# Econometrics Group Assignment

Determinants of Retail Fuel Prices in Hungary


| | |
|---|---|
| Edvárd Müllner | KDWKB8 |
| Sólyom Vigh | BQKGI4 |
| Samu Tósoki | H5NUH7 |


October 31, 2025

## Contents

# 1 Introduction

The aim of this study is to investigate the determinants of retail fuel prices in Hungary at the station level. The dataset contains detailed information on petrol stations' locations, companies, and daily fuel prices for the week of 1–7 October 2024, from which we computed weekly averages for gasoline and diesel as our dependent variables.

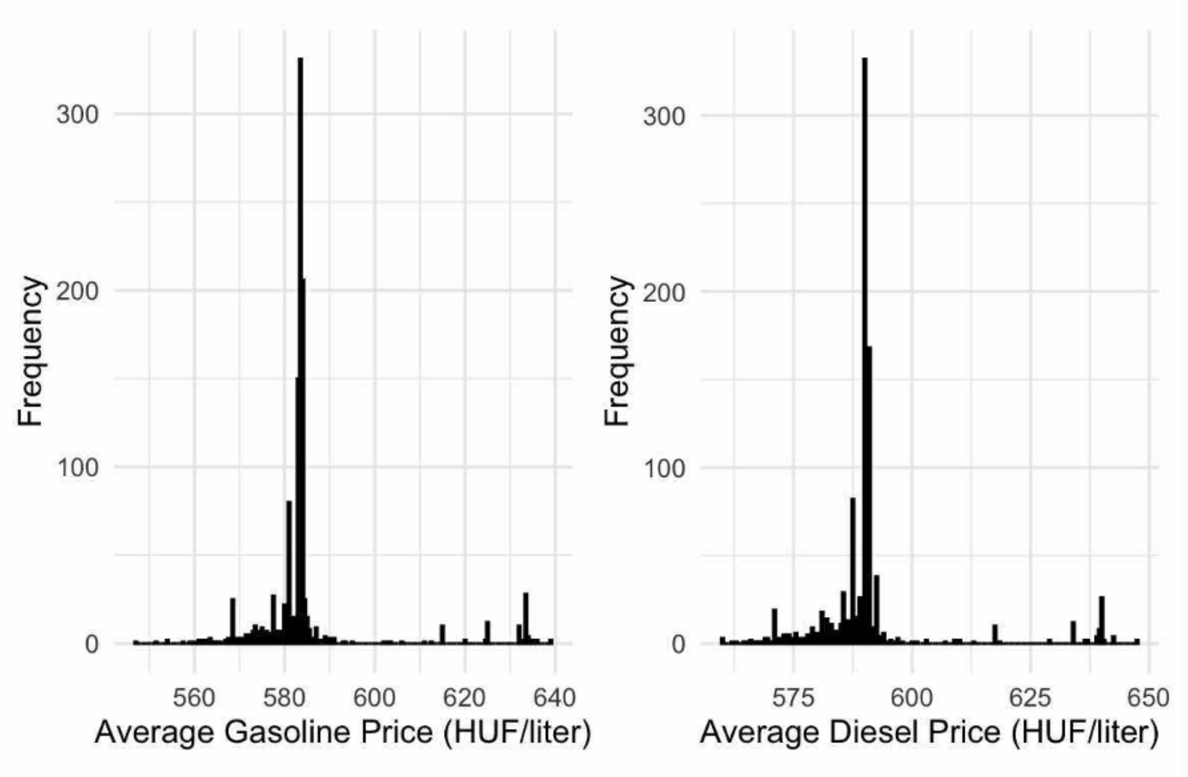


Figure 1: Distribution of average gasoline and diesel prices across stations (HUF/liter). Both fuels show a strong central peak, reflecting price clustering.

The descriptive distribution of prices shows a high degree of concentration around the mean, suggesting that most stations set prices within a narrow range. This indicates limited price dispersion across the country and hints at a highly competitive and transparent market, where cost and tax structures are similar nationwide (from a Hungarian perspective this is an obvious statement, however, in federal countries such as the United States or Germany, fuel taxes and regulations can differ across states or regions.). However, small but systematic deviations from the mean motivate a closer econometric analysis of what drives these local differences.

## 1.1 Economic determinants of fuel prices and price differences

We can say that these four broad categories explain the level and variation of fuel prices. While the first three factors usually account for most of the overall price level, our analysis primarily concentrates on the fourth (market structure and competition intensity) as it is most relevant to understanding local price differences.

- Crude oil prices and global energy markets,

- Exchange rate fluctuations, especially against the US dollar,
- Taxes and regulatory costs, and
- Market structure and competition intensity.

At the retail level, prices reflect not only wholesale cost pass-through but also local competitive and geographical conditions. Basic industrial organization theory implies that retail fuel prices tend to correlate positively with market concentration and the share of premium stations, while overall price dispersion remains modest due to regulatory transparency.

In addition, spatial and demographic factors matter. Urban areas with dense station networks tend to exhibit more competition and lower margins, whereas stations on highways or in rural areas may enjoy local monopoly power. Proximity to international borders introduces cross-border competition, as consumers may arbitrage price differences, especially within the Schengen Area where mobility costs are low.

# 2 Data and Variable Construction

Our dataset originates from weekly retail fuel price data for Hungary (2024.10.01–10.07), containing the variables Date, Settlement, Company, Address, Diesel price, and Gasoline price. From this, we constructed a set of additional variables to capture the economic and spatial mechanisms that could explain station-level price differences.

## 2.1 Brand Category (Supply-Side Structure)

We began by categorising petrol stations based on their operating company into premium, discount, and small categories:

- Premium: Mol, Shell, OMV, Orlen
- Discount: Auchan, Avia, Oil!, Mobil Petrol, Mol Partner+
- Small: all remaining independent operators

This classification reflects a reputation and quality segmentation in the market. Premium brands are associated with stronger brand recognition, higher service quality, and possibly better infrastructure, which allows them to maintain higher prices (brand premium). Discount brands compete on low margins and cost efficiency (i.e., 0–24 self-pay stations), while small operators tend to be locally embedded with more heterogeneous pricing.

This categorical variable proxies for supply-side heterogeneity. Price mark-ups that arise from brand identity, consumer trust, and cost structures.

