Econometrics - FIN403: Practical Assignment 2

Due on Thursday 25 November, 2021

Girolamo Vurro, Filippo Rondoni

Contents

Part 1

In the Table 1 we report the descriptive statistics of the dataset that we analyse.

Table 1: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price	97	0.847	0.344	0.290	0.597	1.117	1.775
log(Price)	97	−0.246	0.405	−1.237	−0.516	0.111	0.574
Quantity	97	4,127.216	2,620.681	170	1,995	5,975	10,940
log(Quantity)	97	8.086	0.765	5.136	7.598	8.695	9.300
Wind speed last 2 days	97	11.979	3.945	5	10	15	25
Wind speed last day 3 & 4	97	20.928	6.524	10	15	25	45
Wave height last 2 days	97	5.093	1.787	2.500	4.000	6.000	12.500
Wave height last 2 days lagged	97	5.021	1.915	3	3.5	6	12
Monday	97	0.186	0.391	0	0	0	1
Tuesday	97	0.196	0.399	0	0	0	1
Wednesday	97	0.206	0.407	0	0	0	1
Thursday	97	0.206	0.407	0	0	0	1
Time	97	49.000	28.145	1	25	73	97

Point a

We regressed the following model:

$$log(price)_t = \beta_1 + \beta_2 mon_t + \beta_3 tues_t + \beta_4 wed_t + \beta_5 thurs_t + \beta_6 t + \epsilon_t$$

We do not reject the null hypothesis that coefficients of the days are equal to 0 individually. There is not a statistical significance that price varies within the week as it is shown in Table 4. Only the coefficient of the time trend is significant with $\alpha = 0.01$ meaning that as the time increases by one unit, the price would fall by 0.3991177 % (*semi-elasticity*).

We also tested whether there is a jointly statistical significance of variation of prices across the week throughout an *F-test* with $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$. We obtained an $F - stat = 0.2296$ and $p - value = 0.9211$, and we did not reject the hypothesis that the coefficients are not jointly significant.

Point b

Adding **wave2** and **wave3** as in the following regression we obtained the coefficients shown in the second column of the Table 4.

$$log(price)_t = \beta_1 + \beta_2 mon_t + \beta_3 tues_t + \beta_4 wed_t + \beta_5 thurs_t + \beta_6 t + \beta_7 wave2_t + \beta_8 wave3_t + \epsilon_t$$

We can observe that **wave2** is statistically significant with an $\alpha = 0.01$, whereas **wave3** is statistically significant with an $\alpha = 0.05$.

In order to asses whether there the two coefficients are jointly statistically significant we computed an *F-test* considering as restricted model the first regression obtained in the previous point.

We obtained an $F - stat = 14.436$ with a *p-value* that is strictly smaller than 0.01. Therefore we rejected the null hypothesis and concluded that the two coefficients are jointly statistically significant.

The coefficients of **wave2** and **wave3** are *semi - elasticities.* They indicate by which percentage increases the price if the average level of waves increases by 1.

Intuitively, they have a positive impact on the prices because if the waves are high (meaning the sea is stormy) it is harder to fish, therefore the supply of fish will be more scarse because it would be harder to fish with higher waves and this would increase the price of the fish.

Point c

The time trend is no more statistically significant. This must be due to the fact that in the first regression the time trend embodies the effects of **`wave2`** and **`wave3`** (that are therefore omitted variables). Indeed, there is a non null correlation between **`t`** and **`wave2`** (-0.294044) as well as between **`t`** and **`wave3`** (-0.3566207). Thus, $\hat{\beta}_6$ in the first model contained an *omitted variable bias*.

Thus, when the last two variables are included in the model, time trend loses its significance. The fact that the increase of time impacted negatively on the price, and given that the waves' hight have both a positive impact on the price (and negative correlation with the time trend), the more time passes and the lower prices become (we can indirectly infer that also level of waves should get lower) as we are probably going towards the good season with less storms.

Point d

In order to asses whether there is AR(1) serial correlation for the errors we performed *Breusch-Godfrey test* on the second model.

