HW1 Econometrics 3

Matthew Aaron Looney

9/11/2017

The Model:

$$y_t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_t \tag{1}$$

where,

ε_t is normally and independently distributed with $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma_t^2$ and $\sigma_t^2 = exp(\alpha_0 + \alpha_1 x_1)$.

The paramater values used are $\beta_0 = 10$, $\beta_1 = \beta_2 = 1$, $\alpha_0 = -2$, and $\alpha_1 = 0.25$.

The design matrix X is given in a csv file.

Problem 1

The following questions are based on Monte Carlo experimental data. Generate 100 samples of y, each of size 20, using the model given in equation 1 (above).

(a)

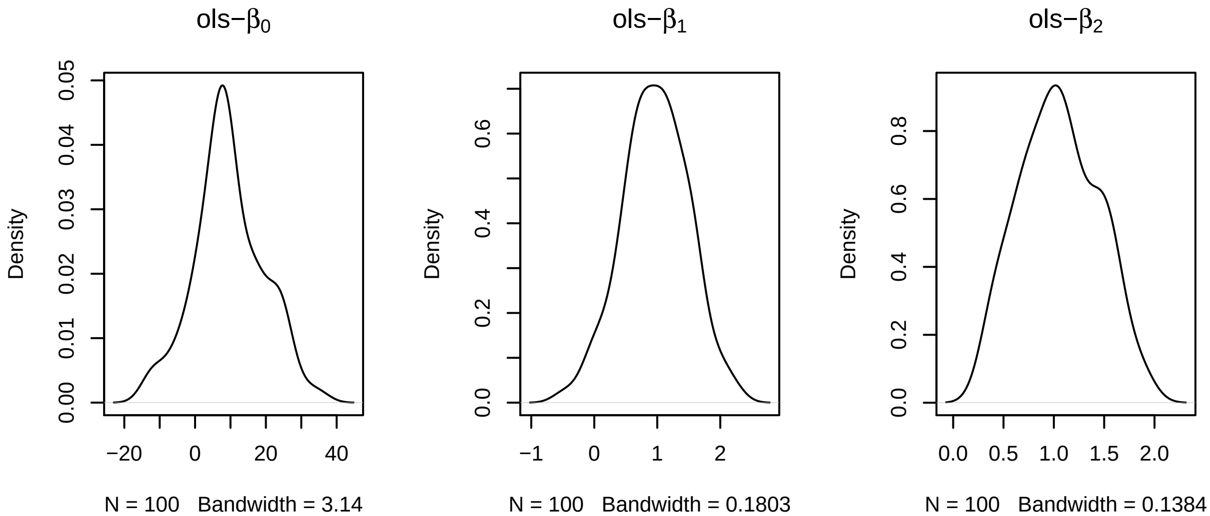
Estiamte the parameters using the least squares principle and provide their covariance matrix. Compare your results with the true parameters. What can you conclude?

Table 1: Summary Statistics: Problem 1, part (a) and (b)

Statistic	N	Mean	St. Dev.	Min	Max
beta_0_ols	100	9.590	9.716	−13.569	35.337
beta_1_ols	100	0.971	0.520	−0.481	2.243
beta_2_ols	100	1.050	0.386	0.344	1.896
var_b0_ols	100	70.313	35.817	17.030	193.850
var_b1_ols	100	0.203	0.103	0.049	0.559
var_b2_ols	100	0.145	0.074	0.035	0.399
se_b0_ols	100	8.139	2.029	4.127	13.923
se_b1_ols	100	0.437	0.109	0.222	0.748
se_b2_ols	100	0.369	0.092	0.187	0.632
t_val_b0_ols	100	1.262	1.271	−1.565	5.098
t_val_b1_ols	100	2.335	1.219	−0.644	5.307
t_val_b2_ols	100	3.038	1.401	0.665	7.349
BP_testStat_ols	100	6.292	4.894	0.061	20.898
GV_HET_Test_ols	100	4.797	3.828	0.001	16.076

BP testStat ols = Breusch Pagan test against heteroskedasticity
GV HET Test ols = General test for LM assumptions
Chi Squared Crit df:2 at alpha:0.05 = 5.99

The results displayed in table 1 (above) show the summary for the 100 sample experiments preformed on the data using OLS. The means of the 100 samples are reported for various statistics. The average t-values for the beta paramaters show the constant is insignificant while β_1 and β_2 are significant. To better understand the distribution of the paramaters over the 100 experiments see density plots below.



(b)

Test for the presence of heteroskedasticity? What do you conclude?

