УДК 330.4

© Щелкалин В. Н., 2011
Харьковский национальный университет радиоэлектроники, г. Харьков

МЕТОД “ГУСЕНИЦА”-SSA – АРПСС – СПОАРУГ И МОДЕЛЬ АРСПСС – СПОАРУГ ДЛЯ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ФИНАНСОВО-ЭКОНОМИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ

Проводиться подальший розвиток методу Бокса-Дженкінса побудови моделей і вдосконалення моделей авторегресії – проінтегрованого ковзного середнього (АРПСС), які були розроблені і успішно використовуються близько сорока років у різних галузях науки і техніки і залишаються донині однією з найбільш ефективних моделей для моделювання і прогнозування фінансово-економічних часовіх рядів, перевершуючи своїх конкурентів по цілому ряду критеріїв, таких як: економність за кількістю параметрів, трудомісткість алгоритмів побудови моделей та ресурсомісткість їх реалізації, формалізацією і автоматизацією їхньої побудови. Запропоновано чергову модифікацію моделей ARIMA і GARCH.

Ключові слова: моделювання, фільтрація, прогнозування, управління, модель АРІКС, метод «Гусениця»-SSA – АРІКС – СПУАРУГ, модель АРСІКС, модель АРСІКС – СПУАРУГ, гетероскедастичність, метод Левенберга-Марквардта.

Постановка проблемы. Научно-технический прогресс постоянно развивает технические, социальные, медицинские в том числе и экономические и прочие системы, усложняя их структуру и увеличивая количество их сложных внутренних взаимосвязей и внешних факторов, от которых они зависят и большинство из которых учесть невозможно. Развитие финансовых рынков сопровождается значительным ростом вовлечённых объёмов капитала, увеличением количества участников рынка. Одним из важных направлений исследований является моделирование динамики доходности и волатильности финансовых рынков. Поэтому актуальной является разработка и использование универсальных математических моделей и методик, позволяющих моделировать, прогнозировать и управлять широким классом процессов, т.к. это позволит повысить эффективность управления и планирования режимов работы сложных экономических систем, сэкономить значительные средства и ресурсы путем всяких математических ухищрений, а также существенно разгружает работу персонала различных организаций, позволяет предвидеть рисковые ситуации и привносит еще много прочей пользы.

Главными требованиями к построению математических моделей в 70 – 90-е годы прошлого века, являлись экономность по количеству параметров, скорость определения модели и её ресурсоёмкость для использования на доступных тогда ЭВМ малой производительности. Однако, современная вычислительная техника и методы математического моделирования предоставляют большие возможности для анализа, моделирования и прогнозирования временных рядов различной природы. Поэтому в настоящее время указанные требования не являются определяющими и современные вычислительные средства и системы позволяют выносить на первый план требования точности моделирования, качества анализа и прогнозирования.

Анализ последних исследований и публикаций. В 70-е годы прошлого века одним из полезных инструментов анализа и прогнозирования финансово-экономических временных рядов считались модели ARIMA и метод Бокса-Дженкинса их построения [1].

Метод сезонной ARMA работает только с предварительно приведенными к стационарному виду временными рядами. Нестационарный ряд обычно характеризуется присутствием большой мощности на низких частотах. Важной является задача элиминирования трендовой составляющей временного ряда. В таких случаях всё, что делалось, — это отфильтровывались нестационарные низкочастотные компоненты и использовался оставшийся ряд для дальнейшего анализа. При этом в качестве фильтра для устранения низкочастотных компонент в моделях ARIMA использовался фильтр первых разностей, максимум вторых. А поэтому метод сезонной ARIMA удовлетворительно моделировал и прогнозировал временные ряды только относительно простой структуры.

В 80-е же годы прошлого века Гренджер и Джойо [2] предложили новый класс ARFIMA моделей, удобно описывающего финансово-экономические временные ряды с эффектами длинной и короткой памяти. 2000-е годы характеризуются применением большого спектра моделей для анализа и прогнозирования финансово-экономических временных рядов, а также ансамблей моделей с различной структурой. С появлением высокоскоростных ЭВМ происходил и переход от ансамблей прогнозирующих моделей к ихним комбинациям. Отличие комбинированных моделей от их ансамблей заключается в одновременном подстраивании параметров моделей.

Наиболее важными характеристиками моделей при анализе и выборе наиболее подходящих математических моделей в экономике являются следующие характеристики:

- способ моделирования трендовой составляющей временного ряда;
- способ нелинейного моделирования временного ряда;
- способ моделирования случайной составляющей временного ряда;
- способ учёта влияния внешних факторов на процесс.

Поэтому приоритетным видом моделей являются комбинированные вероятностно-детерминированные прогнозные модели с нелинейным усложнением, т.к. при этом в модели одновременно учитываются и используются как статистические, так и детерминированные составляющие, что позволяет достичь наилучшего качества прогнозирования [3]. Среди детерминированных моделей приоритетным вариантом является детерминированная модель спектрального разложения, реализующая моделирование на основе разложения по детерминированному ортонормированному базису, отличному от гармонических функций. В то время как наиболее приоритетной среди вероятностных моделей является модель авторегрессии – проинтегрированного скользящего среднего.

Формулирование целей статьи**.** В зарубежной литературе давно известны комбинированные вероятностно-детерминированные модели, представленные в [3]. В данной же статье предлагается очередная модификация моделей ARIMA и GARCH, впервые предложены метод "Гусеница"-SSA – АРПСС – СПОАРУГ, комбинированная вероятностно-детерминированная модель авторегрессии – спектрально проинтегрированного скользящего среднего со спектрально проинтегрированной обобщённой авторегрессионной условной гетероскедастичностью (АРСПСС – СПОАРУГ, на англ. яз. ARSIMA – SIGARCH, на укр. яз. АРСІКС – СІУАРУГ) и метод её построения, которые позволяют более гибко анализировать, моделировать и прогнозировать финансово-экономические временные ряды в сравнении с моделями ARIMA – GARCH.

