DBA5101 Analytics in Managerial Economics

Group Project 1
Estimation of demand function for fish sold at a wet market in Singapore

Group 40 Members
Chung YeonJong (A0280412R)
Lim Ciwen Brendan (A0216513N)
Matthieu Simon Bernard Zekhout (A0195691N)
Tey Ming Chuan (A0220997U)

Introduction

Singapore's vibrant wet markets are renowned for their authentic charm and bustling atmosphere, offering diverse food products such as fish. Within this dynamic marketplace, fish selection is extensive, with each type possessing distinct attributes, such as size, which could influence its pricing.

In this project, we aim to estimate the demand function for fish in the Singapore wet market by leveraging on a given data set that encompasses variables such as price, quantity, date, and onshore and offshore weather conditions. This dataset has log price and quantity per unit weight to account for size differences. However, the dataset may only partially capture some of the nuanced elements as it lacks factors such as freshness that would affect pricing. Nevertheless, we entered this dynamic marketplace, assuming a delicate equilibrium between buyers and sellers at the point of transaction, envisioning a perfectly competitive exchange where supply meets demand and prices are determined. Through our analysis, we will determine variables that count for fixed effects and instrumental, endogenous, and exogenous variables and, consequently, estimate the demand function for fish.

Data Cleaning and Exploratory Data Analysis

The dataset is complete, with no missing values in any cell. From the date column in the dataset, we have confirmed that all data only encompasses Mondays to Fridays and ensured that days of a week ("*Mon*" to "*Fri*") binary variables are correctly encoded. Furthermore, we extracted the month from the date and added dummy binary variables to represent each month. Although we recognize that months extracted from the dataset only span one year, we are still interested in assessing whether seasonality impacts fish sales as it may incite further studies. *Appendix 1* shows a detailed description of the variables provided in the original dataset.

We also conducted exploratory data analysis to uncover any patterns or trends. This analysis included summarising statistics and utilising visualisation tools. *Figure 1* illustrates the behaviour of each variable. Additionally, we employed bar plots to visualise the proportion of 1s and 0s for our weather-related feature variables, as detailed in *Appendix 2*.
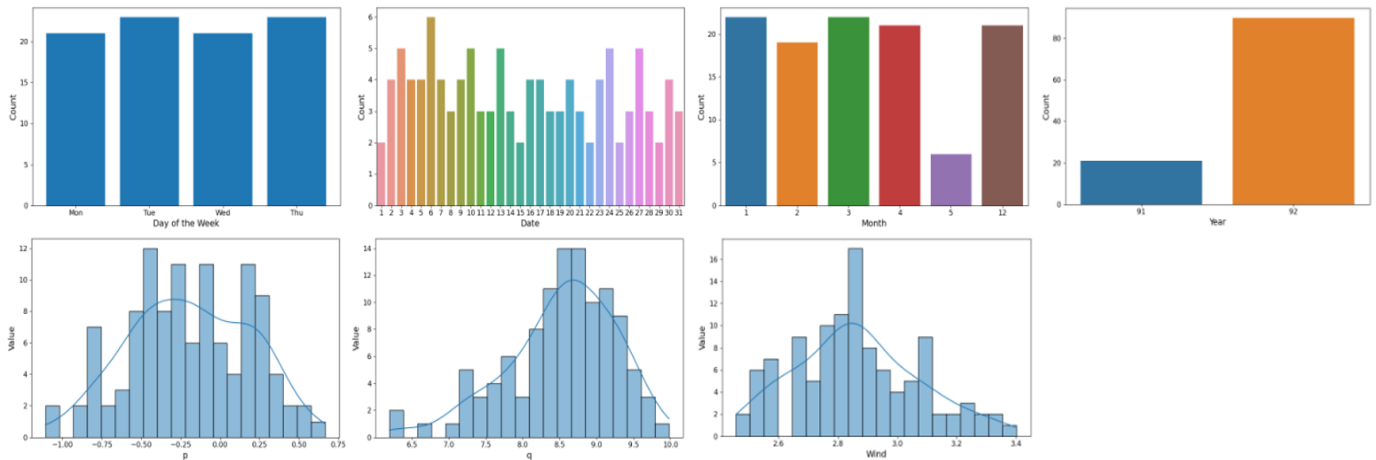

Figure 1. Exploratory Data Analysis

Methodology

To estimate the demand function, we used the variables of log(*price*) represented by p and log(*quantity*) represented by q. Assuming a linear function, we develop the demand function with $q = \alpha - \beta p$. However, due to the simultaneous change of supply and demand when changing one variable, we expect q and p to be endogenous. To address this, we utilise 2 Step Least Squares (2SLS) with selected instrumental variables to obtain an unbiased estimate of β (the slope of the demand function). To increase the robustness of the model, we also considered and tested other variables, assumed to be exogenous, to account for the fixed effects of the demand function. For example, using *Rainy* and *Cold* weather as a fixed effect, we obtain a second model, $q = \alpha + \beta_1 p + \beta_2 Rainy + \beta_3 Cold$.

