DECISION SCIENCES INSTITUTE

Testing the Homoscedasticity Assumption in Linear Regression
in a Business Statistics Course

ABSTRACT

The ability of inexperienced introductory-level undergraduate or graduate business students to properly assess residual plots when studying simple linear regression is in question and the recommendation is to back up exploratory, graphical residual analyses with confirmatory tests. The results from a Monte Carlo study of the empirical powers of six procedures that can be used in the classroom for assessing the homoscedasticity assumption offers guidance for test selection.

KEYWORDS: Regression, Homoscedasticity, Residual analysis, Empirical vs. Practical Power

INTRODUCTION

The subject of regression analysis is fundamental to any introductory business statistics course. And for good reason:

> "Any large organization that is not exploiting both regression and randomization is presumptively missing value. Especially in mature industries, where profit margins narrow, firms 'competing on analytics' will increasingly be driven to use both tools to stay ahead. … Randomization and regression are the twin pillars of Super Crunching." Ian Ayres, *Super Crunchers* (2007)

Introductory business statistics textbooks use graphical residual analysis approaches for assessing the assumptions of linearity, independence, normality, and homoscedasticity when demonstrating how to evaluate the aptness of a fitted simple linear regression model. However, research indicates a deficiency in graphic literacy among inexperienced users. In particular, textbook authors and instructors should not assume that introductory-level business students have sufficient ability to appropriately assess the very important homoscedasticity assumption simply through such an "exploratory" graphical approach.

This paper compares six "confirmatory" procedures that can be used in an introductory business statistics classroom for testing the homoscedasticity assumption when teaching simple linear regression. A Monte Carlo simulation rank orders these procedures with respect to their power to detect a violation in the homoscedasticity assumption and then Tukey's (1959) concept of "practical power" is used for making the actual test selection. More generally, however, it is recommended that a graphical residual analysis approach be coupled with a confirmatory approach for assessing all four simple linear regression modeling assumptions.

ASSESSING THE HOMOSCEDASTICITY ASSUMPTION IN LINEAR REGRESSION

Introductory-level students learning simple linear regression analysis must be made aware of the consequences of a fitted model not meeting its assumptions. In particular, if the assumption of homoscedasticity is violated the ordinary least squares estimator of the slope is no longer efficient, estimates of the variances and standard errors are biased, and inferential methods are no longer valid.

In a seminal paper Anscombe (1973) demonstrated the importance of graphical presentations to enhance understanding of what a data set is conveying and to assist in the model-building process for a regression analysis. Cleveland et al. (1988) and Cook and Weisberg (1994) then showed why the selected aspect ratio of a residual plot is essential to its understanding. They and others have made it clear that inexperienced individuals have difficulty in deciphering the underlying message the residual plots are conveying. Therefore, supplementing a graphical residual analysis of a model's assumptions with appropriate confirmatory approaches would surely enhance model development.

SIX USEFUL TESTS FOR AN INTRODUCTORY BUSINESS STATISTICS COURSE

Six confirmatory approaches to supplement a traditional graphical residual analysis when assessing the homoscedasticity assumption in simple linear regression analysis are presented below for introductory-level classroom use. (Complete descriptions of these six test procedures will be provided to attendees at the DSI session or will be available upon request).

NWRSR – The Neter-Wasserman / Ramsey / Spearman Rho *t* Test

Neter and Wasserman (1974) suggested a procedure for assessing homoscedasticity using a *t*-test of Spearman's rank coefficient of correlation ρ_S based on a secondary analysis between the absolute value of the residuals from an initial linear regression analysis and the initial predictor variable. Their NWRSR procedure yields identical results to Ramsey's RASET or Rank Specification Error Test (1969) which employs the squared residuals in lieu of the absolute residuals proposed by Neter and Wasserman (1974).

NWGQ – The Neter-Wasserman / Goldfeld-Quandt Test

Neter and Wasserman (1974) proposed an extension of a procedure developed by Goldfeld and Quandt (1965). To develop this test the bivariate data set of *n* observations is first divided in half (or as close to half as possible, considering the value of *n* and possibility of ties) based on the ordered values of the independent variable X_{1i}. The NWGQ test requires a simple linear regression analysis on the set of n_I observations with the *smaller* values of X_{1i} and a separate simple linear regression analysis based on the set of n_{II} observations with the *larger* values of X_{1i}. The test statistic F_{NWGQ} is the ratio of the mean square error for the regression model based on the set of n_{II} observations with the *larger* values of X_{1i} to the mean square error for the regression model based on the set of n_I observations with the *smaller* values of X_{1i}.

