

Score tests for Heteroscedasticity: a New Perspective¹

Score Tests per l'Eterschedasticita': una Nuova Prospettiva

Cinzia Carota

Dipartimento di Statistica e Matematica Applicata, Università di Torino
cinzia.carota@unito.it

Riassunto: In questo lavoro si ricavano delle diagnostiche bayesiane per l'eteroschedasticita' dei residui in un modello di regressione multipla che risultano imperniate sulla funzione *score* alla base di molti *score tests* classici. Questo risultato apre una nuova prospettiva di indagine consistente nella ricerca delle condizioni sotto le quali le statistiche caratterizzanti gli *score tests* classici possono essere utilizzate piu' o meno direttamente in ambito bayesiano.

Keywords: Heteroscedasticity, Score function.

1. Introduction

A general Bayesian diagnostic tool for model criticism has been introduced in Carota, Parmigiani and Polson (1996) where a useful linear version of such a tool is also proposed and widely discussed. In the present paper, in order to test for heteroscedasticity in a multiple regression model, we will employ the linear version of a "symmetrized" form of this diagnostic defined as follows:

$$\bar{\Delta} = \frac{1}{2}[KL_1\{p(\mathbf{l}|\mathbf{y}) : p(\mathbf{l})\} + KL_2\{p(\mathbf{l}) : p(\mathbf{l}|\mathbf{y})\}].$$

In this formula - denoted by $\mathbf{y}$ the sample data - KL_1 represents the Kullback-Leibler divergence between posterior and prior distributions (and KL_2 viceversa) of a random parameter $\mathbf{l}$, which is assumed to express the heteroscedasticity of the disturbances in the multiple regression model, as will be shown in next section. Furthermore, by assumption, a) there is a unique value of $\mathbf{l}$, say $\mathbf{l}_0$ or *null value*, such that the homoscedasticity assumption holds, and b) the prior distribution of $\mathbf{l}$ is sharply peaked around $\mathbf{l}_0$ (a plausible circumstance, given that the currently entertained model assumes the homoscedasticity of the disturbances). These two facts justify the linear approximation of $\bar{\Delta}$ (by expanding the integrands in the KL divergences) in a neighborhood of $\mathbf{l}_0$ and the interpretation of small values of $\bar{\Delta}$ as sample evidence in favor of the homoscedasticity assumption, that is as a reinforcement of the prior opinion largely favorable to the current model.

The major consequence of the above "symmetrization" is the disappearance of the Bayes factor in the linear approximation of $\bar{\Delta}$, $\bar{\Delta}_L$:

$$\bar{\Delta}_L = \frac{1}{2}[E(\mathbf{l}|\mathbf{y}) - E(\mathbf{l})]\frac{\partial \mathcal{L}(\mathbf{l})}{\partial \mathbf{l}}|_{\mathbf{l}=\mathbf{l}_0},$$

¹The present paper is partially supported by M.U.R.S.T. (60%)

where, being $\mathcal{L}$ the log-likelihood of $\mathbf{l}$, we are left with the difference between prior and posterior expectations of $\mathbf{l}$ multiplied by the *score function*. For example, one of the advantages of $\bar{\Delta}_L$ with respect to the original diagnostic ($\Delta = KL_1$) and its corresponding linear form is that all the results on the robustness to the choice of the prior about posterior expectations are immediately available as results on the robustness of $\bar{\Delta}_L$.

In the next section we consider two different approaches to the heteroscedasticity problem and in both cases derive the Bayesian diagnostic $\bar{\Delta}_L$ based on the score function; we then compare the obtained results respectively with the Goldfeldt and Quandt (1965) and the Lagrange multiplier test statistics (see Breusch and Pagan, 1979, Godfrey 1978, Cook and Weisberg, 1983). Finally, section 3 prospects some further research directions.

Before we proceed it is worth observing that in this context $\mathbf{l}$ represents the parameter of interest, while all the other parameters, including the regression parameters, in this preliminary phase of model criticism play the role of "nuisance" parameters. Therefore their priors will be "default" priors.

