Simulation Based Selection of Competing Structural Econometric Models

Tong Li*

November 2005

Abstract

This paper proposes a formal model selection test for choosing between two competing structural econometric models. The procedure is based on a novel lack-of-fit criterion, namely, the simulated mean squared error of predictions (SMSEP), taking into account the complexity of structural econometric models. It is asymptotically valid for any fixed number of simulations, and allows for any estimator which has a $\sqrt{n}$ asymptotic normality or is superconsistent with a rate at n. The test is bi-directional and applicable to non-nested models which are both possibly misspecified. The asymptotic distribution of the test statistic is derived. The proposed test is general regardless of whether the optimization criteria for estimation of competing models are the same as the SMSEP criterion used for model selection. An empirical application using timber auction data from Oregon is used to illustrate the usefulness and generality of the proposed testing procedure.

JEL Classification: C12, C15, C52

Keywords: Lack-of-fit, Model selection tests, Non-nested models, Simulated mean squared error of predictions.

*Department of Economics, Vanderbilt University, VU Station B #351819, Nashville, TN 37235-1819. Phone: (615) 322-3582, Fax: (615) 343-8495, Email: tong.li@vanderbilt.edu.

1 Introduction

Model selection is an important component of statistical inference. It involves comparing competing models based on some appropriately defined goodness-of-fit or selection criterion. For the competing models that can be estimated by (conditional) maximum likelihood estimation (MLE), there has been a vast literature on model selection procedures, such as the Akaike (1973, 1974) information criterion (AIC), the Cox test (1961) and the Vuong (1989) likelihood ratio test, to name only a few. Another important development is the use of the encompassing principle in testing non-nested models assuming that one of them is correctly specified. See, e.g., Mizon and Richard (1986), and Wooldridge (1990), among others. For a comprehensive review of the literature, see Gourieroux and Monfort (1994) and Pesaran and Weeks (2001). In light of the development of new estimation methods in econometrics such as the generalized method of mements (GMM) and empirical likelihood estimation methods, which offer robust alternatives to the conventional MLE, recent work in model selection has attempted to develop procedures that can be used for models estimated by other methods than the MLE. For example, see Smith (1992) for extensions of the Cox test and the encompassing test to non-nested regression models that are both estimated by instrumental variables, Rivers and Vuong (2002) for the extension of Vuong's (1989) test to dynamic regression models, Kitamura (2002) for using empirical likelihood ratio-type statistics for testing non-nested conditional models, and Chen, Hong and Shum (2003) for likelihood ratio tests between parametric and (unconditional) moment condition models.

These model selection tests have been found useful in some of structural microeconometric models, which have been developed in the last two decades and applied in such fields of modern economics as labor and industrial organization.[1] For example Vuong's (1989) likelihood ratio test has been used to select structural models both of which are

[1]Heckman (2001) gives an insightful discussion on the development and the issues on identification and inference of structural microeconometric models.

estimated by MLE. See, e.g., Gasmi, Laffont and Vuong (1992) for testing collusive behavior, Wolak (1994) for testing asymmetric information, and Li (2005) for testing binding reservation prices in first-price auctions, to name only a few. Also, Chen, Hong and Shum (2003) develop a test to distinguish between a parametric model which can be estimated by the MLE and an unconditional moment model which can be estimated by the empirical likelihood method, and then apply their procedure to choose between a sequential search model and a non-sequential model. Despite these interesting applications of the aforementioned model selection tests, there are many other situations in which these model selection tests may not be applicable.[2] Such a gap can be mainly attributed to the complexity associated with the nature of structural econometric models. Model selection criteria are formulated in such ways that they are calculated using sample information and compared between competing models. Most of the structural econometric models, however, are constructed based on economic theory which defines maps between the latent variable of interest or/and its distribution and the observables. For instance, in structural auction models, it is assumed that the observed bids are Nash-Bayesian equilibrium strategies which are strictly increasing functions of bidders' private valuations whereas identifying and estimating the private values distribution is one of the main objectives of the structural approach. The presence of latent variables and the complex relationship between the latent and observed variables defined by structural models make the formulation of a well-defined model selection criterion more involved. Moreover, in many cases, structural econometric models are constructed through moment conditions, meaning that

[2] For instance, Laffont, Ossard and Vuong (1995) develop a simulated nonlinear least squares estimator to estimate a structural model of first-price auctions. They encounter a problem of determining between 11 and 18 potential bidders. This problem, significant from an economic viewpoint as having 11 bidders could imply the existence of a large trader and hence asymmetric bidding, calls for a formal test of non-nested models, as the structural models with different numbers of potential bidders are non-nested. While this issue was not further pursued in Laffont, Ossard and Vuong (1995) (see footnote 21 in Laffont, Ossard and Vuong (1995)), and cannot be addressed using the existing model selection methods, it can be resolved using our proposed procedure, as illustrated in the empirical application.

they are estimated not by MLE but by GMM or method of simulated moments (MSM). Therefore, to accommodate these specific features arising from the nature of structural models and the estimation methods, new model selection tests need to be developed.

Developing model selection procedures suitable in distinguishing between competing structural models is especially relevant in using the structural approach to analyze economic data and make policy evaluations. In the structural approach, policy analysis and the resulting recommendations are based on a structural model that is closely derived from economic theory assuming that the involved economic agents are in the environment described by the theory and behave according to the theory. As a result, it is pivotal to validate the structural model under consideration. For example, when analyzing auction data using the structural approach, an econometrician faces choices among different paradigms such as a private value model or a common value model. Even within a chosen paradigm, the econometrician may also need to determine an appropriate parametric functional form for the latent distribution. Furthermore, the researcher sometimes needs to choose between different equilibria if multi-equilibria exist, as is the case for models of two-stage dynamic games which yield a large number of Bayesian perfect equilibria (Laffont and Maskin (1990)).

The goal of this paper is thus to propose a new model selection test in discriminating between competing structural econometric models. Our test is based on a comparison of the predictability of competing structural models. In time series literature, there has been a rich set of papers since Diebold and Mariano (1995) and West (1996) in using predictability for model evaluation. While the proposed test in this paper is related to this literature as it uses predictability as a model selection criterion, it differs significantly in various aspects. First, we formulate the null and alternative hypotheses in terms of comparing (asymptotic) lack-of-fit of competing structural models based on a well-defined population predictability or lack-of-fit criterion that is appropriate for distinguishing between competing structural models. Second, given that structural econometric models usually contain some latent variables that are unobserved, we propose to simulate these latent variables in order to make the predictions on the equilibrium outcomes. Also, since the simulation is used, when formulating the sample analog to those population quantities, we need to correct for the asymptotic bias term caused by the simulation, and hence propose a simulated MSEP (SMSEP) as a consistent sample analog to the population predictability criterion. As a result, while those using simulation based prediction for model evaluations in time series framework usually require that the number of simulations tend to infinity, ours works for any fixed number of simulations. Third, our model selection test allows for any estimators that are $\sqrt{n}$ asymptotically normally distributed, or are superconsistent with the rate n that can arise from some structural microeconometric models such as auction models and job search models (Donald and Paarsch (1993, 1996, 2002), Hong (1998), Chernozhukov and Hong (2004), Hirano and Porter (2003)). Lastly, in a similar spirit to that of Vuong (1989) and Rivers and Vuong (2002), the test is bi-directional and applicable to non-nested structural models which are both possibly misspecified. This adds a considerable advantage to the proposed test because in real applications, structural econometric models can be best considered an approximation but not exact modeling of the true data generating process. Nevertheless, with two possibly misspecified models, our model selection procedure enables one to tell which one is closer to the truth.

