УДК 519.254.1 DOI: 10.26102/2310-6018/2020.29.2.019

Модифицированный метод идентификации логистической кривой Рамсея

Д.В. Иванов
Самарский государственный экономический университет, Самара, Россия

Резюме: Кривые логистики широко используются в различных областях экономики, технологии, биологии, химии. Оценка параметров логистических тенденций по результатам наблюдений за динамическим процессом в экономической системе с целью достоверного анализа экономических показателей и прогнозирования их будущего поведения является одной из основных задач в экономике. Одной из логистических моделей является функция Рамсея. Преимущество этой функции заключается в возможности использовать линейное разностное уравнение для оценки его параметров. В то же время нелинейные преобразования данных не требуются, как для логистических функций Ферхюльста или Гомперца и других S-образных кривых. Оценивание при наличии помехи наблюдения параметров авторегрессии с помощью классического МНК дает смещенные оценки. Предложены модификации двухступенчатого алгоритма оценивания, основанные на применении метода обобщенных полных наименьших квадратов (ОПМНК) и метода расширенных инструментальных переменных (РИВ), для оценки параметров кривой Рамсея. Численные эксперименты показали, что точность оценки параметров с помощью предложенных модификаций выше, чем точность оценки, полученной с применением классического метода наименьших квадратов (МНК).

Ключевые слова: метод полных наименьших квадратов, логистические кривые, функция Рамсея, оценивание параметров.

Для цитирования: Иванов Д.В. Идентификация логистической кривой Рамсея методом полных наименьших квадратов. *Моделирование,оптимизация и информационные технологии*. 2020;8(2). Доступно по: https://moit.vivt.ru/wp-content/uploads/2020/05/Ivanov_2_20_1.pdf DOI: 10.26102/2310-6018/2020.29.2.019

Identification of the Ramsay logistic curve by total least squares

D.V. Ivanov
Samara State University of Economics, Samara, Russia

Abstract: Logistics curves are widely used in various fields of economics, technology, biology, chemistry. Estimating the parameters of logistic trends from the results of observations of the dynamic process in the economic system, with the aim of reliable analysis of economic indicators and predicting their future behavior, is one of the main tasks in the economy. One of the logistic models is the Ramsay function. The advantage of this function is the ability to use a linear difference equation to estimate its parameters. At the same time, non-linear data transformations are not required as for the logistics functions of Ferhulst or Gompertz. Modifications of a two-stage estimation algorithm based on the total least squares method and the extended instrumental variables method are proposed for estimating the parameters of the Ramsey curve.Tests have shown that the accuracy of parameter estimation using the proposed modifications is higher than the accuracy of the estimate obtained using the ordinary least squares method (LS).

Keywords: total least square, logistic curve, Ramsay function, estimation of parameters.

For citation: Ivanov D.V. Identification of the Ramsay logistic curve by total least squares. *Modeling, Optimization and Information Technology*. 2020;8(2). Available from: https://moit.vivt.ru/wp-content/uploads/2020/05/Ivanov_2_20_1.pdf DOI: 10.26102/2310-6018/2020.29.2.019 (In Russ).

Введение

При математическом описании процессов ограниченного роста в различных областях техники, экономики, биологии, химии широко используются S-образные кривые (логистические кривые) [1, 2, 3]. Проблема повышения точности оценивания динамических процессов, описываемых S-образные кривыми, является актуальной. Решение данной проблемы возможно на основе основе современных методов идентификации динамических систем.

Оценивание параметров S-образные кривых по результатам наблюдений динамического процесса для построения прогноза, является одной из важнейших задач. Для описания S-образные кривых используют различные модели, самые распространенные это логистические функции Верхулста (Перла-Рида), Гомперца функция гиперболического тангенса и др. [4].

Постановка задачи

Одной из таких моделей является функция Рамсея, зависящая от двух параметров b и A [5]:

$$y_i = A\left(1-\left(1+b\cdot T_i\right)\exp(-b\cdot T_i)\right)+\xi_i, \tag{1}$$

где $T_i = i\cdot\Delta t$ и Δt интервал дискретизации.

