1992

Maximum Likelihood Estimation in Empirical Models of Auctions

Stephen G. Donald

Harry J. Paarsch

ISSN:0318-725X
ISBN:0-7714-1442-0

RESEARCH REPORT 9211

Maximum Likelihood Estimation in Empirical Models of Auctions

by

Stephen G. Donald
and
Harry J. Paarsch

October 1992

Department of Economics

Social Science Centre

University of Western Ontario

London, Ontario, Canada

N6A 5C2

MAXIMUM LIKELIHOOD ESTIMATION IN EMPIRICAL MODELS OF AUCTIONS*

by

Stephen G. Donald†

and

Harry J. Paarsch‡

November 1991
This Version: September 1992

* Paarsch's work on this project was funded by the SSHRC of Canada. We are grateful to John Cragg and Erwin Diewert for helpful comments and useful suggestions.

† Assistant Professor, Department of Economics, College of Business Administration, University of Florida, Gainesville, Florida, USA 32611.

‡ Assistant Professor, Department of Economics, University of Western Ontario, London, Ontario, CANADA N6A 5C2.

Maximum Likelihood Estimation in Empirical Models of Auctions

Stephen G. Donald

University of Florida, Gainesville, Florida, USA 32611

Harry J. Paarsch

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

In applications of game theory to auctions, researchers assume that players choose strategies based upon a commonly known distribution of the latent characteristics. Rational behaviour, within an assumed class of distributions for the latent process, imposes testable restrictions upon the data generating process of the equilibrium strategies. Unfortunately, the support of the distribution of equilibrium strategies often depends upon all of the parameters of the distribution of the latent characteristics, making the standard application of maximum likelihood estimation procedures inappropriate. We present the maximum likelihood estimator as well as the conditions for its consistency and its asymptotic distribution.

1. Introduction

In empirical applications of game theory to auctions, researchers assume that the distribution of latent (or unobserved) characteristics is common knowledge to the players of the game. For example, in the independent private values model of an auction, the distribution of valuations is known to all bidders. Moreover, each bidder knows that his opponents know the distribution of valuations, and his opponents know that he knows, etc.. Based upon their knowledge of the distribution of latent characteristics, and given their realization from that valuation distribution, players are assumed to choose bids which maximize their expected pay-offs from winning the auction. Given this informational structure, the equilibrium of the game can be characterized by appealing to a particular concept of equilibrium (e.g., Bayesian-Nash).

The goal of some recent empirical research has been to determine if the predictions of game theory are consistent with observed data. For example, Paarsch (1989, 1991, 1992) has proposed the following research strategy. He has noted that the equilibrium strategies of players depend upon the distribution of latent characteristics, and 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 testable restrictions upon the data generating process of the equilibrium strategies. Within such a framework, however, the support of the distribution of the equilibrium strategies often depends upon all of the parameters of the distribution of latent characteristics, even when the support of the latent characteristic distribution depends upon no parameters. The implication of this result is that the standard application of maximum likelihood estimation procedures is inappropriate. In Donald and Paarsch (1993), we developed and evaluated a piecewise pseudo-maximum likelihood estimator which is useful in such situations. In this paper, we derive the maximum likelihood estimator, demonstrate its consistency, and derive its asymptotic distribution which in many cases is non-standard falling within the exponential family. Using Monte Carlo methods, we compare the small sample properties of the proposed estimator with those of the piecewise pseudo-maximum likelihood and non-linear least squares estimators.

2. Empirical Framework

To illustrate the particular class of estimation problems in which we are interested, we model a sealed-bid auction as a non-coöperative game.[1] We consider 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. We 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 $g(c)$ and the cumulative distribution function $G(c)$. Each player is assumed to know his own cost, but not those of his opponents. The costs of players are assumed to be independent draws from $G(c)$, and $G(c)$ is assumed to be common knowledge. We assume that bidders are risk neutral with respect to winning the auction, and that the i^{th} bidder chooses a bid b_i to maximize his expected profit. Finally, we 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

[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.

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

$$[1 - G(\beta^{-1}(b_i))]^m.$$

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 expected profit

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

The lower is b_i, the higher is his probability of winning the auction $[1 - G(\beta^{-1}(b_i))]^m$, 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

$$[1 - G(\beta^{-1}(b_i))]^m - m(b_i - c_i)g(\beta^{-1}(b_i)) \cdot [1 - G(\beta^{-1}(b_i))]^{m-1} \cdot \frac{d\beta^{-1}(b_i)}{db_i} = 0. \quad (2.1)$$

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:

$$\beta'(c)[1 - G(c)]^m - m\beta(c)g(c)[1 - G(c)]^{m-1} = -mcg(c)[1 - G(c)]^{m-1}. \quad (2.3)$$

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

$$\beta(c) = c + \frac{\int_c^\infty [1 - G(u)]^m \, du}{[1 - G(c)]^m}. \quad (2.4)$$

Denoting $c_{(i:n)}$ as the i^{th} 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] In fact, simply imposing $\beta(\infty) = \infty$ is insufficient to guarantee a unique solution since adding any constant a to that solution is also a solution. In this case, $a = 0$ is the appropriate constant.

2.2. Strategy for Interpreting Data

One strategy for interpreting field data (see Paarsch 1989, 1991, and 1992) involves exploiting the fact that (2.4) is a monotonic function of c; the lower is a player's cost, the less he will bid. Because the bidding rules are functions of the random variable c, the bids are also random variables and their densities are related to $g(c)$. For example, the density of $\beta(c)$ is

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

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

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

$$h(w) = \frac{\tilde{g}(\beta^{-1}(w))}{\beta'(\beta^{-1}(w))},$$

where

$$\tilde{g}(z) = n[1 - G(z)]^m g(z)$$

is the density of $z = c_{(1:n)}$.

Consider a family of distributions for c which depend upon the parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$. Without any loss of generality, let $g(c)$ have support upon the interval $[0, \infty)$. 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

$$\left[\int_0^\infty [1 - G(u;\theta)]^m \, du, \infty\right) \equiv [\Im(\theta; m), \infty).$$

$\Im(\theta, m)$ 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 of Wald (1949) used to demonstrate the consistency and asymptotic normality of the maximum likelihood estimator no longer apply.

An alternative strategy could be to abandon estimation by a method like maximum likelihood and to use some other procedure. For example, suppose that the j^{th} raw moment of w,

$$\mathrm{E}[w^j] = \mu_j(\theta, m) = \int_{\Im(\theta,m)}^{\infty} w^j h(w; \theta, m)\, dw \qquad j = 1, 2, \ldots,$$

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

$$w^j = \mu_j(\theta, m) + u_j \qquad j = 1, 2, \ldots,$$

where the expectation of u_j is zero, while its variance depends upon m. The parameter vector θ can then be estimated by non-linear least squares, for example.

A drawback of the non-linear least squares estimator of the parameter vector θ is that it may not be very efficient. Also, one would ideally like to have two estimators defined in different metrics, with different probability limits under alternative distributional assumptions, to provide a basis for specification testing. Thus, knowing how to implement the maximum likelihood estimator when the support depends upon all of the parameters of the distribution appears to be an important and interesting problem.[3]

2.3. Piecewise Pseudo-Maximum Likelihood Estimation

In Donald and Paarsch (1993), we developed a piecewise pseudo-maximum likelihood estimator. The basic idea behind that method of estimation is as follows: Consider a random sample of size T indexed by $t = 1, \ldots, T$. 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 we 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.

[3] Christensen and Kiefer (1991) have researched a similar class of problems which arise in the estimation of structural models of job search. We believe that our work is complementary to theirs.

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 we do not use a first-order condition to concentrate the likelihood function.

In Donald and Paarsch (1993), we demonstrated that under fairly general conditions the piecewise pseudo-maximum likelihood estimator is consistent. We were also able to demonstrate 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.

Despite the computational attractiveness of the piecewise pseudo-maximum likelihood method of estimation, estimating θ_1 remained an issue. In general, several different ways of estimating θ_1 exist. For example, for any m one could use

$$\hat{\theta}_1^m = \theta_1(\hat{\theta}_2^{\text{pple}}, \hat{\underline{w}}(m), m),$$

where $\hat{\theta}_2^{\text{pple}}$ is the piecewise pseudo-maximum likelihood estimator of θ_2. Alternatively, one could use a linear combination of the $\hat{\theta}_1^m$'s, or some order statistic of the $\hat{\theta}_1^m$'s. None of these estimators appeared preferable to the others, and many factors could influence ones choice; e.g., the nature of the lower bound function, the number of auctions with n bidders, etc.. Also, a degree of arbitrariness existed in partitioning the parameter space into (θ_1, θ_2). Several different ways of doing this could exist, and each would lead to different estimates depending upon the partition chosen. In Donald and Paarsch (1993), we attempted to shed some light on these issues by examining the performance of the different estimators using Monte Carlo methods.

Another 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. But when auctions have reserve prices, $\Im(\theta, m)$ is often only defined numerically because it involves the truncated moments of the latent cost (or valuation) distribution, see Paarsch (1991).

Yet another drawback of the method is that introducing continuous covariates into the empirical framework often relied upon special structure in the problem at hand, and was sometimes 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., $G(c;\theta) = G(c;\theta, Z_t)$. 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 (1993) with only trivial modifications to estimate the parameter vector $\alpha = (\theta, \phi)$. If, however, the elements of the Z_t are continuous covariates, such as prices and quantities, then the methods of Donald and Paarsch (1993) cannot be applied directly. One could consider partitioning the covariate space into cells, and then concentrate the likelihood function using the smallest w_t for a particular cell, but if the elements of Z_t are diffusely distributed and if the dimension of Z_t is only moderately large (e.g., three or four), then the number of cells will be huge, so the sample sizes required to get reasonably accurate estimates of $\Im(\alpha, m, Z)$ will be prohibitively large.

In this paper, we derive the maximum likelihood estimator of the parameter vector α using non-linear programming. Within this framework, we can admit $\Im(\alpha, m, Z)$'s which need only be defined numerically, and the presence of continuous covariates. Subsequently, we demonstrate the consistency of this estimator in the case of continuous covariates, and derive its asymptotic distribution when the covariates take on discrete values. The latter problem is interesting in its own right, since the asymptotic theory must be approached in a different way. Using Monte Carlo methods, we compare the small sample properties of the maximum likelihood estimator with those of the piecewise pseudo-maximum likelihood and non-linear least squares estimators.

3. Maximum Likelihood Estimator

We motivate the solution to our problem 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 $[0, \alpha]$, where α is an unknown parameter which the investigator seeks to estimate. The density of w is

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

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

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

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

$$\log L(\alpha; w_1, w_2, \ldots, w_T) = -T \log \alpha.$$

Of course, as α tends to zero the functions $L(\alpha)$ and $\log L(\alpha)$ tend to infinity. An α of zero, however, implies that none of the data should have been observed in the first place. Thus, the optimization problem must be re-written. The simplest way to do this is as a constrained optimization problem. In particular, maximize $L(\alpha)$ or $\log L(\alpha)$ subject to the constraint that all of the observed data be consistent with the resulting estimate. That is,

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

The maximum likelihood estimator is then

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

Abstracting from ties in the data, $T-1$ of the constraints do not bind. Also, the conventional methods used to determine the asymptotic distribution of $\hat{\alpha}$ do not apply. In particular, only the largest w is important in determining the distribution of $\hat{\alpha}$. We shall discuss the problems which arise in performing the asymptotic analysis in detail below. Suffice to say here that we shall define the maximum likelihood estimator of the vector α in similar fashion.

3.1. Optimization Problem

We denote the density of the t^{th} $(t = 1, \ldots, T)$ observation w_t conditional upon m_t and Z_t by

$$h(w_t; \alpha, m_t, Z_t) \qquad 0 < \Im(\alpha, m_t, Z_t) \leq w_t$$

and introduce the following assumption:

Assumption 1.

The true lower bound of the support of the distribution of w satisfies

$$0 < \inf\ \Im(\alpha^0, x) < \sup\ \Im(\alpha^0, x) < \infty$$

where $x = (m, Z)$ and α^0 denotes the true value of α, and where the inf and sup are taken over all values of $x \in X$.

Lemma 1:

Given assumption 1,

$$\lim_{w \to \Im(\alpha^0, x)} h(w; \alpha^0, x) = \frac{n}{m} \frac{1}{\Im(\alpha^0, x)} > 0$$

This result (the proof of which, like all of our results, is in an appendix) shows that the distribution of the winning bid has a strictly positive density at its lower bound. This fact will be important to the proof of Lemma 2 below. There we show that the smallest order statistic for each possible value of x consistently estimates the lower bound for each possible value of x. Moreover, these order statistics are consistent at rate T. This result is useful since later we show that the maximum likelihood estimator in these models depends upon order statistics. In some cases, the maximum likelihood estimator is obtained by solving for the parameters purely as functions of the order statistics. In such cases, the maximum likelihood estimator will also be consistent a rate T. The particular limiting distribution that results will, however, depend upon the number of possible values of x as well as the number of parameters.

In defining the maximum likelihood estimator, we introduce the following two assumptions:

Assumption 2.

The logarithm of the likelihood function is twice continuously differentiable in α.

Assumption 3.

The $w_t = \Im(\alpha, x_t)$ functions are quasi-convex and twice continuously differentiable in α.

We define the maximum likelihood estimator $\hat{\alpha}$ as being the solution to the following optimization problem:

$$\max_{<\alpha>} \sum_{t=1}^{T} \log h(w_t; \alpha, x_t) \quad \text{subject to} \quad \begin{cases} \Im(\alpha, x_1) \leq w_1 \\ \Im(\alpha, x_2) \leq w_2 \\ \vdots \\ \Im(\alpha, x_T) \leq w_T. \end{cases}$$

One can solve for the maximum likelihood estimator by maximizing the following Lagrangean:

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

with respect to the vector α, where $\lambda = (\lambda_1, \ldots, \lambda_T)$ is the vector of T Lagrange multipliers. The maximum likelihood estimator $\hat{\alpha}$ satisfies the following conditions:[4]

$$\begin{aligned} \sum_{t=1}^{T} \left(\nabla_\alpha \log h(w_t; \hat{\alpha}, x_t) - \lambda_t \nabla_\alpha \Im(\hat{\alpha}, x_t) \right) &= 0 \\ \lambda_1 \big(w_1 - \Im(\hat{\alpha}, x_1) \big) &= 0 \\ \lambda_2 \big(w_2 - \Im(\hat{\alpha}, x_2) \big) &= 0 \\ &\vdots \\ \lambda_T \big(w_T - \Im(\hat{\alpha}, x_T) \big) &= 0, \end{aligned}$$

[4] Under the stated conditions, these Kuhn-Tucker conditions are necessary, but insufficient for a global maximum. If the logarithm of the likelihood function is pseudo-concave, then it is well-known (see Mangasarian [1969]) that the Kuhn-Tucker conditions are both necessary and sufficient.

where ∇_α denotes the gradient vector of the function to follow with respect to the vector α. At most, $r = p + q$ of the T constraints will ever bind at one time; i.e., $T - r$ of the Lagrange multipliers will be zero at the optimum. For the binding constraints, the Lagrange multipliers will be non-negative.

Assumption 4.

Given any sequence of r w_t's

$$(\overbrace{w_i, w_j, \ldots, w_k}^{r \text{ times}})$$

which have different covariates, the Jacobian of

$$\begin{aligned} w_i &= \Im(\alpha, x_i) \\ w_j &= \Im(\alpha, x_j) \\ &\vdots \\ w_k &= \Im(\alpha, x_k) \end{aligned} \tag{3.1}$$

is non-singular.

3.2. Consistency of the Estimator

Define the true feasible set of α to satisfy the lower bound constraint for all possible values of $x \in X$ by

$$A^* = \{\alpha \mid \Im(\alpha, x) \leq \Im(\alpha^0, x) \; \forall \; x \in X\}.$$

Note that this set is convex (given Assumption 3), and non-empty since $\alpha^0 \in A^*$. In addition to assumptions 1 to 4, we make the following assumptions:

Assumption 5.

x_t contains discrete variables denoted by the vector x_{1t} which have finite support on X_1 and continuous variables denoted by the vector x_{2t} which have compact support on X_2. Assume that the x_t are independent and identically distributed random variables with probability measure P such that for every nonempty subset A_1 of X_1 and every nonempty open subset of A_2 of X_2, $P(A_1 \times A_2) > 0$.

Denote the true conditional density of w given x as

$$h(w; \alpha^0, x) \text{ on } [\Im(\alpha^0, x), \infty).$$

Assumption 6.

For any $\varepsilon > 0$

$$\inf_{x \in X} \int_{\Im(\alpha^0,x)}^{\Im(\alpha^0,x)+\varepsilon} h(w; \alpha^0, x)\, dw = \delta(\varepsilon) > 0.$$

This assumption simply requires that some lower bound exist upon the probability of w being within ε of the lower bound of its support.

Introduce

$$S_T(\alpha) = \frac{1}{T} \sum_{t=1}^{T} \log h(w_t; \alpha, x_t)$$

and denote the set of feasible values of α by

$$A_T^* = \{\alpha \mid \Im(\alpha, x_t) \le w_t \;\forall\, t = 1, \ldots, T\}.$$

It will be convenient to write the maximum likelihood estimation procedure as solving

$$\max_{<\alpha>} S_T(\alpha) \quad \text{subject to} \quad \alpha \in \bar{A} \cap A_T^*,$$

where $\bar{A}$ is some convex compact set containing α^0. Note also that $\alpha^0 \in A_T^*$ for all T and that $A^* \subset A_T^*$. The proof of consistency of the maximum likelihood estimator $\hat{\alpha}$ will rely on the set A_T^* becoming closer and closer to A^* as T becomes large. The sense in which we say that these two sets are close is given below in Definition 1.

Definition 1.

Define a measure of the difference between A^* and A_T^* where $A^* \subset A_T^*$ as

$$d(A^*, A_T^*) = \sup_{\alpha^1 \in A_T^* \setminus A^*} \; \inf_{\alpha^2 \in A^*} \|\alpha^1 - \alpha^2\|,$$

where $\|\cdot\|$ denotes the Euclidean norm. Also, we say that $A_T^* \xrightarrow{\text{a.s.}} A^*$ if $d(A^*, A_T^*) \xrightarrow{\text{a.s.}} 0$.

The probability measure referred to in the second part of the definition relates to that of the (w_t, x_t) data which determines the set A_T^*, so the interpretation is related to the probability that a sequence of such data generate a sequence of sets A_T^* that become close to the set A^* in the metric defined above. The following theorem shows that this is true.

Theorem 1.

Given assumptions 1 to 6, $A_T^* \cap \bar{A} \xrightarrow{\text{a.s.}} A^* \cap \bar{A}$.

One should note that the proof is by construction. In the model considered in Donald and Paarsch (1993), the result is easy to show since in that case the covariate is the number of opponents m, and we have that for each value of m, $\underline{\hat{w}}(m) \xrightarrow{\text{a.s.}} \Im(\theta^0, m)$. Hence, given that $\Im(\cdot)$ satisfies the conditions of the implicit function theorem, in the limit, the only θ's that satisfy the lower bound constraint are those that satsify $\Im(\theta, m) \leq \Im(\theta^0, m)$ for each m. When we allow for continuous covariates, things become more complicated, although the proof is quite similar using order statistics defined over subsets of the X set.

The following result gives the general consistency result for the maximum likelihood estimator. Note that one of the assumptions is the result proven in Theorem 1.

Theorem 2.

Given assumptions 1 to 6, and

a) $\bar{A}$ is compact;

b) $A_T^* \cap \bar{A} \xrightarrow{\text{a.s.}} A^* \cap \bar{A}$;

c) $S_T(\alpha) \xrightarrow{\text{a.s.}} S(\alpha)$ uniformly over $\bar{A}$, where $S(\alpha)$ is continuous in α;

d) if $S(\alpha) \geq S(\alpha^0)$ for any $\alpha \in \bar{A} \cap A^*$, then $\alpha = \alpha^0$;

$\hat{\alpha} \xrightarrow{\text{a.s.}} \alpha^0$.

The main difference between this proof and others in the literature is that the parameter set is data dependent. Results presented by White and Wooldridge (1991), concerning sieve estimation, are similar in that the parameter set changes, and may be data dependent. The analogue of b) in their case is the requirement that the parameter set become dense in the true parameter set as $T \to \infty$. In that case, the number of terms in the sieve is increasing with the sample size. Here, constraints upon the parameter set become increasingly strong as T becomes large; more constraints are being added with each data point until the set is reduced to the parameters that satisfy the lower bound of the support restriction for all possible values of x.

3.3. Asymptotic Distribution of the Estimator

A natural way to calculate the variance-covariance matrix of $\hat{\alpha}$ would be to consider the behaviour of the Hessian matrix of the Lagrangean

$$\nabla_{\alpha\alpha}\, \mathcal{L}(\hat{\alpha}) = \sum_{t=1}^{T} \left(\nabla_{\alpha\alpha}\, \log h(w_t; \hat{\alpha}, x_t) - \hat{\lambda}_t \nabla_{\alpha\alpha} \Im(\hat{\alpha}, x_t)\right).$$

This is useful when the solution to the optimization problem occurs along a smooth and differentiable part of the contraint set, but typically the solution 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. To see this, consider the simple problem introduced in the first part of this section. There, the properties of the maximum likelihood estimator $\hat{\alpha}$ were solely determined by the behaviour of the largest w in a sample of size T. In this case, the properties will often be determined by the solution to (3.1) for some set of the smallest r order statistics of w_t given x_t.

