Estimation of Demand Function for Wet Market Fish in Singapore

DBA5101 Group Project 1

Submitted by Group 22

Alex Chen Chen (A0280580E)

Toh Yu Qi Chermaine (A0172396W)

Valeria Besse (A0280552H)

Wong Cheuk Wah Jane (A0280543H)

I. Introduction

A wet market is "a market where fresh meat, fish, fruit, vegetables, and sometimes live animals are sold to the public".[1] The term originates from the market's signature slippery floors, as the ice used to keep the products fresh quickly melts. Wet markets are known for containing fresh products delivered directly from suppliers while keeping relatively low costs compared to supermarkets. The number of vendors is limited, and fresh goods are delivered multiple times per week. Some key characteristics of a wet market are the highly differentiated products as well as the wide range of consumer elasticities, varying from small households to restaurants of different price ranges. However, what makes wet markets in Singapore truly unique is the ability to price negotiate with vendors, preventing vendors from fully price discriminating according to their market knowledge and identification of consumer demands. All these factors result in a highly complex market that is often used to study the relationship between buyers and sellers and the ability to estimate a demand function of the market. In this research paper, we will be using a dataset from fish markets in Singapore to attempt to estimate the demand function for fish.

II. Methodology

Fish is often treated as highly differentiated heterogeneous goods due to the large number of fish species, sizes and quality. However, to estimate the demand function of fish sold at the wet market, fish is assumed to be a homogeneous good as there's no data to effectively differentiate the fish variety sold at the market. The capacity limitations in number of vendors in Wet Markets is a barrier to entry, making our model follow the characteristics of imperfect competition. The use of natural logarithm for price and quantity variables allows a better statistical fit as well as greater interpretability, as the coefficient of price represents the elasticity of demand in the model, which can be interpreted as the sensitivity of consumer quantity demanded to a change in price in the good.

In this research paper we focus on three different models: Simple Linear Regression, Multiple Linear Regression and Two Stage Least Squares Regression (2SLS). A simple linear regression using Ordinary Least Squares (OLS) regression allows a baseline linear model of how strong the correlation is between the independent variable (log(price)) and the dependent variable (log(quantity)). A more complex multiple linear regression is then done to further evaluate other independent variables with the goal of reducing the bias on the price variable to get a more accurate estimation of the demand function. Finally, a two stage least squares regression is used to test for endogeneity of the price variable and the use of an instrumental variable to further reduce bias in our model. The instrumental variable is then tested through a Wu-Hausman test to determine its validity.

III. Models and Results

3.1 Correlation Analysis in the Fish Market in Singapore Dataset

By examining the correlations among variables, we understand their potential linear associations for subsequent analysis.

<u>3.1.1 Correlation between *q* and other variables</u>

	Mon	Tue	Wed	Thu	Date	Stormy	Mixed	p	q	Rainy	Cold	Wind
q	0.156964	-0.218871	-0.234749	0.165607	-0.066379	-0.222636	-0.008908	-0.278530	1.000000	0.033979	-0.142820	-0.263995

Most of the variables in the dataset are not strongly correlated with q. The variables *Stormy*, *Wind*, *price (p)*, and the binary variables *Tuesday (Tue)* and *Wednesday (Wed)* may possess some weak explanatory power on q.

<u>3.1.2 Correlation between p and other variables</u>

	Mon	Tue	Wed	Thu	Date	Stormy	Mixed	p	q	Rainy	Cold	Wind
p	-0.074936	-0.018483	-0.009366	0.083350	-0.244170	0.399417	0.065993	1.000000	-0.278530	0.036653	0.241886	0.405123

The independent variable p is moderately correlated with the variables *Stormy* and *Wind*, while weakly correlated with *Date*, q, and *Cold*. These correlations possess the risk of multicollinearity in our model, resulting in either adjustment or omission of the correlated variables to prevent bias on our demand estimation of the fish market.

[1] *wet market*. (2023). https://dictionary.cambridge.org/dictionary/english/wet-market

3.1.3 Correlations of Stormy and Wind with other variables

	Mon	Tue	Wed	Thu	Date	Stormy	Mixed	p	q	Rainy	Cold	Wind
Stormy	0.098818	0.018126	-0.053526	-0.030946	-0.159531	1.000000	-0.422917	0.399417	-0.222636	0.097708	0.392061	0.753123
Wind	0.165220	0.117045	-0.045354	-0.119755	-0.186811	0.753123	-0.046567	0.405123	-0.263995	0.065632	0.450201	1.000000

An important relationship to note is the strong correlation between $Stormy$ and $Wind$ due to both variables being measured by wind speed in Singapore as shown in Appendix 1. Both variables are also moderately correlated with $Cold$ (0.39, 0.45). These results are expected, as they are all weather variables that often appear concurrently on days with more extreme weather conditions. Yet, they raise the risk of multicollinearity when used simultaneously in regression models.

