

УДК 51.77 330.46

МЕТОД РЕГРЕССИИ С ЗАДАНЫМИ ХАРАКТЕРИСТИКАМИ ОСТАТОЧНОГО РЯДА

Мишанин А.А.

Московский физико-технический институт, Долгопрудный, Россия (141700, Московская область, г. Долгопрудный, Институтский пер., 9), e-mail: minar@mail.ru

Предложен новый вычислительный метод для построения регрессий временных стохастических рядов. Для определения регрессии временных рядов часто используется метод наименьших квадратов, при котором сначала производится регрессия стохастического ряда, а затем происходит исследование свойств остаточного ряда с последующим анализом о возможности применения метода и границ доверительных интервалов. В отличие от него в предлагаемом методе заранее задаются характеристики остаточного ряда. После проведения предварительных вычислений над членами стохастического ряда решение задачи сводится к нахождению условного экстремума функции Лагранжа. Для случая, когда на остаточный ряд налагаются жесткие ограничения в виде равенств, решение сводится к решению системы уравнений и может быть получено в квадратурах. При наличии неравенств решение задачи может быть осуществлено посредством методов линейного и выпуклого программирования.

Ключевые слова: оптимальное управление, метод наименьших квадратов, методы оптимизации, регрессия.

REGRESSION METHOD WITH THE SPECIFIED CHARACTERISTICS OF THE RESIDUE SERIES

Mishanin A.A.

Moscow Physical-Technical Institute, Dolgoprudny, Russia (141700, Moscow AREA, Dolgoprudny Institutskiy lane, 9), e-mail: minar@mail.ru

A new numerical method is proposed to construct stochastic time series regressions. The least squares method is often used to determine the time-series regression, wherein at first the regression of stochastic series is performed, then the study of the properties of the residual number is produced, and then the analysis of the application possibility of the method and the confidence bounds is produced. In contrast, the proposed method predetermined characteristics of the residual series. After conducting the preliminary calculations on a time series the problem solution reduces to finding a conditional extremum of the Lagrange function. For the case when the restrictions of the residual series imposed in the form of equations, the solution is reduced to solving a system of equations and can be obtained by quadratures. If there are the inequalities the solution can be obtained by the methods of linear and convex programming.

Keywords: optimal control, the least-squares method, optimization methods, regression.

Введение

Метод наименьших квадратов получил большое распространение для целей нахождения регрессии стохастических рядов. Посредством этого метода можно осуществлять прогнозирование экономических показателей. Однако при использовании данного метода довольно часто опускают исследование возможности и целесообразности его применения, поскольку у каждой теории есть свои ограничения и допущения, при наличии которых она была получена.

Согласно теореме Маркова-Гаусса при рассмотрении модели парной регрессии

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \quad (1.1)$$

метод наименьших квадратов эффективен (имеет наименьшую дисперсию) в классе линейных несмещенных оценок при следующих условиях.

1. Модель данных правильно специфицирована.
2. Все значения регрессора X_i детерминированы и не равны между собой.
3. Ошибки ε_i не имеют систематического характера, т.е. имеют нулевое математическое ожидание $E(\varepsilon_i) = 0$.
4. Дисперсия ошибок постоянна $D(\varepsilon_i) = \sigma^2 = const$.
5. Ошибки различных наблюдений не коррелируют между собой $cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i, j$ (отсутствие автокорреляции).

Что касается проблемы идентификации параметров, то она для системы является сложной и может быть решена до определенной степени посредством феноменологического изучения рассматриваемой системы. Свидетельством истинности выбранной модели может служить лишь то, что она достаточно адекватно описывает реальность.

При этом стоит отметить, что регрессоры, как правило, линейны, поскольку изменения экономических параметров редко носят взрывной нелинейный характер. Кроме того, стохастическая составляющая для широкого круга случайных процессов имеет нормальное или асимптотически нормальное распределение, что приводит к тому, что уравнение регрессии (1.1) можно использовать для прогнозирования большого класса экономических процессов.

Второе условие предписывает иметь как можно больший разброс значений регрессора для более точного описания зависимой переменной. Требование детерминированности регрессора автоматически выполняется в том случае, когда в качестве него используется время. К тому же этот случай является наиболее востребованным в практических целях, поскольку выявляемая временная зависимость переменной чаще гораздо важнее, чем зависимость от неизвестных в будущем значений стохастического регрессора.