As may be expected from the previous discussion, the distribution theory for the estimator can be quite complicated. Because of technical difficulties, we have only analyzed the case of discrete covariates. Hence,

Assumption 7.

x is a discrete random vector with probability mass function $\pi(x)$, with k being the number of points that have $\pi(x) > 0$.

Denote each possible point in the set by $x(i)$, and let $\pi_i = \pi(x(i))$ for $i = 1, ..., k$.

Despite the assumption of discrete covariates, the results are of considerable interest. Indeed, as the following discussion will show, the limiting distributions of the estimators will only fall into the usual normal limiting family in a special case. Allowing for only a slightly more general case will change the nature of the limiting distributions, and introduce the exponential as a limiting distribution.

With discrete covariates, at least three possibilities exist depending upon the relationship betweeen k the number of points with positive support for $\pi(x)$ and r the number of parameters to be estimated. This complication arises because the optimization problem in the discrete case, which we shall call (S), may be written as

$$\max_{<\alpha>} \sum_{i=1}^{k} \hat{\pi}_i S_T(\alpha, x(i)) \quad \text{subject to} \quad \Im(\alpha, x(i)) \leq \underline{\hat{w}}(x(i)) \quad i = 1, \ldots, k$$

where $\underline{\hat{w}}(x(i)) = \min\{w_t : x_t = x(i)\}$ is the smallest order statistic of w_t over all observations that have $x_t = x(i)$, $\hat{\pi}_i = T_i/T$ is the proportion of the sample with $x_t = x(i)$, and

$$S_T(\alpha, x(i)) = \frac{1}{T_i} \sum_{t=1}^{T} \log h(\alpha; w_t, x(i)) I[x_t = x(i)]$$

is the average of the contributions to the logarithm of the likelihood function contribution of observations with $x_t = x(i)$. The fact that the problem involves order statistics will lead to the unusual limiting distributions that appear in the results. The following result, which is proven in Donald and Paarsch (1993), will be used throughout this section and concerns the limiting behaviour of the order statistics. Let T_i denote the number of observations that have $x_t = x(i)$, and note that assumption 7 guarantees that $T_i = O_{\text{a.s.}}(T)$ by the law of large numbers.

Lemma 2.

Under assumptions 1 to 7 and assuming that k is finite

(i) $\underline{\hat{w}}(x(i)) \to \Im(\alpha^0, x(i))$ almost surely;

(ii) $T^{1-\eta}(\underline{\hat{w}}(x(i)) - \Im(\alpha^0, x(i))) \to 0$ for any $\eta > 0$.

That is, the smallest order statistics converge to the lower bound at rate T in this case. The next result, which is contained in Galambos (1978) and discussed in Reiss (1989), gives the limiting distribution of these order statistics, which is related to the Weibull distribution.

Lemma 3.

Suppose that the $\{w_t\}_{t=1}^T$ are drawn randomly from a population with probability density function f and and cumulative distribution function F, on $[\theta, \infty)$ such that for all $z < 0$,

$$\lim_{t \to 0_-} \frac{F(\theta + zt)}{F(\theta - t)} = (-z)^\gamma$$

and $\theta > -\infty$, then

$$\frac{1}{d_T^*}(\min_t\{w_t\} - \theta) \to \mathcal{W}(1, \gamma)$$

where $\mathcal{W}(1, \gamma)$ denotes a random variable that is distributed Weibull with parameters 1 and γ, and $d_T^* = F^{-1}(\frac{1}{T}) - \theta$.

This Lemma provides conditions under which the limiting distribution of the smallest order statistic is

$$[1 - \exp(-z^\gamma)].$$

Note that it is only defined for positive values of z. This gives the well-known fact that extreme order statistics are biased estimators of the lower bound of the distribution, although they generally converge very quickly as shown in the previous result. Note also that if we can find alternative constants d_T such that $d_T^*/d_T \to 1$, then the result will still hold. The γ parameter will depend upon the behaviour of the density function near the lower bound of the support. There may be other types of limiting distributions of smallest order statistics (depending upon the nature of the parent population), but this result is sufficient to characterize the limiting distributions of order statistics in the auction case, because, as the following result shows, the distributions for auctions will satisfy the condition in Lemma 3 with $\gamma = 1$. That is, the limiting distribution of the smallest winning bid for each possible covariate value will be exponential with intensity parameter equal to one. In addition, a convenient form for the normalizing constant can always be found.

Corollary 1.

Under Assumptions 1 to 7,

$$\frac{1}{d_T}(\underline{\hat{w}}(x(i)) - \Im(x(i), \alpha^0)) \xrightarrow{\mathrm{d}} \mathcal{W}(1, 1)$$

where $\mathcal{W}(1,1)$ is an exponential random variable with parameter 1, denoted $\mathcal{E}(1)$, and

$$d_T = \frac{m\Im(\alpha^0, x(i))}{nT_i} = O_p(T^{-1}).$$

The limiting exponential result arises because, as shown in Lemma 1, in the auction models under consideration, the density function is strictly positive at its lower bound. In other situations, where the density does not converge to a strictly positive bounded number at the lower bound of the support, the γ parameter and the rate of convergence will depend upon the behaviour at the lower bound. If the density function converges to zero at the lower bound, then typically the rate of convergence may be slower than rate T, and if the density is unbounded at the lower bound of the support (as occurs with certain Weibull populations with declining hazard rates), then the convergence rate may be even faster than rate T.

To make these notions concrete, consider the following example which has no covariates. Suppose that w is distributed Pareto with parameters α_1 and α_2, so

$$h(w;\alpha_1,\alpha_2) = \frac{\alpha_2\alpha_1^{\alpha_2}}{w^{\alpha_2+1}} \quad 0<\alpha_1<w \text{ and } 0<\alpha_2,$$

then for a sample of size T

$$\frac{\alpha_2 T}{\alpha_1}(\min w_t - \alpha_1) \xrightarrow{\mathrm{d}} \mathcal{E}(1)$$

is distributed exponentially in the limit with parameter 1 because the Pareto density function is strictly positive at the lower bound of the distribution.[5]

Letting the subscript 0 on the function denote population values, introduce the following notation

$$S_0(\alpha, x(i)) = \mathrm{E}_0[\log h(w;\alpha^0, x(i))]$$

where E_0 denotes that the expectation is taken at the true parameter values α^0. Also, define the following population optimization problem (P_i):

$$\max_{<\alpha>} S_0(\alpha, x(i)) \quad \text{subject to} \quad \Im(\alpha, x(i)) \leq \Im(\alpha^0, x(i))$$

[5] The reader will remark that the normalization employed in this example is not the one implied by Lemma 3. We use it because it is equivalent in the limit and has a convenient form, as do the constants d_T in Corollary 1 for the auction case.

as well as the aggregate problem (P)

$$\max_{<\alpha>} \sum_{i=1}^{k} \pi_i S_0(\alpha, x(i)) \quad \text{subject to} \quad \Im(\alpha, x(i)) \leq \Im(\alpha^0, x(i)) \quad i = 1, \ldots, k.$$

We shall assume throughout that we can interchange integration and differentiation. We also introduce the following assumption regarding the problem P_i.

Assumption 8.

The solution to P_i is (α^0, λ_i^0) with $\lambda_i^0 > 0$, so that the constraint binds.

Typically, we find that the expectation of the gradient vector $\nabla_\alpha S_0$ has at least one strictly positive element, with remainder being 0 when evaluated at the true parameters α^0.

Continuing with the Pareto example considered above, one can demonstrate easily that assumption 8 is satisfied with $\lambda^0 = \alpha_2^0/\alpha_1^0 > 0$.

In the remainder of this section we analyze two different cases.

Case 1: $k \leq r$

Partition the vector α into (α_1, α_2) of dimensions k and $(r-k)$ respectively, in such a way that the $k \times k$ matrix whose i^{th} column is

$$\nabla_{\alpha_1} \Im(\alpha, x(i))$$

is non-singular over a neighbourhood of α^0. Of course, when $k = r$ the α_2 component is non-existent. The next result shows that $\hat{\alpha}$ the solution to the sample maximization problem defined above, with probability one, will satisfy the Kuhn-Tucker conditions, with all k constraints binding as T tends to ∞.

Theorem 3.

Given assumptions 1 to 8, and assuming that

$$\nabla_\alpha S_T(\alpha, x(i)) \xrightarrow{\text{a.s.}} \nabla_\alpha S_0(\alpha, x(i))$$

uniformly over a neighbourhood of α^0 for each i, then for large enough T all k constraints bind at the solution $\hat{\alpha}$ with probability 1.

The fact that all k constraints bind make it possible to invert out a subset of parameters α_1 as a function of the remaining parameters by Assumption 4 and the implicit function theorem. The resulting solution will be twice continuously differentiable in both α_2 and the remaining arguments in a neighbourhood of α_2^0. Here, the implicit function will be denoted as

$$\alpha_1 = \psi(\alpha_2, \underline{\hat{w}}, x),$$

the solution to the set of equations

$$\underline{\hat{w}}(x(i)) = \Im(\alpha, x(i)) \quad i = 1, \ldots, k$$

where we introduce the shorthand $\underline{\hat{w}}$ to denote the k vector of $\underline{\hat{w}}(x(i))$'s and x to denote the vector of $x(i)$. Also, note that

$$\alpha_1^0 = \psi(\alpha_2^0, \Im^0, x)$$

where $\Im^0$ denotes the k vector of values of the lower bounds. When $k = r$, one can solve for α_1 just using the constraints, so that it can be written as a function of only the $\underline{\hat{w}}(x(i))$'s, and its distribution will depend on the distributions of the $\underline{\hat{w}}(x(i))$'s. In this case, this condition will give rise to limiting distributions related to those in Corollary 1.

We next introduce the following notation which will be useful in characterizing the results when $k < r$. Define

$$\mathrm{V}[\nabla_\alpha \log h(w; \alpha^0, x(i)] = \begin{pmatrix} \Omega_1^i & \Omega_{12}^i \\ \Omega_{21}^i & \Omega_2^i \end{pmatrix}$$

for each i where the partition is conformable with that of α. In the case where $k < r$, standard mean value expansions will be used to find the limiting distribution. The terms involved will be of the form

$$d_t = \nabla'_{\alpha_2}\psi(\alpha_2)\nabla_{\alpha_1} \log h(w_t, \alpha, x_t) + \nabla_{\alpha_2} \log h(w_t, \alpha, x_t).$$

Note that assumption 8 implies that $\mathrm{E}[d_t] = 0$ for each t. Define for $x_t = x(i)$,

$$\mathrm{V}[d_i] = \nabla'_{\alpha_2}\psi(\alpha_2)\Omega_1^i\nabla_{\alpha_2}\psi(\alpha_2) + \Omega_2^i + \nabla'_{\alpha_2}\psi(\alpha_2)\Omega_{12}^i + \Omega_{21}^i\nabla_{\alpha_2}\psi(\alpha_2)$$

Theorem 4.

Under assumptions 1 to 8, and assuming that $k < r$

$$\sqrt{T}(\hat{\alpha}_1 - \alpha_1^0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_1)$$

and

$$\sqrt{T}(\hat{\alpha}_2 - \alpha_2^0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_2)$$

where

$$V_1 = \sum_{i=1}^{k} \pi_i \mathrm{V}[d_i]$$

and

$$V_2 = \nabla'_{\alpha_2}\psi(\alpha_2)V_1\nabla_{\alpha_2}\psi(\alpha_2).$$

The proof of this result is somewhat similar to that in Donald and Paarsch (1993). Note that the estimation errors in $\underline{\hat{w}}$ do not influence the limiting distribution since those errors disappear at rate T.

When $k = r$, things are very different. Since, by Theorem 3, the parameters are determined by the constraints, no averages are involved and the distribution is related to that of $\underline{\hat{w}}$, which are extreme order statistics. The limiting distributions in this case are related to the $\mathcal{E}(1)$ family, and the estimators will converge at rate T. This unusual result is contained in Theorem 5. To state this theorem succinctly, we first develop some notation. Note that in this case for large enough T, $\hat{\alpha}$ is the solution to

$$\underline{\hat{w}}(x(i)) = \Im(\hat{\alpha}, x(i)),$$

so that, as noted above, we can write

$$\hat{\alpha} = \psi(\underline{\hat{w}}, x)$$

where $\psi(\cdot)$ is a smooth function of $\underline{\hat{w}}$ near the limiting values $\Im^0$. To characterize the limiting distribution we expand the function about $\Im^0$.

Theorem 5.

Under assumptions 1 to 8 and assuming that $k = r$

$$\hat{D}_T^{-1} \hat{J}_T(\hat{\alpha} - \alpha^0) \xrightarrow{\text{d}} (\mathcal{E}_1(1), \ldots, \mathcal{E}_k(1))$$

a vector of independent $\mathcal{E}(1)$ random variables where

$$\hat{D}_T = \text{diag}\{d_T(i)\}$$

of dimension k, where $d_T(i)$ is given in Corollary 1, and

$$\hat{J}_T = \nabla_\alpha \Im(\hat{\alpha})$$

with $\nabla_\alpha \Im(\alpha)$ being the matrix formed by the vectors $\nabla_\alpha \Im\big(\alpha, x(i)\big)$ for $i = 1, \ldots, k$.

Note that the standardization in Theorem 5 will be proportional to T, so that the estimators converge at the rate T, and the limiting distribution is that of a vector of independent exponential $\mathcal{E}(1)$ random variables. This result does not imply that the estimators themselves have one-sided distributions, only that there is a linear transformation of the estimators which has a one-sided distribution. One may be concerned about the fact that the limiting distribution is not normal, since this could make inference difficult. There is no need for this concern since the exponential distribution has a particularly convenient closed-form cumulative distribution function, which should make it even easier to form confidence intervals than would be the case with a normal limiting distribution.

These results can easily be adjusted to the case where only a subset of the parameters influence the lower bound of the distribution. Another case that can easily be examined is where $k < r$ and one can solve for a subset of α_1 as functions of only $\hat{w}$ and x. The result of Theorem 5 would imply that these parameter estimators have $\mathcal{E}(1)$ limiting distributions whereas the remaining parameter estimators will have distributions which fall in the normal limiting family of distributions.

Corollary 2.

Suppose the conditions of Theorem 4 hold, and that a subset of $\hat{\alpha}_1$ can be written as functions of only $\underline{\hat{w}}$ and x, then this subset will have limiting $\mathcal{E}(1)$ distributions, and the remaining parameters will have limiting normal distributions. Moreover, the first subset will converge at rate T and the remainder converge at $\sqrt{T}$.

Note that this corollary can be used to show that the simple Pareto example considered previously in the examples behaves like this.

Case 2: $k > r$

When $k > r$, with finite k, there will generally be more than one way of determining the parameters from the constraints. Moreover, in the population problem (P), the objective function may be tangent to one of the constraints. These facts make it possible for the solution to the sample problem to be such that r constraints bind, or less than r constraints bind, and this will be random from sample to sample. This introduces potential difficulties in the asymptotic analysis. To proceed, we shall make the following assumption which will guarantee that for large T the solution to (P) has at least r constraints binding.

Assumption 9.

In problem (P), the matrix

$$(\nabla_\alpha S_0(\alpha), \{\nabla_\alpha \Im(\alpha)\}_{r-1})$$

has full rank over a neighbourhood of α^0 where

$$\{\nabla_\alpha \Im(\alpha)\}_{r-1}$$

is a collection of derivatives of any $r-1$ distinct lower bounds.

The type of situation that this assumption rules out is illustrated in Figure 1 which applies to the Pareto auction example when $k = 3$ and $r = 2$. The following Lemma then shows that given this assumption the solution to the sample problem for large T will occur where r constraints are binding. The advantage of this is that the asymptotics for the case with r binding constraints and the case with less than r binding constraints are quite different; with the first convergence is at rate T and with the latter convergence is at rate $\sqrt{T}$.

Lemma 4.

In the sample problem (S), assuming that

$$\nabla_\alpha S_T(\alpha, x(i)) \xrightarrow{\text{a.s.}} \nabla_\alpha S_0(\alpha, x(i))$$

uniformly over a neighbourhood of α^0, then the optimal solution occurs at a point such that for large T at least r constraints bind with probability 1.

Suppose there are k constraints and r parameters, then there will be $L = \binom{k}{r}$ possible combinations of constraints at which the solution may occur. Denote the set of possibilities by Ξ. (Note that Assumption 4 guarantees that each possible combination will possess a solution.) We shall let ℓ index each solution and let $\hat{\alpha}(\ell)$ be the solution to the ℓ^{th} set of constraints. Also, let $E(\ell)$ be the event that the solution to (S) is at $\hat{\alpha}(\ell)$. The Lemma shows that for large enough T the solution to problem (S), denoted $\hat{\alpha}$, is such that

$$\hat{\alpha} = \sum_{\ell \in \Xi} \hat{\alpha}(\ell) I[E(\ell)]$$

almost surely, where $I[E]$ denotes the indicator function for the event E. Note that with probability 1 only one of the $I[E(\ell)]$ will be 1.

Using arguments similar to those used previously, one can rule out some of the combinations as being likely to occur with probability 0. For example, consider Figure 2 representing the population problem with $k = 3$ and $r = 2$, and satisfying Assumption 9. In this case, we can rule out an optimum at the solution of C_2 and C_3 as being likely. The reason is that if we maximized the objective function subject to these two constraints then the optimum is actually at B rather than at A. Alternatively, at A it is impossible to find a positive Lagrange multiplier λ_2 to satisfy the first Kuhn-Tucker condition using these two constraints alone. Also note that at the actual optimum B, C_1 is not satisfied. Noting this, we are able to narrow down the set of possible solutions. In the example in Figure 2, there will in fact be two possible solutions. Maximizing the objective subject to either (C_1, C_2) or (C_1, C_3) we are able to satisfy the Kuhn-Tucker conditions; i.e., find positive λ_i's and satisfy all remaining constraints. The following Lemma makes this more precise in general. First, define the solution to the linear equation,

$$\nabla_\alpha S_0(\alpha) - \{\nabla_\alpha \Im(\alpha)\}_{r\ell} \lambda = 0$$

to be λ_0^ℓ, where the notation $\{\cdot\}_{r\ell}$ is the ℓ^{th} possible combination of r elements of the argument.

Lemma 5.

In the sample problem (S), assuming that

$$\nabla_\alpha S_T(\alpha, x(i)) \xrightarrow{\text{a.s.}} \nabla_\alpha S_0(\alpha, x(i))$$

uniformly over a neighbourhood of α^0, then for large enough T $P[E(\ell)] = 0$ if any element of λ_0^ℓ is negative.

Note that Assumption 9 rules out the possibility of any of the λ_0^ℓ being 0. Also, note that in the example in Figure 2, the combination that has (C_2, C_3) will have one of the Lagrange multipliers being negative. Denote the remaining set of possible combinations, not ruled out by Lemma 5, by Ξ_R. Also, define Ξ_ℓ, the ℓ^{th} element of Ξ_R, to be the collection of the indices of constraints used to obtain this solution. For example, in the case considered in Figure 2, there are two possible solutions, so Ξ_R has two elements, and we could define $\Xi_1 = \{1, 2\}$ and $\Xi_2 = \{1, 3\}$.

One potential problem that is raised by this representation is that the events $E(\ell)$ are random. Although Theorem 5 gives a nice characterization of the distribution for any *given* solution (unconditionally), the normalizing matrix $\hat{D}_T^{-1}\hat{J}_T$ will, in general, be different for each possible solution. This problem can be avoided, however, by using the following normalizing random variable,

$$\hat{B}_T = \sum_{\ell \in \Xi_R} I[E(\ell)]\hat{D}_{T\ell}^{-1}\hat{J}_{T\ell}$$

where $\hat{D}_{T\ell}^{-1}\hat{J}_{T\ell}$ is the required normalization for the ℓ^{th} possible solution as in Theorem 5. Since almost surely only one of the $I[E(\ell)]$ will be 1, and $I[E(i)]I[E(j)] = 0$ for $i \neq j$, then we have that

$$\hat{B}_T(\hat{\alpha} - \alpha^0) = \sum_{\ell \in \Xi_R} I[E(\ell)]\hat{D}_{T\ell}^{-1}\hat{J}_{T\ell}(\hat{\alpha}(\ell) - \alpha^0).$$

In order to characterize the limiting distribution of this quantity, we must determine for each value of ℓ, the limiting distribution of

$$\hat{D}_{T\ell}^{-1}\hat{J}_{T\ell}(\hat{\alpha}(\ell) - \alpha^0) = \hat{D}_{T\ell}^{-1}(\underline{\hat{w}}(x(\ell)) - \Im_\ell)$$

conditional on the event $E(\ell)$. This requires more precise information upon what actually determines $E(\ell)$ and its relationship to the above random variables. The following Lemma shows precisely how this is done.

Lemma 6.

