

НИУ ВШЭ - САНКТ-ПЕТЕРБУРГ

ИССЛЕДОВАТЕЛЬСКИЙ ПРОЕКТ
9 КОМАНДА

ВЛИЯНИЕ МОЩНОСТИ ДВИГАТЕЛЯ НА СТОИМОСТЬ АВТОМОБИЛЯ

17 апреля 2024 г.

Верещагина Елизавета
Вшивцева Яна
Юрчик Анна

Содержание

Введение	2
Описание данных	2
Постановка гипотезы	8
Ожидаемые результаты	11
Результаты анализа	11
Анализ ограничений	15
Дискуссионная часть	16
Приложения	19

Введение

Рынку автомобилей присущ сильный разброс в ценах. Именно поэтому при выборе машины люди задумываются о характеристиках, которые могут повлиять на стоимость. Если технические характеристики важны покупателю, он будет ориентироваться на них и анализировать, готов ли он купить такой автомобиль за определенную стоимость. Это ведёт к желанию оценить, насколько больше будет стоимость автомобиля с лучшими характеристиками.

Мощность двигателя принято считать одной из основных технических характеристик автомобиля, именно поэтому можно предположить, что тем выше данный показатель, тем более привлекательна машина для покупателя. Кроме того, производство более мощного двигателя требует более продвинутых технологий, более квалифицированного труда и качественных деталей. Все эти ресурсы, безусловно, требуют дополнительных издержек. Большой спрос со стороны покупателей и издержки со стороны производителя или продавца, в свою очередь, положительно влияют на формирование стоимости товара. На основании этого рассуждения и было выдвинуто предположение о наличии взаимосвязи между стоимостью автомобиля и мощностью двигателя.

Кроме того, логично предположить, что для производства машины с автоматической коробкой передач требуются большие издержки и более качественный труд. Соответственно, эффект на стоимость от мощности автомобиля с автоматической трансмиссией должен быть больше эффекта автомобиля с такими же характеристиками, но механической коробкой передач.

Нас интересует следующий вопрос: увеличивается ли стоимость автомобиля при увеличении мощности двигателя? И усиливается ли этот эффект для автомобиля с автоматической коробкой передач по сравнению с такой же машиной с механической трансмиссией?

Описание данных

Данные для нашего исследования мы взяли с сайта auto.ru. Такой выбор сайта объяснён тем, что через него происходят продажи более полутора миллионов автомобилей каждый год. Для сбора данных было принято решение рассматривать 10 крупнейших городов России, в среднем по 3500 объявлений в каждом городе на начало 2022 года. Такое количество данных мы посчитали необходимым для того, чтобы выборка была достоверной. Благодаря этому подходу данные получились очищен-

ные от пропусков, потому что изначально собирались наблюдения только со всеми имеющимися значениями переменных. Ниже представлены переменные, имеющиеся в результате сбора данных.

- Cost - цена автомобиля (изначально в рублях, но для удобства мы делим переменную на 1000, чтобы получить цену в тысячах рублей).
- Mileage – пробег (в км).
- Year - год выпуска.
- Litres - объем двигателя (в литрах).
- Power - мощность двигателя (в лошадиных силах).
- Engine – переменная указывает на типы двигателя (бензин, дизель и т.д.).
- Transmission – переменная выражает типы трансмиссии (силовой передачи) автомобиля (механика, автомат и т.д.).
- Body – переменная указывает на типы кузова автомобиля (хэтчбек, седан и т.д.).
- Drive – переменная указывает на привод автомобиля (передний/задний/ полный).
- Colour – переменная указывает на цвет автомобиля (черный, желтый и т.д.).
- Mark - переменная отражает марку автомобиля (Audi, BMW и т.д.).

Среди переменных, которые влияют на цену автомобиля мы анализируем и качественные, и количественные.

Основное, что нужно было сделать для начала, - отредактировать данные и привести их к удобному для анализа виду. Это важный шаг для дальнейшего исследования, потому как в неочищенном от выбросов ряду аномальные наблюдения могут сильно изменить результат анализа – существенно завысить или занижить оценку. Чтобы минимизировать риски некорректного вывода, мы должны очистить его от выбросов и слишком высоких или слишком низких значений.

Посмотрим на визуальную составляющую переменной интереса при помощи ящика с усами, сделаем общую оценку наличия выбросов.

Из графика (Рис. 1) видны четкие границы начала выбросов у переменной cost.

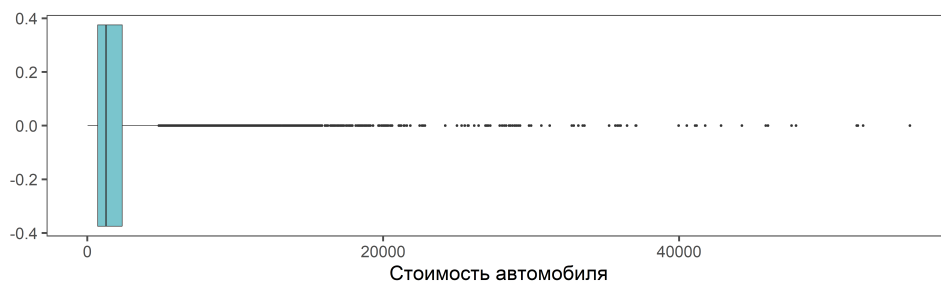


Рис. 1: Ящик с усами для зависимой переменной

Мы пришли к выводу о том, что данные по автомобилям, стоимость которых не превышает примерно 700 тыс. руб. или, наоборот, больше 5 млн., не могут быть использованы: автомобили со стоимостью менее 700 тыс. руб отличаются своей ненадёжностью и скудными характеристиками, также в большинстве случаев они являются поддержанными. По большей части именно эти факты влияют на их стоимость, а не наша переменная интереса. Что касается автомобилей, которые стоят больше, чем верхняя граница - данная категория автомобилей отличается тем, что большую роль в стоимости играет не основная характеристика автомобиля (в частности, мощность двигателя), а более кастомизированные её качества, например, бренд или уровень комфорта в салоне. Так как в нашей модели такие характеристики не являются переменными интереса, данные этих категорий автомобилей мы в рамках исследования использовать не будем.