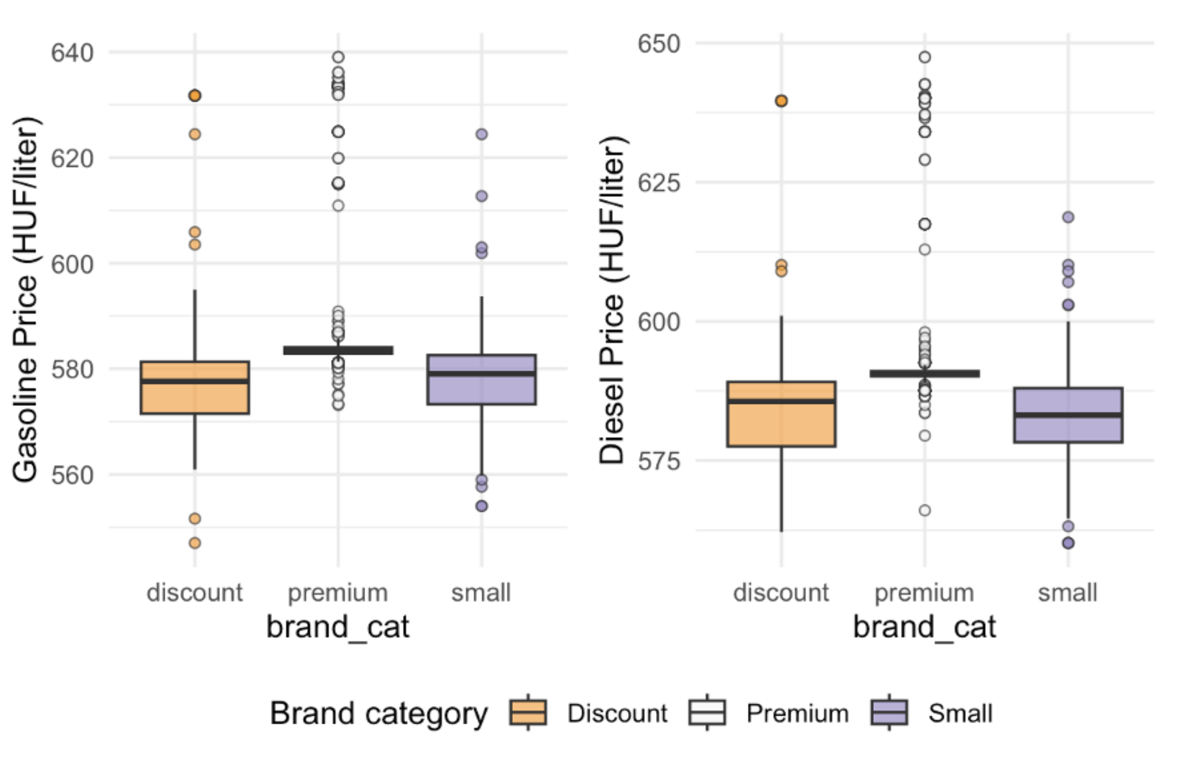


Figure 2: Average gasoline and diesel prices by brand category. Premium brands price slightly above discount and small operators.

As the figure shows, both for gasoline and diesel, premium stations charge visibly higher prices than discount or small operators, confirming the presence of a brand-related markup. However, the difference between discount and small stations is minor, and their price distributions largely overlap. This may reflect the imperfect categorization of the "small" group: since detailed information on these brands was unavailable, we were unable to construct a more nuanced or fragmented classification within this category.

### 2.2 Local Market Structure (Competition Intensity)

Fuel retailing markets are spatially segmented, so the number and distribution of stations in a given settlement directly affects local competition. We created several variables to capture this:

- `brand_in_sett` – the number of stations operated by the same brand in the same settlement. This measures a firm's local brand presence and potential internal cannibalisation. Economically, it can go both ways: more same-brand outlets can indicate dominance (higher prices) or internal competition (lower prices).
- `n_competitors_other_brand` – the number of stations from other brands in the same settlement. This proxies competitive pressure: more rivals should push prices down. We expect a negative relationship with price.
- `hhi_brand` – the Herfindahl–Hirschman Index (HHI) of station ownership within each settlement. This is a standard market concentration measure: higher HHI values imply fewer competitors and more concentrated markets, where prices tend to be higher.

These variables jointly capture local competition intensity and potential market power. By including both the number of competitors and a concentration measure, we can differentiate between dense but fragmented vs. dominated local markets.

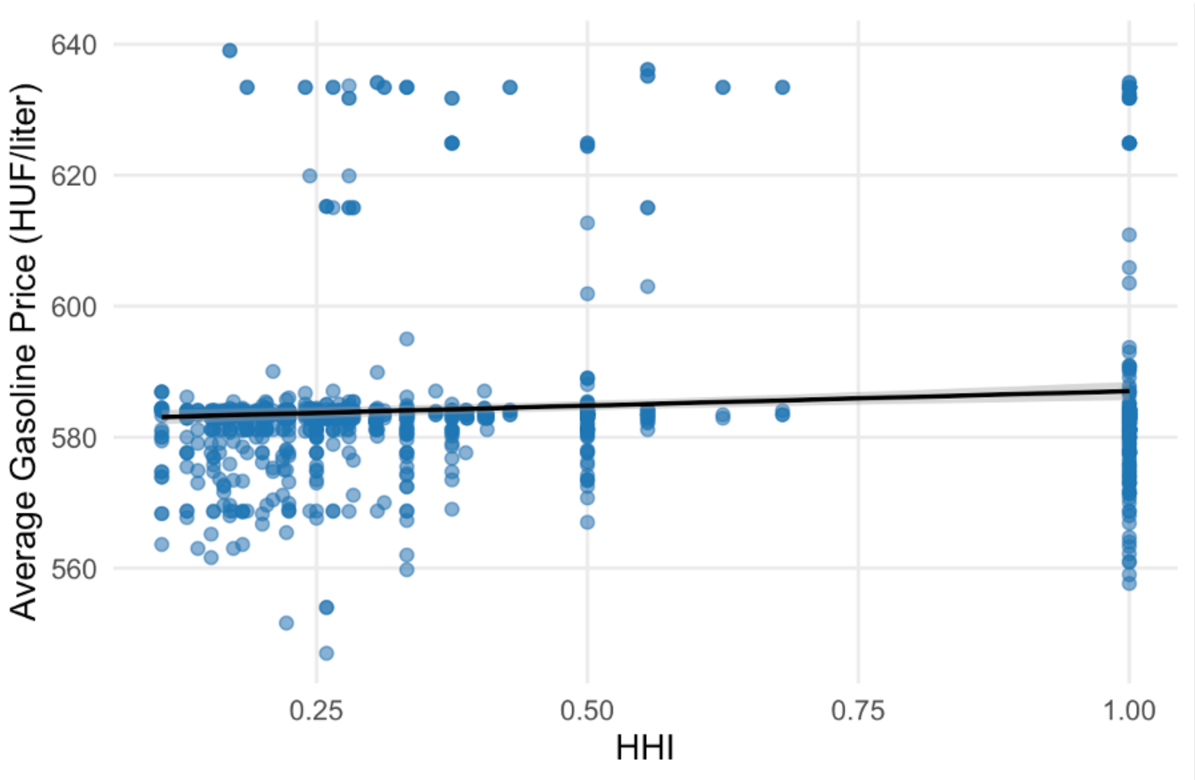


Figure 3: Relationship between market concentration (HHI) and average gasoline prices. Prices tend to rise in more concentrated local markets.

