Journal of Econometrics 46 (1990) 347–364. North-Holland

REGRESSION-BASED TESTS FOR OVERDISPERSION IN THE POISSON MODEL*

A. Colin CAMERON

University of California, Davis, CA 95616, USA

Pravin K. TRIVEDI

Indiana University, Bloomington, IN 47405, USA

Received May 1988, final version received August 1989

A property of the Poisson regression model is mean–variance equality, conditional on explanatory variables. 'Regression-based' tests for this property are proposed in a very general setting. Unlike classical statistical tests, these tests require specification of only the mean–variance relationship under the alternative, rather than the complete distribution whose choice is usually arbitrary. The optimal regression-based test is easily computed as the t-test from an auxiliary regression. If a distribution under the alternative hypothesis is in fact specified and is in the Katz system of distributions or is Cox's local approximation to the Poisson, the score test for the Poisson distribution is equivalent to the optimal regression-based test.

1. Introduction

The benchmark model for the analysis of discrete count data is the Poisson regression model. This model restricts the variance of the data to be equal to the mean, conditional on explanatory variables. Failure of this restriction has consequences similar to those for heteroscedasticity in the linear regression model: provided the regression function is correctly specified, ordinary least-squares parameter estimates are consistent, but variances for parameter estimates are inconsistently estimated and hypothesis tests will be invalid.

Previous studies have tested the assumption of mean–variance equality in the regression context by embedding the Poisson distribution in a more

*This paper is a greatly revised version of a 1985 paper, 'Regression-Based Tests for Overdispersion', supported by National Science Foundation Grant SES82-08180. We thank Takeshi Amemiya, Shiferaw Gurmu, Adrian Pagan, Jim Powell, and George Tauchen for very helpful discussions, and two anonymous referees and an Associate Editor for their comments and suggestions for improvements.

general distribution. The negative binomial distribution is used by Hausman, Hall, and Griliches (1984) and Collings and Margolin (1985); the more general Katz system of distributions is used by Lee (1986); and local alternatives to the Poisson are used by Cox (1983), Cameron and Trivedi (1986), and Dean and Lawless (1989). The score test is a particularly convenient test for this purpose, as it avoids the more burdensome computation of estimates under the alternative.

In an earlier paper [Cameron and Trivedi (1986)] we explored the relationship between score tests of overdispersion and Cox's local approximation to a model with overdispersion, when the null model is Poisson and the alternative belongs to the Katz family of distributions. In the present paper we examine the properties of a family of regression-based tests of overdispersion. The motivation for such tests is the existence of the population regression-type relation between the variance and the mean of an overdispersed random variable. Efficient estimation of this regression is the basis of our test. We also derive variants of this test, examine the connection between them and the standard score test as well as conditional moment tests, and investigate their asymptotic and small-sample properties.

Consider models for which, conditional on covariates, $(y - \mathrm{E}[y])^2 - y$ has expectation zero under the null hypothesis, and expectation equal to an unknown scalar multiple of a function of $\mathrm{E}[y]$ under the alternative hypothesis. The null hypothesis is a moment condition that is a property of the Poisson, while the alternative encompasses mixture models, or models with parameter heterogeneity. If $\mathrm{E}[y]$ were known, the scalar multiple could be estimated by least-squares regression, and the usual t-test for significance would be a test of the null hypothesis of mean–variance equality. Since $\mathrm{E}[y]$ is unknown, we replace it by a consistent estimate based on parameter estimates under the null hypothesis.

This approach leads to a more general class of tests for mean–variance quality than previously considered by others including Cameron and Trivedi (1986). Unlike previous tests, the distribution of the data under the alternative hypothesis need not be specified. Only the relationship between the variance and the mean is specified.

These regression-based tests are therefore applicable to most econometric applications with count data. Statistical theory provides strong reasons for choosing the Poisson model for count data, just as it suggests that a Bernoulli model is appropriate for discrete data. However, the Poisson model may be inappropriate, because of unobserved heterogeneity or because of failure of the independence assumption in the usual derivation of the Poisson. Once we entertain these possibilities, we no longer know the distribution of the data.

The usual approach is to assume a distribution and proceed with classical testing and estimation procedures. For example, if heterogeneity is gamma distributed, then the count data will be negative binomial distributed. Unfortunately, this has the weakness that even if the variance and mean of the assumed negative binomial distribution are correctly specified, if the distribution is not in fact the negative binomial, the maximum-likelihood estimator is inconsistent. For other distributions, such as some members of the Katz system of distributions, not only may the MLE be inconsistent, but it may be computationally burdensome to obtain.

A more robust approach is to use inferential techniques that rely only on specification of the mean and variance of the data. This approach, that of Generalized Linear Models, is very popular in statistics. It is summarized in McCullagh and Nelder (1983) and introduced into the econometrics literature by Gourieroux, Montfort, and Trognon (1984). Our test is in the spirit of this approach. Furthermore, if the Poisson regression model is rejected, the parameters of the model with the variance–mean relationship of the alternative hypothesis can easily be consistently estimated.

Many regression-based tests can be constructed, depending on the weights chosen in the regression. However, an optimal test in this class of regression-based test can be analytically obtained. It is easy to implement, as an asymptotically equivalent test can be computed as the t-test from an ordinary least-squares regression.

When we consider the more parametric world where a distribution under the null hypothesis is specified, this optimal test is sometimes equivalent to a score test. Specifically, the optimal test is equivalent to the score test if the efficient score under the null hypothesis has a particular functional form that arises when the distribution under the alternative hypothesis is in the Katz system of distributions, or is Cox's local approximation to the Poisson. Of course, the regression-based test can be used in a more general setting.

