Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

Testing for heteroscedasticity in regression models

Maria Carapeto [a] & William Holt [b]
[a] Faculty of Finance , Cass Business School , London, UK
[b] Department of Decision Sciences , London Business School , London, UK
Published online: 02 Aug 2010.

Journal of Applied Statistics, Vol. 30, No. 1, 2003, 13–20

Testing for heteroscedasticity in regression models

MARIA CARAPETO[1] **& WILLIAM HOLT**[2], [1]*Faculty of Finance, Cass Business School and* [2]*Department of Decision Sciences, London Business School*

ABSTRACT *A new test for heteroscedasticity in regression models is presented based on the Goldfeld–Quandt methodology. Its appeal derives from the fact that no further regressions are required, enabling widespread use across all types of regression models. The distribution of the test is computed using the Imhof method and its power is assessed by performing a Monte Carlo simulation. We compare our results with those of Griffiths & Surekha (1986) and show that our test is more powerful than the wide range of tests they examined. We introduce an estimation procedure using a neural network to correct the heteroscedastic disturbances.*

1 Introduction

A central assumption of regression theory is that the error term is homoscedastic, i.e. $E[\varepsilon_i^2] = \sigma^2, \forall i$, the variance of each observation is a constant value σ^2. There have been numerous tests developed over the last 30 years to detect heteroscedasticity, including Goldfeld & Quandt (1965), Park (1966), Glejser (1969), Ramsey (1969), Szroeter (1978), Breusch & Pagan (1979) and White (1980). While these tests are commonly used in residuals testing they can be unreliable, difficult to implement or fail to capture complex patterns of heteroscedasticity (see, for example, Goldfeld & Quandt, 1972; and Harrison, 1980).

In this paper we develop a new test based on the Goldfeld–Quandt statistic that may be applied to parametric and non-parametric, linear and nonlinear regression models. Goldfeld & Quandt's test is adapted so that only one regression is required. This is especially important in non-parametric models (e.g. neural networks) where inefficiencies in the estimation procedure would make it difficult to compare estimated residuals from significantly different regression models. (If the test was

Correspondence: Maria Carapeto, Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, UK. E-mail: mcarapeto@city.ac.uk

ISSN 0266-4763 print; 1360-0532 online/03/010013-08
DOI: 10.1080/0266476022000018475

performed *à la* Goldfeld–Quandt, the fact that we cannot guarantee convergence to the global minimum would potentially cause the estimated functional forms to be very different.) We compare the power of our test with other commonly used heteroscedasticity tests and show that it is extremely powerful at detecting both additive and multiplicative heteroscedasticity. Once the presence of heteroscedasticity has been established, we propose a method for detecting its form based on a neural network regression, enabling us to estimate the model within a GLS framework.

2 A new test for heteroscedasticity

Consider the model

$$\mathbf{y} = f(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{y}$ is the $n \times 1$ vector of observations on the dependent variable, $\mathbf{X}$ is the $n \times k$ matrix of independent variables, $\boldsymbol{\varepsilon}$ is the $n \times 1$ vector of errors and f is a known linear or nonlinear data generating function. Following Goldfeld–Quandt, we rank the observations in decreasing order of the variable X_j thought to be responsible for the heteroscedasticity. Two subsamples (1, 2) with dimension m of the first r% of the n observations and of the last r% of the n observations are then taken ($m = r\%n$). Using Harvey & Phillips (1974), r should be no less than 30; however for the sake of homogeneity in each subsample, r should be not more than 40. Of course, the test can still be applied to subsamples of different dimensions.

We wish to test the hypothesis $\mathrm{H}_0 : \sigma_1^2 \leqslant \sigma_2^2$ against an alternative $\mathrm{H}_\mathrm{A} : \sigma_1^2 > \sigma_2^2$. The test statistic is defined by

$$q = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{\hat{\boldsymbol{\varepsilon}}'\mathbf{I}^{\star}\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\mathbf{I}_{\star}\hat{\boldsymbol{\varepsilon}}} = \frac{\boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{I}^{\star}\mathbf{M}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{I}_{\star}\mathbf{M}\boldsymbol{\varepsilon}} \tag{2}$$

for models with a constant intercept. For models without a constant intercept the statistic becomes

$$q = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{\hat{\boldsymbol{\varepsilon}}'\mathbf{A}'\mathbf{I}^{\star}\mathbf{A}\hat{\boldsymbol{\varepsilon}}}{\hat{\boldsymbol{\varepsilon}}'\mathbf{A}'\mathbf{I}_{\star}\mathbf{A}\hat{\boldsymbol{\varepsilon}}} = \frac{\boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{A}'\mathbf{I}^{\star}\mathbf{A}\mathbf{M}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{A}'\mathbf{I}_{\star}\mathbf{A}\mathbf{M}\boldsymbol{\varepsilon}}$$