As the figure shows, there is a positive association between market concentration (HHI) and average fuel prices, although the relationship appears weak and quite noisy. The distribution of HHI values is highly uneven, clustering around 0.25—likely corresponding to "normal" settlements with several competing stations—and 1.0, which probably reflects small or isolated settlements with a single station serving a wider region. While this pattern does not imply that concentration should be excluded from the model, it highlights the need to carefully control for heterogeneity in market structure when interpreting its effect on prices.

### 2.3 Market Structure Shares

We computed brand-category shares for each settlement—that is, the proportion of premium, discount, and small stations in that locality.

These shares provide a finer-grained view of local market dominance. For instance, settlements dominated by premium brands may experience systematically higher prices due to reputation-driven pricing or reduced low-cost competition. These variables help identify whether market outcomes depend not only on the number of competitors but also on who those competitors are.

### 2.4 Demand-Side Controls (Population Density)

We added population density (people per km²) from the KSH database to approximate demand intensity. Higher population density typically correlates with:

- greater potential demand (more customers per station), and
- better access to substitutes (public transport).

The net effect is ambiguous but often negative: stations in dense urban areas can operate with smaller margins due to scale, while rural stations face less competition and higher logistical costs.

Population density captures the demand-side market environment, differentiating between urban and rural pricing behaviour.

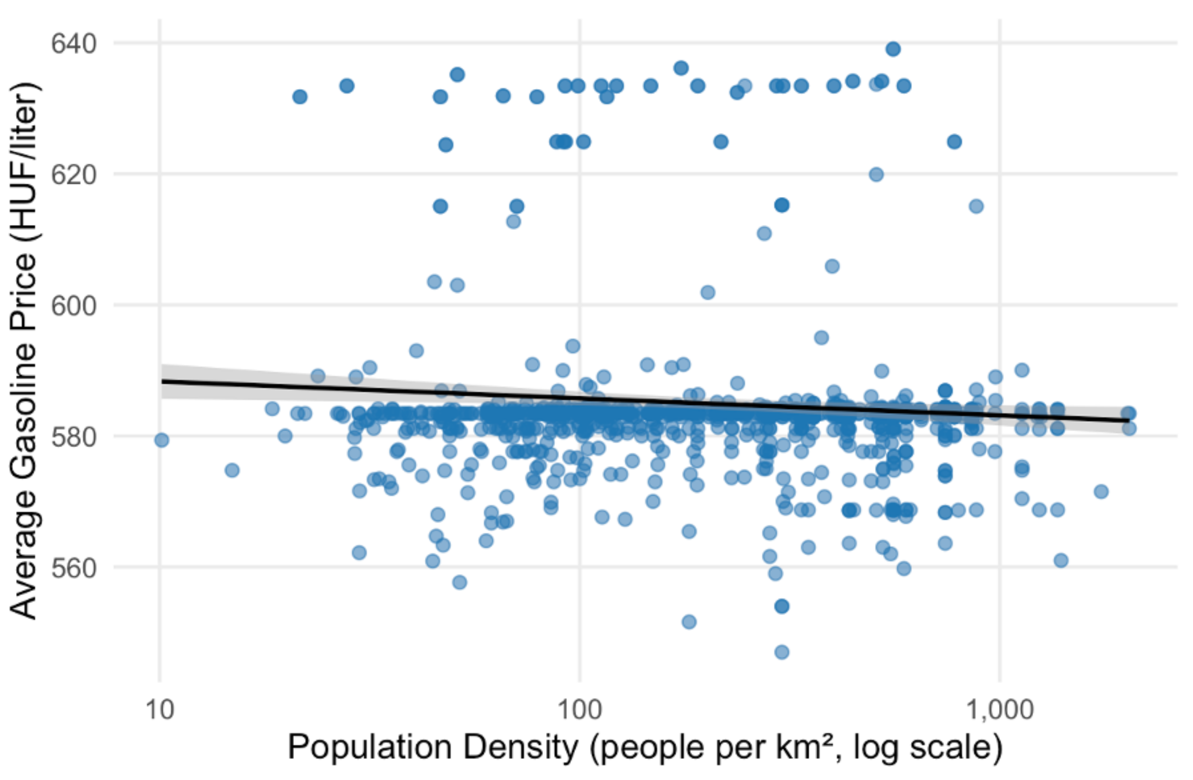


Figure 4: Association between population density and average gasoline prices (log scale). Denser areas tend to have slightly lower prices, consistent with stronger competition.

As the figure shows, there is a weak negative relationship between population density (and gasoline prices. Stations in denser settlements tend to charge slightly lower prices, which likely reflects stronger competition, higher sales turnover, and easier access to substitutes such as public transport. However, the relationship is relatively flat and noisy, suggesting that density alone only marginally explains price variation. The corresponding plot for diesel prices displays a nearly identical pattern, differing mainly in the higher overall price level observed during the sample period.

### 2.5 Spatial and Geographic Variables

We leveraged spatial data to quantify location-based pricing factors using OpenStreetMap geocoding:

- `on_highway` – a dummy variable indicating whether a station is located directly on a motorway. This captures spatial monopoly effects: drivers on motorways have limited alternatives and face switching costs, allowing highway stations to charge higher prices.
- `dist_border_km` – continuous distance from each settlement to the nearest international border.
  `near_border_10km`, `near_border_20km` – dummy variables for being within 10 or 20 km of a border. These variables proxy cross-border competition: fuel markets near borders may experience price pressure from foreign alternatives or different tax regimes.
- Attempted differentiation between Schengen and non-Schengen borders – we also aimed to separate border effects depending on whether the neighbouring country was part of the Schengen Area. The reasoning was that cross-border fuel arbitrage is far more likely across open Schengen borders, while non-Schengen borders imply customs checks and higher crossing costs, making cross-border refuelling less attractive. However, this distinction could not be fully implemented because geocoding of specific border types was not successful.

