Western University
Scholarship@Western

Department of Economics Research Reports | Economics Working Papers Archive

1992

A Comparison of Estimators for Empirical Models of Auctions

Harry J. Paarsch

Follow this and additional works at: https://ir.lib.uwo.ca/economicsresrpt

Part of the Economics Commons

Citation of this paper:

Paarsch, Harry J.. "A Comparison of Estimators for Empirical Models of Auctions." Department of Economics Research Reports, 9210. London, ON: Department of Economics, University of Western Ontario (1992).

032955

ISSN:0318-725X
ISBN:0-7714-1441-2

RESEARCH REPORT 9210

A Comparison of Estimators for Empirical Models of Auctions

by

Harry J. Paarsch

Department of Economics Library
OCT 2 7 1992
University of Western Ontario

October 1992

Department of Economics

Social Science Centre

University of Western Ontario

London, Ontario, Canada

N6A 5C2

A COMPARISON OF ESTIMATORS FOR EMPIRICAL MODELS OF AUCTIONS*

by

Harry J. Paarsch†

May 1992

* This research was supported by a grant from the SSHRC of Canada. Herb Emery and Carol McAusland provided helpful research assistance. I am grateful to John Rust and Quang Vuong for useful discussions of and helpful comments upon the work presented here. I, of course, bear all responsibility for errors.

† Assistant Professor, Department of Economics, University of Western Ontario, London, Ontario, Canada N6A 5C2.

A Comparison of Estimators for Empirical Models of Auctions

Harry J. Paarsch

University of Western Ontario, London, Ontario, Canada N6A 5C2

Using structural econometric models of equilibrium behaviour in games with incomplete information to intrepret field data from auctions has become increasingly widespread. Several different estimation strategies now exist. In this paper, I compare the performance of these different estimators using a stylized empirical model of a procurement auction within the independent private values paradigm.

1. Introduction

Using structural econometric models of equilibrium behaviour in games with incomplete information to interpret field data from auctions has become increasingly widespread. A goal of some recent empirical research has been to determine if the predictions of game theory are consistent with observed data. One proposed research strategy (see, for example, Paarsch [1989, 1992a]) involves noting that the equilibrium strategies of players depend upon the distribution of latent characteristics. This implies that the equilibrium stategies of players are random variables. If the distribution of latent characteristics comes from a particular class of distributions, then rational behaviour within that class of distributions will impose certain testable restrictions upon the data generating process of the equilibrium strategies. Paarsch (1989, 1991, 1992a) as well as Laffont, Ossard, and Vuong (1991) have employed this structural econometric framework to interpret field data from actual auctions using the econometric methods proposed in Paarsch (1989,1992a) and developed in Donald and Paarsch (1991, 1993) as well as Laffont, Ossard, and Vuong (1991). Some of the empirical strategies discussed by Paarsch (1989, 1992a), Donald and Paarsch (1991, 1993) as well as Laffont, Ossard, and Vuong (1991) are quite different. In this paper, I employ a stylized empirical model of a procurement auction within the independent private values paradigm in conjunction with Monte Carlo methods to compare the performance of these different estimators for empirical models of auctions.

This paper is in four more parts. In section 2, I outline a simple model of a procurement auction that contains all of the elements required to construct a structural econometric model of equilibrium behaviour. In section 3, the main methods for estimating this model are described, while a comparison of the estimators is presented in section 4. A summary of the paper's main findings is presented in section 5.

2. Theoretical Model and Empirical Framework

To illustrate the particular class of estimation problems which have been investigated, I consider a sealed-bid auction as a non-coöperative game.[1] I examine auctions at which a known number of bidders N compete to perform a single task for a government agency, with the lowest bidder winning the auction. Each player is assumed to know his own cost of performing the task, but not those of his opponents. I assume that the heterogeneity across agents in the cost of performing the task can be described by a continuous random variable c which has the probability density function $f(c)$ and the cumulative distribution function $F(c)$. The costs of players are assumed to be independent draws from $F(c)$, and $F(c)$ is assumed to be common knowledge. I assume that bidders are risk neutral with respect to winning the auction, and that the ith bidder chooses a bid b_i to maximize his expected profit. Finally, I focus upon symmetric Bayesian-Nash equilibria.

2.1. Deriving the Equilibrium Bid Function

To construct the equilibrium, suppose that the $M = N - 1$ opponents of player i are using a common bidding rule $\beta(c)$ which is increasing and differentiable in c. Since costs are modelled as independent draws from a common distribution, the probability of player i winning with bid b_i equals the probability that each of his opponents bids higher because each has a higher cost:

$$[1 - F(\beta^{-1}(b_i))]^M.$$

Here $\beta^{-1}(b_i)$ denotes the inverse of the bid function. Given that his cost c_i is determined before the bidding, player i's choice of b_i has only two effects upon his

[1] A reader who is unfamiliar with the auction literature will find the surveys by Milgrom (1985, 1987) as well as McAfee and McMillan (1987) helpful.

expected profit

$$(b_i - c_i) \cdot [1 - F(\beta^{-1}(b_i))]^M.$$

The lower is b_i, the higher is his probability of winning the auction $[1 - F(\beta^{-1}(b_i))]^M$, but the lower is his pay-off when he wins $(b_i - c_i)$. Maximizing behaviour implies that the optimal bid solves the first-order condition

$$[1 - F(\beta^{-1}(b_i))]^M - M(b_i - c_i) f(\beta^{-1}(b_i)) \cdot [1 - F(\beta^{-1}(b_i))]^{M-1} \cdot \frac{d\beta^{-1}(b_i)}{db_i} = 0. \quad (2.1)$$

Symmetry among bidders implies

$$b_i = \beta(c_i). \quad (2.2)$$

Substituting (2.2) into (2.1), recalling that $d\beta^{-1}(b_i)/db_i = 1/\beta'(c_i)$, and requiring (2.1) to hold for all feasible c_i's, yields the following differential equation for β:

$$\beta'(c) - \beta(c)\frac{Mf(c)}{[1 - F(c)]} = \frac{-Mcf(c)}{[1 - F(c)]}. \quad (2.3)$$

Integrating (2.3), and imposing the boundary condition $\beta(\infty) = \infty$, yields[2]

$$\beta(c) = c + \frac{\int_c^\infty [1 - F(\xi)]^M \, d\xi}{[1 - F(c)]^M}. \quad (2.4)$$

Denoting $c_{(i:N)}$ as the ith smallest order statistic for a sample of size N from the distribution of c, the winner of the auction will be the player with the lowest cost $c_{(1:N)}$. Because the winning bid function is monotonic in $c_{(1:N)}$, its distribution is related to that of the smallest order statistic for a sample of size N from the distribution of c.

2.2. Strategy for Interpreting Data

Exploiting the fact that (2.4) is a monotonic function of c (the lower is a player's cost, the less he will bid) provides one strategy for interpreting field data; see, for

[2] In fact, simply imposing $\beta(\infty) = \infty$ is insufficient to guarantee a unique solution since adding any constant κ to that solution is also a solution. In this case, $\kappa = 0$ is the appropriate constant.

example, Paarsch (1989,1992a). Because the bidding rules are functions of the random variable c, the bids are also random variables and their densities are related to $f(c)$. For example, the density of $\beta(c)$, denoted $g(b)$, is

$$g(b) = \frac{f(\beta^{-1}(b))}{\beta'(\beta^{-1}(b))},$$

where $\beta'(c)$ is the Jacobian of the transformation of c to $\beta(c)$.

The winning bid is a simple function of the $\{c_i\}_{i=1}^{N}$. Hence, its density is related to $f(c)$. The density of the winning bid $w = \beta(c_{(1:N)})$, denoted $h(w)$, is

$$h(w) = \frac{N[1 - F(\beta^{-1}(w))]^M f(\beta^{-1}(w))}{\beta'(\beta^{-1}(w))} \tag{2.5}$$

where

$$N[1 - F(z)]^M f(z)$$

is the density of $z = c_{(1:N)}$, the lowest cost draw for a sample of size N.

