REGRESSION BASED TESTS OF HETEROSCEDASTICITY IN MODELS WHERE THE VARIANCE DEPENDS ON THE MEAN

A. Colin Cameron

Working Paper Series
#379

Department of Economics
University of California
Davis, California 95616-8578

REGRESSION BASED TESTS OF HETEROSCEDASTICITY IN MODELS WHERE THE VARIANCE DEPENDS ON THE MEAN

A. Colin Cameron

Working Paper Series
#379

Department of Economics
University of California
Davis, California 95616-8578

REGRESSION BASED TESTS OF HETEROSCEDASTICITY IN MODELS WHERE THE VARIANCE DEPENDS ON THE MEAN

A. Colin Cameron

Working Paper Series No. 379
March 1991

Abstract

In many regression models, the variance conditional on regressors depends in part on the conditional mean. For example, generalized linear models and power of the mean models have this property. For such models, this paper proposes tests of the assumed heteroscedasticity against an alternative more general form of heteroscedasticity, conditional on correct specification of the mean. Given parameter estimates under the null hypothesis, the tests are simply implemented by univariate OLS regressions. When the variance is a function of the mean alone, one regression is required, while if the variance is an unknown scalar multiple of a function of the mean alone, two regressions are required. These tests require specification of the first two moments, rather than the complete distributions, under the null and alternative hypotheses. When distributions are specified, the optimal regression based test coincides with the score test for several examples in the linear exponential family, but in other examples the two test procedures may lead to different tests.

Some key words: heteroscedasticity; overdispersion; variance-mean relationships; heterogeneity; score tests; lagrange multiplier tests; moment specification tests; generalized linear models; linear exponential family.

Acknowledgements: The author thanks Pravin Trivedi for very helpful discussions and comments.

1. INTRODUCTION

In many regression models, the variance conditional on regressors depends in part on the conditional mean. For example, generalized linear models and power of the mean models have this property. This paper proposes tests of the assumed conditional variance-mean relationship of such models, against an alternative variance-mean relationship, conditional on correct specification of the mean. Thus, rather than testing homoscedasticity against heteroscedasticity, we test one form of heteroscedasticity against an alternative more general form of heteroscedasticity. In specific examples such as the Poisson and binomial such tests are often called tests of over- or under-dispersion.

The dependence of the conditional variance on the conditional mean, rather than on an unrestricted function of the regressors, can complicate inference. For example, see Carroll and Ruppert (1988). Simplification is achieved here by considering tests based on a particular linear combination of residuals and squared residuals, rather than only squared residuals. Then asymptotic inference using estimated residuals coincides, under the null hypothesis, with inference using unobserved true errors, if the conditional variance depends on the conditional mean alone. If the conditional variance depends on both the conditional mean and additional parameters, allowance needs to be made only for estimation of these additional unknown parameters.

The tests are "regression based" in the sense explained in the body of the paper, and introduced by Cameron and Trivedi (1990) who considered tests of variance-mean equality. Implementation of the tests, given parameter estimates under the null hypothesis, is achieved by univariate OLS regressions. When the variance is a function of the mean alone, one such regression is required. If the variance is an unknown scalar multiple of a function of the mean only, two regressions are required. These cases cover many of the generalized linear models in McCullagh and Nelder (1989).

The tests require specification of the first two moments under the null and alternative hypotheses, rather than the complete distributions under the null and alternative hypotheses. Furthermore, the test statistics are of correct size given only the first two moments assumed under the null hypothesis.

With the additional specification of the third and fourth moments under the null hypothesis, an optimal test in the class of regression-based test can be analytically obtained.

For a number of leading cases of fully parameterized models, this optimal test coincides with score tests. These cases include: score tests of variance constancy in the regression model under normality (Breusch and Pagan (1979), Cook and Weisberg (1983)), score tests of variance-mean equality in the Poisson regression model (Collings and Margolin (1985), Lee (1986), Cameron and Trivedi (1986), Dean and Lawless (1989)), and score tests of extra-binomial variation in binary data (Tarone (1979), Prentice (1986)). More generally, however, the optimal regression-based test does not always coincide with the score test. And even when the regression-based test coincides with a score test, the score test may be implemented by a different method requiring stronger stochastic assumptions.

In section 2, the general theory and test procedures are presented. Specific examples and discussion of the related literature on conditional moment tests is given in section 3. Fully parameterized models are introduced in section 4, where the tests are compared to score tests and the overdispersion test of Cox (1983). Some concluding remarks are made in section 5.

2. Regression Based Tests for Variance-Mean Relationships

2.1 General Framework

The data $\{(y_i, X_i), i = 1,\dots,T\}$ are independent across i. Conditional on the vector of exogenous explanatory variables X_i, the mean of the scalar dependent variable y_i is

$$(2.1) \qquad \mathbb{E}_0[y_i \mid X_i] = \mu_i \equiv \mu(X_i,\beta) ,$$

where the subscript 0 denotes expectation under the null hypothesis, $\mu(\cdot)$ is a differentiable function that is first-order identifiable, i.e., $\mu(X_i, \beta_1) = \mu(X_i, \beta_2) \iff \beta_1 = \beta_2$, and β is a kx1 vector of parameters.

Consider models where the variance of y_i conditional on X_i depends on X_i in part through μ_i. This is a parsimonious model for heteroscedasticity, and encompasses a large range of models such as generalized linear models and power of the mean models. Under the null hypothesis,

$$(2.2) \qquad H_0\text{: } \mathrm{Var}_0(y_i \mid X_i) = v(\mu_i,X_i,\phi) ,$$

where the function $v(\cdot)$ is specified, different subcomponents of X_i may appear in the right-hand sides of (2.1) and (2.2), and ϕ is a nx1 vector of nuisance parameters unrelated to β. In many applications ϕ appears multiplicatively, i.e. $v(\mu_i,X_i,\phi) = f(X_i,\phi)\cdot h(\mu_i,X_i)$, even more simply as a multiplicative scalar, $\phi\cdot h(\mu_i,X_i)$, and in some cases does not appear at all.

Under the alternative hypothesis, the conditional mean is still correctly specified:

$$(2.3) \qquad \mathbb{E}_1[y_i \mid X_i] = \mu_i \equiv \mu(X_i,\beta) .$$

The assumed variance-mean relationship in (2.2) is to be tested against the following specific alternative hypothesis that

$$(2.4) \qquad H_1\colon \mathrm{Var}(y_i \mid X_i) = v(\mu_i,X_i,\phi) + g(\mu_i,X_i,\phi)'\alpha ,$$

where $g(\cdot)$ is a px1 function and α is a px1 parameter vector. In applications $g(\mu_i,X_i,\phi)$ is often a function of μ_i alone or X_i alone.[1] Tests of the variance-mean relationship (2.2) are tests of whether $\alpha = 0$ in (2.4).

A trivial example of (2.4) is that the null hypothesis variance is misspecified by a scalar multiple, in which case $g(\mu_i,X_i,\phi)$ is a scalar equal to $v(\mu_i,X_i,\phi)$. For models with quadratic variance-mean relationships, i.e. for models where $v(\mu_i,X_i,\phi) = v(\mu_i,\phi) = \phi_0 + \phi_1\mu_i + \phi_2\mu_i^2$, (2.4) arises from the following form of parameter heterogeneity. Let y_i conditional on γ_i have mean γ_i and variance $v(\gamma_i,\phi)$, where γ_i is a random variable. If, conditional on X_i, γ_i has mean μ_i and variance $g(\mu_i,X_i,\phi)'\delta$, then the variance of y_i conditional on X_i equals $v(\mu_i,\phi) + (\phi_2+1)\cdot g(\mu_i,X_i,\phi)'\delta$, which is exactly of the form (2.4). Thus (2.4) encompasses any mixture model for the Poisson, binomial, gamma, normal (constant variance) and the geometric, provided only that the mixing distribution has a finite variance.[2] More generally, suppose under H_1, $\mathrm{Var}(y_i \mid X_i) = h(\mu_i,X_i,\phi,\lambda)$ where $h(\mu_i,X_i,\phi,\lambda^*) = v(\mu_i,X_i,\phi)$. Then by first-order Taylor series expansion about $\lambda = \lambda^*$, $h(\mu_i,X_i,\phi,\lambda) = v(\mu_i,X_i,\phi) + \nabla_\lambda h(\mu_i,X_i,\phi,\lambda^*)\cdot(\lambda - \lambda^*)$. This is of the form (2.4), with remainder term in the Taylor series expansion that will disappear asymptotically for local alternatives $\lambda - \lambda^* = O(T^{-1/2})$.

An example of the above framework is testing variance-mean equality. Then $v(\mu_i,X_i,\phi)$ equals μ_i; μ_i is usually set to $\exp(X_i'\beta)$; $g(\mu_i)$ is usually set to 1, μ_i, or μ_i^2; and distributions are usually specified for y_i under both H_0, the Poisson, and H_1, most often the negative binomial obtained by assuming a gamma distribution for γ_i conditional on μ_i. Most previous studies are similarly specialized: a specific variance-mean relationship is studied; often the functional form for the mean is given; a specific functional form for the alternative variance-mean relationship is studied; and the distribution of y_i under both H_0 and H_1 is specified. The framework here is considerably more general.

We begin with the simplest case of no nuisance parameters, i.e. ϕ in (2.2) is known or is of null dimension. Nuisance parameters are deferred to section 2.4.

2.2 Regression Based Tests without Nuisance Parameters

We wish to obtain a test of the variance-mean relationship that applies to as broad a class of models as possible, relies on relatively weak distributional assumptions, and is simple to implement. In this and the next sub-section, the variance is assumed to depend on the mean alone.

From (2.4) the model under H_1 implies:

$$\mathbb{E}_1[\{(y_i - \mu_i)^2 - v(\mu_i)\} \mid X_i] = \alpha \cdot g(\mu_i) ,$$

where $v(\mu_i,X_i)$ and $g(\mu_i,X_i)$ are abbreviated to $v(\mu_i)$ and $g(\mu_i)$. This moment condition suggests testing H_0 by testing the significance of α in the regression of $\{(y_i - \mu_i)^2 - v(\mu_i)\}$ on $g(\mu_i)$. Since the parameters β are unknown, to operationalize this the terms in μ_i in the regression need to be evaluated at an estimate of β, leading to a considerably more complicated asymptotic distribution.

Instead, we introduce a moment condition similar to that above, except that the resulting regression is asymptotically unaffected by replacing β by a root-T consistent estimate. Specifically, since the mean is assumed to be still correctly specified under H_1, the model under H_1 implies:

$$(2.5) \qquad \mathbb{E}_1[\{(y_i - \mu_i)^2 - \nabla_\mu v(\mu_i)(y_i - \mu_i) - v(\mu_i)\} \mid X_i] = \alpha \cdot g(\mu_i) ,$$

where $\nabla_\mu v(\mu_i)$ is the scalar derivative of $v(\mu_i)$ w.r.t. μ_i. This choice of moment condition as the basis of tests is not immediately obvious. An explanation is given after statement of Proposition 1.

