Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support

Keisuke Hirano
University of Miami and UCLA

Jack R. Porter
Harvard University

November 26, 2001

Preliminary and Incomplete

Abstract

In certain auction, search, and related models, the boundary of the support of the observed data depends on some of the parameters of interest. For such nonregular models, standard asymptotic distribution theory does not apply. Previous work has focused on characterizing the nonstandard limiting distributions of particular estimators in these models. In contrast, we study the problem of constructing efficient point estimators. We show that the ML estimator is generally inefficient, but that the Bayes estimator is efficient according to the usual local asymptotic minmax criterion. We provide intuition for this result using Le Cam's limits of experiments framework.

1 Introduction

This paper studies efficient point estimation of structural econometric models in which the boundary of the support of the observed data depends on some of the parameters of interest, and on regressor variables. Some leading examples of such models are certain parametric auction models, search models, and production frontier models. These models violate standard regularity conditions, so that conventional asymptotic theory does not apply.

Previous work on asymptotics for models with parameter-dependent support includes Flinn and Heckman (1982), Smith (1985), Christensen and Kiefer (1991), Donald and Paarsch (1993), Hong (1998), Donald and Paarsch (1999), and Chernozhukov and Hong (2001). These papers derive large-sample properties of specific estimators for various models. Most of these papers focus on ML estimators. Donald and Paarsch (1993) and Donald and Paarsch (1999) also consider some alternative analog-type estimators, and Chernozhukov and Hong (2001) also consider Bayes estimators. From this earlier work, we know that different estimators have different limiting distributions, so we focus on making efficiency comparisons and on identifying a class of optimal estimators.

We use the standard local asymptotic minmax criterion for optimality. This criterion compares estimators by their maximum expected loss over a localized parameter space. In regular finite-dimensional parametric models, this criterion coincides with other familiar optimality definitions (for bowl-shaped loss functions) and leads to the conclusion that the maximum likelihood estimator is optimal. However, inspection of proofs of the efficiency of the ML estimator show that this property is quite closely tied to regularity of the underlying model. Because the models we consider here are nonregular, there is no guarantee that ML will be efficient. In fact, we show that for standard loss functions such as squared error loss, ML is generally *inefficient* in models with parameter-dependent support.

We then consider Bayes estimators, which provide an alternative, likelihood-based method of inference in parametric models. Recent work on Bayesian inference for search and auction models includes Lancaster (1997), Kiefer and Steele (1998), Bajari (1998), Sareen (2000), and Chernozhukov and Hong (2001). In regular parametric models, Bayes estimators are typically asymptotically equivalent to ML (see for example Ibragimov and Hasminskii (1981)). Hence, Bayes estimators are also efficient in regular models. In nonregular models, ML and Bayes estimators are no longer necessarily asymptotically equivalent. We show that, for the class of nonregular models where the boundary of the support depends on at least some of the parameters, Bayes estimators are efficient. Thus Bayes estimators remain efficient under this form of nonregularity, while ML loses its efficiency properties.

We develop intuition for our result on the efficiency of Bayes estimators by applying Le Cam's theory of limits of statistical experiments to various special cases. Using the limits of experiments approach, we can characterize the entire class of attainable limit distributions for estimators in a given model. In the case where the covariates have a discrete distribution, the model we consider is asymptotically equivalent to a simpler model consisting of a vector of draws from shifted exponential distributions. Asymptotic equivalence means that any limit distribution for a statistic in our model of interest can be obtained as the *exact* distribution of a statistic in the shifted exponential model. This result is useful, because the exponential shift experiment has a simple structure which can be exploited to verify optimality of certain estimators. The shifted exponential limit experiment is equivariant under a group of transformations. Under certain conditions on the group of transformations and the loss function, a generalized Bayes procedure with respect to a flat prior is both minimum risk equivariant and minmax. Since Bayes estimators in the original nonregular model have a limiting distribution equal to the distribution of the flat-prior Bayes estimator in the limit experiment, the Bayes procedure in the nonregular model is locally asymptotically minmax.

Our findings are closely related to earlier work on simpler nonregular models without covariates, in particular models for i.i.d. sampling from univariate densities with jumps (Ibragimov and Hasminskii (1981), Pflug (1983), Ghosal and Samanta (1995)). Our work extends these results to the case where there are regressor variables that can shift the support of the outcome variable (as well as affect the shape of the outcome distribution in other ways). Allowing for covariates leads to more complicated limit experiments and requires an extension of the asymptotic results on Bayes estimators in Ibragimov and Hasminskii (1981).

In the next section, we consider a special case of our general model: the experiment of observing n independent and identically distributed draws from a uniform distribution on the interval $[0, \theta]$. This model has been well studied, but is useful for introducing notation, discussing the general limits of experiments framework, and providing intuition for our later results. Moreover, the inefficiency of ML can be seen quite easily in this case. In section 3 we consider a more general model where the support of an outcome can depend on both parameters and covariates. Our first step is to provide a limits of experiments characterization of such models in cases where the covariates are discrete. This directly yields intuition for optimality of the Bayes estimator, using the group structure of the limit experiment. Then, in section 4, we study the asymptotics of Bayes estimators and provide a general efficiency result. Section 5 concludes.

2 Uniform Model

Under standard classical conditions, maximum likelihood estimators are consistent, asymptotically normal, and efficient. A well known example of a model where the classical conditions do not hold, is the experiment of observing n random samples from a uniform distribution on the interval $[0, \theta]$, for $\theta \in \Theta \subset \mathbb{R}_{++}$. Here, maximum likelihood estimation of θ is still consistent, but not asymptotically normal or efficient. We use this example to illustrate the inefficiency of ML, and to show that an efficiency bound can still be obtained in this nonregular case and that Bayes estimators attain the bound. The intuition from the uniform model will carry over in a very direct way to the more general models considered in section 3.