После избавления от выбросов и аномальных значений у переменной интереса и переменной litres с данными всё хорошо, поэтому больше ничего не удалялось.

Описательные статистики до избавления от выбросов представлены в *Таблице*.

В нашей работе мы будем смотреть на значения между 25 квартилем и верхней границей (значение, на 1.5 межквартильного размаха большее 3-го квартиля), то есть отбросим 25% наименьших значений, и определенную часть наибольших значений по переменной cost. Далее будем рассматривать данные в интервале между этими значениями.

Кроме того, максимальное значение переменной litres - 795. Очевидно, это значение не в единицах измерения большинства наблюдений или неверно введенное значение. Понятно, что объём двигателя обычно варьируется в пределах 2-5 литров (с учётом вариации характеристик автомобилей, значения будут в пределах 1-7), поэтому мы убираем все наблюдения со значением этой переменной более 7.

Далее, чтобы было проще работать с категориальной переменной, вы-

делим классы. В результате one-hot кодирования мы получили 4 dummy-переменные: automatic, manual, robot, variator. Для того, чтобы в регрессии использовать только первые две, мы сложили индикаторы последних и получили dummy other, которую далее будем использовать в качестве базового класса. Это было сделано, чтобы оставить в модели только automatic и manual и сравнивать эффекты от этих трансмиссий.

Посмотрим, как изменились основные описательные статистики (Таблица 1).

Переменная	Среднее	Стандартное отклонение	Медиана	Количество наблюдений
Мощность двигателя	160.34	72.22	144.00	16 973
Стоимость	1849.97	991.55	1540.00	16 973
Объём двигателя	2.05	0.78	2.00	16 973
Пробег	104 521.59	75 032.03	94 000.00	16 973

Таблица 1: Описательные статистики для непрерывных переменных

Разница видна по следующим параметрам:

- Довольно сильно изменилось стандартное отклонение каждой из переменных: у стоимости больше, чем в 3 раза
- Сократилась разница между медианным и средним значением, что способно привести к увеличению достоверности полученных корреляций.
- Размах выборки сильно сократился.

Также мы выбрали контрольные переменные: выделили те из них, которые, по нашему мнению, имеют ненулевую ковариацию с переменной интереса (power) и влияют на зависимую переменную (cost). Для подтверждения предположений о зависимостях построены точечные графики непрерывных переменных с зависимой и с переменной интереса, а также ящики с усами, чтобы посмотреть на различия в распределениях между группами (см. Приложение).

Зависимая переменная:

- Cost - цена автомобиля (в тысячах рублей).

Переменная интереса:

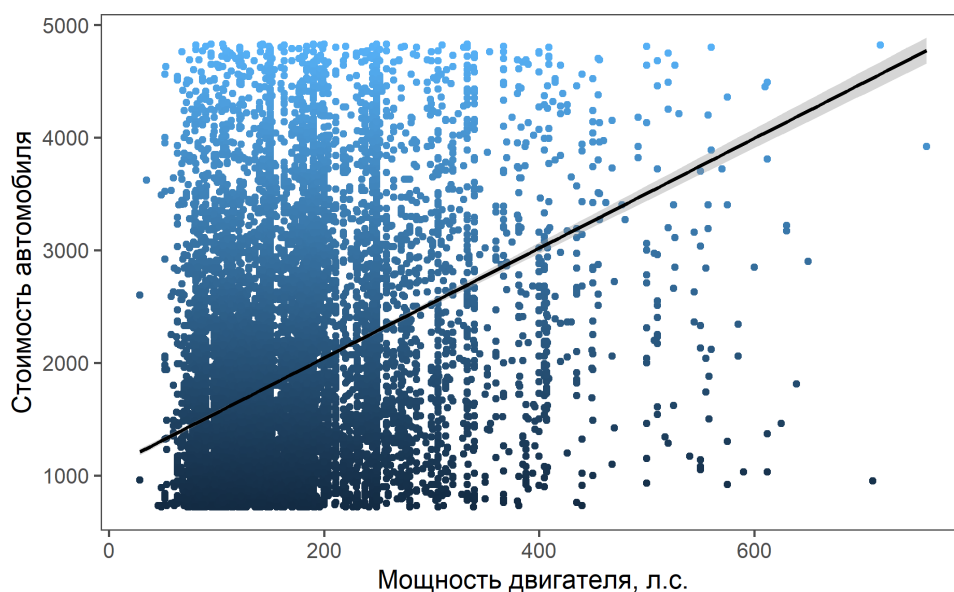


Рис. 2: Точечный график основной зависимости

- Power - мощность двигателя (в лошадиных силах).

Переменные для исследования перекрёстного эффекта:

- automatic - индикаторная переменная для обозначения автоматической коробки передач.
- manual - индикаторная переменная для обозначения механической коробки передач.
- other - индикаторная переменная для обозначения коробки передач, отличной от автоматической и механики (этот класс далее будет использован в качестве базового).

automatic	manual	other	Всего
7 191	6 545	3237	16973

Other включает в себя 1475 наблюдений variator и 1762 robot

Таблица 2: Количество наблюдений по каждому классу трансмиссии

По точечным графикам можно увидеть, что на данных изначальное предположение подтверждается. На графике основной зависимости (Рис.

