ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 30 — 1993

Selecting Estimated Models Using Chi-Square Statistics

Quang VUONG, Weiren WANG *

ABSTRACT. – This paper proposes some tests for choosing between two estimated models using some Pearson type statistics. We allow for arbitrary $\sqrt{n}$-asymptotically normal estimators to be used in forming these statistics. This provides some flexibility in practice. Then Large Sample theory and bootstrap methods are used to construct our tests.

Choix entre modèles estimés à partir de statistiques d'ajustement

RÉSUMÉ. – Cet article propose des tests de choix entre modèles estimés à partir de statistiques de Pearson d'ajustement. Ces statistiques peuvent être construites à partir d'estimateurs $\sqrt{n}$-asymptotiquement normaux, ce qui permet une grande flexibilité dans l'utilisation de nos tests. La théorie asymptotique et les méthodes de bootstrap sont alors utilisées pour construire nos tests.

---

* Q. VUONG: INRA, University of Southern California; W. WANG: University of Kentucky.
Acknowledgement: We would like to thank C. Hsiao, A. Weiss, F. Kwan, two referees and seminar participants at the University of Chicago, the University of Kentucky, Free University and Tilburg University for helpful comments. The second author wants to express his gratitude for partial financial support from a fellowship at CentER for Economic Research at Tilburg University.

1 Introduction

PEARSON [1990] type chi-square statistics have been generally used to test goodness-of-fit, *i.e.*, to test whether a specified parametric model is consistent with observed data. COCHRAN [1952], WATSON [1959], and MOORE [1978, 1986] have provided comprehensive surveys on Pearson chi-square type statistics, *i.e.*, quadratic forms in the cell frequencies. Recently, ANDREWS [1988*a*, 1988*b*] has extended the Pearson chi-square testing method to non-dynamic parametric econometric models, *i.e.*, to models with covariates. Because Pearson chi-square statistics provide natural measures for the discrepancy between the observed data and a specific parametric model, they have also been used for discriminating among competing models. Such a situation is frequent in Social Sciences where many competing models are proposed to fit a given sample. A well known difficulty is that each chi-square statistic tends to become large without an increase in its degrees of freedom as the sample size increases. As a consequence goodness-of-fit tests based on Pearson type chi-square statistics will generally reject the correct specification of every competing model.

To circumvent such a difficulty, a popular method for model selection, which is similar to the use of the AKAIKE [1973] Information Criterion (AIC), consists in considering that the lower the chi-square statistic, the better is the model. Hence the parametric model with smaller value of chi-square statistic is generally chosen. Such a use of chi-square statistics has been suggested by various researchers. *See* MASSY, MONTGOMERY and MORRISON [1970], HECKMAN [1981], and NAKAMURA and NAKAMURA [1985] among others.

The preceding selection rule, however, is not entirely satisfactory. Since chi-square statistics depend on the sample and are therefore random, their actual values are subject to statistical variations. Hence a model with a smaller chi-square statistic is not necessarily better than one with a larger chi-square statistic in terms of goodness-of-fit. To take into account statistical variations, we shall propose some convenient asymptotically standard normal tests for model selection based on Pearson type chi-square statistics. Following VUONG [1989] our tests are testing the null hypothesis that the competing models are as close to the data generating process (DGP) against the alternative hypotheses that one model is closer to the DGP where closeness of a model is measured according to the discrepancy implicit in the Pearson type chi-square statistics. Thus the outcomes of our tests provide information on the strength of the statistical evidence for the choice of a model based on its goodness-of-fit.

Following MOORE [1977, 1978] and ANDREWS [1988*b*], we consider a general class of estimators that is very broad and contains most estimators currently used in practice when forming Pearson type statistics. This covers the case studied in VUONG and WANG [1993] where only the corresponding

minimum chi-square estimator is used. In practice, the use of "unmatched" estimators, *i.e.*, estimators that do not minimize the chosen Pearson type statistic, is quite common and can be found in various works. *See* LINHART and ZUCCHINI [1986], HECKMAN [1981], and NAKAMURA and NAKAMURA [1985]. Such a generalization is useful for many reasons.

First, there frequently exist estimators that are easier to compute than the corresponding minimum chi-square estimator. Second, when the original individual data are available, one uses frequently more efficient estimators, such as the ungrouped maximum likelihood estimator (MLE), *i.e.* the MLE based on individual data, when forming the chi-square type statistic. This is in fact the procedure recommended in many textbooks. See for instance, BISHOP, FIENBERG and HOLLAND [1975]. Third, given an estimation method, a particular weighting matrix is often chosen to obtain a chi-square limiting null distribution for the chi-square statistic. This leads frequently to unmatched estimators. For example, the RAO-ROBSON [1974] statistic has the usual chi-square limiting null distribution when the ungrouped MLE is used. However, the ungrouped MLE does not minimize the quadratic form defining the Rao-Robson statistic. See CHERNOFF and LEHMANN [1954], MOORE and SPRUILL [1975], MOORE [1977, 1978] and ANDREWS [1988*a*].

The paper is organized as follows. Section 2 introduces the basic notations and defines a class of asymptotically normal (AN) estimators. Section 3 investigates the model selection problem based on Pearson type statistics. A large sample test is proposed. In section 4, EFRON [1982] bootstrap method is used to propose alternative and simpler testing procedures for model selection. Section 5 presents some simulation results. Section 6 concludes the paper and mentions some extensions. Proofs and tables are included in Appendixes 1 and 2.

2 Definitions and Assumptions

In this section, we briefly present the basic assumptions on the model and parameter estimators, and we define our general chi-square type statistics.

ASSUMPTION A1: The observed data X_i, $i=1, 2, \ldots$ are independent and identically distributed (iid) with some common true distribution H. Assumption A1 is more suitable for cross-section than time series data. Following RIVERS and VUONG [1991], some of our results can be extended to heterogenous and weakly dependent processes.

The sample space X is partitioned into M mutually disjoint fixed cells $E_1, E_2, \ldots, E_M$. The partition sometimes follows from the qualitative nature of the data. Let $h=(h_1, h_2, \ldots, h_M)$ be the vector of true cell

probabilities. Let a specified model be $F_\theta = \{F(.\,|\,\theta);\ \theta \in \Theta \subset R^k\}$ and denote the vector of its predicted cell probabilities by

$$p(\theta) = [p_1(\theta), p_2(\theta), \ldots, p_M(\theta)]' \qquad \text{where} \quad p_i(\theta) = \int_{E_i} dF(x\,|\,\theta).$$

Because $F(.\,|\,\theta)$ is a joint distribution for X_i, we rule out implicitly econometric models since the latter leave unspecified the distribution of the exogenous variables. This assumption is made for simplicity. We will return to it in section 6.

We impose the following assumption on h and $p(\theta)$:

ASSUMPTION A2: $h_i > 0$, $p_i(\theta) > 0$, and $p_i(\theta)$ is twice continuously differentiable for every $i = 1, 2, \ldots, M$.

Throughout it is assumed that F_θ satisfies some standard additional regularity conditions to ensure the asymptotic results presented subsequently. See for instance MOORE and SPRUILL [1975] and MOORE [1984].

Let n be the sample size. Corresponding to the partition $E_1, E_2, \ldots, E_M$, we can compute the vector of observed cell probabilities

$$f = (f_1, f_2, \ldots, f_M)' \qquad \text{where} \quad f_i = \frac{1}{n}\sum_{j=1}^{n} 1_{E_i}(X_j) \qquad \text{for} \quad i = 1, 2, \ldots, M,$$

and $1_{E_i}(X_j)$ is the indicator function taking values 1 if X_j falls in cell E_i or 0 otherwise. Following MOORE [1978], it is convenient to consider the M-dimensional vector

$$V_n(\theta) = \sqrt{n}\left(\ldots, \frac{f_i - p_i(\theta)}{\sqrt{p_i(\theta)}}, \ldots\right)',$$

which measures the discrepancy between the observed and the expected cell probabilities given θ.

