

# ПОЛИНОМИАЛЬНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ ДАННЫХ ПРИ ВОЗДЕЙСТВИИ КОРРЕЛИРОВАННЫХ ПОМЕХ

*Омельченко А.В., Федоров А.В.*

## **Введение**

Регрессионный анализ является одним из наиболее распространенных статистических методов, которые используются при построении математических зависимостей на основе экспериментальных данных. Чаще всего функции регрессии строятся в классе полиномиальных функций, а коэффициенты полиномов определяются по методу наименьших квадратов (МНК) [1-4]. Однако, в тех случаях, когда ошибки наблюдения коррелированы, МНК не обеспечивает минимально возможных дисперсий оценок коэффициентов полинома, аппроксимирующего функцию регрессии. В этих случаях могут использоваться взвешенный метод наименьших квадратов (ВМНК) и обобщенный метод наименьших квадратов (ОМНК) [3]. Однако остаются открытыми вопросы их эффективности для различных условий, определяющих вид ошибок наблюдения.

В ряде прикладных задач ошибки наблюдения формируются источниками аддитивных помех, которые обладают определенными корреляционными характеристиками. В результате при регрессионном анализе данных ошибки наблюдения оказываются коррелированными.

Поэтому актуальна задача обоснования оптимальных алгоритмов полиномиального регрессионного анализа для различных моделей ошибок наблюдения (помех).

Наиболее разработанными являются модели помех в виде стационарных и однородных нестационарных временных рядов [5, 6]. Для таких рядов существуют хорошо развитые методы оценивания параметров корреляционных функций. Даже в тех случаях, когда данные по предикторной переменной не эквидистантны, зачастую их можно свести к эквидистантной последовательности, что позволяет качественно оценить параметры корреляционных функций ошибок наблюдения и учесть их в алгоритмах регрессионного анализа.

Цели работы: 1) сравнительный анализ эффективности линейных методов оценивания коэффициентов полиномиальной модели в случае коррелированных ошибок наблюдения и обоснование условий их применимости; 2) обоснование алгоритмов оптимального оценивания коэффициентов полиномиальной модели.

Основные результаты работы получены для случая равноотстоящих значений предикторной переменной.

## **Постановка задачи**

Пусть наблюдается  $k$  значений

$$Z_k = a_0 + a_1 x_k + a_1 x_k^2 \dots + a_n x_k^n + \xi_k, \quad k=0,1,\dots,K-1, \quad (1)$$

где  $x_k$  - значения предикторной переменной;  $\xi_k$  - отсчеты ошибок наблюдения с нулевым средним и заданными корреляционными характеристиками;  $a_i, i=0,1,\dots,n$  - неизвестные коэффициенты, подлежащие оцениванию.

Записывая  $k$  уравнений наблюдения в матричной форме, получим

$$\vec{Z} = X \vec{a} + \vec{\xi}, \quad (2)$$

где векторы

$$\vec{Z} = (z_0, \dots, z_{K-1})^{\text{tr}}; \vec{\xi} = (\xi_0, \dots, \xi_{K-1})^{\text{tr}}; \vec{a} = (a_0, \dots, a_n)^{\text{tr}}; \quad (3)$$

и матрица

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{K-1} & x_{K-1}^2 & \dots & x_{K-1}^n \end{bmatrix}. \quad (4)$$

Матрица  $X$  размера  $K \times (n+1)$  называется регрессионной матрицей, а ее столбцы называются регрессорами или факторами [3].

Требуется найти несмещенные оценки коэффициентов  $a_i, i=0,1,\dots,n$  с минимальной дисперсией.

### Методы оценивания

Одним из наиболее распространенных методов оценивания вектора  $\vec{a}$  является метод наименьших квадратов [1-4]. В этом случае решение может быть найдено из нормального уравнения

$$X^{\text{tr}} X \vec{a} = X^{\text{tr}} \vec{Z}. \quad (5)$$

Если матрица  $X$  имеет ранг  $n+1$ , то уравнение (5) имеет единственное решение

$$\hat{\vec{a}} = (X^{\text{tr}} X)^{-1} X^{\text{tr}} \vec{Z}. \quad (6)$$

Будем полагать, что координаты вектора ошибки наблюдения  $\vec{\xi}$  имеют нулевые математические ожидания

$$M[\xi_k] = 0, \quad k = 0, 1, \dots, K-1. \quad (7)$$

Тогда

$$M[\hat{\vec{a}}] = \vec{a}$$

и оценка МНК (6) есть несмещенная оценка вектора  $\vec{a}$ .

Дисперсионная (корреляционная) матрица оценки  $\hat{\vec{a}}$  равна

$$M[(\hat{\vec{a}} - \vec{a})(\hat{\vec{a}} - \vec{a})^{\text{tr}}] = \sigma^2 (X^{\text{tr}} X)^{-1}. \quad (8)$$

Из теоремы Гаусса-Маркова [4] следует, что в случае некоррелированных ошибок наблюдений  $\xi_k$ , удовлетворяющих условию гомоскедастичности

$$M[\xi_k^2] = \sigma^2,$$

оценка (6) обеспечивает минимально возможную дисперсию каждого из полиномиальных коэффициентов.

Имеются результаты, принадлежащие Гренандеру и устанавливающие условия асимптотической эффективности оценок МНК при ошибках наблюдения в виде стационарной случайной последовательности. Показано [5], что необходимым и достаточным условием для этого, в случае полиномиальных регрессоров, является непрерывность спектральной плотности ошибок наблюдения  $g(\lambda)$  в нуле.

Если имеется гетероскедастичность ошибок наблюдения

$$M[\xi_k^2] = \sigma_k^2, \quad k = 0, \dots, K-1,$$

то следует использовать оценку ВМНК [2, 3]

$$\hat{\vec{a}} = (\mathbf{X}^{\text{tr}} \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^{\text{tr}} \mathbf{W} \vec{Z}, \quad (9)$$

где диагональная матрица

$$\mathbf{W} = \text{diag}\{w_0, \dots, w_{K-1}\},$$

а весовые коэффициенты  $w_k = \frac{1}{\sigma_k^2}$ .

Отметим, что последовательность коэффициентов  $w_k, k = 0, \dots, K-1$  может трактоваться как весовое окно. Если ошибки наблюдения коррелированы, то, как установлено в [7] и показано ниже, подбирая это окно, можно уменьшить дисперсии оценок коэффициентов регрессии по сравнению с соответствующими оценками МНК. Это объясняется тем, что весовое окно в регрессионном анализе действует подобно весовым окнам в спектральном анализе, ослабляя влияние интенсивных компонент помехи на результат обработки данных.

Дисперсионная матрица оценки вида (9) равна

$$M[(\hat{\vec{a}} - \vec{a})(\hat{\vec{a}} - \vec{a})^{\text{tr}}] = (\mathbf{X}^{\text{tr}} \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^{\text{tr}} \mathbf{W} \mathbf{R} \mathbf{W} \mathbf{X} (\mathbf{X}^{\text{tr}} \mathbf{W} \mathbf{X})^{-1}, \quad (10)$$

где  $\mathbf{R}$  – корреляционная матрица помехи.