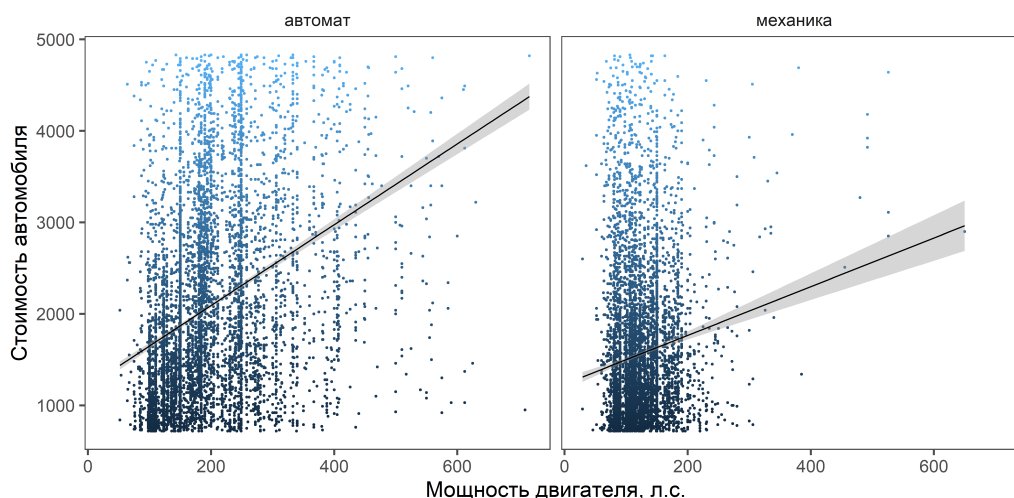


Рис. 3: Точечный график основной зависимости по трансмиссии

2) видно, что линия тренда имеет положительный наклон. При этом на графиках этой зависимости, построенных для каждой из коробок передач (Рис. 3), видна разница между наклонами для автомата и механики.

Кроме того, посмотрим на описательные статистики по механике (Табл. 3) и по автомату (Табл. 4). Видно, что среднее и медиана по стоимости и мощности выше у автомобилей с автоматической коробкой передач.

	Среднее	Стандартное отклонение	Медиана
Мощность двигателя	160.34	72.22	144
Стоимость	1849.97	991.55	1540.00
Объём двигателя	2.05	0.78	2.00
Пробег	104 521.59	75 032.03	94 000

Таблица 3: Описательные статистики при автоматической трансмиссии

Контрольные переменные:¹

- Mileage – пробег (в км): предполагается, что пробег уменьшает стоимость автомобиля, потому что показывает величину износа транс-

¹Описание того, как контрольные категориальные переменные будут использоваться в дальнейшем исследовании, описано в *приложении*.

	Среднее	Стандартное отклонение	Медиана
Мощность двигателя	158.01	84.35	136.00
Стоимость	2064.16	2850.50	1270.00
Объём двигателя	3.13	23.28	1.80
Пробег	116 547.49	84 077.02	106 000.00

Таблица 4: Описательные статистики при механической трансмиссии

портного средства. К тому же пробег может способствовать изнашиванию двигателя и, соответственно, влиять на его мощность.

- Litres - объем двигателя (в литрах): более объёмные двигатели считаются и более мощными, а также на больший объём двигателя требует больших затрат на производство автомобиля, а значит, и его цена будет выше.
- Engine – указывает на типы двигателя (бензин, дизель и т.д.): разные двигатели имеют разные цены, а размах этих цен довольно большой, а значит, различия в этих характеристиках будут существенно влиять на цену автомобиля. Кроме того, мощность разных двигателей различна, то есть характеристика влияет на интересующую нас переменную.
- Drive – указывает на привод автомобиля (передний/задний/полный): привод автомобиля позволяет управлять им более свободно, поэтому и стоимость такого средства увеличивается. Известно, что из-за структуры "внутренностей" автомобиля в переднеприводную машину невозможно поставить по-настоящему мощный двигатель, поэтому мы предполагаем, что привод влияет на переменную интереса.

Постановка гипотезы

Основное регрессионные уравнение²:

$$cost_i = \beta_0 + \beta_1 * power_i + \beta_2 * power_i * automatic_i + \beta_3 * power_i * manual_i + v_i^3$$

²Далее при тестировании гипотез будут добавлены контрольные переменные, здесь записана лишь основная зависимость.

³С учётом всех необходимых предпосылок, которые представлены в приложении. Далее будем считать, что они выполнены для каждой из используемых регрессий.

Мы предполагаем, что при прочих равных средняя стоимость автомобиля увеличивается при увеличении мощности двигателя. Кроме того, мы считаем, что наличие автоматической коробки передач влияет на цену сильнее, чем наличие механической. То есть нас интересует следующий вопрос: действительно ли при прочих равных при увеличении мощности двигателя средняя стоимость автомобиля с автоматической коробкой передач увеличивается в большей степени, чем автомобиля с механической коробкой передач?

Для подтверждения этого предположения нам необходимы следующие условия:

1. Коэффициент при переменной интереса значимо больше нуля.
2. Хотя бы один из коэффициентов основной модели значимо отличен от нуля.
3. Коэффициент при перекрестном эффекте автомата больше коэффициента при перекрестном эффекте механики.
4. Предельный эффект мощности увеличивается, когда трансмиссия автомобиля сменяется с механики на автомат.