We can now define the class of general chi-square type statistics considered in this paper. These statistics are essentially quadratic forms in $V_n(\theta_n)$. Formally, we have:

DEFINITION 1: A general chi-square type statistic is of the form

$$Q_n(\theta_n) = V_n'(\theta_n)\, M(f, \theta_n)\, V_n(\theta_n)$$

where $M(f, \theta)$ and θ_n satisfy assumptions A3 and A4 that follow.

ASSUMPTION A3: Each element of the weighting matrix $M(f, \theta)$ is twice continuously differentiable in $(f, \theta) \in R^M \times \Theta$, and $M(h, \theta)$ is a positive definite matrix for every θ.

ASSUMPTION A4: For some θ_0 in Θ, the estimator θ_n satisfies

$$\sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}} R_0^{-1} \sum_{i=1}^{n} \psi(X_i, \theta_0) + o_p(1) \quad \text{as} \quad n \to \infty$$

where $\psi(X, \theta_0)$ is a measurable function from $X \times \Theta$ to R^k that satisfies $E_H \psi(X, \theta_0) = 0$, and

$$V_0 \equiv E_H \psi(X, \theta_0) \psi(X, \theta_0)' \quad \text{and} \quad R_0 \equiv -E_H(\partial \psi(X, \theta_0)/\partial \theta')$$

are finite and nonsingular. $E_H(.)$ denotes expectation computed under the true data generating process H.

Assumption A4 implies that θ_n is a consistent estimator of some value θ_0 and that $\sqrt{n}(\theta_n - \theta_0)$ is asymptotically normally distributed with zero mean and covariance matrix $R_0^{-1} V_0 (R_0^{-1})'$. Given some suitable regularity conditions, most common estimators θ_n fulfill this assumption. For instance, minimum chi-square estimators, the maximum likelihood (ML) estimator on grouped or ungrouped data, any GMM estimator, and other extremum estimators satisfy assumption A4. *See* AMEMIYA [1985].

Note that the parameter value θ_0 depends on the underlying true distribution H which generates the observations, as well as the estimation method employed. For example, when one uses the minimum chi-square estimator that minimizes $Q_n(\theta)$, then θ_0 is the value minimizing $Q(\theta) \equiv p\lim_H Q_n(\theta)/n$, where the p lim is calculated under the true data-generating process H. This is the case of a "matched" estimation. On the other hand, if θ_n is the MLE of θ based on the ungrouped sample data, then θ_0 is the value at which the Kullback-Leibler information criterion (KLIC) $E_H[-\log f(x|\theta)]$ is minimized, where f is the density function corresponding to $F(.|\theta)$. *See* WHITE [1982]. This is the case of an "unmatched" estimation.

Note also that $Q_n(\theta_n)$ is quite general. It includes some well-known chi-square statistics such as the original Pearson statistic with $M_n = I_M$ (the identity matrix), the Modified Pearson statistic with

$$M_n = \text{diag}(\ldots, p_i(\theta)/f_i, \ldots),$$

the Gauss statistic with $M_n = \text{diag}(\ldots, p_i(\theta), \ldots)$ and the Rao-Robson statistic with M_n being the generalized inverse of the covariance matrix of $V_n(\theta_n)$ when θ_n is the ungrouped MLE. In practice, M_n is chosen by the researcher according to his or her preference or objective.

3 Selecting Estimated Models

As we mentioned earlier, chi-squares statistics are frequently used to discriminate among alternative models. It is easy to see that,

under the present regularity conditions, $Q_n(\theta)/n$ converges to $Q(\theta) \equiv V(\theta)' M(h, \theta) V(\theta)$ in probability as n goes to infinity, where $V(\theta) = [\ldots, (h_i - p_i(\theta))/\sqrt{p_i}(\theta), \ldots]'$. Thus $Q(\theta)$ can be viewed as measuring the departure of a particular member $F(.\,|\,\theta) \in F_\theta$ from the observed sample. It is also worth noting that $Q(\theta) \geqslant 0$ and $Q(\theta) = 0$ if and only if $h = p(\theta)$.

Of special interests to us is the situation in which a researcher has two competing parametric models F_θ and $G_\gamma = \{ G(.\,|\,\gamma);\ \gamma \in \Gamma \subset R^q \}$, select the better of the two models based on their general chi-square type statistics $Q_n(\theta_n)$ and $Q_n(\gamma_n)$ where θ_n and γ_n are general estimators satisfying assumption A4 and the same cells are used in both statistics.

DEFINITION 2 (Equivalent, Better and Worse): Consider two competing parametric models F_θ and G_γ and some chi-square type statistics $Q_n(\theta_n)$ and $Q_n(\gamma_n)$ where θ_n and γ_n are general estimators satisfying A4. Let $Q(.)$ be the probability limit of $Q_n(.)/n$. The hypotheses

$$\tilde{H}_0^e : Q(\theta_0) = Q(\gamma_0)$$
$$\tilde{H}_f : Q(\theta_0) < Q(\gamma_0)$$
$$\tilde{H}_g : Q(\theta_0) > Q(\gamma_0)$$

mean that the estimated models $F(.\,|\,\theta_0)$ and $G(.\,|\,\gamma_0)$ are equivalent, that $F(.\,|\,\theta_0)$ is better than $G(.\,|\,\gamma_0)$, and that $F(.\,|\,\theta_0)$ is worse than $G(.\,|\,\gamma_0)$, respectively.

Definition 2 calls for some remarks. First, it does not require that the same chi-square type discrepancy be used in forming $Q_n(\theta_n)$ and $Q_n(\gamma_n)$. Choosing, however, different discrepancies for evaluating competing models is hardly justified. Second and more importantly, it allows estimators other than the matching minimum chi-square estimators to be used. As pointed out in the introduction, the use of "unmatched" estimators is quite common in practice.

Third, even when the same chi-square discrepancy $Q(.)$ is used, it is important to note that the preceding hypotheses do not bear only on the problem of model selection which is that of choosing between the models F_θ and G_γ and not between the estimated models $F(.\,|\,\theta_n)$ and $G(.\,|\,\gamma_n)$. For, when one uses unmatched estimators, then the probability limits θ_0 and γ_0 of the estimators θ_n and γ_n are not in general equal to the pseudo true values $\theta_* = \operatorname{argmin} Q(\theta)$ and $\gamma_* = \operatorname{argmin} Q(\gamma)$. As a consequence $Q(\theta_0)$ (say) cannot be interpreted as the discrepancy between the model F_θ and the DGP H. In other words, the preceding hypotheses concern the actual data-denerating process in relation to the competing models and the estimation methods employed.

Fourth, it may be interesting to evaluate the relative effects due to specification errors and the choice of estimation methods. One possibility

is to consider the matching minimum chi-square estimator $\tilde{\theta}_n$ and the corresponding value of the Pearson statistic $Q_n(\tilde{\theta}_n)$. Since $Q_n(\tilde{\theta}_n) \leqslant Q_n(\theta_n)$ by definition of $\tilde{\theta}_n$, then we can decompose $Q_n(\theta_n)$ as the sum of two nonnegative terms which are $[Q_n(\theta_n) - Q_n(\tilde{\theta}_n)]$ and $Q_n(\tilde{\theta}_n)$. Because $Q_n(\theta_n)$ combines both effects while $Q_n(\tilde{\theta}_n)$ takes into account specification errors only, then $[Q_n(\theta_n) - Q_n(\tilde{\theta}_n)]$ may be viewed as evaluating effects due to the choice of estimation methods. VUONG and WANG [1993] consider model selection based on $Q_n(\tilde{\theta}_n)$, while we consider here model selection based on $Q_n(\theta_n)$.

