ИНСТРУКЦИЯ К ПРОГРАММЕ ОЦЕНКИ ПОРОГОВОЙ АВТОРЕГРЕССИИ

Автор: К.П. Глущенко
(самая первая версия программы была написана П.С. Ростовцевым)
glu@nsu.ru
http://econom.nsu.ru/users/gluschenko
Версия: 09.2005

0. Введение

Данная программа оценивает пороговые авторегрессии (ПАР; в англоязычной литературе – threshold autoregression, TAR), используя среду Microsoft Excel. Программа организована в виде макроса *TARmodel* в Excel. Программа написана на языке VBA (Visual Basic for Applications).

Пользователям предоставляется неограниченное право скачивать, копировать, использовать и изменять программу (за исключением коммерческого использования). Надеюсь, что вы не забудете включить ссылку на меня.

Чтобы посмотреть или изменить исходный текст программы, воспользуйтесь главным меню Excel: Сервис → Макрос → Макросы → Изменить (в англоязычных версиях Excel – Tools → Macro → Macros → Edit).

1. Алгоритм

Рассматриваемая модель имеет вид

$$\Delta y_t = \begin{cases} \lambda_{(out)}(y_{t-1} - c_{(+)}) + \varepsilon_{(out)t} & \text{если } y_{t-1} > c_{(+)} \\ \lambda_{(in)} y_{t-1} + \varepsilon_{(in)t} & \text{если } c_{(+)} \ge y_{t-1} \ge c_{(-)} \\ \lambda_{(out)}(y_{t-1} - c_{(-)}) + \varepsilon_{(out)t} & \text{если } y_{t-1} < c_{(-)} \end{cases} \qquad (t = 1, \ldots, T),$$

$$-2 < \lambda_{(out)} < 0$$

где $\{y_t\}_{t=0,\ldots,T}$ – временной ряд, Δ – оператор первой разности ($\Delta y_t \equiv y_t - y_{t-1}$), $\varepsilon_{(.)t}$ – остатки регрессии. Оцениваемыми параметрами являются $\lambda_{(out)}$ и $\lambda_{(in)}$, авторегрессионные коэффициенты (заметим, что оценка $\lambda_{(in)}$ обычно не представляет интереса), а также $c_{(+)}$ и $c_{(-)}$, верхний и нижний порог соответственно. Может быть наложено ограничение $\lambda_{(in)} = 0$, означающее, что процесс внутри интервала $[c_{(-)}, c_{(+)}]$ является чистым случайным блужданием.

Пороги c оцениваются так, чтобы выполнялось условие $W \le t_{(out)}/(T+1) \le 1 - W$ (аналогично для $t_{(in)}$), где W – заданное (пользователем) окно, $W \le 0.5$, $t_{(out)}$ – число наблюдений таких, что $y_t \notin [c_{(-)}, c_{(+)}]$, а $t_{(in)}$ – таких, что $y_t \in [c_{(-)}, c_{(+)}]$; $t_{(out)} + t_{(in)} = T+1$. Согласно рекомендации Д. Эндрюса [1], окно должно быть не менее 0.15, т.е. обе части временного ряда должны содержать не менее 15% наблюдений. Однако решение о величине W следует принимать в зависимости от длины анализируемого ряда (если он короткий, лучше увеличить окно, а если длинный, можно уменьшить).

За обсуждением экономического и эконометрического смысла модели следует обратиться к литературе, например, [2, 3, 4] и др. (Обычно y_t – это различие цен или уровней цен между двумя географическими единицами, т.е. $y_t = \ln(p_{rt}/p_{st})$, где p_{it} – цена или уровень цен в i.)

Тестирование на пороговый эффект является по сути тестом спецификации – ПАР против АР(1) (обычной авторегрессии первого порядка, по-английски AR(1)). То есть, тестируемая гипотеза – H_0: порождающий данные процесс есть АР(1) с параметром λ_0, $\Delta y_t = \lambda_0 y_{t-1} + \varepsilon_t$ ($t = 1, \ldots, T$), против альтернативы H_a: процесс есть ТАР с параметрами $\lambda_{(out)}$, $\lambda_{(in)}$, c.

