

РЕГРЕССИОННАЯ МОДЕЛЬ ОЦЕНКИ АВТОМОБИЛЕЙ С ПОМОЩЬЮ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

¹Трифонов Н.Ю., ²Ливинская В.А., ²Коржуков В.В.

¹БГЭУ, г. Минск, Республика Беларусь, n.trifonov@bntu.by

²Белорусско-Российский университет, г. Могилёв

Оценка стоимости автомобилей, находящихся на вторичном рынке, в настоящее время становится всё более востребованной. Это происходит, в первую очередь, из-за стремительного расширения этого рынка. Заказчиками в оценке выступают как физические лица (например, конкретный покупатель или продавец подержанного автомобиля) так и организации, нуждающиеся в оценке транспорта для различных целей. Информация об остаточной стоимости дорожного транспортного средства (автомобиля) востребована при его купле-продаже, постановке на учёт, оценке или переоценке основных средств предприятия, при передаче в залог, при оценке ущерба в результате дорожно-транспортного происшествия, при разводе супругов и иных имущественных спорах, в том числе для нужд судебной экспертизы [1-2]. Набирающая популярность программа покупки старых автомобилей по системе «trade-in» тоже предполагает владение актуальной рыночной информацией о его цене.

Развитость рынка подержанных автомобилей в странах ЕАЭС позволяет использовать для оценки статистические методы сравнительного подхода к оценке стоимости [2-3]. Выборки в несколько десятков объектов сравнения позволяли получать достаточно надёжные результаты. Тем не менее, развитие информационных технологий в части создания более объёмных баз данных на интернет-порталах и сайтах, а также появление инструментов формирования на их основе интернет-данных выборок с заданными характеристиками (т.н. парсеров). Современные технологии позволяют собирать информацию с сайтов-агрегаторов объявлений для её дальнейшего использования. Это позволило поставить задачу эконометрического моделирования рыночной стоимости подержанного автомобиля, обладающего конкретными характеристиками. Для этого необходимо рассмотреть представительные выборки автомобилей различных классов, поскольку ранее [3] было показано, что параметры обесценивания со временем существенно зависят от класса исследуемого автомобиля.

Данная статья посвящена описанию методики и результатам эконометрического моделирования средней цены на вторичном рынке на примере автомобилей, относящихся к одному из наиболее популярных классов - В (т.н. бизнес-класс).

Сбор первичной информации (raw data) занимает обычно до 70 % всего времени, потраченного на моделирование. В данном исследовании он осуществлялся с помощью парсера *Selenium WebDriver* – инструмента для сбора информации с сайта AUTO.ru, содержащего на момент сбора около 560 000 объявлений о продаже.

В результате анализа выборки из 17 742 объявлений (рассматривался город Москва, как наиболее интересный для белорусов сегмент российского рынка) было обнаружено, что 97,7% автомобилей всех классов, представленных на рынке, имеют возраст до 36 лет (с 1983 по 2019 годы) Результат группировки по классам и годам представлен в таблице 1. В дальнейшем анализ будет проводиться по объявлениям для автомобилей с 1983 по 2019 годы выпуска.

Распределение в выборке по возрасту следующее. Больше всего представлено автомобилей, возраст которых не превышает 7 лет (53 %). Автомобилей, возраст которых от 7 до 17 лет на рынке около 22%, от 17 до 27 лет - 18 % и от 27 до 37 лет всего 7 %.

Распределение в выборке по классам следующее. Больше всего на вторичном рынке присутствуют автомобили классов J (43%) и В (24%), причем 55 % автомобилей класса В, представленных на рынке имеют срок эксплуатации от 17 до 27 лет. Следующий по

представительности – класс С, 54 % автомобилей этого класса имеют срок эксплуатации от 7 до 17 лет, 41% авто имеют до 5 лет эксплуатации. Автомобили остальных классов, представленные на рынке в основном (80 % и выше) эксплуатировались до 7 лет.

Таблица 1 - Результат выборочного наблюдения объявлений о продаже автомобилей на вторичном рынке (единиц)

Год выпуска	Класс									Всего
	A	B	C	D	E	F	J	M	S	
1983-1991	7	528	301	207	79	10	37	11	2	1182
1992-2001	98	2138	411	427	48	18	56	16	9	3221
2002-2011	114	1465	784	151	89	95	960	78	57	3794
2012-2019	11	273	249	612	948	500	6013	340	184	9137
Всего	223	3876	1444	1190	1085	613	7029	434	250	16152

Далее описано моделирование цены автомобиля на примере объявлений о продаже автомобилей класса В. В качестве инструмента анализа используется программа *Statistica*.

