11. ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И ИНФОРМАТИКА

11.1. MATHCAD ДЛЯ ЭКОНОМИСТА: ЕСЛИ ИСХОДНЫЕ ДАННЫЕ «НЕНОРМАЛЬНО» РАСПРЕДЕЛЕНЫ

Порунов А.Н., к.э.н., научный сотрудник лаборатории стратегических исследований и операционного проектирования

Самарский государственный технический университет

В статье рассматривается методика реализации в среде Mathcad ненормально распределенного ряда макроэкономического ряда к нормально распределенному на основе преобразования Бокса-Кокса.

ВВЕДЕНИЕ

Очень часто[1] экономисту-аналитику приходится иметь дело со статистическими данными, которые по тем или иным причинам не проходят тест на нормальность. В этой ситуации есть два выхода: либо обратиться к непараметрическим методам, что весьма проблематично для экономиста, поскольку требует изрядной математической подготовки, либо воспользоваться специальными методами, позволяющими преобразовать исходную «ненормальную статистику» в «нормальную», что само по себе так же непросто. Практика статистических исследований показывает, что непараметрические методы наиболее приемлемы, когда объем выборок мал или же наблюдаемые значения переменной являются результатом очень редких событий.

Если же данных много (например, $n > 100$) или исследуются переменные, значения которых определяются бесконечным числом независимых факторов, то не имеет смысла использовать непараметрические статистики. К тому же часто забывают, что непараметрические тесты имеют меньшую статистическую мощность (менее чувствительны), чем их параметрические конкуренты, и если важно провести «тонкий» анализ, обнаружить даже слабые отклонения, то непараметрические методы в этом случае плохой помощник. В этой ситуации лучше обратиться к методам трансформации ненормально распределенных данных в нормально распределенные. Среди множества таких методов преобразований одним из лучших (при неизвестном типе распределения) считается Бокс-Кокс (БК) преобразование.

Авторы этого преобразования известные статистики – Дж. Э.П. Бокс (George Edward Pelham Box), профессор Висконсинского университета в городе Мэдисон (США) и сэр Д. Р. Кокс (Sir David Roxbee Cox) – профессор колледжа Бирбека Лондонского университета. Впервые суть предлагаемого метода была изложена ими в 1964 г., в Журнале Королевского статистического общества [2]. Практические аспекты Бокс-Кокс преобразования (БКП), сегодня достаточно подробно рассмотрены в специальной англоязычной литературе [3-7, 8], чего нельзя сказать о отечественной.

БКП

Пусть некоторая совокупность $\boldsymbol{X}$ представлена вектором непрерывных данных $\boldsymbol{x_i}$, $\boldsymbol{i \in 1, ..., N}$. БКП определяется следующим образом:

$$x(\lambda) = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda}, \lambda \neq 0; \\ \ln(\lambda), \lambda = 0. \end{cases} \quad (1)$$

[1] Математики-экономисты считают, что «очень часто» – мягко сказано, здесь следовало бы сказать «в абсолютном большинстве случаев».

Выражение (1) представляет собой универсальное параметрическое семейство преобразований, которое экономисты часто используют в алгоритмах сезонной (циклической) корректировки для того чтобы сезонная составляющая преобразованного динамического ряда стала (хотя бы в первом приближении) неэволюционирующей по амплитуде, что упрощает ее последующую идентификацию [6]. Широко тиражируемые в литературе по экономической статистике и по этой причине популярные среди экономистов логарифмическое и степенное преобразования представляют лишь частный случай преобразования БК. Так, например, в зависимости от значений $\boldsymbol{\lambda}$ получаем: при $\boldsymbol{\lambda = 0}$ – логарифмическое, при $\boldsymbol{\lambda <> 2}$ – степенное.

Один из способов выбрать оптимальное значение $\boldsymbol{\lambda}$, – использование $\boldsymbol{\lambda}$, максимизирующей логарифм функции правдоподобия. Логарифм функция правдоподобия:

$$f(x,\lambda) = -\frac{N}{2} * \ln\left[\sum_{i=1}^{N} \frac{(x_i(\lambda) - \bar{x}(\lambda))^2}{N}\right] + (\lambda - 1) * \sum_{i=1}^{N} \ln(x_i), \quad (2)$$

где $\bar{x}(\lambda) = \dfrac{1}{N} * \sum_{i=1}^{N} x_i(\lambda)$ – среднеарифметическая БК преобразованных данных.

