UDC 311:33

Fraktal Regress

[1] Igor K. Kochanenko
[2] Ilia L. Pichik

[1] Rostov institute of system integration and high technologies, Russia
278 Krasnoarmeiskaya Str., Rostov-on-Don 344022
Dr. (Tech.), the leading research assistant
E-mail: kik112@yandex.ru
[2] Southern federal university, Russia
200/1, avenue of Strike, Rostov-on-Don, 344090
PhD student
E-mail: iliya37@yandex.ru

Abstract. Procedures of construction of curve regress by criterion of the least fractals, i.e. the greatest probability of the sums of degrees of the least deviations measured intensity from their modelling values are proved. The exponent is defined as fractal dimension of a time number. The difference of results of a well-founded method and a method of the least squares is quantitatively estimated.

Keywords: Fractal; intensity; regress; the forecast; supremum; a lot of information; a variety.

Введение. В области естественных, технических наук, экономики важной задачей является исследование взаимосвязи различных (физических, экономических и т.п.) величин, то есть поиск ответа на вопрос: как влияет изменение одной величины на значение, принимаемые другой; или более широко – взаимосвязи рынка с другими сегментами экономики. Часто ответ можно получить подбором параметров, описывающих некоторую конкретную функциональную зависимость, чем и занимается регрессионный анализ с интересными приложениями в области прогнозирования [1]. Наиболее распространенным приемом статистической обработки экспериментальных данных y_i является получение кривых регрессии методом наименьших квадратов (МНК). Здесь при оценке меры близости к этой кривой используются отклонения модельных и наблюдённых данных (остатки ε_i), которые считаются распределёнными по нормальному закону. Между тем в целом ряде приложений (например, в иконике, финансовом анализе) более информативными являются не y_i, а интенсивности $\lambda_i=dy_i/dt$, где t не обязательно время, это может быть, например, пространственная переменная. В ситуациях, когда условия измерений локально однородны, кластеры постоянной интенсивности можно считать инвариантами относительно изменений условий наблюдений и базой для прогнозирования, пролонгации, обнаружения и распознавания. При этом гипотеза нормального распределения часто не адекватна действительности. В [2] обосновывается рациональность в такой ситуации фрактального распределения. Построение кривой регрессии в случае, когда экспериментально определена интенсивность изучается в данной статье.

Материалы и методы. Критерий МНК связан с гильбертовым функциональным пространством с характерной евклидовой метрикой. Известным свойством такого пространства является то, что произвольная функция $f(z)$ из этого пространства с заданной точностью приближается кусочно-постоянной функцией:

$$f(z)=\sum_1^N \lambda i \chi i(z),\ \chi_i(z)=1 \rightarrow z \in A_i,\ \chi_i(z)=0 \rightarrow z \notin A_i, \qquad (1)$$

A_i–кластеры (области) постоянных λi. N-число кластеров. В работе [2] принимается, что каждый кластер Ai имеет характерное для него число n_i пикселов постоянной интенсивности λi. Причём зависимость $n_i(\lambda_i)$, как обладающая свойствами большой информативности и разнообразия, фрактальна:

$$n_i(\lambda_i) = \alpha A^\alpha / \lambda i^d,\ d = \alpha + 1, \qquad (2)$$

где параметры определяются по соотношениям [3]: фрактальная размерность это $d=ln(G/A)$, G – среднее геометрическое наблюдений интенсивностей, исключая фон; фон это $A_1 == \min\{\lambda_1, \lambda_2, ..., \lambda_{m-1}\}$.

Теперь выражение (1) приобретает вид:

$$f(z)=\sum_1^N [n_i/\alpha A^\alpha]^d \chi i(z)+ \lambda_m \chi_m,\ d > 0,\ i = 1, 2, ..., (m-1)\}, \quad (3)$$

Получается форма кусочно-постоянного выражения только в виде фрактала.

В пространстве R^n обобщением евклидовой метрики является p-метрика, которая при $p=2$ совпадает с евклидовой: $||x-y||_p =||z||_p = (\sum_1^m|z|^p)^{1/p}$,x,y,z из R^n. В работе [4] показано, что в R^n любые две p-метрики d и ρ эквивалентны, т.е. $\kappa_1 d\le\rho\le\kappa_2 d$. Следовательно, кусочно-постоянное приближение возможно и в функциональном пространстве с p-метрикой ($1\le p\le\infty$). При $p=2$ получается евклидова метрика.