While some empirical work has used predictions from structural models to validate a particular choice of the model, because of the lack of a formal test, it has been based on an ad-hoc comparison of the closeness between the predictions and the observed outcomes. The statistical significance of such a closeness is not assessed. In contrast, our testing procedure provides a formal framework in which the statistical significance of the difference in predictability of competing structural models can be assessed. The asymptotic distribution of the test statistic is derived. The proposed test is general regardless of whether the optimization criteria for estimation of competing models are the same as the SMSEP criterion used for model selection. An easy-to-implement bootstrap based test is proposed for practical implementation when at least one of the estimators is obtained without minimizing the selection criterion, in which case the direct estimation of the variance of the asymptotic distribution of the test statistic may be computationally demanding. An empirical application using timber auction data from Oregon is used to illustrate the usefulness and generality of the proposed testing procedure.

It is worth noting that most of the recent work in model selection tests has been based on comparing the Kullback-Leibler Information Criterion (KLIC) between two competing models. See, e.g., Kitamura (2000, 2002), and Chen, Hong and Shum (2003). Our approach is different, as it is based on the simulated mean squared errors of predictions, a lack-of-fit criterion. This is motivated by the fact that many structural econometric models are estimated by GMM or MSM other than the MLE, thus the KLIC cannot be used as a model selection criterion.[3] Our model selection criterion, on the other hand, can be used for any estimation methods that yield estimators with root-n asymptotic normality, or with rate n superconsistency, and hence has an appealing generality.

This paper is organized as follows. Section 2 describes the general model selection framework for structural econometric models using the SMSEP criterion. The hypotheses for model selection are formulated. The asymptotic properties of the proposed test statistic are established. The practical issues arising from the implementation of the test are also discussed. Section 3 is devoted to an empirical application of the proposed test to structural auction models. Section 4 concludes.

2 An SMSEP Criterion and the Resulting Model Selection Test

Two models $\mathcal{M}_1$ and $\mathcal{M}_2$ are estimated using data $\{y_i, \mathbf{x}_i\}$, $i = 1, \ldots, n$, where y is a dependent variable and $\mathbf{x}$ is a $1 \times K$ vector of covariates. Both $\mathcal{M}_j$, $j = 1, 2$ are structural models in the sense that for model $\mathcal{M}_j$, there is a p-dimensional vector of latent variables $\mathbf{v}_j \in V_j \subset I\!R^p$ with the (conditional) probability density function (pdf)

[3] On the other hand, if the structural models considered here are estimated using empirical likelihood or other KLIC based methods, then one can apply the recent model selection tests such as Kitamura (2002).

$f_j(\cdot|\mathbf{x}, \theta_j)$ and the (conditional) cumulative distribution function (cdf) $F_j(\cdot|\mathbf{x}, \theta_j)$, where θ_j is in Θ_j, a compact subset of $I\!R^{K_j}$, such that the observed dependent variable y and the latent variables $\mathbf{v}_j$ have a relationship as a result of the structural model given by $y = H_j(\mathbf{v}_j, f_j(\mathbf{v}_j|\mathbf{x}, \theta_j)) \equiv H_j(\mathbf{v}_j, \mathbf{x}, \theta_j)$.[4] As a result, the function $H_j(\cdot, \mathbf{x}, \theta_j)$ maps $\mathbf{v}_j$ to the equilibrium outcome y under model $\mathcal{M}_j$. For instance, in structural auction models where bidders are assumed to bid optimally according to the Nash-Bayesian equilibrium strategies, the observed bids can be considered as an increasing function of bidders' private valuations. See, e.g., Laffont (1997) for a review on empirical auction models. Note that in addition to θ_j, the parameters that appear in the (conditional) pdf of the latent variables, it is also possible to include in model $\mathcal{M}_j$ some parameters that are not associated with the latent variable density provided that they can be identified and estimated as well. An example of this case is bidders' risk aversion parameter in auction models. Our model selection procedure can be readily adopted to this case, in which we can have $y = H_j(\mathbf{v}_j, f_j(\mathbf{v}_j|\mathbf{x}, \theta_j), \gamma_j) \equiv H_j(\mathbf{v}_j, \mathbf{x}, \theta_j, \gamma_j)$, where γ_j is the parameter vector that is not associated with the latent variable denisty. Thus, for ease of exposition, we will focus on the case where each model $\mathcal{M}_j$ contains only θ_j. We have the following random sampling assumption.

Assumption 1. $\{y_i, \mathbf{x}_i\}$, $i = 1, \ldots, n$, are independently and identically distributed with finite first and second population moments.

Note that we make the random sampling assumption for the sake of exposition. Our proposed selection procedure can be readily extended to (weakly) dependent data, whose data generating process satisfies the mixing conditions, such as those given in Gallant and White (1988).

Let $\hat{\theta}_j$ be an estimator of θ_j using the observations $\{y_i, \mathbf{x}_i\}$. The estimator $\hat{\theta}_j$ can be obtained from any estimation method with $\sqrt{n}$ asymptotic normality. Specifically, we

[4] It is clear from the set-up here that both $\mathcal{M}_j, j = 1, 2$, are allowed to be conditional structural models with $\mathbf{x}$ being the variables that are used for controlling for heterogeneity, as is accounted for by most of the structural models in microeconometric applications.

have the following assumptions.

Assumption 2. For $j = 1, 2$, there is a unique θ_j^* inside the interior of Θ_j, such that $\hat{\theta}_j$ converges to θ_j^* in probability as $n \to \infty$.

Assumption 3. For $j = 1, 2$, there exist $K_j \times 1$ random vectors $U_{j,i}, i = 1, \ldots, n$, with mean zero and bounded second absolute moments such that

$$\sqrt{n}(\hat{\theta}_j - \theta_j^*) = -\frac{1}{\sqrt{n}} A_j \sum_{i=1}^{n} U_{j,i} + o_P(1) \tag{1}$$

where A_j are bounded nonstochastic symmetric $K_j \times K_j$ matrices.

Assumption 2 assumes the (weak) convergence of $\hat{\theta}_j$ to a unique value θ_j^* inside the interior of Θ_j. Since we allow both $\mathcal{M}_j, j = 1, 2$ to be misspecified, $\theta_j^*, j = 1, 2$, are called pseudo-true values as in Gallant and White (1988). Assumption 3 gives an asymptotic linear representation for $\hat{\theta}_j$ that is satisfied by most of the econometric estimators possessing root-n asymptotic normality (see, e.g., Newey and McFadden (1994)). Later this assumption will be changed to accommodate the possibility that one or both estimators are rate-n superconsistent. We also make the following regularity assumption on the equilibrium outcome functions $H_j(\mathbf{v}_j, \mathbf{x}, \theta_j)$, $j = 1, 2$, which is satisfied by most of the structural models studied in the literature.

Assumption 4. For $j = 1, 2$, $H_j(\cdot, \mathbf{x}, \cdot)$ are continuously differentiable on both $\mathbf{v}_j \in V_j$ and $\theta_j \in \Theta_j$.

