МЕТОДЫ АНАЛИЗА ДАННЫХ ПРИ ПОМОЩИ СОСТОЯТЕЛЬНЫХ ОЦЕНОК СПЕКТРАЛЬНЫХ ПЛОТНОСТЕЙ

А. И. Сурмач, Н. В. Семенчук

Гродненский государственный университет имени Янки Купалы
Гродно, Беларусь
E-mail: surmach_ai@mail.ru, senata155@gmail.com

В статье предлагается алгоритм построения оценок спектральных плотностей стационарных случайных процессов с различными спектральными окнами и окнами просмотра данных. Систематизация реальных показателей, приведение ряда к стационарному виду. Применение полученных результатов к реальным данным.

Ключевые слова: стационарные случайные процессы, спектральная плотность, состоятельные оценки, окно просмотра данных, расширенная периодограмма.

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Одним из важных и часто применяемых на практике является класс стационарных случайных процессов.

Стационарный случайный процесс – важный специальный класс случайных процессов, часто встречающийся в приложениях теории вероятностей к различным разделам естествознания и техники. Случайный процесс $X(t)$ называется стационарным, если все его вероятностные характеристики не меняются с течением времени t.

Спектральной плотностью случайного процесса $X(t),\ t \in Z,$ называется функция

$$f(\lambda) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} R(\tau)e^{-i\lambda\tau},$$

$\lambda \in \Pi = [-\pi; \pi],$ при условии, что $\sum\limits_{\tau=-\infty}^{\infty}\left|R(\tau)\right| < \infty.$

Функция $f(\lambda),\ \lambda \in \Pi,$ является периодической с периодом 2π.

Случайный процесс $X(t),\ t \in Z,$ называется стационарным в широком смысле, если $MX^2(t) < \infty$ и

$$m(t) = m = const,\ t \in Z,$$

$$R(t_1, t_2) = R(t_1 - t_2),\ t_1, t_2 \in Z.$$

Рассмотрим последовательность:

$$X(t) = \sum_{j=1}^{p}\beta_j X(t-j) + \varepsilon(t). \tag{1}$$

Последовательность (1) называется последовательностью авторегрессии порядка p. Обозначается AR(p). При этом предполагается, что $\varepsilon(t)$ есть последовательность некоррелированных и одинаково распределенных случайных величин [1].

Большое практическое значение имеют последовательности авторегрессии первого (p = 1) и второго (p = 2) порядков:

Спектральная плотность последовательности (1) имеет вид:

$$f(\lambda)=\frac{\sigma_{\varepsilon}^{2}}{\left|1-\sum_{j=1}^{p}\beta_{j}e^{ij\lambda}j\right|^{2}2\pi},\quad 0\le\lambda\le\pi. \tag{2}$$

Последовательности авторегрессии представляют собой очень большой интерес в различных приложениях. Например, авторегрессия первого порядка использовалась при изучении колебаний уровня Каспийского моря.

При анализе наблюдений по графику оценки спектральной плотности часто бывает удобно определить, можно ли наблюдаемую последовательность рассматривать как последовательность авторегрессии. Это легко можно сделать, имея набор графиков различных процессов авторегрессии.

Рассмотрим последовательность:

$$X(t)=\sum_{k=0}^{q}\alpha_{k}\varepsilon(t-k),\quad \alpha_{0}=1. \tag{3}$$

Последовательность (3) называется последовательностью скользящего среднего порядка q. Обозначается MA(q). При этом предполагается, что $\varepsilon(t)$ есть последовательность некоррелированных и одинаково распределенных случайных величин. Чаще применяются в различных прикладных работах случаи q = 1 и q = 2.

Спектральная плотность последовательности (3) равна [2].

$$f(\lambda)=\frac{\sigma_{\varepsilon}^{2}}{2\pi}\left|1+\alpha_{1}e^{i\lambda}+\ldots+\alpha_{q}e^{i\lambda q}\right|^{2},\quad 0\le\lambda\le\pi. \tag{4}$$

ОЦЕНКИ СПЕКТРАЛЬНЫХ ПЛОТНОСТЕЙ

Пусть $X(0)$, $X(1)$, …, $X(T-1)$ – T последовательных, через равные промежутки времени, наблюдений за стационарным случайным процессом $X(t)$, $t\in Z$ с m = 0 и неизвестной спектральной плотностью $f(\lambda),\ \lambda\in\Pi$.