The test produced a *p-value* strictly smaller than 0.01, therefore we rejected the null hypothesis that $\rho = 0$ concluding that we cannot exclude that there is serial correlation of order up to 1.

Moreover, as the *Durbin-Watson test* has more appeal for small samples, and provided that there is no lagged dependent variable in the model, we also performed it to test autocorrelation. We obtained that the test statistic is 0.745231, while the *p-value* is smaller than 0.01 so we can reject the null hypothesis again conclude that the residuals in this regression model are autocorrelated.

Therefore, **A4** is violated and the Gauss-Markov theorem cannot be applied: the OLS standard errors are not correct. However, the sufficient conditions for OLS to be consistent are **A7** and **A6** that still hold, so it is consistent.

Sufficient conditions for unbiasedness are **A1** and **A2** that still hold, therefore OLS is also unbiased.

Point e

We estimated the robust variance covariance matrix (using *Newey - West*) of the coefficients that takes into account autocorrelation. A good procedure is to use $H = T^{\frac{1}{4}}$ or $H = 0.75T^{\frac{1}{3}}$ lags, so we used 4 lags.

The *t-statistics* on **`wave2`** decreases in comparison with the OLS model because the estimated standard error for this coefficient is now higher (as it takes into account the autocorrelation); whereas the *t-statistic* for **`wave3`** increases slightly as the the estimated value of its coefficient stays the same in both models but in the second the standard error is now smaller than before.

We expect that the new t-statistics are more reliable than before because, provided that there is autocorrelation, they are obtained from the standard errors that are estimated considering autocorrelation.

Table 2: t-statistics and st. errors comparison

		Std. Error	t value
Newey-West	wave2	0.023	3.889
	wave3	0.019	2.430
OLS	wave2	0.022	4.178
	wave3	0.021	2.276

Point f

We obtained the *Prais-Winsten* estimates that are summarized in the Table 3.

Furthermore, in order to assess if **wave2** and **wave3** are jointly statistically significant we performed a *Wald test*: in order to calculate the *Chi-stat* we constructed the matrix $\mathbf{R}$ and the vector $\mathbf{q}$:

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \ \mathbf{q} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

We obtained $\chi^2_2 = 10.0$; as the *p-value* is smaller than 0.05 (0.0067) the null hypothesis $H_0 : \mathbf{R}\beta = \mathbf{q}$ can be rejected and the coefficients of **wave2** and **wave3** are jointly significant.

Whereas, the estimated magnitude of the autocorrelation in the error term is $\hat{\rho} = 0.6874$.

Table 3: Prais-Wisten estimates

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.658	0.239	-2.755	0.007
Monday	0.010	0.065	0.153	0.879
Tuesday	0.003	0.074	0.034	0.973
Wednesday	0.062	0.075	0.837	0.405
Thursday	0.117	0.062	1.891	0.062
Time	-0.001	0.003	-0.249	0.804
Wave height last 2 days	0.050	0.017	2.874	0.005
Wave height last 2 days lagged	0.032	0.017	1.859	0.066

Table 4: Regressions of point a) and b) using OLS

	Dependent variable:	
	lavgprc	
	(1)	(2)
Monday	−0.010	−0.018
	(0.129)	(0.114)
Tuesday	−0.009	−0.009
	(0.127)	(0.112)
Wednesday	0.038	0.050
	(0.126)	(0.112)
Thursday	0.091	0.123
	(0.126)	(0.111)
Time	−0.004***	−0.001
	(0.001)	(0.001)
Wave height last 2 days		0.091***
		(0.022)
Wave height last 2 days 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

Part 2

Point g

We reported the estimated coefficients of the demand equation in table 7 first column.

The coefficient on **log(price)** is an *elasticity* (of the demand): in particular if the price increases by 1% the quantity sold is expected to decrease by 0.548897%. Therefore, **log(quantity)** and **log(price)** are negatively correlated as expected.