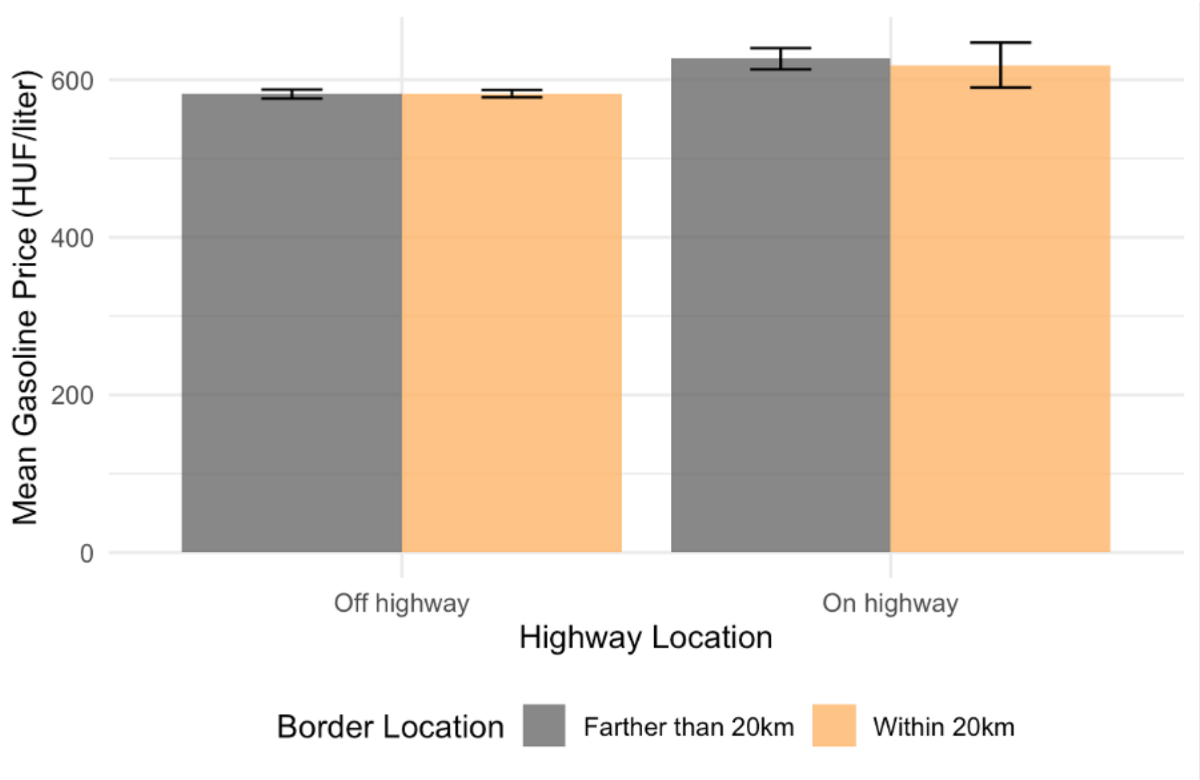


Figure 5: Average gasoline prices by highway and border proximity. Stations on highways charge more on average, while proximity to borders partly offsets this premium.

As the figure shows, highway stations are considerably more expensive on average than off-highway ones, confirming a strong location-based monopoly effect: drivers on motorways have few alternatives and face higher switching costs. In contrast, proximity to international borders alone has only a weak and inconsistent association with prices – near-border stations are sometimes slightly cheaper, but this difference is small relative to overall variation. When both dimensions are considered jointly, the pattern suggests that the highway price premium is somewhat smaller near borders, implying that cross-border competition partially offsets local monopoly power. However, this relationship remains modest and should be interpreted with caution, as border proximity may also capture settlement-level differences (e.g., rural vs. urban context). Together, these results highlight how spatial accessibility and geographic position interact to shape local pricing incentives in Hungary's fuel market.

# 3 Empirical Models

We estimated a series of models to examine how brand type, competition, demand, and geography jointly shape retail fuel prices. The full regression outputs and robustness checks are reported in the Appendix, but here we summarise the main logic behind the modelling steps and highlight the most important findings. Our modelling strategy was incremental: we began with simple brand-based models and progressively added variables that capture competition, demand, and spatial context.

## 3.1 From Brand-Level Pricing to Market Structure

The first models tested whether price differences can be explained purely by brand category. As expected, premium brands charged significantly higher prices than discount or small operators. However, these differences alone explained only a small share of total variation, confirming that brand reputation is not the only driver of prices. Adding local competition measures such as the number of same-brand (brand dominance) and rival stations, as well as the Herfindahl–Hirschman Index (HHI), substantially improved model fit. Prices were generally lower in areas with more competitors, consistent with standard competitive theory. Still, when the same brand operated several stations within one settlement, prices tended to rise slightly—possibly due to stronger brand power.

## 3.2 Incorporating Demand and Geography

Introducing population density helped account for demand-side differences between urban and rural areas. Denser settlements were associated with marginally lower prices, which may reflect higher turnover, better infrastructure, or substitution possibilities (public transport). This variable helped reduce omitted-variable bias from regional composition. Moreover, incorporating an indicator of local income levels or general economic development would have helped to capture further demand-side heterogeneity, as higher-income areas may tolerate higher prices or host more premium stations. Unfortunately, such detailed socioeconomic data were only available in the restricted KSH T-STAR database.

Spatial variables proved especially powerful. Adding the on highway dummy alone dramatically increased explanatory power: highway stations were on average 40–45 HUF (roughly 7%)

more expensive than others. This finding supports the idea that motorway locations behave as local monopolies, where drivers have fewer alternatives and are less price-sensitive. The `near_border_20km` variable was also statistically significant but positively associated with prices. This suggests that stations located within 20 km of an international border tend to charge somewhat higher (0.8-1.4 HUF/liter) prices on average. However, this effect should be interpreted with caution, as the direction and strength of the border influence likely depend on whether the neighboring country is part of the Schengen Area—an aspect we could not fully capture due to incomplete border-type data.

Adding border proximity variables showed weaker direct effects but helped refine our understanding of spatial competition. The inclusion of an interaction term between highway location and border proximity revealed that the highway price premium appears smaller near borders, suggesting that cross-border competition may partly offset local monopoly power. However, this interaction term was not statistically significant, indicating that the moderating effect of border proximity remains suggestive rather than conclusive. Moreover, without distinguishing whether a given border lies within the Schengen Area, we cannot be fully certain about the mechanism behind this relationship.

### 3.3 Summary and Preferred Specification

Among all specifications, the model including brand, competition, demand, highway location, and the highway–border interaction provided the best balance between explanatory power and interpretability. Its $R^2$ was around 0.80, and residual diagnostics indicated stable and robust performance.

Our overall interpretation is that retail fuel prices in Hungary are primarily driven by geographic and market-structure factors rather than demand differences alone. Brand and competition matter, but the most decisive determinants are spatial position and accessibility—particularly whether a station is on a highway and how close it is to an international border.

