РЕГРЕССИОННЫЕ МОДЕЛИ ЗАРАБОТНОЙ ПЛАТЫ В ТОМСКОЙ ОБЛАСТИ

П.А. Савченко

Научный руководитель: доцент, д-р физ.-мат.наук, Ю.Г. Дмитриев

Томский государственный университет, Россия, г. Томск, пр. Ленина, 36, 634050

E-mail: sapi95@rambler.ru

REGRESSION MODELS OF THE SALARYIN THE TOMSK REGION

P.A. Savchenko

Scientific Supervisor: associate professor, Dr. Yu.G. Dmitriyev

Tomsk state university,, Russia, Tomsk, Lenin str., 36, 634050

E-mail: sapi95@rambler.ru

Abstract. *One of important microeconomic problems that the concept of human capital faces is assessment of the influence that different forms of human capital have on the amount of current income (wages). The solution to this problem lies in different modifications of J. Mincer's earning function,. The following paper examines applicability of Mincer earning function (and its modifications) to 2014 data about residents of Tomsk region. It also provides a non-parametric model of wage dependency on different factors using the same data and compares the produced results.*

Введение. Одна из важных микроэкономических проблем - оценка влияния, оказываемого на величину текущих доходов различными формами человеческого капитала. Решение данной задачи осуществляется на основе различных модификаций модели Дж.Минцера, 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; EXP – потенциальный опыт на рынке труда; TEN –профессиональный опыт, накопленный на данном рабочем месте; W –заработная плата по основному месту работу; β_i – коэффициенты при соответствующих переменных [1].В настоящей статье исследуется применимость модели Минцера (и её модификаций) для данных жителей Томской области за 2014 год. А так же строится непараметрическая модель зависимости заработной платы от разных факторов по тем же данным и проводится сравнение полученных результатов.

Постановка задачи. Рассматривается линейная множественная регрессия, назначение которой состоит в анализе связи между несколькими независимыми переменными (регрессорами) и зависимой переменной. Пусть n раз измерены значения факторов $x_1, x_2, x_3, ..., x_k$ и соответствующие значения переменной Y, тогда, предполагается, что $y_i = \beta_0 + \beta_1 x_{i,1} + ... + \beta_k x_{i,k} + \varepsilon_i, i = \overline{1,n}$. Так же предполагается, что выполняются условия Гаусса-Маркова. Тогда с помощью метода наименьших квадратов оценка для вектора параметров β примет вид (в матричной форме): $\hat{\beta} = (X^T X)^{-1} X^T Y$. Для проверки качества уравнения регрессия будет использоваться коэффициент детерминации $R^2 = 1 - \|Y - \hat{Y}\|^2 \big/ \|Y - \overline{Y}\|^2$, который показывает качество подгонки регрессионной модели к наблюдаемым значениям. Еще одной важной характеристикой является средняя ошибка аппроксимации, которая показывает среднее отклонение расчетных значений от фактических: $A = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i / y_i\right|$ [2].

Непараметрическая регрессия, в отличие от параметрических подходов, использует модель, которая не описывается конечным числом параметров. 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. В многомерном случае ядерная оценка регрессии имеет вид: $\hat{r}(x) = \sum_{i=1}^{n} Y_i W_i(x) = \sum_{i=1}^{n} Y_i \prod_{j=1}^{m} K(\frac{x_j - X_{ij}}{h}) \Big/ \sum_{i=1}^{n}\prod_{j=1}^{m} K(\frac{x_j - X_{ij}}{h})$, где j - номер фактора [3].

Выбор оптимальных параметров для ядерной регрессии. Обычно в качестве ядерных функций рассматривают равномерное, треугольное, Епанечниково областью определения которых является отрезок $[-1{,}1]$, в то время как Гауссово (нормальное) ядро имеет бесконечный носитель. Следовательно, только нормальное ядро будет использовать информацию из всех наблюдений. Исходя из этого, для расчета ядерной оценки регрессии заработной платы будет использоваться Гауссово ядро: $K(u) = \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}}$. Ключом к проведению качественного непараметрического оценивания является выбор оптимальной ширины окна для поставленной задачи. Хотя ядерная функция остается важной, её главная 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 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 $A*100\% \le 8\%$, то модель считается адекватной.

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 Томской области. В таблице 1 представлены оценки параметров модели Минцера в общем виде:

Таблица 1

МНК оценки параметров

	β_0	β_1	β_2	β_3	β_4	β_5
$\hat{\beta}$	9,644	0,027	0,006103	-0,0005051	0,023	-0,0003413

Характеристики модели №1: $R^2 = 0{,}120$, $A = 21{,}1\%$. В результате анализа полученных результатов делаем вывод, что общая модель Минцера не применима к имеющимся данным. В таблице 2 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 **2***

МНК оценки параметров

	β_0	β_1	β_2	β_3
$\hat{\beta}$	9,91	0,027	-0,02	0,009789

Характеристики модели №2: $R^2 = 0{,}112$, $A = 5{,}682\%$. Согласно полученным данным можно сделать вывод, что модель Минцера без квадратов факторов можно применять к имеющимся данным.

В данной статье так же рассматривается непараметрический анализ данных по выборке (X, Y), где Y-вектор логарифмов заработной платы $\ln W$, а X- матрица размерности $n \times m$, где $n = 50$-количество наблюдений, $m = 3$- количество факторов(SCH, EXP, TEN). На рисунке 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.

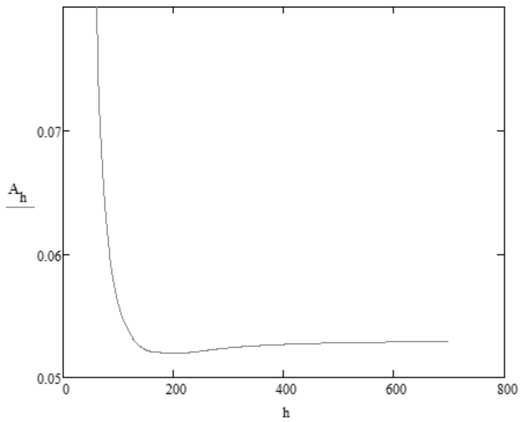

Рис.1. График заивисмости ошибки аппроксимации от ширины окна

Характеристики ядерной оценки: $h = 202$, $A = 4{,}1918\%$.

Заключение. В данной статье были рассмотрены два вида регрессионного анализа: параметрический и непараметрический. Сравнение полученных результатов дает возможность утверждать, что полученная ядерная оценка зависимости заработной платы от различных факторов является наиболее приемлемой для имеющихся данных и может применяться статистами для восстановления данных при опросе в Томской области.

СПИСОК ЛИТЕРАТУРЫ

1. Саградов А. А. Экономическая демография: учебное пособие - М.: Инфра, 2010. – 135 с.
2. Магнус Я.Р. Эконометрика: начальный курс – М.: Дело, 2004. – 74 с.
3. Анатольев С. Непараметрическая регрессия // Квантиль, 2009. – Т. 1. – № 7. – С. 37–51.