$\{\xi_i\}$ - последовательность независимых одинаково распределенных векторов для которой выполняются условия:

-математическое ожидание помехи равно нулю $\mathrm{E}\{\xi_i\}=0$, E - оператор математического ожидания.

-дисперсия помехи ограниченная и постоянная $\mathrm{E}\{\xi_i^2\}=0$.

Требуется по зашумленной последовательности $\{y_i\}$ оценить параметры модели b и A.

Применение метода наименьших квадратов для оценивания параметров b и A непосредственно из выражения (1) сопряжено с трудностями, так как параметры b и A входят в уравнение (1) нелинейно.

Алгоритм идентификации

Для повышения точности оценивания параметров уравнения (1) может быть применено Z-преобразование [5]. Z-преобразования позволяет перейти от нелинейной по параметрам статической модели (1) к динамической линейной по параметрам модели в виде разностного уравнения.

На первом шаге оцениваются параметры, соответствующего разностного уравнения. Для модели (1) на первом шаге оценивается параметр b.

На втором шаге на основе оценок, полученных на первом шаге, оцениваются параметры, входящие в уравнение линейно. Для модели (1) на втором шаге оценивается параметр A.

Шаг 1. Уравнение (1) может быть записано в виде в виде линейного разностного уравнения относительно разностей первого порядка

$$z_i - z_{i-1} = c_1\left(z_{i-1} - z_{i-2}\right) + c_2\left(z_{i-2} - z_{i-3}\right),$$

$$y_i = z_i + \xi_i, \tag{2}$$

где $c_1 = 2\exp(-b \cdot \Delta t)$ и $c_2 = -0.25c_1^2$

Ошибка предсказания для модели (2) может быть записана в виде

$$\varepsilon_i = y_i - y_{i-1} - c_1\left(y_{i-1} - y_{i-2}\right) - c_2\left(y_{i-2} - y_{i-3}\right),$$

$$\varepsilon_i = \xi_i - \xi_{i-1} - c_1\left(\xi_{i-1} - \xi_{i-2}\right) - c_2\left(\xi_{i-2} - \xi_{i-3}\right) \tag{3}$$

Из-за того что ошибка предсказания ε_i не является белым шумом, применение классического метода наименьших квадратов приводит к смещенным оценкам параметров .

Рассмотрим два метода, позволяющие получать несмещенные оценки параметров.

Метод обобщенных полных наименьших квадратов (ОПМНК) [7]. Определим дисперсию для ошибки предсказания:

$$E\left(\varepsilon_i^2\right) = 2\sigma_\xi^2 + c^T H_\xi c + 2c^T h_\xi = 2\sigma_\xi^2(1 + c^T \bar{H} c + c^T \bar{h}),$$

где $c = \begin{pmatrix} c_1 & c_2 \end{pmatrix}^T$, $\bar{h} = \begin{pmatrix} 1 & 0 \end{pmatrix}^T$, $\bar{H} = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$.

Будем искать оценки параметров $\hat{c}$ из условия минимума обобщенного отношения Рэлея [8]:

$$\hat{c}_{GTLS} = \arg\min_c \frac{\sum_{i=1}^{N}\left(y_i - y_{i-1} - c_1\left(y_{i-1} - y_{i-2}\right) - c_2\left(y_{i-2} - y_{i-3}\right)\right)^2}{1 + c^T \bar{H} c + c^T \bar{h}}. \tag{4}$$

Оценки параметров $\hat{c}$ (4) могут быть найдены из решения смещенной нормальной системы уравнений:

$$\hat{c}_{GTLS} = \left(\Phi^T \Phi - \lambda \bar{H}\right)^{-1}\left(\Phi^T Y + \lambda \bar{h}\right), \tag{5}$$

где $\lambda = 2\sigma_\xi^2$, $Y = \begin{pmatrix} y_3 - y_2 \\ \vdots \\ y_N - y_{N-1} \end{pmatrix}$, $\Phi = \begin{pmatrix} y_2 - y_1 & y_1 - y_0 \\ \vdots & \vdots \\ y_{N-1} - y_{N-2} & y_{N-2} - y_{N-3} \end{pmatrix}$.

Система линейных алгебраических уравнений (5) может быть решена стандартными методами решения линейных уравнений, например, с помощью LU разложения.