## 4 Model Diagnostics

### 4.1 Selecting the final model

Model specification followed an incremental approach, beginning with a simple baseline and gradually incorporating structural and locational factors until the preferred model was identified. The process relied on a combination of theoretical reasoning and empirical diagnostics, including changes in information criteria (AIC, BIC), predictive accuracy (RMSE), and formal specification tests (Breusch–Pagan and Ramsey RESET). The initial specification M1 included only the

brand category variables, distinguishing between discount and small stations relative to premium brands. Both categories displayed statistically significant negative coefficients, indicating that non-premium brands systematically price below premium competitors. However, the explanatory power of the model was weak and diagnostic tests suggested possible violations of key OLS assumptions. In particular, the Breusch–Pagan test detected heteroskedasticity in the residuals.

Although a RESET test indicated a possible functional-form misspecification, with only one explanatory variable, it holds no real information. It is testing something trivial: whether a polynomial transformation of your single variable adds explanatory power. To address this, our second model M2 introduced local market-structure controls: the number of stations of the same brand within the settlement , the number of rival stations, and the brand-level Herfindahl–Hirschman Index. A Wald test comparing Model 2 to Model 1 indicated that the inclusion of local market-structure variables significantly improved model fit, reflected by declining AIC and BIC values and a reduction in RMSE, though the diagnostic tests continued to suggest some nonlinearity and heteroskedasticity. M3 extended the specification with settlement-level population density, a proxy for local market scale and demand intensity. The coefficient on this variable was negative and significant, suggesting that prices tend to be lower in more densely populated areas where competitive pressure and throughput are higher. Although the fit of the model improved substantially by AIC and BIC metrics, the RESET test pointed to nonlinearity in the relationship between price and station counts, motivating a functional transformation. Consequently, M4 applied logarithmic forms to the brand and competitor count variables to account for diminishing marginal effects of additional stations. This adjustment led to a further reduction in information criteria, while out-of-sample predicting capability of the model stagnated. Heteroskedasticity persisted, all subsequent models were estimated with robust standard errors to ensure reliable inference. Although a formal Wald comparison between M2 and M3 could not be performed - population-density data were not available for all observations, resulting in differing sample sizes -, the Wald test favoured M4 over M3, confirming that the logarithmic transformation of the station-count variables yields a significantly better fit. Having established a stable market-structure baseline, models 5 through 8 incorporated locational factors, focusing on spatial heterogeneity in fuel pricing. The dummy variable of highway proximity was added first, followed by alternative measures of proximity to international borders, including continuous distance and 10 km and 20 km dummy indicators. The inclusion of the highway dummy yielded a strong positive coefficient, consistent with higher prices at motorway locations where accessibility constraints and traffic intensity confer greater pricing power. The border variables, however, produced smaller and less consistent effects, and models using alternative proximity measures displayed similar fit. These results suggested that border effects might depend on interaction with other factors, rather than operating as a uniform shift across all stations. Though heteroskedastic, including the highway dummy in M5 greatly improved the specification of our model, which is reflected in nearly every metric. Models 6 to 8 showed no statistical improvement over M5. The final two specifications, M9 and M10, therefore introduced an interaction between highway proximity and the 10km as well as the 20 km border dummy respectively. This allowed the model to capture the possibility that the border effect differs for stations located on major motorways. The results aimed to confirmed this hypothesis: while the stand-alone highway effect remained large and positive, the interaction term was negative and significant, indicating that the motorway price premium diminishes near border crossings—consistent with competitive pressure from cross-border refuelling opportunities. Although promising at first, after controlling for their heteroskedasticity by their robust standard errors no interaction effects were proven to be significant. With the interaction effects not holding any additional information, we

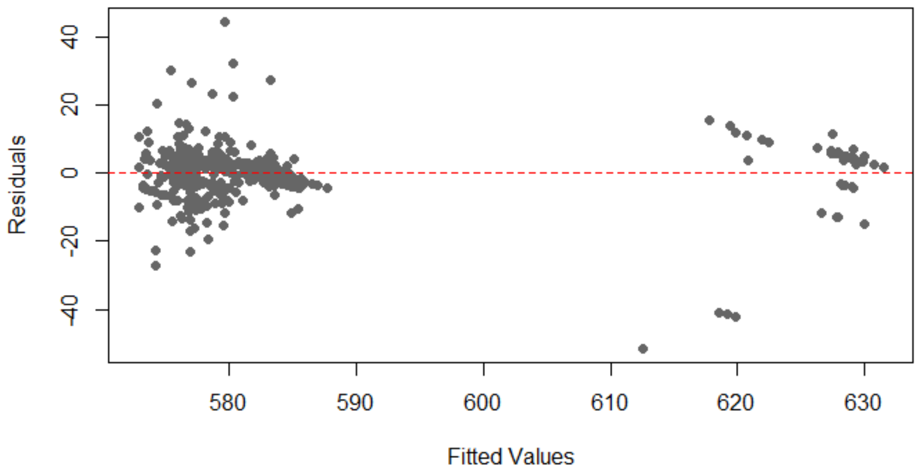


Figure 6: Residuals against fitted values of Model 10

could not find proof of the spatial dispersion of the stations being associated with their fuel pricing. This absence of significant interaction effects after correcting for heteroskedasticity may reflect the pricing structure of the Hungarian fuel market, which is dominated by a few large, vertically integrated companies. These firms typically follow centralized pricing strategies, setting retail prices uniformly across wide geographic areas rather than adjusting them to local market conditions. Such centralized coordination reduces within-brand regional variation and limits the extent to which spatial factors—such as motorway access or border proximity—can influence individual station prices.

### 4.2 Addressing heteroskedasticity in the final model

Diagnostic analysis of the final specification M10 indicated the presence of heteroskedasticity in the residuals. The residuals-versus-fitted values (Figure 6) reveal a clear pattern of non-constant variance, with dispersion increasing around specific ranges of fitted values, suggesting that the variance of the error term is not homogeneous across observations. This visual evidence is consistent with the results of the Breusch–Pagan test, which rejected the null hypothesis of homoskedasticity at conventional significance levels. To ensure valid statistical inference, the model was therefore re-estimated using heteroskedasticity-consistent robust standard errors.