Define

$$(2.6) \qquad y_i^* = (y_i - \mu_i)^2 - \nabla_\mu v(\mu_i)(y_i - \mu_i) - v(\mu_i) .$$

If μ_i is observed, we can test H_0 by testing the significance of α in the regression

$$(2.7) \qquad y_i^* = g(\mu_i)'\alpha + \varepsilon_i,$$

where the error term $\varepsilon_i = y_i^* - \mathbb{E}[y_i^* \mid X_i]$ is possibly heteroscedastic. We therefore consider weighted least squares estimation, with weights $w_i = w(\mu_i, X_i)$.

In implementation, we estimate by OLS (without intercept):

$$(2.8) \qquad \hat{w}_i^{1/2} \cdot \hat{y}_i^* = \hat{w}_i^{1/2} \cdot g(\hat{\mu}_i)'\alpha + u_i ,$$

where $\hat{\mu}_i = \mu(X_i, \hat{\beta}_T)$, $\hat{w}_i = w(\hat{\mu}_i, X_i)$ and $\hat{y}_i^* \equiv (y_i - \hat{\mu}_i)^2 - \nabla_\mu v(\hat{\mu}_i)(y_i - \hat{\mu}_i) - v(\hat{\mu}_i)$. The least squares estimator $\hat{\alpha}_T$ is:

$$(2.9) \qquad \hat{\alpha}_T = (\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}\hat{y}^* ,$$

where the Txp matrix $\hat{G}$ has *ith* row $g(\hat{\mu}_i)'$, $\hat{W}$ is a TxT diagonal matrix with *ith* entry $\hat{w}_i$, and the Tx1 vector $\hat{y}^*$ has *ith* entry $\hat{y}^*_i$.

Tests using $\hat{\alpha}_T$ in (2.9) are called regression based tests, as the motivation is testing $\alpha = 0$ in the regression (2.5). The asymptotic distribution of $\hat{\alpha}_T$ is given in Proposition 1.

<u>Proposition 1</u>: Let $\mathrm{Var}_0(y_i \mid X_i) = v(\mu_i, X_i)$ be a function of μ_i (and X_i) alone, and $\hat{\beta}_T$ be a root-T consistent estimator of β under H_0. Subcomponents of H_1 in (2.4) can be tested using the result that under H_0:

$$(2.10) \qquad \hat{\alpha}_T \overset{A}{\sim} N(0,\ (\hat{G}'\hat{W}\hat{G})^{-1}(\hat{G}'\hat{W}\tilde{\Sigma}\hat{W}\hat{G})(\hat{G}'\hat{W}\hat{G})^{-1}\ ) ,$$

where Σ is a diagonal matrix with *ith* entry

$$(2.11) \qquad \sigma_i^2 = \mathrm{Var}_0(y^*_i \mid X_i) ,$$

and $\tilde{\Sigma}$ is a matrix such that $T^{-1}(\hat{G}'\hat{W}\tilde{\Sigma}\hat{W}\hat{G} - G'W\Sigma WG) \overset{p}{\to} 0$, where G has *ith* entry $g(\mu_i)'$, and W has *ith* entry $w(\mu_i)$. ■

Note that σ_i^2 is the variance of y^*_i, which from (2.6) is quadratic in y_i. It is not the variance of y_i. The notation $\tilde{\Sigma}$ is used to indicate that $\tilde{\Sigma}$ need not be consistent for Σ. $\tilde{\Sigma}$ may be a diagonal matrix with *ith* entry $\tilde{\Sigma}_{ii} = (\hat{y}^*_i - g(\hat{\mu}_i)'\hat{\alpha}_T)^2$, the squared unweighted residual, or $\tilde{\Sigma}_{ii} = (\hat{y}^*_i)^2$, since $\alpha = 0$ under H_0.

The derivation of this proposition, and subsequent ones, is given in the Appendix. It essentially states that (2.8) can be treated as a regular regression equation, ignoring the dependency of $\hat{\mu}_i$ on $\hat{\beta}_T$, and can be explained as follows. Let $\hat{\beta}_T$ be a root-T consistent estimator of β, and $m(y_i,X_i,\beta)$ be a px1 vector function. Then by a Taylor series expansion:

$$T^{-1/2}\sum_{i=1}^{T} m(y_i,X_i,\hat{\beta}_T) = T^{-1/2}\sum_{i=1}^{T} m(y_i,X_i,\beta)$$
$$+ T^{-1}\sum_{i=1}^{T} \nabla_\beta m(y_i,X_i,\beta)\cdot T^{1/2}(\hat{\beta}_T - \beta) + o_p(1) ,$$

where $\nabla_\beta m(y_i,X_i,\beta)$ is the pxk matrix of derivatives with *jth* column $\nabla_{\beta_j} m(y_i,X,\beta)'$. The second term disappears asymptotically if $\mathbb{E}[\nabla_\beta m(y_i,X_i,\beta) \mid X_i] = 0$. If we consider the initially suggested regression, unweighted for simplicity, of $\{(y_i - \hat{\mu}_i)^2 - v(\hat{\mu}_i)\}$ on $g(\hat{\mu}_i)$, the OLS estimator equals $(\hat{G}'\hat{G})^{-1}$ times $\sum_{i=1}^{N} m(y_i,X_i,\hat{\beta}_T)$, where $m(y_i,X_i,\beta) = g(\mu_i)\{(y_i - \mu_i)^2 - v(\mu_i)\}$. Then $\mathbb{E}_0[\nabla_\beta m(y_i,X_i,\beta) \mid X_i] = - g(\mu_i)\cdot\nabla_\mu v(\mu_i)\cdot\nabla_\beta\mu(X_i,\beta) \neq 0$, except in the case of homoscedasticity when $\nabla_\mu v(\mu_i) = 0$. To handle the case of nonconstant variance we note that we can add a multiple $h(\mu_i)$ of $(y_i - \mu_i)$ to the dependent variable in the regression, since $\mathbb{E}[(y_i - \mu_i) \mid X_i] = 0$, in which case $m(y_i,X_i,\beta) = g(\mu_i)\{(y_i - \mu_i)^2 + h(\mu_i)(y_i - \mu_i) - v(\mu_i)\}$. The choice $h(\mu_i) = - \nabla_\mu v(\mu_i)$ ensures $\mathbb{E}_0[\nabla_\beta m(y_i,X,\beta) \mid X_i] = 0$.

The covariance matrix for $\hat{\alpha}_T$ in (2.10) allows for the possibility of a heteroscedastic error in the regression, following the heteroscedastic consistent estimate of White (1980).

Subcomponents of the alternative hypothesis (2.4) can be tested by t-tests for individual components of α equaling zero, or F-tests for subsets of α equaling zero. This can be directly done by estimating (2.8) using the heteroscedastic consistent variance-covariance option available on many packages, which uses $\tilde{\Sigma}_{ii} = (y_i^* - g(\hat{\mu}_i)'\hat{\alpha}_T)^2$. If this is not available, an instrumental variables package can be used along the lines suggested by

Domowitz (1983).

The joint test of whether all components of α equal zero is given by the chi-square test statistic:

$$(2.12) \qquad \hat{d}_T = \hat{y}^{*\prime}\hat{W}\hat{G}(\hat{G}'\hat{W}\tilde{\Sigma}\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}\hat{y}^* .$$

Under H_0, $\hat{d}_T$ is asymptotically chi-square distributed with p degrees of freedom.

This chi-square test statistic, when $\tilde{\Sigma}_{ii} = (\hat{y}_i^*)^2$, can alternatively be computed from an auxiliary regression, as given in Proposition 2.

<u>Proposition 2</u>: Let $\mathrm{Var}_0(y_i \mid X_i) = v(\mu_i, X_i)$ be a function of μ_i (and X_i) alone, and $\hat{\beta}_T$ be a root-T consistent estimator of β under H_0. An overall test of H_0 versus H_1 can be implemented using the result that under H_0:

T times R_u^2 (uncentered R^2) from the auxiliary OLS regression (without intercept) of

$$1 \quad \text{on} \quad g(\hat{\mu}_i)\cdot\hat{w}_i\cdot\{(y_i - \hat{\mu}_i)^2 + \nabla_\mu v(\hat{\mu}_i)(y_i - \hat{\mu}_i) - v(\hat{\mu}_i)\}$$

is asymptotically chi-square distributed with p degrees of freedom. ■

For this and later regressions with 1 as the dependent variable, TR_u^2 can also be computed as T minus the residual sum of squares. Such regressions are auxiliary in the sense that they are a way to compute the test statistic. By contrast regression (2.8) has an intrinsic interpretation.

2.3 Optimal Regression Based Tests without Nuisance Parameters

Different choices of the weighting matrix lead to different test statistics. The optimal test using the regression (2.7), i.e. when β is known, will use the generalized least squares (GLS) estimator. Since inference for the regression (2.8) coincides with that for the regression (2.7) under H_0, we accordingly expect the optimal test within the class of regression based tests to use the feasible GLS estimator in (2.8), with weighting matrix $\hat{W} = \hat{\Sigma}^{-1}$. Define

$$(2.13) \qquad \hat{\alpha}_T^{opt} = (\hat{G}'\hat{\Sigma}^{-1}\hat{G})^{-1}\hat{G}'\hat{\Sigma}^{-1}\hat{y}^* ,$$

the OLS estimator from the regression:

$$(2.14) \qquad \hat{\sigma}_i^{-1}\cdot\hat{y}_i^* = \hat{\sigma}_i^{-1}\cdot g(\hat{\mu}_i)'\cdot\alpha + u_i ,$$

where $\hat{\Sigma}$ is a TxT matrix consistent for Σ, with *ith* diagonal entry $\hat{\sigma}_i^2$ equaling σ_i^2 evaluated at $\hat{\beta}_T$.

We consider the limit distribution of $T^{1/2}\cdot\hat{\alpha}_T$ under local alternatives H_L: $\alpha = T^{-1/2}\delta$, i.e.

$$(2.15) \qquad H_L: \mathrm{Var}(y_i \mid X_i) = v(\mu_i) + g(\mu_i)'(T^{-1/2}\delta), \quad \delta \text{ a finite constant.}$$

The optimality of tests based on $\hat{\alpha}_T^{opt}$ within the class of regression-based tests for H_0 versus H_L is given in Proposition 3.