Метод расширенных инструментальных переменных (РИВ) [9]. Предположим, что существует вектор ψ_i такой, что:

- ψ_i не коррелирован с шумом ε_i;
- ψ_i хорошо коррелирован с регрессором φ_i.

Если размерность вектора инструментальных переменных больше размерности регресионного вектора

$$\mathrm{n}_\psi = \dim \psi_i > 2,$$

тогда оценка $\hat{\theta}_{IV}$ инструментальных переменных может быть найдена из выражения:

$$\hat{c}_{EIV} = \arg\min_{\theta} \left\| \left(\frac{1}{N} \sum_{i=1}^{N} \psi_i \varphi_i^T \right) c - \left(\frac{1}{N} \sum_{i=1}^{N} \psi_i y_i \right) \right\|_2^2 .$$

Минимум выражения (8) является решением системы нормальных уравнений:

$$\hat{c}_{EIV} = \left(\hat{R}_{\psi\varphi}^T \hat{R}_{\psi\varphi} \right)^{-1} \hat{R}_{\psi\varphi}^T \hat{r}_{\psi y}, \qquad (6)$$

где $\hat{R}_{\psi\varphi} = \frac{1}{N} \sum_{i=1}^{N} \psi_i \varphi_i^T$ - матрица размерности $\mathrm{n}_{\psi} \times 2$,

$\hat{r}_{\psi y} = \frac{1}{N} \sum_{i=1}^{N} \psi_i \left(y_i - y_{i-1} \right) \in \mathbb{R}^{n_\psi}$.

Шаг 2. На основе оценки коэффициентов $\hat{c} = \left(\hat{c}_1 \quad \hat{c}_2 \right)^T$ можно определить оценки параметра $\hat{b}$:

$$\hat{b} = -\frac{1}{\Delta t} \ln \left(0.5 \hat{c}_1 \right)$$

или

$$\hat{b} = -\frac{1}{\Delta t} \ln \left(0.25 \left(\hat{c}_1 + 2\sqrt{-\hat{c}_2} \right) \right).$$

Оценки параметра $\hat{A}$ могут быть найдены как решение нормальной системы уравнений:

$$\hat{A} = \left(\Phi_A^T \Phi_A \right)^{-1} \Phi_A^T Y_A,$$

где $\Phi_A = \begin{pmatrix} 1 - \left(1 + \hat{b} \cdot T_1 \right) \exp(-\hat{b} \cdot T_1) \\ \vdots \\ 1 - \left(1 + \hat{b} \cdot T_N \right) \exp(-\hat{b} \cdot T_N) \end{pmatrix}, Y_A = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}.$

Результаты моделирования

Предложенные модификации алгоритма с использованием на первом шаге обобщенного полного метода наименьших квадратов и расширенного метода инструментальных переменных были реализованы в Matlab и сравнивались с алгоритмами, использующими на первом шаге МНК оценку авторегрессионных параметров [4]. В качестве показателей качества модели использовались коэффициент детерминации

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(y_i - \hat{z}_i \right)^2}{\sum_{i=1}^{N} \left(y_i - E\left[y_i \right] \right)^2}$$

и относительная среднеквадратическая погрешность оценивания параметров

$$\delta c = \sqrt{\frac{\left\| \hat{c} - c \right\|^2}{\left\| c \right\|^2}} \cdot 100\,\% .$$

Пример 1. Кривая Рамсея описывается уравнением:

$$y_i = 2\left(1 - \left(1 + 0.025 \cdot T_i \right) \exp(-0.025 \cdot T_i) \right) + \xi_i, \qquad (7)$$

Число наблюдений $N = 15$ интервал дискретизации $\Delta t = 75$.

$\{\xi_i\}$ гауссовски распределенная последовательность с нулевым математическим ожиданием и среднеквадратическим отклонением $\sigma_\xi = 2 \cdot 10^{-2}$.

Разностное уравнение, соответствующее уравнению (10) при заданных параметрах дискретизации:

$$z_i - z_{i-1} = 0.3067(z_{i-1} - z_{i-2}) - 0.0235(z_{i-2} - z_{i-3}), \quad (8)$$
$$y_i = z_i + \xi_i.$$

Результаты идентификации представлены в таблице 1 и на рисунке 1.