В программе используется метод оценки и тестирования, предложенный М. Обстфельдом и А.М. Тэйлором (приложение A в [2, 3]), с небольшими модификациями. Их три: (*а*) Вместо фиксированной величины $W = 0.1$, ширина окна может быть любой в диапазоне [0, 0.5]. (*б*) В качестве шага перебора принято расстояние между соседними наблюдениями вместо какого-либо постоянного значения (например, 0.001). (*г*) В отличие от общепринятого использования симметричных порогов $c_{(-)} = -c_{(+)}$, они асимметричны, хотя и связаны соотношением $c_{(-)} = \ln(2 - \exp(c_{(+)}))$ (отметим, что y_t считается логарифмом некоторой величины); обоснование приведено в [4, приложение A].

Для оценки и тестирования ПАР используется полный перебор по параметру c. Максимизируемая целевая функция – логарифм отношения функций правдоподобия ПАР и АР(1): $LLR = L_{TAR} - L_{AR}$. В свою очередь,

$$L_{AR} = L(\lambda_0, \sigma) = -\frac{1}{2}\sum_{t=1}^{T}(\ln 2\pi + \ln \sigma^2 + \varepsilon_t^2/\sigma^2),$$

$$L_{TAR} = L(\lambda_{(out)}, \lambda_{(in)}, \sigma_{(out)}, \sigma_{(in)}, c) = -\frac{1}{2}\sum_{t\in O}(\ln 2\pi + \ln \sigma_{(out)}^2 + \varepsilon_{(out)t}^2/\sigma_{(out)}^2) -$$

$$-\frac{1}{2}\sum_{t\in I}(\ln 2\pi + \ln \sigma_{(in)}^2 + \varepsilon_{(in)t}^2/\sigma_{(in)}^2),$$

где $O = \{t\colon y_{t-1} \notin [c_{(-)}, c_{(+)}]\}$, $I = \{t\colon y_{t-1} \in [c_{(-)}, c_{(+)}]\}$.

Процедура оценивания и тестирования такова.

1. Оценивается АР(1) с помощью обычного метода наименьших квадратов (ОМНК), получая оценки $\hat{\lambda}_0$ и $\hat{\sigma}$, и рассчитывается $\hat{L}_{AR}$.

2. Для каждого t рассчитывается $c_t = y_t$, если $y_t > 0$, или $c_t = \ln(2 - \exp(y_t))$, если $y_t < 0$. Набор $\{c_t\}$ упорядочивается по возрастанию, что даёт $\{c_k\}$. Рассчитываются $k0 = [W(T-1)] + 1$ и $k1 = T - k0$ (напомним, что W – окно; $[x]$ означает целую часть x.)

3. Для каждого $k = k0,\ldots,k1$ рассчитываются $c_{(+)k} = c_k$ и $c_{(-)k} = \ln(2 - \exp(c_k))$, строятся множества $O_k = \{t\colon y_{t-1} \notin [c_{(-)k}, c_{(+)k}]\}$ и $I_k = \{t\colon y_{t-1} \in [c_{(-)k}, c_{(+)k}]\}$, оцениваются с помощью ОМНК $\Delta y_t = \lambda_{(out)k} y_{t-1} + \varepsilon_{(out)t}$ $(t \in O_k)$ и $\Delta y_t = \lambda_{(in)k} y_{t-1} + \varepsilon_{(in)t}$ $(t \in I_k)$, и рассчитывается LLR_k.

4. Определяется $k^* = \operatorname{argmax}_k(LLR_k)$; $LLR^* = LLR_{k^*}$. В качестве оценок модели берутся $\hat{\lambda}_{(out)} = \lambda_{(out)k^*}$, $\hat{\lambda}_{(in)} = \lambda_{(in)k^*}$, $\hat{c} = c_{k^*}$.