Fifth, consideration of the general hypotheses $\tilde{H}_0^e$, $\tilde{H}_f$, and $\tilde{H}_g$ may be useful by itself. For, one may be interested only in comparing the performance of the estimated models $F(.\,|\theta_n)$ and $G(.\,|\gamma_n)$, irrespective of how the models were actually estimated. In this case, $\tilde{H}_0^e$, $\tilde{H}_f$, and $\tilde{H}_g$ are the relevant hypotheses.

In any case, since θ_n and γ_n are consistent estimators of θ_0 and γ_0 by assumption A4, we can use $[Q_n(\theta_n) - Q_n(\gamma_n)]/n$ to consistently estimate the indicator $Q(\theta_0) - Q(\gamma_0)$ which will be zero under the null hypothesis $\tilde{H}_0^e$. Using a standard Taylor expansion, we can obtain the asymptotic distribution of $[Q_n(\theta_n) - Q_n(\gamma_n)]/\sqrt{n}$, which is normal with zero mean and variance ω^2 under $\tilde{H}_0^e$. The detailed derivation and the expression for ω^2 can be found in the Appendix 1. Hence we define the statistic

$$\text{GCM}_n = \frac{1}{\sqrt{n}} \frac{[Q_n(\theta_n) - Q_n(\gamma_n)]}{\hat{\omega}}$$

where $\hat{\omega}^2$ is a consistent estimator of ω^2 (e.g., its sample analog). (GCM stands for Generally Choosing Models.) We have

THEOREM 1: (Asymptotic Distribution of GCM_n Statistic): Given A1-A4, suppose that $\omega^2 \neq 0$, then

(i) under the null hypothesis $\tilde{H}_0^e$, $\text{GCM}_n \to N(0, 1)$ in distribution,
(ii) under the alternative $\tilde{H}_f$, $\text{GCM}_n \to -\infty$ in probability,
(iii) under the alternative $\tilde{H}_g$, $\text{GCM}_n \to +\infty$ in probability.

Theorem 1 is quite general and gives us a wide variety of asymptotic standard normal tests for model selection based on general chi-square type statistics. Part (ii) and (iii) also implies that the test is consistent. In the next section, we detail the testing procedures based on Theorem 1 by using bootstrap methods.

The preceding tests are based on the standardized difference in *unadjusted* generalized chi-square statistics. In some cases, especially when the sample size is small, one may want to adjust the numerator of the statistic GCM_n by some additive terms of the form $c_n(k, q)$. For instance, this is the case when one wants to take into account the parsimonious nature of the competing models. Provided the correction terms are $o(\sqrt{n})$, as most correction terms are, then Theorem 1 holds for the adjusted statistic.

4 Bootstrap Methods

Implementation of the model selection procedure proposed in section 3 requires the following computations:

(i) Estimation of the parameters θ_n and γ_n,

(ii) Computation of the two chi-squares statistics $Q_n(\theta_n)$ and $Q_n(\gamma_n)$ and the difference $\hat{S}_n \equiv [Q_n(\theta_n) - Q_n(\gamma_n)]/\sqrt{n}$,

(iii) Computation of the variance $\hat{\omega}^2$ of $\hat{S}_n$ and finally, computation of $GCM_n \equiv \hat{S}_n/\hat{\omega}$.

The estimators θ_n and γ_n can be obtained by minimizing some objective function, such as $Q_n(.)$, or by maximizing the likelihood function. Point (ii) is straightforward once θ_n and γ_n are known. Point (iii) is somewhat complicated. In particular, the formula for the variance $\hat{\omega}^2$ involves the calculation of the first and second order partial derivatives of the expected cell probabilities with respect to the parameters θ and γ. Moreover, such a calculation has to be repeated across models and estimation methods. This may not be convenient in applied work. Fortunately, with the help of advanced computers, this can be avoided via some simulations. This is the purpose of this section. Specifically, we shall consider a method for evaluating ω^2 based on EFRON [1982] bootstrap method. In addition, we shall propose two alternative testing procedures for model selection based directly on the bootstrap distribution of the statistic $\hat{S}_n$.

In the preceding section, we have seen that $\hat{S}_n$ is approximately normally distributed with mean zero and variance ω^2. This suggests that ω^2 can be estimated by the sample variance of $\hat{S}_n$ in a (large) number of independent and identical samples of size n. This is the basic idea underlying the bootstrap method which we apply here to the estimation of ω^2. Specifically, we carry out the following steps:

1. Let $\hat{F}$ be the empirical probability distribution of the original data $x_1, x_2, \ldots, x_n$, *i.e.*,

$$\hat{F}: \text{mass } 1/n \text{ at } x_i, \qquad i = 1, 2, \ldots, n.$$

Then draw an i.i.d. "bootstrap sample" $x_1^*, x_2^*, \ldots, x_n^*$ from $\hat{F}$, *i.e.*, draw x_i^* randomly with replacement from the observed values $x_1, x_2, \ldots, x_n$.

2. Using this bootstrap sample $\{x_i^*\}$, estimate the competing models to obtain θ_n^* and γ_n^*. Then calculate the statistic

$$\hat{S}_n^* = [Q_n(\theta_n^*) - Q_n(\gamma_n^*)]/\sqrt{n}.$$

3. Independently repeat steps 1 and 2 a large number of times B, say $B = 1\,000$. Obtain "bootstrap replications" $\hat{S}_n^{*1}, \hat{S}_n^{*2}, \ldots, \hat{S}_n^{*2}$, and compute the sample variance of $\{\hat{S}_n^{*b}, b = 1, \ldots, B\}$:

$$\hat{\omega}_*^2 = \frac{1}{B} \sum_{b=1}^{B} (\hat{S}_n^{*b} - \bar{S}^*)^2,$$

where $\bar{S}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{S}_n^{*b}$ is the average of "bootstrap replications".

The above method (call it Method 1) is quite general. It requires, however, that the individual data be available. If only the frequencies in every cell are available, the bootstrap method can still be used with a slight modification of the resampling procedure. In this case it is natural to assume that the estimators θ_n and γ_n require only grouped data. Then the modified procedure (Method 1') is:

1'. Given the observed cell probabilities (f_i, $i = 1, \ldots, M$) and the total sample size n, we can construct artificial data $a_1, a_2, \ldots, a_n$, such that there are nf_i sample points in cell i. For example, randomly pick up nf_i different points from cell i.

2'. Then draw "bootstrap sample" from this artificial data and calculate the new "bootstrap probabilities" $\{f_i^*, i = 1, \ldots, M\}$.

3'. Same as step 2 using $\{f_i^*, i = 1, \ldots, M\}$ instead of $\{x_i^*, i = 1, \ldots, n\}$, which is possible since the estimation methods use only grouped data.

4'. Indepently repeat steps 2' and 3' for B times and the rest is the same as step 3.

Another way of doing steps 1' and 2' is drawing M-dimensional multinomial variates with sample size n and probabilities $\{f_i, i = 1, \ldots, M\}$.

Once the bootstrap variance $\hat{\omega}_*^2$ is obtained, the test statistic GCM_n is calculated easily using the initial estimates θ_n and γ_n. Under suitable regularity conditions and for a large number of replications (*see* EFRON [1982]), $\hat{\omega}_*^2$ is a consistent estimator of ω^2. Thus, from Theorem 1, a testing procedure for model selection can be based on the comparison of the value of GCM_n to critical values from a standard normal table. For example, at 5% significance level, we compare GCM_n with -1.96 and 1.96. If GCM_n falls between -1.96 and 1.96, we conclude that both estimated models fit the data equally well. If GCM_n is less than -1.96 (or larger than 1.96), then we reject the null hypothesis in favor of the alternative hypothesis that the estimated model $F(.\,|\theta_n)$ (or $G(.\,|\gamma_n)$) is closer to the true distribution.