Таблица 1. Результаты идентификации модели (8).
Table 1. Результаты идентификации модели (8).

	c_1	c_2	δc,%	b	A	R^2
Истинные параметры	0.3067	-0.0235	-	0.0250	2	1.00
МНК	0.2598	0.0036	17.62	0.0272	1.9895	0.9954
ОПМНК (5)	0.2850	-0.0092	8.46	0.0260	1.9946	0.9954
РИВ (6)	-0.2836	0.4134	238.78	0.0240 - 0.0418i	1.8007	0.2042

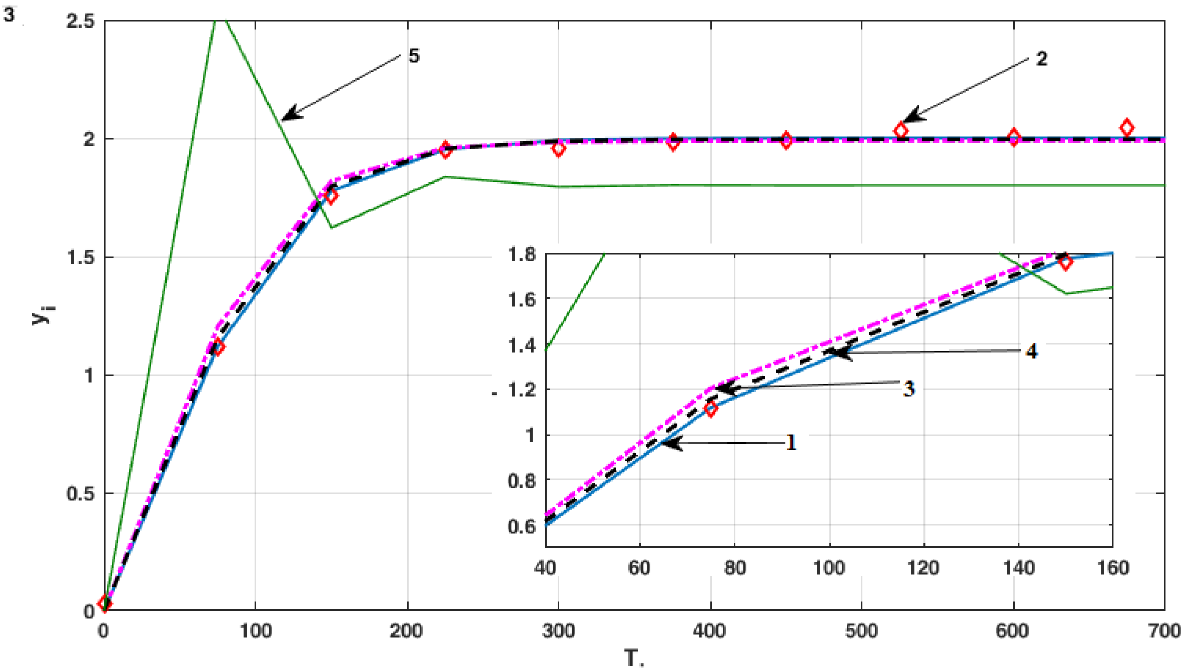

Рисунок 1 – Результаты идентификации модели (8) (1 –истинная кривая без шума, 2 – истинная кривая с шумом, 3 –кривая с МНК оценками, 4 – Кривая с ОПМНК оценками, 5 – Кривая с РИВ оценками)

Figure 1 - The results of the identification of the model (8) (1 - a curve with noise, 2 - a curve with noise, 3 - a curve with OLS estimates, 4 - a curve with GTLS estimates, 5 - a curve with RIV estimates)

Пример 2. Кривая Рамсея описывается уравнением:

$$y_i = 2\big(1-(1+0.025\cdot T_i)\exp(-0.025\cdot T_i)\big)+\xi_i. \quad (9)$$

Число наблюдений $N = 100$ интервал дискретизации $\Delta t = 1$.