Поскольку изначально БК преобразование было ориентировано только на положительные величины, проблему учета отрицательных значений данных снимают, добавляя к исходным значениям некоторое смещение, переводящее все отрицательные величины в положительную область[2]:

$$x(\lambda) = \begin{cases} \dfrac{(x + \text{смещение})^{\lambda} - 1}{\lambda}, \lambda \neq 0; \\ \ln(x + \text{сммещени}), \quad \lambda = 0. \end{cases} \quad (3)$$

При этом должно выполняться следующее условие:
$(x + \text{смещение}) > 0 \ , \ \forall x_i \in X$.

Доверительная оценка $\boldsymbol{\lambda}$ (с использованием статистики отношения правдоподобия) может быть произведена следующим образом:

$$f(x,\lambda) \geq f(x;\tilde{\lambda}) - 0.5\chi^2_{\alpha,1}, \quad (4)$$

где $\tilde{\lambda}$ – оценка максимального правдоподобия для $\boldsymbol{\lambda}$;

$\chi^2_{\alpha,1}$ – верхняя $\boldsymbol{100 * (1 - \alpha)}$ процентиль χ-квадрат распределения с первой степенью свободы.

ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ

Для иллюстрации процедуры БКП в среде Mathcad[3] использовался макроэкономический ряд показателей валового внутреннего продукта (ВВП) Российской Федерации за период с 1985-2009 гг. (в границах современной РФ) составленный автором (табл. 1).

[2] Таким образом получается двухпараматрическое семейство преобразований, которое сегодня называется преобразованием БК.

[3] В большинстве современных математических пакетов сдвиг на константу (смешение) не предусмотрен, т.е. используется алгоритм более простого однопараметрического преобразования.

Таблица 1

ДИНАМИКА РЕАЛЬНОГО ВВП РФ В 1885-2009 гг. В ЦЕНАХ 2008 г.

Млрд. долл.

t	ВВП	t	ВВП	t	ВВП	t	ВВП
1885	76	1917	143	1949	301	1981	1440
1886	73	1918	116	1950	374	1982	1423
1887	80	1919	92	1951	440	1983	1477
1888	86	1920	77	1952	453	1984	1664
1889	79	1921	74	1953	476	1985	1661
1890	75	1922	69	1954	483	1986	1668
1891	65	1923	64	1955	536	1987	1666
1892	93	1924	82	1956	569	1988	1754
1893	92	1925	98	1957	610	1989	1763
1894	95	1926	121	1958	616	1990	1784
1895	106	1927	146	1959	692	1991	1745
1896	93	1928	162	1960	721	1992	1735
1897	105	1929	173	1961	691	1993	1716
1898	94	1930	152	1962	789	1994	1518
1899	89	1931	175	1963	830	1995	1282
1900	90	1932	166	1964	818	1996	1267
1901	87	1933	171	1965	849	1997	1141
1902	86	1934	208	1966	958	1998	1119
1903	99	1935	242	1967	970	1999	1156
1904	95	1936	293	1968	1020	2000	1068
1905	114	1937	289	1969	1062	2001	1111
1906	98	1938	295	1970	1047	2002	1173
1907	88	1939	333	1971	1086	2003	1332
1908	89	1940	359	1972	1203	2004	1585
1909	108	1941	382	1973	1273	2005	1746
1910	111	1942	344	1974	1218	2006	1860
1911	123	1943	225	1975	1253	2007	2001
1912	107	1944	202	1976	1349	2008	2271
1913	118	1945	217	1977	1420	2009	2074
1914	134	1946	194	1978	1469	-	-
1915	158	1947	225	1979	1466	-	-
1916	160	1948	280	1980	1424	-	-

С помощью мастера импорта данных (data import wizard) файл исходных данных, созданный в Excel, импортируется в Mathcad (рис. 1).

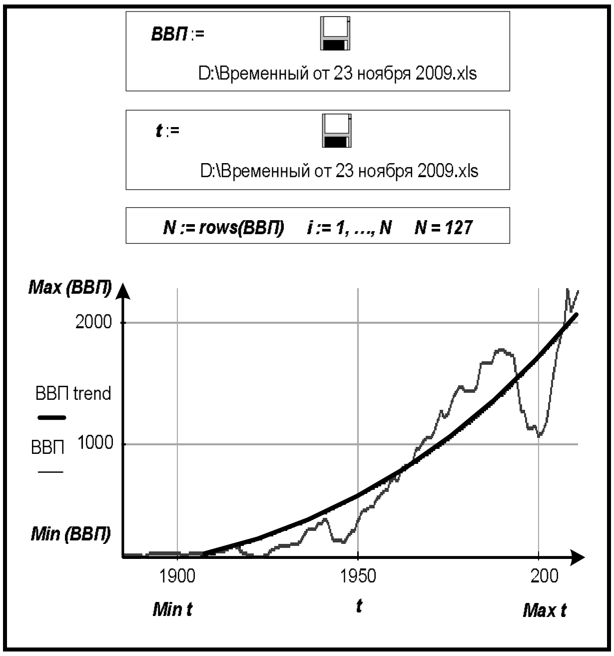

Рис. 1. Макроэкономический ряд ВВП и тренд

Вычтем из ряда тренд – определим ряд остатков (рис. 2).