I have run the Breusch Pagan (BP) test for heteroskedasticity for the 100 samples and averaged over the experiments. The BP test staistic is distributed as Chi-Squared with k-1 degrees of freedom, where k is the total number of papamaters in the model.

The null hypothesis of the Breusch-Pagan test is,

$$\sigma_i^2 = \sigma^2(\alpha_0 + \alpha' z_i)$$

where,

$H_0 : \alpha = 0$

σ_i^2 is the error variance for the ith observation and α_0 and α are regression coefficients.

The test statistic for the Breusch-Pagan test is

$$bp = \frac{1}{v}(u - \bar{u}i)'Z(Z'Z)^{-1}Z'(u - \bar{u}i)$$

where $u = (e_1^2, e_2^2, ..., e_n^2)$, i is a $nx1$ vector of ones, and

$$v = \frac{1}{n}\sum_{i=1}^{n}\left(e_i^2 - \frac{e'e}{n}\right)^2$$

The result from any one of the 100 experiments may or may not indicate heteroskedasticity, however, when we average over the range of the 100 experiments we obtain a test statistic which would suggest, overall we have a heteroskedasticity problem.

This is a modified version of the Breusch-Pagan test, which is less sensitive to the assumption of normality than the original test (Greene 2012, 7th Ed.; p. 276).

(c)

Assuming multiplicative heteroskedasticity, estimate the parameters using GLS and provide their covariance matrix.

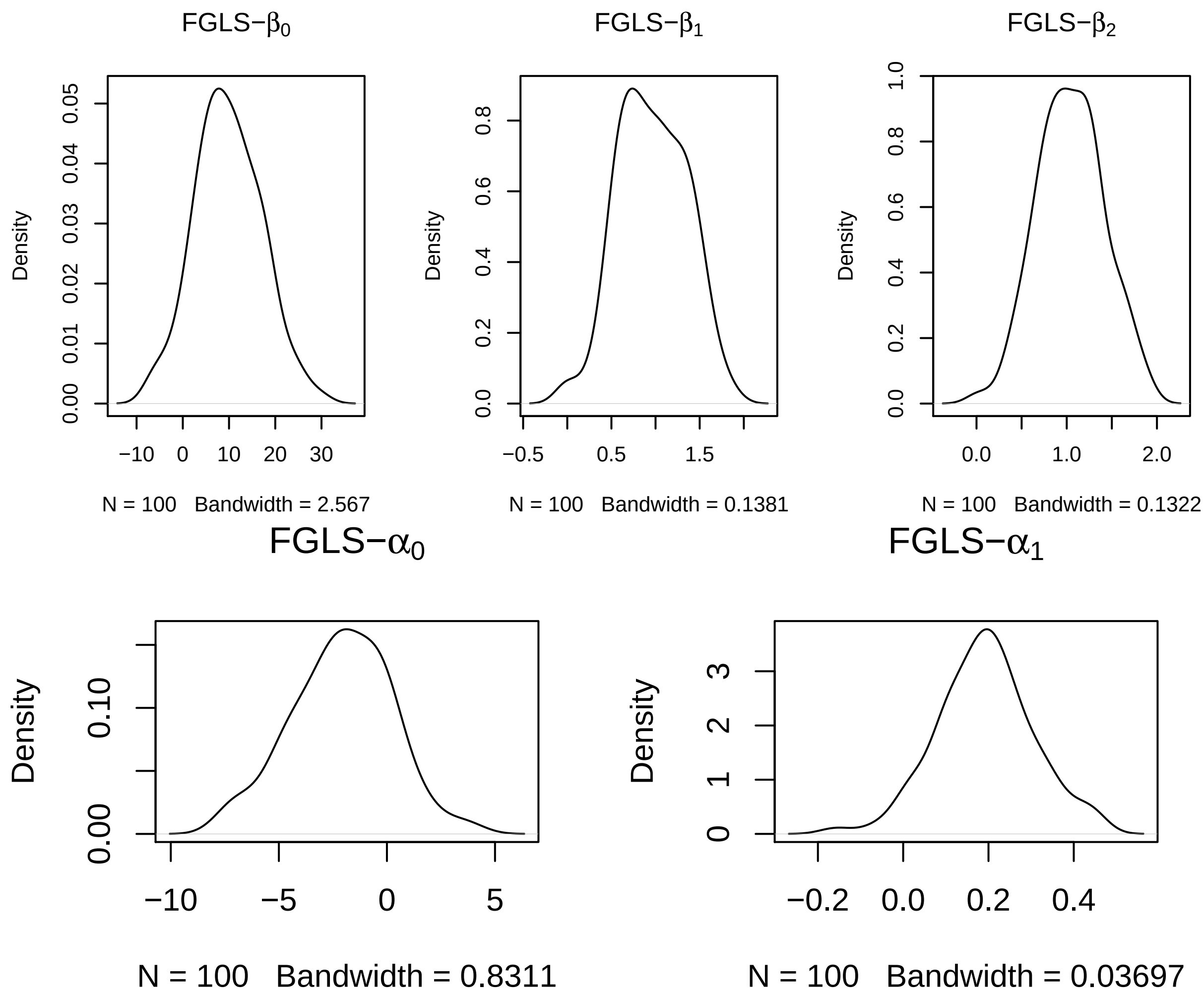
Table 2 (below) shows the summary of results obtained using Feasible generalized least squares (FGLS) to obtain an estimate of the variance-covariance structure. Once we have this structure ($\hat{\Omega}^{-1}$) in hand we can proceed with estimation of the Beta's and the correct variance-covariance.

