.

FUEL PRICES AND TRAFFIC FLOW ON BORDER ROADS: UNCOVERING SMUGGLING PATTERNS

AREF NEMAYANDE, NIMA AZAR, HANANEH MOBALLEGHTOHID

Department of Mathematics, Sharif University Of Technology, Tehran, Iran

Abstract. This study reviews our previous project, "Investigating the Relationship Between Fuel Price Disparities and Traffic Flow on Border Roads," and critically analyzes it through the lens of causal inference. The goal is to evaluate our original findings and methodologies, identify mistakes, and suggest improvements. By revisiting the relationship between fuel prices in Turkey and traffic flow at border crossings with Iran, we aim to better understand the causal mechanisms behind observed smuggling activities.

1. Introduction

Causal inference is a key part of scientific research. It helps researchers determine if one variable (the treatment, X) causes changes in another variable (the outcome, Y), rather than just being related to it. This is different from just finding a correlation, which only shows that two things are connected but doesn't prove that one causes the other. Understanding these cause-and-effect relationships is important for making good policies, and interventions, and understanding how things work.[1]

New statistical methods and computing power have greatly improved our ability to determine causality from observational data. While randomized controlled trials (RCTs) are the best way to establish causal relationships because they randomly assign subjects to eliminate other factors[2], they aren't always practical or ethical. Because of this, we often need to use other methods. Pearl's work emphasizes the importance of counterfactual reasoning and graphical models, providing tools for identifying and estimating causal effects from complex data structures[1].

Causal graphs or causal diagrams are useful tools for identifying confounding variables and depicting causal models. By following simple heuristics for causal diagrams, researchers can determine a sufficient set of variables to control for in order to obtain unbiased estimates of causal effects. Importantly, the correctness of causal inference depends on the accuracy of the causal diagram. Therefore, subject matter knowledge and care are needed to develop the causal model.[10] A causal graph consists of nodes, representing variables, and directed edges (arrows) that indicate the direction of causality from one variable to another.

In the context of fuel prices and smuggling, understanding causality can help create effective policies to prevent illegal trade. For instance, high fuel prices in Turkey are thought to encourage smuggling across its borders, affecting traffic on border roads and creating economic and security problems[3]. By analyzing historical fuel price data from Turkey and traffic flow data from Iranian border crossings in West Azerbaijan (For simple writing, we use *fuel price* instead of *fuel price in Turkey*, and *traffic flow* instead of *traffic*

Key words and phrases. Causal inference, Causal graph, Multivariate regression, Instrumental variables.

flow on border roads of West Azerbaijan from now on), we aim to understand if and how changes in fuel prices drive variations in border traffic, serving as an indicator of smuggling. This study addresses key challenges such as data accuracy and confounding factors, and seeks to establish a causal link between fuel prices and smuggling activities.

Ideally, this would be established through a randomized controlled experiment, where fuel prices are randomly assigned to observe their impact on traffic flow. Such a design would ensure that observed effects are due to the fuel price changes rather than other factors. However, conducting such experiments is impractical, so we must rely on observational data. To draw valid conclusions, we need to apply robust statistical methods and causal inference techniques while considering prior knowledge of confounding factors.

Although correlation analysis revealed a significant relationship between fuel prices and traffic flow, we faced challenges due to potential confounding variables that could affect both fuel prices and traffic flow. We explore a few suggestions such as multiple regression which allows us to control for confounders and estimate the causal effect of fuel prices. Still, we should first consider all confounding factors (which basically would take forever if not impossible). We also suggest exploring lagged effects to capture delayed impacts. By incorporating these techniques, we might improve the accuracy of our causal inferences and better understand how fuel price changes influence smuggling activities through traffic flow.

The paper is divided into two main sections. The Section 2 outlines our previous work, providing a detailed examination of our methodologies and challenges. In the subsection 2.5, the formal form of the problem is presented. The Section 3 introduces an improved methodological framework based on advanced causal inference techniques.

