УДК 519.237.5

СРАВНЕНИЕ МЕТОДОВ ЭКОНОМЕТРИЧЕСКОГО АНАЛИЗА ДАННЫХ ДЛЯ ИДЕНТИФИКАЦИИ ФИНАНСОВЫХ ПУЗЫРЕЙ

Е.А. Гребенюк, А.В. Малинкина

Проведен сравнительный анализ эконометрических методов обнаружения и датирования пузырей на финансовых рынках, основанных на современном подходе к идентификации пузырей. Рассматривались два подхода, основанные на определении финансовых пузырей как периодов, в которые динамика изменения цены описывается нестационарным процессом взрывного типа. Сравнение проведено с использованием результатов моделирования по методу Монте-Карло.

Ключевые слова: взрывной процесс, правосторонний тест на единичный корень, коэффициент корреляции, последовательный анализ.

ВВЕДЕНИЕ

Проблема идентификации финансовых пузырей волнует как экономистов, так и инвесторов. Имеется большое число публикаций, в которых рассматриваются вопросы определения финансовых пузырей и методов их обнаружения и датирования. Эти два вопроса тесно взаимосвязаны, поскольку признаки наличия или возникновения пузыря, поиск которых осуществляется применяемыми методами, напрямую зависят от того, что же понимается под финансовым пузырем.

Большое число проведенных исследований послужило обоснованием возможности применения эконометрического подхода к обнаружению пузырей. Сущность эконометрического подхода заключается в определении пузырей как периодов, в которые цены изменяются по экспоненте. В соответствии с общепринятой классификацией нестационарных процессов [1] такие процессы называются взрывными. Однако реальный процесс может вести себя как «взрывной» ограниченное время, так как иначе значения его будут бесконечно увеличиваться. Начало эконометрическим методам обнаружения пузыря было положено в работах Дибы и Гроссмана [2], применившими правосторонний тест Дики — Фуллера для проверки наличия в процессе единичного корня. Однако Эванс [3] показал, что применяемые ими стандартные процедуры проверки [4] не позволяли различать процессы, в которых периодически возникают и схлопываются пузыри, от стационарных процессов или процессов случайного блуждания, поэтому гипотеза взрывного поведения процесса в момент возникновения и существования пузыря долгое время не получала эмпирического подтверждения. Обоснованием для ее подтверждения послужили результаты, полученные в работах [5—7] и основанные на процедурах последовательной проверки гипотез.

Впервые статистические алгоритмы, позволяющие определять моменты возникновения и «схлопывания» пузыря в реальном времени предложены в работах [5, 6]. В рамках подхода, предложенного Филлипсом, был разработан и исследован ряд алгоритмов, позволяющих обнаруживать финансовые пузыри в реальном времени [8, 9]. Последовательный подход к обнаружению позволяет:

— определить моменты возникновения и схлопывания пузыря в режиме получения текущих наблюдений, в отличие от применяемых ранее методов, которые могли установить факт наличия пузыря в процессе только в апостериорном режиме.

— определить в режиме получения текущих наблюдений моменты возникновения и схлопывания нескольких последовательно возникающих и схлопывающихся пузырей.

Исследуемый в данной статье подход к обнаружению пузырей основан на процедурах последовательной проверки, выполняемых в каждый момент получения нового наблюдения. Процедуры могут различаться: моделями процесса в отсутствие пузыря, моделями пузыря, формулировками нулевой и альтернативной гипотез, процедурами их проверки, используемыми статистиками. К алгоритмам, реализующим этот подход, относятся алгоритмы Филлипса [5—7], алгоритмы Брейтунга с соавторами [9] и предложенный в работе [8] алгоритм последовательного обнаружения изменений свойств коэффициента корреляции.

Исследования, проведенные Филлипсом с соавторами [6], и независимые исследования [9] показали более высокую эффективность предложенных им алгоритмов по сравнению с другими аналогичными алгоритмами. Поэтому при исследовании эффективности алгоритма последовательного обнаружения, предложенного в работе [8], мы проводили сравнение его с алгоритмами Филлипса. В статье изложены основные принципы построения сравниваемых алгоритмов и средствами математического моделирования проведен сравнительный анализ их свойств: точности датирования, доли ложных обнаружений, доли необнаруженных пузырей, а также робастности алгоритмов относительно выбираемых параметров настройки.

1. АЛГОРИТМЫ ОБНАРУЖЕНИЯ И ДАТИРОВАНИЯ ПУЗЫРЕЙ, ОСНОВАННЫЕ НА АНАЛИЗЕ СТЕПЕНИ НЕСТАЦИОНАРНОСТИ ПРОЦЕССОВ ЦЕН И ДИВИДЕНДОВ

1.1. Степень нестационарности процесса

Каждый из процессов изменения цен, дивидендов или их отношений на финансовых рынках в определенный период времени может принадлежать к одному из следующих классов, определяющих степень нестационарности процесса: стационарный процесс, случайное блуждание или процесс с единичным корнем, «взрывной» процесс. Приведем здесь определения классов процессов, более подробные описания можно найти в работах [1, 10]. Пусть процесс y_1, y_2, ..., y_t, ... описывается моделью авторегрессии порядка p:

$$y_t = \beta_1 y_{t-1} + ... + \beta_p y_{t-p} + \varepsilon_t, \qquad (1)$$

тождественными преобразованиями его можно привести к виду

$$\Delta y_t = \gamma y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + \varepsilon_t, \qquad (2)$$

