематические изменения дисперсии отсутствуют, то, следовательно, и абсолютная величина диапазона интервального прогноза не должна с течением времени существенно изменяться. Тем не менее по абсолютной величине интервальные прогнозы существенно изменились, что очевидно связано с неправильной спецификацией статистической модели.

Правда, если мы будем составлять интервальные прогнозы относительно логарифмического ряда данных, то в этом случае разница в их диапазоне относительно первого и последнего наблюдения будет не столь значительной. Так, для июля 1992 г. доля диапазона интервального прогноза составит 3,71% от логарифмического фактического курса доллара, а в июле 2010 г. — 1,97%.

Почему же стационарная модель $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ не позволяет построить прогнозы с оптимальной шириной интервальных прогнозов при переходе к исходному временному

ряду? В главе 1 (см. 1.2) мы уже научились распознавать стационарность временного ряда с помощью построения его графика. Попробуем построить аналогичный график для логарифмических остатков (за период с июля 1992 г. по июнь 2010 г.), полученных в результате решения уравнения регрессии по стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$. В результате получится диаграмма, изображенная на рис. 6.12.

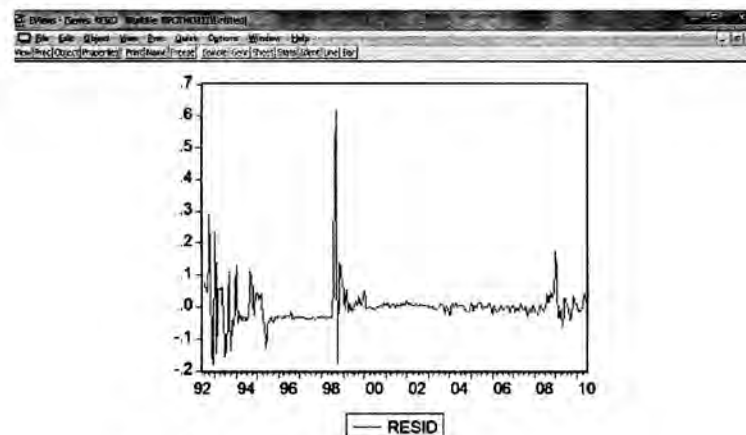


Рис. 6.12. Колебания логарифмических остатков (получены по стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ в период с июля 1992 г. по июнь 2010 г.

Источник: расчеты автора.

Исходя из рис. 6.12 можно сделать следующие выводы: во-первых, большая часть логарифмических остатков, полученных по стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, колеблется примерно в одном диапазоне вокруг нулевого уровня; во-вторых, с течением времени волатильность логарифмических остатков постепенно снижается; в-третьих, на графике видны три значительных всплеска волатильности остатков, однако последний всплеск по сравнению с предыдущими явно незначительный. Отсюда можно сделать вывод, что логарифмические остатки стационарной (точнее сказать, слабо стационарной) статистической модели асимптотически стремятся к относительно узкому диапазону колебаний, т.е. становятся все более стационарными, однако на начальном этапе временного ряда эти колебания еще довольно велики.

Таким образом, чтобы получить оптимальную ширину интервальных прогнозов для стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, необходимо убрать из базы данных часть временного ряда с наиболее волатильными остатками. Для отсеивания наиболее волатильной части остатков будем использовать тест Чоу на точность прогноза. Исходя из рис. 6.12 и с учетом данных табл. 5.4 «Рейтинг наблюдений по величине скачка курса доллара», которые показывают максимальный рост волатильности после дефолта августа 1998 г., проведем тест на точность прогноза относительно сентября 1998 г. В результате получим табл. 6.20, согласно которой нельзя сделать однозначный вывод о наличии структурного изменения. Дело в том, что уровень значимости (*Probability*) *F*-критерия получился больше 0,05, что свидетельствует в пользу нулевой гипотезы об отсутствии структурных изменений. Однако уровень значимости *LR*-статистики равен 0,001354, что существенно меньше 0,05 и однозначно говорит о наличии структурного изменения. В этой ситуации более надежна *LR*-статистика, поскольку *F*-критерий предполагает наличие независимых и нормально распределенных остатков, чего не может быть по определению при решении уравнений авторегрессии. Поэтому нулевая гипотеза о стабильности временного ряда, включающего наблюдения с сентября 1998 г. по июнь 2010 г., отвергается.

Таблица 6.20

Тест Чоу на точность прогноза относительно сентября 1998 г.

| | | | | |
|--|-------------|-------------|-------------|--------|
| Chow Forecast Test: Forecast from 1998M09 to 2010M06 | | | | |
| F-statistic | 0,749907 | Probability | 0,925059 | |
| Log likelihood ratio | 197,9028 | Probability | 0,001354 | |
| Test Equation: | | | | |
| Dependent Variable: LOG(USDOLLAR) | | | | |
| Method: Least Squares | | | | |
| Date: 09/04/10 Time: 14:12 | | | | |
| Sample: 1992M07 1998M08 | | | | |
| Included observations: 74 | | | | |
| Convergence achieved after 7 iterations | | | | |
| Backcast: 1992M06 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| LOG(USDOLLAR(-1)) | 0,941631 | 0,008722 | 107,9660 | 0,0000 |

Окончание

| | | | | |
|----------------------|----------|-----------------------|----------|-----------|
| C | 0,111973 | 0,012170 | 9,200832 | 0,0000 |
| MA(1) | 0,032561 | 0,128010 | 0,254362 | 0,8000 |
| R-squared | 0,994300 | Mean dependent var | | 1,012896 |
| Adjusted R-squared | 0,994139 | S.D. dependent var | | 0,963060 |
| S.E. of regression | 0,073728 | Akaike info criterion | | -2,337169 |
| Sum squared resid | 0,385944 | Schwarz criterion | | -2,243761 |
| Log likelihood | 89,47527 | F-statistic | | 6192,286 |
| Durbin — Watson stat | 1,868236 | Prob (F-statistic) | | 0,000000 |
| Inverted MA Roots | -0,03 | | | |

Далее проведем тест на точность прогноза относительно октября 1998 г. В результате получим табл. 6.21, согласно которой можно сделать однозначный вывод о структурной стабильности выделенного временного ряда. Поскольку уровень значимости (*Probability*) *F*-критерия получился больше 0,05 и уровень значения *LR*-статистики оказался равен 1,0, что однозначно говорит об отсутствии структурного изменения, нулевая гипотеза о стабильности временного ряда, включающего наблюдения с октября 1998 г. по июнь 2010 г., принимается.

Таблица 6.21

Тест Чоу на точность прогноза относительно октября 1998 г.

| | | | |
|--|----------|-------------|----------|
| Chow Forecast Test: Forecast from 1998M10 to 2010M06 | | | |
| F-statistic | 0,108694 | Probability | 1,000000 |
| Log likelihood ratio | 41,68388 | Probability | 1,000000 |
| Test Equation: | | | |
| Dependent Variable: LOG(USDOLLAR) | | | |
| Method: Least Squares | | | |
| Date: 09/04/10 Time: 13:28 | | | |
| Sample: 1992M07 1998M09 | | | |
| Included observations: 75 | | | |
| Convergence achieved after 5 iterations | | | |
| Backcast: 1992M06 | | | |

Окончание

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------------------|-------------|-----------------------|-------------|-----------|
| LOG(USDOLLAR(-1)) | 0,953479 | 0,017370 | 54,89100 | 0,0000 |
| C | 0,113618 | 0,024542 | 4,629456 | 0,0000 |
| MA(1) | 0,467342 | 0,124643 | 3,749435 | 0,0004 |
| R-squared | 0,988761 | Mean dependent var | | 1,036412 |
| Adjusted R-squared | 0,988449 | S.D. dependent var | | 0,977971 |
| S.E. of regression | 0,105110 | Akaike info criterion | | -1,628438 |
| Sum squared resid | 0,795466 | Schwarz criterion | | -1,535738 |
| Log likelihood | 64,06643 | F-statistic | | 3167,056 |
| Durbin — Watson stat | 1,572749 | Prob (F-statistic) | | 0,000000 |
| Inverted MA Roots | -0,47 | | | |

Исходя из результатов теста Чоу на точность прогноза построим статистическую модель $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ на основе данных за период с октября 1998 г. по июнь 2010 г. Вывод итогов после решения этого уравнения представлен в табл. 6.22, из которой следует, что уровень значимости у всех переменных, включенных в модель, оказался равен нулю и все они оказались статистически значимыми, как при 5%-ном, так и при 1%-ном уровне значимости.

Таблица 6.22

Вывод итогов после решения уравнения
 $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$

на основе данных за период с октября 1998 г. по июнь 2010 г.

| |
|--|
| Dependent Variable: LOG(USDOLLAR) |
| Method: Least Squares |
| Date: 09/04/10 Time: 14:40 |
| Sample: 1998M10 2010M06 |
| Included observations: 141 |
| Convergence achieved after 12 iterations |
| Backcast: 1998M09 |

Окончание

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------------------|-------------|-----------------------|-------------|---------------|
| LOG(USDOLLAR(-1)) | 0,880146 | 0,026271 | 33,50306 | 0,0000 |
| C | 0,403217 | 0,087422 | 4,612304 | 0,0000 |
| MA(1) | 0,568750 | 0,067925 | 8,373145 | 0,0000 |
| R-squared | 0,956076 | Mean dependent var | | 3,329758 |
| Adjusted R-squared | 0,955440 | S.D. dependent var | | 0,112608 |
| S.E. of regression | 0,023771 | Akaike info criterion | | -4,619680 |
| Sum squared resid | 0,077976 | Schwarz criterion | | -4,556940 |
| Log likelihood | 328,6874 | F-statistic | | 1501,906 |
| Durbin — Watson stat | 2,125648 | Prob (F-statistic) | | 0,000000 |
| Inverted MA Roots | -0,57 | | | |

Судя по табл. 6.23, уменьшение базы данных способствовало росту точности стационарной статистической модели по ряду параметров. С точки зрения прогнозирования особое значение имеет тот факт, что средняя ошибка по модулю уменьшилась на 1,71 процентных пункта, т.е. весьма существенно. Правда, средняя ошибка по модулю, напротив, выросла на 4,7 коп. Но это объясняется тем фактом, что средний курс доллара за период с октября 1998 г. по июнь 2010 г. оказался равен 28,70 руб. и был в 7,41 раза выше среднего курса доллара за период с июня 1992 г. по сентябрь 1998 г., равного 3,87 руб.

Таблица 6.23

Оценка точности моделей $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, построенных на основе данных за период с октября 1998 г. по июнь 2010 г. и за период с июля 1992 г. по июнь 2010 г.

| Параметр | Точность стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, построенной за период с октября 1998 г. по июнь 2010 г. | Точность стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, построенной за период с июля 1992 г. по июнь 2010 г. | Разница |
|---|--|---|---------|
| Root Mean Squared Error (квадратный корень средней ошибки предсказания) | 0,7558 | 0,8142 | -0,0584 |

Окончание

| Параметр | Точность стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, построенной за период с октября 1998 г. по июнь 2010 г. | Точность стационарной модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, построенной за период с июля 1992 г. по июнь 2010 г. | Разница |
|--|--|---|---------|
| Mean Absolute Error (средняя ошибка в руб. по модулю) | 0,4149 | 0,3679 | 0,047 |
| Mean Absolute Percentage Error (средняя ошибка по модулю, %) | 1,5301 | 3,2386 | -1,7085 |
| Theil Inequality Coefficient (коэффициент неравенства Тейла) | 0,0134 | 0,0177 | -0,0043 |
| Bias Proportion (доля систематической ошибки) | 0,0003 | 0,0046 | -0,0043 |
| Variance Proportion (доля вариации) | 0,0225 | 0,0032 | 0,0193 |
| Covariance Proportion (доля ковариации, т.е. несистематической ошибки) | 0,9772 | 0,9922 | -0,015 |

Следующим шагом будет расчет точечных и интервальных прогнозов для всех наблюдений, на основе которых составлена наша статистическая модель (за период с октября 1998 г. по июнь 2010 г.), а также на июль 2010 г. (курс доллара по этому месяцу не включен в базу данных). При этом точечный прогноз по курсу доллара на июль 2010 г., вычисленный по этой модели, оказался равен 31,02 руб. (фактический курс доллара в июле 2010 г. равнялся 30,19 руб.). Заметим, что этот прогноз оказался на 17 коп.

точнее аналогичного точечного прогноза (31,19 руб.), рассчитанного по модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, построенной по данным за весь период (с июня 1992 г. по июнь 2010 г.).

Далее на основе алгоритма действий № 12 составим по модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, построенной по данным с октября 1998 г. по июнь 2010 г., соответствующие интервальные прогнозы с разными уровнями надежности. Назовем последнюю модель стационарной моделью с оптимизированным временным рядом.

Посмотрим, как у этой модели заданные уровни надежности соотносились с фактической долей точных интервальных прогнозов. После проведения соответствующих подсчетов удалось выяснить, что при 95%-ном уровне надежности из 142 составленных по этой модели интервальных прогнозов в 138 случаях фактический курс доллара оказался в рамках интервального прогноза, т.е. получился точным. Следовательно, при 95%-ном уровне надежности фактическая вероятность точного интервального прогноза у стационарной модели с оптимизированным временным рядом достигла 97,2%, т.е. получилась на 2,2 процентного пункта выше заданного 95%-ного уровня надежности. Судя по табл. 6.24, доля точных прогнозов по этой модели оказалась незначительно ниже заданного уровня надежности лишь при 99,9%-ном уровне. В то время как при 99%-ном уровне надежности и ниже доля точных фактических прогнозов становится выше заданного уровня. Причем эта положительная разница достигает максимума при 40%-ном уровне надежности, когда она равна 26,2 процентного пункта.

Если сравнить данные табл. 6.24 с данными табл. 6.18, то легко сделать вывод, что интервальные прогнозы, составленные по модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ с оптимизированным временным рядом, получились надежнее интервальных прогнозов, рассчитанных на основе аналогичной модели с полным временным рядом.

Таблица 6.24

Доля точных интервальных прогнозов, составленных по модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$, при разных уровнях надежности в период с октября 1998 г. по июль 2010 г. (142 наблюдения)

| t-значение | Уровень надежности, % | Количество точных прогнозов | Доля точных интервальных прогнозов, % | Разница между долей точных интервальных прогнозов и уровнем надежности, проц. пункт |
|------------|-----------------------|-----------------------------|---------------------------------------|---|
| 3,3619 | 99,9 | 141 | 99,3 | -0,6 |
| 2,6117 | 99 | 141 | 99,3 | 0,3 |
| 1,9772 | 95 | 138 | 97,2 | 2,2 |

Окончание

| t-значение | Уровень надежности, % | Количество точных прогнозов | Доля точных интервальных прогнозов, % | Разница между долей точных интервальных прогнозов и уровнем надежности, проц. пункт |
|------------|-----------------------|-----------------------------|---------------------------------------|---|
| 1,6559 | 90 | 130 | 91,5 | 1,5 |
| 1,2877 | 80 | 124 | 87,3 | 7,3 |
| 1,0403 | 70 | 120 | 84,5 | 14,5 |
| 0,8442 | 60 | 107 | 75,4 | 15,4 |
| 0,6763 | 50 | 100 | 70,4 | 20,4 |
| 0,5256 | 40 | 94 | 66,2 | 26,2 |
| 0,3861 | 30 | 69 | 48,6 | 18,6 |
| 0,2538 | 20 | 48 | 33,8 | 13,8 |
| 0,1259 | 10 | 22 | 15,5 | 5,5 |

У стационарной модели, построенной на основе базы данных за период с октября 1998 г. по июль 2010 г., есть еще один весьма ощутимый плюс — у нее более приемлемый диапазон интервальных прогнозов. Так, при прогнозе на октябрь 1998 г. общий диапазон интервального прогноза (верхняя граница интервального прогноза минус нижняя граница интервального прогноза) при 95%-ном уровне надежности составил 1,84 руб. (табл. 6.25), или 11,48% от фактического курса доллара, который тогда равнялся 16,01 руб. В то время как при прогнозе на июль 2010 г. общий диапазон интервального прогноза был равен 2,96 руб., или 9,79% от фактического курса доллара, который тогда равнялся 30,19 руб. Нетрудно также заметить, что за счет уменьшения стандартного отклонения (в структурно стабильном временном ряде, естественно, наблюдается более низкий уровень волатильности) ширина диапазона интервального прогноза в табл. 6.25 существенно меньше, чем в табл. 6.19.