2. The Problem: Analyzing the Relationship Between Fuel Prices and Traffic Flow to Infer Smuggling Activities

2.1. **Background.** Understanding how economic variables relate to illegal activities is important in research. This project focuses on Turkey, which has seen major changes in fuel prices because of different economic and political reasons. We think these changes might affect smuggling in Iran, especially on border roads where illegal fuel trading happens. Our main goal is to study how shifts in fuel prices affect traffic flow on these border roads, using traffic flow as a sign of smuggling activities.

2.2. **Hypothesis.** We believe there is a link between fuel prices and the amount of smuggling. The government sets the gasoline price in Iran, and it remains constant across all parts of the country. When fuel prices in neighboring countries, such as Turkey, rise, the potential for higher profits increases, leading to more smuggling and more traffic on border roads. Conversely, when fuel prices drop, the profits from smuggling decrease, resulting in less smuggling and lower traffic flow.[3]

2.3. **Challenges.** Several challenges were encountered during the study:

- **Data Quality**: Ensuring the accuracy and completeness of the data, particularly traffic data from border crossings, which might be affected by inconsistent reporting or recording errors.

- **Confounding Variables**: Identifying and controlling for variables that could confound the relationship between fuel prices and smuggling activities, such as the exchange rate.
- **Measurement Error**: Addressing potential inaccuracies in the measurement of traffic flow and fuel prices, which could bias the results.

2.4. **Methodology.** We tried several techniques to analyze the data:

- **Correlation Analysis**: In our analysis, we constructed a correlation matrix to examine the relationship between fuel prices in Turkey and traffic volumes across three cross-border roads leading to Turkey from Iran. This matrix was generated by calculating the correlation coefficients between monthly fuel prices and corresponding traffic volumes for each road segment. We aimed to identify patterns that could indicate how changes in fuel prices influence cross-border traffic movements. The matrix revealed a consistent positive correlation between fuel prices and traffic volumes across all analyzed road segments, suggesting that as fuel prices rise, the volume of cross-border traffic from Iran to Turkey tends to increase. This finding aligns with our theory, which posits that a higher difference between the cost of fuel in Iran and Turkey can motivate smuggling through border roads.
- **Simple Linear Regression Analysis**: To further substantiate our findings, we employed simple linear regression models for each road segment. The regression model was specified as follows:

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

  where Y_t is the traffic volume at time t, and X_t is the fuel price at the same time. (ϵ_t is the error term). For each road segment, we tested the null hypothesis that there is no relationship between fuel prices and traffic volume ($\beta = 0$). Using a 95% confidence interval, we rejected the null hypothesis for all road segments where the p-value was less than 0.05. This suggests a statistically significant relationship between fuel prices and traffic volumes. One specific road segment showed particularly strong results, with the relationship between fuel prices and traffic volume being most pronounced. This finding underscores the road's significance in our analysis.
- **Exchange Rate Impact**: When we examined the relationship between fuel prices in Turkey and traffic flow on the border roads of Iran, we reached a disappointing conclusion: we found no significant correlation between these two variables. However, it is important to note that the fuel price in Turkey, when considered in terms of its price in rials, affects smuggling and consequently the traffic flow. To account for the impact of the dollar price in our calculations, we converted the Turkish fuel price to rials. After doing so, we observed that the results improved significantly.
- **Addressing Potential Biases**: Recognizing the limitations of our initial analysis, particularly the potential for confounding variables and selection bias, we expanded our investigation to include non-border crossing roads within the same State (West Azerbaijan). This additional analysis aimed to determine whether the observed correlations at the border were specific to these locations or reflected broader trends. We constructed a second correlation matrix for traffic volumes and

fuel prices along non-border roads, applying the same methodology as our initial analysis. This involved calculating correlation coefficients between monthly averages of fuel prices and traffic volumes. The correlation matrix for non-border roads showed correlations that were either slightly positive or slightly negative. These weaker correlations suggest that the relationship between fuel prices and traffic volumes is less pronounced on non-border roads compared to border crossings. In the regression analysis for these roads, p-values were greater than 0.05, indicating that the relationships were not statistically significant.

