Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support*

Keisuke Hirano
University of Miami
khirano@miami.edu

Jack R. Porter
Harvard University
jporter@harvard.edu

October 16, 2002

*We are grateful to Gary Chamberlain, Guido Imbens, Peter Bickel, numerous seminar participants, a co-editor, and a referee for comments. We thank the National Science Foundation for research support under grants SES-9985257 (Hirano) and SES-0112095 (Porter).

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 maximum likelihood estimator is generally inefficient, but that the Bayes estimator is efficient according to the local asymptotic minmax criterion for conventional loss functions. 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. This parameter-dependent support property arises in certain parametric auction models, search models, and production frontier models. For such models, conventional asymptotic theory does not apply.

Much of the previous work on models with parameter-dependent support, including Flinn and Heckman (1982), Smith (1985), and Christensen and Kiefer (1991), Donald and Paarsch (1993), Hong (1998), Donald and Paarsch (2002), and Chernozhukov and Hong (2001), has focused on obtaining asymptotic distribution theory for maximum likelihood, Bayes, and other estimators. The limit distributions of these estimators are generally nonnormal, making comparisons of different estimators and efficiency considerations more difficult than in the regular case. We build on this important earlier work by directly examining the efficiency issue and 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, for bowl-shaped loss functions, 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 the Blackwell-Le Cam 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 invariant 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 optimal estimation of 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)). Allowing for covariates leads to more complicated limit experiments and requires an extension of the results on asymptotic efficiency in Ibragimov and Hasminskii (1981).

*It is possible to construct loss functions, using the Dirac delta function, such that the Bayes solution under a flat prior is the MLE. However, our focus here is on standard loss functions, such as squared error loss and absolute error loss.

Recent work derives the limiting distributions of specific estimators in more general models 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). Donald and Paarsch (1993) develop the asymptotic properties of ML estimation in models where the support can depend on discrete covariates, and Donald and Paarsch (2002) consider analog-type estimators for such models. Chernozhukov and Hong (2001) extend this work by allowing for continuous covariates and developing limiting distributions for both Bayes and ML estimators. They also consider an expanded class of models which includes a discontinuity in the density between two strictly positive values. Our work also yields a limit distribution theory for Bayes estimators which is consistent with Chernozhukov and Hong (2001), but our main focus is on obtaining asymptotic efficiency results in the nonregular, parameter-dependent support setting with covariates and providing intuition through the limits of experiments theory.

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 the outcome variable can depend on both parameters and covariates. Our first step is to provide a limits of experiments characterization in cases where the covariates are discrete. This directly yields intuition for optimality of the Bayes estimator, using the invariance properties of the limit experiment. Then, in section 4, we study the asymptotics of Bayes estimators and provide an efficiency result that applies to general distributions of the covariates. 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 a random sample of size n from a uniform distribution on the interval $[0, \theta]$, for $\theta \in \Theta \subset \mathbb{R}_{++}$. 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 Estimation

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 $\hat{\theta}_{ML} = Z_{(n)}$. Its limiting distribution is

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

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.

Bayesian estimation of θ provides an alternative approach to maximum likelihood estimation. Given a prior $\pi(\theta)$ on Θ, the posterior distribution $p(\theta|Z_1, \ldots, Z_n)$ is given by Bayes Theorem. Then 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. 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, $\tilde{\theta}_B = Z_{(n)}(n+1)/n$. The limiting distribution can be shown to be

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

The convergence rate is the same as maximum likelihood estimator, but now the limiting distribution is centered to have mean zero. Clearly, 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.† But this does not guarantee that the Bayes estimator will be efficient among all estimators. 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.

2.2 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* $(\mathcal{Z}, \mathcal{A}, P_h : h \in H)$ is a measurable space $(\mathcal{Z}, \mathcal{A})$ along with a collection of probability measures on that space indexed by a parameter h. The experiment is interpreted as the situation where we observed a random variable Z on $(\mathcal{Z}, \mathcal{A})$, 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

†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. Efficiency of bias-corrected ML among all estimators in possibly nonregular models is considered in Hirano and Porter (2002).

distributions of the likelihood ratio process for $\mathcal{E}$, i.e. for every finite subset $I \subset H$ 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 (1996), or Van der Vaart (1991) for a more general version.

Thus, by studying the limit experiment $\mathcal{E}$ we can characterize the set of attainable limit distributions in the original experiment. For example, limit experiments have been used to study the efficiency of maximum likelihood estimation in regular models via local asymptotic normality. We use the limit experiment theory to understand efficiency in the nonregular $U[0, \theta]$ model.

2.3 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

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

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

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

where W is distributed as a shifted exponential with density $f_W(w) = \exp\left(\frac{(h_0-w)}{\theta}\right)\mathbf{1}\{w \geq h_0\}/\theta$.

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 $\exp((h-h_0)/\theta)\mathbf{1}\{W \geq h\}$, 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 $R = \theta^2$. The asymptotic risk of the maximum likelihood estimator is $2\theta^2$. On the other hand, the Bayes estimator, $Z_{(n)}(n+1)/n$, has risk θ^2. So the Bayes estimator for 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 generally will be efficient when the experiment has a shift form, 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 $\delta(0) = \delta(w) - w$, so we can write

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

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

$$E(l(a-h)|w) = \int l(a-h)f(w-h)dh = \int l(y+K)f(y)dy$$

(setting $y = w - h$ and $K = a - w$). 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 a scalar y_i given a vector of covariates 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, we focus on using the limits of experiments framework to provide intuition for the efficiency of Bayes estimators.

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. Below, we concentrate on the discrete covariates case, where it is possible to obtain a useful limit experiment that provides intuition for the optimality of Bayes estimators. We then comment briefly on the case with continuous covariates.

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

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

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 $L \geq k$ and that the $L \times k$ matrix H has full column rank. This is not a pure shift experiment, but it does have a similar equivariance property. For any $c \in \mathbb{R}^k$, 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 estimator 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 of the next section establish this formally.

Extending this argument to the continuous case would be complicated. Taking the limit of the discrete covariate limit experiment as the discrete points of support become dense in $\mathcal{X}$ seems to lead to a stochastic process indexed by elements of $\mathcal{X}$. However, the existence of a limiting stochastic process is not guaranteed due to the dependence of the conditional hazard on the marginal distribution of x. Similar cases do not appear to have been examined in the limit experiment literature, although it is often possible to construct limit experiments, defined on relatively abstract spaces, which have a desired likelihood ratio process.[‡] Even with a limit experiment, extending our efficiency argument would appear to require a more general equivariance notion. We do not attempt to develop that theory here, and instead proceed in the next section with a general efficiency result

[‡]see, for example, Van der Vaart (1996) Example 10.9.

which allows for continuous covariates.

4 Asymptotic Properties of Bayes Estimators

The results of the previous two sections showed that, in various special cases, the limit experiment had an invariance 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 locally asymptotically minmax, using a strategy suggested by Ibragimov and Hasminskii (1981).

As in the previous section we will use the local parameter sequences $\theta_0 + \frac{u}{n}$, for $u \in \mathbb{R}^k$, and $\gamma_0 + \frac{v}{\sqrt{n}}$, for $v \in \mathbb{R}^d$ for some fixed (θ_0, γ_0). Define the local parameter spaces as

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

Let

$$\alpha_0 = \begin{pmatrix} \theta_0 \\ \gamma_0 \end{pmatrix}, \quad h = \begin{pmatrix} u \\ v \end{pmatrix}, \quad \text{and} \quad h_0 = \begin{pmatrix} u_0 \\ v_0 \end{pmatrix}.$$

Also, let φ_n denote a square, diagonal matrix where the first k diagonal elements are $1/n$ and the remaining d diagonal elements are $1/\sqrt{n}$.

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

$$\min_{\tilde{\alpha}} \int l(\varphi_n^{-1}(\tilde{\alpha} - (\alpha_0 + \varphi_n h))) \prod_{i=1}^{n} f(y_i|x_i, \alpha_0 + \varphi_n h) \mathbf{1}(y_i \geq g(x_i, \theta_0 + u/n)) \pi(\alpha_0 + \varphi_n h) dh.$$

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 $(= \alpha_0 + +\varphi_n h)$. Note that if $l(x) = x^2$, then the Bayes estimator is the mean of the posterior distribution, while if $l(x) = |x|$ then the solution is the posterior median. Thus the posterior mean will be shown to be asymptotically efficient for squared error loss, while the posterior median will be shown to be optimal for absolute error loss.

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

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

Assumption 3

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

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

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

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, uniformly on $\mathcal{N}$, $J_{ij}^{\gamma\theta}$ and $J_{ij}^{\theta\theta}$ are bounded and the minimum eigenvalue of I_γ is bounded away from zero, 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*

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

and

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

These assumptions include standard smoothness and moment bounding conditions. As in Chernozhukov and Hong (2001) we do not require discreteness of x. Assumption 5 assures that the boundary function g itself does not contain any jump discontinuities. Assumption 6 is essentially an identification assumption. The second part specifies that the density of y is nonnegligible at the boundary for small deviations of θ in any direction.

These assumptions should be satisfied in many parametric auction, search, and production frontier models as discussed in Donald and Paarsch (1993). To illustrate the applicability of our results, we briefly describe an auction model adapted from Paarsch (1992).

Example: Symmetric Procurement Auction with Independent Private Values and Exponential Cost Distribution.

The task being auctioned has characteristics x observed by all agents. For each of m bidders, a cost c for the task is drawn independently from an exponential distribution conditional on the characteristics x. The cost density and cumulative distribution function are

$$\begin{aligned} f_c(c|x,\theta) &= \frac{1}{h(x,\theta)} \exp\left(-\frac{c}{h(x,\theta)}\right) \mathbf{1}\{c \geq 0\} \\ F_c(c|x,\theta) &= 1 - \exp\left(-\frac{c}{h(x,\theta)}\right) \end{aligned}$$

where $E(c|x,\theta) = h(x,\theta)$, so $h(x,\theta)$ is the function giving the mean cost given characteristics x. The objective is to estimate θ. Given a cost, each agent submits a bid observed by the econometrician. Expected profit maximization leads to a symmetric Bayesian-Nash equilibrium bid function:

$$b = \beta(c|x,\theta) = c + \frac{\int_c^\infty [1 - F_c(u|x,\theta)]^{m-1} du}{[1 - F_c(c|x,\theta)]^{m-1}} = c + \frac{h(x,\theta)}{m-1}.$$

So the bid distribution is a simply a shift of the cost distribution.

$$f_b(b|x,\theta) = \frac{1}{h(x,\theta)} \exp\left(\frac{-1}{h(x,\theta)}\left(b - \frac{h(x,\theta)}{m-1}\right)\right) \mathbf{1}\left\{b \geq \frac{h(x,\theta)}{m-1}\right\}$$

In our notation,

$$\begin{aligned} f(b|x,\theta) &= \frac{1}{h(x,\theta)} \exp\left(\frac{-1}{h(x,\theta)}\left(b - \frac{h(x,\theta)}{m-1}\right)\right) \quad \text{and} \\ g(x,\theta) &= \frac{h(x,\theta)}{m-1} \end{aligned}$$

Then the following conditions would imply Assumptions 1-6:

(a) $\mathcal{X}$ is compact, and x has marginal distribution P_x which does not depend on θ.

