ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 20/21 — 1991

Bayesian Diagnostics for Heterogeneity

Luc BAUWENS, Michel LUBRANO *

ABSTRACT. – In this paper we examine the problem of testing for heterogeneity and heterosckedasticity in a Bayesian framework. We first show that a model with random coefficients is identical to a model with heteroskedastic residuals. We then consider two approaches for testing. The first one is concerned with the point of view of misspecification. The original model is homoskedastic. One is willing to detect any departure from homoskedasticity. We propose diagnostics based on the examination of the Bayesian residuals, after stressing the differences between classical and Bayesian residuals. In the second approach, the starting point is a precise form of the alternative hypothesis, and the model for inference is heteroskedastic. A test for homoskedasticity is then a test for a parameter restriction. This can be done by looking at highest posterior probability regions or by the use of the posterior odds ratio. As a joint product, we develop the posterior analysis of a heteroskedastic regression model for several classes of prior distributions.

Diagnostics bayesiens d'hétérogénéité

RÉSUMÉ. – Dans cet article, nous proposons des tests d'hétérogénéité et d'hétéroscédasticité du point de vue bayesien. Nous commençons par montrer qu'un modèle à coefficients aléatoires est équivalent à un modèle à erreurs hétéroscédastiques. Nous envisageons ensuite deux approches de test. La première approche part du point de vue de la recherche d'erreurs de spécification. Le modèle de départ est homoscédastique. On essaie de détecter tout écart à l'hypothèse d'homoscédasticité. Nous proposons pour ce faire des diagnostocs fondés sur les résidus bayesiens, après avoir mis l'accent sur les différences entre les résidus classiques et les résidus bayesiens. Dans la seconde approche, le point de départ est une forme précise de l'hypothèse alternative, et le modèle d'inférence est hétéroscédastique. Un test d'homoscédasticité est alors un test de restriction paramétrique. Il peut se faire par la méthode de la région de plus haute probabilité *a posteriori* ou par celle du rapport des probabilités *a posteriori*. Un produit joint de notre étude est l'analyse *a posteriori* d'un modèle de régression hétéroscédastique pour quelques classes de distributions *a priori*.

---

* L. BAUWENS: EHESS (GREQE); M. LUBRAND: CNRS (GREQE). We are grateful to Russel Davidson, Jean-François Richard and Herman Van Dijk for helpful discussions and comments. We are also grateful to Jean-Pierre Florens and one anonymous referee for their reports on a previous version of the paper. Remaining errors are solely ours.

1 Introduction

Heterogeneity occurs whenever the "Identically Distributed" assumption is violated. This covers a great variety of situations. In order to be operational let us restrict ourselves to the case of a regression model with random coefficients:

$$y_t = x_t' \beta_t + u_t, \qquad u_t \sim N(0, \sigma^2)$$

where the regression coefficient β_t is allowed to be different for each observation. More precisely β_t is supposed to be drawn from a normal density. In a Bayesian framework it is natural to assume that the β_t are distributed according to a common normal prior density of the form:

$$\beta_t \sim N(\beta, \Sigma).$$

The variation of β_t represents the heterogeneity of the observations. It is possible to specify further a second stage prior density on the parameters β and Σ (it can be called a hierarchical prior, see for instance BERGER [1988, p. 106]). In the absence of repeated observations on the individuals indexed by t it is convenient to integrate out all the β_t and to make inference on the remaining parameters β, Σ and σ^2. Thus given our assumptions, the joint density of (y_t, β_t') is normal. The marginal density of y_t is also normal with:

$$p(y_t | \beta, \sigma^2, \Sigma) = f_N(y_t | x_t' \beta, \sigma^2 + x_t' \Sigma x_t)$$

(*see* the appendix for notations). This is nothing but the data density of a regression model with heteroskedastic errors. In this case the skedastic function is of the form $\sigma_t^2 = \sigma^2 + z_t' \gamma$. A wider class of skedastic functions will be considered in section three. [1]

What are the consequences of heterogeneity and more specifically of heteroskedasticity? In a regression model heteroskedasticity induces bias in the estimator of the variance of the ordinary least square estimator. Although the estimator itself is unbiased, the usual t statistics are no longer correct.

The most general approach for detecting heterogeneity in a maximum likelihood framework is provided by the information matrix test of WHITE [1982]. The idea underlying the test is quite simple. Under the hypothesis of correct specification the two usual ways of computing the information matrix give the same result. One is based on minus the expectation of the

1. In a classical framework the random coefficient β_t is modelled by the auxiliary regression $\beta_t = \beta + \varepsilon_t$ with $\varepsilon_t \sim N(0, \Sigma)$. The latter is combined with the original regression. The data density is then identical with the one obtained here.

second derivative of the log-likelihood function. The other is the expectation of the outer product of the gradient of the same log-likelihood function. In the case of misspecification, the two results differ. The test of the information matrix checks for that difference. However a direct implementation of the test is not at all straightforward. Various artificial regressions have been devised to provide practical results: *see* e.g. the papers by WHITE [1980], CHESHER [1983, 1984], and DAVIDSON-MACKINNON [1988]. As underlined in DAVIDSON-MACKINNON [1988] the proposed classical tests are either clumsy or reject too often the null. In order to circumvent the problem, one has to use double length artitificial regressions.

The idea of this paper is to attack the problem from a Bayesian point of view and so to get small sample results based on exact posterior densities. We shall deal more specifically with the testing of heteroskedasticity since we have shown that heterogeneity and heteroskedasticity can be highly connected. Surprisingly the issue of testing for heteroskedasticity has not been studied in detail in the Bayesian literature. As far as we know, if we except the broad papers by BOX and TIAO [1962] and by ZELLNER [1975], there exists only few papers on the subject: a paper by GEWEKE [1986] dealing with ARCH models and a paper by POIRIER [1987] based on posterior odds.

The paper is articulated around two different points of view:

- The first one is concerned with the problem of misspecification. The original model is homoskedastic. One is willing to detect possible forms of heteroskedasticity. In a classical framework, one usually goes through the examination of the residuals with Lagrange multiplier tests. In the first section of the paper we compare classical and Bayesian residuals and in the second section we propose Bayesian diagnostics based on the examination of the residuals.

- In the second approach the starting point is the alternative hypothesis and the model for inference is heteroskedastic. We develop in section three Bayesian inference for this model. Homoskedasticity is obtained as a particular case and can be tested in two different ways. This is the object of section four. The analogue of a Wald test consists in examining if a parameter representing heteroskedasticity can be set equal to zero (see if zero belongs to a posterior confidence interval of this parameter). The Bayesian analogue of the likelihood ratio test is the posterior odds ratio which compares the averaged likelihood function of a model with and without heteroskedasticity.

2 Comparaison of Classical and Bayesian Residuals

A great number of classical heteroskedasticity diagnostics are based on the examination of the residuals. See references for instance in PAGAN

[1984]. In the next section we shall propose Bayesian diagnostics based on the posterior distribution of the residuals. It is then very useful to compare the properties of classical and Bayesian residuals for the sake of clarity.