5. Оценивается p-значение (наблюдаемая значимость) LLR, используя для получения распределения статистики LLR при нулевой гипотезе модельно-ориентированный бутстреп. Это выполняется таким образом. В каждом эксперименте i ($i = 1,\ldots,N$) генерируются $T + 50$ случайных величин $\varepsilon_{-50}^{(i)},\ldots,\varepsilon_T^{(i)}$, $\varepsilon_t^{(i)} \sim i.i.d.N(0, \hat{\sigma}^2)$. Затем строится имитированный временной ряд: $y_t^{(i)} = (\hat{\lambda}+1)y_{t-1}^{(i)} + \varepsilon_t^{(i)}$, $t = -50,\ldots,T$ и $y_{-51}^{(i)} = 0$, из которого исключаются первые 50 «наблюдений» (чтобы избежать смещения, обусловленного начальным значением). По этому ряду оцениваются модели АР(1) и ПАР как в шагах 1-4, давая реализацию $LLR^{(i)}$. Число реализаций, в которых $LLR^{(i)} > LLR^*$, отнесённое к общему количеству экспериментов N, и есть p-значение LLR^*, т.е. вероятность случайно получить величину LLR, превосходящую LLR^*, в случае выполнения нулевой гипотезы.

2. Входные данные

В качестве исходной информации программа использует данные с листа Excel. Он не обязан быть в том же файле Excel, что содержит программу, но файл с программой должен быть открыт наряду с файлом, содержащим обрабатываемые данные.

Временной ряд должны быть вектором-столбцом, т.е. отдельный ряд размещается в столбце листа, при этом число его строк равно длине ряда (T+1), а порядок строк соответствует порядку временных точек t. За один прогон программа может обрабатывать произвольный набор временных рядов, имеющих одну и ту же длину. Ряды в таком наборе могут располагаться горизонтально (по столбцам) и/или вертикально (по “ярусам”). Ряды не обязательно должны идти непрерывно: программа пропускает пустые ряды. Однако вертикальные пропуски (пустые ярусы) должны иметь ту же длину, что и сами ряды (т.е. число строк в пустом ярусе должно равняться длине ряда).

Ряды не должны занимать первую строку (строка 1) и первый столбец (столбец А) листа: в эти строку и колонку помещаются идентификаторы рядов (названия, коды и т.п.). Строка содержит идентификаторы колонок рядов, а столбец А – идентификаторы ярусов (в первой строке каждого яруса). Начало диапазона данных (прямоугольник листа Excel, содержащий обрабатываемые ряды) может быть сдвинуто от первой строки листа на произвольное число строк, то же самое относится и к столбцам.

Такая организация данных обусловлена тем, что программа была разработана для обработки наборов попарных различий цен между географическими единицами. На рис. 1 приведён пример организации таких данных.