In the sample problem (S), assuming that

$$\nabla_\alpha S_T(\alpha, x(i)) \xrightarrow{\text{a.s.}} \nabla_\alpha S_0(\alpha, x(i))$$

uniformly over a neighbourhood of α^0, then for large enough T $I[E(\ell)] = 1$, and hence $\hat{\alpha} = \hat{\alpha}(\ell)$, if and only if, for every constraint i

$$\Im(\hat{\alpha}(\ell), x(i)) \leq \underline{\hat{w}}(x(i)).$$

In other words, after restricting attention to solutions that are not ruled out by Lemma 5, a particular solution will be the optimum when all of the constraints are satisfied at that particular solution. In the case where the logarithm of the likelihood function is pseudo-concave over a neighbourhood of α^0, then this result would be obvious due to the fact noted in footnote 3. Clearly, the constraints used to determine the ℓ^{th} solution will all be satisfied, so it remains to check any constraint not included in the ℓ^{th} solution. This fact makes characterizing the conditional distribution possible. This is because the joint distribution of all k^{th} order statistics is simply the product of each marginal distribution due to the independence assumption, and the conditional distribution is just the conditional distribution of r components of this conditional upon the fact that certain linear combinations of these r components exceed each of the remaining $(k-r)$ components. This is proved in Theorem 6, which contains the limiting distribution result for the quantity $\hat{B}_T(\hat{\alpha} - \alpha^0)$.

Theorem 6.

Under the assumptions made above

$$\hat{B}_T(\hat{\alpha} - \alpha^0) \xrightarrow{\text{d}} z = (z_1, \ldots, z_r)'$$

which has a joint density function given by

$$\prod_{i=1}^{r} f(z_i) \sum_{\ell \in \Xi_R} \prod_{i \notin \Xi_\ell} [1 - P(k'_{\ell i} z)]$$

where the constants $k_{\ell i}$ are given by

$$k'_{\ell i} = \lim_{T\to\infty} d_T^{-1}(i)\frac{\partial \Im_i}{\partial \alpha'}\hat{J}_{T\ell}^{-1}\hat{D}_{T\ell}.$$

Here, $f(\cdot)$ is the probability density function of an $\mathcal{E}(1)$ random variable, while

$$P(k'_{\ell i}z) = I[k'_{\ell i}z > 0]F(k'_{\ell i}z)$$

and $F(\cdot)$ is the cumulative distribution function for an $\mathcal{E}(1)$ random variable.

The limiting distribution in Theorem 6 is non-standard, and to our knowledge has not appeared in any other problems. The distribution does, however, bear some resemblance to the density function for the smallest order statistic from a finite number of draws from some population.[6] As it stands, the distribution depends upon various unknowns, but these unknowns are all estimable. One may determine consistently the set Ξ_R by using the result of Lemma 6 to see if Lagrange multipliers at the solutions in the sample problem are all positive. Given that this is possible, one may then determine the constants $k_{\ell i}$ using sample estimates.

A simple corollary that follows from this is that the estimator is consistent at the rate T rather than the usual $\sqrt{T}$.

Corollary 3.

Under the conditions of Theorem 6, for any $\delta > 0$

$$T^{1-\delta}(\hat{\alpha} - \alpha^0) = o_p(1)$$

4. Some Monte Carlo Evidence

In this section, we use Monte Carlo methods to compare the small sample properties of the maximum likelihood estimator with those of the piecewise pseudo-maximum

[6] For example, if $w = \min\{w_1, w_2\}$ where w_1 and w_2 are draws from some population with probability density function $f(\cdot)$ and cumulative distribution function $F(\cdot)$, then the density of w is

$$2f(w)[1 - F(w)].$$

likelihood and non-linear least squares estimators. For direct comparability, we have adopted the experimental design used in Donald and Paarsch (1993). Thus, in all of our simulation experiments we assumed that the latent distribution of costs c follows the Pareto law, so

$$g(c) = \frac{\alpha_2 \alpha_1^{\alpha_2}}{c^{\alpha_2+1}} \qquad 0 < \alpha_1 < c,\ 0 < \alpha_2.$$

The density of w is then

$$h(w; \alpha_1, \alpha_2, m) = \frac{\alpha_2(m+1)\left(\frac{\alpha_1\alpha_2 m}{\alpha_2 m - 1}\right)^{\alpha_2(m+1)}}{w^{\alpha_2(m+1)+1}} \qquad \frac{\alpha_1\alpha_2 m}{\alpha_2 m - 1} < w,$$

while the j^{th} raw moment of w is

$$\mathrm{E}[w^j] = \left(\frac{\alpha_1\alpha_2 m}{\alpha_2 m - 1}\right)^j \frac{\alpha_2 n}{\alpha_2 n - j} \qquad j < \alpha_2 n,\ j = 1, 2, \ldots,$$

implying the following empirical specification for the first raw moment:

$$w = \left(\frac{\alpha_1\alpha_2 m}{\alpha_2 m - 1}\right)\frac{\alpha_2 n}{\alpha_2 n - 1} + u_1,$$

where u_1 has a mean of zero and a variance which depends upon m.

We fixed the values of (α_1^0, α_2^0) at $(1, 2)$. This implies that the expected value of c is two, while the variance of c does not exist. This latter implication has no effect upon our work since it is the second raw moment of $z = \min[c_1, \ldots, c_n]$ which is important. Allowing c to have a very diffuse distribution also mimics some of the empirical evidence encountered in field data, see Paarsch (1992). In any case, the second raw moment of z depends upon n and exists in all of our experiments.

We considered three different sample sizes T: 50, 100, and 200. In each of these samples, the number of bidders n could take on four different values: 3, 6, 9, and 12. This implies that the number of opponents m could take on the values 2, 5, 8, and 11. These values of n reflect the amount of competition that is often encountered in field data. Thus, in this model $k = 4 > 2 = r$.

We investigated three different patterns for the design matrix of the m's, the probability distribution of the m's, the $\{\pi(m)\}_{m=1}^{M}$'s. In the first, each m was equally

Table 0
Design Matrices of the T_m's

Sample Size		50				100				200			
	m	2	5	8	11	2	5	8	11	2	5	8	11
Design A	T_m	13	12	12	13	25	25	25	25	50	50	50	50
	$\pi(m)$	0.26	0.24	0.24	0.26	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Design B	T_m	5	10	15	20	10	20	30	40	20	40	60	80
	$\pi(m)$	0.10	0.20	0.30	0.40	0.10	0.20	0.30	0.40	0.10	0.20	0.30	0.40
Design C	T_m	20	15	10	5	40	30	20	10	80	60	40	20
	$\pi(m)$	0.40	0.30	0.20	0.10	0.40	0.30	0.20	0.10	0.40	0.30	0.20	0.10

likely (Design A), while in the second, large m's were more likely than small ones (Design B), and in the third, small m's were more likely than large ones (Design C). In Table 0, we present the T_m's and their corresponding $\pi(m)$'s for the three different designs.

For the piecewise pseudo-maximum likelihood estimator, we partitioned the parameter vector $\alpha = (\alpha_1, \alpha_2)$ in two different ways, concentrating out first α_1 and then α_2. Below, we refer to these partitions as Partition 1 and Partition 2, respectively.

For Partition 1, the piecewise pseudo-maximum likelihood estimator of α_2 is

$$\hat{\alpha}_2^{\text{pple}} = \frac{T}{\sum_{t=1}^{T}(m_t+1)\log\left(\frac{w_t}{\underline{\hat{w}}(m_t)}\right)},$$

but an estimator of α_1 can be defined in at least four different ways. First, consider any of

$$\hat{\alpha}_1^m = \underline{\hat{w}}(m)\left(\frac{\hat{\alpha}_2^{\text{pple}}m - 1}{\hat{\alpha}_2^{\text{pple}}m}\right) \quad m = 1, \ldots, M.$$

Alternative estimators are

$$\hat{\alpha}_1^{\min} = \min[\hat{\alpha}_1^1, \hat{\alpha}_1^2, \ldots, \hat{\alpha}_1^M],$$

$$\hat{\alpha}_1^{\text{a}} = \sum_{m=1}^{M} \frac{T_m \times (m+1)}{N_T} \hat{\alpha}_1^m,$$

and

$$\hat{\alpha}_1^{\mathrm{b}} = \sum_{m=1}^{M} \frac{T_m}{T} \hat{\alpha}_1^m,$$

where $N_T = \sum_{t=1}^{T} n_t$. For Partition 2, the piecewise pseudo-maximum likelihood estimator of α_1 is defined implicitly (see Donald and Paarsch [1993]), and an estimator of α_2 can also be defined in at least four different ways. First, consider any of

$$\hat{\alpha}_2^m = \frac{\hat{\underline{w}}(m)}{(\hat{\underline{w}}(m) - \hat{\alpha}_1^{\mathrm{pple}})m} \quad m = 1, \ldots, M.$$

Alternative estimators are

$$\hat{\alpha}_2^{\min} = \min[\hat{\alpha}_2^1, \hat{\alpha}_2^2, \ldots, \hat{\alpha}_2^M],$$

$$\hat{\alpha}_2^{\mathrm{a}} = \sum_{m=1}^{M} \frac{T_m \times (m+1)}{N_T} \hat{\alpha}_2^m,$$

and

$$\hat{\alpha}_2^{\mathrm{b}} = \sum_{m=1}^{M} \frac{T_m}{T} \hat{\alpha}_2^m.$$

The random numbers for the experiments were generated using the multiplicative congruential method with modulus $(2^{31} - 1)$, multiplier 397204094, and initial seed 2420375. This method generates uniform pseudo-random numbers on the interval $(0, 1)$. (For more details, see Hall et al. 1988, pp. 232-235.) Using the property that the distribution function is distributed uniformly on the interval $(0, 1)$, we applied the inverse distribution function to obtain the pseudo-random w's.

We maximized the logarithm of the likelihood function subject to the T constraints using a slight modification of Schittkowski's (1981a,b) implementation of the (recursive) quadratic approximation method of Wilson (1963), Han (1976, 1977), and Powell (1978), see Vaesson (1984, pp. 57-66).

For Partition 2, we maximized the logarithm of the concentrated likelihood functions using the Newton-Raphson algorithm. We minimized the sum of squared residuals using the Gauss-Newton method. The true parameter values were used as starting values.

The results of the nine experiments are presented in Tables 1 to 9. The abbreviations St.Dev., L.Q., and U.Q. denote respectively the standard deviation, lower quartile, and upper quartile of the estimator's distribution. Also, in these tables the superscript upon an estimator denotes its type. For example, the 5 on $\hat{\alpha}_1^5$ implies that this is an estimator based upon Partition 1, for the case when $m = 5$. The "min" superscript denotes the minimum of all estimators with numeric superscripts and the same subscript. The superscripts "a" and "b" denote the type of averaging of the $\hat{\alpha}_i^m$'s, where the "a" denotes the weights $(T_m \times (m+1)/N_T)$ and where the "b" denotes the weights (T_m/T). An estimator with the superscript "pple" is the piecewise pseudo-maximum likelihood estimator (e.g., $\hat{\alpha}_1^{\text{pple}}$ is the piecewise pseudo-maximum likelihood estimator based upon Partition 2), while one with the superscript "mle" is the maximum likelihood estimator, and one with the superscript "nls" is the non-linear least squares estimator.

As one case see, the rates of bias (when measured using either the mean or the median) for the maximum likelihood estimator are typically less that those of the other estimation methods. What is most stark about the performance of the maximum likelihood estimator is its quick convergence which is suggested by the rate T convergence since $4 = k > r = 2$. Notice in Table 1 that for a sample size of fifty the standard deviation of $\hat{\alpha}_1^{\text{nls}}$ is 0.0236, while that of $\hat{\alpha}_1^{\text{mle}}$ is 0.0042. In Table 3, where the sample size is two hundred, the standard deviation of $\hat{\alpha}_1^{\text{nls}}$ is 0.0113, while that of $\hat{\alpha}_1^{\text{mle}}$ is 0.0005. These results are common across the the nine tables.

5. Summary

We have derived a maximum likelihood estimator which is useful in estimating empirical models of auctions where the support of the distribution often depends upon <u>all</u> of the parameters of interest. Under fairly general conditions, the estimator has been shown to be consistent, and in the case of discrete data we have derived its asymptotic distribution. In addition, we have compared the small sample properties of this estimator with those of the piecewise pseudo-maximum likelihood and non-linear least squares estimators. We find, in the examples considered here, that the maximum likelihood estimator performs extremely well. Essentially, a researcher can obtain the same level of precision with 50 observations using maximum likelihood as would require 2500 observations using non-linear least squares.

A. Appendix

In this appendix, we present the proofs of the theorems and lemmata contained in the paper.

Proof of Lemma 1. Note that

$$h(w) = \frac{\tilde{g}(z)}{\beta'(z)}$$

where we have written it in terms of z, (noting that $\beta^{-1}(w) = z = c_{(1:n)}$), and where

$$\tilde{g}(z) = n[1 - G(z)]^m g(z).$$

The result follows by noting that

$$\beta'(c) = \frac{mg(c) \int_c^\infty [1 - G(\xi)]^m \, d\xi}{[1 - G(c)]^{m+1}},$$

and that $w \to \Im(\alpha, x)$ as $z \to 0$.

Proof of Theorem 1. Let X^i $(i = 1, \ldots, r)$ denote compact subsets of X each with a non-empty interior such that no common boundaries exit; i.e., a strictly positive distance exists between each of them. By assumption 4, $P(X^i) > 0$ for each i. Let

$$T^i = \sum_{t=1}^{T} I(x_t \in X^i),$$

and note that $T^i = O_{\text{a.s.}}(T)$ since $\frac{T^i}{T} \xrightarrow{\text{a.s.}} P(X^i)$ by Kolmogorov's strong law of large numbers.

For each T and each set X^i let

$$t_i = \arg\min \{w_t - \Im(\alpha^0, x_t) \mid x_t \in X^i\},$$

where it is understood that t_i depends on the sample size T. For sufficiently large T, at least one $x_t \in X^i$ exists for all i. Let

$$\tilde{A}_T = \{\alpha \in \bar{A} \mid \Im(\alpha, x_{t_i}) \leq w_{t_i} \quad i = 1, \ldots, p\}.$$

Note that $A_T^* \subset \tilde{A}_T$, so that

$$d(A_T^*, A^*) \leq d(\tilde{A}_T, A^*).$$

If we can show that $d(\tilde{A}_T, A^*) \xrightarrow{\text{a.s.}} 0$, then the result will follow. The following result that

$$\max_i \gamma_i \xrightarrow{\text{a.s.}} 0$$

is useful. To show it let

$$\gamma_i = w_{t_i} - \Im(\alpha^0, x_{t_i}).$$

Fix $\varepsilon > 0$.

$$\begin{aligned}
\Pr(\gamma_i > \varepsilon) &= \Pr\big(w_{t_i} - \Im(\alpha^0, x_{t_i}) > \varepsilon \mid x_{t_i}\big) \\
&= \prod_{x_t \in X^i} \Pr\big(w_t - \Im(\alpha^0, x_t) > \varepsilon \mid x_t\big) \\
&\leq \prod_{x_t \in X^i} \big(1 - \delta(\varepsilon)\big) \\
&= \big(1 - \delta(\varepsilon)\big)^{T^i} \to 0,
\end{aligned}$$

since $T^i \to \infty$, so $\gamma_i \xrightarrow{\text{P}} 0$ for each i. Also,

$$\sum_{T=1}^{\infty} \big(1 - \delta(\varepsilon)\big)^{T^i} < \infty,$$

so that $\gamma_i \xrightarrow{\text{a.s.}} 0$ for each i and the result follows since r is finite.

Let $\underline{w}^0(x) = \Im(\alpha^0, x)$. Also, let

$$\underline{\mathrm{w}}(\underline{\mathrm{x}}) = (\underline{w}_1(x_1), \ldots, \underline{w}_r(x_r)) \in \mathbf{R}^r$$

$$\underline{\mathrm{x}} = (x_1, \ldots, x_r) \in \times_{i=1}^{r} X^i = \underline{\mathrm{X}}$$

$$\underline{\Im}(\alpha, \underline{\mathrm{x}}) = \big(\Im(\alpha, x_1), \ldots, \Im(\alpha, x_r)\big).$$

Let $\underline{\mathrm{x}}(T)$ and $\underline{\mathrm{w}}(T)$ denote the particular values of these variables chosen above; i.e., the i^{th} element of $\underline{\mathrm{w}}(T)$ is w_{t_i}.

Note that $\underline{\mathrm{X}}$ is a compact set. For any $\underline{\mathrm{x}} \in \underline{\mathrm{X}}$ there is an open ball $B(\underline{\mathrm{w}}^0(\underline{\mathrm{x}}), \delta_x)$ such that (by the implicit function theorem and assumption 4) there is a continuous and differentiable solution to

$$\underline{\mathrm{w}}(\underline{\mathrm{x}}) = \underline{\Im}(\alpha, \underline{\mathrm{x}}).$$

Note that by the compactness of $\underline{X}$ and the fact that the individual x_i are not allowed to get close to one another, we have that

$$\underline{\delta} = \inf \{\delta_x \mid \underline{x} \in \underline{X}\},$$

and

$$\inf \{\left|\frac{\partial}{\partial \alpha}\underline{\Im}(\alpha, \underline{x})\right| \mid \underline{x} \in \underline{X}\} > 0.$$

Pick $\bar{T}$ sufficiently large so almost surely

$$\max_i \ \gamma_i < \underline{\delta},$$

for all $T > \bar{T}$. For such T values, by the above argument, a solution for α to the equations

$$\underline{w}(T) = \Im(\alpha, \underline{x}(T)),$$

denoted by $\tilde{\alpha} = \alpha(\underline{w}(T), \underline{x}(T))$. Note that by definition $\alpha^0 = \alpha(\underline{w}^0(T), \underline{x}(T))$. The implicit function is continuous and differentiable and the determinant of the Jacobian is bounded away from zero, the derivative of α with respect to $\underline{w}(T)$ is bounded, then it is easy to show (via a mean value expansion) that

$$\|\tilde{\alpha} - \alpha^0\| \leq B\|\underline{w}(T) - \underline{w}^0(T)\| \leq rB \max_i \ \gamma_i \xrightarrow{\text{a.s.}} 0,$$

where B is some positive finite constant. Thus, $\tilde{\alpha} \xrightarrow{\text{a.s.}} \alpha^0$.

Next, note that we may write $\tilde{A}_T$ for large T as

$$\tilde{A}_T = \{\alpha \in \bar{A} \mid \Im(\alpha, x_{t_i}) \leq \Im(\tilde{\alpha}, x_{t_i}) \quad i = 1, \ldots, r\}.$$

Since $\tilde{\alpha}$ is consistent and $\Im$ is twice continuously differentiable in all its arguments (on a compact set), we have that

$$\sup_{x \in X} \ \Im(\tilde{\alpha}, x) - \Im(\alpha^0, x) \xrightarrow{\text{a.s.}} 0.$$

Now fix an $\varepsilon > 0$. Define the compact set

$$\hat{A} = \{\alpha \in \bar{A} \mid \inf_{\alpha^* \in A^*} \|\alpha^* - \alpha\| \geq \varepsilon\}.$$

Let

$$\delta = \min_{\alpha \in \hat{A}} \max_{x \in X} (\Im(\alpha, x) - \Im(\alpha^0, x)) > 0.$$

For large enough T

$$\sup_{x \in X} \Im(\tilde{\alpha}, x) - \Im(\alpha^0, x) < \delta.$$

Note that if $\alpha \in \hat{A}$, then for some x

$$\Im(\alpha, x) - \Im(\alpha^0, x) > \delta,$$

so that $\alpha \notin \tilde{A}_T$. But $\hat{A}$ includes all points in $\bar{A}$ that have

$$\inf_{\alpha^* \in A^*} \|\alpha^* - \alpha\| \geq \varepsilon,$$

so that if $\alpha \in \tilde{A}_T$ then

$$\inf_{\alpha^* \in A^*} \|\alpha^* - \alpha\| < \varepsilon.$$

Since $\tilde{A}_T$ is compact, it must be that

$$d(\tilde{A}_T, A^*) < \varepsilon,$$

almost surely for large enough T. Since ε is arbitrary and since $A^*_T \subset \tilde{A}_T$ then

$$d(\tilde{A}_T, A^*) < \varepsilon,$$

and the result follows.

Proof of Theorem 2. This proof is very similar to others in the literature; e.g., a slight modification upon Gallant and Nychka (1987). Note that $\hat{\alpha} \in \bar{A}$ for all T, so a subsequence that converges to some $\alpha^* \in \bar{A}$ exists; i.e., $\hat{\alpha}_j \to \alpha^*$. Note that $\alpha^* \in A^*$ since

$$d(A^*_{T_j}, A^*) \xrightarrow{\text{a.s.}} 0,$$

and $\hat{\alpha}_j \in A^*_{T_j}$ for all j. All we need show then is that $\alpha^* = \alpha^0$, which follows from the uniform convergence assumption, the identification assumption, and the fact that $\hat{\alpha}_j$ is the maximizer of the logarithm of the likelihood function over $A^*_{T_j}$. This part of the result is standard.