где $\gamma = \beta_1 + ... + \beta_p - 1$, $\gamma_1 = -(\beta_2 + ... + \beta_p)$, $\gamma_2 = -(\beta_3 + ... + \beta_p)$, ..., $\gamma'_{p-1} = -\beta_p$, ε_t — белый шум. Степень нестационарности или класс процесса y_t определяется значениями коэффициента γ в модели (2): если $\gamma < 0$, то процесс является *стационарным*, если $\gamma = 0$, то y_t — случайное блуждание или «процесс с единичным корнем». Последний термин получил распространение в силу того, что при $\gamma = 0$ характеристическое уравнение процесса (1)

$$1 - \beta Z - ... - \beta_p Z^p = 0 \qquad (3)$$

имеет, по крайней мере, один корень на единичной окружности, если процесс стационарный, то все корни его характеристического уравнения (3) лежат вне единичной окружности. Если $\gamma > 0$, то процесс — «*взрывной*», в этом случае, по крайней мере, один из корней характеристического уравнения (3) лежит внутри единичной окружности. Из уравнения (2) следует, что реальный процесс может вести себя как «взрывной» ограниченное время, так как иначе значения его будут бесконечно увеличиваться.

Филлипсом и Магдалиносом [11] были рассмотрены умеренно взрывные процессы, которые использовались Филлипсом для построения модели пузырей. Модели взрывных процессов, рассматриваемых в работе [11], имеют вид:

$$y_t = \mu + \delta_n y_{t-1} + \varepsilon_t, \qquad (4)$$

или

$$y_t = \mu + \delta_n y_{t-1} + \sum_{i=1}^{J} \alpha_i \Delta y_{t-j} + \varepsilon_t, \qquad (4а)$$

где $\delta_n = 1 + c/k_n$, $c > 0$, μ — константы, $k_n \to \infty$ — последовательность детерминированных величин, стремящихся к бесконечности со скоростью меньшей, чем n, $k_n/n \to 0$, (например, $k_n = n^\alpha$, $\alpha \in (0, 1)$), $\varepsilon_t \sim (0, \sigma^2)$ — последовательность независимых и одинаково распределенных случайных величин, α_1, ..., α_J — коэффициенты авторегрессии в модели (4а), $\Delta y_{t-j} = y_{t-j} - y_{t-j-1}$.

1.1. Алгоритмы Филлипса

В основе разработанных алгоритмов лежит процедура оценивания в каждый момент времени t параметров регрессии вида (4) или (4а) по выборке, изменяющейся в скользящем окне с последующей проверкой гипотезы H_0 о равенстве δ_n нулю против альтернативной гипотезы: H_1:$\delta_n > 0$. Известны два основных варианта алгоритма, которые различаются процедурами построения выборок для оценивания δ_n и используемыми статистиками. Первый предназначен для обнаружения единственного пузыря в процессе, второй может обнаруживать несколько последовательно возникающих и схлопывающихся пузырей.

Рассмотрим первый вариант, названный авторами [4] SADF (supADF — supremum augmented

Dickey — Fuller) — супремум расширенного теста Дики — Фуллера:

$$SADF(r_0) = \sup_{r \in [r_0, 1]} ADF_0^r, \tag{5}$$

где ADF_0^r — значение статистики Дики — Фуллера [4, 10], вычисляемой по выборке, длина которой составляет $T^r = [Tr]$ наблюдений, где T — длина выборки, на которой применяется алгоритм, $r_0 = [l_0/T]$, где l_0 — минимальная длина подвыборки, используемая для оценивания δ_n на первом шаге; $[x]$ означает целую часть x.

Алгоритм SADF заключается в последовательном оценивании регрессии (4) по расширяющейся выборке и определении супремума полученной статистики. Вычисляемая статистика $\sup_{r \in [r_0, 1]} ADF_0^r$ на каждом шаге сравнивается с правосторонним критическим значением предельного распределения этой статистики, если она превышает критическое значение, то это означает, что в процессе присутствует пузырь.

Момент возникновения пузыря определяется как первый момент, когда ADF_0^r — статистика расширенного теста Дики — Фуллера превышает критическое значение предельного распределения $\sup_{r \in [r_0, 1]} ADF_0^r$; момент окончания или «схлопывания» пузыря определяется как первый момент после $[Tr_e + L_T]$, в который статистика ADF_0^r становится ниже критического значения:

$$\hat{r}_e = \inf_{r \in [r_0, 1]} \left\{ r{:}ADF^r > cv_{r_0}^{\beta_T} \right\}$$

$$\text{и } \hat{r}_f = \inf_{r \in [r + L(T)/T, 1]} \left\{ r{:}ADF^r < cv_{r_0}^{\beta_T} \right\}, \tag{6}$$

где L_T — интервал, определяющий минимальную длительность пузыря, $cv_{r_0}^{\beta_T}$ — критическое значение статистики (5), которое зависит от уровня значимости теста β_T, объема выборки T, минимальной доли оцениваемой под выборки r_0.

