

1. Вспомогательные определения

В данном разделе введем некоторые обозначения, которые будем использовать в дальнейшем.

Определение 1.1. Случайный процесс $\{Y_t : t \in \mathbb{Z}\}$ называют стационарным (в широком смысле), если

1. $EY_t \equiv \text{const}$ (среднее постоянно по времени);
2. $\text{cov}(Y_t, Y_{t+h}) = \gamma(h)$ (ковариация зависит только от лага h).

Замечание 1.1. Поскольку $\gamma(0) = \text{cov}(Y_t, Y_t) = DY_t$, то дисперсия также не меняется со временем.

Замечание 1.2. Далее под стационарностью будет подразумеваться именно стационарность в широком смысле.

Определение 1.2. Случайный процесс $\{\varepsilon_t\}$ называют белым шумом $\text{WN}(0, \sigma^2)$, если он стационарный, $E\varepsilon_t = 0$, $\gamma(h) = 0 \ \forall h \neq 0$ и $D\varepsilon_t = \sigma^2$.

Определение 1.3. Моделью $\text{ARMA}(p, q)$, где $p, q \in \mathbb{N} \cup \{0\}$ называют случайный процесс $\{X_t\}$, удовлетворяющий соотношению

$$X_t = \varepsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

где $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$.

Определение 1.4. Спектральной плотностью стационарного процесса называется такая функция $f(\omega)$, что

$$\gamma(h) = 2 \int_0^{1/2} e^{2\pi h \omega i} f(\omega) d\omega.$$

Определение 1.5. Пусть $\{Y_t\}$ — стационарный процесс. Функцию

$$I(\omega) = \frac{1}{n} \left| \sum_{j=1}^n Y_j e^{-2\pi \omega j i} \right|^2$$

называют периодограммой выборки размера n процесса $\{Y_t\}$.

2. Процессы с длинной памятью

Определение 2.1. Говорят, что стационарный процесс $\{Y_t\}$ обладает длинной памятью, если

$$\sum_{h=0}^H |\gamma(h)| \rightarrow \infty,$$

при $H \rightarrow \infty$. Иначе говорят, что $\{Y_t\}$ обладает короткой памятью:

$$\sum_{h=0}^{\infty} |\gamma(h)| < \infty.$$

Существуют и альтернативные определения процессов с длинной памятью, которые можно найти в [1, Section 3.1]. Там же показано, что они согласованы с определением 1.2.1.

Пример 2.1. Процессом с короткой памятью является, например, стационарная модель ARMA(p, q), поскольку $|\gamma(h)| \leq CR^h$, где $C > 0$ и $0 < R < 1$ [2].

Введем понятие дробного интегрирования $(1 - L)^d$, где L — оператор сдвига. Например, для $d = 1$ имеем $(1 - L)Y_t = Y_t - Y_{t-1}$, для $d = 2$ — $(1 - L)^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$, и так далее. Обобщим этот оператор для нецелых d с помощью разложения в ряд Тейлора функции $(1 - x)^d$ в нуле:

$$\begin{aligned} (1 - x)^d &= 1 - dx - \frac{d(1-d)}{2}x^2 - \frac{d(1-d)(2-d)}{3!}x^3 - \dots \\ &= \sum_{j=0}^{\infty} \pi_j(d)x^j = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j x^j, \end{aligned}$$

где $\binom{d}{j}$ — обобщенный биномиальный коэффициент. Коэффициенты $\pi_j(d)$ удовлетворяют соотношению

$$\pi_j(d) = (-1)^j \binom{d}{j} = \frac{j-1-d}{j} \pi_{j-1}(d) = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}, \quad (1)$$

где $\Gamma(x)$ — гамма функция. Заметим, что второе равенство в формуле (1.1) верно для любых d , третье же верно только для $d \notin \mathbb{N} \cup \{0\}$, поскольку гамма функция не определена для неположительных целых чисел.

Определение 2.2. Пусть процесс $\{Y_t\}$ определен соотношением

$$Y_t = (1 - L)^{-d} X_t = \sum_{k=0}^{\infty} \pi_k(-d) X_{t-k}, \quad d < 1/2,$$

где $\pi_k(-d)$ из формулы (1.1), $\{X_t\}$ — стационарная и обратимая модель ARMA(p, d). Процесс $\{Y_t\}$ называют дробно интегрированной моделью ARMA или ARFIMA(p, d, q).