The regression-based test is a nonclassical test that can be viewed as a moment-based specification test in the framework of Tauchen (1985) or Newey (1985). Another popular nonclassical test is the information matrix test. In general this will differ from the regression-based tests, though for a commonly used parameterization of the Poisson model the optimal regression-based test is equivalent to an information matrix test based only on the intercept. Finally, we note that the null hypothesis can also be expressed as $(y - \mathrm{E}[y])^2 - \mathrm{E}[y]$ has expectation zero. This leads to an alternative class of regression-based tests that generally has less power in testing for the Poison regression model.

In section 2, a general theory of regression-based tests for mean–variance equality is presented. In section 3, score tests for the Poisson distribution are presented. In section 4, the regression-based tests are placed in the more general context of moment-based specification tests. Section 5 contains an illustrative application of the test to Kennan's (1985) strike data. In section 6 the small-sample performances of three tests of this paper are investigated in a Monte Carlo framework. Some concluding remarks are made in section 7.

2. Regression-based tests for mean–variance equality

2.1. General framework

The data $\{(y_i, X_i),\ i = 1, \ldots, N\}$ are independent across i. Conditional on the K-dimensional vector of explanatory variables X_i, the mean of the scalar dependent variable y_i is

$$\mathrm{E}[y_i] = \mu_i \equiv \mu(X_i, \beta), \tag{2.1}$$

where μ is a differentiable function that maps onto the positive real line and is first-order identifiable, i.e., $\mu(X_i, \beta_0) = \mu(X_i, \beta_1) \Leftrightarrow \beta_0 = \beta_1$.

We test the assumption of mean–variance equality, a property of the Poisson distribution. Under the null hypothesis,

$$\mathrm{H}_0\colon \quad \mathrm{var}(y_i) = \mu_i. \tag{2.2}$$

The specific alternative hypothesis is that

$$\mathrm{H}_1\colon \quad \mathrm{var}(y_i) = \mu_i + \alpha \cdot g(\mu_i), \tag{2.3}$$

where $g(\cdot)$ is a specified function that maps from R^+ to R^+. Under both H_0 and H_1, (2.1) is assumed to hold. Tests for overdispersion or underdispersion are tests of whether $\alpha = 0$.

A justification for choosing mean–variance relationships of the form (2.3) is that they arise from the following form of parameter heterogeneity. Let y_i conditional on γ_i have mean and variance $\gamma_i(\mu_i, v_i)$, where γ_i conditional on μ_i is random with mean μ_i and variance $\alpha \cdot g(\mu_i)$. Then the variance of y_i (conditional on μ_i) is (2.3). Thus the specification of the alternative hypothesis is general enough to encompass any mixture model, provided only that the mixing distribution has a finite variance. An example of underdispersion is the binomial model with mean μ_i and variance equal to $\mu_i(1 - \mu_i)$. In this case α will be negative.

Most authors restrict the finite variance of the mixture distribution to be a multiple of 1, μ_i, or μ_i^2. We allow more general specification for the variance. Different distributional assumptions for γ_i lead to different unconditional distributions for y_i, sometimes called compound Poisson models. For example, if γ_i is gamma distributed and y_i conditional on γ_i is Poisson distributed, then the unconditional distribution for y_i is the negative binomial distribution of Greenwood and Yule (1920). This is used by Hausman, Hall, and Griliches (1984) and Collings and Margolin (1985) amongst others. The novelty of the tests proposed here is that such distributions need not be assumed.

To summarize, this framework nests virtually all previous studies of overdispersion. In most studies the distribution of y_i under both H_0 and H_1 is specified; attention is restricted to the log-linear model, where $\mu_i = \exp(X_i'\beta)$, and $g(\mu_i)$ is specified to be equal to 1, μ_i, or μ_i^2.

2.2. *Regression-based tests*

From (2.3) the model under H_1 can be written as

$$E\left[(y_i-\mu_i)^2 - y_i\right] = \alpha \cdot g(\mu_i). \tag{2.4}$$

If μ_i is observed, an obvious test for overdispersion is a t-test for $\alpha = 0$ in the least-squares regression

$$(y_i-\mu_i)^2 - y_i = \alpha \cdot g(\mu_i) + \varepsilon_i, \tag{2.5}$$

where the heteroscedastic error term is defined by

$$\varepsilon_i = (y_i-\mu_i)^2 - y_i - E\left[(y_i-\mu_i)^2 - y_i\right]. \tag{2.6}$$

Weighted least-squares estimation of (2.5) with weights $w_i = w(\mu_i)$ yields an estimator for α satisfying

$$\sum_{i=1}^{N} w_i \cdot g(\mu_i) \cdot \left\{(y_i-\mu_i)^2 - y_i - \alpha \cdot g(\mu_i)\right\} = 0. \tag{2.7}$$

Since μ_i is unknown, it is replaced in (2.7) by an estimate based on a consistent estimate for β. Before doing so observe that this will have minimal effect on the estimator for α for the following reason. Eq. (2.7) may be interpreted as a first-order condition for maximizing quasi-likelihood function

$$-Q_N(\alpha,\beta) = \sum_{i=1}^{N} w_i\left[(y_i-\mu(X_i,\beta))^2 - y_i - \alpha \cdot g(\mu(X_i,\beta))\right]^2. \tag{2.8}$$

Given a consistent estimator for β, the criterion to be minimized is $Q_N(\alpha,\hat{\beta})$. It is known that [White (1983), Pagan (1986)], provided

$$E\frac{\partial^2 Q(\alpha^*,\beta^*)}{\partial\alpha\cdot\partial\beta} = 0, \tag{2.9}$$

minimization of $Q_N(\alpha,\hat{\beta})$ yields an estimator $\hat{\alpha}$ which has the same asymptotic properties as the estimator $\tilde{\alpha}$ which minimizes $Q_N(\alpha,\beta^*)$, where β^* is the true value of β.