2.1 Maximum Likelihood Estimator

Let $Z_1, \ldots, Z_n$ be an i.i.d. sample from $U[0, \theta]$, the uniform distribution with density $p(Z|\theta) = \mathbf{1}\{0 \leq Z \leq \theta\}/\theta$, where $\mathbf{1}\{A\}$ is the indicator function for the event A. The likelihood function is

$$p(Z_1, \ldots, Z_n|\theta) = \left(\frac{1}{\theta}\right)^n \mathbf{1}\{0 \leq Z_{(n)} \leq \theta\},$$

where $Z_{(n)}$ denotes the nth order statistic, i.e. the sample maximum. The maximum likelihood estimator is simply $\hat{\theta}_{ML} = Z_{(n)}$. It is straightforward to derive its limiting distribution by direct calculations:

$$n(\hat{\theta}_{ML} - \theta) \rightsquigarrow -Exp\left(\frac{1}{\theta}\right),$$

where $Exp(1/\theta)$ denotes an exponentially distributed random variable with hazard rate $1/\theta$, and $\rightsquigarrow$ denotes convergence in distribution. Clearly, the estimator is not asymptotically normal. Although it converges at rate n, much faster than the usual $\sqrt{n}$ rate, the fact that the limiting distribution lies completely to one side of the true parameter suggests that even better estimators may exist.

2.2 Bayes Estimator

Bayesian estimation of θ provides an alternative approach to maximum likelihood estimation. Given a prior $\pi(\theta)$ on Θ, the posterior distribution is given by Bayes Theorem as:

$$p(\theta|Z_1, \ldots, Z_n) = \frac{\pi(\theta)p(Z_1, \ldots, Z_n|\theta)}{\int \pi(\theta)p(Z_1, \ldots, Z_n|\theta)d\theta}.$$

Given a loss function $l(\theta, a)$, the Bayes estimate chooses a to minimize posterior expected loss

$$E[l(\theta, a)|Z_1, \ldots, Z_n] = \int l(\theta, a)p(\theta|Z_1, \ldots, Z_n)d\theta.$$

Here, a is interpreted as an estimate of θ. For a given prior π and loss l, the corresponding Bayes estimator can be regarded as a decision rule that takes the observed values of $Z_1, \ldots, Z_n$ and produces an estimate of θ. So we can study its frequency properties. For example, suppose we choose squared error loss, $l(\theta, a) = (a - \theta)^2$, and the (improper) prior $\pi(\theta) = \mathbf{1}\{0 < \theta\}/\theta^2$. Then the posterior density can be calculated to be $p(\theta|Z_1, \ldots, Z_n) = (n+1)Z_{(n)}^{n+1}\mathbf{1}\{Z_{(n)} \leq \theta\}/\theta^{n+2}$. The Bayes estimator for squared error loss is the posterior mean, $\tilde{\theta}_B = Z_{(n)}(n+1)/n$. As with the maximum likelihood estimator, the limiting distribution can be derived directly:

$$n(\tilde{\theta}_B - \theta) \rightsquigarrow \theta - Exp\left(\frac{1}{\theta}\right).$$

The convergence rate is the same as maximum likelihood estimator, but now the limiting distribution is centered (in the sense that it has mean zero). It follows that the Bayes estimator will dominate the ML estimator for squared error loss. It can also be shown that using a different prior typically does not change the asymptotic distribution of the Bayes estimator, because the prior is dominated by the likelihood as the sample size increases.

In the case of squared error loss, the Bayes estimator can be interpreted as a bias-corrected version of ML.* However, this interpretation does not generally hold for other loss functions. Moreover, the reasoning so far does not guarantee that the Bayes estimator will be efficient among all estimators even for squared error loss. The limits of experiments framework, described next, allows us to obtain stronger results on the efficiency of Bayes estimators for a wide class of loss functions.

2.3 Limits of Experiments

The limits of experiments theory is an approximation theory for statistical models rather than for estimators within a given model. It provides a parsimonious description of the entire set of attainable limit distributions among estimators in the statistical model. This description, in turn, can often suggest the form of optimal estimators.

An *experiment* is defined as a measurable space (the sample space) along with a collection of probability measures on that space indexed by a parameter h; we denote this by $(\mathcal{Z}, \mathcal{A}, P_h : h \in H)$. The experiment is interpreted as the situation where we observed a random variable Z on a measurable space $(\mathcal{Z}, \mathcal{A})$, distributed as P_h for some h in a parameter space H. We use h to denote a local parameter, related to the original model by $\theta + \psi_n h$ for some fixed θ in the original parameter space and a normalization sequence $\psi_n \longrightarrow 0$. In regular cases $\psi_n = \frac{1}{\sqrt{n}} I$, where I is the identity matrix, while in the nonregular cases we consider here, some of the diagonal elements of ψ_n are $1/n$ rather than $1/\sqrt{n}$.

The *likelihood ratio process* based at $h_0 \in H$ is defined as

$$\left(\frac{dP_h}{dP_{h_0}}(Z) \right)_{h \in H}.$$

Because it depends on the random variable Z, it can be regarded as a stochastic process defined on H. A sequence of experiments $\mathcal{E}_n = (P_{n,h} : h \in H)$ is said to converge to the experiments $\mathcal{E} = (P_h : h \in H)$ if the finite dimensional distributions of the likelihood ratio process converge to the corresponding distributions of the likelihood ratio process for $\mathcal{E}$, i.e. for every finite subset

*Cavanagh, Jones, and Rothenberg (1990) consider bias-corrected ML estimators in regular models under general loss functions. They show that bias-corrected ML (with the bias-correction depending on the loss function) is efficient among asymptotically normal estimators, but do not consider more general estimators or nonregular models.

$I \subset H$ and every $h_0 \in H$,

$$\left(\frac{dP_{n,h}}{dP_{n,h_0}}\right)_{h \in I} \overset{h_0}{\rightsquigarrow} \left(\frac{dP_h}{dP_{h_0}}\right)_{h \in I}.$$

Here $\overset{h_0}{\rightsquigarrow}$ denotes weak convergence under the local parameter sequence $\{\theta + \psi_n h_0\}$. Since the likelihood is a sufficient statistic, it is not surprising that properties of an experiment can be explored through the likelihood ratio process. A key result from the limits of experiments theory is the following:

Theorem 1 *(Asymptotic Representation Theorem) Suppose that a sequence of experiments $\mathcal{E}_n = (P_{n,h} : h \in H)$ converges to an experiment $\mathcal{E}$ such that $\mathcal{E}$, regarded as a set of measures, is dominated by a σ-finite measure. Let T_n be a sequence of statistics in $\mathcal{E}_n$ that converges weakly to a limit law Q_h for every parameter h, where the Q_h concentrate on a fixed Polish set. Then there exists a (possibly randomized) statistic T in $\mathcal{E}$ such that for every $h \in H$, $T_n \overset{h}{\rightsquigarrow} T$.*

Proof: See Van der Vaart (2001).