Предложение 2.1. ARFIMA(p, d, q) является стационарным процессом с нулевым средним при $d < 1/2$. Его спектральная плотность определяется выражением

$$\begin{aligned} f(\omega) &= \sigma^2 4^{-d} \sin^{-2d}(\pi\omega) \frac{|\theta(e^{-2\pi\omega i})|^2}{|\phi(e^{-2\pi\omega i})|^2}, \quad \omega > 0 \\ &\sim \sigma^2 \omega^{-2d} \frac{|\theta(1)|^2}{|\phi(1)|^2}, \quad \omega \rightarrow 0, \end{aligned} \quad (2)$$

Доказательство. См. [3, Proposition 6.1]. □

Замечание 2.1. Из формулы (1.2) видно, что спектральная плотность дробно интегрированного процесса монотонно убывает (возрастает) тогда и только тогда, когда монотонно убывает (возрастает) спектральная плотность процесса $\{X_t\}$.

Следствие 2.1. В условиях предложения 1.2.1 при $0 < d < 1/2$

$$\gamma(h) \sim C_{\gamma,d} h^{2d-1}, \quad h \rightarrow \infty,$$

где

$$C_{\gamma,d} = \sigma^2 \frac{|\theta(1)|^2}{|\phi(1)|^2} \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}.$$

Доказательство. См. [3, Corollary 6.1]. □

Замечание 2.2. Из следствия 1.2.1 сразу следует, что ARFIMA(p, d, q) с $d \in (0, 1/2)$ обладает длинной памятью.

2.1. Возникновение процессов с длинной памятью

Нас интересуют процессы с монотонной спектральной плотностью, поскольку они довольно распространены в реальном мире. Такими процессами являются процессы со степенной спектральной плотностью $f(\omega) \sim \omega^{-\alpha}$, имеющие большое применение в различных областях, например, в физике, биологии, астрофизике, геофизике и экономике.

Процессы с длинной памятью, являющиеся частным случаем процессов со степенной спектральной плотностью, довольно распространены. Например, в работе [4] обнаружена длинная память в таких среднегодовых гидрологических временных рядах, как количество осадков, температура и данных о речном стоке. В работе [5] на наличие длинной памяти исследовалась скорость ветра в Ирландии, в работе [6] исследовался эффект длинной памяти у сейсмических данных. Помимо геофизики, длинная память встречается также в финансах [7, 8].

3. Оценка параметров

Пусть $Y_t = (1 - L)^{-d} X_t$, $d < 1/2$. Будем считать, что $\{X_t\}$ представляет собой модель ARMA(p, q) с нормально распределенным белым шумом $\{\varepsilon_t\}$. Тогда его спектральная плотность $f_X(\omega) = f_X(\omega; \boldsymbol{\psi}, \sigma)$, где

$$\boldsymbol{\psi} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^T.$$

Поставим задачу оценить параметры $\boldsymbol{\varphi}^T = (d, \boldsymbol{\psi}^T)$ и σ^2 .

3.1. Maximum likelihood estimation (MLE)

Поскольку $\{\varepsilon_t\}$ — гауссовский белый шум, вектор

$$\mathbf{Y} = (Y_1, \dots, Y_n)^T \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_n),$$

где $\boldsymbol{\Sigma}_n = (\gamma(|i - j|))_{i,j=1}^n$ — ковариационная матрица \mathbf{Y} . Совместная плотность распределения \mathbf{Y} равна

$$(2\pi)^{-n/2} |\boldsymbol{\Sigma}_n|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{Y}^T \boldsymbol{\Sigma}_n^{-1} \mathbf{Y} \right\}.$$

Рассмотрим логарифм функции правдоподобия. Отбрасывая аддитивные константы, получаем

$$\ell(\boldsymbol{\varphi}, \sigma^2) = -\frac{1}{2} \ln |\boldsymbol{\Sigma}_n| - \frac{1}{2} \mathbf{Y}^T \boldsymbol{\Sigma}_n^{-1} \mathbf{Y}.$$

Положим $\boldsymbol{\Gamma}_n = \boldsymbol{\Sigma}_n / \sigma^2$ и, максимизируя ℓ по σ^2 , получаем

$$\ell_c(\boldsymbol{\varphi}) = -\frac{n}{2} \ln (S(\boldsymbol{\varphi})/n) - \frac{1}{2} \ln g_n(\boldsymbol{\varphi}),$$

где $S(\boldsymbol{\varphi}) = \mathbf{Y}^T \boldsymbol{\Gamma}_n \mathbf{Y}$, $g_n(\boldsymbol{\varphi}) = |\boldsymbol{\Gamma}_n|$. Тогда

$$\hat{\boldsymbol{\varphi}}_{\text{ML}} = \underset{\boldsymbol{\varphi}}{\operatorname{argmax}} \ell_c(\boldsymbol{\varphi}), \quad \hat{\sigma}_{\text{ML}}^2 = S(\hat{\boldsymbol{\varphi}}_{\text{ML}}).$$

Замечание 3.1. В случае ненулевого матожидания $\mathbb{E}Y_t = \mu$, для получения оценок максимального правдоподобия нужно вместо \mathbf{Y} рассматривать $\mathbf{Y} - \mu$.