After analyzing these time series data, we observed a clear pattern: fluctuations in fuel price corresponded significantly with changes in traffic volumes, suggesting a potential causal relationship. However, it is crucial to make sure that causal and associational inferences do not mix. Therefore, we must go further and ask: Does a change in fuel prices cause a change in traffic flow?

2.5. **Causal Inference.** In the context of causal inference, we define our notation as follows: the action or treatment is the fuel price (X), and the outcome of interest is the traffic flow (Y). In this case, we do not distinguish between direct and indirect causes. We argue that this causation is mostly indirect, which indicates the presence of smuggling (Z).

Thus, the hypotheses are as follows:

- Null Hypothesis (H_0): Fuel price does not affect traffic flow.
- Alternative Hypothesis (H_1): Fuel price affects traffic flow.

If the null hypothesis is rejected, we will then ask the following questions: Does this effect indicate that fuel price affects traffic flow? Does this suggest the existence of smuggling? How can we make such inferences?

From our observational data, we want to verify this Causal graph in the first step:

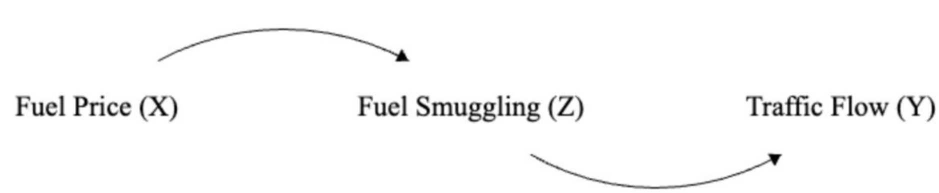

FIGURE 1. First Claim

There are confounding factors that we can consider to draw a more complete causality graph using our background knowledge. For example, important factors such as ”Commodity Exchange” should be included. This leads to a more comprehensive causality graph:

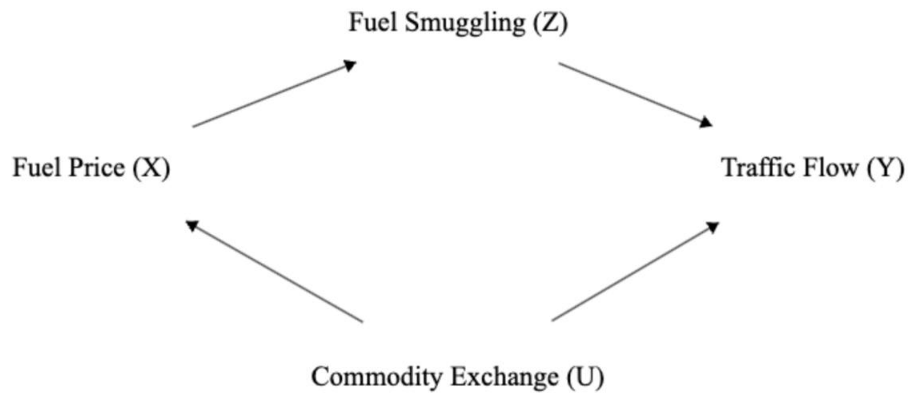

Figure 2. Causal Graph Including Confounding Variable

3. Critical review of the approach

Understanding the causal relationship between fuel prices and traffic flow is crucial for detecting smuggling activities. As we mentioned, we first need to reject the null hypothesis that fuel prices do not affect traffic flow.

We divide this section into two phases, and in each phase, we review our approach and provide suggestions for improvement.

3.1. **Phase 1.**

As we mentioned earlier, ideally, to draw causal conclusions with minimal bias, a randomized controlled experiment would be conducted. This would involve randomly assigning fuel prices and observing the resulting impact on traffic flow, ensuring that the observed effects are due to the manipulation of fuel prices and not confounded by other factors.