Thus, by studying the limit experiment $\mathcal{E}$ we can characterize the set of attainable limit distributions in the original experiment.

Limit experiments are a useful way to understand the efficiency of maximum likelihood estimation in regular models. Regular models are locally asymptotically normal, which means that the limit experiment corresponds to a single observation Z from a shifted multivariate normal distribution $N(h, I_\theta^{-1})$ for $h \in \mathbb{R}^k$ unknown and I_θ known and equal to the Fisher information matrix in the original model. The maximum likelihood estimator of h is simply Z; this is also the generalized Bayes estimator with respect to a flat prior. For any "bowl-shaped" loss function this estimator can be shown to be minmax. Not surprisingly, the sequence of maximum likelihood estimators in $\mathcal{E}_n$ converges to the maximum likelihood estimator of the limit experiment $\mathcal{E}$, showing the asymptotic efficiency of maximum likelihood estimation.

2.4 Limits of Experiments Analysis of the Uniform Model

Now we return to our example of observing a random sample from $U[0, \theta]$. In contrast to the usual regular case, here the appropriate scaling factor for a local parameter sequence is $\psi_n = 1/n$. Thus a local parameter $h \in \mathbb{R}$ corresponds to the sequence of models $U[0, \theta-(h/n)]$. The likelihood ratio has the form

$$\begin{aligned}\frac{dP_{n,h}}{dP_{n,h_0}} &= \frac{(\theta-h/n)^{-n}}{(\theta-h_0/n)^{-n}} \frac{\mathbf{1}\{Z_{(n)} \leq \theta-h/n\}}{\mathbf{1}\{Z_{(n)} \leq \theta-h_0/n\}} \\ &= \frac{(\theta-h_0/n)^n}{(\theta-h/n)^n} \mathbf{1}\{Z_{(n)} \leq \theta-h/n\}\end{aligned}$$

almost surely under P_{n,h_0}. It can be calculated that

$$-n(Z_{(n)}-\theta) \overset{h_0}{\rightsquigarrow} W,$$

where W is distributed as a shifted exponential with density

$$f_W(w) = \exp\left(\frac{(h_0-w)}{\theta}\right) \mathbf{1}\{w \geq h_0\}/\theta.$$

Thus, by straightforward calculations,

$$\frac{dP_{n,h}}{dP_{n,h_0}} \overset{h_0}{\rightsquigarrow} \exp\left(\frac{(h-h_0)}{\theta}\right) \mathbf{1}\{W \geq h\}.$$

Next, consider the situation where we observe a single draw W from the shifted exponential distribution with density f_W. The likelihood ratio for this experiment is

$$[\exp((h-W)/\theta)\mathbf{1}\{W \geq h\}/\theta]/[\exp((h_0-W)/\theta)\mathbf{1}\{W \geq h_0\}/\theta] = \exp((h-h_0)/\theta)\mathbf{1}\{W \geq h\},$$

exactly the same as the limiting likelihood ratio in the uniform case. Hence the finite-dimensional distributions of the likelihood ratio process from the $U[0,\theta]$ experiment converge to the finite-dimensional distributions for an observation from a shifted exponential with hazard rate $1/\theta$.

From the asymptotic representation theorem, we know that estimators of θ have a limiting distribution equal to the distribution of some randomized estimator in the shifted exponential limit experiment. Consider a randomized estimator, T, in the limit experiment. This estimator has maximum risk, $\sup_h E_h l(T-h)$ where the expectation is taken under h. The minmax risk bound in the limit experiment is then

$$R = \inf_T \sup_h E_h l(T-h),$$

where the infimum is taken over all randomized estimators. It follows that this expression is also the asymptotic minmax risk bound for estimators in the original experiment, i.e.

$$\liminf_{n \longrightarrow \infty} \sup_{h \in H} E_h l(n(\hat{\theta} - \theta + h/n)) \geq R$$

provided $\hat{\theta}$ has a limit distribution under every h, and l is lower semicontinuous. So the (*exact*) lower bound for an estimator of the shift from a single observation from a shifted exponential gives the *asymptotic* bound for estimators of θ from a random sample from $U[0, \theta]$. The lower bound, and the form of optimal estimators, will generally depend on the choice of loss function; this finding can be contrasted with the local asymptotic normal case, in which a single estimator, the MLE, is known to be minmax for all bowl-shaped loss functions.

For the shifted exponential experiment with squared error loss, this bound is known from classical decision theory to be $R = \theta^2$. Given this bound, we can analyze the efficiency of the maximum likelihood and Bayes estimators proposed above. From the limiting distribution of the maximum likelihood estimator, $Z_{(n)}$, the asymptotic risk is

$$\int_{-\infty}^{0} w^2 \exp(w/\theta)/\theta dw = 2\theta^2 > R.$$

On the other hand, the Bayes estimator, $Z_{(n)}(n+1)/n$, attains this bound:

$$\int_{-\infty}^{\theta} w^2 \exp((w - \theta)/\theta)/\theta dw = \theta^2 = R.$$

So the Bayes estimator with squared error loss is efficient for squared error loss minimax risk.

