FINAL EXAM - FINANCIAL ECONOMETRICS - 18 March 2022

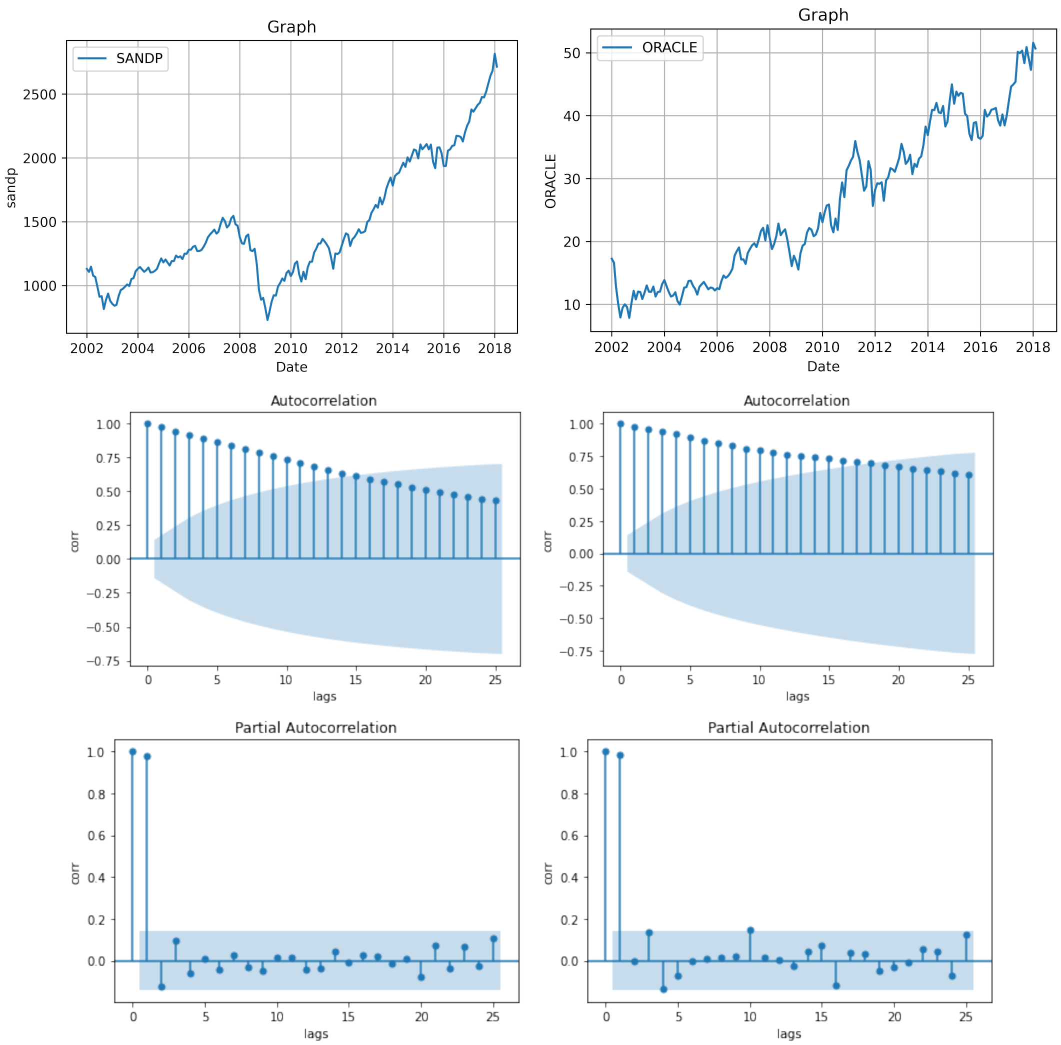
Paulo M. M. Rodrigues and Sónia Félix

Time for completion: 1h 45 min

Unless otherwise stated use 5% for significance level.

Question 1: Consider the price indexes for S&P500 and ORACLE presented in Figure 1, and the corresponding autocorrelations and partial autocorrelations functions.