Although using the bootstrap method to obtain an estimate of ω^2, the basic justification of the preceding testing procedure comes from the asymptotic properties obtained in Theorem 1. In contrast, the next two procedures rely only on the bootstrap methodology, and in particular on two bootstrap methods for assigning approximate confidence intervals to $Q(\theta_0) - Q(\gamma_0)$ based on the bootstrap distribution of $\hat{S}_n^*$. These two

methods are discussed in detail in EFRON [1982], and require steps 1 and 2 defined above or steps 1′ and 2′ if only frequencies are observed. *See* also EFRON [1984] for the comparison of nonnested linear models using a MSE criterion.

The first testing method is based on the percentile method. Let

$$\hat{\mathrm{CDF}}(t) = \frac{\text{number of}\{b : \hat{S}_n^{*b} \leqslant t\}}{B} \equiv \mathrm{Prob}_*(\hat{S}_n^* \leqslant t)$$

be the empirical cumulative distribution function (CDF) of the bootstrap distribution of $\{\hat{S}_n^{*b}, b=1, \ldots, B\}$. For a given significance level α between 0.0 and 1.0, define

$$\hat{S}_L^*(\alpha/2) = \hat{\mathrm{CDF}}^{-1}(\alpha/2) \qquad \text{and} \qquad \hat{S}_U^*(\alpha/2) = \hat{\mathrm{CDF}}^{-1}(1-\alpha/2).$$

The percentile method consists in taking $[\hat{S}_L^*(\alpha/2), \hat{S}_U^*(\alpha/2)]$ as an approximate $1-\alpha$ central confidence interval for $Q(\theta_0) - Q(\gamma_0)$. Thus a test of the null hypothesis $\tilde{H}_0^e$ of equivalence against the alternative hypothesis $\tilde{H}_f$ or $\tilde{H}_g$ at the approximate α significance level is:

(i) accept the null $\tilde{H}_0^e$ of equivalence if $0 \in [\hat{S}_L^*(\alpha/2), \hat{S}_U^*(\alpha/2)]$,

(ii) reject $\tilde{H}_0^e$ in favor of $\tilde{H}_f$ if $\hat{S}_U^*(\alpha/2) < 0$,

(iii) reject $\tilde{H}_0^e$ in favor of $\tilde{H}_g$ if $\hat{S}_L^*(\alpha/2) > 0$.

Or equivalently,

(i) accept $\tilde{H}_0^e$ if $\hat{\mathrm{CDF}}(0) \in [\alpha/2, 1-\alpha/2]$,

(ii) reject $\tilde{H}_0^e$ in favor of $\tilde{H}_f$ if $\hat{\mathrm{CDF}}(0) > 1-\alpha/2$,

(iii) reject $\tilde{H}_0^e$ in favor of $\tilde{H}_g$ if $\hat{\mathrm{CDF}}(0) < \alpha/2$.

The percentile method does not use the value $\hat{S}_n$ for the initial observed sample. More importantly, since the bootstrap distribution is based on replications of the observed sample that produces the value $\hat{S}_n$, the percentile method assumes implicitly that $\hat{S}_n$ is the median of the bootstrap distribution. If this is not a proper assumption, one should incorporate a bias adjustment. This leads to the bias-correction percentile method. We will only present the procedure. Its rationale can be found in Chapter 10 of EFRON [1982].

Define

$$\hat{z} = \Phi^{-1}(\hat{\mathrm{CDF}}(\hat{S}_n)) \qquad \text{and} \qquad z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$$

where Φ is the cumulative distribution function for a standard normal variable. The decision rule for model selection based on the bias-correction percentile method at the α significance level is:

(i) accept $\tilde{H}_0^e$ if $0 \in [\hat{\mathrm{CDF}}^{-1}(\Phi(2\hat{z} - z_{\alpha/2})), \hat{\mathrm{CDF}}^{-1}(\Phi(2\hat{z} + z_{\alpha/2}))]$,

(ii) reject $\tilde{H}_0^e$ in favor of $\tilde{H}_f$ if $\hat{\mathrm{CDF}}^{-1}(\Phi(2\hat{z} + z_{\alpha/2})) < 0$,

(iii) reject $\tilde{H}_0^e$ in favor of $\tilde{H}_g$ if $\hat{\mathrm{CDF}}^{-1}(\Phi(2\hat{z} - z_{\alpha/2})) > 0$.

Let $\hat{z}_0 \equiv \Phi^{-1}(\hat{\text{CDF}}(0))$. It is easy to see that the preceding decision rule is equivalent to:

(i) accept $\tilde{H}_0^e$ if $|\hat{z}_0 - 2\hat{z}| < z_{\alpha/2}$,

(ii) reject $\tilde{H}_0^e$ in favor of $\tilde{H}_f$ if $\hat{z}_0 > 2\hat{z} + z_{\alpha/2}$,

(iii) reject $\tilde{H}_0^e$ in favour of $\tilde{H}_g$ if $\hat{z}_0 < 2\hat{z} - z_{\alpha/2}$.

We will refer to the percentile method without the bias-correction as Method 2, and to the percentile method with the bias-correction as Method 3.

5 An Example

To illustrate the model selection procedures discussed in the preceding section, namely, Methods 1, 2 and 3, we consider an example. The limited Monte Carlo study that we conduct will also give an idea on the relative performance of these methods. We need to define the competing models, the estimation method used for each competing model, and the chi-square type statistic used to measure the departure of each proposed parametric model from the true data generating process. These are now presented.

For our competing models, we consider the problem of choosing between the family of log-normal distributions and the family of exponential distributions. This problem has a long history in the statistical literature. *See*, e.g., Cox [1962] and Atkinson [1970] among others. The log-normal distribution is parameterized by $\alpha = (\alpha_1, \alpha_2)$ and has density

$$f(x; \alpha_1, \alpha_2) = \frac{1}{x(2\pi)^{1/2}\alpha_2} \exp\left(-\frac{(\log x - \alpha_1)^2}{2\alpha_2^2}\right) \qquad \text{for} \quad x > 0$$

and zero otherwise. The exponential distribution with parameter β has density

$$g(x; \beta) = \frac{1}{\beta} \exp(-x/\beta) \qquad \text{for} \quad x > 0$$

and zero otherwise.

The estimator used for each competing model is the ungrouped maximum likelihood estimator (MLE). This choice is particularly convenient here because the ungrouped ML estimator for each model has a closed form and hence is easily computed. Specifically, for the log-normal model,

$$\hat{\alpha}_1 = \frac{1}{n}\sum_{i=1}^{n} \log x_i \qquad \text{and} \qquad \hat{\alpha}_2^2 = \frac{1}{n}\sum_{i=1}^{n} (\log x_1 - \hat{\alpha}_1)^2.$$

For the exponential model, the ungrouped MLE is the sample average, *i.e.*,

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Lastly, we use the original Pearson chi-square statistic to evaluate the discrepancy of a proposed model from the true data generating process. We partition the real line into M intervals $\{(c_{i-1}, c_i), i=0, 1, \ldots, M\}$ where $c_0 = 0$ and $c_M = +\infty$. The choice of the cells is discussed below. The chi-square statistics for the log-normal and exponential models are:

$$Q_n(\alpha) = n \sum_{i=1}^{M} \frac{(f_i - p_i(\alpha))^2}{p_i(\alpha)} \quad \text{and} \quad Q_n(\beta) = n \sum_{i=1}^{M} \frac{(f_i - p_i(\beta))^2}{p_i(\beta)}$$

where $p_i(\alpha)$ and $p_i(\beta)$ are the probabilities of the interval (c_{i-1}, c_i) under $f(x, \alpha)$ and $g(x, \beta)$, respectively.

