CONSISTENCY TESTS FOR HETEROSKEDASTIC AND RISK MODELS

Adrian R. Pagan
University of Rochester

Hernán Sabau
Operadora de Bolsa

Resumen: Este artículo presenta una clase de pruebas de consistencia en la especificación de modelos heteroscedásticos y de riesgo. Las pruebas están relacionadas con otras ya conocidas, tales como las pruebas de momentos de Newey y Tauchen, las de Hausman, las de White, las de multiplicadores de Lagrange de Engle y Pagan, y el análisis de residuos de Pagan y Hall. Se analiza la potencia de las pruebas de consistencia en presencia de desviaciones locales y se reexamina el modelo de Engle, Lilien y Robins.

Abstract: This paper considers a class of consistency tests for the specification of heteroskedastic and risk models. The tests are related to other procedures such as the conditional moment tests of Newey and Tauchen, Hausman's tests, White's tests, the variable addition Lagrange multiplier tests of Engle and Pagan, and the residual analysis of Pagan and Hall. The power of the consistency tests in the presence of local departures is analyzed and the risk premia model of Engle, Lilien and Robins is re-assessed.

1. Introduction

In a previous paper (Pagan and Sabau, 1991) we have analyzed the consistency of the MLE for the parameters of heteroskedastic models of the form

$$y_t \mid F_t \sim \mathrm{N}[x_t\beta,\ h_t(\beta,\alpha) = h_t(\theta)]\ , \tag{1}$$

where the conditioning information set is the o–field generated by $\{x^*_{t-j}, y_{t-j-1}\}_{j=0}^{\infty}$, the x^*_{t-j} being those elements of x_t which are weakly exogenous variables for the parameter vector θ in the sense of Engle *et al.* (1983); the $1 \times k$ vector x_t and the scalar h_t are measurable functions of F_t, and the parameter vector $\theta = (\beta', \alpha')'$ is $p \times 1$, with β being $k \times 1$ and α being $(p-k) \times 1$.

Implicitly assuming that the parameters of interest are functions of β, our attention was centered on the robustness of the MLE $\hat{\beta}$, say, to misspecifications in the conditional variance h_t, extending the work of Carroll and Ruppert (1982). By obtaining a necessary condition for consistency using Domowitz and White's (1982) quasi-ML approach, we concluded that misspecification of h_t will in general induce inconsistency in $\hat{\beta}$. An exception is, of course, the case when h_t is not parameterized in terms of β, in which case GLS and ML are equivalent. When h_t is proportional to $(x_t\beta)^2$ as in Amemiya (1973), or for the Poisson-type model ($h_t = x_t\beta$) used in count data (Hausman *et al.*, 1984), very unlikely conditions would need to be met to avoid inconsistency.

A particularly interesting case occurs when an ARCH regression (Engle, 1982a), or more generally a GARCH regression (Bollerslev, 1986) model has been estimated, but h_t does not follow such a process. Then we found that, if the innovations $u_t = y_t - x_t\beta$ are symmetrically distributed, the consistency of $\hat{\beta}$ hinges on whether h_t is an odd or even function of u_t conditional upon $F_t^u = \{u_{t-j}\}_{j=1}^{\infty}$ (σ– field). Therefore $\hat{\beta}$ was robust to misspecification of the orders of the GARCH process, or to misspecification within the class of symmetric ARCH processes (Engle, 1982a, see also Geweke, 1986, and Engle and Bollerslev, 1986) or when the true conditional distribution is Strudent's t as suggested by the latter authors. However, when the true conditional variance is the sum of a conditionally odd and a conditionally even terms in u_t, it is found that $\hat{\beta}$ will generally be inconsistent. This also happens when the innovations are not symmetrically distributed which emphasizes the need for a careful assessment of the normality assumption, a task undertaken in Sabau (1987b).

If a risk term, defined as a function of the conditional variance, affects the conditional mean as in the ARCH-M model (Domowitz and Hakkio, 1985, Engle *et al.*, 1987) inconsistency will result because misspecification of the conditional variance now implies misspecification of the conditional mean. In our previous paper we did not explicitly consider this type of model since the conclusion is an obvious extension of standard misspecification analysis of regression functions.

The situation recounted above suggests that applied workers using parametric specifications of h_t need to be concerned that any failure to specify h_t correctly may contaminate the estimates of β, which is what they are normally primarily interested in. One possibility is to develop robust estimation procedures, particularly GLS, which remain consistent in the presence of misspecified heteroskedasticity. In fact, there are now semiparametric GLS estimators that are asymptotically as efficient as parametric GLS yet presume no knowledge of the heteroskedasticity (Carroll, 1982, Robinson, 1987, Newey, 1986). This approach does not seem yet a favored one, however, and it is hard to see how it would be done when parameters in the conditional variance feed back into the conditional mean, as in the ARCH-M model, because the information matrix can not be diagonalized as GLS procedures demand. Consequently, if researchers are going to continue to use MLE's in the models described above, it is important that some information be provided along with the MLE's to assess the validity of the chosen specification.

There are, of course, many diagnostic tests for incorrect specifications in the conditional mean and variances of the general linear model –Pagan (1984a) surveys these– and all might be applied. But it is always useful to have some tests that utilize information that is available solely within the estimated model, since this can be provided with computer output. For this reason in section 2 we argue for what we call a set of "consistency tests". Simple examples are that the MLE residuals sum to zero and that the postulated heteroskedastic pattern conforms to the evidence in the squared residuals. These are functions of the output of any MLE program and so could be supplied along with standard results from it. Also in section 2 we relate these "consistency tests" to other procedures in the literature. Section 3 of the paper finds the asymptotic distribution of the proposed test-statistics under the null hypothesis of correct specification and under sequences of local parametric alternatives. In section 4 we make local power comparisons and it is shown that the power of the simpler tests depends upon the extent of the incosistency of the MLE of the conditional mean parameters, so that the tests directly address what it is about the MLE which concerns us. In section 5 we use consistency tests to re-assess the adequacy of the ARCH parameterization used by Engle, Lilien and Robins (1987) in their model for risk premia in the term structure of interest rates, and find some evidence of specification error. Concluding remarks are given in section 6.

2. Consistency Tests of the Heteroskedastic Formulation

The possibility that the MLE of β might inconsistently estimate β_0, the true value of β, motivates the development of tests for the adequacy of the assumed heteroskedastic specification. To appreciate our choices consider the score for θ (normalized by the sample size T), say $d_\theta(\theta) = (d_\beta(\theta)', d_\alpha(\theta)')'$, given by (e.g. Engle, 1982a)

$$d_\beta(\theta) = T^{-1}\sum_{t=1}^{T} h_t^{-1} x_t' u_t + \frac{1}{2} T^{-1} \sum_{t=1}^{T} h_t^{-2} w_t' \varepsilon_t \,, \tag{2a}$$

and

$$d_\alpha(\theta) = \frac{1}{2} T^{-1} \sum_{t=1}^{T} h_t^{-2} z_t' \varepsilon_t \,, \tag{2b}$$

where $w_t = \partial h_t / \partial \beta'$, $z_t = \partial h_t / \partial \alpha'$; and $u_t = y_t - x_t \beta$ and $\varepsilon_t = u_t^2 - h_t$ are the innovations of the conditional mean and conditional variance equations, respectively.

Under the conditions set out in Domowitz and White (1982), $E\,[d_\theta(\theta^*)[= 0$ defines the pseudo-true value $\theta^* = (\beta^{*\prime}, \alpha^{*\prime})'$ of θ (i.e. $\hat{\theta} \overset{as}{\rightarrow} \theta^*$). Consistency obtains when, and only when, $0^* = 0_0$. The necessary and sufficient condition for the consistency of β is $E\,[d_\beta(\beta_0, \alpha^*)] = 0$ (Lemma 2 of Pagan and Sabau, 1991). Unfortunately, $d_\beta(\hat{\theta}) = 0$ by construction and no test can be based on it. However, from (2a), sufficient conditions for $E\,[d_\beta(\beta_0, \alpha^*)] = 0$ are

$$E\,[T^{-1}\sum_{t=1}^{T} h_t^{-1} x_t' u_t] = 0 \,, \quad E\,[T^{-1}\sum_{t=1}^{T} h_t^{-2} w_t' \varepsilon_t] = 0$$

making it appropriate that tests for the adequacy of the assumed model be based upon these separate orthogonality conditions. Note that the MLE of β does not impose any of these restrictions on the data and separate estimators of β may be obtained from each of them (GLS estimators) with the MLE of β being the matrix weighted average of the two (Sabau, 1987a).

At a more basic level, correct specification of the two conditional moments requires $E\,[u_t \mid F_t\,] = 0$, and $E\,[\varepsilon_t \mid F_t\,] = 0$, implying that

$$E\,[\sum_{t=1}^{T} u_t] = 0 \,, \quad E\,[\sum_{t=1}^{T} \varepsilon_t] = 0 \,,$$

and the simplicity of the latter constructs makes $\sum_{t=1}^{T} u_t$ and $\sum_{t=1}^{T} \varepsilon_t$ attractive of the adequacy of the model. We will refer to these as "consistency" tests, due to the fact that no internal information will be needed in their construction, unlike most diagnostic tests which introduce new data supplied by postulating alternative models. Replacing the unknowns β and α by their MLE's, the sample quantity $T^{-1}\sum_{t=1}^{T} \hat{\varepsilon}_t$ for example should tend in probability to zero under the null hypothesis; if it does not, the evidence in the squared residuals $\hat{u}_t^2$ is inconsistent with that in $\hat{h}$.