where $\mathbf{A} = \mathbf{I}_{(n \times n)} - (1/n)\mathbf{S}\mathbf{S}'$ and $\mathbf{S}$ is an $n \times 1$ vector of ones. The $\mathbf{I}^{\star}$ and $\mathbf{I}_{\star}$ are similar to the $n \times n$ identity matrix $\mathbf{I}$, except that some of the diagonal elements are zero. They enable the selection of the particular estimated residuals we are interested in. In the case of $\mathbf{I}^{\star}$, the 1s in the diagonal are associated with the largest m elements of X_j; in $\mathbf{I}_{\star}$, they stand for the smallest m values of X_j.

We use the relation $\hat{\boldsymbol{\varepsilon}} = \mathbf{M}\boldsymbol{\varepsilon}$ with the projection matrix $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ for linear models. For parametric nonlinear models, one should use instead $\mathbf{M} = \mathbf{I} - \mathbf{U}_{y\hat{\beta}}(\mathbf{U}_{\hat{\beta}\hat{\beta}})^{-1}\mathbf{U}'_{y\hat{\beta}}$ and $U(\mathbf{X}, \hat{\boldsymbol{\beta}}, \mathbf{y}) = \frac{1}{2}\Sigma_{i=1}^{n}(y_i - f(\mathbf{X}'_{[i]}, \hat{\boldsymbol{\beta}}))^2$.

For neural network models the projection matrix is given by $\mathbf{M} = \mathbf{I} - \mathbf{T}_{y\hat{\beta}}(\mathbf{U}_{\hat{\beta}\hat{\beta}})^{-1}\mathbf{T}'_{y\hat{\beta}}$ and

$$U(\mathbf{X}, \hat{\boldsymbol{\beta}}, \mathbf{y}; \lambda) = \underbrace{\frac{1}{2}\sum_{i=1}^{n}(y_i - f(\mathbf{X}'_{[i]}, \hat{\boldsymbol{\beta}}))^2}_{T(\mathbf{X}, \hat{\boldsymbol{\beta}}, \mathbf{y})} + \lambda R(\hat{\boldsymbol{\beta}})$$

is the least squares error function, where λ is a regularization parameter and R is the regularization function. (For a derivation of the projection matrix in all types of regression models, see Carapeto *et al.*, 2003.)

Given a particular result q_0 for the statistic we can calculate $P[q > q_0]$ and so obtain a p-value for the test. The p-value can be computed as follows

$$P\left[\frac{\boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{I}^{\star}\mathbf{M}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{I}_{\star}\mathbf{M}\boldsymbol{\varepsilon}} > q_0\right] = P[\boldsymbol{\varepsilon}'\mathbf{M}'(\mathbf{I}^{\star} - q_0\mathbf{I}_{\star})\mathbf{M}\boldsymbol{\varepsilon} > 0] \tag{3}$$

where $\mathbf{M}'\mathbf{I}^{\star}\mathbf{M}$ and $\mathbf{M}'\mathbf{I}_{\star}\mathbf{M}$ are symmetric positive semi-definite quadratic forms. (Note that we can only apply this test in the context of nonlinear models when the second derivatives of the least squares error function with respect to the parameters and the dependent variable are continuous.) The distribution of the quadratic form may be calculated using the Imhof (1961) algorithm, which is a numerical evaluation of the Fourier inversion integral (Durbin & Watson, 1971), under the assumption that the residuals are normally distributed with zero mean. As with all numerical integration methods, the parameters of step size and truncation of the range of integration must be chosen with care. For a discussion of these factors see Press *et al.* (1992). We represent the required p-value as

$$P(Q > 0) = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \frac{\sin\{\theta(t)\}}{t\rho(t)}\,\mathrm{d}t \tag{4}$$

where $Q = \boldsymbol{\varepsilon}'\mathbf{M}'(\mathbf{I}^{\star} - q_0\mathbf{I}_{\star})\mathbf{M}\boldsymbol{\varepsilon}$, $\theta(t) = \frac{1}{2}\Sigma_{j=1}^{n}\tan^{-1}(\lambda_j t)$ and $\rho(t) = \Pi_{j=1}^{n}(1 + \lambda_j^2 t^2)^{1/4}$ with λ_j the eigenvalues of $\mathbf{M}'(\mathbf{I}^{\star} - q_0\mathbf{I}_{\star})\mathbf{M}$.