The model we shall consider under the maintained hypothesis is the linear regression model:

$$y = X\beta + u \tag{1a}$$

$$u \sim N(0, \sigma^2 . I_T) \tag{1b}$$

where y is a vector of T observable random variables for the endogenous variable, X a $T \times k$ matrix of exogenous variables, and u a $T \times 1$ vector of error terms. Alternatively we can define the sampling process in terms of the observable vector y:

$$y \sim N(X\beta, \sigma^2 . I_T) \tag{2a}$$

from which we can derive the definition and distribution of the error term:

$$u = y - X\beta \sim N(0, \sigma^2 . I_T). \tag{2b}$$

Residuals on the contrary are defined only after inference on the parameters and consequently will be different in a classical and in a Bayesian framework.

2.1. Classical Residuals

The classical residuals are obtained by replacing β by its OLS estimator $\hat{\beta}$:

$$\begin{aligned} \hat{u} &= y - X\hat{\beta} \\ &= X\beta + u - X(X'X)^{-1}X'(X\beta + u). \end{aligned} \tag{3}$$

Their expectation taken with respect to the density of u is zero as $\hat{\beta}$ is supposed to be an unbiased estimator of β and their variance for fixed X is:

$$\begin{aligned} \mathrm{Var}_u(\hat{u}) &= \mathrm{Var}_u(y) + X . \mathrm{Var}_u(\hat{\beta}) . X' - 2 . \mathrm{Cov}_u(y, X\hat{\beta}) \\ &= \sigma^2 . [I_T - X(X'X)^{-1}X'] \\ &= \sigma^2 . M_X. \end{aligned} \tag{4}$$

Let us note that this matrix has rank $T-k$ and thus is singular. So the OLS residuals are neither independent nor homoskedastic. However they converge in distribution to u as $T \to \infty$.

The individual residuals $\hat{u}_t = y_t - x_t'\hat{\beta}$ have a zero mean and a variance given by:

$$\begin{aligned} \mathrm{Var}_u(\hat{u}_t) &= \sigma^2 . [1 - x_t'(X'X)^{-1}x_t] \\ &= \sigma^2 . M_{tt}^X. \end{aligned} \tag{5}$$

The distribution of $\hat{u}$ or of $\hat{u}_t$ is normal as it results from the linear combination of two normals. It depends on the unknown parameter σ^2.

The OLS estimator of σ^2 (denoted by s^2) divided by σ^2 is distributed as $\chi^2(T-k)$ and is independent of $\hat{\beta}$. The distribution of a standardized individual residual is then Student with $T-k$ degrees of freedom:

$$\hat{u}_t / \sqrt{s^2 . M^X_{tt}} \sim t(T-k). \tag{6}$$

The squares of the individual residuals are often used in auxiliary regressions. So it is worth computing their characteristics. As the mean of the residuals is zero, the expectation of their square is equal to their variance:

$$E_u(\hat{u}_t^2) = \sigma^2 . M^X_{tt}. \tag{7}$$

The square of the standardized individual residuals has an F distribution with 1 and $T-k$ degrees of freedom:

$$\frac{\hat{u}_t^2}{s^2 . M^X_{tt}} \sim F(1, T-k). \tag{8}$$

By properties of the F distribution, their mean is $(T-k)/(T-k-2)$ and their variance is the square of the mean multiplied by $2.(T-k-1)/(T-k-4)$.

The squares of classical residuals enter in the building of a number of artifical regressions for testing for heteroskedasticity. Apart from a constant term, the various regressors candidates are: dummy variables for the usual GOLDFELD-QUANDT test, powers of y_t for the RESET test (as interpreted in PAGAN [1984]), lags of $\hat{u}_t^2$ for the ARCH test, squares and cross products of the elements of x_t for WHITE's test. The null (homoskedasticity) is rejected if the coefficients of these extra regressors are not zero (see PAGAN [1984]).

2.2. Bayesian Residuals

The logic underlying the definition of Bayesian residuals is radically different (*see* for instance ZELLNER [1975]). Contrary to the situation in the classical framework, β is now a random variable provided with a prior distribution. The process of inference can be described as the passage from the prior distribution of β to its posterior distribution. The posterior analysis is performed conditionally on the sample y. Bayesian residuals [2] are viewed as a random variable for which a posterior distribution has to be found. Under a natural conjugate prior the posterior density of β is Student (*see* the appendix for notation):

$$D''(\beta \mid y) = f_t(\beta \mid \beta_*, M_*, s_*^2, \nu_*) \tag{9}$$

2. Other denomination have been used in the literature. ZELLNER-MOULTON [1985] speak of realized residuals. We are not satisfied with it. In a Bayesian context realized is equivalent to non-random which of course is not acceptable here. DREZE [1977] uses the expression "unobserved realizations of the disturbances" which may be more accurate, but is a bit heavy. Posterior residuals would not be better as is suggests that there could be prior residuals which is not possible as then y is not yet observed.

with:

$$M_* = M_0 + X'X$$
$$\beta_* = M_*^{-1} . (M_0 \beta_0 + X' y)$$
$$s_*^2 = s_0^2 + y' y + \beta_0' M_0 \beta_0 - \beta_*' M_* \beta_*$$
$$\nu_* = \nu_0 + T$$

where β_0, M_0, s_0^2, ν_0 are the parameters of the normal inverted gamma-2 prior distribution. The posterior density of σ^2 is an inverted gamma-2 with scale parameter s_*^2 and ν_* degrees of freedom so that:

$$E(\sigma^2 | y) = s_*^2/(\nu_* - 2).$$

The Bayesian residuals $u = y - X\beta$, where y is understood to be the observed sample, are a linear combination of the elements of the multivariate Student random vector β. Their posterior distribution is a singular Student:

$$D''(u | y) = f_t(u | y - X\beta_*, P_{X^*}^+, s_*^2, \nu_*) \tag{10}$$

with expectation:

$$\begin{aligned} E(u | y) &= y - X . E(\beta | y) \\ &= y - X\beta_* = \theta \end{aligned} \tag{11}$$

and variance:

$$\begin{aligned} V(u | y) &= X . V(\beta | y) . X' \\ &= X M_*^{-1} X' . s_*^2/(\nu_* - 2) \\ &= P_{X^*} . s_*^2/(\nu_* - 2). \end{aligned} \tag{12}$$

In (10) $P_{X^*}^+$ stands for the Moore-Penrose inverse of $P_{X^*} = X M_*^{-1} X'$. By marginalization [3] we get the posterior distribution of individual residuals:

$$D''(u_t | y) = f_t(u_t | \theta_t, (P_{tt}^{X^*})^{-1}, s_*^2, \nu_*) \tag{13}$$

where:

$$\theta_t = y_t - x_t' \beta_* \qquad \text{and} \qquad P_{tt}^{X^*} = x_t' M_*^{-1} x_t.$$

Its variance is given by:

$$\begin{aligned} V(u_t | y) &= x_t' M_*^{-1} x_t . s_*^2/(\nu_* - 2) \\ &= P_{tt}^{X^*} . s_*^2/(\nu_* - 2). \end{aligned} \tag{14}$$

Let us define a standardized Bayesian residual by (see the appendix):

$$\bar{u}_t = (u_t - \theta_t)/\sqrt{P_{tt}^{X^*} . s_*^2/\nu_*}. \tag{15}$$

3. One could directly define $u_t = y_t - x_t' \beta$.

It has a standardized Student distribution with ν_* degrees of freedom. The square of a standardized residual is distributed as a Fisher $(1, \nu_*)$:

$$\tilde{u}_t^2 \sim F(1, \nu_*). \tag{16}$$

The differences between classical and Bayesian residuals are substantial. The classical residual has a zero mean. Under a diffuse prior the Bayesian residual has a posterior mean different from zero and equal to the classical residual. The variance of the classical residual is proportional to the diagonal elements of M_X. Under a diffuse prior the posterior variance of the Bayesian residuals is proportional to the diagonal elements of $P_X = I_T - M_X$. The asymptotic behaviour of the classical and the Bayesian residuals is not the same. For increasing T and under the usual regularity conditions, M_X converges to the identity matrix when P_X converges to zero. This difference comes from the fact that in a classical framework u_t can be seen as an incidental parameter; the variance of its estimator $\hat{u}_t$ is computed with respect to the distribution of y (see 14) the dimension of which is T. In a Bayesian framework the posterior moments of u_t are computed with respect to the distribution of β, the dimension of which is independent of the sample size. Suppose that we know the true realisation of the residuals (say as in a Monte-Carlo experiment). For increasing T the Bayesian posterior distribution of the residuals converges to the true point values. In the classical framework, the distribution of the estimator $\hat{u}_t$ converges to the theoretical *distribution* of u_t.