(b) $\Theta \subset \mathbb{R}^k$ is compact and convex. Let $\mathcal{N}_\theta$ be an open neighborhood of θ_0.

(c) $\bar{h} = \sup_{\theta\in\mathcal{N}_\theta} \sup_{x\in\mathcal{X}} h(x,\theta) < \infty$, and $\underline{\mathrm{h}} = \inf_{\theta\in\mathcal{N}_\theta} \inf_{x\in\mathcal{X}} h(x,\theta) > 0$.

(d) h is twice continuously differentiable in θ for all x.

(e) For some $\delta > 0$, $E\left[\sup_{\theta\in\mathcal{N}_\theta} \|\nabla_\theta h(x,\theta)\|^{2+2\delta}\right] < \infty$ and $E\left[\sup_{\theta\in\mathcal{N}_\theta} \|\nabla_{\theta\theta} h(x,\theta)\|^{1+\delta}\right] < \infty$.

(f) $\inf_{\theta\in\mathcal{N}_\theta} \inf_{\|w\|=1} E_x[|\nabla_\theta h(x,\theta)'w|] > 0$.

Condition (c) bounds h away from zero and infinity on $\mathcal{X}$ and Θ. Conditions (d) and (e) are standard smoothness and moment conditions. Condition (f) is basically a weak stochastic linear independence condition for $\nabla_\theta h(x,\theta)$. These conditions would be satisfied, for instance, with $h(x,\theta)$ $= 1 + x_1\theta_1 + x_2\theta_2$, $\Theta = [0,1]\times[0,1]$, and (x_1,x_2) uniformly distributed on $\mathcal{X} = [0,1]\times[0,1]$. Details of the verification of Assumptions 1-6 are available on request from the authors.[§] A similar set of conditions could be used to verify the assumptions for related auction models given in Paarsch (1992) based on Pareto and Weibull cost distributions.

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

[§] A version of this paper with expanded appendices is available at http://post.economics.harvard.edu/hier.

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

(a) l is continuous and not identically 0.

(b) $l(0) = 0$.

(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. For example, we do not require symmetry or convexity of the loss function. Part (d) of Assumption 7 limits the amount that the loss function can decrease in the tails. In the efficiency result below, we will also require that the limit of the expected posterior loss has a unique minimum. Alternatively, the uniqueness could be guaranteed by additional conditions on the loss function.

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

$$Z_{n,\alpha_0+\varphi_n h_0}(h) \equiv \frac{dP^n_{\alpha_0+\varphi_n h}}{dP^n_{\alpha_0+\varphi_n h_0}}(Y^n, X^n),$$

converge in distribution to a particular process. Let us denote the limiting process as $Z_{\alpha_0,h_0}(h)$, or just $Z_{\alpha_0}(h)$ when we are considering cases in which $h_0 = 0$.

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 locally asymptotically minmax.

Theorem 3 *Suppose Assumptions 1-8 hold. Also, suppose*

$$\psi_{\alpha_0}(s,t) = \int_{\Re^{k+d}} l(s-u, t-v) \frac{Z_{\alpha_0}(u,v)}{\int_{\Re^{k+d}} Z_{\alpha_0}(\bar{u},\bar{v}) d\bar{u} d\bar{v}} dudv.$$

attains its minimum at a unique point, τ_{α_0}. Then

$$(n(\tilde{\theta}_n - \theta_0), \sqrt{n}(\tilde{\gamma}_n - \gamma_0) \rightsquigarrow \tau_{\alpha_0}$$

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

Remarks: Assumptions 1 - 6 are used to establish some intermediate results concerning the properties of the likelihood ratio process. Theorem 3 could be restated assuming these intermediate results as high-level conditions in place of Assumptions 1 - 6. This generalization could be applied to different models than the one considered here. In Appendix B, the result based on the higher level assumptions is discussed in more detail.

5 Conclusion

We have studied optimal estimation of models where the support depends on parameters and covariates. Under the local asymptotic 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 an invariance property that implied minimaxity of flat prior Bayes. Finally, we considered general covariate distributions and proved the asymptotic efficiency of Bayes estimators.

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 $\{P_u : u \in \mathbb{R}^k\}$, which consists of observing a random vector $W = (W_1, \ldots, W_L)$, for $L \geq k$, 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 $L \times k$ matrix $H := [H_1, \ldots, H_L]'$ 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. Assume that for every real number τ, the set $\{a : l(u - a) \leq \tau\}$ is compact for all $u \in \mathbb{R}^k$. 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 $\mathcal{G} = \{g_c : c \in \mathbb{R}^k\}$, where $g_c w = w + Hc$. We can regard $\mathcal{G}$ as the Euclidean space $\mathbb{R}^k$ 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 $\overline{\mathcal{G}} = \{\overline{g}_c : c \in \mathbb{R}^k\}$, 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 $\{P_u : u \in \mathbb{R}^k\}$ is invariant under the action of $\mathcal{G}$ and $\overline{\mathcal{G}}$, and that the loss (since it is of the form $l(u - a)$) is invariant 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 $\eta : \mathbb{R}^k \to \overline{\mathcal{G}}$ which satisfies $\overline{g} \circ \eta(u) = \eta(\overline{g}u)$ for all $\overline{g} \in \overline{\mathcal{G}}$ and all $u \in \mathbb{R}^k$.

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 $\mathbb{R}^k$ it can be readily seen that Lebesgue measure is both a left and right Haar measure, verifying condition 2. For condition 3, note that $\overline{\mathcal{G}} = \mathbb{R}^k$ 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 $\phi(g_c \circ g_d) = \phi(g_{c+d}) = \overline{g}_{c+d} = \overline{g}_c \circ \overline{g}_d = \phi(g_c) \circ \phi(g_d)$. 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 $\overline{g}_c \circ \eta(u) = \overline{g}_c \circ \overline{g}_u$ $= \overline{g}_{c+u} = \eta(c+u) = \eta(\overline{g}_c u)$.

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 k. Thus there are k

linearly independent rows of H. Reorder the elements of W (and the corresponding rows of H) so that the first k elements correspond to the k linearly independent rows of H. Then define $\tilde{H} = [H_1, \ldots, H_k]'$ and

$$t(w) = \left(\tilde{H}^{-1} \begin{pmatrix} w_1 \\ \vdots \\ w_k \end{pmatrix}, \begin{pmatrix} w_{k+1} \\ \vdots \\ w_L \end{pmatrix} - \begin{pmatrix} H'_{k+1} \\ \vdots \\ H'_L \end{pmatrix} \tilde{H}^{-1} \begin{pmatrix} w_1 \\ \vdots \\ w_k \end{pmatrix} \right).$$

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 $\mathbb{R}^k$ and $\mathbb{R}^{L-k}$. (If $L = k$ (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. This follows from the Hunt-Stein Theorem, as in Wesler (1959). The key condition we need to show is that the group $\mathcal{G}$ satisfies a condition known as amenability (see Bondar and Milnes (1981) for various equivalent conditions for amenability). Bondar and Milnes (1981) point out that if a locally compact group is abelian, then it is amenable. The other conditions for the version of the Hunt-Stein Theorem in Wesler (1959) are easily shown.

B Proofs of Theorems

PROOF of Theorem 2: We prove the theorem for a given local parameter h. The extension to any finite number of local parameters $h \in H$ is straightforward. Let

$$\begin{pmatrix} R_n \\ D_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_i \nabla_\gamma \ln f\left(y_i | x_i, \gamma + \frac{v_0}{\sqrt{n}}, \theta + \frac{u_0}{n}\right) \\ \prod_i \mathbf{1}\{y_i \geq g\left(x_i, \theta + \frac{u}{n}\right)\} \end{pmatrix}$$

First, we will show that $R_n | D_n = 1 \rightsquigarrow R$, $R_n \rightsquigarrow R$, and $D_n \rightsquigarrow D$. The joint limiting distribution of (R_n, D_n) and asymptotic independence will follow.

Define $\alpha_n^0 = \alpha + \varphi_n h_0$, $\alpha_n^1 = \alpha + \varphi_n h$, and $\tilde{Y}_{ni} = n^{-1/2} \nabla_\gamma \ln f\left(\tilde{y}_i | \tilde{x}_i, \alpha_n^0\right)$ where the joint density of $(\tilde{y}_i, \tilde{x}_i)$ is given by

$$\tilde{f}(\tilde{y}, \tilde{x}) = \frac{f(\tilde{y}|\tilde{x}, \alpha_n^0) \mathbf{1}\{\tilde{y} \geq g(\tilde{x}, \theta_n^0)\} P_x(\tilde{x})}{\Pr_{\alpha_n^0}(y \geq g(\tilde{x}, \theta_n^1))} \mathbf{1}\{y \geq g(x, \theta_n^1)\}$$

In particular, $(\tilde{y}, \tilde{x})$ is distributed as (y, x) conditional on $y \geq g(\tilde{x}, \theta_n^1)$. Likewise, $\sum_i \tilde{Y}_{ni}$ is distributed as $R_n | D_n = 1$.

