LIKELIHOOD ESTIMATION AND INFERENCE IN A CLASS OF NONREGULAR ECONOMETRIC MODELS

By Victor Chernozhukov and Han Hong[1]

We study inference in structural models with a jump in the conditional density, where location and size of the jump are described by regression curves. Two prominent examples are auction models, where the bid density jumps from zero to a positive value at the lowest cost, and equilibrium job-search models, where the wage density jumps from one positive level to another at the reservation wage. General inference in such models remained a long-standing, unresolved problem, primarily due to nonregularities and computational difficulties caused by discontinuous likelihood functions.

This paper develops likelihood-based estimation and inference methods for these models, focusing on optimal (Bayes) and maximum likelihood procedures. We derive convergence rates and distribution theory, and develop Bayes and Wald inference. We show that Bayes estimators and confidence intervals are attractive both theoretically and computationally, and that Bayes confidence intervals, based on posterior quantiles, provide a valid large sample inference method.

KEYWORDS: Point process, extreme value theory, Bayes, frequentist validity of posterior, computational complexity, epi-convergence, insufficiency of maximum likelihood.

1. INTRODUCTION

THIS PAPER DEVELOPS estimation and inference methods for structural models with a jump in the conditional density, where the location of the jump is described by a parametric regression curve. The jump in the density is very informative about the parameters of this curve and results in a nonregular and difficult inference theory, implying highly discontinuous likelihoods, nonstandard rates of convergence and inference, and considerable implementation difficulties. Aigner, Amemiya, and Poirier (1976) proposed early models of this type in the context of production efficiency analysis. Many recent econometric models also share this interesting structure. For example, in models of procurement auctions, cf. Donald and Paarsch (1993a), the conditional density of winning bids jumps from zero to a positive value at the lowest cost; in models of equilibrium job search (Bowlus, Neumann, and Kiefer (2001)), the wage den-

sity jumps from one positive level to another at the reservation wage. In what follows, we refer to the former model as the one-sided or boundary model, and to the latter model as the two-sided model. In these models, the location of the density jump is linked to the underlying structural parameters. Learning about the location of the jump is thus crucial for learning about the structural parameters.

Several early fundamental papers have developed inference methods for specific cases of these models, including Aigner, Amemiya, and Poirier (1976), Ibragimov and Has'minskii (1981), Flinn and Heckman (1982), Christensen and Kiefer (1991), Smith (1985, 1994), Donald and Paarsch (1993a, 1993b, 1996, 2002), and Bowlus, Neumann, and Kiefer (2001). Ibragimov and Has'minskii (1981, Chapter V), Smith (1985), and Ghosal and Samanta (1995) obtained the limit theory of the likelihood-based optimal (Bayes) estimators (hereafter BEs) and the maximum likelihood estimator (hereafter MLE) in the nonregression case.[2] Paarsch (1992) and Donald and Paarsch (1993a, 1993b, 1996, 2002) introduced and developed the theory of the MLE and related procedures in the one-sided regression model with discrete regressors, demonstrated the prevalence of these models in structural econometric modeling, and stimulated further research in this area.

Nevertheless, the general inference problem posed by Aigner, Amemiya, and Poirier (1976) remained unsolved previously. To our knowledge, little was known about likelihood-based estimation and inference in the general two-sided regression model. In the general one-sided regression model, the problem of the likelihood-based estimation and inference also remained an important unresolved question, with the important exception of the MLE theory for discrete regressors developed by Donald and Paarsch (1993a).[3,4] A general theory of such regression models was expected to have a substantively different and more involved structure than the corresponding theory for the univariate (nonregression) or dummy regressor case.[5] Moreover, there were considerable implementation problems caused by the inherent computational difficulty of classical (maximum likelihood) estimates.

This paper offers solutions to these open questions by providing a the-

[2]Van der Vaart (1999, Chapters 9.4 and 9.5) provides a lucid introduction to this limit theory, focusing on the univariate uniform and Pareto models, including Pareto models with parameter-dependent support and additional shape parameters.