Таблица 6.25

Интервальные и точечные прогнозы по курсу доллара, составленные по стационарной модели, построенной по данным с октября 1998 г. по июль 2010 г.

| Месяц | Факт | Прогноз | Нижняя граница интервального прогноза | Верхняя граница интервального прогноза | Точность интервального прогноза |
|---------|-------|---------|---------------------------------------|--|---------------------------------|
| 1998M10 | 16,01 | 16,7871 | 15,8654 | 17,7088 | В интервале прогноза |
| 1998M11 | 17,88 | 16,7281 | 15,8012 | 17,6550 | Выше интервала прогноза |

Продолжение

| Месяц | Факт | Прогноз | Нижняя граница интервального прогноза | Верхняя граница интервального прогноза | Точность интервального прогноза |
|---------|---------|---------|---------------------------------------|--|---------------------------------|
| 1998M12 | 20,65 | 19,6711 | 18,6350 | 20,7072 | В интервале прогноза |
| ... | ... | ... | ... | ... | ... |
| 2008M01 | 24,4764 | 25,1054 | 23,9041 | 26,3067 | В интервале прогноза |
| 2008M02 | 24,1159 | 24,6121 | 23,4339 | 25,7903 | В интервале прогноза |
| 2008M03 | 23,5156 | 24,362 | 23,1931 | 25,5309 | В интервале прогноза |
| 2008M04 | 23,6471 | 23,6251 | 22,4863 | 24,7639 | В интервале прогноза |
| 2008M05 | 23,7384 | 24,2365 | 23,0695 | 25,4035 | В интервале прогноза |
| 2008M06 | 23,4573 | 24,0206 | 22,8649 | 25,1763 | В интервале прогноза |
| 2008M07 | 23,4456 | 23,73 | 22,5856 | 24,8744 | В интервале прогноза |
| 2008M08 | 24,5769 | 23,8776 | 22,7260 | 25,0292 | В интервале прогноза |
| 2008M09 | 25,2464 | 25,4749 | 24,2562 | 26,6936 | В интервале прогноза |
| 2008M10 | 26,543 | 25,5288 | 24,3118 | 26,7458 | В интервале прогноза |
| 2008M11 | 27,606 | 27,4171 | 26,1157 | 28,7185 | В интервале прогноза |
| 2008M12 | 29,3804 | 27,8679 | 26,5468 | 29,1890 | Выше интервала прогноза |
| 2009M01 | 35,4146 | 30,2186 | 28,7835 | 31,6537 | Выше интервала прогноза |
| 2009M02 | 35,7205 | 37,8279 | 35,9730 | 39,6828 | Ниже интервала прогноза |
| 2009M03 | 34,0134 | 33,7093 | 32,0524 | 35,3662 | В интервале прогноза |
| 2009M04 | 33,2491 | 33,5281 | 31,9005 | 35,1557 | В интервале прогноза |
| 2009M05 | 30,9843 | 32,5416 | 30,9698 | 34,1134 | В интервале прогноза |

Окончание

| Месяц | Факт | Прогноз | Нижняя граница интервального прогноза | Верхняя граница интервального прогноза | Точность интервального прогноза |
|---------|---------|---------|---------------------------------------|--|---------------------------------|
| 2009M06 | 31,2904 | 29,883 | 28,4565 | 31,3095 | В интервале прогноза |
| 2009M07 | 31,7555 | 31,8172 | 30,2965 | 33,3379 | В интервале прогноза |
| 2009M08 | 31,5687 | 31,3657 | 29,8632 | 32,8682 | В интервале прогноза |
| 2009M09 | 30,0922 | 31,3526 | 29,8520 | 32,8532 | В интервале прогноза |
| 2009M10 | 29,0488 | 29,2574 | 27,8652 | 30,6496 | В интервале прогноза |
| 2009M11 | 29,8179 | 28,9144 | 27,5421 | 30,2867 | В интервале прогноза |
| 2009M12 | 30,2442 | 30,2322 | 28,7949 | 31,6695 | В интервале прогноза |
| 2010M01 | 30,4312 | 30,088 | 28,6557 | 31,5203 | В интервале прогноза |
| 2010M02 | 29,9484 | 30,4406 | 28,9906 | 31,8906 | В интервале прогноза |
| 2010M03 | 29,3638 | 29,5469 | 28,1417 | 30,9521 | В интервале прогноза |
| 2010M04 | 29,2886 | 29,2057 | 27,8188 | 30,5926 | В интервале прогноза |
| 2010M05 | 30,4956 | 29,2903 | 27,8995 | 30,6811 | В интервале прогноза |
| 2010M06 | 31,1954 | 31,0042 | 29,5270 | 32,4814 | В интервале прогноза |
| 2010M07 | 30,1869 | 31,0207 | 29,5386 | 32,5028 | В интервале прогноза |

В заключение остановимся на таком важном моменте прогнозирования, как сравнение индекса оптимальности четырех статистических моделей. Дело в том, что наряду с точностью интервальных прогнозов необходимо оценить и их оптимальность, т.е. позаботиться о том, чтобы средняя ширина интервального прогноза (верхняя граница интервального прогноза минус нижняя граница интервального прогноза) не была слишком велика, поскольку это снижает ценность прогноза для инвестора. С этой целью сравним четыре важных параметра по че-

тырем ранее разработанным статистическим моделям (табл. 6.26). Причем наиболее важным среди четырех параметров таблицы является индекс оптимальности интервальных прогнозов, который находится по следующей формуле:

$$\begin{aligned} \text{Индекс оптимальности интервальных прогнозов} = \\ = \text{Средний диапазон интервального} \\ \text{прогноза (руб.) : Точность интервальных прогнозов} \quad (6.10) \\ \text{(при 95\%-ном уровне надежности).} \end{aligned}$$

Следовательно, чем ниже индекс оптимальности интервальных прогнозов, тем лучше конкретная статистическая модель. Таким образом, первое место по этому показателю занимает стационарная модель с оптимизированным временным рядом, в то время как последнее — стационарная модель с полным временным рядом. При этом индекс оптимальности интервальных прогнозов у первой модели равен 0,0276, а это означает, что на один процентный пункт точности интервальных прогнозов (при 95%-ном уровне надежности) у нее приходится 2,76 коп. среднего диапазона интервального прогноза, в то время как один процентный пункт точности интервальных прогнозов у стационарной модели с полным временным рядом обошелся в 5,54 коп. среднего диапазона интервального прогноза, т.е. в последнем случае точность обошлась гораздо дороже.

Заметим также, что представленный в таблице средний диапазон интервального прогноза (руб.) вычислен путем суммирования всех диапазонов интервального прогноза по определенной статистической модели, которые затем делятся на общее количество наблюдений во временном ряде. В свою очередь средний диапазон интервального прогноза (%) находится по следующей формуле:

$$\begin{aligned} \text{Средний диапазон интервального прогноза} \\ \text{(руб.) : Средний фактический} \quad (6.11) \\ \text{курс доллара} \times 100\%. \end{aligned}$$

Судя по табл. 6.26, по всем четырем параметрам наиболее оптимальные показатели у стационарной модели с оптимизированным временным рядом, в то время как наименее оптимальные — у стационарной модели с полным временным рядом. Сравнивая две нестационарные модели, можно прийти к выводу, что модель с оптимизированным временным рядом превосходит модель с полным временным рядом по трем параметрам, незначительно уступая ей лишь по точности интервальных прогнозов (при 95%-ном уровне надежности).

Таблица 6.26

Сравнение четырех статистических моделей по индексу оптимальности и другим параметрам

| Статистическая модель | Средний диапазон интервального прогноза, руб. | Средний диапазон интервального прогноза, % от среднего фактического курса | Точность интервальных прогнозов, % (при 95%-ном уровне надежности) | Индекс оптимальности интервальных прогнозов |
|--|---|---|--|---|
| Нестационарная USDOLLAR = $a \times \text{USDOLLAR}(-1) +$ $+ b \times \text{USDOLLAR}(-2)$ с полным временным рядом | 3,239 | 16,41 | 96,3 | 0,0336 |
| Нестационарная USDOLLAR = $a \times \text{USDOLLAR}(-1) +$ $+ b \times \text{USDOLLAR}(-2)$ с оптимизированным временным рядом | 2,882 | 10,23 | 95,7 | 0,0301 |
| Стационарная $\log(\text{USDollar}) = c +$ $+ a \times \log(\text{USDollar}(-1))$ $+ \text{MA}(1)$ с полным временным рядом | 5,254 | 26,60 | 94,9 | 0,0554 |
| Стационарная $\log(\text{USDollar}) = c +$ $+ a \times \log(\text{USDollar}(-1)) +$ $+ \text{MA}(1)$ с оптимизированным временным рядом | 2,6847 | 9,55 | 97,2 | 0,0276 |

Контрольные вопросы и задания

1. Почему при составлении статистической модели со стационарной ARMA-структурой мы были вынуждены перейти от исходного временного ряда к логарифмическому временному ряду? В каком случае расширенный тест Дикки — Фуллера отвергает нулевую гипотезу о наличии единичного корня? Какой вывод можно сделать в этом случае о стационарности временного ряда?
2. Повторите весь перечень действий, необходимых для построения статистической модели, представляющей собой уравнения авторегрессии (AR) или уравнения авторегрессии со скользящей средней (ARMA). Сколько всего пунктов в этом перечне и можно ли его при необходимости расширить?
3. Каким образом коррелограмма используется для построения моделей авторегрессии и моделей авторегрессии со скользящей средней? Как найти с помощью автокорреляционной и частной автокорреляционной функций величину лага для лаговой переменной AR и для скользящей средней MA?
4. Какой тест используется для проверки модели авторегрессии со скользящей средней на автокорреляцию в остатках? Как проверяется на стационарность ARMA-структура этой статистической модели? К какому значению стремятся функции импульсного и накопленного ответа у стационарной модели? Как изменяется по мере увеличения лага автокорреляция и частная автокорреляция в остатках стационарной статистической модели?
5. Какие выводы можно сделать о стабильности стационарной и нестационарной статистических моделей, если сравнить табл. 6.11 и табл. 5.9? Какая из этих моделей продемонстрировала большую точность в прогнозах после 1998 г.?
6. Сравните точность стационарной и нестационарной статистических моделей в целом за весь период и за различные периоды времени? Какая из этих моделей оказалась точнее за период, начинающийся с 1999 г.? Подкрепите свой вывод конкретными цифрами.
7. Чем объясняется широкий диапазон интервальных прогнозов для большей части наблюдений, полученных по модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1))$? С помощью какого теста мы смогли построить стационарную статистическую модель с оптимизированным временным рядом? Назовите лучшую статистическую модель (из числа уже проанализированных) с точки зрения индекса оптимальности интервальных прогнозов.

ГЛАВА 7

ИСПОЛЬЗОВАНИЕ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ В КАЧЕСТВЕ ИНСТРУМЕНТА ТОРГОВОЙ СИСТЕМЫ

7.1. Использование в торговле модели для прогнозирования курса доллара к рублю с упреждением в один месяц

Заключительная седьмая глава будет посвящена применению статистических моделей в качестве инструмента торговой системы. При этом будут использоваться не только модели, делающие прогноз с упреждением в один месяц, но и модели, прогнозирующие курс валюты с упреждением в две недели, одну неделю и в один день. На материале этой главы читатель научится рассчитывать рекомендуемые курсы покупки и продажи валют, которые можно будет включать в стоп-заявки, используемые в валютной торговле.

Мы уже научились делать интервальные прогнозы по курсу доллара, суть которых заключается в определении двустороннего интервала прогноза (нижняя граница интервального прогноза < фактический курс доллара < верхняя граница интервального прогноза), в котором ожидаемый курс доллара должен оказаться с 95%-ным или иным уровнем надежности (см. шаг 2 «Вычисление интервальных прогнозов» в алгоритме действий № 12 «Как в EViews построить интервальные прогнозы»). Однако инвестора, который собирается в течение ближайшего месяца (или иного инвестиционного периода) продать доллары США (или какую-то иную валюту), интересует еще и другой вопрос, а именно: по какому курсу он должен продать американскую валюту, чтобы цена продажи оказалась выше фактического курса доллара в конце месяца? Соответственно если инвестор в течение месяца (или иного инвестиционного периода) планирует приобрести доллары,

то для него важен ответ на прямо противоположный вопрос: по какому курсу он должен купить эту валюту, чтобы цена этой покупки оказалась бы ниже фактического курса американской валюты в конце месяца?

Теперь предположим, что наш инвестор собрался купить или продать доллары в августе 2010 г. и с этой целью решил составить стоп-заявки, которые будут действительны в течение всего месяца. Следовательно, его инвестиционный период определен сроком с 1 августа по 31 августа 2010 г., а стоп-заявки должны быть готовы уже к концу 31 июля 2010 г. Таким образом, последняя информация по курсу доллара у инвестора будет на конец июля 2010 г. Поэтому вполне очевидно, что фактический курс доллара, установленный Банком России по итогам предыдущих торгов на 1 августа 2010 г., и следует взять в качестве исходного уровня для расчета стоп-заявок. При этом выдвигается предположение, что в течение месяца динамика курса доллара будет меняться в соответствии с нормальным распределением, т.е. будет результатом воздействия множества случайных факторов. В этом случае нам потребуется для расчета цен покупки и продажи использовать нормальное распределение, средним которого будет фактический курс доллара на конец июля 2010 г., а стандартным отклонением — средняя ошибка прогнозируемого индивидуального значения курса доллара на конец августа 2010 г., полученная по итогам решения уравнения регрессии. Методика расчета рекомендуемых курсов покупки и продажи валют представлена в алгоритме действий № 24.

Алгоритм действий № 24

Расчет рекомендуемых курсов покупки и продажи валют с заданным уровнем надежности (по одностороннему t -критерию)

Шаг 1. Расчет точечного прогноза и средней ошибки прогнозируемого курса

Методика расчета рекомендуемых цен покупки и продажи валют отличается от процедуры составления точечных и интервальных прогнозов (см. алгоритм действий № 11 «Как в EViews построить точечный прогноз» и алгоритм действий № 12 «Как в EViews построить интервальные прогнозы») тем, что вместо точечного прогноза в качестве среднего значения для нормального распределения берется фактический курс доллара на начало инвестиционного периода.

Однако чтобы получить оценку ожидаемой волатильности курса доллара, нам необходимо решить уравнение регрессии. Для этого будем использовать стационарную модель с оптимизированным вре-

менным рядом и лагом в один месяц, т.е. модель $\log(\text{USDollar}) = \rho + \alpha \times \log(\text{USDollar}(-1)) + \text{MA}(1)$. С этой целью включим в базу данных этой статистической модели информацию о курсе доллара за период с октября 1998 г. по июль 2010 г., а затем составим точечный прогноз на август 2010 г.