To show $\sum_i \tilde{Y}_{ni} \rightsquigarrow N(0, -E_x[E_\alpha[\nabla_{\gamma\gamma} \ln f(y|x,\alpha)]])$, we verify the conditions of the Lindeberg-Feller CLT: (i) $\|V(\tilde{Y}_{ni})\| < \infty$; (ii) $\sum_i E\|\tilde{Y}_{ni}\|^2 \mathbf{1}\{\|\tilde{Y}_{ni}\| > \varepsilon\} \longrightarrow 0$; (iii) $E[\tilde{Y}_{ni}] = 0$ and $\sum_i V(\tilde{Y}_{ni}) = -E_x[E_\alpha[\nabla_{\gamma\gamma} \ln f(y|x,\alpha)]]$.

$$
\begin{aligned}
E[\tilde{Y}_{ni}] &= \frac{1}{\sqrt{n}} \int\int \nabla_\gamma \ln f\left(\tilde{y}|\tilde{x},\alpha_n^0\right) \frac{f(\tilde{y}|\tilde{x},\alpha_n^0)\mathbf{1}\{\tilde{y} \geq \max(g(\tilde{x},\theta_n^0), g(\tilde{x},\theta_n^1))\}}{\Pr_{\alpha_n^0}(y \geq g(x,\theta_n^1))} P_x(\tilde{x}) d\tilde{y} d\tilde{x} \\
&= \frac{1}{\sqrt{n}} \nabla_\gamma \int\int f\left(\tilde{y}|\tilde{x},\alpha_n^0\right) \frac{\mathbf{1}\{\tilde{y} \geq \max(g(\tilde{x},\theta_n^0), g(\tilde{x},\theta_n^1))\}}{\Pr_{\alpha_n^0}(y \geq g(x,\theta_n^1))} P_x(\tilde{x}) d\tilde{y} d\tilde{x} \\
&= 0 \qquad \text{by Assumptions 2, 3, and 6}
\end{aligned}
$$

$$
\begin{aligned}
\|V(\tilde{Y}_{ni})\| &\leq \frac{1}{n} \int\int \|\nabla_\gamma \ln f\left(\tilde{y}|\tilde{x},\alpha_n^0\right) \nabla_\gamma \ln f\left(\tilde{y}|\tilde{x},\alpha_n^0\right)' \| \frac{f\left(\tilde{y}|\tilde{x},\alpha_n^0\right)}{\Pr_{\alpha_n^0}(y \geq g(x,\theta_n^1))} \\
&\qquad \cdot \mathbf{1}\{\tilde{y} \geq \max(g(\tilde{x},\theta_n^0), g(\tilde{x},\theta_n^1))\} P_x(\tilde{x}) d\tilde{y} d\tilde{x} \\
&\leq \frac{1}{n} \int\int \frac{\|\nabla_\gamma f\left(\tilde{y}|\tilde{x},\alpha_n^0\right)\|^2}{f\left(\tilde{y}|\tilde{x},\alpha_n^0\right) \Pr_{\alpha_n^0}(y \geq g(x,\theta_n^1))} \mathbf{1}\{\tilde{y} \geq g(\tilde{x},\theta_n^0)\} P_x(\tilde{x}) d\tilde{y} d\tilde{x} \\
&< \infty \qquad \text{by Assumptions 3 and 6}
\end{aligned}
$$

Similarly, noting that $\Pr_\alpha(y \geq g(x,\theta)) = 1$, under Assumptions 2 and 3, $\sum_i V(\tilde{Y}_{ni}) \longrightarrow E_x[E_\alpha[\nabla_\gamma \ln f\left(y|x,\alpha\right) \nabla_\gamma \ln f\left(y|x,\alpha\right)']]$ and $\sum_i E\|\tilde{Y}_{ni}\|^2 \mathbf{1}\{\|\tilde{Y}_{ni}\| > \varepsilon\} \longrightarrow 0$.

By the Lindeberg-Feller CLT, $\sum_i \tilde{Y}_{ni} \rightsquigarrow N(0, I_\gamma)$. Similarly, for $Y_{ni} = \frac{1}{\sqrt{n}} \nabla_\gamma \ln f\left(y_i|x_i,\alpha_n^0\right)$, we can show $R_n = \sum_i Y_{ni} \rightsquigarrow N(0, -E_x[E_\alpha[\nabla_{\gamma\gamma} \ln f(y|x,\alpha)]])$, by the Lindeberg-Feller CLT, as above.

Next note $D_n = \prod_i \mathbf{1}\{y_i \geq g\left(x_i,\theta_n^1\right)\} \rightsquigarrow \text{Bernoulli}(\lim_{n\longrightarrow\infty} E_{\alpha_n^0}[\prod_i \mathbf{1}\{y_i \geq g\left(x_i,\theta_n^1\right)\}])$ Now compute $\lim_{n\longrightarrow\infty} E_{\alpha_n^0}[\prod_i \mathbf{1}\{y_i \geq g\left(x_i,\theta_n^1\right)\}]$.

$$
\begin{aligned}
E_{\alpha_n^0}[\mathbf{1}\{y_i \geq g\left(x_i,\theta_n^1\right)\}] &= 1 - E_x[F_{y|x,\alpha_n^0}(g(x,\theta_n^1)|x)] \\
&= 1 - E_x[\mathbf{1}\{g(x,\theta_n^1) - g(x,\theta_n^0) \geq 0\}(g(x,\theta_n^1) - g(x,\theta_n^0)) f(g(x,\theta + \frac{\bar{u}}{n})|x,\alpha_n^0)]
\end{aligned}
$$

By independence, $E_{\alpha_n^0}[\prod_i \mathbf{1}\{y_i \geq g\left(x_i,\theta_n^1\right)\}] \longrightarrow \exp(-E_x[\mathbf{1}\{\nabla_\theta g(x,\theta)'(u-u_0) \geq 0\}\ f(g(x,\theta|x,\alpha) \cdot\nabla_\theta g(x,\theta)'](u-u_0))$ by dominated convergence and Assumptions 2 and 5. So,

$$
D_n \rightsquigarrow \text{Bernoulli}(\exp(-E_x[\mathbf{1}\{\nabla_\theta g(x,\theta)'(u-u_0) \geq 0\} f(g(x,\theta)|x,\alpha) \nabla_\theta g(x,\theta)'](u-u_0))).
$$

Let $S_n^1 = \sum_i \tilde{Y}_{ni}$. Then S_n^1 is distributed as $R_n|D_n = 1$. Similarly, define S_n^0 as the random variable that is distributed as $R_n|D_n = 0$. We have shown $R_n \rightsquigarrow R$, $S_n^1 \rightsquigarrow R$, and $D_n \rightsquigarrow D$, where the distributions of R and D are as given above. Now we can show that R_n and D_n are asymptotically independent. Let m be any continuous, bounded real-valued function.

$$
\begin{aligned}
E_n[m(R_n)] &= E_n[m(R_n)|D_n = 1]\mathrm{Pr}_n(D_n = 1) + E_n[m(R_n)|D_n = 0]\mathrm{Pr}_n(D_n = 0) \\
&= E_n[m(S_n^1)]\mathrm{Pr}_n(D_n = 1) + E_n[m(S_n^0)]\mathrm{Pr}_n(D_n = 0)
\end{aligned}
$$

Note that $\mathrm{Pr}_n(D_n = 0) \longrightarrow \mathrm{Pr}(D = 0) \quad (> 0)$, $\mathrm{Pr}_n(D_n = 1) \longrightarrow \mathrm{Pr}(D = 1)$, $E_n[m(R_n)] \longrightarrow E[m(R)]$, and $E_n[m(S_n^1)] \longrightarrow E[m(R)]$. So, $E_n[m(S_n^0)] \longrightarrow E[m(R)]$. Hence, $S_n^0 \rightsquigarrow R$.

Now joint weak convergence of (R_n, D_n) follows. Let m be a continuous, bounded real-valued function.

$$
\begin{aligned}
E_n[m(R_n, D_n)] &= E_n[m(R_n, 1)|D_n = 1]\mathrm{Pr}_n(D_n = 1) + E_n[m(R_n, 0)|D_n = 0]\mathrm{Pr}_n(D_n = 0) \\
&= E_n[m(S_n^1, 1)]\mathrm{Pr}_n(D_n = 1) + E_n[m(S_n^0, 0)]\mathrm{Pr}_n(D_n = 0) \\
&\longrightarrow E[m(R, 1)]\mathrm{Pr}(D = 1) + E[m(R, 0)]\mathrm{Pr}(D = 0)
\end{aligned}
$$

where the last expression is equal to $E[m(R, D)]$ when the law of (R, D) is defined to be the product of the marginal laws for R and D. We have shown that (R_n, D_n) converges weakly and that it converges to an independent joint limit distribution.

Next we consider terms in the expansion of the log of the likelihood ratio. It will be shown that the likelihood ratio is a product of continuous functions in R_n and D_n, and the result of the theorem will then follow by the weak convergence results just derived above.

Now for some mean values $\bar{u}$ and $\bar{v}$,

$$
\begin{aligned}
&\sum_i \left[\ln f(y_i|x_i,\gamma+\frac{v}{\sqrt{n}},\theta+\frac{u}{n}) - \ln f(y_i|x_i,\gamma+\frac{v_0}{\sqrt{n}},\theta+\frac{u_0}{n})\right] \\
&= \sum_i \left[\nabla_\theta \ln f\left(y_i|x_i,\gamma+\frac{v_0}{\sqrt{n}},\theta+\frac{u_0}{n}\right)'\frac{(u-u_0)}{n} + \nabla_\gamma \ln f\left(y_i|x_i,\gamma+\frac{v_0}{\sqrt{n}},\theta+\frac{u_0}{n}\right)'\frac{(v-v_0)}{\sqrt{n}}\right. \\
&\quad +\frac{1}{2}\frac{(u-u_0)'}{n}\nabla_{\theta\theta}\ln f\left(y_i|x_i,\gamma+\frac{\bar{v}}{\sqrt{n}},\theta+\frac{\bar{u}}{n}\right)\frac{(u-u_0)}{n} \\
&\quad +\frac{1}{2}\frac{(v-v_0)'}{\sqrt{n}}\nabla_{\gamma\theta}\ln f\left(y_i|x_i,\gamma+\frac{\bar{v}}{\sqrt{n}},\theta+\frac{\bar{u}}{n}\right)\frac{(u-u_0)}{n} \\
&\quad \left. +\frac{1}{2}\frac{(v-v_0)'}{\sqrt{n}}\nabla_{\gamma\gamma}\ln f\left(y_i|x_i,\gamma+\frac{\bar{v}}{\sqrt{n}},\theta+\frac{\bar{u}}{n}\right)\frac{(v-v_0)}{\sqrt{n}}\right]
\end{aligned}
$$

By Markov's Inequality and Assumption 3, the terms $\frac{1}{n^2}\sum_i \nabla_{\theta\theta}\ln f\left(y_i|x_i,\gamma+\frac{\bar{v}}{\sqrt{n}},\theta+\frac{\bar{u}}{n}\right)$, $\frac{1}{n^{3/2}}\sum_i \nabla_{\gamma\theta}\ln f\left(y_i|x_i,\gamma+\frac{\bar{v}}{\sqrt{n}},\theta+\frac{\bar{u}}{n}\right)$ are $o_p(1)$.