Если ошибки наблюдений коррелированы, то следует использовать оценку ОМНК [2, 3]

$$\vec{a}^* = (\mathbf{X}^{\text{tr}} \mathbf{R}^{-1} \mathbf{X})^{-1} \mathbf{X}^{\text{tr}} \mathbf{R}^{-1} \vec{Z}. \quad (11)$$

Оценка ОМНК (11) является несмещенной, а ее дисперсионная матрица равна

$$M[(\vec{a}^* - \vec{a})(\vec{a}^* - \vec{a})^{\text{tr}}] = (\mathbf{X}^{\text{tr}} \mathbf{R}^{-1} \mathbf{X})^{-1}. \quad (12)$$

При этом дисперсия каждой из координат оценки ОМНК  $\vec{a}^*$  является минимально возможной в классе линейных оценок.

Таким образом, для оптимального оценивания коэффициентов регрессии в случае коррелированных ошибок наблюдений нужно располагать моделью ошибок наблюдения, позволяющей определить корреляционную матрицу  $\mathbf{R}$ .

Эффективность оценок будем характеризовать совокупностью показателей

$$\eta_i = D_i / D_{i \text{ ОМНК}}, \quad i = 0, \dots, n \quad (13)$$

где  $D_i$  и  $D_{i \text{ ОМНК}}$  - соответственно дисперсия заданной оценки  $i$ -го коэффициента полиномиальной регрессии и дисперсия оценки ОМНК того же коэффициента.

### **Модели помех с дискретным временем**

Рассмотрим модели процессов с нулевым средним, которые будем использовать для описания помехи (ошибки наблюдения) в прикладных задачах регрессионного анализа.

#### Модели «цветных» шумов

В основе классификации шумов лежит скорость изменения спектральной плотности  $G(f) \sim f^\beta$  [8]:

- 1) белый шум имеет постоянную спектральную плотность ( $\beta = 0$ );
- 2) для розового шума спектральная плотность уменьшается на 3 дБ с каждой последующей октавой (спектральная плотность обратно пропорциональна частоте и  $\beta = -1$ );
- 3) коричневый или красный шум - спектральная плотность уменьшается на 6 дБ с каждой последующей октавой (спектральная плотность обратно пропорциональна квадрату частоты и  $\beta = -2$ );
- 4) синий шум - спектральная плотность увеличивается на 3 дБ с каждой последующей октавой (спектральная плотность пропорциональна частоте,  $\beta = 1$ );
- 5) фиолетовый шум - дифференцированный белый шум, его спектральная плотность увеличивается на 6 дБ с каждой последующей октавой (спектральная плотность пропорциональна квадрату частоты и  $\beta = 2$ ).

Ниже уточняются некоторые случаи «цветных» шумов.

#### Модель ARMA

Одной из популярных математических моделей, используемых для анализа и прогнозирования временных рядов, являются процессы авторегрессии-скользящего среднего (autoregressive moving-average model, ARMA).

Считается, что процесс с нулевым средним принадлежит классу ARMA( $p; q$ ), если для его отсчетов справедливо уравнение [6]

$$\xi_k = \sum_{j=1}^p \phi_j \xi_{k-j} + \sum_{i=0}^q \theta_i \eta_{k-i}, \quad (14)$$

где  $p$  и  $q$  - натуральные числа, задающие порядок модели;  $\eta_k$  - процесс белого шума с дисперсией  $\sigma_\eta^2$ ;  $\phi_j, j = \overline{1, p}$  и  $\theta_i, i = \overline{0, q}$  - вещественные числа, называемые соответственно коэффициентами авторегрессии и скользящего среднего, причем  $\theta_0 = 1$ .

Процесс ARMA(p ; q) стационарен, если все корни уравнения

$$1 - \phi_1 z - \dots - \phi_p z^p = 0$$

лежат вне единичного круга  $|z| \leq 1$ .

### Модель ARIMA

Класс процессов авторегрессии проинтегрированного скользящего среднего (АРПСС или ARIMA) является очень важной моделью, которая позволяет описывать однородные нестационарные временные ряды. Эта модель предложена Дж. Боксом и Г. Дженкинсом [6]. Она предназначена для описания нестационарных временных рядов, обладающих свойством стохастического тренда.

Перейдем теперь к описанию модели нестационарного временного ряда АРПСС или ARIMA. При этом весьма полезным является оператор конечной разности назад  $\nabla = 1 - B$ , где  $B$  – оператор сдвига сигнала назад, определяемый как  $Bz_k = z_{k-1}$ .

Оператор  $\nabla$  преобразует нестационарный процесс случайного блуждания  $\xi_k$  в стационарный процесс  $\nabla \xi_k = \eta_k$ . В моделях детерминированных временных рядов  $\nabla$  позволяет элиминировать (исключить) полиномиальный тренд первого порядка.

Процесс ARIMA(p ; v ; q) описывается уравнением

$$\sum_{i=0}^p b_i \xi_{k-i}^v = \sum_{j=0}^q \theta_j \eta_{k-j}, \quad (15)$$

где коэффициенты  $\theta_0 = b_0 = 1$ ;  $\eta_k$  – последовательность независимых отсчетов с дисперсией  $\sigma^2$ ;  $\xi_k^v = \nabla^v \xi_k$  – v-ая конечная разность процесса  $\xi_k$ .

Таким образом, для процесса АРПСС его v-ая разность представляет собой процесс авторегрессии-скользящего среднего (АРСС). Процесс АРПСС характеризуется тремя целочисленными параметрами p, v, q.

Заметим, что классу моделей ARIMA принадлежит и простейшая модель стохастического тренда – процесс случайного блуждания (или просто случайное блуждание). Случайное блуждание определяется аналогично процессу авторегрессии первого порядка, но только у случайного блуждания коэффициент авторегрессии  $\phi_1 = 1$ , так что

$$\xi_k - \xi_{k-1} = \eta_k. \quad (16)$$

Ряд первых разностей случайного блуждания  $\eta_k$  представляет собой белый шум, т.е. процесс ARMA(0;0). Поэтому само случайное блуждание входит в класс моделей ARIMA как модель ARIMA(0;1;0) или в сокращенной записи модель ARI(0;1) и описывает коричневый шум.

### Модель ARFIMA

Модель ARFIMA(p ; d ; q) является обобщением модели ARIMA(p ; v ; q) на случай дробного значения параметра d [8]. Он получил название авторегрессионный дробно-интегрированный процесс скользящего среднего.

Процесс ARFIMA(p;d;q) является стационарным и обратимым для  $d \in (-1/2, 1/2)$ .

