DBA5101

Group Project 1

Estimation of Demand Function of Fish in Singapore Wet Markets

Kenny Lim (A0262753B)

Manan Lohia (A0262838U)

Peiwen Jiang (A0262785R)

Jung Hyun Park (A0262660H)

Janita Bose (A0262823E)

Aim

This paper aims to provide insights on the demand conditions of the wet fish market to the Ministry for Sustainability and Environment (MSE) by identifying variables that effectively affect the demand of fish, for policy making to minimise wastage and improve hygiene.

Executive Summary

In light of COVID, hygiene has become a primary concern amongst citizens and an area of concern governed by MSE are the wet markets. While wet markets possess the charm of Southeast Asian countries, they are often hot and humid. These conditions present a breeding ground for diseases especially in the presence of perishable goods such as fish. It is important for MSE to estimate demand to reduce wastage and simultaneously maintain high standards of hygiene.

Typically, we are unable to determine the effect of price on the quantity of fish as they are simultaneously determined. However, we have identified weather conditions at sea ("Stormy" i.e. high wind speed and wave lengths) as an instrument variable (IV) that allows us to isolate the effects of price on quantity since the IV only affects quantity through price. We found that 1% increase in price leads to approximately 0.94% decrease in quantity demanded. These findings could potentially be used to formulate policies that utilise price as a lever.

Lastly, we also identified other data that could allow us to better understand demand such as fish varieties and the alternative sources of supply e.g. stockpile or import from neighbouring countries.

Demand Function

$$\ln(q) = \beta_0 + \beta_1 \ln(p) + \beta_2 Mon + \beta_3 Tue + \beta_4 Wed + \beta_5 Thur + \beta_6 Rainy + \beta_7 Cold + \epsilon$$

From the data provided, we identified demand side factors such as days of the week and onshore weather conditions. Days of the week were included in the estimate of the demand curve, in the form of dummy variables e.g. Mon, Tues, ... Fri, which accounted for the varying consumption throughout the week. Onshore weather conditions such as "Rainy" and "Cold" weather, affect buyers' motivation to visit the wet market. Finally, we decided to exclude the "Date" variable from the demand function as the timeframe of the data only ranges from Dec to May and was not sufficient to reflect seasonality.

Supply Function

$$\ln(q) = \alpha_0 + \alpha_1 ln(p) + \alpha_2 Stormy + \alpha_3 Mixed + \alpha_4 Wind + \nu$$

Log price and log quantity are also included in the supply function as they are determined simultaneously by both supply and demand shifts. We also included "Mixed" and "Stormy" as one-hot encoded variables as well as "Wind" (likely correlated to "Stormy" and "Mixed"), as they are intuitively critical weather conditions that makes fishing difficult and affects supply quantity.

Naive Model (NM)

We first developed a Naive model[1] containing purely demand side factors. This also ensured that at least one variable ("Stormy" and "Mixed") is omitted from the demand function i.e. necessary condition of identification, to ensure that shifts in demands are not a result of supply factors and vice versa.

Using OLS to estimate demand, we established a baseline for comparison against subsequent models. The results are shown in *Figure 1* below. The coefficient of p is -0.5446 and has a p-value of p is 0.002 which supports our intuition that price and quantity are inversely related.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.6169	0.162	53.327	0.000	8.296	8.937
p	-0.5446	0.175	-3.108	0.002	-0.892	-0.197
Mon	0.0316	0.207	0.153	0.879	-0.378	0.441
Tue	-0.4935	0.204	-2.425	0.017	-0.897	-0.090
Wed	-0.5392	0.206	-2.617	0.010	-0.948	-0.131
Thu	0.0948	0.201	0.471	0.639	-0.304	0.494
Rainy	0.0666	0.177	0.375	0.708	-0.285	0.419
Cold	-0.0616	0.134	-0.458	0.648	-0.328	0.205

Figure 1: OLS of Naive Model

Improved Model (IM)

Reduced Form

$$\ln(\hat{p}) = \hat{\pi}_0 + \hat{\pi}_1 Mon + \hat{\pi}_2 Tue + \hat{\pi}_3 Wed + \hat{\pi}_4 Thur + \hat{\pi}_5 Rainy + \hat{\pi}_6 Cold + \hat{\pi}_7 Stormy + \hat{\pi}_8 Mixed$$

In our improved model, we selected an instrumental variable that is present in the supply function but not the demand function e.g. "Stormy" and "Mixed". Next, we ran OLS using the reduced form equation to estimate log(p).