Similar calculations can be carried out for other risk functions. However, there is a useful heuristic argument that shows that Bayes estimators will generally be efficient, following Berger (1985), Section 6.3. Suppose we observe a single draw for a random variable W with density $f(w-h)$, where h is a location parameter in $\mathbb{R}$. Assume the loss $l(h, a)$ has the form $l(a-h)$. Since the problem is location equivariant, it is natural to focus on equivariant estimators, i.e. estimators which have the form

$$\delta(w + c) = \delta(w) + c$$

Then

$$\delta(0) = \delta(w) - w$$

or

$$\delta(w) = w + \delta(0) = w + K.$$

It can be shown that an equivariant rule has constant risk

$$R(h,\delta) = R(0,\delta) = \int l(w+K)f(w)dw$$

The minimum risk equivariant (MRE) rule minimizes the previous expression. According to the Hunt-Stein theorem (see e.g. Kiefer (1957) and Wesler (1959)), under some conditions the MRE rule turns out to be minmax over all possible decision rules.

Now consider the (generalized) Bayes estimator with respect to the constant prior. This minimizes expected loss with respect to the posterior

$$p(h|w) = c \cdot f(w-h) = f(w-h).$$

The posterior expected loss is

$$E(l(a-h)|w) = \int l(a-h)f(w-h)dh$$

Setting $y = w - h$ this is

$$\int l(y+a-w)f(y)dy = \int l(y+K)f(y)dy$$

(where $K = a - w$) Minimizing this is the same as finding the MRE rule; hence the generalized Bayes estimator is minmax.

Under weak conditions on the prior, the Bayes estimator in the uniform model for a given loss function will have the same limit distribution as the Bayes estimate with flat prior in the shifted exponential experiment, because the prior gets dominated by the likelihood function in the limit. It follows that the Bayes estimator in the uniform experiment will be locally asymptotically minmax, for a fairly arbitrary choice of prior.

3 Limit Experiments for Regression Models with Parameter-Dependent Support

Having developed intuition from the simple uniform case, we examine more general models using the limits of experiment framework. We are interested in econometric models where the conditional density of y_i given x_i has the form

$$f(y_i|x_i,\theta,\gamma)\mathbf{1}(y_i \geq g(x_i,\theta)),$$

where θ and γ are finite-dimensional parameters, and where, for x_i in some set with positive probability, the conditional density of y_i at its support boundary $g(x_i, \theta)$ is strictly positive. A general optimality result will be given in Section 4 along with precise conditions on the model. In this section, our focus is on using the limits of experiments framework to provide intuition for the efficiency result to come.

3.1 Limit Experiment with No Covariates

First, let us consider the special case with no covariates and a scalar parameter. We assume that the y_i are i.i.d. with density

$$f(y_i|\theta)\mathbf{1}(y_i \geq g(\theta)),$$

where $\theta \in \Theta$, a compact subset of $\mathbb{R}$. Let P_θ^n denote the joint law of $y_1, \ldots, y_n$. Assume that $f(g(\theta)|\theta) > 0$, and that g is continuously differentiable with derivative $g' > 0$. As a consequence of the general limit experiment result in Theorem 2 below, we have the following finite-dimensional limit likelihood ratio process: for every $h_0 \in \mathbb{R}$ and every finite set $I \subset \mathbb{R}$,

$$\left(\frac{dP^n_{\theta+h/n}}{dP^n_{\theta+h_0/n}}\right)_{h\in I} \overset{h_0}{\rightsquigarrow} \left(\exp\left(\frac{(h-h_0)}{\lambda}\right)\mathbf{1}(W > h)\right)_{h\in I},$$

where $\lambda = [f(g(\theta)|\theta)g'(\theta)]^{-1}$ and W is a random variable with the shifted exponential density $f_W(w) = \exp\left(-(w-h_0)/\lambda\right)\mathbf{1}(w > h_0)/\lambda$. This is essentially the same likelihood ratio process as in the uniform model. It follows that the experiment consisting of observing one draw from the shifted exponential density

$$f_W(w) = \exp\left(-(w-h)/\lambda\right)\mathbf{1}(w > h)/\lambda,$$

where $\lambda = [f(g(\theta)|\theta)g'(\theta)]^{-1}$, is asymptotically equivalent. By the reasoning we used in the uniform case, the Bayes estimator will be optimal.

3.2 Limit Experiment with Covariates

We next turn to the case with covariates. We assume that (y_i, x_i) is i.i.d. on $\mathcal{Y} \times \mathcal{X}$, where $\mathcal{Y} \subset \mathbb{R}$ and $\mathcal{X} \subset \mathbb{R}^m$. Assume $\mathcal{X}$ is compact, and that x has marginal distribution P_x. The outcome variable y has a conditional density with respect to Lebesgue measure of the form:

$$f(y_i|x_i, \gamma, \theta)\mathbf{1}(y_i \geq g(x_i, \theta)),$$

where $\gamma \in \Gamma \subset \mathbb{R}^d$, $\theta \in \Theta \subset \mathbb{R}^k$, Γ and Θ are compact.

We will use local parameter sequences

$$\theta + \frac{u}{n}, \qquad u \in \mathbb{R}^k,$$

$$\gamma + \frac{v}{\sqrt{n}}, \qquad v \in \mathbb{R}^d.$$

Let $\alpha = (\theta, \gamma)$, $h = (u', v')'$, and $h_0 = (u_0', v_0')'$.

Next we state a result on the limit of the likelihood ratio process for the general model. The assumptions referred to below consist of fairly standard regularity conditions, which will be discussed in detail in Section 4. For now, we focus on using the conclusion of the theorem to provide further intuition.