3 A Monte Carlo experiment

In order to assess the power of the statistic we perform a Monte Carlo study as described in Griffiths & Surekha (1986). We may then compare the power of our statistic with their analysis for the Szroeter (SZ), Goldfeld–Quandt (G-Q), additive Breusch–Pagan (B-P(A)), multiplicative Breusch–Pagan (B-P(M)) and BAMSET tests of heteroscedasticity.

We consider a sample size $n = 50$. The linear model examined is

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{5}$$

with the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1)' = (10, 1)'$. The independent variable x_i was generated according to two distributions: a continuous uniform distribution in the range $[20, 100]$ and a lognormal distribution, where $x_i = \mathrm{e}^{q_i}$ and q_i is randomly generated from $N(3.8, 0.4^2)$. The true disturbances were generated from an $N(0, V)$ distribution. The variance V was given by either an 'additive model', $V(\varepsilon_i) = \sigma_i^2 = \alpha_0^2(1 + \lambda x_i)^2$ or a 'multiplicative model', $V(\varepsilon_i) = \sigma_i^2 = kx_i^{\gamma}$. The parameters γ and λ controlled the severity of the heteroscedasticity. We considered three values of λ, 0.03, 0.04 and 500, and three values of γ, 1.25, 1.35 and 2. In total, we analysed the same eight cases of Griffiths & Surekha with the number of central observations $c = 20\% \times n$. Five thousand replications were generated for both the lognormal and uniform cases. The power of the test was assessed by calculating the proportion of correct rejections of the null hypothesis of no heteroscedasticity at the 5% significance level, i.e. $1 - P$[Type-II error].

TABLE 1. Powers of heteroscedasticity test for selected models and parameter values

		Tests						
Model[a]	Parameter	SZ	G-Q	B-P(A)	B-P(M)	BAMSET	C-H	C-H[b]
AL	$\lambda = 0.04$	0.534	0.464	0.416	0.400	0.275	0.624	0.636
ML	$\gamma = 1.35$	0.557	0.489	0.429	0.421	0.294	0.651	0.658
AU	$\lambda = 0.03$	0.597	0.541	0.470	0.444	0.345	0.695	0.693
MU	$\gamma = 1.25$	0.591	0.536	0.461	0.446	0.343	0.684	0.692
AL	$\lambda = 500$	0.819	0.764	0.703	0.718	0.568	0.901	0.903
ML	$\gamma = 2$	0.819	0.764	0.703	0.718	0.568	0.904	0.910
AU	$\lambda = 500$	0.887	0.847	0.800	0.812	0.692	0.945	0.940
MU	$\gamma = 2$	0.887	0.847	0.800	0.812	0.692	0.951	0.938

Source: Griffiths & Surekha (1986) + authors' calculations.
[a]The code for model type is: M = Multiplicative, A = Additive, U = Uniform and L = Lognormal.
[b]Generating different samples for each model.

The analysis was performed as follows:

Step 1. Generate the vector of independent observations **X** from either a uniform or lognormal distribution.
Step 2. Generate the vector of true disturbances ε from a normal distribution with zero mean and variance given by either the additive or multiplicative models.
Step 3. Calculate **y** from the linear regression model, using **X**, $\boldsymbol{\beta}$ and ε.
Step 4. Estimate the model residuals $\hat{\varepsilon} = \mathbf{My}$.
Step 5. Obtain the ranking for the observations of (the absolute value of) **X**.
Step 6. Generate the $\mathbf{I}^\star$ and $\mathbf{I}_\star$ matrices. For $\mathbf{I}^\star$, $\mathbf{I}(i, i) = 1$, if x_i is one of the highest 40% of observations ($r = 40\%n$). For $\mathbf{I}_\star$, $\mathbf{I}(i, i) = 1$, if x_i is one of the lowest 40% of observations.
Step 7. Calculate the statistic q_0.
Step 8. Compute the p-value using the Imhof algorithm.
Step 9. Reject the null hypothesis of no heteroscedasticity if p-value < 0.05.
Step 10. Repeat this procedure 5000 times.

4 Analysis of Monte Carlo simulation

Table 1 gives the power of the major tests of heteroscedasticity and the new C-H test for a selection of models with weak and strong heteroscedasticity, controlled by the parameters λ and γ. In the Monte Carlo simulation for our test, we used a different sample of 5000 observations from Griffiths & Surekha (1986). However, the magnitude of the differences in the power of our test compared with the others (usually 10% higher) clearly indicates that our test performs better in detecting heteroscedasticity of both additive and multiplicative forms. An additional Monte Carlo simulation was performed (C-H[b]) where the sample of observations of the dependent variable was not fixed for either the lognormal or uniform models. The results follow the same pattern as the previous simulation, consistently outperforming the other statistics, and we may therefore infer that the power of the test is not dependent on the particular sample chosen.

