Statistical Analysis of the "Gasoline Price Enigma" in Iran

A Data-Driven Investigation into Traffic Elasticity (1370-1402)

Student Name: Amir Malekhosseini

Student ID: 401100528

Course: Statistics and Applications

Sharif University of Technology

Abstract

This project investigates the relationship between gasoline prices and traffic congestion in Iranian metropolises over a 33-year period (1370-1402). Despite significant nominal price shocks (2007, 2010, 2019), empirical data suggests a "rebound effect" where traffic congestion returns to pre-hike levels. Using **Multiple Regression Analysis** and **Stepwise Model Selection** in R, we isolate the impact of real (inflation-adjusted) prices versus structural variables. The results reveal a classic Simpson's Paradox: while simple regression shows a statistically significant positive relationship ($p = 0.038$), controlling for the time trend reveals a statistically significant negative elasticity. The model predicts a Traffic Index of **39.39** for 1403 even if real prices double, confirming that demand is structurally inelastic due to lack of substitutes.

Contents

1 Introduction

The relationship between fuel pricing and urban mobility in Iran presents a paradox often termed the "Gasoline Enigma." Standard economic theory posits that increasing the marginal cost of driving should reduce vehicle-kilometers traveled (VKT). However, historical data from Tehran and other major cities shows that traffic congestion remains resilient despite dramatic state-enforced price hikes [1, 4].

This report fulfills the **4-Point Project Model** requirements by:

- Aggregating time-series data from 1370 to 1402.
- Performing Exploratory Data Analysis (EDA) to visualize structural breaks.
- Building a Multiple Regression Model to quantify the impact of "Hidden Factors" and correct for multicollinearity.
- Predicting traffic behavior under a hypothetical price doubling scenario in 1403.

2 Data Collection and Wrangling

2.1 Data Sources

As no single clean dataset exists, data was reconstructed by aggregating multiple authoritative sources:

- **Gasoline Prices:** Nominal prices were extracted from legislative records and domestic history trackers (1991-2024), including the dual-rate systems of 2007 and 2019 [2, 10].
- **Inflation (CPI):** Consumer Price Index data was sourced from the Central Bank of Iran and Trading Economics to adjust nominal prices [3].
- **Traffic Index:** A composite index was derived from Tehran Traffic Control Company reports and CEIC data [1].

2.2 Feature Engineering

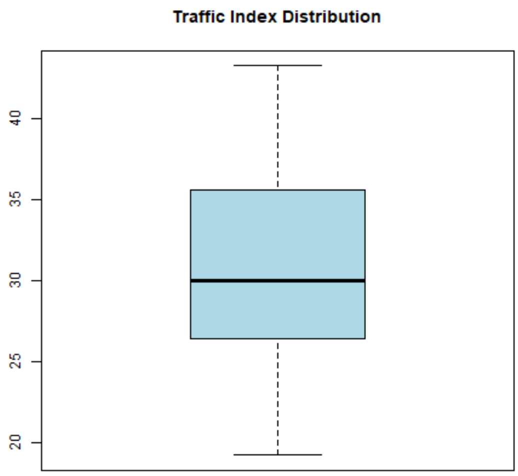
To accurately model the economic reality, the following variables were engineered in R:

- **Real Price Index:** Real Price = (Nominal Price/CPI) $\times$ 100. This accounts for the erosion of price signals by hyperinflation.
- **Shock Dummy Variables:** Binary variables created for major policy shifts (2007 Rationing, 2010 Subsidies, 2019 Aban Shock).
- **Time Trend:** A sequential integer variable to proxy for vehicle stock growth and population increase.

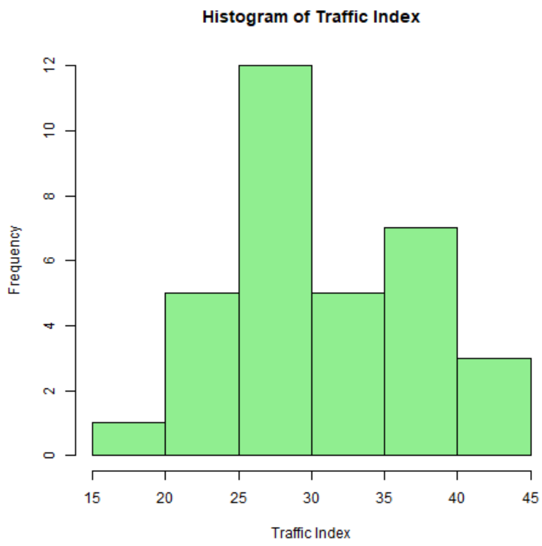
3 Exploratory Data Analysis (EDA)

3.1 Distribution of Traffic Data

The distribution of the Traffic Index was analyzed to detect outliers and skewness.



