

УДК 519.2

МОДЕЛИ ВРЕМЕННЫХ РЕГРЕССИОННЫХ ЗАВИСИМОСТЕЙ И ИХ ПРИЛОЖЕНИЕ К АНАЛИЗУ УРОВНЯ ЖИЗНИ НАСЕЛЕНИЯ

В. К. Клочко, д.т.н., профессор, профессор кафедры АИТУ РГРТУ, Рязань, Россия;
orcid.org/0000-0003-2550-999X, e-mail: klochkovk@mail.ru

В. П. Кузнецов, к.т.н., доцент, доцент кафедры АИТУ РГРТУ, Рязань, Россия;
orcid.org/0000-0001-6189-9011, e-mail: kuznetsovaitu@yandex.ru

В традиционном статистическом анализе построение регрессионных моделей предполагает их стационарность. При этом выборка данных для построения моделей не отражает времени получения данных. Цель работы – расширение возможностей применения регрессионных моделей при прогнозировании за счет включения в них времени. Кроме того, получение выборки измерений в разные моменты времени увеличивает общий ее объем, что позволяет решать проблему малых выборок при построении регрессионных моделей. В работе предлагаются два подхода к построению временных регрессионных моделей. При первом подходе время включается в систему уравнений регрессии вместе с данными, полученными в разные моменты времени. При втором подходе временной обработке подлежат коэффициенты регрессии, найденные независимо в разные моменты времени. В качестве приложения получены линейные регрессионные модели зависимости уровня бедности от среднего дохода населения регионов РФ и от номера года. Данные модели показывают более равномерное изменение регрессионных зависимостей по годам по сравнению с традиционной моделью. Результаты могут найти применение в расчетах Росстата.

Ключевые слова: уравнения регрессии, пространственные зависимости, прогнозирование, уровень бедности, доходы населения.

DOI: 10.21667/1995-4565-2023-83-112-119

Введение

В одномерном статистическом анализе [1] регрессионные модели устанавливают связь между одной зависимой и несколькими независимыми переменными, а также в многомерном статистическом анализе [2] – между векторными переменными. При этом степень влияния той или иной независимой переменной на зависимую устанавливается на основе корреляционного анализа [3]. Такие модели полезны для исследования влияния одних факторов на другие, особенно в тех случаях, когда прямой анализ влияния факторов невозможен из-за малого объема выборки измерений.

Вместе с тем традиционные регрессионные модели имеют недостаток – они являются статическими в силу отсутствия в этих моделях времени. В работе предлагается включить в регрессионные модели время, что позволяет увеличить объем выборки и за счет этого решить проблему малых выборок, а также дает возможность прогнозирования моделей. Прогнозирование регрессионных зависимостей позволяет объективно планировать ту или иную деятельность и правильно принимать решения. В данной работе в качестве иллюстрации применимости предлагаемых моделей проводится анализ доходов населения по данным Росстата в продолжение исследования [4].

Цель работы – расширение возможностей применения регрессионных моделей при прогнозировании за счет включения в них времени, а также решение проблемы малых выборок путем увеличения объема данных.

Временные регрессионные модели

Простейшая однофакторная линейная регрессионная зависимость y от x имеет вид

$$y = a_0 + a_1 x, \quad (1)$$

где параметры a_0 и a_1 находятся методом наименьших квадратов (МНК) на основе выборки измерений (x_i, y_i) , $i = \overline{1, n}$.

Двухфакторная линейная зависимость z от x и y будет

$$z = a_0 + a_1 x + a_2 y, \quad (2)$$

где параметры a_0, a_1, a_2 находятся на основе выборки (x_i, y_i, z_i) , $i = \overline{1, n}$.

Первый подход

Введем в моделях (1) и (2) линейную зависимость параметров моделей от дискретного времени t_k , $k = \overline{1, m}$, или, что удобней, от целочисленного номера k проведения измерений:

$$y = a_{01} + a_{02}k + (a_{11} + a_{12}k)x, \quad (3)$$

$$z = a_{01} + a_{02}k + (a_{11} + a_{12}k)x + (a_{21} + a_{22}k)y, \quad k = \overline{1, m}. \quad (4)$$

Соответствующие модели измерений имеют вид

$$y_{ki} = y + e_{ki}, \quad z_{ki} = z + e_{ki}, \quad i = \overline{1, n}, \quad k = \overline{1, m}, \quad (5)$$

где e_{ki} – ошибки несоответствия моделей измерениям.