The results of the 2SLS - first stage are shown in *Figure 2* below. The coefficients of Stormy and Mixed are jointly significant with p-values of 0.000 and 0.004 respectively. Joint significance of the two IVs implies that supply has significant shift variables and can be used to reliably estimate the demand function.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.3739	0.084	-4.436	0.000	-0.541	-0.207
Stormy	0.4167	0.087	4.808	0.000	0.245	0.589
Mixed	0.2311	0.079	2.914	0.004	0.074	0.388
Mon	-0.1141	0.105	-1.088	0.279	-0.322	0.094
Tue	-0.0762	0.104	-0.736	0.463	-0.282	0.129
Wed	-0.0608	0.105	-0.577	0.565	-0.270	0.148
Thu	0.0334	0.102	0.327	0.744	-0.169	0.236
Rainy	-0.0049	0.090	-0.054	0.957	-0.184	0.175
Cold	0.0635	0.072	0.881	0.380	-0.079	0.206

Figure 2: 2SLS - First Stage of Improved Model

Weak Instruments Test

We further back our claims by performing the weak instruments test[3] to prove that our predicted p is a suitable IV for estimating demand. The weak instruments test shows that the reduced form model is significant with an F-Stat p-value of 0.000189.

2SLS

$$\ln(q) = \beta_0 + \beta_1 \ln(\hat{p}) + \beta_2 Mon + \beta_3 Tue + \beta_4 Wed + \beta_5 Thur + \beta_6 Rainy + \beta_7 Cold + \epsilon_1$$

With our IVs validated, we proceed to implement 2SLS[4]. We use the predicted values of log(p) from the OLS of the reduced form to replace the original observed prices from the dataset i.e. endogenous variable.

The second stage results (see *Figure 3 below*) show that predicted prices of fish are statistically significant, with a p-value of 0.022. This translates to a 2.2% chance the predicted price variable has no effect on the dependent variable, quantity, and our results are produced by chance (highly unlikely). The coefficient of p^ informs us that for every 1% increase in price there is a 0.947% decrease in quantity demanded. Tue and Wed (dummy variables for day of the week) are negative and statistically insignificant, accounting for lower demand on these days compared to Fri.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.5130	0.190	44.791	0.000	8.136	8.890
p	-0.9470	0.408	-2.320	0.022	-1.756	-0.137
Mon	-0.0069	0.214	-0.032	0.974	-0.430	0.417
Tue	-0.5168	0.209	-2.478	0.015	-0.930	-0.103
Wed	-0.5608	0.211	-2.658	0.009	-0.979	-0.142
Thu	0.1085	0.205	0.528	0.599	-0.299	0.516
Rainy	0.0698	0.181	0.386	0.700	-0.289	0.429
Cold	0.0153	0.154	0.100	0.921	-0.290	0.321

Figure 3: 2SLS - Second stage of Improved Model

Hausman Test

We performed the Hausman Test[2] to test the existence of the endogeneity problem. While the results failed to prove endogeneity (likely due to small sample size), we know from our business intuition and domain knowledge that price and quantity are simultaneously determined and observed only at equilibrium levels, and hence cannot be used directly to predict demand.

Forward Selection of other features

In the beginning, we classified Wind as another supply side variable. We tested using Wind alone as an IV instead of Stormy and Mixed as an indicator of weather conditions and the tests suggested it to be significant (See Appendix). However, when used alongside Stormy and Mixed as an IV, the weak instruments test suggested Wind to be statistically insignificant[5]. Intuitively, this made sense as the Stormy and Mixed together encode information about wind speed as well as wave heights, and are hence a better indicator of weather conditions than wind alone.

Dependent Variable	**p**	
Independent Variable	Improved model	Improved model + Wind
Stormy	0.4167*** (0.087)	0.3304* (0.139)
Mixed	0.2311** (0.079)	0.1974* (0.030)
Wind		0.2312 (0.289)

Figure 3: Comparison of Forward Selection Models

Next Steps

In conclusion, our research identified weather conditions at sea such as “Stormy” and “Mixed” as IVs that allows us to effectively estimate the relationship between price and quantity of fish. This model could be utilised to influence policies to minimise wastage and improve hygiene levels in wet markets.

Moving forward, we recommend collecting more data i.e. more features. In this study, we were limited by the features provided and therefore limited in the choice of IVs. Other IVs for consideration include features about the fish sold in the wet market such as the sales of each variety, stockpile of unsold fishes, location of the market, and fish imports etc.