3 Heterogeneity Diagnostics

In this section and in the next one we propose regression diagnostics based on the examination of the residuals. The correct model under H_0 is given by (1.1 *a*)-(1.1 *b*):

$$H_0: \quad y_t = x_t'\beta + u_t, \qquad u_t \sim N(0, \sigma^2).$$

The alternative hypothesis could be a random coefficient model as given in the introduction:

$$H_1: \quad y_t = x_t'\beta_t + u_t, \qquad \beta_t \sim N(\beta, \Sigma), \qquad u_t \sim N(0, \sigma^2)$$

which gives an additive heteroskedastic model:

$$H_1: \quad y_t = x_t'\beta + v_t, \qquad v_t \sim N(0, \sigma^2 + x_t'\Sigma x_t).$$

3.1. Tests Based on the Bayesian Residuals

In the Bayesian approach all the inferential results are given by the posterior density of the parameters. For devising heterogeneity diagnostics, the answer should be contained in the posterior density of the residuals which are expressed as simple linear transformations of the parameters. In particular individual effects should appear as significant differences among the posterior densities of the individual residuals.

In a Bayesian model the predictive density expresses the prior belief the statistician has about the outcome of the experiment, before its realization. As in this section the model is supposed to be correctly specified the predictive density of the residuals illustrates the hypothesis of correct specification.

The test principle used in this section is based on the comparison between these two distributions of the residuals, posterior and predictive. That is to say we shall compare what the residuals really are (posterior) to what they should be (predictive). As this procedure is based solely on inference in the model under H_0, it is the spirit of a classical score test.

3.2. Detecting Individual Effects in the Mean

Let us first compute the predictive density of an individual residual. It is defined as the integral in σ^2 of the product between the sampling density (2*b*) and the prior density on σ^2. We have:

$$
\begin{aligned}
p(u_t) &= \int p(u_t \mid \sigma^2) \cdot \mathrm{D}(\sigma^2) \cdot d\sigma^2 \\
&= \int f_{\mathrm{N}}(u_t \mid 0, \sigma^2) \cdot f_{i\gamma}(\sigma^2 \mid s_0^2, \nu_0) \cdot d\sigma^2 \qquad (17) \\
&= f_t(u_t \mid 0, 1, s_0^2, \nu_0).
\end{aligned}
$$

Its moments are:

$$\mathrm{E}(u_t) = 0 \qquad (18)$$

$$\mathrm{V}(u_t) = s_0^2/(\nu_0 - 2) = \mathrm{E}(\sigma^2). \qquad (19)$$

Let us now turn to the posterior distribution of a Bayesian residual u_t and compute a highest posterior probability confidence interval. From (15) we know that this interval has the following form:

$$\theta_t \pm t_\alpha \cdot \sqrt{\mathrm{P}_{tt}^{X^*} \cdot s_*^2/\nu_*} \qquad (20)$$

where t_α is the value of the Student distribution for the level α. We can now check if this interval contains or not the predictive expectation which is zero. This procedure indicates if a residual has been generated by a zero mean density with posterior probability $(1-\alpha)$. A graphical solution may be quite useful to a global vision. The mean of the Bayesian residual $\theta_t = y_t - x_t' \beta_*$ is plotted together with probability bounds. Then one looks

if zero is outside of the bounds. Note the difference with the classical procedure where one looks if a classical residual falls outside a confidence interval centered on zero.

If zero is not contained in the Bayesian confidence interval, the corresponding residual is called an outlier as in CHALONER, BRUNST [1986]. It can also be interpreted as an individual effect not taken into account under H_0. ZELLNER, MOULTON [1985] have a wider definition of an outlier. They first use graphs of the posterior means together with posterior standard deviations. But they also compute $x_t'(X'X)^{-1}x_t$ for detecting what they call influential observations.

What is the influence of the prior information for this analysis? Or more generally should a misspecification test be done only under a diffuse prior? With an informative prior a Bayesian estimator (the posterior mean) is biased in the classical sense and the posterior mean of the residuals may well not be near zero simply because of the prior. So the test may detect only the presence of a prior information. This fact should be tempered by two remarks:

– A Bayesian model is made of two elements, the prior *and* the data density. An outlier is thus defined for the complete model and not only for the data density.

– From a classical point of view, a Bayesian estimator is certainly strongly biased if the prior contradicts the sample. From a Bayesian point of view, there is a contradiction of information between the prior and the sample, and the posterior density should reflect this conflict by means of large posterior standard deviations. However this is not a situation which is especially sought after. The Bayesian model should be revised in such a case in order to finally reach a situation where the prior is dominated by the sample (unless it serves to solve an identification problem).

3.3. Detecting Individual Effects in the Variance

The detection of heteroskedasticity is a bit more elaborate as it concerns the variance of the residuals and not only the mean. To check for heteroskedasticity or individual effects in the variance, we have to turn to the distribution of the square of the residuals and try to adopt the same methodology as before.

Let us compute the predictive density of the square of a residual. From (17) we can standardize the Student predictive, so:

$$u_t \sim t(\nu_0) \cdot \sqrt{s_0^2/\nu_0}.$$

Then its square is simply a Fisher random variable times a constant:

$$u_t^2 \sim F(1, \nu_0) \cdot s_0^2/\nu_0 \tag{21}$$

and its expectation can be written as:

$$E(u_t^2) = \frac{\nu_0}{\nu_0 - 2} \cdot s_0^2/\nu_0 = s_0^2/(\nu_0 - 2) = E(\sigma^2) \tag{22}$$

which is simply the prior expectation of σ^2.