For later use it is convenient to embed the tests described above in a wider class derived from sets of n first order conditions of the form

$$m(\Phi\,,\theta) = T^{-1}\sum_{t=1}^{T} m_t(\Phi_t\,,\theta) = T^{-1}\sum_{t=1}^{T}\Phi_t u_t = T^{-1}\Phi'\upsilon\ , \tag{3}$$

where $u_t = (u_t\,, \varepsilon_t)'$, $\upsilon = (\upsilon_1',...,\upsilon_T')'$, $\Phi = (\Phi_1,...,\Phi_T)'$, and the Φ_t are $n \times 2$ matrices of measurable functions of F_t. We assume the Φ_t and υ_t to obey the regularity, continuity, dominance and mixing conditions in assumptions (3) - (6) of Newey (1985b). The conditional covariance matrix of u_t is $\Omega_t = \text{diag}\,\{\,h_t\,, 2h_t^2\,\}$. Thus the $m_t(\Phi_t\,,\theta)$ are $n \times 1$ vectors and under correct specification we have

$$E\,[m(\Phi\,,\theta)] = 0\ , \tag{4}$$

because $E\,[m_t \mid F_t] = \Phi_t E\,[\upsilon_t \mid F_t] = 0$. The corresponding covariance matrix is

$$V[m(\Phi\,,\theta)] = E\,[T^{-1}\Phi'\Omega\Phi]\ , \tag{5}$$

where $\Omega = \text{diag}\,\{\,\Omega_t\,\}$, because $E\,[\Phi_t\upsilon_t\upsilon_s'\Phi_s' \mid F_t] = \delta_{ts}\Phi_t\,\Omega_t\Phi_t'$ with δ_{ts} being the Kronecker delta. The two simple consistency tests described above are obtained by setting $\Phi_t = (1\,,0)\ \ \forall\,t$ and $\Phi_t = (0\,,1)\ \forall\,t$ respectively, and a joint consistency test is obtained making $\Phi_t = I_2\ \ \forall\,t$.

Corresponding to the theoretical moments in (4) are sample moments $m(\hat{\Phi}\,,\hat{\theta})$ and these are suitable "consistency test-statistics", since they would tend to be close to zero if the model specification was adequate. Our major problem with implementing these tests lies in the derivation of the limiting distribution of $T^{1/2}m(\hat{\Phi}\,,\hat{\theta})$. Because $\hat{\theta}$ is obtained as the solution to a set of first order conditions, $d_\theta(\hat{\theta}) = 0$, it is apparent that $m(\Phi\,,\hat{\theta})$ are conditional moment restrictions of the sort analyzed in Tauchen (1985) and Newey (1985a, b). The last paper is particularly relevant here and Theorem 1 in the next section is extracted from it.

But before we obtain the distribution of the sample moments $m(\hat{\Phi}\,,\hat{\theta})$, it is useful to look at these tests in somewhat greater detail and to relate them to other procedures in the literature in order to put them into the proper perspective.

The simple consistency tests based on the sum of ML residuals, say m_μ, is an interesting one because many diagnostic tests for specification error (e.g. RESET) arose because such a criterion was not available in the general linear model since the sum of OLS residuals is identically zero whenever an intercept appears among the regressors. It does not seem to have been fully appreciated in the literature that the residuals defined by other estimators need not share this property. Suppose the OLS estimator of β in (1) is $\tilde{\beta}$, $\tilde{u}_t$ are the corresponding residuals. Since $\sum_{t=1}^{T} u_t = 0$ we have that

$$m_\mu = T^{-1}\sum_{t=1}^{T}\hat{u}_t - T^{-1}\sum_{t=1}^{T}\tilde{u}_t = T^{-1}\sum_{t=1}^{T} x_t(\tilde{\beta}-\hat{\beta}) = \bar{x}(\tilde{\beta}-\hat{\beta}) , \tag{6}$$

and consequently the consistency test can be viewed as a specific weighted average of the difference between the ML and OLS estimators, when the weights are the sample means of the regressors. Since $\tilde{\beta}$ is consistent irrespective of the specification of the variance, the test-statistic focuses directly upon the inconsistency in the MLE of β.

An alternative strategy for assessing the adequacy of the maintained model would be to directly compare $\hat{\beta}$ and $\tilde{\beta}$, i.e. to conduct a specification test of the form given by Hausman (1978). In fact this was White's (1980) suggestion, and it can be regarded as a special case of the test-statistics considered in this section viz. when $\Phi_t = (x_t', 0)$. To see this, observe that $m(X, \hat{\beta}) = T^{-1}\sum_{t=1}^{T} x_t'\hat{u}_t = T^{-1}X'\hat{u}$. But $T^{-1}X'\tilde{u} = 0$ and therefore

$$m(X, \hat{\beta}) = T^{-1}X'\hat{u} = T^{-1}X'(\hat{u}-\tilde{u}) = T^{-1}X'X(\tilde{\beta}-\hat{\beta}) , \tag{7}$$

so that $m(X, \hat{\beta})$ is a nonsingular transformation of $(\tilde{\beta}-\hat{\beta})$ and hence has the same distribution.

More generally, suppose that the orthogonality conditions

$$T^{-1}\sum_{t=1}^{T}\Phi_t^*\tilde{\upsilon}_t = 0$$

define the consistent estimator $\tilde{\theta}$ where Φ_t^* includes Φ_t as a submatrix i.e. a general method of moments (GMM) estimator as in Hansen (1982). Using the Mean Value Theorem for Random Functions (Jennrich, 1969), after expansion around θ_0 and grouping terms of $O_p(T^{-1})$ and smaller, we get

$$m(\hat{\Phi}, \hat{\theta}) = T^{-1}\sum_{t=1}^{T}\Phi_t\hat{\upsilon}_t = T^{-1}\sum_{t=1}^{T}\Phi_t\upsilon_t + T^{-1}\sum_{t=1}^{T}\Phi_t\frac{\partial g_t(\theta_0)'}{\partial\theta'}(\hat{\theta}-\theta_0)$$

$$+ A_T(\hat{\theta}-\theta_0) + O_p(T^{-1})$$

where $g_t = g_t(\theta) = (x_t\beta \, , h_t)$, and $A_T = T^{-1}\sum_{t=1}^{T}(\upsilon_t' \otimes I_n)\partial \mathrm{vec}\Phi_t / \partial\theta'$. Note that $A_T \xrightarrow{as} 0$ because $\partial\mathrm{vec}\Phi_t/\partial\theta'$ is measurable function of F_t. Therefore the term $A_T(\hat{\theta} - \theta_0)$ is $O_p(T^{-1})$ and can be relegated to the remainder. The same expansion can be applied to $m(\tilde{\Phi} \, , \tilde{\theta})$which equals zero from the GMM definition of $\tilde{\theta}$. Then under the assumption of a correctly specified model we have

$$m(\hat{\Phi} \, , \hat{\theta}) = m(\hat{\Phi} \, , \hat{\theta}) - m(\tilde{\Phi} \, , \tilde{\theta}) = T^{-1}\sum_{t=1}^{T}\Phi_t\hat{u}_t - T^{-1}\Sigma\tilde{\Phi}_t\tilde{u}_t$$

$$= T^{-1}\sum_{t=1}^{T}\Phi_t \frac{\partial g_t(\theta_0)'}{\partial\theta'}(\tilde{\theta} - \hat{\theta}) + O_p(T^{-1})$$

$$= T^{-1}\Phi' G_\theta(\tilde{\theta} - \hat{\theta}) + O_p(T^{-1}) \tag{8}$$

where $g = (g_1,\ldots,g_T)'$, and $G_\theta = \partial g / \partial\theta'$. Note that the relation will be exact if Φ_t is not a function of 0 and the relevant components of g_t are linear in 0, as in the cases above with a linear mean and the test involving only the mean innovations through $\Phi_t = (1 \, , 0)$ and $\Phi_t = (x_t' \, , 0)$.

Thus whenever $\Phi^* = \Phi$, so that $T^{-1}\Phi' G_\theta$ converges to a nonsingular matrix (a regularity condition for the GMM procedure), we have a Hausman test asymptotically equivalent to our test in (3). When Φ is a proper submatrix of Φ^* the consistency test being based asymptotically on linear combinations of $\tilde{0} - \hat{\theta}$.

Another special case of (8) of particular interest is when $\Phi_t^* = \Phi_t = (h_t^{-1}x_t' \, , 0)$. Then $T^{-1}\Phi' u = 0$ defines the GLS estimator of β in (1), say $\tilde{\beta}_g$, and we have

$$T^{-1}\sum_{t=1}^{T}h_t^{-1}x_t'\hat{u}_t - T^{-1}\sum_{t=1}^{T}h_t^{-1}x_t'\tilde{u}_{gt}$$

$$= (T^{-1}\sum_{t=1}^{T}h_t^{-1}x_t'x_t)(\tilde{\beta}_g - \hat{\beta}) + O_p(T^{-1}) \tag{9}$$

whichalso relates to Hausman's (1978) procedure but now with GLS as the consistent estimator under the alternative. If there is a constant among the regressors then $T^{-1}\sum_{t=1}^{T}h_t^{-1}\tilde{u}_{gt} = 0$, and redefining $\Phi_t = (h_t^{-1} \, , 0)$ we get

$$T^{-1}\sum_{t=1}^{T}\hat{h}_t^{-1}\hat{u}_t - T^{-1}\sum_{t=1}^{T}\tilde{h}_t^{-1}\tilde{u}_{gt} = T^{-1}\sum_{t=1}^{T}h_t^{-1}x_t(\tilde{\beta}_g - \hat{\beta}) + O_p(T^{-1}) \, , \tag{10}$$

which again centers on the inconsistency of $\hat{\beta}$, since $\tilde{\beta}_g$ is robust to variance misspecification. Here the differences of estimators are weighted by the sample mean of $h_t^{-1}x_t$.