Согласно полученному точечному прогнозу, курс доллара по состоянию на 31 августа 2010 г. должен равняться 29,5544 руб. Однако точечный прогноз представляет собой экстраполяцию в будущее текущего тренда курса доллара без учета воздействия на его динамику случайных факторов. Поэтому точечный прогноз при разработке стоп-заявок играет роль своего рода оценочного ориентира, который по итогам каждых торгов нужно пересматривать.

Далее необходимо рассчитать среднюю ошибку индивидуального значения прогнозируемого курса доллара на август 2010 г. Методика ее расчета мы уже знаем (см. шаг 1 «Как найти средние ошибки прогнозируемого курса доллара» в алгоритме действий № 12). В результате удалось вычислить, что средняя ошибка прогнозируемого курса в августе 2010 г. будет равна 0,7122 руб.

Шаг 2. Расчет курсов покупки и продажи валют

После нахождения средней ошибки прогнозируемого индивидуального значения курса доллара на август 2010 г. следующей задачей является расчет цен покупки или продажи доллара с определенным уровнем надежности, чтобы фактический курс американской валюты в конце инвестиционного периода оказался выше или ниже рекомендуемых цен покупки и продажи. При этом границы доверительного интервала для заданного уровня надежности и для степени свободы $n = 142 - k - 1 = 142 - 2 - 1 = 139$ (где 142 — количество наблюдений во временном ряде; k — количество факторных переменных) в Excel находятся с помощью функции СТЬЮДРАСПОБР, вычисляющей t -критерий. Однако в этом случае нас интересует не двустороннее, а одностороннее значение t -критерия Стьюдента. Дело в том, что мы сейчас хотим ответить на два прямо противоположных вопроса. Во-первых, нам необходимо рассчитать, при какой цене продажи доллара (в течение месяца или иного инвестиционного периода) стоимость этой сделки с определенным уровнем надежности окажется выше его фактического курса в конце месяца. Во-вторых, нужно вычислить, при какой цене покупки американской валюты (в течение месяца или иного периода) стоимость этой сделки с определенным уровнем надежности окажется ниже его фактического курса в конце месяца.

Одностороннее t -значение для 90%-ного уровня надежности в Excel находится следующим образом: $\text{СТЮДРАСПОБР}(2 \times (1 - 0,9) = 0,2; 139) = 1,2877$. (Заметим, что при расчете двустороннего t -критерия с 90%-ного уровнем надежности первая цифра, введенная в функцию СТЮДРАСПОБР , была бы равна 0,1; при вычислении одностороннего t -критерия она должна удваиваться.) Соответственно при расчете t -значения, например, с 80%-ным уровнем надежности эта функция приобретает следующий вид: $(2 \times (1 - 0,8) = 0,4; 139) = 0,8442$.

После того как найдено t -значение для 90%-ного уровня надежности, у нас появляется возможность составить рекомендуемые цены покупки и продажи доллара на август 2010 г. (точнее сказать, на инвестиционный период с 1 августа по 31 августа), которые входят по следующей формуле:

$$\hat{Y}_{\text{прод}} = Y_{\text{факт}} - t \times SE_{\text{ind}} = 30,1869 + 1,2877 \times 0,7122 = 31,1040; \quad (7.1)$$

$$\hat{Y}_{\text{покуп}} = Y_{\text{факт}} - t \times SE_{\text{ind}} = 30,1869 - 1,2877 \times 0,7122 = 29,2698, \quad (7.2)$$

где $\hat{Y}_{\text{прод}}$ — рекомендуемая цена продажи доллара;
 $\hat{Y}_{\text{покуп}}$ — рекомендуемая цена покупки доллара;
 $Y_{\text{факт}}$ — фактический курс доллара на конец июля 2010 г.;
 t — значение одностороннего t -критерия для 90%-ного уровня надежности;
 SE_{ind} — средняя ошибка прогнозируемого индивидуального значения курса доллара на август 2010 г.

Соответственно для 80%-ного уровня надежности рекомендуемые цены покупки и продажи доллара на август 2010 г. будут следующими:

$$\hat{Y}_{\text{прод}} = Y_{\text{авг. 2010}} - t \times SE_{\text{ind}} = 29,5544 + 0,8442 \times 0,7122 = 30,7882;$$

$$\hat{Y}_{\text{покуп}} = Y_{\text{авг. 2010}} + t \times SE_{\text{ind}} = 29,5544 - 0,8442 \times 0,7122 = 29,4460.$$

При этом следует иметь в виду, что односторонний t -критерий в отличие от двустороннего t -критерия менее консервативен. При использовании одностороннего t -критерия учитывается только вероятность того, что значение $Y > X$ или, наоборот, $Y < X$. В то время как при использовании двустороннего t -критерия уровень надежности учитывает вероятность того, что $X < Y < X$, а потому задаваемый в этом случае доверительный интервал всегда больше.

Как мы уже говорили, согласно полученному точечному прогнозу, курс доллара по состоянию на 31 августа 2010 г. должен равняться 29,5544 руб. Однако точечный прогноз представляет собой экстраполяцию с упреждением в один месяц текущего тренда курса доллара без учета воздействия на его динамику случайных факторов. Поэтому точечный прогноз при разработке стоп-заявок играет роль ориентира, который по итогам каждого торгов желательно пересматривать.

Делать это можно, например, следующим образом. Если в статистической модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ в течение всего августа по итогам последнего торгового дня менять значения факторной переменной $\text{USDollar}(-1)$, то отклонения точечного прогноза от фактического курса доллара, установившегося в конце августа 2010 г., значительно уменьшатся.

Поскольку при первоначальном прогнозировании в качестве факторной переменной $\text{USDollar}(-1)$ использовалось значение курса доллара на 31 июля 2010 г., то в результате по итогам августа точечный прогноз отклонился от курса доллара на 1,1096 руб. (табл. 7.1). При этом фактический курс доллара оказался в рамках интервального прогноза при 90%-ном уровне надежности.

Но если в качестве лаговой переменной $\text{USDollar}(-1)$ взять значение курса доллара от 26 августа 2010 г., то отклонение точечного прогноза от фактического курса составит всего лишь 8,47 коп. Правда, ценность этого прогноза существенно ниже, поскольку он рассчитан на основе данных, полученных за пять дней до наступления прогнозируемой даты.

Впрочем, точность прогноза зависит даже не столько от актуальной информации по курсу доллара, сколько от близости (до определенного уровня) предыдущего курса доллара к его будущему значению. Судя по табл. 7.1, точность точечного прогноза в период с 3 августа по 5 августа 2010 г. ухудшилась, поскольку на рынке наблюдался падающий тренд, а следовательно, в качестве лаговой переменной $\text{USDollar}(-1)$ использовались более низкие (по сравнению с курсом 31 июля 2010 г.) значения курса доллара. Это объясняется тем, что к концу месяца курс доллара сильно вырос, а потому использование в качестве лаговой переменной более низких значений по курсу доллара, когда в начале месяца на рынке наблюдался падающий тренд, привело к большей погрешности.

Хотя слишком большая близость предыдущего курса доллара к его будущему значению также приводит к некоторому росту погрешности. Об этом свидетельствуют отклонения прогноза от фактического курса доллара, полученные при подстановке в качестве $\text{USDollar}(-1)$ курса доллара по итогам торгов, состоявшихся 28 августа и 31 августа 2010 г. Причем в том случае, когда для прогноза курса доллара на 31 августа

2010 г. использовался его курс, установившийся по итогам того же дня, полученная погрешность оказалась равна 42 коп. Эта погрешность обусловлена параметрами используемого уравнения регрессии, в которых заложен среднестатистический тренд, рассчитанный за период с октября 1998 г. по июль 2010 г. Однако чем больше текущие колебания курса доллара отклоняются от этого тренда, тем более высокую погрешность дают коэффициенты регрессии. Очевидно, что то же самое происходит и в том случае, когда мы сами передвигаем ежемесячную лаговую переменную USDollar(-1), в частности, используя информацию по курсу доллара за последний торговый день, в то время как модель построена на ежемесячных данных на конец месяца. Таким образом, без крайней необходимости, которая может возникнуть из-за пропуска наблюдения (например, в связи с тем, что в выходные или праздничные дни валютные торги не проводились), интервалы во временном ряде, включенном в базу данных статистической модели, не следует нарушать, иначе это сказывается на точности прогнозов.

Воспользовавшись алгоритмом действий № 24, мы рассчитали (исходя из предположения о нормальном распределении остатков) рекомендуемые цены покупки и продажи доллара на август 2010 г. с разными уровнями надежности, поместив их в табл. 7.1. Поскольку рекомендуемые цены покупки или продажи валюты представляют собой стоп-заявки, вступающие в силу при достижении курсом валюты определенного уровня, то инвестору следует учитывать специфику возникающих при этом рисков. В целом по мере повышения уровня надежности при расчете цены покупки и продажи валют повышается и вероятность того, что цена продажи (покупки) валюты будет действительно выше (ниже) ее фактического курса на конец месяца (или иного инвестиционного периода).

Таблица 7.1

Пересмотр точечного прогноза по модели
 $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$
 с учетом данных по курсу доллара за последний торговый день

| Дата | Курс доллара, взятый в качестве USDollar(-1) для прогнозирования | Прогноз по модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ | Отклонение прогноза от фактического курса доллара в конце августа 2010 г. |
|----------|--|---|---|
| 31.07.10 | 30,1869 | 29,55439 | 1,1096 |
| 03.08.10 | 30,1861 | 29,55324 | 1,1108 |
| 04.08.10 | 29,9681 | 29,23831 | 1,4257 |
| 05.08.10 | 29,7958 | 28,9894 | 1,6746 |

Окончание

| Дата | Курс доллара, взятый в качестве USDollar(-1) для прогнозирования | Прогноз по модели $\log(\text{USDollar}) = c + a \times \log(\text{USDollar}(-1)) + \text{MA}(1)$ | Отклонение прогноза от фактического курса доллара в конце августа 2010 г. |
|-----------------|--|---|---|
| 06.08.10 | 29,8633 | 29,08691 | 1,5771 |
| 07.08.10 | 29,8312 | 29,04054 | 1,6335 |
| 10.08.10 | 29,8186 | 29,02234 | 1,6417 |
| 11.08.10 | 30,0239 | 29,31891 | 1,1451 |
| 12.08.10 | 30,205 | 29,58055 | 1,0815 |
| 13.08.10 | 30,4493 | 29,93358 | 0,7304 |
| 14.08.10 | 30,4199 | 29,89108 | 0,7729 |
| 17.08.10 | 30,5199 | 30,03563 | 0,6284 |
| 18.08.10 | 30,4514 | 29,93661 | 0,7274 |
| 19.08.10 | 30,4257 | 29,89947 | 0,7645 |
| 20.08.10 | 30,4636 | 29,95425 | 0,7098 |
| 21.08.10 | 30,5099 | 30,02117 | 0,6428 |
| 24.08.10 | 30,6041 | 30,15735 | 0,5067 |
| 25.08.10 | 30,7559 | 30,37687 | 0,2871 |
| 26.08.10 | 30,8958 | 30,5793 | 0,0847 |
| 27.08.10 | 30,8227 | 30,4735 | 0,1905 |
| 28.08.10 | 30,6969 | 30,29154 | 0,3725 |
| 31.08.10 | 30,664 | 30,24397 | 0,4200 |

Однако по мере роста уровня надежности соответственно растет и риск упущенной прибыли, обусловленный неучастием инвестора в валютных торгах. Дело в том, что чем выше установленный уровень надежности, тем меньше вероятность пересечения курсом валюты рекомендуемой цены ее покупки или продажи, что снижает для инвестора возможность выхода на рынок, хотя и уменьшает риск совершения невыгодной сделки, а также его транзакционные расходы.

Анализируя данные табл. 7.2, нетрудно сделать вывод, что при выборе инвестором цены покупки или продажи с более высоким уровнем надежности происходит снижение риска совершения невыгодной сделки, т.е. уменьшается вероятность того, что цена продажи или покупки окажется ниже фактического курса доллара, который установится на конец месяца. Однако прямо пропорционально снижению риска совершения невыгодной сделки растет риск упущенной прибыли и уменьшается вероятность выхода инвестора на рынок.

Чтобы сделать более сбалансированным соотношение между риском совершения невыгодной сделки и риском упущенной прибыли, инвестор может пойти на изменение структуры планируемой валютной сделки. Например, он имеет возможность не делать ставку на одном уровне цены покупки или продажи, а инвестировать средства по мере достижения курсом доллара цен покупки (или продажи), установленных с 60%-ными, 70%-ными, 80%-ными и 90%-ными уровнями надежности. В этом случае структура совокупного риска по четырем последовательно проведенным валютным транзакциям приобретет следующий вид: риск упущенной прибыли снизится до 75%, вероятность выхода на рынок вырастет до 25%, а риск совершения невыгодной сделки увеличится до 25%.

Как уже говорилось, оценки рисков в табл. 7.2 сделаны на основе предположения о нормальном распределении остатков, однако их фактическое распределение не вполне совпадает с нормальным. Это приводит к определенным отклонениям реальных значений оцениваемого риска от его оценок.

Таблица 7.2

Рекомендуемые цены продажи или покупки и уровень их надежности

| Цены покупки и продажи | Цена продажи с уровнем надежности (не ниже) | | | | Цена покупки с уровнем надежности (не выше) | | | |
|--------------------------------|---|---------|---------|---------|---|---------|---------|---------|
| | 90% | 80% | 70% | 60% | 90% | 80% | 70% | 60% |
| Курс реализации сделки | 31,1040 | 30,7882 | 30,5613 | 30,3677 | 29,2698 | 29,5856 | 29,8125 | 30,0061 |
| Риск невыгодной сделки, % | 10 | 20 | 30 | 40 | 10 | 20 | 30 | 40 |
| Вероятность выхода на рынок, % | 10 | 20 | 30 | 40 | 10 | 20 | 30 | 40 |
| Риск упущенной прибыли, % | 90 | 80 | 70 | 60 | 90 | 80 | 70 | 60 |

Источник: расчеты автора.

Следовательно, прежде чем перейти к тестированию (используя имеющиеся рыночные данные) эффективности рекомендуемых цен покупки и продажи, сначала нужно убедиться в их обоснованности. Иначе

говоря, нужно ответить на вопрос: какая часть цен покупки или продажи, рассчитанных по предложенной методике, в действительности оказалась ниже (или выше) фактического курса доллара в конце месяца?

Чтобы провести эту оценку, не следует ограничиваться расчетом рекомендуемых курсов покупки и продажи доллара на август 2010 г., нужно составить их для всего периода с октября 1998 г. по июль 2010 г., а затем вычислить отклонения рекомендуемых цен от заданного уровня надежности. В таблице 7.3 показан фактический риск того, что рекомендуемая цена продажи валюты, вычисленная с определенным уровнем надежности, в действительности может оказаться ниже курса доллара на конец инвестиционного периода (конец месяца).

Таблица 7.3

Уровень надежности при расчете рекомендуемых цен продажи доллара и риск неудачной сделки

| Односторонний <i>t</i> -критерий | Уровень надежности, % | Количество удачных сделок | Вероятность удачной сделки, % | Разница |
|----------------------------------|-----------------------|---------------------------|-------------------------------|---------|
| 2,3535 | 99 | 138 | 97,2 | -1,8 |
| 1,6559 | 95 | 134 | 94,4 | -0,6 |
| 1,2877 | 90 | 131 | 91,5 | 1,5 |
| 0,8442 | 80 | 128 | 89,4 | 9,4 |
| 0,5256 | 70 | 120 | 84,5 | 14,5 |
| 0,2538 | 60 | 96 | 67,6 | 7,6 |

Источник: расчеты автора.