BF – The Brown-Forsythe Test

Brown and Forsythe (1974) developed a procedure for assessing homoscedasticity using a pooled-variance *t*-test based on a secondary analysis of the residuals obtained from an initial linear regression analysis. To develop the BF test, the data set of *n* observations is divided in half (or as close to half as possible, considering the value of *n* and possibility of ties) based on the ordered values of the independent variable X_{1i}. Group *I* contains the set of n_I residual observations based on the *smaller* values of X_{1i} and Group *II* contains the set of n_{II} residual observations based on the *larger* values of X_{1i} so that $n = n_I + n_{II}$. Determine M_{e_I} and $M_{e_{II}}$, the median residual for Groups *I* and *II*, and then obtain the sets of absolute residual differences $a_{I_i} = \left| e_{I_i} - M_{e_I} \right|$ and $a_{II_i} = \left| e_{II_i} - M_{e_{II}} \right|$ for the two groups. The Brown-Forsythe test statistic t_{BF} is obtained from a pooled-variance *t* test comparing the means of the two sets of absolute residual differences, $\bar{a}_{II} - \bar{a}_I$.

W – The White Test

White's test (White, 1980; Berenson, 2013) is based on a secondary regression analysis with the squared residuals from the initial linear regression analysis used as the dependent variable. The independent variables in the secondary regression analysis consist of the initial predictor variable and its square. The White test statistic W is nr^2, the product of the sample size *n* and r^2, the "unadjusted" coefficient of multiple determination follows a χ^2 distribution.

BPCW – The Breusch-Pagan / Cook-Weisberg Scores Test

The Cook and Weisberg scores test (1983, 1994) is a generalized procedure that reduces to the Breusch and Pagan test (Breusch & Pagan, 1979; Klibanoff et al., 2006). This BPCW test is based on a secondary regression analysis with the squared residuals from the initial linear regression analysis used as the dependent variable. The independent variable in the secondary regression analysis consist of the initial predictor variable. The Breusch-Pagan / Cook-Weisberg test statistic $\chi_{BPCW}{}^2$, computed as the ratio of the sum of squares due to regression in the secondary analysis to twice the square of the biased estimate of the mean square error in the initial linear regression analysis, follows a χ^2 distribution.

GMS – The Glejser / Mendenhall-Sincich Test

Mendenhall and Sincich (2003) adopt a procedure proposed by Glejser (1969) that is based on a secondary regression analysis with the absolute value of the residuals from the initial linear regression analysis used as the dependent variable. The independent variable in the secondary regression analysis is the initial predictor variable. The GMS procedure enables a *t*-test of the population slope from the secondary analysis.

RESULTS OF A MONTE CARLO SIMULATION STUDY

Table 1 presents the set of 24 condition-combinations used for the Monte Carlo power study. The imposition of particular condition-combinations was intended to determine:

1. Whether the six tests are "valid" and hold their specified $\alpha = 0.05$ level of significance under the null hypothesis of homoscedasticity.

2. Whether any test can be declared empirically most powerful in detecting heteroscedasticity when the standard deviation of the random error around the regression line is moderate to large.

Table 1: Controlled Conditions Used in the Monte Carlo Power Study

CONDITION	LEVELS	DESCRIPTION
Distribution	1	Normal
Std. Dev. of Random Error	2	$Y = 1 + 2X + ce$, c=0.5, 1.0
Heteroscedasticity Parameter:		
Case I	2	$Y = 1 + 2X + cXe$, c=0.5, 1.0
Case II	2	$Y = 1 + 2X + (X^c)e$, c=0.75,1.25
Sample Size	4	n = 20, 50, 100, 200
Level of Significance	1	α = 0.05

A total of 5,000 sets of bivariate data were generated for each of the above 24 condition-combinations. Predictor variable *X* was generated from the uniform distribution from 5 to 10. Random error *e* was generated from the standardized normal distribution N(0,1). The constant *c* (*c* = 0.5 or *c* = 1) was used to control the standard deviation of the random error term in the model. In the homoscedasticity case, $Y = 1 + 2X + ce$ (where the random error *ce* has constant variance); in the heteroscedasticity Case I, $Y = 1 + 2X + cXe$ (where the error term *cXe* is correlated with the predictor *X*), or in the heteroscedasticity Case II, $Y = 1 + 2X + (X^c)e$ (where the error term $(X^c)e$ is correlated with the predictor *X*). The sample sizes considered were 20, 50, 100, and 200, which covered a range from “small” sample to “medium” sample to “relatively large” sample. Statistical software R (2008) was used for all the simulation studies.