To effectively identify exogenous and instrumental variables, we analyse the correlation of our variables to p and q. Firstly, *Rainy* and *Cold* weather data was assumed to reflect demand side factors due to the nature of weather on shore, where buyers are located during the sale; hence, they are possible fixed effects of the model. Days of the week and the month of the purchase are also assumed to be a fixed effect to assess for any patterns or seasonality. Finally, the *Wind*, *Stormy*, and *Mixed* are assumed to be supply-side variables as they reflect the severity of the weather offshore, where the fishers are during the fishing process. From these groups of variables, the instrumental variables (IVs) were found to be within the subset of supply-side variables as they are hypothesised not to correlate with the demand on the shore.

Before testing 2SLS, we must establish our constraints. In this process, we constrained binary variables to be grouped (e.g., *Stormy* must be used with *Mixed*); this constraint was enforced to maintain the integrity and generalisability of the model. For instance, a model with *Mon*, *Tue*, and *Thu* but not *Wed* will not be able to generalise to how demand changes across the week and may not improve the business intelligence decisions.

General Reduced form:
$$p = \alpha + \pi_1 Exogenous\ variable + \pi_2 Instrumental\ variable$$

General Structural form:
$$q = \alpha + \beta_1 p + \beta_2 Exogenous\ variable$$

By formulating our reduced and structural form equations, we ran 2SLS with various combinations of IVS and exogenous variables. For the model to be deemed valid, it must exhibit statistical significance, characterised by low p-values, for both the chosen IV in the first stage and p_{hat} in the second stage as per the general reduced and structural form. Next, Hausman's test was conducted on the model to show that endogeneity existed, justifying the use of 2SLS. Finally, for models that contained more than one IV, Sargan's test was conducted to check that the model was not over-identified.

Results

Appendix *3* shows all the models that passed and qualified. Interestingly, we obtained models that showed the month of purchase to be statistically significant, and we can conclude that the month of purchase can qualify as a fixed effect of the demand function. This aligns with the hypothesis that fish

experience seasonality in demand. However, due to a lack of data across all months (notably from June to November), the model will not be generalisable, and we are unable to conclude how the demand changes over the year. As such, we narrowed down to the 3 potential models shown in *Table 1*.

Model	1	2	3
IVs	**Stormy, Mixed**	**Wind**	**Wind**
Exogenous	*Mon, Tue, Wed, Thu*	*Mon, Tue, Wed, Thu*	*Mon, Tue, Wed, Thu Rainy, Cold*
Slope of price	-0.9301 (0.352)	-1.1220 (0.386)	-1.2286 (0.468)
P-value of slope	0.010	0.004	0.010
P-value of F-statistic	0.000403	0.000220	0.00115
R squared	0.191	0.202	0.203

Table 1. 2SLS Model Selection

All 3 models passed the required statistical tests and have instrumental and exogenous variables that make economic sense. A comparison between Model 1 and Model 2 suggests that using *Wind* as IV yields a lower p-value of the slope. In contrast, a comparison between Models 2 and 3 suggests that including more exogenous variables increases the p-value of the slope, which is undesirable. Moreover, the standard error of the estimate increased from 0.386 to 0.468, indicating that using *Rainy* and *Cold* may be inappropriate. Therefore, Model 2 would be the best model due to its low p-value and inclusion of exogenous variables that make it statistically significant and generalisable. Next, to compare how well the 2SLS model fared, model parameters were compared against the Original Least Square (OLS) regression and the base model defined as $q = \alpha + \beta p$.