To formulate a set of hypotheses that are properly defined in the framework of structural econometric models, we define the quantity

$$Q_j(\theta_j^*) = \mathrm{E}_{y,\mathbf{x}}(y - \mathrm{E}_{\mathcal{M}_j}(y|\mathbf{x}, \theta_j^*))^2 \tag{2}$$

where $\mathrm{E}_{y,\mathbf{x}}$ denotes the expectation taken with respect to the true but unknown joint distribution of y and $\mathbf{x}$, and $\mathrm{E}_{\mathcal{M}_j}$ denotes that the expectation is taken with respect to model $\mathcal{M}_j$, which may be misspecified. Thus $y - \mathrm{E}_{\mathcal{M}_j}(y|\mathbf{x}, \theta_j^*)$ represents the prediction error from the conditional model $\mathcal{M}_j$. $Q_j(\theta^*)$ is well-defined and finite because of Assumption 1.

Note that $Q_j(\theta_j^*)$ can be viewed as the (asymptotic) lack-of-fit from model $\mathcal{M}_j$. Then within the classical hypothesis testing framework as adopted in Vuong (1989), Rivers and Vuong (2002), Kitamura (2000, 2002), and Chen, Hong and Shum (2003), we can specify the following set of null and alternative hypotheses

$$H_0 : Q_1(\theta_1^*) = Q_2(\theta_2^*),$$

meaning that $\mathcal{M}_1$ and $\mathcal{M}_2$ are asymptotically equivalent, against

$$H_1 : Q_1(\theta_1^*) < Q_2(\theta_2^*),$$

meaning that $\mathcal{M}_1$ is asymptotically better than $\mathcal{M}_2$ in the sense that the former has a smaller (asymptotic) lack-of-fit than the latter, or

$$H_2 : Q_1(\theta_1^*) > Q_2(\theta_2^*),$$

meaning that $\mathcal{M}_2$ is asymptotically better than $\mathcal{M}_1$.

From the formulation of the null and alternative hypotheses above, it is clear that our model selection is based on a comparison of the asymptotic lack-of-fit, or predictability of the two (conditional) structural models under consideration. In essence, under the null, both structural models have the same asymptotic predictability, while under H_1, model 1 has a better asymptotic predictability than model 2, and under H_2, model 2 is better than model 1 with respect to the asymptotic predictability.

To test H_0 against H_1 or H_2, we need to estimate $Q_j(\theta_j^*)$ using the observations $\{y_i, \mathbf{x}_i\}$, $i = 1, \ldots, n$. Let $g_j(\cdot|\mathbf{x}, \theta_j)$ denote the pdf for y under model $\mathcal{M}_j$. Then $Q_j(\theta_j^*)$ could be estimated consistently by

$$\begin{aligned} \tilde{Q}_j(\hat{\theta}_j) &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \int H_j(\mathbf{v}_j, \mathbf{x}_j, \hat{\theta}_j) f_j(\mathbf{v}_j|\mathbf{x}_j, \hat{\theta}_j) d\mathbf{v}_j)^2 && (3) \\ &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \int y g_j(y|\mathbf{x}_i, \hat{\theta}_j) dy)^2. && (4) \end{aligned}$$

However, (3) can be difficult to compute because the functional form of $H_j(\cdot, \mathbf{x}_j, \theta_j)$ can be complicated, which leads to the computational burden in evaluating the integral. Similarly, (4) can be computationally intractable because the functional form for $g_j(\cdot|\mathbf{x}_j, \theta_j)$, the pdf for the observed equilibrium outcome y under model $\mathcal{M}_j$, can be hard to obtain as the result of the structural model leading to $y = H_j(\mathbf{v}_j, \mathbf{x}_j, \theta_j)$ and so is the integral in (4). To address this issue, we note that $\mathrm{E}_{\mathcal{M}_j}(y|\mathbf{x}_i, \theta_j^*) = \int y g_j(y|\mathbf{x}_i, \theta_j^*)dy$ can be approximated by $\bar{Y}_{j,i}(\theta_j^*) \equiv \sum_{s_j=1}^{S_j} y_{j,i}^{(s_j)}(\theta_j^*)/S_j$, where $y_{j,i}^{(s_j)}(\theta_j^*) \equiv H_j(\mathbf{v}_{j,i;\theta_j^*}^{(s_j)}, \mathbf{x}_i, \theta_j^*)$ and $\mathbf{v}_{j,i;\theta_j^*}^{(s_j)}$, $s_j = 1, \ldots, S_j$, are independent draws from $f_j(\cdot|\mathbf{x}_i, \theta_j^*)$ provided that θ_j^* is known. This is because $\bar{Y}_{j,i}(\theta_j^*)$ is an unbiased simulator of $\mathrm{E}_{\mathcal{M}_j}(y|\mathbf{x}_i, \theta_j^*)$. Noting that θ_j^* are unknown, but can be consistently estimated by $\hat{\theta}_j$, we could replace the integral in (4) by $\bar{Y}_{j,i}(\hat{\theta}_j) \equiv \sum_{s_j=1}^{S_j} y_{j,i}^{(s_j)}(\hat{\theta}_j)/S_j$ where $y_{j,i}^{(s_j)}(\hat{\theta}_j) = H_j(\mathbf{v}_{j,i;\hat{\theta}_j}^{(s_j)}, \mathbf{x}_j, \hat{\theta}_j)$ and $\mathbf{v}_{j,i;\hat{\theta}_j}^{(s_j)}$, $s_j = 1, \ldots, S_j$, are independent draws from $f_j(\cdot|\mathbf{x}_j, \hat{\theta}_j)$. Because of its nonlinearity, however, the following quantity

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{Y}_{j,i}(\hat{\theta}_j))^2$$

does not converge to $Q_j(\theta^*)$ for any fixed number S_j of simulations as the asymptotic bias caused by the simulations does not vanish. To correct for the asymptotic bias caused by the simulations, we define

$$\hat{Q}_j(\theta_j) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{Y}_{j,i}(\theta_j))^2 - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{S_j(S_j-1)}\sum_{s_j=1}^{S_j}(y_{j,i}^{(s_j)}(\theta_j) - \bar{Y}_{j,i}(\theta_j))^2. \tag{5}$$

Then we have the following result regarding the relationship between $\hat{Q}_j(\hat{\theta}_j)$ and its population counterpart $Q_j(\theta_j^*)$.

Proposition 1. Assume Assumptions 1 and 2. For any fixed S_j, as $n \to \infty$, $\hat{Q}_j(\hat{\theta}_j)$ converges to $Q_j(\theta_j^*)$ in probability.