Сформулируем гипотезы в модельной форме:⁴

$$1) E\{cost_i | power_i = power_1 + 1, \\ automatic_i = automatic, manual_i = manual, \dots\} - \\ - E\{cost_i | power_i = power_1, \\ automatic_i = automatic, manual_i = manual, \dots\} > 0$$

2) Гипотезу в модельной форме сформулировать не получится.

$$3) (E\{cost_i | power_i = power_1 + 1, automatic_i = 1, manual_i = 0, \dots\} - \\ - E\{cost_i | power_i = power_1, automatic_i = 1, manual_i = 0, \dots\}) > \\ (E\{cost_i | power_i = power_1 + 1, automatic_i = 0, manual_i = 1, \dots\} - \\ - E\{cost_i | power_i = power_1, automatic_i = 0, manual_i = 1, \dots\})$$

⁴Многоточием будем заменять следующее: $litres_i = litres, mileage_i = mileage, dizel_i = dizel, gaz_i = gaz, gibrid_i = gibrid, FullDr_i = FullDr, FrontDr_i = FrontDr$.

$$4) \left. \frac{\partial E\{cost_i\}}{\partial power_i} \right|_{automatic_i=1} > \left. \frac{\partial E\{cost_i\}}{\partial power_i} \right|_{manual_i=1}$$

И будем проверять следующие гипотезы в нормальной форме:

$$1. \begin{aligned} H_0 : \beta_1 &= 0 \\ H_1 : \beta_1 &> 0 \end{aligned}$$

$$2. \begin{aligned} H_0 : \{\beta_1 = 0, \beta_2 = 0, \beta_3 = 0\} \\ H_1 : H_0 \text{ неверна} \end{aligned}$$

$$3. \begin{aligned} H_0 : \beta_2 &= \beta_3 \\ H_1 : \beta_2 &> \beta_3 \end{aligned}$$

$$4. \begin{aligned} H_0 : ME(power_i)|_{automatic_i=1} &= ME(power_i)|_{manual_i=1} \\ H_1 : ME(power_i)|_{automatic_i=1} &> ME(power_i)|_{manual_i=1} \end{aligned}$$

Предельные эффекты имеют вид

$$\begin{aligned} ME(power)|_{automatic_i=1} &= \beta_1 + \beta_2 \\ ME(power)|_{manual_i=1} &= \beta_1 + \beta_3 \end{aligned}$$

Заметим, что

$$\begin{aligned} ME(power_i)|_{automatic_i=1} - ME(power_i)|_{manual_i=1} &= \\ \beta_1 + \beta_2 - (\beta_1 + \beta_3) &= \beta_2 - \beta_3 \end{aligned}$$

Тогда гипотеза сводится к предыдущей:

$$\begin{aligned} H_0 : \beta_2 - \beta_3 &= 0 \\ H_1 : \beta_2 - \beta_3 &> 0 \end{aligned}$$

И гипотеза о разнице в предельных эффектах эквивалентна гипотезе о разнице коэффициентов.

Ожидаемые результаты

В качестве результатов нашей работы мы ожидаем увидеть подтверждение гипотезы о том, что в среднем при прочих равных стоимость автомобиля увеличивается при увеличении мощности двигателя. То есть коэффициент при переменной интереса “power” должен быть не равен нулю и значим на определенном уровне, нулевая гипотеза об отсутствии взаимосвязи должна быть отвергнута.

Более того, мы включили в модель следующие перекрестные эффекты: $power \cdot automatic$ и $power \cdot manual$. Мы добавили компоненты взаимодействия в регрессию, так как предполагаем, что трансмиссия влияет на отдачу от мощности на цену автомобиля. То есть ожидаемые коэффициенты при этих переменных также должны быть значимы, и вместе с тем предельный эффект мощности у машины с автоматической коробкой должен быть больше предельного эффекта мощности у автомобиля с автоматической.

Чтобы сделать корректные содержательные выводы о влиянии рассмотренных регрессоров, нелинейные спецификации модели должны быть дополнены контрольными переменными. Они в некоторой степени сместят коэффициенты при переменных интереса. Это будет говорить о важности выбранных контрольных переменных, которые помогут более точно оценить эффекты.

Результаты анализа

Рассматриваемые спецификации:

- Модель 1: $cost_i = \beta_0 + \beta_1 * power_i + v_i$,
- Модель 2: $cost_i = \beta_0 + \beta_1 * power_i + \beta_4 * litres_i + \beta_5 * mileage_i + \beta_6 * dizel_i + \beta_7 * gaz_i + \beta_8 * gibrid_i + \beta_9 * FullDr_i + \beta_{10} * FrontDr_i + v_i$,
- Модель 3: $cost_i = \beta_0 + \beta_1 * power_i + \beta_2 * power_i * automatic_i + \beta_3 * power_i * manual_i + v_i$,**
- Модель 4: $cost_i = \beta_0 + \beta_1 * power_i + \beta_2 * power_i * automatic_i + \beta_3 * power_i * manual_i + \beta_4 * litres_i + \beta_5 * mileage_i + \beta_6 * dizel_i + \beta_7 * gaz_i + \beta_8 * gibrid_i + \beta_9 * FullDr_i + \beta_{10} * FrontDr_i + v_i$,**

Таблица 5: Результаты оценивания

	Обозначения в модели	Модель 1	Модель 2	Модель 3	Модель 4
Константа		1068.524*** (19.342)	1582.174*** (44.315)	1188.241*** (22.227)	1672.305*** (45.176)
Мощность двигателя в лошадиных силах	power	4.874*** (0.122)	4.615*** (0.236)	4.910*** (0.177)	4.399*** (0.250)
Мощность*Автомат	power·automatic			−0.410** (0.129)	−0.070 (0.123)
Мощность*Механика	power·manual			−1.917*** (0.136)	−1.290*** (0.130)
Контрольные переменные		Нет	Да	Нет	Да

В скобках даны робастные стандартные ошибки

** p < 0.01, *** p < 0.001

Встроенные пакеты считают p-value для двусторонней альтернативы, когда мы хотим проверить гипотезу с односторонней. Поэтому проведём расчеты вручную.