In the above illustrations, $\tilde{\beta}$ came from a set of first order conditions involving u_t only, and no attention was paid to the second orthogonality condition involving ε_t. This meant that consistency tests related to $\tilde{\beta}-\hat{\beta}$. Similar relations can be found for the α parameters. For simplicity, take the case where $h_t = z_t(\beta)\alpha$, which is the most common in aplied work. Let the simple least squares estimator (SLS) of α obtained from regressing $\tilde{u}_t^2$ on $z_t(\tilde{\beta})$ be $\tilde{\alpha}$. The residuals from this regression are $\tilde{\varepsilon}_t = \tilde{u}_t^2 - z_t(\tilde{\beta})\tilde{\alpha}$. Assuming a constant in z_t we have $T^{-1}\sum_{t=1}^{T}\tilde{\varepsilon}_t = 0$ and, letting $\Phi_t = (0\,,1)$ we have for $m_h = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t$,

$$m_h = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t - T^{-1}\sum_{t=1}^{T}\varepsilon_t$$

$$= T^{-1}\sum_{t=1}^{T}z_t(\tilde{\alpha}-\hat{\alpha}) + O_p(T^{-1})$$

$$= \bar{z}(\tilde{\alpha}-\hat{\alpha}) + O_p(T^{-1}). \tag{11}$$

When $\Phi_t = (0\,, \tilde{z}_t')$, then

$$m(Z\,,\hat{\alpha}) = T^{-1}\tilde{Z}'\hat{\varepsilon} = T^{-1}Z'(\hat{\varepsilon}-\tilde{\varepsilon}) + O_p(T^{-1})$$

$$= T^{-1}Z'Z(\tilde{\alpha}-\hat{\alpha}) + O_p(T^{-1}), \tag{12}$$

which are the counterparts for α of the weighted and full difference of OLS and ML estimators in (6) and (7). Note that in the expansions for (11) and (12) we have taken β as given because we want to concentrate on the α parameters. Substituting any consistent estimator of β in the construction of $\bar{z}$ or $T^{-1}Z'Z$ will have no effect on the asymptotic distribution of the statistics.

We can also consider the simple GLS (SGLS) estimator for α, say $\tilde{\alpha}_g$, obtained from the regression of $\tilde{u}_t^2$ on $z_t(\tilde{\beta})$ in the metric of $\tilde{h}_t^2 = (\tilde{z}_t\alpha)^2$, so that with $\Phi_t = (0\,, h_t^{-2})$ and $\Phi_t = (0\,, h_t^{-2}z_t')$ we get, respectively,

$$T^{-1}\sum_{t=1}^{T}h_t^{-2}\hat{\varepsilon}_t - T^{-1}\sum_{t=1}^{T}\tilde{h}_t^{-2}\tilde{\varepsilon}_{gt} = T^{-1}\sum_{t=1}^{T}h_t^{-2}z_t(\tilde{\alpha}_g-\hat{\alpha}) + O_p(T^{-1})\,, \tag{13}$$

and

$$T^{-1}\sum_{t=1}^{T}\hat{h}_t^{-2}\hat{z}_t'\hat{\varepsilon}_t - T^{-1}\sum_{t=1}^{T}\tilde{h}_t^{-2}\tilde{z}_t'\tilde{\varepsilon}_{gt} = (T^{-1}\sum_{t=1}^{T}h_t^{-2}z_t'z_t)(\tilde{\alpha}_g-\hat{\alpha}) + O_p(T^{-1})\,, \tag{14}$$

as counterparts for α of (10) and (9). It is also easy to combine in a joint form like (8) mean and variance consistency statistics like (6) and (11), (7) and (12), (10) and (13), and (9) and (14).

The consistency tests also arise as a natural application of residual analysis to diagnose the model (Pagan and Hall, 1983). This interpretation is considered later in section 5. Here let us take the closely related issue of variable addition (Pagan, 1984a). When variable addition tests are performed using the LM principle (e.g. Breusch and Pagan, 1980, Engle, 1982b, 1984) they take the form of the general consistency statistic in (3). Suppose a set of variables, say x_{At}, is added to the conditional mean in (1) with coefficients β_A, and the conditional variance is augmented to $h_t(\theta, \theta_A)$. The additions in h_t may be reflecting autonomous changes in the conditional variance specification as well as changes implied by the modification in the conditional mean. Therefore, θ_A includes β_A. It is clear from (2) that the subvector of the score for θ_A is

$$d_A(\theta, \theta_A) = T^{-1}\sum_{t=1}^{T} h_t(\theta, \theta_A)^{-1}\bar{x}_{At}' u_{At} + \frac{1}{2}T^{-1}\sum_{t=1}^{T} h_t(\theta, \theta_A)^{-2} s_{At}^{*\prime} \varepsilon_{At},$$

where $u_{At} = y_t - x_t\beta - x_{At}\beta_A$, $\bar{x}_{At} = (x_{At}, 0)$, $s^*_{At} = \partial h_t(\theta, \theta_A)/\partial\theta_A'$, and $\varepsilon_{At} = u^2_{At} - h_t(\theta, \theta_A)$. Then under $H_0 : \theta_A = 0$,

$$d_A(\theta) = T^{-1}\sum_{t=1}^{T} h_t(\theta)^{-1}\bar{x}_{At}' u_t + \frac{1}{2}T^{-1}\sum_{t=1}^{T} h_t(\theta)^{-2} s_{At}' \varepsilon_t, \tag{15}$$

where $s_{At} = s^*_{At}$ evaluated at $\theta_A = 0$. The LM test for H_0 is based upon (15), and this is clearly a particular case of the general consistency test in (3) by making $\Phi_t = (h_t^{-1}\bar{x}_{At}, \frac{1}{2}h_t^{-2}s_{At})$. We will use this in the next section to provide a $2TR_0^2$ form for the LM tests in heteroskedastic and risk models, thus extending the results in Engle (1982b) to the case where there is a non-diagonal information matrix between β and α.

3. The Distribution of Consistency Test-Statistics

It is convenient to recast the model in (1) in a more general form that explicitly allows for risk terms in the conditional mean, and also clarify our treatment of local alternatives which follows Newey (1985a, b).

For the first of this issues, we will allow for risk terms only in so far as the risk measures are defined by characteristics of the conditional distribution, more specifically as functions of the conditional variance as in the ARCH-M model. The reason for this is that we want to maintain a likelihood framework based on the conditional distribution alone. Inferences in the context of alien risk terms have been discussed by Pagan (1984b) and Pagan and Uliah (1987).

Thus suppose that among the regressors in (1) we include h_t. The effect is that the conditional mean is no longer a linear function of β, but it is usually a nonlinear function of θ. Therefore we replace the model in (1) by the more general

$$y_t \mid F_t \sim \mathrm{N}[\mu_t(\theta), h_t(\theta)] , \tag{16}$$

where $\mu_t = \mu_t(\theta)$ is, of course, a measurable function of F_t. At a theoretical level, this has no effect on our previous arguments although restrictions on the parameters are induced by the requirement that the feedback between the conditional moments does not violate the basic regularity conditions i.e. it is not explosive. It is still the case that $E[d_\beta(\beta_0, \alpha^*)] = \theta$ provides the necessary and sufficient condition for the consistency of $\hat{\beta}$, as Lemmas 1 and 2 of Pagan and Sabau (1991) do not depend on the mean being parameterized as a function of β only. One would expect, however, that the presence of h_t in the conditional mean would make remote the possibility of estimating β consistently under variance misspecification, a point we have already noted.

The amended score corresponding to (16) is

$$\begin{aligned} d_\theta(\theta) &= T^{-1}\sum_{t=1}^{T} h_t^{-1} u_t \frac{\partial \mu_t}{\partial \theta} + \frac{1}{2} T^{-1} \sum_{t=1}^{T} h_t^{-2} \varepsilon_t \frac{\partial h_t}{\partial \theta} \\ &= T^{-1} \sum_{t=1}^{T} \frac{\partial g_t}{\partial \theta} \Omega_t^{-1} \upsilon_t = T^{-1} G_\theta' \Omega^{-1} \upsilon , \end{aligned} \tag{17}$$

where g_t is redefined as $g_t = (\mu_t, h_t)$. The derivatives with respect to θ are more complex than previously and may be best computed numerically. Note that the relations in (6)-(15) are easily modified to cover the risk case.

To consider local alternatives, we reparameterize (16) allowing for an extra set of parameters, say γ_T, depending on sample size so that

$$y_t \mid F_t \sim \mathrm{N}[\mu_t(\theta, \gamma_T), h_t(\theta, \gamma_T)] \tag{18}$$

and the dependence on sample size takes the form $\gamma_T = \gamma_0 + T^{-1/2}\delta$ for fixed γ_0 and δ. The parameterization is such that $\mu_t = \mu_t(\theta) = \mu_t(\theta, \gamma_0)$ and $h_t = h_t(\theta) = h_t(\theta, \gamma_0)$ for all $\theta \in \Theta$, or $g_t = g_t(\theta) = g_t(\theta, \gamma_0)$ with

$$g_t(\theta, \gamma_T) = (\mu_t(\theta, \gamma_T), h_t(\theta, \gamma_T)).$$

That is, at $\delta = 0$ we obtain the null hypothesis in (16). As T increases, (18) approaches (16) at the required rate to keep the noncentrality parameters finite in the distribution of the test-statistics, as shown below in Theorem 1.

Corresponding to the sequence of local alternatives are the mean and variance innovations $u_t(\gamma_T) = u_t(\theta, \gamma_T) = y_t - \mu_t(\theta, \gamma_T)$ and

$$\varepsilon_t(\gamma_T) = \varepsilon_t(\theta, \gamma_T) = u_t(\gamma_T)^2 - h_t(\theta, \gamma_T),$$

with $u_t = u_t(\gamma_0)$ and $\varepsilon_t = \varepsilon_t(\gamma_0)$ under the null hypothesis. Similarly, define $\mathrm{u}_t(\gamma_T) = (u_t(\gamma_T), \varepsilon_t(\gamma_T))'$ and $\upsilon_t = \upsilon_t(\gamma_0)$. The derivatives of the log-likelihood function at observation t are

$$d_{\theta t}(\theta, \gamma_T) = \frac{\partial g_t(\theta, \gamma_T)}{\partial \theta} \Omega_t(\gamma_T)^{-1} \upsilon_t(\gamma_T), \tag{19a}$$

and

$$d_{\gamma t}(\theta, \gamma_T) = \frac{\partial g_t(\theta, \gamma_T)}{\partial \gamma} \Omega_t(\gamma_T)^{-1} \mathrm{u}_t(\gamma_T), \tag{19b}$$

where $\Omega_t(\gamma_T) = \Omega_t(\theta, \gamma_T)$ and $\Omega_t = \Omega_t(\theta, \gamma_0)$. The score for the t^{th} observation under the null hypothesis is $d_{\theta t} = d_{\theta t}(\theta, \gamma_0)$ and

$$d_\theta = d_\theta(\theta) = T^{-1} \sum_{t=1}^{T} d_{\theta t}\ .$$

The negative of the matrix of second derivatives with respect to θ is

$$J_{\theta t}(\theta, \gamma_T) = \frac{\partial g_t(\theta, \gamma_T)}{\partial \theta} \Omega_t(\gamma_T)^{-1} \left(\frac{\partial g_t(\theta, \gamma_T)}{\partial \theta} \right)' + A_{\theta t}\ ,$$

where

$$A_{\theta t} = [\upsilon_t(\gamma_T)' \otimes I_p] \frac{\partial}{\partial \theta'} \left(\frac{\partial g_t(\theta, \gamma_T)}{\delta \theta} \Omega_t(\gamma_T)^{-1} \right),$$

which by using the law of iterated expectations is seen to have zero expected value under the local alternatives in (18). Therefore, the information matrix is given by

$$\mathfrak{I}(\theta_0) = \xi\{T^{-1} \sum_{t=1}^{T} d_{\theta t} d_{\theta t}'\ \} = \xi\{\ T^{-1} \sum_{t=1}^{T} J_{\theta t}\} = \xi\{\ T^{-1} G_\theta' \Omega^{-1} G_\theta\}, \tag{20}$$

where $\xi\{.\} = \lim_{T \to \infty} E\,[.]$, and all the expectations are evaluated at θ_0. It is easlily seen that under the regularity conditions set out in Theorem 1 below, $\mathfrak{l}(\theta_0)$ is the appropriate information matrix for θ under both (16) and (18).