Замечание 3.2. Для вычисления ℓ_c можно использовать алгоритм Левинсона-Дурбина, имеющий временную трудоемкость $O(n^2)$.

3.2. Whittle estimation

Метод максимального правдоподобия применим, когда известно матожидание μ . При неизвестном μ можно использовать его оценку $\bar{\mathbf{Y}}$, однако, помимо этого, существует проблема вычислительной сложности метода при больших n .

Обе эти проблемы можно решить, используя оценку Уиттла (Whittle): вместо логарифма функции правдоподобия рассматривается ее оценка (с точностью до константы) [9]. Пусть $f(\omega, \boldsymbol{\varphi}, \sigma^2)$ — спектральная плотность $\{Y_t\}$, $I(\omega)$ — периодограмма \mathbf{Y} , тогда

$$\ell_W(\boldsymbol{\varphi}, \sigma^2) = -\frac{1}{m} \sum_{j=1}^m \left(\ln f(\omega_j; \boldsymbol{\varphi}, \sigma^2) + \frac{I(\omega_j)}{f(\omega_j; \boldsymbol{\varphi}, \sigma^2)} \right),$$

где $m = \lfloor (n-1)/2 \rfloor$, $\omega_j = j/n$, $j = 1, 2, \dots, m$. Заметим, что $f(\omega; \boldsymbol{\varphi}, \sigma^2) = \sigma^2 g(\omega; \boldsymbol{\varphi})$. Тогда, максимизируя ℓ_W по σ^2 , получаем

$$\hat{\boldsymbol{\varphi}}_W = \underset{\boldsymbol{\varphi}}{\operatorname{argmax}} Q(\boldsymbol{\varphi}), \quad \hat{\sigma}_W^2 = \frac{1}{m} \sum_{j=1}^m \frac{I(\omega_j)}{g(\omega_j; \hat{\boldsymbol{\varphi}}_W)},$$

где

$$Q(\boldsymbol{\varphi}) = -\ln \frac{1}{m} \sum_{j=1}^m \frac{I(\omega_j)}{g(\omega_j; \boldsymbol{\varphi})} - \frac{1}{m} \sum_{j=1}^m \ln g(\omega_j; \boldsymbol{\varphi}).$$

Замечание 3.3. Такой метод оценки параметров можно использовать при неизвестном среднем, поскольку при ее вычислении не используется значение периодограммы в нуле.

Замечание 3.4. Периодограмму временного ряда можно вычислить за $O(n \log n)$ с помощью быстрого преобразования Фурье (FFT), что делает этот метод значительно быстрее MLE для больших n .

3.3. Численное сравнение методов оценки параметров

Сравним качество оценки параметров следующих моделей:

1. $d = q = 0$, $p = 1$ — модель AR(1);
2. $p = q = 0$ — модель ARFIMA(0, d, 0).
3. $p = 1$, $q = 0$ — модель ARFIMA(1, d, 0).

Будем оценивать параметры этих моделей методами, реализация которых есть на языке программирования R [10]:

1. Функция `arima` из пакета `stats`, соответствующая MLE модели ARMA;
2. Функция `arfima` из пакета `arfima` [11], соответствующая MLE модели ARFIMA;
3. Функция `fracdiff` из пакета `fracdiff` [12], соответствующая аппроксимации MLE модели ARFIMA, описанной в работе [5] (обозначим за H&R).

Качественная реализации оценки Уиттла не была найдена, поэтому для сравнения с MLE она была реализована самостоятельно. Поскольку для реальных временных рядов матожидание μ неизвестно, будем рассматривать MLE с известным средним и с его оценкой — выборочным средним (будем обозначать их $\text{MLE}(\mu)$ и $\text{MLE}(\bar{x})$ соответственно). Не умаляя общности, пусть $\mu = 0$.

