ОРИГИНАЛЬНЫЕ СТАТЬИ

Самарская Лука: проблемы региональной и глобальной экологии
2016. – Т. 25, № 1. – С. 13-17.

УДК 519.654

БЛЕСК И НИЩЕТА МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ (приглашение к дискуссии)

© 2016 Л.В. Недорезов

Центр междисциплинарных исследований по проблемам окружающей среды РАН,
г. Санкт-Петербург (Россия)

Поступила 15.01.2016

В работе обсуждаются проблемы использования метода наименьших квадратов (МНК) на практике. Утверждается, что метод базируется на неверных и необоснованных предпосылках, и целесообразно отказаться от его использования для решения практических задач. Предлагается подход к решению задач, не связанный с необходимостью минимизировать какой-либо функционал, и при этом лишенный всех указанных для МНК недостатков.
Ключевые слова: метод наименьших квадратов, анализ отклонений.

Nedorezov L.V. Shine and poverty of Ordinary Least Squares. – In publication problems of use of ordinary least squares (OLS) on a practice are under discussion. It is approved that OLS is based on incorrect and unfounded assumptions, and it is advisable to stop using it for solution of practical problems. One of possible approaches to solution of problems, which doesn't correlate with necessity to minimize any functional, and which haven't pointed out for OLS disadvantages, is presented.
Key words: ordinary least squares, analysis of deviations.

Метод наименьших квадратов (МНК) – один из базовых методов анализа данных различной природы. С его помощью были решены самые разнообразные задачи, что можно найти не только в научных монографиях (см., например, Худсон, 1970; Бард, 1979; Айвазян, Енюков, Мешалкин, 1985; Дрейпер, Смит, 1986, 1987 и мн. др.), но и учебниках (Ивченко, Медведев, 1984; Лакин, 1990 и др.). Поэтому нет необходимости лишний раз останавливаться на блистательных результатах, полученных с помощью МНК, но есть определенная необходимость остановиться на вопросах, относящихся к «нищете» данного метода, на тех проблемах, которые с неизбежностью возникают каждый раз, когда приходится использовать данный метод.

Пусть требуется по одномерному временному ряду $\{x_k^*\}$, $k = 1,2,...,N$, оценить значения параметров нелинейной динамической модели:

$$x_{k+1} = F(x_k, \vec{\alpha}), \quad (1)$$

Недорезов Лев Владимирович, доктор физико-математических наук, профессор, заведующий лабораторией l.v.nedorezov@gmail.com

где $\vec{\alpha}$ – набор параметров модели, F – нелинейная функция, x_k – значение переменной в k-ый момент времени. Эта проблема является типичной, например, для различных экологических задач (в частности, когда x_k – численность популяции в k-ый год; Корзухин, Семевский, 1992; Недорезов, Утюпин, 2011; Недорезов, 2012; McCallum, 2000).

Следуя общепринятым канонам, мы должны сначала построить некий функционал, например, в следующем виде:

$$Q(\vec{\alpha}) = \sum_{k=1}^{N-1} (x_{k+1}^* - F(x_k^*, \vec{\alpha}))^2 . \quad (2)$$

Каждый элемент суммы (2) – это отклонение значения, получаемого с помощью модели (1), от того, которое наблюдается в реальности. Для получения наилучших оценок параметров модели мы должны минимизировать значение функционала Q. Это мы принимаем на веру: нам представляется вполне естественным, что хорошая модель при хороших значениях параметров приводит к минимуму квадратов отклонений теоретических и экспериментальных (эмпирических) данных.

Здесь возникает два важных вопроса. Во-первых, какое отношение функционал (2) имеет к рассматриваемой (биологической, химической, физической…) задаче? И, во-вторых, какой вид должен иметь функционал (2)?

На первый вопрос ответ очевиден – никакого отношения к имеющейся проблеме функционал (2) не имеет. Он имеет отношение только к нашему представлению о хороших моделях, хороших оценках и только. Точнее, фраза должна была бы звучать так: «Нам кажется, что минимизация функционала (2) даст нам наилучшие значения оценок параметров модели. И все остальные возможные значения оценок существенно хуже.» Причем, ключевыми словами в этой фразе являются «Нам кажется…».