Next, note by Markov's LLN,

$$
\begin{aligned}
\frac{1}{n}\sum_i \nabla_{\gamma\gamma}\ln f\left(y_i|x_i,\gamma+\frac{\bar{v}}{\sqrt{n}},\theta+\frac{\bar{u}}{n}\right) &\xrightarrow{p} E\left[\nabla_{\gamma\gamma}\ln f\left(y|x,\alpha\right)\right] \quad \text{and} \\
\frac{1}{n}\sum_i \nabla_{\theta}\ln f\left(y_i|x_i,\gamma+\frac{v_0}{\sqrt{n}},\theta+\frac{u_0}{n}\right) &\xrightarrow{p} E\left[\nabla_{\theta}\ln f\left(y|x,\alpha\right)\right].
\end{aligned}
$$

Under Assumptions 2 and 3, a version of the information matrix equality holds, $I_\gamma = E_x[E_\alpha[\nabla_\gamma \ln f(y|x,\alpha)\nabla_\gamma \ln f(y|x,\alpha)']] = -E_x[E_\alpha[\nabla_{\gamma\gamma}\ln f(y|x,\alpha)]]$. We have

$$
\begin{aligned}
&\sum_i \left[\ln f(y_i|x_i,\gamma+\frac{v}{\sqrt{n}},\theta+\frac{u}{n}) - \ln f(y_i|x_i,\gamma+\frac{v_0}{\sqrt{n}},\theta+\frac{u_0}{n})\right] \\
&= R_n + \frac{1}{2}(v-v_0)'E\left[\nabla_{\gamma\gamma}\ln f\left(y|x,\alpha\right)\right](v-v_0) + E\left[\nabla_\theta \ln f\left(y|x,\alpha\right)\right]'(u-u_0) + o_p(1)
\end{aligned}
$$

Also, note that under Assumptions 2, 3, and 5, $E\left[\nabla_\theta \ln f\left(y|x,\alpha\right)\right] = E[f(g(x,\theta)|x,\alpha)\nabla_\theta g(x,\theta)]$.

Hence,

$$
\begin{aligned}
Z_{n,\alpha+\varphi_n h_0}(h) &= \prod_i \frac{f(y_i|x_i,\theta+\frac{u}{n},\gamma+\frac{v}{\sqrt{n}})\mathbf{1}\{y_i \geq g(x_i,\theta+\frac{u}{n})\}}{f(y_i|x_i,\theta+\frac{u_0}{n},\gamma+\frac{v_0}{\sqrt{n}})\mathbf{1}\{y_i \geq g(x_i,\theta+\frac{u_0}{n})\}} \\
&= \exp\left(R_n + \frac{1}{2}(v-v_0)'E\left[\nabla_{\gamma\gamma}\ln f\left(y|x,\alpha\right)\right](v-v_0)\right)\exp\left(E\left[\nabla_\theta \ln f\left(y|x,\alpha\right)\right]'(u-u_0)+o_p(1)\right)D_n \\
&\overset{h_0}{\rightsquigarrow} \exp\left((v-v_0)'N(-\frac{1}{2}I_\gamma, I_\gamma)\right)\exp\left(E[f(g(x,\theta)|x,\alpha)\nabla_\theta g(x,\theta)'](u-u_0)\right)D.
\end{aligned}
$$

□

To prove Theorem 3, we will need a similar result on the convergence of the likelihood ratio process.

Corollary 4 *For every h_0 and every finite $I \subset H$, uniformly in $\alpha_0 \in \mathcal{N}_0$*

$$(Z_{n,\alpha_0}(h))_{h\in I} \overset{\alpha_0}{\rightsquigarrow} (Z_{\alpha_0}(h))_{h\in I}$$

$$\text{where } (Z_{\alpha_0}(h))_{h\in I} \sim \left(\exp(v'T - \frac{1}{2}v'I_\gamma v)\exp\left(-E[f(g(x,\theta_0)|x,\alpha_0)\nabla_\theta g(x,\theta_0)']u\right) D_h\right)_{h\in I}$$

and $I_\gamma = E_{\alpha_0}\left[\nabla_\gamma \ln f(y|x,\alpha_0)\nabla_\gamma \ln f(y|x,\alpha_0)'\right]$, and 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 $\{h_1,\ldots,h_l\} \subset I$.

$$\begin{aligned} P_{\alpha_0}(D_{h_1}=1,\ldots,D_{h_l}=1) &= \exp(E[\mathbf{1}\{\max\{\nabla_\theta g(x,\theta_0)'u_1,\ldots,\nabla_\theta g(x,\theta_0)'u_l\}>0\} \\ &\quad \cdot f(g(x,\theta_0)|x,\alpha_0)\max\{\nabla_\theta g(x,\theta_0)'u_1,\ldots,\nabla_\theta g(x,\theta_0)'u_l\}]). \end{aligned}$$

PROOF: The verification of the Lindeberg conditions in the proof of Theorem 2 also shows the slightly stronger result that those conditions hold uniformly in $\alpha \in \mathcal{N}_0$. By a uniform version of the Lindeberg-Feller CLT given in Ibragimov and Hasminskii Theorem A.1.15, $\sum_i \tilde{Y}_{ni}$ and $\sum_i Y_{ni}$ converge uniformly in distribution. Since the convergence of $E_{\alpha_n^0}[\prod_i \mathbf{1}\{y_i \ge g\left(x_i,\theta_n^1\right)\}]$ can also be shown uniformly in $\alpha \in \mathcal{N}_0$ under Assumptions 2 and 5, the result follows. □

Some of the details of the following proof are omitted due to space considerations. A more detailed proof is available at http://post.economics.harvard.edu/hier or from the the authors directly.

PROOF of Theorem 3:

To establish the limiting distribution of the Bayes estimator, we generally follow the steps in Ibragimov and Hasminskii's (1981) proof of their Theorem 1.10.2. First, two properties of the likelihood ratio process are established - bounding the tail behavior and bounding small variations. Analogous properties would be taken as primitive assumptions in Ibragimov and Hasminskii; here they are established in Lemmas 5 and 6 from more primitive assumptions (1 - 6). Using the bounds on the likelihood ratio process tails, we can show that the integrals defining expected posterior loss and its limit are well approximated by the corresponding integrals over a large bounded region. Given the bounds on small variations in the likelihood ratio processes, the integrals on the bounded regions are well approximated by the integrands evaluated at a large finite number of points in the region. Given these results, the finite dimensional convergence of the likelihood ratio process (from Corollary 4) translates into finite dimensional convergence of the expected posterior loss to its limit. This convergence can be stengthened to weak convergence after verifying an asymptotic tightness condition. The limiting distribution of the Bayes estimator will follow by an argmax theorem.

For the following lemmas, suppose Assumptions 1 - 6 hold. Let $\mathcal{N}_0$ be the closure of an η-ball around α_0 such that $B(\alpha_0, 2\eta) \subset \mathcal{N}$.

Lemma 5 *There exists constants a, b, $c > 0$ such that*

$$E_{\alpha_0}|Z_{n,\alpha_0}^{1/2}(h_2) - Z_{n,\alpha_0}^{1/2}(h_1)|^2 \leq c(1+R_u)^a \exp(bR_v)(\|u_2-u_1\| + \|v_2-v_1\|^2)$$

for all $h_1, h_2 \in \{h \in \mathcal{U}_n \times \mathcal{V}_n : \|u\| \leq R_u, \|v\| \leq R_v\}$, uniformly for $\alpha_0 \in \mathcal{N}_0$.

PROOF: Define

$$\overline{G}_x = \sup\left\{g\left(x, \theta_0 + \frac{u_1}{n} + t \cdot \frac{(u_2-u_1)}{n}\right) : 0 \leq t \leq 1\right\},$$

$$\underline{G}_x = \inf\left\{g\left(x, \theta_0 + \frac{u_1}{n} + t \cdot \frac{(u_2-u_1)}{n}\right) : 0 \leq t \leq 1\right\}.$$

$$\begin{aligned}
E_{\alpha_0}\left|Z_{n,\alpha_0}^{1/2}(h_2) - Z_{n,\alpha_0}^{1/2}(h_1)\right|^2 &\leq \int \left(p^{1/2}(z|\alpha_0+\varphi_n h_2) - p^{1/2}(z|\alpha_0+\varphi_n h_1)\right)^2 dz \\
&\quad \cdot f^{1/2}(y|\alpha_0+\varphi_n h_1)\mathbf{1}\{y \geq g(x,\theta_0+\frac{u_1}{n})\}P_x(x)dydx\Big]^n\Big) \\
&\leq 2n\left(1 - \int\int f^{1/2}(y|\alpha_0+\varphi_n h_2)\mathbf{1}\{y \geq g(x,\theta_0+\frac{u_2}{n})\}\right. \\
&\quad \left. \cdot f^{1/2}(y|\alpha_0+\varphi_n h_1)\mathbf{1}\{y \geq g(x,\theta_0+\frac{u_1}{n})\}P_x(x)dydx\right) \\
&\leq n\int\int_{g(x,\theta_0)}^{\infty} |f^{1/2}(y|x,\alpha_0+\varphi_n h_2) - f^{1/2}(y|x,\alpha_0+\varphi_n h_1)|^2 dyP_x(x)dx \\
&\quad + n\int\int_{\min(\overline{G}_x, g(x,\theta))}^{g(x,\theta_0)} |f^{1/2}(y|x,\alpha_0+\varphi_n h_2) - f^{1/2}(y|x,\alpha_0+\varphi_n h_1)|^2 dyP_x(x)dx \\
&\quad + n\int\int_{\underline{G}_x}^{\overline{G}_x} |f^{1/2}(y|x,\alpha_0+\varphi_n h_2)\mathbf{1}(y \geq g(x,\theta_0+\frac{u_2}{n})) - \\
&\quad f^{1/2}(y|x,\alpha_0+\varphi_2 h_1)\mathbf{1}(y \geq g(x,\theta_0+\frac{u_1}{n}))|^2 dyP_x(x)dx.
\end{aligned}$$