3.2 Ordinary Least Squares Regression

The following regression models are implemented to compare the statistical significance of the variables in explaining the change in q.

Dependent variable: q

	I	II	III	IV
Intercept	8.419***	8.730***	8.665***	10.114***
	(0.076)	(0.104)	(0.181)	(1.529)
p	-0.541***	-0.447**	-0.449**	-0.434**
	(0.179)	(0.180)	(0.195)	(0.195)
Stormy		-0.232	-0.241	-0.044
		(0.151)	(0.189)	(0.280)
Wind				-0.543
				(0.570)
Mon			0.061	0.103
			(0.207)	(0.212)
Tue		-0.545***	-0.489**	-0.458**
		(0.160)	(0.204)	(0.207)
Wed		-0.600***	-0.554***	-0.553***
		(0.165)	(0.208)	(0.208)
Thu			0.087	0.079
			(0.201)	(0.201)
Mixed			0.003	0.078
			(0.162)	(0.181)
Rainy			0.085	0.089
			(0.178)	(0.178)
Cold			0.005	0.037
			(0.142)	(0.146)
R-squared	0.078	0.236	0.239	0.246
R-squared Adj.	0.069	0.207	0.171	0.170

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

Model I: $q = \beta_0 + \beta_1 p + \epsilon$

Model II: $q = \beta_0 + \beta_1 p + \beta_2 Stormy + \beta_3 Tue + \beta_4 Wed + \epsilon$

Model III: $q = \beta_0 + \beta_1 p + \beta_2 Stormy + \beta_3 Mon + \beta_4 Tue + \beta_5 Wed + \beta_6 Thu + \beta_7 Mixed + \beta_8 Rainy + \beta_9 Cold + \epsilon$

Model IV: $q = \beta_0 + \beta_1 p + \beta_2 Stormy + \beta_3 Wind + \beta_4 Mon + \beta_5 Tue + \beta_6 Wed + \beta_7 Thu + \beta_8 Mixed + \beta_9 Rainy + \beta_{10} Cold + \epsilon$

*Both p and q are always used in the logarithm scale in this research paper because the coefficient of $\log(p)$ to estimate $\log(q)$ represents the price elasticity of demand.

3.2.1 Simple OLS

Model I is a simple linear regression model of price estimating quantity. Used as the baseline model, it compares with the more complex models. The results show a negative relationship between price and quantity with a strong significance level of $p < 0.01$, as expected from an estimation of the demand curve. β_1 = -0.541 means that an increase of 1% in price is estimated to decrease the quantity demanded for fish by 0.541%. The coefficient β_1 is the price elasticity of demand, interpreted as the sensitivity of customers to a change in price. Since it is between 0 and -1, it is considered an inelastic demand as the change in quantity demand is lower than the change in price (comparing in absolute terms).

3.2.2 Multiple Linear Regression OLS

Models II - IV test more complex models using multiple linear regression. Model IV uses all variables in the dataset as reference to other models but suffers from overidentification of variables as well as multicollinearity of the correlated variables stated earlier. Model II emphasizes the variables with stronger significant values, while Model III attempts to reduce multicollinearity by omitting the *Wind* variable. Focusing on the coefficient of the price variable, there is no significant change between the multiple linear regression models, although they are all slightly lower than the baseline simple linear regression model. This suggests that the price elasticity of demand becomes slightly more inelastic as the customers become less price sensitive. The result summary shows that only *Tue*, *Wed*, and p have p-value < 0.05 and are also statistically significant across the models, suggesting that holding the weekday constant can reduce the bias in the price coefficient.

3.3 Two-Stage Least Squares (2SLS) Regression Analysis

3.3.1 Simple Regression with Stormy as the instrument Variable

The variable *Stormy* is used as an instrumental variable in the next models to attempt to reduce the endogeneity bias in our price variable and therefore, get a more accurate estimation of the demand function. As *Stormy* is a supply side variable affecting the supply of the fish on those given days, it can be expected that a shortage of fish in wet markets in Singapore will result in higher fish prices, which then results in less quantity demanded for fish. Therefore, *Stormy* is interpreted as an instrumental variable that indirectly affects quantity demanded q through the price variable p (Appendix 2).

In the first stage of 2SLS, we run a regression with p as the dependent variable and $Stormy$ as the independent variable. The coefficients are used to predict p to get $\hat{p}$, and $\hat{p}$ is then inserted back into the regression to predict q in stage 2.

