Question 1

Question 1(a)

Homoskedasticity refers to when the error term variance in a data set is constant across all the independent variables. Homoskedasticity proves the efficiency of the estimators of the data set, and can be a helpful parameter to define whether the values of standard error, t-value and p-value of the data set are correct.

For this regression model, we expect there to be heteroskedasticity as we are analysing the fraction of average household expenditure on food, which may vary largely from small households and big households with many members. If the variation is large enough, this may skew our regression residuals leading to heteroskedasticity. Moreover, bigger households are likely to have high total consumption expenditure, making it again likely for heteroskedasticity to be present.

Question 1(b)

Heteroskedasticity refers to when the error term variance in a data set is not constant across all of the independent variables. Heteroskedastic error terms would mean that reliable hypothesis tests are unable to be conducted, and values of standard error, t-value and p-value would be incorrect.

Plotting the residuals (residuals against predicted value of 'fraction') for informal evidence of heteroskedasticity, we can see from Figure 1 that our data is roughly equally separated between negative and positive residuals, with some high outliers on either side. Moreover, there is no evident pattern in the residual plot. This suggests that there is no heteroskedasticity in our regression model.

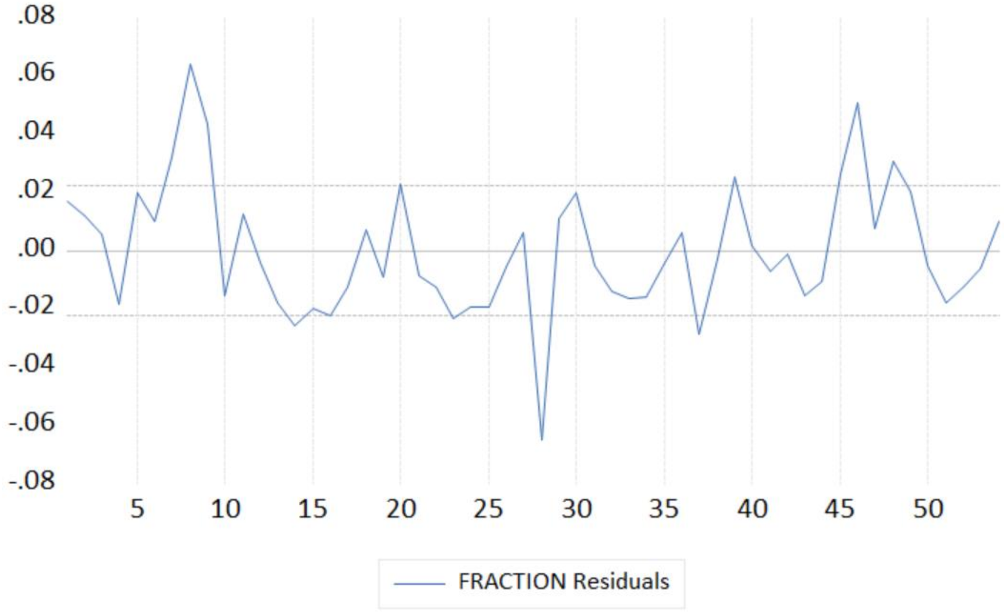

FIGURE 1 RESIDUAL PLOT

Question 1(c)

Model:

$$fraction = \beta_0 + \ \beta_1 total + \ \beta_2 size + u$$

Hypothesis test:

$$\mathrm{H}_0 : Var(u \mid total, size) = E(u^2 \mid total, size) = \ \sigma^2 I_{54}$$
$$\mathrm{H}_1 : Var(\ u|Total, Size) \text{ is a smooth function of } total \text{ and } size$$

Auxiliary Regression:

$$\hat{u}^2 = \propto_0 + \propto_1 \ total + \propto_2 \ size + \propto_3 \ total^2 + \propto_4 \ size^2 + \propto_5 \ total \times size + v$$

White test statistic:

$$W = 54 \times R^2_{\hat{u}^2} \sim \ \chi^2(5) \text{ under } H_0$$

Calculate White test statistic:

$$W_{cal} = 54 \ \times 0.142 = 7.67$$

Using this estimated auxiliary regression for $R^2_{\hat{u}^2}$:

$$\underset{(0.001)}{0.001} + \underset{(0.002)}{0.001} total - \underset{(0.000)}{0.000} size + \underset{(0.001)}{0.000} total^2 + \underset{(0.000)}{0.000} size^2 - \underset{(0.000)}{0.000} total \times size$$

See Appendix for Question 1.