Appendix

A.1 Full comparison between Improved Model and Forward Selection Model

Dependent Variable	**p**	
Independent Variable	Improved model	Improved model + Wind
Intercept	-0.3739 (0.084)	-0.9878 (0.772)
Mon	-0.1141 (0.105)	-0.1315 (0.107)
Tue	-0.0762 (0.104)	-0.0889 (0.105)
Wed	-0.0608 (0.105)	-0.0608 (0.106)
Thur	0.0334 (0.102)	0.0365 (0.102)
Rainy	-0.0049 (0.090)	-0.0065 (0.091)
Cold	0.0635 (0.072)	0.0494 (0.074)
Stormy	0.4167*** (0.087)	0.3304* (0.139)
Mixed	0.2311** (0.079)	0.1974* (0.030)
Wind		0.2312 (0.289)

Appendix

September 28, 2022

In [2]:
```
import pandas as pd
import statsmodels.formula.api as smf
import scipy.stats as stats
import matplotlib.pyplot as plt

class TwoSLS:

    def __init__(self, path):
        self.path = path
        self.data = pd.read_csv(path)

    def OLS(self, model):
        model = smf.ols(formula = model, data = self.data).fit()
        print(model.summary())
        return model

    def SLS(self, structural, reduced):

        first_stage = smf.ols(formula = reduced, data = self.data).fit()
        self.data_stage_2 = self.data.assign(p = first_stage.fittedvalues)

        second_stage = smf.ols(formula = structural, data = self.data_stage_2).fit()
        print('\n Summary of First Stage')
        print(first_stage.summary())
        print('\n Summary of Second Stage')
        print(second_stage.summary())
        return [first_stage, second_stage]

    def weak_instr_test(self, model):
        model = smf.ols(formula = model, data = self.data).fit()
        print('\n Summary of Weak Instruments Test')
        print(model.summary())

    def hausman(self, structural, reduced):
        structural = structural + ' + v'
        OLS = smf.ols(formula = reduced, data = self.data).fit()
        self.data_hausmann = self.data.assign(v = OLS.resid)
```

```
            hausman = smf.ols(formula = structural, data = self.data_hausmann).fit()
            print(hausman.summary())

        def sargan(self,structural, reduced, instrumental, dof):
            first_stage,second_stage = self.SLS(structural, reduced)
            self.data_sargan = self.data.assign(resid = second_stage.resid)

            instrumental = 'resid ~ ' + instrumental
            sargan_results = smf.ols(instrumental, data = self.data_sargan).fit()
            nobs = sargan_results.nobs
            rsqaured = sargan_results.rsquared
            print('\n Summary of Sargan Test')
            print(sargan_results.summary())
            print('\n The p-value of the test statisticis: {}'.format((1 - stats.chi2.cdf(r

    model = TwoSLS('Data-GP1.csv')
```

0.0.1 [1] Naive Model

```
In [6]: model.OLS('q ~ p + Mon + Tue + Wed + Thu + Rainy + Cold')
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      q   R-squared:                       0.223
Model:                            OLS   Adj. R-squared:                  0.170
Method:                 Least Squares   F-statistic:                     4.220
Date:                Tue, 27 Sep 2022   Prob (F-statistic):           0.000396
Time:                        15:13:47   Log-Likelihood:                -109.83
No. Observations:                 111   AIC:                             235.7
Df Residuals:                     103   BIC:                             257.3
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      8.6169      0.162     53.327      0.000       8.296       8.937
p             -0.5446      0.175     -3.108      0.002      -0.892      -0.197
Mon            0.0316      0.207      0.153      0.879      -0.378       0.441
Tue           -0.4935      0.204     -2.425      0.017      -0.897      -0.090
Wed           -0.5392      0.206     -2.617      0.010      -0.948      -0.131
Thu            0.0948      0.201      0.471      0.639      -0.304       0.494
Rainy          0.0666      0.177      0.375      0.708      -0.285       0.419
Cold          -0.0616      0.134     -0.458      0.648      -0.328       0.205
==============================================================================
Omnibus:                       13.824   Durbin-Watson:                   1.491
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               15.126
Skew:                          -0.790   Prob(JB):                     0.000519
```

```
Kurtosis:                       3.880   Cond. No.                         6.74
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Out[6]: <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x29ef5310970>

0.0.2 [2] Hausman Test