Similarly, the matrix of derivatives of the statistic $m(\Phi\,,\theta)$ with respect to θ has an outer product form which under both the null and the sequence of local alternatives is given by (Newey, 1985a)

$$M = M(\theta_0) = \xi\{T^{-1}\sum_{t=1}^{T}\frac{\partial m_t(\theta_0)}{\partial\theta'}\} = -\xi\,\{T^{-1}\sum_{t=1}^{T}m_t(\theta_0)d'_{\theta t}\}.$$

Using (3) and (17) in the last equality we get

$$M(\theta_0) = -\xi\{T^{-1}\sum_{t=1}^{T}\Phi_t\left(\frac{\partial g_t(\theta\,,\gamma_T)}{\partial\theta}\right)'\} = -\xi\{T^{-1}\Phi' G_\theta\}, \tag{21}$$

by the law of iterated expectations, all evaluated under H_0.

We now prove:

THEOREM 1. *Let the random variables* $(y_t\,, x_t)$ *have* pdf *conditional on the information set* $\tilde{F}_t$*(σ–field) given by* $f(y_t\,, x_t \mid \tilde{F}_t) = f(y_t \mid F_t\,;\theta_0\,,\gamma_T)f(x_t \mid \tilde{F}_t)$, *where* $F_t \subset \tilde{F}_t$ *and the* x_t *are weakly exogenous for* θ. *If this pdf together with the function* $g^* = (m(\Phi, \theta)'\,, d_\theta')'$ *obey the regularity, continuity, dominance and mixing conditions in assumptions (1)-(6) of Newey (1985b), then*

$$T^{1/2}m(\hat{\Phi}\,,\hat{\theta}) \xrightarrow{d} \mathrm{N}[\psi\,, Q_\Phi]\,, \tag{22}$$

where $\psi = [\xi\{T^{-1}\Phi' G_\gamma\} - M(\theta_0)\mathfrak{l}(\theta_0)^{-1}\xi\{T^{-1}G_\theta'\Omega^{-1}G_\gamma\}]\,\delta$ *and*

$$Q_\Phi = V_0 - M(\theta_0)\mathfrak{l}(\theta_0)^{-1}M(\theta_0)'\,,$$

and V_0 *is the limit of the covariance matrix in (5) under* H_0. *A consistent estimator of* Q_Φ *is*

$$\hat{Q}_\Phi = T^{-1}\hat{\Phi}'\hat{\Omega}^{1/2}\hat{\Psi}\hat{\Omega}^{1/2}\hat{\Phi},$$

where $\Psi = I_{2T} - \Omega^{-1/2}G_\theta(G_\theta'\Omega^{-1}G_\theta)^{-1}G_\theta'\Omega^{-1/2}$ *and all estimates are under* H_0.

PROOF. The result is Lemma 1 in Newey (1985b) with some specialization. In our context $d_{\theta t}(\theta_0)$ and $m_t(\Phi\,,\theta)$ are martingale differences with respect to the σ– field F_t and so the matrix V simplifies from Newey's corresponding expression to

$$V = \begin{pmatrix} V_0 & -M(\theta_0) \\ -M(\theta_0)' & \mathfrak{l}(\theta_0) \end{pmatrix}.$$

Let L be the matrix selecting $m(\Phi\,,0)$ from g^*, $L = (I_n\,, 0)$

$$C = E\left(\frac{\partial g^*(\theta_0)}{\partial \theta'}\right),$$

$\bar{C} = (0\,, I_p)$, and $P_c = I_{p+n} - C(\bar{C}\,C)^{-1}\bar{C}$. Then using (20) and (21), $LP_c = (I_n\,, -M(\theta_0)\mathfrak{l}(\theta_0)^{-1})$, and therefore the covariance matrix of $T^{1/2}m(\hat{\Phi}\,,\hat{\theta})$ is given by

$$Q_\Phi = LP_cVP_c'L' = V_0 - M(0_0)\mathfrak{l}(\theta_0)^{-1}M(\theta_0)'.$$

Again using (20) and (21) together with (5), and following the argument in Newey (1985a), especially that connected with equation (2.11), yields the consistent estimator $\hat{Q}_\Phi$ of Q_Φ in (23).

Finally, note that

$$E\,[T^{-1}\sum_{t=1}^{T} m_t d_{\gamma t}] = E\,[T^{-1}\sum_{t=1}^{T}\Phi_t \upsilon_t \upsilon_t' \Omega_t^{-1}\left(\frac{\partial g_t}{\partial \gamma}\right)'] = E\,[T^{-1}\Phi' G_\gamma]\,,$$

by using iterated expectations and, similarly

$$E\,[T^{-1}\sum_{t=1}^{T} d_{\theta t} d_{\gamma t}] = E\,[T^{-1}\sum_{t=1}^{T}\frac{\partial g_t}{\partial \theta}\,\Omega_t^{-1}\upsilon_t \upsilon_t' \Omega_t^{-1}\left(\frac{\partial g_t}{\partial \gamma}\right)']$$

$$= E\,[T^{-1}G_\theta'\,\Omega^{-1}G\gamma]$$

Since these two matrices form the matrix K in Newey (1985b), the noncentrality parameter follows directly. ∎

We defer power considerations to the next section. The limiting distribution of $T^{1/2}m(\hat{\Phi}\,,\hat{\theta})$ when the model is correctly specified is established by setting $\delta = 0$ in Theorem 1, a result given in

COROLLARY 1.1. *Under the assumptions of Theorem 1, when the maintained model is correctly specified* ($\delta = 0$),

$$T^{1/2} m(\hat{\Phi}, \hat{\theta}) \xrightarrow{d} \mathrm{N}[0, Q_{\Phi}].$$

PROOF. Set $\delta = 0$ in Theorem 1. ■

Cases of particular interest to this paper are when $\Phi_t = (1, 0)$ and $\Phi_t = (0, 1)$, making the basic consistency statistics

$$m_{\mu} = T^{-1} \sum_{t=1}^{T} \hat{u}_t$$

and

$$m_h = \sum_{t=1}^{T} \hat{\varepsilon}_t \ .$$

Corollary 1.2 specializes Theorem 1 to these situations:

COROLLARY 1.2. *Under the assumptions of Theorem 1,*

$$T^{1/2} m_{\mu} \xrightarrow{d} \mathrm{N}[(\bar{\mu}_{\gamma} - \bar{\mu}_{\theta} V(\hat{\theta}) G_{\theta\gamma})\delta, \sigma^2 - \bar{\mu}_{\theta} V(\hat{\theta}) \bar{\mu}_{\theta}'] \tag{24a}$$

and

$$T^{1/2} m_h \xrightarrow{d} \mathrm{N}[(\bar{h} - \bar{h}_{\theta} V(\hat{\theta}) G_{\theta\gamma})\delta, 2\tau^2 - \bar{h}_{\theta} V(\hat{\theta}) \bar{h}_{\theta}'] \tag{24b}$$

where

$$\bar{\mu}_{\gamma} = \xi \{ T^{-1} \sum_{t=1}^{T} \partial \mu_t / \partial \gamma \}' ,$$

$$\bar{\mu}_{\theta} = \xi \{ T^{-1} \sum_{t=1}^{T} \partial \mu_t / \partial \theta \}' ,$$

$$G_{\theta\gamma} = \xi \{ T^{-1} G_{\theta}' \Omega^{-1} G_{\gamma} \} ,$$

$V(\hat{\theta}') = \mathfrak{l}(\theta_0)^{-1}$ *is the asymptotic covariance matrix of the MLE* $\hat{\theta}$,

$$\sigma^2 = \xi \{ T^{-1} \sum_{t=1}^{T} h_t \} ,$$

$$\bar{h}_\gamma = \xi \{ T^{-1}\sum_{t=1}^{T} \partial h_t / \partial \gamma \}' ,$$

$$\bar{h}_\theta = \xi \{ T^{-1}\sum_{t=1}^{T} \partial h_t / \partial \theta \}' ,$$

and

$$\tau^2 = \xi \{ T^{-1}\sum_{t=1}^{T} h_t^2 \}.$$

PROOF. For $\Phi_t = (1, 0)$ note that $\Phi_t \left(\frac{\partial g_t}{\partial \theta}\right)' = \partial \mu_t / \partial \theta'$, $\Phi_t \left(\frac{\partial g_t}{\partial \gamma}\right)' = \partial \mu_t / \partial \gamma'$, and $\Phi_t \Omega_t \Phi_t' = h_t$.