Из таблицы 7.3 можно сделать вывод, что вероятность удачной сделки при продаже доллара по ценам, рассчитанным с 90%-ным уровнем надежности и при более низких уровнях надежности, выше заданного уровня надежности. Причем при 70%-ном уровне надежности эта положительная разница достигает своего максимума — 14,5 процентных пункта. Правда, при продаже доллара с 95%-ным и 99%-ным уровнями надежности вероятность удачной сделки несколько ниже установленного уровня надежности, и особенно при более высоком уровне надежности.

В таблице 7.4 показан фактический риск того, что рекомендуемая цена покупки валюты, вычисленная с определенным уровнем надежности, в действительности может оказаться выше курса доллара на конец инвестиционного периода (конец месяца). Судя по этой таблице, фактическая вероятность удачной сделки выше установленного уровня надежности при покупке доллара по ценам, рассчитанным как с 99%-ным уровнем надежности, так и с более низкими уровнями на-

дежности. Причем при 70%-ном уровне надежности эта положительная разница в пользу фактической вероятности удачной сделки достигает своего максимума — 17,3 процентного пункта.

Таблица 7.4

Уровень надежности при расчете рекомендуемых цен покупки доллара и риск неудачной сделки

| Односторонний <i>t</i> -критерий | Уровень надежности, % | Количество удачных сделок | Вероятность удачной сделки, % | Разница |
|----------------------------------|-----------------------|---------------------------|-------------------------------|---------|
| 2,3535 | 99 | 141 | 99,3 | 0,3 |
| 1,6559 | 95 | 140 | 98,6 | 3,6 |
| 1,2877 | 90 | 138 | 97,2 | 7,2 |
| 0,8442 | 80 | 134 | 94,4 | 14,4 |
| 0,5256 | 70 | 124 | 87,3 | 17,3 |
| 0,2538 | 60 | 104 | 73,2 | 13,2 |

Если сравнить табл. 7.3 и 7.4, то нетрудно заметить, что вероятность удачной сделки при покупке доллара выше, чем при его продаже. Так, при 99%-ном уровне надежности разница в пользу покупки доллара выше на 2,1 процентных пункта; при 95%-ном — на 4,2 процентного пункта; при 90%-ном — на 5,6 процентного пункта; при 80%-ном — на 4,9 процентного пункта, при 70%-ном — на 2,8 процентного пункта и при 60%-ном уровне надежности — на 5,6 процентного пункта. Очевидно, что эта асимметрия в доходности вызвана как действием ряда фундаментальных факторов, например длительной и на порядок более высокой (по сравнению с ростом цен в США) российской инфляцией, так и политикой дешевого рубля, долгое время проводившейся Банком России.

Оптимальными для валютной торговли являются цены покупки или продажи, установленные в диапазоне от 60 до 90% уровня надежности, поскольку при дальнейшем снижении уровня надежности у инвестора начинает резко падать вероятность удачной сделки, а при превышении 90%-ного уровня надежности значительно вырастает риск упущенной прибыли, в то время как вероятность удачной сделки повышается не столь существенно.

Теперь проверим, насколько эффективным могло быть использование рассчитанных нами цен покупки и продажи в ходе торгов на валютном рынке в августе 2010 г. Поскольку до 10 августа в динамике курса доллара преобладал падающий тренд (рис. 7.1), то первым в торгах смог поучаствовать инвестор (сделка была совершена 4 августа), установивший цену покупки с 60%-ным уровнем надежности. Одна-

ко начиная с 11 августа американская валюта начала устойчиво расти, поэтому на рынок стали выходить игроки, планировавшие продать доллар. В результате 13 августа сделку совершил инвестор, установивший цену продажи с 60%-ным уровнем надежности, 24 августа — инвестор, установивший цену продажи с 70%-ным уровнем надежности, с 26 августа — инвестор, установивший цену продажи с 80%-ным уровнем надежности. Вместе с тем инвестор, установивший цену продажи с 90%-ным уровнем надежности, на рынок выйти не смог, поскольку курс доллара не вырос до этого уровня. Тем более не смогли участвовать в торгах (из-за отсутствия ожидаемого тренда) инвесторы, установившие цену покупки с 70%-ным, 80%-ным и 90%-ными уровнями надежности.



Рис. 7.1. Курс доллара и рекомендуемые цены его покупки и продажи в августе 2010 г.

В таблице 7.5 представлены итоги валютных торгов за период с 1 августа по 31 августа 2010 г. для инвесторов, установивших цены покупки или продажи доллара с разными уровнями надежности. Судя по этой таблице, наиболее высокую курсовую доходность в размере 2,35% по итогам августа получил инвестор, установивший цену продажи доллара с 80%-ным уровнем надежности. На втором месте по этому показателю (с 2,32%-ной доходностью) инвестор, установивший цену покупки с 60%-ным уровнем надежности. Очевидно, что курсовая доходность у последнего инвестора была обеспечена за счет большего риска, однако поскольку тренд на рынке изменился в сторону роста, то этот риск оказался оправданным.

На третьем месте оказался инвестор, придерживавшийся консервативной стратегии «купил и держи», доходность которой в конце месяца

Таблица 7.5
Результаты валютных торгов для инвесторов, установивших цены покупки или продажи доллара с разными уровнями риска, в августе 2010 г.

| Дата продажи или покупки | 31 июля 2010 г. | 4 августа 2010 г. | 13 августа 2010 г. | 24 августа 2010 г. | 26 августа 2010 г. | 31 августа 2010 г. | Полученный кур- совый доход, в % |
|---|--------------------|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-------------------------------------|
| Фактический курс доллара | 30,1869 | 29,9681 | 30,4493 | 30,6041 | 30,8958 | 30,664 | — |
| Купил и продал в конце периода | 30,1869 | — | — | — | — | 30,664 | 1,58 |
| Продажа по цене с 60%-ным уровнем надежности | 30,1869 | — | 30,4493 | — | — | — | 0,27 |
| Продажа по цене с 70%-ным уровнем надежности | 30,1869 | — | — | 30,6041 | — | — | 1,38 |
| Продажа по цене с 80%-ным уровнем надежности | 30,1869 | — | — | — | 30,8958 | — | 2,35 |
| Продажа по цене с 90%-ным уровнем надежности | 30,1869 | — | — | — | — | — | 0,00 |
| Покупка по цене с 60%-ным уровнем надежности | — | 29,9681 | — | — | — | — | 2,32 |
| Покупка по цене с 70%-ным уровнем надежности | — | — | — | — | — | — | 0,00 |
| Покупка по цене с 80%-ным уровнем надежности | — | — | — | — | — | — | 0,00 |
| Покупка по цене с 90%-ным уровнем надежности | — | — | — | — | — | — | 0,00 |

составила 1,58%. Кстати, фактическая вероятность удачной сделки для инвестора, всегда придерживавшегося этой стратегии, по нашим подсчетам, за период с октября 1998 г. по август 2010 г. равна 55,2% (подсчет велся на конец каждого инвестиционного периода). Как видим, последняя цифра свидетельствует в пользу того, что американская валюта чаще растет, чем падает. Для справки заметим, что фактическая вероятность удачной сделки для инвестора, установившего цену покупки доллара с 60%-ным уровнем надежности, за этот же период оказалась существенно выше — 73,2%.

В большей части книги рассматривались статистические модели, делающие прогноз по курсу доллара с упреждением в один месяц, однако теперь перейдем к моделям, которые будут прогнозировать с упреждением в две недели, одну неделю и один день. Причем прогноз будет делаться не только по курсу доллара к рублю, но по и курсу евро к доллару. Нужно сразу отметить, что при переходе к прогнозированию с меньшим сроком упреждения точность наших моделей существенно повысится. И это вполне понятно с точки зрения здравого смысла, а с математической точки зрения объясняется следующим образом. Для оценки тесноты и направления связи между переменными одного временного ряда с определенным лагом используется автокорреляция, которая находится по формуле (1.3)

$$\text{Cov}(X_t, X_{t-T}) = \frac{\sum (X_t - X_{\text{cp}})(X_{t-T} - X_{\text{cp}})}{n}.$$

Однако, как в свое время писал специалист по теории вероятностей и случайным процессам А.А. Свешников: «... при достаточно большом значении интервала времени $T = t_1 - t_2$ отклонение ординаты случайной функции от ее математического ожидания в момент времени t_2 становится практически независимым от значения этого отклонения в момент времени t_1 . В этом случае функция Cov , дающая значение корреляционного момента между $X(t_1)$ и $X(t_2)$, при $T \rightarrow \infty$ будет стремиться к нулю»¹.

Вполне очевидно, что в том случае, когда $T \rightarrow 0$, величина отклонения переменной $X(t)$ от ее математического ожидания в момент времени t становится все более зависимой от значения этого отклонения в момент времени $t - T$. Иначе говоря, корреляция между лаговыми переменными возрастает, а когда в авторегрессионной модели значение $T = t_1 - t_2$ становится равным нулю, то вместо двух лаговых переменных мы получим лишь одну переменную $X(t)$, по которой, правда, будет невозможно прогнозировать.

¹ Свешников А.А. Прикладные методы теории случайных функций. М.: Наука, 1968. С. 33.

7.2. Использование в торговле модели для прогнозирования курса доллара к рублю с упреждением в две недели

Теперь посмотрим, какие результаты может дать расчет цен покупки и продажи валюты для двухнедельного инвестиционного периода. С этой целью построим статистическую модель, по которой можно делать прогноз по курсу доллара к рублю с упреждением в две недели. Полный список действий, которые необходимо выполнить при разработке такой модели, можно найти в главе 6 (см. алгоритм действий № 22 «Перечень действий, необходимых для построения статистической модели, представляющей собой уравнения авторегрессии (AR) или уравнения авторегрессии со скользящей средней (ARMA)»). Однако на подробном описании этих действий мы сейчас не будем останавливаться, поскольку с методикой построения статистических моделей наш читатель уже знаком, поэтому здесь дадим только краткую характеристику этой модели, а также остановимся на наиболее интересных моментах, связанных с ее разработкой.

На основе базы данных по курсу доллара, взятых с интервалом в две недели (на конец периода) с октября 1998 г. по июнь 2010 г., была построена прогностическая модель, по которой можно делать прогнозы с упреждением в две недели. Вывод данных по итогам решения уравнения регрессии можно увидеть в табл. 7.6.

Таблица 7.6

Вывод данных по статистической модели
 $USDOLLAR = c + a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$
 для прогнозирования с упреждением в две недели

| Dependent Variable: USDOLLAR | | | | |
|--|-------------|------------|-------------|--------|
| Method: Least Squares | | | | |
| Date: 09/13/10 Time: 10:06 | | | | |
| Sample (adjusted): 10/27/1998 6/29/2010 | | | | |
| Included observations: 303 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 1,142902 | 0,056586 | 20,19762 | 0,0000 |
| USDOLLAR(-2) | -0,184180 | 0,055182 | -3,337684 | 0,0010 |
| C | 1,200169 | 0,278022 | 4,316806 | 0,0000 |

| | | | |
|-----------------------|-----------|-----------------------|----------|
| R-squared | 0,969392 | Mean dependent var | 28,13939 |
| Adjusted R-squared | 0,969188 | S.D. dependent var | 2,877809 |
| S.E. of regression | 0,505148 | Akaike info criterion | 1,481922 |
| Sum squared resid | 76,55242 | Schwarz criterion | 1,518692 |
| Log likelihood | -221,5112 | F-statistic | 4750,756 |
| Durbin — Watson stat. | 2,018832 | Prob (F-statistic) | 0,000000 |

Подставив в $USDOLLAR = c + a \times USDOLLAR(-1) - b \times USDOLLAR(-2)$ значения коэффициентов из табл. 7.6, получим следующую формулу:

$$USDOLLAR = 1,2002 + 1,1429 \times USDOLLAR(-1) - 0,1842 \times USDOLLAR(-2), \quad (7.3)$$

где USDOLLAR, USDOLLAR(-1), USDOLLAR(-2) — переменные, обозначающие текущий курс доллара, курс доллара с лагом в две недели и лагом в четыре недели.

Интерпретация уравнения (7.3) следующая: в период с октября 1998 г. по июнь 2010 г. рост на 1 руб. курса доллара с лагом в две недели в среднем приводил к повышению прогнозируемого курса доллара на 1,1429 руб.; в свою очередь рост курса доллара с лагом в четыре недели в среднем приводил к снижению прогнозируемого курса доллара на 0,1842 руб.; при исходном уровне курса доллара, равном 1,2002 руб.

Далее оценим точность полученной статистической модели (см. алгоритм действий № 8 «Как оценить точность статистической модели в EViews»), поместив полученные данные в табл. 7.7. Судя по этой таблице, среднее отклонение по модулю курса доллара от его прогноза за весь период составило всего лишь 28,9 коп., а среднее отклонение по модулю в процентах равняется 1,07%.

Для сравнения напомним, что у стационарной модели $\log(USDollar) = c + b \times \log(USDollar(-1)) + MA(1)$ с оптимизированным временным рядом, делающей прогнозы с упреждением в один месяц, среднее отклонение по модулю курса доллара от его прогноза оказалось равно 41,5 коп., а среднее отклонение по модулю в процентах — 1,53% (см. табл. 6.23).

Поскольку исходный уровень временного ряда оказался стационарным, то при построении статистической модели $USDOLLAR = c + a \times USDOLLAR(-1) - b \times USDOLLAR(-2)$ мы не стали переходить

Таблица 7.7

Оценка точности статистической модели
 $USDOLLAR = c + a \times USDOLLAR(-1) - b \times USDOLLAR(-2)$,
 делающей прогноз с упреждением в две недели

| | |
|---------------------------------------|-----------------|
| Forecast: USDOLLARF | |
| Actual: USDOLLAR | |
| Forecast sample: 10/01/1998 7/13/2010 | |
| Adjusted sample: 10/27/1998 7/13/2010 | |
| Included observations: 304 | |
| Root Mean Squared Error | 0,501828 |
| Mean Absolute Error | 0,289323 |
| Mean Absolute Percentage Error | 1,068366 |
| Theil Inequality Coefficient | 0,008869 |
| Bias Proportion | 0,000000 |
| Variance Proportion | 0,007820 |
| Covariance Proportion | 0,992180 |

к логарифмическому временному ряду. О стационарности исходного временного ряда свидетельствуют итоги тестирования исходного временного ряда на стационарность с помощью расширенного теста Дикки — Фуллера (табл. 7.8). Поскольку в результате нам удалось получить уровень значимости (*Prob.*) одностороннего *t*-критерия (*t-Statistic*), равный нулю, то, следовательно, нулевая гипотеза о нестационарности исходного временного ряда отвергается и принимается альтернативная гипотеза о его стационарности.