Второй вариант предназначен для обнаружения нескольких подряд возникающих и схлопывающихся пузырей. Как было замечено Эвансом, мощность стандартного теста на единичный корень [4,10] слишком мала, чтобы обнаруживать периодически схлопывающиеся пузыри, поэтому при наличии нескольких пузырей, интервалы между которыми не слишком велики, алгоритм SADF не способен их различать. В работе [6] Филлипсом с соавторами было предложено обобщение алгоритма SADF — обобщенный supADF-тест (Generalized supADF — GSADF). Алгоритм GSADF также как и алгоритм SADF проверяет гипотезу «процесс с единичным корнем» против альтернативной гипотезы «взрывной процесс», однако использует расширяющиеся выборки, которые могут начинаться в любой из точек интервала $[0, r_2 - r_0]$:

$$GSADF(r_0) = \sup_{\substack{r_1 \in [0, r_2 - r_0] \\ r_2 \in [r_0, 1]}} ADF_{r_1}^{r_2}, \tag{7}$$

где $ADF_{r_1}^{r_2}$ — значение статистики Дики — Фуллера, вычисляемой по выборке, длина которой составляет $T^{r_2} - T^{r_1}$ наблюдений и начинается в момент $T^{r_1} + 1$, где $T^{r_1} = [Tr_1]$, $T^{r_2} = [Tr_2]$. Оценки дат возникновения и схлопывания пузыря для алгоритма GSADF определяются по формулам:

$$\hat{r}_e = \inf_{r_2 \in [r_0, 1]} \left\{ r_2{:} \sup_{r_1 \in [0, r_2 - r_0]} ADF_{r_1}^{r_2} > scv_{r_0}^{\beta_T} \right\}$$

$$\text{и } \hat{r}_f =$$

$$= \inf_{r_1 \in [r_e + L(T)/T, 1]} \left\{ r_2{:} \sup_{r_1 \in [0, r_2 - r_0]} ADF_{r_1}^{r_2} < scv_{r_0}^{\beta_T} \right\}, \tag{8}$$

где $scv_{r_0}^{\beta_T}$ — критическое значение предельного распределения статистики (6).

Выражения для предельных распределений обеих статистик (5) и (7) и их критические значения, которые определяются уровнем значимости теста β_T, объемом выборки T, минимальной долей оцениваемой под выборки r_0, можно найти в работах [6, 7].

Анализ свойств алгоритма по историческим данным, выполненный с помощью моделирования по методу Монте-Карло [7, 12] показал высокую эффективность алгоритмов SADF и GSADF в случае наличия одного пузыря. Экспериментальный анализ выборки с несколькими пузырями показал, что «алгоритм GSADF более эффективен, чем алгоритм SADF при обнаружении нескольких пузырей» [12]. Однако проведенные нами экспериментальные исследования и более поздние результаты [7] не всегда подтверждают этот вывод.

1.2. Алгоритм последовательного обнаружения

Пусть модель исходного процесса в отсутствие пузыря описывается уравнением (4) или (4а), где

$$\delta_n = \begin{cases} 0, \text{если } t = 1, 2, ..., p, \\ c/k_n, \text{если } t = p+1, p+2, ..., T, \end{cases}$$

для некоторого неизвестного момента p, $k_n \to \infty$, $k_n/n \to 0$, $c > 0$, а $\varepsilon_t \sim (0, \sigma^2)$ — последовательность независимых и одинаково распределенных случайных величин, Для определения момента p в текущем режиме получения наблюдений построим статистику:

$$K_{N,t} = \frac{\sum\limits_{i=t-N+1}^{t} (\Delta y_i - \tilde{\Delta}_t)(y_{i-1} - \tilde{y}_t)}{\sqrt{\sum\limits_{i=t-N+1}^{t} (\Delta y_i - \tilde{\Delta}_t)^2 \sum\limits_{i=t-N+1}^{t} (y_{i-1} - \tilde{y}_t)^2}},$$

$$t = N + 1, ..., T, \qquad (9)$$

где $\Delta y_i = y_i - y_{i-1}$, $\tilde{\Delta}_t = \frac{1}{N}\sum\limits_{i=t-N+1}^{t} \Delta y_i$, $\tilde{y}_t = \frac{1}{N}\sum\limits_{i=t-N+1}^{t} y_i$, $t = N + 1, ..., T$, T — объем выборки. Эта статистика представляет собой коэффициент корреляции рядов y_i и Δy_i, $i = 1, ..., t$. Проанализируем поведение коэффициента корреляции в отсутствие и при наличии пузыря. Для упрощения записи введем обозначения: $\hat{y}_i = y_i - \tilde{y}_t$, $\Delta_i = \Delta y_i - \tilde{\Delta}_t$ $\hat{\varepsilon}_i = \varepsilon_i - \frac{1}{N}\sum\limits_{i=t-N+1}^{t} \varepsilon_i$. Если ряд описывается уравнением (4), то

$$\Delta_i = \begin{cases} \hat{\varepsilon}_i, \; t = 1, 2, ..., p, \\ \dfrac{c}{k_n}\hat{y}_i + \hat{\varepsilon}_i, \; t = p + N + 1, ..., T. \end{cases}$$

Перепишем выражение для $K_{N,t}$ с учетом введенных обозначений:

$$K_{N,t} = \begin{cases} \dfrac{\sum\limits_{i=t-N+1}^{t} \hat{\varepsilon}_i \hat{y}_{i-1}}{\sqrt{\sum\limits_{i=t-N+1}^{t} \hat{\varepsilon}_i^2 \sum\limits_{i=t-N+1}^{t} \hat{y}_{i-1}^2}}, \; t = N, N+1, ..., p, \\ \dfrac{\sum\limits_{i=t-N+1}^{t} c/k_n \hat{y}_i^2 + \varepsilon_i \hat{y}_i}{\sqrt{\sum\limits_{i=t-N+1}^{t} (c/k_n \hat{y}_i + \hat{\varepsilon}_i)^2 \sum\limits_{i=t-N+1}^{t} \hat{y}_i^2}}, \\ t = p + N + 1, ..., T. \end{cases}$$