Stage 1

$$\hat{p} = \hat{\pi}_0 + \hat{\pi}_1 Stormy + v$$

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      p   R-squared:                       0.160
Model:                            OLS   Adj. R-squared:                  0.152
Method:                 Least Squares   F-statistic:                     20.69
Date:                Tue, 19 Sep 2023   Prob (F-statistic):           1.41e-05
Time:                        15:13:39   Log-Likelihood:                -40.516
No. Observations:                 111   AIC:                             85.03
Df Residuals:                     109   BIC:                             90.45
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.2903      0.040     -7.336      0.000      -0.369      -0.212
Stormy         0.3353      0.074      4.549      0.000       0.189       0.481
==============================================================================
Omnibus:                        0.924   Durbin-Watson:                   0.634
Prob(Omnibus):                  0.630   Jarque-Bera (JB):                1.036
Skew:                          -0.188   Prob(JB):                        0.596
Kurtosis:                       2.712   Cond. No.                         2.43
==============================================================================
```

Stage 2

$$q = \beta_0 + \beta_1 \hat{p} + \epsilon$$

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      q   R-squared:                       0.050
Model:                            OLS   Adj. R-squared:                  0.041
Method:                 Least Squares   F-statistic:                     5.685
Date:                Tue, 19 Sep 2023   Prob (F-statistic):             0.0188
Time:                        15:13:39   Log-Likelihood:                -121.01
No. Observations:                 111   AIC:                             246.0
Df Residuals:                     109   BIC:                             251.4
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          8.3138      0.112     74.406      0.000       8.092       8.535
p_hat         -1.0824      0.454     -2.384      0.019      -1.982      -0.183
==============================================================================
Omnibus:                       12.308   Durbin-Watson:                   1.625
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               12.886
Skew:                          -0.787   Prob(JB):                      0.00159
Kurtosis:                       3.557   Cond. No.                         6.84
==============================================================================
```

In the first stage regression, the value of F-statistic is > 10 which means that we can reject the null hypothesis that the instrumental variable is weak. The p-values of $Stormy$ is < 0.001, indicating that it is highly statistically significant. Thus, we can conclude that $Stormy$ is a valid instrument for p. In the second stage regression, comparing the coefficient values between OLS estimates and Stage 2 of 2SLS, the coefficient value of $\hat{p}$ is about twice as high with an instrumental variable. It is also worth noting that the coefficient is close to -1, implying a close to unit elasticity of demand. This implies that a 1% increase in price would result in a 1.0824% decrease in quantity demanded. The $\hat{p}$ has a p-value of 0.019, still staying statistically significant at the 5% level.

3.3.2 Multiple Linear Regression with Stormy as the instrument variable and all the Weekdays as the controls

Stage 1

$$\hat{p} = \hat{\pi}_0 + \hat{\pi}_1 Stormy + \hat{\pi}_2 Mon + \hat{\pi}_3 Tue + \hat{\pi}_4 Wed + \hat{\pi}_5 Thu + v$$

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      p   R-squared:                       0.179
Model:                            OLS   Adj. R-squared:                  0.140
Method:                 Least Squares   F-statistic:                     4.575
Date:                Thu, 21 Sep 2023   Prob (F-statistic):           0.000816
Time:                        15:22:56   Log-Likelihood:                -39.223
No. Observations:                 111   AIC:                             90.45
Df Residuals:                     105   BIC:                             106.7
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.2717      0.076     -3.557      0.001      -0.423      -0.120
Stormy         0.3464      0.075      4.639      0.000       0.198       0.494
Mon           -0.1129      0.107     -1.052      0.295      -0.326       0.100
Tue           -0.0411      0.105     -0.394      0.695      -0.248       0.166
Wed           -0.0118      0.107     -0.111      0.912      -0.224       0.200
Thu            0.0496      0.104      0.475      0.636      -0.157       0.257
==============================================================================
Omnibus:                        1.265   Durbin-Watson:                   0.632
Prob(Omnibus):                  0.531   Jarque-Bera (JB):                1.352
Skew:                          -0.222   Prob(JB):                        0.509
Kurtosis:                       2.691   Cond. No.                         5.98
==============================================================================
```

Stage 2

$$q = \beta_0 + \beta_1 \hat{p} + \beta_2 Mon + \beta_3 Tue + \beta_4 Wed + \beta_5 Thu + \epsilon$$

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      q   R-squared:                       0.193
Model:                            OLS   Adj. R-squared:                  0.155
Method:                 Least Squares   F-statistic:                     5.034
Date:                Thu, 21 Sep 2023   Prob (F-statistic):           0.000356
Time:                        15:25:05   Log-Likelihood:                -111.90
No. Observations:                 111   AIC:                             235.8
Df Residuals:                     105   BIC:                             252.1
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          8.5059      0.161     52.882      0.000       8.187       8.825
p_hat         -1.1194      0.415     -2.698      0.008      -1.942      -0.297
Mon           -0.0254      0.208     -0.122      0.903      -0.438       0.387
Tue           -0.5308      0.201     -2.636      0.010      -0.930      -0.132
Wed           -0.5664      0.206     -2.750      0.007      -0.975      -0.158
Thu            0.1093      0.202      0.541      0.590      -0.291       0.510
==============================================================================
Omnibus:                       15.952   Durbin-Watson:                   1.546
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               19.150
Skew:                          -0.811   Prob(JB):                     6.95e-05
Kurtosis:                       4.228   Cond. No.                         7.19
==============================================================================
```