Theorem 2 *Let P_h^n denote the joint law of $(y_1, x_1), \ldots, (y_n, x_n)$ under $\alpha + \varphi_n h$. Under Assumptions 1 - 6, for every h_0 and every finite $I \subset H$,*

$$\left(\frac{dP_h^n}{dP_{h_0}^n}(Y^n, X^n)\right)_{h \in I} \overset{h_0}{\rightsquigarrow}$$

$$\left(\exp((v - v_0)'T - \frac{1}{2}(v - v_0)'I_\gamma(v - v_0)) \exp\left(E[f(g(x,\theta)|x, \theta, \gamma)\nabla_\theta g'](u - u_0)\right) D_h\right)_{h \in I}$$

where $I_\gamma = E_\alpha\left[\nabla_\gamma \ln f(y|x, \alpha)\nabla_\gamma \ln f(y|x, \alpha)'\right]$, and under h_0, T and $(D_h)_{h \in I}$ are independent with $T \sim N(0, I_\gamma)$. $(D_h)_{h \in I}$ are jointly distributed Bernoulli random variables whose distribution is specified by the following marginal probabilities. Let $\{h_1, \ldots, h_l\} \subset I$.

$$\begin{aligned}
&P_\alpha(D_{h_1} = 1, \ldots, D_{h_l} = 1) \\
&= \exp(-E[\mathbf{1}\{\max\{\nabla_\theta g(x,\theta)'(u_1 - u_0), \ldots, \nabla_\theta g(x,\theta)'(u_l - u_0)\} > 0\} \\
&\qquad \cdot f(g(x,\theta)|x, \alpha) \max\{\nabla_\theta g(x,\theta)'(u_1 - u_0), \ldots, \nabla_\theta g(x,\theta)'(u_l - u_0)\}]).
\end{aligned}$$

The limiting likelihood ratio process now depends on the marginal distribution of the covariates, through the expectation terms. To our knowledge, this more complicated likelihood ratio process has not been studied before in the limits of experiments literature. In the continuous covariates case, the corresponding limit experiment would likely only be expressable as a stochastic process with index set $\mathcal{X}$. Here, we concentrate on the simpler discrete covariates case, where it is possible to obtain a useful limit experiment which provides intuition for the optimality of Bayes estimators.

Assume that in the original model, x takes on the values $\{a_1, a_2, \ldots, a_L\}$. Let $p_x(a_j) := Pr(x = a_j)$. Consider the experiment consisting of observing a draw from $(S, W_1, \ldots, W_L)$, where S is distributed as $N(v, I_\gamma^{-1})$, and W_j is a random variable with the shifted exponential density

$$f_{W_j}(w) = \exp\left(-(w - g_j)/\lambda_j\right) \mathbf{1}(w > g_j)/\lambda,$$

with $g_j = \nabla_\theta g(a_j, \theta)' u$ and $\lambda_j = [p_x(a_j) f(g(a_j, \theta)|a_j, \theta, \gamma)]^{-1}$, and $(S, W_1, \ldots, W_L)$ are jointly independent. This experiment can be verified to have the same likelihood ratio process, so it can serve as a limit experiment for the general model with discrete covariates.

This limit experiment is more complicated than in the usual local asymptotic normal case, or the pure exponential shift case. Nevertheless, its structure has enough in common with these more conventional limit experiments that we can obtain some useful intuition. The normally distributed component is independent of the other variables, so we can consider it separately. By standard arguments, the Bayes estimator with a flat prior will be minmax for this component.

The remaining components of the limit experiment correspond to an $L \times 1$ vector of exponential random variables $W = (W_1, \ldots, W_L)$, with known hazard and a vector shift $H'u$, where

$$H = \begin{bmatrix} \nabla_\theta g(a_1, \theta)' \\ \vdots \\ \nabla_\theta g(a_L, \theta)' \end{bmatrix}.$$

We shall refer to this as the *generalized exponential shift* model. Assume that $L \geq m$ and that the $L \times m$ matrix H has full column rank. This is not a pure shift experiment, but it does have a similar equivariance property. For any $c \in \mathbb{R}^m$, consider a transformation of the original data

$$g_c(W) = W + Hc.$$

Notice that if W is distributed according to the generalized exponential shift model with parameter u, then $g_c(W)$ has the same distribution, but with parameter $u + c$. By reasoning similar to that used at the end of Section 2, it can be shown that the Bayes estimator with a flat prior is equivariant. That is, if the Bayes estimate given an observation W is $\tilde{a}$, then the Bayes estimate given $g_c(W)$ is $\tilde{a} + c$. Furthermore, the Bayes estimate is actually the minimum risk equivariant estimator. Under a condition known as *amenability*, which can be verified here, the Hunt-Stein theorem applies, and the minimum risk equivariant estimator is also minmax.

For completeness, we show these steps formally in Appendix A. We can then conclude that in the original problem, estimators which asymptotically have limit distributions equal to the distribution of the Bayes estimator with respect to a flat prior, will be locally asymptotically minmax. An obvious choice is any Bayes estimator; since the prior will typically be dominated as the sample size increases, it will behave like flat-prior Bayes asymptotically. The results to follow establish this formally.

4 Asymptotic Properties of Bayes Estimators

The results of the previous two sections showed that, in various special cases, the limit experiment had an equivariance property which implied that flat-prior Bayes is minmax. (Of course, there may be other estimators which have different risk functions but the same minmax risk.) In looking for local asymptotic minmax estimators for the original model, it is therefore natural to investigate the asymptotic properties of Bayes estimators for general priors. In this section, we show that Bayes estimators behave asymptotically like flat-prior Bayes with respect to the limiting likelihood ratio process. We then show that for the general model (including the case where covariates are continuous), the Bayes estimator is local asymptotic minmax, using a strategy suggested by Ibragimov and Hasminskii (1981).

As in the previous section we will use the local parameter sequences $\theta + \frac{u}{n}$, for $u \in \mathbb{R}^k$, and

$\gamma + \frac{v}{\sqrt{n}}$, for $v \in \mathbb{R}^d$. Define the local parameter spaces as

$$U_n = n(\Theta - \theta_0),$$
$$V_n = \sqrt{n}(\Gamma - \gamma_0).$$

Let

$$h = \begin{pmatrix} u \\ v \end{pmatrix}, \quad \text{and} \quad h_0 = \begin{pmatrix} u_0 \\ v_0 \end{pmatrix}.$$

To define a Bayes estimator, let the *prior* π be a (possibly improper) Lebesgue density on $\Theta \times \Gamma$. The *Bayes estimator* $\tilde{\alpha}_n$ is any solution to

$$\min_{\tilde{\alpha}} \int l(\varphi_n^{-1}(\tilde{\alpha} - \alpha_0)) \prod_{i=1}^{n} f(y_i|x_i, \alpha_0 + \varphi_n h)\mathbf{1}(y_i \geq g(x_i, \theta + u/n))\pi(\alpha_0 + \varphi_n h)dh.$$

This defines the Bayes estimator as minimizing posterior expected loss, where the posterior distribution is with respect to the local parameter h. There is, of course, an equivalent definition in terms of the original parameter α.