Consider a family of distributions for c which depend upon the parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$. Clearly, the parameter vector θ will embed itself in (2.5). Without any loss of generality, let $f(c)$ have support upon the interval $[0, \infty)$. Note that evaluating (2.4) at 0, the lower bound for c, implies that the distribution of the winning bid, assuming that it exists, has support upon

$$\left[\int_0^\infty [1 - F(\xi;\theta)]^M \, d\xi, \infty\right) \equiv [\Im(\theta, M), \infty). \tag{2.6}$$

$\Im(\theta, M)$ is the expectation of $c_{(1:M)}$, the lowest order statistic from a sample of size M; $\Im(\theta, M)$ is the equilibrium amount a player who had a cost draw of zero would bid when playing against M opponents. Because the support of $h(w;\theta, M)$ depends upon the parameters of interest, the standard regularity conditions used to demonstrate the consistency and asymptotic normality of the maximum likelihood estimator of θ no longer apply, and alternative estimation methods must be pursued.

One alternative could be to abandon estimation by a method like maximum likelihood and to use some other procedure. For example, suppose that the kth raw moment of w

$$\mathrm{E}[w^k] = \mu_k(\theta, M) = \int_{\Im(\theta,M)}^{\infty} w^k h(w;\theta, M) \, dw \qquad k = 1, 2, \ldots \tag{2.7}$$

has a closed-form solution, then the observed data w^k can be decomposed as follows:

$$w^k = \mu_k(\theta, M) + u_k \qquad k = 1, 2, \ldots, \tag{2.8}$$

where the expectation of u_k is zero, while its variance depends upon M. The parameter vector θ can then be estimated by non-linear least squares, for example. Often, however, especially in the presence of reservation prices, (2.7) will not have a closed-form solution, and will only be defined numerically.

Donald and Paarsch (1991, 1993) as well as Laffont, Ossard, and Vuong (1991) have developed methods for estimating structural econometric models derived within the above framework. Below, I summarize these estimators.

3. Alternative Estimators

In what follows, I focus upon the winning bid w at an auction with N potential bidders. For the tth $(t = 1, \ldots, T)$ auction, I assume that the pair (w_t, N_t) is observed. Often, additional covariates will be considered important, and typically, institutional rules such as reservation prices will be introduced. For the purposes of this paper, however, I ignore these issues as they do not alter the substance of the main points I make.

3.1. Non-Linear Least Squares

The method of non-linear least squares can be applied directly to estimate the parameters of the empirical specification (2.8), provided (2.7) has a closed-form solution. When $f(c)$ is within the Pareto or Weibull families of distributions, then (2.7) will have closed-form solutions; see, for example, Paarsch (1992a). In such cases, the non-linear least squares estimator minimizes the following objective function:

$$V(\theta) = \sum_{t=1}^{T} \left(w_t - \mu_1(\theta, M_t)\right)^2.$$

The introduction of covariates or the presence of reservation prices can complicate this model. While there are examples of $f(c)$'s which admit regressor and closed-form solutions to (2.7) in the presence of reservation prices, as Laffont, Ossard, and Vuong

(1991) point out, these solutions are not general. Because the statistical properties of the non-linear least squares estimator are well understood and documented (see, for example, Jennrich [1969]), I omit the details here and proceed to describe the new and alternative estimators proposed by Donald and Paarsch (1991, 1993) as well as Laffont, Ossard, and Vuong (1991).

3.2. Piecewise Pseudo-Maximum Likelihood

Donald and Paarsch (1993) develop an alternative approach to estimating the parameter vector θ by non-linear least squares, piecewise pseudo-maximum likelihood estimator. The basic idea behind that method of estimation is as follows: For any particular $M_t = M$, the lower bound function $\underline{w}(M) = \Im(\theta, M)$ can be consistently estimated by the smallest w_t over all of those observations with $M_t = M$, denoted $\hat{\underline{w}}(M)$. Consider a partition of the vector θ into a scalar θ_1 and the remaining $(p-1)$ parameters, denoted θ_2. Suppose that the function $\Im(\theta_1, \theta_2, M_t)$ is monotonic and invertible, so that one can write $\theta_1 = \theta_1(\theta_2, \underline{w}(M_t), M_t)$. Treat $\hat{\underline{w}}(M_t)$ as if it were the lower bound, and substitute it for $\underline{w}(M_t)$ in $\theta_1 = \theta_1(\theta_2, \underline{w}(M_t), M_t)$. θ_1 is now a function of θ_2 and the data. Substitute this function into the logarithm of the likelihood function and then maximize over θ_2. The method is called "piecewise pseudo-maximum likelihood estimation" because the logarithm of the likelihood function is broken up into pieces depending upon the value of the covariate M, and because no first-order condition is used to concentrate the likelihood function.

Donald and Paarsch have demonstrated the consistency of the piecewise pseudo-maximum likelihood estimator. They have also given conditions under which that estimator is distributed asymptotically normal. An interesting feature of the estimator is that its asymptotic distribution does not depend upon that of the preliminary estimator $\hat{\underline{w}}(M)$ because the latter converges at rate T instead of the usual rate $\sqrt{T}$. In small samples, however, some bias can be introduced by the pre-estimation error in $\hat{\underline{w}}(M)$. This problem is most acute when T_M, the number of observations with $M_t = M$, is small, a situation commonly encountered in practice.

A drawback of the method is that it relies upon the presence of a partition of θ so that $\Im(\theta_1, \theta_2, M)$ can be inverted. When auctions have reserve prices, $\Im(\theta, M)$ is often only defined numerically because it involves the truncated moments of the latent cost distribution; see Paarsch (1991).

Another drawback of the method is that introducing continuous covariates into the empirical framework is often impossible. To see this, suppose that, in addition to the covariate M_t, a vector of other covariates Z_t is also considered important in determining bids. One way this might happen is if the distribution of c depends upon Z_t; i.e., $F(c) = F(c; \theta, \phi, Z_t)$. The lower bound will then be a function of the Z_t's too. That is,

$$\Im(\theta, \phi; M_t, Z_t) \leq w_t \quad t = 1, \ldots, T,$$

where the ϕ is an unknown parameter vector of dimension q which relates to the Z_t. When the elements of the Z_t are indicator variables, one can apply the methods of Donald and Paarsch with only trivial modifications to estimate the parameter vector $\alpha = (\theta, \phi)$ On the other hand, if the elements of the Z_t are continuous covariates, such as prices and quantities, then the methods of Donald and Paarsch cannot be applied directly.

3.3. Maximum Likelihood

To circumvent the problems associated with continuous covariates and $\Im(\theta, M)$'s which are only defined numerically, Donald and Paarsch (1991) derived the maximum likelihood estimator of the parameter vector using non-linear programming.

The solution to the maximum likelihood estimation problem can be motivated by presenting the solution to a simpler problem. Consider a random sample of size T for a random variable w_t that is distributed uniformly on the interval $[0, \theta]$, where θ is an unknown parameter which the investigator seeks to estimate. The density of w is

$$h(w; \theta) = \begin{cases} \theta^{-1} & \text{for all } w \in [0, \theta] \\ 0 & \text{otherwise.} \end{cases}$$

The conventional method of deriving the maximum likelihood estimator of θ would involve maximizing the following likelihood function with respect to θ:

$$L(\theta; w_1, w_2, \ldots, w_t) = \prod_{t=1}^{T} h(\theta; w_t) = \theta^{-T},$$

or equivalently maximizing the logarithm of the above likelihood function with respect to θ

$$\log L(\theta; w_1, w_2, \ldots, w_t) = -T \log \theta.$$

Of course, as θ tends to zero the functions $L(\theta)$ and $\log L(\theta)$ tend to infinity. But a θ of zero implies that none of the data should have been observed in the first place. A simple solution to this problem is to re-write this as a constrained optimization problem. In particular, maximize $L(\theta)$ or $\log L(\theta)$ subject to the constraint that all of the observed data be consistent with the resulting estimate. That is,

$$\max_{<\theta>} \; -T \log \theta \quad \text{subject to} \quad \begin{cases} w_1 \leq \theta \\ w_2 \leq \theta \\ \quad \vdots \\ w_T \leq \theta. \end{cases}$$