Для нахождения параметров моделей (3), (4) на основе выборок (x_{ki}, y_{ki}) и (x_{ki}, y_{ki}, z_{ki}) запишем (3) – (5) в матричной форме

$$Y = FA + E \quad \text{или} \quad Z = FA + E, \quad (6)$$

где

$$Y = (y_{11} \dots y_{1n} \ y_{21} \dots y_{2n} \ \dots \ y_{m1} \dots y_{mn})^T, \quad Z = (z_{11} \dots z_{1n} \ z_{21} \dots z_{2n} \ \dots \ z_{m1} \dots z_{mn})^T, \\ E = (e_{11} \dots e_{1n} \ e_{21} \dots e_{2n} \ \dots \ e_{m1} \dots e_{mn})^T,$$

T – символ транспонирования,

$$A = (a_{01} \ a_{02} \ a_{11} \ a_{12})^T \quad \text{или} \quad A = (a_{01} \ a_{02} \ a_{11} \ a_{12} \ a_{21} \ a_{22})^T,$$

$$F = (f_{11}^T \dots f_{1n}^T \ f_{21}^T \dots f_{2n}^T \ \dots \ f_{m1}^T \dots f_{mn}^T)^T,$$

где f_{ki}^T – строка матрицы F :

$$f_{ki}^T = (1 \ k \ x_{ki} \ k \cdot x_{ki}) \quad \text{или} \quad f_{ki}^T = (1 \ k \ x_{ki} \ k \cdot x_{ki} \ y_{ki} \ k \cdot y_{ki}).$$

Критерием МНК нахождения оценок параметров является минимум квадрата евклидовой нормы $J = \|W \cdot E\|^2$ $m \times n$ -вектора E ошибок e_{ki} , взятых с положительными весовыми коэффициентами w_{ki} , определяющими степень доверия k -м, i -м наблюдениям (x_{ki}, y_{ki}) или (x_{ki}, y_{ki}, z_{ki}) , где $W = \text{diag}(w_{11} \dots w_{1n} \ w_{21} \dots w_{2n} \ \dots \ w_{m1} \dots w_{mn})$ – диагональная $m \times m$ -матрица весовых коэффициентов.

Так, для однофакторной модели критерий в развернутом виде запишется как

$$J = \sum_{k=1}^m \sum_{i=1}^n w_{ki} [y_{ki} - (a_{01} + a_{02}k) - (a_{11} + a_{12}k)x_{ki}]^2 \rightarrow \min_A,$$

или в матричной форме

$$J = (Y - FA)^T W (Y - FA) \rightarrow \min_A, \quad (7)$$

справедливой как для Y , так и для Z моделей (6).

Минимизация (7) приводит к МНК-оценкам параметров

$$\hat{A} = (F^T W F)^{-1} F^T W \cdot Y \quad \text{или} \quad \hat{A} = (F^T W F)^{-1} F^T W \cdot Z, \quad (8)$$

$\hat{A} = (\hat{a}_{01} \hat{a}_{02} \hat{a}_{11} \hat{a}_{12})^T$ или $\hat{A} = (\hat{a}_{01} \hat{a}_{02} \hat{a}_{11} \hat{a}_{12} \hat{a}_{21} \hat{a}_{22})^T$ – 4×1 - и 6×1 -векторы оценок параметров.