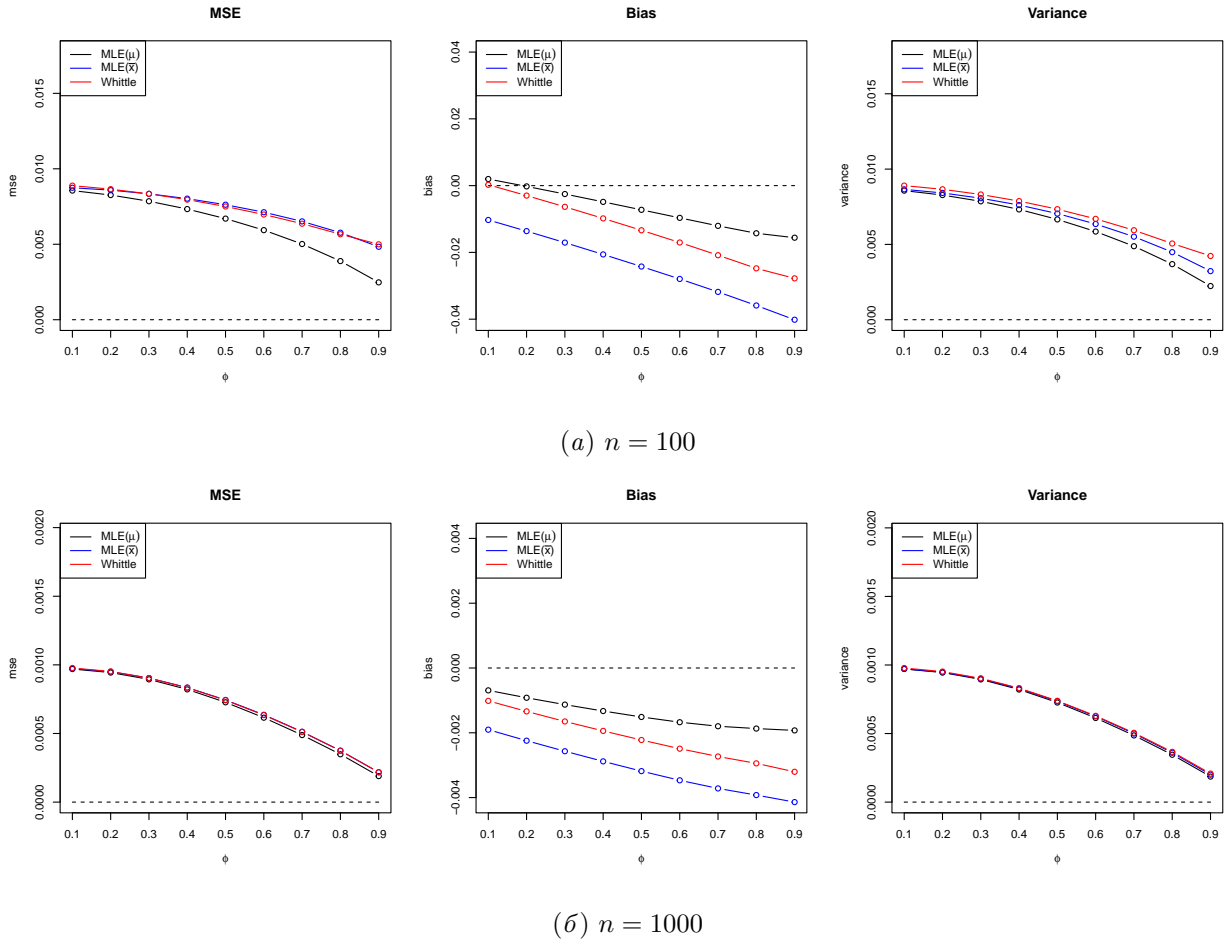


Рис. 1. Среднеквадратичное отклонение, смещение и дисперсия оценок параметра ϕ модели AR(1) (500 повторений)

На рис. 1.1 и 1.2 изображены среднеквадратичное отклонение, смещение и дисперсия оценок параметров ϕ и d моделей AR(1) и ARFIMA(0, d , 0). Отметим, что все

