ALTERNATIVE FORMS OF THE SCORE TEST FOR HETEROGENEITY IN A CENSORED EXPONENTIAL MODEL

Sanjiv Jaggia*

Abstract—Different versions of the score test for neglected heterogeneity for a right censored exponential model are analyzed. These tests depend on how the information matrix is estimated. A test based on the theoretical information matrix is shown to outperform all the other tests. Further, the noncentrality parameter of the test is examined to show how the power of the test is reduced when data are censored.

I. Introduction

Parametric models have been used widely in the analysis of the duration of time until the occurrence of some event of interest. A peculiar feature of such models is that data on durations are seldom complete. It is common for some observations to be censored, typically right censored. Information from the censored observations is extracted for estimation using standard maximum-likelihood techniques. However, the effect of censored data on the performance of diagnostic tests is not very clear. Testing of censored models is important since they may be especially sensitive to specification errors. In this note the effect of the censoring on different versions of the score test for neglected heterogeneity is examined.

In order to implement a score test, the information matrix has to be evaluated under the null hypothesis. With censored data, the theoretical (expected) information cannot be found without making additional assumptions regarding the censoring mechanism. In this note a heterogeneity test for an exponential model is derived that is based on the theoretical information matrix. This test uses the most commonly occurring, type I, right censored data. The Monte Carlo experiments indicate that such a test performs better than the tests, in the extant literature, that are based on the observed (sample) information matrix. Further, the noncentrality parameter of the test is examined to show how the power of the test is affected by the length of the data acquisition period.

The rest of the paper is organized as follows. Section II discusses various versions of the heterogeneity test. A test based on the theoretical information matrix is developed. The Monte Carlo analyses of all tests are presented in section III. Section IV concludes.

II. Background

A common cause of censoring is finite observation periods, which leads to right censored data. Individuals are observed over fixed time periods, and some durations may not have ended when the observation period ends. Typical data consist of \( ti, Xi, \) and \( Ci, \) where \( ti \) is the observed duration, \( Xi \) is a vector of explanatory variables, and \( Ci \) equals 1 if the duration is complete and 0 if censored.

\[
  t_i = \min (T_i, L_i) \quad \text{and} \quad C_i = 1 \quad \text{if} \quad T_i \leq L_i. \tag{1}
\]

The \( T_i \)'s are the actual durations and the \( L_i \)'s are the censoring times. The log-likelihood function is

\[
  L = \sum_{i=1}^{N} \left[c_i \ln \left[f(t_i|X_i)\right] + (1 - C_i) \ln \left[S(t_i|X_i)\right]\right]. \tag{2}
\]

where \( f(t) \) is the density function and \( S(t) \) is the survivor function of \( T. \) For an exponential model, \( f(t_i|X_i) = \mu_i \exp (-\epsilon_i) \) and \( S(t_i|X_i) = \exp (-\epsilon_i), \) where \( \mu_i = \exp (X_i \beta) \) and \( \epsilon_i = \mu_i \delta_i. \) When heterogeneity is present, \( \mu_i = \exp (X_i \beta) + U_i = V_i \exp (X_i \beta), \) where \( V_i = \exp (U_i) \) represents unobserved heterogeneity. Let \( \sigma^2 \) represent the variance of \( V. \) For small \( \sigma^2, \) the density function can be approximated by a second-order Taylor series expansion around the unit mean of \( V \) as

\[
  f(t_i|X_i, V) = f(t_i|X_i, V = 1) + (V - 1)(\partial^2 f(t_i|X_i, V)\partial V^2)|_{V=1} + \frac{1}{2}(V - 1)^2(\partial^2 f(t_i|X_i, V)\partial V^2)|_{V=1}. \tag{3}
\]

Received for publication March 4, 1994. Revision accepted for publication June 16, 1995.

* Suffolk University.