Figure 1: S&P500 and ORACLE prices indexes and corresponding autocorrelations and partial autocorrelations functions



a) [1.0] Based on the plots in Figure 1 what can you say about the stationarity of the S&P and ORACLE time series and their persistence? Justify.

Both series display trending behaviour which suggests time varying means and therefore both series, according to the notion of weak stationarity, are nonstationary. In addition given

the slow decay of the autocorrelation function of both series, this clearly indicates strong persistent behaviour of both series (note that the 1st autocorrelation is very close to 1 in both cases).

```
TABLE 1:  Augmented Dickey-Fuller Results for S&P
(Regression with constant only)
Critical Values: -3.46 (1%), -2.88 (5%), -2.57 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.009
Model:                            OLS   Adj. R-squared:                  0.004
Method:                 Least Squares   F-statistic:                     1.817
Date:                Wed, 16 Mar 2022   Prob (F-statistic):              0.179
Time:                        22:35:55   Log-Likelihood:                -1037.5
No. Observations:                 193   AIC:                             2079.
Df Residuals:                     191   BIC:                             2086.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Level.L1       0.0110      0.008      1.348      0.179      -0.005       0.027
const         -7.8664     12.515     -0.629      0.530     -32.551      16.818
==============================================================================
Omnibus:                       15.778   Durbin-Watson:                   1.878
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               18.844
Skew:                          -0.591   Prob(JB):                     8.09e-05
Kurtosis:                       3.972   Cond. No.                     5.07e+03
==============================================================================
```

TABLE 2: Augmented Dickey-Fuller Results for S&P
(Regression with constant and time trend)
Critical Values: -4.01 (1%), -3.43 (5%), -3.14 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.

OLS Regression Results

Dep. Variable:	y	R-squared:	0.065
Model:	OLS	Adj. R-squared:	0.012
Method:	Least Squares	F-statistic:	1.216
Date:	Wed, 16 Mar 2022	Prob (F-statistic):	0.284
Time:	22:35:55	Log-Likelihood:	-986.59
No. Observations:	185	AIC:	1995.
Df Residuals:	174	BIC:	2031.
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Level.L1	-0.0165	0.017	-0.992	0.323	-0.049	0.016
Diff.L1	0.0605	0.077	0.785	0.434	-0.092	0.213
Diff.L2	-0.0477	0.078	-0.611	0.542	-0.202	0.106
Diff.L3	0.0592	0.077	0.769	0.443	-0.093	0.211
Diff.L4	0.0697	0.076	0.912	0.363	-0.081	0.221
Diff.L5	0.0900	0.077	1.175	0.242	-0.061	0.241
Diff.L6	-0.0678	0.077	-0.886	0.377	-0.219	0.083
Diff.L7	-0.0372	0.077	-0.485	0.628	-0.189	0.114
Diff.L8	0.1438	0.077	1.877	0.062	-0.007	0.295
const	15.4527	15.664	0.986	0.325	-15.464	46.369
trend	0.1837	0.134	1.375	0.171	-0.080	0.447

Omnibus:	12.602	Durbin-Watson:	1.897
Prob(Omnibus):	0.002	Jarque-Bera (JB):	13.973
Skew:	-0.540	Prob(JB):	0.000924
Kurtosis:	3.803	Cond. No.	6.42e+03

TABLE 3: Augmented Dickey-Fuller Results for ORACLE
(Regression with constant only)
Critical Values: -3.47 (1%), -2.88 (5%), -2.57 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.

OLS Regression Results

Dep. Variable:	y	R-squared:	0.027
Model:	OLS	Adj. R-squared:	0.012
Method:	Least Squares	F-statistic:	1.757
Date:	Wed, 16 Mar 2022	Prob (F-statistic):	0.157
Time:	22:34:52	Log-Likelihood:	-369.36
No. Observations:	191	AIC:	746.7
Df Residuals:	187	BIC:	759.7
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Level.L1	0.0021	0.010	0.202	0.840	-0.018	0.023
Diff.L1	-0.0942	0.072	-1.305	0.194	-0.237	0.048
Diff.L2	-0.1448	0.073	-1.984	0.049	-0.289	-0.001
const	0.1836	0.295	0.623	0.534	-0.398	0.765

