DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 7 | Issue 2 Article 15

11-1-2008

Adaptive Estimation of Heteroscedastic Linear Regression Model Using Probability Weighted Moments

Faqir Muhammad
Allama Iqbal Open University, aioufsd@yahoo.com

Muhammad Aslam
Bahauddin Zakariya University, aslamasadi@bzu.edu.pk

G.R. Pasha
Bahauddin Zakariya University, drpasha@bzu.edu.pk

Journal of Modern Applied Statistical Methods
November, 2008, Vol. 7, No. 2, 501-505

1538 – 9472/08/$95.00

Adaptive Estimation of Heteroscedastic Linear Regression Model Using Probability Weighted Moments

Faqir Muhammad
Allama Iqbal Open University

Muhammad Aslam
Bahauddin Zakariya University

G.R. Pasha
Bahauddin Zakariya University

An adaptive estimator is presented by using probability weighted moments as weights rather than conventional estimates of variances for unknown heteroscedastic errors while estimating a heteroscedastic linear regression model. Empirical studies of the data generated by simulations for normal, uniform, and logistically distributed error terms support our proposed estimator to be quite efficient, especially for small samples.

Key words: Adaptive estimator, estimated weighted least squares, heteroscedasticity, probability weighted moments.

Introduction

The basic version of linear regression model assumes homoscedasticity of error terms. If this assumption is not met then the regression disturbances whose variances are not constant across observations are heteroscedastic. In the presence of heteroscedasticity, the method of ordinary least squares (OLS) does not result in biased and inconsistent parameter estimates. However, OLS estimates are no longer best linear unbiased estimators (BLUE). That is, among all the unbiased estimators, OLS does not provide the estimate with the smallest variance. In addition, the standard errors of the estimates become biased and inconsistent when heteroscedasticity is present. This, in turn, leads to bias in test statistics and confidence intervals. Depending on the nature of the heteroscedasticity, significance tests can be too high or too low. These effects are not ignorable as earlier noted by Geary (1966), White (1980) and Pasha (1982), among many others.

When the form of heteroscedasticity is known, using weights to correct for heteroscedasticity is very simple by weighted least squares (WLS). If the form of heteroscedasticity is not known, the standard method of replication is used as given by Fuller and Rao (1978). In this approach, the unknown variance of each residual can be estimated first and these estimates can be used as weights in a second step and the resultant estimates are referred to as estimated weighted least squares (EWLS) estimates.

Pasha (1984) gave a comparison among EWLS and minimum norm quadratic unbiased estimator (MINQUE) and reported EWLS to be better than MINQU-based estimators for estimation of heteroscedastic linear regression model. Pasha and Ord (1994) presented two adaptive estimators, one based on overall test of heteroscedasticity and other on paired comparison procedures following the idea of Bancroft (1964) and Bancroft & Hans (1977). These estimators were also based on EWLS and the attractive performances of these adaptive estimators were reported for efficiency gain.

An adaptive estimator is presented in this article by using probability weighted moments (PWM) as weights for transforming

Faqir Muhammad is Faculty Dean and Professor in the Department of Statistics and Mathematics, Allama Iqbal Open University, Islamabad, Pakistan. E-mail: aioufsd@yahoo.com. Muhammad Aslam is Assistant Professor in the Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan. E-mail: aslamasadi@bzu.edu.pk. G.R. Pasha is Professor in the Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan E-mail: drpasha@bzu.edu.pk.

matrix rather than conventional estimates of unknown error variances as usually used in EWLS. Downton (1966) suggested a linear estimate of the standard deviation of the normal distribution as

$$S_p = \frac{2\sqrt{\pi}}{n(n-1)} \sum_{i=1}^{n} [i - 0.5(n+1)] X_i .$$

Here X_i indicates ordered observations in a sample of size n. The estimate of the standard deviation using PWM is also a function of ordered observations as

$$S_{pw} = \frac{\sqrt{\pi}}{n} \sum_{i=1}^{n} [X_i - 2(1 - \frac{i-0.5}{n}) X_i].$$

The estimate of the mean is $\frac{\sum_{i=1}^{n} X_i}{n}$. The X_i's are the ordered observations and $(i - 0.5)/n$ is the empirical distribution function $F_n(X)$. Such estimator is also used by Muhammad et al. (1993). Greenwood (1979) explained the robustness of the PWM over the conventional moments to outliers by drawing more efficient inferences using PWM.