<u>Proposition 3</u>: Let $\mathrm{Var}_0(y_i \mid X_i) = v(\mu_i,X_i)$ be a function of μ_i (and X_i) alone, $\hat{\beta}_T$ be a root-T consistent estimator of β under H_L, and σ_i^2 defined in (2.11) be correctly specified. Then the optimal regression based test is implemented by the OLS regression (2.14). Under H_0 and given σ_i^2:

$$(2.16) \qquad \hat{\alpha}_T^{opt} \overset{A}{\sim} N(0, (\hat{G}'\hat{\Sigma}\hat{G})^{-1}) .$$

■

The covariance matrix in (2.16) indicates that subcomponents of H_1 given in (2.4) and the overall hypothesis that $\alpha = 0$ can be tested using the usual computer output from this regression. However, σ_i^2 needs to be correctly specified, which requires correct specification of the first four moments of the distribution of y_i under H_0. To guard against possible misspecification of σ_i^2, we can of course apply the tests of section 2.2, using $\hat{w}_i = \hat{\sigma}_i^{-2}$. This conservative implementation of the optimal regression based test requires correct specification of only the first two moments of y_i under H_0.

The optimality is within a very restricted class of tests, those using a weighted sum of the particular linear combination of residuals and squared residuals defined in (2.6). Nonetheless, in the fully parametric case where the entire density is specified, the optimal regression-based test coincides with a score test in several leading examples given in section 4.2 and then enjoys the optimality properties of the classical testing procedures. In other examples, however, the score test differs from the optimal regression based test.

2.4 Regression Based Tests with Nuisance Parameters

In this sub-section we consider the more complex case where the n nuisance parameters ϕ in (2.2) need to be estimated. Regression-based tests again use the regression (2.8), with estimator $\hat{\alpha}_T$ defined in (2.9). The only change is that (2.6) becomes

$$(2.17) \qquad y_i^* = (y_i - \mu_i)^2 - \nabla_\mu v(\mu_i,\phi)(y_i - \mu_i) - v(\mu_i,\phi) ,$$

with corresponding change to $\hat{y}_i^*$.

The asymptotic theory is again unaffected by replacing β by a root-T consistent estimator, but different root-T consistent estimators for $\hat{\phi}_T$ lead to different asymptotic distributions for $\hat{\alpha}_T$. Computation is simpler for a particular choice of $\hat{\phi}_T$, but for completeness we first give a general result.

Let $\hat{\phi}_T$ be a root-T consistent estimator for ϕ such that under H_0,

$$(2.18) \qquad (\hat{\phi}_T - \phi) = (A'A)^{-1} \cdot B'z + o_p(T^{-1/2})$$

where A and B are Txn matrices with *ith* rows $A_i(X,\beta,\phi)'$ and $B_i(X,\beta,\phi)'$ and z is a Tx1 vector with *ith* entry $z_i(y_i,X_i,\beta,\phi)$. This representation is reasonably general in that a first-order Taylor series expansion of the non-linear estimating equations defining a root-T consistent estimator $\hat{\phi}_T$ will yield (2.18) for some A, B and z.

<u>Proposition 4</u>: Let $Var_0(y_i \mid X_i) = v(\mu_i,X_i,\phi)$ be a function of μ_i, X_i and the n nuisance parameters ϕ, and $\hat{\beta}_T$ be a root-T consistent estimator of β under H_0. Let $\hat{\phi}_T$ be the root-T consistent estimator for ϕ defined in (2.18). Subcomponents of H_1 given in (2.4) can be tested using the result that under H_0:

$$(2.19) \qquad \hat{\alpha}_T \overset{A}{\sim} N(0,\ (\hat{G}'\hat{W}\hat{G})^{-1}\ \{\ \hat{G}'\hat{W}\tilde{\Sigma}\hat{W}\hat{G} + \hat{G}'\hat{W}\hat{D}(\hat{A}'\hat{A})^{-1}\hat{B}'\tilde{\Sigma}_{22}\hat{B}(\hat{A}'\hat{A})^{-1}\hat{D}\hat{W}\hat{G}$$
$$- \hat{G}'\hat{W}\tilde{\Sigma}_{12}\hat{B}(\hat{A}'\hat{A})^{-1}\hat{D}\hat{W}\hat{G} - \hat{G}'\hat{W}\hat{D}(\hat{A}'\hat{A})^{-1}\hat{B}'\tilde{\Sigma}_{21}\hat{W}\hat{G}\ \} \cdot (\hat{G}'\hat{W}\hat{G})^{-1}\ ,$$

where D is a Txn matrix with *ith* row $d_i' = \mathbb{E}_0[\nabla_\phi y_i^*] = \nabla_\phi v(\mu_i,\phi)$;
Σ, Σ_{22} and $\Sigma_{12} = \Sigma_{21}'$ are TxT diagonal matrices with *ith* entries $\sigma_i^2 =$ $Var_0(y_i^* \mid X_i)$, $Var_0(z_i \mid X_i)$ and $Cov_0(y_i^*,z_i \mid X_i)$;
$\hat{D}$ is a Txn matrix with *ith* row $\hat{d}_i' = \nabla_\phi v(\hat{\mu}_i,\hat{\phi}_T)$;

and $\tilde{\Sigma}$, $\tilde{\Sigma}_{22}$ and $\tilde{\Sigma}_{12}$ are matrices such that $T^{-1}(\hat{G}'\hat{W}\tilde{\Sigma}\hat{W}\hat{G} - G'W\Sigma WG) \xrightarrow{p} 0$, $\lim_{T\to\infty} T^{-1}(\hat{B}'\tilde{\Sigma}_{22}\hat{B} - B'\Sigma_{22}B) \xrightarrow{p} 0$ and $\lim_{T\to\infty} T^{-1}(\hat{G}'\hat{W}\tilde{\Sigma}_{12}\hat{B} - G'W\Sigma_{12}B) \xrightarrow{p} 0$. ■

The overall test of $\alpha = 0$ is given by the test statistic:

$$\hat{d}_T = \hat{y}^{*\prime}\hat{W}\hat{G}\cdot\{\ \hat{G}'\hat{W}\tilde{\Sigma}\hat{W}\hat{G} + \hat{G}'\hat{W}\hat{D}(\hat{A}'\hat{A})^{-1}\hat{B}'\tilde{\Sigma}_{22}\hat{B}(\hat{A}'\hat{A})^{-1}\hat{D}'\hat{W}\hat{G}$$
$$- \hat{G}'\hat{W}\tilde{\Sigma}_{12}\hat{B}(\hat{A}'\hat{A})^{-1}\hat{D}'\hat{W}\hat{G} - \hat{G}'\hat{W}\hat{D}(\hat{A}'\hat{A})^{-1}\hat{B}'\tilde{\Sigma}_{21}\hat{W}\hat{G}\ \}\cdot\hat{G}'\hat{W}y^* . \tag{2.20}$$

Under H_0, $\hat{d}_T$ is asymptotically chi-square distributed with p degrees of freedom. Computation of tests of subcomponents of H_1 or the overall test in (2.20) will generally require access to matrix routines.

We now consider a specialization that permits use of regression routines only. Suppose $(\hat{\phi}_T - \phi) = (D'WD)^{-1}(D'Wy^*) + o_p(T^{-1/2})$, in which case $A'A = D'WD$, $B = D'W$, and $z = y^*$ so that $\Sigma = \Sigma_{22} = \Sigma_{12}$. Then the term in braces in (2.19) and (2.20) becomes:

$$\{\hat{G}'\hat{W}\tilde{\Sigma}\hat{W}\hat{G} + \hat{G}'\hat{W}\hat{D}(\hat{D}'\hat{W}\hat{D})^{-1}\hat{D}'\hat{W}\tilde{\Sigma}\hat{W}\hat{D}(\hat{D}'\hat{W}\hat{D})^{-1}\hat{D}\hat{W}\hat{G} - \hat{G}'\hat{W}\tilde{\Sigma}\hat{W}\hat{D}(\hat{D}'\hat{W}\hat{D}^{-1})\hat{D}\hat{W}\hat{G}$$
$$- \hat{G}'\hat{W}\hat{D}(\hat{D}'\hat{W}\hat{D})^{-1}\hat{D}'\hat{W}\tilde{\Sigma}\hat{W}\hat{G}\ \}$$
$$= \{\hat{G}'\hat{W} - \hat{G}'\hat{W}\hat{D}(\hat{D}'\hat{W}\hat{D})^{-1}\hat{D}'\hat{W}\}\cdot\tilde{\Sigma}\cdot\{\hat{W}\hat{G} - \hat{W}\hat{D}(\hat{D}'\hat{W}\hat{D})^{-1}\hat{D}\hat{W}\hat{G}\}$$
$$= \hat{r}'\hat{W}^{1/2}\tilde{\Sigma}\hat{W}^{1/2}\hat{r}\ , \tag{2.21}$$

where $\hat{r} = \{\hat{W}^{1/2}\hat{G} - \hat{W}^{1/2}\hat{D}\cdot(\hat{D}'\hat{W}\hat{D})^{-1}\hat{D}'\hat{W}\hat{G}\}$ is the $T\text{x}n$ matrix of residuals from the regression of $\hat{w}_i^{1/2}\cdot g(\hat{\mu}_i)$ on $\hat{w}_i^{1/2}\hat{d}_i$. The covariance matrix in (2.19) becomes $(\hat{G}'\hat{W}\hat{G})^{-1}\hat{r}'\hat{W}^{1/2}\tilde{\Sigma}\hat{W}^{1/2}\hat{r}(\hat{G}'\hat{W}\hat{G})^{-1}$, and (2.20) is similarly simplified.

If in addition $\hat{\phi}_T$ is defined to be the estimator that solves the first-order conditions $\hat{D}'\hat{W}\hat{y}^* = 0$, then $\hat{r}'\hat{W}^{1/2}\hat{y}^* = \hat{G}'\hat{W}\hat{y}^*$. Combining this with (2.21) with $\tilde{\Sigma}_{ii} = (\hat{y}_i^*)^2$ permits computation of the overall chisquare test by the following auxiliary regressions.

Proposition 5: Let $\mathrm{Var}_0(y_i \mid X_i) = v(\mu_i, X_i, \phi)$ be a function of μ_i, X_i and nuisance parameters ϕ, and $\hat{\beta}_T$ be a root-T consistent estimator of β under H_0. Let $\hat{\phi}_T$ be the root-T consistent estimator that solves the first-order conditions $\hat{D}'\hat{W}\hat{y}^* = 0$, where $\hat{D}$ is a Txn matrix with *ith* row $\nabla_\phi v(\hat{\mu}_i, \hat{\phi})$ and $\hat{W}$ is the same TxT weighting matrix as used to calculate $\hat{\alpha}_T$. An overall test for H_0 versus H_1 can be performed as follows.