С учетом найденных по формулам (8) оценок параметров уравнения регрессии для k -х моментов времени записываются как

$$\hat{y} = \hat{a}_{01} + \hat{a}_{02}k + (\hat{a}_{11} + \hat{a}_{12}k)x, \quad \hat{z} = \hat{a}_{01} + \hat{a}_{02}k + (\hat{a}_{11} + \hat{a}_{12}k)x + (\hat{a}_{21} + \hat{a}_{22}k)y, \quad k = \overline{1, m}.$$

Для нелинейной модели, например $y = a_0 + a_1x + a_2y + a_3xy$, меняется размерность строки матрицы F :

$$f_{ki}^T = (1 \quad k \quad x_{ki} \quad k \cdot x_{ki} \quad y_{ki} \quad k \cdot y_{ki} \quad x_{ki}y_{ki} \quad k \cdot x_{ki}y_{ki}),$$

а также вектора параметров $A = (a_{01} a_{02} a_{11} a_{12} a_{21} a_{22} a_{31} a_{32})^T$. Соответствующее уравнение регрессии с оценками параметров будет

$$\hat{z} = \hat{a}_{01} + \hat{a}_{02}k + (\hat{a}_{11} + \hat{a}_{12}k)x + (\hat{a}_{21} + \hat{a}_{22}k)y + (\hat{a}_{31} + \hat{a}_{32}k)xy, \quad k = \overline{1, m}.$$

Прогноз зависимостей (3), (4) на N -й момент времени ($N > m$) определяется формулами

$$\hat{y} = \hat{a}_{01} + \hat{a}_{02}N + (\hat{a}_{11} + \hat{a}_{12}N)x, \quad \hat{z} = \hat{a}_{01} + \hat{a}_{02}N + (\hat{a}_{11} + \hat{a}_{12}N)x + (\hat{a}_{21} + \hat{a}_{22}N)y.$$

Второй подход

Второй подход основан на построении уравнений регрессии независимо в каждый k -й фиксированный момент времени по обычной методике:

$$\hat{y} = \hat{a}_0(k) + \hat{a}_1(k)x \quad \text{или} \quad \hat{z} = \hat{a}_0(k) + \hat{a}_1(k)x + \hat{a}_2(k)y$$

и запоминании коэффициентов.

Затем найденные оценки параметров $\hat{a}_0(k), \hat{a}_1(k), \hat{a}_2(k)$, $k = \overline{1, m}$, используются в качестве выборок. На основе m выборок оценок параметров методом МНК находятся тренды – линейные зависимости вида

$$a_0^*(k) = b_{00} + b_{01} \cdot k, \quad a_1^*(k) = b_{10} + b_{11} \cdot k, \quad a_2^*(k) = b_{20} + b_{21} \cdot k,$$

где b_{ij} – параметры трендов (МНК-оценки).

Уравнения регрессии принимают вид

$$y^* = a_0^*(k) + a_1^*(k) \cdot x, \quad z^* = a_0^*(k) + a_1^*(k) \cdot x + a_2^*(k) \cdot y, \quad k = \overline{1, m},$$

и используются для разных k -х моментов времени, в том числе при прогнозировании ($k > m$).

Показатели точности моделей и доверительный интервал

Одним из показателей точности модели является среднее абсолютное относительное отклонение измерений от модельных значений [5], вычисляемое для (1) и (2) в каждый k -й момент времени по формуле

$$\varepsilon_k = \frac{1}{n} \sum_{i=1}^n \frac{|y_{ki} - \hat{y}_{ki}|}{|y_{ki}|} \cdot 100\% \quad \text{или} \quad \varepsilon_k = \frac{1}{n} \sum_{i=1}^n \frac{|z_{ki} - \hat{z}_{ki}|}{|z_{ki}|} \cdot 100\%, \quad (9)$$

где

$$\hat{y}_{ki} = \hat{a}_{01} + \hat{a}_{02}k + (\hat{a}_{11} + \hat{a}_{12}k)x_{ki} \quad \text{или} \quad \hat{z}_{ki} = \hat{a}_{01} + \hat{a}_{02}k + (\hat{a}_{11} + \hat{a}_{12}k)x_{ki} + (\hat{a}_{21} + \hat{a}_{22}k)y_{ki}.$$

Усредненный по k показатель точности $\varepsilon_{cp} = \frac{1}{m} \sum_{k=1}^m \varepsilon_k$.