Associational assumptions, even when untested, can be verified in principle with sufficiently large samples and fine measurements. In contrast, causal assumptions cannot be verified even in principle without resorting to experimental control.[1]

To make a completely unbiased causal inference, ideally, we would design an experiment with the following test and control groups: the test group would consist of roads that have been randomly assigned to experience increasing fuel prices (to see if it motivates more fuel smuggling), while the control group would not experience these increases. It is essential to ensure that both the test and control groups are unaware of the experiment. However, designing such an experiment is quite impossible, so we are limited to using observational data.

Using observational data presents challenges in establishing causality due to the presence of confounding variables that can influence both fuel prices and traffic flow. Therefore, to draw valid causal inferences from observational data, it is crucial to apply robust statistical methods and causal inference techniques, while also considering prior knowledge of confounding factors. By carefully addressing these issues, we can reduce bias and improve the reliability of our conclusions regarding the impact of fuel price on traffic flow, potentially illuminating smuggling activities.

3.1.1. *Causal Graph.* We initially checked the correlation between fuel prices and traffic flow and found a significant relationship. We then used simple regression to further explore this relationship.

It is well known that correlation doesn't imply causation. In addition, both correlation and simple regression only capture linear relationships and are sensitive to confounding variables that may influence both the independent and dependent variables.

We identified potential confoundings, such as the US dollar which could influence fuel prices and traffic flow. Ignoring these confoundings can lead to biased estimates and incorrect causal inferences.

To improve our results, we need to consider a comprehensive causality graph.[6] We can then adjust these confounding factors in our models or methods to account for their effects.

A causal graph can help us visualize and better understand the relationships between variables. A more accurate and detailed graph will lead to better results in subsequent analyses. Here is one suggested graph:

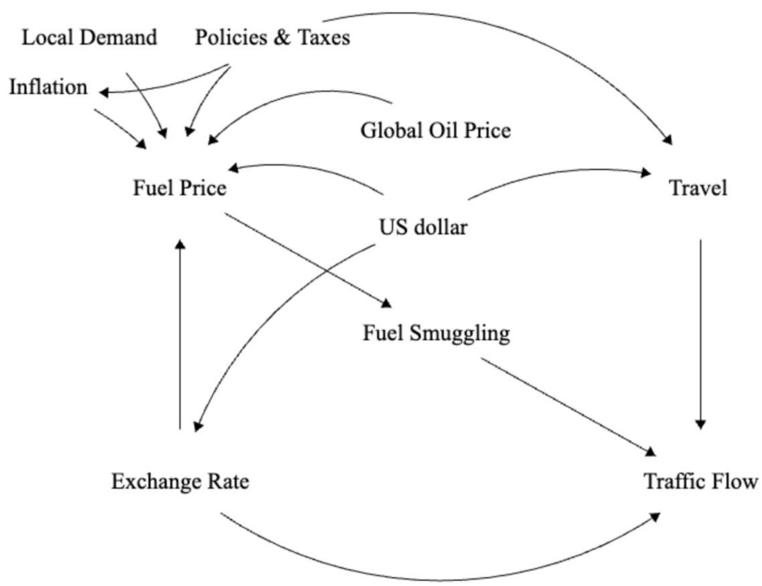

Figure 3. Causal Graph

Here is a short description of a suggested causal graph with discussion definitions of variables and their influences.

- **Fuel Price (X)**: The cost of fuel for consumers in Turkey, measured in Iranian RIAL. Affected by local demand, policies, taxes, global oil prices, and inflation.
- **Traffic Flow (Y)**: The volume of vehicles on the cross-border roads between Iran and Turkey. Impacted by travel from Iran, fuel smuggling activities, and changes in fuel prices.
- **Fuel Smuggling (Z)**: The illegal transportation and sale of fuel to avoid taxes or exploit price differentials between regions. Higher local fuel prices can increase smuggling activities.