Изложение основного материала исследования**.** Суть метода сначала заключалась в многомерной декомпозиции экзогенных временных рядов и прогнозируемого временного ряда, на базисные латентные компоненты, включая их степени и комбинации, получаемые методами главных (PCA), гладких (SCA) и независимых компонентов (ICA) в отборе из этих базисных компонентов конструктивных методом быстрого ортогонального поиска (FOS), одного из самых эффективных и экономных по временными затратами методов в зарубежной литературе, и отсечении деструктивных, тем самым сформировав передаточную функцию математической модели процесса; в дальнейшей идентификации шумовой части математической модели процесса благодаря сезонным моделям авторегрессии – скользящего среднего, и одновременной параметрической идентификации полученной структуры модели методом Левенберга-Марквардта [4]. Как стало известно позже, пердлоденный метод схож с декомпозиционным методом моделирования (ДММ) [3]. Метод “Гусеница”-SSA тоже использует декомпозицию временных рядов по сингулярным значениям (SVD). Известны публикации использования метода “Гусеница”-SSA в различных отраслях науки и техники как метода достаточно хорошо описывающего нестационарные временные ряды с линейными, параболическими или экспоненциальными трендами с не всегда устойчивой колебательной составляющей, однако метод для моделирования использует неоптимальный с точки зрения точности воспроизведения некоторых временных рядов ортогональный базис векторов траекторной матрицы. Поэтому было предложено совместно использовать рекуррентную модель прогнозирования метода “Гусеница”-SSA и модель авторегрессии – скользящего среднего, обученных на конкурентной основе с учётом обобщённого критерия точности и адекватности. Использование такой комбинации было продиктовано тем, что отдельно эти подходы имеют ряд недостатков, а совместное их использование приводит к синергии, повышая их эффективность, робастность и адекватность. Однако при выделении тренда методом “Гусеница”-SSA, как и любым другим методом, остаточная составляющая ряда в большинстве случаев остаётся нестационарной, поэтому далее метод “Гусеница”-SSA использовался уже в комбинации с моделью авторегрессии – проинтегрированного скользящего среднего (АРПСС). При этом совместное использование приведенных выше методов подразумевает, что параметрическая идентификация осуществляет вычисление оценок параметров модели АРПСС и нелинейного обобщения авторегрессии на главных компонентах, минимизируя сумму квадратов отклонений ошибок моделирования с учетом временного ряда, получаемого методом “Гусеница”-SSA, который в течение моделирования не имеет модели и соответственно параметров, а метод Левенберга-Марквардта, вычисляя параметры остальных аддитивных частей модели помогает определить внешний вид временного ряда метода “Гусеница”-SSA, выступающего в роли помощника определения детерминированной (трендовой) составляющей процесса.

Предлагаемый подход как раз и является одним из вариантов приоритетных на сегодняшний день комбинированных вероятностно-детерминированных подходов. Здесь также реализуется так называемый трендовый подход, когда процесс моделируется как отклонение фактических значений от тренда (который представлен здесь временным рядом, получаемым методом “Гусеница”-SSA) и который обеспечивает устойчивость получаемой модели и достаточную точность моделирования, тогда как ранее вероятностной моделью АРПСС пытались описать весь процесс. Таким образом, произведена успешная попытка спустя более тридцати лет после создания метода Бокса-Дженкинса и метода “Гусеница”-SSA объединить их. Однако для удовлетворения таких требований к моделям, как: скорость обучения, трудоёмкость, ресурсоёмкость, наглядность модели, простота использования и интерпретируемость, затратный по времени и ресурсам метод “Гусеница”-SSA в дальнейшем было предложено использовать лишь для предварительной структурной идентификации и грубой параметрической идентификации так называемого интегрирующего полинома от оператора задержки предлагаемой модели, а также для грубой структурной и параметрической идентификации полинома от оператора задержки, наличие которого отличает более общую полиномиальную модель от модели Бокса-Дженкинса, структура и коэффициенты которого сначала равны таковыми рекуррентной модели прогнозирования метода “Гусеница”-SSA.

Метод “Гусеница”-SSA также предлагается использовать для предварительного обобщённого коинтегрирования временных рядов при моделировании многосвязных процессов, а также для разделения отдельно на финитный и отдельно на апериодический регуляторы в случае использования предлагаемой модели в теории автоматического управления [5]. Также возможно нелинейное усложнение передаточной функции модели одним из способов: FOS, GMDH, RBF, LARS, построенных на главных компонентах, их степенях и сочетаниях. Поэтому, в первую очередь, предлагаемая автором модель направлена на теорию автоматического управления, моделирование и прогнозирование технических систем и технологических процессов, в связи с тем, что их передаточные функции более детерминированные и имеют сложную нелинейную структуру.

Описание предлагаемых математических моделей. Математическая модель процессов, зависящих от нескольких экзогенных факторов в операторной форме может быть представлена в виде модели сезонной авторегрессии – проинтегрированного скользящего среднего (САРПССЭ) [6]:

$$z_t^y = \sum_{i=1}^{N} \frac{b_{n_{b^i}}^i(L)}{a_{n_{a^i}}^i(L)} \cdot z_{t-m_i}^{x^i} + \frac{c_{n_c^\Sigma}^\Pi(L)}{d_{n_d^\nabla}^\nabla(L)} \cdot e_t \quad \textbf{(1)}$$

где L – оператор сдвига по времени на одну единицу назад, такой что $L^i x_t = x_{t-i}$, N – количество экзогенных переменных; z_t^y – пронормированный от 0 до 1 по формуле $z_t^y = \dfrac{y_t - y_t^{\min}}{y_t^{\max} - y_t^{\min}}$ или каким-либо другим способом временной ряд y_t моделируемого и прогнозируемого процесса с вычтенным средним значением; $z_{t-m_i}^{x^i}$ – пронормированный таким же образом i-й экзогенный временной ряд x_t^i с вычтенным средним значением; m_i – задержка i-го экзогенного временного ряда x_t^i по времени относительно прогнозируемого временного ряда y_t; $a_{n_{a^i}}^i(L)$, $b_{n_{b^i}}^i(L)$ – полиномы от L степеней n_{a^i} и n_{b^i} соответственно; $c_{n_c^\Sigma}^\Pi(L) = c_{n_c^1}^1(L^{s_1}) \cdot c_{n_c^2}^2(L^{s_2}) \times \ldots \times c_{n_c^{n_s}}^{n_s}(L^{s_{n_s}}) = \prod_{i=1}^{n_s} c_{n_c^i}^i(L^{s_i})$ – полином L^{s_i} степени n_c^i, определяющий составляющую скользящего среднего периодической компоненты с периодом s_i, $n_c^\Sigma = \sum_{i=1}^{n_s} n_c^i \cdot s_i$;

$$d_{n_d^\nabla}^\nabla(L) = d_{n_d^\Sigma}^\Pi(L) \nabla_{s_1}^{D_1} \nabla_{s_2}^{D_2} \ldots \nabla_{s_{n_s}}^{D_{n_s}} = d_{n_d^1}^1(L^{s_1}) \cdot d_{n_d^2}^2(L^{s_2}) \cdot \ldots \cdot d_{n_d^{n_s}}^{n_s}(L^{s_{n_s}}) \nabla_{s_1}^{D_1} \nabla_{s_2}^{D_2} \ldots \nabla_{s_{n_s}}^{D_{n_s}} =$$

$$= \prod_{i=1}^{n_s} d_{n_d^i}^i(L^{s_i}) \nabla_{s_1}^{D_1} \nabla_{s_2}^{D_2} \ldots \nabla_{s_{n_s}}^{D_{n_s}},$$

$$n_d^\nabla = n_d^\Sigma + \sum_{i=1}^{n} D_i \cdot s_i,$$

$$d_{n_d^\Sigma}^\Pi(L) = d_{n_d^1}^1(L^{s_1}) \cdot d_{n_d^2}^2(L^{s_2}) \times \ldots \times d_{n_d^{n_s}}^{n_s}(L^{s_{n_s}}) =$$

$= \prod_{i=1}^{n_s} d_{n_d^i}^i(L^{s_i})$ – полином от L^{s_i} степени n_d^i, определяющий составляющую авторегрессии сезонной компоненты с периодом s_i, $n_d^\Sigma = \sum_{i=1}^{n_s} n_d^i \cdot s_i$; e_t – остаточные ошибки модели: D_i – порядок взятия разности s_i; ∇_{s_i} и L^{s_i} – упрощающие операторы такие, что $\nabla_{s_i} y_t = (1 - L^{s_i}) \cdot y_t = y_t - y_{t-s_i}$.

