ASSIGNMENT – 3 (Heteroscedasticity)

This study is conducted for the state of **West Bengal using IHDS-**II data for finding out how various factors influence the food share in expenditure.

In the first case, we use only the squared residuals versus predicted values to catch heteroscedasticity in our models.

1(a) $\text{food-share}_i = \alpha_0 + \alpha_1*(\text{lnincomepc})_i + \alpha_2*(\text{lnn})_i + \delta_1*n1_i + \delta_2*n2_i + \delta_3*n3_i + \delta_4*n4_i + \delta_5*n5_i + \delta_7*n7_i + \delta_8*n8_i + \varepsilon_i$

1(b) $\text{food-share}_i = \gamma_0 + \gamma_1*(\text{incomepc})_i + \alpha_2*(\text{lnn})_i + \delta_1*n1_i + \delta_2*n2_i + \delta_3*n3_i + \delta_4*n4_i + \delta_5*n5_i + \delta_7*n7_i + \delta_8*n8_i + \varepsilon_i$

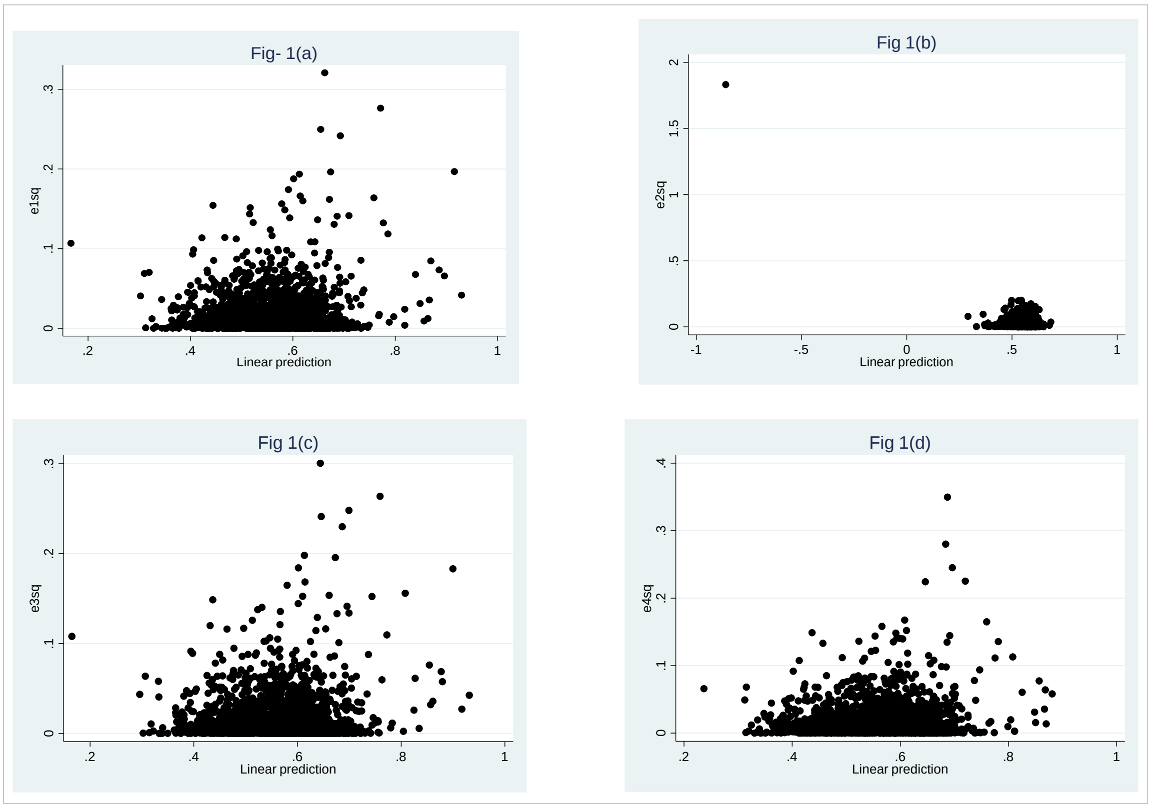
1(c) $\text{food-share}_i = \alpha_0 + \alpha_1*(\text{lnincomepc})_i + \theta_1*hs1_i + \theta_2*hs2_i + \theta_3*hs3_i + \theta_4*hs4_i + \theta_5*hs5_i + \theta_6*hs6_i + \theta_7*hs7_i + \delta_1*n1_i + \delta_2*n2_i + \delta_3*n3_i + \delta_4*n4_i + \delta_5*n5_i + \delta_7*n7_i + \delta_8*n8_i + \varepsilon_i$

