新能源车价格影响因素分析报告

徐之皓 2020111509 孔祥毅 2020111498

2022 年 12 月

摘要：为深入研究新能源车价影响因素，本报告利用当前市场在售新能源车的销售价格数据，通过无交互作用与有交互作用的方差分析，确定 7 个影响新能源车价的主要因素，并建立对新能源车的估价模型。同时，通过描述性统计，回归诊断，关联实际，本报告的研究成果可以为新能源购车者提供科学、稳定的价格参考依据，给各企业市场调研部提供模型参考，同时预测未来上市的潜在新能源车的价格。

关键词：新能源车；回归分析；方差分析；交互作用；车价预测

一、研究背景

新能源车指的是采用非常规燃料来为汽车提供动力，具有新结构、新技术的汽车。近年来，随着国际原油价格波动不断，大力发展新能源汽车产业、抢占市场份额成为了各汽车制造商和业界新势力争夺的焦点。而在目前国内的新能源车市场中，各品牌各种类的新能源车层出不穷，价格也从几万元到上百万不等。找出是什么样的因素在影响着新能源车的销售价格，以及为何会产生价格上的巨大差异，是一件值得探究的事情。本报告利用懂车帝、易车网、汽车之家等网站的官方数据，确定影响新能源车销售价格的重要因素及具体数据，并运用回归分析量化了这些因素的影响。通过对影响新能源车销售价格的因素的了解和探究，购车者可以更加理性地购买爱车，车评机构可以通过理论模型来评估新能源车价值。

二、数据来源和相关说明

爬取数据为“汽车之家”网站（www.autohome.com.cn）2022 年前后上市的销量较好的新能源车数据共 214 条，清洗后最终得到 101 个合格的新能源车样本。我们希望基于这些公开的市场数据，建立恰当的回归模型，从数量上刻画新能源车价格同各个影响因素之间的关系。数据包含如下信息：

表 1 新能源车数据集变量说明

变量类型	变量	水平数
连续型	口碑	
	指导价（万元）	
	轴距（毫米）	
	最大马力（Ps）	
	续航里程（公里）	
	快充时间（小时）	
	智能系统	
	上市热度（万点击量/天）	
离散型	厂商	27 个（保时捷，北京奔驰，比亚迪等）
	增程汽油机	2 个（有/无）

	驱动形式	4 个（后置后驱，前置前驱，三电机四驱，双电机四驱）
	车身结构	4 个（SUV，两厢车，三厢车，掀背车）
	电池种类	3 个（磷酸铁锂电池，三元锂+磷酸铁锂，三元锂电池）