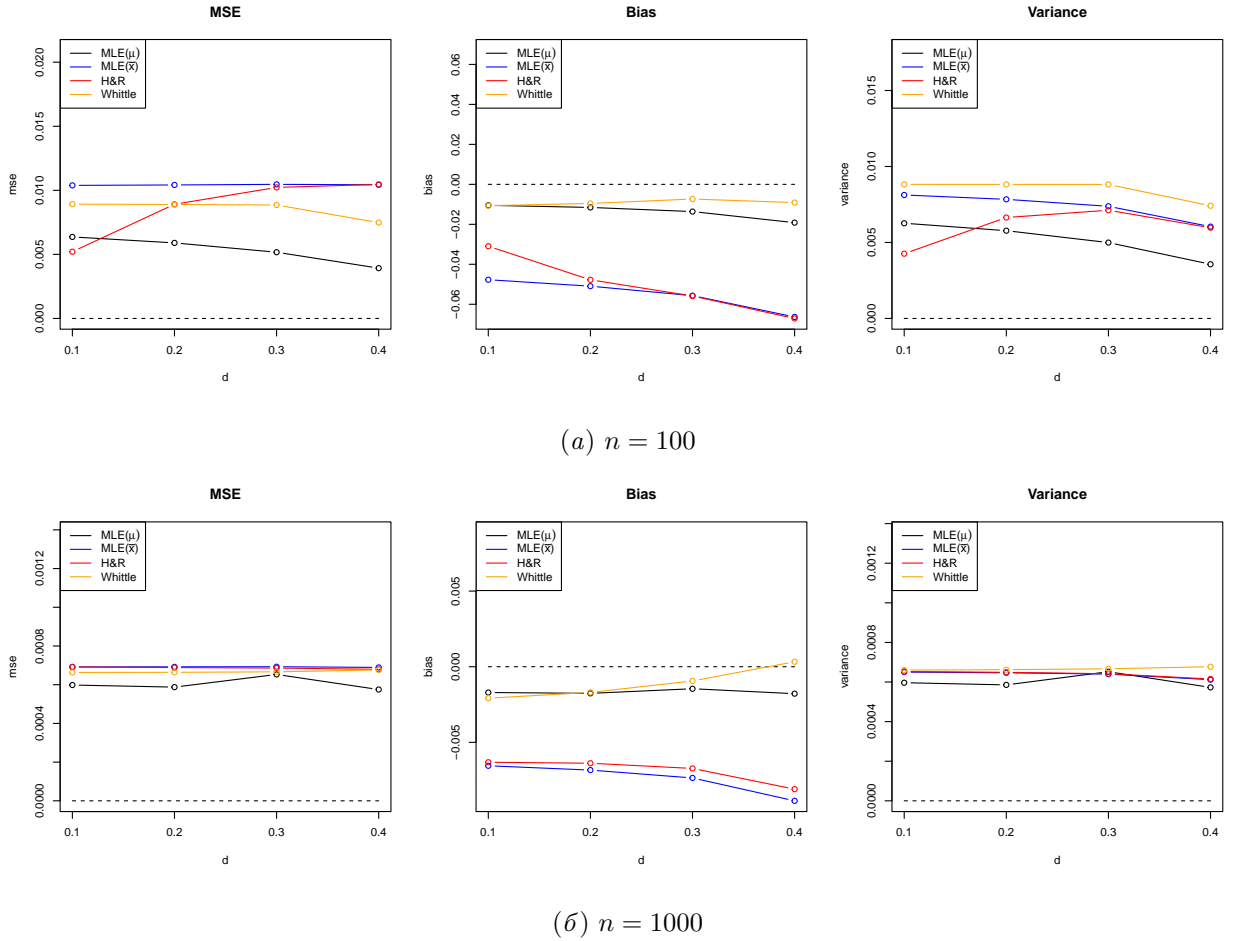


Рис. 2. Среднеквадратичное отклонение, смещение и дисперсия оценок параметра d модели $ARFIMA(0, d, 0)$ (500 повторений)

оценки имеют в большинстве своем отрицательное смещение и отличаются между собой в основном степенью смещения. Как и ожидалось, оценка параметров методом максимального правдоподобия с известным средним дает оценку с наименьшим **MSE**. С другой стороны, если использовать вместо известного среднего выборочное среднее, оценки становятся сильно смещенными. Whittle же, в свою очередь, дает менее смещенную оценку, чем $MLE(\bar{x})$, а в случае оценки d имеет смещение даже меньше, чем у $MLE(\mu)$. Однако оценки Whittle обладают наибольшей дисперсией среди всех рассмотренных методов, но разница не такая значительная, как в случае смещения.