[3]Also, another important and more recent addition to the literature is the work of Hirano and Porter (2003) discussed below.

[4]There is also a literature on the ad hoc "linear programming" estimators of linear boundary functions, started by Smith (1994); see, e.g., Portnoy and Jurečková (2000), Knight (2002), and Chernozhukov (2000). Asymptotics of these estimators differs from that of MLE, except for some special homoscedastic cases.

[5]See Aigner, Amemiya, and Poirier (1976) and Donald and Paarsch (1996) for pertinent discussions. In fact, we show in this paper that in contrast to the previous cases, the limit likelihood is a function of the multivariate Poisson process with no finite-dimensional sufficient statistics and complex correlation structure.

ory for the likelihood-based estimation and inference methods in both the two- and one-sided models with general regressors. These methods cover likelihood-based optimal[6] (Bayes) and maximum-likelihood procedures. Our results demonstrate that, among these methods, there are computationally and theoretically attractive ways to obtain parameter estimates and carry out statistical inference. The inference results of this paper also enable the construction of confidence intervals using both Bayes and Wald inference.

We show that Bayes confidence intervals based on posterior quantiles are valid in large samples and also perform well in small samples. These confidence intervals are simple to implement in practice since they require no knowledge of the (rather complex) asymptotic theory. Moreover, the proof of the validity of Bayes inference may itself be of independent interest, since it applies generally to cases that have no finite-dimensional asymptotically sufficient statistics. The estimation methods are also attractive due to their familiar finite-sample and hence large sample optimality; see, e.g., Theorem I.9.1 and Corollary I.9.1 in Ibragimov and Has'minskii (1981). They are computationally attractive when carried out through Markov chain Monte Carlo procedures (MCMC); see, e.g., Robert and Casella (1998). In contrast, the computation of the MLE is subject to the curse of dimensionality due to numerous discontinuities in the likelihood function.

To facilitate the results on the likelihood-based estimation and inference, we first develop a complete large sample theory of the likelihood for these models. We show that the limit likelihood is a function of a multivariate Poisson process with a complex correlation structure and no finite-dimensional sufficient statistics. An interesting and important feature of the limit result is that the MLE is generally not an asymptotically sufficient statistic in these models (in contrast to the nonregression case or dummy regression case). This completely rules out the Fisher type "reduction" of all likelihood-based procedures to some transformations of the MLE. That is, the likelihood contains more information asymptotically than the MLE does, and likelihood-based procedures, including optimal estimators, are generally not functions of the MLE even asymptotically, as they are in the nonregression or dummy regression case (or regular models). This property, which may be called incompleteness, motivates the study of the entire likelihood and a wide class of likelihood-based procedures.

Although the asymptotic insufficiency of the MLE does not allow one to single out the MLE as the estimator on which optimal procedures can be based generally, it does *not* imply that the MLE is an inferior estimation procedure either. The MLE can be approximated by Bayes estimators under loss functions that approximate the delta function, e.g., $0 - 1$ loss $1[|u| > \epsilon]/\epsilon$. Hence

---

[6]The optimality of Bayes estimators is their general property under very weak conditions both in finite and large samples; cf. Theorem 9.1 and Corollary 9.1 in Ibragimov and Has'minskii, (1981, p. 93).

the MLE is approximately optimal under loss functions that approximate the delta function, and may perform better under these alternative loss functions than other Bayesian estimators (BEs) such as posterior means or posterior medians. Thus, the MLE generally cannot be dominated by any other *given* BE when risk comparisons are made across different loss functions. Moreover, the MLE is invariant to reparameterization, which makes it a valuable estimation approach, given its approximate optimality.

