FISH PRICES AT THE FULTON FISH MARKET IN NEW YORK CITY

FIN-403 Econometrics

Assignment 2
Group 12

Professor
Prof. Andreas Fuster

Teaching Assistant
Federico Baldi Lanfranchi

Authors
Charalampos Salis (367920)
Ribeiro de Carvalho Paulo (314985)

December 2, 2023

Part 1

a)

The output of the linear regression model is presented in Table 1(1). Since none of the coefficients of the dummy variables is statistically significant, we conclude that there is no significant evidence in the data that the price varies systematically within a week.

Table 1: Regression results for the logarithm of the average price

	Dependent variable:	
	Log Average Price	
	(1)	(2)
Monday (0/1)	−0.010	−0.018
	(0.129)	(0.114)
Tuesday (0/1)	−0.009	−0.009
	(0.127)	(0.112)
Wednesday (0/1)	0.038	0.050
	(0.126)	(0.112)
Thursday (0/1)	0.091	0.123
	(0.126)	(0.111)
Time (linear trend)	−0.004***	−0.001
	(0.001)	(0.001)
Wave height (avg max 2 days)		0.091***
		(0.022)
Wave height (avg max 3 & 4 day lagged)		0.047**
		(0.021)
Constant	−0.073	−0.920***
	(0.115)	(0.190)
Observations	97	97
R^2	0.085	0.309
Adjusted R^2	0.035	0.255
Residual Std. Error	0.397 (df = 91)	0.349 (df = 89)
F Statistic	1.701 (df = 5; 91)	5.699*** (df = 7; 89)
Note:		*p<0.1; **p<0.05; ***p<0.01

b)

The results of the new model with the added variables wave2 and wave3 are presented in Table 1(2). It is apparent that, individually, both variables are statistically significant at 5% significance level. To test joint significance, we perform an F-test, comparing our model with the restricted model (i.e., without wave2 and wave3). The p-value of this test is close to 0 ($3.72 \cdot 10^{-6}$), which implies that there is evidence of joint significance in the data.

Stormy seas could influence the price of the fish through the law of supply and demand. More specifically, since rough seas render fishing more challenging, bad weather might affect the supply of fish and, consequently, their price.

c)

The coefficient of the time trend became statistically insignificant in the second model. A possible explanation is omitted variable bias, i.e., our model was not representative of reality. When omitted variable bias is present, it can lead to incorrect conclusions in hypothesis testing (a coefficient that should be statistically significant may appear insignificant or vice versa).

d)

The plot of the time series of the residuals from the regression model fitted in (b) is presented in Figure 1. The plot suggests that there is positive autocorrelation in the time series of the residuals, while also exhibiting mean-reversion.

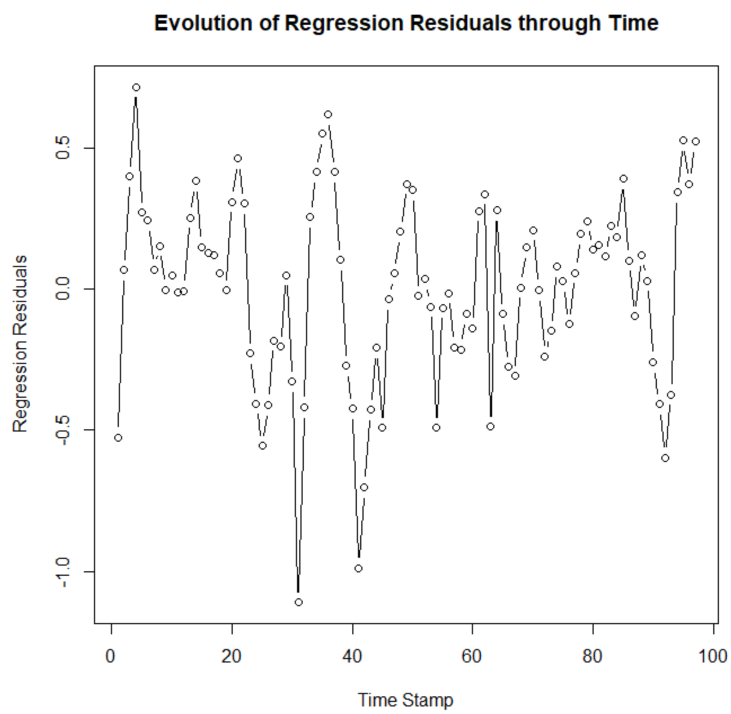

Figure 1: Evolution of the regression residuals of the model presented in (b) through time.