- **Commodity Exchange** (U_2): The volume of goods traded from Iran to Turkey or through Turkey to nearby countries. It affects the local fuel price due to the impact on import costs.
- **Travel** (T): The tendency of Iranian people to travel to Turkey by any means. Influenced by the cost of travel and policies affecting travel.
- **US Dollar** (U_1): The value of the US Dollar relative to the Iranian RIAL. Changes in the US Dollar value relative to the Iranian RIAL affect the cost of fuel for Iranians. A stronger US Dollar makes fuel in Turkey more expensive when priced in RIAL. Also influences travel expenses for Iranians and can impact the commodity exchange between Iran and Turkey, affecting trade dynamics.
- **Policies & Taxes** (U_3): Governmental regulations, subsidies, duties, and taxes imposed on fuel within Turkey. These policies can significantly affect the price of fuel and the overall cost of travel.
- **Inflation** (I): The rate at which the general level of prices for goods and services is rising in Turkey. Higher inflation increases the cost of fuel along with other commodities.
- **Local Demand** (W_1): The demand for fuel within Turkey, encompasses the consumption needs of households, businesses, and transportation services. Drives the fuel price in conjunction with other factors.
- **Global Oil Price** (W_2): The international market price of crude oil. Influences the cost of fuel in Turkey due to its reliance on oil imports.

3.1.2. *Multivariate Regression.* After identifying the types of variables and establishing the causal graph, the next step is to address the adverse effects of confounders. Numerous studies illustrate how ignoring confounders can lead to incorrect conclusions.[7] Multiple regression is a powerful tool that allows us to control for the impact of confounders on our variables. In a multiple regression model, we can assess the effect of an independent variable on a dependent variable while holding the effects of other variables constant, thereby preventing unwanted biases.

In a multiple regression model, the response variable is expressed as a linear combination of the predictor variables. Similar to a simple regression model, our goal is to estimate the coefficients that quantify the relationship between the variables. Suppose U represents a confounding variable, such as the "US dollar". The regression equation can be written as follows:

$$Y = \alpha + \beta X + \gamma U + \epsilon$$

where ϵ is the error term. The null hypothesis in this model is:

$$H_0 : \ \beta = \gamma = 0$$

This hypothesis states that there is no linear relationship between the response variable Y and the predictor variables X and U. The alternative hypothesis is:

$$H_1 : \ \beta = 0 \text{ or } \gamma = 0$$

By rejecting the null hypothesis, we conclude that there is a linear relationship between the variables. In our example, the coefficient indicates the effect of the fuel price X on traffic flow Y, holding the confounding variable U constant. This coefficient provides an estimate of the causal effect of fuel prices on traffic flow.

We can extend this model by including additional potential confounding variables. The extended model can be written as:

$$Y = \alpha + \beta X + \gamma_1 U_1 + \gamma_2 U_2 + \ldots + \gamma_n U_n + \epsilon$$

Where $U_1, U_2, \ldots U_n$ represents additional confounding variables. This extension allows us to account for multiple confounding factors simultaneously, providing a more comprehensive analysis of the relationship between fuel prices and traffic flow.

One of the challenges of this method is that some variables cannot be measured or may not have available data. The availability and quality of data on confounders can limit the analysis. Additionally, the regression model can only capture linear relationships between the variables.

3.1.3. *Backdoor Criterion.* Although adding as many measurable factors as possible can make the results more reliable, we must be careful when adjusting variables. For example, the article *Causal Inference Is Not Just a Statistics Problem*[4] introduces four methods for data generation and examines how adjusting a specific factor in each scenario changes the results. Each of the four datasets is generated based on a distinct causal mechanism. The author discussed that even examining the correlation between independent variable X and the known factor V does not help us determine whether adjusting for V is appropriate. For example, if we assume V is a collider variable, we should not adjust for it. A collider is a variable influenced by both the independent variable X and the dependent variable Y. Adjusting for this variable biases the results.