5 Computing the test statistic

Computing the value of the statistic and its p-value is straightforward. The code in Appendix 1 implements the statistic in the mathematical computer language

MATLAB for a real example. The program can be thought of a high-level pseudocode that is easily readable and can be translated into other programming languages, such as Visual Basic. Although MATLAB does not require vector and matrix size allocation, we have included these lines for clarity.

To demonstrate the usefulness of the statistic we obtain UK economic household consumption expenditure data from Economic Trends Annual Supplement 2000 over the period 1984Q1 to 2000Q1. We regress the first differences of household final consumption of durables on the first differences of total household final consumption and three quarterly dummy variables for the second, third and fourth quarters. The inclusion of these dummy variables is required as the data are not seasonally adjusted. We suspect that, as the order of magnitude of the figures increased so strongly over the period, the variance of the error term will also increase.

Applying our test to this dataset with $r = 37.5\%$ produces a statistic of 2.25 with a p-value of 0.03, which corroborates our suspicion of heteroscedasticity being present in this case. However, using alternative reputable tests, such as Szroeter and Goldfeld–Quandt, we cannot reject the null hypothesis of homoscedasticity, implying that we could be making a Type-II error in our diagnosis. The Szroeter statistic is 0.44 with a p-value of 0.33 and the Goldfeld–Quandt statistic is 1.57 with a p-value of 0.17. Note that for the Goldfeld–Quandt statistic an alternative version of the test was used, as sorting the absolute value of the observations for the first regressor would have made it impossible to estimate the required regressions. This is because at least one of the dummy variables would have been zero in each subsample of 24 observations. We have therefore defined our subsamples as the first and last 24 observations.

6 Detecting the form of heteroscedasticity and model estimation procedure

Following Section 4, once it has been established that heteroscedasticity is present in the residuals, it is important to find the variable(s) that are responsible, the functional form that relates them to the variance of the residuals, and to correct the problem. We propose a multivariate procedure that may tackle any form of heteroscedastic relationship. This considers the most general class of function $g(.)$ that relates the variance of the errors to a group of variables $\mathbf{Z}$, which also include the regressors $\mathbf{X}$ and some external variables $\mathbf{W}$. Specifically, we can write

$$E[\varepsilon_i^2] = \sigma^2 g(\mathbf{Z}'_{[i]}), \quad \mathbf{Z}'_{[i]} = [\mathbf{X}'_{[i]}\mathbf{W}'_{[i]}] \tag{6}$$

The most natural way to estimate this function is through a neural network procedure which, being a nonlinear non-parametric class of regression models, is able to detect even a very complex pattern of heteroscedasticity. However, neural networks must be used with care in the context of diagnostic checking as there is a danger of model misspecification. Klenin (1996) demonstrated that a small network may be successfully used to check the specification of another neural model. We therefore suggest using the smallest possible network (i.e. few nonlinear nodes) to estimate the heteroscedasticity relationship and applying a rigorous model selection procedure, such as Moody & Utans (1992). Once the function has been estimated the estimation procedure is trivial using GLS.

7 Conclusions

We have introduced a methodology based on the Goldfeld–Quandt test to detect the presence of heteroscedasticity. As only one regression is required, it may be widely applied across regression models of all types. Our test is especially appealing for neural network models where there are no residual diagnostic checks available. The test was shown to be extremely powerful at spotting both additive and multiplicative heteroscedasticity.

We have suggested a method to assess the form of heteroscedasticity employing a neural network regression, and have shown how this may be used to estimate the regression model using a GLS procedure.

Acknowledgements

We would like to thank Andrew Scott for useful comments on an earlier draft of this paper and an anonymous referee for suggesting the application. Financial support from Fundacao para a Ciencia e a Techologia, BT Laboratories and UK Economic and Social Research Council (ESRC) is gratefully acknowledged.

REFERENCES

BREUSCH, T. & PAGAN, A. (1979) A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, 47, pp. 1287–1294.

CARAPETO, M., HOLT, W. & REFENES, A.-P. N. (2003) On model complexity and selection, *Journal of Statistical Computation and Simulation*, forthcoming.

DURBIN, J. & WATSON, G. (1971) Testing for serial correlation in least squares regression, III, *Biometrika*, 58, pp. 1–19.