1(d) $\text{food-share}_i = \alpha_0 + \alpha_1*(\text{lnincomepc})_i + \alpha_2*(\text{lnn})_i + \delta_1*n1_i + \delta_2*n2_i + \delta_3*n3_i + \delta_4*n4_i + \delta_5*n5_i + \delta_7*n7_i + \delta_8*n8_i + \gamma_{relgr1}*relgr1_i + \gamma_{relgr2}*relgr2_i + \gamma_{relgr3}*relgr3_i + \beta_{casgr1}*casgr1_i + \beta_{casgr2}*casgr2_i + \beta_{casgr3}*casgr3_i + \beta_{casgr5}*casgr5_i + \beta_{casgr6}*casgr6_i + \varepsilon_i$

Variable Description:

food-share	food expenditure share in consumption
lnincomepc	Logarithm of annual per capita income of the household
lnn	Logarithm of household size
n1	Proportion of female children to the household size
n2	Proportion of male children to the household size
n3	Proportion of adult males(age 21-60) to the household size
n4	Proportion of adult females(age 21-60) to the household size
n5	Proportion of male teenagers to the household size
n7	Proportion of elderly males(above age 60) to the household size
n8	Proportion of elderly females(above age 60) to the household size
hs1	Household size 1
hs2	Household size 2
hs3	Household size 3
hs4	Household size 4
hs5	Household size 5
hs6	Household size 6
hs7	Household size 7
relgr1	Hindu family
relgr2	Muslim family
relgr3	Christian family
casgr1	Caste - Brahmin
casgr2	Caste – General/Forward(except Brahmin)
casgr3	Caste- OBC
casgr5	Caste- ST
casgr6	Caste- All other castes

In each of these models, we apply OLS method to find the estimated coefficients and then predict all the residuals from our model. Then, the residuals are squared and plotted against the predicted food-share values. Our aim is the detect any pattern in the scatter plot which implies heteroscedasticity.



Graphical analysis

In figure 1(a) and 1(c), we find that the squared residuals on an average, increases with the fitted values. The pattern is comparatively clearer in 1(c) which indicates the presence of heteroscedasticity. Figure 1(b) is not very clear as the residuals are clustered around a specific range.

The last model which includes caste dummy variables and religion dummy variables shows a very different scatter plot. In Figure1(d), the residuals on an average increase in the beginning and after a certain range falls with the fitted food-share values.

We can conclude that none of the figures shows a random pattern of squared residual errors and it is important to perform a test of heteroscedasticity. In this analysis, the special case of White test is used to detect heteroscedasticity.

Q (2) Special case of White test

We perform the special case of White test in each of our models to catch heteroscedasticity. The intuition behind this test is to find if the residuals is related to the predicted dependent variable values. So, we first run the original regression using OLS and obtain the squared residuals,e_i^2 and the predicted food share in expenditure ,$\hat{Y}_i$.

We then run an auxiliary regression $e_i^2 = \alpha_0 + \alpha_1\hat{Y}_i + \alpha_3 \hat{Y}_i^2 + \eta_i$

The null hypothesis is $H_0 : \alpha_1 = \alpha_3 = 0$ (homoskedasticity) against the alternative H_1: at least one of α_1, α_3 is non-zero(heteroscedasticity).

Then, the F-test is used with df(2,n-3) and conclusions are drawn accordingly.

Model 1(a)

$$\text{food-share}_i = \alpha_0 + \alpha_1*(\text{lnincomepc})_i + \alpha_2*(\text{lnn})_i + \delta_1*n1_i + \delta_2*n2_i + \delta_3*n3_i + \delta_4*n4_i + \delta_5*n5_i + \delta_7*n7_i + \delta_8*n8_i + \varepsilon_i$$