1 See Kiefer (1988) and Lancaster (1990) for a survey of duration models.

2 Horowitz and Neumann (1989) show that the standard diagnostics may lead to erroneous conclusions when data are censored. See also Jaggia and Trivedi (1994).
Since $V$ is not observable,  
\[
f(t_i|x_i) = E_{\theta_i}[f(t_i|x_i,V)] = \{\mu_i \exp(-\epsilon_i) [1 + (\sigma^2/2)(\epsilon_i^2 - 2\epsilon_i)]\}. 
\]
The corresponding likelihood function is  
\[
L = \prod_{i=1}^{N} \{C_i[\ln(\mu_i) + \ln(1 + (\sigma^2/2)(\epsilon_i^2 - 2\epsilon_i))] \\
- \epsilon_i + (1 - C_i)[\ln(1 + (\sigma^2\epsilon_i^2)/2)].
\]

A score test of heterogeneity can be constructed that tests for $\sigma^2 = 0$. Score or Lagrange multiplier (LM) tests are based on the estimation of only the null model. These tests are easy to implement as opposed to the Wald or likelihood ratio tests, for which the alternative model has to be estimated. Further, a score test is particularly suitable for this application since $\sigma^2$ is being tested on the boundary of the parameter space. Consider  
\[
H_0: \quad \sigma^2 = 0.
\]
Let $\Theta = (\sigma^2, \beta')$. Let $s_1(\Theta_0)$ and $I(\Theta_0)$ denote the score vector and the information matrix evaluated under the null hypothesis, where $s_1(\Theta) = \partial L / \partial \sigma^2, s_2(\Theta) = \partial L / \partial \beta'$, and $I(\Theta) = -E[\partial^2 L / \partial \Theta \partial \Theta']$. Here,  
\[
\frac{\partial L}{\partial \sigma^2}\bigg|_{\Theta_0} = \frac{1}{2} \sum (\epsilon_i^2 - 2C_i \epsilon_i) = s_1(\Theta_0).
\]
The test of neglected heterogeneity is  
\[
LM = s_1'(\hat{\Theta}_0) I(\hat{\Theta}_0) s_1(\hat{\Theta}_0) 
\]
where $(-)$ denotes evaluation at the maximum-likelihood estimates and $I^{11} = [I_{11} - I_{12}(I_{22})^{-1}I_{21}]^{-1}$ is the partitioned inverse of $I(\Theta)$. LM is asymptotically $\chi^2(1)$ distributed under $H_0$.

In order to implement the test, $I^{11}$ has to be evaluated. When all the observations are complete, the theoretical information matrix is easily computed and $I^{11}$ equals $1/N$. When data are censored, such a matrix cannot be found without making additional assumptions regarding the censoring mechanism. Here $-I(\Theta) = E(H_{0i})$, where $H_{0i} = \partial^2 L / \partial \Theta \partial \Theta'$ and the expectation is taken with respect to the joint distribution of $t_i$ and $C_i$ conditional on $X_i$. Moreover, since $C_i$ assumes only two values,  
\[
-I(\Theta) = E(H_{0i}) = 1 \quad P(C_i = 1) \\
+ E(H_{0i})|C_i = 0) \quad P(C_i = 0)
\]
where  
\[
E(H_{0i})|C_i = 1) = \int H_{0if}(t_i|C_i = 1) \, dt
\]
and $E(H_{0i})|C_i = 0) = H_{0i}$ evaluated at $C_i = 0$ and $t_i = L_i$. Notice that $L_i$ does not have to be the same for all individuals as all individual durations may not start on the same date, even though they are all observed until a common predetermined date. In such a situation, $L_i$ will be different for various individuals and, hence, is random. With $L_i$'s known, the following simplification aids in computing equation (9):

\[
E(H_{0i})|C_i = 1) = \frac{1}{P(T_i < L_i)} \int H_{0if}(t_i) \, dt.
\]

Also, $P(C_i = 0) = P(T_i > L_i) = 1 - P(C_i = 1) = \exp(-\mu L_i)$.

In this note, the above simplifications are used to derive the score test based on the outer product of the sample scores as an estimate of the information matrix evaluated under the maximum-likelihood estimates and the expectation is taken with respect to the joint distribution of $t_i$ and $C_i$.

\[
H_{0i} = \frac{1}{2} \sum (\epsilon_i^2 - 2C_i \epsilon_i) = s_1(\hat{\Theta}_0).
\]

Note that if $L_i \rightarrow \infty$, implying no censoring, $\exp(-\epsilon_i) = 0$. Hence $P(C_i = 1) = 1/N$, as derived with uncensored data.