Можно показать, что до момента времени $t = p$ значение статистики $K_{N,t}$ с ростом N приближается к нулю, а после, при $t > p$ — к единице. Это асимптотические значения. В интервале $p < t < p + N + 1$ формула для вычисления $K_{N,t}$ представляет собой сумму вида:

$$K_{N,t} = \frac{\sum\limits_{i=t-N+1}^{p} \hat{\varepsilon}_i \hat{y}_{i-1}}{\sqrt{\sum\limits_{i=t-N+1}^{p} \hat{\varepsilon}_i^2 \sum\limits_{i=t-N+1}^{p} \hat{y}_{i-1}^2}} + \frac{\sum\limits_{i=p+1}^{t} (\alpha\hat{y}_i^2 + \varepsilon_i \hat{y}_i)}{\sqrt{\sum\limits_{i=p+1}^{t} (\alpha\hat{y}_i + \hat{\varepsilon}_i)^2 \sum\limits_{i=p+1}^{t} \hat{y}_i^2}}.$$

Если ряды y_i и Δy_i, $i = 1, ..., t$, имеют нормальное распределение, то плотность распределения коэффициента корреляции удовлетворяет уравнению [13]:

$$f_{(\rho, N)} = \frac{(N-2)(1-r^2)^{(N-4)/2}(1-\rho^2)^{(N-1)/2}\Gamma(N-1)}{\sqrt{2\pi}\Gamma(N-1/2)(1-\rho r)^{N-3/2}} \times$$
$$\times \, {}_2F_1\Big(\frac{1}{2}, \frac{1}{2}, \frac{2N-1}{2}, \frac{r\rho+1}{2}\Big),$$

где ρ — значение коэффициента корреляции генеральной совокупности, N — объем выборки, по которой рассчитывался выборочный коэффициент корреляции, r — значение выборочного коэффициента корреляции,

$${}_2F_1\Big(\frac{1}{2}, \frac{1}{2}, \frac{2N-1}{2}, \frac{r\rho+1}{2}\Big) =$$
$$= \frac{\Gamma(N-1/2)}{\Gamma(1/2)\Gamma(N-1)}\int\limits_0^1 t^{-1/2}(1-t)^{N-2}\Big(1 - t\frac{\rho r+1}{2}\Big)^{-1/2} dt$$

— гипергеометрическая функция [14].

Для определения момента возникновения пузыря будем пользоваться алгоритмом последовательного обнаружения изменений в последовательности коэффициентов корреляции: $K_{N,t}$, $K_{N,t+1}$, ..., $K_{N,t+r}$, ... Этот алгоритм рассматривался и подробно описан в работе [8], поэтому здесь мы только кратко перечислим основные шаги, входные и выходные данные.

Входная информация. Набор исторических данных (цен и дивидендов или отношения цен к дивидендам или просто цен) финансового актива, объемом не менее 200—300 наблюдений, включающий в себя периоды, содержащие пузыри, а также моменты возникновения и схлопывания этих пузырей, которые могут быть получены по оценкам экспертов.

Выходная информация. Значение в каждый момент t признака «наличие пузыря»: 1, если пузырь в текущий момент времени существует в процессе, 0 — если пузырь отсутствует, а также моменты возникновения и схлопывания пузырей, обнаруженных до момента t.

Основные шаги алгоритма.

1. Выбор настроечных параметров по историческим данным. Настроечные параметры алгоритма: длина интервала для расчета коэффициента корреляции N; оценки значений коэффициента корреляции генеральной совокупности $\hat{\rho}_0$ (в отсутствие пузыря) и $\hat{\rho}_1$ (при наличии пузыря). С ростом интервала оценивания N эти оценки приближаются к асимптотическим значениям 0 и 1. При конечных N оценки $\hat{\rho}_0$ лежат в интервале $-1 < \hat{\rho}_0 < 0$, а оценки $\hat{\rho}_1$ — в интервале $0 < \hat{\rho}_1 < 1$, и их значения зависят от N. Поэтому при выборе настроечных параметров сначала выбирается значение N. Определение объема выборки N для расчета статистики $K_{n,t}$ основывается на следующих соображениях:

— если пузыри встречаются часто и длительность их невелика, то N желательно уменьшать;

— при больших N усредненные значения $\hat{y}_i = y_i - \tilde{y}_t$ и $\Delta_i = \Delta y_i - \tilde{\Delta}_t$ плохо аппроксимируют значения рядов y_i и Δy_i в окне длины N, с другой стороны при малых значениях N увеличивается дисперсия значений выборочного коэффициента корреляции r, что снижает качество обнаружения.

После выбора значения N по историческим данным вычисляются оценки:

— $\hat{\rho}_1$ как средние значения выборочного коэффициента корреляции в интервале наличия пузыря;

— $\hat{\rho}_0$ как средние значения выборочного коэффициента корреляции научастках, на которых отсутствует пузырь.

2. Задание начальных значений, вычисление порогов[1]**.** По объему выборки N и полученным на шаге 1 оценкам $\hat{\rho}_0$, $\hat{\rho}_1$ коэффициентов генеральной совокупности ρ_0, ρ_1, вычисляются значения верхнего и нижнего порогов по формулам (П1) и (П2) в работе [8] соответственно. Устанавливается значение признака «наличие пузыря», а значение решающей функции g полагают равным нулю.