The results of linear regression are presented in the table below

Linear regression of model 1(a)

food share	Coef.	Std Err.	t-value	p-value	[95% Conf	Interval]	Sig
ln(incomepc)	-0.078	0.003	-24.56	0	-0.084	-0.072	***
ln(n)	-0.031	0.008	-4.05	0	-0.046	-0.016	***
n1	0.087	0.032	2.7	0.007	0.024	0.149	***
n2	0.078	0.032	2.41	0.016	0.015	0.141	**
n3	0.084	0.033	2.57	0.01	0.02	0.149	**
n4	-0.011	0.031	-0.35	0.724	-0.073	0.05	
n5	0.042	0.036	1.17	0.244	-0.029	0.113	
o.n6	0	.	.	.	.	.	
n7	0.022	0.035	0.63	0.531	-0.046	0.09	
n8	0.025	0.033	0.76	0.448	-0.04	0.091	
Constant	1.326	0.041	32.44	0	1.246	1.406	***

Mean dependent var	0.564	SD dependent var	0.154
R-squared	0.258	Number of observations	2142
F-test	82.373	Prob > F	0
Akaike crit. (AIC)	-2542.21	Bayesian crit. (BIC)	-2485.515

*** *p<0.01*, ** *p<0.05*, * *p<0.1*

Now, we use the squared residuals from this above model - e_{1i}^2 and regress it on $\hat{Y}_i$ and $\hat{Y}_i^2$.

The auxiliary regression is : $e_{1i}^2 = \alpha_0 + \alpha_1 \hat{Y}_i + \alpha_3 \hat{Y}_i^2 + \eta_i$.

Auxiliary Regression

e1sq	Coef.	Std Error	t-value	p-value	[95% Confidence	Interval]	Sig
foodsharef	-0.377	0.058	-6.5	0	-0.491	-0.263	***
foodsharefsq	0.361	0.051	7.02	0	0.26	0.462	***
Constant	0.113	0.016	6.94	0	0.081	0.145	***

Mean dependent var	0.018	SD dependent var	0.027
R-squared	0.029	Number of observations	2142
F-test	31.694	Prob > F	0
Akaike crit. (AIC)	-9508.781	Bayesian crit. (BIC)	-9491.772

*** *p<0.01, ** p<0.05, * p<0.1*

$F(2, 2139) = 31.694$ and the LM statistic = $n.R^2_u = 2142 * 0.029 = 62.118$.

The LM statistic follows χ^2 with 2 df.

The p-value of the F-test is $0 < 0.05$. So, on the basis of the given evidence, we reject the null hypothesis of homoscedasticity at 5% level of significance.

This implies that the error variance in our model is not constant and it varies with the predicted values which is also consistent with our previous graphical analysis where we find that the errors tend to increase with the fitted values on an average. However, it was difficult to comment on heteroscedasticity just with our scatter plot and it was important to run the special case of White test to confirm our beliefs of heteroscedasticity.

Model – 1(b)

$\text{food-share}_i = \gamma_0 + \gamma_1 * (\text{incomepc})_i + \alpha_2 * (\ln n)_i + \delta_1 * n1_i + \delta_2 * n2_i + \delta_3 * n3_i + \delta_4 * n4_i + \delta_5 * n5_i + \delta_7 * n7_i + \delta_8 * n8_i + \varepsilon_i$

The results of our linear regression are presented in the table below.

Linear regression of model 1(b)

food share	Coef.	Std Error	t-value	p-value	[95% Confidence Interval]		Sig
income(pc)	-5.19E-07	4.44E-08	-11.7	0	0	0	***
ln(n)	-0.032	0.008	-3.79	0	-0.048	-0.015	***
n1	0.11	0.035	3.14	0.002	0.042	0.179	***
n2	0.104	0.035	2.93	0.003	0.034	0.173	***
n3	-0.026	0.036	-0.73	0.463	-0.096	0.044	
n4	-0.103	0.034	-3	0.003	-0.17	-0.036	***
n5	0.052	0.04	1.3	0.192	-0.026	0.13	
o.n6	0	.	.	.	.	.	
n7	-0.049	0.038	-1.28	0.201	-0.123	0.026	
n8	-0.038	0.037	-1.03	0.303	-0.11	0.034	
Constant	0.639	0.032	19.69	0	0.575	0.703	***
Mean dependent var	0.564		SD dependent var	0.154			
R-squared	0.106		Number of obs	2142			
F-test	27.943		Prob > F	0			
Akaike crit. (AIC)	-2141.838		Bayesian crit. (BIC)	-2085.143			

**** p<0.01, ** p<0.05, * p<0.1*

The auxiliary regression is : $e_{2i}^2 = \alpha_0 + \alpha_1 \hat{Y}_i + \alpha_3 \hat{Y}_i^2 + \eta_i$.