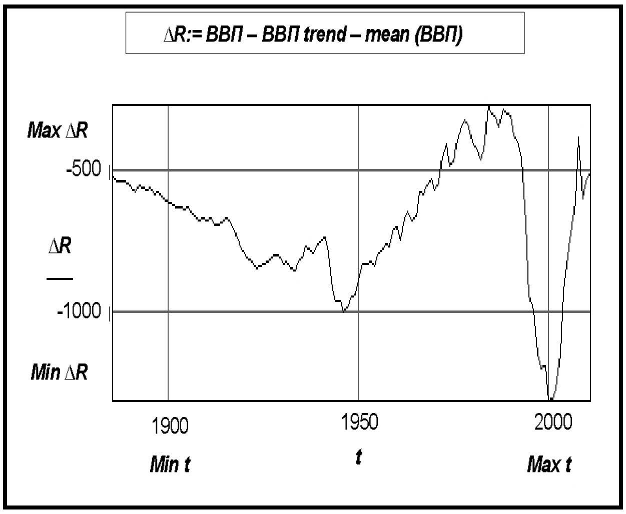

Рис. 2. Динамика ряда остатков

Для проверки близости распределения ряда остатков к нормальному, построим гистограмму распределения остатков (рис. 3).

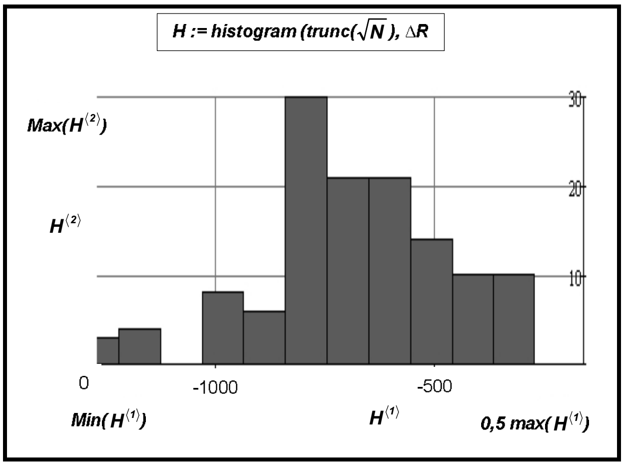

Рис. 3. Гистограмма распределения ряда остатков

Как видно из гистограммы, характер распределения ряда остатков далек от нормального. Как показывает практика, может оказаться, «…что преобразование квадратного корня еще слабовато (не поджимает справа хвост распределения), а логарифмическое – уже слишком сильное (хвостик появляется слева). Раньше пришлось бы выбирать из этих двух, но преобразование Бокса-Кокса в этом случае (**λ** между 0 и 0,5) найдет промежуточное решение. Поэтому, если истинное нормализующее преобразование неизвестно, преобразвание Бокса-Кокса считается лучшим» [8].

Поскольку БКП применяется только к положительным уровням ряда, выберем величину смещения так, чтобы ***(ΔR + Смещение) > 0*** при любых значениях ряда остатков ***ΔR***. Примем величину смещения равной:

$$\text{Смещение} := 1{,}2\ \min(\Delta R)\ ,$$

тогда ряд остатков с учетом смещения:

$$\Delta Rg := \Delta R - 1{,}2\ \min(\Delta R)\ .$$

Пусть показатель степени изменяется в пределах:

$$\lambda := -1,\ -1 + 0{,}1,\ ...,\ 15\ ,$$

с шагом 0,1, тогда лог-функцию правдоподобия можно определить следующим образом:

$$FP(\Delta Rg,\lambda) := \frac{-N}{2} * \\ * \ln\left[\frac{\sum_{i=1}^{N}\frac{(\Delta Rgi)^{\lambda}-1}{\lambda} - \frac{1}{N} * \sum_{i=1}^{N}\frac{(\Delta Rgi)^{\lambda}-1}{\lambda}}{N}\right] + \\ + (\lambda - 1) * \sum_{i=1}^{N}\ln(\Delta Rgi).$$