In (15) and (16) we have derived the posterior density of the standardized residual $\bar{u}_t$ and of its square $\bar{u}_t^2$. We have:

$$\bar{u}_t^2 = (u_t^2 - 2 \cdot u_t \theta_t + \theta_t^2) \cdot \frac{\nu_*}{s_*^2 \cdot P_{tt}^{X^*}} \sim F(1, \nu_*).$$

Consequently:

$$u_t^2 = \bar{u}_t^2 \cdot s_*^2 \cdot P_{tt}^{X^*}/\nu_* + 2 \cdot u_t \theta_t - \theta_t^2.$$

From the posterior distribution of u_t we have:

$$2 u_t \theta_t - \theta_t^2 \sim t(\theta_t^2, [4 \cdot \theta_t^2 \cdot P_{tt}^{X^*}]^{-1}, s_*^2, \nu_*). \tag{23}$$

Then the square of a residual is equal to a Fisher random variable with one and ν_* degrees of freedom multiplied by $\nu_*/(\nu_* - 2)$ times the posterior variance of the residuals plus a Student random variable with mean equal to the square of the posterior mean of this residual. The posterior expectation of this variable is quite easy to compute:

$$\begin{aligned} E(u_t^2 \mid y) &= P_{tt}^{X^*} \cdot s_*^2/(\nu_* - 2) + \theta_t^2 \\ &= \mathrm{Var}(u_t \mid y) + \theta_t^2 \\ &= E(\sigma^2 \mid y) \cdot P_{tt}^{X^*} + \theta_t^2 \end{aligned} \tag{24}$$

Some comments can be done about (24). For an increasing number of observations this expectation converges toward θ_t^2. So the dominant element is not $E(\sigma^2 \mid y)$. Small sample variation around $E(\sigma^2 \mid y)$ is allowed by $P_{tt}^{X^*}$. The dominant element is the square of the expectation of the Bayesian residual.

Individual effects can be detected with the examination of squared Bayesian residuals by constructing confidence intervals. A simulation method can be used for this purpose as an analytical confidence interval is not available: use the Student distribution (13) and take the square of the drawings. In the case of correct specification (homoskedasticity), these confidence intervals should be located all in the same area (overlap) as they should all contain the same point value represented by $E(\sigma^2)$, the predictive expectation of u_t^2. If we are non-informative we can replace this value by the posterior expectation of σ^2. On the contrary outliers are synonyms of heterogeneity.

3.4. Detecting Heteroskedasticity with Auxiliary Regressions

The previous procedure to detect a particular pattern in the variance of the Bayesian residuals is quite general, but may not be convenient as it requires the numerical evaluation of T posterior confidence intervals. We propose here a method based on an auxiliary regression at the cost of some extra hypothesis.

We can always approximate a random variable by its conditional expectation plus an orthogonal error term. We do this operation for the predictive and the posterior expectation of the squared Bayesian residuals.

From the predictive expectation (22) we have:

$$u_t^2 = \frac{s_0^2}{\nu_0 - 2} + \varepsilon_{1t}$$

where the conditioning variable is just a constant.

From the posterior expectation (24) we have the alternative approximation:

$$u_t^2 = \frac{s_*^2}{\nu_* - 2} \cdot P_{tt}^{X^*} + \theta_t^2 + \varepsilon_{2t}. \tag{25}$$

In the case of correct specification these two approximations should coincide. So in term of the observables we get:

$$\theta_t^2 = \frac{s_0^2}{\nu_0 - 2} - \frac{s_*^2}{\nu_* - 2} \cdot P_{tt}^{X^*} + \varepsilon_{1t} - \varepsilon_{2t}. \tag{26}$$

Introducing parameters we arrive at the following auxiliary regression:

$$\theta_t^2 = \alpha_0 - \alpha_1 \cdot P_{tt}^{X^*} + \varepsilon_t. \tag{27}$$

This means that if the specification is correct, the conditional expectation of the observable θ_t^2 is constant apart from a small sample deviation due to the variation of the exogenous variable X (for large T $P_{tt}^{X^*}$ vanishes). So the posterior distribution of α_0 in (27) and the posterior distribution of σ^2 in the model under H_0 should be very much alike in the case of correct specification. Otherwise other variables not contained in (27) are necessary to explain the variation of θ_t^2 which is then due to heteroskedasticity. This test however is not very feasible. The distribution of $\varepsilon_t = \varepsilon_{1t} - \varepsilon_{2t}$ is quite complicated. Apart from multiplicative terms, it is the difference between two Fisher random variables and a Student random variable. By construction its expectation is zero. Bayesian inference for α_0 does not seem tractable if we want to base it on the exact distribution of ε_t. A normal approximation for ε_t is not convenient as it would yield a Student posterior density for α_0 which should be compared to the inverted gamma-2 posterior density of σ^2.

The alternative solution is to make a precise hypothesis on the form of heteroskedasticity considered. Let us suppose that under H_1 we have:

$$u_t \sim N(0, \sigma_t^2) \tag{28}$$

provided with the following incomplete prior density:

$$D'(\sigma_t^2) = f_{i\gamma}(\sigma_t^2 \mid \sigma^2 + z_t'\gamma, \upsilon_0) \tag{29}$$

where z_t is a set of extra exogenous variables and (σ^2, γ) unknown parameters. This case is a bit more flexible than the random coefficient model as z_t is not restricted to be composed of the squares and cross products of x_t and the only restriction on (σ^2, γ) is that $\sigma^2 + z_t'\gamma$ be positive. The predictive density of the residuals under H_1 is Student:

$$p(u_t \mid z_t, \sigma^2, \gamma) = f_t(u_t \mid 0, 1, \sigma^2 + z_t'\gamma, \upsilon_0)$$

and the predictive density of the square of the residuals is Fisher:

$$u_t^2 \mid z_t, \sigma^2, \gamma \sim F(1, \upsilon_0).(\sigma^2 + z_t'\gamma)/\upsilon_0 \tag{30}$$

with expectation:

$$E(u_t^2 \mid z_t, \sigma^2, \gamma) = (\sigma^2 + z_t'\gamma)/(\upsilon_0 - 2). \tag{31}$$

We shall follow the same path as before but using (31) instead of (22). We thus build the following linear approximation of u_t^2 from this conditional expectation:

$$u_t^2 = \sigma^2 + z_t'\gamma + \varepsilon_{1t}. \tag{32}$$

From the posterior expectation of the square of the residuals evaluated under the null hypothesis we build the same approximation as in (25). Under the alternative hypothesis H_1 it should be equivalent to (32). Merging these two expressions we have:

$$P_{tt}^{X^*}.s_*^2/(\nu_* - 2) + \theta_t^2 = \sigma^2 + z_t'\gamma + \varepsilon_{1t} - \varepsilon_{2t}$$

which gives the auxiliary regression we are looking for:

$$\theta_t^2 = \alpha_0 + \alpha_1 P_{tt}^{X^*} + z_t'\gamma + \varepsilon_t. \tag{33}$$

The distribution of $\varepsilon_t = \varepsilon_{1t} - \varepsilon_{2t}$ is the same as before. But this time homoskedasticity will be rejected if inference on γ does not provide a multiple confidence interval containing zero. The posterior distribution of γ can be symmetric without harm in practical cases. [4] So a normal approximation for the distribution of ε_t does not spoil anything.

The auxiliary regression (33) is certainly a very convenient Bayesian test for detecting heteroskdasticity. It has been implemented in the "Bayesian Interactive Program" currently developped at GREQE. This auxiliary

4. The domain of definition of γ will be precised in the next section.

regression is similar but nevertheless different from the classical ones surveyed in PAGAN [1984] due to the presence of the term in $P_{tt}^{X^*}$. However for an increasing number of observations, this term vanishes and the two approaches coincide. This term can be seen as a small sample corrector factor.

4 Posterior Analysis of a Heteroskedastic Regression Model

In the previous sections, we examined the problem of heteroskedasticity from the point of view of H_0, conducting inference in the standard regression model. The detection of possible misspecification due to heteroskedasticity was done by inspection of the residuals. Now we decide to adopt the symmetrical point of view, that of H_1. By assuming some parametric form of heteroskedasticity in a regression model, it is possible to draw inference on the parameters of the regression function (say β), as well as on the parameters inducing the variability of the variance among the observations (say γ), through their posterior distribution. If the variance becomes constant for a particular value (say 0) of the parameter γ, H_0 (the homoskedastic model) is nested within H_1 (the heteroskedastic model). In this section we shall detail inference for the heteroskedastic model and its relationship with the random coefficient model. Testing will occupy section five.