Для проверки первой гипотезы воспользуемся односторонним z-тестом. Посчитаем t-статистику⁵.

$$t_{st} = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \approx 17.6,$$

$$1 - \Phi_0^{-1}(17.6) \approx 0$$

Можно сделать вывод, что коэффициент при переменной интереса значимо больше нуля и наше предположение подтвердилось: мощность действительно увеличивает среднюю стоимость автомобиля при прочих равных.

Чтобы проверить вторую гипотезу, обратимся к тесту Вальда и посчитаем w - статистику.

$$w_{st} = 549.57,$$

$$1 - \chi_3^2(549.57) \approx 0$$

Тест Вальда показал, что можно отвергнуть модель с ограничениями, а значит, коэффициенты при переменной интереса и перекрестных эффектах значимо отличны от нуля. У нас есть ещё одно подтверждение верности поставленной гипотезы. Однако для гарантии правильности

⁵Все расчёты представлены в коде в Приложении

всего предположения нужно проверить также третью из сформулированных гипотез: гипотезу о разнице между коэффициентами. Для этого нам снова нужно посчитать t-статистику:

$$t_{st} = \frac{(\hat{\beta}_2 - \hat{\beta}_3) - (\beta_2 - \beta_3)}{SE((\hat{\beta}_2 - \hat{\beta}_3))} = \frac{\hat{\beta}_2 - \hat{\beta}_3}{SE(\hat{\beta}_2 - \hat{\beta}_3)}$$

$$\begin{aligned} SE((\hat{\beta}_2 - \hat{\beta}_3)) &= \sqrt{Var((\hat{\beta}_2 - \hat{\beta}_3))} = \\ &= \sqrt{Var(\hat{\beta}_2) + Var(\hat{\beta}_3) - 2cov(\hat{\beta}_2; \hat{\beta}_3)} \approx 0.23 \end{aligned}$$

$$t_{st} \approx 5.24$$

$$1 - \Phi_0^{-1}(5.24) \approx 0$$

Гипотеза о различиях в коэффициентах подтвердилась (заметим, что тогда и гипотеза о различиях в предельных эффектах в таком случае подтверждается). То есть мы можем говорить об увеличении эффекта.

Трансмиссия	Средний предельный эффект	p-value	Нижняя граница доверительного интервала	Верхняя граница доверительного интервала
Механика	3.11	0.00	2.59	3.63
Автомат	4.33	0.00	3.86	4.80

Таблица 6: Средний предельный эффект для переменной интереса

Посмотрим на предельные эффекты от трансмиссий. Как видно в таблице (Таблица 4), p-value для каждого из эффектов принимает довольно низкие значения, поэтому можно говорить об их значимости на уровне менее 1%. Более того, они сильно различаются даже при ошибке, что видно на графике (Рис. 3).

Все сделанные предположения подтвердились, то есть эффект мощности на цену можно считать значимым, притом есть влияние коробки передач на оценку влияния мощности на стоимость.

Основные выводы

Итак, при оценке параметров регрессии мы получили результаты, которые ожидали увидеть. При этом введение контрольных переменных дало ощутимый эффект на оценку коэффициентов. Контрольные

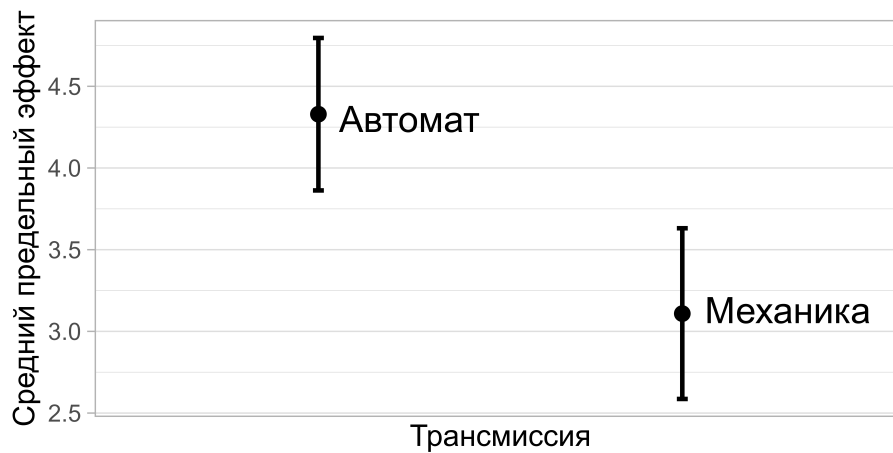


Рис. 4: Средний предельный эффект для переменной интереса

переменные уменьшают оценку параметра при переменной интереса, что говорит о следующем: в моделях без контрольных переменных параметр переоценен. Чтобы минимизировать смещение оценки от реального значения исследуемого влияния, мы должны включить эти переменные в модель, поэтому далее мы будем интерпретировать модель с контрольными переменными. Именно она позволяет наиболее качественно исследовать влияние переменной интереса на двух группах машин: с механической передачей и автоматом. Оба параметра отрицательны, но кросс-эффект для механики сильнее.

Анализируя выходные данные, можно заключить, что увеличение мощности двигателя приводит к росту стоимости автомобиля: при прочих равных при увеличении мощности двигателя на единицу, средняя стоимость автомобиля увеличивается \approx на 4.5 тыс. руб. Этот эффект наблюдается для коробок передач Робот и Вариатор. Эти трансмиссии являются наиболее модернизированными, и стоимость на них, соответственно, в среднем выше, чем у таких же автомобилей с автоматической и механической коробками передач. Поэтому эффект от последних меньше, чем в среднем.