We can also test formally for AR(1) autocorrelation using the Durbin-Watson and the Breusch-Godfrey tests. Both tests suggest that there is evidence of autocorrelation in the data. Since there is significant evidence that autocorrelation is present, the OLS estimator is not BLUE anymore. Assuming that strict exogeneity holds (i.e., our model properly captures the way causation works), the estimators continue to be unbiased. The standard errors are also wrong and need to be corrected.

e)

At a significance level of 5%, both wave2 and wave3 variables remain statistically significant after the Newey-West correction of the standard errors. Since we found significant evidence of autocorrelation in (d), we expect that the new tests are more reliable than the usual OLS t-statistics.

f)

By performing the same F-test with (b) of the Prais-Wisten model, presented in Table 2(1), we reject the null hypothesis that both coefficients are zero. Hence, there is evidence in the data that the two variables are jointly significant. On the other hand, in order to estimate the magnitude of the autocorrelation in the error term, we regress e_t on e_{t-1} and obtain that: $\hat{\rho} = 0.687$, as presented in Table 2(2).

Table 2

	Dependent variable:	
	Log Average Price	Residuals
	(1)	(2)
Monday (0/1)	0.010	
	(0.911)	
Tuesday (0/1)	0.003	
	(0.980)	
Wednesday (0/1)	0.062	
	(0.540)	
Thursday (0/1)	0.117	
	(0.168)	
Time (linear trend)	−0.001	
	(0.855)	
Wave height (avg max 2 days)	0.050*	
	(0.038)	
Wave height (avg max 3 & 4 day lagged)	0.032	
	(0.176)	
Lagged Residuals		0.687***
		(0.076)
Constant	−-0.658*	
	(0.046)	
Observations	97	96
R^2	0.134	0.460
Adjusted R^2	0.066	0.454
Residual Std. Error	0.362 (df = 89)	0.256 (df = 95)
F Statistic	1.973 (df = 7; 89)	80.922*** (df = 1; 95)
Note:		*p<0.1; **p<0.05; ***p<0.01

Part 2

g)

The output of the model is presented in Table 3. The coefficient of log average price is -0.549, which implies that if the log average price increases by one, then we expect the total quantity to decrease by 54.9%. From an economic perspective, the coefficient of log(price) is negative, which implies that as price increases, we expect the quantity to decrease and is correct from an economical (supply and demand) standpoint. However, this is problematic from a modelling perspective, since this is the case of simultaneity and reverse causality.

Table 3

	Dependent variable:
	Log Total Quantity
Log Average Price	-0.549^{***}
	(0.184)
Monday (0/1)	-0.318
	(0.227)
Tuesday (0/1)	-0.684^{***}
	(0.224)
Wednesday (0/1)	-0.535^{**}
	(0.221)
Thursday (0/1)	0.068
	(0.221)
Time (linear trend)	-0.001
	(0.003)
Constant	8.301^{***}
	(0.203)
Observations	97
R^2	0.219
Adjusted R^2	0.167
Residual Std. Error	0.698 (df = 90)
F Statistic	4.201^{***} (df = 6; 90)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

h)

A candidate instrument variable must satisfy two conditions:

1. Relevance condition: It is logical to assume that the wave influences the price of the product (since rough seas make fishing more challenging, bad weather might affect the supply of fish and, consequently, their price).
2. Exclusion restriction: This assumption is difficult to establish without a thorough understanding of potential confounding variables, which requires a combination of further statistical analysis, economic reasoning and sensitivity analyses.

i)

The results of the first-stage regression are presented in Table 4(1). Following the rule of thumb that the t-value must be greater than or equal to 3.16 (since we only have one instrument), we conclude that there is significant evidence in the data that the wave2 variable is a strong instrument (assuming that the 2 assumptions described above hold), since it has a t-value of 4.758.

Table 4: Results for different first-stage regressions with log average price as endogenous variable.