2. Different approaches to heteroscedasticity

Consider a partitioned linear model: $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}$, where $\mathbf{y}' = (\mathbf{y}'_1 : \mathbf{y}'_2)$, $\mathbf{X}' = (\mathbf{X}'_1 : \mathbf{X}'_2)$ and $\mathbf{u}' = (\mathbf{u}'_1 : \mathbf{u}'_2)$ respectively represent two vectors of observations on the dependent variable y , whose dimensions are $(n_1 \times 1)$ and $(n_2 \times 1)$, $n_1 + n_2 = n$; two matrices of predetermined values of k independent variables, both with rank k ; and two vectors of uncorrelated Gaussian errors with zero means; while $\mathbf{b}$ is a common vector of random parameters. We take up the situation in which the variances in the two groups are unequal and both unknown: $\sigma_1^2 \neq \sigma_2^2$. The likelihood function:

$$L(\mathbf{b}, \sigma_1, \sigma_2 | \mathbf{X}, \mathbf{y}) \propto \frac{1}{\sigma_1^{n_1} \sigma_2^{n_2}} \exp \left[-\frac{1}{2\sigma_1^2} (\mathbf{y}_1 - \mathbf{X}_1 \mathbf{b})' (\mathbf{y}_1 - \mathbf{X}_1 \mathbf{b}) - \frac{1}{2\sigma_2^2} (\mathbf{y}_2 - \mathbf{X}_2 \mathbf{b})' (\mathbf{y}_2 - \mathbf{X}_2 \mathbf{b}) \right],$$

if we change variables from $\mathbf{b}, \sigma_1^2$ and σ_2^2 to $\mathbf{b}, \sigma_1^2$ and $l = \sigma_1^2 / \sigma_2^2$, $l \in (0, \infty)$, (see Zellner, 1971), can be written as follows:

$$\frac{l^{n_2/2}}{\sigma_1^{n_1+n_2+1}} \exp \left\{ -\frac{1}{2\sigma_1^2} [(\mathbf{y}_1 - \mathbf{X}_1 \mathbf{b})' (\mathbf{y}_1 - \mathbf{X}_1 \mathbf{b}) + l(\mathbf{y}_2 - \mathbf{X}_2 \mathbf{b})' (\mathbf{y}_2 - \mathbf{X}_2 \mathbf{b})] \right\}.$$

In this way the parameter of interest turns out to be l ($\mathbf{l}$ reduces to a scalar) whose *null value* is $l = l_0 = 1$, while $\mathbf{b}$ and σ_1 are the nuisance parameters. Assuming that $p(\mathbf{b}, \sigma_1, l) = p(\mathbf{b}, \sigma_1)p(l) \propto \frac{1}{\sigma_1} p(l)$ and omitting additive constants, we obtain that the log of the integrated likelihood with respect to $\mathbf{b}$ and σ_1 is:

$$\begin{aligned} \mathcal{L}(l | \mathbf{X}, \mathbf{y}) &\equiv \log \int L(\mathbf{b}, \sigma_1, l | \mathbf{X}, \mathbf{y}) p(\mathbf{b}, \sigma_1) d\mathbf{b} d\sigma_1 \propto \\ &\frac{n_2}{2} \log l - \frac{1}{2} \log |\mathbf{X}'_1 \mathbf{X}_1 + l \mathbf{X}'_2 \mathbf{X}_2| - \frac{n_1 + n_2 - k}{2} \times \\ &\log \{ \mathbf{y}'_1 \mathbf{y}_1 + l \mathbf{y}'_2 \mathbf{y}_2 - (\mathbf{X}'_1 \mathbf{y}_1 + l \mathbf{X}'_2 \mathbf{y}_2)' (\mathbf{X}'_1 \mathbf{X}_1 + l \mathbf{X}'_2 \mathbf{X}_2)^{-1} (\mathbf{X}'_1 \mathbf{y}_1 + l \mathbf{X}'_2 \mathbf{y}_2) \}, \end{aligned}$$