It is straightforward to verify that in the present case

$$\mathrm{E}_{*} \left. \frac{\partial^2 Q_N(\alpha^*, \beta^*)}{\partial \alpha \cdot \partial \beta} \right|_{\alpha^*=0} = 0, \tag{2.10}$$

and hence under H_0 or local alternatives to H_0 the asymptotic properties of $\hat{\alpha}$ will be the same as those of $\tilde{\alpha}$. (In section 4 we shall analyze a variant for which the above condition does not hold.)

We replace μ_i in (2.7) by $\hat{\mu}_i = \mu(X_i, \hat{\beta})$, where the obvious choice for $\hat{\beta}$ is the Poisson maximum-likelihood estimator, which satisfies

$$\sum_{i=1}^{N} \mu_i^{-1} \cdot (y_i - \mu_i) \cdot \frac{\partial \mu_i}{\partial \beta} = 0. \tag{2.11}$$

Note that the Poisson MLE is consistent under both H_0 and H_1 [see McCullagh and Nelder (1983) or Gourieroux, Montfort, and Trognon (1984)].

The resulting two-step estimator for α can be written as

$$\hat{\alpha}_w = \left(\hat{g}' \hat{W} \hat{g} \right)^{-1} \hat{g}' \hat{W} y^*,$$

where the $N \times 1$ vectors $\hat{g}$ and y^* have ith entries $g(\hat{\mu}_i)$ and $(y_i - \hat{\mu}_i)^2 - y_i$, respectively, and $\hat{W}$ is a diagonal matrix with ith entry $\hat{w}_i$ such that $\hat{W}$ is consistent for W.

We consider the limit distribution of $N^{1/2} \cdot \hat{\alpha}_w$ under local alternatives H_L: $\alpha = N^{-1/2}\gamma$, or more formally

$$H_L: \quad \operatorname{var}(y_i) = \mu_i + \left(N^{-1/2}\gamma\right) \cdot g(\mu_i), \quad \gamma \text{ a finite constant.} \tag{2.12}$$

Then, under H_L,

$$N^{1/2} \cdot \hat{\alpha}_w \xrightarrow{\mathrm{d}} \mathrm{N}\left(\gamma, \left[\lim_{N\to\infty} N^{-1} g' W g \right]^{-2} \cdot \lim_{N\to\infty} N^{-1} g' W \Sigma W g \right), \tag{2.13}$$

where g has ith entry $g(\mu_i)$ and Σ is a diagonal matrix with ith entry

$$\sigma_i^2 = \mathrm{E}\left[\left\{ (y_i - \mu_i)^2 - y_i \right\}^2 \right]. \tag{2.14}$$

The result (2.13) is obtained from the general theory of estimators defined implicitly by a set of moment equations. We consider the joint distribution of

α and β that simultaneously solve (2.7) and (2.11) and, since α does not appear in (2.11), use the simplification suggested for example by Newey (1984). An alternative derivation of (2.13) in Cameron and Trivedi (1985) establishes that (2.13) holds more generally when β is estimated by any consistent estimator of $O_p(N^{-1/2})$.

Note that σ_i^2 equals the variance of ε_i under H_0. This means that replacing μ by $\hat{\mu}$ in (2.5) makes no difference to the asymptotic distribution of the weighted least-squares estimators, a consequence of (2.10). This considerably simplifies the analysis.

Specializing to H_0, (2.13) yields the t-test statistic for $\alpha = 0$,

$$T_w = \left[\hat{g}'\hat{W}\hat{\Sigma}\hat{W}\hat{g}\right]^{-1/2} \cdot \hat{g}'\hat{W}y^*, \tag{2.15}$$

where $\hat{\Sigma}$ is a consistent estimate of Σ.

For the Poisson distribution $\sigma_i^2 = 2\mu_i^2$. It is then natural to use $\hat{\sigma}_i^2 = 2\hat{\mu}_i^2$, where $\hat{\mu}_i = \hat{\mu}(X_i, \hat{\beta})$. Other estimators of Σ are discussed in section 2.4.

Under H_0, T_w is asymptotically distributed as $N(0,1)$ and can be used for either one-sided or two-sided tests of overdispersion ($\alpha > 0$) or underdispersion ($\alpha < 0$). Under H_L, $(T_w)^2$ is asymptotically distributed as noncentral chi-square with noncentrality parameter

$$\lambda_w = \gamma^2 \cdot [g'Wg]^2 \cdot [g'W\Sigma Wg]^{-1}. \tag{2.16}$$

2.3. *Optimal regression-based test*

Different choices of the weighting matrix lead to different test statistics. The optimal test within the class of regression-based tests based on (2.5) maximizes the local power by maximizing λ_w in (2.16). This corresponds to $W = \Sigma^{-1}$. Then we may use

$$T_{\text{opt}} = \left[\hat{g}'\hat{\Sigma}^{-1}\hat{g}\right]^{-1/2} \cdot \hat{g}'\hat{\Sigma}^{-1}y^*, \tag{2.17}$$

which is distributed as $N(0,1)$ under H_0.