	Dependent variable:		
	Log Average Price		
	(1)	(2)	(3)
Wave height (avg max 2 days)	0.103^{***}		0.093^{***}
	(0.022)		(0.022)
Max wind speed (3-day lagged)		0.017^{**}	0.009
		(0.007)	(0.006)
Monday (0/1)	−0.036	−0.031	−0.045
	(0.116)	(0.126)	(0.116)
Tuesday (0/1)	0.007	−0.086	−0.036
	(0.114)	(0.127)	(0.118)
Wednesday (0/1)	0.083	−0.001	0.057
	(0.113)	(0.123)	(0.114)
Thursday (0/1)	0.136	0.098	0.136
	(0.113)	(0.122)	(0.113)
Time (linear trend)	−0.002	-0.003^{**}	−0.002
	(0.001)	(0.001)	(0.001)
Constant	-0.706^{***}	-0.444^{**}	-0.846^{***}
	(0.169)	(0.183)	(0.194)
Observations	97	97	97
R^2	0.269	0.148	0.286
Adjusted R^2	0.221	0.091	0.229
Residual Std. Error	0.357 (df = 90)	0.386 (df = 90)	0.355 (df = 89)
F Statistic	5.528^{***} (df = 6; 90)	2.602^{**} (df = 6; 90)	5.084^{***} (df = 7; 89)
Note:			*p<0.1; **p<0.05; ***p<0.01

j)

The results of the IV regression using wave2 as instrument are presented in Table 5(1). It is apparent that after the addition of the instrumental variable, we expect that the price will reduce the quantity of a product even further.

The difference between OLS and IV estimates typically arises when there is endogeneity in the model and the OLS estimator is biased (as it fails to account for the correlation between the endogenous variable and the error term).

Table 5: Results for the regression of log total quantity using different instruments for the log average price.

	Dependent variable:		
	Log Total Quantity		
	(1)	(2)	(3)
Log average price (endogenous)	-0.960^{**}	-1.960^{**}	-1.127^{***}
	(0.422)	(0.907)	(0.415)
Monday (0/1)	-0.322	-0.332	-0.323
	(0.233)	(0.292)	(0.239)
Tuesday (0/1)	-0.687^{***}	-0.696^{**}	-0.689^{***}
	(0.230)	(0.288)	(0.236)
Wednesday (0/1)	-0.520^{**}	-0.482^{*}	-0.514^{**}
	(0.227)	(0.286)	(0.233)
Thursday (0/1)	0.106	0.196	0.121
	(0.230)	(0.295)	(0.236)
Time (linear trend)	-0.003	-0.007	-0.004
	(0.003)	(0.005)	(0.003)
Constant	8.271^{***}	8.198^{***}	8.259^{***}
	(0.210)	(0.268)	(0.215)
Observations	97	97	97
R^2	0.175	-0.291	0.133
Adjusted R^2	0.120	-0.377	0.075
Residual Std. Error (df = 90)	0.717	0.897	0.735
Note:	*p<0.1; **p<0.05; ***p<0.01		

The standard errors for the OLS and IV regression are 0.184138 and 0.421982, respectively. We know that the IV estimator is less precise than the OLS one, especially if x and z are only weakly correlated and we only have a small sample size. On the other hand, in OLS, standard errors may be smaller due to biased estimates (due to endogeneity).

k)

The results of the regression using speed3 as instrument are presented in Table 5(2). By examining the first-stage regression (Table 4(2)), we obtain a t-value equal to 2.565, which implies that speed3 is not a very strong instrument.

l)

The results of the regression using both wave2 and speed3 as instruments are presented in Table 5(3). Based on the results of the respective first-stage regression (Table 4(3)), we can perform an F-test to determine whether the two instruments are jointly significant. We reject the null hypothesis (p-value $= 1.68 \cdot 10^{-05}$) and conclude that there is significant evidence in the data that the two instruments are jointly significant.

m)

By performing the Sargan test we obtain that we fail to reject the null hypothesis (Sargan p-value: 0.1611). Hence, we conclude that there is evidence that our moments are consistent with the data, under the assumption that some instruments are valid, supporting the overidentifying restriction.

n)

By performing the Wu-Hausman test we obtain that we fail to reject the null hypothesis (W-H p-value: 0.097). Hence, we conclude that there is evidence in the data that we could proceed with OLS.

It is also important to mention that our decision to perform an IV regression should not be based solely on the results of a single test and requires further knowledge about the problem setting. In this case, we know theoretically (from the law of supply and demand) that we have simultaneity and reverse causality, so the use of instruments instead of the standard OLS regression is justified.