| | *Dependent variable:* | |
|---|---|---|
| | gasoline_avg | |
| | M10 | M10robust |
| | (1) | (2) |
| brand_catdiscount | $-7.596^{***}$ | $-7.596^{***}$ |
| | (0.682) | (1.134) |
| brand_catsmall | $-4.893^{***}$ | $-4.893^{***}$ |
| | (0.647) | (0.857) |
| log(brand_in_sett + 1e-06) | $1.806^{***}$ | $1.806^{***}$ |
| | (0.522) | (0.506) |
| log(n_competitors_other_brand + 1e-07) | $0.256^{**}$ | $0.256^{**}$ |
| | (0.116) | (0.128) |
| hhi_brand | $8.200^{***}$ | $8.200^{***}$ |
| | (2.892) | (2.991) |
| log(pop_density) | $-0.721^{**}$ | $-0.721^{**}$ |
| | (0.285) | (0.352) |
| on_highway1 | $43.744^{***}$ | $43.744^{***}$ |
| | (0.947) | (1.848) |
| near_border_20km1 | 0.843 | $0.843^{**}$ |
| | (0.539) | (0.388) |
| on_highway1:near_border_20km1 | $-9.023^{***}$ | $-9.023$ |
| | (2.716) | (9.825) |
| Constant | $583.252^{***}$ | $583.252^{***}$ |
| | (2.057) | (2.207) |
| Observations | 804 | 804 |
| $R^2$ | 0.798 | 0.798 |
| Adjusted $R^2$ | 0.796 | 0.796 |
| Residual Std. Error (df = 794) | 6.175 | 6.175 |
| F Statistic (df = 9; 794) | $349.399^{***}$ | $349.399^{***}$ |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | |



Table 1: Model 10 with regular and with heteroskedasticity-consistent robust standard errors

As shown in Table ??, the coefficient estimates remain stable between the original and robust versions of the model, indicating that heteroskedasticity affects mainly the precision of inference rather than the magnitude or direction of effects. Accordingly, all final results are reported with robust standard errors, which correct for non-constant error variance while preserving the efficiency and interpretability of the estimated coefficients.

### 4.3 Addressing multicollinearity in the final model

Multicollinearity was assessed in the final specification using the Generalized Variance Inflation Factor (GVIF), which accommodates categorical predictors and interaction terms. The results indicate that collinearity does not pose a major concern. All adjusted values remain well below the conventional threshold of 5, suggesting acceptable levels of variable independence. The moderate GVIF observed for the HHI (4.499) reflects its conceptual overlap with the count of brand-specific stations, but remains within tolerable limits. The highest unadjusted GVIF values are associated with the logged station-count variables, which partly capture similar dimensions of market structure; however, their theoretical distinctness and the relatively low adjusted GVIF values support their joint inclusion. In general, the variance inflation diagnostics confirm that the final model's parameter estimates are not unduly influenced by multicollinearity.

| **Variable** | **GVIF** | **Df** | **GVIF**$^{1/(2 \cdot Df)}$ |
|---|---|---|---|
| brand_cat | 1.178227 | 2 | 1.041855 |
| log(brand_in_sett + 1e-06) | 1.762115 | 1 | 1.327447 |
| log(n_competitors_other_brand + 1e-07) | 18.438491 | 1 | 4.294006 |
| hhi_brand | 20.239651 | 1 | 4.498850 |
| log(pop_density) | 1.687258 | 1 | 1.298945 |
| on_highway | 1.304635 | 1 | 1.142206 |
| near_border_20km | 1.062566 | 1 | 1.030808 |
| on_highway:near_border_20km | 1.152074 | 1 | 1.073347 |

Table 2: Generalized Variance Inflation Factors (GVIF) for Model 10

## 5 Discussion: Diesel vs. Gasoline

The difference between the average prices of diesel and gasoline varies over time. Sometimes diesel is cheaper, but lately the opposite is more common. However, price differences do not emerge at the station level, but rather because of different demand dynamics and other supply side tendencies that arise on the wholesale level. The underlying demand side differences were visible in our analysis as one of the competition intensity variables, namely the `n_competitors_other_brands` was not significant in the models built on diesel, but were significant in the models built on gasoline.This is probably because diesel prices are less elastic due to its different demand structure. The main reason for this is that diesel products have a more stable demand due to the heavy reliance on it by the freight transport industry. While this variable showed this difference, our other variable designed to capture competition intensity effects (HHI) was significant in both of the models. Our explanation for this was that HHI is a better proxy to capture competition intensity because it contains both the number of competitors and their relative market shares. Therefore, it shows significant effect on diesel prices too, however the number of competitor stations bears with less explanatory power and its significance disappears on the less price-elastic market.

## 6 Conclusion

Our analysis shows that retail fuel pricing in Hungary is strongly associated with geography and market structure. Spatial location emerges as the most influential factor shaping price variation across stations. The results show a spatial monopoly effect of stations located on highways, which charge substantially higher prices, around 40–45 HUF, reflecting limited consumer choice and higher switching costs. Brand-related differences also play a significant role: premium networks such as Mol, Shell, OMV, and Orlen consistently apply a higher markup relative to discount and small operators. The observed brand premium of approximately 7.6 HUF for gasoline underlines the competitive segmentation within the retail market.

Local competition effects aligned with our expectations: higher market concentration corresponds to higher prices, whereas densely populated areas show slightly lower price levels due to higher competitive pressure. The final models' fits were robust with an $R^2$ around 0.8.


## Acknowledgement

We gratefully acknowledge the use of ChatGPT for geocoding assistance and for support with writing and formatting this document using the Writefull for LaTeX tool.

## Appendix

Table 3: Regression results with robust (HC1) standard errors for gasoline models (M1–M5)

| | *Dependent variable: gasoline_avg* | | | | |
|---|---|---|---|---|---|
| | M1 | M2 | M3 | M4 | M5 |
| | (1) | (2) | (3) | (4) | (5) |
| brand_catdiscount | −7.230*** | −4.845*** | −4.176** | −4.241*** | −7.579*** |
| | (1.549) | (1.558) | (1.632) | (1.494) | (1.138) |
| brand_catsmall | −8.789*** | −6.087*** | −5.936*** | −5.323*** | −4.844*** |
| | (0.926) | (0.937) | (0.998) | (0.975) | (0.858) |
| brand_in_sett | | 2.598*** | 3.824*** | | |
| | | (0.409) | (0.617) | | |
| n_competitors_other_brand | | −0.506*** | −0.588*** | | |
| | | (0.091) | (0.107) | | |
| log(brand_in_sett + 1e-06) | | | | 10.316*** | 1.864*** |
| | | | | (1.309) | (0.502) |
| log(n_competitors_other_brand + 1e-07) | | | | 1.026*** | 0.268** |
| | | | | (0.260) | (0.131) |
| hhi_brand | | 3.826** | 1.622 | 32.052*** | 8.377*** |
| | | (1.811) | (1.913) | (6.263) | (3.039) |
| pop_density | | | −0.005*** | | |
| | | | (0.001) | | |
| log(pop_density) | | | | −2.632*** | −0.736** |
| | | | | (0.582) | (0.350) |
| on_highway1 | | | | | 42.707*** |
| | | | | | (1.964) |
| Constant | 586.796*** | 582.093*** | 583.438*** | 583.936*** | 583.452*** |
| | (0.464) | (1.173) | (1.435) | (3.559) | (2.228) |
| Observations | 960 | 960 | 804 | 804 | 804 |
| $R^2$ | 0.075 | 0.154 | 0.183 | 0.220 | 0.795 |
| Adjusted $R^2$ | 0.073 | 0.150 | 0.177 | 0.214 | 0.794 |
| Residual Std. Error | 12.461 (df = 957) | 11.932 (df = 954) | 12.407 (df = 797) | 12.125 (df = 797) | 6.214 (df = 796) |
| F Statistic | 38.612*** | 34.780*** | 29.804*** | 37.443*** | 442.024*** |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Table 4: Regression results with robust (HC1) standard errors for gasoline models (M6–M10)