This optimal test is easily computed as the square root of the explained sum of squares from the OLS regression

$$\begin{aligned}&\left(\sqrt{2} \cdot \hat{\mu}_i\right)^{-1} \cdot \left\{(y_i - \hat{\mu}_i)^2 - y_i\right\} \\ &= \left(\sqrt{2} \cdot \hat{\mu}_i\right)^{-1} \cdot g(\hat{\mu}_i) \cdot \alpha + \text{error}. \end{aligned} \tag{2.18}$$

Alternatively, the standard t-test from this regression is asymptotically equivalent to T_{opt}.

2.4. Comment

The regression-based tests rely on asymptotic theory that does not actually require that y be Poisson-distributed under H_0. The result (2.17) requires assumptions on only the first four moments of the distribution of y, and the result (2.13) requires assumptions on only the first two moments.

Thus T_w in (2.15) could be used to test mean–variance equality for any distribution. It is sufficient to let $\hat{\Sigma}$ be any estimate such that $N^{-1}\hat{g}'\hat{W}\hat{\Sigma}\hat{W}\hat{g}$ is consistent for $\lim_{N\to\infty} N^{-1}g'W\Sigma Wg$. Following Eicker (1967) and White (1980), an obvious choice is the diagonal matrix with ith entry $\hat{\sigma}_i^2$ equal to $[(y_i-\hat{\mu}_i)^2-y_i]^2$. This is justified by the general theory of estimators defined by a set of moment equations. More specifically, Newey's (1984, p. 202) conditions for the use of Eicker–White estimator are satisfied here. This approach generates a 't-test' based on the weighted least-squares estimate of α in the regression (2.5) and its Eicker–White heteroskedasticity consistent standard error. For later reference we shall call this T_{EW} test.

Another estimator of σ_i^2 is obtained by using information on the first two moments of y_i to write (2.14) as follows:

$$\sigma_i^2 = \mathrm{E}\left[(y_i-\mu_i)^4\right] - 2\cdot\mathrm{E}\left[(y_i-\mu_i)^3\right] - \mu_i^2 + \mu_i. \tag{2.14a}$$

When expectations are taken with respect to the null-hypothesis Poisson distribution, $\sigma_i^2 = 2\mu_i^2$, since the third and fourth central moments of y_i are, respectively, μ_i and $(3\mu_i^2+\mu_i)$.

An alternative estimator of σ_i^2 is obtained by substituting a consistent estimator of μ_i in (2.14a) and replacing the term under the expectations operator by the sample mean. This method is attractive because it makes weaker distributional assumptions than the alternative, but its feasibility depends upon obtaining positive estimates of σ^2 which is not guaranteed; see section 5 for an example.

3. Score tests for the Poisson model

The optimal regression-based test can be shown to be equivalent to the following class of score tests, where the Poisson is embedded in a more general distribution with mean μ_i and variance $\mu_i + \alpha\cdot g(\mu_i)$. Suppose the efficient score takes the following form:

$$\left.\frac{\partial \ln \mathscr{L}}{\partial \beta}\right|_{\alpha=0} = \sum_{i=1}^{N} \mu_i^{-1}\cdot(y_i-\mu_i)\cdot\frac{\partial\mu_i}{\partial\beta}, \tag{3.1}$$

$$\left.\frac{\partial \ln \mathscr{L}}{\partial \alpha}\right|_{\alpha=0} = \sum_{i=1}^{N} \left(2\mu_i^2\right)^{-1}\cdot g(\mu_i)\cdot\left\{(y_i-\mu_i)^2-y_i\right\}. \tag{3.2}$$

(3.1) will always hold, since they are the first-order conditions for the Poisson maximum-likelihood estimator, while (3.2) holds in some examples presented below. Score tests given (3.1) and (3.2) are particularly convenient as the information matrix under H_0 is block-diagonal, since for the Poisson $\mathrm{E}[\{(y_i-\mu_i)^2-y_i\}\cdot(y_i-\mu_i)]=0$. Also for the Poisson, $\mathrm{E}[\{(y_i-\mu_i)^2-y_i\}^2] = 2\mu_i^2$, so that from (3.2), $I_{\alpha\alpha}=\sum_{i=1}^{N}(2\mu_i^2)^{-1}\cdot g^2(\mu_i)$ under H_0. The score test based on (3.1) and (3.2) is therefore:

$$T_{\text{score}}=\left[\sum_{i=1}^{N}\left(2\hat{\mu}_i^2\right)^{-1}\cdot g^2(\hat{\mu}_i)\right]^{-1/2} \times\left[\sum_{i=1}^{N}\left(2\hat{\mu}_i^2\right)^{-1}\cdot g(\hat{\mu}_i)\cdot\left\{(y_i-\hat{\mu}_i)^2-y_i\right\}\right]. \tag{3.3}$$

This is identical to the regression-based statistic T_{opt} in (2.17).

There are at least two choices of distribution under the alternative hypothesis for which the efficient score under H_0 takes the form (3.2).

Firstly, this is the case for the Katz system of distributions, defined in Katz (1963). Score tests for this system are presented in Lee (1986) for $\mu=\exp(X'\beta)$ and $g(\mu)=\mu$ or μ^2, discussed briefly in Collings and Margolin (1985) for the case $g(\mu)=\mu^2$ and presented in a more general setting in Cameron and Trivedi (1986). The Katz system includes the binomial and negative binomial distributions. The latter has often been used to model overdispersion since at least Greenwood and Yule (1920).