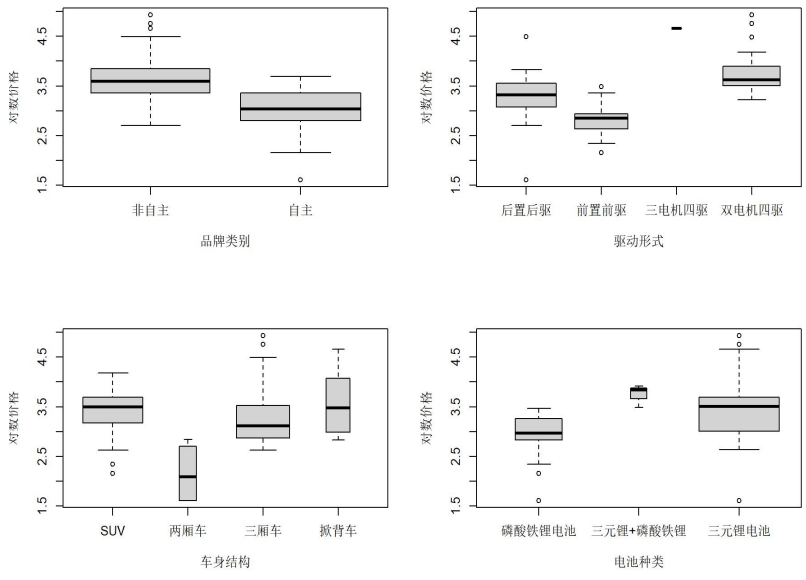
三、数据清洗与整理

根据经验，为减少异方差现象，使结论更加可靠且准确，我们采用因变量对数变换，即对“指导价”取对数价格，并验证了其大致符合正态分布，为我们的分析提供了前提基础。由于厂商数量相对于总体样本量过多，其对应变量解释力过强（单因素方差分析调整后判决系数达到 85%以上），我们根据市场情况将厂商这一解释变量改为品牌类别，将原 27 家厂商中的自主品牌、进口、内外合资等 5 类品牌归为自主与非自主两大类。随后结合实际情况，我们将口碑、续航里程中的缺失值用平均数填充（常规处理法），而快充时间用最大值填充（缺失值基本为不支持快充），进而得到了一份清洗过后的数据集。

四、描述性统计分析

为了获取应变量与自变量关系的整体概念，我们先对数据进行简单的描述性分析。对离散型变量我们在 R 中采用盒状图呈现，对连续型变量如轴距、续航里程等则采用散点图描绘。

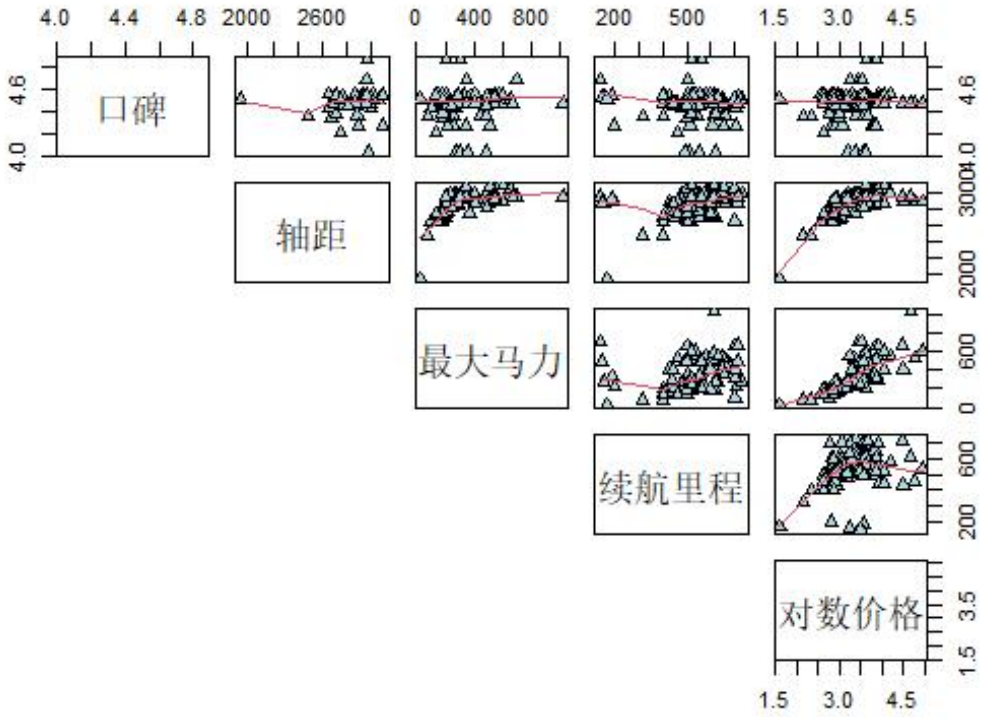
图 1 离散变量对对数价格的影响



从图 1 可以看出，自主品牌、前置前驱、两厢、使用磷酸锂铁电池的新能源车，价格一

般比非自主品牌，其它驱动形式、车身结构与电池种类的新能源车要便宜。这说明合资品牌一般面向高端市场，有较高的品牌溢价，自主品牌则相反。

图 2 连续变量散点图



从图 2 对连续变量的散点图中，我们不难发现：

1、对数价格和口碑并没有必然的联系，在口碑较好的新能源车中同时存在低价与高价车。

2、轴距、最大马力都与对数价格成一定的线性关系，即轴距越长，马力越大的车往往车价也更贵。

3、相对于轴距、最大马力，新能源车的续航里程对于价格的影响则并没有较为显著的线性关系，其实是因为其与电池种类和性能有较大关系。

五、模型分析与预测

（一）无交互作用模型

定义 lm2 为全模型，即包含（品牌类别+口碑+轴距+增程汽油机+最大马力+驱动形式+车身结构+电池种类+续航里程+快充时间+智能系统+上市平均热度）所有影响因素的线性模型，对其进行方差分析结果如下：