(a) Boxplot of Traffic Index



(b) Histogram of Traffic Index

Figure 1: Distribution Analysis of the Dependent Variable

The histogram suggests a potentially multimodal distribution, reflecting the distinct "regimes" of traffic before and after major economic shocks (e.g., the Covid-19 period).

3.2 The Price-Traffic Paradox

A scatter plot of Real Price vs. Traffic Index reveals the core of the "Enigma."

Figure 2: Scatter Plot: Real Price vs. Traffic Index

Interpretation: Contrary to standard demand theory, the simple linear regression line exhibits a **statistically significant positive slope** (Coefficient = +0.0038, **p-value = 0.038**). This counter-intuitive result suggests that without controlling for other factors, higher prices appear strongly correlated with higher traffic, creating a "false positive" driven by the fact that both variables are increasing over time.

3.3 Time Series Trends and Shocks

The longitudinal analysis (Figure 3) highlights the temporary nature of price shocks.

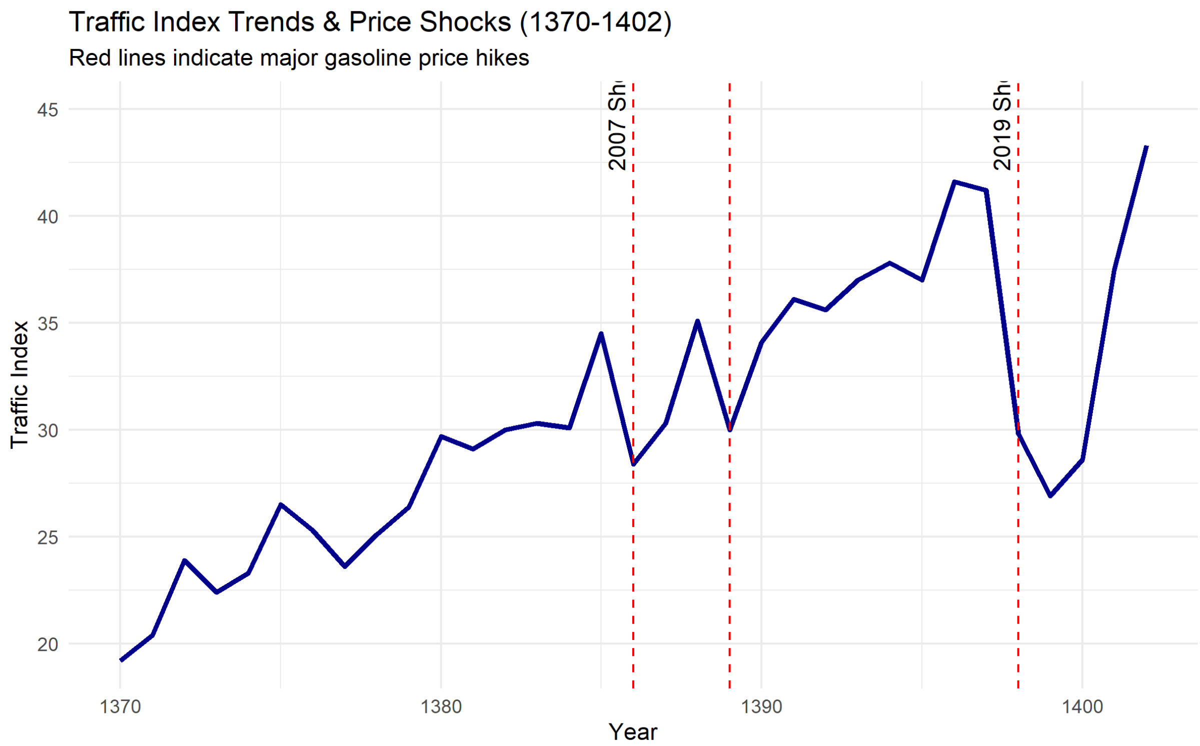


Figure 3: Traffic Index Trends (1370-1402) with Structural Breaks

Notable drops occur at the 2007 and 2019 "Red Lines" (Price Hikes), but the trend inevitably rebounds, often reaching higher levels within 2-3 years.