	Base OLS	**OLS Model 2**	**2SLS Model 2**
Slope of price	-0.5409 (0.179)	-0.5625 (0.168)	-1.1220 (0.386)
P-value of slope	0.003	0.001	0.004
P-value of F-statistic	0.00308	7.08e-05	0.000220
R squared	0.078	0.220	0.202

Table 2. Final Model Comparison

Table 2 shows the comparison between the 2SLS of Model 2 and OLS. We observed changes to the slope of the demand function as we account for fixed effects and endogeneity. Comparing Model 2's OLS against Model 2's 2SLS, we observed a significant difference in the slope magnitude, indicating that the 2SLS method was necessary to account for the significant endogenous effects of price and quantity, even though the standard error increased. Nevertheless, we found the p-value of the slope of the 2SLS model to be statistically significant. While some parameters of the 2SLS do not perform as well as the OLS in Model 2, we conclude that the slope of the demand function is unbiased as it accounts for endogenous effects. Hence, our estimated demand function can be shown as:

$$q = 8.5054 - 1.1220\,p - 0.0256\,Mon - 0.5308\,Tue - 0.5664\,Wed - 0.1094\,Thu$$

In this demand function, we can account for the day of the week from Monday to Friday and the simultaneous nature of price and quantity to accurately estimate the slope of the demand function.

Discussion

The slope coefficient of a -1.1220 indicates that a 1% increase in the price of fish corresponds to 1.1220% decrease in the quantity demanded. Given that the demand elasticity value is derived as $E = -\frac{\Delta q/q}{\Delta p/p}$, the slope coefficient can also be directly extracted from the $\log(q)$ $\log(p)$ graph, allowing us to derive the elasticity of fish demand $E = 1.1220$ of the demand in the Singapore wet market. Since the magnitude of E is larger than 1, fish can be considered an elastic product, meaning that customers are responsive to price changes (The Investopedia Team, 2023). Consequently, sellers should exercise caution when raising prices as it could prompt consumers to explore alternatives and lead to a significant drop in sales. Conversely, during times of surplus, reducing prices can boost demand. Nevertheless, it is important to note that correlation does not imply causation, and these relationships are associations observed within the given data.

Conclusion

Our analysis revealed fish as an elastic product where customers are highly responsive to price fluctuations. As we conclude this study, it is important to acknowledge that the identified relationships are specific to the given data. Moreover, it is essential to remember that finding correlations does not establish causation. To make more comprehensive and generalised conclusions about these dynamics, we need to conduct further data collection and analysis.

References

Graddy, K. (2006). Markets: The Fulton Fish Market. *Journal of Economic Perspectives*, *20*(2), 207–220. Retrieved from: https://doi.org/10.1257/jep.20.2.207

Graddy, K., & Kennedy, P. (2010). When Are Supply And Demand Determined Recursively Rather Than Simultaneously? *Eastern Economic Journal*, *36*(2), 188–197. Retrieved from: https://doi.org/10.1057/eej.2009.3

The Investopedia Team. (2023). Price elasticity of demand meaning, types, and factors that impact it. *Investopedia*. Retrieved from: https://www.investopedia.com/terms/p/priceelasticity.asp

Appendices

Appendix 1: Observed variables in dataset

Mon	1 if Monday, 0 otherwise
Tue	1 if Tuesday, 0 otherwise
Wed	1 if Wednesday, 0 otherwise
Thu	1 if Thursday, 0 otherwise
Date	Date in yymmdd format
Stormy	1 if wave height greater than 1.5m, and wind speed greater than 18 knots. Based on moving averages of the last three days' wind speed and wave height.
Mixed	1 if wave height greater than 1m, and wind speed greater than 13 knots, excluding Stormy days. Based on moving averages of the last three days' wind speed and wave height.
p	**Log** average daily prices per kg of fish
q	**Log** average daily quantity sold of fish in kg
Rainy	1 if rainy weather on shore
Cold	1 if cold weather on shore
Wind	Wind speed in knots

Extracted data was taken from the "*Date*" to extract model "*Date*", "*Month*" and "*Year*".