(1) Obtain the nx1 vector of residuals $\hat{r}_i$ from the multivariate OLS regression (without intercept):

$$\hat{w}_i^{1/2} \cdot g(\hat{\mu}_i) \; = \; \hat{w}_i^{1/2} \cdot \; \nabla_\phi v(\hat{\mu}_i, \hat{\phi}) \cdot \gamma \; + \; r_i \; .$$

(2) Obtain R_u^2 (the uncentered R^2), equals T - the residual sum of squares, from the auxiliary OLS regression (without intercept) of

$$1 \quad \text{on} \quad \hat{r}_i \hat{w}_i^{1/2} \{(y_i - \hat{\mu}_i)^2 - \nabla_\mu v(\hat{\mu}_i, \hat{\phi}_T)(y_i - \hat{\mu}_i) - v(\hat{\mu}_i, \hat{\phi}_T)\} \; .$$

Then under H_0, T times R_u^2 is asymptotically chi-square distributed with p degrees of freedom. ■

The regression (1) can be implemented by n univariate OLS regressions, and the regression (2) is a univariate OLS regression. The potentially difficult part is the estimation of $\hat{\phi}_T$. Given the estimator $\hat{\beta}_T$ and hence $\hat{\mu}_i$, $\hat{\phi}_T$ solves the first-order conditions

$$(2.22) \quad 0 = \sum_{i=1}^{T} \nabla_\phi v(\hat{\mu}_i, \hat{\phi}_T)' \, w(\hat{\mu}_i, \hat{\phi}_T) \{(y_i - \hat{\mu}_i)^2 - \nabla_\mu v(\hat{\mu}_i, \hat{\phi}_T)(y_i - \hat{\mu}_i) - v(\hat{\mu}_i, \hat{\phi}_T)\}.$$

$\hat{\phi}_T$ is easily computed when ϕ appears multiplicatively, i.e. $v(\mu_i, X_i, \phi) = f(X_i, \phi) \cdot h(\mu_i, X_i)$. Then $\nabla_\phi v(\mu_i, X_i, \phi) = h(\mu_i) \cdot \nabla_\phi f(X_i, \phi)$ and $\nabla_\mu v(\mu_i, \phi) =$

$f(X_i,\phi)\cdot\nabla_\mu h(\mu_i,X_i)$, and (2.22) becomes:

$$(2.23)\quad 0 = \sum_{i=1}^{T} h(\hat{\mu}_i)\nabla_\phi f(\hat{\phi}_T)'\, w(\hat{\mu}_i)\{(y_i - \hat{\mu}_i)^2 - [h(\hat{\mu}_i) + \nabla_\mu h(\hat{\mu}_i)(y_i - \hat{\mu}_i)]f(\hat{\phi}_T)\}$$

$$= \sum_{i=1}^{T} h^*(y_i,\hat{\mu}_i)h(\hat{\mu}_i)w(\hat{\mu}_i)\{h^*(\hat{\mu}_i)^{-1}(y_i - \hat{\mu}_i)^2 - f(\hat{\phi}_T)\}\nabla_\phi f(\hat{\phi}_T)' \quad ,$$

where $h^*(y_i,\hat{\mu}_i) = h(\hat{\mu}_i) + \nabla_\mu h(\hat{\mu}_i)(y_i - \hat{\mu}_i)$, dependence on X_i is suppressed, and we assume w_i is multiplicative in ϕ or does not depend on ϕ. Then $\hat{\phi}_T$ can be computed by nonlinear weighted least squares regression of $h^*(\hat{\mu}_i)^{-1}(y_i - \hat{\mu}_i)^2$ on $f(\phi_T)$, with weights $h^*(y_i,\hat{\mu}_i)h(\hat{\mu}_i)w(\hat{\mu}_i)$.

The most common case of nuisance parameter is a specialization of (2.23), where ϕ appears as a scalar multiplicative constant, i.e. $v(\mu_i,\phi) = \phi\cdot h(\mu_i)$, so that $f(\phi,X_i) = \phi$ and $\nabla_\phi f(\phi,X_i) = 1$. Then (2.23) can be solved, yielding:

$$(2.24)\qquad \hat{\phi}_T = \Big(\sum_{i=1}^{T} h(\hat{\mu}_i)w(\hat{\mu}_i)[h(\hat{\mu}_i) + \nabla_\mu h(\hat{\mu}_i)(y_i - \hat{\mu}_i)]\Big)^{-1} \cdot \sum_{i=1}^{T} h(\hat{\mu}_i)w(\hat{\mu}_i)(y_i - \hat{\mu}_i)^2 \,.$$

$\hat{\phi}_T$ can be computed by finding the average of each of the scalar quantities in the two sums, or by the LS regression mentioned after (2.23). This estimator differs from customary estimators, except when $h(\mu_i)$ is constant. For example, for generalized linear models with multiplicative nuisance parameter, McCullagh and Nelder (1989) suggest the estimator based on the Pearson chi-square statistic $\tilde{\phi}_T = (T-k)^{-1}\sum_{i=1}^{T} (h(\hat{\mu}_i))^{-1}(y_i - \hat{\mu}_i)^2$. This is close to $\hat{\phi}_T$ when $w(\hat{\mu}_i)$ equals a constant multiple of $h(\hat{\mu}_i)^{-2}$, in which case

$$(2.25)\qquad \hat{\phi}_T = \Big(\sum_{i=1}^{T} [1 + h(\hat{\mu}_i)^{-1}\nabla_\mu h(\hat{\mu}_i)(y_i - \hat{\mu}_i)]\Big)^{-1} \cdot \sum_{i=1}^{T} h(\hat{\mu}_i)^{-1}(y_i - \hat{\mu}_i)^2 \,.$$

A result for the optimal regression-based test with nuisance parameters is not given, as it will vary with the estimator of the nuisance parameters. Nonetheless, a natural choice for the weights function in applying procedures 4 and 5 is $w_i = (\mathrm{Var}_0(y_i^* \mid X_i))^{-1}$.

3. Examples and Discussion

3.1 Examples of Regression-Based Tests

Examples of regression-based tests are given for several leading variance-mean relationships. The estimator $\hat{\beta}_T$ will often be suggested by the particular application. Alternatively, and at times coincidently, a weighted least squares estimator may be used.

Subcomponents of H_1 in (2.4) may be tested using Propositions 1 (no nuisance parameters) and 4 (nuisance parameters). The overall test of $\alpha = 0$ can also be implemented using these propositions, or by using the auxiliary regressions in Propositions 2 (no nuisance parameter) and 5 (a particular estimator of the nuisance parameter). Use of Proposition 3 is deferred to examples in section 4, where the first four moments of y under H_0 are specified.

Example 3.1: Variance Mean Equality. In this case $v(\mu_i,X_i,\phi) = \mu_i$ and $\nabla_\mu v(\mu_i,X_i,\phi) = 1$. Propositions 1 and 2 are applicable, as there are no nuisance parameters. Subcomponents of α can be tested by OLS regression (without intercept) of $\hat{w}_i^{1/2} \cdot \{(y - \hat{\mu}_i)^2 - y_i\}$ on $\hat{w}_i^{1/2} \cdot g(\hat{\mu}_i)$, using a heteroscedastic consistent estimate of the covariance matrix. This test was proposed by Cameron and Trivedi (1990). The joint test for $\alpha = 0$ can also be implemented by auxiliary regression of 1 on $g(\hat{\mu}_i) \cdot \hat{w}_i \cdot \{(y - \hat{\mu}_i)^2 - y_i\}$.

Example 3.2: Constant Variance. In this case $v(\mu_i, X_i, \phi) = \phi$, $\nabla_\mu v(\mu_i, X_i, \phi) = 0$, and $\nabla_\phi v(\mu_i, X_i, \phi) = 1$. The obvious nuisance parameter estimator is $\hat{\phi}_T = T^{-1} \sum_{i=1}^{T} (y - \hat{\mu}_i)^2$, which is (2.24) with $\hat{w}_i$ a constant. In using Proposition 5, the first regression is of $\hat{g}_i = g(\hat{\mu}_i)$ on a constant, so that $\hat{r}_i = \hat{g}_i - \bar{g}$, where $\bar{g} = T^{-1} \sum_{i=1}^{T} g(\hat{\mu}_i)$, and the chisquare test uses TR_u^2 from the auxiliary regression of 1 on $(\hat{g}_i - \bar{g}) \cdot \{(y_i - \hat{\mu}_i)^2 - \hat{\phi}_T\}$.

Example 3.3: Variance Proportional to the Mean. In this case $v(\mu_i, X_i, \phi) = \phi\mu_i$, $\nabla_\mu v(\mu_i, X_i, \phi) = \phi$ and $\nabla_\mu v(\mu_i, X_i, \phi) = \mu_i$. The nuisance parameter estimate is $\hat{\phi}_T = (\sum_{i=1}^{T} \hat{\mu}_i \hat{w}_i y_i)^{-1} \cdot \sum_{i=1}^{T} \hat{\mu}_i \hat{w}_i (y_i - \hat{\mu}_i)^2$. In using Proposition 5, the first regression is of $\hat{w}_i^{1/2} \hat{g}_i$ on $\hat{w}_i^{1/2} \hat{\mu}_i$, with residual $\hat{r}_i$, and the chisquare test uses TR_u^2 from the auxiliary regression of 1 on $\hat{r}_i \hat{w}_i^{1/2} \cdot \{(y_i - \hat{\mu}_i)^2 - \hat{\phi}_T y_i\}$.

Example 3.4: Constant Coefficient of Variation. In this case $v(\mu_i, X_i, \phi) = \phi \cdot \mu_i^2$, $\nabla_\mu v(\mu_i, X_i, \phi) = 2\phi\mu_i$ and $\nabla_\phi v(\mu_i, X_i, \phi) = \mu_i^2$. The nuisance parameter estimator $\hat{\phi}_T = (\sum_{i=1}^{T} \hat{\mu}_i^2 \hat{w}_i [\hat{\mu}_i^2 + 2\hat{\mu}_i (y_i - \hat{\mu}_i)])^{-1} \cdot \sum_{i=1}^{T} \hat{\mu}_i^2 \hat{w}_i (y_i - \hat{\mu}_i)^2$. Proposition 5 is implemented by first regressing $\hat{w}_i^{1/2} \hat{g}_i$ on $\hat{w}_i^{1/2} \hat{\mu}_i^2$, giving residuals $\hat{r}_i$, and then computing TR_u^2 from the auxiliary regression of 1 on $\hat{r}_i \hat{w}_i^{1/2} \{(y_i - \hat{\mu}_i)^2 - 2\hat{\phi}_T \hat{\mu}_i (y_i - \hat{\mu}_i) - \hat{\phi}_T \hat{\mu}_i^2\}$.