In our limited Monte Carlo study, we consider various sets of experiments in which the data are generated from a mixture of an exponential distribution and a log-normal distribution. These two distributions are calibrated so that they have the same population means and variances, namely one and one. Hence the data generating process has the density

$$h(\pi) = \pi \text{ Exponential}(1) + (1-\pi) \text{ Log-normal}(-0.346\,6, 0.832\,6),$$

where π is set to some specific value for each set of experiments. In each set of experiments, several random samples are drawn from this mixture of distributions. The sample size varies from 100 to 1,000, and for each sample size the number of replications is 1,000.

Throughout, the chosen partition has four cells defined by the values $c_0 = 0$, $c_1 = 0{,}1$, $c_2 = 1.0$, $c_3 = 3.0$ and $c_4 = +\infty$. Note that because the log-normal distribution has two parameters, four is the minimum number of cells for which a perfect fit is not always achieved when fitting this distribution by minimum chi-square methods. The power of our tests for model selection is likely to improve by increasing the number of cells. Note also that the shapes of the log-normal and exponential densities differ greatly around the origin. This motivates the choice of $c_1 = 0.1$. The value c_2 is equal to the common population mean, while c_3, which is two standard deviations away from the mean, is used to control for large deviations.

We choose five different values for π which are 0.0, 1.0, 0.5357, 0.25, and 0.75. Although our proposed model selection procedure does not require that the data generating process belong to either of the competing models, we consider the two limiting cases $\pi = 0.0$ and $\pi = 1.0$ for they correspond to the correctly specified cases. The value $\pi = 0.535\,7$ is determi-

ned to be the value for which the estimated log-normal distribution and the estimated exponential distribution are approximately at equal distance from the mixture $h(\pi)$ according to the Pearson discrepancy and the above cells. Thus this set of experiments corresponds approximately to the null hypothesis of our proposed model selection test GCM_n. Finally, to investigate the cases where both competing models are misspecified but not at equal distance from the data generating process, we consider the cases where $\pi = .25$ and $\pi = .75$. The former case corresponds to a data generating process which is log-normal but slightly contaminated by an exponential distribution. The second case is interpreted similarly with an exponential distribution slightly contaminated by a log-normal distribution.

The results of our five sets of experiments are presented in Tables 1-5 (*see* Appendix 2). The first half of each table gives the average values of the ungrouped ML estimators $\hat{\alpha}$ and $\hat{\beta}$, the Pearson goodness-of-fit statistics $Q_n(\hat{\alpha})$ and $Q_n(\hat{\beta})$, and the model selection statistic CGM_n with its bootstrap estimated variance $\hat{\omega}_*^2$ (*see* Method 1). The values in parentheses are standard errors. The second half of each table gives in percentage the number of times our proposed model selection procedures based on the three methods described in the previous section, favor the log-normal model, the exponential model, or are indecisive. The tests are conducted at the 5% nominal significance level. In the first two sets of experiments ($\pi = 0.0$ and $\pi = 1.0$) where one model is correctly specified, we use the labels "correct" and "incorrect" when a choice is made. Finally, in the case where $\pi = .5357$, we give in addition the 2.5%, 5.0%, 95%, and 97.5% fractiles of the observed distribution of the GCM_n statistic. This allows a comparison with the asymptotic $N(0, 1)$ approximation under our null hypothesis of equivalence.

Tables 1 and 2 report the cases when one model is correctly specified. It is well-known that the MLE is consistent for the true parameter value under correct specification. For example, in Table 1, the log-normal model is correctly specified, and the MLE of $\alpha = (\alpha_1, \alpha_2)$ approaches the true values $\alpha_0 = (-0.3466, 0.8326)$ as the sample size increases from 100 to 1,000. However, the Pearson chi-square statistic $Q_n(\hat{\alpha})$ for this model does not have a standard chi-square limiting distribution even under correct specification because the ungrouped MLE is used. In fact, the limiting distribution is somewhere between a $\chi^2(1)$ and a $\chi^2(2)$. *See* CHERNOFF and LEHMANN [1954]. In Table 1, $Q_n(\hat{\alpha})$ has a mean around 1.79 which lies between 1 (mean of $\chi^2(1)$) and 2 (mean of $\chi^2(2)$). For the misspecified model, which is the exponential model here, the MLE $\hat{\beta}$ converges to the pseudo-true parameter β_* which minimizes the KLIC. The corresponding Pearson chi-square statistic $Q_n(\hat{\beta})$, as we expect, increases at the rate of n. The bootstrap estimator of ω also converges as the sample size becomes larger. The test statistic for model selection GCM_n approximately increases at a rate $\sqrt{n}$. In Table 2, where the exponential model is correctly specified, one can observe similar results.

The second half of Table 1 summarizes the results for our three model selection procedures. Method 1 performs quite well and for small sample

sizes ($n=100$ or 250), this method seems to dominate the other two methods in selecting the correct model, which is the log-normal model in this case. However, as the sample size increases to $n=500$ or 1,000, the three methods perform equally well. The three methods also select the correct model almost 100% of the times, as expected.

The second half of Table 2 reports somewhat different results. Except at sample size 1,000 where all three methods perform equally well, in smaller samples Method 3 now seems to dominate the other two methods. All three methods, however, do not work as well when the exponential model is correctly specified (Table 2) as when the log-normal model is correctly specified (Table 1). This can be explained by the fact that the log-normal model has one more parameter than the exponential model, and hence is more difficult to reject even when it is misspecified.

For Tables 3, 4 and 5, the data was generated neither from the log-normal model nor from the exponential model, but from a mixture of these two models. Hence, the log-normal model and the exponential model are both incorrectly specified. The MLE's of α and β converge to their pseudo-true values α_0 and β_0. For instance, in Table 3, $\alpha=(\alpha_1, \alpha_2)$ converges to $\hat{\alpha}=(-0.4723, 1.104)$ and β converges to $\hat{\beta}=0.9994$. The bootstrap estimator $\hat{\omega}_*$ approaches 0.5026. Both chi-square statistics $Q_n(\hat{\alpha})$ and $Q_n(\hat{\beta})$ increase approximately at the rate of n. The same comments apply to Table 4 and Table 5.

In Table 3, the data generating process is chosen such that both the log-normal model and the exponential model are approximately equally close to it. The test statistic GCM_n is expected to have a limiting standard normal distribution N(0, 1). This is roughly confirmed in Table 3. For example, for $n=1{,}000$, GCM_n has a mean of 0.9842 and a standard error of 0.9449. The fractiles reported in Table 3 show that the finite sample distribution of GCM_n is slightly skewed to the right. The three procedures for model selection perform very well. All three of them conclude that both models fit equally well the data with a probability of around 95%, which is 1 minus the nominal size of the test.

With a few exceptions, Tables 4 and 5 reproduce the qualitative results of Tables 1 and 3, respectively, although in a weaker form. When the log-normal model is closer to the true data generating process (Table 4), Method 1 slightly dominates the other two methods. On the other hand, when the exponential is closer to the true data generating process (Table 5), Method 3 seems to dominate especially at small sample sizes. As noted earlier, selecting the exponential model appears more difficult than selecting the log-normal model.

From our limited Monte Carlo study, it is difficult to say which method absolutely dominates the other methods. Moreover, all methods require about the same amount of computation. Although no clear cut conclusion can be made, our study has shown that the three methods work relatively well.

6 Conclusions and Extensions

In this paper, we have studied the problem of selecting estimated models using chi-square type statistics. In particular, we have proposed some convenient asymptotically standard normal tests based on chi-square type statistics that use estimators in a quite general class. The tests are designed to determine whether the estimated competing models are as close to the true distribution against the alternative hypothesis that one estimated model is closer, where closeness is measured according to the discrepancy implicit in the chi-square type statistic used.