Стоит отметить, что условие детерминированности регрессоров может быть ослаблено до условия предопределенности, т.е. регрессором может выступать случайная величина, но к моменту наблюдения ее значение уже определено. Также вместо условия детерминированности регрессоров достаточно более слабого условия независимости регрессоров и остаточного ряда.

Третье условие несмещенности оценки в модели линейной регрессии удовлетворяется достаточно просто за счет соответствующего подбора константы в уравнении регрессора. В результате для удовлетворения условий теоремы Маркова-Гаусса наиболее важной оказывается проверка четвертого и пятого условий об отсутствии автокорреляции и гомоскедастичности остаточного ряда, поскольку нарушение условий гомоскедастичности и

автокорреляции остаточного ряда приводит к тому, что метод наименьших квадратов приводит к смещенным, неэффективным и несостоятельным оценкам.

Проверка остаточного ряда на гомоскедастичность проводится при помощи критерия Гельфанда-Кванта, по которому все наблюдения разбиваются на две группы и сравниваются дисперсии остаточных рядов этих групп. При сильном расхождении данных дисперсий делается вывод о гетероскедастичности остаточного ряда.

Наиболее часто встречающийся вид автокорреляции – это автокорреляция первого порядка:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \xi_i,$$

где ρ - коэффициент автокорреляции первого порядка.

В этом случае

$$E(\varepsilon_i \varepsilon_{i-1}) = \rho E(\varepsilon_i^2) + E(\varepsilon_i \xi_i) = \rho E(\varepsilon_i^2) \Rightarrow \text{Corr}(\varepsilon_i, \varepsilon_{i-1}) = \rho.$$

Наиболее общепотребительным критерием для выявления автокорреляции первого порядка является критерий Дарбина-Уотсона:

$$DW = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n \varepsilon_i^2} \quad (1.2).$$

Можно показать, что

$$DW = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n \varepsilon_i^2} \approx 2(1 - \rho).$$

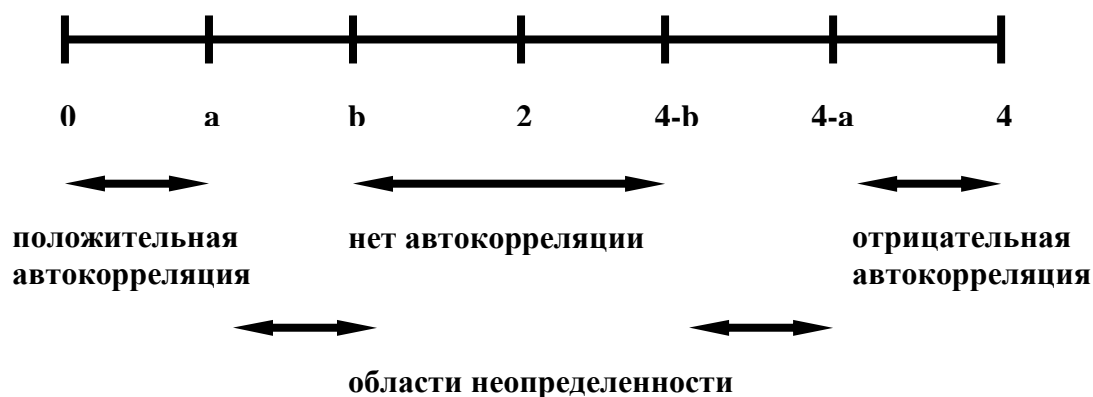


Рисунок 1. Интерпретация критерия Дарбина-Уотсона.

В случае отсутствия автокорреляции $DW = 2$, при положительной автокорреляции DW стремится к нулю, а при отрицательной – к четырем. Практическое применение критерия Дарбина-Уотсона можно изобразить на рисунке (рис. 1), из которого следует наличие областей неопределенности критерия. Теоретические значения a , b зависят от количества заданных наблюдений n , числа независимых переменных k и уровня значимости α .

Для выявления автокорреляции также применяется h – критерий Дарбина-Уотсона:

$$h = \left(1 - \frac{1}{2} DW \right) \sqrt{\frac{n}{1 - nV}},$$

где n - количество наблюдений в модели, V - дисперсия лаговой результативной переменной. При увеличении объема выборки h – статистика стремится к нормальному распределению с нулевым математическим ожиданием и единичной дисперсией.