Example 3.5: Variance Quadratic in the Mean. We consider the simplest case where $v(\mu_i, X_i, \phi) = \mu_i + \phi \cdot \mu_i^2$ and ϕ is known. Here $\nabla_\mu v(\mu_i, X_i, \phi) = (1 + 2\phi\mu_i)$. Subcomponents of α are tested by OLS regression (without intercept) of $\hat{w}_i^{1/2} \cdot \{(y - \hat{\mu}_i)^2 - (1 + 2\phi\hat{\mu}_i)(y - \hat{\mu}_i) - (\hat{\mu}_i + \phi\hat{\mu}_i^2)\}$ on $\hat{w}_i^{1/2} \cdot g(\hat{\mu}_i)$, using a heteroscedastic consistent estimate of the covariance matrix. The joint test of $\alpha = 0$ can also be implemented by the auxiliary regression of 1 on $g(\hat{\mu}_i) \cdot \hat{w}_i \cdot \{(y - \hat{\mu}_i)^2 - (1 + 2\phi\hat{\mu}_i)(y - \hat{\mu}_i) - (\hat{\mu}_i + \phi\hat{\mu}_i^2)\}$.

Example 3.6: Variance a Power of the Mean. In this case $v(\mu_i,X_i,\phi) = \phi_1\mu_i^{\phi_2}$, $\nabla_\mu v(\mu_i,X_i,\phi) = \phi_1\phi_2\mu_i^{\phi_2-1}$, $\nabla_{\phi_1} v(\mu_i,X_i,\phi) = \mu_i^{\phi_2}$, and $\nabla_{\phi_2} v(\mu_i,X_i,\phi) = \phi_1\mu_i^{\phi_2}\log \mu_i$. Application of Proposition 5 is straightforward when ϕ_2 is known, while the more awkward Proposition 4 is needed if ϕ_2 is estimated.

The above examples cover many commonly used parametric models: the Poisson (example 3.1), normal with constant variance σ^2 (example 3.2 with $\phi = \sigma^2$), gamma (example 3.4), binomial with m trials (example 3.5 with $\phi = m^{-1}$), geometric (example 3.5 with $\phi = 1$), normal with variance a power of the mean (example 3.6). Example 3.3 is a test of whether the simplest correction for over- or under- dispersion in the Poisson model, that the variance is a multiple of the mean, is adequate. Other examples are also easily constructed.

It should be stressed that these tests are valid under quite general distributional assumptions, the essential assumptions being those on the first two moments of y_i under H_0. This point is illustrated in section 4.2.

3.2 Discussion

The approach taken here is to set up tests of variance-mean relationship as a regression problem. This is similar in spirit to Carroll and Ruppert (1988, p.10), for example, who state that "We view heteroscedasticity of variance as a regression problem, i.e., systematic and smooth change of variability as predictors are perturbed. Looked at this way, there are many similarities with modeling the mean vector." The tests proposed here exploit the simplifications that arise by choosing the regressor to be y_i^* defined in (2.6), and that for testing inference need be performed only under the simpler null hypothesis. The spirit of the regression based test is that of a Wald test, regression coefficients being obtained under H_1, but unlike a Wald test inference is performed under H_0.

The regression based test approach of this paper was introduced by Cameron and Trivedi (1990), who considered regression-based tests for variance-mean equality. In that case, where $v(\mu_i) = \mu_i$, they noted that H_1 could be expressed as either $\mathbb{E}_1[\{(y_i - \mu_i)^2 - \mu_i\} \mid X_i] = \alpha \cdot g(\mu_i)$ or as $\mathbb{E}_1[\{(y_i - \mu_i)^2 - y_i\} \mid X_i] = \alpha \cdot g(\mu_i)$, and that the less obvious latter representation lead to simpler tests, which coincided with score tests for the Poisson against the Katz system. The results here explain why the latter representation "works", as it equals (2.5) with $v(\mu_i) = \mu_i$ and $\nabla_\mu v(\mu_i) = 1$.

The regression based test of the joint hypothesis $\alpha = 0$ is a chisquare test based on the limit normal distribution of $T^{-1/2}\hat{G}'\hat{W}\hat{y}^* = T^{-1/2}\sum_{i=1}^{T} \hat{g}_i\hat{w}_i\hat{y}_i^*$. This is a special case of a conditional moment (CM) specification test, Newey (1985) and Tauchen (1985), of the departure of $T^{-1/2}\sum_{i=1}^{T} m(y_i,X_i,\hat{\theta}_T)$ from zero, where $m(\cdot)$ is a vector function such that $\mathbb{E}_0[m(y_i,X_i,\theta) \mid X_i] = 0$. Convenient auxiliary regressions for implementation of CM chisquare tests are given by Newey (1985) and Tauchen (1985) when $\hat{\theta}_T$ is the MLE and the likelihood function is specified, and by White (1990) in a number of special cases. Wooldridge (1990) proposes "robust regression-based" tests, where "robust" means validity under quite general distributional assumptions comparable to those here and "regression-based" means computation by auxiliary regressions, a different meaning than that used here.

The CM test literature is generally silent on the choice of $m(y_i,X_i,\theta)$ except in the fully parametric case when the optimal choice is a score test. This paper can be viewed as motivating the particular choice of $m(y_i,X_i,\theta) = g_iw_iy_i^*$ for testing variance-mean relationships. Furthermore, the choice permits very simple computation, since then $\mathbb{E}_0[\nabla_\theta m(y_i,X_i,\theta) \mid X_i] = 0$. Finally, the distribution of the test is obtained under quite general distributional assumptions.

Wooldridge (1991) has proposed CM tests of variance-mean relationships, using $m(y_i,X_i,\theta) = g(\mu_i,\phi)\cdot v(\mu_i,\phi)^{-2}\cdot\{(y_i - \mu_i)^2 - v(\mu_i,\phi)\}$. These tests are more difficult to implement than those here, as $\mathbb{E}_0[\nabla_\beta\{(y_i - \mu_i)^2 - v(\mu_i,\phi)\} \mid X_i] \neq 0$, unless $v(\mu_i,\phi)$ is constant in μ_i.[3] Proposition 3 suggests that in some cases better weighting functions than $v(\mu_i,\phi)^{-2}$ might be used.[4] These CM tests are implemented by the method of Wooldridge (1990). This transforms a test based on $T^{-1/2}\sum_{i=1}^{T}\hat{m}_i$ to one based on $T^{-1/2}\sum_{i=1}^{T}\hat{m}_i^*$, where $\hat{m}_i^*$ is such that in general, $T^{-1/2}\sum_{i=1}^{T}\hat{m}_i \neq T^{-1/2}\sum_{i=1}^{T}\hat{m}_i^*$. As a consequence, this implementation method will test for misspecification in directions different to $g(\mu_i,\phi)$, as noted by Wooldridge (1990, p.26). If we instead consider $m(y_i,X_i,\theta) = g(\mu_i,\phi)\cdot w(\mu_i,\phi)\cdot y_i^*$, where y_i^* is defined in (2.17), then it is possible to use the method of Wooldridge (1990) to test for misspecification in the direction of $g(\mu_i,\phi)$, provided that the estimator $\hat{\phi}_T$ is chosen so that $T^{-1/2}\sum_{i=1}^{T}\hat{d}_i\hat{w}_i\hat{y}_i^* = o_p(1)$. Proposition 5 here then coincides with the test obtained by application of procedure 2.1 of Wooldridge (1990).

The preceding results have not assumed a particular distribution for y_i under H_0, beyond the first two moments (or four moments for the optimal test in Proposition 3). To permit comparison with other more parametric tests we turn to fully parameterized models under H_0.

4. Tests for Fully Parameterized Models

4.1 Optimal Regression Based Tests for Generalized Linear Models

Generalized linear models assume a density of the form:

$$f(y,\theta,\phi) = \exp\{a(\phi)^{-1}(y\theta - b(\theta)) + c(y,\phi)\} , \tag{4.1}$$

for some specific functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$. If ϕ is known, this is a member of the linear exponential family (LEF), which includes the normal, Poisson, binomial, gamma and inverse Gaussian, where θ is the canonical or natural parameter. For further details, see McCullagh and Nelder (1989), or Gourieroux, Montfort and Trognon (1984) who use an alternative equivalent representation of (4.1) in terms of the mean parameter μ.

Differentiation of the logarithm of the density yields:

$$(4.2) \qquad E[y] = \mu = \nabla_{\theta} b(\theta) ,$$

$$(4.3) \qquad E[(y - \mu)^2] = v(\mu,\phi) = a(\phi)\cdot\nabla_{\theta}^2 b(\theta) .$$

Since (4.3) is of the form (2.2), this is the natural family of models with variance-mean relationships to investigate. Indeed, it is the different variance-mean relationships that determine the different members of the linear exponential family.

Propositions 1 and 2 are easily applied. The obvious root-T consistent estimator $\hat{\beta}_T$ is the MLE, defined by the first-order conditions:

$$(4.4) \qquad \sum_{i=1}^{T} v(\mu(X_i,\hat{\beta}_T),\phi)^{-1}\cdot(y_i - \mu(X_i,\hat{\beta}_T)) \cdot \nabla_{\beta}\mu(X_i,\hat{\beta}_T)' = 0 ,$$

where $v(\mu_i,\phi)$ is assumed to be non-zero. This weighted least squares estimator can actually be used to implement the theory in section 2 for any data generating process, since it is consistent for β under both H_0 and H_1 of section 2, see McCullagh and Nelder (1989) or Gourieroux, Montfort and Trognon (1984).

To implement the optimal regression based test of Proposition 3, we need only obtain $\sigma_i^2 = \mathrm{Var}(y_i^*)$. This requires the first four central moments of y under H_0. For the LEF,

$$(4.5) \qquad E[(y - \mu)^3] = v(\mu,\phi)\cdot\nabla_\mu v(\mu,\phi) \ ,$$

$$(4.6) \qquad E[(y - \mu)^4] = v(\mu,\phi)\cdot\{(\nabla_\mu v(\mu,\phi))^2 + v(\mu)\cdot\nabla_\mu^2 v(\mu,\phi) + 3v(\mu,\phi)\} \ ,$$

using the recursion formula $E[(y - \mu)^{k+1}] = \mathrm{Var}(y)\cdot\{(\nabla_\mu E[(y - \mu)^k] + k\cdot E[(y - \mu)^{k-1}]\}$ obtained in the Appendix. Hence:

$$(4.7) \qquad \sigma_i^2 = (v(\mu_i,\phi))^2\cdot\{(\nabla_\mu^2 v(\mu_i,\phi))^2 + 2\} \ .$$

For members of the LEF with quadratic variance function, σ_i^2 reduces to $2(v(\mu_i,\phi))^2$.

Proposition 3 only applies in the case where ϕ is known. Nonetheless, when ϕ needs to be estimated a natural choice for the weight function in applying procedures 4 and 5 is still $w_i = \sigma_i^{-2}$. Thus, if examples 3.3 and 3.4 are for LEF models, the weight function w_i would be chosen to be μ_i^2 and μ_i^4, respectively, in which case the estimators $\hat{\phi}_T$ are of the form (2.25).