表 2 全模型 lm2 方差分析

```
Anova Table (Type III tests)

Response: 对数价格
                Sum Sq Df F value    Pr(>F)
(Intercept)     0.4617  1 10.4115 0.0017931 **
品牌类别        0.5810  1 13.1002 0.0005065 ***
口碑            0.2516  1  5.6739 0.0195057 *
轴距            1.1484  1 25.8945 2.207e-06 ***
增程汽油机      0.0066  1  0.1486 0.7008344
最大马力        0.3765  1  8.4905 0.0045868 **
驱动形式        0.0521  3  0.3917 0.7592778
车身结构        0.1064  3  0.7995 0.4975980
电池种类        0.1088  2  1.2266 0.2985558
续航里程        0.1527  1  3.4430 0.0670713 .
快充时间.小时. 1.0451  1 23.5669 5.594e-06 ***
智能系统        0.3454  1  7.7875 0.0065256 **
上市平均热度    0.0211  1  0.4768 0.4917997
Residuals       3.6809 83
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

从全模型的 summary 可以观察到对模型没有显著影响的影响因素（显著性<0.05）。剔除部分因子后我们得到了选模型 lm3，其中所有变量显著性水平均小于 0.05，模型通过检验。

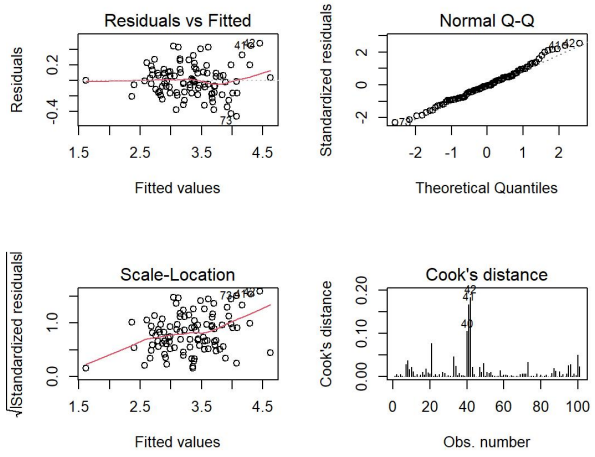
表 3 选模型 lm3 方差分析

```
Anova Table (Type III tests)

Response: 对数价格
                Sum Sq Df F value    Pr(>F)
(Intercept)     0.4427  1 10.0448 0.0020684 **
品牌类别        0.9674  1 21.9497 9.535e-06 ***
口碑            0.2875  1  6.5225 0.0122779 *
轴距            1.4228  1 32.2827 1.518e-07 ***
最大马力        1.7595  1 39.9221 9.012e-09 ***
续航里程        0.1925  1  4.3678 0.0393541 *
快充时间.小时. 1.0769  1 24.4338 3.393e-06 ***
智能系统        0.6248  1 14.1765 0.0002914 ***
Residuals       4.0988 93
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

