ИСПОЛЬЗОВАНИЕ ТРЕНДОВО-ФАКТОРНОЙ МОДЕЛИ ДЛЯ ПОВЫШЕНИЯ ТОЧНО-СТИ ПРОГНОЗА

УДК 330.43

Ирина Владленовна Орлова,
к.э.н., профессор, профессор каф. Системного анализа и моделирования экономических процессов Финансового университета при Правительстве РФ
Тел.: (499) 277-21-44
Эл. почта: IVOrlova@fa.ru

Виктор Борисович Турундаевский,
к.э.н., доцент, профессор каф. Прикладной математики Московского государственного университета экономики, статистики и информатики (МЭСИ)
Тел.: (495) 442-60-98
Эл. почта: vik_turund@mail.ru

В статье предложен метод повышения точности трендово-факторной модели в предположении, что приращение эндогенной переменной зависит не только от времени, но и от отклонений от своих трендов экзогенных переменных, доказаны соответствующие формулы, отражающие структуру дисперсии остатков трендово-факторной модели.

Ключевые слова: *временные ряды, тренд, множественная регрессия, блочные матрицы, дисперсия остатков*

Irina V. Orlova,
PhD, Professor, Dept. System analysis and modeling of economic processes of the Financial University under the Government of the Russian Federation
Tel.: (499) 277-21-44
E-mail: IVOrlova@fa.ru

Viktor B. Turundaevsky,
PhD, Associate Professor, Dept. Applied Mathematics, Moscow State University of Economics, Statistics and Informatics (MESI)
Tel.: (495) 442-60-98
E-mail: vik_turund@mail.ru

USE OF THE TREND-FACTOR MODEL TO IMPROVE THE ACCURACY FORECASTS

In this paper we propose a method to improve the accuracy of the trend-factor model on the assumption that the increase in endogenous variable-screens depend not only on time but also deviations from their trend of exogenous variables, proved by the corresponding formulas that reflect the structure of the dispersion, these remnants of the trend-factor model.

Keywords: *time series, trend, multiple regression, block matrix, the variance of residuals*

1. В работе получена оценка доверительного интервала при прогнозировании временных рядов по многофакторной регрессионной модели при высокой степени коррелированности определяющих переменных.

Прогнозирование временных рядов $y(t)$ обычно проводится по одномерной модели, поскольку определяющие переменные $x_j(t)$, как и результирующая переменная $y(t)$ обычно растут с увеличением t, в результате чего коэффициенты корреляции между ними становятся высокими, а в этом случае, как известно, определить степень влияния каждого из показателей $x_j(t)$ на $y(t)$ с помощью МНК (метода наименьших квадратов) невозможно. Быстрый рост ошибки прогноза в одномерных регрессионных моделях существенно ограничивает возможности этого метода. Имея временные ряды зависимой переменной $y(t)$ и экзогенных переменных $x_j(t)$, можно добиться значительного сокращения доверительного интервала прогноза, если рассматривать многофакторные регрессионные модели финансово-экономических показателей $y(t)$, построенные по отклонениям экзогенных переменных $x_j(t)$ от своих трендов.

2. Регрессионная трендово-факторная модель имеет вид

$$y(t) = \hat{y}(t) + \gamma_1 dx_1(t) + \ldots + \gamma_n dx_n(t) + \varepsilon(t), \tag{1}$$

где $y(t)$ – фактические уровни временного ряда,
$\hat{y}(t)$ – расчетные значения уровней временного ряда,
$dx_j(t) = x_j(t) - \hat{x}_j(t)$ – отклонения $x_j(t)$ от тренда,
$\varepsilon(t)$ – случайная величина, имеющая нормальное распределение с нулевым средним и не зависящая от t.

Будем считать, что $dx_j(t)$ не зависят от t. Соотношение (1) отражает гипотезу, что отклонения $y(t)$ от тренда $\hat{y}(t)$ объясняются с точностью до $\varepsilon(t)$ отклонениями $x_j(t)$ от своих трендов $\hat{x}_j(t)$.

Отклонения $dx_j(t)$ можно определить с помощью МНК ($j = \overline{1,n}$).