Постановка задачи

На основании вышесказанного предлагается исследовать следующую задачу. Необходимо построить регрессор заданной функциональной зависимости таким образом, чтобы остаточный ряд обладал необходимыми свойствами; например, остаточный ряд должен иметь нормальное распределение.

Полная проверка остаточного ряда на соответствие его распределения нормальному заключается в проверке равенства нулю всех нечетных центральных моментов распределения остаточного ряда (при условии бесконечности количества его членов) и равенства $(2n-1)!!$ всех четных центральных моментов. На практике для оценки нормальности конечного ряда обычно используют коэффициенты асимметрии

$$\gamma = \frac{E(\xi - E(\xi))^3}{\sigma^3}$$

и эксцесса

$$\delta = \frac{E(\xi - E(\xi))^4}{\sigma^4} - 3.$$

Дополняя эти условиями требованием несмещенности оценки $\bar{\varepsilon}_i = 0$, получаем следующую задачу. Имеется временной ряд случайной величины. При заданном виде регрессора (спецификации) $f(t, a_1, \dots, a_n)$ найти значения параметров a_1, \dots, a_n так, чтобы математическое ожидание, коэффициенты асимметрии и эксцесса остаточного ряда

$\varepsilon = (\xi - f)$ обращались бы в нуль, а дисперсия остаточного ряда была бы минимально возможной. Иначе говоря, возникает задача по поиску условного экстремума следующей функции Лагранжа:

$$L(a_1, \dots, a_n, \lambda_1, \lambda_2, \lambda_3) = E((\xi - f)^2) + \lambda_1 E(\xi - f) + \\ + \lambda_2 E((\xi - f)^3) + \lambda_3 \left[E((\xi - f)^4) - 3 \left(E((\xi - f)^2) \right)^2 \right],$$

где $\lambda_1, \lambda_2, \lambda_3$ - множители Лагранжа. В результате поиска седловой точки функции Лагранжа мы получаем систему $(n+3)$ уравнений, из которых одно линейное относительно f , одно уравнение четвертой степени и $(n+1)$ кубических уравнений, так что в общем случае (на пространстве комплексных чисел) система имеет $(3^{(n+1)} * 4)$ решений.

Стоит отметить, что система имеет 3 уравнения относительно параметров a_1, \dots, a_n и n уравнений относительно $a_1, \dots, a_n, \lambda_1, \lambda_2, \lambda_3$. Из этого следует, что при $n < 3$ система не имеет решения, поскольку она переопределена относительно параметров a_1, \dots, a_n , для разрешения этой задачи необходимо ввести $(3-n)$ дополнительных условий.

При $n = 3$ система распадается на две подсистемы относительно только параметров a_1, \dots, a_n и только множителей Лагранжа $\lambda_1, \lambda_2, \lambda_3$, этот вырожденный случай имеет 12 решений. Простейшим случаем трехпараметрической функции является квадратный полином $f = a_1 t^2 + a_2 t + a_3$.

Применим данный метод к статистическим данным по ценам на нефть марки **Brent** за 2011 год. Полное семейство регрессий состоит из шести комплексно сопряженных пар решений, действительные части которых изображены ниже (рис. 2). В легенде графика каждое решение описывается среднеквадратическим отклонением соответствующего остаточного ряда. Математические ожидания, коэффициенты асимметрии и эксцесса всех остаточных рядов равны нулю (с точностью до $1 * 10^{-15}$) в соответствии с условиями постановки задачи.

Стоит отметить, что использование метода наименьших квадратов может приводить к неверным результатам, поскольку его целью является минимизация дисперсии. Однако дисперсия является мерой неопределенности случайного процесса и согласно неравенству Рао-Крамера обратно пропорционально зависит от количества получаемой об объекте

исследования информации. Иначе говоря, дисперсию случайного процесса можно рассматривать не как производную величину, полученную в результате применения метода наименьших квадратов, а как заданный параметр задачи, принимающий какое-то определенное значение (спектр значений или интервал значений). Таким образом, в приведенном примере (рис. 2) можно выбрать не регрессию с минимальной дисперсией, а такую регрессию, дисперсия которой, например, с точки зрения эксперта, соответствует реальной ситуации.