В таблице 1.1 представлены значения среднеквадратичной ошибки и смещения оценок параметров d и ϕ модели $ARFIMA(1, d)$. Заметим, что в оценках присутствует смещение, и оно уменьшается с увеличением ϕ . Также для $\phi = 0.1$ и $\phi = 0.5$ оценки d имеют отрицательное смещение, а оценки ϕ , наоборот, положительное. Также в таб-

Таблица 1. Смещение и среднеквадратичное отклонение оценок параметров d и ϕ модели ARFIMA(1, d , 0) ($n = 100, 500$ повторений)

d	ϕ	MSE										Bias									
		MLE(μ)					MLE(\bar{x})					H&R					Whittle				
		\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$	\hat{d}	$\hat{\phi}$				
0.1	0.1	0.071	0.075	0.205	0.179	0.009	0.018	0.069	0.067	-0.096	0.083	-0.301	0.26	-0.054	0.035	-0.086	0.068				
0.2	0.1	0.064	0.07	0.224	0.201	0.025	0.032	0.077	0.073	-0.095	0.083	-0.336	0.296	-0.119	0.099	-0.094	0.074				
0.3	0.1	0.081	0.086	0.303	0.272	0.049	0.055	0.084	0.081	-0.12	0.108	-0.427	0.386	-0.179	0.161	-0.109	0.09				
0.4	0.1	0.072	0.08	0.367	0.328	0.081	0.089	0.179	0.194	-0.119	0.111	-0.499	0.456	-0.243	0.23	-0.26	0.241				
0.1	0.5	0.057	0.048	0.099	0.059	0.01	0.015	0.057	0.054	-0.102	0.065	-0.258	0.185	-0.07	0.033	-0.066	0.024				
0.2	0.5	0.055	0.045	0.107	0.062	0.031	0.025	0.074	0.058	-0.108	0.071	-0.277	0.201	-0.154	0.108	-0.153	0.107				
0.3	0.5	0.051	0.041	0.119	0.068	0.063	0.043	0.098	0.062	-0.114	0.079	-0.304	0.225	-0.232	0.174	-0.209	0.161				
0.4	0.5	0.052	0.043	0.129	0.074	0.104	0.065	0.104	0.066	-0.131	0.098	-0.326	0.244	-0.306	0.235	-0.228	0.177				
0.1	0.9	0.029	0.028	0.013	0.008	0.007	0.007	0.034	0.025	0.07	-0.085	0.003	-0.044	0.001	-0.043	0.049	-0.069				
0.2	0.9	0.019	0.021	0.011	0.006	0.009	0.004	0.026	0.019	0.038	-0.063	-0.013	-0.034	-0.037	-0.026	0.02	-0.056				
0.3	0.9	0.013	0.016	0.009	0.004	0.014	0.003	0.022	0.015	0.007	-0.043	-0.033	-0.022	-0.076	-0.011	-0.024	-0.039				
0.4	0.9	0.009	0.018	0.009	0.004	0.025	0.002	0.028	0.01	-0.019	-0.036	-0.059	-0.011	-0.121	0.003	-0.095	-0.016				