By Assumptions 1 and 2 and the fundamental theorem of calculus, we can write the first term in the preceding display as

$$\begin{aligned}
& n\int\int_{g(x,\theta_0)}^{\infty}|\int_0^1 \varphi_n(h_2-h_1)\nabla_\alpha f^{1/2}(y|x,\alpha_0+\varphi_n(h_1+t(h_2-h_1)))dt|^2 P_x(x)dx \\
\leq\ & n\int_0^1\int\int_{g(x,\theta_0)}^{\infty}|\varphi_n(h_2-h_1)'\nabla_\alpha f^{1/2}(y|x,\alpha_0+\varphi_2(h_1+t(h_2-h_1)))|^2 dyP_x(x)dxdt \\
& \text{(by Jensen's inequality)} \\
\leq\ & A\frac{\|u_2-u_1\|^2}{n}\max_{i,j}\sup_{t\in[0,1]} J_{ij}^{\theta\theta}(\alpha_0+\varphi_n(h_1+t(h_2-h_1))) \\
& +B\frac{\|v_2-v_1\|}{\sqrt{n}}\|u_2-u_1\|\max_{i,j}\sup_{t\in[0,1]} J_{ij}^{\gamma\theta}(\alpha_0+\varphi_n(h_1+t(h_2-h_1))) \\
& +C\|v_2-v_1\|^2\max_{i,j}\sup_{t\in[0,1]} J_{ij}^{\gamma\gamma}(\alpha_0+\varphi_n(h_1+t(h_2-h_1))).
\end{aligned}$$

This term has the desired form by Assumption 4.

Now consider the third term and apply the inequality $(\sqrt{r}-\sqrt{s})^2 \leq |r-s|$ for $r,s\geq 0$.

$$\begin{aligned}
& \int\int_{\underline{G}_x}^{\overline{G}_x}|f^{1/2}(y|x,\alpha_0+\varphi_n h_2)\mathbf{1}(y\geq g(x,\theta_0+\frac{u_2}{n})) \\
& \qquad -f^{1/2}(y|x,\alpha_0+\varphi_n h_1)\mathbf{1}(y\geq g(x,\theta_0+\frac{u_1}{n}))|^2 dyP_x(x)dx \\
\leq\ & \int(\overline{G}_x-\underline{G}_x)\left[\sup_{y\in[\underline{G}_x,\overline{G}_x]} f(y|x,\alpha_0+\varphi_n h_2)+\sup_{y\in[\underline{G}_x,\overline{G}_x]} f(y|x,\alpha_0+\varphi_n h_1)\right]P_x(x)dx \\
\leq\ & c\frac{||u_2-u_1||}{n}
\end{aligned}$$

where the last inequality follows by Assumption 2 and noting that the first derivative of g is uniformly continuous on $\bar{\mathcal{N}}$ and hence bounded on $\bar{\mathcal{N}}$.

The second term follows in exactly the same way, so we have verified the condition. □

Lemma 6 *For all (u,v) and $\alpha_0\in\mathcal{N}_0$, $E_{\alpha_0}Z_{n,\alpha_0}^{1/2}(u,v)\leq\exp[-g_n(\|u\|,\|v\|)]$, where g_n is a sequence of functions from $[0,\infty)\times[0,\infty)$ into $[0,\infty)$ such that:*

(a) for each $n\geq 1$, g_n is increasing to infinity in each of its arguments.

(b) For any $N_u,N_v\geq 0$, we have $\lim_{n\to\infty,\max\{x,y\}\to\infty} x^{N_u}e^{N_v y}\exp[-g_n(x,y)]=0$.

PROOF: Note that

$$\begin{aligned}
E_{\alpha_0} Z_{n,\alpha_0}^{1/2}(h) &= \Big(\int\int f(y|x,\alpha_0+\varphi_n h)^{1/2} \mathbf{1}\{y \geq g(x,\theta_0+\frac{u}{n})\} f(y|x,\alpha_0)^{1/2} \\
&\qquad \cdot \mathbf{1}\{y \geq g(x,\theta_0)\} dy P_x(x) dx \Big)^n \\
&= \Big(1 - \frac{1}{2} \int\int \big[f(y|x,\alpha_0+\varphi_n h)^{1/2} \mathbf{1}\{y \geq g(x,\theta_0+\frac{u}{n})\} \\
&\qquad - f(y|x,\alpha_0)^{1/2} \mathbf{1}\{y \geq g(x,\theta_0)\} \big]^2 dy P_x(x) dx \Big)^n \\
&\leq \exp\Big(-\frac{n}{2} \int\int \big[f(y|x,\alpha_0+\varphi_n h)^{1/2} \mathbf{1}\{y \geq g(x,\theta_0+\frac{u}{n})\} \\
&\qquad - f(y|x,\alpha_0)^{1/2} \mathbf{1}\{y \geq g(x,\theta_0)\} \big]^2 dy P_x(x) dx \Big)
\end{aligned}$$

where the last inequality follows by $1 - \rho \leq \exp(-\rho)$.

Define $\overline{G}'_x = \sup\{g(x,\theta_0 + t\frac{u}{n}) : 0 \leq t \leq 1\}$, $\underline{G}'_x = \inf\{g(x,\theta_0 + t\frac{u}{n}) : 0 \leq t \leq 1\}$. Write

$$\begin{aligned}
&\int\int \Big[f(y|x,\alpha_0+\varphi_n h)^{1/2} \mathbf{1}\{y \geq g(x,\theta_0+\frac{u}{n})\} - f(y|x,\alpha_0)^{1/2} \mathbf{1}\{y \geq g(x,\theta_0)\} \Big]^2 dy P_x(x) dx \\
&= \int\int_{\overline{G}'_x}^{\infty} |f^{1/2}(y|x,\alpha_0+\varphi_n h) - f^{1/2}(y|x,\alpha_0)|^2 dy P_x(x) dx \\
&+ \int\int_{\underline{G}'_x}^{\overline{G}'_x} |f^{1/2}(y|x,\alpha_0+\varphi_n h) \mathbf{1}\{y \geq g(x,\theta_0+u/n)\} - f^{1/2}(y|x,\alpha_0) \mathbf{1}\{y \geq g(x,\theta_0)\}|^2 dy P_x(x) dx.
\end{aligned}$$

Let $C_x = \{y : \min\{g(x,\theta_0+u/n), g(x,\theta_0)\} \leq y \leq \max\{g(x,\theta_0+u/n), g(x,\theta_0)\}$. Uniform convergence of $nE[|g(x,\theta_0+u/n) - g(x,\theta_0)| \inf_{C_x} \min\{f(y|x,\alpha_0+\varphi_n h), f(y|x,\alpha_0)\}]$ to $E\left[f(g(x,\theta_0)|x,\alpha_0)\right.$ $\left.\cdot|\nabla_\theta g(x,\theta_0)'u|\right]$ in u for $\|u\| = 1$ and $\alpha_0 \in \mathcal{N}_0$ follows by Assumptions 2 and 5, so that

$$\begin{aligned}
&\lim_{n \longrightarrow \infty} \inf_{\alpha_0 \in \mathcal{N}_0} \inf_{\|u\|=1} E[n|g(x,\theta_0+u/n) - g(x,\theta_0)| \inf_{C_x} \min\{f(y|x,\alpha_0+\varphi_n h), f(y|x,\alpha_0)\}] \\
&= \inf_{\alpha_0 \in \mathcal{N}_0} \inf_{\|u\|=1} E_\alpha \left[f(g(x,\theta_0)|x,\alpha_0) |\nabla_\theta g(x,\theta_0)'u| \right].
\end{aligned}$$

The last inequality below then follows by Assumption 6.

$$\begin{aligned}
&\int\int_{\underline{G}'_x}^{\overline{G}'_x} |f^{1/2}(y|x,\alpha_0+\varphi_n h) \mathbf{1}(y \geq g(x,\theta_0+u/n)) \\
&\qquad - f^{1/2}(y|x,\alpha_0) \mathbf{1}(y \geq g(x,\theta_0))|^2 dy P_x(x) dx \\
&\geq E[|g(x,\theta_0+u/n) - g(x,\theta_0)| \inf_{C_x} \min\{f(y|x,\alpha_0+\varphi_n h), f(y|x,\alpha_0)\}] \\
&\geq C \frac{\|u\|}{n} \qquad (1)
\end{aligned}$$

Now consider the first term. As in the proof of Lemma 5, as $n \longrightarrow \infty$,

$$n \int \int_{\overline{G}'_x}^{\infty} |f^{1/2}(y|x, \alpha_0 + \varphi_n h) - f^{1/2}(y|x, \alpha_0)|^2 dy P_x(x) dx \longrightarrow \sum_{i=1}^{d} \sum_{j=1}^{d} v_i v_j \frac{1}{4} J_{ij}^{\gamma\gamma}(\alpha_0)$$

where the convergence is uniform in $\alpha_0 \in \mathcal{N}_0$, since $J^{\theta\theta}$, $J^{\gamma\theta}$, and $J^{\gamma\gamma}$ are continuous, hence uniformly continuous, on $\bar{\mathcal{N}}$.