Proof of Corollary 1. Using Lemma 1 we can show that for all $z < 0$

$$\lim_{t \to 0_-} \frac{F(\Im(x(i), \alpha^0) + zt)}{F(\Im(x(i), \alpha^0) - t)} = -z$$

using L'Hôpital's rule, and also $\Im(x(i), \alpha^0) > -\infty$ by Assumption 1, so the result follows by Lemma 3. It is also easy to show by L'Hôpital's rule that d_T is such that

$$\frac{d_T^*}{d_T} \to 1$$

almost surely where d_T^* is given in Lemma 3.

Proof of Theorem 3. Note that for the corresponding population problem the Kuhn-Tucker conditions are satisfied at

$$(\alpha^0, \pi(x(1))\lambda_1^0, \pi(x(2))\lambda_2^0, ..., \pi(x(k))\lambda_k^0)$$

and by Assumptions 7 and 8 each of the Lagrange multipliers are positive, so each of the k constraints binds. Also, note that by Assumption 4, the matrix formed by the k vectors $\nabla_{\alpha_1}\Im(\alpha, x(i))$ has full rank in a neighbourhood of α^0 which implies that the Kuhn-Tucker conditions are sufficient over this neighbourhood for a solution. Since $\hat{\alpha}$ is consistent, and since we have uniform convergence of the gradient vector, then the solution to the sample problem must also be such that all k constraints bind with Lagrange multipliers converging almost surely to the population values.

Proof of Theorem 4. The proof is very similar to that in Donald and Paarsch (1993), Propositions 2 and 3. The result follows from standard mean value expansions, noting that since k constraints bind

$$\hat{\alpha}_1 = \psi(\hat{\alpha}_2, \underline{\hat{w}}, x).$$

One can then expand the first-order condition

$$\sqrt{T}\nabla_{\alpha_2} S_T(\psi(\hat{\alpha}_2, \underline{\hat{w}}, x, \hat{\alpha}_2)) = 0$$

about α_2^0, and ignore the pre-estimation error in $\underline{\hat{w}}$ since by Lemma 2 this is $O_p(T^{-1})$. The result for α_1 follows from the δ method applied to the function $\psi(\alpha_2, \underline{\hat{w}}, x)$, which is twice continuously differentiable in a neighbourhood of the true values.

Proof of Theorem 5. By Theorem 3, the solution is such that all k constraints bind, so that $\hat{\alpha}$ can be determined by

$$\hat{\alpha} = \psi(\hat{\underline{w}}, x)$$

for some twice continuously differentiable function ψ. Note that $\alpha^0 = \psi(\Im^0, x)$, so an expansion of $\psi(\hat{\underline{w}}, x)$ about $\Im^0$ yields

$$\hat{\alpha} - \alpha^0 = J_T^{*-1}(\hat{\underline{w}} - \Im^0)$$

where $J_T^* = \nabla_\alpha \Im(\alpha^*)$ for some α^* lying between $\hat{\alpha}$ and α^0. Since $\hat{\alpha} \to \alpha^0$, then $\alpha^* \to \alpha^0$ and J_T^* is almost surely invertible by Assumption 3. Therefore,

$$\hat{D}_T^{-1} J_T^*(\hat{\alpha} - \alpha^0) = \hat{D}_T^{-1}(\hat{\underline{w}} - \Im^0)$$

which is distributed jointly asymptotically as a vector of independent $\mathcal{E}(1)$ random variables by Corollary 1. The result then follows by showing that $\hat{J}_T - J_T^* = o_p(1)$ which follows since both $\hat{J}_T$ and J_T^* converge to $\nabla \Im(\alpha^0, x)$ because $\hat{\alpha} \xrightarrow{P} \alpha^0$ and $\alpha^* \xrightarrow{P} \alpha^0$.

Proof of Lemma 4. Due to the uniform convergence of the gradient ∇S over a neighbourhood of α^0 and the fact that $\hat{\alpha} \to \alpha^0$ almost surely the result follows by Assumption 9 using a similar argument to that used in the proof of Theorem 3.

Proof of Lemma 5. Given the uniform convergence of the gradient ∇S over a neighbourhood of α^0 and consistency of $\hat{\alpha}$ it must be the case that the solution to

$$\nabla S(\hat{\alpha}) - \{\nabla_\alpha \Im(\hat{\alpha})\}_{r\ell} \hat{\lambda}^\ell = 0$$

which exists for large T by Assumption 4, is such that $\hat{\lambda}^\ell \to \lambda_0^\ell$ almost surely. For those combinations that have an element of λ_0^ℓ that is negative, this implies that for large enough T the corresponding element of $\hat{\lambda}^\ell$ must be negative and hence the Kuhn-Tucker conditions can not be satisfied at such a solution for large enough T. Hence, for large enough T for such combinations $P[E(\ell)] = 0$.

Proof of Lemma 6. For large enough T, all of the $\hat{\alpha}(\ell)$ are within a neighbourhood of α^0 using the result in Theorem 5 which applies to any of the $\hat{\alpha}(\ell)$. The uniform convergence and continuity of the derivatives over this neighbourhood and the fact that the constraint set is convex by Assumption 3, imply that if more than one $\hat{\alpha}(\ell)$ satisfy all of the Kuhn-Tucker conditions, and these two values differ, then there will be a violation of Assumption 9. The event that the two $\hat{\alpha}(\ell)$ solutions are identical occurs with probability 0, so the result holds.

Proof of Theorem 6. It is easy to see that

$$\hat{B}_T(\hat{\alpha}-\alpha^0) - B_T^*(\hat{\alpha}-\alpha^0) \xrightarrow{\text{P}} 0$$

where

$$B_T^* = \sum_{\ell\in\Xi_R} I[E(\ell)] D_{T\ell}^{-1} J_{T\ell}^*$$

with $D_{T\ell}^{-1}$ being the true normalization given in Corollary 1, and $J_{T\ell}^*$ being given in the mean value expansions in Theorem 5. Thus, it suffices to find the asymptotic distribution of $B_T^*(\hat{\alpha}-\alpha^0)$. Using the mean value expansions as in Theorem 5,

$$B_T^*(\hat{\alpha}-\alpha^0) = \sum_{\ell\in\Xi_R} I[E(\ell)] D_{T\ell}^{-1}(\underline{\hat{w}}(\ell) - \Im^0(\ell))$$

where $\underline{\hat{w}}(\ell)$ and $\Im^0(\ell)$ are the vectors formed using the ℓ^{th} combination of smallest order statistics and true lower bounds. Using similar expansions the event $E(\ell)$ occurs when

$$k_{\ell i}^{*\prime} D_{T\ell}^{-1}(\underline{\hat{w}}(\ell) - \Im^0(\ell)) \leq d_T^{-1}(i)(\underline{\hat{w}}_i - \Im_i^0)$$

for each $i \notin \Xi_\ell$ where

$$k_{\ell i}^{*\prime} = d_T^{-1}(i)\nabla_\alpha \Im(\alpha^*, x(i)) J_{T\ell}^{*^{-1}} D_{T\ell}$$

for some α^* between $\hat{\alpha}$ and α^0 . Note that $k_{\ell i}^* \xrightarrow{\text{P}} k_{\ell i}$ since $\hat{\alpha} \xrightarrow{\text{P}} \alpha^0$. The vector consisting of

$$d_T^{-1}(i)(\underline{\hat{w}}_i - \Im_i^0)$$

is distributed as a k vector of independent $\mathcal{E}(1)$ random variables. Thus, the asymptotic distribution reduces to finding the asymptotic distribution of

$$Z = \sum_{\ell\in\Xi_R} \bar{z}(\ell)$$

where $\bar{z}(\ell)$ is the ℓ^{th} combination of the $\mathcal{E}(1)$ random variables and the event $E(\ell)$ occurs when

$$k'_{\ell i}\bar{z}(\ell) \leq \bar{z}_i$$

for all $i \notin \Xi_\ell$. The probability that $Z \leq z$ then can be computed as

$$\sum_{\ell \in \Xi_R} P[E(\ell)]P[\bar{z}(\ell) \leq z | E(\ell)]$$

$$= \sum_{\ell \in \Xi_R} \int_0^{z_1} \cdots \int_0^{z_r} (\prod_{i \notin \Xi_\ell} \int_{k'_{\ell i}\bar{z}(\ell)}^{\infty}) \prod_{j=1}^{k} f(\bar{z}_j) \prod_{j=1}^{k} d\bar{z}_j$$

$$= \sum_{\ell \in \Xi_R} \int_0^{z_1} \cdots \int_0^{z_r} \prod_{i \in \Xi_\ell} f(\bar{z}_i) \prod_{j \notin \Xi_\ell} [1 - P(k'_{\ell i}\bar{z}(\ell))] \prod_{i \in \Xi_\ell} d\bar{z}_j$$

where the notation

$$\prod_{i=1}^{k} d\bar{z}_i = d\bar{z}_1 d\bar{z}_2 \ldots d\bar{z}_k$$

is used. This result implies that the density is as given in the Theorem. Note that the function P arises since $k'_{\ell i}\bar{z}(\ell)$ may be negative and the density for the $\mathcal{E}(1)$ variable is only defined over positive values.

B. Bibliography

Christensen, B. and N. Kiefer, "The Exact Likelihood Function for an Empirical Job Search Model," Econometric Theory 7 (1991), 464–486.

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

Galambos J., The Asymptotic Theory of Extreme Order Statistics (New York: John Wiley, 1978).

Gallant, R. and D. Nychka, "Semi-nonparametric Maximum Likelihood Estimation," Econometrica 55 (March 1987), 363–390.

Gill, P., W. Murray, and M. Wright, Practical Optimization (New York: Academic Press, 1981).

Han, S., "Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems," Mathematical Programming 11 (1976), 263–282.

Han, S., "A Globally Convergent Method for Nonlinear Programming," Journal of Optimization Theory and Applications 22 (July 1977), 297–309.

Mangasarian, O., Nonlinear Programming (New York: McGraw-Hill, 1969).

McAfee, R. and J. McMillan, "Auctions and Bidding," Journal of Economic Literature 25 (June 1987), 699–738.

Milgrom, P., "The Economics of Competitive Bidding: A Selective Survey," in L. Hurwicz, D. Schmeidler and H. Sonnenschein, eds., Social Goals and Social Organization (Cambridge: Cambridge University Press, 1985).

Milgrom, P., "Auction Theory," in T. Bewley, ed., Advances in Economic Theory: Fifth World Congress (Cambridge: Cambridge University Press, 1987).

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

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

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

Powell, M., "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," in G. Watson, ed., Numerical Analysis. Proceedings of the Biennial Conference held at Dundee, June 1977. Lecture Notes in Mathematics, Vol. 630 (Berlin: Springer, 1978).

Reiss, R.-D., Approximate Distributions of Order Statistics: with Applications to Nonparametric Statistics (New York: Springer-Verlag, 1989).

Schittkowski, K., "The Nonlinear Programming Method of Wilson, Han, and Powell with an Augmented Lagrangian Type Line Search Function. Part 1: Convergence Analysis," Numerische Mathematik, 38 (1981a), 83–114.

Schittkowski, K., "The Nonlinear Programming Method of Wilson, Han, and Powell with an Augmented Lagrangian Type Line Search Function. Part 2: An Efficient Implementation with Linear Least Squares Subproblems," Numerische Mathematik, 38 (1981b), 115–127.

Vaessen, W., UBC NLP: Nonlinear Function Optimization (Vancouver, Canada: University of British Columbia Computing Centre, 1984).

Wald, A., "Note on the Consistency of the Maximum Likelihood Estimate," Annals of Mathematical Statistics, 60 (1949), 595–601.

Wilson, R., A Simplicial Algorithm for Concave Programming Ph.D. dissertation (Cambridge, U.S.A.: Harvard University, 1963).

White, H. and J. Wooldridge, "Some Results on Sieve Estimation with Dependent Observations," in W. Barnett, J. Powell, and G. Tauchen, eds., Nonparametric and Semiparametric Methods in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics (Cambridge, England: Cambridge University Press, 1991.)

Table 1
Experiment 1: Design A, Sample Size = 50

Estimator	Mean	St.Dev.	L.Q.	Median	U.Q.
$\hat{\alpha}_1^2$	1.0394	0.0469	1.0098	1.0390	1.0733
$\hat{\alpha}_1^5$	1.0156	0.0165	1.0054	1.0160	1.0267
$\hat{\alpha}_1^8$	1.0098	0.0100	1.0034	1.0098	1.0164
$\hat{\alpha}_1^{11}$	1.0070	0.0072	1.0024	1.0069	1.0118
$\hat{\alpha}_1^{min}$	1.0012	0.0191	1.0009	1.0060	1.0109
$\hat{\alpha}_1^a$	1.0181	0.0195	1.0059	1.0187	1.0318
$\hat{\alpha}_1^b$	1.0128	0.0131	1.0045	1.0132	1.0219
$\hat{\alpha}_1^{pple}$	1.0083	0.0098	1.0017	1.0086	1.0155
$\hat{\alpha}_1^{nls}$	1.0008	0.0236	0.9854	1.0007	1.0172
$\hat{\alpha}_1^{mle}$	1.0009	0.0042	0.9991	1.0004	1.0018
$\hat{\alpha}_2^2$	1.9764	0.0897	1.9215	1.9849	2.0395
$\hat{\alpha}_2^5$	2.0474	0.2114	1.8980	2.0397	2.1917
$\hat{\alpha}_2^8$	2.1752	0.3646	1.9165	2.1338	2.3973
$\hat{\alpha}_2^{11}$	2.3552	0.5608	1.9497	2.2648	2.6537
$\hat{\alpha}_2^{min}$	1.8941	0.1669	1.8033	1.9305	2.0160
$\hat{\alpha}_2^a$	2.1397	0.2796	1.9332	2.1113	2.3122
$\hat{\alpha}_2^b$	2.2049	0.3636	1.9348	2.1590	2.4215
$\hat{\alpha}_2^{pple}$	2.2141	0.3256	1.9875	2.1806	2.4124
$\hat{\alpha}_2^{nls}$	2.0477	0.2738	1.8591	2.0362	2.2169
$\hat{\alpha}_2^{mle}$	1.9815	0.0925	1.9461	1.9817	2.0063

Table 2
Experiment 2: Design A, Sample Size = 100

Estimator	Mean	St.Dev.	L.Q.	Median	U.Q.
$\hat{\alpha}_1^2$	1.0189	0.0343	0.9969	1.0196	1.0413
$\hat{\alpha}_1^5$	1.0076	0.0113	1.0002	1.0074	1.0150
$\hat{\alpha}_1^8$	1.0047	0.0068	1.0003	1.0045	1.0093
$\hat{\alpha}_1^{11}$	1.0034	0.0050	1.0002	1.0033	1.0069
$\hat{\alpha}_1^{min}$	0.9975	0.0165	0.9960	1.0030	1.0064
$\hat{\alpha}_1^{a}$	1.0086	0.0140	0.9998	1.0088	1.0181
$\hat{\alpha}_1^{b}$	1.0062	0.0095	1.0001	1.0063	1.0124
$\hat{\alpha}_1^{pple}$	1.0042	0.0072	0.9997	1.0045	1.0091
$\hat{\alpha}_1^{nls}$	1.0015	0.0172	0.9902	1.0024	1.0130
$\hat{\alpha}_1^{mle}$	1.0003	0.0013	0.9997	1.0001	1.0008
$\hat{\alpha}_2^2$	1.9866	0.0560	1.9535	1.9908	2.0232
$\hat{\alpha}_2^5$	2.0219	0.1427	1.9237	2.0209	2.1063
$\hat{\alpha}_2^8$	2.0861	0.2431	1.9245	2.0640	2.2284
$\hat{\alpha}_2^{11}$	2.1634	0.3587	1.9163	2.1151	2.3584
$\hat{\alpha}_2^{min}$	1.9247	0.1311	1.8692	1.9602	2.0190
$\hat{\alpha}_2^{a}$	2.0645	0.1899	1.9352	2.0449	2.1790
$\hat{\alpha}_2^{b}$	2.0943	0.2421	1.9313	2.0680	2.2321
$\hat{\alpha}_2^{pple}$	2.0969	0.2163	1.9470	2.0772	2.2299
$\hat{\alpha}_2^{nls}$	2.0342	0.2017	1.9011	2.0312	2.1689
$\hat{\alpha}_2^{mle}$	1.9908	0.0294	1.9774	1.9926	2.0008

Table 3
Experiment 3: Design A, Sample Size = 200

Estimator	Mean	St.Dev.	L.Q.	Median	U.Q.
$\hat{\alpha}_1^2$	1.0109	0.0226	0.9956	1.0110	1.0274
$\hat{\alpha}_1^5$	1.0042	0.0076	0.9988	1.0040	1.0096
$\hat{\alpha}_1^8$	1.0026	0.0045	0.9994	1.0029	1.0056
$\hat{\alpha}_1^{11}$	1.0019	0.0033	0.9998	1.0020	1.0041
$\hat{\alpha}_1^{min}$	0.9976	0.0111	0.9956	1.0016	1.0040
$\hat{\alpha}_1^{a}$	1.0049	0.0094	0.9987	1.0051	1.0117
$\hat{\alpha}_1^{b}$	1.0035	0.0064	0.9992	1.0036	1.0080
$\hat{\alpha}_1^{pple}$	1.0024	0.0047	0.9991	1.0027	1.0056
$\hat{\alpha}_1^{nls}$	1.0001	0.0113	0.9927	1.0004	1.0077
$\hat{\alpha}_1^{mle}$	1.0001	0.0005	1.0000	1.0001	1.0003
$\hat{\alpha}_2^2$	1.9955	0.0349	1.9710	1.9983	2.0206
$\hat{\alpha}_2^5$	2.0177	0.0892	1.9555	2.0179	2.0794
$\hat{\alpha}_2^8$	2.0521	0.1517	1.9429	2.0466	2.1503
$\hat{\alpha}_2^{11}$	2.0910	0.2165	1.9356	2.0780	2.2286
$\hat{\alpha}_2^{min}$	1.9538	0.0908	1.9134	1.9829	2.0187
$\hat{\alpha}_2^{a}$	2.0391	0.1193	1.9531	2.0335	2.1158
$\hat{\alpha}_2^{b}$	2.0551	0.1505	1.9451	2.0467	2.1512
$\hat{\alpha}_2^{pple}$	2.0563	0.1422	1.9545	2.0434	2.1513
$\hat{\alpha}_2^{nls}$	2.0137	0.1381	1.9239	2.0162	2.1062
$\hat{\alpha}_2^{mle}$	1.9946	0.0107	1.9927	1.9949	1.9998

Table 4
Experiment 4: Design B, Sample Size = 50

Estimator	Mean	St.Dev.	L.Q.	Median	U.Q.
$\hat{\alpha}_1^2$	1.0616	0.0601	1.0215	1.0596	1.0982
$\hat{\alpha}_1^5$	1.0172	0.0173	1.0059	1.0168	1.0285
$\hat{\alpha}_1^8$	1.0088	0.0098	1.0026	1.0090	1.0152
$\hat{\alpha}_1^{11}$	1.0058	0.0067	1.0016	1.0059	1.0104
$\hat{\alpha}_1^{\text{min}}$	1.0019	0.0167	1.0008	1.0055	1.0101
$\hat{\alpha}_1^{\text{a}}$	1.0146	0.0138	1.0060	1.0146	1.0245
$\hat{\alpha}_1^{\text{b}}$	1.0101	0.0100	1.0038	1.0104	1.0173
$\hat{\alpha}_1^{\text{pple}}$	1.0060	0.0085	1.0007	1.0063	1.0118
$\hat{\alpha}_1^{\text{nls}}$	1.0021	0.0263	0.9857	1.0047	1.0203
$\hat{\alpha}_1^{\text{mle}}$	0.9996	0.0044	0.9978	0.9999	1.0014
$\hat{\alpha}_2^2$	1.8644	0.1497	1.7817	1.8951	1.9749
$\hat{\alpha}_2^5$	1.9779	0.1995	1.8441	1.9775	2.1178
$\hat{\alpha}_2^8$	2.1130	0.2934	1.9100	2.0869	2.2881
$\hat{\alpha}_2^{11}$	2.2613	0.4595	1.9453	2.1970	2.5141
$\hat{\alpha}_2^{\text{min}}$	1.8038	0.1676	1.6971	1.8148	1.9309
$\hat{\alpha}_2^{\text{a}}$	2.1205	0.2983	1.9154	2.0911	2.2969
$\hat{\alpha}_2^{\text{b}}$	2.1658	0.3456	1.9283	2.1256	2.3668
$\hat{\alpha}_2^{\text{pple}}$	2.2148	0.3264	1.9993	2.1841	2.4113
$\hat{\alpha}_2^{\text{nls}}$	2.1013	0.3812	1.8265	2.0774	2.3681
$\hat{\alpha}_2^{\text{mle}}$	1.9575	0.1354	1.9028	1.9682	2.0038