Посмотрим на коэффициенты перед перекрестными эффектами: механическая коробка передач уменьшает эффект от мощности на 1.29 тыс. руб., а автоматическая - на 0.07 тыс. руб. То есть можно заключить, что итоговый эффект от мощности автомобиля следующий: автомобиль с автоматической трансмиссией в среднем при изменении мощности на единицу стоит больше на 4.329 тыс.руб. при прочих равных, а автомобиль с механической коробкой передач - на 3.109 тыс. руб. больше.

Можно сделать вывод, что в зависимости от типа коробки передач влияние мощности машины на ее цену изменяется: автоматическая трансмиссия даёт больший эффект, чем механическая.

Анализ ограничений

Для рассмотренных моделей существуют определенные ограничения.

Во-первых, мы использовали данные по *российскому рынку автомобилей на начало 2022 года* в качестве генеральной совокупности, то есть перенесение результатов на рынки других стран или другой временной промежуток может нести угрозу внешней валидности. Ценообразование на рынках разных стран различается, и численные зависимости могут быть другие. Кроме того, несколько лет назад многие из автомобилей, представленных в нашем исследовании, могли быть более востребованы (а некоторые, возможно, будут более востребованы через какое-то время) на рынке подержанных авто, чем сейчас, а значит, и стоять будут дороже. Это связано с развитием технологий, модой на определенные модели и возможностями покупки моделей на рынке новых машин.

Говоря о рынке новых автомобилей, стоит отметить: он имеет иное ценообразование в сравнении с *рынком подержанных машин, который мы рассматривали* в рамках исследования. На рынке подержанных автомобилей цены сильно ниже тех, что устанавливает производитель для новых моделей, поэтому результаты оценки изменятся.

Кроме того, в качестве контрольных переменных мы использовали тип и объём двигателя, пробег и привод автомобиля. В качестве пропущенной переменной также можно рассмотреть квалификацию работника, занятого на производстве конкретного предприятия. Этот показатель может влиять на цену машины и на мощность ее двигателя: более квалифицированные работники могут разработать и собрать более качественный и сложный по функциональности двигатель, что ведёт и к более высокой мощности. При этом хорошие специалисты требуют более высокой зарплаты, соответственно, издержки на производство автомобиля будут выше, себестоимость выше и цена при перепродаже также будет больше. Однако нельзя не заметить, что квалификацию работника мы не можем использовать в исследовании: численно измерить это довольно сложно. Конечно, можно посмотреть на качество образования сотрудников, но таких данных у нас нет. С учётом того, что оба влияния (и на стоимость, и на мощность) от переменной положительны, коэффициент в итоговой регрессии смещён в большую сторону.

Далее, в исследовании использованы автомобили, произведенные в

России, и иномарки, собранные за рубежом и привезенные в страну. Рассматривая цены на них, мы не учитываем налоги, не делаем поправки на валютный курс и инфляцию. То есть мы используем цену, сложившуюся в результате влияния внешних факторов. Из-за этого результаты исследования могут быть смещены, и неясно, в какую сторону.

Нельзя не обратить внимание на то, что многие при перепродаже сбрасывают пробег автомобиля, который мы используем в качестве контрольной переменной. Это значит, что какие-то из наблюдений вполне могут содержать неверное фактическое значение этого показателя. В таком случае эффект контрольной переменной приносит некорректную поправку коэффициента при переменной интереса.

Говоря о самой переменной интереса, важно отметить: мы предполагаем, что мощность измерена на момент продажи (с учётом всех характеристик, в том числе возраста автомобиля и пробега). Строго говоря, это необязательно так.

Еще одной угрозой может являться и то, что в нашем исследовании не рассматривается «наценка за бренд». Многие крупные производители автомобилей, такие как BMW или Mercedes-Benz, завышают цену на свои продукты, и это связано не с мощностью машины. Это, безусловно, скажется на рынке подержанных автомобилей и даст определенный эффект. Однако тяжело оценить это, используя инструменты, которые у нас есть, так как мы не знаем, какой конкретно процент от стоимости является наценкой производителя.

Наконец, мы использовали выборку из автомобилей, среди которых есть и старые машины, произведенные достаточно давно, в 90-ые годы, и относительно новые, произведенные в 2000-ые годы. Сложно не согласиться с тем, что в разные времена уровни технического прогресса и пределы качества различались. Это означает, что высокий для тех времён уровень технологий стал средним в современном контексте. Следовательно, раньше отдача от единицы мощности могла быть иной, и сложившаяся на современном рынке цена должна быть пересчитана с учётом инфляции. Но с учётом мотивации работы нам необходимы реальные числа, установленные людьми на рынке. Поэтому описанная угроза не является серьёзной проблемой в рамках исследования.

Дискуссионная часть

Итак, результаты исследования показали, что мощность положительно влияет на стоимость автомобиля. Действительно, производство более мощного двигателя требует более продвинутых технологий, более ква-

лифицированного труда и качественных деталей. Все эти ресурсы, безусловно, требуют дополнительных издержек, что увеличивает стоимость продукта.

Притом наш анализ строился на следующем: у автомобиля с автоматической трансмиссией в среднем эффект от мощности больше, чем у такого же автомобиля с механической коробкой передач. Это, в свою очередь, объясняется рядом причин: во-первых, механическая коробка имеет более простую конструкцию, а соответственно, и меньшее количество деталей и используемых механизмов, поэтому отдачу на цену при перепродаже иметь будет относительно небольшую. Во-вторых, она более экономична. Это происходит из-за специфики автоматической коробки. Она имеет сложную конструкцию, способствующую более эффективному использованию топлива. Данный факт объясняет ее большую стоимость по сравнению с механической. И наконец, автоматическая коробка упрощает процесс управления автомобилем: нет сцепления, передачи переключаются без вмешательства водителя. То есть во многом автомобиль делает работу за человека, что требует высокой технологичности, детализированности и повышает стоимость такого автомобиля.