4.2 Score Tests for Specific Examples of Generalized Linear Models

The preceding sections give tests that can be applied to a very wide range of models. A natural alternative test is the score (or lagrange multiplier) test. Score tests may be easily implemented, requiring only H_0 density parameter estimates. They are not necessarily easily derived, as the H_1 density may be very cumbersome. The H_1 density may also be overly restrictive, for example, permitting over- but not under-dispersion.

In this section we compare existing score tests for particular LEF models with the optimal regression-based test. From (2.13), the regression based test for $\alpha = 0$ is determined by $G'\Sigma^{-1}y^*$, evaluated at $\hat{\mu}_i$. For models in the LEF, the *ith* entry in Σ is given in (4.7), so that $G'\Sigma^{-1}y^*$ equals:

$$(4.8) \qquad \sum_{i=1}^{T} g(\mu_i)\cdot(v(\mu_i,\phi))^{-2}\cdot\{(\nabla_\mu^2 v(\mu_i,\phi))^2 + 2\}^{-1}\cdot \{(y_i - \mu_i)^2 - \nabla_\mu v(\mu_i,\phi)(y_i - \mu_i) - v(\mu_i,\phi)\}\ .$$

We say that the optimal regression based test coincides with a score test for LEF models embedded in a more general H_1 distribution with mean μ_i and variance $v(\mu_i) + g(\mu_i)'\alpha$ if $(\nabla_\alpha \mathcal{L}(\beta,\phi,\alpha=0))'$, the efficient score evaluated at $\alpha = 0$, equals (4.8) to a multiplicative constant, where $\mathcal{L}(\beta,\phi,\alpha)$ is the log-likelihood for the fully parameterized H_1 model. For the examples given below, the score is found to be of the form (4.8).

It should be noted that even when (4.8) holds, the implementation of a score test may differ from that for a regression based test. In particular, the distribution of $\nabla_\alpha \mathcal{L}(\hat{\beta}_T,\hat{\phi}_T,\alpha=0)$ would typically be obtained under the assumption that the H_0 model has an LEF density, whereas the regression based test may be implemented under much weaker assumptions using propositions 1 and 2. We return to this point below.

Example 4.1: Poisson model under H_0 tested against the Katz system (which includes the negative binomial) under H_1. The score test is equivalent to the optimal regression based test. See Cameron and Trivedi (1990).

Example 4.2: Normal model under H_0 with constant variance σ^2 tested against a normal model with variance $h(\mu_i,X_i,\alpha)$, where $h(\mu_i,X_i,\alpha=0) = \sigma^2$. The efficient score, Breusch and Pagan (1979) and Cook and Weisberg (1983), is:

$$(4.9) \qquad 0.5 \cdot \sum_{i=1}^{T} \nabla_{\alpha} h(\mu_i, X_i, \alpha=0)' \cdot \sigma^{-4} \cdot \{(y_i - \mu_i)^2 - \sigma^2\} .$$

The regression based test is obtained by noting that $Var_1(y_i \mid X_i) = h(\mu_i, X_i, \alpha) = h(\mu_i, X_i, \alpha=0) + \nabla_{\alpha} h(\mu_i, X_i, \alpha=0) \cdot \alpha = \sigma^2 + \nabla_{\alpha} h(\mu_i, X_i, \alpha=0) \cdot \alpha$. So $g(\mu_i, X_i) = \nabla_{\alpha} h(\mu_i, X_i, \alpha=0)'$. Also $y_i^* = \{(y_i - \mu_i)^2 - \sigma^2\}$, and $\sigma_i^2 = 2\sigma^4$. Thus the optimal regression based test equals the score test.

Example 4.3: Binomial model with m trials. The score test against the extended beta-binomial is obtained by Prentice (1986), generalizing the score test against the beta-binomial of Tarone (1979). For the binomial $v(\mu_i) = \mu_i(1 - m^{-1}\mu_i)$, so that $\nabla_{\mu} v(\mu_i, \phi) = (1 - 2m^{-1}\mu_i)$, $\nabla_{\mu}^2 v(\mu_i, \phi) = 0$, and $\sigma_i^2 = \mu_i(1 - m^{-1}\mu_i)$. For the extended beta-binomial, $Var_1(y_i \mid X_i) = \mu_i(1 - m^{-1}\mu_i) \cdot (1 + (m-1)\alpha)$, so that $g(\mu_i) = (m-1)\mu_i(1 - m^{-1}\mu_i)$. For the optimal regression based test, $G'\Sigma^{-1}y^*$ equals:

$$(4.10) \sum_{i=1}^{T} \{2\mu_i(1 - m^{-1}\mu_i)\}^{-1} \cdot \{(y_i - \mu_i)^2 - (1 - 2m^{-1}\mu_i)(y_i - \mu_i) - \mu_i(1 - m^{-1}\mu_i)\}.$$

By rearranging the equation immediately above equation (6) of Prentice (1986), see the Appendix, (4.10) equals the efficient score $\nabla_{\alpha} \mathcal{L}(\beta, \phi, \alpha=0)$. The regression based approach indicates that this test can be used against more general alternatives than $g(\mu_i) = (m-1)\mu_i(1 - m^{-1}\mu_i)$, and against more general alternative distributions than the extended beta-binomial.

Even when the criterion function for the score test coincides with that for a regression based test, implementation may differ. As an example, consider testing for variance constancy under normality. The (chisquare) score test statistic is usually calculated as one-half the explained sum of squares from the auxiliary OLS regression (including an intercept) of

$\hat{\sigma}^{-2}(y_i - \hat{\mu}_i)^2$ on $g(\hat{\mu}_i)$. Let $\hat{y}^*$ have ith entry $\{(y_i - \hat{\mu}_i)^2 - \hat{\sigma}^2\}$, $\bar{G}$ be a Txp matrix with ith row $(g(\hat{\mu}_i) - \bar{g})'$, where $\bar{g} = T^{-1}\sum_{i=1}^{T} g(\hat{\mu}_i)$, and $\hat{\Sigma}$ have ith diagonal entry $2\hat{\sigma}^4$, where $\hat{\sigma}^2 = T^{-1}\sum_{i=1}^{T}(y_i - \hat{\mu}_i)^2$. Then this score test statistic statistic equals $\hat{y}^{*\prime}\bar{G}(\bar{G}'\hat{\Sigma}\bar{G})^{-1}\bar{G}'\hat{y}^*$, see e.g. Cook and Weisberg (1983, eq. 8). By contrast, the regression based test in example 3.2 equals $\hat{y}^{*\prime}\bar{G}(\bar{G}'\tilde{\Sigma}\bar{G})^{-1}\bar{G}'\hat{y}^*$, where $\tilde{\Sigma}$ has ith entry $(\hat{y}_i^*)^2$. The difference between the two is that the score test implementation assumes $\mathrm{Var}_0(y_i^*) = \mathrm{Var}_0((y_i - \mu_i)^2 - \sigma^2)) = 2\sigma^4$. The regression based test is of correct asymptotic level for all distributions, but differs from the usual correction to the score test for nonnormal kurtosis, e.g. Carroll and Ruppert (1988, p.98), which uses $\hat{y}^{*\prime}\bar{G}(\bar{G}'\bar{\Sigma}\bar{G})^{-1}\bar{G}'\hat{y}^*$ where $\bar{\Sigma}$ has constant ith entry $T^{-1}\sum_{i=1}^{T}(\hat{y}_i^*)^2$.

The algebra in obtaining the efficient score in examples other than 4.2 is quite lengthy and tedious, whereas the general regression-based test is easily obtained. Another relatively simple test is the overdispersion test of Cox (1983).

4.3 Cox's Overdispersion Test.

For models with variance-mean relationships of the sort considered here, Cox (1983) proposed a test for overdispersion that is a score test based on an approximation to the H_1 density of y_i which is asymptotically valid for local alternatives to the variance-mean relationship of the H_0 density.

Specifically, suppose y_i has density $f(y,\gamma_i)$, where the scalar parameter γ_i is itself a random variable distributed with mean $\mu_i = \mu(X_i,\beta)$ and variance $g(\mu_i)'\alpha$. Under H_0, $\alpha = 0$ and under H_1, $\alpha > 0$. The restriction made by Cox is to consider only local alternatives: $\alpha = T^{-1/2}\delta$. The resulting variance-mean relationship for y_i conditional on μ_i will be exactly H_L given in (2.15). Cox obtains an approximation to the log-likelihood function, denoted $\mathcal{L}^*(\beta,\delta)$, and tests H_0 by a score test for $\delta = 0$ based on this approximate log-likelihood $\mathcal{L}^*(\beta,\delta)$. Cox obtains:

$$(4.11) \qquad (\nabla_\delta \mathcal{L}^*(\beta,\delta{=}0))' = 0.5 \cdot T^{-1/2} \sum_{i=1}^{T} g(\mu_i) H(y_i,\mu_i) ,$$

where

$$(4.12) \qquad H(y_i, \mu_i) = \nabla_\mu^2 \log f(y_i,\mu_i) + (\nabla_\mu \log f(y_i,\mu_i))^2,$$

and $f(y_i,\mu_i)$ is the H_0 density. Comparison with (4.8) reveals that in general this will lead to a test different to a regression based test.

Specializing to the LEF density given in (4.1),

$$(4.13) \qquad H(y_i,\mu_i) = (v(\mu_i,\phi))^{-2}\{(y_i - \mu_i)^2 - \nabla_\mu v(\mu_i,\phi)(y_i - \mu_i) - v(\mu_i,\phi)\}^2 ,$$

since $\nabla_\mu \log f(y,\mu,\phi) = v(\mu,\phi)^{-1} \cdot (y - \mu)$ and $\nabla_\mu^2 \log f(y,\mu,\phi) = - v(\mu,\phi)^{-1} - \nabla_\mu v(\mu,\phi) \cdot v(\mu,\phi)^{-2} \cdot (y - \mu)$. Substituting (4.13) into (4.11) and comparing with (4.8) reveals that for models in the LEF with quadratic variance function, the optimal regression based test coincides with Cox's test. For other models in the LEF, Cox's test is a regression based test, but is less powerful than the optimal regression based test since it weights by $(v(\mu_i,\phi))^{-2}$ instead of σ_i^2 defined in (4.7).

4.4 Tests for Parametric Models not in the LEF: An Example

The regression based test is very restrictive, and can only coincide with a score test if the score is the particular linear combination of residuals and squared residuals given in (2.5). As a simple illustration of where the regression based test differs from the score test, suppose that under H_0, y_i

is normally distributed with mean μ_i and variance μ_i, an example of a power of the mean model.