For $\Phi_t = (1, 0)$ note that $\Phi_t \left(\frac{\partial g_t}{\partial \theta}\right)' = \partial h_t / \partial \theta'$, $\Phi_t \left(\frac{\partial g_t}{\partial \gamma}\right)' = \partial h_t / \partial \gamma'$, and $\Phi_t \Omega_t \Phi_t' = 2h_t^2$. ■

Note that when the conditional mean does not depend on α as in (1), the variance of $T^{1/2}m_\mu$ simplifies to $\sigma^2 - \bar{x} V(\hat{\beta})\bar{x}'$, with $\bar{x}$ being the sample mean of x_t. Similarly, if h_t does not depend on β, the variance of $T^{1/2}m_h$ simplifies to $2\tau^2 - \bar{z} V(\hat{\alpha})\bar{z}'$, with $\bar{z}$ being the sample mean of z_t. When the information matrix is block-diagonal between β and α the variances of the statistics are $\sigma^2 - \bar{x}V(\hat{\beta})\bar{x}' - \bar{\mu}_\alpha V(\hat{\beta})\bar{\mu}_\alpha'$ and $2\tau^2 - \bar{z}V(\hat{\alpha})\bar{z}' - \bar{w}V(\hat{\alpha})\bar{w}'$, respectively, with $\bar{\mu}_a$ the sample mean of $\partial \mu_t / \partial \alpha$ and $\bar{w}$ the sample mean of w_t. An interesting case is the ARCH model, for then $\mathrm{I}(\theta_0)$ is diagonal and also

$$\bar{w} = -T^{-1}\sum_{t=1}^{T} \alpha_j u_{t-j} x_{t-j} \xrightarrow{as} 0, \tag{25}$$

resulting in the simple variances $\sigma^2 - \bar{x}V(\hat{\beta})\bar{x}'$ and $2\tau^2 - \bar{z}V(\hat{\alpha})\bar{z}'$. This is easily seen to apply as well to the more general GARCH model.

Of course many other test-statistics might be used e.g. there are some advantages to employing the sum of standardized residuals $T^{-1}\sum_{t=1}^{T} \hat{h}_t^{-1}\hat{u}_t$ or $T^{-1}\sum_{t=1}^{T} \hat{h}_t^{-2}\hat{\varepsilon}_t$ which relate to the differences between ML and GLS estimators and thus have smaller sampling variation under H_0 for large T. All of these can be handled with Theorem 1 by appropriate choices of Φ_t, and the asymptotic distributions obtained as in Corollary 1.2. The variances are consistently estimated by replacing expectations with sample moments. It is important to note, however, that the fact that θ needs to be estimated means that the asymptotic variance of the test-statistics is less than it would be if θ were known, so that inference needs to allow for this extra source of uncertainty. The situation is therefore analogous to that which occurs when testing for serial correlation in the presence of lagged dependent variables.

A joint test can also be constructed with $\Phi_t = I_2$. Proceeding as in Corollary 1.2, it is easily seen that $\mathrm{cov}[T^{1/2}m_{\mu}, T^{1/2}m_h] = -\bar{\mu}_{\theta}V(\hat{\theta})\bar{h}_{\theta}'$, because V_0 is diagonal by construction. The statistics are rarely independent, but here againg the GARCH model is one exception. This is seen by combining (25) with $\mathfrak{l}_{\beta\alpha}(\theta_0) = 0$ and $\bar{\mu}_{\alpha} = 0$.

Although the actual computation of the test-statistic is already clear from Theorem 1, it is worthwhile producing a simplified calculation. This can be done by means of an uncentered coefficient of determination, say R_0^2, of an auxiliary regression as in Engle (1982b, 1984), Newey (1985a), and Davidson and MacKinnon (1984), and will be specially attractive when using multi-dimensional tests. For this purpose we produce

THEOREM 2. *Under the assumptions of Theorem 1*

$$s = Tm(\hat{\Phi}, \hat{\theta})'\hat{Q}_{\Phi}^{-1}m(\hat{\Phi}, \hat{\theta}) \xrightarrow{d} \chi^2[n; \lambda_{\Phi}^2], \tag{26}$$

with

$$\lambda_{\Phi}^2 = \delta'\xi\{T^{-1}G_{\gamma}'\Omega^{-1/2}\Gamma_{\Phi}\Omega^{-1/2}G_{\gamma}\}\delta$$

and

$$\Gamma_{\Phi} = \Psi\Omega^{1/2}\Phi(\Phi'\Omega^{1/2}\Psi\Omega^{1/2}\Phi)^{-1}\Phi'\Omega^{1/2}\Psi.$$

Moreover, if $(G_{\theta}, \Omega\Phi)$ has full column rank then $s - s^* \xrightarrow{as} 0$ where $s^* = 2TR_0^2$, and R_0^2 is the uncentered coefficient of determination of the regression of $\hat{\upsilon}$ on $\hat{G}_{\theta}$ and $\hat{\Omega}\hat{\Phi}$ in the metric of $\hat{\Omega}$.

PROOF. From Theorem 1 we have that

$$s = Tm(\hat{\Phi}, \hat{\theta})'\hat{Q}m(\hat{\Phi}, \hat{\theta}) \xrightarrow{d} \chi^2[n; \psi'Q^{-1}\psi].$$

But using (3) and (23) we get $s = \hat{\upsilon}'\hat{\Phi}(\hat{\Phi}'\hat{\Omega}^{1/2}\hat{\Psi}\hat{\Omega}^{1/2}\hat{\Phi})^{-1}\hat{\Phi}'\hat{\upsilon} \xrightarrow{as} 2TR_0^2 = s^*$ by an argument identical to Engle (1982b), noting $T^{-1}\hat{G}_{\theta}'\hat{\Omega}^{-1}\hat{\upsilon} \xrightarrow{as} 0$ and $\frac{1}{2}T^{-1}\hat{u}'\hat{\Omega}^{-1}\hat{u} \xrightarrow{as} 1$. The noncentrality parameter follows using (20) and (21) in the expression for ψ, so

$$\psi = \xi\{T^{-1}[\Phi'G_{\gamma} - \Phi'G_{\theta}(G_{\theta}'\Omega^{-1}G_{\theta})^{-1}G_{\theta}'\Omega^{-1}G_{\gamma}]\}\delta = \xi\{T^{-1}\Phi'\Omega^{1/2}\Psi\Omega^{-1/2}G_{\gamma}\}\delta,$$

and the result follows using(23). ■

The immediate result under the null hypothesis is:

COROLLARY 2.1. *Under the assumptions of Theorem 2 and the model being correctly specified* ($\delta = 0$), $s \xrightarrow{d} \chi^2_n$ *and* $s - s^* \xrightarrow{as} 0$.

PROOF. Setting $\delta = 0$ in Theorem 2 results in $\lambda^2_{\Phi} = 0$. ■

Note that our auxiliary regression differs from Newey's which has single length, a dependent variable of unity and is devised for more general environments. The same can be said about the double-length auxiliary regression procedure put forward by Davidson and MacKinnon (1984), which differs from ours but is asymptotically equivalent for our specialization. The basic difference with Engle's (1982b) criteria is that we need to incorporate a double-length auxiliary regression to allow for a possible non-diagonal information matrix between β and α, a case explicitly excluded from his Theorem 1. In fact, we can apply Theorem 2 directly to the LM test for additional variables being properly excluded in either or both conditional moments of y_t. The LM test is based on the subvector of the score given in (15), and denoting $G_A = \partial g(\theta, \theta_A) / \partial\theta_A$ we have

COROLLARY 2.2. *Under the assumptions of Theorem 2 the LM test for* $H_0 : \theta_A = 0$ *in the model* $y_t | F_t \sim \mathrm{N}[\mu_t(\theta, \theta_A), h_t(\theta, \theta_A)]$ *is given by*

$$s_{LM} = Tm(\hat{\Omega}^{-1}\hat{G}_A, \hat{\theta})'\hat{Q}_{\Omega^{-1}G_A} m(\hat{\Omega}^{-1}\hat{G}_A, \hat{\theta}) \xrightarrow{d} \chi^2[n\, ; \lambda^2_{\Omega^{-1}G_A}] \tag{27}$$

and an asymptotically equivalent statistic is $s^*_{LM} = 2TR^2_0$ *from the regression of* $\hat{\upsilon}$ *on* $\hat{G}_\theta$ *and* $\hat{G}_A$ *in the metric of* $\hat{\Omega}$, *with all estimates under* H_0 *and where* n *is the dimension of* θ_A.

PROOF. The subvector of the score for θ_A modified from (15) to consider the more general specification of the conditional mean is

$$T^{-1}\sum_{t=1}^{T} \frac{\partial g_t}{\partial\theta_A}\Omega_t^{-1}\upsilon_t = T^{-1}G_A'\,\Omega^{-1}\upsilon$$

under H_0, and with all functions evaluated under this hypothesis. Therefore, set $\Phi = \Omega^{-1}G_A$ in Theorem 2. ■

Note that Theorem 2 and its corollaries are general enough to accommodate wide classes of heteroskedastic and risk models, and allow for separate or joint testing of either conditional moment. When $\iota_{\alpha\beta}(\theta_0) = 0$, as in the GARCH case, tests for the conditional variance can be obtained from a single length regression with $\hat{\varepsilon}_t$ as the dependent variable. This is the univariate specialization of the tests provided in Kraft and Engle (1982) for the multivariate case. The single-length test for the conditional mean is an adequate one, which follows from Theorem 1 of Engle (1982b). However, this is not the LM test in this context for it is not using the information about β in the conditional variance and therefore no optimality claims can be made. The LM test for the conditional mean must be computed from the double length regression to incorporate all the relevant information about β in the likelihood function and thus have the usual optimal properties of the LM, Wald and LR procedures. And in the general case of a non-diagonal information matrix, the double-length regression must be performed to obtain the proper LM statistic even if the test is designed for a single conditional moment.

The LM test is designed to achieve high power in a given direction, namely that of the additional variables. But it can also be used as a very general test for misspecification by proper definition of the additional variables to approximate a wide range of departures from the null hypothesis. This is the idea used in the RESET test (Ramsey, 1969), and it coul be used here as well by adding powers of $\hat{y}_t$ to the conditional mean and/or powers of $\hat{h}_t$ to the conditional variance.