However, assuming that there is no measurement error for the independent variable, we could infer that not just **log(price)** affect **log(quantity)** (negatively), but **log(quantity)** may also affect **log(price)** positively as we could assume that if the quantity sold (demanded) increases, the prices should also increase; this would lead to a biased and inconsistent estimate of the causal effect of price on the demand for fish as **A2** would be violated because the correlation between the dependent variable and the errors is not equal to 0 (**simultaneity and reverse causality**).

Point h

In order to be a valid instrument, **`wave2`** needs to fulfill two conditions:

1. the **relevance condition**: it should be correlated with **log(price)**;
2. the **exclusion restriction**: it only affects **log(quantity)** through its effect on **log(price)**.

The relevance condition should be fulfilled as we have seen in the previous part that **`wave2`** had a statistically significant impact on the prices; in order that the exclusion restriction is satisfied we should make an economic argument, namely we should assume that the the demand of fish (therefore the total quantity sold) would be unaffected (directly) by the hight of the waves so by the conditions of the sea. Therefore it should have no impact on our dependent variable. To sum up, the hight of waves impacts the quantity **supplied**, thus the price, but should not impact the **demand** for fish (provided that there are enough fish stocks).

Point i

We reported the results of the first stage regression using OLS in the table 7 in the second column. With a single instrument, this is equivalent to a *t-statistic* of 3.16.

We obtain a *t-statistic* of 4.758 relatively to the coefficient $\hat{\gamma_2}$. Knowing that with a single instrument the rule of thumb to consider it as "strong instrument" is to have a *t-statistic* bigger than 3.16, we conclude that **`wave2`** is a strong instrument for **log(price)**.

Point j

We reported the new IV estimates in the third column of the table 7.

In particular we obtained $\hat{\beta}_{2,IV} = -0.960347$ which is bigger in absolute value than the OLS estimate $\hat{\beta_2} = -0.548897$.

According to the new IV estimate, an increase of 1% of the price reduces the total quantity by 0.960347% (*elasticity*).

The OLS estimate was not able to isolate the effect ("good variation") of the price on the quantity sold but indeed it was underestimating (in absolute values) this effect. The OLS estimates included also the effect that the quantity sold had on the prices which is intuitively positive as when quantity sold increases, also prices should increase and this produced an attenuated effect (smaller coefficient in absolute value) if compared to the one obtained with the IV model, where the sea conditions help us to isolate the good variation in price that we can use to find the true effect of price on quantity.

The magnitude of the IV standard errors is in general larger than the OLS ones as can be also seen in the table 7 comparing the standard errors (in brackets) of column three with column one.

In particular the magnitude of IV standard error (0.421982) relative to the coefficient of **log(price)** estimated with IV is higher than the magnitude of the standard errors obtained with the OLS (0.184138), as a consequence now the CI are larger than before, therefore we might have a more "imprecise estimate", so maybe we should not overinterpret. But this is also due to the fact that the the correlation between

Part 2 [Point j] continued on next page...

instrument and the endogenous variable is necessairly smaller than 1 by contruction (we saw in the slides that in the simplest case the new variance is multiplied by $\frac{1}{\rho_{xz}^2} \geq 1$):

$$V[\beta_{IV}] = \hat{\sigma^2}(X'Z(Z'Z)^{-1}Z'X)^{-1}$$

This might be caused by a measurament error in the price of the fish, or to the fact that exclusion restriction is not satisfied: this would mean that there is a direct effect of the hight of the waves on the quantity sold.

Point k

We report the IV results in the fourth column of table 7 using as instrument **speed3** instead of **wave2**.

From the first stage regression:

$$log(price)_t = \gamma_1 + \gamma_2 speed3_t + \gamma_3 mon_t + \gamma_4 tues_t + \gamma_5 wed_t +_6 thurs_t + \gamma_7 t + +\epsilon_t$$

We obtain a *t-statistic* of 2.565 relative to $\hat{\gamma_2}$ which is lower than 3.16. So, by the rule of thumb we conclude that **speed3** is not a strong instrument.

Point l

We reported the IV results in the fifth column of table 7.