Выражение (1) в более компактной форме можно представить в виде:

$$\tilde{a}(q) \cdot z_t^y = \sum_{i=1}^{k} \tilde{b}^i(q) \cdot z_{t-m_i}^{x^i} + \tilde{c}(q) \cdot e_t,$$

где $\tilde{a}(L) = d_{n_d^\nabla}^\nabla(L) \cdot \prod_{i=1}^{N} a_{n_{a^i}}^i(L)$,

$$\tilde{b}^i(L) = b_{n_{b^i}}^i(L) \cdot d_{n_d^\nabla}^\nabla(L) \cdot \prod_{j=1, i \neq j}^{N} a_{n_{a^j}}^j(L),$$

$$\tilde{c}(q) = c_{n_c^\Sigma}^\Pi(q) \cdot \prod_{i=1}^{N} a_{n_{a^i}}^i(q), \; i = \overline{1, N}.$$

Выражение для прогноза с упреждением l при помощи предложенной модели совместного использования модели метода “Гусеница”-SSA и сезонной модели авторегрессии – проинтегрированного скользящего среднего с экзогенными переменными, после приведения ее от дробно-разностного (рационального) к разностному уравнению принимает вид:

$$\hat{y}_t(l) = \hat{y}_t^{SSA}(l) + \sum_{j=1}^{n_d^{\Sigma}+\sum_{i=1}^{N} n_{a^i}} \widetilde{a}_j \cdot h_{t+l-j}^{y} +$$

$$+\sum_{i=1}^{N}\left(\widetilde{b}_0^i \cdot x_{t+l-m_i}^i - \sum_{j=1}^{n_d^{\Sigma}+n_{b^i}+\sum_{p=1}^{N} n_{a^p}} \widetilde{b}_j^i \cdot x_{t+l-m_i-j}^i\right) - \sum_{j=1}^{n_c^{\Sigma}+\sum_{i=1}^{N} n_{a^i}} \widetilde{c}_j \cdot e_{t+l-j};$$

$$\hat{x}_t^i(l) = \hat{x}^k{}_t^{SSA}(l) + \sum_{j=1}^{n_d^{x^k\,\Sigma}} \widetilde{a}_{x^k\,j} \cdot z_{t+l-j}^{x^k} + \sum_{j=1}^{n_c^{x^k\,\Sigma}} \widetilde{c}_{x^k\,j} \cdot e_{x^k\,t+l-j}, k = \overline{1,N},$$

(2)

где $a_{n_{a^i}}^i(L)$, $b_{n_{b^i}}^i(L)$ – полиномы от L степеней n_{a^i} и n_{b^i} соответственно;

$$y_{t+j} = \begin{cases} y_{t+j}, j \le 0, \\ \hat{y}_t(j), j > 0, \end{cases} \qquad x_{t+j}^i = \begin{cases} x_{t+j}^i, j \le 0, \\ \hat{x}_t^i(j), j > 0, \end{cases}$$

$i = \overline{1,N}$; $\qquad e_{t+j} = \begin{cases} e_{t+j}, j \le 0, \\ 0, j > 0. \end{cases}$;

$h_i^y = y_i - \widetilde{w}_i^{N+1}$; $\qquad \hat{y}_t^{SSA}(i) = \sum_{j=1}^{L-1} f_j^y \cdot \widetilde{w}^y{}_{t+i-j}^{N+1}$,

$i = \overline{1,l}$; $\qquad \widetilde{w}^y{}_i^{N+1} = \begin{cases} \hat{y}_i, i > t, \\ \widetilde{w}^y{}_i^{N+1}, i \le t; \end{cases}$

$$(f_{L-1}^y, f_{L-2}^y, \ldots, f_1^y)^T = \frac{1}{1-\sum_{i=1}^{R}\left(u_L^i\right)^2} \sum_{i=1}^{R} u_L^i \cdot \begin{pmatrix} u_1^i & u_2^i & \ldots & u_{L-1}^i \end{pmatrix}^T$$

, где $\begin{pmatrix} u_1^i & u_2^i & \ldots & u_{L-1}^i \end{pmatrix}^T$ – вектор, состоящий из первых $(L-1)$ элементов i-того собственного вектора U^i сингулярного разложения траекторной матрицы экзогенных и моделируемого процессов

$$X = \begin{pmatrix} X_1 & X_2 & \ldots & X_N & Y \end{pmatrix} = \begin{pmatrix} X_{1,1} & X_{1,2} & \ldots & X_{1,K} & X_{2,1} & X_{2,2} & \ldots \\ \ldots & X_{2,K} & \ldots & X_{N,1} & X_{N,2} & \ldots & X_{N,K} & Y_1 & Y_2 & \ldots & Y_K \end{pmatrix}$$

; $Y_j = \begin{pmatrix} y_{j-1} & y_{j-2} & \ldots & y_{j+L-2} \end{pmatrix}^T$, $X_{i,j} = \begin{pmatrix} x_{j-1}^i & x_{j-2}^i & \ldots & x_{j+L-2}^i \end{pmatrix}$, $i = \overline{1,N}$, $j = \overline{1,K}$ – преобразования исходного $(N+1)$-мерного временного ряда (N – количество экзогенных временных рядов и один прогнозируемый) в последовательность L^y – мерных векторов (L^y – ширина окна), число которых равно $(N+1)\cdot K$, $K = n - L^y + 1$, где n – длина временных рядов; u_L^i – последний элемент вектора U^i; R – количество элементов сингулярного разложения; $\hat{y}_t^{SSA}(i) = \sum_{j=1}^{L-1} f_j^y \cdot \widetilde{w}^y{}_{t+i-j}^{N+1}$ – модель рекуррентного прогнозирования метода “Гусеница”-SSA временного ряда y_t, $t = \overline{0, n-1}$; которая в свою очередь нередко может быть экономно записана сезонной моделью АРПСС (АРП) или при помощи модели распределенных лагов Алмон, где $\widetilde{w}^y{}_0^{N+1}$, $\widetilde{w}^y{}_1^{N+1}$, …, $\widetilde{w}^y{}_{n-1}^{N+1}$ – ряд соответствующий преобразованию прогнозируемого временного ряда y_t; ряды $\widetilde{w}^{x^i}{}_0$, $\widetilde{w}^{x^i}{}_1$, …, $\widetilde{w}^{x^i}{}_{n-1}$, $i = \overline{1,N}$ соответствуют преобразованиям i-ых временных рядов, соответствующих экзогенным временным рядам, при помощи сингулярного спектрального анализа на этапе диагонального усреднения, переводящего матрицу $\widetilde{Z}^i$, $i = \overline{1, N+1}$ состоящую из K столбцов от $(i-1)\cdot K$-го до $i \cdot K - 1$-го матрицы Z в ряды $\widetilde{w}^{x^i}{}_0$, $\widetilde{w}^{x^i}{}_1$, …, $\widetilde{w}^{x^i}{}_{n-1}$, по формуле