Now by the Fatou's Lemma and the Cauchy-Schwarz Inequality,

$$\begin{aligned}
&\int \int_{\overline{G}'_x}^{\infty} |f^{1/2}(y|x, \alpha_0 + \varphi_n h) - f^{1/2}(y|x, \alpha_0)|^2 dy P_x(x) dx \\
&= \int \int_{\overline{G}'_x}^{\infty} |(\varphi_n h)' \nabla_\alpha f^{1/2}(y|x, \alpha_0)|^2 dy P_x(x) dx + o(||\varphi_n h||^2) \\
&\geq \int \int_{g(x,\theta_0)}^{\infty} |(\varphi_n h)' \nabla_\alpha f^{1/2}(y|x, \alpha_0)|^2 dy P_x(x) dx + o(||\varphi_n h||^2) \\
&\quad - \int \int_{g(x,\theta_0)}^{\overline{G}'_x} |(\varphi_n h)' \nabla_\alpha f^{1/2}(y|x, \alpha_0)|^2 dy P_x(x) dx \\
&= \frac{1}{n} \left[\frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{k} u_i u_j \frac{1}{4} J_{ij}^{\theta\theta}(\alpha_0) + 2 \sum_{i=1}^{k} \sum_{j=1}^{d} u_i \frac{v_j}{\sqrt{n}} \frac{1}{4} J_{ij}^{\gamma\theta}(\alpha_0) + \sum_{i=1}^{d} \sum_{j=1}^{d} v_i v_j \frac{1}{4} J_{ij}^{\gamma\gamma}(\alpha_0) \right] \\
&\quad - \int \int_{g(x,\theta_0)}^{\overline{G}'_x} |(\varphi_n h)' \nabla_\alpha f^{1/2}(y|x, \alpha_0)|^2 dy P_x(x) dx + o(||\varphi_n h||^2) \\
&\geq \frac{1}{n} \left[-C \frac{\|u\|}{2} + B \|v\|^2 \right]
\end{aligned}$$

The last inequality following for sufficiently large n and some $B > 0$ (C the same constant as used for second term) by Assumption 4 and noting

$$\begin{aligned}
&\int \int_{g(x,\theta_0)}^{\overline{G}'_x} |(\varphi_n h)' \nabla_\alpha f^{1/2}(y|x, \alpha_0)|^2 dy P_x(x) dx + o(||\varphi_n h||^2) \\
&\leq \frac{1}{n} \|h\|^2 \int \int_{g(x,\theta_0)}^{\overline{G}'_x} \sup_{y \in [g(x,\theta_0), \overline{G}'_x]} \|\nabla_\alpha f^{1/2}(y|x, \alpha_0)\|^2 dy P_x(x) dx + o(||\varphi_n h||^2) \\
&\leq c \cdot \frac{1}{n} \|h\|^2 \int |\overline{G}'_x - g(x, \theta_0)| P_x(x) dx + o(||\varphi_n h||^2) \;=\; o\left(\frac{1}{n}\right)
\end{aligned}$$

We have shown that there exists $a > 0$ such that for large enough n

$$\begin{aligned}
&\int \int \left[f(y|x, \alpha_0 + \varphi_n h)^{1/2} \mathbf{1}\{y \geq g(x, \theta_0 + \frac{u}{n})\} - f(y|x, \alpha_0)^{1/2} \mathbf{1}\{y \geq g(x, \theta_0)\} \right]^2 dy P_x(x) dx \\
&\geq \frac{a}{n} (\|u\| + \|v\|^2)
\end{aligned}$$

suggesting $g_n(x, y) = \left(\frac{a}{2}\right) (\|x\| + \|y\|^2)$. $\square$

We pause here to note that Theorem 3 could be restated in terms of higher-level assumptions so that its results could be applicable directly to other models. In particular, to obtain the limiting distribution of the Bayes estimator part of the conclusion, Assumptions 1 - 6 could be replaced by the conclusions of Lemmas 5 and 6 and Corollary 4. If the conclusions of Lemmas 5′, 6′, and 2′ (stated below) are additionally assumed, then the efficiency result of Theorem 3 also follows. The rest of the proof is modularized to depend only on these likelihood ratio process properties without reference to Assumptions 1 - 6.

Before proving finite-dimensional convergence and asymptotic tightness of the expected posterior loss, we need to establish a few useful intermediate results. The following lemmas include generalizations or extensions of Ibragimov and Hasminskii (1981) Lemma 1.5.1, Lemma 1.5.2, Theorem 1.5.2, and Lemma A.1.22. For the following lemmas, suppose Assumptions 1 - 8 hold. The proofs are omitted due to space considerations and are available at http://post.economics.harvard.edu/hier or from the authors directly.

Lemma 7 *For all δ and η small enough and $\alpha_0 \in \mathcal{N}_0$,*

$$P_{n,\alpha_0}\left(\int_0^\eta \cdots \int_0^\eta \int_0^\delta \cdots \int_0^\delta Z_{n,\alpha_0}(u,v)\pi(\theta + \frac{u}{n}, \gamma + \frac{v}{\sqrt{n}})dudv < \frac{\pi(\theta,\gamma)}{4}\delta^k \eta^d\right)$$
$$< 2A^{1/2}(k^{1/4} + d^{1/2})(\delta^{1/2} + \eta)$$

where k is the dimension of Θ and d is the dimension of Γ.

Define $\Gamma_H^u = \{u : H \leq \|u\| < H+1\} \cap U_n$, $\Gamma_J^v = \{v : J \leq \|v\| < J+1\} \cap V_n$,

$$I_{n,\alpha_0,HJ} = \int_{\Gamma_H^u \times \Gamma_J^v} Z_{n,\alpha_0}(u,v)\pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}})dudv$$

$$Q_{n,\alpha_0,HJ} = \frac{I_{n,\alpha_0,HJ}}{\int_{U_n \times V_n} Z_{n,\alpha_0}(u,v)\pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}})dudv}.$$

Lemma 8 *There exist constants $B, C, b > 0$ such that for $\alpha_0 \in \mathcal{N}_0$,*

$$P_{n,\alpha_0}(I_{n,\alpha_0,HJ} > e^{-bg_n(H,J)}) \leq B(1 + H^B)\exp(CJ)e^{-bg_n(H,J)}$$

$$E_{n,\alpha_0}[Q_{n,\alpha_0,HJ}] \leq B(1 + H^B)\exp(CJ)e^{-bg_n(H,J)}.$$

Lemma 9 *For any N, uniformly in $\alpha_0 \in \mathcal{N}_0$,* $\lim_{n,H\longrightarrow\infty} H^N P_{n,\alpha_0}(\|\varphi_n^{-1}(\tilde{\alpha}_n - \alpha_0)\| > H) = 0.$

Lemma 10 *The distributions of the integrals $\int_{\|u,v\|\leq M} l(s-u,t-v)Z_{n,\alpha_0}(u,v)dudv$ and $\int_{\|u,v\|\leq M} Z_{n,\alpha_0}(u,v)dudv$ converge uniformly for $\alpha_0 \in \mathcal{N}_0$ to the distributions of the integrals $\int_{\|u,v\|\leq M} l(s-u,t-v)Z_{\alpha_0}(u,v)dudv$ and $\int_{\|u,v\|\leq M} Z_{\alpha_0}(u,v)dudv$.*

PROOF: Note that $\sup_{n,u,v} E_{n,\alpha_0} Z_{n,\alpha_0}(u,v) \leq 1 < \infty$, so

$$\sup_n E_{n,\alpha_0} \int_{\|u,v\|\leq M} |l(s-u,t-v)||Z_{n,\alpha_0}(u,v)|dudv < \infty.$$

Also, $E_{n,\alpha_0}[Z_{n,\alpha_0}^{1/2}(u_1,v_1) + Z_{n,\alpha_0}^{1/2}(u_2,v_2)]^2 \leq 4\sup_{u,v} E_{n,\alpha_0} Z_{n,\alpha_0}(u,v)$, and

$$E_{n,\alpha_0}|Z_{n,\alpha_0}(u_1,v_1) - Z_{n,\alpha_0}(u_2,v_2)| \leq 2[E_{n,\alpha_0}|Z_{n,\alpha_0}^{1/2}(u_1,v_1) - Z_{n,\alpha_0}^{1/2}(u_2,v_2)|^2]^{1/2}[\sup_{u,v} E_{n,\alpha_0} Z_{n,\alpha_0}(u,v)]^{1/2}.$$

For fixed M, $\|u_1,v_1\| \leq M$, and $\|u_2,v_2\| \leq M$,

$$\sup_n E_{n,\alpha_0}|Z_{n,\alpha_0}(u_1,v_1) - Z_{n,\alpha_0}(u_2,v_2)| \leq B'(\|u_1-u_2\|^{1/2} + \|v_1-v_2\|).$$

By marginal convergence of Z_{n,α_0} the above conditions also hold for Z_{α_0}. Partition $\Re^{k+d}$ into cubes with edges of length δ parallel to the coordinate axes. Let Δ_j be the intersection of the j^{th} cube with the set $\{(u,v) : \|u,v\| \leq M\}$. Choose a point (u_j,v_j) in each Δ_j. Noting that $|e^{itc} - e^{itb}| \leq |t||\int_b^c e^{ity}dy| \leq |t||c-b|$, as $\delta \longrightarrow 0$,

$$\begin{aligned} &\Bigg| E_{n,\alpha_0} \exp\left[i\lambda \sum_j Z_{n,\alpha_0}(u_j,v_j) \int_{\Delta_j} l(s-u,t-v)dudv \right] \\ &\qquad - E_{n,\alpha_0} \exp\left[i\lambda \int_{\|u,v\|\leq M} l(s-u,t-v)Z_{n,\alpha_0}(u,v)dudv \right] \Bigg| \\ &\leq \ |\lambda| \sum_j \int_{\Delta_j} l(s-u,t-v)E_{n,\alpha_0}|Z_{n,\alpha_0}(u,v) - Z_{n,\alpha_0}(u_j,v_j)|dudv \\ &\leq \ B'|\lambda| \sup_{\|u,v\|\leq M} |l(s-u,t-v)|((k\delta)^{1/2} + (d\delta))\mathrm{mes}\{(u,v) : \|u,v\| \leq M\} \longrightarrow 0 \end{aligned}$$

The same result holds with Z_{α_0} replacing Z_{n,α_0}. By the (uniform) marginal convergence of Z_{n,α_0} in Corollary 4, the characteristic functions converge, $E_{n,\alpha_0} \exp[i\lambda \int_{\|u,v\|\leq M} l(s-u,t-v) \cdot Z_{n,\alpha_0}(u,v)dudv] \longrightarrow E_\alpha \exp[i\lambda \int_{\|u,v\|\leq M} l(s-u,t-v)Z_{\alpha_0}(u,v)dudv]$ giving the conclusion of the lemma. □

Lemma 11 *For any* $\{h_1, \ldots, h_m\} \subset \Re^{k+d}$, $(\psi_{n,\alpha_0}(h_1), \ldots, \psi_{n,\alpha_0}(h_m))$ $\rightsquigarrow (\psi_{\alpha_0}(h_1), \ldots, \psi_{\alpha_0}(h_m))$ *uniformly for* $\alpha_0 \in \mathcal{N}_0$.

PROOF:

From lemma 9, $P_{n,\alpha_0}(\|\varphi_n^{-1}(\tilde{\alpha}_n - \alpha_0)\| > M) \longrightarrow 0$, uniformly for n large enough as $M \longrightarrow \infty$.