В качестве первой оценки спектральной плотности $f(\lambda),\ \lambda\in\Pi,$ рассмотрим периодограмму:

$$I_{T}(\lambda)=\frac{1}{2\pi T}\left|\sum_{t=0}^{T-1}X(t)e^{-i\lambda t}\right|^{2}. \tag{5}$$

В качестве классической оценки спектральной плотности $f(\lambda),\ \lambda\in\Pi,$ рассмотрим расширенную периодограмму:

$$I_{T}^{(h)}(\lambda)=\frac{1}{2\pi H_{2}^{(T)}(0)}d_{T}(\lambda)d_{T}(-\lambda), \tag{6}$$

где

$$d_{T}(\lambda)=\sum_{t=0}^{T-1}h_{T}(t)X(t)e^{-i\lambda t},$$

$$H_{2}^{(T)}(\lambda)=\sum_{t=0}^{T-1}(h_{T}(t))^{k}e^{-i\lambda t},$$

функция $h_T(t)=h\left(\frac{t}{T}\right)$, $h:[0,1]\to\mathbf{R}$, – функция окна просмотра данных, $k\in\mathbf{N}$, $T\in\mathbf{N}$.

В качестве состоятельной оценки спектральной плотности $f(\lambda),\ \lambda\in\Pi$, в точке $\lambda=\lambda_j=\frac{2\pi j}{T}$ рассмотрим оценку, построенную путем осреднения расширенной периодограммы $I_T^{(h)}(\lambda_{j+k})$, т.е. статистику вида [3]:

$$\hat{f}_T(\lambda_j)=\sum_{k=-[\frac{T}{2}]+1}^{[\frac{T}{2}]}\varphi_T(k)I_T^{(h)}(\lambda_{j+k}). \tag{7}$$

АНАЛИЗ РЕАЛЬНЫХ ДАННЫХ

Для анализа при помощи состоятельных классических оценок спектральных плотностей были выбраны данные ОАО «Молочный мир» по количеству отгруженного молока в упаковке с 01.01.2014 по 24.12.2014 года за каждый рабочий день.

Получили временной ряд из 256 значений. Графически ряд представляется на рис. 1.

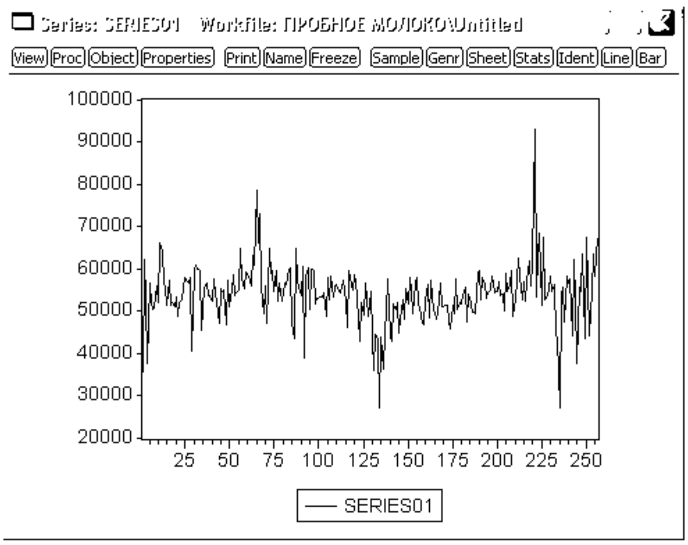

Рис. 1. Временной ряд данных по отгрузке молока

Рассмотрим полученный ряд. По данному графику временного ряда можно предположить, что он стационарный. Для проверки ряда на стационарность будем пользоваться программой Eviews 5.5.

Для проверки стационарности ряда проведем тест Дики – Фуллера (рис. 2).

Т.к. p-значение $p = 0{,}0001 < 0{,}05$, то гипотеза о том, что ряд нестационарный, отвергается.