Итак, если искать приближение кривой регрессии в пространстве интенсивностей по критерию максимума суммы в правой части (3), т.е.

$$\max f(z) = \max \sum_{i=1}^{N} (\alpha \cdot A^\alpha / n_i)^{-d} \cdot \chi_i(z), \quad (4)$$

то, учитывая, что n_i/N определяет долю постоянных интенсивностей λi можно интерпретировать (4) как наибольшую вероятность получения набора минимальных отклонений экспериментально оценённых интенсивностей от модельных, по аналогии с МНК можно назвать такой метод приближения методом наименьших фракталов (МНФ). В главном описанное выше основано на подходе работы автора [2], где задача узнавания изображения основывается на оценках фрактальной размерности пространства “интенсивность – координата” и формы изображения.

В статье [5] интенсивность использовалась для прогноза. Но там практически предполагалось наличие фрактального временного ряда и кривая регрессии получалась МНК. Кривая регрессии МНК строилась для оценки фрактальной размерности. В [5] прогнозирование валютного кризиса проводилось по величине фрактальной размерности с использованием подхода, теоретически обоснованного в работе [3], где доказаны диапазоны фрактальных размерностей, при которых динамическая система будет устойчивой.

Известно, что для гиперболических распределений, которым подчиняются фракталы, соотношения $p(\lambda i)=\alpha A^\alpha/\lambda i^d$ справедливы для значений аргумента, лишь начиная с некоторого порога аргумента λi_0 – точка супремума. Здесь $A=\alpha/ \lambda i_0$. Например, понимая, что в случае регрессии λi это разница между модельным и измеренным значениями интенсивности, то, $\lambda i_0=\Delta_0=(d-1)/P_{\max}$, стандартные вероятности $P_{\max}=0.9$; 0.97– $\sup[p(b)]$, b-параметр кривой регрессии. Таким образом, в случае фрактального распределения не экстремум, а супремум в точке λi_0 даёт возможность оценить производные по параметрам кривой регрессии:

$$\frac{d}{db}\sum_{i=1}^{N} A1(\lambda_i - b)^{-d}, \quad DIFb =_{db}^{d} (\cdot) = N(dA1\Delta_0^{-d}), \quad \Delta_0 = \lambda i_0, \quad A1=A\cdot\Delta_0^d.$$

Эти производные в МНК, где, как известно, находится экстремум функции, равны нулю. Вспоминая график гиперболического распределения, характерного для фракталов, супремум следует искать в точках наименьших значений аргумента (Δ_0).

Обсуждение. Для иллюстрации алгоритма получения фрактальной кривой регрессии, не уменьшая общности, в качестве модельной кривой регрессии взята линейная модель $y_i=a+bx_i$, i=1,2…N.

Используя простые преобразования, можно показать, что $\Delta_0=(d-1)/p$. Если считать приемлемыми, соответствующие супремуму наибольшие вероятности $p=0.9$ или $p=0.97$, то, например, при $d=1.1$ получаем $\Delta_0(0.9)=0.103$ и $\Delta_0(0.97)=0.111$, которые будут использоваться в количественном анализе ниже.

Таким образом, при известных производных по параметрам, задача на безусловный супремум сводится к уравнению (в случае нелинейной регрессии к системе уравнений):

$$\sum_{i=1}^{N} \left| \lambda_i - b \right|^{-d} = A3, \quad A3=DIFb/\ d\cdot A1, \tag{5}$$

DIFb- производная от критерия по параметру *b*.