The maximum likelihood estimator is then

$$\hat{\theta} = \max[w_1, w_2, \ldots, w_T].$$

In the model considered above, Donald and Paarsch (1991) define the maximum likelihood estimator as the solution to the following optimization problem:

$$\max_{<\theta>} \sum_{t=1}^{T} \log h(w_t; \theta, M_t) \quad \text{subject to} \quad \begin{cases} \Im(\theta, M_1) \leq w_1 \\ \Im(\theta, M_2) \leq w_2 \\ \quad \vdots \\ \Im(\theta, M_T) \leq w_T. \end{cases}$$

Given the assumptions maintained by Donald and Paarsch, one can solve for the maximum likelihood estimator by maximizing the following Lagrangean:

$$\mathcal{L}(\theta, \lambda) = \sum_{t=1}^{T} \ell_t(\theta, \lambda_t) = \sum_{t=1}^{T} \Big(\log h(w_t; \theta, M_t) + \lambda_t \big(w_t - \Im(\theta, M_t)\big) \Big)$$

with respect to θ, where $\lambda = (\lambda_1, \ldots, \lambda_T)$ is the vector of T Lagrange multipliers. The maximum likelihood estimator $\hat{\theta}$, which Donald and Paarsch demonstrate to be consistent, satisfies the following conditions:

$$\begin{aligned} \sum_{t=1}^{T} \big(\nabla_\theta \log h(w_t; \hat{\theta}, M_t) - \lambda_t \nabla_\theta \Im(\hat{\theta}, M_t)\big) &= \mathbf{0} \\ \lambda_1 \big(w_1 - \Im(\hat{\theta}, M_1)\big) &= 0 \\ \lambda_2 \big(w_2 - \Im(\hat{\theta}, M_2)\big) &= 0 \\ &\vdots \\ \lambda_T \big(w_T - \Im(\hat{\theta}, M_T)\big) &= 0. \end{aligned}$$

At most, p of the T constraints will ever bind at one time; i.e., $T-p$ of the Lagrange multipliers will be zero at the optimum. For the binding constraints, the λ_t's will be non-negative.

A natural way to examine the asymptotic behaviour of the maximum likelihood estimator would be to consider the Hessian matrix of the Lagrangean

$$\begin{aligned}\nabla_{\theta\theta}\ \mathcal{L}(\theta) &= \sum_{t=1}^{T} \nabla_{\theta\theta}\ \ell_t(\theta, \lambda_t) \\ &= \sum_{t=1}^{T} \left(\nabla_{\theta\theta}\ \log h(w_t; \theta, M_t) - \lambda_t \nabla_{\theta\theta} \Im(\theta, M_t)\right).\end{aligned}$$

This is useful when the solution to an optimization problem occurs along a smooth and differentiable part of the contraint set, but as Donald and Paarsch point out the solution typically obtains at the intersection of the constraints. In this case, the Hessian is ill-defined. Moreover, the properties of the perturbed optimum are determined solely by the constraints. As may be expected, the distribution theory for the estimator is quite complicated. Indeed, the limiting distributions of the estimators will fall into the usual normal limiting family only in a special case. Allowing for only a slightly more general case will change the nature of the limiting distributions. In particular, the distribution of the estimator often falls within the Weibull limiting family. One attractive feature of this estimator is that it often converges at rate T as opposed to rate $\sqrt{T}$ which is typically the case in statistics, a fact arising from the use of extreme value statistics instead of averages to define the estimator. This quick rate of convergence can be very important in applications to auctions where sample sizes of 50 are common, and those of 200 would be considered large.

3.4. Simulated Non-Linear Least Squares

As mentioned above, the solution to (2.7) is often only defined numerically. In such cases, employing the method of non-linear least squares to estimate θ is impossible. Laffont, Ossard, and Vuong (1991) propose a general method for estimating empirical models of auctions which they call, "simulated non-linear least squares." This method is closely related to that of simulated moments developed by McFadden (1989) and Pakes and Pollard (1989), and appears to represent a major step forward in the structural econométric analysis of auctions.

Laffont, Ossard, and Vuong note that within the independent private values paradigm, the equilibrium bid function for the tth auction can be written as the following conditional expectation:

$$\beta_t(c) = \mathrm{E}[c_{(2:N_t)} \mid c_{(1:N_t)} = c]$$

which implies that

$$\begin{aligned} \mathrm{E}[w_t] &= \mu_1(\theta, M_t) \\ &= \mathrm{E}[c_{(2:N_t)}] \\ &= \int_0^\infty \cdots \int_0^\infty c_{(2:N_t)} f_t(c_1;\theta) \ldots f_t(c_{N_t};\theta)\, dc_1 \ldots dc_{N_t}. \end{aligned} \tag{3.1}$$

Here, the t subscript on $f_t(\cdot)$ is used to signify that the density $f(\cdot)$ can vary from observation to observation depending upon, for example, covariates. Equation (3.1) is used as the basis of the proposed simulated non-linear least squares estimator. The idea is that for each of $t = 1, \ldots, T$ observations, S independent samples of size N_t are drawn; (3.1) is then approximated by the sample mean of this function for each of the simulated samples. Following Gourieroux and Monfort (1990), Laffont, Ossard, and Vuong use the concept of the "importance function" (common in Bayesian Monte Carlo integration) to get around the problem that the parameter vector to be estimated θ is part of the function to be simulated $\mu_1(\theta, M)$. Letting $\psi_t(\cdot)$ denote the importance function — which the researcher must specify *ex ante* — Laffont, Ossard, and Vuong estimate

$$\begin{aligned} \mu_1(\theta, M_t) &= \int_0^\infty \cdots \int_0^\infty c_{(2:N_t)} f_t(c_1;\theta) \ldots f_t(c_{N_t};\theta)\, dc_1 \ldots dc_{N_t} \\ &= \int_0^\infty \cdots \int_0^\infty c_{(2:N_t)} \frac{f_t(c_1;\theta) \ldots f_t(c_{N_t};\theta)}{\psi_t(c_1) \ldots \psi_t(c_{N_t})}\, \psi_t(c_1) \ldots \psi_t(c_{N_t})\, dc_1 \ldots dc_{N_t} \end{aligned}$$

by

$$\bar{m}(\theta, M_t) = \frac{1}{S} \sum_{s=1}^{S} m_s(\theta, M_t)$$

where

$$m_s(\theta, M_t) = c^s_{(2:N_t)t} \frac{f_t(c^s_{1t};\theta) \ldots f_t(c^s_{N_t t};\theta)}{\psi_t(c^s_{1t}) \ldots \psi_t(c^s_{N_t t})}.$$

Here, each of the c_{it}^s's represents an independent draw for a random variable having the probability density function $\psi_t(\cdot)$, $i = 1, \ldots, N_t$; $s = 1, \ldots, S$; and $t = 1, \ldots, T$.

The principal advantage of this method is that it eliminates the numerical evaluation of integrals that arise when calculating

$$\mathrm{E}[w] = \int_0^\infty \left[z + \frac{\int_z^\infty [1 - \int_0^\xi f(u;\theta)\, du]^M \, d\xi}{[1 - \int_0^z f(u;\theta)\, du]^M} \right] N[1 - \int_0^z f(u;\theta)\, du]^M f(z)\, dz.$$

The method developed by Laffont, Ossard, and Vuong appears to be quite general. In particular, it appears useful for applications within other paradigms where the equilibrium bid function can also be expressed as an expectation; e.g., the affiliated private values model of Milgrom and Weber (1982).[3]

Laffont, Ossard, and Vuong note that using $\bar{m}(\theta, M_t)$ as $\mu_1(\theta, M_t)$ when minimizing the objective function

$$\sum_{t=1}^{T} \left(w_t - \bar{m}(\theta, M_t)\right)^2$$

with respect to θ will produce an inconsistent estimator for any finite number of simulations S because $\bar{m}(\theta, M_t)$ estimates $\mu_1(\theta, M_t)$ with error. They derive the simulated non-linear least squares estimator of θ, which is consistent for a finite fixed S simulations, by minimizing the following objective function:

$$Q(\theta) = \sum_{t=1}^{T} \left[\left(w_t - \bar{m}(\theta, M_t)\right)^2 - \frac{1}{S(S-1)} \sum_{s=1}^{S} \left(m_s(\theta, M_t) - \bar{m}(\theta, M_t)\right)^2 \right] \tag{3.2}$$

with respect to θ. The bias introduced by the pre-estimation error in $\bar{m}(\theta, M_t)$ is corrected by introducing the second term of $Q(\theta)$, which represents an estimate of the sample variance of $\bar{m}(\theta, M_t)$. Laffont, Ossard, and Vuong show that the simulated non-linear least squares estimator is distributed asymptotically normal, and they derive an estimator of the variance-covariance matrix of this estimator.