Первый шаг в анализе данных – визуализация. Очевидно, цена авто на вторичном рынке сильно зависит от возраста автомобиля. На рисунке 1 представлено корреляционное поле для предиктора «цена» (в российских рублях) и одного из количественных регрессоров – «возраст» (автомобиля в годах) для класса В. В этом графике учтено, что из выборки были предварительно удалены результаты некоторых аномальных наблюдений для автомобилей возрастом до 10 лет.

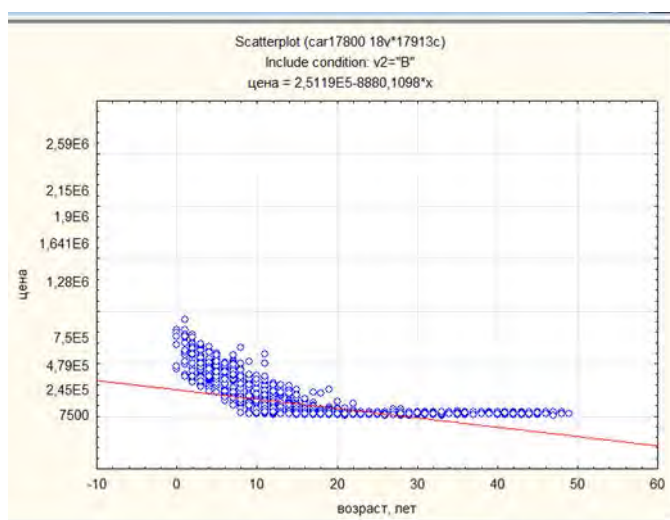


Рисунок 1 – Корреляционное поле «цена»-«возраст» для анализируемой выборки

Для отбора категориальных факторов строились частотные таблицы. Выяснилось: 53,3% автомобилей в выборке имеют задний привод и 46,53 % - передний привод; 99,5% автомобилей в выборке имеет механическую коробку передач (присутствовали также автоматическая и роботизированная коробки передач); 96,93% автомобилей в выборке имеют бензиновый двигатель.

В качестве инструмента моделирования применялся множественный регрессионный анализ, реализованный в программе Statistica. На первом этапе был произведен отбор наиболее значимых признаков, в модуле Data Mining, (feature selection). Он позволил сделать выводы о включения в модель тех же факторов, что и частотный анализ. Факторы были проранжированы по степени влияния на результативный признак (цена) с помощью F-

критерия. Наибольшее влияние оказывает фактор «страна производства», далее по степени влияния - «коробка передач», «возраст», «привод», «ЛС» (мощность двигателя в л.с.).

Для использования информации, содержащейся в категориальных факторах, использовался аппарат фиктивных переменных. Для доказательства статистических различий у объектов, соответствующих различным значениям категориальных признаков был использован программный модуль Nonparametrics.

Так, для разделения всей совокупности объявлений на два класса по возрасту (до 10 лет и более) была введена бинарная переменная Fage. На рисунке 2 приведен результат сравнения цены в полученных таким образом двух выборках с помощью непараметрического критерия Колмогорова. Нулевая гипотеза об отсутствии различий в средней цене авто обеих групп отвергается на уровне значимости 0,05.