Null Hypothesis: SERIES01 has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic based on SIC, MAXLAG=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.503995	0.0000
Test critical values:	1% level	-3.456093	
	5% level	-2.872765	
	10% level	-2.572826	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SERIES01)
Method: Least Squares
Date: 01/14/15 Time: 10:55
Sample (adjusted): 4 256
Included observations: 253 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SERIES01(-1)	-0.458178	0.083245	-5.503995	0.0000
D(SERIES01(-1))	-0.366228	0.078352	-4.674150	0.0000
D(SERIES01(-2))	-0.216749	0.061108	-3.546996	0.0005
C	24770.43	4502.383	5.501626	0.0000

R-squared	0.415735	Mean dependent var	57.13834
Adjusted R-squared	0.408695	S.D. dependent var	8305.491
S.E. of regression	6386.619	Akaike info criterion	20.37748
Sum squared resid	1.02E+10	Schwarz criterion	20.43335
Log likelihood	-2573.751	F-statistic	59.05873
Durbin-Watson stat	1.969446	Prob(F-statistic)	0.000000

Рис. 2. Тест Дики – Фуллера

Сделали предположение, что это модель авторегрессии – скользящего среднего первого порядка (рис. 3).

Dependent Variable: SERIES01
Method: Least Squares
Date: 01/14/15 Time: 10:59
Sample (adjusted): 2 256
Included observations: 255 after adjustments
Convergence achieved after 11 iterations
Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	54239.43	1169.397	46.38240	0.0000
AR(1)	0.911451	0.049376	18.45929	0.0000
MA(1)	-0.748966	0.077951	-9.608165	0.0000

R-squared	0.126407	Mean dependent var	53981.96
Adjusted R-squared	0.119474	S.D. dependent var	6865.182
S.E. of regression	6442.036	Akaike info criterion	20.39077
Sum squared resid	1.05E+10	Schwarz criterion	20.43243
Log likelihood	-2596.823	F-statistic	18.23195
Durbin-Watson stat	1.966250	Prob(F-statistic)	0.000000

Inverted AR Roots	.91
Inverted MA Roots	.75

Рис. 3. Авторегрессия – скользящее среднее первого порядка

Все коэффициенты значимы. Коэффициент детерминации неплохой, что говорит о хорошем качестве модели. Статистика Дарбина – Уотсона почти равна 2, что говорит о том, что остатки гомоскедастичны и модель не является ложной. В целом уравнение тоже значимо. Далее проверим остатки. График остатков на рис. 4.

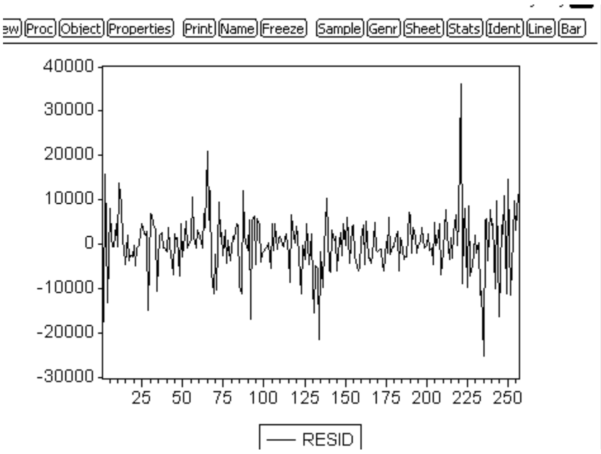

Рис. 4. График остатков

ЧАКФ, АКФ и Q-статистика подтверждают, что остатки являются «белым шумом» (рис. 5).