The consistency tests of Corollary 1.2 may be constructed as variable addition tests and the same applies to some of the estimator-difference tests in (6)-(14) and this establishes a relation between variable addition and variable transformation tests as in Breusch and Godfrey (1986). Note, however, that the tests in which the sums of residuals are weighted by the inverse conditional variances will, in general, result in $(G_\theta, \Omega\Phi)$ having less than full column rank and thus cannot be constructed by variable addition. Nevertheless, the tests for these choices of Φ are well defined in general and will have power as long as the ML and GLS estimators differ.

4. Some Power Considerations

In order to analyze the local power of the tests, it is convenient to decompose the vector in (22) for the quadratic form producing the noncentrality parameter into $\psi = \psi_\mu + \psi_h$, where ψ_μ summarizes the inconsistency arising from a misspecified conditional mean, and ψ_h summarizes the inconsistency arising from a misspecified conditional variance. Partioning $\Phi_t = (\Phi_{1t}, \Phi_{2t})$ we have after simple algebraic manipulation

$$\psi_\mu = \xi \; \{ \; T^{-1}\sum_{t=1}^{T}\left(\Phi_{1t} - M(\theta_0)V(\hat{\theta})h_t^{-1}\frac{\partial\mu_t}{\partial\theta}\right)\frac{\partial\mu_t}{\partial\gamma'} \; \} \; \delta, \tag{28a}$$

and

$$\psi_h = \xi \{ T^{-1}\sum_{t=1}^{T}\left(\Phi_{2t} - \frac{1}{2}M(\theta_0)V(\hat{\theta})h_t^{-2}\frac{\partial h_t}{\partial \theta}\right)\frac{\partial h_t}{\partial \gamma'} \} \delta, \tag{28b}$$

and we also decompose $M(\theta_0) = M_\mu(\theta_0) + M_h(\theta_0)$, where

$$M_\mu(\theta_0) = \xi \{ T^{-1}\sum_{t=1}^{T}\Phi_{1t}\frac{\partial \mu_t}{\partial \theta'} \}$$

and $M_h(\theta_0) = \xi \{ T^{-1}\sum_{t=1}^{T} \Phi_{2t}\frac{\partial h_t}{\partial \theta'} \}$.

Therefore, since Q_Φ is positive definite, the tests will have power against misspecification in μ_t whenever $\psi_\mu \neq 0$, and against misspecification in h_t whenever $\psi_h \neq 0$. It is of particular interest to examine the power of tests designed for one conditional moment when misspecification appears only in the other.

Let us consider first a mean test with a correctly specified conditional mean. Then $\Phi_{2t} = 0$ and so $M(\theta_0) = M_\mu(\theta_0)$, and also $\partial\mu_t / \partial\gamma = 0$. From (28) we get $\psi_\mu = 0$ and therefore

$$\psi = \psi_h = -\frac{1}{2}M_\mu(\theta_0)V(\hat{\theta})\xi \{ T^{-1}\sum_{t=1}^{T}h_t^{-2}\frac{\partial h_t}{\partial \theta}\frac{\partial h_t}{\partial \gamma'} \}\delta .$$

Note that this is possible, in general, only in models without risk terms because if these are present misspecification in h_t will contaminate μ_t thus making $\psi_\mu \neq 0$ and causing the test to have more power. But still if there are no risk terms, only in special cases will ψ_h vanish and so the inconsistency in $\hat{\beta}$ induced by conditional variance error analyzed in our previous paper will be acting through ψ_h. Now if there are no risk terms we also have $\partial\mu_t / \partial\alpha = 0$ and so we may partition $M_\mu(\theta_0) = (M_{\mu 1}(\theta_0) , 0)$, resulting in

$$\psi_h = -\frac{1}{2} M_{\mu 1}(\theta_0)(V(\hat{\beta}),\ \mathrm{cov}(\hat{\beta} , \hat{\alpha}))\, \xi \{ T^{-1}\sum_{t=1}^{T} h_t^{-2}\frac{\partial h_t}{\partial \theta}\frac{\partial h_t}{\partial \gamma'} \}\delta,$$

which reduces to

$$\psi_h = -\frac{1}{2}M_{\mu 1}(\theta_0)V(\hat{\beta})\xi \{ T^{-1}\sum_{t=1}^{T} h_t^{-2}\frac{\partial h_t}{\partial \beta}\frac{\partial h_t}{\partial \gamma'} \}\delta$$

when the information matrix is diagonal. This of course applies to the simple heteroskedasticity ($h_t = h_t(\alpha)$) and GARCH models. For the former, it is evident that $\psi_h = 0$. For the latter, whether ψ_h is or not zero will depend on the nature of the true conditional variance, represented here by the departure $\partial h_t / \partial\gamma$. It is clear that in the cases where we established consistency of $\hat{\beta}$ in our previous paper, the expected value in ψ_h will vanish and the test will have no power.

The other case is a variance test with a correctly specified conditional variance. Then $\Phi_{1t} = 0$ and $\partial h_t / \partial\gamma = 0$, so $M(\theta_0) = M_h(\theta_0)$ and $\psi_h = 0$, resulting in $\psi = \psi_\mu = -M(\theta_0)V(\hat{\theta})\xi\ \{ T^{-1}\sum_{t=1}^{T} h_t^{-1} \frac{\partial\mu_t}{\partial\theta} \frac{\partial\mu_t}{\partial\gamma'} \}\delta$. The expectation will be zero only in very special cases. For example, when the mis-specification is in the form of autocorrelated errors and neither conditional moment depends on lagged y's. Hence the variance test is capable of detecting misspecification in the conditional mean although whether it does or not depends upon the context.

The above cases place properly into perspective our terming the tests as "consistency" tests, for any departure not affecting the consistency of the subset of parameters on which the test focuses will be part of the implicit null hypothesis. However, performing groups of consistency tests singly and jointly will provide a valuable tool for assessing the model. Because of Theorem 2 performing a wide range of tests has a small computational cost.

To assess the power of general consistency tests against specific departures the optimality properties of LM tests provide a benchmark. For the test of $H_0 : \theta_A = 0$ against $H_1 : \theta_A \neq 0$, Corollary 2.2 gives the relevant test-statistic as (27) with NCP

$$\lambda^2_{\Omega^{-1}G_A} = \delta'\xi\ \{T^{-1}G_\gamma'\Omega^{-1/2}\Gamma_{\Omega^{-1}G_A}\Omega^{-1/2}G_\gamma\}\delta, \tag{29}$$

where

$$\Gamma_{\Omega^{-1}G_A} = \Psi\Omega^{-1/2}G_A(G_A'\Omega^{-1/2}\Psi\Omega^{-1/2}G_A)^{-1}G_A'\Omega^{-1/2}\Psi.$$

Of course, the optimality properties of LM tests apply generally only when the departure for which the test has been designed coincides with the actual departure from H_0, so that $G_\gamma = G_A$, and then (29) results in the optimal NCP

$$\lambda^2_{LM} = \delta'\xi\ \{T^{-1}G_\gamma'\Omega^{-1/2}\Psi\Omega^{-1/2}G_\gamma\}\delta. \tag{30}$$

It is interesting to note that this LM test is the optimal test in the class presented in (3) for departures of the form $\gamma_T = \gamma_0 + T^{-1/2}\delta$. This we prove by specializing Theorem 3.1 of Newey (1985a) to our situation:

THEOREM 3. *Under the assumptions of Theorem 1, the LM test for $\delta = 0$ is the optimal consistency test in the sense that its NCP, given in (30), is no smaller than that of any other test of the form given in (3).*

PROOF. From Theorem 3.1 of Newey (1985a) applied to our case with u_t being a martingale difference with respect to F_t, the optimal test with basic statistic $m(\Phi, 0)$ has $\Phi_t = E[d_{\gamma t}\upsilon_t' \mid F_t]E[\upsilon_t\upsilon_t' \mid F_t]^{-1}$, where the expectations are taken under H_0. But then $E[\upsilon_t u_t' \mid F_t] = \Omega_t$, and from (19) $E[d_{\gamma t}\upsilon_t' \mid F_t] = \partial g_t / \partial\gamma$, which produces the optimal test statistic based on $T^{-1}G_\gamma'\Omega^{-1}\upsilon$, the subvector of the score for the test of $\delta = 0$. ■

Therefore, to evaluate the power of a general consistency test in a specific direction all that needs to be done is to assess how well the relevant Φ projects onto the space spanned by $\Omega^{-1}G_\gamma$. The smaller the distance of Φ from this space, the greater the power in the specific direction under analysis. The situation resembles that of choice of instruments to achieve efficiency in estimation.

To illustrate the argument let us consider the linear ARCH(1) model given by

$$y_t \mid F_t \sim \mathrm{N}[x_t\beta\, ,\, h_t = \alpha_0 + \alpha_1 u_{t-1}^2 = z_t\alpha], \tag{31}$$

where $z_t = (1\, ,\, u_{t-1}^2)$ and $\alpha = (\alpha_0\, ,\, \alpha_1)'$. Suppose the true conditional variance h_t^* departs locally in the direction of $h_t^{**} = \alpha_0 + \alpha_1 y_{t-1}^2$. Using $y_{t-1} = x_{t-1}\beta + u_{t-1}$ and $w_t = -2\alpha_1 u_{t-1}x_{t-1}\beta$ we may write

$$h_t^{**} = h_t - w_t\beta + \alpha_1(x_{t-1}\beta)^2,$$

and we consider separately the two departure directions by making

$$h_t^* = h_t - T^{-1/2}\delta_1 w_t\beta + T^{-1/2}\delta_2(x_{t-1}\beta)^2 = h_t - T^{-1/2}z_{at}\delta, \tag{32}$$

where $\delta = (\delta_1\, ,\, \delta_2)'$ and $z_{at} = (-w_t\beta\, ,\, (x_{t-1}\beta)^2)'$. The alternative h_t^{**} is interesting empirically in view of the results obtained by Weiss (1984). It is also interesting theoretically because the two departure directions in (32) are, respectively, a conditionally odd and a conditionally even function of u_{t-1}.