$\{\xi_i\}$ гауссовски распределенная последовательность с нулевым математическим ожиданием и среднеквадратическим отклонением $\sigma_\xi = 10^{-4}$.

Разностное уравнение, соответствующее уравнению (9) при заданных параметрах дискретизации:

$$z_i - z_{i-1} = 1.9025(z_{i-1} - z_{i-2}) - 0.9048(z_{i-2} - z_{i-3}), \quad (10)$$

$$y_i = z_i + \xi_i.$$

Результаты идентификации представлены в таблице 2 и на рисунке 2.

Таблица 2. Результаты идентификации модели (10).

Table 2. Результаты идентификации модели (10).

	c_1	c_2	δc ,%	b	A	R^2
Истинные параметры	1.9025	-0.9048	-	0.050	2	1.00
МНК	1.8309	-0.8332	4.82	0.0884	1.6196	0.8823
ПМНК (5)	1.8316	-0.8340	4.75	0.0879	1.6286	0.8841
РИВ (6)	1.8955	-0.8980	0.46	0.0537	1.8007	0.9954

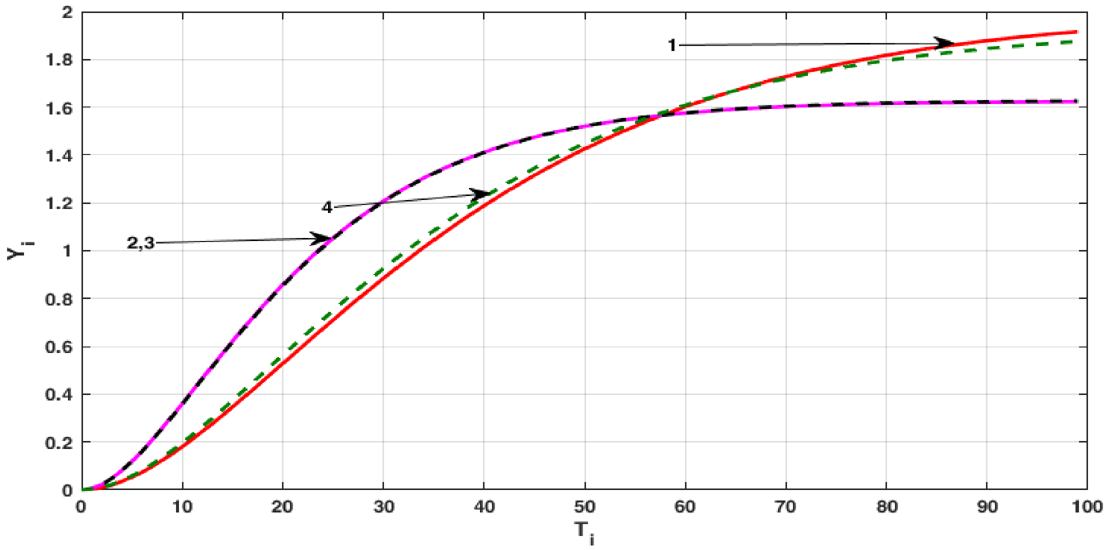

Рисунок 2 – Результаты идентификации модели (10) (1 – истинная кривая, 2, 3 –кривые с МНК оценками и ОПМНК оценками, 4 – Кривая с РИВ оценками)

Figure 2 - The results of the identification of the model (10) (1 - a curve with noise, 2, 3 - a curve with OLS estimates and GTLS estimates, 4- a curve with RIV estimates)

Заключение

Представленные результаты показывают, что предложенные модификации двухэтапного алгоритма с ОПМНК или РИВ на первом шаге способны давать более точные оценки чем алгоритм с МНК оценкой коэффициентов на первом шаге. Предложенные модификации дополняют друг друга: при малой частоте дискретизации более точные оценки дает обобщенный метод полных наименьших квадратов, при большой частоте дискретизации метод расширенных инструментальных переменных.