The work of Laffont, Ossard, and Vuong appears to be a major contribution. At this point it would appear useful to evaluate the relative merits of the approaches considered by Paarsch (1989,1991,1992a), Donald and Paarsch (1991, 1993) as well as Laffont, Ossard, and Vuong (1991).

[3] Elsewhere, I am attempting this sort of an analysis; see Paarsch (1992b).

4. Comparison of the Estimators

The main difference between the methods of Paarsch (1989,1991,1992a) and Donald and Paarsch (1991, 1993), and those of Laffont, Ossard, and Vuong (1991) is the use of distributions. For computational parsimony, Paarsch and Donald and Paarsch typically choose distributions for which the winning bid has a closed-form regression function or a density function which is easy to evaluate. In some models, structure is exploited to introduce covariates, but as Laffont, Ossard, and Vuong point out, these solutions are not general. In addition, introducing institutional rules such as reservation prices is often quite difficult. Laffont, Ossard, and Vuong develop a framework within which only the density $f(\cdot)$ need have a closed-form representation on a computer, while the importance function $\psi(\cdot)$ must have a cumulative distribution function whose inverse can be calculated efficiently. Additional covariates and institutional rules can be admitted in a straightforward way. Thus, the method of Laffont, Ossard, and Vuong appears to be quite general. Because this generality obtains at the cost of approximating the exact solution by simulation on a computer, it is interesting to examine the costs incurred in terms of inefficiency as well as small sample bias when simulated approximations are used. I shall address these issues using as a benchmark a stylized empirical model for which the winning bid's regression function has a closed-form solution, and for which the winning bid's density is easy to calculate. I am then able to investigate the effects that simulation has upon the performance of the simulated non-linear least square estimator vis-à-vis the maximum likelihood and non-linear least squares estimators.

In the benchmark model, I model c as an exponentially distributed random variable. Thus,

$$f(c;\theta) = \frac{1}{\theta}\exp(-c/\theta) \quad 0 < c,\ 0 < \theta.$$

In this case,

$$\beta(c) = c + \frac{\theta}{M}$$

and

$$w = \beta(z) = z + \frac{\theta}{M}$$

where $z = c_{(1:N)}$ has probability density function

$$\frac{N}{\theta}\exp(-zN/\theta),$$

so the tth observed w has the following probability density function:

$$h(w_t;\theta,M_t) = \frac{N_t}{\theta}\exp\left(\frac{-w_tN_t}{\theta}+\frac{N_t}{M_t}\right) \qquad \frac{\theta}{M_t} \leq w_t. \tag{4.1}$$

Now

$$\mathrm{E}[w_t] = \mu_1(\theta,M_t) = \frac{\theta}{N_t}+\frac{\theta}{M_t} = \theta\left(\frac{2N_t-1}{N_tM_t}\right),$$

so an empirical specification like (2.8) is

$$w_t = \theta\left(\frac{2N_t-1}{N_tM_t}\right)+u_{1t}. \tag{4.2}$$

Note that (4.2) is a linear model, and the small sample behaviour of the estimator for θ can be investigated easily. Also, departures in the behaviour of the simulated non-linear least squares estimator from the behaviour of benchmark estimators can be attributed solely to simulation error.

There are least two issues of concern when implementing optimization estimators based upon simulation. First, how much simulation should be done; i.e., how big should S be? And second, how large must the sample size T be for the estimator's asymptotic behaviour to settle down? To address these issues, I consider simulation sample sizes of 25, 50, and 100, and samples sizes of 50, 100, and 200. The choice of sample sizes was determined by the amount of data that is typically available concerning auctions. In particular, samples size of 50 are relatively common, whereas those of 200 would be considered large. The choice of simulation sample sizes, on the other hand, was determined by computational costs. In particular, since the researcher must solve the estimation problem repeatedly when simulation methods are used, it is useful to know how little simulation he can get away with, and still preserve a tolerable degree of precision. Note that Laffont, Ossard, and Vuong (1991) use $S = 20$ in their application of the simulated non-linear least squares estimator.

To reduce any uncertainty which might be introduced by the choice of importance function, I chose to consider as the importance function the true density at the true value of the parameter. For the experiments considered here, I set θ equal to one.

To reduce uncertainty which might be introduced by sample covariate design, I chose to consider a paradigm within which the number of bidders did not vary. I

chose $N = 5$ (or $M = 4$) because this corresponded to situations often encountered in practice. Under these assumptions, the piecewise pseudo-maximum likelihood and the maximum likelihood estimator are identical. Moreover, the small sample distributions of the maximum likelihood and non-linear least squares estimators can be calculated exactly. In particular, the maximum likelihood estimator of θ is

$$\hat{\theta}^{\text{mle}} = M w_{(1:T)} = M\left[z_{(1:T)} + \frac{\theta}{M}\right]$$

where $z_{(1:T)}$ is distributed exponentially with hazard rate parameter NT/θ. The non-linear least squares estimator of θ is

$$\begin{aligned}\hat{\theta}^{\text{nls}} = \frac{NM\sum_{t=1}^{T} w_t}{(2N-1)T} &= \frac{NM}{(2N-1)T}\left[\sum_{t=1}^{T}\left(w_t - \frac{\theta}{M}\right) + \frac{\theta T}{M}\right] \\ &= \frac{NM}{(2N-1)T}\left[\sum_{t=1}^{T} z_t + \frac{\theta T}{M}\right]\end{aligned}$$

where z_t is distributed exponentially with hazard rate parameter N/θ implying that $\sum_{t=1}^{T} z_t$ is distributed gamma with parameters T and N/θ.

The random numbers for the experiments were generated using the multiplicative congruential method with modulus $(2^{31} - 1)$, multiplier 397204094.[4] This method generates uniform pseudo-random numbers on the interval $(0, 1)$. Using the property that the distribution function is distributed uniformly on the interval $(0, 1)$, I applied the inverse distribution function to obtain the pseudo-random w's. The initial seed used to generate the $\{w_t\}_{t=1}^{T}$ was 2420375, while the initial seed to generate the simulation samples was 123457.

In calculating the empirical distributions of the simulated non-linear least squares estimator, I encountered several numerical problems. First, the objective function $Q(\theta)$ admits local optima. This property is illustrated in Figure 1 where $S = 5$ and $T = 10$. Thus, using a Newton-Raphson type search was ruled out. Instead, I grid-searched the parameter space from zero to six. In doing this, I encountered another

[4] This is by no means the only random number generator available as Marsaglia (1972) and others have demonstrated; see Knuth (1969) for a survey. I have chosen to use this random number generator because most researchers use random number generators like it. Also, among the class of linear congruential generators available, this one has relatively attractive properties; see Fishman and Moore (1982) for details.

property of the objective function $Q(\theta)$: the global optimum is often not near one (the true value of the parameter generating the data), but rather somewhere near four. This property is illustrated in Figure 2 where $S = 25$ and $T = 50$. In Figures 3 and 4, the expected value of the non-linear and simulated non-linear least squares objective functions are presented. In both figures, $S = 25$, $S = 50$, and $S = 100$, while $T = 50$ in Figure 3 and $T = 200$ in Figure 4. For both figures, I used 1000 replications to calculate the average values of the objective functions at each value of θ from one to ten using intervals of 0.1. Notice that although a local minimum to the function exists at one in Figures 3 and 4 when $S = 25$, the global minimum is somewhere near four; the local and global minima are reversed when $S = 50$ and $S = 100$.