Calculate White critical value @ 5 level of significance:

$$\chi^2(5) = 11.07$$

Decision rule:

$$\text{Reject } \mathrm{H}_0 \text{ if } W_{\mathrm{cal}} > \chi^2(5)$$

As $W_{\mathrm{cal}} = 7.62 < \chi^2(5) = 11.07$, there insufficient evident to reject the null hypothesis that are homoskedastic errors ($E(u^2 \mid total, size) = \ \sigma^2 I_{54}$) for our linear regression model at the 5% level of significance.

Question 1(d)

The White test with fitted values has an auxiliary regression:

$$\hat{u}^2 = \alpha_0 + \alpha_1 \widehat{fraction} + \alpha_2 \widehat{fraction}^2 + v$$

Given that the White calc is: $W_{cal} = 54 \ \times 0.058 = 3.133$,

We do not reject the null hypothesis that errors are homoskedastic as $W_{cal} = 3.133 < \chi^2(5) = 11.07$ at 5% level of significance. From this, there is no reason to believe that the White test with fitted values will be a better idea. However, given that the Akaike Information Criterion (AIC) is -11.241 compared to the ordinary White test where AIC is -

11.224 for the auxiliary regression, there may be reason to prefer this White test over the previous one. However, given that the adjusted R^2 declines from 0.053 to 0.021, we believe that there its hardly justificatory to choose this model over the other, nonetheless.

See Appendix for Question 1.

Question 1(e)

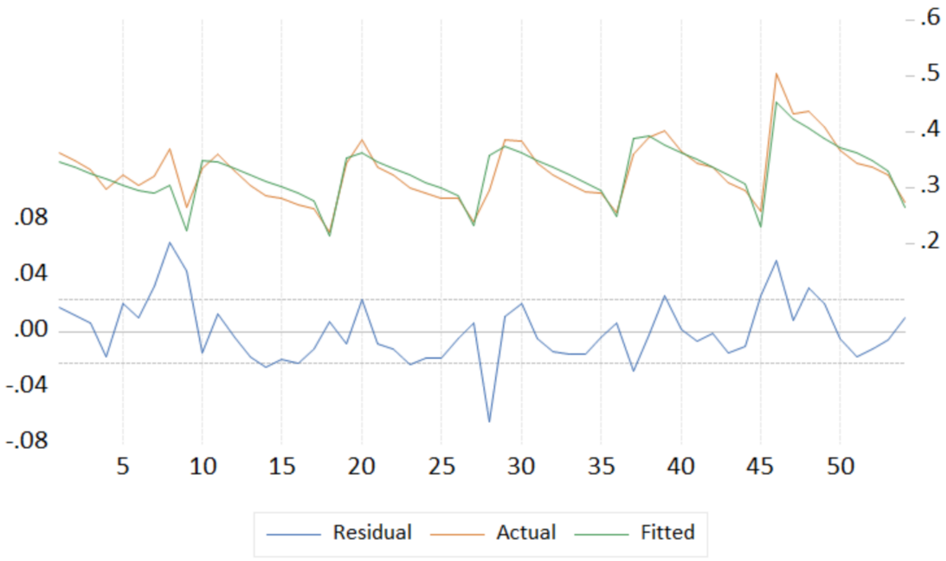

FIGURE 2 LINE AND RESIDUAL PLOT

The line graph of the residuals (the top graph in Figure 2) shows that there is a cyclical pattern, where positive residuals are followed by positive residuals and negative residuals are followed by negative residuals, as indicated by the peaks and troughs. The residuals in the series carry over into the future periods, which can be observed by the cyclical pattern. It is unlikely that this pattern is white noise, thus we can say that the graph indicates serially correlation.