Omnibus:	1.709	Durbin-Watson:	2.011
Prob(Omnibus):	0.425	Jarque-Bera (JB):	1.434
Skew:	-0.035	Prob(JB):	0.488
Kurtosis:	3.419	Cond. No.	68.9

```
TABLE  4:  Augmented Dickey-Fuller Results for ORACLE
(Regression with constant and time trend)
Critical Values: -4.01 (1%), -3.43 (5%), -3.14 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.100
Model:                            OLS   Adj. R-squared:                  0.076
Method:                 Least Squares   F-statistic:                     4.108
Date:                Wed, 16 Mar 2022   Prob (F-statistic):            0.00147
Time:                        22:34:52   Log-Likelihood:                -358.98
No. Observations:                 190   AIC:                             730.0
Df Residuals:                     184   BIC:                             749.4
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Level.L1      -0.1494      0.043     -3.495      0.001      -0.234      -0.065
Diff.L1       -0.0305      0.075     -0.409      0.683      -0.178       0.117
Diff.L2       -0.0936      0.072     -1.301      0.195      -0.236       0.048
Diff.L3        0.0951      0.072      1.320      0.189      -0.047       0.237
const          0.9650      0.347      2.779      0.006       0.280       1.650
trend          0.0329      0.009      3.596      0.000       0.015       0.051
==============================================================================
Omnibus:                        2.057   Durbin-Watson:                   2.015
Prob(Omnibus):                  0.358   Jarque-Bera (JB):                1.883
Skew:                          -0.023   Prob(JB):                        0.390
Kurtosis:                       3.486   Cond. No.                         337.
==============================================================================
```

b) [2.0] Given the outputs provided in Tables 1 - 4, how would you classify the orders of integration of the two series? Justify.

The S&P is a unit root nonstationary series (this result is consistent regardless of the test used (in both Table 1 and Table 2 the null of a unit root is not rejected)). Regarding ORACLE the situation is different, the results in Table 3 (test regression with constant only) do not reject the null hypothesis, but the results in Table 4 (test regression with constant and trendo) reject the null of a unit root, hence we opt for the latter given that a trend seems to be a relevant variable to be considered in the test regression. Hence, S&P is I(1) and ORACLE is I(0).

c) [1.0] Given the Augmented Dickey Fuller regression output in Table 4, what is the corresponding autoregressive order of the ORACLE variable in levels?

Since the results in Table 4 indicate an AR(3) in first differences in levels ORACLE should follow AR(4)dynamics.

d) [1.5] Based on your conclusions in b) discuss the validity of the regression results provided in Table 5 (see below). Note that this output corresponds to a regression of the type:

$$ORACLE_t = \alpha + \beta S\&P_t + \varepsilon_t$$

.

Note that since S&P is I(1) and ORACLE is I(0) this corresponds to an unbalanced regression and consequently the results should be interpreted with care as the consequences are similar to those of a spurious regression. This is visible from Table 5 as the R^2 is large, the regressors are highly significant and and the DW is close to zero. Hence, this regression is meaningless.

TABLE 5: OLS Regression Results

Dep. Variable:	ORACLE	R-squared:	0.779
Model:	OLS	Adj. R-squared:	0.778
Method:	Least Squares	F-statistic:	676.2
Date:	Wed, 16 Mar 2022	Prob (F-statistic):	8.02e-65
Time:	21:29:45	Log-Likelihood:	-611.11
No. Observations:	194	AIC:	1226.
Df Residuals:	192	BIC:	1233.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-7.0546	1.334	-5.289	0.000	-9.686	-4.424
SANDP	0.0225	0.001	26.003	0.000	0.021	0.024

Omnibus:	61.272	Durbin-Watson:	0.058
Prob(Omnibus):	0.000	Jarque-Bera (JB):	11.203
Skew:	0.169	Prob(JB):	0.00369
Kurtosis:	1.872	Cond. No.	5.05e+03

The output in Table 6 corresponds to a regression of the type:

$$\Delta \log ORACLE_t = \delta_0 + \delta_1 \Delta \log S\&P_t + u_t$$

which corresponds to a regression of log returns of ORACLE on the log returns of the S&P500.