4.1. Heteroskedastic Models

Let us consider the following model:

$$y_t = x_t' \beta + u_t, \qquad u_t \sim N(0, \sigma_t^2) \tag{34}$$

similar to (1) except that now we allow the variance of u_t to take different values at different points of time. We shall suppose as in BREUSCH and PAGAN [1979] that:

$$\sigma_t^2 = \sigma^2 . h(z_t' \gamma)$$

with $h(z_t' \gamma) > 0$, $\forall z_t \in \mathbb{R}^l$ and $\forall \gamma \in G \subseteq \mathbb{R}^l$. It may be necessary to truncate the space $\mathbb{R}^l$ of γ to G to ensure the positivity of the function h. Notice however that what really matters is that for any possible value of z_t, the function $h(\gamma' z_t)$ be strictly positive. The condition that e. g. G be the positive orthant of $\mathbb{R}^l$ may not be necessary, but only sufficient.

The variables in x_t and z_t are assumed to be weakly exogenous and may coincide, except that z_t is assumed not to include a constant, so that when γ is equal to 0, $h(0) = 1$ and $\sigma_t^2 = \sigma^2$. Two well-known examples that

encompass most cases used in practice are:

$$\sigma^2 . h(z_t' \gamma) = \sigma^2 . \exp(z_t' \gamma)$$

which corresponds to multiplicative heteroskedasticity and:

$$\sigma^2 . h(z_t' \gamma) = \sigma^2 + z_t' \tilde{\gamma}$$

which corresponds to additive heteroskedasticity [5] and where $\tilde{\gamma} = \gamma / \sigma^2$. A closely related case is the ARCH linear model of ENGLE [1982] analysed by GEWEKE [1986, 1989] from a Bayesian viewpoint.

The random coefficient model given in the introduction is a particular case of additive heteroskedasticity with the following skedastic function:

$$\sigma^2 . h(z_t' \gamma) = \sigma^2 + x_t' \Sigma x_t .$$

So z_t is defined with the squares and cross products of x_t and γ is an obvious reparametrization of Σ. Then G, the parametric space of γ is defined by the condition that Σ is positive definite. GRIFFITHS *et al.* [1979] detail Bayesian inference for a restricted version of this model (only two regressors and diagonal Σ).

4.2. Likelihood Function

Let $\theta = (\beta' \sigma^2 \gamma')'$ be the $k+1+l$ vector of parameters and H_γ the $T \times T$ diagonal matrix with typical element $1/h(z_t' \gamma)$. This matrix can be viewed as the relative precision matrix of the error terms. The likelihood function can be written as:

$$(35) \quad L(\theta | y) \propto \sigma^{-T} . |H_\gamma|^{1/2} . \exp\left\{ - \frac{1}{2\sigma^2} [(\beta - \hat{\beta}_\gamma)' X' H_\gamma X (\beta - \hat{\beta}_\gamma) + \hat{s}_\gamma^2] \right\}$$

where

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

is the Weighted Least Squares estimator and:

$$\hat{s}_\gamma^2 = y' [H_\gamma - H_\gamma X (X' H_\gamma X)^{-1} X' H_\gamma] y .$$

Conditionaly on γ this likelihood function is very similar to that of an ordinary regression model.

4.3. Prior Distributions

The parameters γ and σ^2 will be considered as nuisance parameters, so that we shall not bother too much about defining informative prior densities

5. An alternative parametrisation is to write $\sigma_t^2 = h(\alpha_0 + \alpha' z_t)$, so that $\sigma^2 = h(\alpha_0)$.

for these parameters, except if this could simplify computations. For the parameter β, we wish however to allow the use of an informative prior density, as well as a non-informative one. The use of Jeffreys' invariance principle does not yield a simple diffuse prior because the nuisance parameters appear in all the non-zero elements of the information matrix of (35).

– The prior on σ^2 and γ is rather easy to devise. As most of the time we want to be non-informative on these parameters we choose a prior of the form:

$$D'(\sigma^2, \gamma) = D'(\sigma^2) . D'(\gamma). \tag{36}$$

For σ^2 we can choose either an inverted gamma-2:

$$D'(\sigma^2) = f_{i\gamma}(\sigma^2 \mid s_0^2, \nu_0) \tag{37}$$

or a diffuse prior:

$$D'(\sigma^2) \propto 1/\sigma^2. \tag{38}$$

The prior on γ can be chosen freely as we shall integrate numerically this parameter. For the clarity of the exposition we decide to be non-informative with the restricted uniform prior:

$$D'(\gamma) \propto I_G \tag{39}$$

where $I_G = 1$ if $\gamma \in G$, 0 if not.

– The case of the prior on β needs some discussion. We have the choice between a normal prior conditional on σ^2 as in POIRIER [1987]:

$$D'(\beta \mid \sigma^2) = f_N(\beta \mid \beta_0, \sigma^2 . M_0^{-1}) \tag{40}$$

and an independent Student prior:

$$D'(\beta) = f_t(\beta \mid \beta_0, M_0, 1, \lambda_0). \tag{41}$$

The advantage of formulation (40) over (41) is that it leads to a slightly simpler posterior density for β given γ; its drawback is the dependence of the prior covariance matrix of β on σ^2, since $V(\beta \mid \sigma^2) = \sigma^2 . M_0^{-1}$; this dependence may induce pathological posterior results as explained in BAUWENS [1990] [6]. Apart from the difficulty of choosing an informative prior on σ^2, it would be strange to define such a prior and at the same time to remain non-informative on γ. The distinction between (40) and (41) vanishes if we decide to be non-informative on both β and σ^2.

4.4. Posterior Analysis

It is obvious, from (35), that conditionally on γ, we can apply the standard results of a Bayesian analysis of the linear regression model, as

6. Here, of course, this argument applies conditionally on any value of γ

e. g. in ZELLNER [1971, chapter 3]. The main difficulty is therefore to compute the marginal posterior densities of β and of γ. This has to be done by numerical integration techniques.

The first proposition states the results obtained with the Normal prior (40) on β conditional on σ^2 together with the inverted gamma-2 prior (37) on σ^2 and the diffuse prior (39) on γ.

PROPOSITION 1: A normal prior on β conditional on σ^2 with an inverted gamma-2 prior on σ^2 and a diffuse prior on γ gives the following posterior density:

$$D''(\beta \mid \gamma, y) = f_t(\beta \mid \tilde{\beta}_\gamma, \tilde{M}_\gamma, \tilde{s}_\gamma, \nu_0 + T - k) \tag{42}$$

where:

$$\tilde{\beta}_\gamma = \tilde{M}_\gamma^{-1}(M_0 \beta_0 + X' H_\gamma X \hat{\beta}_\gamma)$$

$$\tilde{M}_\gamma = M_0 + X' H_\gamma X$$

$$\tilde{s}_\gamma^2 = s_0^2 + \hat{s}_\gamma^2 + \beta_0' M_0 \beta_0 + \hat{\beta}_\gamma' X' H_\gamma X \hat{\beta}_\gamma - \tilde{\beta}_\gamma' \tilde{M}_\gamma \tilde{\beta}_\gamma$$

and:

$$D''(\sigma^2 \mid \gamma, y) = f_{i\gamma}(\sigma^2 \mid \tilde{s}_\gamma^2, \nu_0 + T - k) \tag{43}$$

$$D''(\gamma \mid y) \propto |H_\gamma|^{1/2} |\tilde{M}_\gamma|^{-1/2} (\tilde{s}_\gamma^2)^{-(\nu_0 + T - k)/2} . I_G. \tag{44}$$

If the prior on β is diffuse ($M_0 = 0$, $s_0 = 0$ and $\nu_0 = 0$), the posterior mean of β given γ is the weighted least squares (WLS) estimator defined in classical econometrics and the posterior covariance matrix given γ and σ^2 is the corresponding formula for the sampling covariance matrix of the WLS estimator.