Параметр  $d$  связан с широко известным параметром Хёрста  $H$  соотношением  $d = H - 1/2$ . Если показатель Хёрста изменяется от  $1/2$  до  $1$ , т.е.  $d \in (0, 1/2)$ , то процесс характеризуется долговременной памятью. Это означает, что последующие значения процесса сильно зависят от прошедших значений. Показатель Хёрста, колеблющийся от  $0$  до  $1/2$ , т.е.  $d \in (-1/2, 0)$ , означает антиперсистентный процесс, который меняется быстрее, чем «белый шум». Можно сказать, что чем выше показатель Хёрста, тем меньше на временном ряду «зазубрин».

Известно [8], что корреляционная функция процесса ARFIMA(0;d;0) описывается выражением

$$r[k] = \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k+1-d)\Gamma(1-k-d)} \sigma_\eta^2, \quad (17)$$

а спектральная плотность

$$g(f) = \sigma_\eta^2 (2 \sin \pi f)^{-2d}, \quad |f| \leq 0,5. \quad (18)$$

В теории фрактальных шумов значения  $d \in (0, 1/2)$  соответствуют персистентному процессу синего шума, а  $d \in (-1/2, 0)$  эквивалентно антиперсистентной системе розового шума. Белый шум соответствует  $d=0$ .

### **Регрессионный анализ при воздействии аддитивной помехи в виде процесса авторегрессии**

Если процесс авторегрессии порядка  $p$  распределен нормально, то совместная плотность вероятности распределения его отсчетов описывается выражением [6]

$$p(x_0, \dots, x_{K-1}) = p(x_0, \dots, x_{p-1}) \cdot \frac{1}{(2\pi)^{(K-p)/2} \sigma^{(K-p)}} \times \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=p}^{K-1} \left[\sum_{s=0}^p b_s x_{k-s}\right]^2\right\}, \quad (19)$$

где коэффициенты  $b_0 = 1$ ;  $b_s = -\phi_s$ ,  $s = \overline{1, p}$ ; совместная плотность вероятности первых  $p$  отсчетов

$$p(x_0, \dots, x_{p-1}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2} \sum_{i,j=0}^{p-1} \sigma_{ij}^{-1} x_i x_j\right\}, \quad (20)$$

а  $\sigma_{ij}^{-1}$  – элементы матрицы, обратной корреляционной матрице  $p$  первых отсчетов помехи.

С использованием декомпозиции Холецкого представим матрицу  $\Sigma$  в виде

$$\Sigma = \tilde{H} \tilde{H}^{\text{tr}}, \quad (21)$$



Из (29) получим квазиоптимальную (КО) оценку параметров регрессии

$$\hat{\mathbf{a}} = [(\mathbf{E}^{\text{tr}} \mathbf{E})^{-1} \mathbf{E}^{\text{tr}} \bar{\mathbf{y}}], \quad (31)$$

где вектор декоррелированных наблюдений

$$\bar{\mathbf{y}} = \mathbf{B}_y \bar{\mathbf{z}},$$

регрессионная матрица

$$\mathbf{E} = \mathbf{B}_y \mathbf{X}.$$

Несложно показать, что оценка (31) несмещенная.

### **Оценивание коэффициентов линейной и квадратичной регрессии при ошибках наблюдения в виде процесса авторегрессии**

Исследуем эффективность МНК, ВМНК, ОМНК и КО оценок коэффициентов регрессии при воздействии различных помех. В качестве предикторной переменной, задающей матрицу  $\mathbf{X}$  согласно (4), будем использовать целочисленную переменную  $x_k = k$ ,  $k = 0, K - 1$ .

Для нахождения дисперсий оценок ОМНК коэффициентов регрессии можно воспользоваться выражением для дисперсионной матрицы (12), диагональные элементы которой и являются дисперсиями оценок коэффициентов полиномиальной функции. Применительно к помехам типа АР-процессов выражение для дисперсионной матрицы (12) удобно представить в виде

$$\mathbf{M}[(\hat{\mathbf{a}} - \bar{\mathbf{a}})(\hat{\mathbf{a}} - \bar{\mathbf{a}})^{\text{tr}}] = (\mathbf{E}^{\text{tr}} \mathbf{E})^{-1}, \quad (32)$$

где согласно (27) матрица  $\mathbf{E} = \mathbf{B}\mathbf{X}$ ;  $\mathbf{X}$  – регрессионная матрица;  $\mathbf{B}$  – нижняя треугольная матрица вида (24).

Для оценок МНК и КО также можно получить выражения для дисперсионных матриц, аналогичные (32). Так дисперсионная матрица вектора оценки МНК

$$\mathbf{M}[(\hat{\mathbf{a}} - \bar{\mathbf{a}})(\hat{\mathbf{a}} - \bar{\mathbf{a}})^{\text{tr}}] = (\mathbf{X}^{\text{tr}} \mathbf{X})^{-1} \mathbf{X}^{\text{tr}} \mathbf{R} \mathbf{X} (\mathbf{X}^{\text{tr}} \mathbf{X})^{-1}, \quad (33)$$

где  $\mathbf{R}$  – корреляционная матрица отсчетов помехи размерности  $K \times K$ , а дисперсионная матрица вектора КО оценок

$$\mathbf{M}[(\hat{\mathbf{a}} - \bar{\mathbf{a}})(\hat{\mathbf{a}} - \bar{\mathbf{a}})^{\text{tr}}] = (\mathbf{E}^{\text{tr}} \mathbf{E})^{-1} \mathbf{E}^{\text{tr}} \mathbf{B}_y \mathbf{R} \mathbf{B}_y^{\text{tr}} \mathbf{E} (\mathbf{E}^{\text{tr}} \mathbf{E})^{-1}, \quad (34)$$

где матрица  $\mathbf{E} = \mathbf{B}_y \mathbf{X}$ , а матрица  $\mathbf{B}_y$  размерности  $(K - p) \times K$  имеет ленточную структуру вида (30).

### **Исследование эффективности оценок коэффициентов регрессии для помехи в виде процесса авторегрессии первого порядка**

Для процесса авторегрессии первого порядка имеем лишь один коэффициент  $\phi_1 \neq 0$ . Корреляционная функция такого процесса

$$r[k] = \sigma^2 \phi_1^k, \quad k = 0, 1, \dots,$$

а спектральная плотность мощности

$$g(f) = \frac{2\sigma_{\eta}^2}{|1 - \phi_1 e^{-j2\pi f}|^2} = \frac{2\sigma_{\eta}^2}{(1 - \phi_1)^2 + 4\phi_1 \sin^2 \pi f}, \quad 0 \leq f \leq 1/2.$$

При  $\phi_1 > 0$  спектральная плотность мощности является убывающей функцией частоты  $f$ , а при  $\phi_1 < 0$  - возрастающей.

Весовое окно в алгоритме оценивания ВМНК описывалось весовой функцией ортогональных многочленов Хана [9]

$$w_k = \frac{\Gamma(k + \alpha + 1)\Gamma(K - k + \alpha)}{\Gamma(k + 1)\Gamma(K - k)}, \quad k = \overline{0, K - 1}, \quad (35)$$

где параметр  $\alpha \geq -1$ .