Table 5
Experiment 5: Design B, Sample Size = 100

Estimator	Mean	St.Dev.	L.Q.	Median	U.Q.
$\hat{\alpha}_1^2$	1.0284	0.0368	1.0048	1.0274	1.0533
$\hat{\alpha}_1^5$	1.0082	0.0118	1.0005	1.0082	1.0161
$\hat{\alpha}_1^8$	1.0043	0.0067	1.0000	1.0042	1.0086
$\hat{\alpha}_1^{11}$	1.0028	0.0048	0.9998	1.0028	1.0061
$\hat{\alpha}_1^{min}$	0.9985	0.0142	0.9983	1.0026	1.0060
$\hat{\alpha}_1^{a}$	1.0069	0.0095	1.0009	1.0069	1.0135
$\hat{\alpha}_1^{b}$	1.0048	0.0071	1.0003	1.0049	1.0096
$\hat{\alpha}_1^{pple}$	1.0030	0.0061	0.9991	1.0031	1.0072
$\hat{\alpha}_1^{nls}$	1.0006	0.0200	0.9885	1.0020	1.0156
$\hat{\alpha}_1^{mle}$	0.9997	0.0020	0.9989	0.9998	1.0006
$\hat{\alpha}_2^2$	1.9296	0.0870	1.8836	1.9466	1.9905
$\hat{\alpha}_2^5$	1.9850	0.1237	1.9015	1.9878	2.0651
$\hat{\alpha}_2^8$	2.0521	0.2007	1.9191	2.0354	2.1708
$\hat{\alpha}_2^{11}$	2.1210	0.2915	1.9291	2.0881	2.2858
$\hat{\alpha}_2^{min}$	1.8834	0.1158	1.8187	1.9002	1.9702
$\hat{\alpha}_2^{a}$	2.0540	0.1984	1.9188	2.0371	2.1704
$\hat{\alpha}_2^{b}$	2.0758	0.2277	1.9234	2.0563	2.2063
$\hat{\alpha}_2^{pple}$	2.0953	0.2149	1.9479	2.0756	2.2277
$\hat{\alpha}_2^{nls}$	2.0470	0.2836	1.8565	2.0343	2.2359
$\hat{\alpha}_2^{mle}$	1.9769	0.0596	1.9473	1.9834	2.0027

Table 6
Experiment 6: Design B, Sample Size = 200

Estimator	Mean	St.Dev.	L.Q.	Median	U.Q.
$\hat{\alpha}_1^2$	1.0161	0.0246	1.0005	1.0160	1.0315
$\hat{\alpha}_1^5$	1.0046	0.0077	0.9993	1.0047	1.0102
$\hat{\alpha}_1^8$	1.0025	0.0046	0.9994	1.0025	1.0056
$\hat{\alpha}_1^{11}$	1.0016	0.0032	0.9995	1.0017	1.0038
$\hat{\alpha}_1^{min}$	0.9982	0.0102	0.9977	1.0015	1.0038
$\hat{\alpha}_1^{a}$	1.0039	0.0065	0.9996	1.0038	1.0084
$\hat{\alpha}_1^{b}$	1.0027	0.0048	0.9996	1.0027	1.0061
$\hat{\alpha}_1^{pple}$	1.0018	0.0040	0.9991	1.0019	1.0046
$\hat{\alpha}_1^{nls}$	1.0005	0.0135	0.9921	1.0021	1.0102
$\hat{\alpha}_1^{mle}$	0.9998	0.0008	0.9997	1.0000	1.0001
$\hat{\alpha}_2^2$	1.9634	0.0530	1.9377	1.9718	2.0003
$\hat{\alpha}_2^5$	1.9982	0.0810	1.9386	1.9964	2.0558
$\hat{\alpha}_2^8$	2.0327	0.1262	1.9423	2.0313	2.1188
$\hat{\alpha}_2^{11}$	2.0700	0.1840	1.9386	2.0595	2.1859
$\hat{\alpha}_2^{min}$	1.9322	0.0793	1.8855	1.9487	1.9938
$\hat{\alpha}_2^{a}$	2.0338	0.1274	1.9430	2.0302	2.1176
$\hat{\alpha}_2^{b}$	2.0457	0.1454	1.9401	2.0412	2.1395
$\hat{\alpha}_2^{pple}$	2.0566	0.1430	1.9561	2.0437	2.1509
$\hat{\alpha}_2^{nls}$	2.0301	0.1955	1.9002	2.0326	2.1579
$\hat{\alpha}_2^{mle}$	1.9871	0.0247	1.9785	1.9932	1.9970

Table 7

Experiment 7: Design C, Sample Size = 50

Estimator	Mean	St.Dev.	L.Q.	Median	U.Q.
$\hat{\alpha}_1^2$	1.0343	0.0456	1.0060	1.0357	1.0648
$\hat{\alpha}_1^5$	1.0140	0.0159	1.0041	1.0141	1.0248
$\hat{\alpha}_1^8$	1.0106	0.0104	1.0036	1.0105	1.0174
$\hat{\alpha}_1^{11}$	1.0122	0.0107	1.0052	1.0109	1.0172
$\hat{\alpha}_1^{\mathrm{min}}$	1.0017	0.0207	1.0010	1.0075	1.0125
$\hat{\alpha}_1^{\mathrm{a}}$	1.0213	0.0252	1.0057	1.0222	1.0381
$\hat{\alpha}_1^{\mathrm{b}}$	1.0167	0.0177	1.0057	1.0174	1.0286
$\hat{\alpha}_1^{\mathrm{pple}}$	1.0127	0.0127	1.0046	1.0134	1.0220
$\hat{\alpha}_1^{\mathrm{nls}}$	1.0018	0.0264	0.9846	0.9999	1.0183
$\hat{\alpha}_1^{\mathrm{mle}}$	0.9997	0.0050	0.9983	1.0002	1.0016
$\hat{\alpha}_2^2$	2.0300	0.0914	1.9698	2.0312	2.0938
$\hat{\alpha}_2^5$	2.1743	0.2733	1.9719	2.1631	2.3652
$\hat{\alpha}_2^8$	2.3539	0.5242	1.9655	2.2761	2.6606
$\hat{\alpha}_2^{11}$	2.4680	0.9338	1.8133	2.2506	2.8571
$\hat{\alpha}_2^{\mathrm{min}}$	1.9032	0.2472	1.7656	1.9929	2.0781
$\hat{\alpha}_2^{\mathrm{a}}$	2.1819	0.2846	1.9674	2.1511	2.3521
$\hat{\alpha}_2^{\mathrm{b}}$	2.2580	0.4019	1.9551	2.2088	2.4817
$\hat{\alpha}_2^{\mathrm{pple}}$	2.2126	0.3285	1.9901	2.1799	2.3982
$\hat{\alpha}_2^{\mathrm{nls}}$	2.0457	0.2430	1.8817	2.0291	2.1965
$\hat{\alpha}_2^{\mathrm{mle}}$	1.9618	0.0773	1.9419	1.9761	2.0003

Table 8
Experiment 8: Design C, Sample Size = 100

Estimator	Mean	St.Dev.	L.Q.	Median	U.Q.
$\hat{\alpha}_1^2$	1.0161	0.0333	0.9949	1.0162	1.0384
$\hat{\alpha}_1^5$	1.0067	0.0111	0.9996	1.0064	1.0141
$\hat{\alpha}_1^8$	1.0052	0.0072	1.0005	1.0050	1.0102
$\hat{\alpha}_1^{11}$	1.0058	0.0063	1.0017	1.0055	1.0094
$\hat{\alpha}_1^{min}$	0.9976	0.0175	0.9948	1.0034	1.0076
$\hat{\alpha}_1^{a}$	1.0101	0.0183	0.9981	1.0101	1.0223
$\hat{\alpha}_1^{b}$	1.0079	0.0129	0.9996	1.0078	1.0167
$\hat{\alpha}_1^{pple}$	1.0068	0.0094	1.0004	1.0071	1.0131
$\hat{\alpha}_1^{nls}$	1.0012	0.0187	0.9885	1.0011	1.0137
$\hat{\alpha}_1^{mle}$	0.9999	0.0014	0.9991	0.9999	1.0006
$\hat{\alpha}_2^2$	2.0170	0.0608	1.9788	2.0163	2.0566
$\hat{\alpha}_2^5$	2.0919	0.1896	1.9603	2.0810	2.2167
$\hat{\alpha}_2^8$	2.1733	0.3346	1.9343	2.1193	2.3882
$\hat{\alpha}_2^{11}$	2.2257	0.5366	1.8691	2.1311	2.4596
$\hat{\alpha}_2^{min}$	1.9296	0.1816	1.8512	1.9957	2.0524
$\hat{\alpha}_2^{a}$	2.0916	0.1920	1.9586	2.0697	2.2059
$\hat{\alpha}_2^{b}$	2.1281	0.2640	1.9401	2.0925	2.2746
$\hat{\alpha}_2^{pple}$	2.0948	0.2146	1.9474	2.0716	2.2321
$\hat{\alpha}_2^{nls}$	2.0225	0.1747	1.8981	2.0142	2.1431
$\hat{\alpha}_2^{mle}$	1.9842	0.0204	1.9736	1.9906	1.9979

Table 9

Experiment 9: Design C, Sample Size = 200

Estimator	Mean	St.Dev.	L.Q.	Median	U.Q.
$\hat{\alpha}_1^2$	1.0096	0.0225	0.9940	1.0091	1.0253
$\hat{\alpha}_1^5$	1.0039	0.0076	0.9985	1.0040	1.0093
$\hat{\alpha}_1^8$	1.0029	0.0046	0.9998	1.0030	1.0060
$\hat{\alpha}_1^{11}$	1.0031	0.0038	1.0004	1.0032	1.0056
$\hat{\alpha}_1^{min}$	0.9976	0.0116	0.9940	1.0021	1.0046
$\hat{\alpha}_1^{a}$	1.0059	0.0124	0.9971	1.0059	1.0147
$\hat{\alpha}_1^{b}$	1.0046	0.0087	0.9986	1.0047	1.0106
$\hat{\alpha}_1^{pple}$	1.0038	0.0062	0.9997	1.0040	1.0080
$\hat{\alpha}_1^{nls}$	1.0000	0.0122	0.9915	1.0007	1.0081
$\hat{\alpha}_1^{mle}$	1.0000	0.0005	0.9997	1.0000	1.0002
$\hat{\alpha}_2^2$	2.0117	0.0398	1.9831	2.0129	2.0380
$\hat{\alpha}_2^5$	2.0513	0.1190	1.9638	2.0454	2.1299
$\hat{\alpha}_2^8$	2.0961	0.2090	1.9441	2.0855	2.2285
$\hat{\alpha}_2^{11}$	2.1184	0.3012	1.9034	2.0803	2.3001
$\hat{\alpha}_2^{min}$	1.9526	0.1237	1.8900	2.0070	2.0374
$\hat{\alpha}_2^{a}$	2.0511	0.1201	1.9610	2.0440	2.1273
$\hat{\alpha}_2^{b}$	2.0703	0.1624	1.9499	2.0585	2.1733
$\hat{\alpha}_2^{pple}$	2.0560	0.1425	1.9549	2.0450	2.1501
$\hat{\alpha}_2^{nls}$	2.0101	0.1160	1.9312	2.0133	2.0919
$\hat{\alpha}_2^{mle}$	1.9926	0.0080	1.9893	1.9949	1.9970

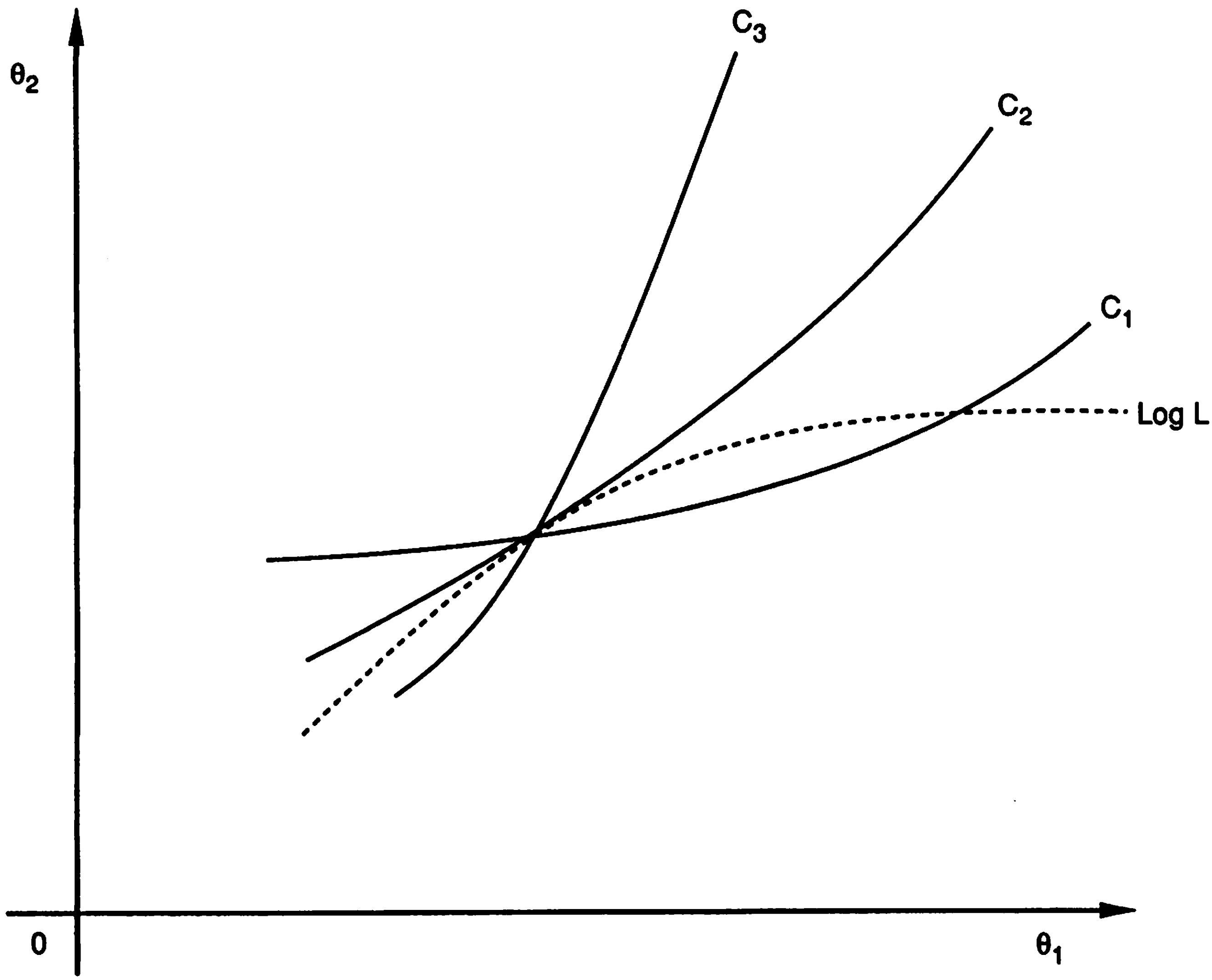

Figure 1

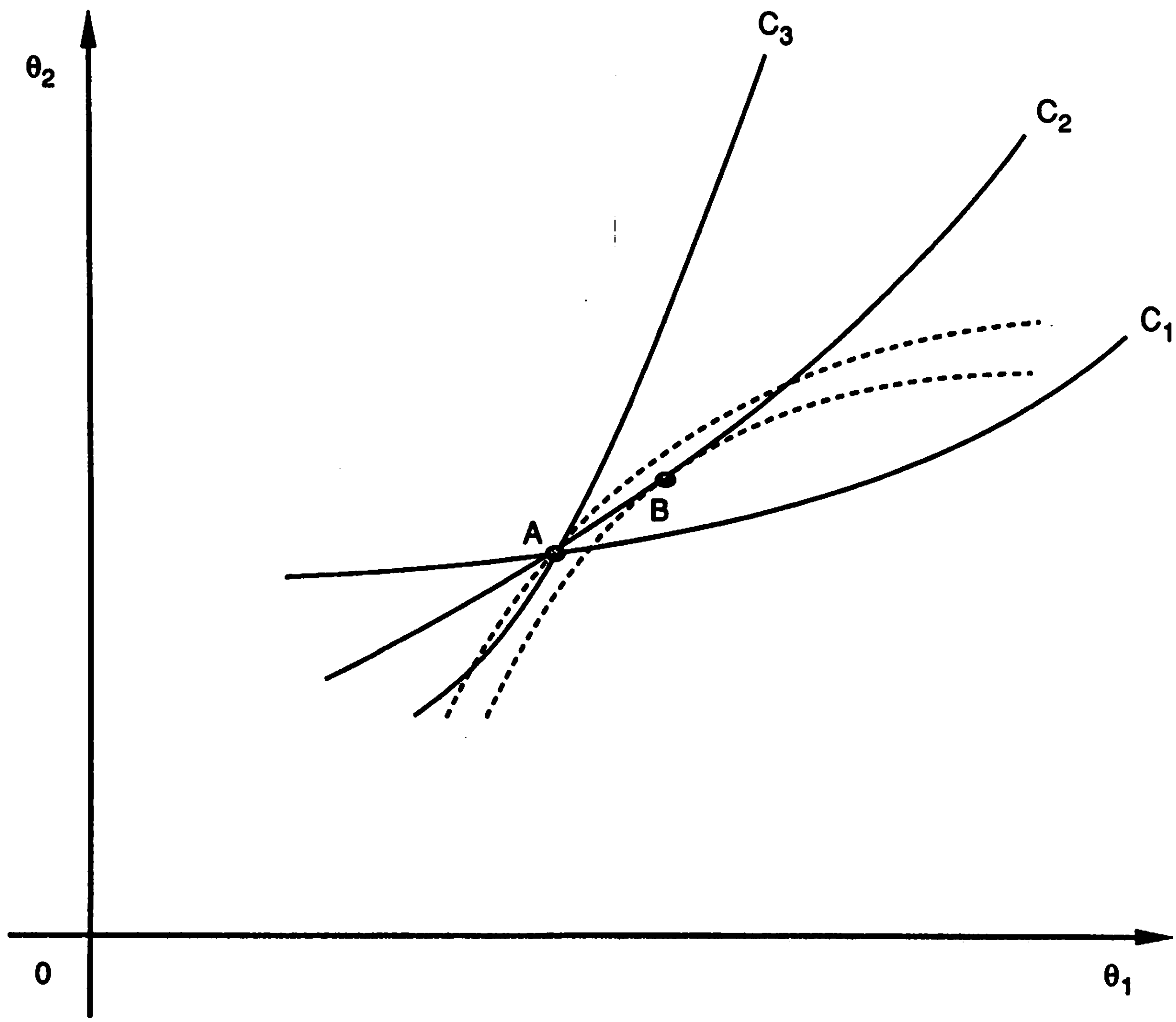

Figure 2

Figure 1a: Empirical Distributions of the alpha_1 Estimators, Experiment 1, Design A, Sample Size = 50

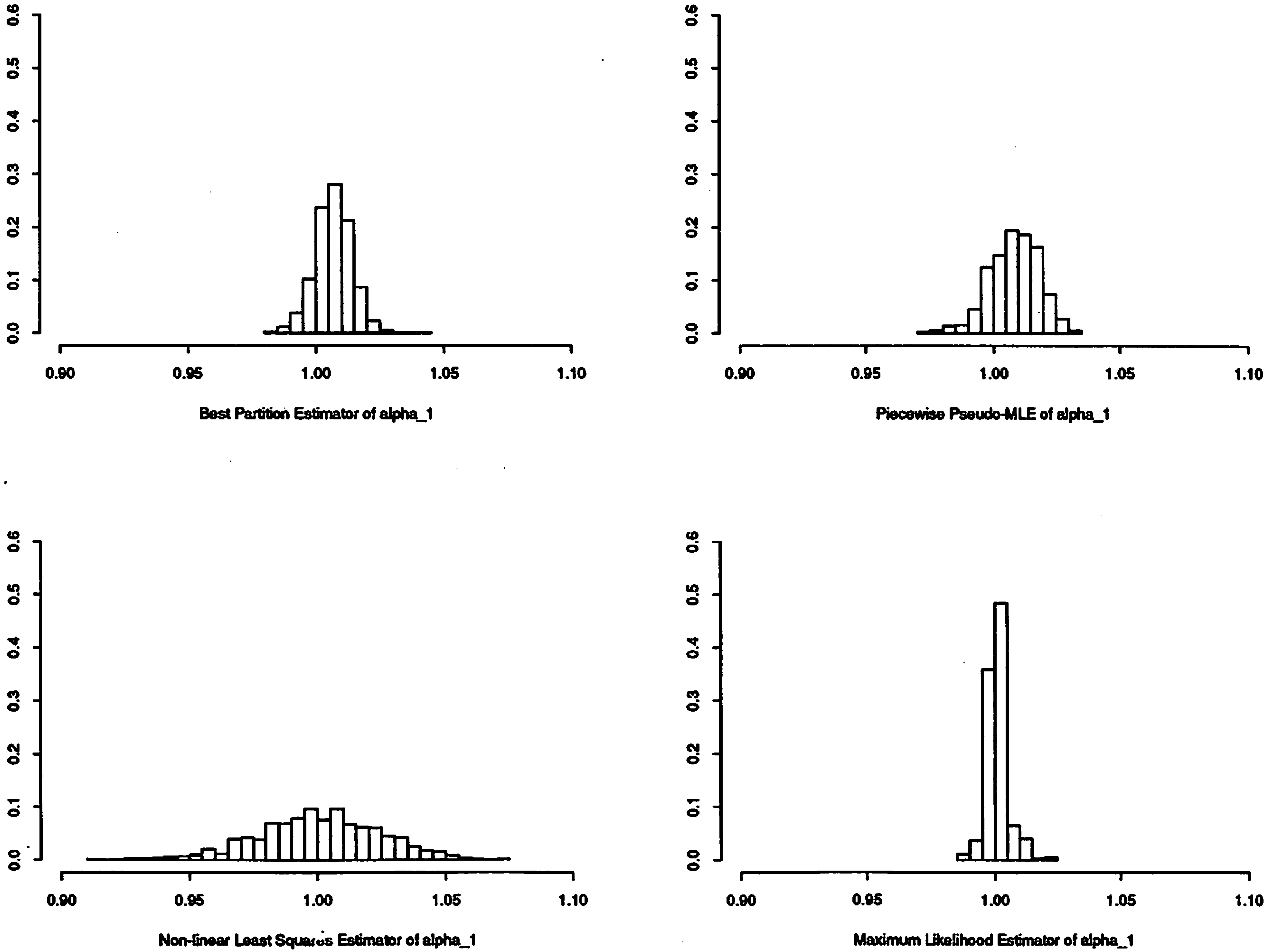

Figure 1b: Empirical Distributions of the alpha_2 Estimators, Experiment 1, Design A, Sample Size = 50