3. Вычисление значения выборочного коэффициента корреляции. По объему выборки N, значениям рядов наблюдений y_t, их разностям Δy_t в каждый момент времени t, в скользящем окне $[t - N + 1, t]$ вычисляется выборочный коэффициент корреляции (9).

4. Вычисление статистики алгоритма

$$s_{t-1}^{t} = \ln\frac{f_{\theta_1}(r_t)}{f_{\theta_0}(r_t)} =$$

$$= \ln\frac{(1-\hat{\rho}_1^2)^{(N-1)/2}(1-\hat{\rho}_0 r_t)^{N-3/2}\times}{(1-\hat{\rho}_1 r_t)^{N-3/2}(1-\hat{\rho}_0^2)^{(1-1)/2}\times} \to$$

$$\to \frac{\times\left[1+\frac{\hat{\rho}_1 r_t+1}{4(2N-1)}+\frac{9(\hat{\rho}_1 r_t+1)^2}{16(2N-1)(2N-1)}+\ldots\right]}{\times\left[1+\frac{\hat{\rho}_0 r_t+1}{4(2N-1)}+\frac{9(\hat{\rho}_0 r_t+1)^2}{16(2N-1)(2N-1)}+\ldots\right]},$$

где r_t — выборочный коэффициент корреляции в момент времени t, вычисленный по под выборке, $[t - N + 1, t]$, $\hat{\rho}_0$, $\hat{\rho}_1$ — оценки значений коэффициентов корреляции генеральной совокупности на участках без пузыря и с пузырем, соответственно. По вычисленному значению статистики s_{t-1}^{t} решающая функция g_t пересчитывается по формуле $g_t = \max\{hn, g_{t-1} + s_{t-1}^{t}\}$, где hn — нижний порог алгоритма, и сравнивается с верхним порогом hh. Если $g_t > hh$, а $g_{t-1} < hh$, то на участке обнаружен пузырь, и значение признака «наличие пузыря» изменяется на 1. Если $g_t < hh$, а $g_{t-1} > hh$, то это означает «схлопывание» пузыря, и значение признака «наличие пузыря» полагают равным нулю, в остальных случаях значение признака не меняется. При получении нового наблюдения переходят к шагу 3.

4. ИССЛЕДОВАНИЕ И СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЛГОРИТМОВ ИДЕНТИФИКАЦИИ ФИНАНСОВЫХ ПУЗЫРЕЙ

4.1. Описание процедуры моделирования

На настоящий момент качество алгоритмов, предложенных Филлипсом с соавторами, в целом превосходит качество алгоритмов обнаружения и датирования пузырей, предложенных другими исследователями [8, 9]. Поэтому для анализа предложенного алгоритма мы ориентировались на результаты численного моделирования, приведенные в их последних публикациях [7, 12], повторили исследования с использованием их моделей, расширив список проверяемых параметров, и провели соответствующие эксперименты для разработан-

[1] Предполагается, что в момент начала наблюдения известно, существует в процессе пузырь или нет. Если существует, то признак устанавливается равным единице, если нет — нулю.

ного нами алгоритма последовательного анализа (на который в дальнейшем изложении мы будем ссылаться как на SEQCORR — sequential correlation). Алгоритмы SEQCORR, SADF и GSADF сравнивались по ряду параметров:

— по доли ложных обнаружений;

— по точности обнаружения — значениям запаздывания и отклонения длительности пузыря от фактической;

— по доли истинных обнаружений пузырей: обнаружение считается истинным, когда начало обнаруженного пузыря лежит в интервале, в котором моделировался пузырь, и число обнаруженных пузырей совпадает с числом пузырей в имитационной модели.

— по доли ложных «схлопываний» — когда в интервале, в котором моделировался один пузырь, обнаружено несколько пузырей;

— по числу пропущенных пузырей.

Цели проведенного моделирования:

— обоснование применимости для обнаружения и датирования пузырей алгоритма SECCORR;

— сравнение его с алгоритмами SADF и GSADF по перечисленным параметрам;

— исследование чувствительности алгоритмов к изменениям параметров исходного процесса, таких как скорость роста пузыря ($\alpha = 0{,}5$; 0,55; 0,6), длина интервала наблюдения до момента возникновения пузыря, длина интервала между пузырями, длительность пузырей.

4.2. Сравнительный анализ алгоритмов GSADF, SADF и SEQCORR при наличии одного пузыря

Анализ алгоритмов проводился по моделям, используемым в работах [7, 12]. Анализировались случаи, когда анализируемый процесс содержит один пузырь или два пузыря.

В случае одного пузыря процесс моделировался в виде:

$$X_t = X_{t-1}1\{t < t_e\} + \delta_T X_{t-1}1\{t_e \le t \le t_f\} + X_{t-1}1\{t > t_f\} + \varepsilon_t,$$

где $\delta_T = 1 + cT^{-\alpha}$, t_e и t_f — моменты начала и схлопывания пузыря соответственно, $\varepsilon_t \sim N(0, \sigma^2)$ — последовательность независимых и одинаково распределенных случайных величин. Каждый эксперимент повторялся 5000 раз, постоянные параметры эксперимента: $X_0 = 100$, $\sigma = 6{,}79$, $c = 1$, $T = 100$, параметры, изменяющиеся в зависимости от условий эксперимента: t_e, t_f и α.