Let us now use the independent Student prior on β (41) together with the diffuse priors (38) and (39) on γ and σ^2.

PROPOSITION 2: A Student prior density on β and a non-informative prior on σ^2 and γ give the following posterior density:

$$D''(\beta \mid \gamma, y) \propto [1 + (\beta - \beta_0)' M_0 (\beta - \beta_0)]^{-(\lambda_0 + k)/2} \times [\hat{s}_\gamma^2 + (\beta - \hat{\beta}_\gamma)' X' H_\gamma X (\beta - \hat{\beta}_\gamma)]^{-T/2} \tag{45}$$

$$D''(\sigma^2 \mid \gamma, y) = f_{i\gamma}(\sigma^2 \mid \hat{s}_\gamma^2, T - k) \tag{46}$$

$$D''(\gamma \mid y) \propto |H_\gamma|^{1/2} K_\gamma^{-1} . I_G \tag{47}$$

where K_γ is the normalizing constant defined by the integral in β of the right hand side of (45).

Note that the posterior density of β, conditional on γ but marginal with respect to σ^2, is in the form of a 2-0 poly-t density as defined by DREZE [1977]; the normalizing constant and the moments of such a density can be computed by *univariate* numerical integration, *whatever the dimension of* β, using a technique described by RICHARD and TOMPA [1980].

A non-Bayesian proceeds by conditioning on the ML estimators of the nuisance parameters (or an approximation thereof), while a Bayesian proceeds ideally by marginalizing the conditional posterior density (42) or (45),

or some of its characteristics of interest, with respect to these parameters. This has to be done by numerical integration with respect to γ. Usually, unless γ is a scalar, one will resort to Monte-Carlo integration and will have to find an efficient importance function for γ. As a result of such computations one can get also any desired posterior result about γ or functions of γ.

A Monte-Carlo procedure to find the posterior moments of β and γ can be sketched as follows. The first task is to find an importance function for γ. Two routes can be followed:

– In section 3.4 we devised a test based on an augmented regression which makes an explicit use of $z_t'\gamma$. The marginal posterior density of γ in this regression, multiplied by I_G can be used as an importance function.

– Alternatively we can decide to maximize the kernel of the posterior density of γ on the space defined by G and build a normal or a Student approximation around the posterior mode of γ.

More refined importance functions could be required if the posterior density of γ is strongly asymmetrical. Once we have chosen an importance function, N random drawings of γ noted γ_i are used to compute the marginal posterior moments of γ and β by importance sampling. The weights W_i are defined by the ratio of the kernel of the posterior density of γ to the kernel of the importance function both evaluated at γ_i. The moments of order r of γ are computed as:

$$E(\gamma^r \mid y) \simeq \sum_{i=1}^{N} \gamma_i^r . W_i / \sum_{i=1}^{N} W_i. \tag{48}$$

The marginal moments of β are obtained with formulae linking unconditional moments to conditional moments. So:

$$E(\beta \mid y) = E_\gamma [E(\beta \mid \gamma, y)] \tag{49}$$

which is evaluated by:

$$E(\beta \mid y) \simeq \sum_{i=1}^{N} E(\beta \mid \gamma_i, y) . W_i / \sum_{i=1}^{N} W_i \tag{50}$$

and:

$$V(\beta \mid y) = E_\gamma [V(\beta \mid \gamma, y)] + V_\gamma [E(\beta \mid \gamma, y)]. \tag{51}$$

The first term is evaluated by:

$$E_\gamma [V(\beta \mid \gamma, y)] \simeq \sum_{i=1}^{N} V(\beta \mid \gamma_i, y) . W_i / \sum_{i=1}^{N} W_i \tag{52}$$

and the second by:

$$\begin{aligned} V_\gamma [E(\beta \mid \gamma, y)] &= E_\gamma [E(\beta \mid \gamma, y) . E(\beta \mid \gamma, y)'] - E(\beta \mid y) . E(\beta \mid y)' \\ &\simeq \sum_{i=1}^{N} E(\beta \mid \gamma_t, y) . E(\beta \mid \gamma_i, y)' . W_i / \sum_{i=1}^{N} W_i \\ &\quad - E(\beta \mid y) . E(\beta \mid y)'. \end{aligned} \tag{53}$$

Marginal densities can be computed in a similar fashion. The conditional moments of β given γ have an analytical expression when the prior on β is normal conditional on σ^2 and have to be evaluated numerically when the prior on β is independent Student.

5 Testing for Heteroskedasticity

We consider two ways of testing H_0 (*i.e.* $\gamma=0$) versus H_1 (*i.e.* $\gamma\neq 0$):

- H_0 can be rejected at the level α if 0 does not belong to a $(1-\alpha)\%$ highest posterior density region of the parameter γ indexing the variance (in the spirit of a Wald test in a classical approach);
- a second way to test H_0 is to use the posterior odds ratio (which can be interpreted as the Bayesian counterpart of the likelihood ratio) like in POIRIER [1987]. However this approach raises some issues which are discussed below.

5.1. Posterior Confidence Interval for γ

Of particular interest for forming a judgment on the issue of heteroskedasticity are in the spirit of a Wald test of the restriction $\gamma=0$:

– the expectation and covariance matrix of γ noted respectively $E(\gamma|y)$ and $V(\gamma|y)$,

– the univariate marginal distributions of the elements of γ,

– the joint probability that $\gamma\leqslant 0$ (if 0 is not on the boundary of G),

i.e. informations helpful to judge wether the point 0 is far from or close to the center of the posterior density of γ. We want to get posterior confidence intervals for γ and see if zero belongs to such intervals. As γ is a multivariate random vector it is more convenient to consider the following transformation ξ of γ defined by:

$$\xi=(\gamma-E(\gamma|y))'\,V(\gamma|y)^{-1}(\gamma-E(\gamma|y)) \tag{54}$$

and see if:

$$\xi_*=E(\gamma|y)'\,V(\gamma|y)^{-1}E(\gamma|y) \tag{55}$$

belongs to a posterior confidence interval of ξ. The marginal posterior density of ξ and its fractiles can be obtained directly as a result of the Monte-Carlo integration sketched in section four if we use in (54) and (55) the posterior moments of γ obtained with the auxilary regression (33). Otherwise we have to proceed in two separate rounds and get a first evaluation of these moments with a first Monte Carlo integration and

use a second round of Monte Carlo integration to evaluate the distribution of ξ. If ξ_* is sufficiently far away in the right tail of this distribution, one can reject $H_0 : \gamma = 0$. This procedure will not necessarily provide a HPD (highest posterior density) region, unless the marginal posterior density of γ is symmetrical around $E(\gamma | y)$.