然后进行模型诊断。在 R 中绘制残差图，结果如下：

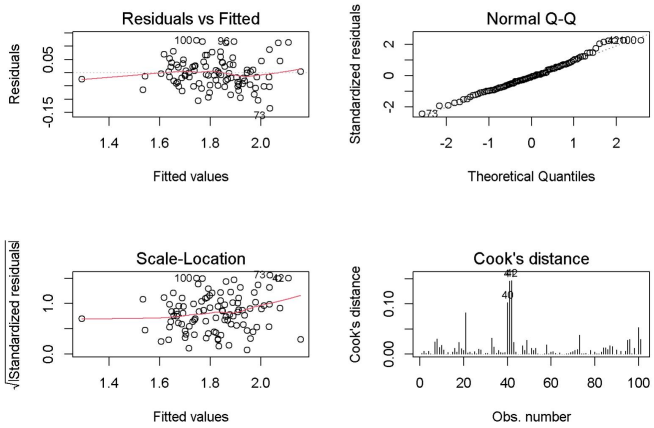
图 3 选模型 lm3 残差图



按照回归假设，残差应为独立同分布的正态变量。从图 3 中我们不难发现，lm3 的误差

方差有随着应变量 y 增大的趋势，根据经验，我们对 y 取平方根变换以消除方差增大趋势，得到了选模型 lm4。此时可以看到异方差问题得到了较好的解决（图 4）。

图 4 选模型 lm4 残差图



我们对调整后的模型进行多重共线性检验，发现各自变量的方差膨胀因子（VIF）均远小于 10，且在贝叶斯信息准则（BIC）下所有变量均被选入。我们还使用 gvlma()函数对线性模型各项假设进行了综合检验，结果如下：

表 4 lm4 综合检验

检验项目	P 值	决策
Global Stat	0.7587	接受假设
Skewness	0.3603	接受假设
Kurtosis	0.7535	接受假设
Link Function	0.5020	接受假设
Heteroscedasticity	0.4844	接受假设