$$\widetilde{w}^y{}_i^k = \begin{cases} \frac{1}{i+1}\sum_{j=1}^{k+1} \widetilde{z}_{j,i-j+2}^i, i = \overline{0, \min(L,K)-2}; \\ \frac{1}{\min(L,K)} \sum_{j=1}^{\min(L,K)} \widetilde{z}_{j,i-j+2}^i, i = \overline{\min(L,K)-1, \max(L,K)-1}; \\ \frac{1}{n-i} \sum_{j=k-\max(L,K)+2}^{n-\max(L,K)+1} \widetilde{z}_{j,i-j+2}^i, i = \overline{\max(L,K), n-1}, \end{cases}$$

$k = \overline{1, N+1}$; где $Z = \widetilde{Z}^i + \ldots + \widetilde{Z}^j$ – сумма $\widetilde{Z}^i = \left(U^i \cdot \left(U^i\right)^T \cdot X_1 \quad U^i \cdot \left(U^i\right)^T \cdot X_2 \quad \ldots \quad U^i \cdot \left(U^i\right)^T \cdot X_N \quad U^i \cdot \left(U^i\right)^T \cdot Y\right)$ матриц разложения, отобранных стандартным анализом собственных чисел траекторной матрицы в методе “Гусеница”-SSA. Имеется и модификация рекуррентного метода SSA-прогнозирования – векторное

SSA-прогнозирование [7], которое в ряде случаев позволяет получать более точные прогнозы. $\hat{x}^{k\,SSA}_{t}(i)=\sum_{j=1}^{L^y-1} f_j^{x^k}\cdot \widetilde{w}^{x^k\,N+1}_{t+i-j}$, $k=\overline{1,N}$, тоже самое, только для k-того экзогенного временного ряда, а также все переменные и параметры с индексом x^i, $i=\overline{1,N}$ аналогично интерпретируются как для временного ряда y_t, $t=\overline{1,n}$;

С нелинейным усложнением передаточной функции первое уравнение в (2) принимает вид:

$$\hat{z}_t^y(l)=\hat{y}_t^{SSA}(l)+\sum_{i=1}^{r} g_i\cdot p_t^i+\sum_{j=1}^{n_d^{\Sigma}+\sum_{i=1}^{N} n_{a^i}} \widetilde{a}_j\cdot h_{t+l-j}^y+$$

$$+\sum_{i=1}^{N}\left(\widetilde{b}_0^i\cdot z_{t+l-m_i}^{x^i}-\sum_{j=1}^{n_d^{\Sigma}+n_{b^i}+\sum_{p=1}^{N} n_{a^p}} \widetilde{b}_j^i\cdot z_{t+l-m_i-j}^{x^i}\right)-\sum_{j=1}^{n_c^{\Sigma}+\sum_{i=1}^{N} n_{a^i}} \widetilde{c}_j\cdot e_{t+l-j},$$

$$\hat{x}_t^i(l)=\hat{x}^{k\,SSA}_{t}(l)+\sum_{j=1}^{n_d^{x^k\,\Sigma}} \widetilde{a}_{x^k\,j}\cdot z_{t+l-j}^{x^k}+\sum_{j=1}^{n_c^{x^k\,\Sigma}} \widetilde{c}_{x^k\,j}\cdot e_{x^k\,t+l-j}, k=\overline{1,N},$$

где $g_i\cdot p_t^i$ – члены полинома Колмогорова–Габора:

$$\sum_{i=1}^{r} g_i\cdot p_i^t=FOS\left(\sum_{i=1}^{M}\widetilde{g}_i\cdot\widetilde{x}_t^i+\sum_{i=1}^{M}\sum_{j=1}^{M}\widetilde{g}_{ij}\cdot\widetilde{x}_t^i\cdot\widetilde{x}_t^j+\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\widetilde{g}_{ijk}\cdot\widetilde{x}_t^i\cdot\widetilde{x}_t^j\cdot\widetilde{x}_t^k+\ldots\right)$$

, $g_n=\widetilde{g}_{ij\ldots k}, n=\overline{1,r}$ или

$\sum_{i=1}^{r} g_i\cdot p_i^t=FOS\left(\sum_{i=1}^{M_2}\widetilde{g}_i\cdot\varphi_i\left(\widetilde{\widetilde{x}}_t\right)\right)$, где

$\varphi_i\left(\widetilde{\widetilde{x}}_t\right)=\frac{1}{(2\pi)^{-\frac{M}{2}}\cdot\left|\Sigma_i\right|^{-\frac{1}{2}}}\cdot e^{-\frac{1}{2}\left(\widetilde{\widetilde{x}}_t-\overline{c}_i\right)\Sigma_i^{-1}\cdot\left(\widetilde{\widetilde{x}}_t-\overline{c}_i\right)^T}$, $\overline{c}_i$ – вектор математических ожиданий временных рядов, представляющих главные компоненты, Σ_i – ковариационные матрицы, $i=\overline{1,M}$; или

$\sum_{i=1}^{r} g_i\cdot p_i^t=GMDH\left(\widetilde{x}_t^1,\widetilde{x}_t^2,\ldots,\widetilde{x}_t^M\right)$ – структурная идентификация методом МГУА, $p_l^t=\widetilde{x}_t^i\cdot\widetilde{x}_t^j\cdot\ldots\cdot\widetilde{x}_t^k$, составленного из главных компонент обозначенных $\widetilde{x}_t^i, i=\overline{1,M}$, а также их степеней и сочетаний, отобранных при помощи FOS–алгоритма, определяющего нелинейную часть предложенной модели; FOS – функция структурного упрощения модели, записанной в её аргументе, с учетом характера поведения временных рядов, при помощи алгоритма быстрого ортогонального поиска; $h_i^y=z_i^y-\widetilde{w}_i^{N+1}-\sum_{i=1}^{r} g_i\cdot p_t^i$.