Пусть тренд $\hat{y}(t) = \gamma_{-2}t^2 + \gamma_{-1}t + \gamma_0$ является квадратичной параболой с неизвестными коэффициентами. Тогда уравнение (1) можно записать в виде

$$y(t) = \gamma_{-2}t^2 + \gamma_{-1}t + \gamma_0 + \gamma_1 dx_1(t) + \ldots + \gamma_n dx_n(t) + \varepsilon(t). \tag{2}$$

Коэффициенты уравнения (2) будем оценивать методом наименьших квадратов.

Для прогнозирования $y(t)$ воспользуемся уравнением (2). Значения отклонений $dx_j(t)$ для периода упреждения, как правило, неизвестны. В этом случае, считая $dx_j(t)$ реализациями случайных величин с нулевым средним, получаем

$$M(y(t)) = \gamma_{-2}t^2_L + \gamma_{-1}t_L + \gamma_0, \tag{3}$$

где $t_L = t_N + L$ – время прогноза.

Как видим, прогнозируемое значение $y(t_L)$ равно прогнозу по тренду, однако, ошибка прогноза за счет включения в модель переменных $dx_j(t)$ будет меньше, чем при прогнозировании по одномерному временному ряду.

3. Ниже будет дана количественная оценка повышения точности прогноза.

При оценке погрешности прогноза по одномерному временному ряду доверительный интервал прогноза равен [1]

$$\hat{y}_L \pm t_\alpha \cdot S_p,$$

где $t_\alpha - \alpha$ – процентная точка распределения Стьюдента. Дисперсия ошибок прогноза S_p^2 состоит из двух слагаемых:

$$S_p^2 = S_{\hat{y}}^2 + S_{\varepsilon 1}^2,$$

где $S_{\varepsilon 1}^2$ – дисперсия отклонений от тренда $\varepsilon_1(t) = y(t) - \widehat{y}(t)$,
$S_{\widehat{y}}^2$ – дисперсия ошибки тренда $\widehat{y}(t)$.

Величина $S_{\widehat{y}}^2$ вычисляется по формуле [1]

$$S_{\widehat{y}}^2 = S_{\varepsilon 1}^2 \cdot t_L' \cdot \left(T' \cdot T\right)^{-1} \cdot t_L \tag{4}$$

где

$$T = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \dots & \dots & \dots \\ 1 & t_N & t_N^2 \end{pmatrix}, \; t_L' = \left(1 \; t_L \; t_L^2\right). \tag{5}$$

Если $y(t)$ прогнозировать по уравнению (2), то ошибка прогноза вычисляется аналогично. Дисперсия ошибки прогноза равна $S_p^2 = S_{\widehat{y,x}}^2 + S_\varepsilon^2$ при этом S_ε^2 будет дисперсией отклонений $y(t)$ от уравнения регрессии (2)

$$(\varepsilon(t) = y(t) - \hat{y}(t) - \gamma_1 dx_1(t) + \ldots + \gamma_n dx_n(t)),$$

величина $S_{\widehat{y,x}}^2$ вычисляется по формуле

$$S_{\widehat{y,x}}^2 = S_\varepsilon^2 \cdot \overline{t}_{L,x}{}' \cdot (T_x' \cdot T_x)^{-1} \cdot \overline{t}_{L,x}, \tag{6}$$

где

$$T_x = \begin{pmatrix} 1 \; t_1 \; t_1^2 \; dx_1(t_1) dx_2(t_1) \ldots dx_n(t_1) \\ 1 \; t_2 \; t_2^2 \; dx_1(t_2) dx_2(t_2) \ldots dx_n(t_2) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ 1 \; t_N \; t_N^2 \; dx_1(t_N) dx_2(t_N) \ldots dx_n(t_N) \end{pmatrix} \tag{7}$$

$$\overline{t}_{L,x} = \left(1 \;\; t_L \;\; t_L^2 \;\; dx_1\left(t_L\right) \; \ldots dx_n\left(t_L\right)\right)$$

Введем обозначения

$$D = \begin{pmatrix} dx_1(t_1) & dx_2(t_1) \ldots dx_n(t_1) \\ dx_1(t_2) & dx_2(t_2) \ldots dx_n(t_2) \\ \dots\dots\dots\dots\dots\dots \\ dx_1(t_N) & dx_2(t_N) \ldots dx_n(t_N) \end{pmatrix} \tag{8}$$

$$\overline{d}'_L = (dx_1(t_L) \;\; dx_2(t_L) \; \ldots dx_n(t_L)) \tag{9}$$

Ниже будет показано, что

$$S_{\widehat{y,x}}^2 \approx S_\varepsilon^2 \cdot \overline{t_L}' \cdot \left(T' \cdot T\right)^{-1} \cdot \overline{t}_L + S_\varepsilon^2 \cdot \overline{d}'_L \cdot \left(D' \cdot D\right)^{-1} \cdot \overline{d}_L. \tag{10}$$