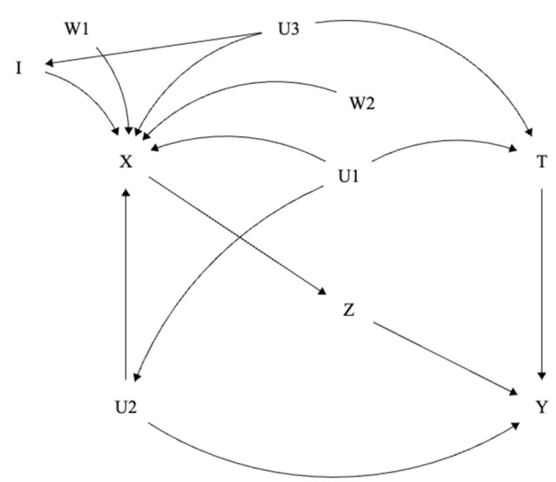

Figure 4. Causal Graph (Variables names)

To solve this problem, we can refer to the "backdoor criterion"[5]. In general, the backdoor criterion is a crucial concept used to identify a set of variables that, when conditioned upon, allow for the estimation of causal effects.

The challenge is to select a subset of factors for measurement and adjustment, ensuring that if we compare treated vs. untreated subjects having the same values of the selected factors, we get the correct treatment effect in that subpopulation. Such a set of factors is called a “sufficient set” or a set “appropriate for adjustment”. Formally, a set of variables S satisfies the backdoor criterion relative to an independent variable X and a dependent variable Y if:

(1) No variable in S is a descendant of X.
(2) S blocks every path between X and Y that contains a backdoor path (a path that starts with an arrow into X).

In this criterion, a set S of nodes is said to block a path p if either (i) p contains at least one arrow-emitting node that is in S, or (ii) p contains at least one collision node that is outside S and has no descendant in S.[1]

To correctly estimate the causal effect, we need to identify and control for variables that satisfy the backdoor criterion. For example, the set $\{U_1, U_2, U_3\}$ is sufficient for adjustment.

3.1.4. *Instrumental Variables.* If certain confounders, such as ”Policies & Taxes” cannot be measured, we cannot include them in the regression model. To address this issue, we can consider using instrumental variables (IV). IVs are a method used in econometrics and statistics that allow researchers to estimate the causal effect of an exposure on an outcome while controlling for confounding bias. The idea of IVs is to find another variable that causes the treatment and is correlated with the outcome only through the treatment.

Let W be our instrumental variable, like ”Local Demand”. We should argue two instrumental variables assumptions:[12], [8]

(1) $Cov(W, X) \neq 0$. This means the instrument must be correlated with the treatment variable (strong first stage).
(2) $Cov(W, \epsilon) \neq 0$. The instrument must not be correlated with the error term in the outcome equation. This means the instrument affects the dependent variable, Y, only through X and not directly.

The first assumption is fortunately verifiable from the data. However, the second IV condition is not easily verifiable and can only be argued in favor of it.

The local demand for gas in Turkey is likely to affect fuel prices. If the demand for gas increases, fuel prices might rise due to higher consumption and limited supply. Conversely, if the demand decreases, prices may fall. The critical assumption here is that W (local demand) affects Y (traffic flow) only through its impact on X (fuel price) and not directly.

After identifying an appropriate instrumental variable that satisfies these criteria, we can estimate the causal effect of the exposure on the outcome using methods such as Two-Stage Least-Squares (2SLS) Regression.[13] Here is a brief implementation of this method:

(1) **First Stage Regression**: Regress the treatment variable (fuel prices or X) on the instrument (local demand or W).

$$X = \alpha_0 + \alpha_1 W + u$$

(2) **Second Stage Regression**: Regress the outcome variable (traffic flow or Y) on the predicted values of the treatment variable ($\hat{X}$) from the first stage.

$$Y = \beta_0 + \beta_1 \hat{X} + \epsilon$$

(3) **Interpret the Results**: The coefficient β_1 In the second stage regression provides an unbiased estimate of the causal effect of fuel prices on traffic flow.