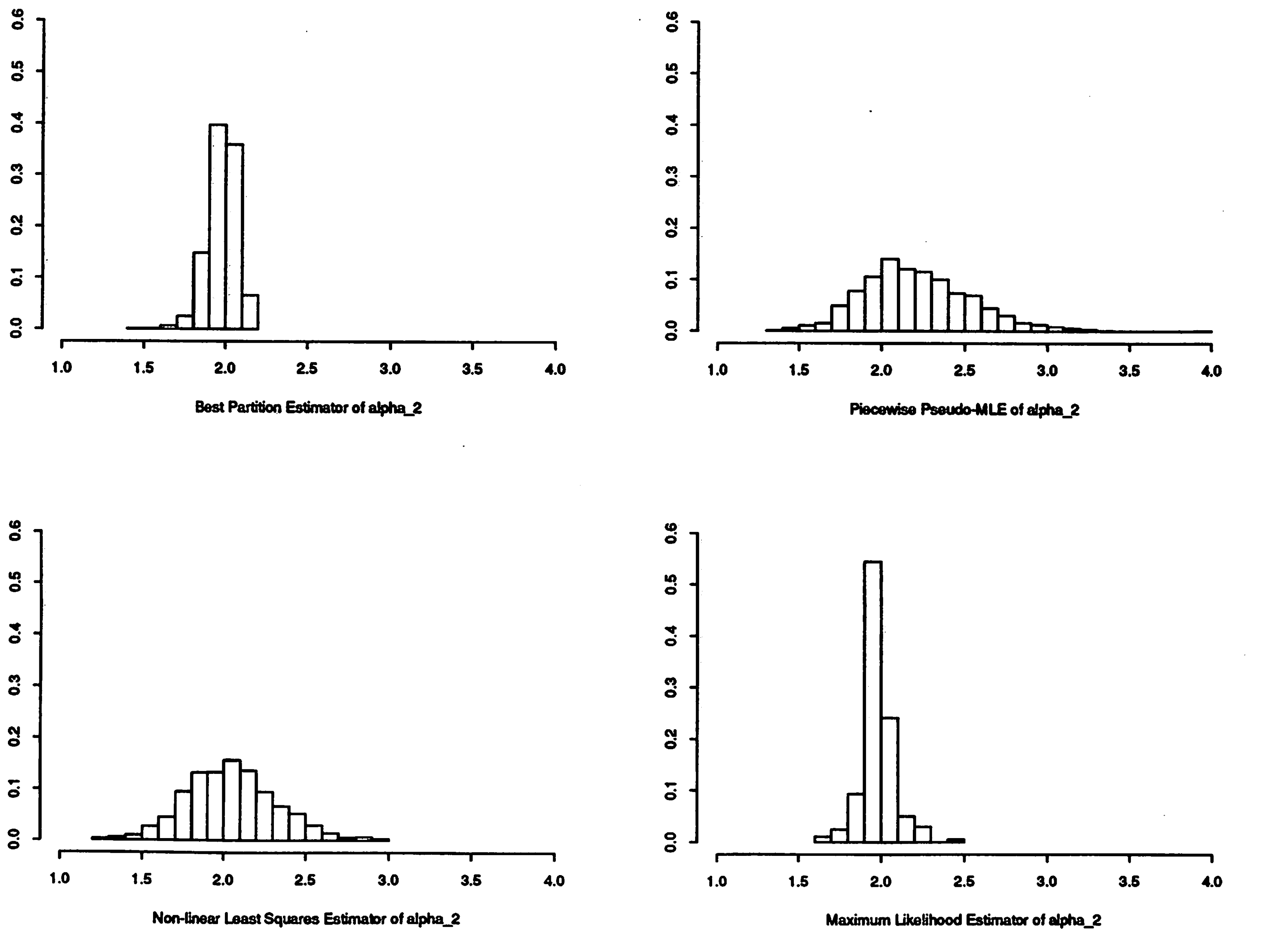

Figure 2a: Empirical Distributions of the alpha_1 Estimators, Experiment 2, Design A, Sample Size = 100

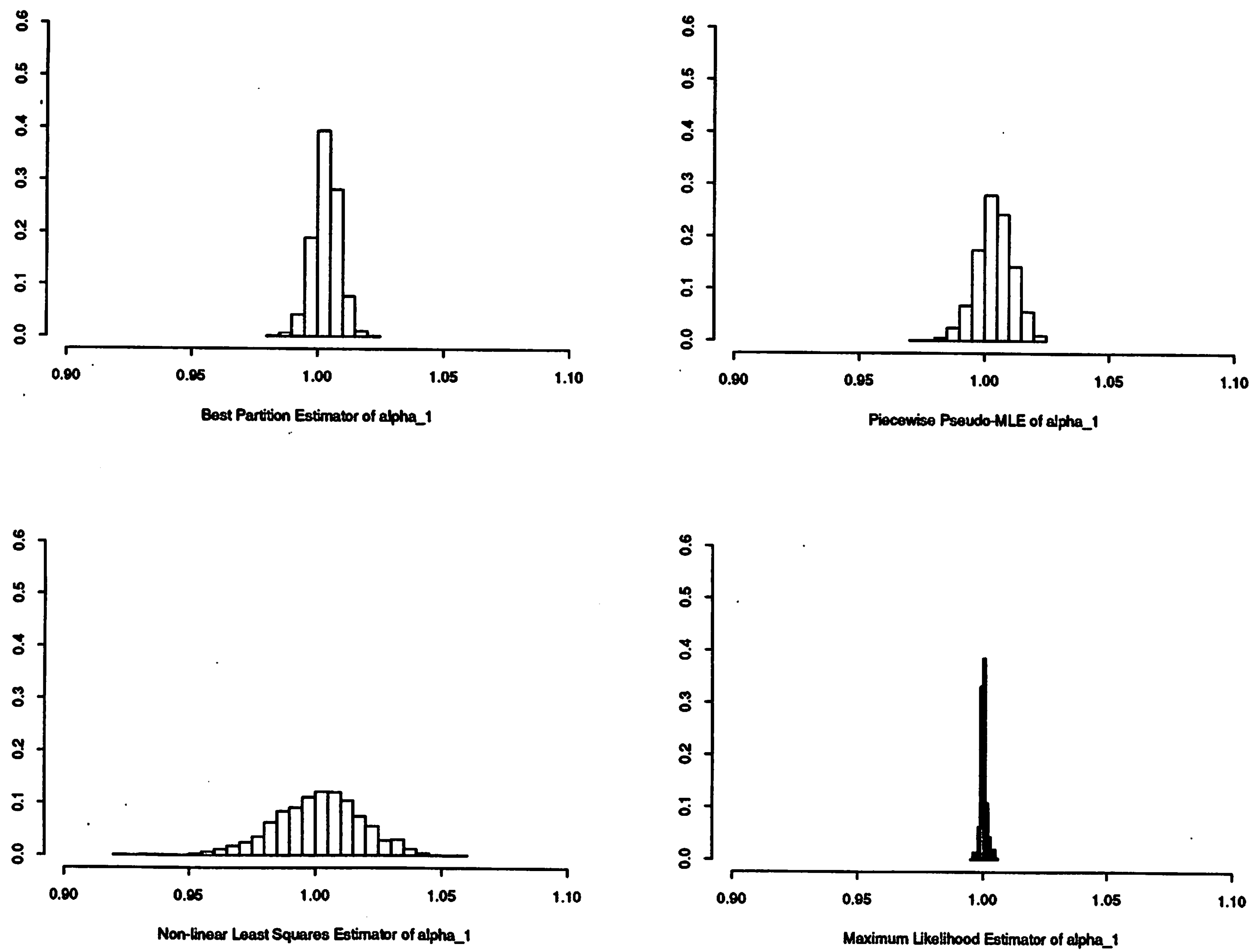

Figure 2b: Empirical Distributions of the alpha_2 Estimators, Experiment 2, Design A, Sample Size = 100

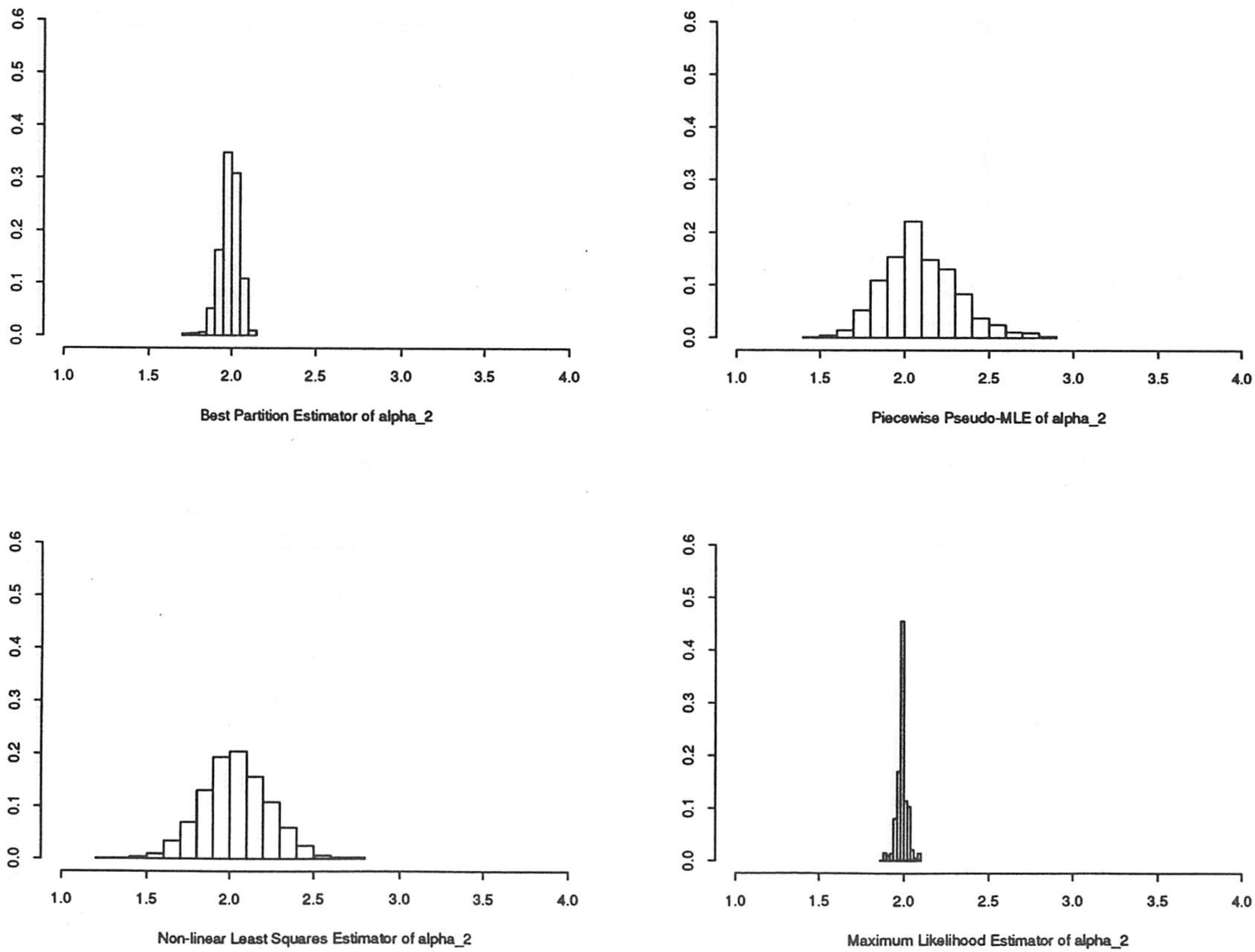

Figure 3a: Empirical Distributions of the alpha_1 Estimators, Experiment 3, Design A, Sample Size = 200

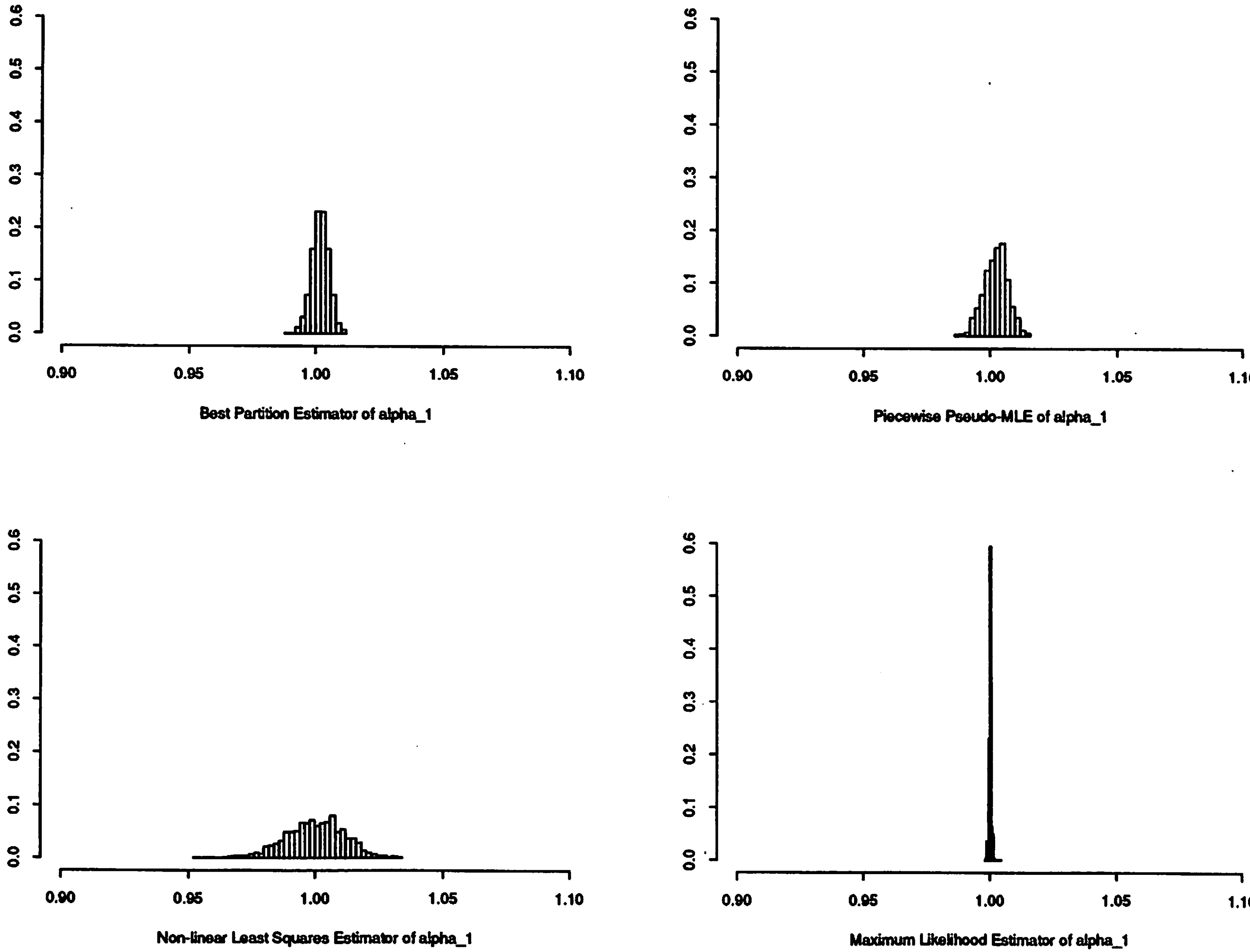

Figure 3b: Empirical Distributions of the alpha_2 Estimators, Experiment 3, Design A, Sample Size = 200

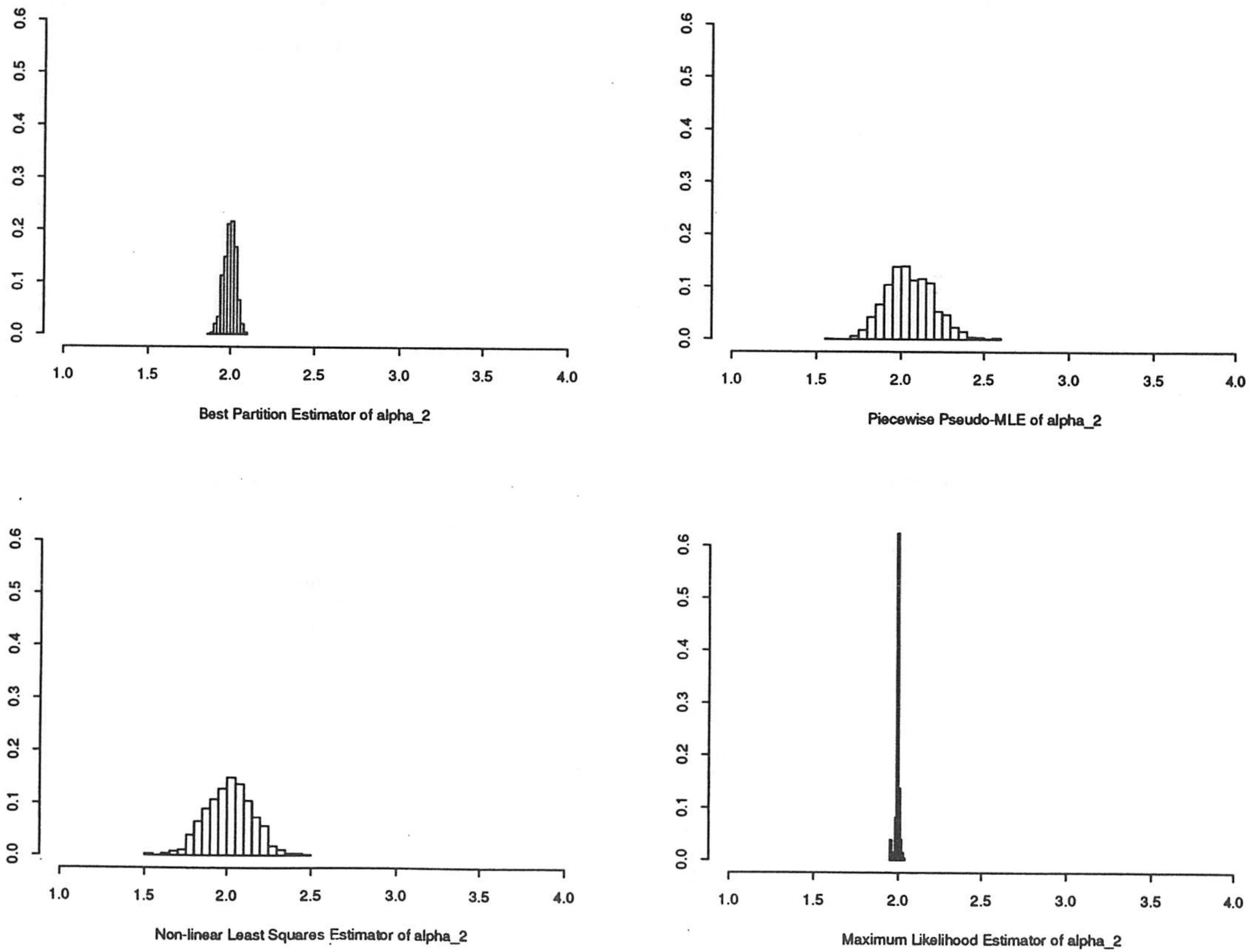

Figure 4a: Empirical Distributions of the alpha_1 Estimators, Experiment 4, Design B, Sample Size = 50