To facilitate the implementation of our proposed tests, we have used a bootstrap estimate of the asymptotic variance of the numerator of our test statistic. We have also considered two testing procedures that are directly based on the bootstrap method. The three procedures are fairly simple, and mainly require the computation of estimators and chi-square statistics. Several Monte Carlo experiments were conducted and showed that the three procedures perform relatively well. It was also found that they were comparable, and that none of them absolutely dominates the others.

Our work can be extended in several directions. One direction is to consider econometric models explicitly. For econometric models, only the conditional distribution of the endogenous variables y given the exogenous variables z is specified to belong to a conditional parametric probability model $f(y \mid z; \theta)$, while the marginal distribution of the exogenous variables is left unspecified. Without knowing this marginal distribution, one cannot associate a given parameter value θ with a joint distribution for the observed data (y_i, z_i). Hence when the full sample space $X = Y \times Z$ is partitioned into mutually disjoint cells, the predicted probability in each cell cannot be calculated. This expected probability can, however, be consistently estimated by substituting the empirical marginal distribution for the true marginal distribution of z. Specifically, following ANDREWS [1988*a*], we can consider

$$p_i(\theta) = \frac{1}{n} \sum_{j=1}^{n} \int_Y 1_{E_i}(y, z_j) f(y \mid z_j : \theta)\, dv(y) \qquad \text{for} \quad i = 1, 2, \ldots, M,$$

where $v(y)$ is some σ-finite measure on Y. Given these "expected" cell frequencies, chi-square type statistics can be constructed given any estimator in the general class considered in this paper. Then, as in section 3, the resulting test statistic for model selection can be shown to be $\sqrt{n}$-asymptotically normal. Derivation of the asymptotic variance of the statistic is; however, much more tedious. Fortunately, Methods 1, 2 and 3 discussed in section 4 still apply.

A second extension is to use random cells instead of fixed cells. *See, e.g.*, ANDREWS [1988 *b*] for various interesting examples. Moreover, with appropriate random cells, the asymptotic distribution of the goodness-of-fit statistics may become independent of the true parameter θ_0 under correct specification of the parametric model. *See, e.g.*, ROY [1956] and WATSON [1959]. Recent work on goodness-of-fit statistics under correct specification has, however, shown that the asymptotic distribution of such statistics will not change when cell boundaries are random provided they converge in probability to some fixed values. *See* CHIBISOV [1971], MOORE and SPRUILL [1975] and ANDREWS [1988 *a*]. In view of this latter result, it is expected that our test statistics will remain asymptotically normally distributed with the same asymptotic variance ω^2 under similar conditions.

APPENDIX 1

Proof of Theorem 1 and derivation of ω^2: We first state without proofs two easy lemmas that can be found in VUONG and WANG [1993].

LEMMA 2: Given A3-A4, the weighting matrix $\mathbf{M}_n \equiv \mathbf{M}(f, \theta_n)$ satisfies

$$\mathbf{M}_n = \mathbf{M}_0 + \frac{1}{\sqrt{n}}\mathbf{L}_n + o_p(1/\sqrt{n}),$$

$$\mathbf{L}_n = \sum_{j=1}^{\mathbf{M}} \frac{\partial \mathbf{M}_0}{\partial h_j}\sqrt{n}(f_j - h_j) + \sum_{j=1}^{k} \frac{\partial \mathbf{M}_0}{\partial \theta_j}\sqrt{n}(\theta_{nj} - \theta_{0j}),$$

where $\mathbf{M}_0 \equiv \mathbf{M}(h, \theta_0)$, $\partial \mathbf{M}_0/\partial h_j$ and $\partial \mathbf{M}_0/\partial \theta_j$ are evaluated at (h, θ_0).

LEMMA 3: Under A1-A4,

$$\frac{1}{\sqrt{n}}\mathbf{Q}_n(\theta_n) = \sqrt{n}\, b'\mathbf{M}_0 b + b'\mathbf{L}_{nb} + 2b'\mathbf{M}_0\mathbf{D}_1\mathbf{U}_n - b'\mathbf{M}_0\mathbf{D}_2\mathbf{B}\sqrt{n}(\theta_n - \theta_0) + o_p(1).$$

where $b = \left(\ldots \frac{h_i - p_{0i}}{\sqrt{p_{0i}}}, \ldots \right)'$, $p_0 = p(\theta_0)$ for some $\theta_0 \in \Theta$,

$$\mathbf{D}_1 = \operatorname{diag}\left(\ldots, \frac{\sqrt{h_i}}{\sqrt{p_{0i}}}, \ldots \right),$$

$$\mathbf{D}_2 = \operatorname{diag}\left(\ldots, \frac{h_i + p_{0i}}{p_{0i}} \frac{\sqrt{h_i}}{\sqrt{p_{0i}}}, \ldots \right) = \mathbf{D}_1(\mathbf{D}_1^2 + \mathbf{I}_{\mathbf{M}}),$$

$$\mathbf{B} = \operatorname{diag}\left(\ldots, \frac{1}{\sqrt{h_i}}, \ldots \right) \frac{\partial p_0}{\partial \theta'}, \text{ and } \partial p_0/\partial \theta' \text{ is evaluated at } \theta_0,$$

$$\mathbf{U}_n = \sqrt{n}\left(\ldots, \frac{f_i - h_i}{\sqrt{h_i}}, \ldots \right)' = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [e(\mathbf{X}_i) - q_{\mathbf{H}}],$$

$$e(\mathbf{X}_i) = [1_{\mathbf{E}_1}(\mathbf{X}_i)/\sqrt{h_1}, \ldots, 1_{\mathbf{E}_{\mathbf{M}}}(\mathbf{X}_i)/\sqrt{h_{\mathbf{M}}}]',$$

$$q_{\mathbf{H}} = (\sqrt{h_1}, \sqrt{h_2}, \ldots, \sqrt{h_{\mathbf{M}}})'.$$

Both Lemma 2 and Lemma 3 are simply Taylor expansions of $\mathbf{M}_n$ at (h, θ_0) and $\mathbf{Q}_n(\theta_n)/\sqrt{n}$ at θ_0, respectively. Lemma 3 is also a more detailed expansion of Theorem 5.3 of MOORE [1984]. However, Moore inadvertently ignored the term $b'\mathbf{L}_n b$, and the condition $\mathbf{M}_n \to \mathbf{M}_*$ in probability under H seems too weak for his stated result.

To obtain the asymptotic variance of $[Q_n(\theta_n) - Q_n(\gamma_n)]/\sqrt{n}$, we define the following $1 \times M$ row vectors

$$C'_{1f} = b'_{1f} + 2b'_f M_{f0} D_{1f}, \qquad C'_{2f} = (b'_{2f} - b'_f M_{f0} D_{2f} B_f) R_0^{-1},$$

$$b'_{1f} = \left(\ldots, b'_f \frac{\partial M_{f0}}{\partial h_i} b_f, \ldots \right), \qquad b'_{2f} = \left(\ldots, b'_f \frac{\partial M_{f0}}{\partial \theta_i} b_f, \ldots \right),$$

and the $M \times M$ matrix

$$W_f = E_H [e(X_i) - q_H] \psi(X_i; \theta_0)' = E_H e(X_i) \psi(X_i; \theta_0)'.$$

The subscript f indicates that matrices are now attached to the model F_θ. Similar vectors and matrices are defined for the model G_γ with the subscript f replaced by g. Using Lemma 2 and Lemma 3, we can easily obtain

LEMMA 4: Given A1-A4,
(i) for model F_θ,

$$\frac{1}{\sqrt{n}} Q_n(\theta_n) = \sqrt{n}\, Q(\theta_0) + C'_{1f} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [e(X_i) - q_H] + C'_{2f} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_f(X_i; \theta_0) + o_p(1),$$

(ii) for model G_γ, a similar equality holds with θ and f replaced by γ and g, respectively.