GLEJSER, H. (1969) A new test for heteroscedasticity, *Journal of the American Statistical Association*, 64, pp. 316–323.

GOLDFELD, S. & QUANDT, E. (1965) Some tests for homoscedasticity, *Journal of the American Statistical Association*, 60, pp. 539–547.

GOLDFELD, S. & QUANDT, E. (1972) *Nonlinear Methods in Econometrics* (North Holland).

GRIFFITHS, W. E. & SUREKHA, K. (1986) A Monte Carlo evaluation of the power of some tests for heteroscedasticity, *Journal of Econometrics*, 31, pp. 219–231.

HARRISON, M. (1980) The small sample performance of the Szroeter bounds test for heteroscedasticity and a simple test for use when Szroeter's test is inconclusive, *Oxford Bulletin of Economics and Statistics*, 42, pp. 235–250.

HARVEY, A. & PHILLIPS, G. (1974) A comparison of the power of some tests for heteroscedasticity in the general linear model, *Journal of Econometrics*, 2, pp. 307–316.

IMHOF, J. (1961) Computing the distribution of quadratic forms in normal variables, *Biometrika*, 48, pp. 419–426.

KLENIN, M. (1996) Neural networks for risk analysis in stock price forecasts, *Proceedings of the Fifth International Conference on Neural Networks in the Capital Markets* (World Scientific).

MOODY, J. & UTANS, J. (1992) Principled architecture selection for neural networks: application to corporate bond rating prediction. *In:* J. MOODY, S. HANSON & R. LIPPMANN (Eds), *Advances in Neural Information Processing Systems 4*, pp. 683–690 (Morgan Kaufmann).

PARK, R. (1966) Estimation with heteroscedastic error terms, *Econometrica*, 34, p. 888.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. & FLANNERY, B. P. (1992) *Numerical Recipes in C*, 2nd edn (Cambridge University Press).

RAMSEY, J. B. (1969) Tests for specification error in the general linear model, *Journal of the Royal Statistical Society, series B*, 31, pp. 250–271.

SZROETER, J. (1978) A class of parametric tests for heteroscedasticity in linear economic models, *Econometrica*, 46, pp. 1311–1327.

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

Appendix 1

```
load data.txt % Load data : n by k matrix, first col is first differences
of consumption durables, second col is first differences of total
consumption, third col is dummy 2, fourth col is dummy 3, fifth col is
dummy 4

n = length(data); % Number of observations.
ehat = ones(n,1); % Residual errors.

X = ones(n,5); % Matrix for x variables
X( : ,2) = data( : ,2); % X_2 is dx
X( : ,3) = data( : ,3); % X_3 is d2
X( : ,4) = data( : ,4); % X_4 is d3
X( : ,5) = data( : ,5); % X_5 is d4
I1 = zeros(n,n); % I^*.
I2 = zeros(n,n); % I_*.
I = eye(n); % Identity matrix.

r = 37.5; % Proportion of n in each subsample.
m = (r/100)*n; % No. of observations in each subsample.

M = I - X*inv(X'*X)*X'; % Projection matrix.
y = data( : ,1); % Dependent variable.
ehat = M*y; % Estimated errors.
betahat = inv(X'*X)*X'*y; % Estimated regression parameters.

x = data( : ,2); % Variable thought to be causing
heterscedasticity.
x = abs(x);

xsort = sort(x); % Sort x into ASCENDING order.
val1 = xsort(n-m);
val2 = xsort(m);

x = data( : ,2); % Variable thought to be causing
heteroscedasticity.
for t = 1 : n
  if abs(x(t)) > val1
    I1(t,t) = 1;
  else
    I1(t,t) = 0;
  end
  if abs(x(t)) < = val2
    12(t,t) = 1;
  else
    I2(t,t) = 0;
  end
end

q0 = (ehat'*I1*ehat)/(ehat'*I2*ehat)

q0_ pvalue = imhofm(M,I1,I2,q0); % Get the critical value from Imhof.
```

The *p*-value for the statistic can be computed using the Imhof (1961) method, implemented as:

```
function [p] = imhof(M,I1,I2,q0) % Computes p-value for
heteroscedasticity statistic.
a = eig(M'*(I1-q0*I2)*M);
step = 0.001;
top = 1;
integral = 0;
for t = step : step : top % 0 - infinity.
  theta = 0.5* sum(atan(a*t));
  rho = prod(1 + (a.^2)*(t^2))^0.25;
  integral = integral + (sin(theta))/(t*rho);
end
integral = integral*step;
p = 1/2 + (1/pi)*integral;
```