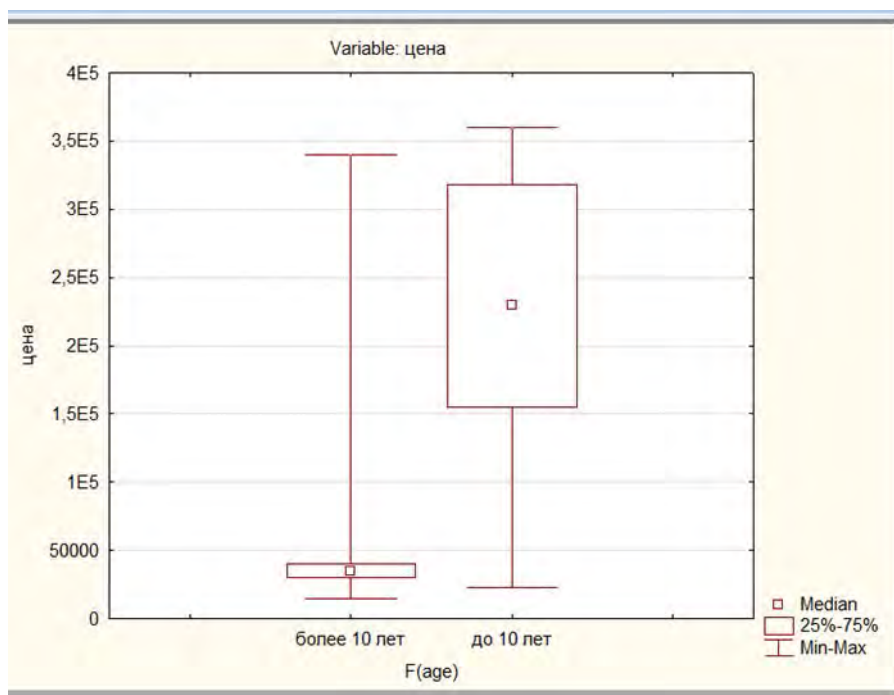


Рисунок 2 – Сравнение статистической значимости различия средней цены автомобилей в выборках с возрастом до и после 10 лет

Аналогичные исследования были проделаны и для других признаков (фиктивных переменных): «страна производства» (бинарная переменная, равная 1 при производстве автомобиля в России, 0 - в противном случае), «тип двигателя» (бинарная переменная Ftop, равная 1 для автомобиля с бензиновым двигателем, 0 – для автомобиля с дизельным), «коробка передач» (бинарная переменная FKOR, равная 1 для автомобиля с механической коробкой передач, 0 – в противном случае), «привод» (бинарная переменная FPRIV, равная 1 для автомобиля с задним приводом, 0 – в противном случае).

Поскольку возраст автомобиля – это единственный параметр, изменение которого вызывает изменение цены конкретной модели на вторичном рынке, для спецификации регрессионной модели было, помимо исследования связи «цена»-«возраст», дополнительно построено корреляционное поле для предиктора «логарифм цены» и регрессора «возраст». Окончательный выбор был сделан в пользу полулогарифмической модели. Результат оценки выборки автомобилей моложе 10 лет представлен на рисунке 3.

Summary for Dependent Variable: ln(цена) (dCAR)						
Regression Summary for Dependent Variable: ln(цена) (dCAR)						
R= ,92014658 R²= ,84666973 Adjusted R²= ,84404870						
F(6,351)=323,03 p<0,0000 Std.Error of estimate: ,28798						
Include condition: v6<=10						
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(351)	p-level
N=358						
Intercept			13,42952	0,302260	44,4304	0,000000
страна произ-ва	-0,242082	0,024407	-0,37753	0,038064	-9,9185	0,000000
возраст, лет	-0,384169	0,024506	-0,10439	0,006659	-15,6767	0,000000
Лс, силы	0,362663	0,026953	0,00906	0,000673	13,4556	0,000000
Ftop(бензин)	-0,056452	0,020966	-0,77892	0,289284	-2,6926	0,007430
FKOR(мех)	-0,071388	0,024940	-0,10968	0,038318	-2,8624	0,004457
F(PRIV(задний))	-0,253938	0,023573	-0,77000	0,071479	-10,7725	0,000000

Рисунок 3 – Регрессионное уравнение для средней цены автомобилей бизнес-класса на вторичном рынке со сроком эксплуатации менее 10 лет

Уравнение является статистически значимым ($p < 0,05$), Значение коэффициента детерминации $R^2 = 0,85$ показывает, что около 85% вариации средней цены автомобилей класса В возраста до 10 лет определяется вариацией выбранных факторов.

Для окончательного вывода о возможности использовать модель для предсказания средней цены на вторичном рынке, был проведен анализ остатков на нормальное распределение (в соответствии с предпосылками МНК Гаусса-Маркова). Гистограмма остатков, представленная на рисунке 4, позволяет не отвергать гипотезу о нормальном распределении остатков.