| | *Dependent variable: gasoline_avg* | | | | |
|---|---|---|---|---|---|
| | M6 | M7 | M8 | M9 | M10 |
| | (6) | (7) | (8) | (9) | (10) |
| brand_catdiscount | −7.583*** | −7.543*** | −7.559*** | −7.543*** | −7.596*** |
| | (1.138) | (1.137) | (1.145) | (1.137) | (1.134) |
| brand_catsmall | −4.852*** | −4.808*** | −4.854*** | −4.808*** | −4.893*** |
| | (0.864) | (0.858) | (0.856) | (0.858) | (0.857) |
| log(brand_in_sett + 1e-06) | 1.869*** | 1.881*** | 1.876*** | 1.881*** | 1.806*** |
| | (0.502) | (0.502) | (0.500) | (0.502) | (0.506) |
| log(n_competitors_other_brand + 1e-07) | 0.268** | 0.249* | 0.275** | 0.249* | 0.256** |
| | (0.131) | (0.130) | (0.130) | (0.130) | (0.128) |
| hhi_brand | 8.359*** | 7.879*** | 8.612*** | 7.879*** | 8.200*** |
| | (3.051) | (3.028) | (3.016) | (3.028) | (2.991) |
| log(pop_density) | −0.749** | −0.718** | −0.736** | −0.718** | −0.721** |
| | (0.358) | (0.348) | (0.349) | (0.348) | (0.352) |
| on_highway1 | 42.718*** | 42.825*** | 42.752*** | 42.825*** | 43.744*** |
| | (1.970) | (1.966) | (1.985) | (1.966) | (1.848) |
| dist_border_km | −0.002 | | | | |
| | (0.008) | | | | |
| near_border_10km1 | | 1.219* | | 1.219* | |
| | | (0.718) | | (0.718) | |
| near_border_20km1 | | | 0.494 | | 0.843** |
| | | | (0.564) | | (0.388) |
| on_highway1:near_border_20km1 | | | | | −9.023 |
| | | | | | (9.825) |
| Constant | 583.641*** | 583.427*** | 583.238*** | 583.427*** | 583.252*** |
| | (2.433) | (2.224) | (2.196) | (2.224) | (2.207) |
| Observations | 804 | 804 | 804 | 804 | 804 |
| $R^2$ | 0.795 | 0.796 | 0.796 | 0.796 | 0.798 |
| Adjusted $R^2$ | 0.793 | 0.794 | 0.794 | 0.794 | 0.796 |
| Residual Std. Error | 6.217 (df = 795) | 6.210 (df = 795) | 6.214 (df = 795) | 6.210 (df = 795) | 6.175 (df = 794) |
| F Statistic | 386.343*** | 387.500*** | 386.812*** | 387.500*** | 349.399*** |

*Note:* *p<0.1; **p<0.05; ***p<0.01

| Model | AIC | BIC | Testing RMSE | BP_p | RESET_p |
|---|---|---|---|---|---|
| M_gasoline1 | 7572.786 | 7592.254 | 11.553 | 0.042 | 1.000 |
| M_gasoline2 | 7492.533 | 7526.602 | 11.100 | 0.000 | 0.000 |
| M_gasoline3 | 6339.940 | 6377.457 | 11.501 | 0.000 | 0.000 |
| M_gasoline4 | 6303.035 | 6340.552 | 11.501 | 0.000 | 0.000 |
| M_gasoline5 | 5229.059 | 5271.265 | 5.809 | 0.000 | 0.000 |
| M_gasoline6 | 5230.963 | 5277.859 | 5.817 | 0.000 | 0.000 |
| M_gasoline7 | 5229.050 | 5275.946 | 5.824 | 0.000 | 0.000 |
| M_gasoline8 | 5230.186 | 5277.082 | 5.804 | 0.000 | 0.000 |
| M_gasoline9 | 5229.050 | 5275.946 | 5.824 | 0.000 | 0.000 |
| M_gasoline10 | 5221.089 | 5272.675 | 5.799 | 0.000 | 0.000 |

Table 5: Model comparison (M1–M10) including AIC, BIC, out-of-sample RMSE, and diagnostic test results

Table 6: Regression results with robust (HC1) standard errors for diesel models (M1–M5)

| | *Dependent variable: diesel_avg* | | | | |
|---|---|---|---|---|---|
| | M1 | M2 | M3 | M4 | M5 |
| | (1) | (2) | (3) | (4) | (5) |
| brand_catdiscount | −7.810*** | −5.604*** | −4.981*** | −4.994*** | −8.200*** |
| | (1.565) | (1.591) | (1.660) | (1.524) | (1.100) |
| brand_catsmall | −10.594*** | −8.040*** | −7.893*** | −7.209*** | −7.004*** |
| | (0.899) | (0.877) | (0.943) | (0.975) | (0.837) |
| brand_in_sett | | 2.482*** | 3.612*** | | |
| | | (0.410) | (0.614) | | |
| n_competitors_other_brand | | −0.457*** | −0.527*** | | |
| | | (0.096) | (0.111) | | |
| log(brand_in_sett + 1e-06) | | | | 9.952*** | 1.749*** |
| | | | | (1.325) | (0.529) |
| log(n_competitors_other_brand + 1e-07) | | | | 0.879*** | 0.176 |
| | | | | (0.257) | (0.127) |
| hhi_brand | | 4.446** | 2.396 | 28.970*** | 6.761** |
| | | (1.811) | (1.909) | (6.195) | (2.996) |
| pop_density | | | −0.005*** | | |
| | | | (0.001) | | |
| log(pop_density) | | | | −2.360*** | −0.611* |
| | | | | (0.580) | (0.335) |
| on_highway1 | | | | | 42.555*** |
| | | | | | (1.919) |
| Constant | 593.634*** | 588.640*** | 589.879*** | 590.526*** | 590.191*** |
| | (0.462) | (1.180) | (1.434) | (3.540) | (2.109) |
| Observations | 970 | 970 | 814 | 814 | 814 |
| $R^2$ | 0.101 | 0.175 | 0.197 | 0.231 | 0.800 |
| Adjusted $R^2$ | 0.099 | 0.170 | 0.191 | 0.225 | 0.799 |
| Residual Std. Error | 12.438 (df = 967) | 11.937 (df = 964) | 12.432 (df = 807) | 12.167 (df = 807) | 6.204 (df = 806) |
| F Statistic | 54.492*** | 40.807*** | 33.047*** | 40.435*** | 461.505*** |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Table 7: Regression results with robust (HC1) standard errors for diesel models (M6–M10)