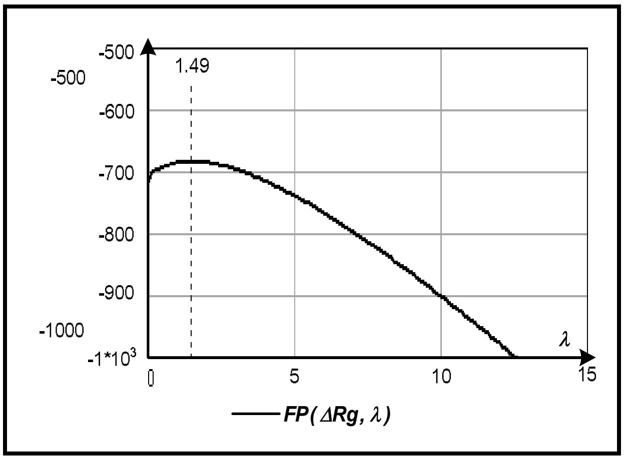

Рис. 4. График логарифмической функции правдоподобия

Для того чтобы найти оптимальное значение λ_{opt}, итеративно подставляем значения λ при которых логарифмическая функция правдоподобия $FP(\Delta R, \lambda)$ достигает максимума. Ориентируясь по графику логарифмической функции правдоподобия, возьмем «вилку» из значений:

$$FP(\Delta Rg, 1{,}48) = -682{,}903;$$

$$FP(\Delta Rg, 1{,}49) = -682{,}902\ ;$$

$$FP(\Delta Rg, 1{,}50) = -682{,}903\ .$$

Промежуточное значение $FP(\Delta R, 1{,}49)$ соответствует максимуму функции $FP(\Delta R,\lambda)$, т.е. в данном случае $\lambda_{opt} = 1{,}49.$

БК преобразование будет иметь следующий вид:

$$BC := \frac{\Delta Rg^{\lambda opt} - 1}{\lambda^{opt}}\ .$$

Проведем сортировку ряда BC:

$$\Delta Rn := sort(BC)\ .$$

Это позволит нам отразить кривую плотности нормального распределения на гистограмме (рис. 5):

$$H := histjgram(trunc(\sqrt{N}) - 1, BC)\ .$$

Классическая форма функции плотности нормального распределения (гаусин) в принятых обозначениях будет иметь следующий вид:

$$Nn(\Delta Rn) := \frac{1}{\sqrt{2 * \pi}} * \\ * \exp\left[\frac{-1}{2} * \left(\frac{\Delta Rn - mean(\Delta Rn)}{Stdev(\Delta Rn)}\right)^2\right].$$

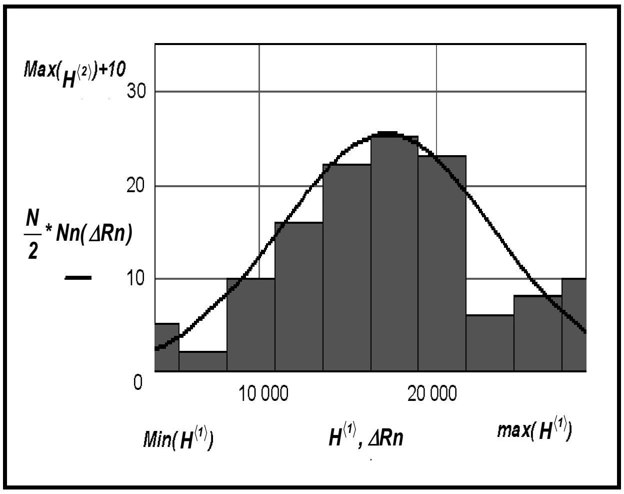

Рис. 5. Ряд остатков после БКП

Гистограмма (см. рис. 5) показывает, что характер распределения остатков после преобразования по методу Бокса-Кокса, близок к нормальному. Оценим ряд остатков на нормальность распределения, на основе показателей эксцесса и асимметрии[4].

Коэффициент асимметрии:

$$skew(BC) = -0{,}0334\ .$$

Эксцесса:

$$kurt(BC) = -0{,}1163\ .$$

Рассчитаем вспомогательные величины σA и σE:

$$\sigma A := \sqrt{\frac{6 * (N-2)}{(N+1) * (N+3)}} = 0.2123\ ;$$

$$\sigma E := \sqrt{\frac{24 * N * (N-2) * (N-3)}{(N+1)^2 * (N+3) * (N+5)}} = 0.4099\ .$$

Для ряда с распределением близким к нормальному должны выполняться следующие условия:

$$|\ skew(BC)\ | = 0.0334 < 1.5 * \sigma A = 0.3185$$

и

$$|\ kurt(BC) - \frac{6}{N+1}\ | = 0.16 < 1.5 * \sigma E = 0.6149\ .$$

В данном случае условия выполняются.