Secondly, this is the case for local alternatives to the Poisson suggested by Cox (1983). Specifically, under the alternative hypothesis y_i is Poisson-distributed with mean λ_i, where λ_i is stochastic with mean μ_i, variance $\alpha\cdot g(\mu_i)$, and the efficient score is based on the local approximation to this density proposed by Cox. Score tests are presented in Cox (1983), Cameron and Trivedi (1986), and for the case $g(\mu)=\mu^2$ in Dean and Lawless (1989).

Detailed derivation of the efficient scores and tests in these two cases is given in Cameron and Trivedi (1986). They study the same model as here, except that $g(\mu)$ is restricted to equal μ^k, for k a positive integer. Their results generalize immediately, simply replace μ^k by $g(\mu)$ throughout.

The score test in principle uses more information about the distribution of y_i under the alternative than does the regression-based test. For some alternative hypotheses, the efficient score under H_0 may not be of the form (3.2), in which case the score test and the regression-based test will differ.

The small-sample properties of T_{opt} and some variants for the Poisson case with $g(\mu)=\mu^2$ are investigated in section 6. Together with the investigations of Collings and Margolin (1985) and Dean and Lawless (1989), these results provide useful evidence about the small-sample properties of the regression-based test in settings where it coincides with the score test.

4. Discussion

4.1. Specification-test interpretation

The regression-based tests proposed in this paper can be interpreted as tests based on an auxiliary-criterion function or moment function of the type analyzed in Tauchen (1985) and Newey (1985). In the terminology of Tauchen, an auxiliary criterion is a moment restriction that holds under H_0 but may not hold under H_1.

The restriction considered in this paper in $E[(y_i - \mu_i)^2 - y_i]$ equals 0. Tauchen's approach would suggest basing a test of H_0 on

$$\hat{\tau}_N = N^{-1} \cdot \sum_{i=1}^{N} \left\{ (y_i - \hat{\mu}_i)^2 - y_i \right\}.$$

Clearly we could more generally consider an auxiliary criterion:

$$\hat{\tau}_N = N^{-1} \cdot \sum_{i=1}^{N} \hat{w}_i \cdot \left\{ (y_i - \hat{\mu}_i)^2 - y_i \right\}, \tag{4.1}$$

based on $E[w_i \cdot \{(y_i - \mu_i)^2 - y_i\}] = 0$, where the weights w_i may be a function of μ_i. Different choices of weight w_i lead to different test statistics.

The regression-based approach guides the choice of weights w_i. The test T_{opt} in (2.17) suggests that the best choice for tests against alternatives of the form (2.3) is $w_i = (2\mu_i^2)^{-1} \cdot g(\mu_i)$. This choice of weights means that (4.1) is identical to $\partial \ln \mathscr{L} / \partial \alpha |_{\alpha=0, \beta=\hat{\beta}}$, where $\mathscr{L}$ is the likelihood for the Katz system of distributions.

As already noted, a test statistic asymptotically equivalent to that based on (4.1) can be computed as the t-test from the OLS regression of $\left(\sqrt{2} \cdot \hat{\mu}_i\right)^{-1} \cdot [(y_i - \hat{\mu}_i)^2 - y_i]$ on $\left(\sqrt{2} \cdot \hat{\mu}_i\right)^{-1} \cdot g(\hat{\mu}_i)$. Tauchen's results instead suggest an asymptotically equivalent t-test for significance of the intercept in a regression of $(2 \cdot \hat{\mu}_i^2)^{-1} \cdot \{(y_i - \hat{\mu}_i)^2 - y_i\}$ on an intercept and the vector $\hat{\mu}_i^{-1} \cdot (y_i - \hat{\mu}_i) \cdot \partial \mu_i / \partial \beta |_{\beta=\hat{\beta}}$.

4.2. Information-matrix test

The information-matrix test [see White (1982)] is also a moment condition test, as noted by Tauchen (1985). It is based on the quantity $\sum_{i=1}^{N} H(y_i, \beta)$, where

$$H(y,\beta)=\frac{\partial \log f(y,\beta)}{\partial\beta}\cdot\frac{\partial \log f(y,\beta)}{\partial\beta'}+\frac{\partial^2 \log f(y,\beta)}{\partial\beta\partial\beta'},$$

where $f(y,\beta)=f(y,\mu)$ evaluated at $\mu=\mu(X,\beta)$ is the density of y under H_0. For the Poisson

$$H(y,\beta)=\mu^{-2}\cdot\left\{(y-\mu)^2-y\right\}\frac{\partial\mu}{\partial\beta}\cdot\frac{\partial\mu}{\partial\beta'}+\mu^{-1}\cdot(y-\mu)\cdot\frac{\partial^2\mu}{\partial\beta\partial\beta'}.$$

In the special case $\mu=\exp(X'\beta)$, $H(y,\beta)$ simplifies to $[(y-\mu)^2-\mu]XX'$. Then the information-matrix test based on the intercept term, i.e., using only the first entry in $H(y,\beta)$, uses the auxiliary-criterion function:

$$\hat{\tau}_N=N^{-1}\cdot\sum_{i=1}^{N}\left\{(y_i-\hat{\mu}_i)^2-\hat{\mu}_i\right\}.$$

Since $(1/N)\Sigma_N^{i=1}\hat{\mu}_i=\bar{y}$ in this example, this leads to a test identical to the optimal regression test or score test of the form (3.3), against alternatives $\operatorname{var}(y_i)=\mu_i+\alpha\cdot\mu_i^2$. Thus, for this example the information-matrix test can be interpreted as a test for a particular form of random-parameter heterogeneity, a link discussed more generally in Chesher (1984).