From this lemma it follows that

$$\frac{1}{\sqrt{n}} [Q_n(\theta_n) - Q_n(\gamma_n)] = \sqrt{n} [Q(\theta_0) - Q(\gamma_0)] + (C'_{1f} - C'_{1g}, C'_{2f} - C'_{2g}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{bmatrix} e(X_i) - q_H \\ \psi_f(X_i; \theta_0) \\ \psi_g(X_i; \gamma_0) \end{bmatrix}$$

From the multivariate central limit theorem and assumption A4, we can now immediately obtain the asymptotic distribution of

$$[Q_n(\theta_n) - Q_n(\gamma_n)]/\sqrt{n}$$

under the null hypothesis of equivalence $\tilde{H}_0^e$. Define

$$C'_{fg} = (C'_{1f} - C'_{1g}, C'_{2f} - C'_{2g}), \qquad W_{fg} = E_H \psi_f(X_i; \theta_o) \psi_g(X_i; \gamma_o)',$$

$$\Sigma_u = I_M - q_H q'_H,$$

$$W_{12} = \begin{bmatrix} \Sigma_u & W_f & W_g \\ W'_f & V_{of} & W_{fg} \\ W'_g & W'_{fg} & V_{og} \end{bmatrix}$$

Let $\omega^2 = C'_{fg} W_{12} C_{fg}$, we then have $\frac{1}{\sqrt{n}} \cdot \frac{Q_n(\theta_n) - Q_n(\gamma_n)}{\hat{\omega}} \xrightarrow{D} N(0, 1)$

APPENDIX 2

TABLE 1

Data Generating Process = Log-Normal (− 0.3466, 0.8326).

		100	250	500	1,000
$\hat{\alpha}_1$		−.3361 (.0817)	−.3461 (.0543)	−.3471 (.0380)	−.3495 (.0236)
$\hat{\alpha}_2$		.8251 (.0559)	.8303 (.0393)	.8338 (.0251)	.8341 (.0203)
$\hat{\beta}$		1.0093 (.1016)	.9995 (.0655)	1.0019 (.0437)	.9984 (.0309)
$\hat{\omega}_*$		.3495 (.2018)	.3349 (.1050)	.3149 (.0774)	.3076 (.050)
$Q_n(\hat{\alpha})$		1.730 (1.880)	1.740 (1.430)	1.740 (1.660)	1.790 (1.750)
$Q_n(\hat{\beta})$		11.79 (3.000)	27.32 (4.510)	53.29 (6.750)	105.15 (9.060)
GCM_n		−3.707 (1.864)	−5.420 (2.073)	−7.871 (2.416)	−11.028 (2.708)
Model selection based on GCM_n	Incorrect	0.0%	0.0%	0.0%	0.0%
	Indecisive	18.0%	3.0%	0.0%	0.0%
	Correct	82.0%	97.0%	100.0%	100.0%
Non-bias-corrected percentile method	Incorrect	0.00%	0.0%	0.0%	0.0%
	Indecisive	31.0%	6.5%	0.5%	0.0%
	Correct	69.0%	93.5%	99.5%	100.0%
Bias-corrected percentile method	Incorrect	0.00%	0.0%	0.0%	0.0%
	Indecisive	42.5%	8.5%	0.0%	0.0%
	Correct	57.5%	91.5%	100.0%	100.0%

TABLE 2

Data Generating Process = Exponential (1.0)

		100	250	500	1,000
$\hat{\alpha}_1$		−.5830 (.1381)	−.5769 (.0753)	−.5790 (.0591)	−.5750 (.0414)
$\hat{\alpha}_2$		1.267 (.1270)	1.270 (.0854)	1.285 (.0588)	1.278 (.0406)
$\hat{\beta}$		.9955 (.0976)	.9966 (.0600)	1.0026 (.0440)	1.0010 (.0321)
$\hat{\omega}_*$		.5614 (.1176)	.5240 (.0686)	.5140 (.0408)	.5070 (.0275)
$Q_n(\hat{\alpha})$		7.150 (4.910)	15.94 (7.960)	30.45 (11.24)	60.26 (16.15)
$Q_n(\hat{\beta})$		2.040 (1.740)	2.190 (2.040)	2.150 (1.990)	2.240 (2.010)
GCM_n		.9171 (.8843)	1.6652 (.9863)	2.4648 (.9820)	3.622 (1.003)
Model selection based on GCM_n	Incorrect	0.0%	0.0%	0.0%	0.0%
	Indecisive	85.5%	58.5%	30.5%	5.0%
	Correct	14.5%	41.5%	69.5%	95.0%
Non-bias-corrected percentile method	Incorrect	0.00%	0.0%	0.0%	0.0%
	Indecisive	89.0%	58.0%	31.5%	6.5%
	Correct	11.0%	42.0%	68.5%	93.5%
Bias-corrected percentile method	Incorrect	0.50%	0.0%	0.0%	0.0%
	Indecisive	68.0%	50.0%	24.5%	5.0%
	Correct	31.5%	50.0%	75.5%	95.0%

TABLE 3

Data Gen. Process = .5357*Exp (1.0) + .4643*log-normal (− 0.347, 0.833)

		100	250	500	1,000
$\hat{\alpha}_1$		−.4728 (.1107)	−.4743 (.0710)	−.4720 (.0486)	−.4723 (.0328)
$\hat{\alpha}_2$		1.097 (.1125)	1.102 (.0778)	1.106 (.0597)	1.104 (.0369)
$\hat{\beta}$		.9994 (.1027)	.9954 (.0640)	1.0003 (.0446)	.9994 (.0301)
$\hat{\omega}_*$		.5016 (.0760)	.5040 (.0512)	.4986 (.0469)	.5026 (.0251)
$Q_n(\hat{\alpha})$		4.650 (3.650)	8.270 (5.460)	15.18 (7.370)	26.04 (9.780)
$Q_n(\hat{\beta})$		3.770 (2.770)	7.380 (4.180)	12.55 (5.530)	24.36 (7.340)
GCM_n		.1109 (1.044)	.0737 (1.039)	.2100 (1.020)	.0942 (.9449)
GCM_n	2.5% fractile		−2.895	−2.114	−1.932	−1.811
	5.0% fractile		−1.674	−1.770	−1.643	−1.446
	95.0% fractile		1.6595	1.4666	1.7269	1.5126
	97.5% fractile		2.0570	1.9239	2.0849	2.1395
Model selection based on GCM_n	Favor log-*n*		3.0%	2.5%	1.0%	1.0%
	Equivalent		94.0%	95.0%	94.0%	95.5%
	Favor exp		3.0%	2.5%	5.00%	3.5%
Non-bias-corrected percentile method	Favor log-*n*		3.00%	2.00%	3.50%	1.50%
	Equivalent		96.0%	94.5%	93.0%	95.0%
	Favor exp		1.00%	3.50%	3.50%	3.50%
Bias-corrected percentile method	Favor log-*n*		2.50%	0.50%	1.50%	2.00%
	Equivalent		95.5%	95.0%	95.0%	94.5%
	Favor exp		2.00%	4.50%	3.50%	3.50%

TABLE 4

Data Gen. Process = 0.25*Exp (1.0) + 0.75*log-normal (− 0.347, 0.833)