```
In [10]: model.hausman('q ~ p + Mon + Tue + Wed + Thu + Rainy + Cold', 'p~ Stormy + Mixed + Mon
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      q   R-squared:                       0.233
Model:                            OLS   Adj. R-squared:                  0.173
Method:                 Least Squares   F-statistic:                     3.869
Date:                Tue, 27 Sep 2022   Prob (F-statistic):           0.000514
Time:                        15:21:29   Log-Likelihood:                -109.12
No. Observations:                 111   AIC:                             236.2
Df Residuals:                     102   BIC:                             260.6
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      8.5276      0.179     47.612      0.000       8.172       8.883
p             -0.9258      0.375     -2.466      0.015      -1.671      -0.181
Mon           -0.0011      0.208     -0.005      0.996      -0.414       0.412
Tue           -0.5105      0.204     -2.506      0.014      -0.915      -0.106
Wed           -0.5548      0.206     -2.691      0.008      -0.964      -0.146
Thu            0.1104      0.201      0.549      0.585      -0.289       0.510
Rainy          0.0729      0.177      0.411      0.682      -0.279       0.425
Cold          -0.0144      0.140     -0.102      0.919      -0.293       0.264
v              0.4790      0.417      1.148      0.254      -0.349       1.307
==============================================================================
Omnibus:                       14.250   Durbin-Watson:                   1.479
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               15.792
Skew:                          -0.799   Prob(JB):                     0.000372
Kurtosis:                       3.927   Cond. No.                         10.8
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

0.0.3 [3] Weak Instruments Test

```
In [18]: model.weak_instr_test('p ~ Mon + Tue + Wed + Thu + Stormy + Mixed + Wind')
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      q   R-squared:                       0.193
Model:                            OLS   Adj. R-squared:                  0.154
Method:                 Least Squares   F-statistic:                     5.018
Date:                Tue, 13 Sep 2022   Prob (F-statistic):           0.000366
Time:                        13:04:41   Log-Likelihood:                -111.94
No. Observations:                 111   AIC:                             235.9
Df Residuals:                     105   BIC:                             252.1
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      8.4957      0.163     52.162      0.000       8.173       8.819
p_fitted      -0.9546      0.356     -2.685      0.008      -1.659      -0.250
Mon            0.0992      0.207      0.480      0.632      -0.310       0.509
Tue           -0.4520      0.202     -2.238      0.027      -0.852      -0.052
Wed           -0.5071      0.206     -2.458      0.016      -0.916      -0.098
Thu            0.0660      0.201      0.328      0.744      -0.333       0.465
==============================================================================
Omnibus:                       14.914   Durbin-Watson:                   1.539
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               17.348
Skew:                          -0.785   Prob(JB):                     0.000171
Kurtosis:                       4.134   Cond. No.                         6.48
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

0.0.4 [4] 2SLS on Improved Model

```
In [12]: model.SLS('q ~ p + Mon + Tue + Wed + Thu + Rainy + Cold', 'p ~ Stormy + Mixed + Mon +
```

```
 Summary of First Stage
                            OLS Regression Results
==============================================================================
Dep. Variable:                      p   R-squared:                       0.251
Model:                            OLS   Adj. R-squared:                  0.192
Method:                 Least Squares   F-statistic:                     4.265
Date:                Tue, 27 Sep 2022   Prob (F-statistic):           0.000189
Time:                        15:37:00   Log-Likelihood:                -34.145
No. Observations:                 111   AIC:                             86.29
```

```
Df Residuals:                     102   BIC:                             110.7
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.3739      0.084     -4.436      0.000      -0.541      -0.207
Stormy         0.4167      0.087      4.808      0.000       0.245       0.589
Mixed          0.2311      0.079      2.914      0.004       0.074       0.388
Mon           -0.1141      0.105     -1.088      0.279      -0.322       0.094
Tue           -0.0762      0.104     -0.736      0.463      -0.282       0.129
Wed           -0.0608      0.105     -0.577      0.565      -0.270       0.148
Thu            0.0334      0.102      0.327      0.744      -0.169       0.236
Rainy         -0.0049      0.090     -0.054      0.957      -0.184       0.175
Cold           0.0635      0.072      0.881      0.380      -0.079       0.206
==============================================================================
Omnibus:                        2.185   Durbin-Watson:                   0.672
Prob(Omnibus):                  0.335   Jarque-Bera (JB):                2.123
Skew:                          -0.332   Prob(JB):                        0.346
Kurtosis:                       2.862   Cond. No.                         7.10
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