Другим показателем точности является выборочное среднеквадратическое отклонение (СКО) измерений относительно модельных значений:

$$s_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{ki} - \hat{y}_{ki})^2} \quad \text{или} \quad s_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_{ki} - \hat{z}_{ki})^2}, \quad (10)$$

а также усредненное по k выборочное СКО $s_{cp} = \frac{1}{m} \sum_{k=1}^m s_k$.

Доверительный интервал для измерений берется относительно модельных значений и находится для каждого k -го момента времени ($k = \overline{1, m}$):

$$\hat{y} \pm \gamma \cdot s_{cp} \text{ или } \hat{z} \pm \gamma \cdot s_{cp}, \quad (11)$$

где коэффициент γ определяется правилом нескольких сигма, например двух-трех, или с учетом более строго критерия Стьюдента для заданной доверительной вероятности [5].

При расчете доверительного интервала неопределенность возрастает с увеличением момента N времени прогноза ($N > m$), что выражается увеличением коэффициента γ в (11). При построении доверительного интервала для будущего измерения к дисперсии ошибки прогноза D_N прибавляется дисперсия ошибки измерения (в предположении независимости ошибок). Величину D_N можно рассчитать [5] на основе линейной модели временного ряда

$$y = b_0 + b_1 k \text{ или } z = b_0 + b_1 k, \quad k = \overline{1, m},$$

где b_0 и b_1 – МНК-коэффициенты, по формуле

$$D_N = s_{cp}^2 \cdot h_N^T (H_m^T H_m)^{-1} h_N, \quad N > m, \quad (12)$$

где s_{cp}^2 – выборочная усредненная дисперсия ошибок измерения, $H_m = (h_1^T h_2^T \dots h_m^T)^T$ – $m \times 2$ -матрица, k -я строка которой $h_k^T = (1 \ k)$ и $h_N^T = (1 \ N)$.

Тогда доверительный интервал прогноза (11) с учетом (12) будет

$$\hat{y} \pm \gamma \cdot s_{cp} \sqrt{1 + h_N^T (H_m^T H_m)^{-1} h_N}, \quad \hat{z} \pm \gamma \cdot s_{cp} \sqrt{1 + h_N^T (H_m^T H_m)^{-1} h_N}.$$

Приложение к анализу доходов населения

Одной из проблем статистических расчетов доходов населения как в России, так и за рубежом является проблема малых выборок при оценке уровня бедности или уровня жизни по данным опроса доходов населения в малочисленных районах и областях. Расчет по малым выборкам приводит к большим ошибкам оценок уровня жизни. Об актуальности решения этой проблемы свидетельствуют обзор [6] многих публикации на данную тему, а также данные семинара, проведенного Всемирным Банком онлайн в Росстате весной 2021 года. Анализ публикаций в [6] показывает, что основой ряда предлагаемых моделей является статическая однофакторная регрессионная модель со случайной аддитивной составляющей, статистические характеристики которой определяются с учетом соседних областей. При этом недостающие данные выборки генерируются датчиком псевдослучайных чисел.

На наш взгляд имитация недостающих выборочных данных псевдослучайными числами не может привести к заметному снижению неопределенности при решении проблемы малых выборок. Предлагается решать проблему малых выборок за счет включения в общую выборку данных, полученных за несколько лет. Математически это сводится к построению временных регрессионных моделей.

В таблице 1 представлены данные о среднедушевых доходах (в рублях) и доле бедных (в процентах) по 16 субъектам ЦФО РФ за 2016 – 2020 годы, взятые на сайте Росстата [7].

Данные за 2016 – 2019 годы использованы для нахождения однофакторной пространственно-временной зависимости уровня бедности (y) от среднедушевого дохода (x), а данные за 2020 год оставлены для тестирования полученной модели.