		100	250	500	1,000
$\hat{\alpha}_1$		−.4016 (.0937)	−.4126 (.0586)	−.4078 (.0434)	−.4033 (.0294)
$\hat{\alpha}_2$		.9649 (.1105)	.9703 (.0602)	.9725 (.0486)	.9699 (.0345)
$\hat{\beta}$		1.002 (.0980)	.9932 (.0638)	.9986 (.0427)	1.0005 (.289)
$\hat{\omega}_*$		.4931 (.1403)	.4698 (.0541)	.4638 (.0372)	.4620 (.0310)
$Q_n(\hat{\alpha})$		3.130 (2.590)	3.960 (2.740)	6.030 (4.130)	9.970 (5.270)
$Q_n(\hat{\beta})$		7.320 (3.320)	15.63 (5.140)	30.14 (6.420)	61.40 (9.990)
GCM_n		−1.126 (1.415)	−1.671 (1.162)	−2.393 (1.053)	−3.589 (1.171)
Model selection based on GCM_n	Favor log-*n*		22.0%	33.0%	65.0%	94.0%
	Equivalent		77.5%	67.0%	35.0%	6.0%
	Favor exp		0.5%	0.0%	0.0%	0.0%
Non-bias-corrected percentile method	Favor log-*n*		20.0%	40.5%	64.0%	89.0%
	Equivalent		70.0%	59.5%	36.0%	11.0%
	Favor exp		0.00%	0.00%	0.00%	0.00%
Bias-corrected percentile method	Favor log-*n*		20.0%	28.0%	54.0%	87.5%
	Equivalent		80.0%	72.0%	46.0%	12.5%
	Favor exp		0.00%	0.00%	0.00%	0.00%

TABLE 5

Data Gen. Process = 0.75*Exp (1.0) + 0,25* log-normal (− 0.347, 0.833)

	100	250	500	1,000
$\hat{\alpha}_1$	− .5220 (.1158)	− .5171 (.0715)	− .5323 (.0585)	− .5215 (.0375)
$\hat{\alpha}_2$	1.171 (.1188)	1.181 (.0823)	1.194 (.0607)	1.193 (.0399)
$\hat{\beta}$	.9917 (.1015)	.9984 (.0618)	.9894 (.0466)	.9982 (.0309)
$\hat{\omega}_*$	.5227 (.1184)	.5108 (.0548)	.4969 (.0351)	.5023 (.0235)
$Q_n(\hat{\alpha})$	5.740 (4.240)	11.37 (6.690)	22.53 (8.770)	41.73 (12.60)
$Q_n(\hat{\beta})$	2.620 (2.410)	3.920 (2.980)	4.910 (3.550)	8.510 (4.860)
GCM_n	.5482 (1.003)	.9013 (.9879)	1.5761 (.9453)	2.0816 (.9309)

Model selection based on GCM_n	Favor log-*n*	0.20%	0.05%	0.00%	0.00%
	Equivalent	93.5%	88.0%	62.5%	43.0%
	Favor exp	4.50%	11.5%	37.5%	57.0%
Non-bias-corrected percentile method	Favor log-*n*	2.00%	0.50%	0.00%	0.00%
	Equivalent	93.5%	81.0%	75.0%	39.5%
	Favor exp	4.50%	18.5%	25.0%	60.5%
Bias-corrected percentile method	Favor log-*n*	1.00%	0.00%	0.00%	0.00%
	Equivalent	89.0%	80.0%	66.0%	37.0%
	Favor exp	10.0%	20.0%	34.0%	63.0%

• References

AKAIKE, H. (1973). – "Information Theory and an Extension of the Likelihood Ratio Principle", *Proceedings of the Second International Symposium of Information Theory*, ed. by PETROV, B. N. and CSAKI, F. Budapest: Akademiai Kiado, pp. 257-281.

AMEMIYA, T. (1985). – *Advanced Econometrics*. Cambridge: Harvard University Press.

ANDREWS, D. W. K. (1988*a*). – "Chi-Square Diagnostic Tests for Econometric Models: Theory", *Econometrica*, 56, pp. 1419-1453.

ANDREWS, D. W. K. (1988*b*). – "Chi-Square Diagnostic Tests for Econommetric Models: Introduction and Applications", *Journal of Econometrics*, 37, pp. 135-156.

ATKINSON, A. C. (1970). – "A Method for Discriminating Between Models", *Journal of the Royal Statistical Society*, Series B, 32, pp. 323-353.

BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (1975). – *Discrete Multivariate Analysis*. Cambridge: MIT Press.

CHERNOFF, H. and LEHMANN, E. L. (1954). – "The Use of Maximum of Likelihood Estimates in χ^2 Test for Goodness of Fit", *Ann. Math. Statist.*, 25, pp. 579-586.

CHIBISOV, D. M. (1971). – "Certain Chi-Square Type Tests for Continuous Distributions", *Theory of Probability and Its Applications*, 16, pp. 1-22.

COCHRAN, W. G. (1952). – "The χ^2 Test of Goodness of Fit", *Ann. Math. Statist.*, 23, pp. 315-345.

COX, D. R: (1962). – "Further Results on Tests of Separate Families of Hypotheses", *Journal of the Royal Statistical Society*, Series B, 24, pp. 406-424.

EFRON, B. (1982). – "The Jackknife, the Bootstrap and Other Resampling Plans", *CBMS-NSF Regional Conference Series in Applied Mathematics*, 38.

EFRON, B. (1984). – "Comparing Non-Nested Linear Models", *Journal of the American Statistical Association*, 79, pp. 791-803.

HECKMAN, J. J. (1981). – "Heterogeneity and State Dependence", in *Studies in Labor Markets*, University of Chicago Press, Chicago, ed. by ROSEN S..

LINHART, H. and ZUCCHINI, W. (1986). – *Model Selection*, John Wiley and Sons.

MASSY, W. F., MONTGOMERY, D. B. and MORRISON, D. G. (1970). – "Stochastic Models of Buying Behavior", Cambridge, Massachusetts: MIT Press.

MOORE, D. S. (1977). – "Generalized Inverses, Wald's Method and the Construction of Chi-Squared Tests of Fit", *Journal of the American Statistical Association*, 7, pp. 131-137.

MOORE, D. S. (1978). – "Chi-Square Tests", in *Studies in Statistics*, ed. by HOGG, R. V. Volume 19, The Mathematical Association of America.

MOORE, D. S. (1984). – "Measures of Lack of Fit from Tests of Chi-Squared Type", *Journal of Statistical Planning and Inference*, 10, pp. 151-166.

MOORE, D. S. (1986). – "Tests of Chi-squared Type", *Goodness-of-Fit techniques*, ed. by D'AGOSTINO, R. B. and STEPHENS, M. A.

MOORE, D. S. and SPRUILL, M. C. (1975). – "Unified Large-Samples Theory of General Chi-Squared Statistics for Tests of Fit", *Ann. Statist.*, 3, pp. 599-616.

NAKAMURA, A. and NAKAMURA, M. (1985). – *The Second Paycheck: A Socioeconomic Analysis of Earnings*, Orlando, Florida: Academic Press.

PEARSON, K. (1900). – "On the Criterion that a Given System of Deviation from the Probable in the Case of a Correlated System of Variables is such that It Can be Reasonably Supposed to have Arisen from Random Sampling", *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 50, pp. 157-175.

RAO, K. C. and ROBSON, D. S. (1974). – "A Chi-Square Statistic for Goodness-of-Fit Tests within the Exponential Family", *Communications in Statistics*, 3, pp. 1139-1153.

RIVERS, D. and VUONG, Q. H. – "Model Selection Tests for Nonlinear Dynamic Models", WP 9108, INRA-ESR, Toulouse.

ROY, A. R. (1956). – "On χ^2 Statistics with Variable Intervals", Technical Report No. 1, Stanford University, Department of Statistics.

VUONG, Q. H. (1989). – "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses", *Econometrica*, 57, pp. 257-306.

VUONG, Q. H., and WANG, W. (1993). – "Minimum Chi-Square Estimation and Tests for Model Selection", *Journal of Econometrics*, 56, pp. 141-168.

WATSON, G. S. (1959). – "Some Recent Results in Chi-Square Goodness-of-Fit Tests", *Biometrics*, 15, pp. 440-468.

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