Рассмотрим соотношение (10). Согласно предположению $dx_j(t)$ не зависят от $t(j = \overline{1,n})$, следовательно, в формуле (10) от t зависит лишь первое слагаемое, т.е. $S_{\widehat{y,x}}^2$ можно представить в виде $S_{\widehat{y,x}}^2 \approx S_\varepsilon^2(\varphi(t_L) + a)$,

где $a = \overline{d}'_L (D'D)^{-1} \overline{d}_L$, (11)

$\varphi(t_L) = \overline{t}'_L (T'T)^{-1} \overline{t}_L$ – монотонно возрастающая функция t_L.

Дисперсия ошибки прогноза примет вид

$$S_p^2 = S_\varepsilon^2(\varphi(t_L) + a) + S_\varepsilon^2 = S_\varepsilon^2\left(\varphi(t_L) + a + 1\right). \tag{12}$$

Дисперсию ошибки прогноза по одномерному временному ряду, используя обозначения (11), можно записать в виде

$$S_p^2 = S_{\varepsilon 1}^2(\varphi(t_L) + 1). \tag{13}$$

Так как $S_\varepsilon^2 < S_{\varepsilon 1}^2$, то дисперсия ошибки прогноза (12) будет меньше дисперсии ошибки прогноза (13).

4. Докажем соотношение (10).

В соответствии с (5) и (8) запишем матрицу T_x в виде блочной матрицы $T_x = (T, D)$.

Тогда матрица $\left(T_x' \cdot T_x\right)$ примет вид

$$T_x' \cdot T_x = \begin{pmatrix} T'T, \, T'D \\ D'T, \, D'D \end{pmatrix}.$$

Обратная матрица $\left(T_x{}' \cdot T_x\right)^{-1}$ равна [2]

$$\left(T_x' \cdot T_x\right)^{-1} = \begin{pmatrix} T_{11}, \, T_{12} \\ T_{21}, \, T_{22} \end{pmatrix}, \text{ где}$$

$$T_{11} = (T'T)^{-1} - (T'T)^{-1} T'D(D'T(T'T)^{-1} T'D - D'D)^{-1} D'T(T'T)^{-1}$$
$$T_{21} = (D'T(T'T)^{-1} T'D) - D'D)^{-1} D'T(T'T)^{-1}$$
$$T_{12} = (T'D(D'D)^{-1} D'T - T'T)^{-1} T'D(D'D)^{-1}$$
$$T_{22} = (D'D)^{-1} - (D'D)^{-1} D'T(T'D(D'D)^{-1} D'T - T'T)^{-1} T'D(D'D)^{-1}$$

Обозначим через V_{x1T} матрицу ковариаций векторов $\overline{d}_x = (dx_1, ..., dx_n)$ и $\overline{t} = (1, t, t^2)$, $V_{x,t} = \dfrac{1}{N} D'T$ через $\gamma_{x,t}$ – матрицу коэффициентов регрессии d_{xi} по t^{j-1}, $\gamma_{x,t} = (T'T)^{-1} T'D$ через R_t – ковариационную матрицу вектора $\overline{t} = (1, t, t^2)$, $R_t = \dfrac{1}{N} T'T$ через R_x – ковариационную матрицу вектора $\overline{d}_x$, $R_x = \dfrac{1}{N} D'D$ через $\gamma_{x,t}$ – матрицу $\gamma_{x,t} = (D'D)^{-1} D'T$. Тогда матрица $\left(T_x{}' \cdot T_x\right)^{-1}$ примет вид

$$(T_x' \cdot T_x)^{-1} = \frac{1}{N} \begin{pmatrix} R_t^{-1} - \gamma_{x,t}(V_{x,t}\gamma_{x,t} - R_x)^{-1} \gamma_{x,t}', \; (V_{x,t}'\gamma_{t,x} - R_t)^{-1} \gamma_{t,x}' \\ (V_{x,t}\gamma_{x,t} - R_x)^{-1} \gamma_{x,t}', \; R_t^{-1} - \gamma_{t,x}(V_{x,t}'\gamma_{t,x} - R_t)^{-1} \gamma_{t,x}' \end{pmatrix}.$$