Let $L \geq M$,

$$\begin{aligned} & \psi_{n,\alpha_0}(s,t) - \frac{\int_{\|u,v\| \leq M} l(s-u, t-v) Z_{n,\alpha_0}(u,v) \pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}}) du dv}{\int_{\|\tilde{u},\tilde{v}\| \leq L} Z_{n,\alpha_0}(\tilde{u},\tilde{v}) \pi(\theta_0 + \frac{\tilde{u}}{n}, \gamma_0 + \frac{\tilde{v}}{\sqrt{n}}) d\tilde{u} d\tilde{v}} \\ = & \frac{\int_{\|u,v\| > M} l(s-u, t-v) Z_{n,\alpha_0}(u,v) \pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}}) du dv}{\int Z_{n,\alpha_0}(\tilde{u},\tilde{v}) \pi(\theta_0 + \frac{\tilde{u}}{n}, \gamma_0 + \frac{\tilde{v}}{\sqrt{n}}) d\tilde{u} d\tilde{v}} \qquad (2) \\ & - \frac{\int_{\|u,v\| \leq M} l(s-u, t-v) Z_{n,\alpha_0}(u,v) \pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}}) du dv}{\int_{\|\tilde{u},\tilde{v}\| \leq L} Z_{n,\alpha_0}(\tilde{u},\tilde{v}) \pi(\theta_0 + \frac{\tilde{u}}{n}, \gamma_0 + \frac{\tilde{v}}{\sqrt{n}}) d\tilde{u} d\tilde{v}} \\ & \cdot \frac{\int_{\|u,v\| > L} Z_{n,\alpha_0}(u,v) \pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}}) du dv}{\int Z_{n,\alpha_0}(\tilde{u},\tilde{v}) \pi(\theta_0 + \frac{\tilde{u}}{n}, \gamma_0 + \frac{\tilde{v}}{\sqrt{n}}) d\tilde{u} d\tilde{v}} \qquad (3) \end{aligned}$$

For fixed s, t, by Assumption 7 and Lemma 8,

$$\begin{aligned} & E_{n,\alpha_0} \left| \frac{\int_{\|u,v\| > M} l(s-u, t-v) Z_{n,\alpha_0}(u,v) \pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}}) du dv}{\int Z_{n,\alpha_0}(\tilde{u},\tilde{v}) \pi(\theta_0 + \frac{\tilde{u}}{n}, \gamma_0 + \frac{\tilde{v}}{\sqrt{n}}) d\tilde{u} d\tilde{v}} \right| \\ \leq \; & B e^{-a g_n(M/2,0)/2} + B' e^{-a g_n(0,M/2)/2} \end{aligned}$$

Similarly, the same bound holds for term (3), using

$$\frac{\int_{\|u,v\| \leq M} Z_{n,\alpha_0}(u,v) \pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}}) du dv}{\int_{\|\tilde{u},\tilde{v}\| \leq L} Z_{n,\alpha_0}(\tilde{u},\tilde{v}) \pi(\theta_0 + \frac{\tilde{u}}{n}, \gamma_0 + \frac{\tilde{v}}{\sqrt{n}}) d\tilde{u} d\tilde{v}} \leq 1.$$

It follows by Markov's Inequality that uniformly for $\alpha_0 \in \mathcal{N}_0$

$$\psi_{n,\alpha_0}(s,t) - \frac{\int_{\|u,v\| \leq M} l(s-u, t-v) Z_{n,\alpha_0}(u,v) \pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}}) du dv}{\int_{\|\tilde{u},\tilde{v}\| \leq L} Z_{n,\alpha_0}(\tilde{u},\tilde{v}) \pi(\theta_0 + \frac{\tilde{u}}{n}, \gamma_0 + \frac{\tilde{v}}{\sqrt{n}}) d\tilde{u} d\tilde{v}} \xrightarrow{p} 0 \qquad (4)$$

as $M \longrightarrow \infty$ uniformly for large enough n.

The prior π is positive and continuous at α_0, so $\pi(\theta_0 + (\tilde{u}/n), \gamma_0 + (\tilde{v}/\sqrt{n})) \longrightarrow \pi(\theta_0, \gamma_0)$ uniformly on $\|u,v\| \leq M$ and $\alpha_0 \in \mathcal{N}_0$ as $n \longrightarrow \infty$. So, applying the Cramer-Wold device and Lemma 10, we can obtain uniform convergence of the marginal distributions of

$$\frac{\int_{\|u,v\| \leq M} l(s-u, t-v) Z_{n,\alpha_0}(u,v) \pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}}) du dv}{\int_{\|\tilde{u},\tilde{v}\| \leq L} Z_{n,\alpha_0}(\tilde{u},\tilde{v}) \pi(\theta_0 + \frac{\tilde{u}}{n}, \gamma_0 + \frac{\tilde{v}}{\sqrt{n}}) d\tilde{u} d\tilde{v}}$$

for $M, L < \infty$.

Then by another application of Markov's Inequality, as in (4),

$$\psi_{\alpha_0}(s,t) - \frac{\int_{\|u,v\|\le M} l(s-u,t-v) Z_{\alpha_0}(u,v)dudv}{\int_{\|\tilde{u},\tilde{v}\|\le L} Z_{\alpha_0}(\tilde{u},\tilde{v})d\tilde{u}d\tilde{v}} \xrightarrow{p} 0$$

as $M \longrightarrow \infty$. The conclusion of the lemma follows. □

Lemma 12 *For every $\varepsilon, \eta > 0$, there exists $\delta > 0$ such that:*

$$\lim \sup_{n\longrightarrow\infty} \sup_{\alpha_0 \in \mathcal{N}_0} P_{n,\alpha_0}\left(\sup_{\|s_1-s_2,t_1-t_2\|\le\delta, \|s_1,t_1\|\le M, \|s_2,t_2\|\le M} |\psi_{n,\alpha_0}(s_1,t_1) - \psi_{n,\alpha_0}(s_2,t_2)| \ge \varepsilon \right) \le \eta$$

PROOF:

Note that the part of the proof of Lemma 11 that shows $\psi_{n,\alpha_0}(s,t)$ can be approximated by a ratio of integrals on a compact space can be straightforwardly extended from fixed s, t to uniform convergence on $\{(s,t) : \|s,t\| \le M\}$.

By Assumption 7, $l(s-u,t-v)$ is uniformly continuous on $\|s,t\| \le M$, $\|u,v\| \le N$, so we can choose δ small enough to make $\sup_{\|s_1-s_2,t_1-t_2\|\le\delta, \|s_1,t_1\|\le M, \|s_2,t_2\|\le M, \|u,v\|\le N} [l(s_1-u,t_1-v) - l(s_2-u,t_2-v)]$ as small as desired. Since

$$\sup_{\alpha_0 \in \mathcal{N}_0} E_{n,\alpha_0} \left| \frac{\int_{\|u,v\|\le N} Z_{n,\alpha_0}(u,v)\pi(\theta_0 + \frac{u}{n}, \gamma_0 + \frac{v}{\sqrt{n}})dudv}{\int_{\|\tilde{u},\tilde{v}\|\le L} Z_{n,\alpha_0}(\tilde{u},\tilde{v})\pi(\theta_0 + \frac{\tilde{u}}{n}, \gamma_0 + \frac{\tilde{v}}{\sqrt{n}})d\tilde{u}d\tilde{v}} \right| < \infty$$

we can then use Markov's Inequality to arrive at the conclusion of the lemma. □

Lemmas 11 and 12 establish weak convergence of the stochastic process ψ_{n,α_0} to ψ_{α_0}. They are also sufficient to ensure that almost all sample paths of ψ_{α_0} are continuous. Then by Lemmas 9 and 9′ (below), we can apply an argmax theorem, such as Theorem 3.2.2 in Van der Vaart and Wellner (1996), to establish the limiting distribution of the Bayes estimator.

Now we turn to showing asymptotic efficiency. Define the limiting risk function $L(\alpha) = lim_{n\longrightarrow\infty} E_{n,\alpha}[l(\varphi_n^{-1}(\tilde{\alpha}_n - \alpha))]$. We want to show that the convergence to this limiting risk function is uniform in $\alpha \in \mathcal{N}_0$. Note that conditions given in the first part of this proof are sufficient for uniform convergence in distribution of the Bayes estimator for $\alpha_0 \in \mathcal{N}_0$. Also, using Assumption 7 and

Lemma 9, by choosing H large, we can make the bound on $E_{n,\alpha_0}[l(\varphi_n^{-1}(\tilde{\alpha}_n-\alpha_0))\mathbf{1}\{\|\varphi_n^{-1}(\tilde{\alpha}_n-\alpha_0)\| > H\}]$ as small as desired. Let $M = B(1+H^b)$ and $l_M(h) = \min\{M, l(h)\}$.

$$\begin{aligned}
&\sup_{\alpha_0\in\mathcal{N}_0} |E_{n,\alpha_0} l(\varphi_n^{-1}(\tilde{\alpha}_n-\alpha_0)) - E_{\alpha_0} l(\tau_{\alpha_0})| \\
&\quad\leq \sup_{\alpha_0\in\mathcal{N}_0} |E_{n,\alpha_0} l(\varphi_n^{-1}(\tilde{\alpha}_n-\alpha_0)) - E_{n,\alpha_0} l_M(\varphi_n^{-1}(\tilde{\alpha}_n-\alpha_0))| \\
&\qquad + \sup_{\alpha_0\in\mathcal{N}_0} |E_{n,\alpha_0} l_M(\varphi_n^{-1}(\tilde{\alpha}_n-\alpha_0)) - E_{\alpha_0} l_M(\tau_{\alpha_0})| + \sup_{\alpha_0\in\mathcal{N}_0} |E_{\alpha_0} l_M(\tau_{\alpha_0}) - E_{\alpha_0} l(\tau_{\alpha_0})|
\end{aligned}$$

Since $|E_{n,\alpha_0} l(\varphi_n^{-1}(\tilde{\alpha}_n-\alpha_0)) - E_{n,\alpha_0} l_M(\varphi_n^{-1}(\tilde{\alpha}_n-\alpha_0))| \leq E_{n,\alpha_0}[l(\varphi_n^{-1}(\tilde{\alpha}_n-\alpha_0))\ \mathbf{1}\{\|\varphi_n^{-1}(\tilde{\alpha}_n-\alpha_0)\| > H\}]$, the above bound is sufficient to make the first term as small as desired by choosing H and n sufficiently large, and a similar argument can be made for the third term using Lemma $9'$. Since l_M is countinuous and bounded, the second term can be made as small as desired by the uniform convergence in distribution of the Bayes estimator.