The asymptotic insufficiency of the MLE and the absence of other finite-dimensional sufficient statistics have important consequences for inference. They imply, in particular, that the large sample validity of Bayes inference has to be established without relying on previous works on Bayes inference in non-regular (nonregression) models, e.g., Ghosal (1999). The previous arguments cannot be immediately extended because they deal with simple boundary models in which MLEs (such as sample minima or maxima) are asymptotically sufficient. In these models, the leading term in the expansion of the posterior is an exponential distribution *shifted* by the MLE, i.e., it has an exponential shift form. This distribution coincides with that of the MLE, allowing a simple analytical demonstration of the equivalence of the Wald inference based on the MLE and the posterior inference, thus establishing the validity of the posterior for frequentist inference. In sharp contrast, in our case the limit posterior is a function of a multivariate Poisson process, generally has no finite-dimensional sufficient statistics, and neither coincides with the distribution of the MLE nor is shifted by the MLE. All of these reasons preclude the applicability of the previous argument and also imply that the Wald inference and the posterior inference are not asymptotically equivalent. Instead, our proof is based on the asymptotic average risk optimality of posterior quantiles under the check loss function, which allows us to establish the quantile unbiasedness of posterior quantiles and then prove their inference validity. This argument is generic and thus may be of independent interest.

Our work is also related to a recent paper of Hirano and Porter (2003), who present a thorough, important analysis of the efficiency problem in boundary regression models. Hirano and Porter (2003) provide an insightful construction of an exponential-shift experiment framework and apply a minimax analysis within this framework, which explains and characterizes the efficiency gains of BEs over the MLE under the common mean squared and absolute deviation criteria. In contrast, we consider likelihood-based estimation and inference in a general class of models that covers both two-sided and one-sided regression models. We derive the limit likelihood, the large sample properties and limit distributions of likelihood-based estimators, all of them given explicitly (and conveniently) in terms of a Poisson process; develop Bayes and Wald inference; and investigate the finite-sample properties of these estimation and inference procedures.[7]

---

[7]Hirano and Porter (2003) also discuss the limit likelihood for boundary models as an auxiliary result, extending our previous results on boundary models without nuisance parameters;

The remainder of the paper is organized as follows. Section 2 provides a nontechnical, practical discussion of the inference and estimation methods. Section 3 contains a theoretical discussion of large sample theory. Section 4 illustrates the estimation and inference methods developed in the paper in a simulation study based on an auction model described in Donald and Paarsch (2002).

2. THE SETUP AND INFORMAL OVERVIEW OF PROCEDURES AND RESULTS

This section describes the model and provides a nontechnical discussion of the assumptions, results, and inference procedures developed in the paper.

2.1. *The Model*

It is convenient to describe the framework we consider in terms of a regression model, where the disturbance has a discontinuous density. Let (Y_i, X_i), $i = 1, \ldots, n$, denote an i.i.d. sample of size n generated by the model

$$(2.1) \qquad Y_i = g(X_i, \beta) + \epsilon_i,$$

where Y_i is the dependent variable, X_i is a vector of covariates that has distribution function F_X, and the disturbance ϵ_i has conditional density $f(\epsilon | X_i, \beta, \alpha)$. The central assumption of the model is that the conditional density $f(\epsilon | X_i, \beta, \alpha)$ has a jump (or discontinuity) normalized to be at 0, which may depend on the parameters β and α:

$$(2.2) \qquad \lim_{\epsilon \uparrow 0} f(\epsilon | x, \beta, \alpha) = q(x, \beta, \alpha), \quad \lim_{\epsilon \downarrow 0} f(\epsilon | x, \beta, \alpha) = p(x, \beta, \alpha),$$

$$p(x, \beta, \alpha) > q(x, \beta, \alpha) + \eta,$$

for some $\eta > 0$, $\forall x \in \mathbf{X} = \text{support}(X)$,

for all $(\beta, \alpha) \in \mathcal{B} \times \mathcal{A}$, a compact subset of $\mathbb{R}^{d_\beta} \times \mathbb{R}^{d_\alpha}$.