Полученные в исследовании результаты представляют лишь часть возможных взаимосвязей. Как было упомянуто ранее, исследование валидно для российского рынка, поэтому мы не можем перенести полученные результаты на рынки других стран. При этом их изучение может быть полезным не только с точки зрения возможного нахождения изученной связи, но и для сравнения рынков разных стран.

Расширение исследования может заключаться в применении спецификаций данной модели в исследовании зарубежных автомобильных рынков, при этом важно учитывать специфику отдельного рынка. Тогда можно будет выяснить, насколько обнаруженная связь актуальна для других регионов и в каком контексте она может существовать.

Расширить исследование возможно и в рамках Российского рынка. Возможные экономические и политические изменения текущего года сильно влияют на рынки крупного производства, в том числе и в автомобильной сфере. Уход иностранных поставщиков, увеличение отечественного производства, рост цен на детали и комплектующие зарубежных марок меняют не только представленный ассортимент автомобилей, но и предпочтения покупателей, что в конечном счете также влияет на цену.

Результаты модели в контексте перечисленных шоков могут значительно измениться, именно поэтому для дальнейшего исследования на тему стоимостей автомобилей будет интересно изучить последствия данных изменений. Такой перерасчет модели имеет смысл делать через некоторое время с пересмотром контекста.

Расширение исследования позволит оценить, насколько найденная связь сохраняет актуальность в условиях значительных изменений, и если сохраняет, то при каких новых вводных.

Приложения

Данные, с которыми мы работали, можно скачать по ссылке:
https://drive.google.com/uc?export=download&id=1XKRSG8IwrAvhBMY_f0H4JWJkOZYokQJ4

Работа с категориальными переменными:

Помимо исследуемой переменной *transmission*, в модель были включены ещё две категориальные переменные: *Fuel* (тип топлива в автомобиле) и *Drive* (привод автомобиля). Распределение наблюдений по классам показано в Таблице 7 и Таблице 8 соответственно. Для обеих переменных были созданы *dummys* внутри базовой функции *lm()*. В модели используются следующие обозначения:

Переменная *Fuel*: *dizel* (дизельное топливо), *gaz* (газовый тип топлива), *gibrid* (гибридное топливо), *petrol* (стандартный бензин), последний из классов мы использовали в качестве базового.

Переменная *Drive*: *FullDr* (полный привод), *FrontDr* (передний привод), *BackDr* (задний привод), последний мы использовали в качестве базового класса.

Таким образом, в модели используются следующие бинарные переменные: *dizel*, *gaz*, *gibrid*, *FullDr*, *FrontDr*.

Бензин	Газ	Гибрид	Дизель	Всего
13 160	69	308	3436	16 973

Таблица 7: Количество наблюдений по каждому классу двигателя

Задний	Передний	Полный	Всего
925	9 917	6 131	16 973

Таблица 8: Количество наблюдений по каждому классу привода

Необходимые для всех моделей предпосылки:

- $E\{v_i|x\} = 0$
- Q_{xx} — матрица полного ранга, где $Q_{xx} = E\{x \cdot x'\}$
- $Var(u_i|x) < \inf$

	Var	Mean	SD	Min	Max	Median
Мощность двигателя	7115.66	158.01	84.35	13	900	136.00
Стоимость	8 125 370.68	2064.16	2850.50	30.00	55 590.00	1270.00
Объём двигателя	541.74	3.13	23.28	0.70	1020.00	1.80
Пробег	7×10^9	116 547.49	84 077.02	0	1170000	106 000.00

Таблица 9: Описательные статистики для непрерывных переменных до чистки данных

– $\{power_i, automatic_i, manual_i, litres_i, mileage_i, dizel_i, gaz_i, gibrid_i, FullDr_i, FrontDr_i, v_i\}_{i=1,2,\dots,n} \sim iid$

, где x - вектор $(power_i, automatic_i, manual_i, litres_i, mileage_i, dizel_i, gaz_i, gibrid_i, FullDr_i, FrontDr_i)$

Код для построения основной модели:⁶

```
1 super.nelinreg.contr <- lm(cost ~ power + power:Automatic +
2                             power:Manual + drive + engine +
3                             litres + mileage,
4                             data = dummies.auto)
5 cov4 <- vcovHC(super.nelinreg.contr, type = 'HC0')
```

Код для проверки гипотез:

```
1 pow.std <- sqrt(cov4[2, 2]) # SE(beta_1_hat)
2 pm.std <- sqrt(cov4[11, 11]) # SE(beta_2_hat)
3 pa.std <- sqrt(cov4[10, 10]) # SE(beta_3_hat)
4 cov.am <- cov4[10, 11]
5
6 beta_1 <- super.nelinreg.contr$coefficients[2]
7 beta_2 <- super.nelinreg.contr$coefficients[10]
8 beta_3 <- super.nelinreg.contr$coefficients[11]
9
10 t.stat.pow <- beta_1/pow.std # 1 hypothesis
11 pnorm(q=t.stat.pow,
12       lower.tail = F)
13
14 car::linearHypothesis(super.nelinreg.contr, # 2 hypothesis
15                       c("power = 0",
16                         "power:Automatic = 0",
17                         "power:Manual = 0"),
18                       df = Inf,
19                       test = "Chisq",
20                       white.adjust = "hc0")
```

⁶Полный код можно открыть [по ссылке](#)

```

21
22 std.sum <- pm.std + pa.std - 2*cov.am # 3 hypothesis
23 t.stat.am <- (beta_2 - beta_3)/std.sum
24 pnorm(t.stat.am,
25       lower.tail = F)