In Table 1, I present the exact small sample results for the maximum likelihood and non-linear least squares estimators as well as the results of the Monte Carlo simulations for the simulated non-linear squares estimator. The estimators are denoted respectively "mle", "nls", and "snls". Also, the subscripts "25", "50", and "100" on the simulated non-linear least squares estimators denote the simulation sample sizes 25, 50, and 100 respectively. The abbreviations St.Dev., L.Q., and U.Q. denote respectively the standard deviation, lower quartile, and upper quartile of the estimator's distribution. In parentheses beside each of the simulated non-linear least squares estimators, I present the proportion of the 1000 observations which were less than two. In Figure 5, I present the population density functions and empirical distributions for the estimators.

Notice that both the maximum likelihood and the non-linear least squares estimators do well at estimating θ which is in marked contrast to the performance of the simulated non-linear least squares estimator. As both S and T increase, the simulated non-linear least squares estimator's performance improves. This result obtains because although the function $\bar{m}(\cdot)$ may approximate the function $\mu_1(\cdot)$ quite well, the function $Q(\theta)$ does not approximate the function $V(\theta)$ well. Thus, using $\hat{\theta}_S^{\text{snls}}$, the minimum of $Q(\theta)$, to approximate $\hat{\theta}^{\text{nls}}$, the minimum of $V(\theta)$, does not provide an accurate estimate of θ, although the behaviour of the simulated non-linear least squares estimator appears to improve for large S. In this case, the term

$$\frac{1}{S(S-1)} \sum_{t=1}^{T} \sum_{s=1}^{S} \left(m_s(\theta, M_t) - \bar{m}(\theta, M_t)\right)^2$$

become relatively small, and in the limit disappears, so the term

$$\sum_{t=1}^{T} \left(w_t - \bar{m}(\theta, M_t)\right)^2$$

approximates the function $V(\theta)$ well.

5. Summary

The distressing fact which emerges from the above tables and figures is that the simulated non-linear least squares estimator only finds a neighbourhood of the true parameter value some of the time. Also, the amount of simulation required to get accurate estimates is quite high, and in some problems it could be prohibitive. In any case, S should be much larger than 20, the value which was used in the application presented by Laffont, Ossard, and Vuong.

Notice that none of the consistency problems which plague the simulated non-linear least squares estimator arise when either the maximum likelihood or the non-linear least squares estimators are used. Of course, this comparison is a little unfair. In all cases where one would want to use the simulated non-linear least squares estimator, the non-linear least squares estimator would be infeasible, and in some cases the maximum likelihood estimator would also be infeasible. Although the maximum likelihood estimator requires much more structure than the simulated non-linear least squares estimator, it has one redeeming feature: the maximum likelihood estimator often coverges at rate T, while the simulated non-linear least squares estimator only converges at rate $\sqrt{T}$. In practical terms, this means that an empirical researcher can obtain as much accuracy with a sample size of 50 when using the method of maximum likelihood as he could with a sample size of 2500 when using the method of simulated non-linear least squares. This is of particular relevance when investigating field data from actual auctions because sample sizes of 50 are common, and those of 200 would be considered large.

A. Appendix

In this appendix, I present the first and second derivatives of the simulated non-linear least squares objective function. Recall that the simulated non-linear least squares objective function is

$$Q(\theta) = \sum_{t=1}^{T} \left[(w_t - \bar{m}_t)^2 - \frac{1}{S(S-1)} \sum_{s=1}^{S} (m_{st} - \bar{m}_t)^2 \right]$$

where

$$m_{st} = c^s_{(2:N_t)t} \frac{f_t(c^s_{1t}) \dots f_t(c^s_{N_t t})}{\psi(c^s_{1t}) \dots \psi(c^s_{N_t t})}$$

and

$$\bar{m}_t = \frac{1}{S} \sum_{s=1}^{S} m_{st}.$$

Let

$$\mathcal{F}^s_t = f_t(c^s_{1t}) \dots f_t(c^s_{N_t t}) = \frac{1}{\theta^{N_t}} \exp\left(- \sum_{i=1}^{N_t} c^s_{it}/\theta \right)$$

and

$$\mathcal{G}^s_t = \psi_t(c^s_{1t}) \dots \psi_t(c^s_{N_t t}),$$

a constant. The first derivative of $Q(\theta)$ with respect to θ is

$$\frac{dQ(\theta)}{d\theta} = \sum_{t=1}^{T} -2 \left[(w_t - \bar{m}_t) \frac{d\bar{m}_t}{d\theta} + \frac{1}{S(S-1)} \sum_{s=1}^{S} (m_{st} - \bar{m}_t) \left(\frac{dm_{st}}{d\theta} - \frac{d\bar{m}_t}{d\theta} \right) \right]$$

where

$$\frac{dm_{st}}{d\theta} = \frac{c^s_{(2:N_t)t}}{\mathcal{G}^s_t} \frac{d\mathcal{F}^s_t}{d\theta}$$

and

$$\frac{d\bar{m}_t}{d\theta} = \frac{1}{S} \sum_{s=1}^{S} \frac{dm_{st}}{d\theta} = \frac{1}{S} \sum_{s=1}^{S} \frac{c^s_{(2:N_t)t}}{\mathcal{G}^s_t} \frac{d\mathcal{F}^s_t}{d\theta}$$

with

$$\frac{d\mathcal{F}^s_t}{d\theta} = -\mathcal{F}^s_t \left[\frac{N_t}{\theta} - \frac{\sum_{i=1}^{N_t} c^s_{it}}{\theta^2} \right].$$

The second derivative of $Q(\theta)$ with respect to θ is

$$\frac{d^2Q(\theta)}{d\theta^2} = \sum_{t=1}^{T} -2\left[(w_t - \bar{m}_t)\frac{d^2\bar{m}_t}{d\theta^2} - \left(\frac{d\bar{m}_t}{d\theta}\right)^2 + \right.$$
$$\left. \frac{1}{S(S-1)}\sum_{s=1}^{S}\left\{(m_{st} - \bar{m}_t)\left(\frac{d^2m_{st}}{d\theta^2} - \frac{d^2\bar{m}_t}{d\theta^2}\right) + \left(\frac{dm_{st}}{d\theta} - \frac{d\bar{m}_t}{d\theta}\right)^2\right\}\right]$$

where

$$\frac{d^2m_{st}}{d\theta^2} = \frac{c^s_{(2:N_t)t}}{\mathcal{G}^s_t}\frac{d^2\mathcal{F}^s_t}{d\theta^2}$$

and

$$\frac{d^2\bar{m}_t}{d\theta^2} = \frac{1}{S}\sum_{s=1}^{S}\frac{d^2m_{st}}{d\theta^2} = \frac{1}{S}\sum_{s=1}^{S}\frac{c^s_{(2:N_t)t}}{\mathcal{G}^s_t}\frac{d^2\mathcal{F}^s_t}{d\theta^2}$$

with

$$\frac{d^2\mathcal{F}^s_t}{d\theta^2} = -\frac{d\mathcal{F}^s_t}{d\theta}\left[\frac{N_t}{\theta} - \frac{\sum_{i=1}^{N_t} c^s_{it}}{\theta^2}\right] + \mathcal{F}^s_t\left[\frac{N_t}{\theta^2} - \frac{2\sum_{i=1}^{N_t} c^s_{it}}{\theta^3}\right]$$
$$= \mathcal{F}^s_t\left\{\left[\frac{N_t}{\theta} - \frac{\sum_{i=1}^{N_t} c^s_{it}}{\theta^2}\right]^2 + \left[\frac{N_t}{\theta^2} - \frac{2\sum_{i=1}^{N_t} c^s_{it}}{\theta^3}\right]\right\}.$$

B. Bibliography

Donald, S. and H. Paarsch. "Maximum Likelihood Estimation in Empirical Models of Auctions." Mimeograph. Vancouver, Canada: Department of Economics, The University of British Columbia, 1991.

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

Fishman, G. and L. Moore. "A Statistical Evaluation of Multiplicative Congruential Random Number Generators with Modulus $2^{31} - 1$," *Journal of the American Statistical Association*, 77 (1982), 129–136.

Gouriéroux, C. and A. Monfort. "Simulation Based Inference in Models with Heterogeneity," *Annales d'Économie et de Statistique*, 20 (1990/1991), 69–107.