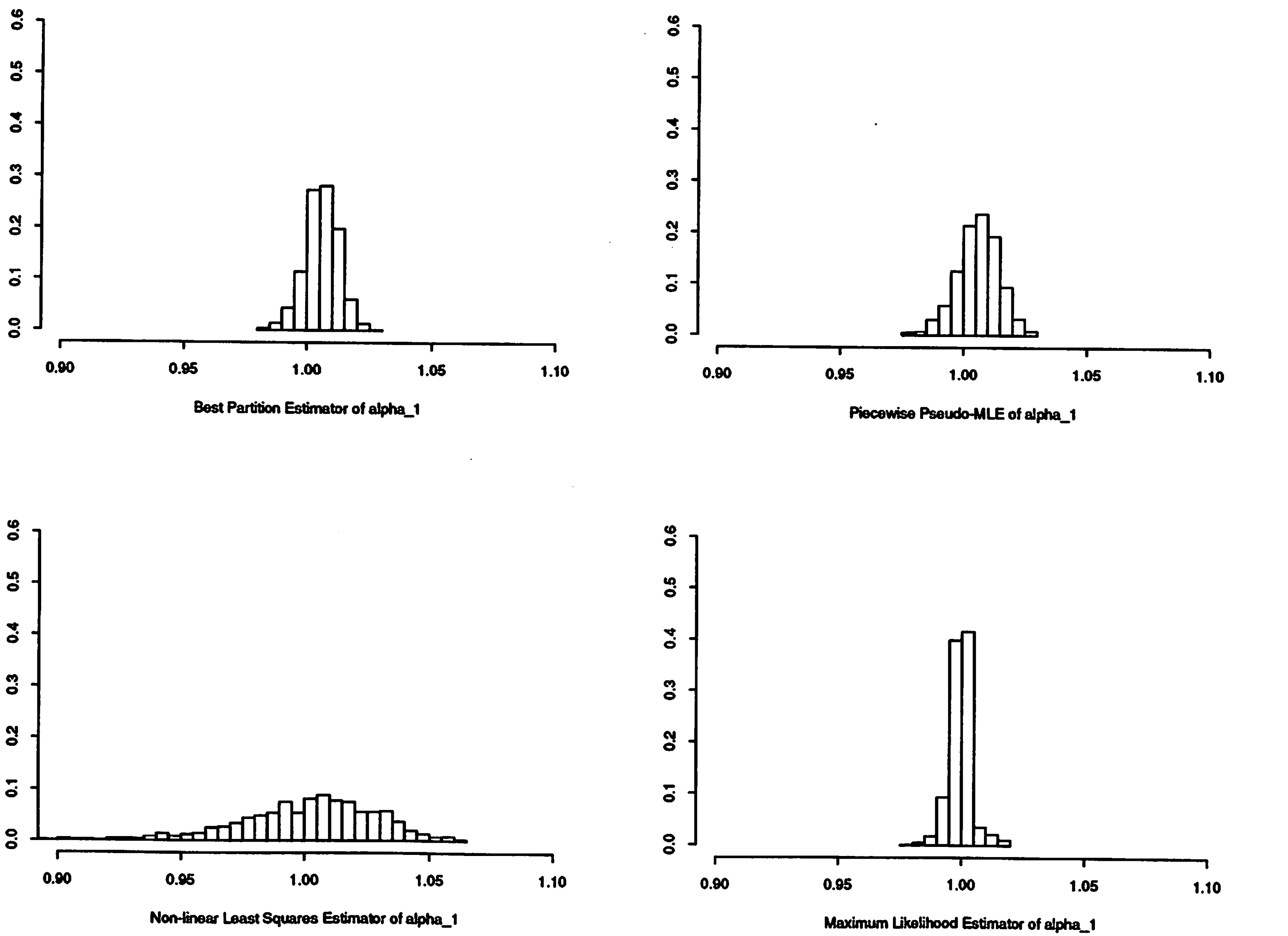

Figure 4b: Empirical Distributions of the alpha_2 Estimators, Experiment 4, Design B, Sample Size = 50

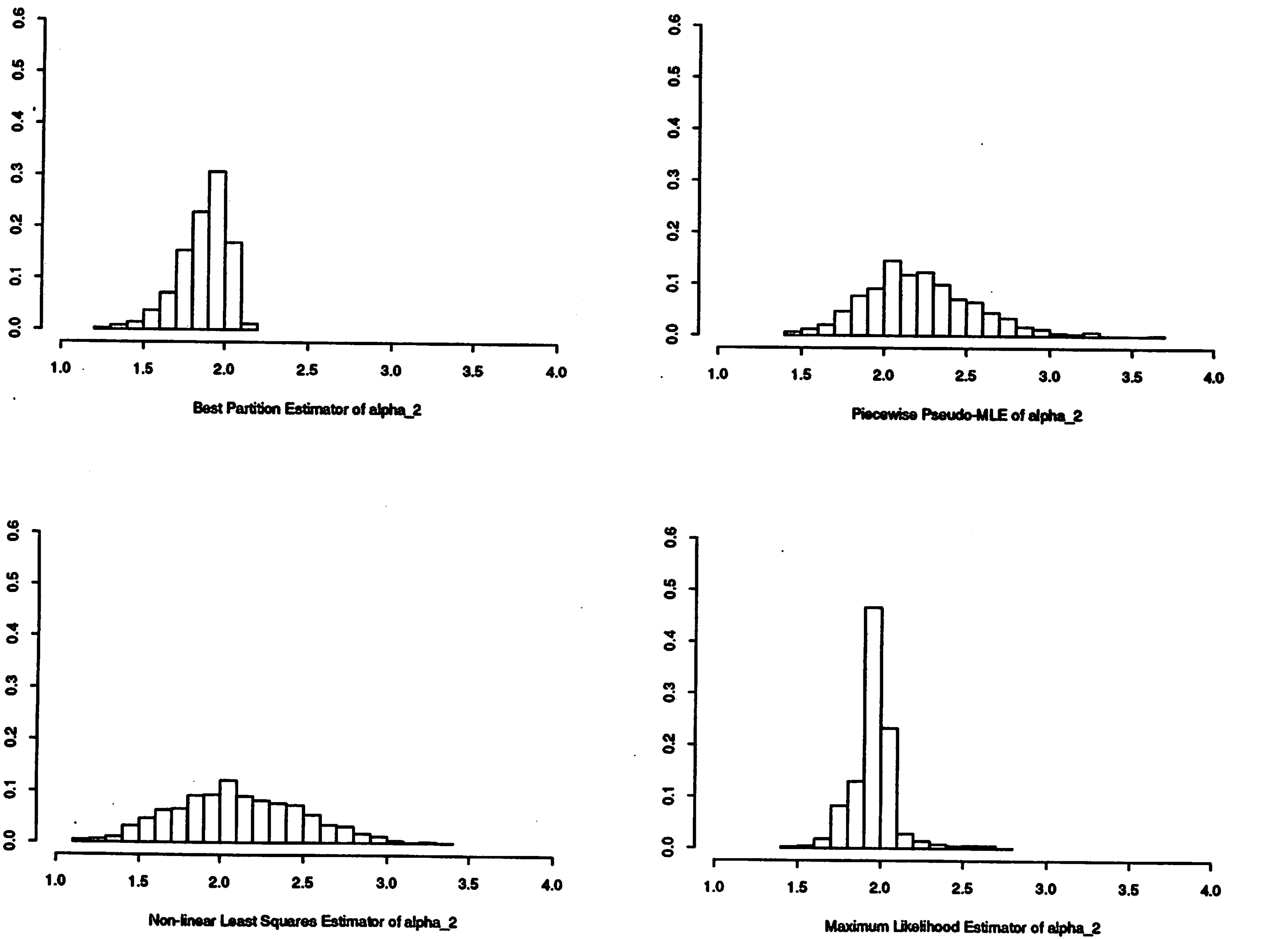

Figure 5a: Empirical Distributions of the alpha_1 Estimators, Experiment 5, Design B, Sample Size = 100

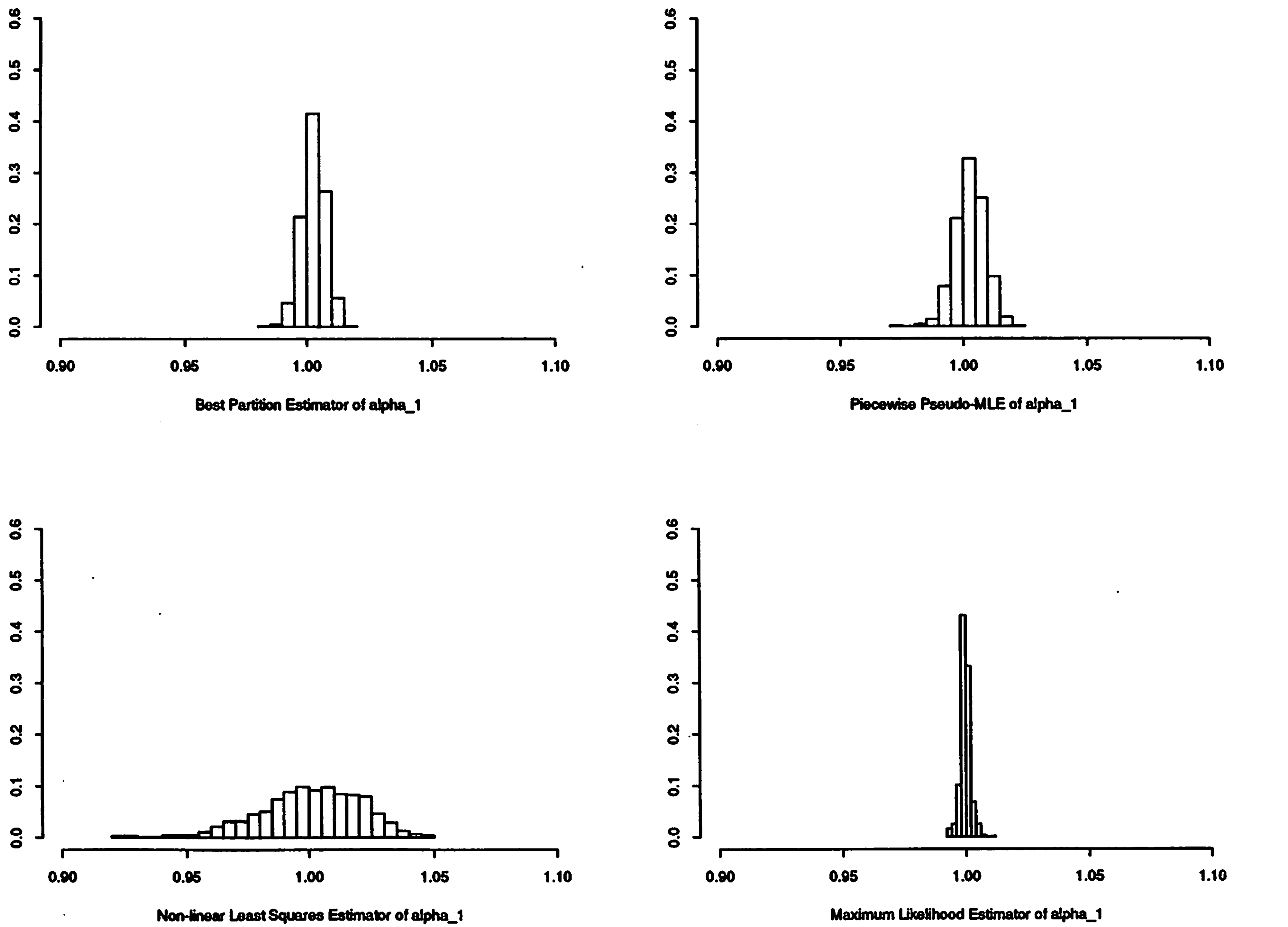

Figure 5b: Empirical Distributions of the alpha_2 Estimators, Experiment 5, Design B, Sample Size = 100

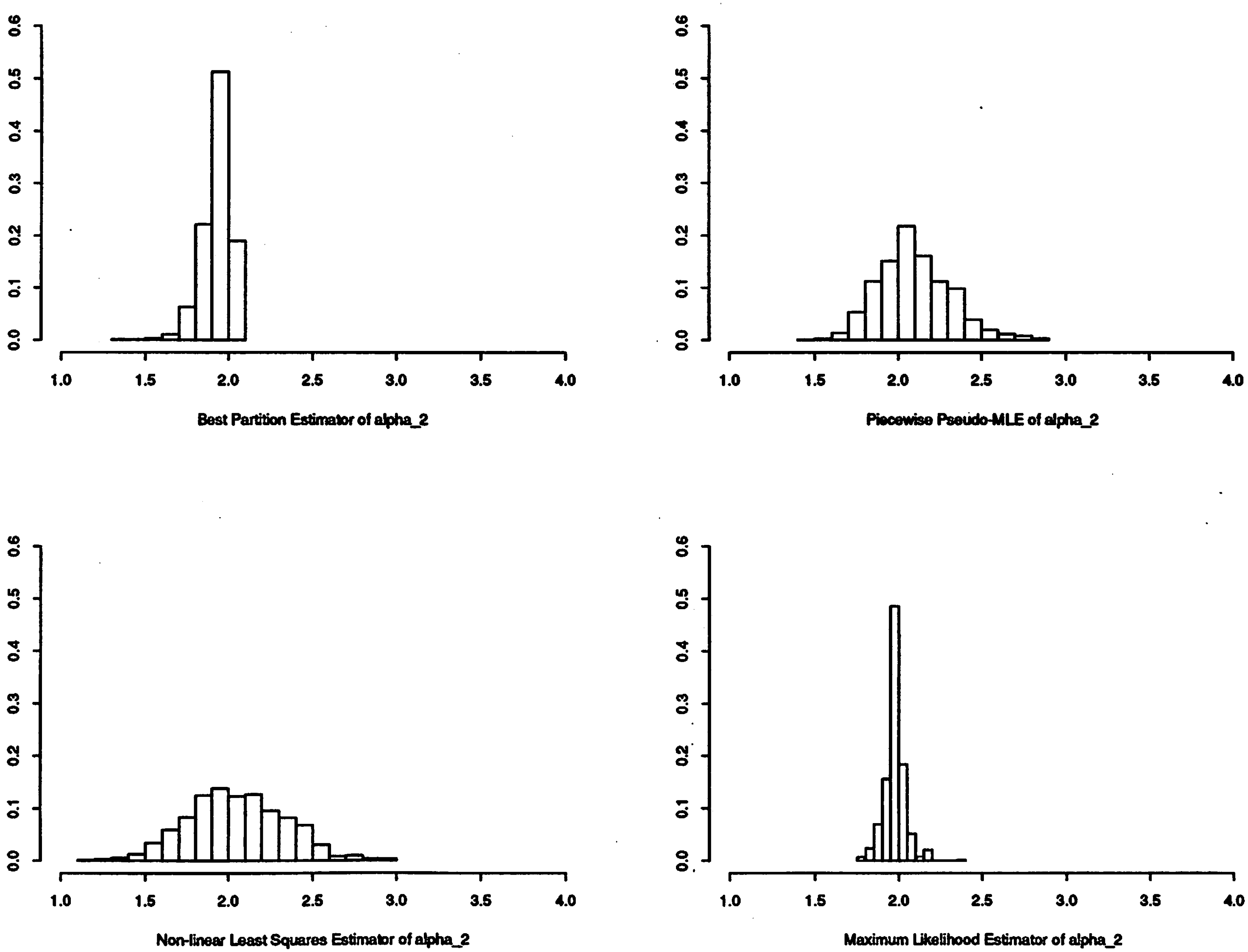

Figure 6a: Empirical Distributions of the alpha_1 Estimators, Experiment 6, Design B, Sample Size = 200

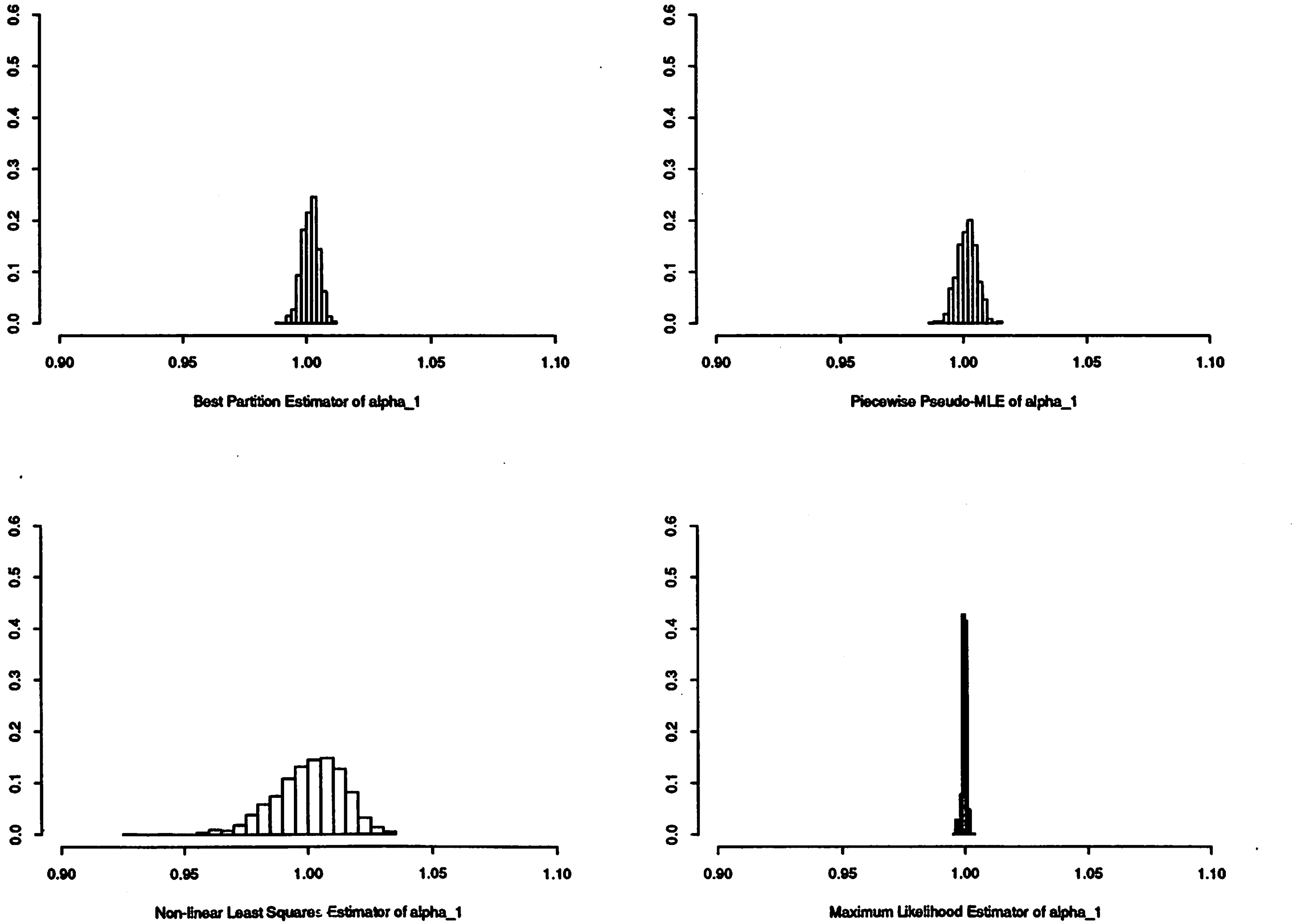

Figure 6b: Empirical Distributions of the alpha_2 Estimators, Experiment 6, Design B, Sample Size = 200

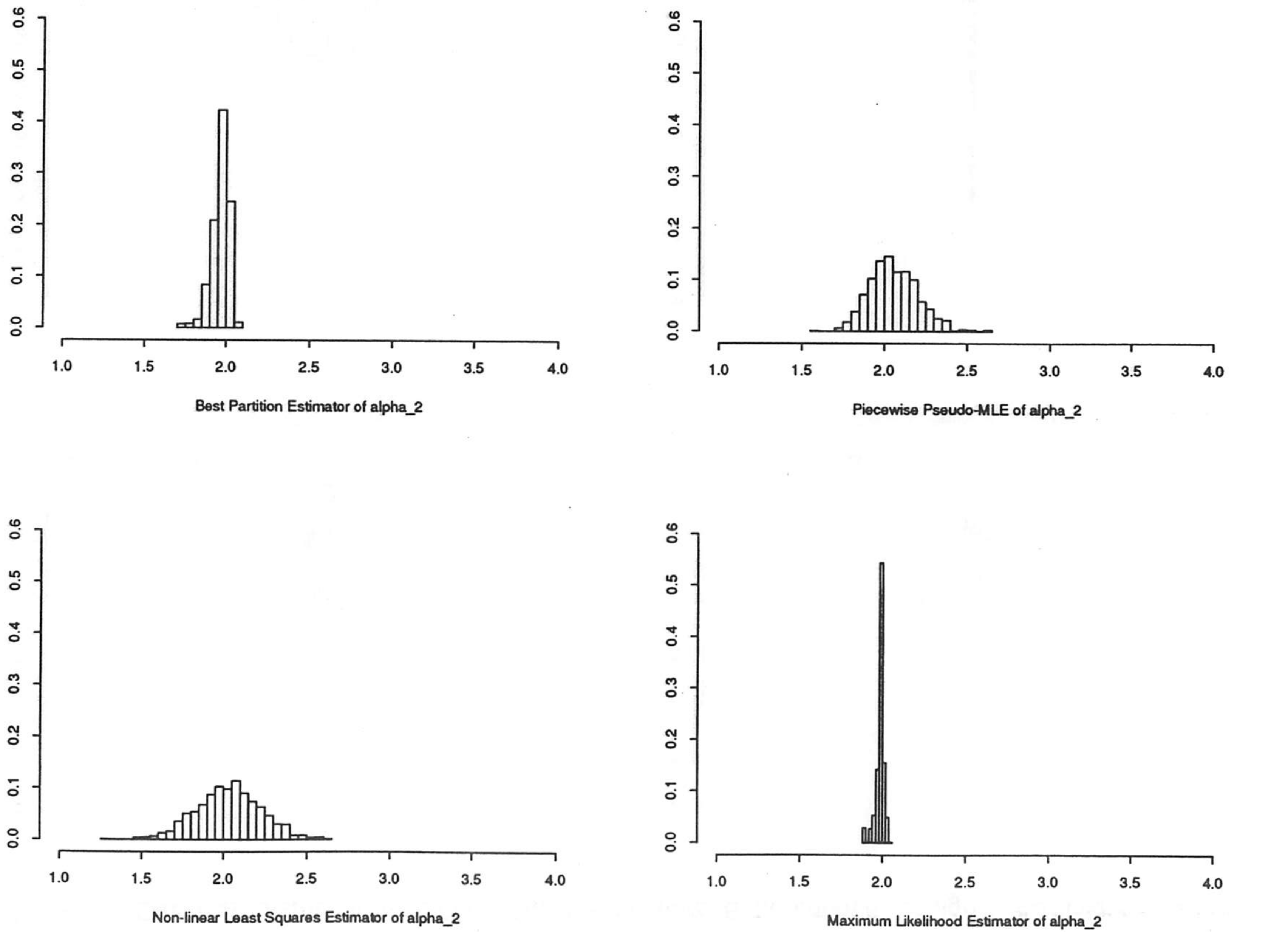

Figure 7a: Empirical Distributions of the alpha_1 Estimators, Experiment 7, Design C, Sample Size = 50