We assume the following six conditions on the model.

Assumption 1 *(y_i, x_i) is i.i.d. on $\mathcal{Y} \times \mathcal{X}$, where $\mathcal{Y} \subset \mathbb{R}$ and $\mathcal{X} \subset \mathbb{R}^d$. Assume $\mathcal{X}$ is compact. x has marginal distribution P_x. y has a conditional density with respect to Lebesgue measure of the form:*

$$f(y_i|x_i, \gamma, \theta)\mathbf{1}(y_i \geq g(x_i, \theta)),$$

where $\gamma \in \Gamma \subset \mathbb{R}^k$, $\theta \in \Theta \subset \mathbb{R}^m$, $\Gamma \times \Theta$ is compact and convex.

Assumption 2 *$f(y|x, \alpha)$ is continuous in y and twice continuously differentiable in α for all y and x, $g(x, \theta)$ is continuously differentiable in θ for all x. Also, for α in some open neighborhood (α_0), $f(y|x, \alpha) > 0$ and $f(y|x, \alpha) < \infty$ uniformly in y, x.*

Assumption 3

$$\int\int \sup_{\alpha \in \mathcal{N}} \|\nabla_\alpha f(y|x, \alpha)\| \mathbf{1}(y \geq g(x, \theta))dydP_x(x) < \infty$$

$$\int\int \sup_{\tilde{\alpha}, \alpha \in \mathcal{N}} \frac{||\nabla_\alpha f(y|x, \tilde{\alpha})||^2}{f(y|x, \tilde{\alpha})^2}\mathbf{1}(y \geq g(x, \theta))f(y|x, \alpha)dydP_x(x) < \infty$$

$$\int\int \sup_{\tilde{\alpha},\alpha\in\mathcal{N}} \frac{\|\nabla_{\alpha\alpha} f(y|x,\tilde{\alpha})\|^{1+\delta}}{f(y|x,\tilde{\alpha})} \mathbf{1}(y \geq g(x,\theta)) f(y|x,\alpha) dy dP_x(x) < \infty$$

for some $\delta > 0$.

Assumption 4 *The function* $J_{ij}^{\gamma\gamma}(\alpha) = E_\alpha[\nabla_{\gamma_i} \ln f(y|x,\alpha) \nabla_{\gamma_j} \ln f(y|x,\alpha)]$ *and similarly defined* $J_{ij}^{\gamma\theta}$ *and* $J_{ij}^{\theta\theta}$ *are continuous in* α. *For each* i, j, $J_{ij}^{\gamma\gamma}(\alpha)$ *has a majorant that is a product of a polynomial in* $\|\theta\|$ *and exponential in* $\|\gamma\|$. *Also for each* α, *there exists an open neighborhood* $\mathcal{N}$ *of* α *such that* $J_{ij}^{\gamma\theta}$ *and* $J_{ij}^{\theta\theta}$ *are bounded on* $\mathcal{N}$ *and the minimum eigenvalue of* I_γ *is bounded away from zero on* $\mathcal{N}$, *where* I_γ *is the matrix with elements* $J_{ij}^{\gamma\gamma}$.

Assumption 5

$$E_x[\sup_{\alpha\in\mathcal{N}} \|\nabla_\theta g(x,\theta)\|] < \infty$$

Assumption 6 *There exists* $\varepsilon > 0$ *such that*

$$\inf_{\alpha,\tilde{\alpha}\in\mathcal{N}} Pr_\alpha(y \geq g(x,\tilde{\theta})) \geq \varepsilon$$

and

$$\inf_{\|w\|=1} E_\alpha\left[f(g(x,\theta)|x,\alpha)|\nabla_\theta g(x,\theta)'w|\right] \geq \varepsilon.$$

We also make the following assumptions about the loss function and the prior.

Assumption 7 *The* loss function $l : \mathbb{R}^{k+d} \to [0,\infty)$ *satisfies*

(a) l *is continuous and not identically 0.*

(b) $l(0) = 0$, $l(x) = l(-x)$ *for all* $x \in \mathbb{R}^{k+d}$.

(c) l *has a polynomial majorant:*

$$l(x) \leq B_0(1 + ||x||^b)$$

for some $B_0, b > 0$, *all* $x \in \mathbb{R}^{k+d}$.

(d) There exist numbers $H_0, \eta > 0$ *such that for all* $H \geq H_0$,

$$\sup\{l(x) : x \leq H^\eta\} - \inf\{l(x) : x \geq H\} \leq 0.$$

Assumption 8 *The prior* π *is continuous and positive at* α_0, *with a polynomial majorant.*

The assumptions on the loss function and the prior are fairly weak, and allow for most choices of prior and loss of which we are aware.

Recall that Theorem 2, given in the previous section, shows that under Assumptions 1-6, the finite-dimensional distributions of the likelihood ratio process

$$Z_{n,\alpha+\varphi_n h_0}(h) \equiv \frac{dP_h^n}{dP_{h_0}^n}(Y^n, X^n),$$

converge in distribution to a particular process. Let us denote the limiting process as $Z_{h_0}(h)$.

The next result is the main result of the paper. It strengthens the finite-dimensional convergence of the likelihood ratio process to convergence in distribution of the Bayes estimator, and then shows that the Bayes estimator is local asymptotic minmax.

Theorem 3 *Suppose Assumptions 1-8 hold. Also, suppose* $\psi_{h_0}(s,t) = \int l(s-u, t-v)\xi_{h_0}(u,v)dudv$ *attains its minimum at a unique point,* τ*, where* $\xi_{h_0}(u,v) = Z_{h_0}(u,v)/\int Z_{h_0}(s,t)dsdt$*. Then*

$$(n(\tilde{\theta}_n - \theta_0), \sqrt{n}(\tilde{\gamma}_n - \gamma_0)) \rightsquigarrow \tau.$$

Moreover, $\tilde{\alpha}_n$ *is asymptotically efficient at* α_0 *with respect to loss* l*.*

Remarks: We do not require that the covariates have finite support in this result. In order to obtain a general efficiency result, we use a risk-continuity argument as in Ibragimov and Hasminskii (1981).