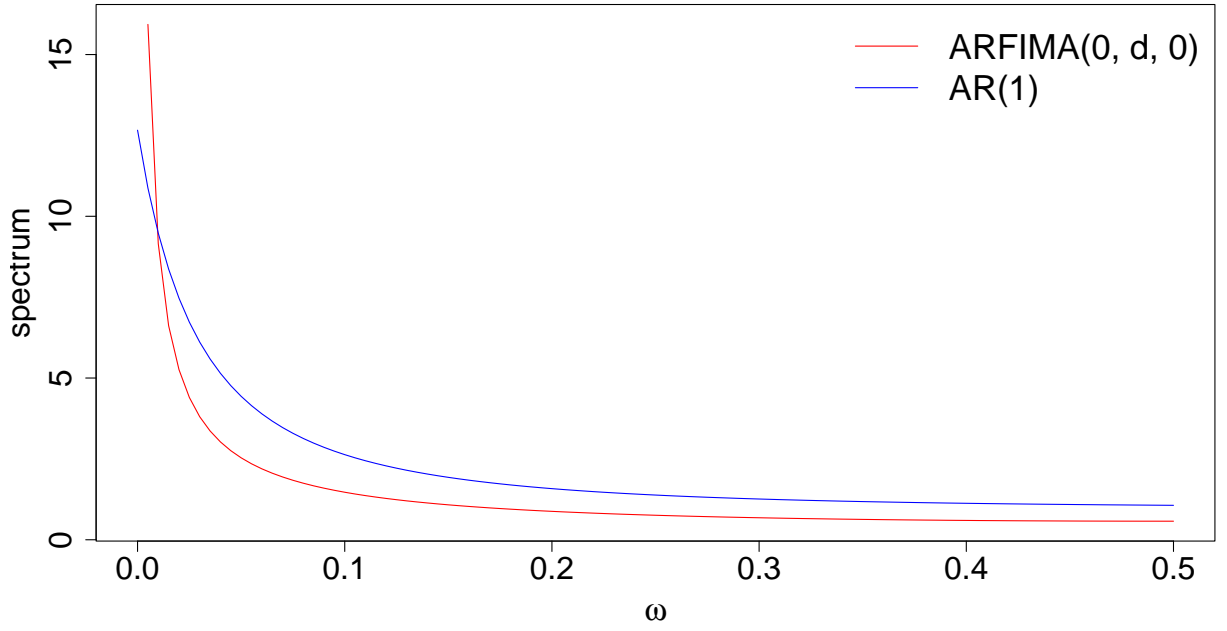


Рис. 3. Спектральная плотность процессов с одинаковой дисперсией

лице синим цветом выделена лучшая по строке оценка d , а красным — лучшая оценка ϕ . Видно, что метод Н&R в большинстве случаев дает оценки с наименьшим MSE, однако оценки имеют большое смещение при $\phi = 0.1$ и $\phi = 0.5$. Наиболее смещенные оценки получаются в методе $\text{MLE}(\bar{x})$, причем это перестает быть таковым только при $\phi = 0.9$. Оценки методом Whittle по смещению находятся посередине, однако имеют наибольшую дисперсию среди остальных.

4. Сравнение Monte Carlo SSA при разных моделях шума

Пусть $\{X_t\}$ — модель $\text{AR}(1)$, а $\{Y_t\}$ — модель $\text{ARFIMA}(0, d, 0)$. Будем считать дисперсию белого шума одинаковой для обоих процессов и равной σ^2 . Дисперсии $\{X_t\}$ и $\{Y_t\}$ соответственно равны

$$\text{D}X_t = \frac{\sigma^2}{1 - \phi^2}, \quad \text{D}Y_t = \sigma^2 \frac{\Gamma(1 - 2d)}{\Gamma(1 - d)^2}.$$

Тогда дисперсии процессов равны тогда и только тогда, когда

$$\phi = \pm \sqrt{1 - \frac{\Gamma(1 - d)^2}{\Gamma(1 - 2d)}}.$$

Пусть $d = 0.4$. Тогда при $\phi \approx 0.719$ процессы $\{X_t\}$ и $\{Y_t\}$ имеют одинаковую дисперсию. На рис. 2.1 изображены спектральные плотности процессов. На нем видно, что процесс $\{Y_t\}$ имеет меньшее значение плотности для всех значений ω , за исключением близких к нулю.

Предполагается, что если рассмотреть сигнал с частотой $\omega : f_Y(\omega) < f_X(\omega)$, то критерий MC-SSA с суррогатными реализациями модели ARFIMA(0, d , 0) даст мощность больше, чем тот же критерий с суррогатными реализациями модели AR(1). Убедимся в этом. Пусть длина ряда $N = 100$ и

$$S = \{A \cos(2\pi n\omega)\}_{n=1}^N, \quad A = 1, \quad \omega = 0.075.$$

Тогда $H_0 : S = 0$ и $H_0 : S \neq 0$.