When comparing the FGLS results with OLS we notice the average of the paramater estimates are similar enough to be considered the same, as we expect. However, we observe the variance-covariance of the FGLS is much smaller, indicating a much better job at recovering the nature of the error structure and addressing the heteroskedasticity problem embedded in the data. Since we have generated the data with known structure and paramaters we can verify the accuracy of our techniques ability to uncover the true paramaters. The FGLS technique does an excellent job at recovering both the beta's of our data and the alpha paramaters of our error structure. Using the FGLS technique we also notice all of our paramater estimates are significant at the 95% level, a change from the OLS procdure where only two of the three paramaters were significant.

To allow us to understand the distribution of the FGLS paramater estimates over the range of the 100 experiments, I provide density plot's (below).

Table 2: Summary Statistics: Problem 1, part (c)

Statistic	N	Mean	St. Dev.	Min	Max
beta_0_ols	100	9.590	9.716	−13.569	35.337
beta_1_ols	100	0.971	0.520	−0.481	2.243
beta_2_ols	100	1.050	0.386	0.344	1.896
var_b0_ols	100	70.313	35.817	17.030	193.850
var_b1_ols	100	0.203	0.103	0.049	0.559
var_b2_ols	100	0.145	0.074	0.035	0.399
se_b0_ols	100	8.139	2.029	4.127	13.923
se_b1_ols	100	0.437	0.109	0.222	0.748
se_b2_ols	100	0.369	0.092	0.187	0.632
t_val_b0_ols	100	1.262	1.271	−1.565	5.098
t_val_b1_ols	100	2.335	1.219	−0.644	5.307
t_val_b2_ols	100	3.038	1.401	0.665	7.349
beta_0_fgls	100	9.836	7.165	−6.479	29.558
beta_1_fgls	100	0.965	0.385	−0.008	1.857
beta_2_fgls	100	1.044	0.369	0.022	1.865
alpha_0_fgls	100	−2.100	2.320	−7.556	3.861
alpha_1_fgls	100	0.192	0.113	−0.157	0.452
var_b0_fgls	100	21.369	20.497	4.819	100.744
var_b1_fgls	100	0.052	0.046	0.013	0.225
var_b2_fgls	100	0.022	0.013	0.008	0.096
se_b0_fgls	100	4.272	1.774	2.195	10.037
se_b1_fgls	100	0.214	0.082	0.115	0.474
se_b2_fgls	100	0.142	0.038	0.088	0.309
t_val_b0_fgls	100	2.573	2.236	−2.263	10.344
t_val_b1_fgls	100	4.911	2.391	−0.045	12.458
t_val_b2_fgls	100	7.864	3.571	0.168	18.313
var_b0_HCCM_0	100	73.048	62.861	8.697	357.686
var_b1_HCCM_0	100	0.233	0.196	0.031	1.094
var_b2_HCCM_0	100	0.085	0.049	0.024	0.280
var_b0_HCCM_3	100	126.622	115.448	12.261	645.598
var_b1_HCCM_3	100	0.395	0.358	0.048	2.010
var_b2_HCCM_3	100	0.164	0.123	0.038	0.704


FGLS−β0
FGLS−β1
FGLS−β2
Density
N = 100 Bandwidth = 2.567
N = 100 Bandwidth = 0.1381
N = 100 Bandwidth = 0.1322
FGLS−α0
FGLS−α1
N = 100 Bandwidth = 0.8311
N = 100 Bandwidth = 0.03697

(d)

Write a Matlab code to estimate the parameters using maximum likelihood method and provide their covariance matrix.

$$\ln L = -0.5n\log(2\pi) - 0.5\sum(\sigma^2) - 0.5\sum\left[\frac{(y - X'\beta)^2}{\sigma^2}\right] \tag{2}$$

where,

$\sigma^2 \simeq \exp(\alpha_0 + \alpha_1 x_1)$

Table 3 (below) gives the summary statistics for the Maximum Likelihood Estimation method (MLE). Again we notice the MLE does a good job at uncovering the true paramaters of our model. The MLE has the added benefit of obtaining all paramater estimates at one time, assuming the MLE assumptions are met.