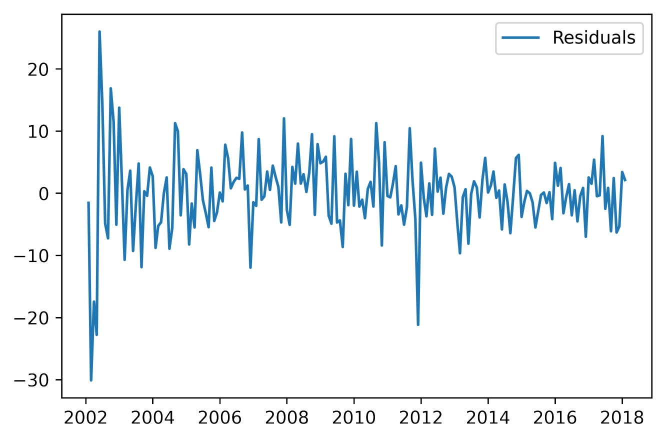
TABLE 6: OLS Regression Results

Dep. Variable:	ret_oracle	R-squared:	0.331
Model:	OLS	Adj. R-squared:	0.328
Method:	Least Squares	F-statistic:	94.51
Date:	Wed, 16 Mar 2022	Prob (F-statistic):	2.10e-18
Time:	21:52:14	Log-Likelihood:	-631.36
No. Observations:	193	AIC:	1267.
Df Residuals:	191	BIC:	1273.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0634	0.464	0.137	0.892	-0.852	0.979
ret_sandp	1.0892	0.112	9.722	0.000	0.868	1.310

In Figure 2 (see below) we plot the residuals of this regression.

Figure 2: Plot of Residuals



e) [1.5] We computed the following statistics from the residuals:

```
[('Jarque-Bera', 149.449),
 ('Chi^2 two-tail prob.', 0.000),
 ('Skew', -0.446),
 ('Kurtosis', 7.218)]
```

What information can we extract from these statistics, and what conclusion can be drawn with respect to the properties of the residuals.

Thse results are useful for understanding whether residuals are normally distributed. In the case of a normal distribution skewness=0 and kurtosis =3, hence from the results provided we see that the residuals display some negative asymmetry and leptokurtosis as Kurtosis >3. The non-normality is then confirmed by the Jarque-Bera test as the null hypothesis is rejected.

f) [1.5] We also performed an Engle test on the squared residuals using 5 lags:

```
[('lm', 25.444),
 ('lmpval', 0.000)]
```

Indicate how this test is implemented and what conclusion we can draw from the results.

The Engle test is a test for ARCH effects. In this case its implementation considers the residuals of the model estimated in Table 6, $\hat{\varepsilon}_t$ and compute the auxiliary regression,

$$\hat{\varepsilon}_t^2 = \phi_0 + \phi_1\hat{\varepsilon}_{t-1}^2 + \phi_2\hat{\varepsilon}_{t-2}^2 + ... + \phi_5\hat{\varepsilon}_{t-5}^2 + u_t$$

$H_0 : \phi_1 = \phi_2 = ... = \phi_5 = 0$ and $H_A : \phi_1 \neq 0$ or $\phi_2 \neq 0$ or $\phi_5 \neq 0$. **Engle Test:** $LM = T * R^2$

From the results we observe that the null is rejected and therefore there is evidence of ARCH effects in the residuals.

g) [1.0] Given the regression output in Table 7 (see below), indicate what model this corresponds to and whether you would use it?