Таблица 7.8

Тестирование исходного временного ряда на стационарность
 при построении статистической модели
 $USDOLLAR = c + a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$
 для прогнозирования с упреждением в две недели

Null Hypothesis: USDOLLAR has a unit root

Exogenous: Constant

Lag Length: 2 (Automatic based on SIC, MAXLAG=15)

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey — Fuller test statistic | -3,873740 | 0,0025 |

| | | | |
|---------------------------------------|-----------|-----------|--|
| Test critical values: | 1% level | -3,451847 | |
| | 5% level | -2,870899 | |
| | 10% level | -2,571828 | |
| *MacKinnon (1996) one-sided p-values. | | | |

Чтобы проверить качество полученной статистической стационарной модели, посмотрим, во-первых, как изменяются с увеличением лага автокорреляция и частная автокорреляция в остатках; во-вторых, насколько соответствуют фактические значения корелограммы остатков их теоретическим значениям. Судя по рис. 7.2, по мере роста величины лага уровень автокорреляции постепенно снижается, асимптотически стремясь к нулю, а частная автокорреляция упала почти до нуля уже со второго лага. Если сравнить фактический уровень автокорреляции и частной автокорреляции (вертикальные линии) с их теоретическими значениями (верхняя линия), то они практически не отличаются. Все это свидетельствует о хорошем качестве полученной стационарной модели.

Тестирование на импульсный ответ ARMA-структуры модели $USDOLLAR = c + a \times USDOLLAR(-1) - b \times USDOLLAR(-2)$ так-

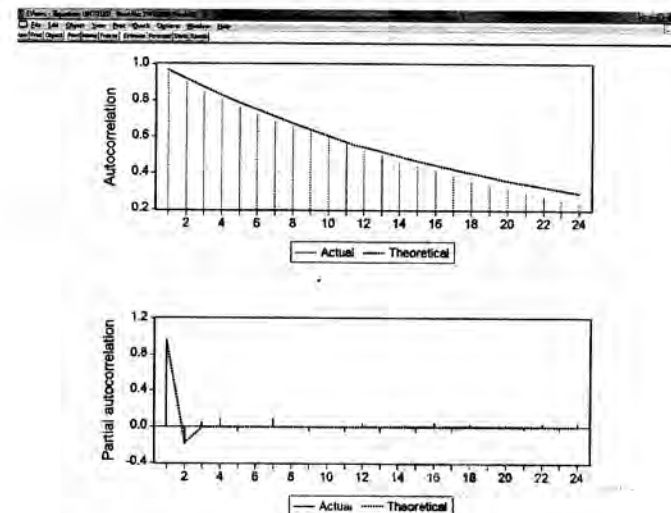


Рис. 7.2. Фактический уровень автокорреляции и частной автокорреляции статистической модели $USDOLLAR = c + a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$ (вертикальные линии) и их теоретические значения (точечная линия)

же показало ее стационарность. Рисунок 7.3 показывает, что величина импульсного ответа — по мере увеличения периодов тестирования на внешние шоки (инновационную неопределенность) — асимптотически стремится к нулю. Если проанализировать динамику накопленного импульсного ответа, то по мере увеличения периодов тестирования его величина стабилизируется на определенном уровне, что также свидетельствует о стационарности построенной статистической модели.

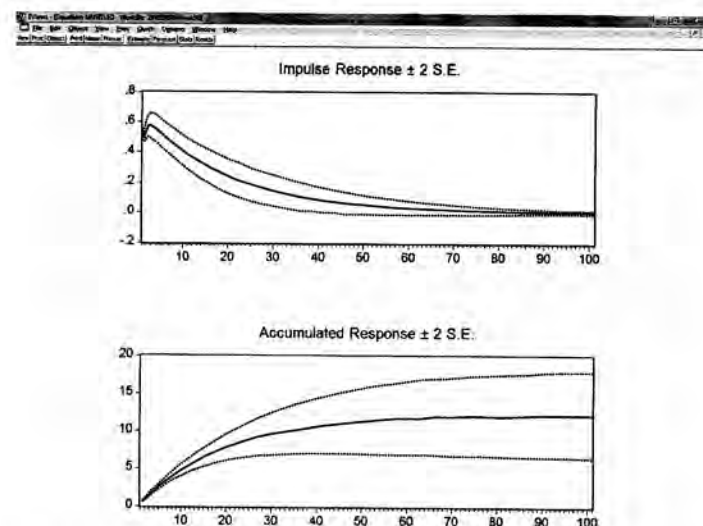


Рис. 7.3. Величина импульсного ответа ARMA-структуры модели $USDOLLAR = c + a \times USDOLLAR(-1) - b \times USDOLLAR(-2)$ по мере увеличения периодов тестирования на внешние шоки (инновационную неопределенность) асимптотически стремится к нулю

Убедившись в достаточно высоком качестве статистической модели $USDOLLAR = c + a \times USDOLLAR(-1) - b \times USDOLLAR(-2)$, мы составили с ее помощью прогноз с упреждением в две недели на 13 июля 2010 г. При этом использовались данные по курсу доллара, взятые с интервалом две недели (на конец этого периода) с октября 1998 г. по 29 июня 2010 г. Согласно полученному точечному прогнозу, курс доллара на 13 июля 2010 г. должен был равняться 30,82 руб., но в действительности американская валюта в этот день стоила 30,88 руб., т.е. ее курс отклонился всего лишь на 6 коп. Таким образом, прогноз курса американской валюты оказался точным при интервальном прогнозе, составленном с 20%-ным уровнем надежности.

Полученная в результате составления прогноза средняя ошибка индивидуального прогнозного значения курса доллара оказалась равна 0,5075 руб. Ее мы использовали для составления рекомендуемых цен покупки и продажи, воспользовавшись алгоритмом действий № 24. При этом для расчета рекомендуемых цен покупки и продажи в качестве среднего значения для нормального распределения был взят фактический курс доллара от 26 июня 2010 г., т.е. его последнее значение перед началом инвестиционного периода, начавшегося с 29 июня и закончившегося 13 июля 2010 г. Рассчитанные нами рекомендуемые цены продажи и покупки доллара на рубли представлены в табл. 7.9.

Таблица 7.9

Рекомендуемые цены продажи и покупки доллара с 29 июня по 13 июля 2010 г.

| Цены покупки и продажи | Цена продажи с уровнем надежности (не ниже), % | | | | Цена покупки с уровнем надежности (не выше), % | | | |
|------------------------|--|---------|---------|---------|--|---------|---------|---------|
| | 90 | 80 | 70 | 60 | 90 | 80 | 70 | 60 |
| Курс реализации сделки | 31,7272 | 31,5035 | 31,3424 | 31,2047 | 30,4250 | 30,6487 | 30,8098 | 30,9475 |

Однако, прежде чем перейти к тестированию по рыночным данным эффективности рекомендуемых цен покупки и продажи, сначала нужно убедиться в их обоснованности. Нужно ответить на вопрос: какая доля цен покупок или продаж, рассчитанных по предложенной методике, в действительности оказалась ниже (или выше) фактического курса доллара в конце месяца?

Чтобы провести эту оценку, нужно рассчитать рекомендуемые курсы покупки и продажи доллара не только на период с 29 июня по 13 июля 2010 г., но и на весь период с октября 1998 г. по июнь 2010 г., а также отклонения рекомендуемых цен от заданного уровня надежности. В таблице 7.10 показан фактический риск того, что рекомендуемая цена продажи валюты, вычисленная с определенным уровнем надежности, в действительности может оказаться ниже курса доллара на конец инвестиционного периода (конец двухнедельного периода).

Судя по табл. 7.10, вероятность удачной сделки при продаже доллара по ценам, рассчитанным с 90%-ным уровнем надежности и при более низких уровнях надежности, выше заданного уровня надежности. Причем при 60–70%-ных уровнях надежности эта положительная разница достигает своего максимума — 15,8–15,9 процентного пункта. Правда, при продаже доллара с 95%-ным и 99%-ными уровнями надежности вероятность удачной сделки несколько ниже установленного уровня надежности.

Таблица 7.10

Уровень надежности при расчете рекомендуемых цен продажи доллара и риск неудачной сделки

| Односторонний t-критерий | Уровень надежности, % | Количество удачных сделок | Вероятность удачной сделки, % | Разница |
|--------------------------|-----------------------|---------------------------|-------------------------------|---------|
| 2,3388 | 99 | 295 | 96,4 | -2,6 |
| 1,6499 | 95 | 287 | 93,8 | -1,2 |
| 1,2844 | 90 | 283 | 92,5 | 2,5 |
| 0,8428 | 80 | 277 | 90,5 | 10,5 |
| 0,5250 | 70 | 263 | 85,9 | 15,9 |
| 0,2536 | 60 | 232 | 75,8 | 15,8 |

В таблице 7.11 показан фактический риск того, что рекомендуемая цена покупки валюты, вычисленная с определенным уровнем надежности, в действительности может оказаться выше курса доллара на конец инвестиционного периода (конец двухнедельного периода). Судя по этой таблице, фактическая вероятность удачной сделки при покупке доллара с 99%-ным уровнем надежности оказалась равна установленному уровню надежности. В то время как при более низких уровнях надежности фактическая вероятность удачной сделки оказалась выше заданного уровня. Причем при 70%-ном уровне надежности эта положительная разница в пользу фактической вероятности удачной сделки достигает своего максимума — 17,3 процентного пункта.

Таблица 7.11

Уровень надежности при расчете рекомендуемых цен покупки доллара и риск неудачной сделки

| Односторонний t-критерий | Уровень надежности, % | Количество удачных сделок | Вероятность удачной сделки, % | Разница |
|--------------------------|-----------------------|---------------------------|-------------------------------|---------|
| 2,3388 | 99 | 303 | 99,0 | 0,0 |
| 1,6499 | 95 | 300 | 98,0 | 3,0 |
| 1,2844 | 90 | 293 | 95,8 | 5,8 |
| 0,8428 | 80 | 282 | 92,2 | 12,2 |
| 0,5250 | 70 | 267 | 87,3 | 17,3 |
| 0,2536 | 60 | 219 | 71,6 | 11,6 |

Сравнив табл. 7.10 и 7.11, легко заметить, что вероятность удачной сделки при покупке доллара несколько выше, чем при его продаже, за исключением 60%-ного уровня надежности, где ситуация прямо противоположная. Теперь проверим, насколько эффективным могло быть использование рассчитанных нами цен покупки и продажи в ходе торгов на валютном рынке с 29 июня по 13 июля 2010 г. Поскольку до 2 июля в динамике курса доллара преобладал растущий тренд (см. рис. 7.3), то первыми в торгах смогли участвовать инвесторы, установившие цену продажи доллара с 60%-ным и 70%-ным уровнями надежности. Однако с 3 июля американская валюта начала падать, поэтому на рынок стали выходить игроки, планировавшие купить доллар с 60%-ным и 70%-ным уровнями надежности.

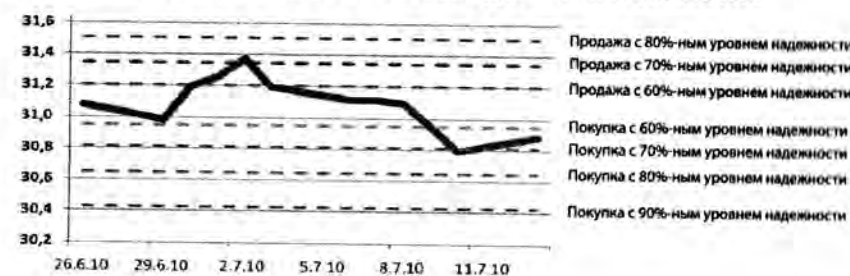


Рис. 7.4. Курс доллара и рекомендуемые цены его покупки и продажи с 29 июня по 13 июля 2010 г.

В таблице 7.12 представлены итоги валютных торгов за период с 29 июня по 13 июля 2010 г. для инвесторов, установивших цены покупки или продажи доллара с разными уровнями надежности. Судя по этой таблице, наиболее высокую курсовую доходность в размере 0,95% по итогам двухнедельного инвестиционного периода получил инвестор, установивший цену продажи доллара с 70%-ным уровнем надежности. На втором месте по этому показателю оказался инвестор, установивший цену продажи доллара с 60%-ным уровнем надежности: доходность — 0,58%. Третье место по доходности досталось инвесторам, установившим цену покупки доллара с 60%-ным и 70%-ным уровнями надежности, поскольку доходность у обоих оказалась равна 0,28%. В то время как инвестор, придерживавшийся стратегии «купил и держи», в течение двух недель понес убытки в размере 0,62%. При этом заметим, что фактическая вероятность удачной сделки для инвестора, придерживавшегося этой стратегии, по нашим подсчетам, за период с октября 1998 г. по июнь 2010 г. составила 52,0% (из 306 сделок 152 были удачными, если вести подсчет доходности на конец каждого инвестиционного периода).

Таблица 7.12
Результаты валютных торгов с 29 июня по 13 июля 2010 г. для инвесторов, установивших цены покупки или продажи доллара с разными уровнями риска

| Дата продажи или покупки | 26 июня 2010 г. | 1 июля 2010 г. | 2 июля 2010 г. | 10 июля 2010 г. | 13 июля 2010 г. | Полученный курсовой доход, в % |
|---|--------------------|-------------------|-------------------|--------------------|--------------------|-----------------------------------|
| Фактический курс доллара | 31,0761 | 31,2554 | 31,3703 | 30,7953 | 30,8823 | — |
| Купил и продал в конце периода | — | — | — | — | —0,62 | 1,58 |
| Продажа по цене с 60%-ным уровнем надежности | — | 31,2554 | — | — | 0,58 | 0,27 |
| Продажа по цене с 70%-ным уровнем надежности | — | — | 31,3703 | — | 0,95 | 1,38 |
| Продажа по цене с 80%-ным уровнем надежности | — | — | — | — | 0,0 | 2,35 |
| Продажа по цене с 90%-ным уровнем надежности | — | — | — | — | 0,0 | 0,00 |
| Покупка по цене с 60%-ным уровнем надежности | — | — | — | 30,7953 | 0,28 | 2,32 |
| Покупка по цене с 70%-ным уровнем надежности | — | — | — | 30,7953 | 0,28 | 0,00 |
| Покупка по цене с 80%-ным уровнем надежности | — | — | — | — | 0,0 | 0,00 |
| Покупка по цене с 90%-ным уровнем надежности | — | — | — | — | 0,0 | 0,00 |

7.3. Использование в торговле модели для прогнозирования курса доллара к рублю с упреждением в одну неделю

Сейчас рассмотрим, насколько эффективна для использования в валютных торгах статистическая модель, по которой можно делать прогноз по курсу доллара к рублю с упреждением в одну неделю. При этом на полном изложении процедуры построения такой прогностической модели особо останавливаться не будем, а расскажем о наиболее важных особенностях этой модели.

На основе базы данных по курсу доллара, взятых с интервалом в одну неделю (на конец периода) с 1 октября 1998 г. по 20 июля 2010 г., нами была построена прогностическая модель, по которой можно делать прогнозы с упреждением в одну неделю. Вывод данных по итогам решения уравнения регрессии можно увидеть в табл. 7.13.

Таблица 7.13

Вывод данных по статистической модели
 $USDOLLAR = c + a \times USDOLLAR(-1)$
 для прогнозирования с упреждением в одну неделю

| Dependent Variable: USDOLLAR | | | | |
|--|-------------|-----------------------|-------------|----------|
| Method: Least Squares | | | | |
| Date: 09/13/10 Time: 17:50 | | | | |
| Sample (adjusted): 10/06/1998 7/20/2010 | | | | |
| Included observations: 611 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| USDOLLAR(-1) | 0,979579 | 0,004944 | 198,1178 | 0,0000 |
| C | 0,597035 | 0,139541 | 4,278574 | 0,0000 |
| R-squared | 0,984721 | Mean dependent var | | 28,08436 |
| Adjusted R-squared | 0,984696 | S.D. dependent var | | 2,978022 |
| S.E. of regression | 0,368405 | Akaike info criterion | | 0,844003 |
| Sum squared resid | 82,65506 | Schwarz criterion | | 0,858455 |
| Log likelihood | -255,8429 | F-statistic | | 39250,65 |
| Durbin — Watson stat. | 2,043630 | Prob (F-statistic) | | 0,000000 |

Подставив в $USDOLLAR = c + a \times USDOLLAR(-1)$ коэффициенты из табл. 7.13, получим следующую формулу:

$$USDOLLAR = 0,5970 + 0,9796 \times USDOLLAR(-1), \quad (7.4)$$

где $USDOLLAR$, $USDOLLAR(-1)$ — переменные, обозначающие текущий курс доллара и курс доллара с лагом в одну неделю.