Auxiliary Regression

e2sq	Coef.	Std Error	t-value	p-value	[95% Confidence Interval]		Sig
foodsharef2	-1.013	0.015	-65.94	0	-1.043	-0.982	***
foodsharef2sq	0.852	0.017	49.22	0	0.818	0.886	***
Constant	0.319	0.007	48.16	0	0.306	0.332	***
Mean dependent var	0.021		SD dependent var	0.048			
R-squared	0.672		Number of observations	2158			

F-test	2210.348	Prob > F	0
Akaike crit. (AIC)	-9427.097	Bayesian crit. (BIC)	-9410.067

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Here, $F(2, 2155) = 2210.348$ and the p- value is 0. So, here also we reject the null hypothesis.

In this case, the graphical analysis was not quite clear as the errors were clustered around a specific region. The reason may be because we did not take logarithm of one independent variable i.e., annual income per capita. So, the special case of White test points out the presence of heteroscedasticity in our model.

Model – 1(c)

$$\text{food-share}_i = \alpha_0 + \alpha_1 * (\text{lnincomepc})_i + \theta_1 * hs1_i + \theta_2 * hs2_i + \theta_3 * hs3_i + \theta_4 * hs4_i + \theta_5 * hs5_i + \theta_6 * hs6_i + \theta_7 * hs7_i + \delta_1 * n1_i + \delta_2 * n2_i + \delta_3 * n3_i + \delta_4 * n4_i + \delta_5 * n5_i + \delta_7 * n7_i + \delta_8 * n8_i + \varepsilon_i$$

The results of regression are presented in this table below.

Linear regression of model 1(c)

food share	Coef.	Std Err.	t-value	p-value	[95% Confidence Interval]		Sig
ln(incomepc)	-0.077	0.003	-24.14	0	-0.083	-0.07	***
hhszgr1	0.112	0.024	4.77	0	0.066	0.158	***
hhszgr2	0.04	0.016	2.51	0.012	0.009	0.071	**
hhszgr3	0.002	0.013	0.14	0.89	-0.024	0.028	
hhszgr4	-0.006	0.013	-0.45	0.656	-0.031	0.02	
hhszgr5	0.006	0.013	0.43	0.67	-0.021	0.032	
hhszgr6	0.006	0.015	0.38	0.705	-0.023	0.035	
hhszgr7	-0.016	0.017	-0.95	0.342	-0.05	0.017	
n1	0.089	0.032	2.78	0.006	0.026	0.151	***
n2	0.083	0.032	2.58	0.01	0.02	0.146	**
n3	0.091	0.033	2.78	0.005	0.027	0.155	***
n4	-0.018	0.031	-0.57	0.569	-0.08	0.044	
n5	0.047	0.036	1.3	0.194	-0.024	0.117	
o.n6	0	.	.	.	.	.	
n7	0.015	0.035	0.43	0.664	-0.054	0.084	
n8	-0.005	0.034	-0.14	0.886	-0.072	0.062	
Constant	1.265	0.04	31.85	0	1.187	1.343	***
Mean dependent var	0.564		SD dependent var	0.154			
R-squared	0.266		Number of obs	2142			
F-test	51.477		Prob > F	0			

Akaike crit. (AIC)	-2554.653	Bayesian crit. (BIC)	-2463.941

*** *p<0.01, ** p<0.05, * p<0.1*

The auxiliary regression is : $e_{3i} = \alpha_0 + \alpha_1\hat{Y}_i + \alpha_3\hat{Y}_i^2 + \eta_i$.

Auxiliary Regression

e3sq	Coef.	Std Error	t-value	p-value	[95% Confidence Interval]		Sig
foodsharef3	-0.319	0.057	-5.61	0	-0.431	-0.207	***
foodsharef3sq	0.309	0.05	6.12	0	0.21	0.408	***
Constant	0.097	0.016	6.09	0	0.066	0.129	***
Mean dependent var	0.018			SD dependent var	0.027		
R-squared	0.023			Number of observations	2142		
F-test	25.532			Prob > F	0		
Akaike crit. (AIC)	-9506.724			Bayesian crit. (BIC)	-9489.716		

*** *p<0.01, ** p<0.05, * p<0.1*

Here, F(2,2139) = 25.532 and the p-value is 0. Here, similarly like before, we reject the null of homoscedasticity. In our graphical analysis, we found the residual square vs fitted values of this model quite similar to model 1(a). But, in this model it is more evident that the squared residuals increase with our fitted food-share values. So, our test merely confirms our conjecture of heteroscedasticity.