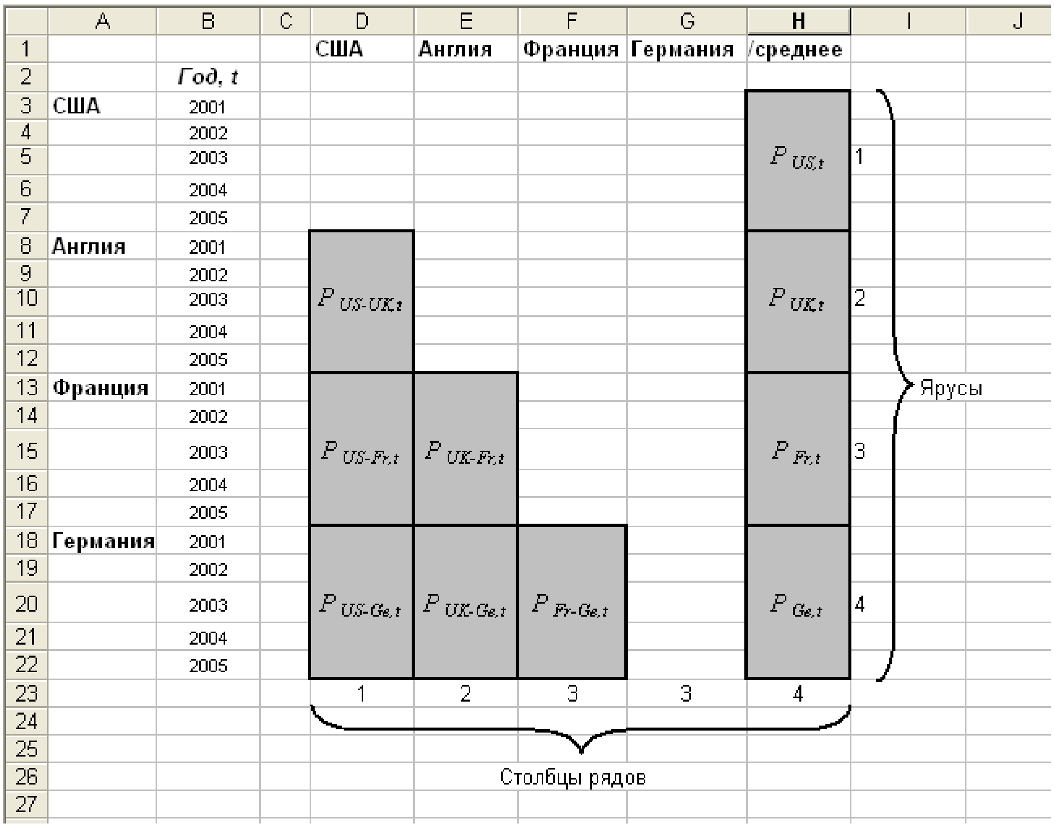

Рис. 1. Пример организации данных (каждый серый прямоугольник представляет отдельный временной ряд)

Для обработки за один прогон можно выбрать любой набор рядов, попадающий в прямоугольный диапазон. Например, если выбран диапазон H3:H22, будут обрабатываться четыре ряда цен, отнесённых к средней по всем странам цене; в случае диапазона D8:F22 будут обработаны шесть рядов попарных сравнений. Все показанные ряды обрабатываются при задании диапазона D3:H22, и единственный ряд $\{P_{UK\text{-}Ge,t}\}$ будет обработан, если задать диапазон E18:E22.

Можно также оценивать ПАР не по всему ряду, а по его части. Для этого нужно задать диапазон, соответствующий обрабатываемой части ряда. Например, при указании диапазона H9:H12 оценивание будет производиться по ряду $\{P_{UK,t}\}$ за 2002–2005 гг.; диапазон H8:H11 даёт оценки для ряда без последнего года, а диапазон H9:H11 – оценки по средней части ряда $\{P_{UK,t}\}$, без 2001 и 2005 гг. Такие оценивания могут быть выполнены и по нескольким рядам за один прогон, но очевидно, что это должны быть ряды из одного яруса. Например, диапазон D14:H17 задаёт оценивание для трёх рядов $\{P_{US\text{-}Fr,t}\}$, $\{P_{UK\text{-}Fr,t}\}$ и $\{P_{Fr,t}\}$ без начального года. Но чтобы получить оценки для всех четырёх рядов, включающих Францию, нужен ещё один прогон – с указанием диапазона F19:F22.

Чтобы задать, какие ряды должны быть обработаны, нужно перед запуском программы выделить необходимый ряд или набор рядов. Пример приведён на рис. 2. В этом примере выделен диапазон D3:I17, при этом будут обработаны шесть рядов (не включены ряды, относящиеся к Германии).