模型各项检验通过。最终得到的无交互作用的选模型 lm4 结果如下：

表 5 选模型 lm4 回归分析结果

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -5.405e-02  2.278e-01  -0.237 0.813020
品牌类别自主    -6.479e-02  1.328e-02  -4.878 4.41e-06 ***
口碑             1.268e-01  4.547e-02   2.789 0.006418 **
轴距             3.516e-04  4.794e-05   7.334 8.12e-11 ***
最大马力         2.635e-04  4.506e-05   5.847 7.38e-08 ***
续航里程         1.087e-04  4.700e-05   2.313 0.022943 *
快充时间.小时.   1.450e-01  3.141e-02   4.616 1.25e-05 ***
智能系统         2.443e-02  6.849e-03   3.567 0.000574 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05615 on 93 degrees of freedom
Multiple R-squared:  0.884,     Adjusted R-squared:  0.8752
F-statistic: 101.2 on 7 and 93 DF,  p-value: < 2.2e-16
```

调整后判决系数达到 87.52%，说明拟合优度较好。因此，非自主品牌新能源车价格方程为

$$\sqrt{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

自主品牌新能源车价格方程为

$$\sqrt{y} = \beta_0 + \beta^* + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

其中 y 为对数价格，β_0为截距项，β_1~β_6为回归系数，$\beta^* = -0.006479$ 是自主品牌带来的漂移项，x_1~x_6分别代表口碑、轴距、最大马力、续航里程、快充时间、智能系统。

（二）模型预测

使用的数据为价格在 15 到 30 万区间的 5 台新能源车数据，包含自主与非自主品牌。最终预测效果如下：

表 6 lm4 预测效果

指导价	预测价	预测偏差
24.98	39.17	14.19
18.53	21.55	3.02
27.98	19.15	-8.83
19.78	16.65	-3.13
15.24	16.51	1.27

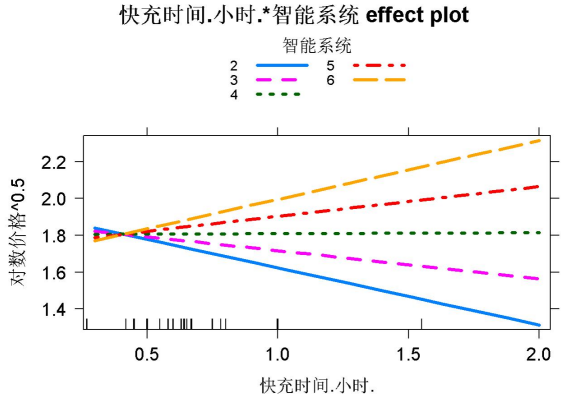
预测 MSE 为 59.97，RMSE 为 7.74。可以看出模型在少数预测上产生了较大偏差，并且指导价越高的样本上越容易产生较大的误差。

（三）有交互作用模型

对于选模型 lm4，我们考虑添加交互作用项，以期能在模型拟合和预测上有更好的表现。

经过尝试，我们发现有显著交互作用的是快充时间与智能系统，交互项 P 值 1.42e-05。交互作用的影响如图所示。

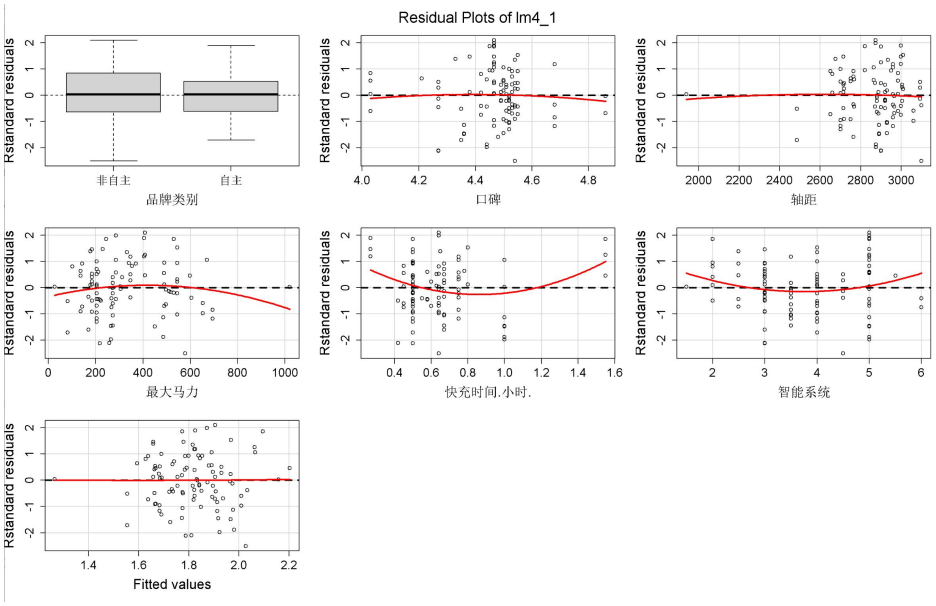
图 5 快充时间与智能系统交互作用示意图



可以看出随着智能系统评分增加，快充时间对对数价格的影响逐渐由负相关转变为正相关。这一现象可这样解释为：如果新能源车搭载有较好的智能系统，顾客会更愿意为较长的充电时间买单。而接下来我们对这一改进模型进行残差分析。

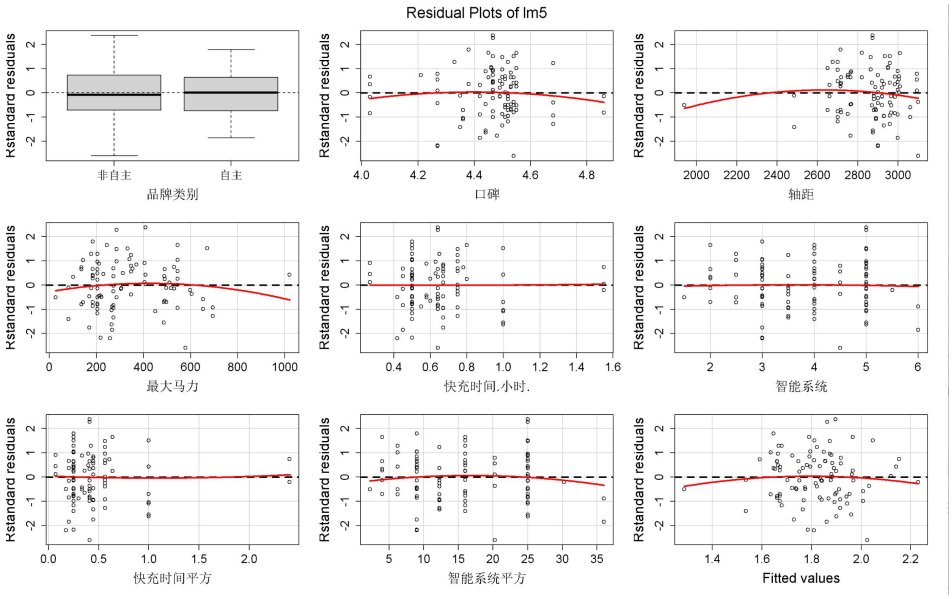