Kiefer (1984) proposes a test that uses the sample variance of $s_1(\Theta)$. Consider the mean score $(1/N) \int \partial L / \partial \sigma^2 = (1/N) s_2$. The suggested variance of the mean score is $(1/N^2) \Sigma (s_i - s) \Sigma^2$, where $s_i$ is the sample mean of $s$. As the covariance between the scores is ignored, the above variance is overestimated, which would result in an underrejection of $H_0$. Kiefer recommends a one-tailed test as it is based on testing for $\sigma^2 = 0$ against $\sigma^2 > 0$.

For a more satisfactory test, one has to allow for the possible correlation of $s_1(\Theta)$ with $s_2(\Theta)$, even though the allowance is based on the sample information matrix. Under the usual regularity conditions, one can use either the sample Hessian of the log-likelihood function or simply the outer product of the sample scores as an estimate of $I(\Theta)$. The score test based on the outer product of the sample scores is identical to White’s information matrix test as suggested by Lancaster and Cheshier (1985).

### III. Monte Carlo Experiments

The Monte Carlo experiments are conducted to study the size and the power properties of the various versions of the heterogeneity test described above. LM is based on the theoretical information matrix. LM is based on the outer product of the sample scores, LM is based on the sample Hessian, and LM is Kiefer’s test and is one-tailed.

A sample size of 200 is used in most experiments. Further, $\mu = \exp(\beta_0 + X_i \beta_1 + X_2 \beta_2 + U)$ is considered where $\beta' = (-2, -2, 1)$. $X_i$ and $X_2$ are $N(0, 1)$ distributed and are held fixed for all experiments. Further, since the score test developed in this paper is based on an approximation (see equations (3) and (4)), the lognormal as well as the gamma heterogeneity distributions are considered to check the power of these tests. Both distributions are generated with a unit mean and a

5 The details of the derivation are available from the author on request. The information matrix for the Weibull model can be derived similarly, though some of its components involve the numerical evaluation of incomplete digamma and trigamma functions.

4 See Godfrey (1988) for an exposition and the required regularity conditions.
Censoring.

The performance of LM with larger samples, the rejection rate of LM falls within the above interval. For LM, the proportion of rejections is higher than the nominal level. The performance of LM cannot be evaluated accurately as the estimated information matrix is not always nonnegative definite. Thus the resulting test statistic becomes meaningless sometimes owing to sampling fluctuations (the estimated $I^T$ is negative in about 19% of the 500 replications). As expected, Kiefer’s test underrejects.

In order to check whether the size of these tests statistics improves with larger samples, samples of 500 and 1000 observations are also considered (see tables 2 and 3). It is found that even though the nominal size of LM gets closer, it is still significantly higher than the chosen level of 5%. For implementing LM, the estimated $I^T$ is negative in only about 2.8% and 0.33% of the replications with samples of 500 and 1000, respectively. Even though the positive-definiteness problem is alleviated with larger samples, the rejection rate of LM is still significantly higher than 5%.

The power of the above tests, against distributional alternatives of lognormal and gamma heterogeneity, is shown in tables 4 and 5. Once again, LM is found to be most powerful with both lognormal as well as gamma heterogeneity. Even though the number of rejections of LM seems similar to that of LM, a comparison of the power of the two tests cannot be made as the size of these tests is shown to be different. On a size-corrected basis, the power of LM would be less than that of LM.

It is interesting to note that the number of rejections with censored data are less than that with uncensored data for all tests. This decrease in power can be explained through the noncentrality parameter of the test. Given $H_0$: $\sigma^2 = 0$, let the sequence of alternative hypotheses be given by $H_\ell$: $\sigma^2 = \sqrt{\ell}$. Then, asymptotically, the test is distributed as $\chi^2(1, \eta)$, where the noncentrality parameter $\eta = (1/N)^2 I^T_l$. When all observations are complete, $I^T_0 = 1/N$ and $\eta = I^T_0$. With the above-censored data, $I^T_\ell$ depends on parameters $\beta$ as well as censoring times $L$. Given $\beta = [-2, -2, 1]$, evaluated at the mean $X$, along with $L = 20$, $\eta$ equals 0.499$r^2$. This value is much smaller than the one derived with complete observations. Therefore the power of the tests is affected by the length of the data acquisition period, and it becomes smaller as the period is reduced. With $L = 5$, $\eta$ equals 0.018$r^2$.