5.2. Posterior Odds

Posterior odds ratios are perhaps the most traditional way of comparing hypotheses in the Bayesian setup. They are recommended by BERGER and DELAMPADY [1987] as an indispensible ingredient of Bayesian cooking, [7] especially for testing a precise hypothesis, like $\gamma = 0$ in our case. POIRIER [1987] derives the posterior odds ratio for essentially the same hypothesis as ours and suggests efficient methods to compute it.

In traditional Bayesian testing or comparing hypotheses, a prior point probability is affected to an hypothesis, say:

$$\Pr(H_0) = \pi \qquad \text{and} \qquad \Pr(H_1) = 1 - \pi.$$

The posterior probabilities of H_0 and H_1 are then evaluated given the observed sample by a direct application of Bayes theorem:

$$\Pr(H_0 | y) = l(y | H_0) . \pi / [l(y | H_0) . \pi + l(y | H_1) . (1 - \pi)]$$
$$\Pr(H_1 | y) = 1 - \Pr(H_0 | y).$$

The Bayesian decision rule which minimizes the expected loss is to choose in favour of the hypothesis which has the highest posterior probability, under a symmetric loss function (*see* ZELLNER [1971, chap. 11]).

In an alternative presentation, the ratio:

$$\Pr(H_0) / \Pr(H_1)$$

is called the prior odds ratio. Its posterior counterpart is the posterior odds ratio defined by:

$$\frac{\Pr(H_0 | y)}{\Pr(H_1 | y)} = \frac{\Pr(H_0)}{\Pr(H_1)} \cdot \frac{l(y | H_0)}{l(y | H_1)} = \frac{\Pr(H_0)}{\Pr(H_1)} \cdot B$$

where B is called the Bayes factor. The latter is equal to the ratio of the likelihood of the sample under H_0 and under H_1. An attractive decision rule is to favour H_0 if the posterior odds ratio is greater than the prior odds ratio. It is equivalent to favouring H_0 if the Bayes factor is greater than one. The Bayes factor provides the evidence contained in the data in

---

7. In fact, these authors recommend to report the Bayes factor, which is easily recovered from the posterior odds ratio (since the latter is equal to the prior odds ratio times the Bayes factor), and a posterior confidence region C (HPD region): "The Bayes factor communicates the evidence in the data against θ_0 (the null), and C indicates the magnitude of the possible discrepancy of θ from θ_0" (p. 328).

favour of H_0 with respect to H_1 since it is the ratio of likelihood function of the observed sample given each hypothesis. This decision rule is a somewhat likelihooder approach as it is based on the Bayes factor alone. It coincides with the Bayesian decision rule if the prior odds is equal to one.

The computation of the Bayes factor is not anodyne. When the null hypothesis is composite the likelihood function has to be marginalized using a prior density (possibly non-informative) in order to get rid of the parameters which are not the object of the test. For instance in a model parametrized in (γ, θ), if we want to test $\gamma = 0$, whatever the value of θ, this parameter has to be integrated out. Moreover if the alternative hypothesis is $\gamma \neq 0$ the prior density on γ must be proper for integrating this parameter under H_1. The Bayes factor is computed as:

$$B = \frac{\int l(y \mid \gamma = 0, \theta) . D'(\theta) . d\theta}{\int l(y \mid \gamma, \theta) . D'(\theta) . D'(\gamma) . d\theta . d\gamma}$$

Let us discuss the specification of the prior odds. We assign the prior probability π to the hypothesis $\gamma = 0$ and $1 - \pi$ to $\gamma \neq 0$. This has to be coherent with the prior information given previously on γ with $D'(\gamma)$. This means that a positive probability is given to a point. The remaining probability has to be spread over G in accordance with $D'(\gamma)$. So the complete prior on γ is expressed as:

$$D'_1(\gamma) = \pi . 1_{(\gamma = 0)} + (1 - \pi) . D'(\gamma) . 1_{(\gamma \neq 0)} \tag{56}$$

where $0 < \pi < 1$, $1_{(\gamma = 0)}$ is equal to 1 if $\gamma = 0$, and to 0 if $\gamma \neq 0$, while $1_{(\gamma \neq 0)} = 1 - 1_{(\gamma = 0)}$.

The Bayes factor can easily be recovered, as the kernel of the marginal likelihood function has already been calculated in propositions 1 and 2, and is implicitely given in (44) and (47) for different prior densities on (β, σ^2). For a normal gamma prior on (β, σ^2) we have:

$$l(y \mid \gamma) \propto |H_\gamma|^{1/2} |\tilde{M}_\gamma|^{-1/2} (\tilde{s}^2_\gamma)^{-(v_0 + T - k)/2} \tag{57}$$

and the corresponding Bayes factor is:

$$B = \frac{|H_{\gamma=0}|^{1/2} |\tilde{M}_{\gamma=0}|^{-1/2} (\tilde{s}^2_{\gamma=0})^{-(v_0 + \tau - k)/2}}{\int |H_\gamma|^{1/2} |\tilde{M}_\gamma|^{-1/2} (\tilde{s}^2_\gamma)^{-(v_0 + T - k)/2} . D'(\gamma) . d\gamma} \tag{58}$$

The integral of the denominator is obtained as a by-product of the evaluation of the posterior moments of γ in (48). Provided that the importance function for γ is properly normalized as a density, then:

$$\int l(y \mid \gamma) . D'(\gamma) . d\gamma \simeq \frac{1}{N} \sum_{i=1}^{N} W_i .$$

An important remark is to be done. Inference on γ and the computation of ξ_* in (55) have been done using the smooth prior $D'(\gamma)$ while the posterior odds ratio required implicitly the use of the more complicated prior $D_1'(\gamma)$ which assigns a positive probability to the single point $\gamma=0$. Therefore the two test procedures are not strictly comparable. If we desire to render them compatible we have to alter the posterior density of γ into:

$$(59) \quad D''(\gamma|y) \propto \pi . l(y|\gamma=0) . 1_{(\gamma=0)} + (1-\pi) . l(y|\gamma) . D'(\gamma) . 1_{(\gamma \neq 0)}$$

and to compute ξ_* from the latter, rather than from (44) or (47).

A major difficulty of the posterior odds approach is the need to assign a positive probability π to a single point ($\gamma=0$). How can one choose the value of π? A sensitivity analysis with respect to the value of π should at least be performed for scientific reporting. BERGER and DELAMPADY [1987] recommend moreover to compute upper bounds for the Bayes factor, *i.e.* to compute the largest possible value of the Bayes factor with respect to a wide class of prior densities for γ. As a final comment, we can express our point of view on the posterior odds approach by saying that it seems better suited for testing non-nested hypotheses than nested hypotheses. The reason is fundamentally that this methodology permits to treat symmetrically H_0 and H_1.

6 Conclusion

In this paper we have presented two main broad strategies for detecting heteroskedasticity. Facing an empirical problem which one should we use? Clearly they do not correspond to the same spirit. The first kind of diagnostics should be used for detecting a possible misspecification due to heteroskedasticity inside a Bayesian modelling strategy à la Hendry-Richard-Sargan. The second approach developed in the paper is completely different. The model of interest is recognized from the start as heteroskedastic. This can occur for instance in microeconomic surveys. Inside this framework the null hypothesis of homoskedasticity can be tested as can be tested for instance the equilibrium hypothesis inside a disequilibrium model.