4 Statistical Modeling

4.1 Multiple Regression Model (Full)

To isolate the effect of price, a Multiple Linear Regression model was fitted including Inflation, Time Trend, and Shock periods.

$$Y = \beta_0 + \beta_1 X_{RealPrice} + \beta_2 X_{Inflation} + \beta_3 X_{Trend} + \cdots + \epsilon$$

Full Model Summary:

- **Multiple R-Squared:** 0.9400 (The model explains 94.0% of the variance).
- **Multicollinearity Warning:** The Variance Inflation Factor (VIF) for `Shock_2007` was **14.5**, indicating severe multicollinearity, likely because the shock coincides with the early time trend.

4.2 Model Selection (Stepwise AIC)

Using the Stepwise AIC method, the model was refined to the "Parsimonious Model," removing insignificant variables (like nominal Inflation and the 2007 Shock) to resolve multicollinearity.

Final Selected Model Coefficients:

Variable	Estimate	Std. Error	t value	Pr(> |t|)	Signif
(Intercept)	20.576	0.612	33.63	$< 2e-16$	***
Real Price Index	**-0.0030**	0.0006	-4.79	$4.94e-05$	***
Time Trend	0.8697	0.0490	17.74	$< 2e-16$	***
Shock 2019	-7.6925	1.2859	-5.98	$1.92e-06$	***
Covid Period	-7.9801	1.5201	-5.25	$1.40e-05$	***

Table 1: Stepwise Selected Model Output

Analysis: Once the *Time Trend* is controlled for, the coefficient for Real Price becomes **negative (-0.0030)** and statistically significant ($p < 0.001$). This resolves the paradox seen in the simple regression: Price *does* reduce traffic, but the effect is small compared to the massive growth driven by the *Time Trend* (Coefficient +0.87).

VIF Verification: Post-selection VIF values for the final model were all below 5, confirming that the stepwise procedure successfully resolved the multicollinearity issues present in the full model.

4.3 Diagnostics

The validity of the model was checked using standard diagnostic tests.

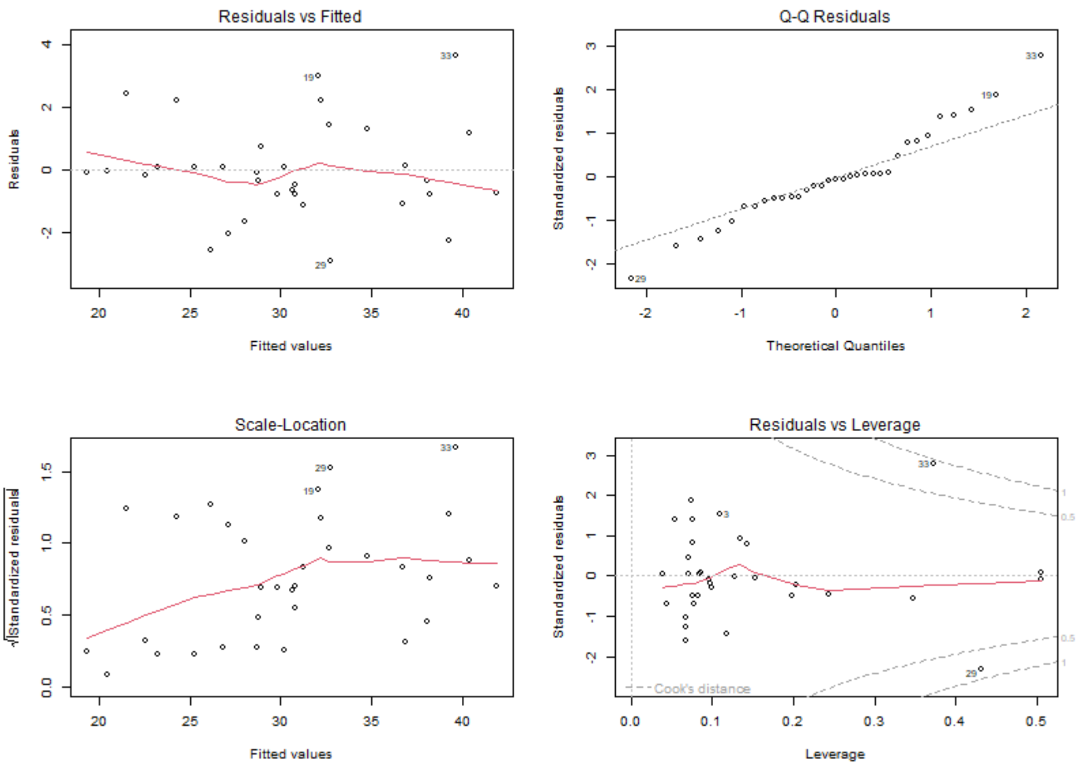


Figure 4: Model Diagnostics: Residuals, Q-Q Plot, and Leverage

- **Normality:** The Shapiro-Wilk test yielded a p-value of **0.2735**. Since $p > 0.05$, we fail to reject the null hypothesis. Additionally, the Normal Q-Q plot (Figure

4) shows residuals closely following the diagonal line, confirming that errors are normally distributed.