При различных значениях параметра  $\alpha$  выражение (35) позволяет строить симметричные окна различной формы. На рис. 1 показан вид весового окна (35) длиной  $K = 100$  для следующих случаев:  $\alpha = -0,5$  - пунктирная кривая;  $\alpha = 0,5$  - штрих пунктирная кривая;  $\alpha = 1$  - сплошная кривая. В случае  $\alpha = 0$  приходим к равномерному весовому окну.

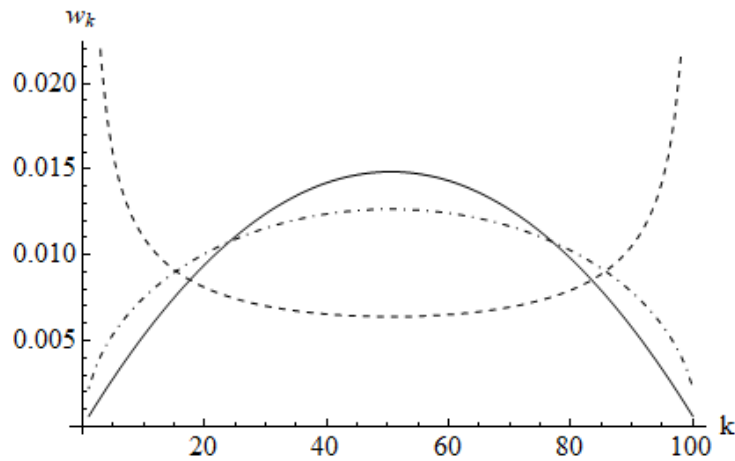
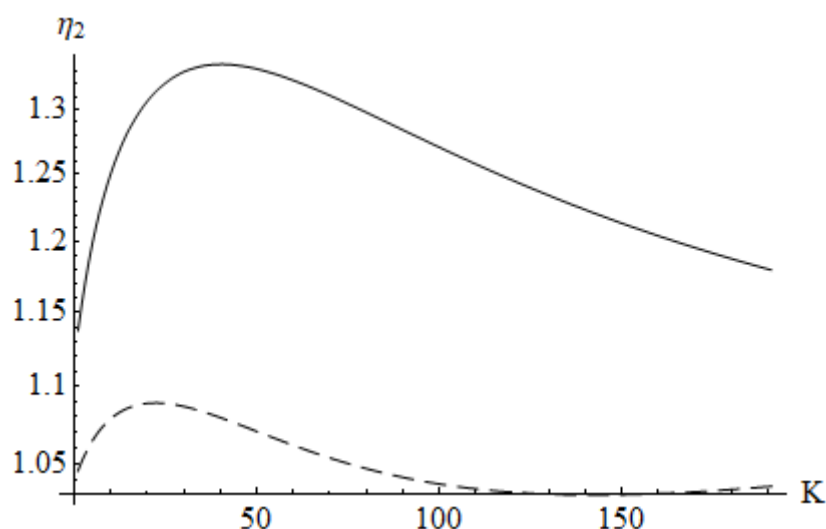


Рис. 1. Вид весовых окон

В результате вычислений было установлено, что КО оценка (31) в рассматриваемом случае медленно сходится к оптимальной оценке ОМНК и не может быть рекомендована к использованию.

При значениях коэффициента авторегрессии  $\phi_1 > 0$  оценки ВМНК с параметром окна  $\alpha > 0$  имеют большую дисперсию, чем оценки МНК. Однако, использование при этом в методе ВМНК окна с  $\alpha < 0$  может обеспечить меньшие дисперсии оценок по сравнению с оценками МНК. В качестве иллюстрации на рис. 2 изображены зависимости показателя эффективности оценки  $\eta_2$  от числа отсчетов  $\alpha = 0,5$  для случая полиномиальной модели второго порядка и модели помехи в виде процесса авторегрессии первого порядка с единичной дисперсией и коэффициентом  $\phi_1 = 0,9$ . В данном случае параметр окна  $\alpha = -0,5$ . Сплошной кривой здесь показана зависимость для оценки МНК, а пунктирной - для оценки ВМНК.



*Рис. 2. Зависимость показателя эффективности  $\eta_2$  от объема наблюдения  $K$  для модели помехи в виде процесса авторегрессии первого порядка с  $\phi_1 = 0,9$*

Отметим, что зависимости показателей  $\eta_0$ ,  $\eta_1$  и  $\eta_2$  от  $K$  имеют схожий вид, как при порядке полинома  $n=2$ , так и при  $n=1$ . Это свойство сохранялось во всех проанализированных нами случаях, как для процессов авторегрессии первого порядка, так и процессов авторегрессии второго порядка. Поэтому здесь и ниже приводятся лишь зависимости показателя  $\eta_2$  от  $K$ .

При значениях коэффициента авторегрессии  $\phi_1 < 0$  ситуация противоположная описанной выше. Здесь использование оценок ВМНК с параметром окна  $\alpha > 0$  может обеспечить меньшие дисперсии оценок по сравнению с оценками МНК. Так на рис. 3 изображены зависимости показателя эффективности  $\eta_2$  от числа отсчетов  $K$  для случая полиномиальной модели второго порядка и модели помехи в виде процесса авторегрессии первого порядка с коэффициентом  $\phi_1 = -0,9$ . В данном случае параметр окна  $\alpha = 0,5$ . Сплошной кривой здесь показана зависимость для оценки МНК, а пунктирной – для оценки ВМНК. Колебательный характер оценок в окрестности малых значений  $K$  обусловлен, по всей видимости, частотными свойствами весового окна.

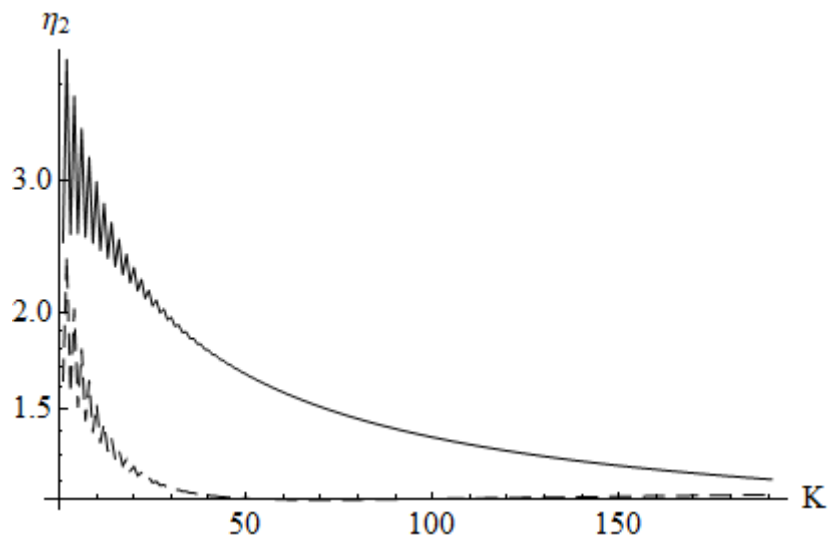


Рис. 3. Зависимость показателя эффективности  $\eta_2$  от объема наблюдения  $K$  для модели помехи в виде процесса авторегрессии первого порядка с  $\phi_1 = -0,9$

### Исследование эффективности оценок коэффициентов регрессии для помехи в виде процесса авторегрессии второго порядка

Ниже представлены результаты исследования эффективности оценок коэффициентов линейной и квадратичной регрессии.