In general, however, there is no such relationship between the information-matrix tests and regression-based tests. For other parameterizations of the mean of the Poisson model, or for different alternative hypotheses for the relationship between mean and variance, the two tests will be different.

4.3. Alternative regression-based tests

The regression-based tests are based on the moment condition (2.4). We could alternatively consider tests based on re-expressing (2.3) as

$$\mathrm{E}\left[(y_i-\mu_i)^2-\mu_i\right]=\alpha\cdot g(\mu_i). \tag{4.2}$$

This suggests a range of tests based on weighted least-squares estimation of

$$(y_i-\mu_i)^2-\mu_i=\alpha\cdot g(\mu_i)+\nu_i, \tag{4.3}$$

where ν_i is a heteroskedastic error. Empirically μ_i is replaced by $\hat{\mu}_i$.

The resulting test statistics that correspond to T_w and T_{opt} in section 2 are presented in Cameron and Trivedi (1985). Changing the last term in the dependent variable from y_i in section 2 to $\hat{\mu}_i$ here leads to a much more cumbersome asymptotic theory, which for brevity is not reproduced here. However, using an approach suggested by an anonymous referee, the essential problem, sometimes referred to as the 'Durbin problem', can be explained as follows. The condition (2.9) will not hold in this case since the criterion to be maximized is

$$-Q_N(\alpha,\beta) = \sum_{i=1}^{N} w_i\Big[(y_i-\mu(X_i,\beta))^2 - \mu(X_i,\beta) - \alpha\cdot g(\mu(X_i,\beta))\Big]^2 \tag{4.4}$$

and

$$\frac{\partial^2 Q}{\partial\alpha\cdot\partial\beta} = -2\cdot\sum w_i\frac{\partial\mu_i}{\partial\beta}\Big\{g'(\mu_i)\big[(y_i-\mu_i)^2-\mu_i-\alpha\cdot g(\mu_i)\big] - 2g(\mu_i)(y_i-\mu_i) - g(\mu_i) - \alpha\cdot g(\mu_i)g'(\mu_i)\Big\}, \tag{4.5}$$

so that when $\alpha^* = 0$,

$$\mathrm{E}_*\frac{\partial^2 Q}{\partial\alpha\cdot\partial\beta} = 2\cdot\sum w_i\frac{\partial\mu_i}{\partial\beta}g(\mu_i) \neq 0.$$

Replacing μ_i by $\hat{\mu}_i$ in (4.3) adds an extra source of error that in this case does not disappear asymptotically.

This alternative approach is perhaps a more obvious test than that of section 2. This test can be viewed as the natural generalization of the diagnostic check of plotting $(y_i-\hat{\mu}_i)^2$ against $\hat{\mu}_i$, rather than against y_i. An interesting exercise is to compare the two approaches.

For the Poisson model, the optimal test in this new class of tests is

$$T_2 = \big[\hat{g}'\hat{W}_2\hat{g}\big]^{-1/2}\big[\hat{g}'\hat{W}_2 y_2^*\big], \tag{4.6}$$

where the $N\times 1$ vectors g and y_2^* have ith entries $g(\hat{\mu}_i)$ and $(y_i-\hat{\mu}_i)^2-\hat{\mu}_i$, and $\hat{W}_2$ is a square matrix consistent for W_2 given by

$$W_2 = \Big[D - \Delta\big(\Delta' D_\mu^{-1}\Delta\big)^{-1}\Delta'\Big]^{-1}, \tag{4.7}$$

where D and D_μ are diagonal matrices with ith entries $(2\mu_i^2 + \mu_i)$ and μ_i, respectively, and Δ is a $N \times K$ matrix with ith row $\partial\mu_i/\partial\beta'$.

This test can be shown to be locally less powerful than or, at best, as powerful as T_{opt} [see Cameron and Trivedi (1985)]. In the commonly studied case where $\mu_i = \exp(X_i'\beta)$, X_i includes an intercept, and $g(\mu_i) = \mu_i^2$, it can be shown that the two estimates of α are numerically identical, but lead to different tests. But in general the estimates of α will differ.

For tests of mean–variance equality in models other than the Poisson, the optimal test in this alternative class based on (4.3) may be more efficient than the corresponding generalization of (2.17) discussed in section 2.4. For example, the optimal test in this alternative class can be shown to be identical to the score test for testing the normal distribution with mean μ_i and variance μ_i against the normal distribution with mean μ_i and variance $(\mu_i + \alpha \cdot g(\mu_i))$. For details, see Cameron and Trivedi (1985). Of course, the usual reason for testing mean–variance equality is that the null hypothesis model is the Poisson, in which case it is unambiguously best to use tests based on the moment condition (2.4).

5. An illustration

To illustrate the application of the regression-based test we estimate a simple model of strikes using the data provided by Kennan (1985). In his model of strikes the duration of strike is explained in terms of a variable Z which denotes the cyclical deviation of aggregate output from trend. He provides data on Z and the duration and number of strikes ending each month during January 1968 to December 1976 period. With five monthly observations missing, this yields 103 observations. For illustration we estimated a Poisson model of the number of strikes ending each month using a constant and Z as explanators. The mean number of strikes each month is 5.5, the standard deviation is 3.7, and the range of observations is 1 to 18. The estimate of the Poisson model, and the regression for testing overdispersion of the form (2.3) with $g(\mu) = \mu^2$ were as follows (t-ratios in parenthesis):

$$\hat{\mu} = \underset{(56.60)}{1.697} + \underset{(4.61)}{2.759} \cdot Z, \tag{5.1}$$

$$(y - \hat{\mu})^2 - y = \underset{(4.34)}{0.2319} \cdot \hat{\mu}^2. \tag{5.2}$$

There is evidence of significant overdispersion. It is relevant and interesting to note that, when σ_i^2 ($i = 1, \ldots, N$) were estimated from (2.14a), several negative values were found, so the regression-based test could not be carried out under the weaker assumptions of the method of moments. In general this estimator of σ_i^2 may not be positive, so its use in data analysis is problematic.