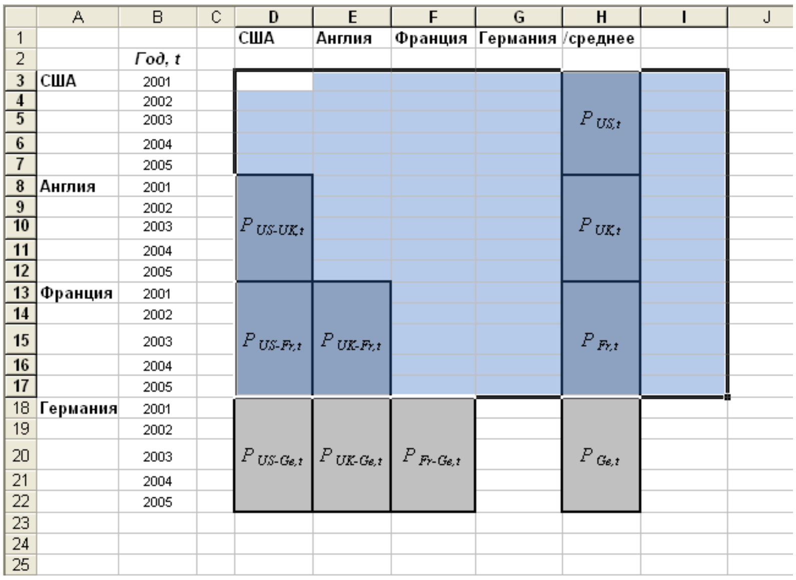

Рис. 2. Задание рядов, которые должны быть обработаны за один прогон

Другой способ состоит в том, что диапазон на листе Excel, содержащий данные, которые следует обработать, задаётся в диалоговом окне программы (см. следующий раздел).

3. Запуск программы

Через главное меню Excel программа вызывается последовательностью Сервис → Макрос → Макросы → Выполнить (Tools → Macro → Macros → Run). Если активны другие макросы, перед нажатием Выполнить (Run) выделите TARmodel. Другой способ – нажать <Ctrl>t. После этого появляется диалоговое окно программы, показанное на рис. 3.

При появлении диалоговое окно содержит настройки по умолчанию. Вы можете изменить все или некоторые из них. Элементы диалогового окна следующие.

Диапазон определяет диапазон на листе Excel, содержащий обрабатываемые ряды. По умолчанию это диапазон, выделенный перед запуском программы. Если такового нет, берётся адрес активной ячейки, и тогда вы должны вручную ввести координаты диапазона рядов. Координаты выделенного диапазона вы можете изменить.

Длина ряда определяет размер выборки (T+1). По умолчанию берётся число строк в выделенном диапазоне (если такового нет, значение в этом окошке равно 1). Если координаты диапазона введены вручную или изменены, нужно ввести фактическую длину ряда –

непосредственно или пользуясь кнопочками счётчика справа от окошка. Длина ряда по умолчанию также должна быть изменена на фактическую, когда диапазон исходных данных содержит более одного яруса (в примере на рис. 3 длина ряда должна быть изменена на 5.)