 Summary of Second Stage
                            OLS Regression Results
==============================================================================
Dep. Variable:                      q   R-squared:                       0.192
Model:                            OLS   Adj. R-squared:                  0.137
Method:                 Least Squares   F-statistic:                     3.501
Date:                Tue, 27 Sep 2022   Prob (F-statistic):            0.00208
Time:                        15:37:00   Log-Likelihood:                -111.98
No. Observations:                 111   AIC:                             240.0
Df Residuals:                     103   BIC:                             261.6
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      8.5130      0.190     44.791      0.000       8.136       8.890
p             -0.9470      0.408     -2.320      0.022      -1.756      -0.137
Mon           -0.0069      0.214     -0.032      0.974      -0.430       0.417
Tue           -0.5168      0.209     -2.478      0.015      -0.930      -0.103
Wed           -0.5608      0.211     -2.658      0.009      -0.979      -0.142
Thu            0.1085      0.205      0.528      0.599      -0.299       0.516
Rainy          0.0698      0.181      0.386      0.700      -0.289       0.429
Cold           0.0153      0.154      0.100      0.921      -0.290       0.321
```

```
==============================================================================
Omnibus:                       14.766   Durbin-Watson:                   1.520
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               17.077
Skew:                          -0.782   Prob(JB):                     0.000196
Kurtosis:                       4.116   Cond. No.                         8.23
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
Out[12]: [<statsmodels.regression.linear_model.RegressionResultsWrapper at 0x29ef53de490>,
          <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x29efa4ad280>]
```

0.0.5 [5] Forward Selection

```
In [13]: model.SLS('q ~ p + Mon + Tue + Wed + Thu + Rainy + Cold', 'p ~ Stormy + Mixed + Mon +
```

```
 Summary of First Stage
                            OLS Regression Results
==============================================================================
Dep. Variable:                      p   R-squared:                       0.255
Model:                            OLS   Adj. R-squared:                  0.189
Method:                 Least Squares   F-statistic:                     3.849
Date:                Tue, 27 Sep 2022   Prob (F-statistic):           0.000323
Time:                        15:38:51   Log-Likelihood:                -33.795
No. Observations:                 111   AIC:                             87.59
Df Residuals:                     101   BIC:                             114.7
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.9878      0.772     -1.279      0.204      -2.520       0.544
Stormy         0.3304      0.139      2.385      0.019       0.056       0.605
Mixed          0.1974      0.090      2.195      0.030       0.019       0.376
Mon           -0.1315      0.107     -1.225      0.223      -0.344       0.081
Tue           -0.0889      0.105     -0.848      0.399      -0.297       0.119
Wed           -0.0608      0.106     -0.576      0.566      -0.270       0.149
Thu            0.0365      0.102      0.357      0.722      -0.167       0.240
Rainy         -0.0065      0.091     -0.071      0.943      -0.186       0.173
Cold           0.0494      0.074      0.665      0.508      -0.098       0.197
Wind           0.2312      0.289      0.800      0.426      -0.342       0.805
==============================================================================
Omnibus:                        2.463   Durbin-Watson:                   0.677
Prob(Omnibus):                  0.292   Jarque-Bera (JB):                2.355
Skew:                          -0.353   Prob(JB):                        0.308
```

```
Kurtosis:                       2.895   Cond. No.                         80.2
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

 Summary of Second Stage
                            OLS Regression Results
==============================================================================
Dep. Variable:                      q   R-squared:                       0.198
Model:                            OLS   Adj. R-squared:                  0.143
Method:                 Least Squares   F-statistic:                     3.628
Date:                Tue, 27 Sep 2022   Prob (F-statistic):            0.00155
Time:                        15:38:51   Log-Likelihood:                -111.60
No. Observations:                 111   AIC:                             239.2
Df Residuals:                     103   BIC:                             260.9
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      8.5007      0.189     45.062      0.000       8.127       8.875
p             -0.9945      0.401     -2.477      0.015      -1.791      -0.198
Mon           -0.0114      0.213     -0.054      0.957      -0.433       0.410
Tue           -0.5196      0.208     -2.500      0.014      -0.932      -0.107
Wed           -0.5633      0.210     -2.680      0.009      -0.980      -0.146
Thu            0.1101      0.205      0.538      0.592      -0.296       0.516
Rainy          0.0702      0.180      0.389      0.698      -0.287       0.428
Cold           0.0244      0.153      0.160      0.873      -0.279       0.328
==============================================================================
Omnibus:                       14.243   Durbin-Watson:                   1.542
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               16.354
Skew:                          -0.761   Prob(JB):                     0.000281
Kurtosis:                       4.105   Cond. No.                         8.12
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
Out[13]: [<statsmodels.regression.linear_model.RegressionResultsWrapper at 0x29ef53101f0>,
          <statsmodels.regression.linear_model.RegressionResultsWrapper at 0x29efb47e400>]
```