Next we establish weak convergence of $\tau_{\alpha'}$ to τ_{α_0} when $\alpha' \longrightarrow \alpha_0$. To show the weak convergence, we follow a similar argument to the one used to establish the limiting distribution of the Bayes estimator. Corresponding lemmas are denoted with a prime. This result is used to show continuity of the limiting risk function, which will lead to the desired efficiency result.

Lemma 5′ *There exists a, b, $c > 0$ such that*

$$E_\alpha |Z_{\alpha_0}^{1/2}(h_2) - Z_{\alpha_0}^{1/2}(h_1)|^2 \leq c(1+R_u)^a \exp(bR_v)(\|u_2-u_1\| + \|v_2-v_1\|^2)$$

for all $h_1, h_2 \in \{h : \|u\| \leq R_u, \|v\| \leq R_v\}$ and any α_0.

PROOF: The lemma follows by Lemma5 and the finite dimensional convergence in Corollary 4.
□

Lemma 6′ *For all (u,v) and $\alpha_0 \in \mathcal{N}$, $E_{\alpha_0} Z_{\alpha_0}^{1/2}(u,v) \leq \exp[-G(\|u\|,\|v\|)]$, where G is a sequence of functions from $[0,\infty)\times[0,\infty)$ into $[0,\infty)$ such that:*

(a) G is increasing to infinity in each of its arguments.

(b) For any $N_u, N_v \geq 0$, we have $\lim_{\max\{x,y\}\to\infty} x^{N_u} e^{N_v y} \exp[-G(x,y)] = 0$.

PROOF:

$$E_{\alpha_0} Z_{\alpha_0}^{1/2}(u,v) = \exp\big(-\frac{1}{8} v' I_\gamma v - \frac{1}{2} E_{\alpha_0}\left[f(g(x,\theta_0)|x,\alpha_0) |\nabla_\theta g(x,\theta_0)' u| \right] \big)$$

By Assumptions 4 and 6, there exists $a > 0$, so that we can set $G(x,y) = a(\|x\| + \|y\|^2)$ to satisfy the conclusion of the lemma. □

Lemma 2′ *The finite dimensional distributions of $Z_{\alpha'}(h)$ converge to those of $Z_{\alpha_0}(h)$ as $\alpha' \longrightarrow \alpha_0$.*

PROOF: Let $T_{\alpha'}$, $(D_{\alpha',h_1}, \ldots, D_{\alpha',h_m})$, T_{α_0}, and $(D_{\alpha_0,h_1}, \ldots, D_{\alpha_0,h_m})$ be the normal and joint Bernoulli random variables corresponding to $(Z_{n,\alpha'}(h_1), \ldots, Z_{n,\alpha'}(h_m))$ and $(Z_{n,\alpha_0}(h_1), \ldots, Z_{n,\alpha_0}(h_m))$ as given in Corollary 4. The conclusion of the lemma will follow by the continuous mapping theorem and noting that for $\{j_1, \ldots, j_l\} \subset \{h_1, \ldots, h_m\}$, $P_{\alpha'}(D_{\alpha',j_1} = 1, \ldots, D_{\alpha',j_l} = 1) \longrightarrow P_{\alpha_0}(D_{\alpha_0,j_1} = 1, \ldots, D_{\alpha_0,j_l} = 1)$ and $var(T_{\alpha'}) \longrightarrow var(T_{\alpha_0})$ as $\alpha' \longrightarrow \alpha_0$. The variances converge by Assumption 3. The probabilities converge by the continuity of $\mathbf{1}\{\max\{\nabla_\theta g(x,\theta')' u_{j_1}, \ldots, \nabla_\theta g(x,\theta')' u_{j_l}\} > 0\}\ f(g(x,\theta')|x,\alpha')\ \max\{\nabla_\theta g(x,\theta')' u_{j_1}, \ldots, \nabla_\theta g(x,\theta')' u_{j_l}\}])$ with respect to α' at α_0 and Assumption 5. □

The proofs of Lemmas 7′, 8′, and 10′ follow the arguments of the proofs of Lemmas 7, 8, 10 and are omitted here due to space considerations.

Lemma 7′ *For all δ and η small enough,*

$$P_\alpha\left(\int_0^\eta \cdots \int_0^\eta \int_0^\delta \cdots \int_0^\delta Z_\alpha(u,v) du dv < \frac{1}{2}\delta^k \eta^d \right) < 2A^{1/2}(k^{1/4} + d^{1/2})(\delta^{1/2} + \eta)$$

Define

$$I_{\alpha,HJ} = \int_{\Gamma_H^u \times \Gamma_J^v} Z_\alpha(u,v) du dv,\ \ Q_{\alpha,HJ} = \frac{I_{\alpha,HJ}}{\int Z_{n,\alpha}(u,v) du dv}.$$

Lemma 8′ *There exist constants $B, C, b > 0$ such that*

$$P_\alpha(I_{\alpha,HJ} > e^{-bG(H,J)}) \leq B(1 + H^B)\exp(CJ)e^{-bG(H,J)}$$

$$E_\alpha[Q_{\alpha,HJ}] \leq B(1+H^B)\exp(CJ)e^{-bG(H,J)}$$

Lemma 9′ *Suppose* $\varphi_n^{-1}(\tilde{\alpha}_n - \alpha) \rightsquigarrow \tau_\alpha$. *For any* N

$$\lim_{H \longrightarrow \infty} H^N P_\alpha(\|\tau_\alpha\| > H) = 0$$

PROOF: The result follows by the Portmanteau Theorem, Lemma 9, and noting that the H-bounds in the proof of Lemma 9 are uniform in n for large enough n. □

Lemma 10′ *The distributions of the integrals* $\int_{\|u,v\|\leq M} l(s-u,t-v)Z_{\alpha'}(u,v)dudv$ *and* $\int_{\|u,v\|\leq M} Z_{\alpha'}(u,v)dudv$ *converge to the distributions of the integrals* $\int_{\|u,v\|\leq M} l(s-u,t-v)Z_{\alpha_0}(u,v)dudv$ *and* $\int_{\|u,v\|\leq M} Z_{\alpha_0}(u,v)dudv$ *as* $\alpha' \longrightarrow \alpha_0$.

Lemma 11′ $(\psi_{\alpha'}(s_1,t_1),\ldots,\psi_{\alpha'}(s_m,t_m)) \xrightarrow{p} (\psi_{\alpha_0}(s_1,t_1),\ldots,\psi_{\alpha_0}(s_m,t_m))$ *as* $\alpha' \longrightarrow \alpha_0$.

PROOF:

In the proof of Lemma 11, we have already shown that

$$\psi_{\alpha'}(s,t) - \frac{\int_{\|u,v\|\leq M} l(s-u,t-v)Z_{\alpha'}(u,v)dudv}{\int_{\|\tilde{u},\tilde{v}\|\leq L} Z_{\alpha'}(\tilde{u},\tilde{v})d\tilde{u}d\tilde{v}} \xrightarrow{p} 0$$

as $M \longrightarrow \infty$, and similarly for $\psi_{\alpha_0}(s,t)$.

Using the Cramer-Wold device and Lemma 10′, the marginal distributions of

$$\frac{\int_{\|u,v\|\leq M} l(s-u,t-v)Z_{\alpha'}(u,v)dudv}{\int_{\|\tilde{u},\tilde{v}\|\leq L} Z_{\alpha'}(\tilde{u},\tilde{v})d\tilde{u}d\tilde{v}}$$

converge to the marginal distributions of

$$\frac{\int_{\|u,v\|\leq M} l(s-u,t-v)Z_{\alpha_0}(u,v)dudv}{\int_{\|\tilde{u},\tilde{v}\|\leq L} Z_{\alpha_0}(\tilde{u},\tilde{v})d\tilde{u}d\tilde{v}}$$

for $M, L < \infty$. The conclusion of the lemma follows. □

Lemma 12′ *For every $\varepsilon, \eta > 0$, there exists $\delta > 0$ such that:*

$$\lim \sup_{\alpha' \longrightarrow \alpha_0} P_{\alpha'} \left(\sup_{\|s_1 - s_2, t_1 - t_2\| \leq \delta, \|s_1, t_1\| \leq M, \|s_2, t_2\| \leq M} |\psi_{\alpha'}(s_1, t_1) - \psi_{\alpha'}(s_2, t_2)| \geq \varepsilon \right) \leq \eta$$

PROOF: By the same argument as in the proof of Lemma 12 applied to $\psi_{\alpha'}$ rather than ψ_{n,α_0}.

□

Since $\varphi_n^{-1}(\tilde{\alpha}_n - \alpha) \rightsquigarrow \tau_\alpha$, $L(\alpha) = E_\alpha[l(\tau_\alpha)]$. Moreover, for $\alpha' \longrightarrow \alpha_0$, $\tau_{\alpha'} \rightsquigarrow \tau_{\alpha_0}$. Thus for $\alpha_0 \in \mathcal{N}_0$, L is continuous. Also by Assumption 7 and Lemma 9, L is bounded on $\alpha_0 \in \mathcal{N}_0$.

Continuity and boundedness of the risk function can be used to establish that for any nonempty open subset A of $\mathcal{N}_0$ $\liminf_{n \longrightarrow \infty} \sup_{\alpha \in \mathcal{N}_0} E_{n,\alpha}[l(\varphi_n^{-1}(\hat{\alpha}_n - \alpha)] \geq \sup_{\alpha \in A} L(\alpha)$ by Theorem 1.9.1 in Ibragimov and Hasminskii (1981). Recalling the uniform convergence to the limiting risk function shown earlier, asymptotic efficiency of $\tilde{\alpha}_n$ on $\mathcal{N}_0$ follows. Hence for any point α_0 in $\mathcal{N}_0$, $\tilde{\alpha}_n$ is asymptotically efficient. □

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 (2002): "Superconsistent Estimation and Inference in Structural Econometric Models using Extreme Order Statistics," *Journal of Econometrics*, 109, 305–340.

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.

Hirano, K., and J. Porter (2002): "Efficiency in Asymptotic Shift Experiments," manuscript.

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.

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

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. (1991): "An Asymptotic Representation Theorem," *International Statistical Review*, 59, 97–121.

——— (1996): "Limits of Experiments," manuscript.

Van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

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