$$DIFb = A1 \cdot d \cdot N \cdot (\Delta_0)^{-(1+Dd)}.$$

Следует заметить, что в случае нелинейной модели $y_i=a+bx_i+cx^2$ уравнение (5) дополняется близким по структуре уравнением, но с изменённой правой частью: $A3=DIFc/\ d\cdot A1$, где производная по параметру *c* равна:

$$DIFc = \cdot 2d \sum_{i=1}^{N} x_i \cdot (\Delta_0)^{-(1+d)}.$$

Результаты. По аналогии с МНК можно считать, что соотношение (5) получено на основании необходимого условия супремума функции (2). Решение уравнения (5) просто получить, воспользовавшись стандартной функцией root системы компьютерной математики Mathcad:

$$a = root[\sum_{i=1}^{N} \left| \lambda_i - b \right|^{-d} - A3, b].$$

Интересно оценить разницу в результатах (5) и метода наименьших квадратов. В качестве исходных данных были использованы материалы статьи [6], где приведены количественные данные сценарного прогноза населения России до 2055г. Ожидаемые значения циклов персистентности, оценённые по приведенной в [7] теореме, оказались существенно меньше заявленного в [6] горизонта прогноза до 2055 г. Это значит, что результаты анализа сценариев привели к количественным оценкам, коррелированность между которыми исчезает через малый промежуток времени, и, следовательно, их лучше не использовать для долговременного прогноза до 2055 г. и далее. Но, тем не менее, их вполне можно применить здесь для оценки модельных кривых линейной регрессии в виде y=a+bt; такие оценки были получены двумя методами: МНК и МНФ. Различия в полученных оценках параметров достаточно существенны, они не превышают 35%.

Выводы. Существенные различия в кривых регрессии МНК и МНФ свидетельствуют о необходимости при работе с регрессией соблюдать осторожность и обязательно проверить на адекватность предлагаемые модели. Наблюдения изображений, окружающей действительности, анализ видеоинформации часто в качестве наиболее информативных признаков выделяют контуры яркостей (интенсивностей). В связи с этим при статистическом изучении взаимосвязи величин различной природы, для построения адекватных моделей регрессии рационально проверять целесообразность использования в качестве исходных данных временные ряды интенсивностей, которые могут привести к более адекватному результату. И при положительных результатах проверки применять обоснованные здесь процедуры метода наименьших фракталов.

Примечания:

1.Kochanenko I. Fractal Characteristics of Forcast Properties of Non-linear Regress. // Lecture Notes in Information Technology, Jornal of Information Engineering Research Institute, USA, 2012, Vol.13, p.192-195.

2. Коханенко И.К. Анализ видеоинформации на основе фрактальной кластеризации. // Оптический журнал. 2010, Т. 77, №8, С.47-58.

3. Коханенко И.К. Фракталы в оценке эволюции сложных систем.// Автоматика и телемеханика. 2002. № 8. С. 54–62.

4. Кроновер Р.М. Фракталы и хаос в динамических системах. М: Постмаркет. 2000. 352с.

5. Мансуров А.К. Прогнозирование валютных кризисов с помощью методов фрактального анализа. //Проблемы прогнозирования. 2008. №1. С. 145-159.

6. Никитина С.Ю., Щербов С.Я. Вероятностный прогноз численности населения России. // Вопросы статистики. 2007. №7. С. 6-9.

7. Коханенко И.К., Москаев В.А., Пищик И.Л. Метод фрактальной оценки времени разладки прогноза // Обозрение прикл. и промышл. матем. 2012, т. 19, вып. 4. С. 574-575.

УДК 311:33

Фрактальная регрессия

[1] Игорь Константинович Коханенко
[2] Илья Львович Пищик

[1] РИСИиНТ, Россия
344022, г.Ростов-на-Д., ул.Красноармейская, 278
Доктор технических наук, ведущий научный сотрудник
E-mail: kik112@yandex.ru
[2]Южный федеральный университет, Россия
344090, г. Ростов-на-Дону, пр. Стачки, 200/1
Аспирант
E-mail: iliya37@yandex.ru

Аннотация. Обосновываются процедуры построения кривой регрессии по критерию наименьших фракталов, т.е. наибольшей вероятности сумм степеней наименьших отклонений измеренных интенсивностей от их модельных значений. Показатель степени определяется как фрактальная размерность временного ряда. Количественно оценивается разница результатов обоснованного метода и метода наименьших квадратов.

Ключевые слова: фрактал; интенсивность; регрессия; прогноз; супремум; информативность; разнообразие.