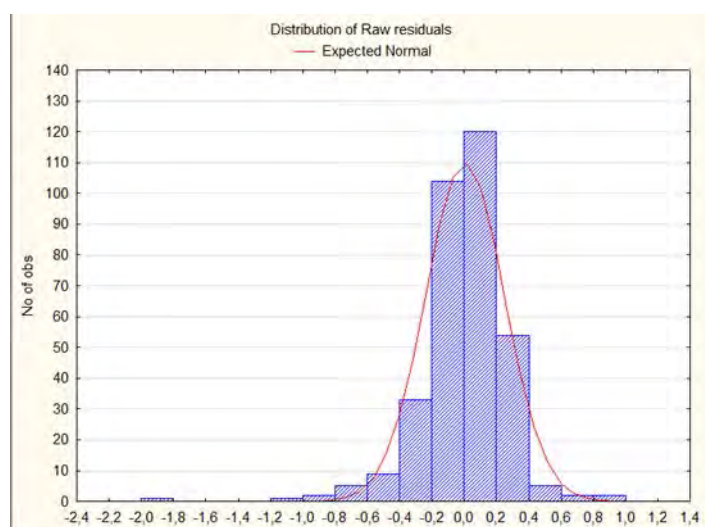


Рисунок 4 – Демонстрация нормальности распределения остатков

Также была проверена гипотезы об отсутствии гетероскедастичности. Проверка была выполнена с помощью графического анализа остатков. При этом зависимость между возрастом и вектором остатков не наблюдалась, что позволило не отклонять эту гипотезу.

Как итог, для предсказания цены автомобилей класса В с возрастом до 10 лет, согласно сложившейся конъюнктуре цен предложения, может быть использована следующая модель:

$$\ln \text{цена} = 13,43 - 0,38 \times \text{страна_произ} - 0,104 \times \text{возраст} + 0,09 \times \text{ЛС} - 0,78 \times \text{Ftop(бензин)} - 0,1096 \times \text{FKOR} - 0,77 \times \text{FPRIV(задний)}$$

При этом

$$\text{страна_произ} = \begin{cases} 1, & \text{если марка автомобиля производства РФ,} \\ 0, & \text{в противном случае;} \end{cases}$$

$ЛС$ – мощность двигателя в лошадиных силах ;

возраст - в годах (разность между 2019 и годом выпуска авто);

$$FKOR(\text{мех}) = \begin{cases} 1, & \text{если у автомобиля механическая коробка передач;} \\ 0, & \text{в противном случае.} \end{cases}$$

$$PRIV(\text{задний}) = \begin{cases} 1, & \text{если у автомобиля задний привод;} \\ 0, & \text{в противном случае.} \end{cases}$$

Согласно этому уравнению, ежегодно автомобиль этого класса теряет в цене 10,4 %. При этом автомобили производства РФ на 37,7% в среднем дешевле аналогичных иномарок. В выборку попали автомобили со средним возрастом 5 лет, средний объем двигателя 1,55л, средняя мощность 103 л.с. Было выполнено предсказание цены автомобиля с такими значениями характеристик, а также механической коробкой передач и передним приводом, произведенным в России. С вероятностью 0,95 логарифм цены для автомобиля с заданными характеристиками будет варьироваться от 12,51 (что соответствует цене 271837Р) до 12,63 (что соответствует цене 307208,4 Р).

Для автомобиля с теми же характеристиками, но иномарки ($\text{страна_произ} = 0$), с вероятностью 0,95 средняя цена на рынке варьируется от $EXP(12,902) = 401343,1$ Р до $EXP(13,00) = 442740,5$ Р.

Аналогичным образом проводилась оценка цены для автомобилей класса В в возрастном диапазоне от 10 лет до 20 лет. Для автомобилей класса В возраста от 10 до 20 лет значимыми оказались факторы страна производства, возраст, мощность двигателя, коробка передач, привод. Уравнение множественной регрессии в этом случае выглядит так:

$$\ln \text{цена} = 12,317 - 1,129 \times \text{страна_произ} - \text{ва} - 0,03 \times \text{возраст} + \\ + 0,006 \times ЛС - 0,5988 \times FKOR(\text{мех}) - 0,1 \times FPRIV(\text{задний})$$

Согласно данной модели, ежегодно автомобиль этого класса теряет в цене 3,5%. Страна производитель - РФ, механическая коробка передач и задний привод оказывают большее влияние на снижение цены, чем у более молодых машин.

Таким образом, с помощью современных информационных технологий для автомобилей класса В была получена эконометрическая модель рыночной стоимости в зависимости от возраста и других ценообразующих характеристик в виде двух регрессионных уравнений, использование которых в практике оценочной деятельности позволяет существенно повысить достоверность расчётов и уменьшить их трудоёмкость.

ЛИТЕРАТУРА

1. Оценка стоимости машин, оборудования и транспортных средств: учебник / А.П. Ковалев [и др.]. – Москва: Интерреклама, 2003. – 488 с.
2. Трифонов, Н.Ю. Теория оценки стоимости: учебное пособие / Н.Ю. Трифонов. – Минск: Вышэйшая школа, 2017. – 208 с.
3. Трифонов, Н.Ю. Характеристика накопленного износа автомобилей методами финансовой математики / Н.Ю. Трифонов, С.В. Скрыган // Белорусский экономический журнал. – 2014. – № 3. – С. 133-143.