

Weighted Least Squares Test and Applications: Lecture XIX

Charles B. Moss ¹

¹University of Florida

February 27, 2012

- 1 Readings
- 2 Heteroscedasticity and Weighted Least Squares
 - Weighted Least Squares
 - Differential Demand Model
 - Dutch Consumption Data
 - Systematic Variance
 - Difference in Variance
- 3 Testing for Heteroscedasticity
 - White's General Test
 - Breusch-Pagan/Godfrey LM Test
- 4 Feasible Generalized Least Squares
- 5 Readings – Next Time

Readings

- * Cameron, A. C. and P. K. Trivedi. 2010. *Microeconometrics Using Stata Revised Edition*. Stata Press. (Chapter 5).
- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition (Chapter 9 Section 9.1 9.4 pp. 257 274)
- * Hansen, B.E. 2011. *Econometrics* Draft Textbook Website: <http://www.ssc.wisc.edu/bhansen>. (Chapter 9.2)

Heteroscedasticity and Weighted Least Squares

- Weighted least squares is a correction for heteroscedasticity based potentially on something in the sample.
 - Large firms may have larger variances.
 - The variance in the consumption of goods may be related to the level of consumer income.
 - In the production specification, the variance may be related to a risk-increasing input (i.e., nitrogen).
- Returning to the heteroscedasticity specification

$$V[\epsilon_i | X] = \sigma_i^2, i = 1, \dots, n. \quad (1)$$

Weighted Least Squares

- If we assume a constant variance

$$\begin{aligned}
 E[\epsilon\epsilon'|X] &= \sigma^2\Omega = \sigma^2 \begin{bmatrix} \omega_1 & 0 & 0 & \cdots & 0 \\ 0 & \omega_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \omega_n \end{bmatrix} \\
 &= \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}
 \end{aligned} \tag{2}$$

- Where we can adjust for the arbitrary scaling

$$\text{tr}(\Omega) = \sum_{i=1}^n \omega_i = n \tag{3}$$

- Returning to the differential demand model, this time for the consumer

$$\bar{w}_{it} d \ln (q_{it}) = \theta_i d \ln (Q_t) + \pi_{1i} d \ln (p_{1t}) + \pi_{2i} d \ln (p_{2t}) + \pi_{3i} d \ln (p_{3t}) + \pi_{4i} d \ln (p_{4t}) + \pi_{5i} d \ln (p_{5t}) \quad (4)$$

- Where $\bar{w}_{it}$ is the average budget share,
- $d \ln (q_{it})$ is the change in the logarithm of demand for good i ,
- $d \ln (Q_t)$ is the change in logarithm of overall consumer income, and
- $d \ln (p_{jt})$ is the logarithmic change in price paid for good j .

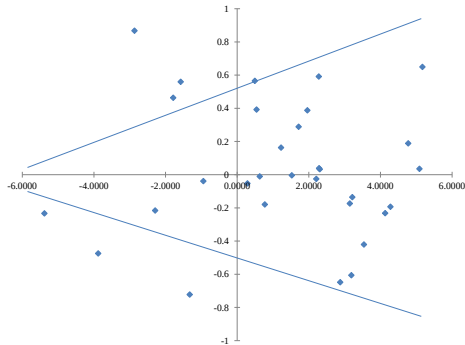
Dutch Consumption Data

	RHS -1*	RHS-2*	RHS-3*	RHS-4*	RHS-5*	dQ	P1	P2	P3	P4	P5
1922-23	0.0659	0.2634	0.5708	-0.9879	0.3764	0.2885	-4.36	3.68	-6.74	-6.98	-7.77
1923-24	0.3538	-0.3695	-0.7077	0.3163	-1.1680	-1.5751	0.98	2.22	4.03	3.04	-5.88
1924-25	-0.3803	-0.5578	-0.5705	-2.7734	-1.0998	-5.3819	10.29	-0.58	3.24	9.67	0.08
1925-26	0.5378	1.1012	0.8803	1.2772	0.9795	4.7759	-4.67	-1.82	-11.18	-13.51	-4.42
1926-27	0.0163	0.0618	1.2425	1.1612	-0.1756	2.3061	-1.51	-0.99	-2.00	-4.14	-8.28
1927-28	0.2905	-0.0990	0.1056	0.2179	1.4460	1.9610	-1.95	0.21	2.94	2.99	0.04
1928-29	-0.6469	0.4793	0.4324	0.3251	-1.9139	-1.3239	-5.30	-2.80	-0.15	-5.20	6.45
1929-30	0.9331	0.9470	0.9746	0.3283	-0.9020	2.2810	-5.80	-7.53	-11.20	-10.78	1.00
1930-31	-0.3961	-0.2819	1.2489	0.8564	2.1124	3.5397	-12.27	-7.55	-14.70	-2.42	-21.15
1931-32	-0.4693	0.0477	1.3016	0.8689	1.1292	2.8781	-7.46	-7.11	-18.77	-10.31	-25.00
1932-33	-0.3673	-0.5602	-1.7140	1.2614	-2.5005	-3.8805	3.79	3.92	6.27	-24.04	2.78
1933-34	1.1229	-1.7431	0.2239	-1.2477	-1.2220	-2.8661	3.77	0.17	-7.02	0.27	5.92
1934-35	0.3356	0.5809	-0.5448	0.2706	-0.1010	0.5412	-1.57	-2.62	-0.27	-6.47	-7.88
1935-36	0.4006	-0.9744	-1.2234	-0.2995	0.3104	-1.7864	0.79	-0.53	2.63	-5.67	-2.68
1936-37	0.0036	0.9027	0.5251	-0.4028	-1.9709	-0.9423	5.87	2.85	3.58	8.20	12.23
1937-38	0.4168	0.7496	0.0771	-1.2785	0.5289	0.4939	-1.07	0.03	1.71	18.37	4.98
1938-39	0.6533	1.2541	0.6148	2.0157	0.6358	5.1738	-4.11	1.60	3.98	-12.54	-2.52
1949-50	-0.5092	-1.1658	-5.3467	3.1960	7.0151	3.1893	2.72	6.56	8.56	11.35	25.70
1950-51	0.1383	-1.1777	1.9936	-0.4541	1.7098	2.2100	9.26	17.39	1.53	5.81	7.22