It is not surprising that Instrumental variable analysis is not without challenges:

One of the main challenges is the identification of valid instrumental variables, which should be independent of the outcome except through the exposure of interest and should be strongly correlated with the exposure. Additionally, the instrumental variable must be measurable.

Another challenge is weak instruments, which occur when the correlation between the instrumental variable and the exposure is not strong enough.

By carefully addressing these challenges and selecting appropriate instrumental variables, we can improve the robustness of our causal inferences regarding the impact of fuel prices on traffic flow.

3.1.5. *Effect of lagged variables.* In our initial analysis, we examined the relationship between fuel prices in Turkey and traffic volumes across various cross-border roads leading from Iran, using direct measures of these variables. However, logically, we recognize that the impact of a change in fuel prices on traffic flow takes time and effort. There is often a lagged effect, meaning that changes in fuel prices may take some time to manifest in traffic volumes. Therefore, it is crucial to account for this temporal delay in our analysis.

From the initial charts we constructed, it became evident that the correlation between fuel prices and traffic flow is stronger when considering fuel prices with one or more period lags. This suggests that the impact of fuel price changes on traffic flow may take time and that past fuel prices could be more influential in determining current traffic volumes.

We should test different lags to identify the most appropriate lag structure. Common choices include one-period, two-period, or even longer lags. For example, a lag of one month captures short-term adjustments, while longer lags reflect more gradual changes in smuggling activities.

We can modify the regression models to include lagged terms of the fuel price variable. For example, the updated simple linear regression model (discussed in the second section) could be specified as follows:

$$Y_t = \alpha + \beta X_{t-1} + \epsilon_t$$

where Y_t is the traffic volume at time t, and X_{t-1} is the fuel price at the previous time $t-1$. This modification can be employed in other models developed and discussed earlier. Notably, when we used the lagged version of the price series, we found that the correlation between fuel price and traffic flow increased significantly.

By including multiple lagged terms $(X_{t-1}, X_{t-2}, \ldots, X_{t-n})$ for the fuel prices, the model attempts to account for the cumulative effect of past fuel prices on the current traffic volume. This method allows for the modeling of more complex relationships and can lead to improved predictive accuracy compared to simpler models that do not consider past values of the independent variables. For example, the linear regression model could be specified as follows:

$$Y_t = \alpha + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_n X_{t-n} + \epsilon_t$$

Where Y_t is the traffic volume at time $X_{t-1}, X_{t-2}, \ldots, X_{t-n}$ are the fuel prices at previous times $t-1, t-2, \ldots, t-n$.

This approach is widely used in various fields, including economics, finance, and environmental science, to analyze time-dependent phenomena.

3.1.6. *Regression and Time Series Problems.* When using a linear regression model for time series data, several challenges can arise that affect the accuracy and reliability of the results. One significant issue is autocorrelation, where residuals are correlated over time, violating the assumption of independence and leading to biased standard errors and incorrect inferences. Non-stationarity, where the statistical properties of the series change over time, can result in spurious relationships and misleading conclusions.

3.2. **Phase 2.** Until now, we have discussed our initial attempt at recognizing the causal relationship between fuel prices and traffic flow. The next step was to investigate smuggling activity along border roads. Here is what we aimed to infer:

3.2.1. *Border Condition Isolation.* We tested our hypothesis on some non-border roads to isolate the border or non-border conditions while keeping other factors constant. We selected some non-border roads in West Azerbaijan and tested our hypothesis again. The first observation was that there was very little correlation and high p-values in the regression model. This suggests that the border condition is crucial for the association and even the causation relationship between fuel prices and traffic flow.