Model – 1(d)

$$\text{food-share}_i = \alpha_0 + \alpha_1*(\text{lnincomepc})_i + \alpha_2*(\text{lnn})_i + \delta_1*n1_i + \delta_2*n2_i + \delta_3*n3_i + \delta_4*n4_i + \delta_5*n5_i + \delta_7*n7_i + \delta_8*n8_i + \gamma_{relgr1}*relgr1_i + \gamma_{relgr2}*relgr2_i + \gamma_{relgr3}*relgr3_i + \beta_{casgr1}*casgr1_i + \beta_{casgr2}*casgr2_i + \beta_{casgr3}*casgr3_i + \beta_{casgr5}*casgr5_i + \beta_{casgr6}*casgr6_i + \mathcal{E}_i$$

This model is an extended version of our model 1(a) in which we have introduced dummy variables for religion and caste . The results are presented below.

Linear regression of model 1(d)

food share	Coef.	Std Err.	t-value	p-value	[95% Confidence Interval]		Sig
ln(incomepc)	-0.068	0.003	-20.93	0	-0.075	-0.062	***
ln(n)	-0.026	0.007	-3.43	0.001	-0.04	-0.011	***
n1	0.089	0.031	2.86	0.004	0.028	0.151	***
n2	0.076	0.032	2.42	0.015	0.015	0.138	**
n3	0.126	0.032	3.87	0	0.062	0.189	***
n4	0.02	0.031	0.66	0.51	-0.04	0.081	

n5	0.05	0.035	1.4	0.161	-0.02	0.119	
o.n6	0	.	.	.	.	.	
n7	0.07	0.034	2.03	0.042	0.002	0.137	**
n8	0.06	0.033	1.81	0.07	-0.005	0.124	*
relgr1	0.001	0.032	0.04	0.969	-0.061	0.064	
relgr2	0.063	0.033	1.91	0.056	-0.002	0.127	*
relgr3	-0.004	0.045	-0.09	0.927	-0.092	0.084	
casgr1	-0.082	0.012	-6.84	0	-0.106	-0.059	***
casgr2	-0.06	0.008	-7.98	0	-0.075	-0.046	***
casgr3	-0.046	0.011	-4.1	0	-0.068	-0.024	***
casgr5	0.057	0.022	2.63	0.009	0.014	0.099	***
casgr6	-0.107	0.023	-4.75	0	-0.151	-0.063	***
Constant	1.223	0.053	22.94	0	1.118	1.327	***

Mean dependent var	0.564	SD dependent var	0.154
R-squared	0.298	Number of obs	2142
F-test	53.124	Prob > F	0
Akaike crit. (AIC)	-2645.93	Bayesian crit. (BIC)	-2543.876

*** *p<0.01, ** p<0.05, * p<0.1*

The auxiliary regression is : $e_{3i}^2 = \alpha_0 + \alpha_1 \hat{Y}_i + \alpha_3 \hat{Y}_i^2 + \eta_i$.

Auxiliary Regression

e4sq	Coef.	St Error	t-value	p-value	[95% Confidence Interval]		Sig
foodsharef4	-0.177	0.059	-3.02	0.003	-0.292	-0.062	***
foodsharef4 sq	0.183	0.053	3.48	0.001	0.08	0.286	***
Constant	0.057	0.016	3.52	0	0.025	0.089	***

Mean dependent var	0.017	SD dependent var	0.026
R-squared	0.013	Number of observations	2142
F-test	13.867	Prob > F	0
Akaike crit. (AIC)	-9634.506	Bayesian crit. (BIC)	-9617.498

*** *p<0.01, ** p<0.05, * p<0.1*

Here, the F-statistic value for testing heteroscedasticity is 13.867 and the associated p-value is 0 <0.05.

So, we reject our null hypothesis of homoscedasticity. Figure 1(d) is different from the other scatter plots as in this case, the squared residuals shows a pattern of increasing in the beginning and then falling after a certain point. It is thus clear that the residuals do not show randomness.