Dutch Consumption Data - Continued

	RHS -1*	RHS-2*	RHS-3*	RHS-4*	RHS-5*	dQ	P1	P2	P3	P4	P5
1951-52	0.1364	0.9855	0.2353	-0.0205	-0.7059	0.6309	11.05	1.16	2.65	-1.77	8.20
1952-53	-0.0664	0.3023	0.9068	0.4318	3.5144	5.0889	-0.27	-0.25	1.68	1.22	-6.76
1953-54	-0.2710	1.6126	1.5022	-0.1572	1.5925	4.2790	1.61	3.79	1.65	12.42	0.94
1954-55	-0.3590	1.6293	-0.6179	-0.2900	0.4108	0.7732	1.33	-3.32	6.74	0.61	-0.37
1955-56	-0.2353	2.5566	0.3922	-0.7595	1.1909	3.1449	1.03	-4.46	0.71	16.59	6.92
1956-57	-0.3836	-1.9758	0.5935	-0.3619	-0.1628	-2.2906	4.52	4.29	6.18	7.28	1.91
1957-58	0.1317	0.4209	0.1391	0.6807	0.9149	2.2873	-0.57	-0.01	-2.02	-5.97	-2.15
1958-59	0.2538	0.8093	-0.1398	0.8093	-0.0147	1.7179	-1.25	-0.73	2.84	1.55	3.63
1959-60	-0.2003	1.2512	0.2873	0.4196	1.4591	3.2168	3.40	-2.67	0.95	1.50	-5.12
1960-61	-0.1861	1.8689	0.9228	1.0702	0.4575	4.1334	3.66	-1.78	1.99	1.64	4.27
1961-62	0.0627	0.6323	0.5470	-0.5104	0.4947	1.2264	3.85	-0.20	-0.38	15.49	-0.41
1962-63	-0.0332	0.7450	0.1819	0.6208	0.0122	1.5266	6.29	1.88	6.56	-2.25	3.80

Systematic Variance



Effect on Least Squares

- Creating this weighting is a little more difficult because of the negative numbers in $d \ln(Q_t)$

$$\omega_i = \frac{d \ln(Q_t) + 6.000}{\sum_{s=1}^{31} (d \ln(Q_s) + 6.000)} 31 \quad (5)$$

- If we estimated this model with ordinary least squares, the variance of the parameters would be

$$\text{Asy. Var} \left[\hat{\beta} \right] = \frac{\sigma^2}{n} \left(\text{plim} \frac{1}{n} X'X \right)^{-1} \left(\text{plim} \frac{1}{n} X' \Omega X \right) \left(\text{plim} \frac{1}{n} X'X \right)$$

$$V \left[\hat{\beta} \middle| X \right] = (X'X)^{-1} \left(\sigma^2 \sum_{i=1}^n \omega_i x_i x_i' \right) (X'X)^{-1}$$

Effect on Least Squares – Continued

- Defining

$$Q_n^* = \frac{X' \Omega X}{n} = \frac{1}{n} \sum_{i=1}^n \omega_i x_i x_i' \quad (7)$$

- The distribution of $\hat{\beta}$ converges to

$$\hat{\beta} \overset{a}{\sim} N \left[\beta, \frac{\sigma^2}{n} Q^{-1} Q^* Q^{-1} \right], \ni: Q = \frac{X' X}{n} \quad (8)$$