В качестве помехи использовался процесс авторегрессии второго порядка с единичной дисперсией и коэффициентами  $\phi_1 = 1,864$ ;  $\phi_2 = -0,960$ .

Спектральная плотность мощности процесса авторегрессии второго процесса выражается как [6]

$$g(f) = \frac{2\sigma_{\eta}^2}{|1 - \phi_1 e^{-j2\pi f} - \phi_2 e^{-j4\pi f}|^2}, \quad 0 \leq f \leq 1/2.$$

Для рассматриваемого случая вид спектральной плотности процесса авторегрессии представлен на рис. 4.

На рис. 5 – рис. 7 показаны зависимости дисперсий оценок ОМНК коэффициентов регрессии от объема наблюдений  $K$  для полинома порядка  $n=1$  (пунктирная кривая) и полинома порядка  $n=2$  (сплошная кривая). Большие значения дисперсий оценок коэффициентов регрессии при  $n=2$ , чем при  $n=1$ , при тех же значениях  $K$  являются платой за увеличение степени полинома.

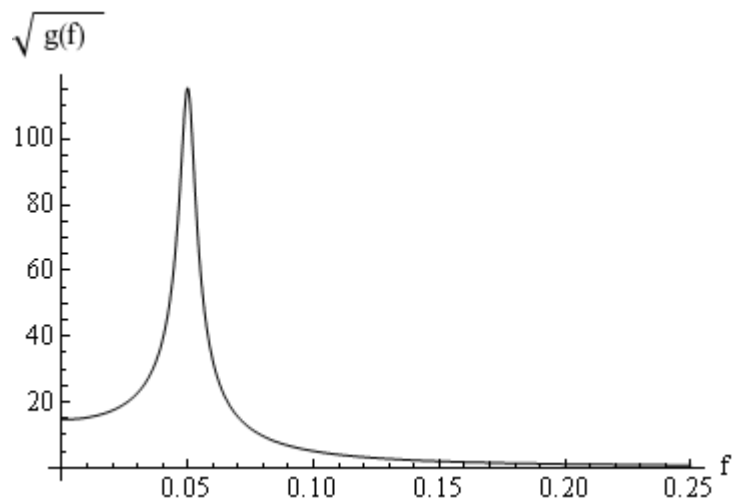


Рис. 4. Вид спектральной плотности процесса авторегрессии второго порядка

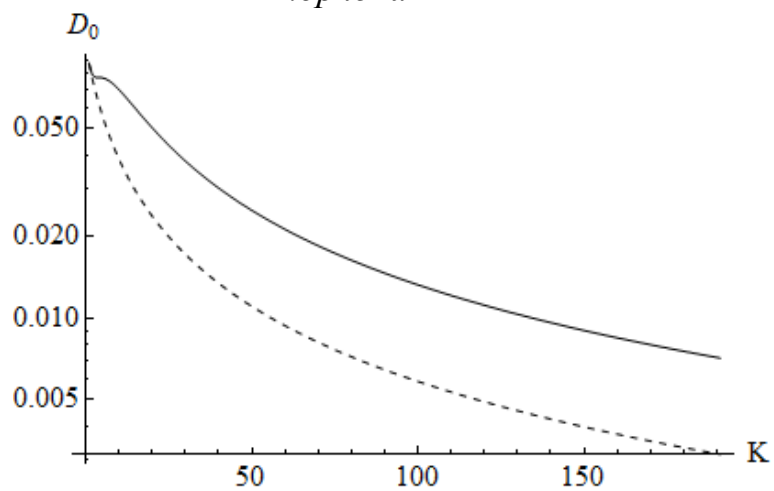


Рис. 5. Зависимости дисперсии оценок коэффициента  $a_0$  для моделей первого и второго порядков от объема наблюдения

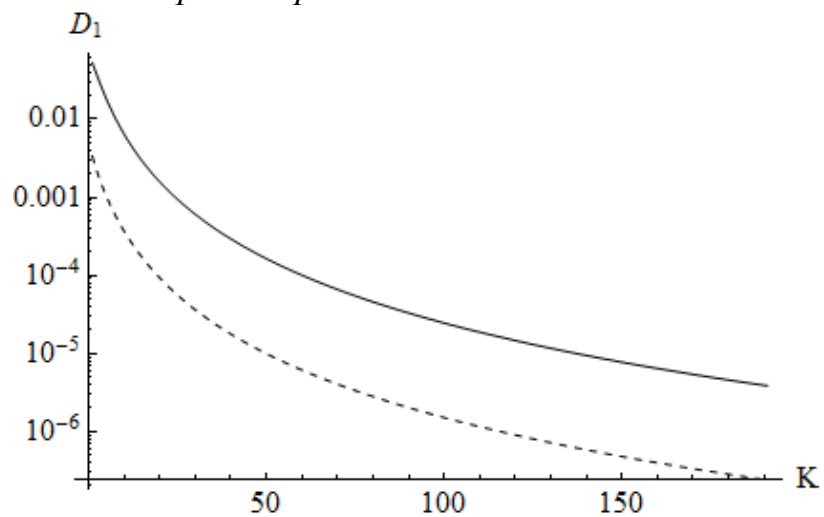


Рис. 6. Зависимости дисперсии оценок коэффициента  $a_1$  для моделей первого и второго порядков от объема наблюдения

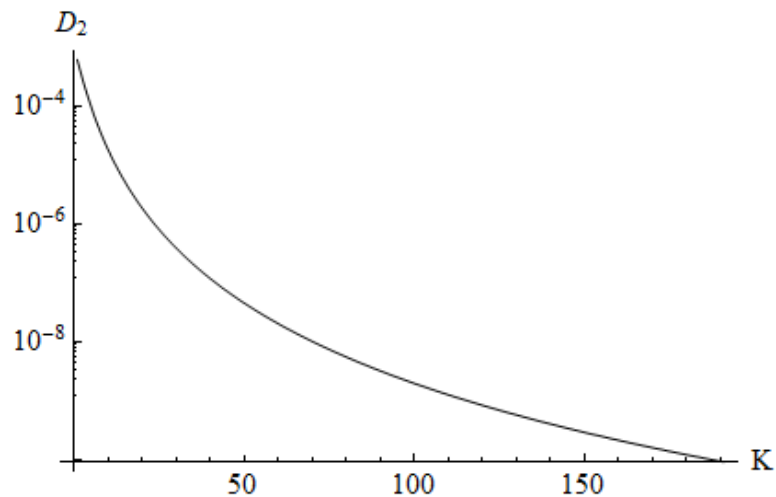


Рис. 7. Зависимость дисперсии оценок коэффициента  $a_2$  для модели второго порядка от объема наблюдения

На зависимостях, представленных на рис. 5 – рис. 7, прослеживается разная скорость спадания дисперсий оценок различных коэффициентов с увеличением числа наблюдений  $K$ . Так для коэффициентов нулевого порядка  $M[\hat{a}_0 - a_0]^2 \sim K^{-1}$ ; для коэффициентов первого порядка  $M[\hat{a}_1 - a_1]^2 \sim K^{-3}$ ; для коэффициентов второго порядка  $M[\hat{a}_2 - a_2]^2 \sim K^{-5}$ .