A heteroscedastic linear regression model and a usual EWLS estimator are given below. In addition, a new estimator based on probability weighted moments, denoted as PEWLS estimator, is presented. Finally, empirical results, an application for this approach and conclusions are put forth.

Methodology

Linear Regression Model with Heteroscedastic Errors and EWLS

Consider the following heteroscedastic linear regression model:

$$y_{ij} = x'_i \beta + u_{ij},\ i=1, 2, \ldots, k, j=1, 2, \ldots, n_i,\ \sum_{i}^{k} n_i = n \qquad (2.1)$$

where y_{ij} is the jth response at the ith design point x_i, x_i are known p-vectors, β is a p-vector of unknown parameters and u_{ij} are the mutually independent with $E(u_{ij}) = 0$ and $E(u_{ij}^2) = \sigma_i^2$, $j = 1, 2, \ldots, n_i$. The variances σ_i^2's are unknown and heteroscedastic. A matrix form of model (2.1) is

$$y = X\beta + u, \qquad (2.2)$$

where

$$y = (y_{11} \ . \ . \ . \ y_{1n_1} \ . \ . \ . \ y_{k1} \ . \ . \ . \ y_{kn_k})'_{n \times 1},$$

$$u = (u_{11} \ . \ . \ . \ u_{1n_1} \ . \ . \ . \ u_{k1} \ . \ . \ . \ u_{kn_k})'_{n \times 1},$$

and

$$X = (x_{11} \ . \ . \ . \ x_{1n_1} \ . \ . \ . \ x_{k1} \ . \ . \ . \ x_{kn_k})'_{n \times p},$$
$$x_{ij} = x_i, \quad j = 1, 2, \ldots, n_i,$$

with heteroscedastic error terms of covariance matrix Ω having typical ith diagonal elements σ_i^2.

The usual OLS estimator for β in (2.2) is

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'y$$

Fuller and Rao (1978) presented EWLS estimator of β as

$$\hat{\beta}_{EWLS} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}y), \qquad (2.3)$$

where

$$\hat{\Omega} = \text{diag}\{\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_n^2\},$$

$$\hat{\sigma}_i^2 = n_i^{-1} \sum_{j=1}^{ni} (y_{ij} - x'_i \hat{\beta}_{OLS})^2 .$$

PWM-based Adaptive Estimator (PEWLS)

Probability weighted moments are used as weights in transforming matrix $\hat{\Omega}$ in (2.3) and propose a new estimator as

$$\hat{\beta}_{PEWLS} = (X'\hat{\Phi}^{-1}X)^{-1}(X'\hat{\Phi}^{-1}y),$$

where

$$\hat{\Phi} = diag\{\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_n\},$$

$$\hat{\phi}_i = \frac{\sqrt{\pi}}{n}\sum_{i=1}^{n}[Y_i - 2(1 - \frac{i-0.5}{n})Y_i].$$

The estimate of the mean is $\frac{\sum_{i=1}^{n} Y_i}{n}$. The Y_i's are the ordered observations and $(i - 0.5)/n$ is the empirical distribution function $F_n(Y)$.