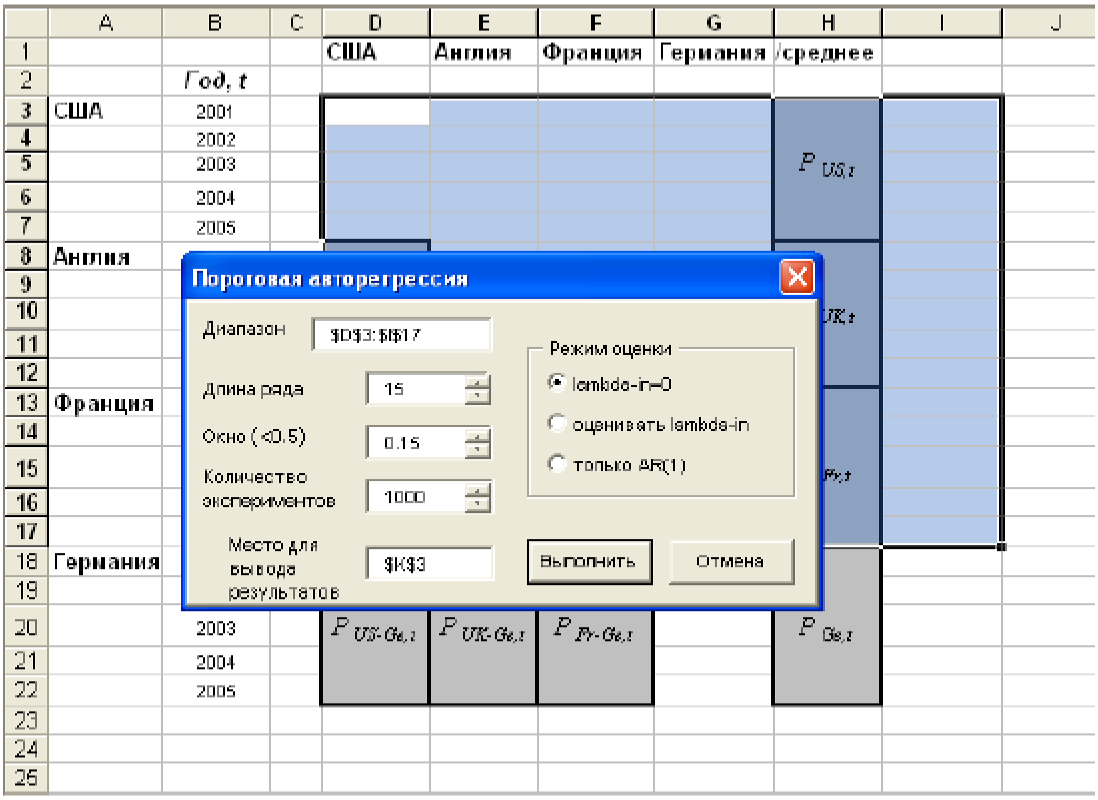

Рис. 3. Диалоговое окно программы

Окно задаёт величину W (см. раздел 1), т.е. минимальную долю ряда, которая может быть внутри или вне интервала $[c_{(-)}, c_{(+)}]$. Значение по умолчанию – 0.15.

Количество экспериментов задаёт число экспериментов N в методе Монте-Карло при оценке p-значения LLR. По умолчанию принимается 1000. Поскольку p-значение довольно чувствительно к N, желательно, чтобы N было как можно больше. Но, с другой стороны, именно величина N определяет время работы программы. Поэтому выбор должен зависеть от количества обрабатываемых рядов. Во всяком случае, N=1000 представляется недостаточным, я бы порекомендовал по меньшей мере 10000.

Место для вывода результатов задаёт верхнюю левую ячейку выходной таблицы. По умолчанию выходная таблица начинается с той же строки, что и обрабатываемые ряды, и со столбца, на один правее последнего столбца диапазона исходных данных. Вы можете задать другое положение таблицы при условии, что она не будет пересекаться с диапазоном исходных данных.

Режим оценки содержит три переключателя. Режим по умолчанию – lambda-in=0, означающий, что оценивание будет производиться при ограничении $\lambda_{(in)} = 0$. Это ограничение снимается в режиме оценивать lambda-in. И, наконец, оценивается только модель AP(1) вместо и AP(1), и ПАР, если выбран режим только AR(1).

Программа проверяет правильность настроек при вводе их. Когда настройки в диалоговом окне установлены надлежащим образом, нажмите OK. После этого диалоговое окно раздвигается, как показано на рис. 4. Дополнительная часть окна показывает структуру диапазона исходных данных и даёт возможность исправить настройки.