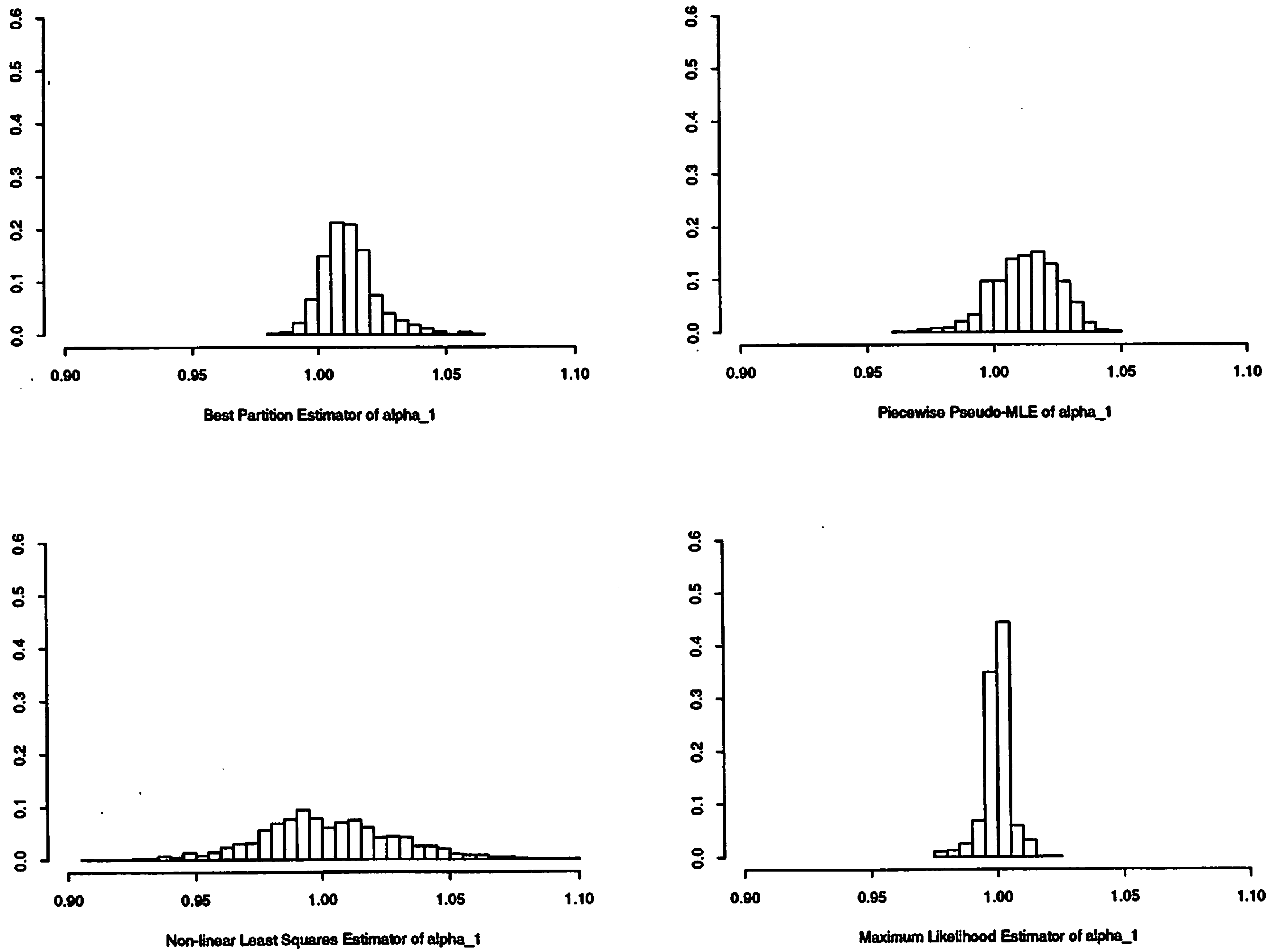

Figure 7b: Empirical Distributions of the alpha_2 Estimators, Experiment 7, Design C, Sample Size = 50

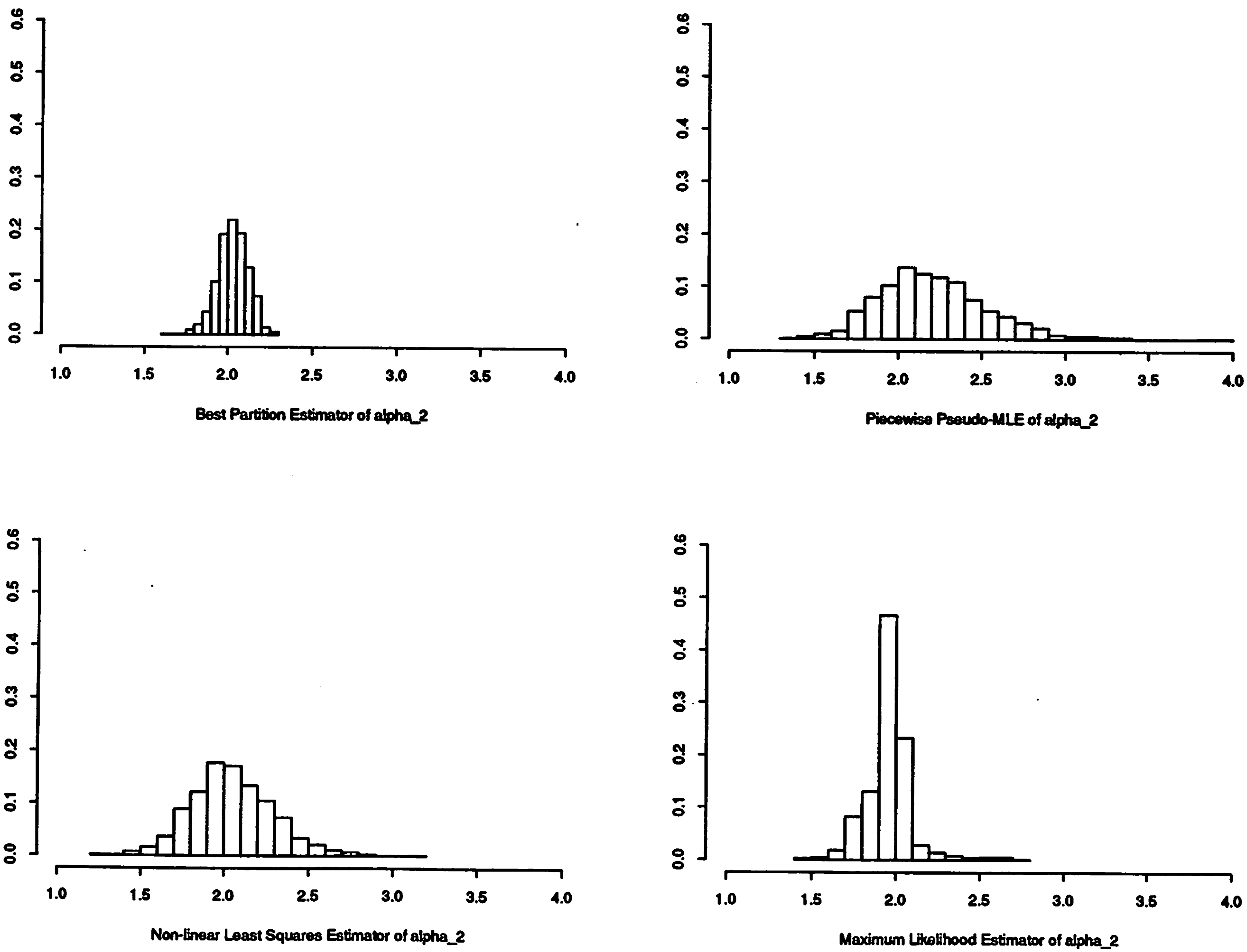

Figure 8a: Empirical Distributions of the alpha_1 Estimators, Experiment 8, Design C, Sample Size = 100

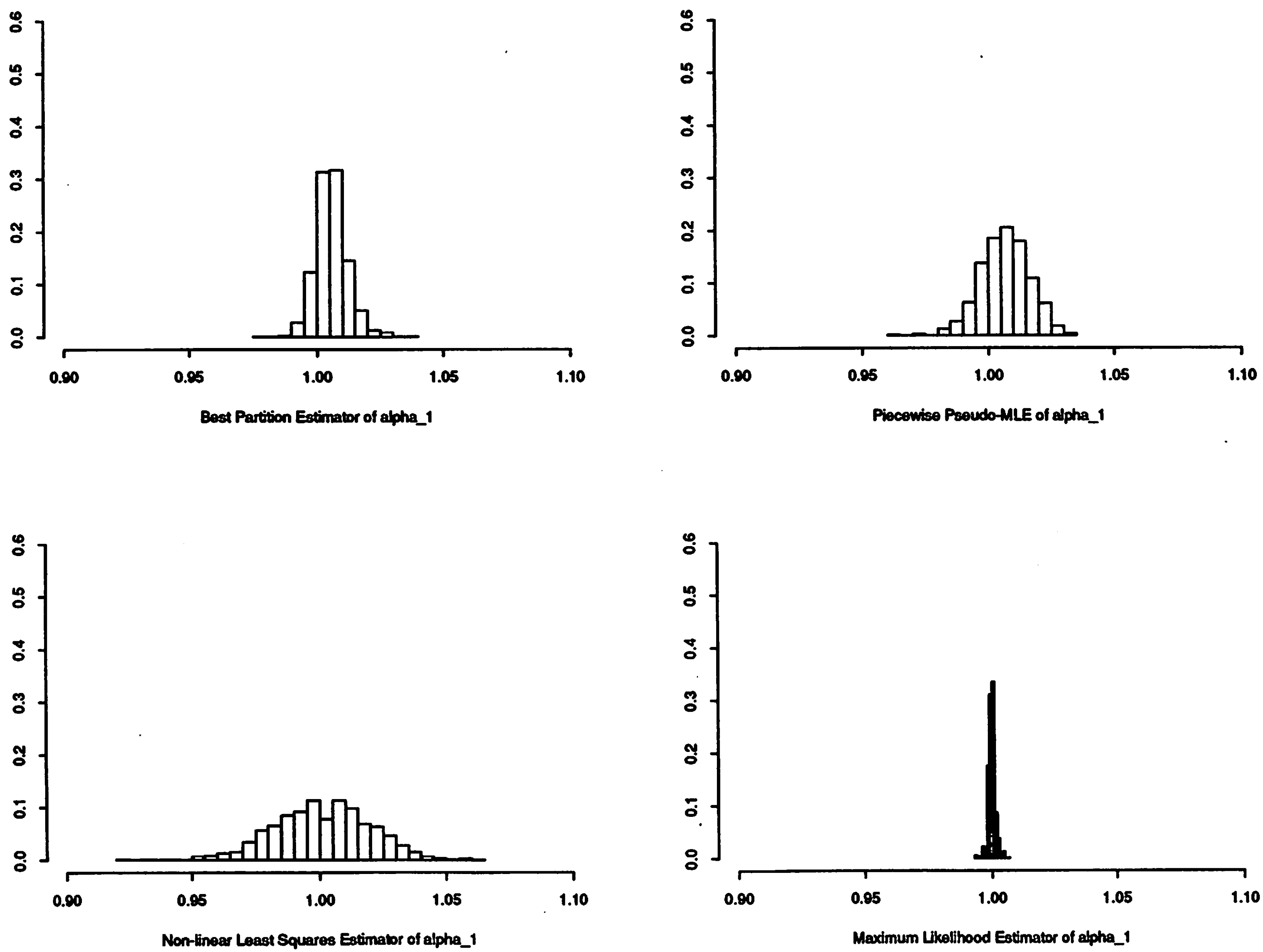

Figure 8b: Empirical Distributions of the alpha_2 Estimators, Experiment 8, Design C, Sample Size = 100

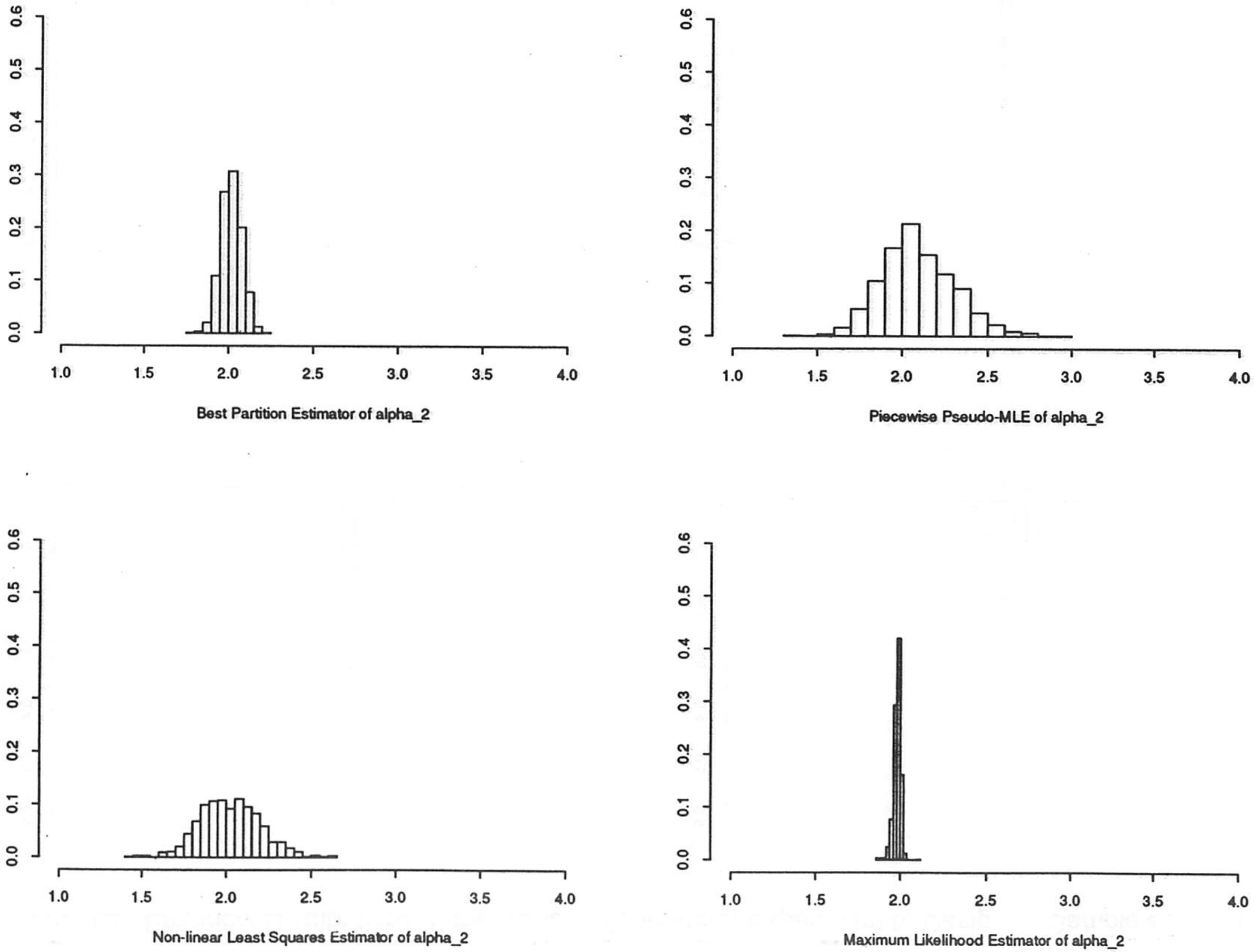

Figure 9a: Empirical Distributions of the alpha_1 Estimators, Experiment 9, Design C, Sample Size = 200

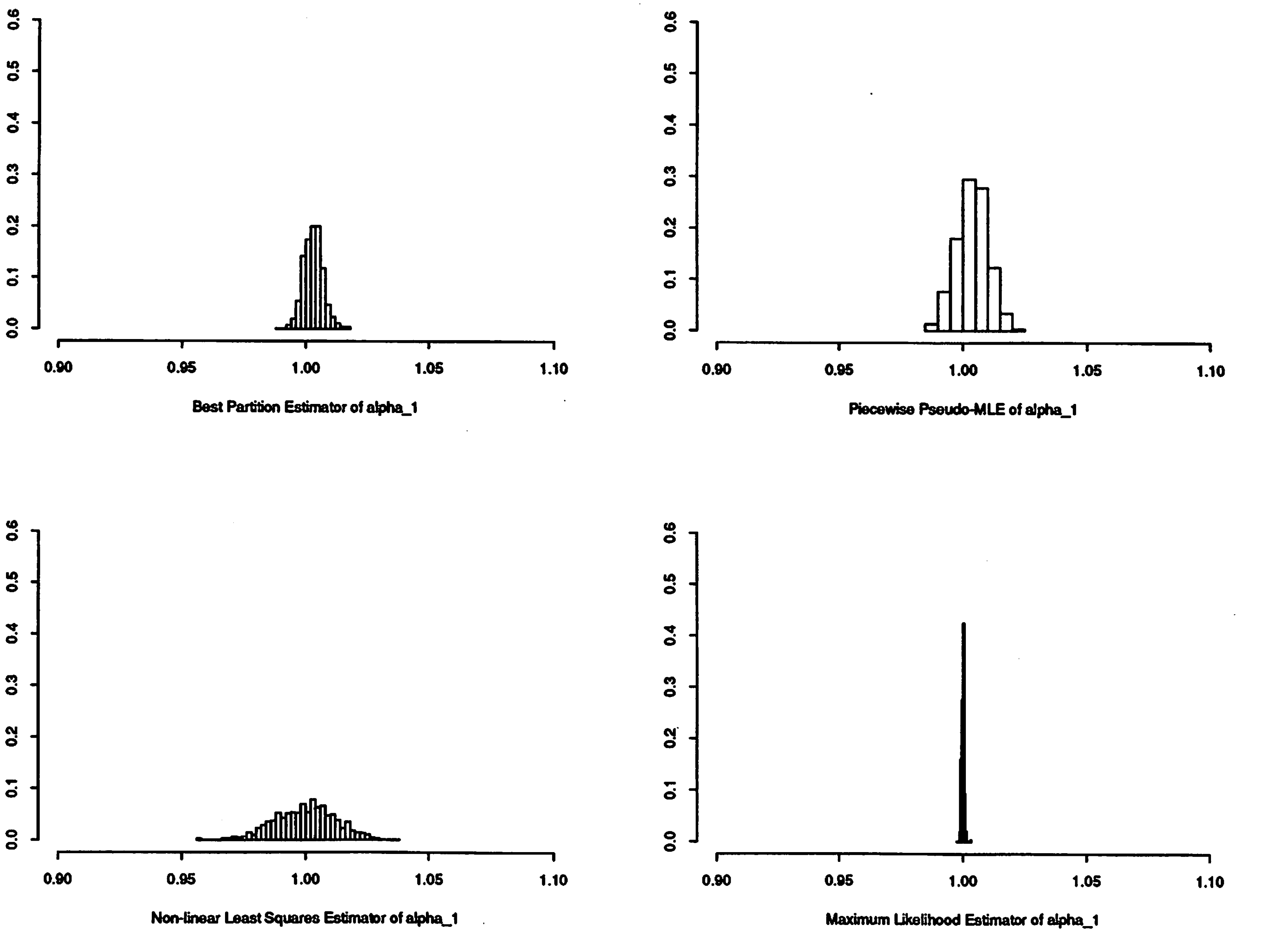

Figure 9b: Empirical Distributions of the alpha_2 Estimators, Experiment 9, Design C, Sample Size = 200