Hence, in this model the location of the discontinuity in the density of Y conditional on X is given by the regression function $g(X, \beta)$, which is described by the parameter β. Thus, there are two sets of parameters, collected into a vector $\gamma = (\beta', \alpha')'$, where β affects the regression curve and possibly the disturbance distribution, and α affects the shape of the disturbance distribution only.

---

cf. Chernozhukov and Hong (2001b).

We consider two regression models: the two-sided model, where the conditional density jumps from one positive value to another, and the one-sided model, where the conditional density jumps from zero to a positive constant. The one-sided model is a *limit* case of the two-sided model. In addition, Aigner, Amemiya, and Poirier (1976) suggested that the two-sided model may be applied to one-sided models in the presence of outliers, using an additional side to model the outliers.[8] More generally, the two-sided model approximates models with a sharp change in the density, where the location of the change depends on parameters and regressors. The finite sample distribution of the parameter estimates in such models is approximated by that in the model with a density jump. The two-sided model also naturally arises in equilibrium search models; see, e.g., Bowlus, Neumann, and Kiefer (2001).

The key feature of the regression model (2.1)–(2.2) is that the conditional density of Y given X jumps at the location $g(X, \beta)$, which depends on the parameter β and covariates X. This feature generates sharp discontinuities in the likelihood, which create statistical nonregularities and computational difficulties. The discontinuities are highly informative about β and imply estimability at rate n. (The simplest univariate example is the uniform model $U(0, \beta)$, where β can be estimated at rate n.) On the other hand, inference about α is standard in many ways.

Note that the classification of parameters into α and β is motivated statistically, as in Donald and Paarsch (1993a) and van der Vaart (1999, Chapters 9.3 and 9.4). Parameters of structural economic models usually coincide with the boundary parameters β, as indicated earlier. If they do not, and Wald inference is to be used, then one needs to reparameterize them into α and β, as indicated in Donald and Paarsch (1993a). In the following, we briefly review a structural auction model that illustrates the plausibility of our regularity conditions and provides an example for the Monte Carlo work.

EXAMPLE (An independent private value procurement auction): Consider an econometric model of an independent private value procurement auction, formulated in Paarsch (1992) and Donald and Paarsch (2002). In this model, Y_i is the winning bid for the auction i; the covariates $X_i = (Z_i, m_i)$ describe variation across auctions; m_i denotes the number of bidders in the ith auction minus 1; and Z_i denotes other observed characteristics of the auctions.

The bidders' privately observed costs V follow an i.i.d. Pareto distribution given X; i.e., the density of V given X is described by

$$f_V(v|X) = \frac{\theta_2 \theta_1^{\theta_2}}{v^{\theta_2+1}}, \quad v \geq \theta_1 > 0, \ \theta_2 > 0.$$

---

[8]Another approach that can deal with outliers is developed in Chernozhukov (2000) and is based on near-extreme quantile regression.

The parameters θ_2 and θ_1 are functions of X and β, e.g., $\theta_1(X, \beta) = \exp(\beta_1' Z)$ and $\theta_2(X, \beta) = \exp(\beta_2' Z)$, but this dependence will be suppressed for convenience.

Assuming a Bayes–Nash Equilibrium solution concept, the equilibrium bidding function satisfies

$$\sigma(v) = v + \frac{\int_v^\infty (1 - F_V(\xi|X))^m \, d\xi}{(1 - F_V(v|X))^m},$$

which is the cost plus the expected net revenue conditional on winning the auction. Evaluating $\sigma(v)$ at $v = \theta_1$ gives the conditional support for the winning bid. As shown in Paarsch (1992), this implies the following conditional density function of the winning bid Y, which is the first order statistic generated by the specified bidding rule:

$$f_Y(y|X) = \frac{\theta_2 m \left[\frac{\theta_1 \theta_2 (m-1)}{[\theta_2(m-1)-1]} \right]^{\theta_2 m}}{y^{\theta_2 m + 1}} \cdot \mathbf{1}\left(y \geq \left[\frac{\theta_1 \theta_2 (m-1)}{\theta_2 (m-1) - 1} \right] \right).$$

Therefore, this is an example of a one-sided regression model (2.1), where $Y_i = g(X_i, \beta) + \epsilon_i$, $g(X, \beta) = \theta_1(X, \beta) \cdot \theta_2(X, \beta) \cdot (m-1)/(\theta_2(X, \beta)(m-1) - 1)$, and ϵ_i has density $f(\epsilon|X, \beta) = f_Y(g(X, \beta) + \epsilon|X, \beta)$ conditional on X.