Рассмотрим первый подход, использующий векторно-матричную обработку всех исходных данных за 2016 – 2019 годы. В результате обработки получено уравнение регрессии

$$\hat{y} = \hat{a}_{01} + \hat{a}_{02}k + (\hat{a}_{11} + \hat{a}_{12}k)x. \quad (13)$$

Здесь $\hat{a}_{01} = 40,0$; $\hat{a}_{02} = -2,19$; $\hat{a}_{11} = -0,0011$; $\hat{a}_{12} = 0,0001$. Матрица весов W была назначена единичной. При $k = \overline{1, 4}$ из уравнения (13) формируются уравнения регрессии для 2016 – 2019 годов; при $k = 5$ из уравнения (13) выводится прогнозное уравнение регрессии для 2020 года. Оно имеет вид:

$$\hat{y} = 29,05 - 0,00060x. \quad (14)$$

Таблица 1 – Среднедушевой доход и доля бедных в субъектах ЦФО

Table 1 – Average per capita income and share of the poor in the subjects of Central Federal

District

Субъек- ты ЦФО (обла- сти)	2016		2017		2018		2019		2020	
	Сред- ний доход	Доля бед- ных	Сред- ний доход	Доля бед- ных	Сред- ний доход	Доля бед- ных	Сред- ний доход	Доля бед- ных	Сред- ний доход	Доля бед- ных
Белг.	29 799	8,1	30 342	7,8	30 778	7,5	32 352	7,8	32 835	7,2
Брянс.	24 006	14,4	25 107	14,2	26 585	13,6	28 371	13,8	28 310	13,6
Влад.	22 365	14,6	23 554	13,2	23 539	13,1	25 358	12,6	25 803	12,5
Вор.	29 284	9,4	29 498	9,0	30 289	8,9	32 022	8,9	31 997	8,5
Иван.	23 676	14,9	24 860	14,3	24 503	14,7	25 794	14,2	26 219	13,7
Кал.	28 800	10,3	28 715	10,0	29 129	10,4	31 394	10,2	32 183	9,7
Костр.	23 174	13,7	24 093	13,4	23 716	12,7	25 285	12,6	25 260	12,5
Курск.	25 355	10,5	26 112	10,3	27 275	9,9	29 149	9,9	29 671	9,9
Лип.	27 934	9,2	28 956	8,8	30 010	8,7	32 479	8,6	32 201	8,4
Орл.	23 006	14,2	23 979	13,9	24 895	13,5	26 064	13,6	26 705	12,9
Рязан.	23 799	12,8	24 272	13,3	25 441	13,0	26 886	12,7	27 248	12,8
Смол.	23 614	17,8	24 766	16,9	25 888	16,4	27 388	16,3	27 998	15,6
Тамб.	25 769	10,6	26 058	10,4	26 828	9,8	28 154	10,7	27 771	10,8
Твер.	24 063	12,8	24 353	12,4	25 125	12,2	27 211	11,7	27 530	11,4
Тульс.	26 815	10,0	27 226	9,9	27 208	10,0	28 557	10,3	29 208	10,1
Ярос.	26 768	11,0	27 200	10,7	27 055	10,2	28 658	10,3	29 475	9,9

На рисунке 1 показано наложение прогнозного уравнения регрессии 1-го подхода на корреляционное поле, построенное по данным 2020 года, а также построен доверительный интервал (95 %) для выходной переменной y (уровня бедности).