Hall, B., R. Schnake, and C. Cummins. *TSP Reference Manual, Version 4.1.* Palo Alto, CA: TSP International, 1988.

Jennrich, R.. "Asymptotic Properties of Non-Linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40 (1969), 633–643.

Knuth, D.. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms.* Reading, MA: Addison-Wesley, 1969.

Laffont, J.-J., H. Ossard, and Q. Vuong. "Econometrics of First-Price Auctions. " Document de Travail 07. Toulouse, France: Institut d'Economie Industrielle, Université des Sciences Sociale de Toulouse, 1991.

Lerman, S. and C. Manski. "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in *Structural Analysis of Discrete Data with Econometric Applications* edited by C. Manski and D. McFadden. Cambridge: MIT Press, 1981.

Marsaglia, G.. "Random Numbers Fall Mainly in the Planes," *Proceedings of the National Academy of Sciences of the United States of America*, 61 (1968), 25–28.

Marsaglia, G.. "The Structure of Linear Congruential Sequences," in *Applications of Number Theory to Numerical Analysis* edited by S. Zaremba. New York: Academic Press, 1972.

McAfee, R. and J. McMillan. "Auctions and Bidding." *Journal of Economic Literature*, XXV (1987), 699–738.

McFadden, D.. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57 (1989), 995–1026.

Milgrom, P.. "The Economics of Competitive Bidding: A Selective Survey" in *Social Goals and Social Organization: Essays in Memory of Elisha Pazner* edited by L. Hurwicz, D. Schmeidler, and H. Sonnenschein. Cambridge: Cambridge University Press, 1985.

Milgrom, P.. "Auction Theory" in *Advances in Economic Theory* edited by T. Bewley. Cambridge: Cambridge University Press, 1987.

Milgrom, P. and R. Weber. "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50 (1982), 1089–1122.

Paarsch, H.. "Empirical Models of Auctions within the Independent Private Values Paradigm and an Application to British Columbian Timber Sales." Discussion Paper 89–14. Vancouver, Canada: Department of Economics, The University of British Columbia, 1989.

Paarsch, H.. "Empirical Models of Auctions and an Application to British Columbian Timber Sales." Discussion Paper 91–19. Vancouver, Canada: Department of Economics, The University of British Columbia, 1991.

Paarsch, H.. "Deciding between the Common and Private Value Paradigms in Empirical Models of Auctions." *Journal of Econometrics*, 51 (1992), 191–215.

Paarsch, H.. "Estimating Structural Econometric Models of First-Price Sealed-Bid Auctions within the Affiliated Private Values Paradigm," in progress, (1992b).

Pakes, A. and D. Pollard. "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57 (1989), 1027–1057.

Table 1.
Comparison of ML, NLS, and SNLS Estimators.
N = 5; S = 25, 50, 100; T = 50, 100, 200.

Estimator	Mean	St.Dev.	L.Q.	Median	U.Q.
Sample Size = 50					
$\hat{\theta}^{mle}$	1.004	0.004	1.005	1.011	1.022
$\hat{\theta}^{nls}$	1.000	0.047	0.956	0.997	1.041
$\hat{\theta}^{snls}_{25}$ (0.200)	2.924	1.055	2.888	3.291	3.469
$\hat{\theta}^{snls}_{50}$ (0.303)	2.901	1.301	1.082	3.387	3.849
$\hat{\theta}^{snls}_{100}$ (0.537)	2.435	1.548	1.016	1.139	3.896
Sample Size = 100					
$\hat{\theta}^{mle}$	1.002	0.002	1.002	1.006	1.011
$\hat{\theta}^{nls}$	1.000	0.033	0.969	0.999	1.029
$\hat{\theta}^{snls}_{25}$ (0.197)	3.054	1.100	3.175	3.366	3.474
$\hat{\theta}^{snls}_{50}$ (0.301)	2.967	1.310	1.076	3.624	3.861
$\hat{\theta}^{snls}_{100}$ (0.595)	2.360	1.631	1.009	1.075	4.214
Sample Size = 200					
$\hat{\theta}^{mle}$	1.001	0.001	1.001	1.003	1.006
$\hat{\theta}^{nls}$	1.000	0.024	0.978	0.999	1.021
$\hat{\theta}^{snls}_{25}$ (0.123)	3.427	0.983	3.350	3.455	4.097
$\hat{\theta}^{snls}_{50}$ (0.414)	2.765	1.498	1.027	3.697	3.877
$\hat{\theta}^{snls}_{100}$ (0.681)	2.082	1.554	1.005	1.046	4.158

Figure 1: Non-Linear Least Squares and Simulated Non-Linear Least Squares Objective Functions, S = 5, T = 10.

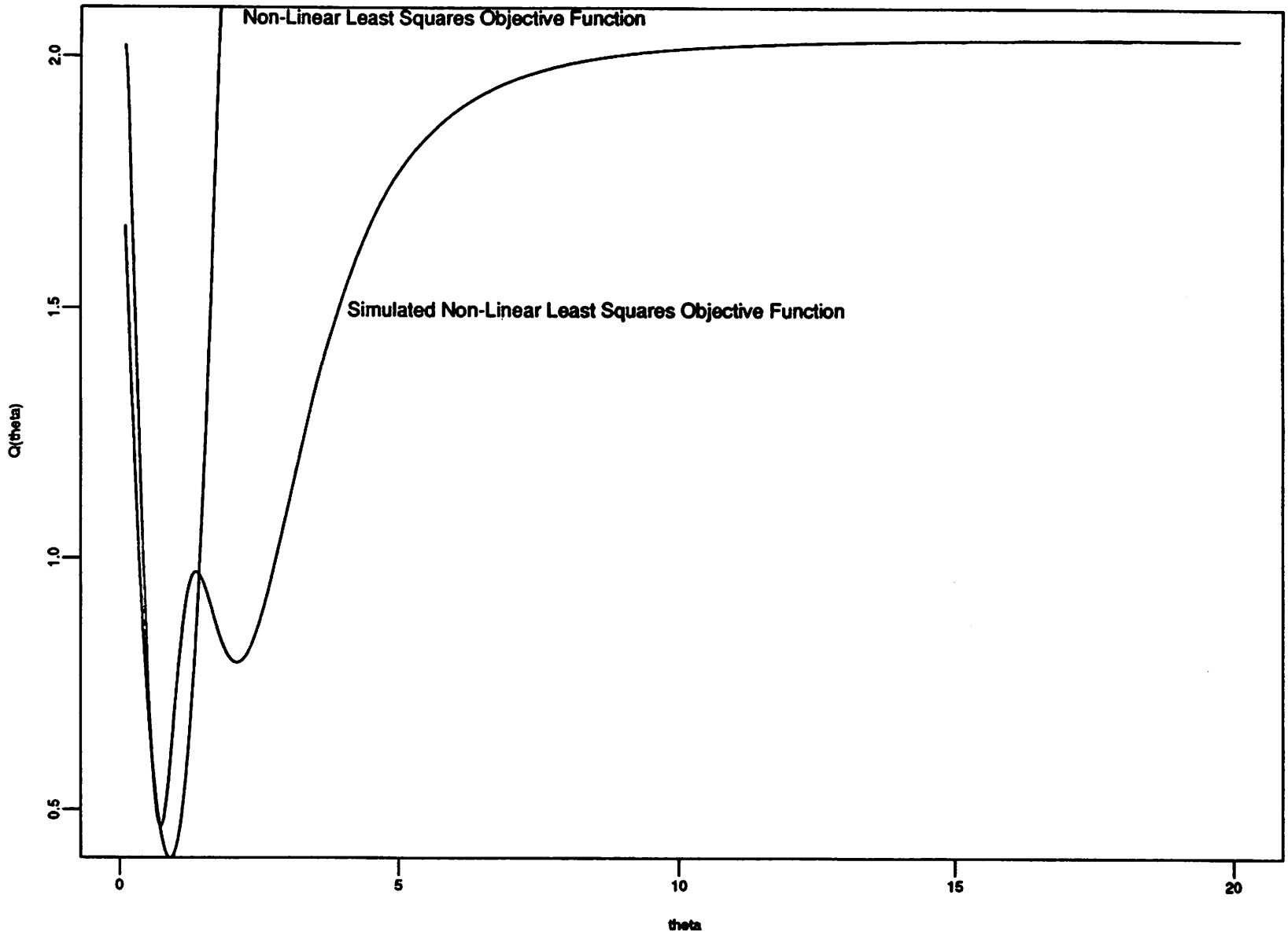

Figure 2: Non-Linear Least Squares and Simulated Non-Linear Least Squares Objective Functions, S = 25, T = 50.