The strike model was also estimated by maximum likelihood based on the negative binomial model with overdispersion of the form (2.3) with $g(\mu)=\mu^2$ [denoted NEGBIN II in Cameron and Trivedi (1986)]. This yielded $\hat{\alpha}=0.233$ with t-value of 3.345 and

$$\hat{\mu} = \underset{(25.33)}{1.697} + \underset{(2.12)}{2.850} \cdot Z. \tag{5.3}$$

The estimate of α is close to that provided by the regression-based test. In general, substitution of $\hat{\alpha}$ estimated thus into the likelihood function of negative binomial distributed observations and maximizing the resulting expression with respect to the remaining parameters provides a feasible two-step quasi-likelihood estimation method [Gourieroux et al. (1984)]. The 't-values' of the two estimates differ because the Poisson estimate is consistent under the null and the alternative, whereas the NEGBIN estimate is efficient under the alternative.

6. A Monte Carlo investigation

The objective of this investigation was to examine the size and power characteristics of T_{opt}, T_{EW}, and T_{μ}, the last of which is a variant of T_{opt} obtained by replacing y^* in the regression (2.17) by y_2^*.

6.1. Design of the experiments

Six models and two sample sizes, $N=50$ or 100, were used and each simulation experiment was based on 500 replications. Table 1 gives details of the combinations of (β_0, β_1). All models included a constant term and one explanatory variable x which was taken as a random draw from a uniform [0,1] distribution and then held fixed for all remaining experiments.

In all cases we set $\mu=\exp(\beta_0+\beta_1 x+v)$ where v, the heterogeneity term, is a random draw from a specified distribution. This term is identically zero for three cases, viz., models 1, 2, and 3. In the case of models 4 and 5 it is a random draw from Gamma $[\mu, \phi]$ distribution, the parameter ϕ having been chosen to yield modest overdispersion in the case of model 4 and greater overdispersion in the case of model 5. This can be seen by examining the typical range of the random variable y reported in table 1. In the case of models 4 and 5 the y variable will have negative binomial distribution with $\mathrm{E}(y|x,v)=\mu$, $\operatorname{var}(y|x,v)=\mu+(1/\phi)\mu^2$, so that $g(\mu)=\mu^2$ is appropriate. In the case of model 6 the heterogeneity term is a random draw from a lognormal

Table 1
Parameter combinations.

Model	(β_0, β_1)	ϕ	Range of y	Over-dispersion
1	(0.1, 1.0)	0.0	[0, 6]	Nil
2	(1.0, 1.0)	0.0	[0, 10]	Nil
3	(1.5, 1.0)	0.0	[2, 17]	Nil
4	(1.0, 1.0)	2.0	[0, 12]	Low
5	(1.0, 1.0)	1.1	[0, 18]	High
6	(1.0, 1.0)	n.a.	[0, 14]	Moderate

Table 2
Percentage rejections of H_0: $\alpha = 0$ vs. H_1: $\alpha > 0$ using nominal 5% significance level and critical value of 1.645.

	T_{opt}		T_{μ}		T_{EW}	
Model	$N = 50$	$N = 100$	$N = 50$	$N = 100$	$N = 50$	$N = 100$
1	1.4	1.4	0.0	0.8	4.2	4.6
2	1.6	2.4	1.4	2.2	3.8	6.2
3	1.6	0.2	1.2	0.2	3.6	3.2
4	36.4	74.6	21.0	59.2	36.0	68.8
5	53.8	99.0	44.6	98.4	49.0	98.6
6	67.6	87.8	60.8	81.8	85.6	87.6

distribution with mean $\exp(\tau/2)$ and variance $\omega(\omega-1)$, where $\omega = \exp(\tau^2)$. No closed-form expression for the Poisson–Lognormal mixture is available [Shaban (1988)], but it can be shown that $E(y) = \mu\omega$ and $\text{var}(y) = E(y)[1 + (\omega - 1)E(y)]$, so $g(\mu) = \mu^2$ is appropriate in this case also. In the experiments we set $\tau = 0.31$ which implies $E[\exp(v)] = 1.168$, $\text{var}[\exp(v)] = 0.496$, and $(\omega - 1) = 0.101$. This imparts only a moderate degree of overdispersion and constitutes a stringent test of the hypothesis of zero overdispersion.

6.2. The results

In models 1, 2, and 3, μ is increased by varying β_0, but there is no overdispersion. Ideally the actual and nominal test sizes would coincide. However, the actual size of the two tests, T_{opt} and T_{μ}, for all three models and for both sample sizes turns out to be considerably smaller than the nominal size of 5% (see table 2). The match between the actual and asymptotic size match is worst for T_{μ} and best for T_{EW}. An indication of a departure of the empirical from the asymptotic distribution of T_{opt} is the presence of slight negative skewness in both distributions and for both sample sizes. This suggests that small-sample adjustments to T_{opt} along the lines of Dean and Lawless (1989) are desirable.

All comparisons of empirical power are subject to the qualification that the three tests appear to have different sizes and hence their relative power is not strictly comparable. For any one test, however, we may examine the improvement in power as either the sample size or the degree of overdispersion increases.