Furthermore, we applied an *F-test* to the first stage regression:

$$log(price)_t = \gamma_1 + \gamma_2 wave2_t + \gamma_3 speed3_t + \gamma_4 mon_t + \gamma_5 tues_t + \gamma_6 wed_t + \gamma_7 thurs_t + \gamma_8 t + +\epsilon_t$$

With null hypotesis that both the coefficients of the instrumental variables are 0. We obtained an $F - statistic = 12.469$ wich is strictly bigger than 10. By the rule of thumb we conclude that the instruments jointly fulfill the criterion for being sufficiently "strong instruments".

Point m

Performing the Sargan test we obtained a $\chi_1^2 = 1.964$ with the corresponding $p - value = 0.161$ therefore we could not reject the null hypothesis under which all the restrictions are valid (i.e. all instruments are indeed exogenous).

Point n

Performing the Wu-Hausman test we obtained $t - value = 2.821$ and a $p - value = 0.097$ therefore with $\alpha = 5\%$ we cannot reject the null hypothesis, i.e. we cannot reject the hypothesis that **log(price)** is exogenous.

From this result we could conclude that we could also utilize the OLS model

We summarized the diagnostics that we presented in the last three points in table 5.

Table 5: Diagnostic tests for the model with two instruments

	df1	df2	statistic	p-value
Weak instruments	2	89	12.469	0.00002
Wu-Hausman	1	89	2.821	0.097
Sargan	1		1.964	0.161

Bonus question

In order to adjust the IV estimates of the model in item (l) we estimated the variance and covariances matrix of the coefficients using Newey-West keeping 4 lags (using the rule of thumb mentioned in point e)). Therefore we calculated again the coefficients that are unchanged in comparison as the ones obtained in point l) as they are not affected by the serial autocorrelation (in the formula to calculate them we do not need the $V[\epsilon]$). However, now we obtained standard errors estimates that take into account the serial autocorrelation. Notice that the coefficient of **log(price)** slightly looses significance as the CI is now larger.

In table 6 we report the new adjusted estimates:

Table 6: Adjusted IV estimates for serial autocorrelation

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.259	0.245	33.648	0
log(Price)	-1.127	0.492	-2.291	0.024
Monday	-0.323	0.187	-1.725	0.088
Tuesday	-0.689	0.202	-3.406	0.001
Wednesday	-0.514	0.211	-2.433	0.017
Thursday	0.121	0.166	0.726	0.469
Time	-0.004	0.004	-0.975	0.332

Table 7:

	Dependent variable:				
	log(Quantity)	log(Price)	log(Quantity)		
	OLS	*OLS*	*instrumental variable*		
	(1)	(2)	(3)	(4)	(5)
log(Price)	−0.549***		−0.960**	−1.960**	−1.127***
	(0.184)		(0.422)	(0.907)	(0.415)
Wave height last 2 days		0.103***			
		(0.022)			
Monday	−0.318	−0.036	−0.322	−0.332	−0.323
	(0.227)	(0.116)	(0.233)	(0.292)	(0.239)
Tuesday	−0.684***	0.007	−0.687***	−0.696**	−0.689***
	(0.224)	(0.114)	(0.230)	(0.288)	(0.236)
Wednesday	−0.535**	0.083	−0.520**	−0.482*	−0.514**
	(0.221)	(0.113)	(0.227)	(0.286)	(0.233)
Thursday	0.068	0.136	0.106	0.196	0.121
	(0.221)	(0.113)	(0.230)	(0.295)	(0.236)
Time	−0.001	−0.002	−0.003	−0.007	−0.004
	(0.003)	(0.001)	(0.003)	(0.005)	(0.003)
Constant	8.301***	−0.706***	8.271***	8.198***	8.259***
	(0.203)	(0.169)	(0.210)	(0.268)	(0.215)
Observations	97	97	97	97	97
R^2	0.219	0.269	0.175	−0.291	0.133
Adjusted R^2	0.167	0.221	0.120	−0.377	0.075
Residual Std. Error (df = 90)	0.698	0.357	0.717	0.897	0.735
F Statistic (df = 6; 90)	4.201***	5.528***			
Note:			*p<0.1; **p<0.05; ***p<0.01		