Robust standard errors (including F-test and LM-test)

Q 3. Here, we are comparing the usual standard errors and robust standard errors in model 1(d) . Heteroscedasticity causes the standard errors to be biased . OLS assumes the errors are both identically and independently distributed and robust standard errors relax either or both of these assumptions. So, in presence of heteroscedasticity, robust standard errors tend to be more trustworthy.

The regression equation is : $\hat{Y}$= 1.223 – 0.068*lnincomepc – 0.026*lnn + 0.089*n1 + 0.076*n2

+ 0.126*n3 + 0.020*n4 + 0.050*n5 + 0.070*n7 + 0.060*n8

+ 0.001*relgr1 + 0.063*relgr2 – 0.004*relgr3 – 0.082*casgr1

- 0.060*casgr2 – 0.046*casgr3 + 0.057*casgr5 – 0.107*casgr6

n = 2142 , R^2 = 0.2983

	Usual Standard Errors				Robust Standard Errors			
food share	**Std Err**	**t-value**	**p-value**	**Sig**	**Std Err**	**t-value**	**p-value**	**Sig**
ln(income pc)	0.003	-20.93	0	***	0.004	-17.06	0	***
ln(n)	0.007	-3.43	0.001	***	0.008	-3.36	0.001	***
n1	0.031	2.86	0.004	***	0.03	2.94	0.003	***
n2	0.032	2.42	0.015	**	0.032	2.4	0.016	**
n3	0.032	3.87	0	***	0.034	3.73	0	***
n4	0.031	0.66	0.51		0.032	0.63	0.527	
n5	0.035	1.4	0.161		0.037	1.33	0.185	
o.n6	.	.	.		.	.	.	
n7	0.034	2.03	0.042	**	0.037	1.89	0.058	*
n8	0.033	1.81	0.07	*	0.035	1.71	0.088	*
relgr1	0.032	0.04	0.969		0.029	0.04	0.966	
relgr2	0.033	1.91	0.056	*	0.029	2.13	0.033	**
relgr3	0.045	-0.09	0.927		0.037	-0.11	0.912	
casgr1	0.012	-6.84	0	***	0.011	-7.22	0	***
casgr2	0.008	-7.98	0	***	0.008	-7.62	0	***
casgr3	0.011	-4.1	0	***	0.012	-3.88	0	***
casgr5	0.022	2.63	0.009	***	0.019	2.94	0.003	***
casgr6	0.023	-4.75	0	***	0.022	-4.8	0	***
Constant	0.053	22.94	0	***	0.056	21.76	0	***

We can clearly see that there is not much difference in standard errors and the robust t-statistic does not change the significance of any of the coefficients.

F-test to check if the caste dummy variables are significant or not in both the regressions.

Suppose we want to test the null hypothesis that after other factors are controlled for, there are no differences in average food-share in expenditure by caste.

The null hypothesis is $H_0 : \beta_{casgr1} = 0, \beta_{casgr2} = 0, \beta_{casgr3} = 0, \beta_{casgr5} = 0, \beta_{casgr6} = 0$

In the usual case , the F-statistic is $F(5,2124) = 21.23$ and associated p-value = $0.000 < 0.05$

So, we reject the null hypothesis at 5% significance level using the p-value.

Now, let us consider the case where we use the robust standard errors instead of usual standard errors and check the joint significance of the caste dummies.

Here, F-statistic value is $F(5,2124) = 20.94$ and the associated p-value = $0.000 < 0.05$.

So, we reject the null hypothesis at 5% significance level using the p-value.

LM test to check the significance of caste dummies

The regression model is food-share = $\alpha_0 + \alpha_1$*(lnincomepc) + α_2*(lnn) + δ_1*n1 + δ_2*n2

+ δ_3*n3 + δ_4*n4 + δ_5*n5 + δ_7*n7 + δ_8*n8

+ γ_{relgr1}*relgr1 + γ_{relgr2}*relgr2 + γ_{relgr3}*relgr3 + β_{casgr1}*casgr1

+ β_{casgr2} *casgr2 + β_{casgr3}*casgr3 + β_{casgr5}*casgr5+ β_{casgr6} *casgr6

+ ε

We use the LM statistic to test the null that $H_0 : \beta_{casgr1} = 0, \beta_{casgr2} = 0, \beta_{casgr3} = 0, \beta_{casgr5} = 0, \beta_{casgr6} = 0$

against the alternative that caste have an influence on our dependent variable controlling other factors.