Thus the power of T_{opt} is relatively low (0.364) for model 4 and $N=50$ where overdispersion is 'low' – the range of observations is only [0, 12] compared with [0, 10] when there is no overdispersion. The powers of T_{opt} and T_{EW} are very similar in this case. When the sample size is increased from 50 to 100, the power of T_{opt} rises sharply from 0.364 to 0.746 and that of T_{EW} rises to 0.688. For both sample sizes the power of T_{μ} is considerably less than that of the other two. The fact that T_{opt} may have low power when overdispersion is low or moderate is not a very discouraging result since the efficiency of the Poisson regression is known to be high in this case [Cox (1983)].

In model 5 the typical range of observations is [0, 18]. In this case power increases to 0.538 for $N=50$ and 0.990 for $N=100$, indicating a very satisfactory performance of the test. As expected the power of T_{μ} is always lower than that of T_{opt}, but when overdispersion is high even this test performs very well. The high power of T_{EW} is an encouraging feature since the test is based on relatively weak distributional assumptions.

In model 6 the typical range of observations is [1, 14]. In this case we find that the performance of the estimators is somewhat similar to that in the case of model 4.

The overdispersion tests of model 6 perform reasonably well against the Poisson–Lognormal alternative, because in that case the form of the variance function is exactly that which is used in the regression. It is a reasonable conjecture that whenever the true variance function can be approximated by the function used here the test will have high power. In any case the test proposed here is valid against very general types of local departures from the null as shown in Cameron and Trivedi (1986).

7. Conclusion

In this paper we propose tests for the Poisson regression model that specify a relationship between the mean and variance under the alternative hypothesis, but do not require specification of the distribution under the alternative. The tests are easy to implement – the optimal regression-based statistic is computed as the t-test from an OLS regression. Monte Carlo analysis suggests that the tests perform particularly well in samples of size 100 and reasonably well in samples of size 50, but that there may be scope for small-sample adjustments which would improve the size characteristics of T_{opt} and T_{μ}. By contrast T_{EW} appears to need a smaller adjustment.

Should the null hypothesis be rejected, we might 'accept' the alternative hypothesis. The Poisson MLE of β is still consistent (provided the mean is correctly specified), but the usual computer output gives incorrect standard errors. We can obtain correct standard errors using the variance–mean relationship under the alternative hypothesis. Better still, we can obtain more efficient estimators for β, as detailed in McCullagh and Nelder (1983) and Gourieroux, Montfort, and Trognon (1984). These methods require an estimator of α, consistent under the alternative hypothesis. Several estimators have been suggested. Yet another is the estimated coefficient from the regression (2.18) used to compute the optimal regression-based test.

Finally, we note that the motivation and essential theory of these tests exploits only the relationship between mean and variance. In principle this approach can be extended beyond tests of mean–variance equality to tests of other relationships between mean and variance, i.e., to generalized linear models other than the Poisson.

References

Cameron, A.C. and P.K. Trivedi, 1985, Regression based tests for overdispersion, Econometric workshop technical paper no. 9 (Stanford University, Stanford, CA).

Cameron, A.C. and P.K. Trivedi, 1986, Econometric models based on count data: Comparisons and applications of some estimators and tests, Journal of Applied Econometrics 1, 29–54.

Chester, A., 1984, Testing for neglected heterogeneity, Econometrica 52, 865–872.

Collings, B.J. and B.H. Margolin, 1985, Testing goodness of fit for the Poisson assumption when observations are not identically distributed, Journal of the American Statistical Association 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, Journal of the American Statistical Association 84, 467–472.

Eicker, F., 1967, Limit theorems for regressions with unequal and dependent errors, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1, 59–82.

Gourieroux, C., A. Montfort, and A. Trognon, 1984, Pseudo maximum likelihood methods: Applications to Poisson models, Econometrica 52, 701–720.

Greenwood, M. and G.U. Yule, 1920, An enquiry into the nature of frequency distributions and multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents, Journal of the Royal Statistical Society A 83, 255–279.

Hausman, J., B.H. Hall, and Z. Griliches, 1984, Econometric models for count data with an application to the patents–R&D relationship, Econometrica 52, 909–938.

Katz, L., 1965, Unified treatment of a broad class of discrete distributions, in: G. Patil, ed., Proceedings of the international symposium on classical and contagious discrete distributions (Statistical Publishing Company, Calcutta) 175–182.

Kennan, J., 1985, The duration of contract strikes in US manufacturing, Journal of Econometrics 28, 5–28.

Lee, L.-F., 1986, Specification tests for Poisson regression models, International Economic Review 27, 689–706.

McCullagh, P. and J.A. Nelder, 1983, Generalized linear models (Chapman and Hall, London).

Newey, W.K., 1984, A method of moments interpretation of sequential estimators, Economics Letters 14, 201–206.

Newey, W.K., 1985, Maximum likelihood specification testing and conditional moment tests, Econometrica 53, 1047–1070.

Pagan, A.R., 1986, Two stage and related estimators and their applications, Review of Economic Studies 53, 517–538.

Shaban, S.A., 1988, Poisson–lognormal distributions, Ch. 7 in: E.L. Crow and K. Shimizu, eds., Log-normal distribution, theory and practice (Marcel Dekker, New York).

Tauchen, G., 1985, Diagnostic testing and evaluation of maximum likelihood models, Journal of 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., 1982, Maximum likelihood estimation of misspecified models, Econometrica 50, 1–25.

White, H., 1983, Estimation, inference and specification analysis, Discussion paper no. 83-26 (University of California, San Diego, CA).