When we compare the results obtained using the MLE procedure we notice all paramater estimates are similar, however, when we view the variance-covariance matrix we see that the FGLS seems to do a better job over the MLE procedure. This may be due to the optimization routine used for the MLE or the convergence tolerence, or both. More expermineting with the optimization variables could yield better results. At any rate the MLE estimation stratagy does a better job that OLS and just about as good a job as FGLS, in the current context.

To help us better understand the distribution of the MLE paramater estimates over the range of the 100 experiments, I provide density plot's (below).

Table 3: Summary Statistics: Problem 1, part (d)

Statistic	N	Mean	St. Dev.	Min	Max
beta_0_mle	100	9.294	7.485	−14.598	26.088
beta_1_mle	100	1.001	0.381	0.029	1.898
beta_2_mle	100	1.037	0.382	−0.214	1.880
alpha_0_mle	100	−2.742	2.681	−10.509	5.598
alpha_1_mle	100	0.271	0.129	−0.174	0.656
var_b0_mle	100	36.725	17.487	11.604	85.507
var_b1_mle	100	0.113	0.052	0.038	0.291
var_b2_mle	100	0.082	0.032	0.026	0.176
var_alpha_0_mle	100	5.661	2.746	1.795	14.500
var_alpha_1_mle	100	0.013	0.006	0.004	0.033
se_b0_mle	100	5.900	1.391	3.406	9.247
se_b1_mle	100	0.329	0.073	0.195	0.539
se_b2_mle	100	0.281	0.056	0.162	0.419
se_alpha_0_mle	100	2.317	0.544	1.340	3.808
se_alpha_1_mle	100	0.110	0.026	0.062	0.182
t_val_b0_mle	100	1.712	1.489	−2.522	6.053
t_val_b1_mle	100	3.212	1.464	0.113	7.151
t_val_b2_mle	100	3.770	1.468	−0.607	8.509
t_val_alpha_0_mle	100	−1.258	1.180	−4.173	1.470
t_val_alpha_1_mle	100	2.617	1.312	−0.958	5.546