As justified in Proposition 1, for $j = 1, 2$ and any fixed S_j, $\hat{Q}_j(\hat{\theta}_j)$ consistently estimate $Q_j(\theta_j^*)$. Furthermore, S_1 does not necessarily equal S_2. As a result, we propose to use

$\hat{Q}_j(\hat{\theta}_j)$ in practice to estimate $Q_j(\theta_j^*)$ and hence to test H_0 against H_1 or H_2. $\hat{Q}_j(\hat{\theta}_j)$ is thus the SMSEP we propose for model $\mathcal{M}_j$. It can be viewed as an in-sample SMSEP as it is calculated from the same sample that is used in the estimation. Alternatively, we can consider an out-of-sample SMSEP in the sense that the original data set is split into two parts, one part is used for estimation of the competing models, and the other part is used for calculating $\hat{Q}_j(\hat{\theta}_j)$ and hence for model selection test. Since within the framework considered here, the asymptotic properties of the tests based on in-sample and out-of-sample are the same, we will focus on the in-sample test based on (5) for ease of exposition. It is worth noting that using $\hat{Q}_j(\hat{\theta}_j)$ has the computational advantage as it can be readily obtained from the sample information with the help of simulations. Besides, as given in Proposition 1, it converges to the population lack-of-fit criterion as the sample size approaches infinity for any fixed number of simulations. This feature makes it a basis for constructing our test statistic below. Note that bias corrections similar to (5) were first used in Laffont, Ossard and Vuong (1995) and subsequently in Li and Vuong (1997) in constructing objective functions to be minimized that produce simulated nonlinear least squares estimators which are consistent for a fixed number of simulations in estimating structural auction models. A novelty of this paper is to use (5) for a different purpose, that is to use it as a consistent sample analog to the population lack-of-fit criterion in constructing a general test statistic for choosing between rival structural econometric models, not limited to auction models, as long as the structural models under consideration allow one to generate predictions from simulations.

In order to propose our test statistic, we define $T_n \equiv \sqrt{n}(\hat{Q}_1(\hat{\theta}_1) - \hat{Q}_2(\hat{\theta}_2))$. Then the next theorem establishes asymptotic properties of T_n under our specified hypotheses H_0, H_1 and H_2.

Theorem 1. Assume Assumptions 1-4.

(i) Under H_0, $T_n \Rightarrow N(0, \sigma^2)$, where

$$\sigma^2 = \lim_{n\to\infty}[(1, -B_{1,n}, B_{2,n})V_n(1, -B_{1,n}, B_{2,n})'],$$

$$V_n = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}\begin{pmatrix} C_i \\ U_{1,i} \\ U_{2,i} \end{pmatrix},$$

$$\begin{aligned} C_i &= (y_i - \bar{Y}_{1,i}(\theta_1^*))^2 - \frac{1}{S_1(S_1-1)}\sum_{s_1=1}^{S_1}(y_{1,i}^{(s_1)}(\theta_1^*) - \bar{Y}_{1,i}(\theta_1^*))^2 \\ &- (y_i - \bar{Y}_{2,i}(\theta_2^*))^2 + \frac{1}{S_2(S_2-1)}\sum_{s_2=1}^{S_2}(y_{2,i}^{(s_2)}(\theta_2^*) - \bar{Y}_{2,i}(\theta_2^*))^2, \end{aligned}$$

$$B_{j,n} = A_j \frac{\partial \hat{Q}_j}{\partial \theta_j'}|_{\theta_j^*},$$

and A_j and $U_{j,i}$ are defined in (1) of Assumption 3.

(ii) Under H_1, $T_n \rightarrow_p -\infty$.

(iii) Under H_2, $T_n \rightarrow_p \infty$.

Theorem 1 is valid in a general sense in that while the SMSEP criteria $\hat{Q}_j(\hat{\theta}_j), j = 1, 2$, are used in constructing T_n, the estimation methods that are used to obtain $\hat{\theta}_j, j = 1, 2$, can be any resulting in estimators with $\sqrt{n}$ asymptotic normality. The estimators include those commonly used in practice such as the GMM estimators, the MSM estimators surveyed in Gourieroux and Monfort (1996), as well as some semiparametric estimators surveyed in Powell (1994). Moreover, the criteria that are optimized in estimation can be different from the SMSEP used as our model selection criterion. Such a general feature of our selection procedure leads to the consequence that the asymptotic variances of $\hat{\theta}_j$, $j = 1, 2$ in general contribute to the asymptotic variance σ^2 of T_n, as reflected in the presence of $A_j, U_{j,i}$ like terms in σ^2. On the other hand, in some applications one or both $\hat{\theta}_j, j = 1, 2$ can be obtained by minimizing the same SMSEP criterion defined in (5) which is used as our model selection criterion. In these situations, the expression for σ^2 can simplify, as indicated in the following corollary.

Corollary 1. Assume Assumptions 1-4.

(i) If $\hat{\theta}_1$ is obtained by minimizing (5), then under H_0, $T_n \Rightarrow N(0, \sigma^2)$, where

$$\sigma^2 = \lim_{n\to\infty} [(1, \mathbf{0}, B_{2,n}) V_n (1, \mathbf{0}, B_{2,n})'],$$

where $B_{2,n}$ and V_n are defined in Theorem 1.

(ii) If $\hat{\theta}_2$ is obtained by minimizing (5), then under H_0, $T_n \Rightarrow N(0, \sigma^2)$, where

$$\sigma^2 = \lim_{n\to\infty} [(1, -B_{1,n}, \mathbf{0}) V_n (1, -B_{1,n}, \mathbf{0})'],$$

where $B_{1,n}$ and V_n are defined in Theorem 1.

(iii) If both $\hat{\theta}_1$ and $\hat{\theta}_2$ are obtained by minimizing (5), then under H_0, $T_n \Rightarrow N(0, \sigma^2)$, where

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(C_i),$$

where C_i is defined in Theorem 1.

Corollary 1 gives simplified expressions for σ^2 when one or both $\hat{\theta}_j$, $j = 1, 2$, are obtained from minimizing (5), the SMSEP criterion. This can occur when one or both structural models are specified using the first moment conditions, and one or both estimators are simulated nonlinear least squares estimators resulting from minimizing (5). Related examples are Laffont, Ossard and Vuong (1995) and Li and Vuong (1997). Most interestingly, if both estimators are obtained from minimizing (5), then σ^2 is the same as if θ_j^*, $j = 1, 2$, were known. As a result, σ^2 does not depend on the asymptotic variances of $\hat{\theta}_j$, $j = 1, 2$, meaning that the sampling variability attributed to the estimation of θ_j^* is (asymptotically) irrelevant in using T_n to test H_0.

As can be seen from Theorem 1 and Corollary 1, in order to propose a test statistic that is operational, one needs a consistent estimator for σ^2, the asymptotic variance of T_n. Provided that one can find such a consistent estimator, say $\hat{\sigma}^2$, we have the following result.

Corollary 2. Assume Assumptions 1-4. Let $\hat{T}_n = T_n/\hat{\sigma}$.

(i) Under H_0, $\hat{T}_n \Rightarrow N(0,1)$.

(ii) Under H_1, $\hat{T}_n \rightarrow_p -\infty$.

(iii) Under H_2, $\hat{T}_n \rightarrow_p \infty$.

As stated in Corollary 2, our test statistic $\hat{T}_n$ has a nice asymptotic property in that under H_0, it has a standard normal distribution asymptotically. Therefore, given a consistent estimate $\hat{\sigma}$ for σ, our model selection procedure involves computing $\hat{T}_n$ and then comparing it with critical values from a standard normal distribution. Specifically, let α denote the specified asymptotic significance level of the test and $Z_{\alpha/2} \equiv \Phi^{-1}(1-\alpha/2)$, where $\Phi^{-1}(\cdot)$ denotes the inverse cumulative standard normal distribution. If $|\hat{T}_n| \leq Z_{\alpha/2}$, then we accept H_0. Otherwise, if $\hat{T}_n < -Z_{\alpha/2}$, we reject H_0 in favor of H_1; if $\hat{T}_n > Z_{\alpha/2}$, we reject H_0 in favor of H_2.