Процесс нахождения таких комбинированных моделей (2) совместного использования модели АРПСС и метода “Гусеница”-SSA может быть продолжительным в связи с ресурсоемкостью метода “Гусеница”-SSA. Поэтому метод “Гусеница”-SSA анализа предложено использовать лишь для предварительной структурной идентификации и грубой параметрической идентификации интегрирующего полинома $w(L)$ (отсюда и название модель авторегрессии – спектрально проинтегрированного скользящего среднего) от оператора задержки L, который также может быть интерпретирован как оператор перевода в пространство состояний, модели

$f(L)\cdot w(L)\cdot y_t=\frac{c(L)}{d(L)}\cdot e_t$ – модель АРСПСС, (3)

а рекуррентный метод SSA-прогнозирования для грубой структурной и параметрической идентификации полинома $f(L)$, структура и коэффициенты которого сначала равны таковыми рекуррентной модели прогнозирования метода “Гусеница”-SSA и наличие которого отличает более общую полиномиальную модель от модели Бокса-Дженкинса, а также совместно с $w(L)$ определяющих долгосрочную память модели, описывая более широкий класс процессов долгосрочной памяти, чем при фрактальном интегрировании в модели ARFIMA, которая в свою очередь была придумана для преодоления недостатка ARIMA моделей при моделировании и прогнозировании процессов в длинной памятью – потери (искажения) долгосрочной информации при взятии приращений. Полиномы $d(L)$ и $c(L)$, в свою очередь, определяют краткосрочные зависимости процесса.

При анализе и прогнозировании финансово-экономических временных рядов, зависящих от нескольких других необходима сбалансированность динамических свойств переменных, стоящих в левой и правой частях уравнения модели. В этом случае идеи метода “Гусеница”-SSA выступают для предварительной обобщённой коинтеграции временных рядов и модель разбивается следующим образом:

$$\hat{\tilde{w}}_t^y = \frac{b^y_{n_{b^y}}(L)}{a^y_{n_{a^y}}(L)} \cdot z_t^y + \sum_{i=1}^{N} \frac{b^{w^y\,i}_{n^{w^y}_{b^i}}(L)}{a^{w^y\,i}_{n^{w^y}_{a^i}}(L)} \cdot z^{x^i}_{t-m_i} + \frac{c^{w^y\,\Pi}_{n^{w^y\,\Sigma}_{c}}(L)}{d^{w^y\,\nabla}_{n^{w^y\,\nabla}_{d}}(L)} \cdot e_t^{w^y};$$

$$f^y(L) \cdot \hat{\tilde{w}}_t^y = \sum_{i=1}^{N} \frac{b^i_{n_{b^i}}(L)}{a^i_{n_{a^i}}(L)} \cdot z^{x^j}_{t-m_i} + \frac{c^{\Pi}_{n^{\Sigma}_{c}}(L)}{d^{\nabla}_{n^{\nabla}_{d}}(L)} \cdot e_t;$$

$$f^{x_i}(L) \cdot \omega(L) \cdot z_t^{x^j} = \frac{c^{x^j\,\Pi}_{n_{x^j}{}^{\Sigma}_{c}}(L)}{d^{x^j\,\nabla}_{n_{x^j}{}^{\nabla}_{d}}(L)} \cdot e_t^{x^j},\ j = \overline{1, N},$$

(4)

$\hat{\tilde{w}}_i^y$ – аппроксимация временным рядом $\hat{y}_t$ и экзогенными временными рядами x^j_t, $j = \overline{1, N}$ временного ряда $\tilde{w}_t^y$ при помощи модели сезонной АРПССЭ или АРПЭ, изначально полученного методом “Гусеница”-SSA, а впоследствии подстраиваемого оптимизационным методом при конкурентном обучении модели; $\omega(L)$ – интегрирующий полином, переводящий временной ряд x_t^j во временной ряд $\hat{\tilde{w}}_t^{x^j}$ – аппроксимация временного ряда $\tilde{w}_{x^k}{}^{N+1}_{t}$ моделью АРПСС; начальные грубые значения коэффициентов полиномов $f^y(L)$, $f^{x_i}(L)$ и их количество можно брать равными коэффициентам f_j^y и $f_j^{x_i}$, $j = \overline{1, N}$ моделей рекуррентного SSA-прогнозирования $\hat{y}_t^{SSA}(i) = \sum_{j=1}^{L^y-1} f_j^y \cdot \tilde{w}^{y\,N+1}_{t+i-j}$ и $\hat{x}^{k\,SSA}_{t}(i) = \sum_{j=1}^{L^{x^k}-1} f_j^{x^k} \cdot \tilde{w}^{x^k\,N+1}_{t+i-j}$ соответственно; L^y и L^{x_i} – соответствующие длины окон; а затем итерационно подстраивать вместе с остальными коэффициентами модели (4) при помощи метода Левенберга-Марквардта. Модель (3) существенно выигрывает по времени обучения комбинированную модель совместного использования сезонной модели АРССЭ и метода “Гусеница”-SSA (2), но несколько уступает ей по статистическим свойствам и с учётом способа её построения названа сезонной моделью авторегрессии – спектрально проинтегрированного скользящего среднего с экзогенными переменными (АРСПССЭ, ARSIMAX) и может быть записана следующим образом:

$$\hat{z}_t^y(l) = \sum_{j=1}^{L^y + n_{b^y} + \sum_{i=1}^{N} n^{w^y}_{a^i} + n^{w^y\nabla}_{d} + \sum_{i=1}^{N} n_{a^i} + n_d^\nabla} \tilde{\tilde{a}}_j \cdot z^y_{t+l-j} +$$

$$\boxed{+\sum_{i=1}^{N}\left(\tilde{\tilde{b}}_0^i \cdot z^{x^i}_{t+l-m_i} - \sum_{j=1}^{n^{\tilde{\tilde{b}}}} \tilde{\tilde{b}}_j^i \cdot z^{x^i}_{t+l-m_i-j}\right) - \sum_{j=1}^{n_c^\Sigma + n_{a^y} + \sum_{i=1}^{N} n^{w^y}_{a^i} + n^{w^y\nabla}_{d} + \sum_{i=1}^{N} n_{a^i} + n_d^\nabla} \tilde{\tilde{c}}_j \cdot e_{t+l-j} -}$$

$$- \sum_{j=1}^{L^y + n^{w^y\Sigma}_{c} + n_{a^y} + \sum_{i=1}^{N} n^{w^y}_{a^i} + + \sum_{i=1}^{N} n_{a^i} + n_d^\nabla} \tilde{\tilde{d}}_j \cdot e^{w^y}{}_{t+l-j};$$

$$\hat{z}_t^{x^k}(l) = \sum_{j=1}^{L^{x^k} + n^{\nabla}_{x^k d}} \tilde{\tilde{a}}_j^{x^k} \cdot z^{x^j}_{t+l-j} + \sum_{j=1}^{n^{x^k\Sigma}_{c}} \tilde{\tilde{c}}_j^{x^k} \cdot e^{x^k}_{t+l-j},$$