5 Conclusion

We have studied optimal estimation of models where the support depends on parameters and covariates. Under the local asympototic minmax criterion, Bayes estimators are efficient in these models. We provided intuition for this result by first examining the Uniform$[0, \theta]$ model. Then we considered a general model with discrete covariates. For this model, we provided further intuition for efficiency of Bayes, by showing that the limit experiment had a group structure that implied minimaxity of flat prior Bayes. Finally, we extended the result to continuous covariates using a risk-continuity argument.

Throughout the paper we have focused on point estimation under a given loss function. However, the limits of experiments theory can also be informative about optimal testing (see, for example, Ploberger (1998)), and other aspects of inference such as construction of confidence intervals and predictive intervals. We leave such topics for future work.

A Generalized Exponential Shift Model

In this section we examine the exponential shift model in further detail. Consider the experiment $\{P_u : u \in \mathbb{R}^m\}$, which consists of observing a random vector $W = (W_1, \ldots, W_L)$, for $L \geq m$, where P_u specifies that the components W_j are independently distribution with shifted exponential densities

$$f_j(w_j|u) = \exp\left(-(w_j - H_j' u)/\lambda_j\right) \mathbf{1}(w_j > H_j' u)/\lambda_j.$$

We assume that the λ_j and H_j are known, with $\lambda_j > 0$, and that the $L \times m$ matrix

$$H := \begin{bmatrix} H_1' \\ \vdots \\ H_L' \end{bmatrix}$$

has full column rank.

Let the loss function $l(u, a)$ for estimating u have the form $l(u, a) = l(u - a)$, with $l(0) = 0$, $l \geq 0$, and l continuous and strictly convex. l nonnegative and continuous. Assume that for every real number τ, the set

$$\{a : l(u - a) \leq \tau\}$$

is compact for all $u \in \mathbb{R}^m$. Let $\tilde{u}$ be the generalized Bayes estimator corresponding to the flat prior for u: $\tilde{u}$ solves

$$\min_{\tilde{u}} \int l(u - \tilde{u}) \prod_{j=1}^{L} f_j(W_j|u) du.$$

Assume $\tilde{u}$ exists and is unique. Then we claim that $\tilde{u}$ is minmax for loss l. To provide a formal justification for this claim, we set up the experiment as a group family. On the sample space $\mathbb{R}^L$, define the group of transformations $\mathcal{G} = \{g_c : c \in \mathbb{R}^m\}$, where

$$g_c w = w + Hc.$$

We can regard $\mathcal{G}$ as the Euclidean space $\mathbb{R}^m$ with the usual topology. The composition operator is $g_c \circ g_d = g_{c+d}$, and the identity element is $e = g_0$. The inverse is $g_c^{-1} = g_{-c}$. We define associated groups $\overline{\mathcal{G}}$ and $\tilde{\mathcal{G}}$ on the parameter space and action space respectively. Here $\overline{\mathcal{G}} = \{\overline{g}_c : c \in \mathbb{R}^m\}$, with

$$\overline{g}_c u = u + c$$

and $\tilde{\mathcal{G}} = \overline{\mathcal{G}}$. All three groups are abelian: $g_c \circ g_d = g_d \circ g_c$. It can be readily seen that the experiment $\{P_u : u \in \mathbb{R}^m\}$ is equivariant under the action of $\mathcal{G}$ and $\overline{\mathcal{G}}$, and that the loss (since it is of the form $l(u-a)$) is equivariant under $\tilde{\mathcal{G}}$.

Next, we will show that the generalized Bayes estimator with respect to the right Haar measure associated with group $\overline{\mathcal{G}}$ and given loss l, is the minimum risk equivariant (MRE) estimator. This can be verified using Theorem 6.59 of Schervish (1995). To apply that result we need to verify the following conditions:

1. The experiment is invariant under the action of $\mathcal{G}, \overline{\mathcal{G}}$.

2. The left Haar measure λ and the right Haar measure ρ exist.

3. (a) $\overline{\mathcal{G}}$ is a topological group.

   (b) λ is σ-finite and not identically 0.

   (c) The function $f : \overline{\mathcal{G}} \times \overline{\mathcal{G}} \to \overline{\mathcal{G}}$ defined by $f(g, h) = g^{-1} \circ h$ is continuous.

4. The mapping $\phi : \mathcal{G} \to \overline{\mathcal{G}}$ defined by $\phi(g) = \overline{g}$ is a group isomorphism.

5. There is a bimeasurable (measurable, one-to-one and onto, with measurable inverse) mapping $\eta : \mathbb{R}^m \to \overline{\mathcal{G}}$ which satisfies $\overline{g} \circ \eta(u) = \eta(\overline{g}u)$ for all $\overline{g} \in \overline{\mathcal{G}}$ and all $u \in \mathbb{R}^m$.

6. There exists a bimeasurable function $t : \mathbb{R}^L \to \mathcal{G} \times \mathcal{Y}$ for some space $\mathcal{Y}$ (where we write $t(w) = (h, y)$) such that, for every $g \in \mathcal{G}$ and $w \in \mathbb{R}^L$,

$$t(w) = (h, y) \implies t(gw) = (g \circ h, y).$$

7. For every u, the distribution of on $\mathcal{G} \times \mathcal{Y}$ induced from P_u by t has a density with respect to $\lambda \times v$, where v is some measure on $\mathcal{Y}$.

Condition 1 is immediate from the definition of the groups and the translation nature of the experiment. Since $\overline{\mathcal{G}}$ is the translation group on $\mathbb{R}^m$ it can be readily seen that Lebesgue measure is both a left and right Haar measure, verifying condition 2. For condition 3, note that $\overline{\mathcal{G}} = \mathbb{R}^m$ and hence it is a topological space. Lebesgue measure is σ-finite and not identically 0. Also note that $f(g_c, g_d) = g_c^{-1} \circ g_d = g_{d-c}$ which is easily seen to be continuous.