Исследования эффективности оценок коэффициентов регрессии показали следующее: 1) с ростом  $K$  квазиоптимальные оценки быстро сходятся к оптимальным оценкам ОМНК; 2) при правильном выборе параметра окна  $\alpha$  оценки ВМНК эффективнее соответствующих оценок МНК. Эти выводы проиллюстрированы на рис. 8, где значение параметра весового окна  $\alpha = 1$  (короткий пунктир соответствует квазиоптимальным оценкам, длинный пунктир – ВМНК, сплошная кривая соответствует оценкам МНК).

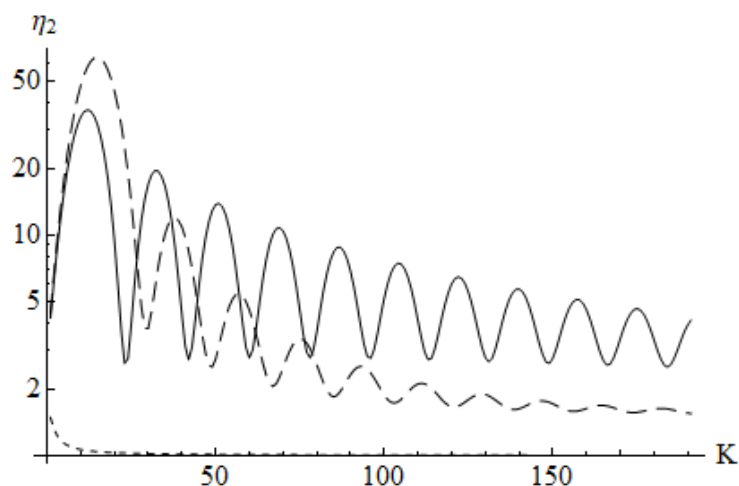


Рис. 8. Зависимость показателя эффективности  $\eta_2$  от объема наблюдения

Отметим, что зависимости показателей  $\eta_0$  и  $\eta_1$  от  $K$  подобны зависимостям, приведенным на рис. 8, как для полиномиальных моделей первого, так и второго порядков.

Как видно из рис. 8 дисперсии оценок МНК и ОМНК коэффициентов полиномиальной регрессии не являются монотонно убывающими функциями числа наблюдаемых отсчетов  $K$ .

### **Регрессионный анализ на фоне аддитивной помехи в виде обобщенного процесса авторегрессии**

В данном случае помеха является процессом  $ARIMA(p; v; 0)$ , который кратко обозначается также как процесс  $ARI(p; v)$  и описывается уравнением

$$\nabla^v \xi_k = \sum_{i=1}^p \phi_i \nabla^v \xi_{k-i} + \eta_k, \quad (36)$$

где  $\eta_k, k \in Z$  – стационарная некоррелированная последовательность с дисперсией  $\sigma^2$ .

С формальной точки зрения модель  $ARI(p; v)$  является частным случаем процесса авторегрессии порядка  $p + v$ . Поэтому можно считать, что процесс  $ARI(p; v)$  описывается уравнением авторегрессии

$$\xi_k = \sum_{i=1}^{p+v} \phi_i \xi_{k-i} + \eta_k. \quad (37)$$

При оценивании коэффициентов полиномиальной регрессии на фоне процесса  $ARIMA(p; v)$  можно получить корректные (состоятельные) линейные оценки лишь коэффициентов  $a_i$ , с индексами  $i \geq v$ . Эти оценки полностью определяются конечными разностями назад  $\nabla^v z_j, j = \overline{v, K-1}$  или, что тоже самое, конечными разностями вперед  $\Delta^v z_j, j = \overline{0, K-v-1}$ .

Запишем оценку ОМНК параметров полиномиальной регрессии наблюдаемой на фоне процесса  $ARI(p; v)$ . При этом будем предполагать, что наблюдается последовательность

$$z_k = \sum_{r=0}^n a_r k^r + \xi_k, \quad k = \overline{0, K-1}, \quad (38)$$

где  $\xi_k$  – процесс  $ARI(p; v)$ , и требуется оценить коэффициенты  $a_r, r = \overline{v, n}$ .

После взятия конечной разности порядка  $v$  получим уравнение наблюдения

$$\Delta^v z_k = \sum_{r=0}^n a_r \Delta^v k^r + \Delta^v \zeta_k, \quad k = \overline{0, K-v-1}. \quad (39)$$

В рамках сформулированной постановки оценку ОМНК параметров полиномиальной регрессии представим в виде

$$\bar{a}^* = [(E^{tr} E)^{-1} E^{tr} \bar{y}], \quad (40)$$

где координатами вектора  $\vec{a}^*$  являются оценки коэффициентов при старших степенных функциях

$$\vec{a}^* = (a_v^*, \dots, a_n^*);$$

вектор наблюдений

$$\vec{y} = B\vec{z}^*, \quad \vec{z}^* = (\Delta^v z_0, \dots, \Delta^v z_{K-v-1});$$

регрессионная матрица

$$E = BX,$$

где матрица  $X = [x_{kj}]$  размерности  $(K - v) \times (n + 1 - v)$  имеет элементы

$$x_{kj} = \Delta^v k^{j+v};$$

матрица  $B$  определяется значениями коэффициентов AP-процесса и имеет структуру вида (24).

Таким образом, после взятия конечной разности процесс  $ARI(p; v)$  превращается в процесс AP порядка  $p$  и его анализ ничем не отличается от анализа последнего. Поэтому для оценивания параметров полиномиальных функций, наблюдаемых на фоне процессов  $ARI(p; v)$ , может использоваться алгоритм МНК, ВМНК или ОМНК, построенный в пространстве конечных разностей от наблюдаемого ряда данных.