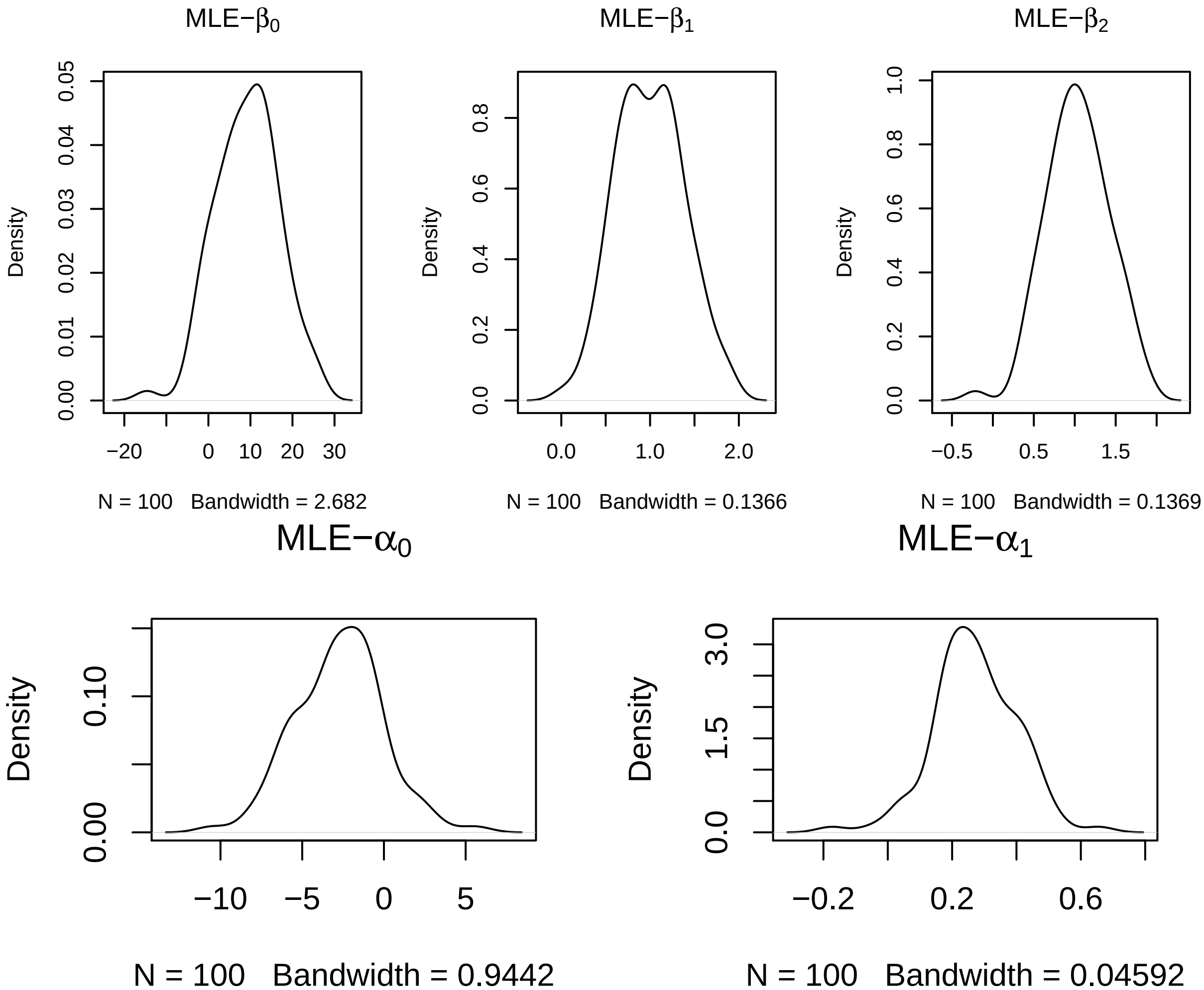

MLE−β0
Density
N = 100 Bandwidth = 2.682
MLE−β1
Density
N = 100 Bandwidth = 0.1366
MLE−β2
Density
N = 100 Bandwidth = 0.1369
MLE−α0
Density
N = 100 Bandwidth = 0.9442
MLE−α1
Density
N = 100 Bandwidth = 0.04592

Table 4: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
beta_0_ols	100	9.530	2.594	3.907	16.808
beta_1_ols	100	0.990	0.192	0.477	1.481
beta_2_ols	100	1.032	0.175	0.550	1.489
var_b0_ols	100	36.308	24.120	7.328	148.470
var_b1_ols	100	0.105	0.070	0.021	0.428
var_b2_ols	100	0.075	0.050	0.015	0.306
se_b0_ols	100	5.767	1.754	2.707	12.185
se_b1_ols	100	0.310	0.094	0.145	0.654
se_b2_ols	100	0.262	0.080	0.123	0.553
t_val_b0_ols	100	1.803	0.711	0.462	3.630
t_val_b1_ols	100	3.479	1.272	1.367	7.728
t_val_b2_ols	100	4.290	1.449	1.426	9.489
DW_Test	100	3.290	0.357	2.252	3.871
rho_fgls	100	0.755	0.159	0.257	0.971
sigma2_fgls	100	6.055	2.114	2.008	12.177
beta_0_fgls	100	8.951	21.059	−66.922	64.418
beta_1_fgls	100	0.987	1.097	−1.709	4.750
beta_2_fgls	100	1.057	0.339	−0.083	1.801
var_b0_fgls	100	8.444	12.104	1.360	81.390
var_b1_fgls	100	0.057	0.090	0.005	0.604
var_b2_fgls	100	0.106	0.177	0.004	1.188
se_b0_fgls	100	2.497	1.494	1.166	9.022
se_b1_fgls	100	0.197	0.135	0.068	0.777
se_b2_fgls	100	0.262	0.194	0.065	1.090
t_val_b0_fgls	100	4.674	6.749	−14.133	17.114
t_val_b1_fgls	100	6.831	5.535	−4.868	17.325
t_val_b2_fgls	100	5.880	3.585	−0.165	16.919
var_b0_HCCM_0	100	27.372	15.930	4.489	81.518
var_b1_HCCM_0	100	0.079	0.046	0.010	0.262
var_b2_HCCM_0	100	0.056	0.057	0.007	0.413
var_b0_HCCM_3	100	43.248	25.773	6.306	143.734
var_b1_HCCM_3	100	0.121	0.071	0.015	0.418
var_b2_HCCM_3	100	0.108	0.118	0.012	0.851

good 1 = meats
good 2 = dairy
good 3 = beans