Ответ на второй вопрос несколько сложнее: анализ литературных данных показывает, что исследователи используют самые разные функционалы для нахождения оценок параметров. Действительно, откуда следует, что мы должны использовать квадраты отклонений? Следующий функционал выглядит ничуть не хуже функционала (2):

$$Q(\vec{\alpha}) = \sum_{k=1}^{N-1} \left| x_{k+1}^* - F(x_k^*, \vec{\alpha}) \right|^{\gamma} . \quad (3)$$

В (3) γ – положительное число, не обязательно равное 2. Если исследователи желают учесть влияние малых отклонений на получаемые оценки параметров, то в выражения (2) или (3) вводят так называемые «веса» (взвешенный МНК):

$$Q(\vec{\alpha}) = \sum_{k=1}^{N-1} w_k (x_{k+1}^* - F(x_k^*, \vec{\alpha}))^2 . \quad (4)$$

В (4) $w_k = const \geq 0$. Впрочем, эти веса могут зависеть и от величин отклонений. Если к указанным функционалам (2)-(4) добавить другие, в которых также используются отклонения теоретических и экспериментальных данных, но после различных нелинейных преобразований (в том числе, логарифмирования, двойного логарифмирования и др.), отклонения выборочных значений от траекторий модели (1) (глобальное приближение; Wood, 2001a, b) и так далее, то станет очевидным – никто не знает какой именно функционал следует использовать в том или ином конкретном случае. Как и неизвестно какими критериями следует пользоваться при выборе вида минимизируемого функционала.

Предположим, что для функционала (2) были найдены оценки значений параметров $\vec{\alpha}^*$, при которых функционал имеет глобальный минимум. На следующем шаге анализа соответствия модели и имеющихся данных проводится исследование совокупности отклонений $\{e_k\}$, где

$$e_k = x^*_{k+1} - F(x^*_k, \vec{\alpha}^*).$$

Многие исследователи полагают, что распределение $\{e_k\}$ должно быть Нормальным с нулевым средним, а в последовательности отклонений не должно быть сериальной корреляции (Бард, 1979; Дрейпер, Смит, 1986, 1987). Если при выбранном уровне значимости какое-либо требование нарушается, то модель признается непригодной для описания рассматриваемого временного ряда. В частности, наличие сериальной корреляции может быть связано с тем, что какой-то важный процесс не учитывается в модели (или учитывается неправильно), и модель необходимо модифицировать.

Иными словами, в рамках традиционного подхода мы по одной-единственной точке пространства параметров модели судим о пригодности или непригодности модели для аппроксимации данных. Тем более, что эта точка находится при минимизации функционала, не имеющего никакого отношения ни к задаче, ни к имеющимся данным, ни к модели.

Требование Нормальности распределения $\{e_k\}$ представляется одновременно и слишком сильным, и неверным. Если, к примеру, величины $\{x^*_k\}$ измеряются в граммах, то «промахнуться» при измеренях на пару тонн с положительной вероятностью или получить отрицательное значение представляются не просто маловероятными, а невозможными событиями. Интересно, каким нужно быть исследователем, чтобы получить отрицательный вес? Именно поэтому априорное предположение о Нормальности распределения представляется ненормальным.

Более естественным является предположение о симметрии (относительно нуля) отклонений $\{e_k\}$ (что означает, что ошибки в обе стороны происходят с одинаковыми вероятностями). Для проверки симметрии можно использовать критерии однородности двух выборок. Кроме этого необходимо потребовать, чтобы плотность распределения отклонений $\{e_k\}$ имела монотонно убывающую ветвь в области положительных значений аргумента (соответственно, монотонно возрастающую ветвь в области отрицательных значений). Это соответствует тому, что большие отклонения происходят с меньшими вероятностями.

Таким образом, из сказанного выше следует, что ко всем результатам, полученным с помощью традиционного метода наименьших квадратов (МНК), следует относиться с осторожностью. МНК – это логический тупик. Как ни улучшай «подъездные пути» к этому тупику (сколь ни улучшай асимптотические свойства оценок в тех или иных частных случаях) тупиком он и останется.

Где же выход из сложившегося тупика? Представляется вполне логичным отказаться от построения каких-либо функционалов типа (2)-(4). Вместо этого «просканировать» интересующую область пространства параметров следующим образом: каждой выбранной (например, случайным образом) точке пространства параметров соответствует вполне определенный набор отклонений $\{e_k\}$, если эти отклонения удовлетворяют неким требованиям (симметрии относительно нуля и др.), то такую точку считать принадлежащей допустимому множеству. Понятно, что это допустимое множество зависит от выбранных уровней значимости.