Интерпретация уравнения (7.4) следующая: в период с 1 октября 1998 г. по 20 июня 2010 г. рост на 1 руб. курса доллара с лагом в одну неделю в среднем приводил к повышению прогнозируемого курса доллара на 0,9796 руб. при исходном уровне курса доллара, равном 0,5970 руб.

Теперь оценим точность полученной статистической модели (см. алгоритм действий № 8 «Как оценить точность статистической модели в EViews»), поместив результаты этой оценки в табл. 7.14. Судя по этой таблице, среднее отклонение по модулю курса доллара от его прогноза за весь период составило 19,7 коп., а среднее отклонение по модулю в процентах равняется 0,72%. В то время как у модели $USDOLLAR = c + a \times USDOLLAR(-1) + b \times USDOLLAR(-2)$, делающей прогнозы с упреждением в две недели, среднее отклонение по модулю курса доллара от его прогноза оказалось равно 28,9 коп., а среднее отклонение по модулю в процентах — 1,07% (см. табл. 7.7).

Таблица 7.14

Оценка точности статистической модели
 $USDOLLAR = c + a \times USDOLLAR(-1)$,
 делающей прогноз с упреждением в одну неделю

| | |
|---------------------------------------|----------|
| Forecast: USDOLLARF | |
| Actual: USDOLLAR | |
| Forecast sample: 10/01/1998 7/27/2010 | |
| Adjusted sample: 10/06/1998 7/27/2010 | |
| Included observations: 612 | |
| Root Mean Squared Error | 0,367636 |
| Mean Absolute Error | 0,196738 |
| Mean Absolute Percentage Error | 0,724011 |
| Theil Inequality Coefficient | 0,006508 |
| Bias Proportion | 0,000001 |
| Variance Proportion | 0,003736 |
| Covariance Proportion | 0,996263 |

Поскольку модель, построенная с использованием исходного уровня временного ряда, оказалась стационарной, то в этом случае мы не стали переходить к логарифмическому временному ряду. О стационарности исходного временного ряда свидетельствуют итоги тестирования исходного временного ряда на стационарность с помощью расширенного теста Дикки — Фуллера (табл. 7.15). При этом уровень значимости (*Prob. **) одностороннего *t*-критерия (*t*-Statistic) получился весьма близким к нулю, а потому нулевая гипотеза о нестационарности исходного временного ряда отвергается и принимается альтернативная гипотеза о его стационарности.

Таблица 7.15

Тестирование исходного временного ряда на стационарность
 при построении статистической модели
 $USDOLLAR = c + a \times USDOLLAR(-1)$
 для прогнозирования с упреждением в одну неделю

| | | |
|---|-------------|-----------|
| Null Hypothesis: USDOLLAR has a unit root | | |
| Exogenous: Constant | | |
| Lag Length: 3 (Automatic based on SIC, MAXLAG = 18) | | |
| | t-Statistic | Prob.* |
| Augmented Dickey — Fuller test statistic | −3,661206 | 0,0049 |
| Test critical values: | 1% level | −3,440876 |
| | 5% level | −2,866075 |
| | 10% level | −2,569244 |
| *MacKinnon (1996) one-sided p-values. | | |

Для проверки качества этой статистической модели посмотрим, во-первых, как изменяется с увеличением лага автокорреляция и частная автокорреляция в остатках, во-вторых, насколько соответствуют фактические значения коррелограммы остатков их теоретическим значениям. Судя по рис. 7.5, по мере роста величины лага уровень автокорреляции постепенно снижается, асимптотически стремясь к нулю, а частная автокорреляция упала почти до нуля уже со второго лага. Если сравнить фактический уровень автокорреляции и частной автокорреляции (нижние вертикальные линии) с их теоретическими значениями (верхняя линия), то можно заметить, что они в основном совпадают. Вместе с тем следует отметить, что фактический уровень автокорреляции после пятого лага заметно ниже ее теоретического уровня, а фактический уровень частной автокорреляции после 10-го лага в некоторых случаях выше теоретического нулевого уровня.

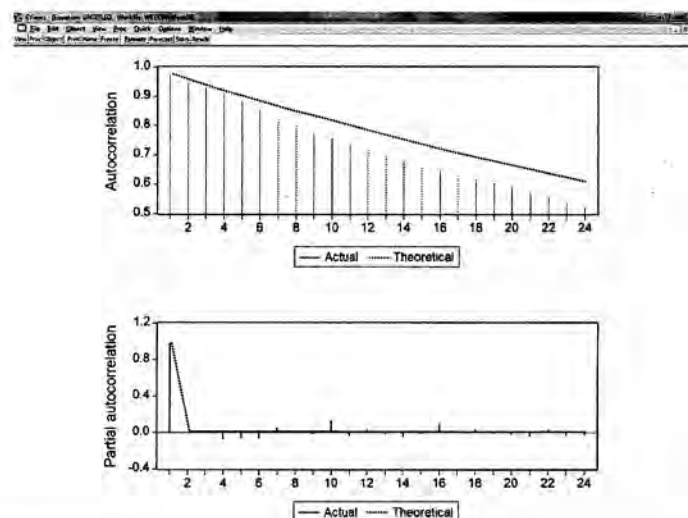


Рис. 7.5. Зависимость автокорреляции и частной автокорреляции в остатках с увеличением временного лага для статистической модели $USDOLLAR = c + a \times USDOLLAR(-1)$ с упреждением в одну неделю

Как и коррелограмма, тестирование на импульсный ответ ARMA-структуры модели $USDOLLAR = c + a \times USDOLLAR(-1)$ также показало ее стационарность. Рисунок 7.6 показывает, что величина импульсного ответа по мере увеличения периодов тестирования на внешние шоки (инновационную неопределенность) асимптотически стремится к нулю. Если проанализировать динамику накопленного импульсного ответа, то по мере увеличения периодов тестирования его величина стабилизируется на определенном уровне, что также свидетельствует о стационарности построенной статистической модели. Правда, происходит это довольно медленно, поэтому, чтобы наглядно продемонстрировать последнюю тенденцию, пришлось увеличить время тестирования до 200 периодов.

Убедившись в адекватности статистической модели $USDOLLAR = c + a \times USDOLLAR(-1)$, составим с ее помощью прогноз с упреждением в одну неделю на 27 июля 2010 г. При этом используем данные по курсу доллара, взятые с интервалом в одну неделю (на конец этого периода) с 1 октября 1998 г. по 20 июня 2010 г. Согласно полученному точечному прогнозу, курс доллара на 27 июля 2010 г. должен был равняться 30,55 руб., хотя в действительности американская валюта в этот день стоила 30,30 руб., т.е. ее курс отклонился на 25 коп. Таким образом, прогноз курса американской валюты оказался точным при интервальном прогнозе, составленном с 50%-ным уровнем надежности.

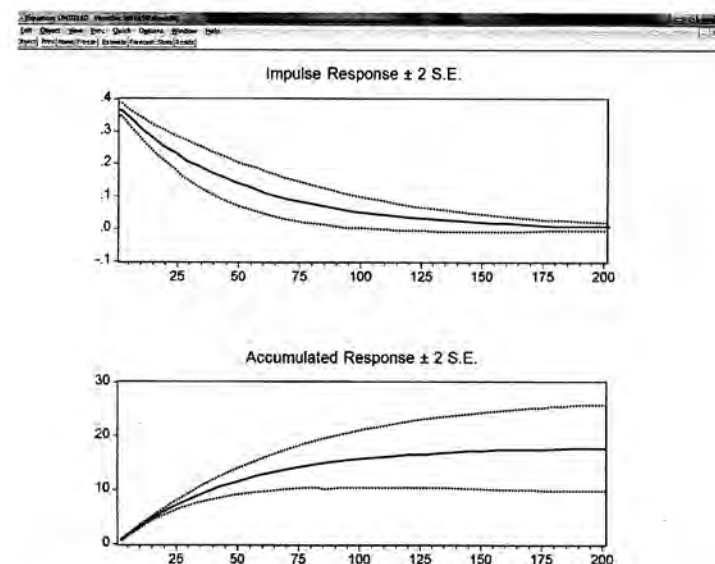


Рис. 7.6. Графики импульсного и накопленного импульсного ответов ARMA-структуры модели $USDOLLAR = c + a \times USDOLLAR(-1)$

Полученная в результате составления прогноза средняя ошибка индивидуального прогнозного значения курса доллара оказалась равна 0,3689 руб. Ее мы применили для формирования рекомендуемых цен покупки и продажи, воспользовавшись алгоритмом действий № 24. Для расчета рекомендуемых цен покупки и продажи в качестве среднего значения при нормальном распределении был взят фактический курс доллара от 17 июля 2010 г., равный 30,46 руб., т.е. его последнее значение перед началом инвестиционного периода, начавшегося с 20 июля и закончившегося 27 июля 2010 г. Рассчитанные нами рекомендуемые цены продажи и покупки представлены в табл. 7.16.

Таблица 7.16
Рекомендуемые цены продажи и покупки с 20 июля по 27 июля 2010 г.

| Цены покупки и продажи | Цена продажи с уровнем надежности (не ниже), % | | | | Цена покупки с уровнем надежности (не выше), % | | | |
|------------------------|--|---------|---------|---------|--|---------|---------|---------|
| | 90 | 80 | 70 | 60 | 90 | 80 | 70 | 60 |
| Курс реализации сделки | 30,9348 | 30,7722 | 30,6551 | 30,5550 | 29,9882 | 30,1508 | 30,2679 | 30,3680 |

Прежде чем перейти к тестированию по рыночным данным эффективности рекомендуемых цен покупки и продажи, необходимо сначала убедиться в их обоснованности. Это означает, что нужно рассмотреть, какая доля цен покупки или продажи, рассчитанных по указанной методике, в действительности оказалась ниже (или выше) фактического курса доллара в конце месяца.

С этой целью мы рассчитали рекомендуемые курсы покупки и продажи доллара не только на период с 20 июля по 27 июля 2010 г., но и на весь период с 1 октября 1998 г. по 20 июля 2010 г., а затем сопоставили отклонения рекомендуемых цен от заданного уровня надежности. В таблице 7.17 показан фактический риск того, что рекомендуемая цена продажи валюты, вычисленная с определенным уровнем надежности, в действительности может оказаться ниже курса доллара на конец инвестиционного периода (конец недели). Из этой таблицы можно сделать вывод, что вероятность удачной сделки при продаже доллара по ценам, рассчитанным с 95%-ным уровнем надежности и при более низких уровнях надежности, выше заданного уровня надежности. Причем при 60%-ном уровне надежности эта положительная разница достигает своего максимума — 16,8 процентного пункта. Правда, при продаже доллара с 99%-ным уровнем надежности вероятность удачной сделки несколько ниже установленного уровня надежности.

Таблица 7.17

Уровень надежности при расчете рекомендуемых цен продажи доллара и риск неудачной сделки

| Односторонний <i>t</i> -критерий | Уровень надежности, % | Количество удачных сделок | Вероятность удачной сделки, % | Разница |
|----------------------------------|-----------------------|---------------------------|-------------------------------|---------|
| 2,3388 | 99 | 595 | 97,4 | -1,6 |
| 1,6499 | 95 | 587 | 96,1 | 1,1 |
| 1,2844 | 90 | 579 | 94,8 | 4,8 |
| 0,8428 | 80 | 563 | 92,1 | 12,1 |
| 0,5250 | 70 | 526 | 86,1 | 16,1 |
| 0,2536 | 60 | 469 | 76,8 | 16,8 |

В таблице 7.18 показан фактический риск того, что рекомендуемая цена покупки валюты, вычисленная с определенным уровнем надежности, в действительности может оказаться выше курса доллара на конец инвестиционного периода (в конце недели). Судя по этой таблице, фактическая вероятность удачной сделки оказалась выше установленного уровня надежности. Правда, при 99%-ном уровне надежности эта положительная разница оказалась весьма незначительной, в то время

как при более низких уровнях надежности фактическая вероятность удачной сделки быстро нарастает. Причем при 70%-ном уровне надежности эта положительная разница в пользу фактической вероятности удачной сделки достигает своего максимума — 19,0 процентных пунктов.

Таблица 7.18

Уровень надежности при расчете рекомендуемых цен продажи доллара и риск неудачной сделки

| Односторонний <i>t</i> -критерий | Уровень надежности, % | Количество удачных сделок | Вероятность удачной сделки, % | Разница |
|----------------------------------|-----------------------|---------------------------|-------------------------------|---------|
| 2,3388 | 99 | 606 | 99,2 | 0,2 |
| 1,6499 | 95 | 599 | 98,0 | 3,0 |
| 1,2844 | 90 | 593 | 97,1 | 7,1 |
| 0,8428 | 80 | 573 | 93,8 | 13,8 |
| 0,5250 | 70 | 544 | 89,0 | 19,0 |
| 0,2536 | 60 | 445 | 72,8 | 12,8 |

Сравнив табл. 7.17 и 7.18, легко заметить, что вероятность удачной сделки при покупке доллара несколько выше, чем при его продаже, за исключением 60%-ного уровня надежности, где она, напротив, несколько меньше. Теперь проверим, насколько эффективным могло быть использование рассчитанных нами цен покупки и продажи в ходе торгов на валютном рынке с 20 июля по 27 июля 2010 г.

Поскольку 20 июля курс доллара значительно вырос (рис. 7.7), то первым в торгах смог участвовать инвестор, установивший цену продажи

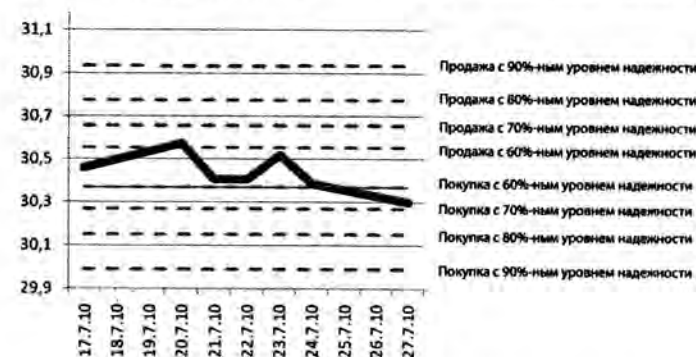


Рис. 7.7. Курс доллара и рекомендуемые цены его покупки и продажи с 20 июля по 27 июля 2010 г.

доллара с 60%-ным уровнем надежности. Однако затем американская валюта стала преимущественно падать, поэтому на рынок смог выйти игрок, планировавший купить доллар по цене с 60%-ным уровнем надежности.

В таблице 7.19 представлены итоги валютных торгов за период с 20 июля по 27 июля 2010 г. для инвесторов, установивших цены покупки или продажи доллара с разными уровнями надежности. Судя по этой таблице, положительную курсовую доходность в размере 0,37% по итогам недельного инвестиционного периода получил инвестор, установивший цену продажи доллара с 60%-ным уровнем надежности, в то время как инвестор, установивший цену покупки доллара с 60%-ным уровнем надежности, еще не успел заработать на своей покупке, поскольку приобрел валюту в последний день инвестиционного периода. В свою очередь инвестор, придерживавшийся стратегии «купил и держи», понес убытки в размере 0,53%. Заметим, что фактическая вероятность удачной сделки для инвестора, придерживавшегося этой стратегии, по нашим подсчетам, за период с 1 октября 1998 г. по 20 июля 2010 г. составила 48,5% (из 612 сделок 297 были удачными, если вести подсчет доходности на конец каждого инвестиционного периода).