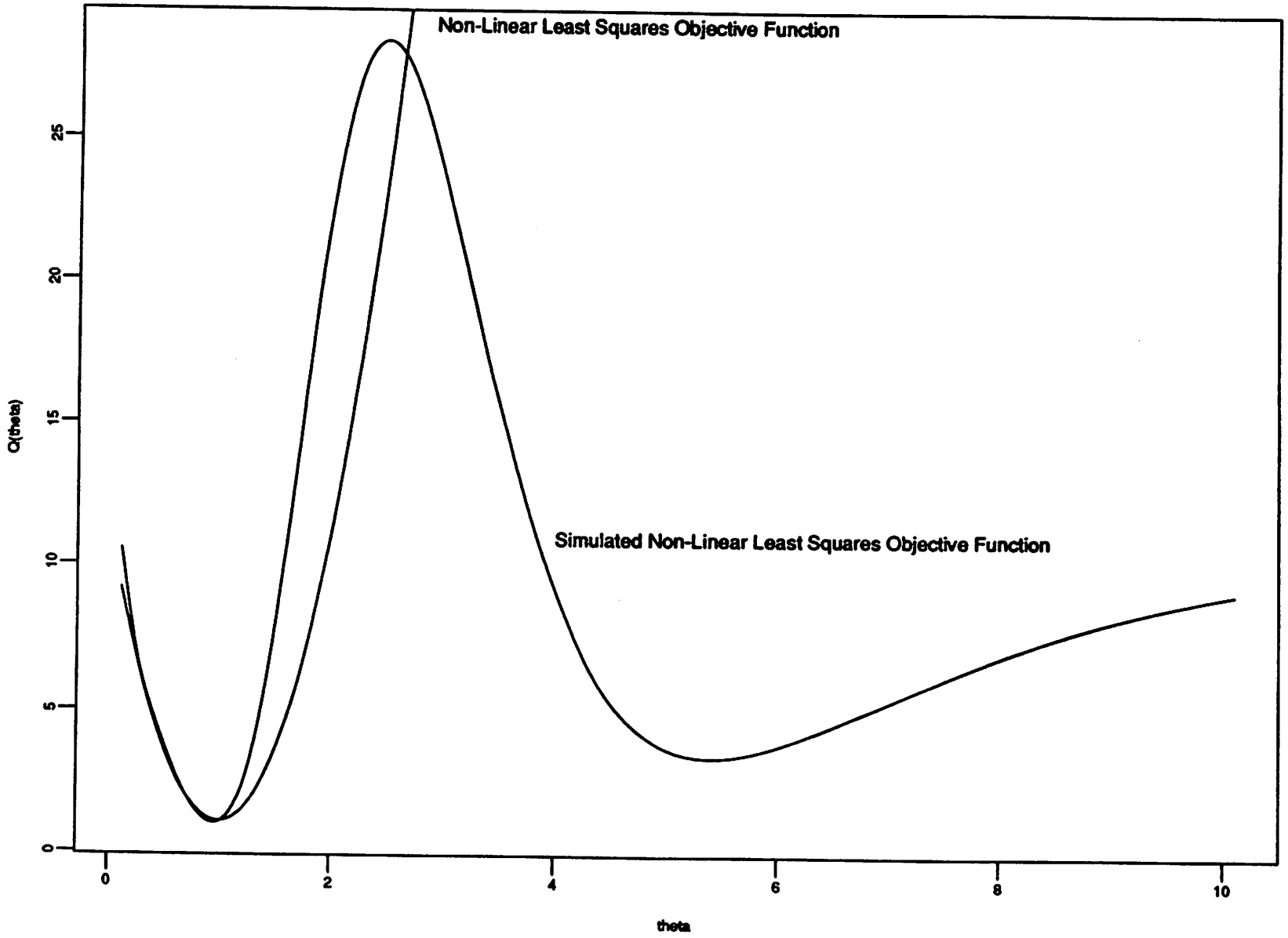

Figure 3: Non-Linear Least Squares and Simulated Non-Linear Least Squares Objective Functions, S = 200, T = 200.

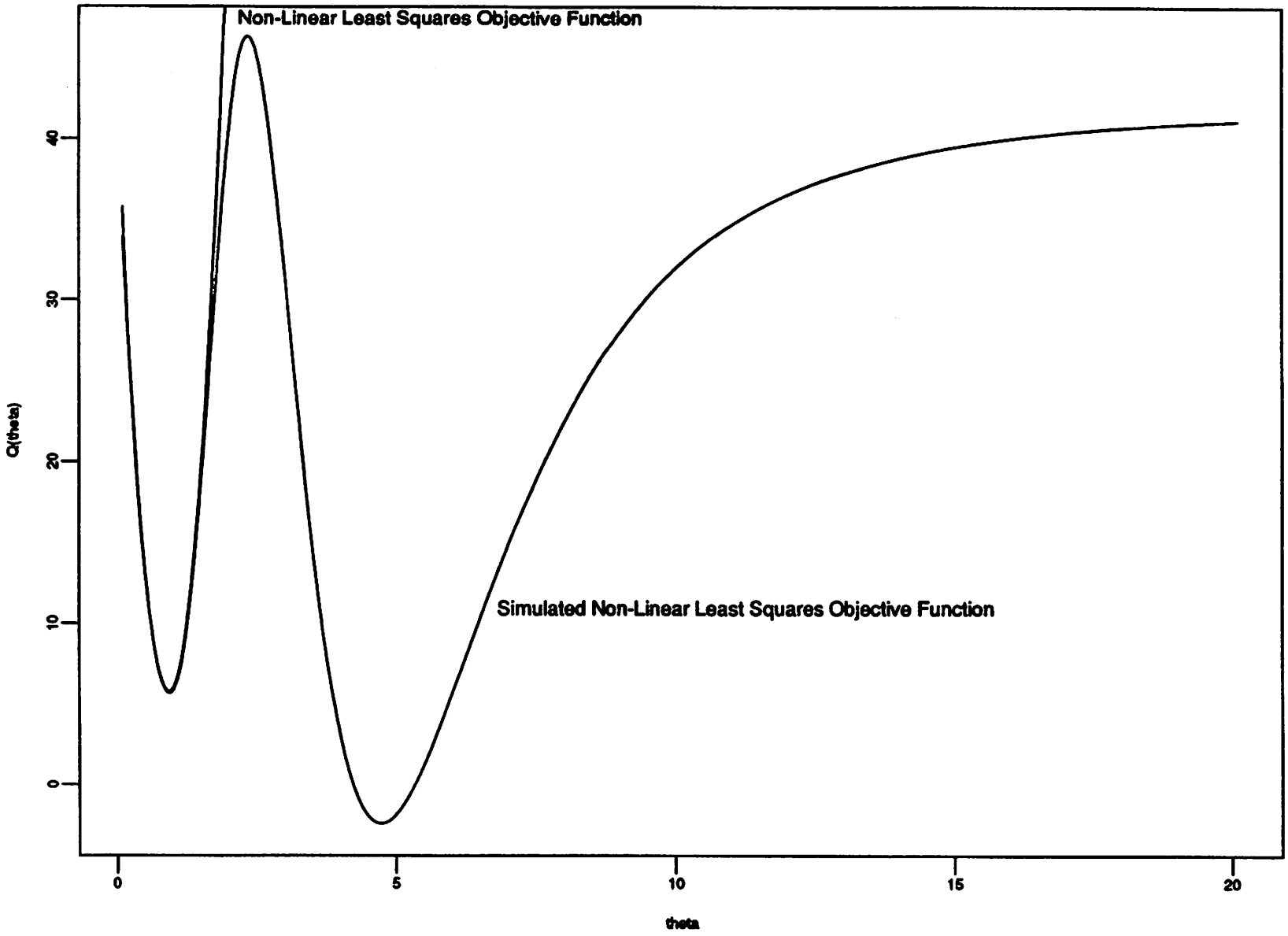

Figure 4: Density Functions and Histograms of the Estimators: S = 25, 50, 100, 200; T = 50, 100, 200.