В качестве критериев эффективности нами рассматривались величины:

1) доля ложных обнаружений (начало обнаруженного пузыря не попадает в интервал $[t_e, t_f]$);

2) среднее смещение оценки момента обнаружения $\hat{t}_e$ пузыря в интервале $[t_e, t_f]$ относительно момента t_e: $\delta_e = |\hat{t}_e - t_e|/T$;

3) среднее смещение оценки момента обнаружения схлопывания пузыря $\hat{t}_f$ относительно t_f: $\delta_f = |\hat{t}_f - t_f|/T$;

4) доля экспериментов, в которых был обнаружен один пузырь;

5) доля экспериментов, в которых было обнаружено больше одного пузыря (все пузыри, для которых оценки момента обнаружения $\hat{t}_e$ лежат в интервале $[t_e, t_f)$, считаем обнаруженными);

6) доля экспериментов, в которых пузыри не были обнаружены.

Для исследования чувствительности алгоритмов проводилась серия экспериментов, в которых значения параметров α, t_e и t_f изменялись (табл. 1—3).

Таблица 1

Результаты исследования качества работы алгоритмов в зависимости от скорости роста пузыря

α	Алгоритм	Критерий эффективности					
		1	2	3	4	5	6
0,6	SADF	24,76	0,08	0,09	73,36	3,86	22,74
	GSADF	22,3	0,08	0,05	64,5	16,3	10,72
	SEQCORR	6,52	0,07	0,11	84,1	0,06	15,84
0,55	SADF	21,46	0,07	0,12	84,84	1,94	16,89
	GSADF	19,34	0,07	0,12	76,62	13,54	8,74
	SEQCORR	6,14	0,06	0,14	91,3	0,04	8,66
0,5	SADF	15,7	0,06	0,14	92,22	0,48	7,28
	GSADF	16,14	0,05	0,1	87,36	7,47	4,74
	SEQCORR	6,34	0,05	0,17	94,96	0	5,04

По результатам, представленным в табл. 1—3 можно сделать следующие выводы.

- Доля ложных обнаружений для алгоритмов SADF и GSADF снижается с увеличением скорости роста пузыря (уменьшением α), практически мало зависит от длительности пузыря, растет по мере увеличения интервала от момента начала наблюдения до момента возникновения пузыря. Эта доля в среднем в 3—4 раза выше, чем у алгоритма SEQCORR, которой малочувствителен к изменению скорости роста пузыря и даты его возникновения относительно начала наблюдения, но увеличивает число ложных обнаружений при уменьшении длительности пузыря.
- Запаздывание в обнаружении малочувствительно к изменению исследуемых параметров. У алгоритма SEQCORR среднее запаздывание в обнаружении возникновения пузыря в среднем на 10 % меньше, чем у алгоритмов SADF и GSADF, но среднее запаздывание в обнаружении окончания пузыря выше в среднем на 30 %, чем у алгоритмов SADF и GSADF.
- Доля обнаружения одного пузыря в рассматриваемом интервале $[t_e, t_f]$ у всех алгоритмов повышается с ростом скорости пузыря и ростом длительности пузыря, а снижается с увеличением интервала от момента начала наблюдения до момента возникновения пузыря. Средняя доля обнаружения одного пузыря составляет 75,3, 67,5 и 84,7 % для алгоритмов SADF, GSADF и SEQCORR соответственно. У алгоритма GSADF доля обнаружения двух пузырей значительно выше, чем у алгоритмов GSADF и SEQCORR. Обнаружение двух и более пузырей означает, что в интервале наличия пузыря происходят ложные «схлопывания». Алгоритм GSADF можно считать более эффективным, чем алгоритм SADF по числу обнаружений, только если рассматривать обнаружение двух пузырей в интервале наличия пузыря как одно обнаружение, без учета ложных «схлопываний».
- Доля обнаружения двух пузырей составляет 3,8, 15 и 0,1 % алгоритмов SADF, GSADF и SEQCORR, соответственно.
- Доля пропущенных пузырей у всех алгоритмов снижается с ростом скорости и ростом длительности пузыря и увеличивается по мере увеличения интервала от момента начала наблюдения от момента возникновения пузыря. Эта

Таблица 2

Результаты исследования качества работы алгоритмов в зависимости от длительности пузыря (при $t_f - t_e$ = 10; 15; 20)

Длина пузыря	Алгоритм	Критерий эффективности					
		1	2	3	4	5	6
10	SADF GSADF SEQCORR	22,54 21 10,08	0,07 0,06 0,06	0,06 0,03 0,11	53,6 54,68 64,98	2,28 10,18 0,04	44,12 34,38 34,98
15	SADF GSADF SEQCORR	24,76 22,3 6,52	0,08 0,08 0,07	0,09 0,05 0,11	73,36 64,5 84,1	3,86 16,3 0,06	22,74 10,72 15,84
20	SADF GSADF SEQCORR	21,92 20,34 5,68	0,09 0,09 0,08	0,03 0,09 0,14	82,24 68,88 89,88	4,88 19,18 0,2	12,78 8,09 9,92

Таблица 3

Результаты исследования зависимости качества обнаружения от момента возникновения пузыря относительно точки начала наблюдения (при $t_f - t_e$ = 15)

t_e	Алгоритм	Критерий эффективности					
		1	2	3	4	5	6
21	SADF GSADF SEQCORR	13,6 17,38 6,18	0,09 0,09 0,07	0,06 0,05 0,12	75,6 62,92 88,6	10,42 21 0,06	10,13 13,06 11,0
41	SADF GSADF SEQCORR	24,76 22,3 6,52	0,08 0,08 0,07	0,09 0,05 0,11	73,36 64,5 84,1	3,86 16,3 0,06	22,74 10,72 15,24
61	SADF GSADF SEQCORR	28,38 27,12 6,82	0,09 0,08 0,07	0,07 0,08 0,12	68,92 63,36 80,42	2,42 14,62 0,14	28,56 20,06 19,44

доля самая высокая у алгоритма SADF, алгоритмы GSADF и SECCORR сравнимы по этому показателю, но у алгоритма GSADF его значение в среднем на 4 % меньше, чем у алгоритма SECCORR.

