МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

МАТЕРИАЛЫ

IV Международной молодежной научной конференции

«МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННЫХ, ТЕХНИЧЕСКИХ И ЭКОНОМИЧЕСКИХ СИСТЕМ»

Томск, 20–21 мая 2016 г.

Под общей редакцией
кандидата технических наук И.С. Шмырина

Томск
Издательский Дом Томского государственного университета
2016

Кривая E1 показывает поведение E при отклонении априорной догадки от истинного на 100 единиц; E2 — при отклонении на 1 единицу; E3 — на 0.5 единицы; E4 — на 0.1 единицы; E5 — нет отклонения.

На графике видно, что при любом из рассмотренных отклонениях центра симметрии от истинного существует выигрыш в точности в смысле СКО.

3. Построение адаптивной оценки

Предположим, что $F(x)$ неизвестна. Тогда оценим λ по имеющейся выборке и введем адаптивную оценку:

$$\lambda_n = \frac{F_n(x)(1-F_n(x)-F_n(2a-x))+F_n(\min(x,2a-x))}{2F_n(\min(x,2a-x))+1-F_n(x)-F(2a-x)+(n-1)(1-F_n(x)-F_n(2a-x))^2}, \quad (6)$$

$$\hat{F}_n^G(x) = F_n(x) - \lambda_n(F_n(x) - 1 + F_n(2\theta - x)). \quad (7)$$

Анализ оценки (7) будет проведен автором в последующих работах.

Заключение

В настоящей работе рассмотрена оценка функции с учетом симметрии при неизвестном центре, проиллюстрирован выигрыш в точности при использовании данной оценки по сравнению с эмпирической функцией распределения. Также построена адаптивная оценка функции распределения путем оценивания весового коэффициента по исходным данным.

ЛИТЕРАТУРА

1. *Shuster E.U.* Estimating the distribution function of a symmetric distribution. Biometrika (1975), 62, 3, p. 631–635.

2. *Дмитриев Ю.Г., Устинов Ю.К.* Статистическое оценивание распределений вероятностей с использованием дополнительной информации, Издательство Томского университета, Томск , 1988.

3. *Shuster E.U.* On the Goodness-of-Fit Problem for Continuous Symmetric Distributions. Journal of the American Statistical Association, September 1973, Volume 58, Number 343, p. 713–715.

РЕГРЕССИОННЫЕ МОДЕЛИ ЗАРАБОТНОЙ ПЛАТЫ В РОССИЙСКОЙ ФЕДЕРАЦИИ

Ю.Г. Дмитриев, П.А. Савченко
Томский государственный университет
dmit70@mail.ru, sapi95@ramble.ru

Введение

Одна из важнейших микроэкономических проблем, стоящих перед концепцией человеческого капитала — оценка влияния, оказываемого на величину текущих доходов (заработной платы) различными формами человеческого капитала: продолжительностью обучения, общим профессиональным опытом и так называемым специфическим человеческим капиталом — продолжительностью работы на определенной фирме. Решение данной задачи осуществляется на основе различных модификаций модели Дж. Минцера [1], впервые обоснованной в 1958 году:

$$\ln W = \beta_0 + \beta_1 SCH + \beta_2 EXP + \beta_3 EXP^2 + \beta_4 TEN + \beta_5 TEN^2,$$

где

- SCH — число лет обучения, скорректированное по достигнутому уровню образования (начальное и неполное среднее — 8 лет, полное среднее — 10 лет, профессиональное-техническое — 11,5 лет, среднее специальное — 13 лет, высшее — 15 лет; послевузовское (аспирантура) — 18 лет);
- EXP — потенциальный опыт на рынке труда (рассчитывается по условной формуле $EXP = \text{возраст} - SCH - 6$ лет);

- *TEN* – «специфический человеческий капитал» или профессиональный опыт, накопленный на данном рабочем месте (или с данным работодателем); *W* — заработная плата по основному месту работу; β_i – коэффициенты при соответствующих переменных, характеризующие норму отдачи от инвестиций в образование, профессиональный опыт и «специфический человеческий капитал» (β_0, β_1, β_2, β_4 имеют знак «плюс»; β_3 — знак «минус»).

Основным недостатком модели Минцера, отмечаемым многими исследователями, является постоянство коэффициента β_1, т.е. при увеличении длительности обучения *SCH* на один год заработная плата индивида должна меняться одинаково вне зависимости от того, какое образование он уже имеет. Однако можно предположить, что обучение в аспирантуре увеличивает заработную плату сильнее, чем обучение в техникуме [2].

В настоящей статье исследуется применимость модели Минцера (и её модификаций) для данных о жителях РФ за 2014 г., а также строится непараметрическая модель зависимости заработной платы от разных факторов по тем же данным и проводится сравнение полученных результатов.