It now remains to discuss how to consistenly estimate σ^2. In principle a consistent estimator of σ^2 can be obtained by replacing the quantities in Theorem 1 or Corollary 1 by their empirical counterparts. In particular, in case (iii) of Corollary 1, when both estimators are obtained from minimizing (5) and their sampling variation is asymptotically irrelevant to σ^2, σ^2 can be straightfowardly estimated from the sample variation of $C_i, i = 1, \ldots, n$. This simplicity, however, disappears as soon as at least one of the estimators is obtained from some optimization procedure other than minimizing (5), as computational complications can arise from the need to estimate the terms B_j in Theorem 1 that are associated with the derivatives of $\hat{Q}_j(\cdot)$. To overcome the computational burden in estimating σ^2, we propose to use a bootstrap procedure as a computationally convenient alternative. Specifically, in the general case where at least one of the estimators $\hat{\theta}_j$ are obtained without minimizing (5), the bootstrap procedure in estimating σ^2 consists in the following steps.

Step 1. Resample $\{y_i, \mathbf{x}_i\}; i = 1, \ldots, n$, with replications to get $\{y_i^{(b)}, \mathbf{x}_i^{(b)}\}; i = 1, \ldots, n$ with $b = 1, \ldots, B$.

Step 2. Use the sample $\{y_i^{(b)}, \mathbf{x}_i^{(b)}\}; i = 1, \ldots, n$ to get estimates $\hat{\theta}_j^{(b)}, j = 1, 2; b = 1, \ldots, B$, the same way as obtaining $\hat{\theta}_j, j = 1, 2$, using the original data.

Step 3. For the simulated sample $\{y_{j,i}^{(s_j)}(\hat{\theta}_j^{(b)})\}$; $s_j = 1, \ldots, S_j$, resample it with replications to get $\{y_{j,i}^{(s_j,m)}(\hat{\theta}_j^{(b)})\}$; $s_j = 1, \ldots, S_j$ with $m = 1, \ldots, M$. For each $b = 1, \ldots, B$, calculate $T_n^{(b)} = \sqrt{n}(\hat{Q}_1^{(b)}(\hat{\theta}_1^{(b)}) - \hat{Q}_2^{(b)}(\hat{\theta}_2^{(b)}))$, where $\hat{Q}_j^{(b)}(\hat{\theta}^{(b)}), j = 1, 2$, are the bootstrapped version of $\hat{Q}_j(\hat{\theta}_j)$, and defined as follows

$$\hat{Q}_j^{(b)}(\hat{\theta}_j^{(b)}) = \frac{1}{n}\sum_{i=1}^{n}(y_i^{(b)} - \bar{Y}_{j,i}(\hat{\theta}_j^{(b)}))^2 - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{S_j(S_j-1)}\left[\frac{1}{M}\sum_{m=1}^{M}\sum_{s_j=1}^{S_j}(y_{j,i}^{(s_j,m)}(\hat{\theta}_j^{(b)}) - \bar{Y}_{j,i}(\hat{\theta}_j^{(b)}))^2\right].$$

Step 4. Calculate $\hat{\sigma}_B^2 \equiv \sum_{b=1}^{B}(T_n^{(b)} - T_n)^2/(B-1)$.

Note that the nonparametric boostrap is used in step 1, as under the null, both models can be misspecified. Also, in step 3, to account for the variation associated with the simulation, the simulated sample $\{y_{j,i}^{(s_j)}(\hat{\theta}_j^{(b)})\}$ needs to be resampled in order to get the boostraped version of the test statistic.

By now we have maintained Assumption 3 that assumes the root-n asymptotic normality for the estimators obtained under competing models $\mathcal{M}_j$, $j = 1, 2$. Maintaining this assumption simplifies the presentation and discussion. While most of the estimators that are used in estimating structural models satisfy this assumption, another class of estimators, relevant to some structural models where the support of the dependent variable also depends on the structural parameters, can have n consistency, a rate faster than root-n. These estimators include those based on likelihood (Donald and Paarsch (1993, 1996)), Hong (1998), Chernozhukov and Hong (2004), Hirano and Porter (2003)), and those based on the extreme order statistics (Donald and Paarsch (2002)). It is worth noting that when one or both competing models are estimated by these rate-n consistent estimators, Theorem 1 not only remains valid, but also simplifies in a similar way to that in Corollary 1. The next corollary gives the corresponding results.

Corollary 3. Assume Assumptions 1-2.

(i) If $\hat{\theta}_1$ is rate-n superconsistent, but $\hat{\theta}_2$ is root-n and satisfies Assumption 3, then under H_0, $T_n \Rightarrow N(0, \sigma^2)$, where

$$\sigma^2 = \lim_{n\to\infty} [(1, B_{2,n}) W_{2,n} (1, B_{2,n})'],$$

where $B_{2,n}$ is defined in Theorem 1, and

$$W_{2,n} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var} \begin{pmatrix} C_i \\ U_{2,i} \end{pmatrix}.$$

(ii) If $\hat{\theta}_2$ is rate-n superconsistent, but $\hat{\theta}_1$ is root-n and satisfies Assumption 3, then under H_0, $T_n \Rightarrow N(0, \sigma^2)$, where

$$\sigma^2 = \lim_{n\to\infty} [(1, -B_{1,n}) W_{1,n} (1, -B_{1,n})'],$$

where $B_{1,n}$ is defined in Theorem 1, and

$$W_{1,n} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var} \begin{pmatrix} C_i \\ U_{1,i} \end{pmatrix}.$$

(iii) If both $\hat{\theta}_1$ and $\hat{\theta}_2$ are n superconsistent, then under H_0, $T_n \Rightarrow N(0, \sigma^2)$, where

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(C_i),$$

where C_i is defined in Theorem 1.

As reflected in Theorem 1 and Corollary 3, our proposed model selection procedure can be used when the competing models are estimated by estimators that are either rate-n superconsistent, or have root-n asymptotic normality. Thus, it has generality and wide applicability. Moreover, when both models are estimated by rate-n superconsistent estimators, Corollary 3 indicates that the sampling variability attributed to the estimation of θ_j^*, $j = 1, 2$, does not affect (asymptotically) σ^2, thus calculation of σ^2 greatly simplifies in the same way as in the case when both estimators are obtained from minimizing (5), though the reasons are different.

3 An Empirical Application

To illustrate the usefulness and feasibility of our proposed model selection procedure, we present an application from analyzing the timber sale auctions in Oregon organized by Orgeon Department of Forest (ODF). This data set has been analyzed in Li (2003), which estimates a structural model within an independent private value (IPV) paradigm. A particular feature of the timber auctions in Oregon, as noted in Li (2003), is the presence of the publicly announced reserve prices. As is well known, a structural auction model derived from the game theory assumes that bidders draw their bids dependent of the number of potential bidders. Specifically, within the IPV paradigm, as shown by Riley and Samuelson (1981) among others, the symmetric Nash-Bayesian equilibrium strategy b_m for the m-th bidder with a private value v_m above the reserve price p is given by

$$b_m = v_m - \frac{1}{(F(v_m))^{N-1}} \int_p^{v_m} F^{N-1}(x)dx, \tag{6}$$

where N is the number of potential bidders and $F(\cdot)$ is the private value distribution.