Consider the simple tests based on

$$m_\mu = T^{-1}\sum_{t=1}^{T}\hat{u}_t \quad , \quad m_h = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t\, .$$

From Corollary 1.2 and (25) the variances are $\sigma_\mu^2 = \bar{z}\alpha - xV(\hat{\beta})\bar{x}'$ and $\sigma_h^2 = 2\tau^2 - \bar{z}\,V(\hat{\alpha})\bar{z}'$, respectively, where $\tau^2 = \alpha'\xi\{T^{-1}Z'Z\}\alpha$. Since the mean is correctly specified the term ψ_μ in (28) is zero. To compute the noncentrality parameter we need

$$T^{-1}G_\theta'\Omega^{-1}G_\gamma = \frac{1}{2}T^{-1}\begin{pmatrix} -\sum_{t=1}^{T} h_t^{-2}w_t' w_t\beta & \sum_{t=1}^{T} h_t^{-2}w_t'(x_{t-1}\beta)^2 \\ -\sum_{t=1}^{T} h_t^{-2}z_t' w_t\beta & \sum_{t=1}^{T} h_t^{-2}z_t' x_{t-1}\beta^2 \end{pmatrix},$$

and because w_t is conditionally odd in u_{t-1} while $h_t^{-2}z_t$ and $h_t^{-2}(x_{t-1}\beta)^2$ are conditionally even in u_{t-1}, it follows from Lemma 3 of Pagan and Sabau (1991) that the off-diagonal elements have zero expectation. Let $\hat\theta_v$ be the estimator of θ obtained using (nonlinear) GLS in the variance equation $\hat{u}_t^2 = h_t(\theta) + e_t$, $e_t = \varepsilon_t + \{\hat{u}_t^2 - u_t^2\}$. Sabau (1987a) has shown that for ARCH models this is a consistent estimator of θ_0 if the model is correctly specified. Partition $\hat\theta_v = (\hat\beta_v', \hat\alpha_v')'$ and let $\hat\beta_m$ be the feasible GLS estimator for β obtained from the mean equation $y_t = x_t\beta + u_t$. Sabau has also shown that

$$V(\hat\beta_v) = \xi\,\{\frac{1}{2}T^{-1}\sum_{t=1}^{T} h_t^{-2}w_t' w_t\}^{-1}, \quad V(\hat\beta_m) = \xi\,\{T^{-1}\sum_{t=1}^{T} h_t^{-1}x_t' x_t\}^{-1}$$

are the asymptotic covariance matricesof these estimators and further that the MLE $\hat\beta$ is asymptotically a matrix weighted average of them. Therefore the covariance matrix of $\hat\beta$ obeys

$$V(\hat\beta)^{-1} = V(\hat\beta_m)^{-1} + V(\hat\beta_v)^{-1}. \tag{33}$$

Thus letting $d = \xi\,\{\frac{1}{2}T^{-1}\sum_{t=1}^{T} h_t^{-2}z_t'(x_{t-1}\beta)^2\}$ we have that the $G_{\theta\gamma}$ factor in (24) is given by

$$G_{\theta\gamma} = \mathrm{diag}\,\{V(\tilde\beta_v)^{-1}\beta, d\},$$

while $\bar\mu_\gamma = 0$ because of the correctly specified mean and $\bar{h}_\gamma\delta = \delta_1 c_1$, where

$$c_1 = 1\,\xi\,\{T^{-1}\sum_{t=1}^{T}(x_{t-1}\beta)^2\},$$

in view of (25). Finally, using the fact that the information matrix is diagonal between β and α (Engle, 1982a), we get

$$s_{\mu} = T\frac{m_{\mu}^{2}}{\hat{\sigma}_{\mu}^{2}} \xrightarrow{d} \chi^{2}[1\,;\lambda_{\mu}^{2} = \delta_{1}^{2}\xi\,\{\bar{x}\,V(\hat{\beta})V(\hat{\beta}_{v})^{-1}\beta\}^{2}/\,\sigma_{\mu}^{2}]\,, \tag{34a}$$

and

$$s_{h} = T\frac{m_{h}^{2}}{\hat{\sigma}_{h}^{2}} \xrightarrow{d} \chi^{2}[1\,;\lambda_{\mu}^{2} = \delta_{2}^{2}\xi\,\{c_{1} - \bar{z}\,V(\hat{\alpha})d\}^{2}/\,\sigma_{h}^{2}]\,, \tag{34b}$$

where $\hat{\sigma}_{\mu}^{2}$ and $\hat{\sigma}_{h}^{2}$ are obtained replacing parameter estimates in the corresponding expressions for σ_{μ}^{2} and o_{h}^{2}.

The power for the mean test depends upon δ_1, which summarizes misspecification in the odd direction and the factor $V(\hat{\beta})V(\hat{\beta}_v)^{-1}$ is just the (generalized) proportional contribution to the efficiency of the MLE for β obtained from the information in the conditional variance. Therefore, the power of s_{μ} will increase when the variance is very informative about β relative to the mean. This is the natural thing to expect because only if the signal from the variance is very clear shall we get a good deal of information from it. And of course in the case of a misspecified conditional variance the inconsistency will grow larger. Thus power will be good in the cases when there may be a substancial gain in efficiency over estimating β by OLS, which was one of the arguments put forward by Engle (1982a) in favour of ML estimation.

The power of the variance test, on the other hand, comes completely from the even term, as reflected in δ_2. Clearly, power will grow as the presence of the lagged squared mean is clearer in the true conditional variance.

The LM test against the alternative h_t^{**} can be obtained as TR_0^2 from the regression of $\hat{\varepsilon}_t$ on $(\hat{z}_t, \hat{z}_{at})$ in the metric of $2(\hat{z}_t\hat{\alpha})^2$. The test may be computed from a single length regression because of the diagonality of the information matrix and the absence of information about α in the conditional mean (Engle, 1983, Kraft and Engle 1982). The same argument would also apply to the construction of s_h from an auxiliary regression. Note however that LM and consistency tests on β will require the double length regression in order to use the information in the conditional variance. Proceeding as above and using (27), the noncentrality parameter for the LM test is found to be

$$\lambda_{LM}^{2} = \delta_{1}^{2}\beta' V(\hat{\beta}_{v})^{-1}[V(\hat{\beta}_{v}) - V(\hat{\beta})]V(\hat{\beta}_{v})^{-1}]\beta + \delta_{2}^{2}[c_{2} - d'V(\hat{\alpha})d]\,, \tag{35}$$

where $c_2 = \xi\,\{\,T^{-1}\sum_{t=1}^{T} h_t^{-2}(x_{t-1}\beta)^2\,\}$ and the relation in (33) has been used. The first term comes from the odd misspecification direction and the second term comes from the even misspecification direction. It is now apparent that the power of the ML test that comes from the odd term is increased by a contribution to overall efficiency of both moments. But in contrast to the mean consistency test, when the conditional variance provides most of the information and $V(\hat{\beta}_v)$ approaches $V(\hat{\beta})$, power becomes null. In fact it is easy to see that the Φ matrix for the simple mean test has a null projection on the space spanned by $\Omega^{-1}G_\gamma$ when only h_t is misspecified i.e. the first row of $\left(\frac{\partial g_t}{\partial \gamma}\right)'$, is zero. Thus the power of the simple mean test may be very far from optimal in any specific direction, though it has the attraction of a wide range. For the variance test, the projection of Φ onto the space spanned by $\Omega^{-1}G_\gamma$ is that of a vector of ones onto the space spanned by the vector with typical element $h_t^{-2}(x_{t-1}\beta)^2$.

5. Re-Interpretation and Application

Before looking at an application of the above tests, it is useful to reinterpret them in the framework of Pagan and Hall (1983). By definition the errors u_t and ε_t have $E[u_t \mid F_t] = 0$ and $E[\varepsilon_t \mid F_t] = 0$, constraints that may be written as

$$E[u_t \mid F_t] = \gamma_1 , \tag{36a}$$

and

$$E[\varepsilon_t \mid F_t] = \gamma_2 , \tag{36b}$$

where $\gamma_1 = \gamma_2 = 0$. From (36) we could derive the estimating equations

$$\hat{u}_t = \gamma_1 + u_t + \{ \hat{u}_t - u_t \}, \tag{37a}$$

and

$$\hat{\varepsilon}_t = \gamma_2 + \varepsilon_t + [\hat{\varepsilon}_t - \varepsilon_t \}. \tag{37b}$$

Regressing $\hat{u}_t$ and $\hat{\varepsilon}_t$ against a constant yields

$$\hat{\gamma}_1 = T^{-1}\sum_{t=1}^{T} \hat{\mathbf{s}}_t \quad , \quad \hat{\gamma}_2 = T^{-1}\sum_{t=1}^{T} \hat{\mathbf{s}}_t$$

which correspond to the m_μ and m_h statistics analyzed in Corollary 1.2. The variances of the $\hat{\gamma}_j$ depend upon two factors: the errors u_t and ε_t and the difference between these and their estimated values i.e. the terms given in curly brackets. As shown in Corollary 1.2, the OLS standard errors accompanying $\hat{\gamma}_j$ overstate the true standard errors. This means that any t-statistic for $\gamma_j = 0$ from the regressions described above will be biased in favour of the null hypothesis. Still, such regressions can provide quick checks of the adequacy of the variance and mean specifications since, if $H_0 : \gamma_j = 0$ is rejected, this conclusion would not be reversed if the correct standard errors were employed.