```

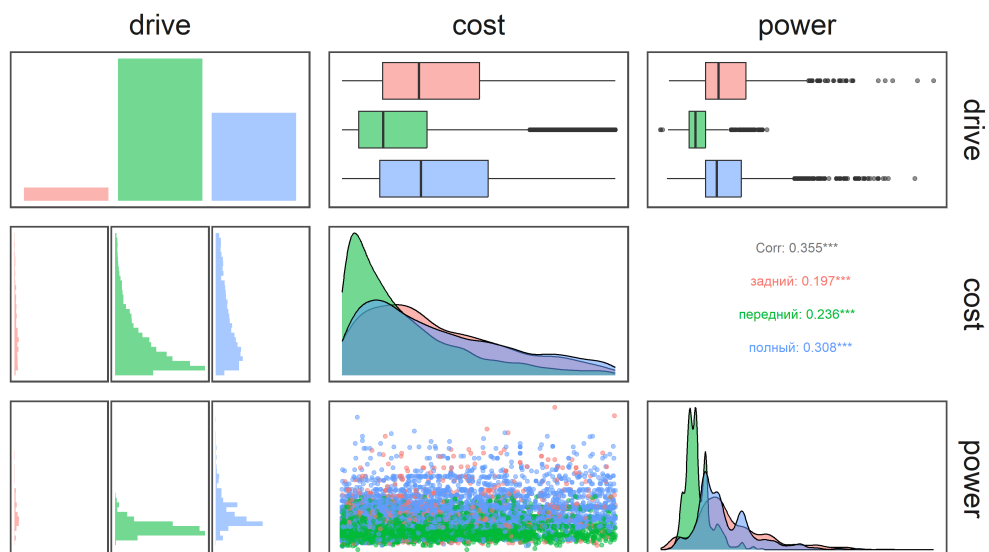


Рис. 5: Зависимая переменная и переменная интереса: в разрезе по классам переменной drive)

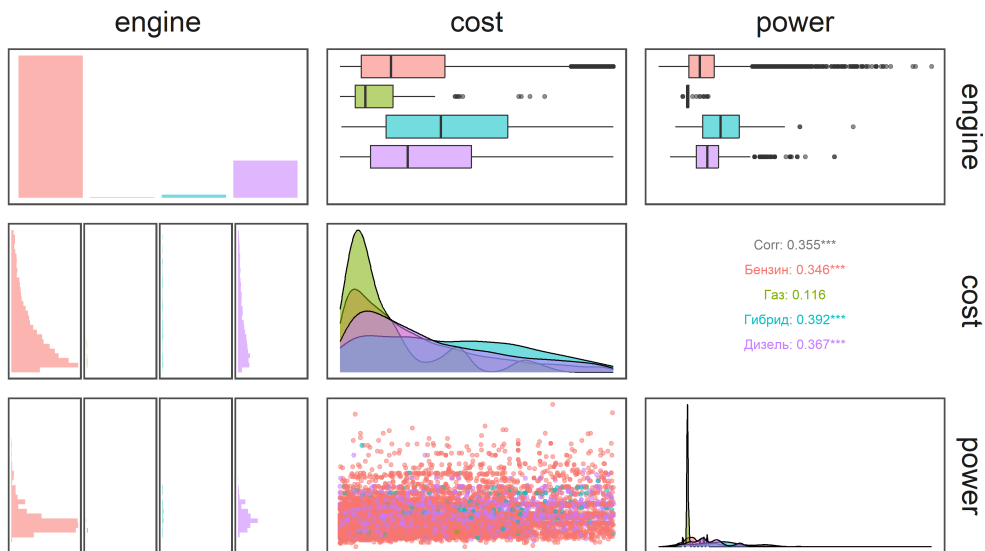


Рис. 6: Зависимая переменная и переменная интереса: в разрезе по классам переменной engine

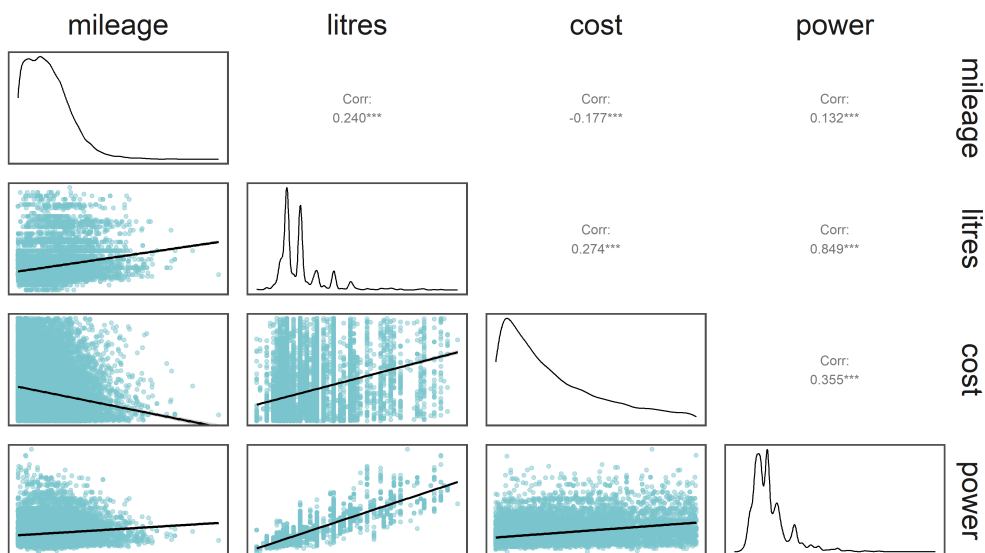


Рис. 7: Зависимая переменная и переменная интереса: точечные графики для непрерывных контрольных переменных