7.4. Использование в торговле модели для прогнозирования курса евро к доллару с упреждением в один день

До сих пор мы делали прогнозы относительно курса доллара к рублю. А теперь попробуем оценить, насколько эффективно будет использование в торгах статистической модели, по которой можно делать прогноз по курсу евро к доллару с упреждением в один день. На основе базы данных по курсу евро к доллару, взятых с интервалом в один день (цена закрытия) с 5 января 1999 г. по 13 сентября 2010 г., нами была построена прогностическая модель, по которой можно делать прогнозы с упреждением в один день. Поскольку ARMA-структура статистической модели, полученная по исходному временному ряду, оказалась нестационарной, мы решили построить ее на основе логарифмического временного ряда. Данные по итогам решения уравнения регрессии, полученного из логарифмированного временного ряда, можно увидеть в табл. 7.20. Поскольку коэффициент a получился меньше единицы (хотя эта разница и незначительна), то можно говорить о стационарной ARMA-структуре этой модели.

Таблица 7.19

Результаты валютных торгов с 20 по 27 июля 2010 г. для инвесторов, установивших цены покупки или продажи доллара с разными уровнями риска

| Дата продажи или покупки | 17 июля 2010 г. | 20 июля 2010 г. | 27 июля 2010 г. | Полученный курсовой доход, в % |
|--|-----------------|-----------------|-----------------|--------------------------------|
| Фактический курс доллара | 30,4615 | 30,5739 | 30,3006 | - |
| Покупка в начале и продажа в конце периода | - | - | - | -0,53 |
| Продажа по цене с 60%-ным уровнем надежности | - | 30,5739 | - | 0,37 |
| Продажа по цене с 70%-ным уровнем надежности | - | - | - | 0,0 |
| Продажа по цене с 80%-ным уровнем надежности | - | - | - | 0,0 |
| Продажа по цене с 90%-ным уровнем надежности | - | - | - | 0,0 |
| Покупка по цене с 60%-ным уровнем надежности | - | - | 30,3006 | 0,0 |
| Покупка по цене с 70%-ным уровнем надежности | - | - | - | 0,0 |
| Покупка по цене с 80%-ным уровнем надежности | - | - | - | 0,0 |
| Покупка по цене с 90%-ным уровнем надежности | - | - | - | 0,0 |

Таблица 7.20

Вывод данных по статистической модели
 $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$
 для прогнозирования с упреждением в один день

| Dependent Variable: LOG(EURUSD) | | | | |
|---|-------------|-----------------------|-------------|-----------|
| Method: Least Squares | | | | |
| Date: 09/21/10 Time: 14:14 | | | | |
| Sample (adjusted): 1/05/1999 9/13/2010 | | | | |
| Included observations: 2994 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| LOG(EURUSD(-1)) | 0,999686 | 0,000530 | 1886,272 | 0,0000 |
| R-squared | 0,998452 | Mean dependent var | | 0,155595 |
| Adjusted R-squared | 0,998452 | S.D. dependent var | | 0,169631 |
| S.E. of regression | 0,006674 | Akaike info criterion | | -7,180916 |
| Sum squared resid | 0,133308 | Schwarz criterion | | -7,178911 |
| Log likelihood | 10750,83 | Durbin - Watson stat. | | 1,994655 |

Источник: расчеты автора по данным Европейского центрального банка.

Подставив в $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$ коэффициенты из табл. 7.20, получим следующую формулу:

$$\log(\text{EURUSD}) = 0,9996 \times \log(\text{EURUSD}(-1)), \quad (7.5)$$

где EURUSD, EURUSD(-1) — переменные, обозначающие текущий курс евро к доллару и курс евро к доллару с лагом в один день.

Однако интерпретация формулы (7.5) не столь очевидна, поскольку она относится к логарифмическому ряду. Поэтому с помощью потенцирования этой формулы можно перейти от логарифмов к исходному временному ряду, как мы это уже делали при преобразовании формулы (6.5) в формулу (6.6). В результате исходная линейная функция (7.5), решенная относительно логарифмического временного ряда, станет степенной функцией, которую можно применять относительно исходного временного ряда:

$$\text{EURUSD} = \text{EURUSD}(-1)^{0,9996}. \quad (7.6)$$

При этом интерпретация формулы (7.6) будет следующей: в период с 5 января 1999 г. по 13 сентября 2010 г. рост на 1% курса евро к доллару в предыдущем торговом дне в среднем способствовал повышению курса евро к доллару в следующем торговом дне на 0,9996%.

Далее оценим точность полученной статистической модели (см. алгоритм действий № 8 «Как оценить точность статистической модели в EViews»), поместив результаты этой оценки в табл. 7.21. Судя по этой таблице, среднее отклонение по модулю курса евро к доллару от его прогноза за весь период составило лишь 0,58 цента, а среднее отклонение по модулю в процентах равняется 0,50%.

Таблица 7.21

Оценка точности статистической модели
 $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$,
 делающей прогноз с упреждением в один день

| | |
|--------------------------------------|----------|
| Forecast: EURUSDF | |
| Actual: EURUSD | |
| Forecast sample: 1/04/1999 9/14/2010 | |
| Adjusted sample: 1/05/1999 9/14/2010 | |
| Included observations: 2995 | |
| Root Mean Squared Error | 0,007943 |
| Mean Absolute Error | 0,005837 |
| Mean Absolute Percentage Error | 0,497382 |
| Theil Inequality Coefficient | 0,003307 |
| Bias Proportion | 0,000170 |
| Variance Proportion | 0,000100 |
| Covariance Proportion | 0,999730 |

Теперь посмотрим, является ли стационарным логарифмический временной ряд, на основе которого построена наша статистическая модель. С этой целью проведем тестирование логарифмического временного ряда с помощью расширенного теста Дикки — Фуллера (табл. 7.22). При этом уровень значимости (*Prob.* *) одностороннего *t*-критерия получился равным 0,2908, а потому нулевая гипотеза о нестационарности логарифмического временного ряда не отвергается.

Таким образом, мы получили статистическую модель со стационарной ARMA-структурой, построенной на основе нестационарного

Таблица 7.22

Тестирование логарифмического временного ряда на стационарность
при построении статистической модели
 $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$
для прогнозирования с упреждением в один день

| | | | | |
|---|-------------|-----------------------|-------------|-----------|
| Null Hypothesis: EURUSD1 has a unit root | | | | |
| Exogenous: Constant, Linear Trend | | | | |
| Lag Length: 0 (Automatic based on SIC, MAXLAG = 28) | | | | |
| | | | t-Statistic | Prob.* |
| Augmented Dickey — Fuller test statistic | | | -2,577379 | 0,2908 |
| Test critical values: | 1% level | | -3,961107 | |
| | 5% level | | -3,411308 | |
| | 10% level | | -3,127496 | |
| *MacKinnon (1996) one-sided p-values. | | | | |
| Augmented Dickey — Fuller Test Equation | | | | |
| Dependent Variable: D(EURUSD1) | | | | |
| Method: Least Squares | | | | |
| Date: 09/24/10 Time: 15:41 | | | | |
| Sample (adjusted): 1/05/1999 9/13/2010 | | | | |
| Included observations: 2994 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| EURUSD1(-1) | -0,003387 | 0,001314 | -2,577379 | 0,0100 |
| C | -0,000380 | 0,000271 | -1,399129 | 0,1619 |
| @TREND(1/04/1999) | 6,24E-07 | 2,58E-07 | 2,418451 | 0,0156 |
| R-squared | 0,002292 | Mean dependent var | | 2,75E-05 |
| Adjusted R-squared | 0,001624 | S.D. dependent var | | 0,006674 |
| S.E. of regression | 0,006669 | Akaike info criterion | | -7,181774 |
| Sum squared resid | 0,133015 | Schwarz criterion | | -7,175758 |
| Log likelihood | 10754,12 | F-statistic | | 3,434905 |
| Durbin — Watson stat | 1,992902 | Prob (F-statistic) | | 0,032356 |

логарифмического временного ряда. Посмотрим, получим ли мы в результате стационарные остатки, что весьма важно для получения надежных прогнозов. С этой целью проведем с помощью расширенного теста Дикки — Фуллера тестирование остатков, полученных после решения уравнения регрессии $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$. Судя по табл. 7.23, можно сделать вывод, что мы получили стационарные остатки, поскольку уровень значимости теста (Prob.) оказался равен нулю.

Таблица 7.23

Тестирование остатков, полученных после решения уравнения регрессии
 $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$

| | | | |
|--|-----------|-------------|--------|
| Null Hypothesis: D(RESID) has a unit root | | | |
| Exogenous: None | | | |
| Lag Length: 18 (Automatic based on SIC, MAXLAG = 28) | | | |
| | | t-Statistic | Prob.* |
| Augmented Dickey — Fuller test statistic | | -22,28603 | 0,0000 |
| Test critical values: | 1% level | -2,565740 | |
| | 5% level | -1,940930 | |
| | 10% level | -1,616628 | |

*MacKinnon (1996) one-sided p-values.

Для проверки качества модели $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$ посмотрим, во-первых, как изменяется с увеличением лага автокорреляция и частная автокорреляция в остатках, во-вторых, насколько соответствуют фактические значения коррелограммы остатков их теоретическим значениям. Судя по рис. 7.8, по мере роста величины лага уровень автокорреляции постепенно снижается, асимптотически стремясь к нулю, а частная автокорреляция падает почти до нуля, начиная со 2-го лага. Правда, при этом фактический уровень автокорреляции (нижние вертикальные линии) падает гораздо быстрее его теоретических значений (верхняя пологая точечная линия). Что же касается динамики фактического уровня частной автокорреляции (вертикальные линии), то она практически совпадает с его теоретическими значениями (верхняя точечная линия).

Тестирование на импульсный ответ ARMA-структуры модели $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$ хотя и показало ее стационар-

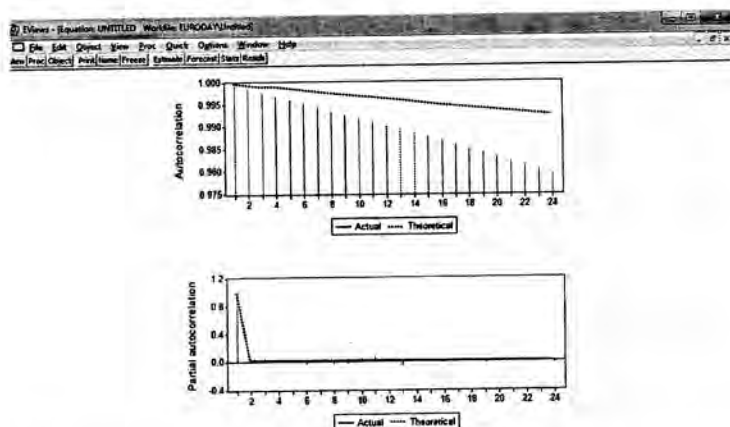


Рис. 7.8. Зависимость автокорреляции и частной автокорреляции в остатках с увеличением временного лага для статистической модели $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$ с упреждением в один день

ность, однако вместе с тем выявило тот факт, что для достижения стабильности модели требуется весьма длительное тестовое время. Как хорошо видно из рис. 7.9, величина импульсного ответа и величина накопленного импульсного ответа по мере увеличения периодов тестирования на внешние шоки (инновационную неопределенность) асимптотически стремятся: первая — к нулю, а вторая — к опреде-

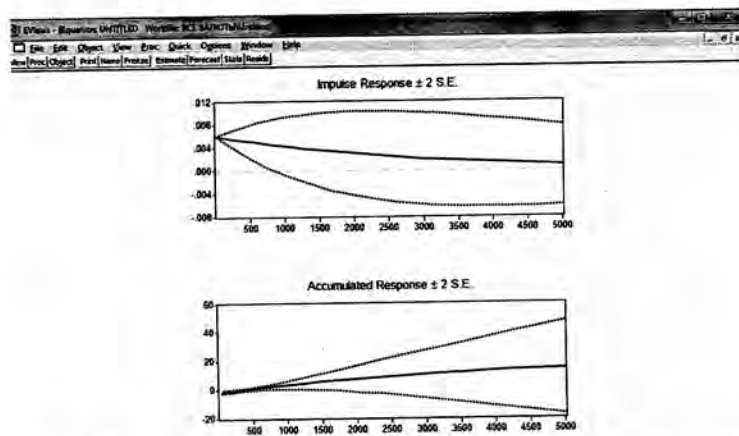


Рис. 7.9. Графики импульсного и накопленного импульсного ответов ARMA-структуры модели $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$

ленному пределу. Однако, чтобы показать на рисунке обе эти тенденции, мы были вынуждены увеличить время тестирования до 5000 периодов.

Убедившись, что статистическая модель $\log(\text{EURUSD}) = a \times \log(\text{EURUSD}(-1))$ в целом адекватна, составим с ее помощью прогноз с упреждением в один день на 14 сентября 2010 г. При этом используются данные курса евро к доллару, взятые с интервалом в один день (цена закрытия) с 5 января 1999 г. по 13 сентября 2010 г. Согласно полученному точечному прогнозу, курс евро на 14 сентября 2010 г. должен был равняться 1,2800 дол., хотя в действительности единая европейская валюта в этот день стоила 1,2850 дол., т.е. ее курс отклонился на 0,5 цента. Таким образом, прогноз курса евро к доллару оказался точным при интервальном прогнозе, составленном с 50%-ным уровнем надежности.

Полученная в результате составления прогноза средняя ошибка индивидуального прогнозного значения курса евро оказалась равна 0,0085, дол., или 0,85 цента. Ее мы использовали для составления рекомендуемых цен покупки и продажи, воспользовавшись алгоритмом действий № 24. При этом для расчета рекомендуемых цен покупки и продажи в качестве среднего значения для нормального распределения был взят фактический курс евро на конец 13 сентября 2010 г. (точнее сказать, на 24 часа по GMT, т.е. на 24 часа по Гринвичу), равный 1,2997 дол.

На основе этих данных мы составили рекомендуемые цены покупки и продажи евро относительно к доллару на 14 сентября, при этом торги велись с интервалом в один час по ценам открытия. Рассчитанные нами рекомендуемые цены продажи и покупки представлены в табл. 7.24.

Таблица 7.24

Рекомендуемые цены продажи и покупки евро на 14 сентября 2010 г.

| Цены покупки и продажи | Цена продажи с уровнем надежности (не ниже), % | | | | Цена покупки с уровнем надежности (не выше), % | | | |
|------------------------|--|--------|--------|--------|--|--------|--------|--------|
| | 90 | 80 | 70 | 60 | 90 | 80 | 70 | 60 |
| Курс реализации сделки | 1,3107 | 1,3069 | 1,3042 | 1,3019 | 1,2887 | 1,2925 | 1,2952 | 1,2975 |

Прежде чем перейти к тестированию по рыночным данным эффективности рекомендуемых цен покупки и продажи, необходимо сначала убедиться в их обоснованности. Нужно посмотреть, какая доля цен покупки или продажи, рассчитанных по этой методике, оказалась ниже (или выше) фактического курса евро к доллару в конце торгового дня.