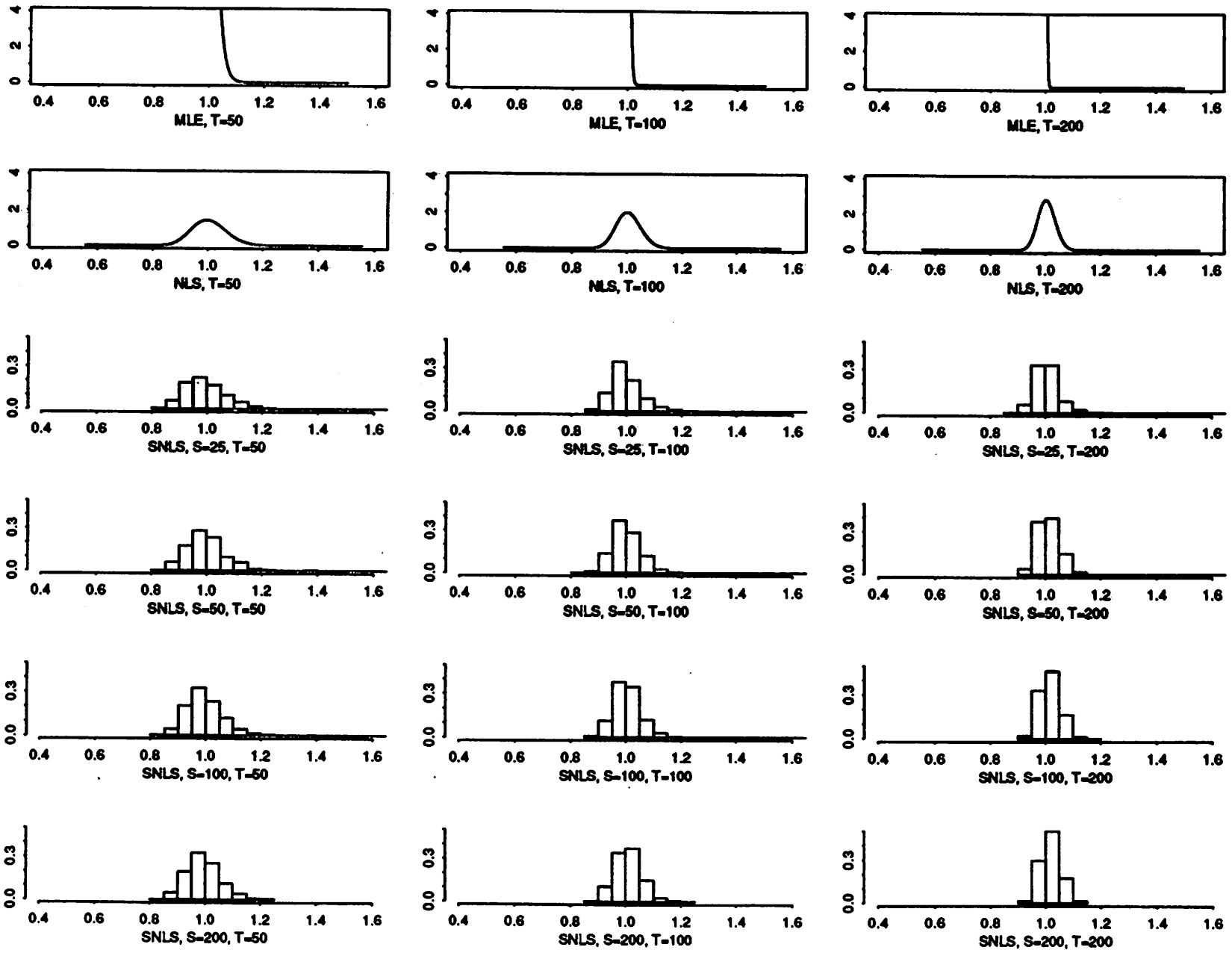
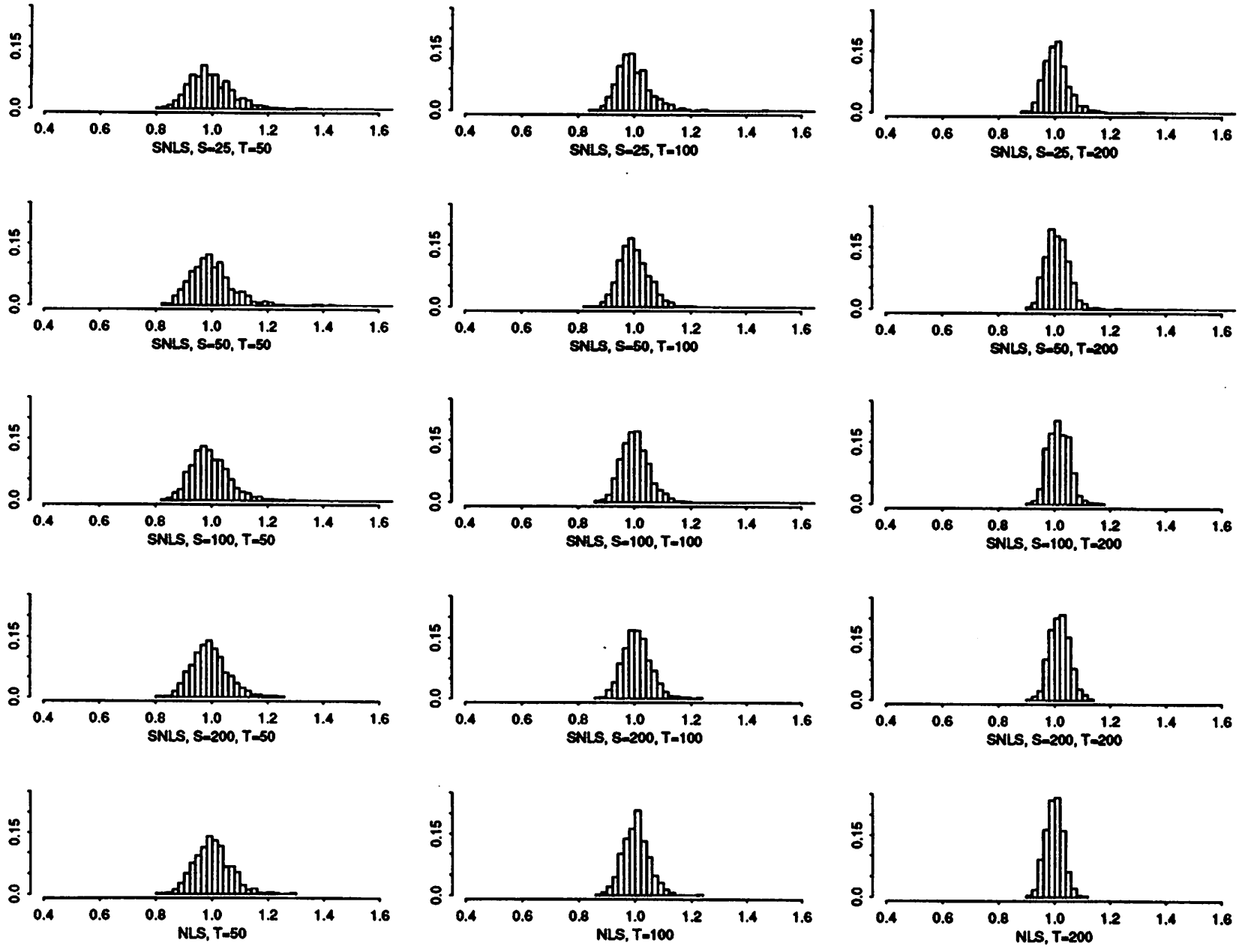
Figure 5: Histograms of the Estimators: S = 25, 50, 100, 200; T = 50, 100, 200.