To verify condition 4, write

$$\begin{aligned} \phi(g_c \circ g_d) &= \phi(g_{c+d}) \\ &= \overline{g}_{c+d} \\ &= \overline{g}_c \circ \overline{g}_d \\ &= \phi(g_c) \circ \phi(g_d) \end{aligned}$$

So ϕ is a group homomorphism, and since it is clearly one-to-one and onto, is a group isomorphism.

To verify condition 5, let $\eta(u) = \overline{g}_u$. This is clearly bimeasurable. We have

$$\begin{aligned} \overline{g}_c \circ \eta(u) &= \overline{g}_c \circ \overline{g}_u \\ &= \overline{g}_{c+u} \\ &= \eta(c+u) \\ &= \eta(\overline{g}_c u). \end{aligned}$$

To verify conditions 6 and 7, we need to construct a maximal invariant (a statistic that identifies orbits of $\mathcal{G}$). By assumption H has full column rank and hence row rank of m. Thus there are m linearly independent rows of H. Reorder the elements of W (and the corresponding rows of H) so that the first m elements correspond to the m linearly independent rows of H. Then define

$$\tilde{H} = \begin{bmatrix} H_1' \\ \vdots \\ H_m' \end{bmatrix}.$$

$$t(w) = \left(\tilde{H}^{-1} \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}, \begin{pmatrix} w_{m+1} \\ \vdots \\ w_L \end{pmatrix} - \begin{pmatrix} H_{m+1}' \\ \vdots \\ H_L' \end{pmatrix} \tilde{H}^{-1} \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix} \right).$$

It can be seen that t satisfies the requirements of condition 6 and that the distribution of t has a density with respect to the product of Lebesgue measures on $\mathbb{R}^m$ and $\mathbb{R}^{L-m}$. (If $L = m$ (so that there is only a single orbit) we can modify the argument by having $\mathcal{Y}$ be an arbitrary singleton set with counting measure.)

Therefore we conclude that $\tilde{u}$ is the MRE.

Next, we want to show that the MRE is in fact minmax over all possible estimators. To apply the generalized version of the Hunt-Stein theorem due to Wesler (1959), we need to verify the following conditions:

1. The distributions P_u are dominated by a σ-finite measure.

2. The action space is a separable metric space, and for each $u \in \mathbb{R}^m$, $l(u, a)$ is nonnegative and continuous in a, and for every real number τ, the set
$$\{a : l(u, a) \leq \tau\}$$
is compact.

3. $\mathcal{G}$ satisfies a condition known as amenability (see Bondar and Milnes (1981) for various equivalent conditions for amenability).

4. $\mathcal{G}$ is a locally compact, σ-compact topological group with its Borel σ-algebra generated by the compact subsets of $\mathcal{G}$.

The first two conditions are immediate. To show amenability, note that Bondar and Milnes (1981) point out that if a locally compact group is abelian, then it is amenable. The group $\mathcal{G}$ is a Euclidean space and hence satisfies condition 4.

B Proofs of Theorems

References

BAJARI, P. (1998): "Econometrics of the First Price Auction with Asymmetric Bidders," manuscript, Stanford University.

BERGER, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

BONDAR, J. V., AND P. MILNES (1981): "Amenability: A Survey for Statistical Applications of Hunt-Stein and Related Conditions on Groups," *Z. W. verw. Gebiete*, pp. 103–128.

CAVANAGH, C. L., L. E. JONES, AND T. J. ROTHENBERG (1990): "Efficient Estimation under Asymmetric Loss," manuscript, University of California.

CHERNOZHUKOV, V., AND H. HONG (2001): "Likelihood Inference with Density Jump," manuscript.

CHRISTENSEN, B. J., AND N. M. KIEFER (1991): "The Exact Likelihood Function for an Empirical Job Search Model," *Econometric Theory*, 7, 464–486.

DONALD, S., AND H. PAARSCH (1993): "Maximum Likelihood Estimation When the Support of the Distribution Depends Upon Some or All of the Unknown Parameters," manuscript.

DONALD, S. G., AND H. J. PAARSCH (1999): "Superconsistent Estimation and Inference in Structural Econometric Models using Extreme Order Statistics," manuscript, University of Iowa.

FLINN, C., AND J. HECKMAN (1982): "New Methods for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, 18, 115–168.

GHOSAL, S., AND T. SAMANTA (1995): "Asymptotic Behaviour of Bayes Estimates and Posterior Distributions in Multiparameter Nonregular Cases," *Mathematical Methods of Statistics*, 4, 361–388.

Hong, H. (1998): "Maximum Likelihood Estimation for Job Search, Auction, and Frontier Production Function Models," manuscript.

Ibragimov, I., and R. Hasminskii (1981): *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

Kiefer, J. (1957): "Invariance, Minmax Sequential Estimation, and Continuous Time Processes," *Annals of Mathematical Statistics*, 28, 573–601.

Kiefer, N. M., and M. F. J. Steele (1998): "Bayesian Analysis of the Prototypal Search Model," *Journal of Business and Economics Statistics*, 16, 178–186.

Lancaster, T. (1997): "Exact Structural Inference in Optimal Job Search Models," *Journal of Business and Economics Statistics*, 15, 165–179.

Pflug, G. C. (1983): "The Limiting Log-Likelihood Process for Discontinuous Density Families," *Z. W. verw. Gebiete*, 64, 15–35.

Ploberger, W. (1998): "A Complete Class of Tests When the Likelihood is Locally Asymptotically Quadratic," manuscript, University of Rochester.

Sareen, S. (2000): "Evaluating Data in Structural Parametric Auction, Job-Search, and Roy Models," manuscript, Bank of Canada.

Schervish, M. J. (1995): *Theory of Statistics.* Springer-Verlag, New York.

Smith, R. L. (1985): "Maximum Likelihood Estimation in a Class of Nonregular Cases," *Biometrika*, 72, 67–90.

Van der Vaart, A. W. (2001): "Limits of Experiments," manuscript.

Wesler, O. (1959): "Invariance Theory and a Modified Minimax Principle," *Annals of Mathematical Statistics*, 30, 1–20.