Порунов Аркадий Николаевич

[4] Для более строгой проверки статистической гипотезы необходимо использовать такие критерии согласия как Пирсона, Колмогорова или омега-квадрат.

Литература

1. Приведение данных к нормальному распределению: преобразование Бокса-Кокса [Электронный ресурс] // Тематический форум. http://molbiol.ru/forums/index.php?showtopic=201368
2. Box G.E.P., Cox D.R. An analysis of transformations. (With discussion) J. Roy. Statist. Soc. Ser. B 26 1964 211-252. http://www.ams.org/mathscinet-getitem?mr=192611
3. Box-Cox Transformation.
4. Box-Cox Transformations: An Overview.Pengfei Li. Department of Statistics, University of Connecticut. Apr 11, 2005 http://www.stat.uconn.edu/~studentjournal/index_files/pengfi_s05.pdf
5. Cand`es E.J., Guo F. New Multiscale Transforms, Minimum Total Variation Synthesis: Applications to Edge-Preserving Image Reconstruction. 2001. http://wwwstat.stanford.du/~olshen/manuscripts/selenite/node6.htm
6. Carroll R.J., Ruppert D. On prediction and the power transformation family. Biometrika 68: 609-615.
7. Davidson R., MacKinnon J.G. Estimation and Inference in Econometrics. Oxford University Press. 1993.
8. Definition of Box-Cox Transformation http://economics.about.com/cs/economicsglossary/g/box_cox.htm.

Ключевые слова

Преобразование Бокса-Кокса; макроэкономический ряд; непараметрические методы; параметрические методы; статистика.

РЕЦЕНЗИЯ

В последние годы в практику экономического анализа вошло большое количество программных реализаций алгоритмов статистической обработки в различных математических пакетах таких как Mathematica, Statistica, SPSS и др. Но незаслуженно выпал из этого перечня универсальный математический пакет Mathcad. Такое положение можно объяснить двумя причинами, – тем, что само по себе практическое использование этого пакета требует от пользователя ясного понимания специфики и возможностей применяемых методов обработки данных, всего того, что еще сравнительно недавно составляло основу подготовки и определяло в дальнейшем главное конкурентное преимущество отечественных ученых. Последнее особенно важно, если преследуются не только исследовательские, но и учебные цели.

В этой связи приведенный в рецензируемой статье пример практической реализации преобразования Бокса-Кокса в среде Mathcad особенно актуален, поскольку в некоторой степени заполняет брешь в ряду примеров статистической обработки экономических рядов на базе открытых алгоритмов в среде Mathcad.

В статье достаточно прозрачно показана процедура Бокс-Кокс преобразования с использованием метода максимального правдоподобия, что позволяет детально познакомиться с его возможностями и основным приемам реализации.

Заключение. Рецензируемая статья выполнена на актуальную тему, несомненно, обладает практической значимостью и рекомендуется к опубликованию.

Новиков А.В., д.т.н., профессор Самарского государственного технического университета

11.1. MATHCAD FOR THE ECONOMIST: IF INITIAL DATA ARE «ABNORMALLY» DISTRIBUTED

A.N. Porunov, the Scientific Employee of Laboratory of Strategic Researches and Operational Designing

Samara State Technical University

In clause the technique of realization in Mathcad environment abnormally distributed of some macroeconomic of some to normally distributed on the basis of Box-Cox transformation is considered.

Literature

1. Reduction of data to normal distribution: transformation Boksa-Coksa [the Electronic resource]//the Thematic forum. http://molbiol.ru/forums/index.php? showtopic=201368
2. Box, G.E.P.; Cox, D.R. An analysis of transformations. (With discussion) J. Roy. Statist. Soc. Ser. B 26 1964 211-252. http://www.ams.org/mathscinet-getitem?mr=192611
3. Box-Cox Transformations: An Overview.Pengfei Li. Department of Statistics, University of Connecticut. Apr 11, 2005 http://www.stat.uconn.edu/~studentjournal/index_files/pengfi_s05.pdf
4. Carroll, RJ and Ruppert, D. On prediction and the power transformation family. Biometrika 68: 609-615.
5. Box-Cox Transformation.
6. http://wwwstat.stanford.du/~olshen/manuscripts/selenite/node6.htm
7. Davidson, Russell, and James G. MacKinnon. 1993. Estimation and Inference in Econometrics. Oxford University Press.
8. Definition of Box-Cox Transformation http://economics.about.com/cs/economicsglossary/g/box_cox.htm

Keywords

Transformation of Boxing-coke; a macroeconomic number; nonparametric methods; parametrical methods; statistics.