Results

A Monte Carlo study was performed on the model used by Jacquez, et al. (1968) among others in their numerical work.

$$y_{ij} = 1 + x_i + u_{ij}\ ; \quad i = 1, 2, 3, \ldots, k; \quad j = 1, 2, 3, \ldots, n_i \tag{4.1}$$

The u_{ij} are independently distributed with zero mean and variance σ_i^2. Different versions for the model (4.1) were used according to the following formations: n_i were set to be equal to m; m = 5, 10. k was chosen as k = 6, 8, 10. x_i were selected as ; for k = 6, x_i were (1, 2, 4, 7, 9, 10), for k = 8, x_i were (1, 2, 4, 5, 6, 7, 9, 10), and for k = 10, x_i were (1, 2, 3, 4, 5, 6, 7, 8, 9, 10). For each pair (m, k), two σ-pattern (data generating process: DGP) were chosen; DGP-I: $\sigma_i = (x_i + 8)/9$, and DGP-II: $\sigma_i = (0.5\, x_i + 1)/3$.

Different data sets are generated for each pair of (m, k) and σ-pattern for normal, uniform, and logistically distributed error terms. For each pair of (m, k) and σ-pattern, 2,000 simulations are run. On the basis of the generated data, in Table 4.1 and Table 4.2, the efficiency of the EWLS estimator relative to the PEWLS estimator for β, is compared as R.E = $MSE(\hat{\beta}_{PEWLS}) / MSE(\hat{\beta}_{EWLS})$.

The mean values of standard error of estimates of the regressions are compared by computing the ratios SE (PEWLS)/SE (EWLS). These ratios are shown in Table 4.3 and 4.4.

Table 4.1 shows the relative efficiencies under DGP-I. For normally distributed errors, PEWLS performs better than EWLS for all the pairs (m, k) in terms of efficiency. But for small samples (m = 5, k = 6), PEWLS is more efficient and the gain in efficiency reaches to 20% while comparing with that of EWLS. For m = 10, both estimators tend to become equal efficient as k increases from 6 to 10. For uniform and logistic errors, no substantial efficiency is observed while using PEWLS.

Table 4.2 (DGP-II) shows the same trend of efficiency as shown by Table 4.1 for all the tried error patterns. It is noted again that when m = 5 is fixed, the new proposed estimator shows more efficient behavior for small values of k, namely, for k = 6.

Table 4.3 and 4.4 show that the results of the adaptive estimator PEWLS are brightly encouraging with respect to the standard error of estimate for the fitted model even for all the selected pairs of (m, k) and the error patterns. For normal errors and small samples (m = 5), the results are quite impressive by using PWELS as compared to its competitor for all chosen k. The standard errors of estimates of the fitted model are about double for EWLS as compared to that of our proposed PEWLS (e.g., for k = 6, 8). Almost similar are the findings for the other tried error distributions so far. Same fashion of less standard error of estimates is observed for DGP-II in Table 4.4. These findings show that by using the proposed adaptive estimator, one can find better regression estimates as compared to that by using EWLS.

Application

To illustrate the computations of the proposed PEWLS estimators and to compare its performance with the EWLS, already available in the literature, take the example of compensation per employee ($) in Nondurable Manufacturing Industries of US Department of Commerce as quoted by Gujarati (2003, p. 392). This example is used to compare these findings in practical data with findings already available in the literature.

Table 5.1 reports the performance of OLS, EWLS and the proposed PEWLS estimators. First, OLS estimates are found and the presence of heteroscedasticity is noted by

Table 4.1: Relative Efficiency of PEWLS and EWLS Estimators of β (DGP-I)

k	Normal		Uniform		Logistic	
	$m = 5$	$m = 10$	$m = 5$	$m = 10$	$m = 5$	$m = 10$
6	0.8088	0.9779	0.9887	0.9344	0.9885	1.0000
8	0.8526	0.9839	0.9921	0.9617	0.9891	1.0051
10	0.9400	0.9899	0.9989	0.9625	0.9911	0.9656

Table 4.2: Relative Efficiency of PEWLS and EWLS Estimators of β (DGP-II)

k	Normal		Uniform		Logistic	
	$m = 5$	$m = 10$	$m = 5$	$m = 10$	$m = 5$	$m = 10$
6	0.8918	0.9915	1.0000	0.9268	0.9831	0.9943
8	0.9112	0.9652	0.9471	0.9915	0.9962	0.9952
10	1.0031	0.9705	1.0252	0.9966	0.9986	1.0000