ЛИТЕРАТУРА

1. .Shields, L.B., Gertz, T.A., Wilson, K.C., Figg, G.L., Hester, S.T., & Honaker, J.T. The S-curve discontinuity theory predicts the path towards a “well” society and increased longevity. *Medical Hypotheses*, 2018;121:99-102. DOI:10.1016/j.mehy.2018.09.006.
2. Semenichev, V.K., Kurkin, E.I., Semenichev, E.V., Danilova, A.A., Fisun, G.A., & Kasatkina, E.I. Non-renewable recourses life cycles modeling aspects. *Procedia Computer Science*, 2015;65:872-879. DOI: 10.1016/j.procs.2015.09.046.
3. Tomić, D. Empirical evidence of an S-curve in Croatia. *Ekonomska Istraživanja*, 2019;32(1):2212-2230. DOI: 10.1080/1331677X.2019.1645718.
4. Семёнычев В.К., Коробецкая А.А., Кожухова В.Н. *Предложения эконометрического инструментария моделирования и прогнозирования эволюционных процессов*: монография. – Самара: САГМУ, 2015.

2. Ramsay, J.O.. A comparative study of several robust estimates of slope, intercept and scale in linear regression. *Journal of the American Statistical Association*,1977;72(359):608-615. DOI: 10.1080/01621459.1977.10480624.
3. Refaat, E.A. *Lecture notes on z-transform*. Morrisville: Lulu Press. 2005.
4. Markovsky I., Van Huffel S. Overview of total least squares methods. *Signal Processing*, 2007;87(10):2283–2302.
5. Ivanov, D.V., Chertykovtseva, N.V., Terekhova (Zharkova), A.A., & Andreeva, E.A. Identification of exponential trend models with fractional white noise. *Journal of Physics Conference Series, 2019, 1368,* 042061. DOI: 10.1088/1742-6596/1368/4/042061.
6. Söderström T., Stoica P. *Instrumental Variable. Methods for System Identification*. Berlin: Springer, 1983.

REFERENCES

1. Shields, L.B., Gertz, T.A., Wilson, K.C., Figg, G.L., Hester, S.T., & Honaker, J.T. The S-curve discontinuity theory predicts the path towards a “well” society and increased longevity. Medical Hypotheses, 2018;121:99-102, DOI:10.1016/j.mehy.2018.09.006.
2. Semenichev, V.K., Kurkin, E.I., Semenichev, E.V., Danilova, A.A., Fisun, G.A., & Kasatkina, E.I. Non-renewable recourses life cycles modeling aspects. *Procedia Computer Science*, 2015;65:872-879. DOI: 10.1016/j.procs.2015.09.046.
3. Tomić, D. Empirical evidence of an S-curve in Croatia. *Ekonomska Istraživanja*, 2019; 32(1):2212-2230. DOI: 10.1080/1331677X.2019.1645718.
4. Semenychev, V.K., Korobetskaya, A.A., & Kozhukhova, V.N. *Proposals of econometric tools for modeling and forecasting evolutionary processes*: Monograph. Samara: SAGMU. 2015. (In Russ)
5. Ramsay, J.O.. A comparative study of several robust estimates of slope, intercept and scale in linear regression. *Journal of the American Statistical Association*,1977;72(359):608-615. DOI: 10.1080/01621459.1977.10480624.

6. Refaat, E.A. *Lecture notes on z-transform.* Morrisville: Lulu Press. 2005.
7. Markovsky I., Van Huffel S. Overview of total least squares methods. *Signal Processing*, 200;87(10):2283–2302.
8. Ivanov, D.V., Chertykovtseva, N.V., Terekhova (Zharkova), A.A., & Andreeva, E.A. Identification of exponential trend models with fractional white noise. *Journal of Physics Conference Series, 2019, 1368,* 042061.DOI: 10.1088/1742-6596/1368/4/042061.
9. Söderström T., Stoica P. *Instrumental Variable. Methods for System Identification*. Berlin: Springer, 1983.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Иванов Дмитрий Владимирович к.ф.-м.н., доцент, кафедра «Высшая математика и экономико-математические методы», ФГБОУ ВО Самарский государственный экономический университет, Самара, Российская Федерация.
e-mail: *dvi85@list.ru*
ORCID: 0000-0002-5021-5259

Dmitriy V. Ivanov PhD, Associate professor, «Higher Mathematics And Economic-Mathematical Methods» Department, Samara State University Of Economics , Samara, Russian Federation.