The regression based test against H_1: $\mathrm{Var}(y_i) = \mu_i + g(\mu_i)'\alpha$ is given in example 3.1. The optimal regression based test will weight by $\sigma_i^2 = \mathrm{Var}_0((y_i - \mu_i)^2 - y_i) = 2\mu_i(\mu_i+1)$. The joint test of $\alpha = 0$ is based on the closeness to zero of:

$$\sum_{i=1}^{T} g(\hat{\mu}_i)\cdot(2\hat{\mu}_i(\hat{\mu}_i+1))^{-1}\cdot\{(y_i - \hat{\mu}_i)^2 - y_i\} . \tag{4.14}$$

The score test against the alternative $N(\mu_i,\ \mu_i+g(\mu_i)'\alpha)$ distribution is based on the efficient score evaluated at $\alpha = 0$. From the Appendix this is:

$$\sum_{i=1}^{T} g(\hat{\mu}_i)\cdot(2\hat{\mu}_i^2)^{-1}\cdot\{(y_i - \hat{\mu}_i)^2 - \hat{\mu}_i\} . \tag{4.15}$$

Clearly the two tests differ.

These tests both differ from Cox's overdispersion test, a score test against the local alternative to the $N(\mu_i,\ \mu_i)$. After some algebra given in the Appendix this test will be a test of the closeness to zero of:

$$\sum_{i=1}^{T} g(\hat{\mu}_i)\cdot(8\hat{\mu}_i^4)^{-1}\cdot\{(y_i - \hat{\mu}_i)^4 + 4\hat{\mu}_i(y - \hat{\mu}_i)^3 + (4\hat{\mu}_i^2 - 3\hat{\mu}_i)(y - \hat{\mu}_i)^2 - 6\hat{\mu}_i^2(y - \hat{\mu}_i) - 4\hat{\mu}_i^3\} . \tag{4.16}$$

5. Conclusion

In this paper heteroscedasticity tests are proposed for regression models that specify a relationship between the variance and mean under the null hypothesis, but do not require complete specification of the distribution.

These tests are regression based in the sense that a linear combination of the first two conditional moments under H_1 (or equivalently of the errors) leads naturally to a regression that can be run to test the H_0 variance-mean relationship. The particular linear combination chosen is one that simplifies implementation. The tests can be implemented directly from the regression (Propositions 1, 3 and 4), or from an auxiliary regression (Propositions 2 and 5). Implementation is simplest when the H_0 variance depends on the mean alone, but the more general case is also accommodated.

The tests have the advantages of simple derivation and implementation, and of being of correct asymptotic size under minimal distributional assumptions. These advantages are at the potential expense of low power, though the optimal regression based test does coincide with the score test for several leading examples in the linear exponential family.

Footnotes

[1] The function $g(\cdot)$ may also depend on parameters in addition to β and ϕ. Since the ensuing asymptotic theory is unaffected if these additional parameters are replaced by root-T consistent estimates, they are suppressed for ease of exposition.

[2] These and other models in the linear (or natural) exponential family with quadratic variance-mean relationships are detailed in Morris (1982).

[3] In the case where the variance is constant, the test of Wooldridge (1991) and the overall chisquare test given in example 3.2 are equivalent. Otherwise the two procedures differ.

[4] Wooldridge (1991) does not explicitly state what alternatives his CM test is testing against, limiting discussion of power.

APPENDIX

Proposition 1: The estimator is $T^{1/2}\cdot\hat{\alpha}_T = (T^{-1}\hat{G}'\hat{W}\hat{G})^{-1}T^{-1/2}\hat{G}'\hat{W}\hat{y}^*$. Let g_i, w_i, v_i and y_i^* denote $g(\mu_i)$, $w(\mu_i)$, $v(\mu_i)$ and $(y_i - \mu_i)^2 - \nabla_\mu v(\mu_i)(y_i - \mu_i) - v(\mu_i)$; and let $\hat{g}_i$, $\hat{w}_i$, $\hat{v}_i$ and $\hat{y}_i^*$ denote similar quantities evaluated at $\hat{\mu}_i = \mu(X_i,\hat{\beta}_T)$, where $\hat{\beta}_T$ is root-T consistent for β. $\mathbb{E}_0[y_i^*] = 0$ from (2.1) and (2.2). $\mathbb{E}_0[\nabla_\mu y_i^*] = \mathbb{E}_0[-2(y_i - \mu_i) - \nabla_\mu^2 v(\mu_i)(y_i - \mu_i) + \nabla_\mu v(\mu_i) - \nabla_\mu v(\mu_i)] = 0$ from (2.1). By first-order Taylor series expansion:

$$
\begin{aligned}
T^{-1/2}\hat{G}'\hat{W}\hat{y}^* &= T^{-1/2}\sum_{i=1}^{T}\hat{g}_i\hat{w}_i\hat{y}_i^* \\
&= T^{-1/2}\sum_{i=1}^{T} g_i w_i y_i^* + T^{-1}\sum_{i=1}^{T}\nabla_\mu(g_i w_i y_i^*)\nabla_\beta\mu_i \cdot T^{1/2}(\hat{\beta}_T - \beta) + o_p(1) \\
&= T^{-1/2}\sum_{i=1}^{T} g_i w_i y_i^* + o_p(1)
\end{aligned}
$$

since $T^{1/2}(\hat{\beta}_T - \beta) = O_p(1)$ by assumption and $T^{-1}\sum_{i=1}^{T}\nabla_\mu(g_i w_i y_i^*)\nabla_\beta\mu_i = o_p(1)$, applying a LLN and using $\mathbb{E}_0[\nabla_\mu(g_i w_i y_i^*)] = \nabla_\mu g_i w_i \mathbb{E}_0[y_i^*] + g_i w_i \mathbb{E}_0[\nabla_\mu y_i^*] = 0$, since $\mathbb{E}_0[y_i^*] = 0$ and $\mathbb{E}_0[\nabla_\mu y_i^*] = 0$.

Applying a CLT under H_0, $T^{-1/2}G'Wy^* \xrightarrow{d} N(0, \lim_{T\to\infty} T^{-1}G'W\Sigma WG)$. Formal conditions for the CLT can be obtained from White (1980), for example. Sufficient conditions will include boundedness of g_i, w_i, and $(2+\delta)$-th absolute central moment of y_i^* and hence $(4+\delta)$-th central moment of y_i. Under more general conditions $(T^{-1}\hat{G}'\hat{W}\hat{G} - T^{-1}G'WG) \xrightarrow{p} 0$. Combining, $T^{1/2}\hat{\alpha}_T \xrightarrow{d} N(0, (\lim_{T\to\infty} T^{-1}G'WG)^{-1}\cdot \lim_{T\to\infty} T^{-1}G'W\Sigma WG\cdot(T^{-1}G'W\Sigma WG)^{-1}$. Proposition 1 follows by replacing the limit matrices by consistent estimators, following White (1980).

Proposition 2: Consider the regression of 1 on a px1 vector z_i, and let 1 be a Tx1 vector of ones and Z be a Txp matrix with *ith* row z_i'. Then R_u^2 equals $1'Z(Z'Z)^{-1}Z'1 \,/\, 1'1 = 1'Z(Z'Z)^{-1}Z'1 \,/\, T$. If $z_i = \hat{g}_i\hat{w}_i\hat{y}_i^*$, then this equals $\hat{d}_T / T$, where $\hat{d}_T$ is defined in (2.12) with $\tilde{\Sigma}_{ii} = (\hat{y}^*_i)^2$.

Proposition 3: First amend the proof of Proposition 1 to inference under H_L. Now $\mathbb{E}_L[y^*_i] = g_i'\alpha$ from (2.4), while $\mathbb{E}_L[\nabla_\mu y^*_i] = 0$. Again:

$$T^{-1/2}\hat{G}'\hat{W}\hat{y}^* = T^{-1/2}\sum_{i=1}^{T} g_i w_i y^*_i + T^{-1}\sum_{i=1}^{T} \nabla_\mu(g_i w_i y^*_i)\nabla_\beta\mu_i \cdot T^{1/2}(\hat{\beta}_T - \beta) + o_p(1)$$

Again $T^{-1}\sum_{i=1}^{T} \nabla_\mu(g_i w_i y^*_i)\nabla_\beta\mu_i = o_p(1)$, this time since $\mathbb{E}_L[\nabla_\mu g_i w_i y^*_i] = (\nabla_\mu(g_i w_i))\cdot g_i'\alpha = (\nabla_\mu(g_i w_i))\cdot T^{-1/2}g_i'\delta$. Applying a CLT under H_L, $T^{-1/2}G'Wy^* \overset{d}{\to} N(\lim_{T\to\infty} T^{-1}G'WG\cdot\delta, \lim_{T\to\infty} T^{-1}G'W\Sigma WG)$. Again $(T^{-1}\hat{G}'\hat{W}\hat{G} - T^{-1}G'WG) \overset{p}{\to} 0$.

Thus under H_L, $T^{1/2}\cdot\hat{\alpha}_T$ has the same limit distribution as in proposition 1, except that the mean equals $T^{-1/2}\delta$ rather than zero.

For the overall test of $\alpha = 0$, $\hat{d}_T \overset{d}{\to} \chi^2(p,\lambda)$ under H_L, where the non-centrality parameter $\lambda = \delta^2 (\lim_{T\to\infty} T^{-1}G'WG)^{-1}\cdot(\lim_{T\to\infty} T^{-1}G'W\Sigma WG)\cdot(\lim_{T\to\infty} T^{-1}G'WG)^{-1}$, and hence power, is maximized when $W = \Sigma^{-1}$.

Proposition 4: Proceed as in proposition 1, except now $y^*_i = (y_i - \mu_i)^2 - \nabla_\mu v(\mu_i,\phi)(y_i - \mu_i) - v(\mu_i,\phi)$ additionally depends on ϕ. Again, $\mathbb{E}_0[\nabla_\mu y^*_i] = 0$, while $\mathbb{E}_0[\nabla_\phi y^*_i] = \mathbb{E}_0[(\nabla^2_{\mu\phi}v(\mu_i,\phi))(y_i - \mu_i) - \nabla_\phi v(\mu_i,\phi)] = \nabla_\phi v(\mu_i,\phi) \neq 0$. Letting $\hat{\beta}_T$ and $\hat{\phi}_T$ be root-T consistent estimators, the Taylor series expansion yields:

$$T^{-1/2}\hat{G}'\hat{W}\hat{y}^* = T^{-1/2}\sum_{i=1}^{T} g_i w_i y^*_i + T^{-1}\sum_{i=1}^{T} \nabla_\phi(g_i w_i y^*_i) \cdot T^{1/2}(\hat{\phi}_T - \phi) + o_p(1)$$

$$= T^{-1/2}G'Wy^* + (T^{-1}G'WD) \cdot (T^{-1}A'A)^{-1} \cdot (T^{-1/2}B'z) + o_p(1)$$