我们使用 car 包中的 residualPlots()函数，该函数可以绘制出模型各个自变量以及拟合值与残差的关系。如果图中红色的拟合线与黑色虚线偏离较大，则可认为线性假设不成立，应考虑添加二次项等其它办法。图 6 中显示了各个自变量以及拟合值与标准化残差的关系图。

图 6 lm4_1 残差图



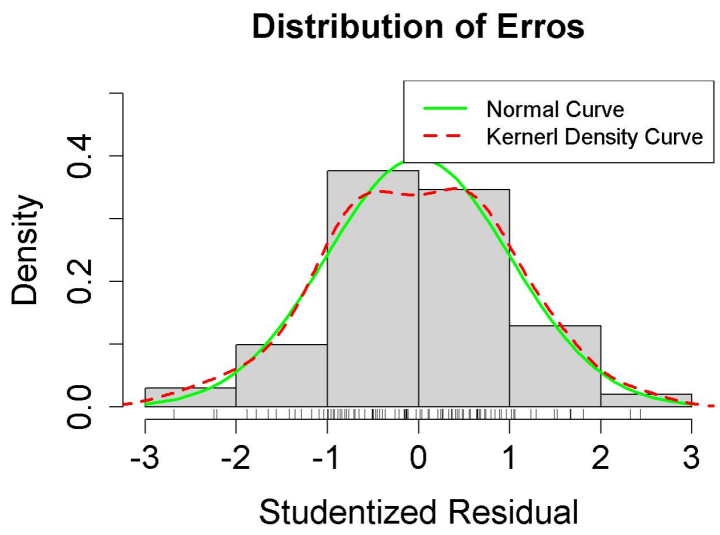
可以看出拟合曲线与黑色虚线并不完全接近，因此考虑添加二次项。由图，继续添加了快充时间、智能系统的二次项（快充时间平方、智能系统平方）后，我们得到有交互作用模型 lm5。可以看到 lm5 的残差分布已经比较符合线性假设。

图 7 lm5 残差图



我们使用自定义的 residstuplot()函数来检验残差的正态性。原理是：如果残差满足正态假设，其学生化后的密度函数（红色虚线）应该贴近标准正态密度函数曲线（绿线）。绘制结果如下：

图 8 lm5 残差分布



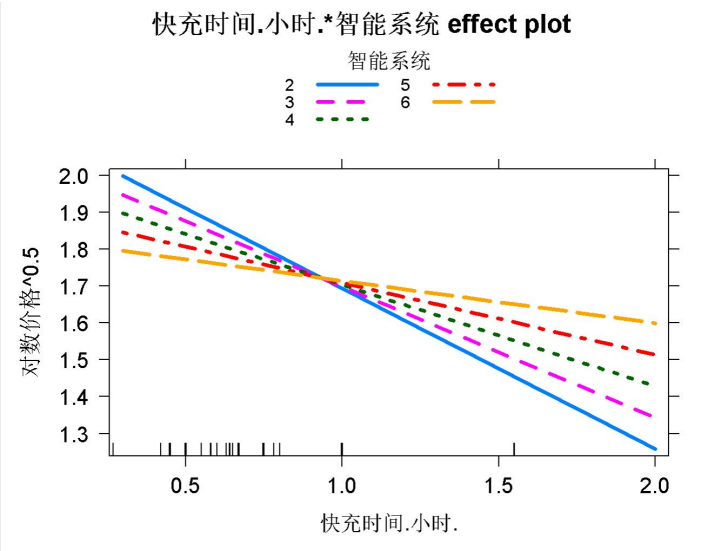
可见红线与绿线较为贴近，残差基本满足正态性假设。通过计算，我们发现在 AIC、BIC 准则下 lm5 均要优于 lm4。

表 7 lm4、lm5 信息准则值

模型	AIC	BIC
lm4	-285.4211	-261.885
lm5	-308.9079	-280.1416

重新绘制交互作用图：

图 9 lm5 交互作用示意图



从可知，随着快充时间的增加，智能系统每增加一个单位引起的 y 的改变在减少。与之前相比，现在的模型认为智能系统高分依然能削弱较长的快充时间对定价的不利影响，但即使是在最高分情况下快充时间与对数价格仍然呈现负相关关系，这也更加符合直觉。

最终得到的有交互作用模型 lm5 结果如下：

表 8 lm5 分析结果

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              1.903e-01  2.230e-01   0.853 0.395733
品牌类别自主            -7.212e-02  1.239e-02  -5.819 8.75e-08 ***
口碑                     1.573e-01  4.084e-02   3.852 0.000218 ***
轴距                     3.990e-04  4.345e-05   9.184 1.30e-14 ***
最大马力                 2.693e-04  3.975e-05   6.775 1.20e-09 ***
快充时间.小时.          -5.954e-01  1.659e-01  -3.588 0.000538 ***
智能系统                -7.472e-02  4.305e-02  -1.735 0.086045 .
快充时间平方             1.955e-01  6.656e-02   2.937 0.004194 **
智能系统平方             6.657e-03  4.943e-03   1.347 0.181413
快充时间.小时.:智能系统  7.998e-02  3.992e-02   2.003 0.048112 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04954 on 91 degrees of freedom
Multiple R-squared:  0.9116,    Adjusted R-squared:  0.9029
F-statistic: 104.3 on 9 and 91 DF,  p-value: < 2.2e-16
```