4.3. Сравнительный анализ алгоритмов GSADF, SADF и SEQCORR при наличии двух пузырей

Модель процесса с двумя пузырями имеет вид:

$$X_t = X_{t-1}1\{t < t_{e_1}\} + \delta_T X_{t-1}1\{t_{e_1} \le t \le t_{f_1} \cup$$
$$\cup\ t_{e_1} \le t \le t_{f_2}\} + X_{t-1}1\{t_{f_1} < t < t_{e_2} \cup t_{f_2} < t\} + \varepsilon_t,$$

где $\delta_T = 1 + cT^{-\alpha}$, t_{e_1} и t_{e_2} — моменты начала, t_{f_1} и t_{f_2} — моменты схлопывания первого и второго пузырей соответственно, $\varepsilon_t \sim N(0, \sigma^2)$ — последовательность независимых и одинаково распределенных случайных величин. Каждый эксперимент повторялся 5000 раз, постоянные параметры эксперимента: $X_0 = 100$, $\sigma = 6{,}79$, $c = 1$, $\alpha = 0{,}6$, $T = 100$, $t_{e_1} = 21$, $t_{e_2} = 61$, изменяющиеся в зависимости от эксперимента значений t_{f_1} и t_{f_2}. Проверялась способность алгоритмов обнаруживать два пузыря при различных соотношениях длительностей первого и второго пузырей. Для двух пузырей рассматривались критерии эффективности:

1) доля ложных обнаружений (начало обнаруженного пузыря не попадает в интервал $[t_{e_i}, t_{f_i})$, $i = 1, 2$);

2) доля экспериментов, в которых обнаружен только один пузырь;

3) доля экспериментов, в которых обнаружено ровно два пузыря;

4) доля экспериментов, в которых обнаружено больше двух пузырей (все пузыри, для которых оценки момента обнаружения $\hat{t}_{e_i}$ лежат в интервале $[t_{e_i}, t_{f_i})$, $i = 1, 2$, считались обнаруженными);

5) доля экспериментов, в которых пузыри не были обнаружены.

Таблица 4

Результаты исследования качества алгоритмов при наличии двух пузырей

Длина: первого пузыря	Длина: второго пузыря	Алгоритм	Критерий эффективности: 1	2	3	4	5
20	10	SADF	17,86	4,62	79,28	14,18	1,92
		GSADF	15,42	3,68	65,36	29,6	1,36
		SEQCORR	2,1	16,28	82,1	0,32	1,3
	15	SADF	15,84	5,28	79,54	14,22	0,96
		GSADF	13,04	3,9	65,64	29,82	0,64
		SEQCORR	1,94	16,74	82,38	0,32	0,56
	20	SADF	11,92	5,66	79,58	14,22	0,54
		GSADF	10,24	4,06	65,74	29,6	0,34
		SEQCORR	0,9	17	82,4	0,32	0,28
15	10	SADF	23,68	10,34	72,58	12,24	4,84
		GSADF	19,66	9,28	61,42	26,28	3,02
		SEQCORR	3,38	9,38	87,44	0,2	2,98
	15	SADF	21,44	11,32	73,94	12,56	2,18
		GSADF	17,7	9,72	61,82	26,9	1,5
		SEQCORR	2,88	10,24	88,12	0,2	1,44
	20	SADF	15,5	12,08	74,12	12,56	1,24
		GSADF	15,38	24,66	50,76	21,32	3,26
		SEQCORR	2,14	10,8	88,16	0,2	0,84
10	10	SADF	30,22	28,98	52,16	7,24	11,62
		GSADF	22,96	25,34	50,92	17,36	7,08
		SEQCORR	9,04	27,1	65,14	0,1	7,58
	15	SADF	26,58	30,22	56,46	7,56	5,76
		GSADF	21,32	26,08	52,06	18,44	3,42
		SEQCORR	7,56	29,96	66,5	0,1	3,44
	20	SADF	17,94	32,52	56,78	7,64	3,06
		GSADF	16,48	26,9	52,5	18,74	1,86
		SEQCORR	6,74	31,08	66,5	0,1	2,26

В табл. 4 приведены результаты экспериментов.

По результатам экспериментов с двумя пузырями можно сделать следующие выводы.