This is an ARCH(5) and it would not be usable as one of the coefficients is negative, violating one of the assumptions for validity of the model.

```
                      TABLE 7: Constant Mean - ARCH Model Results
==============================================================================
Dep. Variable:                Residuals   R-squared:                       0.000
Mean Model:               Constant Mean   Adj. R-squared:                  0.000
Vol Model:                         ARCH   Log-Likelihood:               -593.625
Distribution:                    Normal   AIC:                           1201.25
Method:              Maximum Likelihood   BIC:                           1224.09
                                          No. Observations:                  193
Date:                  Wed, Mar 16 2022   Df Residuals:                      192
Time:                          22:11:20   Df Model:                            1
                                 Mean Model
=========================================================================
                 coef    std err          t      P>|t|  95.0% Conf. Int.
-------------------------------------------------------------------------
mu            -0.0180      0.324 -5.555e-02      0.956 [ -0.653,  0.617]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|       95.0% Conf. Int.
----------------------------------------------------------------------------
omega         11.7331      3.294      3.562  3.676e-04       [  5.278, 18.189]
alpha[1]       0.0000  2.106e-02      0.000      1.000 [-4.127e-02,4.127e-02]
alpha[2]       0.0806  4.763e-02      1.692  9.069e-02   [-1.277e-02,  0.174]
```

```
alpha[3]       -0.4328      0.135      3.203  1.359e-03      [  0.168,  0.698]
alpha[4]        0.0000  8.755e-02      0.000      1.000      [ -0.172,  0.172]
alpha[5]        0.1234  6.233e-02      1.979  4.780e-02    [1.198e-03,  0.246]
==============================================================================
Covariance estimator: robust
```

h) [1.5] Given the outputs in Tables 8 and 9 (see below), indicate the main difference between the two models and which of the two you would recommend for modeling the volatility in the residuals.

Table 8 corresponds to a GARCH(1,1) and Table 9 to a GJR(2,2). Given the significance of gamma in Table 9 this suggests asymmetric behaviour of the shocks and therefore ne shoudl opt for the GJR.

i) [1.5] Given the output in Table 9 (see below) how persistent is the volatility based on this model.

The persistence is different depending on whether shocks are positive and negative. For negative shocks persistence is (0.1515+0+0.1429+0+0.8285) whereas when shocks are positive (0.1515+0+0+0.8285)

TABLE 8: Zero Mean - GARCH Model Results

Dep. Variable:	Residuals	R-squared:	0.000
Mean Model:	Zero Mean	Adj. R-squared:	0.005
Vol Model:	GARCH	Log-Likelihood:	-598.066
Distribution:	Normal	AIC:	1206.13
Method:	Maximum Likelihood	BIC:	1222.45
		No. Observations:	193
Date:	Wed, Mar 16 2022	Df Residuals:	193
Time:	22:44:16	Df Model:	0

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	3.2493	3.142	1.034	0.301	[-2.909, 9.407]
alpha[1]	0.0981	0.152	0.647	0.517	[-0.199, 0.395]
alpha[2]	2.2297e-10	0.147	1.518e-09	1.000	[-0.288, 0.288]
beta[1]	1.4920e-08	0.279	5.354e-08	1.000	[-0.546, 0.546]
beta[2]	0.7556	9.956e-02	7.590	3.202e-14	[0.560, 0.951]

Covariance estimator: robust

TABLE 9: Zero Mean - GJR-GARCH Model Results

Dep. Variable:	Residuals	R-squared:	0.000
Mean Model:	Zero Mean	Adj. R-squared:	0.005
Vol Model:	GJR-GARCH	Log-Likelihood:	-596.147
Distribution:	Normal	AIC:	1204.29
Method:	Maximum Likelihood	BIC:	1223.87
		No. Observations:	193
Date:	Wed, Mar 16 2022	Df Residuals:	193
Time:	22:44:53	Df Model:	0

Volatility Model

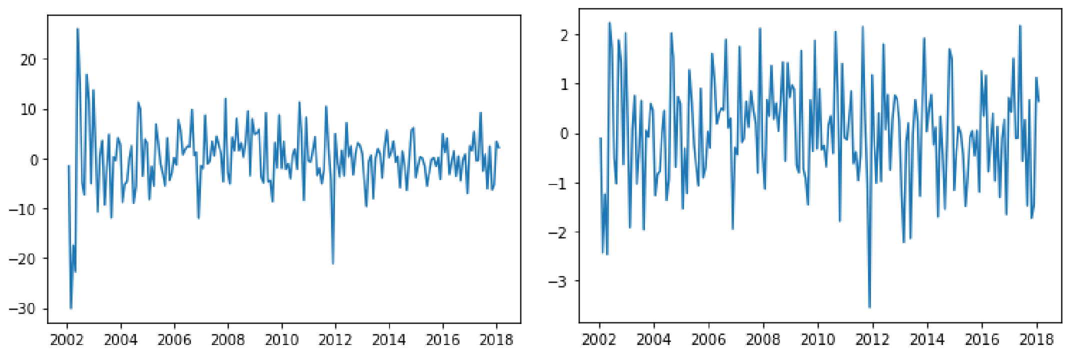
	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	1.6541	1.203	1.374	0.169	[-0.705, 4.013]
alpha[1]	0.1515	9.780e-02	1.549	0.121	[-4.020e-02, 0.343]
alpha[2]	0.0000	0.169	0.000	1.000	[-0.332, 0.332]
gamma[1]	-0.1429	6.854e-02	-2.085	3.708e-02	[-0.277,-8.563e-03]
beta[1]	0.0000	0.105	0.000	1.000	[-0.207, 0.207]
beta[2]	0.8285	0.258	3.211	1.321e-03	[0.323, 1.334]