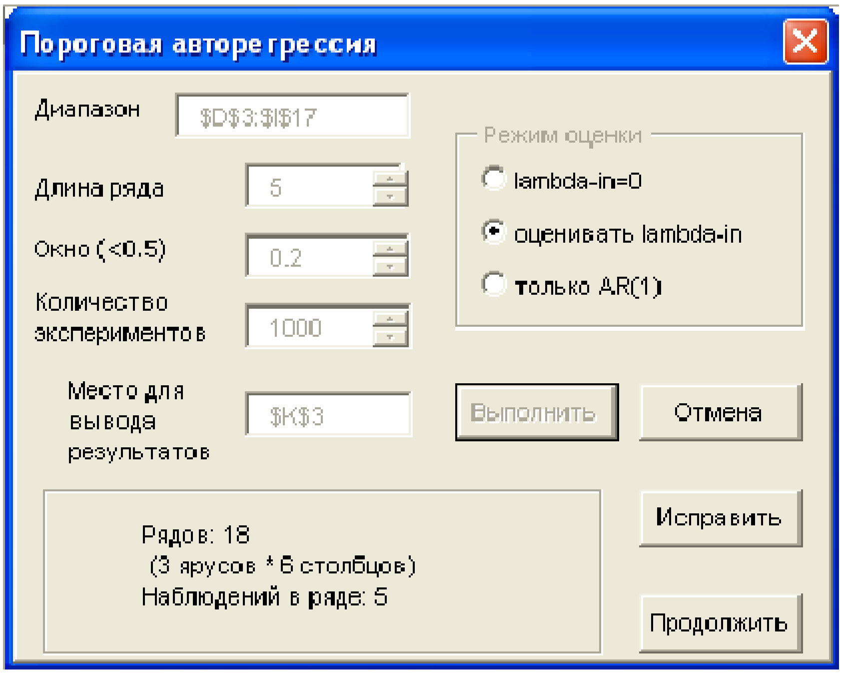

Рис. 4. Раздвинутое диалоговое окно

Рис. 4 использует тот же пример, что и рис. 3. Как видно, величина *W* (окно) изменена, как и длина ряда. Теперь программа понимает, что ряды расположены один за другим и один под другим, предполагая, что диапазон исходных данных содержит 18 рядов (включая пустые, которые будут игнорироваться при выполнении программы).

Вы можете вернуться к заданию настроек, нажав Исправить, или начать оценивания, нажав Продолжить

При выполнении программы из неё можно в любой момент выйти, нажав клавишу Esc.

4. Выходные данные

Если задан набор временных рядов, они обрабатываются слева направо и сверху вниз, при этом пустые ряды пропускаются. Пример выходной таблицы представлен на рис. 5.

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
1																							
2																							
3	N exp	r	s	p(unit root)	AR lambda	s.d.	t-stat	AR T-half	LLR	p(LLR)	TAR lambda-out	s.d.	t-stat	TAR T-half	c(-1)	C(-1), %	c	C, %	t out	t in	TAR lambda-in	s.d.	t-stat
4	10000	США	/среднее		-0.698384	0.105	-6.675	0.5783	20.7	0.001	-1.901853444	0.402	-0.244	6.709829	0.055	5.699	0.059	6.1	18	65	-0.286290883	0.146	-1.957
5	10000	Англия	США		-0.468958	0.092	-5.081	1.09517	12.6	0.035	-0.636832009	0.140	-4.537	0.684326	0.017	1.6826	0.018	1.8	58	25	-1.715142755	0.541	-3.170
6	10000	Англия	/среднее		-0.400633	0.088	-4.575	1.35412	20	0.002	-1.429829883	0.418	-1.364	0.820909	0.081	8.412	0.081	8.5	13	70	-0.316614984	0.119	-2.656
7	10000	Франция	США		-0.376303	0.087	-4.316	1.46825	6.94	0.336	-0.825088301	0.146	-5.633	0.397567	0.045	4.5679	0.046	4.7	38	45	-0.106563386	0.310	-0.344
8	10000	Франция	Англия		-0.275724	0.075	-3.665	2.14874	58.3	0.000	-0.798027834	0.238	-3.356	0.433318	0.102	10.788	0.103	10.8	33	50	-0.023347652	0.099	-0.235
9	10000	Франция	/среднее		-0.014178	0.015	-0.922	48.5425	8.38	0.250	-0.212414926	0.116	-1.831	2.902822	0.253	28.732	0.253	28.8	20	63	-0.007338084	0.022	-0.339
10	10000	Германия	США		-0.010484	0.014	-0.734	65.7669	9.15	0.199	-0.119351616	0.055	-2.159	5.453693	0.183	20.031	0.183	20.1	54	29	0.050059591	0.037	1.367
11	10000	Германия	Англия		-0.008916	0.008	-1.067	77.395	29.4	0.052	-0.092295148	0.127	-0.729	7.157949	0.429	53.573	0.430	53.7	12	71	-0.003407009	0.007	-0.471
12	10000	Германия	Франция		-0.00604	0.011	-0.526	114.407	13.2	0.186	-0.450226178	0.155	-2.904	1.158628	0.405	49.869	0.405	50.0	27	56	0.003715117	0.017	0.216
13	10000	Германия	/среднее		-0.00303	0.016	-0.193	228.381	7.68	0.351	-0.276259299	0.135	-2.040	2.143829	0.187	20.571	0.188	20.7	22	61	0.03673584	0.021	1.770
14																							

Рис. 5. Пример выходной таблицы

Выходная таблица содержит следующие столбцы.