С этой целью мы рассчитали рекомендуемые курсы покупки и продажи евро за период с 5 января 1999 г. по 13 сентября 2010 г., а затем отклонения рекомендуемых цен от заданного уровня надежности. В таблице 7.25 показан фактический риск того, что рекомендуемая цена продажи единой европейской валюты, вычисленная с определенным уровнем надежности, в действительности может оказаться ниже курса евро к доллару на конец торгового дня (цена закрытия).

Из таблицы 7.25 можно сделать вывод, что вероятность удачной сделки при продаже евро по ценам, рассчитанным с 95%-ным уровнем надежности и при более низких уровнях надежности, выше заданного уровня надежности. При 60%-ном уровне надежности эта положительная разница достигает своего максимума — 4,1 процентного пункта. Правда, при продаже евро с 99%-ным уровнем надежности вероятность удачной сделки чуть ниже установленного уровня надежности.

Всего проведено 2996 наблюдений.

Таблица 7.25

Уровень надежности при расчете рекомендуемых цен продажи евро в долларах и риск неудачной сделки

| Односторонний <i>t</i> -критерий | Уровень надежности, % | Количество удачных сделок | Вероятность удачной сделки, % | Разница |
|----------------------------------|-----------------------|---------------------------|-------------------------------|---------|
| 2,3276 | 99 | 2952 | 98,5 | -0,5 |
| 1,6454 | 95 | 2863 | 95,6 | 0,6 |
| 1,2818 | 90 | 2744 | 91,6 | 1,6 |
| 0,8417 | 80 | 2480 | 82,8 | 2,8 |
| 0,5245 | 70 | 2215 | 73,9 | 3,9 |
| 0,2534 | 60 | 1920 | 64,1 | 4,1 |

В таблице 7.26 показан фактический риск того, что рекомендуемая цена покупки евро валюты, вычисленная с определенным уровнем надежности, в действительности может оказаться выше ее курса на конец торгового дня. Судя по этой таблице, фактическая вероятность удачной сделки оказалась выше установленного уровня надежности при 95%-ном и более низких уровнях надежности. Причем при 70%-ном уровне надежности эта положительная разница в пользу фактической вероятности удачной сделки достигает своего максимума — 4,3 процентного пункта. Правда, при 99%-ном уровне фактический уровень надежности оказался немного ниже заданного уровня надежности.

Таблица 7.26

Уровень надежности при расчете рекомендуемых цен покупки евро относительно доллара и риск неудачной сделки

| Односторонний <i>t</i> -критерий | Уровень надежности, % | Количество удачных сделок | Вероятность удачной сделки, % | Разница |
|----------------------------------|-----------------------|---------------------------|-------------------------------|---------|
| 2,3276 | 99 | 2955 | 98,6 | -0,4 |
| 1,6454 | 95 | 2874 | 95,9 | 0,9 |
| 1,2818 | 90 | 2749 | 91,8 | 1,8 |
| 0,8417 | 80 | 2469 | 82,4 | 2,4 |
| 0,5245 | 70 | 2227 | 74,3 | 4,3 |
| 0,2534 | 60 | 1899 | 63,4 | 3,4 |

Сравнив табл. 7.25 и 7.26, нетрудно прийти к выводу, что вероятности удачной сделки как при покупке, так и при продаже евро относительно доллара друг от друга практически не отличаются. В то время как вероятность удачной сделки при покупке доллара на рубли, как правило, несколько выше, чем при его продаже. Очевидно, это объясняется тем, что уровень инфляции в США и странах еврозоны является вполне сопоставимым, а также гораздо меньшим вмешательством Федеральной резервной системы США и Европейского центрального банка (по сравнению с Банком России) в процесс регулирования курса своих валют.

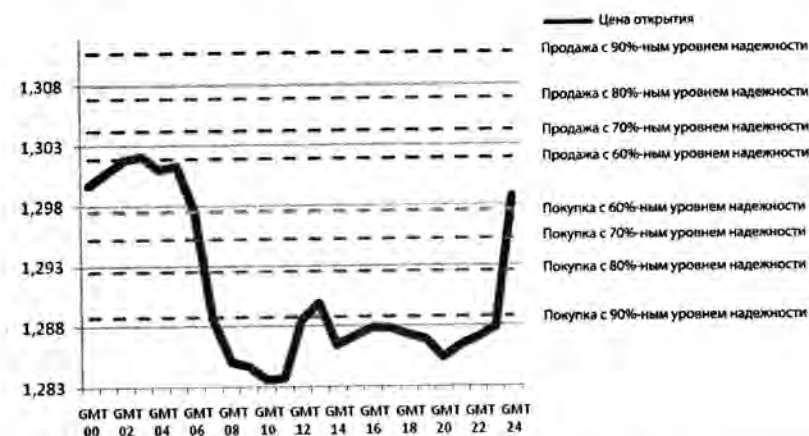


Рис. 7.10. Курс евро к доллару и цены его продажи и покупки 14 сентября 2010 г.
Источник: расчеты автора по данным <http://www.fxstreet.com>.

Таблица 7.27

Результаты валютных торгов 14 сентября 2010 г. для инвесторов, установивших цены покупки или продажи евро к доллару с разными уровнями риска

| Время продажи или покупки | GMT 00 | GMT 03 | GMT 07 | GMT 24 | Полученный курсовой доход, в % |
|--|--------|--------|--------|--------|--------------------------------|
| Фактический курс евро к доллару | 1,2997 | 1,3021 | 1,2886 | 1,2987 | — |
| Купил в начале и продал в конце периода | — | — | — | — | -0,08 |
| Продажа по цене с 60%-ным уровнем надежности | — | 1,3021 | — | — | 0,18 |
| Продажа по цене с 70%-ным уровнем надежности | — | — | — | — | 0,00 |
| Продажа по цене с 80%-ным уровнем надежности | — | — | — | — | 0,00 |
| Продажа по цене с 90%-ным уровнем надежности | — | — | — | — | 0,00 |
| Покупка по цене с 60%-ным уровнем надежности | — | — | 1,2886 | — | 0,78 |
| Покупка по цене с 70%-ным уровнем надежности | — | — | 1,2886 | — | 0,78 |
| Покупка по цене с 80%-ным уровнем надежности | — | — | 1,2886 | — | 0,78 |
| Покупка по цене с 90%-ным уровнем надежности | — | — | 1,2886 | — | 0,78 |

Источник: расчеты автора по данным <http://www.fxstreet.com>.

Примечание: GMT — среднее время по Гринвичу.

Поскольку 14 сентября с нуля часов до трех часов утра по Гринвичу курс евро к доллару довольно устойчиво рос, то первым в торгах смог участвовать инвестор, установивший цену продажи евро с 60%-ным уровнем надежности, который продал евро по курсу 1,3021 дол. Однако затем тренд на рынке изменился, и курс евро стал снижаться. При этом в 7 часов утра курс евро настолько резко понизился, что его смогли приобрести инвесторы, установившие цену покупки евро с 60%-ными, 70%-ными, 80%-ными и 90%-ными уровнями надежности, которые приобрели евро валюту по курсу 1,2886 дол. (рис. 7.10).

Судя по таблице 7.27, по результатам торгов в паре «евро — доллар», состоявшихся 14 сентября, наиболее высокую доходность получили инвесторы, установившие цену покупки евро с 60%-ными, 70%-ными, 80%-ными и 90%-ными уровнями надежности, — 0,78%. В то время как инвестор, установивший цену продажи евро с 60%-ным уровнем надежности, получил лишь 0,18%. В свою очередь инвестор, придерживавшийся стратегии «купил и держи», понес убытки в размере 0,08%. При этом фактическая вероятность удачной сделки для инвестора, придерживавшегося этой стратегии, по нашим подсчетам, за период с 5 января 1999 г. по 13 сентября 2010 г. составила 49,9% (из 2996 торговых дней удачными оказались лишь 1495 дней).

Контрольные вопросы и задания

1. Какую роль играет точечный прогноз при разработке стоп-заявок? Как точечный прогноз можно пересматривать по итогам последних торгов? Назовите плюсы и минусы такого пересмотра прогноза.
2. Почему точечный прогноз не стоит брать в качестве среднего для нормального распределения при расчете рекомендуемых цен покупки или продажи валюты на прогнозируемый период? Какое значение следует взять в качестве среднего для нормального распределения при расчете рекомендуемых цен покупки или продажи?
3. Каким образом при расчете рекомендуемых цен покупки и продажи находится средняя ошибка прогнозируемого индивидуального значения курса доллара? Какой t -критерий используется при расчете рекомендуемых цен покупки или продажи? Чем односторонний t -критерий отличается от двустороннего t -критерия?
4. Какой риск по мере повышения уровня надежности рекомендуемых цен покупки и продажи валюты у инвестора снижается, а какой,

- напротив, повышается? Можно ли оба эти риска одновременно свести к минимуму?
5. Повышается ли точность статистических моделей при уменьшении срока прогнозирования? Как этот процесс можно объяснить с точки зрения математики? Повышается ли при этом ценность прогноза?
 6. Объясните, почему при построении статистических моделей для прогнозирования курса доллара к рублю с упреждением в две и в одну неделю мы не стали переходить к логарифмическому временному ряду, в то время как при построении модели для прогнозирования курса евро к доллару нам пришлось перейти к логарифмическому временному ряду?
 7. Почему при одном и том же уровне надежности, в соответствии с которым составлялись цены покупки и продажи доллара на рубли, вероятность удачной сделки при покупке доллара выше, чем при его продаже? В чем причина асимметрии в доходности? Можно ли сказать, что такая же асимметрия в доходности существует относительно пары «евро — доллар»? Подтвердите свой вывод конкретными цифрами.

Краткое заключение

Ну вот, уважаемый читатель, вы уже читаете последнюю страницу книги, посвященной на первый взгляд довольно трудному делу — прогнозированию курса доллара. Однако «не так страшен черт, как его малюют» — гласит народная пословица.

Надеюсь, наша книга убедила читателей, что причудливо меняющиеся курсы валют вполне поддаются прогнозированию. И хотя эти прогнозы нельзя давать со 100%-ным уровнем надежности, однако прогнозирование курса валюты с 95%-ным уровнем надежности является вполне посильной для инвестора задачей, если он, конечно, владеет соответствующей методикой. Хотелось бы также заметить, что 100%-ный уровень надежности в прогнозах невозможен не только на волатильных финансовых рынках, но и во многих областях человеческой деятельности, поскольку в жизни всегда есть место риску и разного рода случайностям.

Вполне очевидно, что интервальные прогнозы по курсам валют позволяют инвестору оценить как ожидаемые доходы, так и возможные убытки на прогнозируемый период. Соответственно он может либо захеджироваться от этих рисков, либо быть готовым понести возможные убытки в обмен на потенциально возможную прибыль. Причем если в первом случае, зная величину риска, инвестор может правильно оценить реальную цену покупки инструмента для хеджирования, то во втором случае он идет на риск не вслепую, а заранее оценив целесообразность такого шага. Благодаря надежному интервальному прогнозу (тестированию его надежности посвящена значительная часть этой книги) участник рынка еще до инвестирования средств в валюту имеет информацию как о максимальном уровне своих возможных убытков, так и о максимальной величине своих возможных доходов.

Еще одним важным направлением в сфере валютного прогнозирования является составление рекомендуемых цен продажи и покупки валют, что позволяет инвестору составлять стоп-заявки на продажу или покупку валют с определенным уровнем надежности. При желании уровень надежности можно повысить за счет повышения цены продажи или снижения цены покупки валюты, однако при этом у инвестора соответственно повышается риск упущенной прибыли. Умение составлять рекомендуемые цены покупки и продажи позволяет инвестору рассчитать приемлемый для него уровень риска и обеспечить эффективное участие в валютной игре.

С уважением,

Владимир Брюков, независимый аналитик

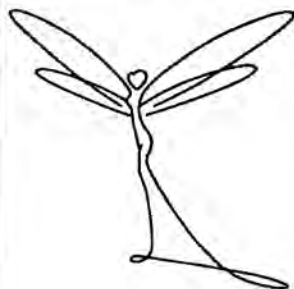
Свои вопросы к автору по содержанию этой книги, а также предложения по ее улучшению читатели могут прислать на адреса его электронной почты:

bryukov@bk.ru, bryukov@yandex.ru, vladimir2006.23.12@rambler.ru

Литература

1. *Андерсон Т.* Статистический анализ временных рядов / пер. под ред. Ю.К. Беляева. М. : Мир, 1976.
2. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ : пер. с англ. Кн. 1. 2-е изд.. М. : Финансы и статистика, 1986.
3. *Елисеева И.И., Юзбашев М.М.* Общая теория статистики : учебник / под ред. И.И. Елисеевой. М. : Финансы и статистика, 1995.
4. *Макарова Н.И., Трофимец В.Я.* Статистика в Excel. М. : Финансы и статистика, 2003.
5. *Молчанов И.Н., Герасимова И.А.* Компьютерный практикум по начальному курсу эконометрики (реализация на EViews) : практикум. Ростов н/Д, 2001.
6. *Турунцева М.Ю.* Анализ временных рядов. М. : МИЭФ ГУ-ВШЭ, 2003.
7. *Четыркин Е.М., Калихман И.Л.* Вероятность и статистика. М. : Финансы и статистика, 1982.
8. *Эконометрика* : учебник / под ред. И.И. Елисеевой, И.И. Елисеева, С.В. Курышева, Т.В. Костеева и др. 2-е изд., испр. и доп. М. : Финансы и статистика, 2006.

Свадебный салон «ДеМУАЗЕЛЬ»



ДеМУАЗЕЛЬ

Предлагает коллекцию свадебных и вечерних платьев, выполненную в самых современных традициях с европейским неподражаемым вкусом и качеством, американской роскошью и российской неповторимостью каждого платья.

Адрес салона:

Москва, ул. Садовая-Черногрозская, д. 20
(Садовое кольцо, ст. м. Курская или Красные Ворота)

Телефон: (495) 621 59 00

Сайт: <http://www.dmsl.ru>



Брюков Владимир Георгиевич,

независимый финансовый аналитик, с 2003 года занимается банковской журналистикой. С 2005 года особое место в его публикациях занимают статистические методы анализа валютных и финансовых рынков. Теме валютного прогнозирования, в первую очередь, прогнозу по курсу доллара США, посвящены многие его статьи, опубликованные в журналах «Валютный спекулянт», «Инвестиционный банкинг» и в ряде других изданиях. В этих публикациях обобщаются результаты проведенного автором исследования валютного рынка, предлагаются оптимальные методы прогнозирования по курсам валют с учетом последних достижений современной статистической науки.

С середины 2009 года В.Г. Брюков регулярно публикует на сайте <http://bankir.ru> (Банкир.Ру) прогнозы по курсам пятнадцати ведущих мировых валют на будущий месяц. Насколько точными при этом получаются прогнозы, наши читатели могут убедиться сами, посетив на этом сайте рубрику «Валютный рынок». Сотрудничество с этим известным и авторитетным порталом, а также большой интерес, проявленный читателями к теме валютного прогнозирования, стали для автора стимулом, способствовавшим написанию данной книги.

www.bankzakon.ru

www.bankzakon.ru — это справочная система, предназначенная для накопления и предоставления информации сотрудникам кредитных организаций. Справочная система представляет собой банк данных, содержащий вопросы и ответы практически во всех областях деятельности кредитных организаций на территории Российской Федерации. Ценность справочной системы состоит в том, что вопросы формулировались непосредственно самими сотрудниками банков при возникновении у них сложностей в трактовании или применении нормативных документов регулирующих органов, а ответы давались нашими авторами.

ISBN 978-5-406-01441-7



9 785406 014417