where the term in $(\hat{\beta}_T - \beta)$ again drops out, since $\mathbb{E}_0[\nabla_\mu y^*_i] = 0$; we use $\mathbb{E}[\nabla_\phi(g_i w_i y^*_i)] = \nabla_\phi(g_i w_i)\mathbb{E}_0[y^*_i] + g_i w_i \mathbb{E}_0[\nabla_\phi y^*_i] = \nabla_\phi v(\mu_i,\phi)$; D is a Txn diagonal matrix with *ith* row $d_i' = \nabla_\phi v(\mu_i,\phi)$; and we have used $T^{1/2}(\hat{\phi}_T - \phi) = T^{1/2}(A'A)^{-1}\cdot B'z + o_p(1)$. Applying a CLT, tedious algebra yields $T^{-1/2}G'Wy^* \overset{d}{\to} N(0, V_{11} + V_{22} + V_{12} + V_{21})$, where

$$V_{11} = (\lim_{T\to\infty} T^{-1}G'W\Sigma WG)$$

$$V_{22} = (\lim_{T\to\infty} T^{-1}G'WD)(\lim_{T\to\infty} T^{-1}A'A)^{-1}(\lim_{T\to\infty} T^{-1}B'\Sigma_{22}B)(\lim_{T\to\infty} T^{-1}A'A)^{-1}(\lim_{T\to\infty} T^{-1}D'WG)$$

$$V_{12} = (\lim_{T\to\infty} T^{-1}G'W\Sigma_{12}B)(\lim_{T\to\infty} T^{-1}A'A)^{-1}(\lim_{T\to\infty} T^{-1}D'WG)\ , \qquad V_{21} = V_{12}',$$

and where Σ, Σ_{22} and $\Sigma_{12} = \Sigma_{21}'$ are TxT diagonal matrices with *ith* entries $\mathrm{Var}_0(y_i^* \mid X_i)$, $\Sigma_{22} = \mathrm{Var}_0(z_i \mid X_i)$, and $\Sigma_{12} = \mathrm{Cov}_0(y_i^*, z_i \mid X_i)$. Again $(T^{-1}\hat{G}'\hat{W}\hat{G} - T^{-1}G'WG) \xrightarrow{p} 0$. Combining and replacing the limit sums by consistent estimates yields Proposition 4.

<u>Proposition 5</u>: $(\hat{\phi}_T - \phi) = (D'WD)^{-1}(D'Wy^*) + o_p(T^{-1/2})$, since by a first-order Taylor series expansion $T^{-1/2}\hat{D}'\hat{W}\hat{y}^* = T^{-1/2}D'Wy^* + (T^{-1}D'WD)^{-1}\cdot T^{-1/2}(\hat{\phi}_T - \phi)$, using $\mathbb{E}_0[\nabla_\phi y_i^*] = d_i'$. Substituting $A'A = D'WD$, $B = D'W$, and $z = y^*$ yields an expression for $(V_{11} + V_{22} + V_{12} + V_{21})$ defined in the proof of Proposition 4 which from (2.21) will be consistently estimated by $T^{-1}\hat{r}'\hat{W}^{1/2}\tilde{\Sigma}\hat{W}^{1/2}\hat{r}$. Also $\hat{r}'\hat{W}^{1/2}\hat{y}^* = \{\hat{G}'\hat{W}^{1/2} - \hat{G}'\hat{W}\hat{D}(\hat{D}'\hat{W}\hat{D})^{-1}\hat{D}'\hat{W}^{1/2}\}\hat{W}^{1/2}\hat{y}^* = \hat{G}'\hat{W}\hat{y}^*$, since $\hat{D}'\hat{W}\hat{y}^* = 0$. So in (2.20), $\hat{d}_T = \hat{y}^{*\prime}\hat{W}^{1/2}\hat{r}(\hat{r}'\hat{W}^{1/2}\tilde{\Sigma}\hat{W}^{1/2}\hat{r})\hat{r}'\hat{W}^{1/2}\hat{y}^*$, which equals TR_u^2 from regression (2) when $\tilde{\Sigma}_{ii} = (\hat{y}_i^*)^2$.

<u>Moments of the LEF</u>: For the LEF density, $\nabla_\theta f(y) = a(\phi)^{-1}(y - \nabla_\theta b(\theta))f(y)$.

$$\begin{aligned}
\text{So}\quad \nabla_\theta \mathbb{E}[(y-\mu)^k] &= \nabla_\theta \int (y - \nabla_\theta b(\theta))^k f(y)\,dy \\
&= \int - k\,\nabla_\theta^2 b(\theta)(y - b(\theta))^{k-1} f(y)\,dy \\
&\quad + \int (y - \nabla_\theta b(\theta))^k a(\phi)^{-1}(y - \nabla_\theta b(\theta))f(y)\,dy \\
&= -k\,\nabla_\theta^2 b(\theta)\,\mathbb{E}[(y-\mu)^{k-1}] + a(\phi)^{-1}\mathbb{E}[(y-\mu)^{k+1}]\ .
\end{aligned}$$

$$\begin{aligned}
\text{So}\quad \mathbb{E}[(y-\mu)^{k+1}] &= a(\phi)\nabla_\theta \mathbb{E}[(y-\mu)^k] + k a(\phi)\nabla_\theta^2 b(\theta)\,\mathbb{E}[(y-\mu)^{k-1}] \\
&= \mathrm{Var}(y)\nabla_\mu \mathbb{E}[(y-\mu)^k] + k\mathrm{Var}(y)\mathbb{E}[(y-\mu)^{k-1}]\ ,
\end{aligned}$$

using (4.2) and (4.3).

Score Test for Extended Beta-Binomial (4.10): From Prentice (1986), the score statistic contribution, given in the equation above equation (6), for n trials with probability of success p (and failure q = (1 - p)) is:

$$0.5\cdot\{y(y-1)p^{-1} + (n-y)(n-y-1)q^{-1} - n(n-1)\}$$
$$= (2pq)^{-1}\{y(y-1)(1-p) + (n-y)(n-y-1)p - n(n-1)p(1-p)\}$$
$$= (2pq)^{-1}\{y^2 - y - 2nyp + 2yp + n^2p^2 - np^2\}$$
$$= (2pq)^{-1}\{(y-np)^2 - y + 2yp - np^2\}$$
$$= (2pq)^{-1}\{(y-np)^2 - (1-2p)(y-np) - np(1-p)\}$$
$$= n(2\mu(1-n^{-1}\mu))^{-1}\{(y-\mu)^2 - (1-2n^{-1}\mu)(y-\mu) - \mu(1-n^{-1}\mu)\},$$

which to a scalar multiple is the term in (4.10). Note that the final form of the score test statistic given in Prentice (1986, equation 6) for the i.i.d. case actually drops the middle term $-(1-2n^{-1}\mu)(y-\mu)$.

Derivation of (4.15): For $y \sim N(\mu, \mu + \alpha g(\mu))$,

$$\log f(y) = -0.5\log 2\pi - 0.5\log(\mu + \alpha g(\mu)) - 0.5\cdot(\mu + \alpha g(\mu))^{-1}(y-\mu)^2.$$

So $\nabla_\alpha \log f(y) = -0.5\, g(\mu)(\mu + \alpha g(\mu))^{-1} + 0.5\cdot g(\mu)(\mu + \alpha g(\mu))^{-2}(y-\mu)^2$,

which evaluated at $\alpha = 0$ equals $g(\mu)(2\mu^2)^{-1}\{(y-\mu)^2 - \mu\}$.

Derivation of (4.16): For $y \sim N(\mu, \mu + \alpha g(\mu))$,

$$\log f(y) = -0.5\log 2\pi - 0.5\log(\mu) - (2\mu)^{-1}(y-\mu)^2.$$

So $\nabla_\mu \log f(y) = (2\mu^2)^{-1}\{(y-\mu)^2 + 2\mu(y-\mu) - \mu\}$

and $\nabla^2_\mu \log f(y) = -(4\mu^3)^{-1}\{(y-\mu)^2 + 2\mu(y-\mu) - \mu\} + (2\mu^2)^{-1}\{-2\mu - 1\}$.

Combining as in (4.12),

$$H(y,\mu) = (4\mu^4)^{-1}\cdot\{(y-\mu)^4 + 4\mu(y-\mu)^3 + (4\mu^2 - 3\mu)(y-\mu)^2 - 6\mu^2(y-\mu) - 4\mu^3\}.$$

REFERENCES

Breusch, T.S. and A.R. Pagan, 1979, "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica*, 47, 1287-1294.

Cameron, A.C. and P.K. Trivedi, 1986, "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests", *J. Applied Econometrics*, 1, 29-54.

Cameron, A.C. and P.K. Trivedi, 1990, "Regression Based Tests for Overdispersion in the Poisson Model", *Journal of Econometrics*, 46, 347-364.

Carroll, R.J. and D. Ruppert, 1988, *Transformation and Weighting in Regression"*, New York: Cambridge.

Cook R.D. and S. Weisberg, 1983, "Diagnostics for Heteroscedasticity in Regression", *Biometrika*, 70, 1-10.

Collings, B.J. & B.H. Margolin, 1985, "Testing Goodness of Fit For the Poisson Assumption When Observations Are Not Identically Distributed", *J. Am. Statist. Assoc.*, 80, 411-418.

Cox, D.R., 1983, "Some Remarks on Overdispersion," *Biometrika*, 70, 269-274.

Dean, C. and J.F. Lawless, 1989, "Tests for Detecting Overdispersion in Poisson Regression Models", *J. Am. Statist. Assoc.,* 84, 467-472.

Domowitz, I., 1983, "Comment on Diagnostic Tests as Residual Analysis", *Econometric Reviews*, 2, 219-222.

Gourieroux, C., A. Montfort and A. Trognon, 1984, "Pseudo Maximum Likelihood Methods: Theory", *Econometrica*, 52, 681-700.

Lee, L-F, 1986, "Specification Tests for Poisson Regression Models", *International Economic Review*, 27, 689-706.

Morris, C.N., 1982, "Natural Exponential Families with Quadratic Variance Functions", *Annals of Statistics*, 10, 65–80.

McCullagh, P. and J.A. Nelder, 1989, *Generalized Linear Models*, London: Chapman and Hall.

Newey, W.K., 1985, "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica*, 53, 1047–1070.

Prentice, R.L., 1986, "Binary Regression Using an Extended Beta-Binomial Distribution, With Discussion of Correlation Induced by Covariate Measurement Errors", *J. Am. Statist. Assoc.*, 81, 321–327.

Tarone, R.E., 1979, "Testing the Goodness-of-Fit of the Binomial Distribution", *Biometrika*, 66, 585–590.

Tauchen, G., 1985, "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *J. Econometrics*, 30, 415–443.

White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity", *Econometrica*, 48, 817–838.

White, H., 1990, *Estimation, Inference and Specification Analysis*, Cambridge University Press, forthcoming.

Wooldridge, J.M., 1990, "A Unified Approach to Robust, Regression-Based Specification Tests", *Econometric Theory*, 6, 17–43.

Wooldridge, J.M., 1991, "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances", *Journal of Econometrics*, 47, 5–46.