### **Исследование точности оценок коэффициентов квадратичной регрессии на фоне процесса ARFIMA(0,d,0)**

Исследование точности оценивания полиномиальных коэффициентов производилось с использованием выражений для дисперсионных матриц оценок МНК (33), ВМНК (10) и ОМНК (32). При этом использовался полиномиальный базис на дискретной сетке  $\{0, 1, \dots, K - 1\}$ . Отсчеты корреляционной функции процесса  $ARFIMA(0; d; 0)$  находились согласно (17).

На рис. 9 - рис. 11 показаны зависимости дисперсий оценок ОМНК коэффициентов полиномиальной регрессии второго порядка от объема наблюдения. Здесь штрих пунктирная кривая соответствует случаю помехи  $ARFIMA(0; -0,4; 0)$ ; пунктирная кривая - случаю помехи  $ARFIMA(0; 0; 0)$ ; сплошная кривая - случаю помехи  $ARFIMA(0; 0,4; 0)$ .

Из анализа результатов исследований следует, что для модели белого шума ( $d = 0$ ) оценки МНК и ОМНК совпадают и с ростом числа наблюдений  $K$  их дисперсии уменьшаются в соответствии с выражением

$$M[(\hat{a}_i - a_i)^2] \sim K^{-(2i+1)}.$$

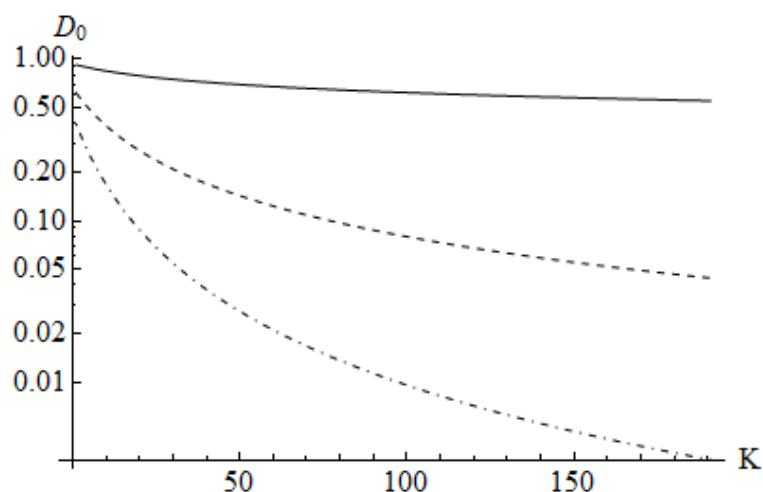


Рис. 9. Зависимости дисперсий оценок коэффициента  $a_0$  для модели второго порядка от объема наблюдения

Для модели персистентных процессов ( $d > 0$ ) скорость уменьшения дисперсий несколько меньше, а для модели антиперсистентных процессов ( $d < 0$ ) – несколько выше, чем для белого шума.

В результате исследований было также установлено, что при  $d > 0$  метод МНК обеспечивает дисперсии оценок коэффициентов регрессии достаточно близкие к оптимальным оценкам ОМНК. Дисперсии же оценок ВМНК могут быть как больше соответствующих дисперсий оценок МНК, так и меньше. При  $d > 0$  можно подобрать параметр весового окна  $\alpha < 0$  так, что оценки ВМНК будут эффективнее оценок МНК (в смысле критерия минимума дисперсий оценок коэффициентов полиномов).

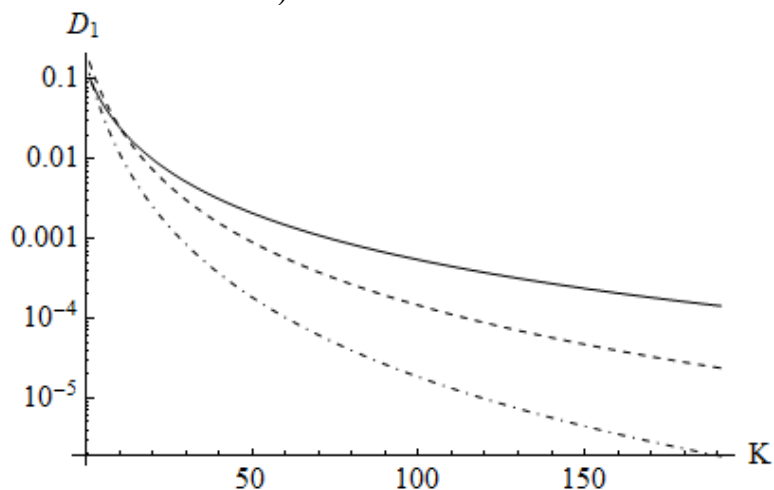


Рис. 10. Зависимости дисперсий оценок коэффициента  $a_1$  для модели второго порядка от объема наблюдения

При  $d < 0$  метод МНК значительно сильнее проигрывает методу ОМНК по дисперсиям оценок коэффициентов регрессии, чем при  $d > 0$ . В этом случае также можно подобрать параметр весового окна (35)  $\alpha > 0$  так, что оценки ВМНК будут эффективнее оценок МНК.

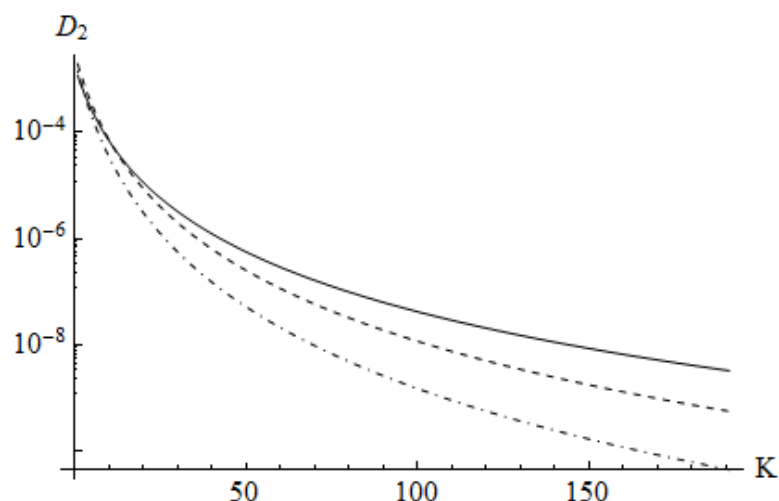


Рис. 11. Зависимости дисперсий оценок коэффициента  $a_2$  для модели второго порядка от объема наблюдения

Для полиномиальной модели второго порядка данные выводы проиллюстрированы на рис. 12 и рис. 13, где сплошные кривые соответствуют методу МНК, а пунктирные - методу ВМНК с параметром окна  $\alpha = -0,5$  для модели помехи ARFIMA(0;0,4;0) и  $\alpha = 0,5$  для модели помехи ARFIMA(0;-0,4;0), т.е. при выполнении условия  $\text{sign}[\alpha] = -\text{sign}[d]$ .

В результате анализа результатов исследований сделан вывод, что для персистентного процесса ARFIMA применение ОМНК обеспечивает незначительный выигрыш по дисперсиям оценок коэффициентов полинома. Для антиперсистентного процесса применение ОМНК может дать значимый выигрыш.