First, we estimate the restricted model by regressing Y on all regressors except the caste dummies and obtain our residuals, $\tilde{u}$. Then, the residuals are regressed on all the independent variables including the excluded ones in the beginning.

uhat	Coef.	St.Err.	t-value	p-value	[95% Confidence Interval]		Sig
lnincomepc	0.008	0.003	2.6	0.009	0.002	0.015	***
lnn	0.004	0.007	0.56	0.574	-0.01	0.019	
n1	0.007	0.031	0.23	0.817	-0.054	0.069	
n2	0.005	0.032	0.15	0.884	-0.057	0.066	
n3	0.035	0.032	1.07	0.286	-0.029	0.098	
n4	0.024	0.031	0.77	0.444	-0.037	0.084	
n5	0.006	0.035	0.16	0.875	-0.064	0.075	
o.n6	0	.	.	.	.	.	
n7	0.044	0.034	1.28	0.201	-0.023	0.111	
n8	0.024	0.033	0.73	0.464	-0.04	0.089	
relgr1	0.06	0.032	1.89	0.058	-0.002	0.123	*
relgr2	0.098	0.033	2.98	0.003	0.033	0.162	***
relgr3	0.027	0.045	0.61	0.543	-0.061	0.115	
casgr1	-0.082	0.012	-6.84	0	-0.106	-0.059	***
casgr2	-0.06	0.008	-7.98	0	-0.075	-0.046	***
casgr3	-0.046	0.011	-4.1	0	-0.068	-0.024	***
casgr5	0.057	0.022	2.63	0.009	0.014	0.099	***
casgr6	-0.107	0.023	-4.75	0	-0.151	-0.063	***
Constant	-0.139	0.053	-2.61	0.009	-0.244	-0.035	***

Mean dependent var	0	SD dependent var	0.133
R-squared	0.048	Number of observations	2142
F-test	6.243	Prob > F	0
Akaike crit. (AIC)	-2645.927	Bayesian crit. (BIC)	-2543.876

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The LM statistic is $n*R^2_u = 101.9348$ and the p-value is 3.893e-14 <0.05.

So, we reject the null hypothesis at 5% level of significance.

Heteroscedasticity-robust LM statistic

To obtain the heteroscedasticity-robust LM statistic, we need to perform a series of steps.

Firstly, we obtain the residuals, $u^{\sim}$ from our restricted model.

Secondly, we regress each of the excluded independent variables under null i.e., all the 5 caste dummies on each of the included independent variables which leads to 5 sets of residuals($r_1^{\sim}$, $r_2^{\sim}$, $r_3^{\sim}$, $r_4^{\sim}$, $r_5^{\sim}$) .

Thirdly, we find the products between these 5 set of residuals and $u^{\sim}$ (p1,p2,p3,p5,p6)

Fourthly, we generate a new variable one which takes only value 1.

Lastly, we regress one on p1, p2, p3, p5, p6 without constant term.

The heteroscedasticity robust LM statistic is n- SSR where SSR is the usual sum of squared residuals from the final regression. Under H_0, LM is distributed approximately as χ^2 with 5 degrees of freedom here.

Once, the robust LM statistic is obtained, we can take our decision whether to accept or reject the null on the basis of p-value or critical value.

Here, LM = n – SSR = 2142 – 2050.3718 = 91.6282 which is greater than the critical value $\chi^2(5)$ at 0.05 level of significance = 11.07.

So, here also we reject our null hypothesis on the basis of robust LM statistic.

Q 4 . **Comparing the models by checking statistical significance of coefficients using usual standard errors and heteroscedasticity-robust standard errors**

Model- 1(d)

In this model we are regressing food expenditure share on log(annual per capita income), ln(household-size) , religion dummy variables and caste dummy variables.

Firstly, we apply our usual OLS technique and use the standard errors and t-values of coefficients and comment on their statistical significance. Then, we account for heteroscedasticity and use heteroscedasticity robust standard errors and t-values to find significance and compare them with our previous results.