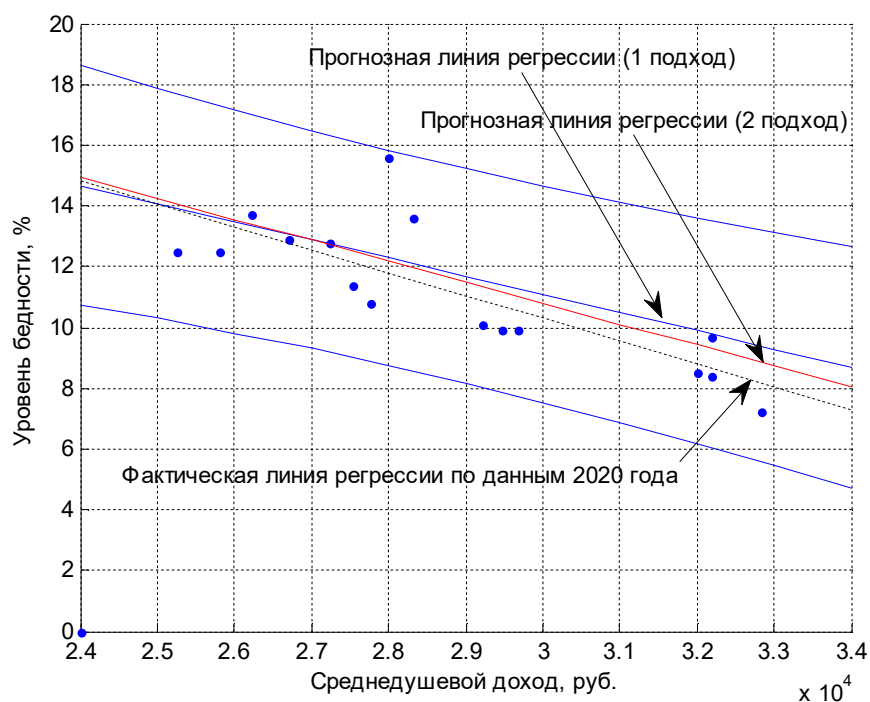


Рисунок 1 – Результаты тестирования

Figure 1 – Test results

Результаты тестирования можно признать убедительными, так как все точки корреляционного поля оказались внутри доверительного интервала.

Данные за 2020 год были подвергнуты самостоятельному регрессионному анализу; получено следующее уравнение регрессии:

$$\hat{y} = 32,90 - 0,000753x. \quad (15)$$

Соответствующая уравнению (15) линия регрессии показана на рисунке 1 (фактическая линия регрессии).

Точность прогнозирования оценим сначала с помощью среднего относительного (по модулю) отклонения исходных данных от модельных значений (9). Для прогнозного уравнения регрессии (14) этот показатель оказался равным 11,83 %, для фактического уравнения регрессии (15) – 8,53 %.

Точность прогнозирования оценим также с помощью СКО исходных данных от модельных значений (10). Для прогнозного уравнения регрессии (14) СКО оказалось равным 1,49; для фактического уравнения регрессии (15) СКО приняло значение 1,31.

Рассмотрим второй подход, использующий декомпозицию исходных данных во времени. На основании исходных данных 2016-2019 годов находятся ежегодные уравнения регрессии:

$$2016 \text{ год: } \hat{y} = 36,063 - 0,000937x; \quad 2017 \text{ год: } \hat{y} = 37,166 - 0,000970x;$$

$$2018 \text{ год: } \hat{y} = 34,133 - 0,000840x; \quad 2019 \text{ год: } \hat{y} = 32,454 - 0,000753x.$$

В качестве примера на рисунке 2 приведены результаты регрессионного анализа для 2016 года, полученные с помощью пакета Excel.

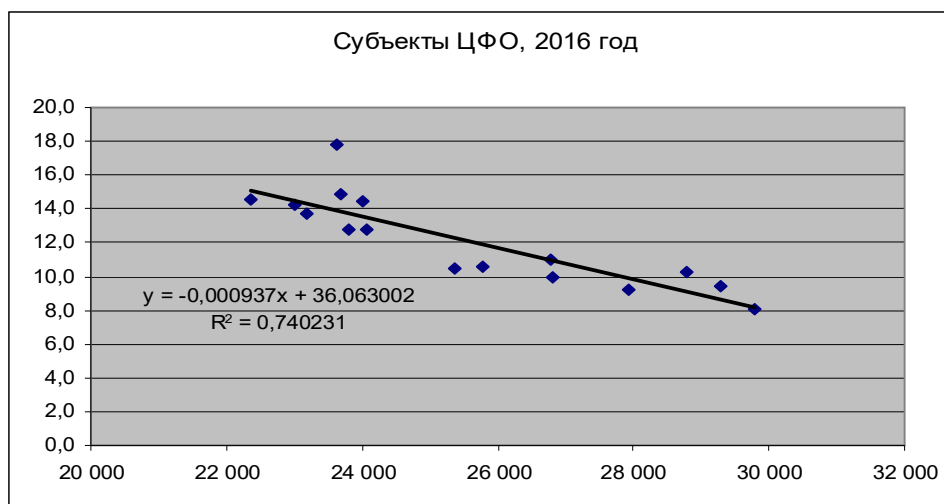


Рисунок 2 – Регрессионный анализ по итогам 2016 года
Figure 2 – Regression analysis for 2016

Из коэффициентов уравнений регрессии формируются временные ряды с линейными трендами. В качестве примера на рисунке 3 приведен временной ряд и линейный тренд для коэффициента a_0 .