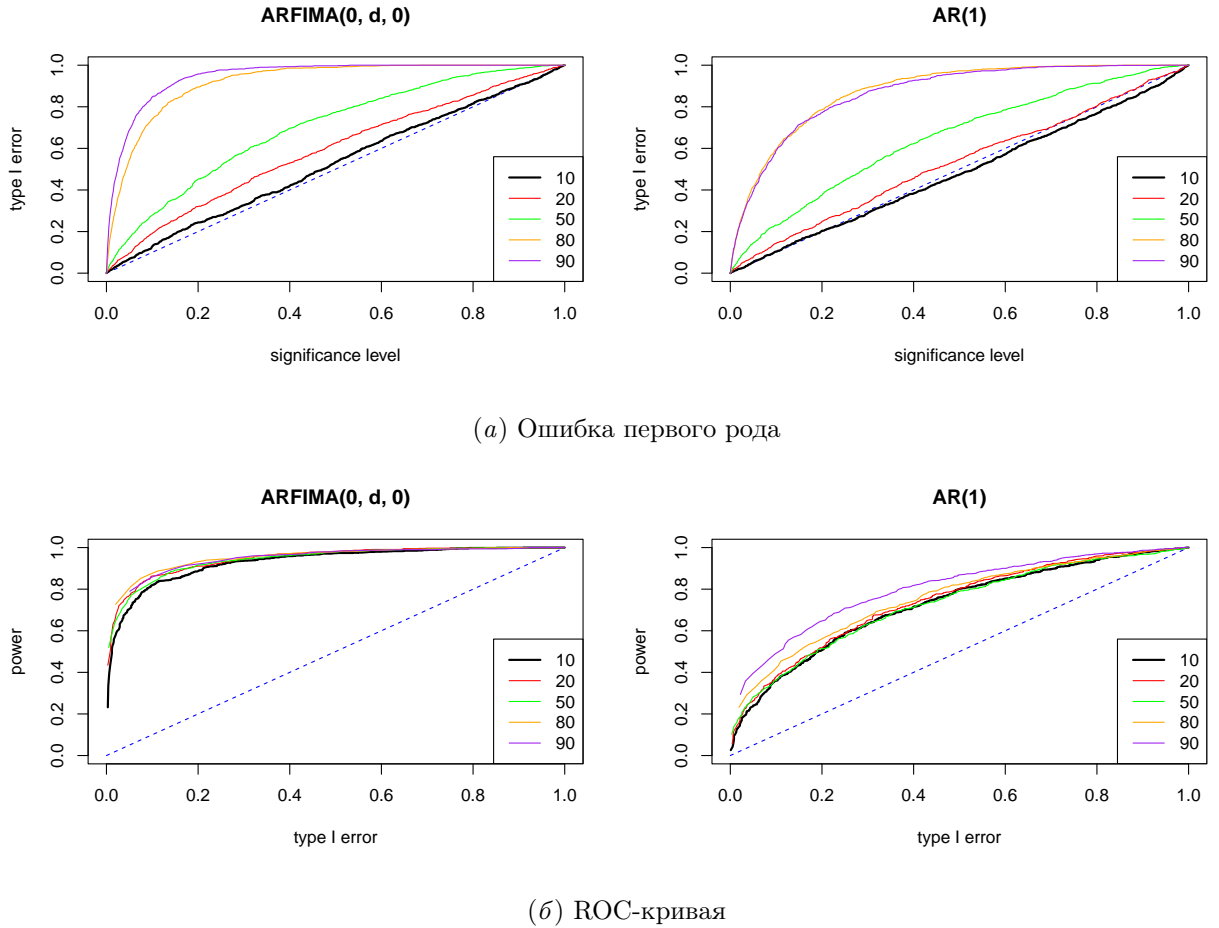


Рис. 4. Сравнение мощностей MC-SSA

На рис. 2.2 изображены график ошибок первого рода и ROC-кривые критериев MC-SSA для разных длин окна L . По рис. 2.2, б видно, что, действительно, замена модели

AR(1) на ARFIMA(0, d , 0) увеличила мощность критерия против данной альтернативы. Также по рис. 2.2, *a* заметно, что такой критерий стал более радикальным, а значит может требовать большего числа суррогатных данных G для поправки критерия.

Список литературы

1. Palma Wilfredo. Long-Memory Time Series: Theory and Methods. — Wiley, 2006.
2. Time Series Analysis: Forecasting and Control / Box G., Jenkins G., Reinsel G., and Ljung G. — Fifth ed. — 2016.
3. Hassler Uwe. Time Series Analysis with Long Memory in View. — Wiley, 2018.
4. Hipel Keith W., McLeod Ian. Time series modelling of water resources and environmental systems. — Elsevier, 1994.
5. Haslett John, Raftery Adrian E. Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource // Journal of the Royal Statistical Society. Series C (Applied Statistics). — 1989. — Vol. 38, no. 1. — P. 1–50.
6. Long memory effects and forecasting of earthquake and volcano seismic data / Mariani Maria C., Bhuiyan Md Al Masum, Tweneboah Osei K. and Gonzalez-Huizar Hector // Physica A: Statistical Mechanics and its Applications. — 2020. — Vol. 559. — P. 125049.
7. Barkoulas J., Labys W. C., Onochie J. I. Fractional dynamics in international commodity prices // Journal of Futures Markets. — 1997. — Vol. 17. — P. 161–189.
8. Guglielmo Maria Caporale Luis Gil-Alana, Plastun Alex. Long memory and data frequency in financial markets // Journal of Statistical Computation and Simulation. — 2019. — Vol. 89, no. 10. — P. 1763–1779.
9. Whittle P. The Analysis of Multiple Stationary Time Series // Journal of the Royal Statistical Society. Series B (Methodological). — 1953. — P. 125–139.
10. R Core Team. — R: A Language and Environment for Statistical Computing. — R Foundation for Statistical Computing, Vienna, Austria, 2024. — Access mode: <https://www.R-project.org/>.
11. Veenstra J.Q. — arfima: Fractional ARIMA (and Other Long Memory) Time Series Modeling : 2012.
12. Maechler Martin. — fracdiff: Fractionally Differenced ARIMA aka ARFIMA(P,d,q) Models : 1999.