Appendix 2: Observed variables in dataset
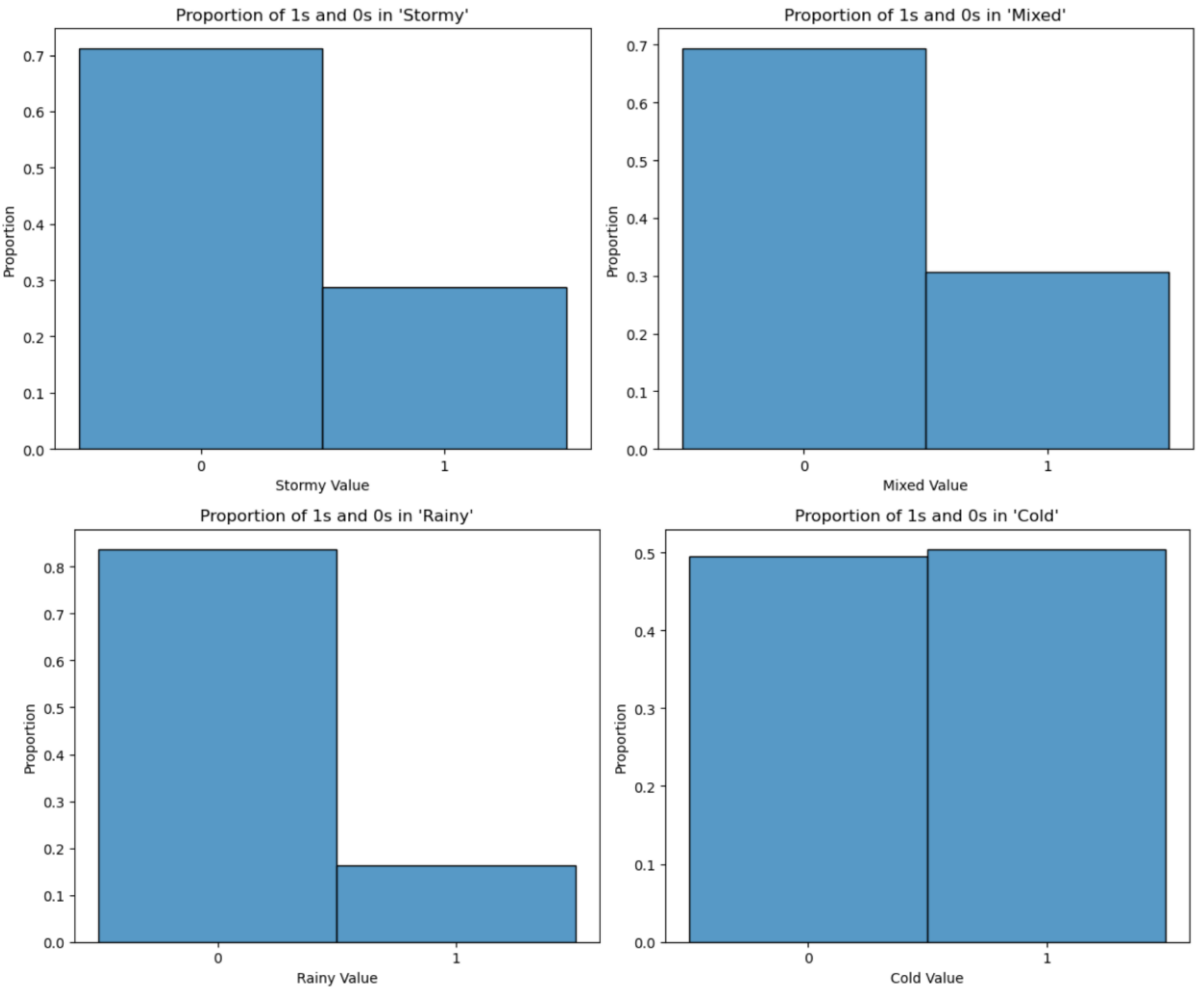
Appendix 3: 2SLS Models that passed all tests

Model	IVs	Exogenous	Slope of price	P-value of slope	P-value of F-statistic	R squared
1	*Stormy, Mixed*	*DoW*	-0.9301	0.010	0.000403	0.191
2	*Wind*	*DoW*	-1.1220	0.004	0.000220	0.202
3	*Wind*	*DoW, Rainy, Cold*	-1.2286	0.010	0.00115	0.207
4	*Wind*	*MoY*	-1.7966	0.003	0.0823	0.100
5	*Stormy, Mixed*	*MoY*	-1.2798	0.005	0.104	0.095
6	*Stormy, Mixed*	*MoY, Rainy, Cold*	-1.2269	0.010	0.125	0.113
7	*Wind*	*MoY, Rainy, Cold*	-1.7417	0.012	0.140	0.110
8	*Stormy, Mixed*	*MoY, DoW, Rainy, Cold*	-1.0981	0.013	0.00802	0.230
9	*Wind*	*MoY, DoY, Rainy, Cold*	-1.4924	0.010	0.00661	0.234
10	*Stormy, Mixed*	*DoW, MoY*	-1.1402	0.007	0.00356	0.222
11	*Wind*	*DoW, MoY*	-1.4957	0.004	0.00237	0.231

*DoW: Days of the Week

*MoY: Months of a Year

Appendix 4: GitHub code

- GitHub URL: https://github.com/mariotey/DBA5101