In the final and most complex model, the first stage regression uses *Stormy* again as an instrumental variable to estimate $\hat{p}$, but includes the exogenous binary variables *Mon – Thu* to hold the day of the week constant. Holding weekdays unchanged, a 1% increase in price decreases the quantity sold by around 1.12 %, which is slightly higher than the coefficient in the simple 2SLS model. Since the absolute value of the coefficient β_1 is now greater than 1, the elasticity of demand of the function is slightly elastic, showing a greater price sensitivity in consumers. This varies greatly from the β_1 coefficients in the OLS models. Although the F-statistic drops below 10 in the first stage, the p-value of $\hat{p}$ in the second stage is < 0.01, becoming strongly statistically significant. The results suggest that keeping the day of the week constant can remove bias from the price variable and result in a better explanatory variable to estimate quantity demanded. Although Tuesday and Wednesday were the only significant days in our model, we kept all weekdays to increase interpretability and allow our model to be more generalizable to other fish markets around the world where Mondays and Thursdays may also have statistical significance in their datasets due to differences in market structures and supply chain logistics.

3.4 Endogeneity Test (Wu-Hausman Test)

The Wu-Hausman test is used to compare the efficiency of the OLS multilinear regression model compared to the 2SLS multilinear regression model. The null hypothesis is that the price variable is exogenous (not correlated with error term) and therefore, OLS assumptions have not been violated.

```
Wu-Hausman test of exogeneity
H0: All endogenous variables are exogenous
Statistic: 12.5483
P-value: 0.0006
Distributed: F(1,105)
```

The results show that the F-statistic is highly significant (> 10), rejecting the null hypothesis and supporting the use of the instrumental variable *Stormy* to reduce the endogeneity bias of *p*. The test suggests that the 2SLS multilinear regression is the most efficient and unbiased model to estimate the demand function for fish in the wet market.

IV. Limitations

A limitation in the research paper is the inability to differentiate different customer groups. Households and businesses (restaurants) tend to have very different spending patterns and demand elasticities. Households tend to buy lower quantities with a more elastic demand, while business tend to buy higher quantities with a more inelastic demand. Merging the consumer groups results in an aggregated demand function that may not be representative of the demand function of either consumer group.

One of the main characteristics of a wet market is the ability to negotiate prices. Vendors engage in price discrimination, resulting in the possibility of similar fish being sold at completely different prices. The ability to set prices and price discriminate along with the heterogeneity nature of fish would result in a model resembling monopolistic competition. The limited timeframe as well as relatively small datasets may also serve as limitations in generalizing our results to other food markets. The timeframe is only between 1991 to 1992, limiting the ability to effectively measure how different seasons or months can influence the quantity demanded for fish.

V. Conclusion

Our results showed the need of the instrumental variable *Stormy* in our 2SLS regression model due to the endogeneity of the *p* variable when estimating quantity *q*. The weather conditions at the shore have an important effect on availability and scarcity of fish in wet markets, driving up prices due to the pricing power of vendors. The change in price will then affect the consumers' demand for fish, especially private households that have lower price elasticities and certain expectations of the prices of fish. By including all the weekdays in our final model, we hope to increase generalizability of our model in markets outside of Singapore where different market structures could result in other days of the week being significant. The generalizability of the model may also depend on the size of the markets, as an increase in vendors may reduce the pricing power of merchants, resulting in less price discrimination due to the availability of substitutes. All these cases result in a highly complex market that may vary based on the market regulations in the country and supply volatility such as weather outcomes. Fish being a highly differentiated and perishable good adds to this complexity, resulting in a demand function that may be more price sensitive and fluctuating than fish sold in regular grocery stores.

APPENDIX

1. Variables in the dataset and their explanations

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

2. Moderate and strong Correlation between variables