$k = \overline{1, N}$, где $\tilde{\tilde{a}}_j$, $j = \overline{1, L^y + n_{b^y} + \sum_{i=1}^{N} n^{w^y}_{a^i} + n^{w^y\nabla}_{d} + \sum_{i=1}^{N} n_{a^i} + n_d^\nabla}$ – коэффициенты полинома

$$\tilde{\tilde{a}}(L) = f^y(L) \cdot b^y_{n_{b^y}}(L) \cdot \prod_{i=1}^{N} a^{w^y i}_{n^{w^y}_{a^i}}(L) \cdot d^\nabla_{n_d^\nabla}(L)$$

; $\tilde{\tilde{b}}_j^i$, $i = \overline{1, N}$, $j = 1, 2, \ldots, n^{\tilde{\tilde{b}}}$; $n^{\tilde{\tilde{b}}} = = L^y +$

$$+\max\left(n^{w^y}_{b^i} + \sum_{j=1, j\neq i}^{N} n^{w^y}_{a^j} + n_{a^y} + n^{w^y\nabla}_{d} + \sum_{j=1, j\neq i}^{N} n_{a^j} + n_d^\nabla, j = \overline{1, N}, n_{b^i} + n_{a^y} + \sum_{j=1, j\neq i}^{N} n^{w^y}_{a^j} + n^{w^y\nabla}_{d} + \sum_{j=1, j\neq i}^{N} n_{a^j} + n_d^\nabla\right)$$

– коэффициенты полинома

$$\tilde{\tilde{b}}^i(L) = f^y(L) \cdot \sum_{i=1}^{N}\left(b^{w^y i}_{n^{w^y}_{b^i}}(L) \cdot \prod_{j=1, j\neq i}^{N} a^{w^y j}_{n^{w^y}_{a^j}}(L) \cdot a^y_{n_{a^y}}(L) \cdot d^{w^y\nabla}_{n^{w^y\nabla}_{d}}(L) \cdot \prod_{j=1, j\neq i}^{N} a^j_{n_{a^j}}(L) \cdot d^\nabla_{n_d^\nabla}(L)\right) -$$

$$-\sum_{i=1}^{N}\left(b^i_{n_{b^i}}(L) \cdot a^y_{n_{a^y}}(L) \cdot \prod_{j=1, j\neq i}^{N} a^{w^y j}_{n^{w^y}_{a^j}}(L) \cdot d^{w^y\nabla}_{n^{w^y\nabla}_{d}}(L) \cdot \prod_{j=1, j\neq i}^{N} a^j_{n_{a^j}}(L) \cdot \cdot d^{w^y\nabla}_{n^{w^y\nabla}_{d}}(L)\right)$$

$\tilde{\tilde{c}}_j$, $j = \overline{1, n_c^\Sigma + n_{a^y} + \sum_{i=1}^{N} n^{w^y}_{a^i} + n^{w^y\nabla}_{d} + \sum_{i=1}^{N} n_{a^i} + n_d^\nabla}$ – коэффициенты полинома

$$\tilde{\tilde{c}}(L) = c^\Pi_{n_c^\Sigma}(L) \cdot a^y_{n_{a^y}}(L) \cdot \prod_{i=1}^{N} a^{w^y i}_{n^{w^y}_{a^i}}(L) \cdot d^{w^y\nabla}_{n^{w^y\nabla}_{d}}(L) \cdot \prod_{i=1}^{N} a^i_{n_{a^i}}(L) \cdot d^\nabla_{n_d^\nabla}(L)$$

, $i = \overline{1, N}$; $\tilde{\tilde{d}}_j$, $j = \overline{1, L^y + n^{w^y\Sigma}_{c} + n_{a^y} + \sum_{i=1}^{N} n^{w^y}_{a^i} + + \sum_{i=1}^{N} n_{a^i} + n_d^\nabla}$ – коэффициенты полинома

$$\tilde{\tilde{d}}(L) = f^y(L) \cdot c^{w^y\Pi}_{n^{w^y\Sigma}_{c}}(L) \cdot a^y_{n_{a^y}}(L) \cdot \prod_{i=1}^{N} a^{w^y i}_{n^{w^y}_{a^i}}(L) \cdot \prod_{i=1}^{N} a^i_{n_{a^i}}(L) \cdot d^\nabla_{n_d^\nabla}(L)$$

, $i = \overline{1, N}$; $\tilde{\tilde{a}}_j^{x^i}$, $j = \overline{1, L^{x^k} + L^\omega + n^{x^k\nabla}_{d}}$ – коэффициенты полинома $\tilde{\tilde{a}}^{x^i}(L) = f^{x^i}(L) \cdot \omega(L) \cdot d^{x^i\nabla}_{n^{x^i\nabla}_{d}}(L)$; $\tilde{\tilde{c}}^{x^i}{}_j$, $j = \overline{1, n^{x^k\Sigma}_{c}}$ – коэффициенты полинома $\tilde{\tilde{c}}^{x^i}(L) = c^{x^i\Pi}_{n^{x^i\Sigma}_{c}}(L)$.

Для учета гетероскедастичности процесса (изменения дисперсии процесса во времени) применяется обобщенная модель с авторегрессионной условной гетероскедастичностью $\mathrm{GARCH}(m,r)$, имеющая вид [8]:

$$\sigma_t^2 = w + \theta(L)\varepsilon_t^2 + \varphi(L)\sigma_t^2,$$

где σ_t^2 – временной ряд изменения дисперсии процесса y_t, $\theta(L) = \theta_1 L + \theta_2 L^2 + \ldots + \theta_p L^m$, $\varphi(L) = \varphi_1 L + \varphi_2 L^2 + \ldots + \varphi_r L^r$, ε_t^2 – остаточные члены модели. Модель $\mathrm{GARCH}(m,r)$ может быть записана через модель АРСС(s,m) следующим образом [9]:

$$\varepsilon_t^2 = \frac{w + (1-\varphi(L))}{(1-\theta(L)-\varphi(L))} v_t,$$

где $s = \max(r,m)$, $v_t = \varepsilon_t^2 - \sigma_t^2$.

Фрактально проинтегрированный процесс GARCH может быть записан следующим образом:

$$(1-L)^{2-H}\varepsilon_t^2 = \frac{w + (1-\varphi(L))}{(1-\theta(L)-\varphi(L))\cdot(1-L)^{-1}} v_t,$$

где H - показатель Херста.