Difference in Variance

- It is working out the difference between the variance under Homoscedasticity and Heteroscedasticity

$$\begin{aligned} \text{Est. Var} \left[\hat{\beta} \middle| X \right] - \text{Var} \left[\hat{\beta} \middle| X \right] &= s^2 (X'X)^{-1} - \sigma^2 (X'X)^{-1} (X'\Omega X) (X'X)^{-1} \\ D &= \frac{\sigma^2}{n} \left(\frac{X'X}{n} \right)^{-1} \left[\frac{X'X}{n} - \frac{X'\Omega X}{n} \right] \left(\frac{X'X}{n} \right)^{-1} \end{aligned} \quad (9)$$

Difference in Variance

- As developed by Greene (p.272) the difference hinges on

$$\begin{aligned}\Delta &= \frac{X'X}{n} - \frac{X'\Omega X}{n} \\ &= \sum_{i=1}^n \left(\frac{1}{n} \right) x_i x_i' - \sum_{i=1}^n \left(\frac{\omega_i}{n} \right) x_i x_i' \\ &= \frac{1}{n} \sum_{i=1}^n (1 - \omega_i) x_i x_i'\end{aligned}\tag{10}$$

If the heteroscedasticity is not correlated with the variables in the model, then at least in large samples, the ordinary least squares computations, although not the optimal way to use the data, will not be misleading.

White's General Test

- I have been applying a variation on the White's test - The hypothesis for the test is

$$\begin{aligned}H_0 : \sigma_i^2 &= \sigma^2 \text{ for all } i \\H_1 : &\text{Not } H_0.\end{aligned}\tag{11}$$

Estimate the residuals for the regression and then construct the estimated error - squared.

- "A Simple operation version of his test is carried out by obtaining nR^2 in the regression of e_i^2 on a constant and all unique variables contained in x and all squares and cross-products of the variables in x ."

White's General Test - Continued

- "The statistic is asymptotically distributed chi-squared with $P - 1$ degrees of freedom, where P is the number of regressors in the equation, including the constant."
- Basically, can the estimated variance be explained by the variables from the regression.

Breusch-Pagan/Godfrey LM Test

- The Breusch and Pagan test is a Lagrange Multiplier test of the hypothesis

$$\sigma_i^2 = \sigma^2 f(\alpha_0 + \alpha' z_i) \quad (12)$$

where z_i is a vector of independent variables.

- Let g be a vector of

$$g_i = \frac{\hat{\epsilon}_i^2}{\hat{\epsilon}'\hat{\epsilon}/n} - 1. \quad (13)$$

Breusch-Pagan/Godfrey LM Test – Continued

- The Breusch-Pagan test then becomes

$$LM = \frac{1}{2} \left[g'Z (Z'Z)^{-1} Z'g \right]. \quad (14)$$

Under the null hypothesis of homoscedasticity this statistic has a limiting chi-squared distribution with the degrees of freedom equal to the number of variables in z_i .

Feasible Generalized Least Squares

- The Feasible Generalized Least Squares estimator is then

$$\hat{\beta} = \left(X' \hat{\Omega}^{-1} X \right)^{-1} X' \hat{\Omega}^{-1} y \quad (15)$$

- The asymptotic variance then becomes

$$\text{Asy. Var} \left[\hat{\beta} \right] = \sigma^2 \left(X' \hat{\Omega}^{-1} X \right)^{-1} X' \hat{\Omega}^{-1} \Omega \hat{\Omega}^{-1} X \left(X' \hat{\Omega}^{-1} X \right)^{-1} \quad (16)$$

if the estimate of the variance is correct $\Rightarrow \hat{\Omega}^{-1} \Omega \rightarrow I$ and the variance matrix simply becomes

$$V \left[\hat{\beta} \right] = \sigma^2 \left(X' \hat{\Omega}^{-1} X \right)^{-1} \quad (17)$$

Production Economics Results

Parameter	Ordinary Least Squares		Generalized Least Squares	
	Estimate	Std. Dev.	Estimate	Std.Dev.
α_0	93.44560	14.01369	99.54622	309.83546
α_1	0.03240	0.13397	0.00220	2.73342
α_2	0.24683	0.27178	0.25376	3.62521
α_3	0.00252	0.22487	-0.09517	3.45816
A_{11}	0.00066	0.00078	0.00057	0.01422
A_{12}	-0.00194	0.00191	-0.00154	0.02866
A_{13}	0.00120	0.00140	0.00135	0.02181
A_{22}	0.00147	0.00194	0.00092	0.03603
A_{23}	-0.00030	0.00100	-0.00049	0.01738
A_{33}	-0.00066	0.00180	0.00021	0.03129

Readings – Next Time

- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition (Chapter 10 Section 10.1 10.5 pp. 290 313)
- * Theil, H. 1971. *Principles of Econometrics*. John Wiley & Sons. (Chapter 7).