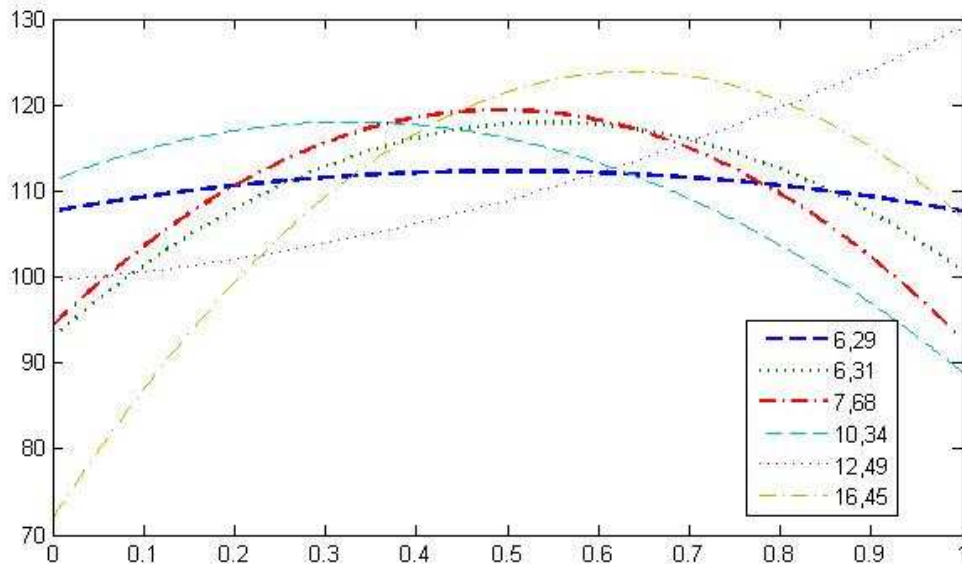


Рисунок 2. Семейство регрессий с нормальным остаточным рядом.

Ниже вместе с графиком исходных данных приведены регрессии, полученные методом наименьших квадратов и методом нормального остаточного ряда с минимальной дисперсией (рис. 3).



Рисунок 3. Исходные данные и регрессии, полученные методом наименьших квадратов и методом нормального остаточного ряда с наименьшей дисперсией.

Из теории метода наименьших квадратов известно, что чем выше степень полинома, тем больше вероятность того, что полином будет отражать случайные колебания наблюдаемой величины, не отражая ее реального тренда. Поэтому тренд наблюдаемого экономического показателя желательно искать в линейном виде $f = a_1t + a_2$. Однако в этом случае для корректности поставленной задачи необходимо вместо трех оставить два условия, налагаемых на ряд. Представляется интересным разобрать случай, когда при несмещенной оценке $\bar{\varepsilon}_i = 0$ остаточный ряд не имеет авторегрессии, что свидетельствует о правильно подобранном регрессоре. Как уже отмечалось выше, как правило, в экономических показателях встречается авторегрессия первого порядка, которая определяется посредством критерия Дарбина-Уотсона (1.2). Потребуем точного выполнения критерия $DW = 2$.

Нетрудно показать, что в этом случае при достаточно больших временных рядах

$$a_1 = \frac{\text{cov}(x, t)}{Dt} \pm \frac{\sqrt{(\text{cov}(x, t))^2 - Dt * Dx}}{Dt}, \quad (1.3)$$

где первое слагаемое представляет значение коэффициента a_1 , полученное регрессией методом наименьших квадратов, а второе слагаемое в силу неравенства Коши-Буняковского является мнимым и лежит в плоскости, перпендикулярной плоскости действительной части решений. Как видно из формулы (1.3), равенство нулю мнимой части достигается только в случае полной корреляции между x и t . В наличии мнимых частей у решений нет никаких противоречий с постановкой задачи, главное, чтобы действительные части хорошо описывали поведение временного ряда. Более того, это в некоторых случаях может объяснять гетероскедастичность остаточного ряда, поскольку при постоянной общей дисперсии дисперсия действительной части будет уменьшаться на размер увеличения дисперсии мнимой части решения и наоборот.

Действительная часть коэффициента a_2 также совпадает со значением коэффициента a_2 , полученным методом наименьших квадратов. В результате проекция линейной регрессии, построенной при равенстве нулю коэффициента авторегрессии первого порядка, на множество действительных чисел совпадает с регрессией, полученной методом

наименьших квадратов. Полученный результат можно распространить на случай отсутствия авторегрессии n -го порядка.