| | *Dependent variable: diesel_avg* | | | | |
|---|---|---|---|---|---|
| | M6 | M7 | M8 | M9 | M10 |
| | (6) | (7) | (8) | (9) | (10) |
| brand_catdiscount | −8.214*** | −8.161*** | −8.170*** | −8.161*** | −8.234*** |
| | (1.100) | (1.097) | (1.101) | (1.097) | (1.088) |
| brand_catsmall | −7.017*** | −6.970*** | −7.021*** | −6.970*** | −7.007*** |
| | (0.836) | (0.838) | (0.831) | (0.838) | (0.825) |
| log(brand_in_sett + 1e-06) | 1.766*** | 1.769*** | 1.775*** | 1.769*** | 1.636*** |
| | (0.532) | (0.529) | (0.527) | (0.529) | (0.520) |
| log(n_competitors_other_brand + 1e-07) | 0.176 | 0.155 | 0.190 | 0.155 | 0.161 |
| | (0.127) | (0.126) | (0.127) | (0.126) | (0.123) |
| hhi_brand | 6.705** | 6.191** | 7.204** | 6.191** | 6.504** |
| | (2.994) | (2.980) | (2.985) | (2.980) | (2.911) |
| log(pop_density) | −0.652* | −0.593* | −0.614* | −0.593* | −0.576* |
| | (0.341) | (0.333) | (0.334) | (0.333) | (0.335) |
| on_highway1 | 42.582*** | 42.682*** | 42.626*** | 42.682*** | 43.872*** |
| | (1.923) | (1.922) | (1.937) | (1.922) | (1.873) |
| dist_border_km | −0.007 | | | | |
| | (0.008) | | | | |
| near_border_10km1 | | 1.342* | | 1.342* | |
| | | (0.727) | | (0.727) | |
| near_border_20km1 | | | 0.929* | | 1.337*** |
| | | | (0.550) | | (0.422) |
| on_highway1:near_border_20km1 | | | | | −9.543 |
| | | | | | (7.722) |
| Constant | 590.768*** | 590.178*** | 589.807*** | 590.178*** | 589.811*** |
| | (2.238) | (2.104) | (2.093) | (2.104) | (2.107) |
| Observations | 814 | 814 | 814 | 814 | 814 |
| $R^2$ | 0.801 | 0.801 | 0.801 | 0.801 | 0.805 |
| Adjusted $R^2$ | 0.799 | 0.799 | 0.799 | 0.799 | 0.802 |
| Residual Std. Error | 6.205 (df = 805) | 6.199 (df = 805) | 6.196 (df = 805) | 6.199 (df = 805) | 6.146 (df = 804) |
| F Statistic | 403.853*** | 404.854*** | 405.272*** | 404.854*** | 367.698*** |

*Note:* *p<0.1; **p<0.05; ***p<0.01

| Model | AIC | BIC | Testing RMSE | BP_p | RESET_p |
|---|---|---|---|---|---|
| M_diesel1 | 7647.943 | 7667.452 | 13.352 | 0.025 | 1.000 |
| M_diesel2 | 7571.303 | 7605.444 | 12.932 | 0.000 | 0.000 |
| M_diesel3 | 6422.028 | 6459.643 | 13.217 | 0.000 | 0.000 |
| M_diesel4 | 6386.902 | 6424.517 | 13.206 | 0.000 | 0.000 |
| M_diesel5 | 5291.463 | 5333.780 | 8.428 | 0.000 | 0.000 |
| M_diesel6 | 5292.596 | 5339.615 | 8.443 | 0.000 | 0.000 |
| M_diesel7 | 5290.981 | 5338.001 | 8.451 | 0.000 | 0.000 |
| M_diesel8 | 5290.308 | 5337.327 | 8.428 | 0.000 | 0.000 |
| M_diesel9 | 5290.981 | 5338.001 | 8.451 | 0.000 | 0.000 |
| M_diesel10 | 5278.108 | 5329.829 | 8.480 | 0.000 | 0.000 |

Table 8: Model comparison for diesel (M1–M10) including AIC, BIC, out-of-sample RMSE, and diagnostic test results

| | *Dependent variable:* | |
|---|---|---|
| | gasoline_avg | diesel_avg |
| | (1) | (2) |
| brand_catdiscount | $-7.596^{***}$ | $-8.234^{***}$ |
| | (1.134) | (1.088) |
| brand_catsmall | $-4.893^{***}$ | $-7.007^{***}$ |
| | (0.857) | (0.825) |
| log(brand_in_sett + 1e-06) | $1.806^{***}$ | $1.636^{***}$ |
| | (0.506) | (0.520) |
| log(n_competitors_other_brand + 1e-07) | $0.256^{**}$ | 0.161 |
| | (0.128) | (0.123) |
| hhi_brand | $8.200^{***}$ | $6.504^{**}$ |
| | (2.991) | (2.911) |
| log(pop_density) | $-0.721^{**}$ | $-0.576^{*}$ |
| | (0.352) | (0.335) |
| on_highway1 | $43.744^{***}$ | $43.872^{***}$ |
| | (1.848) | (1.873) |
| near_border_20km1 | $0.843^{**}$ | $1.337^{***}$ |
| | (0.388) | (0.422) |
| on_highway1:near_border_20km1 | $-9.023$ | $-9.543$ |
| | (9.825) | (7.722) |
| Constant | $583.252^{***}$ | $589.811^{***}$ |
| | (2.207) | (2.107) |
| Observations | 804 | 814 |
| $R^2$ | 0.798 | 0.805 |
| Adjusted $R^2$ | 0.796 | 0.802 |
| Residual Std. Error | 6.175 (df = 794) | 6.146 (df = 804) |
| F Statistic | $349.399^{***}$ (df = 9; 794) | $367.698^{***}$ (df = 9; 804) |
| *Note:* | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Table 9: Regression results with robust (HC1) standard errors