Для коэффициента a_1 линейный тренд имеет вид:

$$\hat{a}_1 = -0,00105 + 0,00008k, \quad k = \overline{1, 4}.$$

Для $k = 5$ по уравнениям трендов выполняется прогноз коэффициентов $\hat{a}_0$, $\hat{a}_1$ на год вперед и формируется прогнозируемое уравнение регрессии для 2020 года:

$$\hat{y} = 31,490 - 0,000690x.$$

Соответствующая линия регрессии для 2-го подхода приведена на рисунке 1 (прогнозная линия регрессии). Визуальный анализ показывает близкое совпадение результатов прогноза первого и второго подходов. При этом прогнозная линия регрессии второго подхода располагается ближе к фактической линии регрессии по сравнению с первым подходом. Такой же

вывод следует из количественных оценок точности прогнозирования. Так, среднее относительное (по модулю) отклонение исходных данных от модельных значений оказалось равным 10,58 %, а СКО исходных данных от модельных значений – 1,43. Однако второй подход сопровождается большим объемом вычислений из-за ежегодной процедуры регрессионного анализа и трендовой обработки коэффициентов уравнений регрессии.

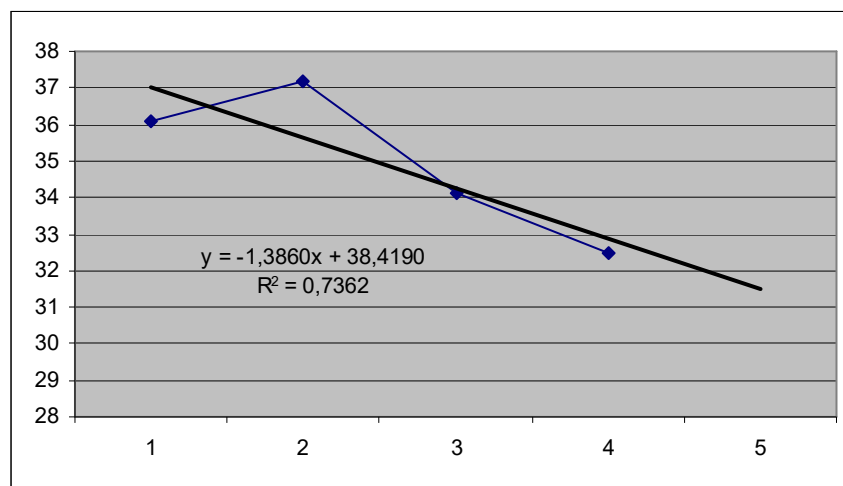


Рисунок 3 – Временной ряд и линейный тренд для коэффициента a_0

Figure 3 – Time Series and Linear Trend for Coefficient a_0

Выводы

Предложены два подхода, позволяющие расширить возможности построения уравнений регрессии за счет временной обработки данных. При первом подходе время включается в систему уравнений регрессии вместе с данными, полученными в разные моменты времени. Параметры регрессионной модели находятся матричным методом наименьших квадратов с весами по совокупности всех данных. При втором подходе временной обработке подлежат коэффициенты регрессии, найденные в разные моменты времени независимо, и эти коэффициенты становятся функцией времени.

Подходы апробированы на официальных данных среднедушевого дохода населения 16 регионов ЦФО РФ. Сравнение показало, что предложенные подходы для линейных моделей получают эквивалентными и в сравнении с традиционным подходом к расчету уравнений регрессии дают равномерное распределение регрессионных зависимостей по годам с возможностью их прогнозирования. Данные подходы можно рекомендовать Росстату для расчета уровня бедности как в регионах РФ, а так и в муниципальных образованиях с малым объемом выборки. Перспектива исследований связана с построением многофакторных линейных и нелинейных временных регрессионных моделей и их внедрением в практику расчетов дохода населения регионов и муниципальных образований РФ.