调整后判决系数达到 90.29%，也说明拟合优度优于 lm4。因此，非自主品牌新能源车价格方程为

$$\sqrt{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_5^2 + \beta_8 x_6^2 + \beta_9 x_5 x_6$$

自主品牌新能源车价格方程为

$$\sqrt{y} = \beta_0 + \beta^* + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_5^2 + \beta_8 x_6^2 + \beta_9 x_5 x_6$$

其中 y 为对数价格，β_0为截距项，β_1~β_9为回归系数，$\beta^* = -0.006479$ 是自主品牌带来的漂移项，x_1~x_6分别代表口碑、轴距、最大马力、续航里程、快充时间、智能系统。

（四）有交互作用模型预测

为了更进一步检验模型是否得到了改进，我们继续使用之前的预测数据集进行预测。

表 9 lm5 预测效果

指导价	预测价	预测偏差
24.98	34.90	9.92
18.53	19.35	0.82
27.98	19.56	-8.42
19.78	17.43	-2.35
15.24	17.71	2.47

预测 MSE 为 36.32，RMSE 为 6.03，虽然在第一、第三个样本上仍然有一定偏离，但绝对误差都减小了，说明模型准确度相比于 lm4 已得到了不小的提升。

六、结语

报告利用当前市场在售新能源车数据，确定影响新能源车销售价格的重要因素，并量化这些因素对销售价格的影响。报告中，我们通过无交互项线性回归、有交互项线性模型、有交互项非线性模型逐步提高模型拟合优度，通过多样的可视化工具展示了分析结果，并在预测集上检验了模型效果。最后调整后 R^2 提高到 90.29%，说明拟合优度很高。我们的模型认为，口碑、轴距、最大马力、续航里程、快充时间、智能系统和是否是自主品牌都对新能源车价格产生了显著影响，快充时间、智能系统之间存在显著交互作用。这为厂商合理定价、顾客理性购车、政府市场调控都提供了有力依据。

大力发展新能源车是我国汽车行业实现对欧美传统车企实现“弯道超车”策略的重要支点，目前我国在市场规模，新能源车销量、电池生产、电机电控、智能化等领域已经处于世界领先地位，但从我们的分析结果看，自主品牌还是主要采取高性价比，薄利多销的策略，在定价上依然无法完全弥补天然劣势。希望有朝一日，自主品牌能在定价权上与合资品牌平起平坐。