Если допустимое множество пусто, то это дает основание для утверждения о том, что модель непригодна для аппроксимации данных. Если множество не пусто, то среди элементов множества следует отыскивать такие, которые обладают экстремальными свойствами, то есть (с позиций статистических критериев) лучше других соответствуют имеющимся данным. Заметим, что при анализе реальных данных наблюдались ситуации, когда МНК-оценки не удовлетворяли статистическим критериям (симметрии, Нормальности, отсутствия сериальной корреляции), в то время как допустимое множество было не пусто. Более того, в некоторых случаях МНК-оценки были достаточно близки к допустимому множеству (Недорезов, 2012, 2015; Nedorezov, 2012, 2015).

Необходимо еще раз подчеркнуть – это лишь один из возможных вариантов отказа от использования МНК. Но при таком подходе устраняется еще один недостаток МНК. Если имеется несколько связанных временных рядов (например, временные ряды для численностей хищников и жертв), которые нужно использовать для оценки параметров модели (например, модели Лотки – Вольтерра), то и в этом, более сложном случае, можно обойтись без построения (минимизируемого) функционала. А именно, следовать основному принципу: точка пространства параметров модели принадлежит допустимому множеству, если обе соответствующие последовательности отклонений удовлетворяют выбранным статистическим критериям (Недорезов, 2016).

После того, как построено допустимое множество, нужно провести дополнительное исследование свойств элементов этого множества. Как было сказано выше, для оценок параметров модели целесообразно выбирать элементы допустимого множества, обладающие экстремальными свойствами.

Необходимо пояснить, что же именно подразумевается под словами «экстремальные свойства». Если при тестировании некоей гипотезы (например, о симметрии распределения) мы не можем отклонить нулевую гипотезу (о равенстве функций распределения положительных отклонений e_k и отрицательных отклонений, взятых со знаком минус) с уровнем значимости 5%, то это вовсе не означает, что мы обязаны принять нулевую гипотезу. Это означает только то, что означает: у нас нет оснований для отклонения. Для той же гипотезы более сильный результат наблюдается, если мы не можем отклонить гипотезу с 20% уровнем значимости. Наконец, если мы не можем отклонить нулевую гипотезу с 95% уровнем значимости, то это означает, что мы обязаны принять эту гипотезу.

Как легко догадаться, все те претензии к МНК, рассмотренные выше, имеют самое непосредственное отношение и к методу максимального правдоподобия…

СПИСОК ЛИТЕРАТУРЫ

Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей. М.: Финансы и статистика, 1985. 487 с.

Бард Й. Нелинейное оценивание параметров. М.: Статистика, 1979. 349 с.

Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Том 1. М.: Финансы и статистика, 1986. 366 с. – **Дрейпер Н., Смит Г.** Прикладной регрессионный анализ. Том 2. М.: Финансы и статистика, 1987. 351 с.

Ивченко Г.И., Медведев Ю.И. Математическая статистика. Учебное пособие для втузов. М.: Высшая школа, 1984. 248 с.

Корзухин М.Д., Семевский Ф.Н. Синэкология леса. СПб: Гидрометеоиздат, 1992. 192 с.

Лакин Г.Ф. Биометрия. М.: Высшая школа, 1990. 352 с.

Недорезов Л.В. Хаос и порядок в популяционной динамике: моделирование, анализ, прогноз. Саарбрюкен: LAP Lambert Academic Publishing, 2012. 352 с. – **Недорезов Л.В.** Аппроксимация временных рядов по динамике Paramecium caudatum моделями Ферхюльста и Гомпертца: нетрадиционный подход // Биофизика. 2015. Т. 60, вып. 3. С. 564-573. – **Недорезов Л.В.** Динамика системы «рысь-заяц»: применение модели Лотки-Вольтерра // Биофизика. 2016. Т. 61, вып. 1. С. 178-184. – **Недорезов Л.В., Утюпин Ю.В.** Непрерывно-дискретные модели популяционной динамики: аналитический обзор. Новосибирск: ГПНТБ СО РАН, 2011. 234 с.

Худсон Д. Статистика для физиков. М.: Мир, 1970. 296 с.

McCallum H. Population parameters estimation for ecological models. Brisbane: Blackwell Sciences, 2000. 348 p.

Nedorezov L.V. Gause Experiments *vs*. Mathematical Models. Population Dynamics: Analysis, Modelling, Forecast. 2012. V. 1, N 1. P. 47-58. – **Nedorezov L.V.** Paramecia aurelia dynamics: non-traditional approach to estimation of model parameters (on an example of Verhulst and Gompertz models) // Ecological Modelling. 2015. V. 317. P. 1-5.

Wood S.N. Minimizing model fitting objectives that contain spurious local minima by bootstrap restarting// Biometrics. 2001a. V. 57. P. 240-244. – **Wood S.N.** Partially specified ecological models// Ecological Monographs. 2001b. V. 71. P. 1-25.