As in Li (2003), we consider 108 lots with different species grades and in different regions. Table 1 gives summary statistics on the data such as the appraised volumes measured in thousand board feet (MBF), the reserve prices, the regional dummies to indicate where the lots are located, the bids per MBF and the log grades. For more details on these variables, see Li (2003). Also, following Li (2003), we assume that the private value density at the ℓ-th lot be specified as

$$f_\ell(v_{m\ell}|z_\ell) = \frac{1}{\exp(\gamma_\ell)} \exp\left[-\frac{1}{\exp(\gamma_\ell)} v_{m\ell}\right], \tag{7}$$

where $v_{m\ell}$ is the private value for the m-th bidder at the ℓ-th auction, $\gamma_\ell = \gamma_0 + \gamma_1 \text{grade}_\ell + \gamma_2 \text{region1}_\ell + \gamma_3 \text{region2\&3}_\ell$, and z_ℓ denotes the heterogeneity vector consisting of variables "grade", "region1" and "region2&3", where "grade" is for log grade to measure the quality, "region1" and "region2&3" are both regional dummies.

Table 1

Summary Statistics of the Timber Sale Data

Variable	Number of Observations	Mean	S.D.	Min	Max
Bid	451	331.62	136.70	119.67	2578.3
Winning Bid	108	382.5	231.13	157.86	2578.3
Reserve Price	108	273.19	77.32	118.32	463.96
Volume	108	3165.22	2894.31	256.74	20211
Grade	108	2.1653	0.3837	1.2727	3.0199
Region1	108	0.8448	0.3625	0	1
Regions2&3	108	0.1397	0.3471	0	1
Number of Submitted Bids	108	4.1759	2.0178	1	10

As indicated from (7), to conduct the structural analysis, one needs to know the number of potential bidders. With the timber auction data in our case, however, we only observe the number of actual bidders, which is not the same as the number of potential bidders due to the fact that the bidders whose valuations are below the reserve prices will not submit their bids. In essence, when reserve prices are binding, the number of potential bidders, if assumed to be a constant across auctions, can be regarded as a structural parameter that cannot be identified from the bidding model but from elsewhere.[5] To resolve the issue of not observing the number of potential bidders, Li (2003) assumes that the number of potential bidders is 10, which is the maximum number of actual bidders in the data set. To illustrate the application of the proposed model selection procedure, we consider another alternative assumption about the number of potential bidders which is $N = 50$. This assumption comes from the fact that there are in total 50 different bidders in the data.[6] Note that the resulting structural models from (7) with different number of

[5]Determining the number of potential bidders in auctions with the presence of reserve prices is indeed a common problem facing empirical economists when analyzing auction data. See also the discussion in footnote 1.

[6]My discussion with the expert at ODF also confirms that there are about 50 firms that could be

bidders $N = 10$ and $N = 50$ are non-nested. Assuming that $N = 10$, Li (2003) estimates the structural model under the specification (8) for the private value distribution using an estimation method based on the indirect inference principle originally suggested by Smith (1993), Gourieroux, Monfort and Renault (1993), and Gallant and Tauchen (1996).[7] Now assuming $N = 50$, we re-estimate the structural model using the same method. Table 2 reports the estimation results.[8] It is interesting to note that only comparing the estimates from these two models with $N = 10$ and $N = 50$, respectively, does not allow us to distinguish between these two models, as the two sets of estimates are similar in both magnitudes and significance levels. Thus, to determine the number of potential bidders that better describes the bidding process, we apply our model selection procedure and obtain that the test statistic is $\hat{T}_n = 11.96$, with the formulation of H_1 as the model with $N = 50$ being preferred and H_2 as the model with $N = 10$ being preferred.[9] As a result, at the 95% significance level, $N = 10$ is preferred to $N = 50$. Also, note that it is possible that neither $N = 10$ nor $N = 50$ could be a correct description of the true number of potential bidders. For instance, we maintain the assumption that the number of bidders is a constant across the auctions. In reality, however, the number of potential bidders may vary across auctions. Nevertheless, since our model selection test allows both models to be misspecified, we can conclude from the test that $N = 10$ is a better approximation than $N = 50$ for the number of potential bidders. In other words, the competition effect is better measured by $N = 10$ than $N = 50$. This application demonstrates the usefulness of our model selection procedure in selecting the competing structural econometric models.[10]

---

potentially interested in timber auctions.

[7]The procedure proposed in Li (2003) consists in the OLS estimation at the first step and simulation and parameter calibration at the second step. See Li (2003) for details.

[8]For completeness and comparison, we also include the results reported in Li (2003) for the case of $N = 10$.

[9]We obtain this $\hat{T}_n$ by setting $S_1 = S_2 = 100$ in calculating T_n, and setting $B = 800$ and $M = 100$ in obtaining $\hat{\sigma}$.

[10]Note that the use of the indirect inference type estimators in estimating our structural models and

Table 2

Estimates for Structural Parameters in Timber Sale Auctions

	Parameter	γ_0	γ_1	γ_2	γ_3
N = 50	Estimate	4.6932	0.2576	0.2250	-0.0678
	Standard Error	0.3247	0.1082	0.2032	0.2352
N = 10	Estimate	4.9042	0.2642	0.2070	-0.0865
	Standard Error	0.3017	0.0962	0.1927	0.2203

4 Conclusion

This paper develops a general framework for testing between competing non-nested structural econometric models. Our method allows for any estimators that are either root-n asymptotically normally distributed or superconsistent, and can be used for distiguishing between two models that are both possibly misspecified. The statistical significance of the difference between two models under consideration is assessed through a simulation based lack-of-fit criterion, taking into account the complex nature of structural econometric models. As such, our approach provides a new model selection method for choosing between competing structural models.

We apply our testing procedure to determine the number of potential bidders in the timber auctions in Oregon. Such an application illustrates the usefulness and generality of our test, and also demonstrates the importance of developing model selection tests in structural econometric models.

As previously mentioned, this paper is motivated by the need to develop a general model selection test for structural models when the estimation methods used do not allow one to use the existing procedures. On the other hand, because of its generality, our proposed method can also be applied to the cases in which the existing model selection

the complex feature of the structural model itself make it difficult to apply the existing model selection procedures here. This application demonstrates the generality of our proposed selection procedure.

tests work as well. For instance, when two competing structural models are estimated by the MLE, we can use the Vuong (1989) likelihood ratio test as well as our test for model selection. Thus, it would be interesting in this case to compare the asymptotic properties of both tests as well as their finite sample performances. This is left for future research.