Under the null hypothesis of homoscedasticity the empirical α levels were obtained for each of the six test procedures using the condition combinations *c* = 0.5 or 1.0 and *n* = 20, 50, 100, or 200. Since the standard error associated with a stated 0.05 α level of significance is 0.00308 (based on 5,000 repetitions), it was found that the NWGQ, W, and GMS tests are “valid” for the conditions considered (Gibbons, 1976) since their empirical α levels or proportion of incorrect rejections never exceeded three standard errors of the nominal α level. On the other hand, the NWRSR test was liberal when *c* = 0.5 and *n* = 20 and valid otherwise while the BPCW test was conservative for very small sample size (*n* = 20) and valid otherwise. The BF test, however, was conservative for both smaller sample sizes (*n* = 20 and *n* = 50) when *c* = 1.0 and for the very small sample size (*n* = 20) when *c* = 0.5 and, therefore, cannot be recommended for use under such conditions.

With respect to empirical power, the Cochran Q test (1950) showed that there were real differences among the six tests at the α = 0.05 level of significance for each of the sixteen condition-combinations *c* = 0.5 or 1.0 (in Case I), c=0.75 or 1.25 (in Case II) and *n* = 20, 50, 100, or 200. Using the Benjamini-Hochberg (1995) multiple comparisons approach for each of these condition-combinations, these six test procedures were ranked based on their empirical power under heteroscedasticity Case I and Case II.

In summation, with respect to empirical power the BPCW test is best with larger data sets (i.e., $n \geq 100$) while the NWGQ test is clearly best for smaller sized samples (*n* < 100) and second best with larger data sets. The GMS test is always a good test and superior to the W test, the NWRSR test, and the BF test, three procedures that cannot be recommended based on empirical power.

CONCLUSIONS AND RECOMMENDATIONS

When evaluating the worth of a statistical procedure, Tukey (1959) defined *practical power* as the product of statistical power and the utility of the statistical technique (i.e., the ease in which the test can be learned and likelihood it would be used). Given the potential limitations for each of the six tests of homoscedasticity that were described in Section 3, it appears that the GMS test would have most practical power when teaching an introductory-level undergraduate business statistics course while either the NWGQ test or BPCW test can be recommended for teaching an introductory-level MBA course – the NWGQ test displays more statistical power with smaller sized samples and the BPCW test performers better with larger data sets.

REFERENCES

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, *27*, 17-21.

Ayres, I. (2007). *Super Crunchers*. New York: Bantam.

Benjamini, Y., & Hochberg,Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, 57*, 289-300.

Berenson, M. L. (2013). Using Excel for White's test – an important technique for evaluating the equality of variance assumption and model specification in a regression analysis. *Decision Sciences Journal of Innovative Education, 11,* 243-262.

Breusch, T., & Pagan, A. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica, 47,* 1287-1294.

Brown, M., & Forsythe, A. (1974). Robust tests for equality of variances. *Journal of the American Statistical Association, 69*, 364-367.

Cleveland, W. S., McGill, M. E., & McGill, R. (1988). The shape parameter of a two-variable graph. *Journal of the American Statistical Association, 83,* 289-300.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika, 37,* 256-266.

Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika, 70,* 1-10.

Cook, R. D., & Weisberg, S. (1994). *An introduction to regression graphics*. New York: Wiley.

Gibbons, J. D. (1976). *Nonparametric methods for quantitative analysis*. New York: Holt Rinehart and Winston.

Glejser, H. (1969). A new test for heteroscedasticity. *Journal of the American Statistical Association, 64*, 315-323.

Goldfeld, S., & Quandt, R. (1965). Some tests for homoscedasticity. *Journal of the American Statistical Association, 60,* 539-547.

Klibanoff, P., Sandroni, A., Moselle, B., & Saraniti, B. (2006). *Managerial statistics: A case-based approach*. Marion, OH: Thomson/South-Western (Cengage Learning).

Mendenhall, W., & Sincich, T. (2003). *A second course in statistics: Regression analysis,* 6th ed. Upper Saddle River, NJ: Prentice Hall (Pearson Higher Education).

Neter, J., & Wasserman, W. (1974). *Applied linear statistical models*. Homewood, IL: Richard D. Irwin.

Ramsey, J. B. (1969). “Tests for specification error in the general linear model,” *Journal of the Royal Statistical Society, Series B, 31,* 350-371.

R Development Core Team. (2008). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0, URL http://www.R-project.org.

Tukey, J. W. (1959). A quick, compact two-sample test to Duckworth’s specifications. *Technometrics*, *1,* 31-48.

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