Question 1(f)

Linear regression model:

$$fraction_i = \beta_0 + \beta_1 total_i + \beta_2 size_i + u_i$$

$$u_i = \rho_1 u_{i-1} + e_i$$

$$e_i \sim N(0, \sigma^2)$$

Hypothesis test:

$$\mathrm{H}_0 : \rho_1 = 0$$
$$\mathrm{H}_1 : \rho_1 \neq 0$$

Auxiliary Regression:

$$\hat{u} = \propto_0 + \propto_1 total_i + \propto_2 size_i + \alpha_3 u_{i-1} + v_i$$

Bruesch-Godfrey test statistic:

$$BG = (54 - 1) \times R^2_{\hat{u}} \sim \chi^2(1) \text{ under } H_0$$

Calculate White test statistic:

$$W_{cal} = 53 \times 0.132 = 6.98$$

Using this estimated auxiliary regression for $R^2_{\hat{u}}$:

$$\underset{(0.008)}{-0.004} + \underset{(0.001)}{0.000} total_i + \underset{(0.001)}{0.000} size_i + \underset{(0.136)}{0.369} u_{i-1}$$

See Appendix for Question 1.

Calculate White critical value @ 5 level of significance:

$$\chi^2(1) = 3.84$$

Decision rule:

$$\text{Reject } \mathrm{H}_0 \text{ if } W_{\mathrm{cal}} > \chi^2(1)$$

As $W_{\mathrm{cal}} = 6.98 > \chi^2(1) = 3.84$, there is sufficient evidence to reject the null hypothesis that there is no first order autocorrelation in our regression model for the alternative that there is at 5% level of significance.

Question 1(g)

There are two ways to correct for autocorrelation in our model (as per our results in 1(c) and 1(f)).

From 1(c), we know that there is no evidence to suggest heteroskedasticity in our linear model, and from 1(f), that there is evidence to suggest there is autocorrelation however at the 5% level of significance. To correct for autocorrelation, therefore, we may have two methods. Our first method is to use HAC standard errors instead (which uses autocorrelation-robust standard errors). Our regression from this:

$$\widehat{fraction} = \underset{(0.010)}{0.335} - \underset{(0.009)}{0.154} total + \underset{(0.003)}{0.015} size$$

A simple OLS regression of fraction on total and size is:

$$\widehat{fraction} = \underset{(0.008)}{0.345} - \underset{(0.011)}{0.154} total + \underset{(0.002)}{0.015} size$$

Plotting the correlogram

Sample: 1 54
Included observations: 54

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.347	0.347	6.8793	0.009
		2	0.110	-0.012	7.5853	0.023
		3	0.134	0.114	8.6578	0.034
		4	-0.011	-0.106	8.6647	0.070
		5	-0.137	-0.122	9.8180	0.081
		6	0.019	0.114	9.8405	0.132
		7	-0.003	-0.033	9.8413	0.198
		8	-0.123	-0.101	10.828	0.212
		9	0.197	0.301	13.436	0.144
		10	0.105	-0.101	14.198	0.164
		11	-0.031	-0.022	14.266	0.219
		12	0.031	0.006	14.335	0.280
		13	-0.041	-0.125	14.459	0.342
		14	-0.137	0.020	15.868	0.321
		15	-0.046	-0.007	16.034	0.380
		16	-0.085	-0.141	16.614	0.411
		17	-0.188	-0.002	19.491	0.301
		18	-0.006	0.017	19.495	0.362
		19	0.019	-0.031	19.526	0.424
		20	-0.205	-0.193	23.271	0.276
		21	-0.098	-0.011	24.153	0.286
		22	-0.044	-0.014	24.338	0.330
		23	-0.152	-0.056	26.579	0.274
		24	-0.060	0.009	26.938	0.307

FIGURE 3 CORRELOGRAM OF HAC OLS REGRESSION

Sample: 1 54
Included observations: 54

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.347	0.347	6.8793	0.009
		2	0.110	-0.012	7.5853	0.023
		3	0.134	0.114	8.6578	0.034
		4	-0.011	-0.106	8.6647	0.070
		5	-0.137	-0.122	9.8180	0.081
		6	0.019	0.114	9.8405	0.132
		7	-0.003	-0.033	9.8413	0.198
		8	-0.123	-0.101	10.828	0.212
		9	0.197	0.301	13.436	0.144
		10	0.105	-0.101	14.198	0.164
		11	-0.031	-0.022	14.266	0.219
		12	0.031	0.006	14.335	0.280
		13	-0.041	-0.125	14.459	0.342
		14	-0.137	0.020	15.868	0.321
		15	-0.046	-0.007	16.034	0.380
		16	-0.085	-0.141	16.614	0.411
		17	-0.188	-0.002	19.491	0.301
		18	-0.006	0.017	19.495	0.362
		19	0.019	-0.031	19.526	0.424
		20	-0.205	-0.193	23.271	0.276
		21	-0.098	-0.011	24.153	0.286
		22	-0.044	-0.014	24.338	0.330
		23	-0.152	-0.056	26.579	0.274
		24	-0.060	0.009	26.938	0.307