Covariance estimator: robust

j) [1.5] Figure 3 presents plots of the residuals and standardized residuals. Indicate why the standardized residuals may be important.

Standardized residuals may be important to analyse the quality of fit of the volatility modeled considered, as if it is well specified the standardized residuals should behave similar to white noise.

Figure 3: Residuals (left plot) and Standardized Residuals (right plot)



k) [1.5] Given the following results of the Engle test computed from the standardized residuals, what do you conclude?

Engle test on the squared residuals:

[('lm', 11.183), ('lmpval', 0.0479)]

Given the results provided we conclude that the volatility model is not able yet to capture all the volatility in the residuals and should therefore be improved. Note that the Engle test rejects the null of no ARCH effects in the standardized residuals.

Question 2: Consider the following regression model to study the impact of airline market power on airline pricing:

$$\ln P_{it} = \beta_0 + \alpha_i + \delta_t + \beta_1 marketshare_{it} + \beta_2 \ln miles_i + \beta_3 (\ln miles_i)^2 + \beta_4 \ln passeng_{it} + \varepsilon_{it}$$

where $\ln P_{it}$ is the logarithm of the average price charged on route i in quarter t, *marketshare* is the market share of the largest carrier for each of the routes, $\ln miles$ is the logarithm of the route distance (in miles), and $\ln passeng$ is the logarithm of the average number of passengers. The terms α_i and δ_t are route and time specific effects, respectively.

1. Explain the importance of including the term α_i in this model. Which characteristics of a route can be captured by this term? Are these likely to be correlated with $marketshare_{it}$?

   Including the term α_i is a way of accounting for time-invariant characteristics of a route i that affect prices and might also be correlated with the explanatory variables. Omitting this term may lead to an omitted variable bias problem.
   Characteristics of the cities near the airports on a route could affect the demand for air travel, such as the population, education levels, or types of employers. Of course, each of these can be time-varying, although for a short stretch of time they are roughly constant.
   Perhaps, the quality of the freeway system and access to trains, along with geographical features are roughly time-invariant and could certainly be correlated with market concentration.

   Consider the estimates provided in the table below.

2. In the OLS and random effects model, the logarithm of the route distance (in miles) and its square are included. Carefully explain why these variables are not included in the fixed effects model.

	Dependent variable: logarithm of average price		
	(1)	(2)	(3)
	Pooled OLS	Random effects	Fixed effects
marketshare	0.3028***	0.0999***	0.0647**
	(0.0298)	(0.0251)	(0.0265)
ln miles	-0.9683***	-1.0638***	
	(0.1266)	(0.2508)	
$(\ln miles)^2$	0.1065***	0.1095***	
	(0.0096)	(0.0190)	
ln passeng	-0.0773***	-0.2236***	-0.3163***
	(0.0056)	(0.0070)	(0.0086)
intercept	7.0392***	8.5440***	6.9592***
	(0.4185)	(0.8269)	(0.0546)
N	4596	4596	4596
adj. R^2	0.422		0.048

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3. Should the authors be more confident in the random or fixed effects estimates? Why? What does it imply regarding the correlation between a_i and the explanatory variables included in the model?
4. Indicate how to perform an autocorrelation test in the context of model (3) in the Table above (within estimation).