**IV. Conclusion**

In this note various versions of the heterogeneity test for a right censored exponential model are analyzed. Different versions of the test depend on the choice of estimate of the information matrix. A test based on the theoretical information matrix is derived that performs better than the ones that are based on the observed information matrix. Further, it is shown that the power of the test decreases when the observation period (censoring times) is reduced.

**REFERENCES**


---


7 See Cox and Hinkley (1974).
NOTES


**TESTS OF THE SPECIFICATION OF UNIVARIATE AND BIVARIATE ORDERED PROBIT**

J. S. Butler and Patrali Chatterjee*

Abstract—This note presents tests of the specification of univariate and bivariate ordered probit. The test is sensitive to deviations from either normality or the exogeneity of the explanatory variables. As an example, the ownership of dogs and televisions, both sources of time-intensive entertainment, is studied. The specification for dogs is not rejected, the specification for televisions is rejected at the 2.0% level, and the specification for both together is rejected at the 1.3% level.

I. Introduction and Literature Review

Ordered probit models represent situations in which a discrete outcome represents greater affinity, preference, or propensity for a good or outcome. Examples include children or, in this paper, dogs and televisions. The underlying propensity could represent a tendency or quality; examples include discrete quality measures and contract provisions negotiated as a function of bargaining strength.

The assumptions of the ordered probit model include a list of explanatory variables that affect the dependent variable and are exogenous, i.e., uncorrelated with either the normally distributed latent disturbance or the prediction error from the model. Maximum-likelihood estimation (MLE) maintains and efficiently employs the assumptions, but is inconsistent if any of the assumptions are invalid. This note proposes tests of the assumptions of normality and exogeneity using estimation by the generalized method of moments (GMM).

The null hypothesis of the GMM test of the specification is a joint hypothesis that the latent dependent variable is distributed normally or as a bivariate ordered probit, and the effect of normality. If the test does not reject the joint null hypothesis, however, MLE could be used with stronger assurance that specification error is not present.

Ordered probit models are applied rarely in bivariate models. We report here estimation in such a model after testing the model specification. We examine ownership of dogs and televisions, both sources of time-intensive entertainment.

See Maddala (1983) for the earliest uses of ordered probit models in economics. The papers cited here use ordered probit in two-equation models. Jimenez and Kugler (1987), Frazis (1993), and Butler et al. (1994) use ordered probit models as the first stage of a selection bias model. All are studying aspects of education: the effect on earnings of in-service training in a Colombian program, the effect on earnings of schooling choice concerning college, and the effect on grades in intermediate microeconomics of calculus classes. A different two-stage model with ordered probit in the first stage is used by Kao and Wu (1990), who study the default risk of bonds (first stage) and the yield on bonds (second stage). Amel and Liang (1994) model the entry of banks into new markets by probit or ordered probit and, as a second stage, the market performance of banks.

Gustaffson and Stafford (1992) study the decisions of Swedish mothers to work and to receive public child care subsidies. They use ordered probits to model the decision to work in three ranges. Their model does not allow correlated disturbances.

Calhoun (1989, 1991) uses bivariate ordered probit models to study the relationship between desired and excess fertility. The dependent variables are children ever born CEB and desired family size DFS. DFS and CEB can be estimated as a bivariate ordered probit, and the DFS can be censored in that it can be reported as the number of children ever born, even if the DFS is less than CEB. The censored model then takes DFS as reported if DFS exceeds CEB or as CEB or less if DFS is reported to be CEB. That avoids asking about unwanted births (DFS less than CEB). Calhoun (1989, 1991) thus estimates a censored model not used in this paper, but does not test the specification.

II. The Ordered Probit Model and the Bivariate Ordered Probit Model

We begin by specifying the bivariate ordered probit model, then describing the univariate test, and finally describing the bivariate test.

We indicate the two ordered probit indexes by subscripts \(a\) and \(b\), functions of single indices \(z_a\) and \(z_b\), which are functions of exogenous variables \(X\) and coefficients \(\beta_a\) and \(\beta_b\). The exogenous variables need not be the same in the two equations, and a simultaneous-equations model in the two indexes can be estimated if each equation includes at least one regressor omitted from the other. Subscripts indicating individual observations are suppressed. The unobserved propensities are defined as

\[
y_a^* = X'\beta_a + \epsilon_a = z_a + \epsilon_a \tag{1}
\]

\[
y_b^* = X'\beta_b + \epsilon_b = z_b + \epsilon_b. \tag{2}
\]