1. Постановка задачи

Пусть *n* раз измерены значения факторов $x_1, x_2, x_3, ..., x_m$ и соответствующие значения переменной *Y*, тогда предполагается, что

$$y_i = \beta_0 + \beta_1 x_{i,1} + ... + \beta_k x_{i,m} + \varepsilon_i,\ i = \overline{1,n}$$

(второй индекс при *x* относится к номеру фактора, а первый — к номеру наблюдения). Также предполагается, что выполняются условия Гаусса-Маркова:

1) $M\varepsilon_i = 0,\ \forall i = \overline{1,n}$,

2) $M\varepsilon_i \varepsilon_j = \begin{cases} \sigma^2, & i = j, \\ 0, & i \neq j, \end{cases} \ \forall i, j = \overline{1,n}$,

3) $D\varepsilon_i = \sigma^2,\ \forall i = \overline{1,n}$.

На основании метода наименьших квадратов несмещенная оценка для вектора параметров β в матричной форме примет вид [3]: $\hat{\beta} = (X^T X)^{-1} X^T Y$, где X — матрица планирования эксперимента размерности $n \times (m+1)$:

$$X = \begin{bmatrix} 1 & x_{11} & ... & x_{1m} \\ ... & ... & ... & ... \\ 1 & x_{n1} & ... & x_{nm} \end{bmatrix}.$$

Для проверки качества уравнения регрессии будет использоваться коэффициент детерминации $R^2 = 1 - \frac{\|Y - \hat{Y}\|^2}{\|Y - \overline{Y}\|^2}$ (где $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$), который показывает качество подгонки регрессионной модели к наблюдаемым значениям. Если $R^2 = 0$, то регрессия Y на $x_1, x_2, x_3, ..., x_k$ не улучшает качество предсказания по сравнению с тривиальным предсказанием $\hat{y}_i = \overline{y}$. Другой крайний случай $R^2 = 1$ означает точную подгонку: все $\varepsilon_i = 0$, то есть все точки наблюдений лежат на регрессионной плоскости.

Еще одной важной характеристикой является средняя ошибка аппроксимации, которая показывает среднее отклонение расчетных значений от фактических [3]:

$$A = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|.$$

Непараметрическая регрессия, в отличие от параметрических подходов, использует модель, которая не описывается конечным числом параметров [4].

Идея ядерного сглаживания состоит в представлении последовательности весов $\{W_i(x)\}$, $i = \overline{1,n}$, где форма весовой функции $W_i(x)$ описывается посредством функции плотности со скалярным параметром h (ширина окна), который регулирует размер и форму весов около x. Эту функцию формы принято называть ядром K. Ядро — ограниченная, вещественная функция K с единичным интегралом.

Для сглаживания в многомерном случае в основном применяют произведение одномерных ядер. В этом случае ядерная оценка регрессии имеет вид:

$$\hat{r}(x) = \sum_{i=1}^{n} Y_i W_i(x) = \frac{\sum_{i=1}^{n} Y_i \prod_{j=1}^{m} K\left(\frac{x_j - X_{ij}}{h}\right)}{\sum_{i=1}^{n}\prod_{j=1}^{m} K\left(\frac{x_j - X_{ij}}{h}\right)},$$

где j — номер фактора.

2. Выбор оптимальных параметров для ядерной регрессии

Обычно в качестве ядерных функций рассматривают равномерное, треугольное, Епанечниково, Гауссово (нормальное) ядро. Важно отметить, что областью определения первых трех ядер является отрезок $[-1,1]$, в то время как последнее имеет бесконечный носитель. Следовательно, при использовании равномерного, треугольного или Епанечникова ядра оценка будет использовать информацию в ограниченном окне в окрестности X, а оценка, использующая гауссово ядро, будет использовать информацию из всех наблюдений. Исходя из этого, для расчета ядерной оценки регрессии заработной платы будет использоваться нормальное ядро [4]:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}}.$$

Ключом к проведению качественного непараметрического оценивания является выбор оптимальной ширины окна для поставленной задачи. Хотя ядерная функция остается важной, всё же главная её роль состоит в обеспечении дифференцируемости и гладкости получающихся оценок. Ширина окна, с другой стороны, определяет поведение оценки в конечных выборках, что ядерная функция сделать просто не в состоянии. Так как данные в имеющейся выборки масштабированы, то в данной работе параметр размытости ищется как минимум средней ошибки аппроксимации:

$$A = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i(h)}{y_i}\right| \to \min_h,$$

если $A \cdot 100\% \le 8\%$, то модель считается адекватной [5].

3. Результаты регрессионного анализа

В настоящей работе регрессионный анализ строится по данным 23 волны Российского мониторинга экономического положения и здоровья населения (РМЭЗ), где объем выборки равен 50 респондентам, проживающим в одном городе (Ленинградская область, Волосовский район). В табл. 1 представлены оценки параметров модели Минцера в общем виде ($\ln W = Y$, $EXP^2 = EXP2$, $TEN^2 = TEN2$), в табл. 2 — её характеристики, которые были вычислены в программе Mathcad Prime 3.0:

Таблица 1

МНК оценки параметров

	β_0	β_1	β_2	β_3	β_4	β_5
$\hat{\beta}$	9,341	0,044	0,00542	– 0,0003195	0,022	– 0,000352

Таблица 2

Характеристики модели №1 (Минцера)

	R^2	A
Характеристики модели	0,122	20,7%

В результате анализа полученных в табл. 1, 2 результатов делаем вывод, что общая модель Минцера не применима к имеющимся данным.