Table 4.3: Ratios of Standard Error of Estimates of PEWLS and EWLS (DGP-I)

k	Normal		Uniform		Logistic	
	$m = 5$	$m = 10$	$m = 5$	$m = 10$	$m = 5$	$m = 10$
6	0.5721	0.9244	0.6592	0.9068	0.8470	0.9650
8	0.5944	0.9287	0.6720	0.9325	0.8169	0.9661
10	0.6515	0.9365	0.6263	0.9317	0.7474	0.9317

Table 4.3: Ratios of Standard Error of Estimates of PEWLS and EWLS (DGP-II)

k	Normal		Uniform		Logistic	
	$m = 5$	$m = 10$	$m = 5$	$m = 10$	$m = 5$	$m = 10$
6	0.6770	0.9477	0.7011	0.9616	0.6329	0.9899
8	0.6839	0.9234	0.6531	0.9577	0.7378	0.9965
10	0.6833	0.9307	0.7139	0.9603	0.6859	1.0011

using White's test (1980) with p-value 0.07. It is noted that the proposed estimator bear lower standard errors among all the remaining estimators presenting an adequate reliability for its adaptation. It is further noted that the proposed estimator give better R^2 and much improved standard errors of regression that confirms the adequacy of the fitted model. Similarly, the proposed adaptive estimator gives lowest Akaike Information Criteria (AIC) values that indicate the right specification of the weighting mechanism.

Conclusion

It was found that use of probability weighted moments as estimates of unknown heteroscedastic weights rather than conventional estimates of variances for unknown heteroscedastic errors while estimating a heteroscedastic linear regression model, makes more efficient estimations. This new formulation, considerably, contributes in reducing standard errors of estimates for fitted models. The gain in efficiency and the reduction

Table 5.1: Comparative Statistics

Estimators	Estimation of β_0			Estimation of β_1			R^2	S.E. of Regression	AIC
	$\hat{\beta}_0$	SE	*t*-statistic	$\hat{\beta}_1$	SE	*t*-statistic			
OLS	3417.70	81.04	42.17	148.81	14.40	10.33	0.9385	111.56	12.46
EWLS	3406.20	80.86	42.13	154.24	16.93	9.11	0.9645	126.54	12.71
PEWLS	3437.40	79.39	43.29	142.99	17.69	10.44	0.9842	103.87	12.31

in standard errors of estimates of regression model are appealing, especially, for small samples and thus make our new adaptation more attractive for many of practical situations of small samples.

References

Downton, F. (1966). Linear estimates with polynomial coefficients. *Biometrika 53*, 129.

Geary, R. C. (1966). A note on residual heretovariance and estimation efficiency in regression. *American Statistician, 20*, 30-31.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research, 15*, 1049-1054.

Gujarati, D. N. (2003). *Basic econometrics.* (4th ed.). NY: McGraw-Hill.

Horn, S. D., Horn, R. A., and Duncan, D. B. (1975). Estimating the heteroscedastic variances in linear models. *Journal of the American Statistical Association, 70*, 380-385.

Pasha, G. R. (1984). A comparative empirical study of WLS and MINQU-based estimators for small sample. *Karachi Journal of Mathematics,* 2, 31-42.

White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and direct test for heteroscedasticity. *Econometrica, 48*, 817-838.

Jacquez, J. A., Mathur, F. J., & Crawford, C. R. (1968). Linear regression with non-constant unknown error variances. *Biometrics, 24*, 607-627.

Muhammad, F., Ahmad, S., & Abdullah, M. (1993). Use of probability weighted moments in the analysis of means. *Biometrical Journal, 35*, 371-378.

Pasha, G. R. (1982). *Estimation methods for regression models with unequal error variances.* Unpublished Ph. D. dissertation, University of Warwick.

Pasha, G. R., & Ord, J. K. (1994). Adaptive estimators for heteroscedastic linear regression models, *Pakistan Journal of Statistics, 10*, 47-54.