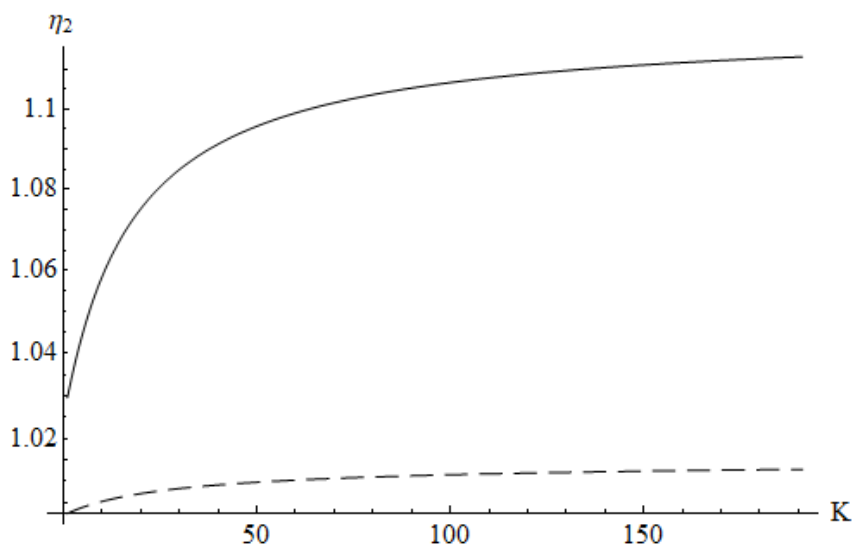


Рис. 12. Зависимость показателя эффективности  $\eta_2$  от объема наблюдения для модели помехи ARFIMA(0;0,4;0)

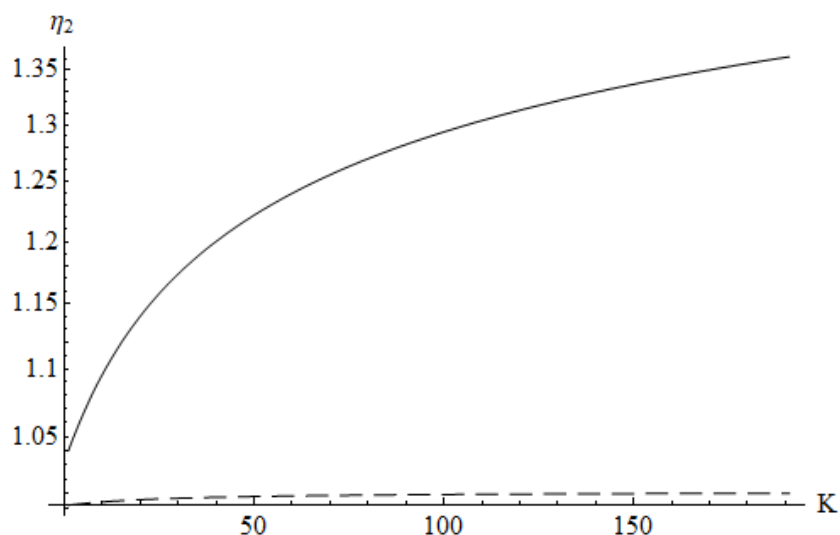


Рис. 13. Зависимость показателя эффективности  $\eta_2$  от объема наблюдения для модели помехи  $ARFIMA(0; -0,4; 0)$

### Выводы

В работе выполнен сравнительный анализ оценок МНК, ВМНК и ОМНК коэффициентов полиномиальной регрессии при воздействии различных помех.

Обоснованы оптимальные оценки коэффициентов линейной регрессии для помех в виде процессов авторегрессии.

Впервые исследованы оценки ОМНК коэффициентов полиномиальной регрессии при воздействии помех в виде процессов  $ARI(p; v)$  и  $ARFIMA(0; d; 0)$ .

В результате исследований установлено следующее:

1. Если помехи описываются процессами авторегрессии первого или второго порядков и существует возможность оценить их параметры, то следует использовать оценки ОМНК коэффициентов полиномиальной регрессии, которые во многих случаях являются значительно эффективнее соответствующих оценок МНК. Если же оценить параметры модели помехи нельзя, то можно использовать оценки ВМНК коэффициентов полиномиальной регрессии, которые при правильном выборе весового окна обеспечивают значительный выигрыш в точности оценивания. Для этой цели могут использоваться весовые окна ортогональных многочленов Хана.

2. В случае коррелированных ошибок наблюдения, дисперсии оценок МНК и ВМНК коэффициентов полиномиальной регрессии не являются монотонно убывающими функциями числа наблюдаемых отсчетов.

3. При оценивании коэффициентов полиномиальной регрессии на фоне помех в виде процесса  $ARI$  порядка  $v$  следует пользоваться конечными разностями от последовательности наблюдаемых данных того же порядка  $v$ . После этого может использоваться любой из алгоритмов оценивания коэффициентов полиномиальной регрессии, построенный в пространстве конечных разностей от наблюдаемой последовательности данных.

4. При воздействии помехи в виде антиперсистентного процесса ARFIMA применение оценок ОМНК обеспечивает больший выигрыш по сравнению с оценками МНК, чем при воздействии помехи в виде персистентного процесса ARFIMA.

#### Литература

1. Дрейпер Н. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. – М.: Издательский дом “Вильямс”, 2007. – 912 с.
2. Большаков А.А. Методы обработки многомерных данных и временных рядов / А.А. Большаков, Р.Н. Каримов. – М.: Горячая линия. – Телеком, 2007. – 522с.
3. Себер Дж. Линейный регрессионный анализ / Себер Дж.; [пер. с англ. В.П. Носко под ред. М.Б. Малютова] – М.: Мир. – 1980. – 456 с.
4. Демиденко Е.З. Линейная и нелинейная регрессии / Демиденко Е.З. – М.: Изд. «Финансы и статистика». – 1981. - 302 с.
5. Андерсон, Т. Статистический анализ временных рядов / Андерсон Т. – М.: Мир, 1976. – 755 с.
6. Бокс Дж., Анализ временных рядов. Прогноз и управление. Выпуск 1 / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1974. – 406 с.
7. Омельченко А.В. Весовые функции полиномиального регрессионного анализа / А.В. Омельченко, А.В. Федоров // Вісник Харк. нац. ун-ту. Сер. «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління». – 2008. – № 833, вип. 10. – С. 193–205.
8. Петерс Э. Фрактальный анализ финансовых рынков: Применение теории Хаоса в инвестициях и экономике / Петерс Э. -М.: Интернет-трейдинг, 2004. – 304 с.
9. Шарапудинов И.И. Многочлены, ортогональные на сетках. Теория и приложения. / Шарапудинов И.И. - Махачкала: ДГПУ, 1997. – 315 с.