However, more than this approach is needed for making any definitive inferences. Our initial method has several issues: We cannot completely isolate the border condition. For example, the road we selected for our project was an intercity road between two crowded cities in West Azerbaijan, which may have unique characteristics influencing the traffic flow. The selection of an intercity road between two densely populated cities introduces confounding factors that might influence traffic flow independently of the border condition. These factors could include population density, because as population density increases, the demand for transportation services, including both public transit and private vehicle use, also rises. Urbanization level since this involves significant infrastructure development, including new roads, bridges, and tunnels. These developments can alter existing traffic routes and create new ones to accommodate growing transportation needs. And economic activities, which are not accounted for in the initial model. These factors might influence roads between more populated cities and not on border roads, which can make our analysis biased. Since it is impossible to remove these factors from roads, we should test our hypothesis on a broader range of roads.

3.2.2. *Path in a Causality Graph.* We can interpret what we have been trying to do as follows: examining the path in a causality graph. Some articles have studied paths and mediations in causal graphs.[1], [9] Broadly speaking, we aim to demonstrate that there is one and only one causal path between the fuel price node and the traffic flow node in our graph. Of course, claiming there is only one path is an exaggeration, as there could be various paths that we have ignored. One might also argue that there is a path of length one between these two variables (direct effect), which we have not shown in the graph. We justify this omission by stating that the effects of these ignored paths between the two variables are little, and the significant effect is through the path that involves the smuggling variable.

In essence, we first argued in our work that there is a causal relationship between our two variables X and Y. This can indicate a causal path between them. By examining various paths and using our prior knowledge, one significant path is the one that involves smuggling (Z). Hence, we claim that the only causal path between these two variables is through Z, indicating the existence of smuggling. To confirm (or attempt to confirm) our claim, we introduce the new variable "traffic flow on non-border," or K, which differs from our previous variable, Y, by not being at the border. We then argued that if there were any other path between X and Y besides the one we mentioned, the same path should exist between X and our new variable K. This should result in a relationship. However, we could not find such a relationship between X and K.

3.2.3. *Mediator Analysis.* Mediator analysis is a statistical approach used to understand the mechanism or process through which an independent variable influences a dependent variable through one or more intervening variables, known as mediators. The main objective is to decompose the total effect of an independent variable on a dependent variable into direct and indirect effects.

In our work, a mediator analysis could help identify whether a smuggling variable or other factors mediate the relationship between fuel prices and traffic flow. We can use this technique to examine how the total effect is distributed between the direct effect and various indirect effects. This approach can be effective in our research.[11]

References

[1] Pearl, J. (2003). Statistics and causal inference: A review. Test 12, 281–345.
[2] Imbens, G. W., & Rubin, D. B. (2015). Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press.
[3] Ghoddusi, H. Rafizadeh, N. Rahmati, M. (2018) Price elasticity of gasoline smuggling: A semi-structural estimation approach, Energy Economics, Volume 71, 2018, Pages 171-185.
[4] D'Agostino McGowan, L., Gerke, T., & Barrett, M. (2024). Causal Inference Is Not Just a Statistics Problem. Journal of Statistics and Data Science Education, 32(2), 150–155.
[5] Pearl, J. (1995). Causal diagrams for empirical research. Biometrika, 82(4), 669-688.
[6] Pearl, Judea (2000). Causality. Cambridge, MA: MIT Press. ISBN 9780521773621.
[7] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2013). An Introduction to Statistical Learning: with Applications in R. New York: Springer, . APA.
[8] Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. International journal of epidemiology, 29(4), 722-729.
[9] Gunzler, D., Chen, T., Wu, P., & Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. Shanghai archives of psychiatry, 25(6), 390.
[10] Cummiskey K, Lübke K. Causality in statistics and data science education. AStA Wirtsch Sozialstat Arch. 2022;16(3-4):277–86.
[11] Rijnhart, J.J.M., Lamp, S.J., Valente, M.J. et al. Mediation analysis methods used in observational research: a scoping review and recommendations. BMC Med Res Methodol 21, 226 (2021).
[12] 08 - Instrumental Variables — Causal Inference for the Brave and True
[13] Two-Stage Least-Squares Regression