Предлагаемая спектрально проинтегрированная обобщенная модель с авторегрессионной условной гетероскедастичностью имеет вид:

$$f^{\varepsilon^2}(L)\cdot\omega^{\varepsilon^2}(L)\cdot\varepsilon_t^2 = \frac{w + (1-\varphi(L))}{(1-\theta(L)-\varphi(L))}\cdot v_t, \quad (5)$$

где $\omega^{\varepsilon^2}(L)$ – интегрирующий полином от оператора задержки, с помощью которого аппроксимируется временной ряд дисперсии шумов ε_t^2 в преобразованный сглаженный методом «Гусеница»-SSA временной ряд $w_t^{\varepsilon_t^2}$, а предварительная структурная идентификация и грубая параметрическая идентификация полинома от оператора задержки $f^{\varepsilon^2}(L)$, с помощью которого аппроксимируется сам ряд ε_t^2 производится при определении коэффициентов рекуррентной прогнозирующей формулы метода «Гусеница»-SSA; w – среднее значение или уровень временного ряда ε_t^2.

Совместно модели (4) и (5) образуют модель авторегрессии – спектрально проинтегрированного скользящего среднего со спектрально проинтегрированной обобщённой авторегрессионной условной гетероскедастичностью (АРСПСС – СПОАРУГ, на англ. яз. ARSIMA – SIGARCH).

Для построения предложенных моделей используются стандартные алгоритмы методов “Гусеница”-SSA и Бокса-Дженкинса.

Результаты исследований. Тестирование предложенной модели (3) проводилось на следующих данных, представленных на рисунке 1.

По автокорреляционной функции можно судить о наличии долгосрочной памяти

процесса.На рисунке 3 изображён график исследуемого процесса с трендовой

компонентой, построенной методом “Гусеница”-SSA

В результате совместного применения метода “Гусеница”-SSA и метода Бокса-Дженкинса была построена модель процесса:

$$\left(1 - 0.1558L - 0.1509L^2 - 0.1374L^3 - 0.1188L^4 - 0.0982L^5 - 0.0789L^6 - 0.0627L^7 - \right.$$
$$\left. - 0.0501L^8 - 0.0416L^9 - 0.0367L^{10} - 0.0349L^{11} - 0.0348L^{12} - \ldots\right)\cdot\nabla y_t =$$
$$= \frac{(-0.3341 + 0.023L^{10})}{(1 - 1.8831 + 0.9839)\nabla}\cdot\frac{(-0.39 + 0.8047L + 0.5951L^2 - 0.1958L^3)}{(1 + 0.6288L - 0.1067L^2 + 0.0649L^3 + 0.2264L^4 + 0.0987L^5 + 0.1262L^6)\nabla} e_t$$

и получены с её помощью одношаговые прогнозы с 95%-и доверительными интервалами:

Использование предлагаемой модели и метода её построения позволило снизить среднеквадратическую ошибку прогнозирования с 0.0361 до 0.0134 по сравнению с применением классической модели АРПСС.

Таким образом, для получения адекватных моделей сложных финансово-экономических процессов, высококачественных прогнозов необходимо комбинировать модели с разными структурами, включая нелинейные модели, которые являются взаимодополняющими при их конкурентном обучении.

Предлагаемый метод “Гусеница”-SSA – АРПСС – СПОАРУГ является модификацией метода “Гусеница”-SSA с автоматическим отделением долгосрочной памяти от краткосрочной и периодических составляющих и может быть интерпретирована как развитие моделей в пространстве состояний, а предложенная модель АРСПСС – СПОАРУГ, как и модель ARFIMA – FIGARCH, является очередной модификацией модели ARIMA – GARCH, а метод построения предлагаемой модели является развитием метода Бокса-Дженкинса, но для моделирования более широкого класса процессов.

Предложенная модель АРСПСС – СПОАРУГ и методика её построения является некоторым промежуточным подходом между классическими регрессионными и современными нейросетевыми и более формализованная по выбору структуры, являясь при этом оптимальной по детализации с учетом существующих на сегодняшний день математических и машинных как достоинств и достижений так и недостатков и ограничений.

Подводя итоги описанным выше преимуществам предлагаемой методики, еще раз следует отметить, что основная идея состоит в эффекте синергии, который возникает в результате комбинированного применения двух методов: метода “Гусеница”-SSA и метода Бокса-Дженкинса.

Основным преимуществом предлагаемой методики построения модели адекватной исследуемому процессу является ее строгая формализация и, следовательно, возможность полной автоматизации всех этапов построения и использования модели.