- Средняя доля ложных обнаружений у алгоритмов SADF, GSADF и SEQCORR составляет 20,1, 16,9 и 4,1 % соответственно. Она увеличивается при сокращении длительности первого пузыря для всех алгоритмов. Алгоритм SEQCORR наиболее чувствителен к уменьшению длительности первого пузыря, но доля ложных обнаружений для этого алгоритма даже в наихудшем случае не превышает доли ложных обнаружений алгоритма GSADF (10,24 %).
- Средняя доля обнаружения двух пузырей повышается при увеличении длительности первого пузыря. Она составляет 69,4, 58,5 и 78,8 %для алгоритмов SADF, GSADF и SEQCORR соответственно, что противоречит утверждению о том, что алгоритм GSADF более эффективен, чем алгоритм SADF при обнаружении двух пузырей. Так же, как и в случае обнаружения одного пузыря, эффективность алгоритма GSADF снижается из-за ложных «схлопываний» в интервале наличия пузыря. Для корректировки ложных «схлопываний» при датировании пузырей алгоритмами GSADF и SADF в работах [5—7, 12] предложена процедура коррекции длительности пузыря (см. формулы (6) и (8)). Однако наши эксперименты не подтверждают ее эффективности.
- Доля обнаружения трех и более пузырей составляет 11,4, 24,2 и 0,2 % для алгоритмов SADF, GSADF и SEQCORR, соответственно.
- Доля пропущенных пузырей у всех алгоритмов увеличивается по мере сокращения длительности пузыря. Она самая высокая у алгоритма GSADF, алгоритмы SADF и SECCORR сравнимы по этому показателю.

ЗАКЛЮЧЕНИЕ

Проведено сравнение алгоритмов обнаружения пузырей, предложенных в работах, проводимых Филлипсом с соавторами [5—7, 12], с алгоритмом последовательного обнаружения коэффициента корреляции [8]. Поскольку сравниваемые алгоритмы обладают одинаковым набором функций, т. е. могут обнаруживать моменты, возникновения и схлопывания пузырей в реальном времени, то процедуры сравнения были направлены на оценку качества выполнения этих функций. Поэтому для сравнения применялось численное моделирование, позволяющее набирать соответствующую статистику. Алгоритм SEQCORR сравнивался с алгоритмом GSADF на реальных данных (отношение цен и дивидендов фондового индекса S & P500) проводилось в работе [8]. Как показал проведенный эксперимент, алгоритм SEQCORR обнаруживает и датирует все события с января 1871 г. по декабрь 2010 г., определенные экономистами как финансовые пузыри, причем датировка событий приближается к оценке, полученной экономистами, а алгоритм GSADF /пропускает одно событие. Результаты сравнения, проведенные с помощью моделирования по методу Монте-Карло, показали, что алгоритм SEQCORR ничем не уступает, а по части критериев даже превосходит алгоритмы, предложенные в работах [5—7].

ЛИТЕРАТУРА

1. *Mills T.C., Markellos R.N.* The Econometric Modelling of Financial Time Series. — Cambrige: Cambrige University Press, 3rd Edition, 2008. — P. 472.
2. *Diba B.T., and Grossman H.I.* Explosive rational bubbles in stock prices? // The American Economic Review. — 1988. — Vol. 78. — P. 520—530.
3. *Evans G.W.* Pitfalls in Testing for Explosive Bubbles in Asset Prices // American Economic Review. — 1991. — Vol. 81. — P. 922—930.
4. *Канторович Г.Г.* Лекции: Анализ временных рядов // Экономический журнал Высшей школы экономики. — 2002. — Т. 6, № 3. — С. 379—401.
5. *Phillips P.C.B., Wu Y., Yu J.* Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? // International Economic Review. — 2011. — Vol. 52, N 1.
6. *Phillips P.C.B., Shi S.P., and Yu J.* Testing for Multiple Bubbles 1: Historical Episodes of Exuberance and Collapse in the S&P 500 / Working Paper. — 2013.
7. *Phillips P.C.B., Shi S.P., and Yu J.* Testing for multiple bubbles: historical episodes of exuberance and collapse in the s & p 500 // International Economic Review. — 2015. — Vol. 56, N 4. — P. 1043—1077.
8. *Гребенюк Е.А., Малинкина А.В.* Применение методов эконометрического анализа данных для идентификации и датирования «пузырей» на финансовых рынках // Проблемы управления. — 2014. — № 5. — С. 50—58.
9. *Homm U., Breithing J.* Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods // Journal of Financial Econometrics. — 2012. — Vol. 10, N 1. — P. 198—231.
10. *Суслов В.И., Ибрагимов Н.М.* и др. Эконометрия. — Новосибирск: Изд-во СОРАН, 2005. — 744 с.
11. *Phillips P.C.B., Magdalinos T.* Limit theory for moderate deviations from a unit root// Journal of Econometrics. — 2007. — 136. — P. 115—130. — URL: http://korora.econ.yale.edu/phillips/pubs/art/p1172.pdf (дата обращения 3.05.2017).
12. *Phillips P.C.B., Shi S.P., and Yu J.* Testing for multiple bubbles: limit theory of real time detectors / September 2013, cowles foundation discussion paper no. 1915.
13. *Kenney J.F., and Keeping E.S.* Mathematics of Statistics. Pt. 2: 2nd ed. — Princeton, NJ: Van Nostrand, 1951.
14. *Лебедев Н.Н.* Специальные функции и их приложения. — М.-Л.: Физматгиз,1963.

Статья представлена к публикации членом редколлегии В.В. Клочковым.

Гребенюк Елена Алексеевна — д-р техн. наук,
гл. науч. сотрудник, Институт проблем управления
им. В.А. Трапезникова РАН, г. Москва,
✉ lngrebenuk@rambler.ru,

Малинкина Антонина Валерьевна — гл. специалист,
ОАО «Банк Москвы», ✉ malinkinaav@gmail.com.