В табл. 3, 4 приведены результаты анализа модернизированной модели Минцера, в которой из набора факторов исключаются квадраты EXP^2, TEN^2:

$$\ln W_i = \beta_0 + \beta_1 SCH_i + \beta_2 EXP_i + \beta_3 TEN_i + \varepsilon_i,\ i = \overline{1,n}.$$

Таблица 3

МНК оценки параметров

	β_0	β_1	β_2	β_3
$\hat{\beta}$	9,497	0,052	0,011	0,008305

Таблица 4

Характеристики модели №2

	R^2	A
Характеристики модели	0,114	5,687%

Согласно полученным результатам можно сделать вывод, что модель Минцера без квадратов факторов (EXP^2 — потенциальный опыт работы, TEN^2 — профессиональный опыт) можно применять к имеющимся данным.

В данной статье также рассматривается непараметрический анализ данных по выборке (X,Y), где Y — вектор логарифмов заработной платы $\ln W$, а X — матрица факторов размерности $n \times m$, где $n = 50$ — количество наблюдений, $m = 3$ — количество факторов (*SCH*, *EXP*, *TEN*). На рис. 1 представлен график зависимости ошибки аппроксимации A от ширины окна h. По этому графику видно, что существует минимум функции, который достигается в искомом оптимальном h.

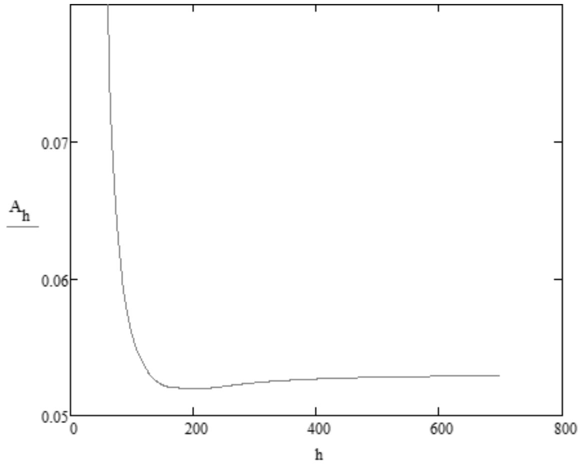

Рис. 1. График заивисмости ошибки аппроксимации от ширины окна

В табл. 5 представлены вычисленные значения оптимальной ширины окна и минимальной ошибки аппроксимации для имеющихся данных:

Таблица 5

Характеристики ядерной оценки

	h	A
Характеристики $\hat{r}(x)$	206	5,193%

Заключение

В работе рассмотрены параметрическая и непараметрическая регрессионные модели для анализа заработной платы в РФ от различных факторов. Сравнение результатов проведенного анализа показало, что ядерная оценка зависимости заработной платы является более приемлемой для имеющихся данных. Полученная ядерная оценка может применяться для восстановления данных по заработной плате при опросе в данной области путем прогнозирования.

ЛИТЕРАТУРА

1. *Саградов А.А.* Экономическая демография: учебное пособие // М. : Инфра, Москва, 2010, с. 135–137.
2. *Демидова А.А.* Непараметрический анализ зависимости заработной платы россиян от образования и опыта работы //Москва, ВШЭ
3. *Магнус Я.Р.* Эконометрика: начальный курс // Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий // Москва: Изд-во «Дело».
4. *Анатольев С.* Непараметрическая регрессия // журнал «Квантиль», 2009.
5. *Расин, Джефри* Непараметрическая эконометрика: вводный курс // журнал «Квантиль», 2008.

ИССЛЕДОВАНИЕ ТАРИФНЫХ СТАВОК ПРИ СТРАХОВАНИИ АВТОТРАНСПОРТА: СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Е.В. Жукова, Г.М. Кошкин
Томский государственный университет
water_melon93@mail.ru, kgm@mail.tsu.ru

Введение

Страхование — одно из наиболее гибких и стабильных видов деятельности по защите интересов субъекта, нарушаемых в результате непредвиденных негативных явлений. Эффективность финансовой деятельности страховой компании зависит от правильного расчета нетто-премий для различных видов страхования различных категорий населения [1]. В данной работе проводится анализ влияния количества заключенных договоров и страховых выплат на изменение страховых тарифов и прибыли страховой компании в автостраховании с помощью статистического моделирования.

1. Постановка задачи

Объектами страхования являются автомобили, застрахованные по КАСКО. Рассмотрим два отдела страховой компании (А и В). Количество заключенных договоров и страховых случаев соответствует данным реальных отделов компании “Росгосстрах”. Другие данные, связанные со страховыми суммами и выплатами, являются конфиденциальной информацией, поэтому анализ динамики КАСКО проводится с помощью статистического моделирования. По реальным средним величинам страховых сумм и выплат, которые равны 910739 руб. и 53634 руб. соответственно [2], в работе моделируются выборки по страховым суммам и выплатам, подчиняющиеся нормальному закону. Задача состоит в анализе динамики прибыли страховой компании при изменении различных факторов.