Пусть $A = (V_{x,t}\gamma_{x,t} - R_x)^{-1} \gamma_{x,t}'$. Так как матрица $\left(T_x{}' \cdot T_x\right)^{-1}$ симметрична, то $(V_{x,t}'\gamma_{t,x} - R_t)^{-1} \gamma_{t,x}' = A'$.

Тогда матрицу $\left(T_x{}' \cdot T_x\right)^{-1}$ можно записать в виде

$$\left(T_x{}' \cdot T_x\right)^{-1} = \frac{1}{N} \begin{pmatrix} R_t^{-1} - \gamma_{x,t} A & A' \\ A & R_x^{-1} - \gamma_{t,x} A' \end{pmatrix}.$$

Согласно предположению, $d_{xj}(t)$ не зависят от t, поэтому матрица коэффициентов ковариации $V_{x,t} \approx 0$ и матрицы коэффициентов регрессии $\gamma_{x,t} \approx 0$, $\gamma_{t,x} \approx 0$ и, следовательно, $A \approx 0$. В результате

$$\left(T_x{}' \cdot T_x\right)^{-1} \approx \frac{1}{N} \begin{pmatrix} R_t^{-1} & 0 \\ 0 & R_x^{-1} \end{pmatrix} = \begin{pmatrix} (T'T)^{-1} & 0 \\ 0 & (D'D)^{-1} \end{pmatrix}. \tag{14}$$

Подставляя (14) в (6), получаем искомое соотношение (10):

$$S_{\widehat{yx}}^2 = S_\varepsilon^2(\overline{t}_L{}', \overline{d}_L{}') \begin{pmatrix} (T'T)^{-1} & 0 \\ 0 & (D'D)^{-1} \end{pmatrix} (\overline{t}_L, \overline{d}_L) =$$
$$= S_\varepsilon^2 \overline{t}_L{}' (T'T)^{-1} \overline{t}_L + S_\varepsilon^2 \overline{d}_L{}' (D'D)^{-1} \overline{d}_L.$$

5. Пример.

На основе статистических данных за 16 месяцев, приведенных в табл. 1, построить прогноз объема реализации продукции одного из продуктов фирмы на два месяца вперед.

Таблица 1

Время	(мес.)	1	2	3	4	5	6	7	8
Объем реализации (тыс. руб.)		121	137	148	191	274	370	432	445
Индекс потребительских расходов (%)		100	98,4	101,2	103,5	104,1	107	107,4	108,5
Время	(мес.)	9	10	11	12	13	14	15	16
Объем реализации (тыс. руб.)		432	367	321	307	254	228	176	134
Индекс потребительских расходов (%)		108	109	110,1	110,7	110,3	112	112,3	112,9

Решение.

Проверим наличие тренда в исходных временных рядах с помощью Мастера диаграмм [3] в MS Excel (Рис. 1 и Рис. 2).

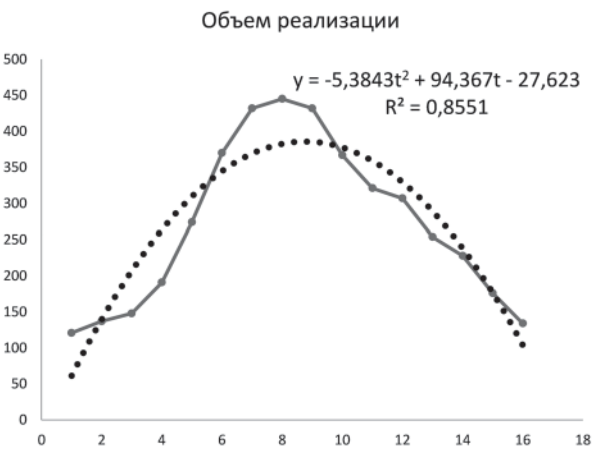

Рис. 1. Временной ряд «Объем реализации» содержит тренд – полином второго порядка