Appendix

Proof of Proposition 1:

By using a strong law of large numbers such as Theorem 3.3.1 in Amemiya (1985), we have

$$\hat{Q}_j(\theta_j^*) - \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}_{y,\mathbf{x},y_j^{(s_j)}}[q_{i,j,S_j}(\theta_j^*)] \to_{a.s.} 0,$$

where

$$q_{i,j,S_j}(\theta_j) = (y_i - \bar{Y}_{j,i}(\theta_j))^2 - \frac{1}{S_j(S_j-1)}\sum_{s_j=1}^{S_j}(y_{j,i}^{(s_j)}(\theta_j) - \bar{Y}_{j,i}(\theta_j))^2.$$

On the other hand,

$$\begin{aligned}
\mathrm{E}_{y,\mathbf{x},y_j^{(s_j)}}[q_{i,j,S_j}(\theta_j^*)] &= \mathrm{E}_{y,\mathbf{x},y_j^{(s_j)}}[(y_i - \bar{Y}_{j,i}(\theta_j^*))^2] - \mathrm{E}_{y,\mathbf{x},y_j^{(s_j)}}[\frac{1}{S_j(S_j-1)}\sum_{s_j=1}^{S_j}(y_{j,i}^{(s_j)}(\theta_j^*) - \bar{Y}_{j,i}(\theta_j^*))^2] \\
&= \mathrm{E}_{y,\mathbf{x},y_j^{(s_j)}}[(y_i - \mathrm{E}_{\mathcal{M}_j}(y|\mathbf{x}_i,\theta_j^*))^2] + \mathrm{E}_{\mathbf{x},y_j^{(s_j)}}[\bar{Y}_{j,i}(\theta_j^*) - \mathrm{E}_{\mathcal{M}_j}(y|\mathbf{x}_i,\theta_j^*)]^2 \\
&- \frac{1}{S_j}\mathrm{E}_{\mathbf{x}}\mathrm{Var}_{\mathcal{M}_j} y_{j,i}^{(s_j)}(\theta_j^*) \\
&= \mathrm{E}_{y,\mathbf{x}}[(y_i - \mathrm{E}_{\mathcal{M}_j}(y|\mathbf{x}_i,\theta_j^*))^2],
\end{aligned}$$

where $\mathrm{Var}_{\mathcal{M}_j}(\cdot)$ denotes the conditional variance given $\mathbf{x}$ under model $\mathcal{M}_j$, the second equality follows from the unbiased estimation of $\mathrm{Var}_{\mathcal{M}_j}\bar{Y}_{j,i}(\theta_j^*)$, and the conditional independence of y_i and the simulations $y_{j,i}^{(s_j)}$ given $\mathbf{x}_i$ leading to $\mathrm{E}_{y,\mathbf{x},y_j^{(s_j)}}[(y_i - \mathrm{E}_{\mathcal{M}_j}(\bar{Y}_{j,i}|\mathbf{x}_i,\theta_j^*))(\bar{Y}_{j,i} - \mathrm{E}_{\mathcal{M}_j}(\bar{Y}_{j,i}|\mathbf{x}_i,\theta_j^*))] = 0$. As a result,

$$\hat{Q}_j(\theta_j^*) \to_{a.s.} Q_j(\theta_j^*). \tag{A.1}$$

Then Proposition 1 follows from (A.1) above and Assumption 2 that $\hat{\theta}_j$ converges to θ_j^* in probability as $n \to \infty$. □

Proof of Theorem 1:

A Taylor expansion of $\hat{Q}_j(\hat{\theta}_j)$ around θ_j^* yields

$$\hat{Q}_j(\hat{\theta}_j) = \hat{Q}_j(\theta_j^*) + \frac{\partial \hat{Q}_j}{\partial \theta_j'}|_{\bar{\theta}_j}(\hat{\theta}_j - \theta_j^*),$$

where $\bar{\theta}_j$ is a value between $\hat{\theta}_j$ and θ_j^*, $j = 1, 2$. We then have

$$\begin{aligned}
\sqrt{n}\hat{Q}_j(\hat{\theta}_j) &= \sqrt{n}\hat{Q}_j(\theta_j^*) + \frac{\partial \hat{Q}_j}{\partial \theta_j'}|_{\theta_j^*}\sqrt{n}(\hat{\theta}_j - \theta_j^*) \\
&+ \left(\frac{\partial \hat{Q}_j}{\partial \theta_j'}|_{\bar{\theta}_j} - \frac{\partial \hat{Q}_j}{\partial \theta_j'}|_{\theta_j^*}\right)\sqrt{n}(\hat{\theta}_j - \theta_j^*) \\
&= \sqrt{n}\hat{Q}_j(\theta_j^*) + \frac{\partial \hat{Q}_j}{\partial \theta_j'}|_{\theta_j^*}\sqrt{n}(\hat{\theta}_j - \theta_j^*) + o_P(1), \qquad \text{(A.2)}
\end{aligned}$$

where the second equality follows from that $\bar{\theta}_j - \theta_j^* \to 0$ in probability because $\hat{\theta}_j - \theta_j^* \to 0$ in probability as assumed in Assumption 2, and that $\sqrt{n}(\hat{\theta}_j - \theta_j^*) = O_P(1)$ from Assumption 3, as well as $\partial\hat{Q}_j(\bar{\theta}_j)/\partial\theta_j' - \partial\hat{Q}_j(\theta_j^*)/\partial\theta_j' \to 0$ in probability which is a result of Assumption 4. It then follows that

$$\begin{aligned}
\sqrt{n}\{\hat{Q}_1(\hat{\theta}_1) - \hat{Q}_2(\hat{\theta}_2) - (Q_1(\theta_1^*) - Q_2(\theta_2^*))\} &= \sqrt{n}\{\hat{Q}_1(\theta_1^*) - \hat{Q}_2(\theta_2^*) - (Q_1(\theta_1^*) - Q_2(\theta_2^*))\} \\
&+ \frac{\partial \hat{Q}_1}{\partial \theta_1'}|_{\theta_1^*}\sqrt{n}(\hat{\theta}_1 - \theta_1^*) \\
&- \frac{\partial \hat{Q}_2}{\partial \theta_2'}|_{\theta_2^*}\sqrt{n}(\hat{\theta}_2 - \theta_2^*) + o_P(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(C_i - A_1U_{1,i} + A_2U_{2,i}) \\
&- \sqrt{n}(Q_1(\theta_1^*) - Q_2(\theta_2^*)) + o_P(1), \qquad \text{(A.3)}
\end{aligned}$$

where the second equality follows from Assumption 3 and the definition of $\hat{Q}_j(\theta_j^*)$, $j = 1, 2$. Then (i), (ii) and (iii) follow from (A.3) and application of central limit theorem after some algebra. $\square$

Proof of Corollary 1:

If $\hat{\theta}_j$ is obtained by minimizing (5), then $\partial\hat{Q}_j(\hat{\theta}_j)/\partial\theta_j' = 0$ by the first-order condition of the minimization problem. On the other hand, noting that θ_j^* is the probability limit of $\hat{\theta}_j$, $\lim_{n\to\infty}\partial\hat{Q}_j(\theta_j^*)/\partial\theta_j' = 0$. As a result, $\lim_{n\to\infty} B_j = 0$. Then (i), (ii), (iii) follow directly. $\square$

Proof of Corollary 2:

The result directly follows from Theorem 1 and the assumption that $\hat{\sigma}^2$ is a consistent estimator for σ^2. □

Proof of Corollary 3:

If $\hat{\theta}_j$ is rate-n superconsistent, then $\sqrt{n}(\hat{\theta}_j - \theta_j^*) = o_P(1)$. As a result, (A.2) becomes

$$\sqrt{n}\hat{Q}_j(\hat{\theta}_j) = \sqrt{n}\hat{Q}_j(\theta_j^*) + o_P(1). \tag{A.4}$$

(i) Now if $\hat{\theta}_1$ is superconsistent, but $\hat{\theta}_2$ is root-n and satisfies Assumption 3, then (A.4) holds for $\hat{\theta}_1$ while (A.2) holds for $\hat{\theta}_2$. It then follows that (A.3) becomes

$$\begin{aligned}\sqrt{n}\{\hat{Q}_1(\hat{\theta}_1) - \hat{Q}_2(\hat{\theta}_2) - (Q_1(\theta_1^*) - Q_2(\theta_2^*))\} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(C_i + A_2 U_{2,i}) \\ &- \sqrt{n}(Q_1(\theta_1^*) - Q_2(\theta_2^*)) + o_P(1). \quad (A.5)\end{aligned}$$

Then the result follows from (A.5) after some algebra.