whose first derivative evaluated at the null value $l_0 = 1$ is the *score function*, here denoted by S :

$$S = \frac{n_2}{2} - \frac{1}{2} \text{tr}\{((\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_1 \mathbf{X}_1 + I)^{-1}\} - \frac{n_1 + n_2 - k}{2} \frac{\hat{\mathbf{u}}'_2 \hat{\mathbf{u}}_2}{\hat{\mathbf{u}}' \hat{\mathbf{u}}},$$

with $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}}$, $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\hat{\mathbf{u}}_2 = \mathbf{y}_2 - \mathbf{X}_2\hat{\mathbf{b}}$. On dividing by $\hat{\mathbf{u}}'_2 \hat{\mathbf{u}}_2$ numerator and denominator of the last term in S we obtain that $\hat{\mathbf{u}}'_2 \hat{\mathbf{u}}_2 / \hat{\mathbf{u}}' \hat{\mathbf{u}} = 1/(GQ + 1)$, where $GQ = \hat{\mathbf{u}}'_1 \hat{\mathbf{u}}_1 / \hat{\mathbf{u}}'_2 \hat{\mathbf{u}}_2$ mimics the Goldfeld and Quandt (1965) test statistic for heteroscedasticity, the difference being that here we do not fit two separate regressions for the two sets of observations. This is not necessary because we do not need to know the distributional properties of GQ and no costs in terms of degrees of freedom are therefore involved. In order to obtain $\hat{\Delta}_L$ the score function S must be multiplied by the difference between prior and posterior expectations of the parameter l . Though not necessary, it is quite natural to assume that the prior expectation coincides with the null value, while to obtain the posterior expectation (the posterior distribution of l is $p(l|\mathbf{X}, \mathbf{y}) \propto p(l) \exp\{\mathcal{L}(l|\mathbf{X}, \mathbf{y})\}$) numerical integration techniques are required. Note that if the regression coefficient vectors were not identical in the two data sets, as in Goldfeld and Quandt (1965), and if, for large m and r , the prior would be as follows:

$$p(l) \propto \begin{cases} l^m & \text{if } 0 < l \leq 1 \\ l^{-r} & \text{if } 1 < l < \infty \end{cases}$$

then the posterior expectation could be immediately evaluated by suitably referring to the standard F tables.

More generally, we can assume that $\mathbf{u}$ follows a multivariate normal distribution with mean zero and diagonal covariance matrix $\sigma^2 \mathbf{W}$, whose diagonal entries are $w_1, w_2, \dots, w_n$, with all $w_i > 0$ and $w_i = w(\mathbf{z}_i, \mathbf{l})$. Now $\mathbf{l}$ is a q -dimensional vector of unknown parameters and $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})'$ denotes a known vector, while the unknown constant σ^2 represents the part of $\text{Var}(y_i)$ common to all the observations. If w_i is a twice differentiable function of $\mathbf{l}$ and there is a unique value $\mathbf{l}_0$ of $\mathbf{l}$ such that $w_i = w(\mathbf{z}_i, \mathbf{l}_0) = 1$, then, after some algebra, the corresponding score function, S_W , can be expressed as:

$$S_W = \frac{1}{2}[\mathbf{d} + \mathbf{D}'(\mathbf{v} - \mathbf{1})]$$

where $\mathbf{d}$ is a q by 1 vector with typical element $d_s = \text{tr}\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\ddot{\mathbf{W}}_s\mathbf{X}\}$ and $\ddot{\mathbf{W}}_s$ denotes the diagonal matrix of the derivatives of w_i , $i = 1, \dots, n$, with respect to l_s ; $\mathbf{D}$ is defined as follows:

$$D = \begin{bmatrix} \frac{\partial w_1}{\partial l_1} & \cdots & \frac{\partial w_1}{\partial l_q} \\ \vdots & \vdots & \vdots \\ \frac{\partial w_n}{\partial l_1} & \cdots & \frac{\partial w_n}{\partial l_q} \end{bmatrix};$$