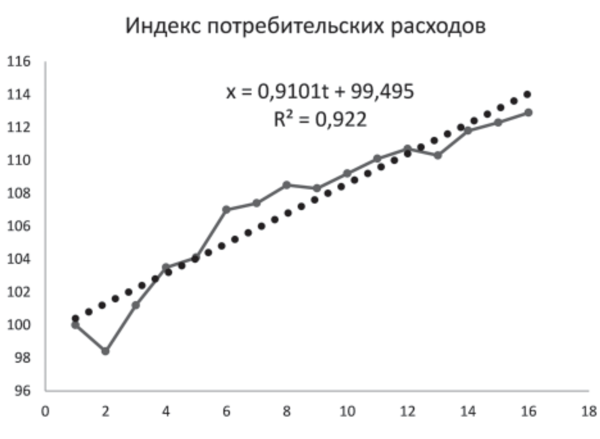

Рис. 2. Временной ряд «Индекс потребительских расходов» содержит тренд – полином первого порядка

Используя надстройку MS Excel Анализ данных [4], мы получим не только трендовую модель $\hat{y}(t) = -27,623 + 94,367t - 5,384$, с высоким коэффициентом детерминации (0,855), но и стандартную ошибку S_{ε}^2 равную 46,849 (Рис. 3).

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,925
R-квадрат	0,855
Нормированный R-квадрат	0,833
Стандартная ошибка	**46,849**
Наблюдения	16

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	2	168323,881	84161,940	38,345	3,53008E-06
Остаток	13	28533,057	2194,851		
Итого	15	196856,938			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	**-27,623**	40,038	-0,690	0,502
t	**94,367**	10,840	8,705	0,000
t^2	**-5,384**	0,620	-8,686	0,000

Рис. 3. Фрагмент протокола регрессионного анализа построения трендовой модели *Y*

Вычислим $dx_i(t) = x_i(t) - \hat{x}_i(t),\ i = 1,2,...n$ – отклонения от тренда для фактора «Индекс потребительских расходов» (Рис. 4).

Остатки	–0,41	–2,92	–1,03	0,36	0,05	2,04	1,53	1,72
	0,61	0,6	0,59	0,28	–1,03	–0,44	–0,85	–1,16

Рис. 4. Остатки – отклонения от линейного тренда для фактора «Индекс потребительских расходов»

Построим регрессионную трендово-факторную модель, добавив к исходным данным ряд остатков (рис. 5).

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,944
R-квадрат	0,891
Нормированный R-квадрат	0,864
Стандартная ошибка	**42,239**
Наблюдения	16

	Коэффициенты	*Стандартная ошибка*
Y-пересечение	**37,928**	48,778
t	**72,516**	14,666
t^2	**-4,099**	0,852
Остатки	**26,362**	13,193

Рис. 5. Фрагмент протокола регрессионного анализа построения трендово-факторной модели

Сравнивая стандартные ошибки двух моделей 46,849 (Рис. 3) и 42,239 (Рис. 5), мы видим, что стандартная ошибка трендово-факторной модели уменьшилась в этом примере на 9,8%.

Литература

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998.

2. Гантмахер Ф. Р. Теория матриц – 5-е изд. – М.: Физматлит, 2004.

3. Орлова И.В., Турундаевский В.Б. Некоторые особенности, возникающие при изучении нелинейной регрессии с использованием Excel и других программ. / Экономика, статистика и информатика. Вестник УМО. 2014. № 1. С. 158–161

4. Орлова И.В., Половников В.А. Экономико-математические методы и модели: компьютерное моделирование: учеб. пособие. – 3-е изд., перераб. и доп. – М.: Вузовский учебник: ИНФРА-М, 2011. – 389 с.

References

1. Ayvazian S.A., Mkhitaryan V.S. Applied Statistics and Econometrics fundamentals. M.: UNITY, 1998.

2. Gantmakher F.R. Matrix theory – 5th ed. – M .: Fizmatlit, 2004.

3. Orlova I.V., Turundaevsky V.B. Some features, fusstic in the study of non-linear regression using excel and other programs. / Ekonomika, statistika i informatika. Vestneyk UMO. 2014. № 1. S. 158–161

4. Orlova I.V., Polovnikov V.A. Economic-mathematical methods and models: computer simulation: Proc. allowance. – 3rd ed., Rev. and add. – M.: INFRA-M, 2011. – 389 p.