Date: 01/14/15 Time: 11:08
Sample: 1 256
Included observations: 256

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.030	-0.030	0.2300	0.632
		2	-0.010	-0.011	0.2585	0.879
		3	0.106	0.105	3.1813	0.365
		4	-0.074	-0.069	4.6319	0.327
		5	-0.043	-0.045	5.1159	0.402
		6	0.007	-0.008	5.1272	0.528
		7	0.069	0.085	6.3914	0.495
		8	-0.032	-0.025	6.6654	0.573
		9	0.080	0.075	8.3802	0.496
		10	0.021	0.006	8.4941	0.581
		11	0.016	0.036	8.5593	0.662
		12	-0.067	-0.083	9.7615	0.637
		13	0.004	0.009	9.7660	0.713
		14	-0.075	-0.083	11.284	0.664
		15	-0.008	0.016	11.303	0.731
		16	0.046	0.020	11.873	0.753
		17	-0.083	-0.068	13.799	0.681
		18	-0.007	-0.034	13.813	0.741
		19	-0.003	-0.006	13.816	0.794
		20	-0.048	-0.039	14.472	0.806
		21	0.026	0.044	14.655	0.840
		22	-0.010	-0.022	14.682	0.876
		23	-0.002	0.018	14.683	0.906
		24	0.039	0.035	15.121	0.917
		25	0.007	0.014	15.136	0.938
		26	0.012	0.011	15.178	0.954
		27	-0.072	-0.073	16.691	0.938
		28	0.042	0.044	17.192	0.945
		29	0.004	0.006	17.195	0.959
		30	0.053	0.074	18.005	0.958
		31	0.061	0.031	19.103	0.953
		32	-0.032	-0.042	19.411	0.961
		33	0.024	0.015	19.578	0.969
		34	-0.002	-0.006	19.579	0.977
		35	0.027	0.036	19.794	0.982
		36	-0.027	-0.016	20.020	0.986

Рис. 5. ЧАКФ и АКФ

Т.о. можно сделать вывод, что данный временной ряд является стационарным и поэтому его можно использовать для реализации алгоритма.

Апробация алгоритма на модельных данных.

Построим реализацию процесса на рис. 6.

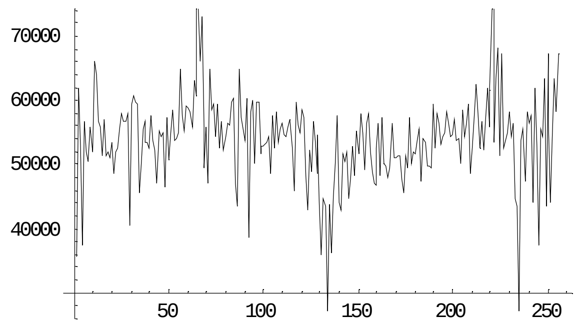

Рис. 6. Реализация процесса

Построили расширенную периодограмму с окном просмотра данных Рисса, Бохнера, Парзена (рис. 7).

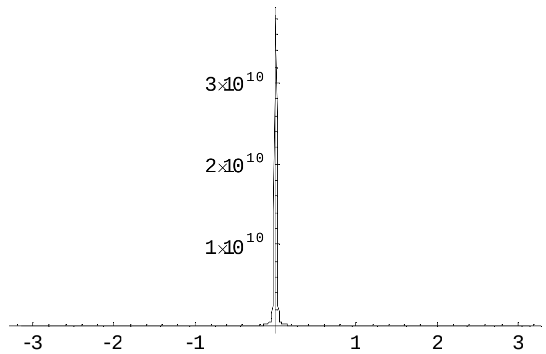

Рис. 7. Расширенная периодограмма

И ее состоятельную классическую оценку спектральной плотности (рис. 8).

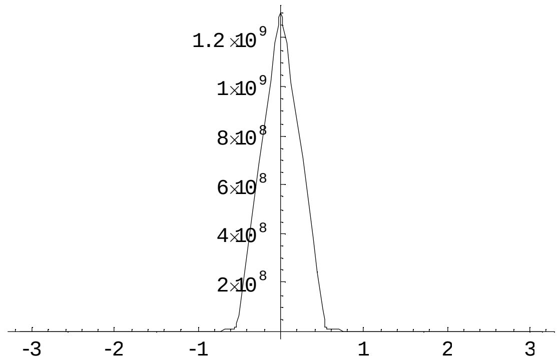

Рис. 8. Состоятельная классическая оценка спектральной плотности
для периодограммы на рис. 7

По состоятельной оценке можно сделать вывод, что отгрузка молока изменяется в соответствии с процессом авторегрессии первого порядка.

ЛИТЕРАТУРА

1. *Труш, Н. Н. Мирская, Е. И.* Случайные процессы. Преобразование Фурье наблюдений / Н. Н. Труш, Е. И. Мирская. Мн. : БГУ, 2000. С. 18–24.

2. *Журбенко, И. Г.* Спектральный анализ стационарных случайных процессов / И. Г. Журбенко, Н. Н. Труш // Вестник БГУ. Сер. 1: Физ. Мат. Мех. 1981. № 1. С. 20–26.

3. *Андерсон, Т.* Статистический анализ временных рядов / Т. Андерсон. М. : Мир, 1976. С. 256–263.