Библиографический список

1. Эконометрика: учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева и др.; Под ред. И.И. Елисеевой. М.: Финансы и статистика, 2005. 576 с.
2. Чураков Е. П. Введение в многомерные статистические методы: учебное пособие. СПб.: Издательство «Лань», 2016. 148 с.
3. Елисеева И. И., Юзбашев М. М. Общая теория статистики: учебник / под ред. И. И. Елисеевой. М.: Финансы и статистика, 2003. 480 с.
4. Ключко В. К., Кузнецов В. П. Математические модели повышения точности расчета показателей уровня жизни населения // Вестник Рязанского государственного радиотехнического университета. 2020. № 73. С. 105-115.

5. Модели неопределенности в теории и приложениях: учеб. пособие / **В. К. Ключко**. М.: КУРС. 2022. 204 с.

6. **Molina Isabel, Rao J. N. K., Guadarrama Maria**. Small Area Estimation Methods for Poverty Mapping: A Selective Review // *Statistics and Applications*. 2019, vol. 17, no. 1, pp 11-22.

7. https://rosstat.gov.ru/storage/mediabank/urov_31g.doc.

UDC 512.2

MODELS OF TEMPORAL REGRESSION DEPENDENCIES AND THEIR APPLICATION TO LIVING STANDARD OF POPULATION

V. K. Klochko, Dr.Sc. (Tech.), full professor, Department of automation and information technologies in control, professor of the department, RSREU, Ryazan, Russia;
orcid.org/0000-0003-2550-999X, e-mail: klochkovk@mail.ru

V. P. Kuznetsov, Ph.D. (Tech.), associate professor, Department of automation and information technologies in control, associate professor of the department, Ryazan, Russia;
orcid.org/0000-0000-0000-000X, e-mail: kuznetsovaityu@yandex.ru

*In traditional statistical analysis, the construction of regression models assumes their stationary nature. In this case, the selection of data for building models does not reflect the time of data reception. **The aim of the work** is to expand the possibilities of using regression models in forecasting by including time in them. In addition, obtaining a sample of measurements at different points in time increases its total volume, which allows solving the problem of small samples when building regression models. The paper proposes two approaches to constructing time regression models. In the first approach, time is included in the system of regression equations along with the data obtained at different points in time. In the second approach, the regression coefficients found independently at different points in time are subject to time processing. As an application, linear regression models of the dependence of poverty level on average population income of the regions of Russian Federation and on the year number were obtained. The model data show a more uniform change in regression dependencies over years compared to a traditional model. The results can be used in Rosstat calculations.*

Key words: regression equations, temporal dependencies, forecasting, poverty level, incomes of the population.

DOI: 10.21667/1995-4565-2023-83-112-119

References

1. *Ekonomika: uchebnik* / **I. I. Eliseeva, S. V. Kuryshcheva, T. V. Kosteeva** i dr.; Pod red. I. I. Eliseevoy. Moscow: Finansy i statistika. 2005, 576 p. (in Russian).

2. **Churakov E. P.** *Vvedenie v mnogomernye statisticheskie metody: uchebnoe posobie*. SPb.: Izdatel'stvo «Lan». 2016, 148 p. (in Russian).

3. **Eliseeva I. I., Yuzbashev M. M.** *Obshchaya teoriya statistiki: uchebnik* / pod red. I. I. Eliseevoy. Moscow: Finansy i statistika. 2003, 480 p. (in Russian).

4. **Klochko V. K., Kuznetsov V. P.** Matematicheskie modeli povysheniya tochnosti rascheta pokazatelej urovnya zhizni naseleniya. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2020, no. 73, pp. 105-115. (in Russian).

5. *Modely neopredelennosti v teorii i prilozheniyah*. Uchebnoe posobie. **V. K. Klochko**. Moscow: KURS. 2022, 204 p. (in Russian).

6. **Molina Isabel, Rao J. N. K., Guadarrama Maria**. Small Area Estimation Methods for Poverty Mapping: A Selective Review. *Statistics and Applications*. 2019, vol. 17, no. 1, pp 11-22.

7. https://rosstat.gov.ru/storage/mediabank/urov_31g.doc.