Применительно к уже упоминавшимся выше статистическим данным по ценам на нефть марки **Brent** за 2011 год для метода наименьших квадратов получим следующее уравнение регрессии:

$$Brent = 1.5687t + 109.9438,$$

а для уравнения регрессии, при котором у остаточного ряда нет авторегрессии первого порядка, получим

$$Brent = (1,5680 \pm 23,3849i)t + (109,9442 \mp 11,6925i).$$

Погрешность обусловлена тем, что при выводе формулы (1.3) были отброшены члены малости порядка $1/n$. Среднеквадратические отклонения остаточных рядов для этих методов соответственно равны 6.9113 и 9.6845.

Таким образом, метод наименьших квадратов дает не только несмещенную, эффективную, состоятельную и оптимальную линейную оценку, но и дает оптимальную оценку в пространстве действительных решений для случая отсутствия авторегрессии. Однако нельзя не отметить тот факт, что уравнение регрессии с ненулевой мнимой частью не может соответствовать стационарному случаю. Это следует из того, что линейная регрессия с ненулевой мнимой частью с течением времени все дальше будет удаляться от плоскости действительных значений и это приведет к тому, что дисперсия остаточного ряда будет стремиться к бесконечности.

Поскольку, как отмечалось выше, мнимая часть регрессии обращается в нуль только при полной корреляции между x и t , можно сделать вывод о том, что условия несмещенности оценки и отсутствия автокорреляции для стационарной модели практически несовместны. Принимая во внимание тот факт, что применяемый для регистрации автокорреляции критерий Дарбина-Уотсона не эффективен в случае гетероскедастичности, становится логичным заменить условие несмещенности оценки критерием Гельфанда-Кванта для проверки на гетероскедастичность, причем будем требовать точного выполнения равенства дисперсии первой и второй части временного ряда.

Численный расчет такого вида регрессии приводит к двум парам уравнений:

$$Brent_1 = (1,5524 \mp 9,2789i)t + (109,9500 \pm 10,8562i)$$

$$Brent_2 = (-1,7619 \pm 22,8259i)t + (109,3703 \mp 14,2606i),$$

для которых математическое ожидание и среднеквадратическое отклонение остаточных рядов соответственно равны:

$$Brent_1 : 0,0019 \mp 6,2167i; \quad 7,4170 ,$$

$$Brent_2 : 2,2388 \pm 2,8477i; \quad 9.6203 .$$

Нетрудно видеть, что действительная часть первой пары уравнений регрессии практически полностью совпала с уравнением регрессии, полученной методом наименьших квадратов. Таким образом, метод наименьших квадратов, помимо вышеперечисленных свойств, дает оптимальную оценку в пространстве действительных решений для случая отсутствия авторегрессии и гетероскедастичности.

На основе рассмотренных выше примеров можно заключить, что предлагаемый метод является мощным инструментом в прогнозировании экономических событий, поскольку позволяет исследователю конструировать различные стохастические модели. При этом исследование сводится к решению системы уравнений, что с точки зрения затрат машинного времени выгодно отличает данный метод от многих вычислительных методов. Следует отметить, что условия, накладываемые на ряд, могут быть не только жесткими в виде равенств, но и в виде неравенств. В этом случае регрессия может быть найдена методами линейного и выпуклого программирования.

Список литературы

1. Льюнг Л. Идентификация систем. Теория для пользователя. - М. : Наука, 1991. - 432 с.
2. Мишанин А.А. Применение аппарата теории вероятностей к описанию экономических процессов // Труды ИСА РАН Динамика неоднородных систем. – 2010. - Т. 53 (3). - С. 181-199.
3. Мишанин А.А. Применение стохастических методов к вопросам безопасности экономики // Фундаментальные проблемы системной безопасности : материалы V Международной научной конференции, посвященной 90-летию со дня рождения выдающегося ученого, генерального конструктора ракетно-космических систем академика В.Ф. УТКИНА. – Елец : ЕГУ им. И.А. Бунина, 2014. - С. 511-516.
4. Рао С.Р. Линейные стохастические методы и их применения. – М. : Наука, 1968.
5. Уотшем Т. Дж., Парамоу К. Количественные методы в финансах. – М. : Финансы; Юнити, 1999. – 527 с.
6. Ширяев А.Н. Вероятность. – М., 1979.

Рецензенты:

Дикусар В.В., д.ф.-м.н., профессор, ведущий сотрудник, ВЦ РАН им. А.А. Дородницына,
г.Москва.

Умнов А.Е., д.т.н., профессор, старший научный сотрудник, Государственный университет
Московский физико-технический институт, Московская область, г. Долгопрудный.