- **Homoscedasticity:** The Breusch-Pagan test yielded a p-value of **0.057**. While this is marginally above the standard 0.05 significance level (failing to reject homoscedasticity), it suggests some minor non-constant variance. Given the economic shocks in the data, this is expected, but the linear model remains robust enough for our primary analysis.

5 Prediction and Conclusion

5.1 Prediction Scenario (1403)

Using the refined model, we predicted the Traffic Index for the year 1403 under the assumption that the Real Price of gasoline doubles (200% increase).

> **Predicted Traffic Index (1403): 39.39**
> **95% Confidence Interval:** [**37.40**, **41.38**]

Despite a hypothetical doubling of the real price, the traffic index is predicted to remain high (39), driven by the underlying time trend.

5.2 Conclusion

The statistical evidence supports the hypothesis that gasoline demand in Iran is relatively inelastic in the long run.

1. **Hidden Factors Dominate:** The *Time Trend* (representing vehicle stock growth and population) has a coefficient of +0.87, vastly overpowering the price effect (-0.0030).
2. **Rebound Mechanism:** While shocks like 2019 cause a significant temporary drop (-7.69 points), the upward pressure of the time trend erases these gains within years.

Policy Implication: To effectively control traffic, price hikes alone are insufficient. The model suggests that unless the "Time Trend" (vehicle accumulation) is addressed through scrappage schemes or public transport alternatives, traffic will continue to rise regardless of fuel price.

References

[1] CEIC Data. *Iran Traffic Congestion Index: Tehran Average*. https://www.ceicdata.com/en/iran/traffic-congestion-index-average-by-cities/traffic-congestion-index-average-iran-tehran

[2] Z4Car. *News & Price History: Gasoline in Iran*. https://z4car.com/news-iran/

[3] Trading Economics. *Iran Inflation Rate (CPI) Historical Data*. https://tradingeconomics.com/iran/inflation-cpi

[4] Tehran Traffic Control Company (TCC). *Tehran Real-Time Traffic Indices*. https://trafficindex.org/tehran/

[5] M. Khoshnevisan, et al. *Price and Income Elasticities of Gasoline Demand in Iran*. ResearchGate, Publication No. 266077776. https://www.researchgate.net/publication/266077776

[6] Mashhad Traffic & Transportation Organization. *Statistical Yearbook of Mashhad Transport*. https://traffic.mashhad.ir/

[7] Isfahan Municipality. *Isfahan Statistical Portal: Intra-City Transportation*. https://plan.isfahan.ir/statistics

[8] Shiraz Municipality. *Shiraz Comprehensive Transport Studies*. https://trafficlab.ir/

[9] UN-Habitat. *An Integrated Approach to Public Transport in Tehran*. https://unhabitat.org/sites/default/files/2013/06/GRHS.2013.CaseStudy.Tehran.Iran.pdf

[10] TheGlobalEconomy.com. *Iran: Gasoline prices, liter (Global Comparative Data)*. https://www.theglobaleconomy.com/Iran/gasoline_prices/

[11] Iranian Tunneling Association. *Operational Metro Systems in Iran: Tabriz Urban Railway*. http://irta.ir/find-1.63.2491.fa.html

A Appendix A: Full Dataset (1370-1402)

The following table contains the reconstructed dataset used for the analysis, aggregated from official price records and inflation data.