Probability distributions

We have used in the text several probability distributions. It is useful to recall here their notation and properties.

Inverted Gamma-2 Distribution

A random variable σ^2 in $\mathbb{R}_+$ has an inverted gamma-2 distribution with scale parameter s and ν degrees of freedom if its probability density is defined by:

$$f_{i\gamma}(\sigma^2 | s, \nu) = k_1^{-1} . \left(\frac{s}{2}\right)^{\nu/2} . (\sigma^2)^{-(\nu+2)/2} . \exp - \frac{s}{2\sigma^2}$$

with $k_1 = \Gamma(\nu/2)$ and ν, $s > 0$. The first two moments of σ^2 are given by:

$$E(\sigma^2) = \frac{s}{(\nu-2)} \quad \text{provided } \nu > 2$$

$$\text{Var}(\sigma^2) = \frac{2s^2}{(\nu-4)(\nu-2)^2} \quad \text{provided } \nu > 4.$$

Multivariate Normal Distribution

A random vector y of $\mathbb{R}^k$ has a multivariate normal distribution with mean μ and variance Σ if its probability density function is given by:

$$f_N(y | \mu, \Sigma) = k_2^{-1} . |\Sigma|^{-1/2} . \exp - \frac{1}{2} (y-\mu)' \Sigma^{-1} (y-\mu)$$

with $k_2 = (2\pi)^{k/2}$, $\mu \in \mathbb{R}^k$ and $\Sigma \in \mathscr{C}^k$ where $\mathscr{C}^k$ is the cone of positive definite matrices.

Univariate Student Density

A random variable x of $\mathbb{R}$ has a univariate Student distribution if its probability density is defined by:

$$f_t(x | \mu, h, s, \nu) \propto [s + h . (x-\mu)^2]^{-(\nu+1)/2}$$

(*see* next paragraph for normalizing constants and moments). This notation of the univariate Student density is convenient for posterior analysis in regression models, but is not very often found in textbooks. It can be rewritten into:

$$f_t(x | \mu, h, s, \nu) \propto \left[1 + \frac{h}{s} . (x-\mu)^2\right]^{-(\nu+1)/2}$$

or equivalently $x \sim t(\mu, h, s, \nu)$.

The usual notation for the standardized Student density is:

$$f_t(t \mid \nu) \propto \left[1 + \frac{t^2}{\nu}\right]^{-(\nu+1)/2}$$

or equivalently $t \sim t(\nu)$.

In order to standardize the initial density we have to operate the following change of variable:

$$t = (x - \mu) . (h\nu/s)^{1/2}.$$

A confidence interval at a $1-\alpha$ level is obtained as:

$$\mu \pm t_\alpha . (s/h\nu)^{1/2}$$

where t_α is the fractile of the Student distribution such that $\Pr(-t_\alpha < t < t_\alpha) = 1 - \alpha$.

Multivariate Student Distribution

A random vector y of $\mathbb{R}^k$ has a multivariate Student distribution with parameters μ, M, s and ν degrees of freedom if its probability density function is defined by:

$$f_t(y \mid \mu, M, s, \nu) = k_3^{-1} . |M|^{1/2} . s^{\nu/2} . [s + (y-\mu)' M (y-\mu)]^{-(\nu+k)/2}$$

with $k_3 = \pi^{k/2} . \Gamma(\nu/2)/\Gamma((\nu+k)/2)$, s, $\nu > 0$ and $M \in \mathscr{C}^k$. The first two moments are:

$$E(y) = \mu \quad \text{provided that } \nu > 1$$

$$\text{Var}(y) = \frac{s}{\nu - 2} . M^{-1} \quad \text{provided that } \nu > 2.$$

This density can be viewed as an infinite mixture of normal densities mixted by an inverted gamma-2. We have:

$$f_t(y \mid \mu, M, s, \nu) = \int f_N(y \mid \mu, M^{-1} . \sigma^2) . f_{i\gamma}(\sigma^2 \mid s, \nu) . d\sigma^2.$$

• References

BAUWENS, L. (1990). – "The Pathology of the Natural Conjugate Prior Density in the Regression Model", *GREQE Discussion Paper*, 90-A-14.

BERGER, J. O. (1988). – *Statistical Decision Theory and Bayesian Analysis*, Springer Verlag, New York.

BERGER, J. O. and DELAMPADY, M. (1987). – "Testing Precise Hypothesis", *Statistical Science*, 2, pp. 317-352.

BOX, G.E.P. and TIAO, C. G. (1962). – "A Further Look at Robustness via Bayes Theorem", *Biometrica*, 49, pp. 419-433.

BREUSCH, T. S. and PAGAN, A. R. (1979). – "A Simple Test for Heteroskedasticity and Random Coefficient Variation", *Econometrica*, 47, pp. 1287-1294.

CHALONER, K. and BRUNST, R. (1986). – "A Bayesian Approach to Residual Analysis and Outlier Detection", *University of Minnesota, Dept. of Applied Statistics*, Technical report 480.

CHESHER, A. (1983). – "The Information Matrix Test: Simplified Calculation via a Score Test Interpretation", *Economics Letters*, 13, pp. 45-48.

CHESHER, A. (1984). – "Testing for Neglected Heterogeneity", *Econometrica*, 52, pp. 865-872.

DAVIDSON, R. and MACKINNON, J. G. (1988). – "A New Form of the Information Matrix Test", *GREQE Discussion Paper*, 89-A-06.

DREZE, J. H. (1977). – "Bayesian Regression Analysis Using Poly-T Densities", *Journal of Econometrics*, 6, pp. 329-354.

ENGLE, R. (1982). – "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50, pp. 987-1008.

GEWEKE, J. (1986). – "Exact Inference in Models with Autoregressive Conditional Heteroscedasticity", in E. BERNDT, H. WHITE and W. BARNETT, Eds., *Dynamic Econometric Modelling*, Cambridge University Press, Cambridge.

GEWEKE, J. (1989). – "Exact Predictive Densities for Linear Models with ARCH Disturbances", *Journal of Econometrics*, 40, pp. 63-86.

GRIFFITHS, W. E., DRYNAN, R. G. and PRAKASH, S. (1979). – "Bayesian Estimation of a Random coefficient model", *Journal of Econometrics*, 10, pp. 201-220.

PAGAN, R. A. (1984). – "Model Evaluation by Variable Addition", in D. F. HENDRY and K. F. WALLIS, Eds., *Econometrics and Quantitative Economics*, Blackwell, Oxford.

POIRIER, D. J. (1987). – "Bayesian Diagnostic Testing in the General Linear Normal Regression Model", Paper presented at the third Valencia International Meeting on Bayesian Statistics in Altea, Spain.

RICHARD, J. F. and TOMPA, H. (1980). – "On the Evaluation of Poly-T Density Functions", *Journal of Econometrics*, 12, pp. 335-351.

WHITE, H. (1980). – "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48, pp. 817-838.

WHITE, H. (1982). – "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50, pp. 1-25.

ZELLNER, A. (1971). – *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

ZELLNER, A. (1975). – "Bayesian Analysis of Regression Error Term", *Journal of the American Statistical Association*, 70, pp. 138-144.

ZELLNER, A. and MOULTON B. R. (1985). – "Bayesian Regression Diagnostics with Application to International Consumption and Income Data", *Journal of Econometrics*, 29, pp. 187-211.