FIGURE 3 CORRELOGRAM OF OLS REGRESSION

It can be seen from Figure 3 and 4 that there is no difference in autocorrelation between the regressions. Moreover, from Figure 5 and 6 it can be seen that in a test of joint significance of all regressors, or a test of individual significance of each regressor, at 5% level of significance, all are significant. The only difference is that the HAC standard errors are smaller than the normal OLS standard errors.

Dependent Variable: FRACTION
Method: Least Squares
Date: 05/27/22 Time: 11:27
Sample: 1 54
Included observations: 54
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 4.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.344533	0.009504	36.25155	0.0000
TOTAL	-0.154188	0.009132	-16.88380	0.0000
SIZE	0.014982	0.002645	5.663901	0.0000

R-squared	0.829157	Mean dependent var	0.327143
Adjusted R-squared	0.822457	S.D. dependent var	0.053398
S.E. of regression	0.022500	Akaike info criterion	-4.696687
Sum squared resid	0.025818	Schwarz criterion	-4.586188
Log likelihood	129.8106	Hannan-Quinn criter.	-4.654072
F-statistic	123.7596	Durbin-Watson stat	1.290344
Prob(F-statistic)	0.000000	Wald F-statistic	172.5995
Prob(Wald F-statistic)	0.000000		

FIGURE 5 OLS REGRESSION OF FRACTION ON TOTAL AND SIZE

Dependent Variable: FRACTION
Method: Least Squares
Date: 05/27/22 Time: 02:45
Sample: 1 54
Included observations: 54

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.344533	0.007838	43.95497	0.0000
TOTAL	-0.154188	0.011358	-13.57484	0.0000
SIZE	0.014982	0.001505	9.953125	0.0000

R-squared	0.829157	Mean dependent var	0.327143
Adjusted R-squared	0.822457	S.D. dependent var	0.053398
S.E. of regression	0.022500	Akaike info criterion	-4.696687
Sum squared resid	0.025818	Schwarz criterion	-4.586188
Log likelihood	129.8106	Hannan-Quinn criter.	-4.654072
F-statistic	123.7596	Durbin-Watson stat	1.290344
Prob(F-statistic)	0.000000		

FIGURE 6 OLS REGRESSION OF FRACTION ON TOTAL AND SIZE WITH HAC

The second method is to estimate the regression for the model:

$$fraction_i = \beta_0 + \beta_1 total_i + \beta_2 size_i + \alpha_1 fraction_{i-1} + u_i$$

As our results from 1(g) suggest the presence of first order autocorrelation.

This gives us an estimated equation of

$$\widehat{fraction} = \underset{(0.012)}{0.333} - \underset{(0.012)}{0.138}\, total + \underset{(0.002)}{0.016}\, size + \underset{(0.131)}{0.423}\, fraction_{i-1}$$

From Figure 7, it can be seen that our regression shows statistical significance for all regressors, individually and jointly, at the 5% level of significance. Our standard errors are also larger, however, than our normal OLS regression. However, from our correlogram from Figure 8, there is reason to prefer this model as our first-order AC is no longer significant compared to the ordinary OLS regression one in Figure 4.

Dependent Variable: FRACTION
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)
Date: 05/27/22 Time: 12:04
Sample (adjusted): 2 54
Included observations: 53 after adjustments
Convergence achieved after 7 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.333432	0.012416	26.85484	0.0000
TOTAL	-0.137669	0.011723	-11.74327	0.0000
SIZE	0.015597	0.002308	6.758742	0.0000
AR(1)	0.422745	0.131357	3.218299	0.0023

R-squared	0.855797	Mean dependent var	0.326480
Adjusted R-squared	0.846968	S.D. dependent var	0.053684
S.E. of regression	0.021001	Akaike info criterion	-4.816042
Sum squared resid	0.021611	Schwarz criterion	-4.667340
Log likelihood	131.6251	Hannan-Quinn criter.	-4.758858
F-statistic	96.93282	Durbin-Watson stat	1.997113
Prob(F-statistic)	0.000000		