N exp показывает текущий номер эксперимента Монте-Карло *i* при обработке данного ряда; после завершения оценивания эта ячейка содержит общее число экспериментов *N*.

r и **s** – идентификаторы колонки и яруса ряда. Например, если все 10 рядов с рис. 1 обрабатываются за один прогон (когда задан входной диапазон D3:H22), эти колонки выходной таблицы выглядят как на рис. 5.

p(unit root) всегда пустая. Эта колонка предназначена для p-значений теста на единичный корень, которые пользователь получает с помощью какой-нибудь другой программы (к примеру, программы МакКиннона [5], EViews и т.п.).

AR lambda – оценка λ_0.

s.d. – стандартное отклонение оценки λ_0.

t-stat – значение t-статистики оценки λ_0.

AR T-half – время полураспада возмущений в модели АР(1); оно рассчитывается как $\ln 0.5/\ln(|1+\lambda_0|)$, его единица измерения совпадает с периодичностью соответствующего временного ряда (месяцы, кварталы и т.п.).

LLR – оценка величины статистики LLR для TAR.

p(LLR) –p-значение оценки LLR.

TAR lambda-out – оценка $\lambda_{(out)}$.

s.d. – стандартное отклонение оценки $\lambda_{(out)}$.

t-stat значение t-статистики оценки $\lambda_{(out)}$.

TAR T-half – время полураспада возмущений в модели ПАР; оно рассчитывается как $\ln 0.5/\ln(|1+\lambda_{(out)}|)$.

c(–1) – значение c, ближайшее к оценке c. Поскольку перебор ведётся по дискретным точкам , мы не знаем точную точку максимума *LLR*. Истинное значение c лежит где-то между оценкой c и соседней точкой перебора. Величина в этом столбце даёт представление о том, насколько велика может быть неточность оценки c.

C(–1), % – величина $c(-1)$, выраженная в процентах; она рассчитывается как $\ln(e^{c(-1)} - 1)\cdot 100$ (напомним, что $\{y_t\}$ считаются логарифмами).

c – оценка c (точнее, $c_{(+)}$).

C, % – величина c, выраженная в процентах; она рассчитывается как $(e^c - 1)\cdot 100$.

t out – $t_{(out)}$, т.е. число наблюдений таких, что $y_t \notin [c_{(-)}, c_{(+)}]$.

t in – $t_{(in)}$, т.е. число наблюдений таких, что $y_t \in [c_{(-)}, c_{(+)}]$.

TAR lambda-in – оценка $\lambda_{(in)}$; если задан режим lambda-in=0, в этом столбце всегда нуль.

s.d. – стандартное отклонение оценки $\lambda_{(in)}$.

t-stat – значение t-статистики оценки $\lambda_{(in)}$; если задан режим lambda-in=0, в этом столбце всегда нуль.

Литература

1. Andrews D. W. K. Tests for Parameter Instability and Structural Change with Unknown Change Point // *Journal of Econometrics*. – 1993. – V. 61, № 4. – P. 821-856.

2. Obstfeld M., Taylor A. M. Non-Linear Aspects of Good-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited. *NBER Working Paper* No. 6053, 1997. (www.nber.org/papers/w6053).

3. Obstfeld M., Taylor A. M. Non-Linear Aspects of Good-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited // *Journal of Japanese and International Economies*. – 1997. – V. 11. – P. 441-479.

4. Gluschenko K. The Law of One Price in The Russian Economy. *LICOS Discussion Paper* No. 152/2004, 2004 (www.econ.kuleuven.ac.be/licos/DP/DP2004/DP152.pdf).

MacKinnon, J. G. (1996). Numerical Distribution Functions for Unit Root and Cointegration Tests. *Journal of Applied Econometrics*, **11** (6), 601-618.