The argument above extends to the general class of consistency tests embedded in (3). Here the restriction $E[u_t \mid F_t] = 0$, may be expressed as

$$E[\upsilon_t \mid F_t] = \Phi_t'\gamma \tag{38}$$

where $\gamma = 0$. From (38) we could derive the estimating equations

$$\hat{u}_t = \hat{\Phi}_t'\gamma + u_t + \{\hat{\upsilon}_t - \upsilon_t\} + \{\Phi_t - \hat{\Phi}_t\}'\gamma,$$

so the regression of $\hat{u}_t$ and $\hat{\varepsilon}_t$ would be against $\hat{\Phi}_{1t}$ and $\hat{\Phi}_{2t}$ respectively ($\Phi_t' = (\Phi_{1t}', \Phi_{2t}')$ is partitioned conformably with u_t). To ensure that Φ_t is a function only of information available after MLE it will need to be made a function of either h_t or perhaps past values of $\hat{u}_t$ and $\hat{\varepsilon}_t$. Selection of $\Phi_{jt} = h_t^{-1}$ produces tests based upon $\sum_{t=1}^{T} \hat{h}_t^{-1}\hat{u}_t$ and $\sum_{t=1}^{T} (\hat{h}_t^{-1}\hat{u}_t^2 - 1)$, statistics that appear as outputs in many ARCH programs. Setting $\Phi_{2t} = h_t$ would effectively produce a test based upon whether the coefficient of $\hat{h}_t$ in the regression of $\hat{u}_t$ against $\hat{h}_t$ was equal to unity; the value the coefficient would be if the model was correctly specified. The LM test normally involves constructing Φ_t from data outside of the model (as well as imposing the restriction that $\gamma_1 = \gamma_2 = 0$), and it has the potential to be the best diagnostic if the chosen Φ_t correlates well with the specification error. Conversely, it may be very poor if this is not true.

Fundamentally, what the re-interpretation offered above is meant to do is to focus attention upon the innovations $\hat{u}_t$ and $\hat{\varepsilon}_t$. An analysis of these quatities should be the primary mode of detecting specification errors in heteroskedastic models. The LM tests correlates them with other series, just as the various consistency tests do, but it is not always the case that this is the best procedure. Sometimes greater insight may be available from graphical methods or even the recursive estimates of the associated $\hat{\gamma}_j$.

Engle, Lilien and Robins (1987) (ELR henceforth) investigated the influence of risk premia upon the excess holding yield for 60-day Treasury bills. In terms of (16) y_t is the excess holding yield and $\mu_t(\theta) = x_t\beta$, with x_t containing an intercept, the yield differential between 30 and 60-day bills, and risk premium $\log h_t$. The variable h_t is modelled as the ARCH process

$$\alpha_0 + \alpha_1 \sum_{j=1}^{4} \left(\frac{5-j}{10}\right) u_{t-j}^2 .$$

MLE was applied to this model and a number of LM tests were performed with prescribed Φ_t to assess the adequacy of their preferred equation (eq. (22), p. 402).

A range of consistency tests was computed for ELR's preferred equation. Using the variances estimated in Corollary 1.2, the ratios of m_μ and m_h to their asymptotic standard errors were respectively 0.42 and -1.28, which does not suggest that ELR's chosen model is incorrect. However, inspection of the residuals $\hat{u}_t$ revealed that they were highly non-normal, making a robust estimator of var(m_h) desirable. Because the term involving the covariance matrix of $\hat{\theta}$ had already been computed using robust estimators of its components, the only modification needed was the replacement of $2\hat{h}_t$ by $\hat{u}_t^4 - \hat{u}_t^2$ in the variance formula. When this is done the t-statistic for m_h becomes 2.15, providing some marginal evidence that ELR's selected equation is deficient.

A second set of consistency tests was then performed. These involved testing if the coefficient of $\hat{h}_t$ in the regression of $\hat{u}_t$ and $\hat{\varepsilon}_t$ against $\hat{h}_t$ were zero i.e. γ_1 and γ_2 in a modified version of (36) were estimated and compared to zero. The regression t-statistics, which are biased in favor of the null hypothesis that $\gamma_j = 0$, were -2.82 and -8.64 respectively, which constitutes very strong evidence against the specification adopted by ELR. Hence, it appears that their decision to model the risk premium as an ARCH process is in error.

What is particularly interesting about this situation is that the consistency tests have disclosed a problem in the specification of ELR's model that was not apparent from the broad range of LM tests that ELR employed in their paper. In fact, the estimated coefficient of $\hat{h}_t$ in the regression of $\hat{\varepsilon}_t$ against $\hat{h}_t$ was -0.67 instead of zero i.e. the regression of $\hat{u}_t^2$ against $\hat{h}_t$ would yield a coefficient of 0.32 rather than the theoretical value of unity it sould have if the ARCH specification was correct. Exactly what one might do to improve the specification of the risk premium for the excess holding yield is an open question, but this example should have served to highlight the fact that the consistency tests advocated in this paper can provide crucial information about the adequacy of any modelling exercise involving heteroskedastic error terms.

6. Concluding Remarks

In this paper we have proposed a class of consistency tests for heteroskedastic and risk models. The tests are easy to compute, have a clear intuitive explanation in terms of residual analysis, and may be constructed without using any information external to the model. Thus performing a set of alternative consistency tests may produce valuable information for model specification. Local power has been analyzed and we have provided an example in which a wide range of LM tests for specific departures failed to detect model inadequacy, whereas the simplest vesions of the consistency tests point to some source of specification error. Furthermore, by suitable definition of the consistency tests, they can also be used to test for specific departures, as the LM test for variable additions has been shown to belong to the class. We have also related the consistency tests to other testing procedures in the literature and thus have provided a simple and general framework for model diagnostic in heteroskedastic and risk models.

References

Amemiya, T. (1973). "Regression Analysis when the Variance of the Dependent Variable is Proportional to the Square of its Expectation", *Journal of the American Statistical Association*, vol. 68, pp. 928-934.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, vol. 31, pp. 307-327.

Breusch, and L. G. Godfrey (1986). "Data Transformation Tests", *The Economic Journal*, vol. 96, pp. 47-58.

———, T. S., and A. R. Pagan (1980). "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics", *Review of Economic Studies*, vol. 47, pp. 239-253.

Carroll, R. J. (1982). "Adapting for Heteroscedasticity in Linear Models", *Annals of Statistics*, vol. 10, pp. 1224-1233.

———, and D. Ruppert (1982). "A Comparison between Maximum Likelihood and Generalized Least Squares in a Heteroscedastic Linear Model", *Journal of the American Statistical Association*, vol. 77, pp. 878-882.

Davidson, R., and J. G. MacKinnon (1984). "Model Specification Tests Based on Artificial Linear Regressions", *International Economic Review*, vol. 25, pp. 485-502.

Domowitz, I., and C. S. Hakkio (1985). "Conditional Variance and the Risk Premium in the Foreign Exchange Market", *Journal of International Economics*, vol. 19, pp. 47-66.

———, and H. White (1982). "Misspecified Models With Dependent Observations", *Journal of Econometrics*, vol. 20, pp. 35-58.

Engle, R. F. (1982a). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, vol. 50, pp. 987-1007.

——— (1982b). "A General Approach to Lagrange Multiplier Model Diagnostics", *Journal of Econometrics*, vol. 20, pp. 83-104.

——— (1983). "Estimates of the Variance of U.S. Inflation Based upon the ARCH Model", *Journal of Money, Credit, and Banking*, vol. 15, pp. 286-301.

——— (1984). "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics", in Z. Griliches and M. D. Intrilligator (eds.), *Handbook of Econometrics*, vol. 2, Elsevier Science Publishers.

———, and T. Bollerslev (1986). "Modelling the Persistence of Conditional Variances", *Econometric Reviews*, vol. 5, pp. 1-50.

———, D. F. Hendry, and J. F. Richard (1983). "Exogeneity", *Econometrica*, vol. 51, pp. 277-304.

———, D. M. Lilien, and R. P. Robins (1987). "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model", *Econometrica*, vol. 55, pp. 391-407.

Geweke, J. (1986). "Comment", *Econometric Reviews*, vol. 5, pp. 57-61.

Hansen, L. P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, vol. 50, pp. 1029-1054.

Hausman, J. A. (1978). "Specification Tests in Econometrics", *Econometrica*, vol. 46, pp. 1251-1271.

______, Z. Griliches and B. H. Holland (1984). "Econometric Models for Count Data with an Application to the Patents R and D Relationship", *Econometrica*, vol. 52, pp. 909-938.

Jennrich, R. I. (1969). "Asymptotic Properties of Non-Linear Least Squares Estimators", *Annals of Mathematical Statistics*, vol. 40, pp. 633-643.

Kraft, D. F., and R. F. Engle (1982). "Autoregressive Conditional Heteroscedasticity in Multiple Time Series Models", *UCSD Discussion Paper*, pp. 82-23.

Newey, W. D. (1985a). "Maximum Likelihood Specification Testing and Conditional Moments", *Econometrica*, vol. 53, pp. 1047-1070.

______ (1985b). "Generalized Method of Moments Specification Testing", *Journal of Econometrics*, vol. 29, pp. 229-256.

______ (1986). "Efficient Estimation of Models with Conditional Moment Restrictions", MIT, manuscript.

Pagan, A. R. (1984a). "Model Evaluation by Variable Addition", in D. F. Hendry and K. F. Wallis (eds.), *Econometrics and Quantitative Economics*, Basil Blackwell.

______ (1984b). "Econometric Issues in the Analysis of Regressions with Generated Regressors", *International Economic Review*, vol. 25, pp. 221-247.

______, and A. D. Hall (1983). "Diagnostic Tests as Residual Analysis", *Econometric Reviews*, vol. 2, pp. 159-218.

______, and H. C. Sabau (1991). "On the Inconsistency of the MLE in Certain Heteroskedastic Regression Models", *Estudios Económicos*, vol. 6, pp. 159-172.

______, and A. Ullah (1987). "The Econometric Analysis of Models with Risk Terms", University of Rochester, manuscript.

Ramsey, J. B. (1969). "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis", *Journal of the Royal Statistical Society*, Series B, vol. 31, pp. 350-371.

Robinson, P. M. (1987). "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form", *Econometrica*, vol. 55, pp. 875-891.

Sabau, H. C. (1987a). "The Structure of the Maximum Likelihood Estimator in Heteroskedastic Normal Conditional Models", ANU, manuscript.

______ (1987b). "Testing Normality in Heteroskedastic Models", ANU, manuscript.

Tauchen, G. (1985). "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *Journal of Econometrics*, vol. 30, pp. 415-443.

Weiss, A. A. (1984). "ARMA Model with ARCH Errors", *Journal of Time Series Analysis*, vol. 5, pp. 129-143.

White, H. (1980). "Using Least Squares to Approximate Unknown Regression Functions", *International Economic Review*, vol. 21, pp. 149-170.