Inverted AR Roots	.42

FIGURE 7 OLS REGRESSION WITH AR(1) USING CLS

Sample (adjusted): 2 54
Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	-0.000	-0.000	1.E-05	
		2	-0.064	-0.064	0.2335	0.629
		3	0.144	0.144	1.4350	0.488
		4	-0.061	-0.068	1.6564	0.647
		5	-0.197	-0.183	4.0246	0.403
		6	0.021	-0.004	4.0526	0.542
		7	0.008	0.006	4.0568	0.669
		8	-0.218	-0.180	7.1322	0.415
		9	0.245	0.244	11.099	0.196
		10	0.088	0.027	11.625	0.235
		11	-0.092	-0.029	12.207	0.271
		12	0.066	0.002	12.517	0.326
		13	-0.016	-0.091	12.536	0.404
		14	-0.141	-0.047	14.020	0.372
		15	0.012	0.037	14.030	0.447
		16	-0.011	-0.079	14.040	0.523
		17	-0.128	-0.007	15.365	0.498
		18	0.019	-0.041	15.396	0.567
		19	0.135	0.068	16.966	0.525
		20	-0.215	-0.206	21.060	0.333
		21	-0.024	-0.048	21.112	0.391
		22	0.069	-0.002	21.562	0.425
		23	-0.149	-0.073	23.721	0.362
		24	0.039	0.066	23.873	0.411

*Probabilities may not be valid for this equation specification.

FIGURE 8 CORRELOGRAM OF REGRESSION WITH AR(1)

Appendix for Question 1

Question 1(c)

Dependent Variable: UHAT^2
Method: Least Squares
Date: 05/27/22 Time: 02:46
Sample: 1 54
Included observations: 54

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000847	0.000581	1.458101	0.1513
TOTAL	0.000653	0.001721	0.379244	0.7062
SIZE	-0.000299	0.000246	-1.212696	0.2312
TOTAL^2	0.000405	0.001373	0.294836	0.7694
SIZE^2	5.16E-05	2.59E-05	1.995120	0.0517
TOTAL*SIZE	-0.000333	0.000203	-1.639722	0.1076

R-squared	0.142432	Mean dependent var	0.000478
Adjusted R-squared	0.053102	S.D. dependent var	0.000862
S.E. of regression	0.000839	Akaike info criterion	-11.22399
Sum squared resid	3.38E-05	Schwarz criterion	-11.00299
Log likelihood	309.0477	Hannan-Quinn criter.	-11.13876
F-statistic	1.594442	Durbin-Watson stat	1.895795
Prob(F-statistic)	0.179688		

Question 1(d)

Dependent Variable: UHAT^2
Method: Least Squares
Date: 05/27/22 Time: 03:06
Sample: 1 54
Included observations: 54

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005476	0.003443	1.590725	0.1179
FRACTIONHAT	-0.033388	0.021372	-1.562269	0.1244
FRACTIONHAT^2	0.054183	0.032904	1.646684	0.1058

R-squared	0.058016	Mean dependent var	0.000478
Adjusted R-squared	0.021075	S.D. dependent var	0.000862
S.E. of regression	0.000853	Akaike info criterion	-11.24121
Sum squared resid	3.71E-05	Schwarz criterion	-11.13071
Log likelihood	306.5128	Hannan-Quinn criter.	-11.19860
F-statistic	1.570515	Durbin-Watson stat	1.825278
Prob(F-statistic)	0.217828		

Question 1(f)

Dependent Variable: UHAT
Method: Least Squares
Date: 05/27/22 Time: 11:18
Sample (adjusted): 2 54
Included observations: 53 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004279	0.007769	-0.550734	0.5843
TOTAL	0.008024	0.011150	0.719654	0.4752
SIZE	3.37E-05	0.001442	0.023391	0.9814
UHAT(-1)	0.369080	0.135729	2.719251	0.0090

R-squared	0.131792	Mean dependent var	-0.000323
Adjusted R-squared	0.078636	S.D. dependent var	0.022153
S.E. of regression	0.021264	Akaike info criterion	-4.791152
Sum squared resid	0.022155	Schwarz criterion	-4.642450
Log likelihood	130.9655	Hannan-Quinn criter.	-4.733968
F-statistic	2.479354	Durbin-Watson stat	1.979655
Prob(F-statistic)	0.072087		