(ii) Now if $\hat{\theta}_2$ is superconsistent, but $\hat{\theta}_1$ is root-n and satisfies Assumption 3, then (A.4) holds for $\hat{\theta}_2$ while (A.2) holds for $\hat{\theta}_1$. The result follows from an argument similar to that of (i).

(iii) Now if both $\hat{\theta}_1$ and $\hat{\theta}_2$ are superconsistent, then (A.4) holds for both $\hat{\theta}_1$ and $\hat{\theta}_2$. As a result, (A.3) becomes

$$\begin{aligned}\sqrt{n}\{\hat{Q}_1(\hat{\theta}_1) - \hat{Q}_2(\hat{\theta}_2) - (Q_1(\theta_1^*) - Q_2(\theta_2^*))\} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} C_i \\ &- \sqrt{n}(Q_1(\theta_1^*) - Q_2(\theta_2^*)) + o_P(1). \quad (A.6)\end{aligned}$$

Then the result follows from (A.6) after some algebra. □

References

Akaike, H. (1973): "Information Theory and an Extension of the Likelihood Ratio Principle," in *Proceedings of the Second International Symposium of Information Theory*, Edited by B.N. Petrov and F. Csaki, 257-281, Budapest: Akademiai Kiado.

Akaike, H. (1974): "A New Look at the Statistical Model Identification," *IEEE Transactions and Automatic Control AC-19*, 716-723.

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

Chen, X., **H. Hong**, and **M. Shum** (2003): "Likelihood Ratio Tests between Parametric and Moment Condition Models," Working Paper, Princeton University.

Chernozhukov, V., and **H. Hong** (2004): "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models," *Econometrica*, 72, 1145-1480.

Cox, D.R. (1961): "Tests of Separate Families of Hypotheses," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 105-123.

Diebold, F.X. and **R.S. Mariano** (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

Donald, S. and **H. Paarsch** (1993): "Piecewise Pseudo Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review*, 34, 121-148.

Donald, S. and **H. Paarsch** (1996): "Identification, Estimation, and Testing in Parametric Empirical Models of Auctions Within the Independent Private Value Paradigm," *Econometric Theory*, 12, 512-567.

Donald, S. and **H. Paarsch** (2002): "Superconsistent Estimation and Inference in Structural Econometric Models Using Extreme Order Statistics," *Journal of Econometrics*, Vol. 109 (2), 305-340.

Gallant, A.R. and **G.E. Tauchen** (1996): "Which Moments to Match?" *Econometric Theory*, 12, 657-681.

Gallant, A.R. and **H. White** (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, New York: Basil Blackwell.

Gasmi, F., **J.J. Laffont**, and **Q. Vuong** (1992): "Econometric Analysis of Collusive Behavior in a Soft Drink Industry," *Journal of Economics and Management Strategy*, 1, 277-311.

Gourieroux, C. and **A. Monfort** (1994): "Testing Non-Nested Hypotheses," in R.F. Engle and D. McFadden (eds), *Handbook of Econometrics*, Vol. 4, 2585-2640, Amsterdam: North Holland.

Gourieroux, C. and **A. Monfort** (1996): *Simulation-Based Econometric Methods*, Oxford: Oxford University Press.

Gourieroux, C., **A. Monfort**, and **E. Renault** (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85-S118.

Heckman, J. (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, 109, 673-748.

Hirano, K., and **J. Porter** (2003): "Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support," *Econometrica*, 71, 1307-1338.

Hong, H. (1998): "Nonregular Maximum Likelihood Estimation in Auction, Job Search and Production Frontier Models," Mimeo, Princeton University.

Kitamura, Y. (2000): "Comparing Misspecified Dynamic Econometric Models Using Nonparametric Likelihood," Working Paper, University of Wisconsin.

Kitamura, Y. (2002): "Econometric Comparison of Conditional Models," Working Paper, University of Pennsylvania.

Laffont, J.J. (1997): "Game Theory and Empirical Economics: The Case of Auction Data," *European Economic Review*, 41, 1-35.

Laffont, J.J. and **E. Maskin** (1990): "The Efficient Market Hypothesis and Insider Trading on the Stock Market," *Journal of Political Economy*, 98, 70-93.

Laffont, J.J., **H. Ossard**, and **Q. Vuong** (1995): "Econometrics of First-Price Auctions," *Econometrica*, 63, 953-980.

Li, T. (2003): "Inference of Structural Econometric Models: A Unified Approach," Mimeo, Indiana University.

Li, T. (2005): "Econometrics of First-Price Auctions with Entry and Binding Reservation Prices," *Journal of Econometrics*, 126, 173-200.

Li, T. and **Q. Vuong** (1997): "Using All Bids in Parametric Estimation of First-Price Auctions," *Economics Letters*, 55, 321-325.

Mizon, G.E. and **J.F. Richard** (1986): "The Encompassing Principle and its Application to Testing Non-nested Hypotheses," *Econometrica*, 54, 657-678.

Newey, W.K. and **D. McFadden** (1994): "Large Sample Estimation and Hypothesis Testing," in R.F. Engle and D. McFadden (eds), *Handbook of Econometrics*, Vol. 4, 2111-2245, Amsterdam: North Holland.

Pesaran, M.H. and **M. Weeks** (2001): "Non-Nested Hypothesis Testing: An Overview," in B. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Oxford: Blackwell.

Powell, J.L. (1994): "Estimation of Semiparametric Models," in R.F. Engle and D. McFadden (eds), *Handbook of Econometrics*, Vol. 4, 2443-2521, Amsterdam: North Holland.

Riley, J. and **W. Samuelson** (1981): "Optimal Auctions," *American Economic Review*, 71, 381-392.

Rivers, D. and **Q. Vuong** (2002): "Model Selection Tests for Nonlinear Dynamic Models," *Econometrics Journal*, 5, 1-39.

Smith, A. (1993): "Estimating Nonlinear Time Series Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63-S84.

Smith, R. (1992): "Non-nested Tests for Competing Models Estimated by Generalized Method of Moments," *Econometrica*, 60, 973-980.

Vuong, Q. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 307-333.

West, K.D. (1996): "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, 1067-84.

Wolak, F. (1994): "An Econometric Analysis of the Asymmetric Information Regulator Utility Interation," *Annales d'Economie et de Statistique*, 34, 13-69.

Wooldridge, J.M. (1990): "An Encompassing Approach to Conditional Mean Tests with Applications to Testing Non Nested Hypotheses," *Journal of Econometrics*, 45, 331-350.