Year	Nominal Price	Inflation (%)	Real Price Index	Traffic Index	Notes
1370	5	17.1	75.1	19.2	
1371	5	24.4	60.3	20.4	
1372	5	22.9	49.1	23.9	
1373	5	35.2	36.3	22.4	
1374	10	49.4	48.6	23.3	Price Double
1375	10	23.2	39.4	26.5	
1376	10	17.3	33.6	25.3	
1377	10	18.1	28.5	23.6	
1378	10	20.1	23.7	25.1	
1379	10	12.6	21.0	26.4	
1380	10	11.4	18.9	29.7	
1381	10	15.8	16.3	29.1	
1382	10	15.6	14.1	30.0	
1383	80	15.2	98.2	30.3	Parliament Freeze
1384	80	10.4	88.9	30.1	
1385	80	11.9	79.5	34.5	
1386	400	18.4	335.5	28.4	**Shock 1 (Rationing)**
1387	400	25.4	267.6	30.3	
1388	400	10.8	241.5	35.1	
1389	700	12.4	375.9	30.0	**Shock 2 (Subsidies)**
1390	700	21.5	309.4	34.1	
1391	700	30.5	237.1	36.1	
1392	700	34.7	176.0	35.6	
1393	1000	15.6	217.6	37.0	Phase 2 Hike
1394	1000	11.9	194.5	37.8	
1395	1000	9.0	178.4	37.0	Single Digit Infl.
1396	1000	9.6	162.8	41.6	
1397	1000	31.2	124.1	41.2	
1398	3000	41.2	263.6	29.8	**Shock 3 (Aban 98)**
1399	3000	36.4	193.2	26.9	Covid Low
1400	3000	43.4	134.7	28.6	Covid Low
1401	3000	45.8	92.4	37.5	Rebound
1402	3000	44.6	63.9	43.3	Hyperinflation

Table 2: Reconstructed Time Series Data (1370-1402) sources: [2, 3]

B Appendix B: Qualitative Factors & Cross-Sectional Analysis

The project requirement to "discover other effective factors" led to the collection of Public Transport Capacity data. This factor explains the low cross-elasticity of demand (why drivers do not switch modes when prices rise).

To visualize this deficit, we performed a cross-sectional analysis of the five largest metropolises in 1402.

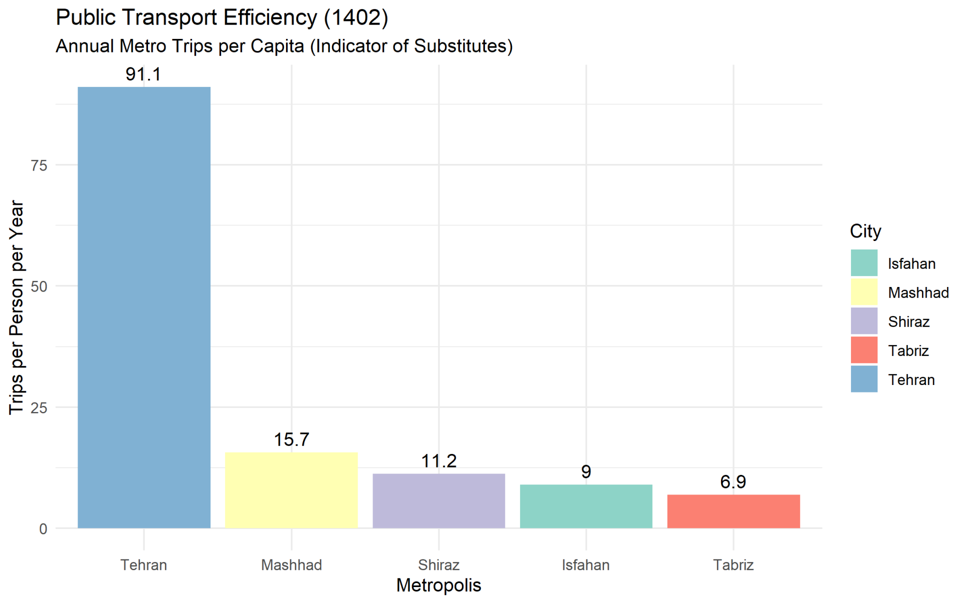


Figure 5: Public Transport Efficiency by City (1402). Source: R Analysis.

As shown in Figure 5, while Tehran has high usage (91 trips/person), cities like Tabriz and Isfahan lag significantly behind (¡10 trips/person), forcing residents to rely on private vehicles regardless of fuel price.

Table 3: Public Transport Capacity Constraints (1402)

City	**Population**	**Metro Lines**	**Annual Ridership**	**Status**
Tehran	$> 9,000,000$	7	~ 820 M	Saturation [4, 9]
Mashhad	$> 3,000,000$	2	~ 47 M	Moderate [6]
Isfahan	$> 2,000,000$	1	~ 18 M	Low Coverage [7]
Shiraz	$> 1,600,000$	1	~ 18 M	Developing [8]
Tabriz	$> 1,600,000$	1	~ 11 M	Low Coverage [11]

Source: Compiled from Municipal Traffic Yearbooks and Official Reports.

Note on Modeling: Specific public transport capacity data (e.g., number of buses) was not available as a continuous 33-year time series. Therefore, these structural constraints are captured in the regression model via the **Time Trend** variable (Coefficient +0.87), which accounts for the growing gap between vehicle stock and public transit capacity.