and $\mathbf{v} - \mathbf{1}$ is a q -dimensional vector with typical element $(\hat{u}_s^2 / \tilde{\sigma}^2 - 1)$. Here $\tilde{\sigma}^2 = \sum \hat{u}_s^2 / (T - k)$, i. e. the unbiased estimation of the common variance σ^2 instead of the maximum likelihood estimation appearing in the classical score function (see for example Godfrey, 1978, Cook and Weisberg, 1983). As to the difference between prior and

posterior expectation, for a given prior distribution $p(\mathbf{l})$, the posterior is:

$$p(\mathbf{l}|\mathbf{X}, \mathbf{y}) \propto p(\mathbf{l}) \prod_1^n w_i^{-1/2} |\mathbf{X}'_w \mathbf{X}_w|^{-1/2} \{\mathbf{y}'_w [I - \mathbf{X}_w (\mathbf{X}'_w \mathbf{X}_w)^{-1} \mathbf{X}'_w] \mathbf{y}_w\}^{-(n-k)/2}$$

where $\mathbf{X}_w = \mathbf{W}^{-1/2} \mathbf{X}$ and $\mathbf{y}_w = \mathbf{W}^{-1/2} \mathbf{y}$. For further progress an explicit form for w_i must be chosen. Specific families are: $w_i = (\sum l_s z_{is})^t$, with t a pre-specified integer; $w_i = \exp\{\sum \lambda_s z_{is}^{a_s}\}$, which, if by convention we take $z^a = \log z$ when $a = 0$, yields important sub-families for $\mathbf{a} = \mathbf{0}$ and $\mathbf{a} = \mathbf{1}$ (see Cook and Weisberg, 1983); and $w(\mathbf{z}_i, \mathbf{l}) = w(\mathbf{x}'_i \mathbf{b}, l)$, in which $\mathbf{x}_i$ is the i -th row of $\mathbf{X}$ and $\mathbf{l}$ reduces to l , so that the variance is constrained to depend on the expected response.

3. Conclusions

A further possibility, which we do not explore here, is embedding the problem of heteroscedasticity in the framework of the mixed linear model. In an interesting (albeit fairly short) branch of literature, a number of authors have suggested that parameter heterogeneity can be reasonably viewed as due to stochastic variation. The most widely cited studies are Hildreth and Houch (1968), many papers by Swamy (for a survey see Swamy, 1973) and Hsiao (1975). In this framework, under suitably conditions, we expect even stronger links of $\bar{\Delta}_L$ with the corresponding Lagrange multiplier tests.

References

- Breusch T.S., Pagan A.R. (1979) A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, 47, n.5, 1287-1294.
- Carota, C., Parmigiani, G., Polson, N. G. (1996) Diagnostic Measures for Model Criticism, *Journal of the American Statistical Association*, 91, 753-762.
- Cook D., Weisberg S. (1983) Diagnostics for heteroscedasticity in regression, *Biometrika*, 71, n.1, 1-10.
- Godfrey L. (1978) Testing for multiplicative heteroskedasticity, *Journal of Econometrics*, 8, 227-236.
- Goldfeld S., Quandt R. (1965) Some Tests for Homoscedasticity, *Journal of the American Statistical Association*, 60, 539-547.
- Hildreth C. Houch C. (1968) Some Estimators for Linear Models with Random Coefficients, *Journal of the American Statistical Association*, 63, 584-595.
- Hsiao, C. (1975) Some Estimation Methods for a Random Coefficient Model, *Econometrica*, 43, 305-325.
- Swamy P. (1973) Criteria, Constraints and Multicollinearity in Random Coefficient Regression Models, *Annals of Economics and Social Measurements*, 2, 429-450.
- Zellner A. (1971) *An Introduction to Bayesian Inference in Econometrics*, Wiley, N. Y.