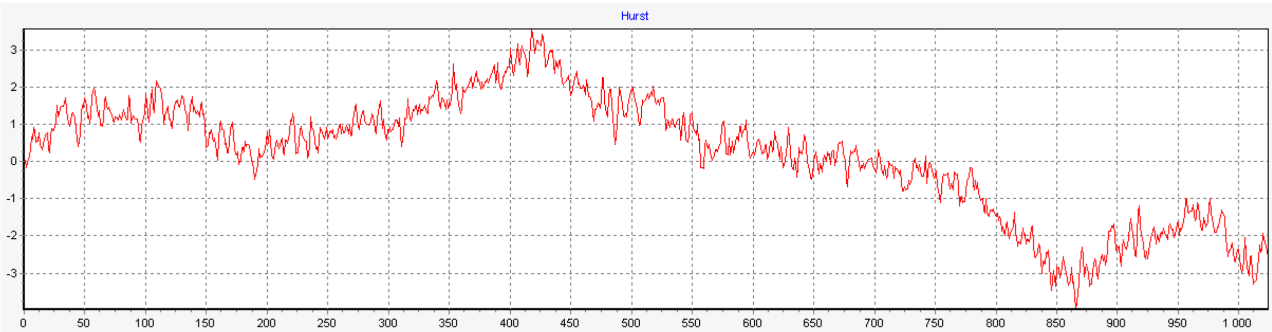

Рис.1. График исследуемого процесса с краткосрочной и долгосрочной памятью

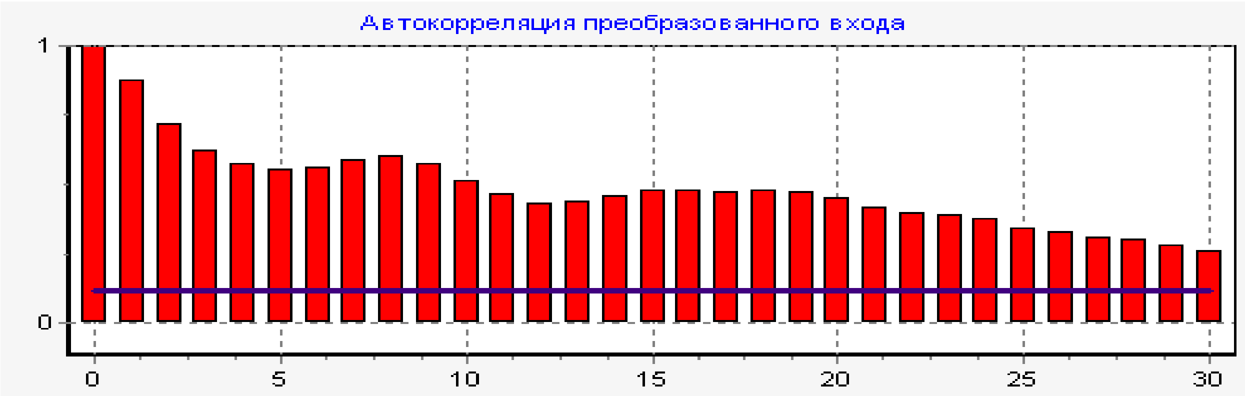

Рис.2. Автокорреляционная функция исследуемого процесса

Рис.3. График исследуемого процесса с трендом

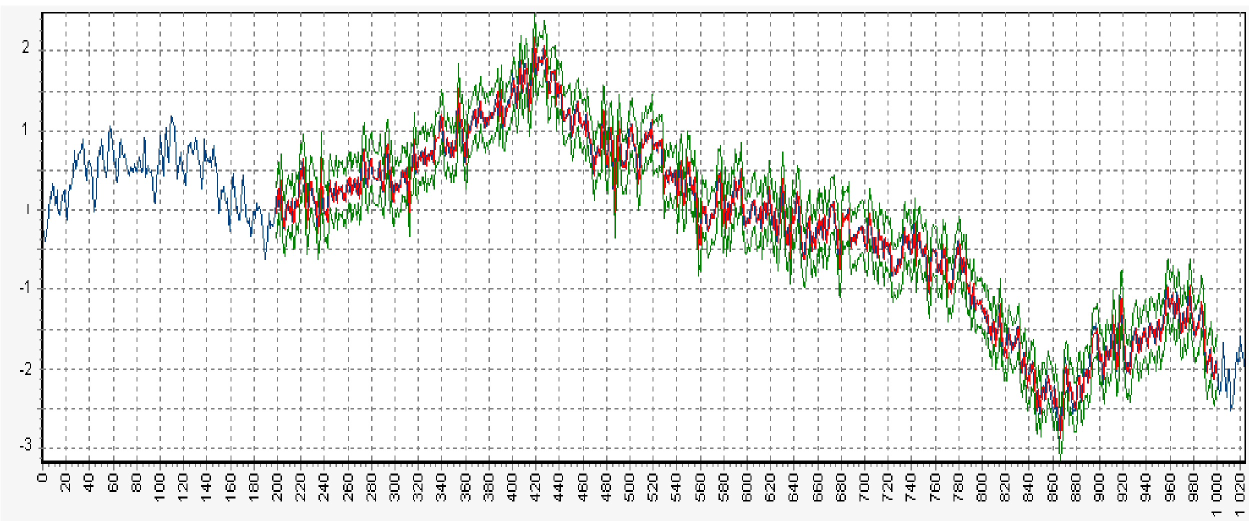

Рис.4. График исследуемого процесса с одношаговыми прогнозами и 95%-и доверительными интервалами

Список литературы

1. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Пер. с англ. – М. : Мир, 1974. Вып. II. – 197 с.
2. Granger C.W.J., Joyeux R. An Introduction to Long-Memory Time Series Models and Fractional Differencing // Journal of Time Series Analysis. 1980. N 1(1). P. 15-29.
3. Седов А.В. Моделирование объектов с дискретно-распределёнными параметрами: декомпозиионный подход / А.В. Седов; Южный научный центр РАН. – М. : Наука, 2010. – 438 с.
4. Щелкалин В.Н., Тевяшев А.Д. «Автоматизированная система анализа и оперативного прогнозирования процессов потребления целевых продуктов в жилищно-коммунальном хозяйстве». Международный конкурс инновационных проектов “Харьковские инициативы”, 2010.
5. Щелкалин В.Н., Тевяшев А.Д. Модель авторегрессии – спектрально проинтегрированного скользящего среднего со спектрально проинтегрированной обобщенной авторегрессионной условной гетероскедастичностью для моделирования, фильтрации, прогнозирования и управления процессами в современных системах автоматизации. // Труды Международной научно-практической конференции «Передовые информационные технологии, средства и системы автоматизации и их внедрение на российских предприятиях» AITA-2011. Москва, 4 – 8 апреля 2011 г. М.: Институт проблем управления им. В.А. Трапезникова РАН, 2011. с. 996 – 1022.
6. Евдокимов А. Г., Тевяшев А. Д. Оперативное управление потокораспределением в инженерных сетях. – Х. : Вища школа, 1980. – 144 с.
7. Голяндина Н. Э. Метод «Гусеница»-SSA: прогноз временных рядов: Учеб. пособие. – СПб., 2004. – 52 с.
8. Перцовский О.Е. Моделирование валютных рынков на основе процессов с длинной памятью: Препринт WP2/2004/03 – М.: ГУ ВШЭ, 2003. – 52 с.
9. Bollerslev T. Generalized autoregressive conditional heteroscedasticity // Journal of econometrics. – 1986/ – V. 31. – PP. 307-327.

Аннотация

Виталий Щелкалин

МЕТОД “ГУСЕНИЦА”-SSA – АРПСС – СПОАРУГ И МОДЕЛЬ АРСПСС – СПОАРУГ ДЛЯ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ФИНАНСОВО-ЭКОНОМИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ

В работе производится дальнейшее развитие метода Бокса-Дженкинса построения моделей и усовершенствование моделей авторегрессии – проинтегрированного скользящего среднего (АРПСС), разработанных и успешно используемых около сорока лет в различных отраслях науки и техники и остающихся по сей день одними из наиболее эффективных моделей для моделирования и прогнозирования финансово-экономических временных рядов, превосходя своих конкурентов по целому ряду критериев, таких как: экономность по количеству параметров, трудоемкость алгоритмов построения моделей и ресурсоемкость их реализации, формализация и автоматизируемость их построения. Предложена очередная модификация моделей ARIMA и GARCH.

Ключевые слова: *моделирование, фильтрация, прогнозирование, управление, модель АРПСС, метод «Гусеница»-SSA – АРПСС – СПОАРУГ, модель АРСПСС, модель АРСПСС – СПОАРУГ, гетероскедастичность, метод Левенберга-Марквардта.*

Summary

Vitaliy Shchelkalin

"CATERPILLAR"-SSA – ARIMA – SIGARCH METHOD AND ARSIMA – SIGARCH MODEL FOR ANALYSIS AND FORECASTING OF FINANCIAL AND ECONOMIC TIME SERIES

In this paper the further development of the Box-Jenkins method for modeling and improvement of autoregressive - integrated moving average (ARIMA) models developed and successfully used about forty years in various branches of science and technology and remaining in present time as one of the most efficient models for modeling, forecasting of financial and economic time series, exceeding their own rivals on whole row of criterions, such as: economy on the number of parameters, the complexity of algorithms for modeling and resource intensity of their implementation, on formalization and automation of their construction is produced. Proposed another modification of the models ARIMA and GARCH.

Key words: *modeling, filtering, forecasting, management, ARIMA model, "Caterpillar»-SSA – ARIMA – SIGARCH method, ARSIMA model, ARSIMA – SIGARCH model, heteroskedasticity, Levenberg-Marquardt method.*