Our hypothesised model is $\text{food-share}_i = \alpha_0 + \alpha_1 * (\text{lnincomepc})_i + \alpha_2 * (\text{lnn})_i + \delta_1 * n1_i + \delta_2 * n2_i$

$$+ \delta_3 * n3_i + \delta_4 * n4_i + \delta_5 * n5_i + \delta_7 * n7_i + \delta_8 * n8_i$$

$$+ \gamma_{relgr1} * relgr1_i + \gamma_{relgr2} * relgr2_i + \gamma_{relgr3} * relgr3_i + \beta_{casgr1} * casgr1_i$$

$$+ \beta_{casgr2} * casgr2_i + \beta_{casgr3} * casgr3_i + \beta_{casgr5} * casgr5_i + \beta_{casgr6} * casgr6_i$$

$$+ \varepsilon_i$$

We apply OLS to get our estimated model as :

$$\hat{Y} = 1.223 - 0.068*(\text{lnincomepc}) - 0.026*(\text{lnn}) + 0.089*n1 + 0.076*n2$$

$$+ 0.126*n3 + 0.020*n4 + 0.050*n5 + 0.070*n7 + 0.060*n8$$

$$+ 0.001*relgr1 + 0.063*relgr2 - 0.004*relgr3 - 0.082*casgr1$$

$$- 0.060*casgr2 - 0.046*casgr3 + 0.057*casgr5 - 0.107*casgr6$$

$$n = 2142 ,\ R^2 = 0.2983$$

The slope coefficients are the same when we use the heteroscedasticity robust case. So, we present a table below comparing the standard errors, t-values and significance at 5% level.

foodshare	Usual Standard Errors				Robust Standard Errors			
	Std Err	t-value	p-value	Sig	Std Err	t-value	p-value	Sig
ln(income pc)	0.003	-20.93	0	***	0.004	-17.06	0	***
ln(n)	0.007	-3.43	0.001	***	0.008	-3.36	0.001	***
n1	0.031	2.86	0.004	***	0.03	2.94	0.003	***
n2	0.032	2.42	0.015	**	0.032	2.4	0.016	**
n3	0.032	3.87	0	***	0.034	3.73	0	***
n4	0.031	0.66	0.51		0.032	0.63	0.527	
n5	0.035	1.4	0.161		0.037	1.33	0.185	
o.n6	.	.	.		.	.	.	
n7	0.034	2.03	0.042	**	0.037	1.89	0.058	*
n8	0.033	1.81	0.07	*	0.035	1.71	0.088	*
relgr1	0.032	0.04	0.969		0.029	0.04	0.966	
relgr2	0.033	1.91	0.056	*	0.029	2.13	0.033	**
relgr3	0.045	-0.09	0.927		0.037	-0.11	0.912	
casgr1	0.012	-6.84	0	***	0.011	-7.22	0	***
casgr2	0.008	-7.98	0	***	0.008	-7.62	0	***
casgr3	0.011	-4.1	0	***	0.012	-3.88	0	***
casgr5	0.022	2.63	0.009	***	0.019	2.94	0.003	***
casgr6	0.023	-4.75	0	***	0.022	-4.8	0	***
Constant	0.053	22.94	0	***	0.056	21.76	0	***

**** p<0.01, ** p<0.05, * p<0.1*

Comparison

First of all, we can see that there is almost a negligible difference in the two sets of standard errors in this model. In some cases, the heteroscedastic robust standard errors are greater than the usual and in other cases it is less in value. The standard errors are same in two cases – n2 (proportion of male children in a household) and casgr2(caste dummy variable for general except brahmins).

The t-values are also similar because of the similarity in the two sets of standard errors. The largest absolute difference occurs for the log of per-capita annual income variable.

n7(proportion of elderly male i.e., above age 60 in a household) is no longer significant at 5% level and relgr2(religion dummy variable for Muslims), which was previously not significant at 5% level is now significant. This means that, keeping other factors constant, the proportion of elderly men in a household has no influence in the household's food expenditure share which is probably because elderly people tend to eat less compared to other household members. The insignificance of elderly female coefficient supports the results. Now, relgr2 denotes whether the household is a Muslim family or not. The significance of relgr2 means that the average food expenditure share in a Muslim household is significantly different from the base group (i.e., other households except Hindus and Christians like the Sikhs, Jains, etc)

Apart from these two, the results are the same.

It may be concluded that in our analysis, using the heteroscedasticity robust standard errors fail to give us any significantly different insights as compared to our previous results.