The main regularity conditions C0–C5 are collected in Appendix A. They generalize the conditions of Ibragimov and Has'minskii (1981) and Smith (1985). These conditions are flexible enough to cover various auction models, including the one stated above, frontier production function models, and equilibrium search models.[9]

2.2. *Definitions of Estimation Procedures and Informal Overview of Results*

Define the likelihood function as[10]

$$(2.3) \qquad L_n(\gamma) \equiv \prod_{i \leq n} f(Y_i - g(X_i, \beta)|X_i; \gamma).$$

Bayes estimators (BEs) are likelihood-based optimal estimators that minimize the average expected risk, where the risk is computed under different

---

[9]See Chernozhukov and Hong (2001a) for an example of verification of these conditions in the auction model that underlies our Monte Carlo simulations.

[10]The likelihood can be made unconditional by multiplying through with the density (probability mass) function of $\{X_i, i \leq n\}$. This additional term is omitted because it cancels out in the definition of the likelihood ratio.

parameter values and then averaged over these parameter values. The estimators are generally of the following form:

$$(2.4) \qquad \widehat{\gamma} \equiv \arg\inf_{\gamma \in \mathcal{G}} \int_{\mathcal{G}} \rho_n(\gamma - \bar{\gamma}) \frac{L_n(\bar{\gamma})\mu(\bar{\gamma})}{\int_{\mathcal{G}} L_n(\gamma')\mu(\gamma') \, d\gamma'} \, d\bar{\gamma},$$

where $\rho_n(\gamma) \equiv \rho(n\beta, \sqrt{n}\alpha)$ is a loss function, $\mu(\cdot)$ is the weight density (prior density) on $\mathcal{G}$, and $L_n(\gamma)\mu(\gamma)/\int_{\mathcal{G}} L_n(\gamma')\mu(\gamma') \, d\gamma'$ is the posterior density. The optimality properties of Bayes procedures carry over to the limit; see, e.g., Theorem I.9.1 in Ibragimov and Has'minskii (1981).

The loss function ρ_n is made explicitly dependent on the sample size for purposes of asymptotic analysis, as in Ibragimov and Has'minskii (1981), but this may be ignored in practice. Convexity and standard conditions are imposed on the loss function ρ and the prior μ, and collected as D1–D3 in Appendix A. Examples of such loss functions include: (A) $\rho(z) = z'z$, a quadratic loss function, (B) $\rho(z) = \sum_{j=1}^{d} |z_j|$, an absolute deviation loss function, and (C) $\rho(z; \tau) = \sum_{j=1}^{d}(1(z_j > 0) - \tau)z_j$, $\tau \in (0, 1)$, a variant of the Koenker and Bassett (1978) check loss function. Solutions of (2.4) with loss functions (A), (B), and (C) generate BEs $\widehat{\gamma}$ that are, respectively, (A) a vector of posterior means, (B) a vector of posterior medians (for each parameter component), and (C) a vector of posterior τth quantiles.[11] Since BEs become very difficult to compute when ρ is not convex, we focus on convex loss functions for pragmatic reasons. However, the proofs of our main results apply more generally to other loss functions specified in Ibragimov and Has'minskii (1981).

In practice, $\widehat{\gamma}$ can be computed using the Metropolis–Hastings algorithm and related Markov chain Monte Carlo (MCMC) methods, which produce a sequence of draws $(\gamma^{(1)}, \ldots, \gamma^{(b)})$, whose marginal distribution is given by the posterior. Appropriate algorithms and implementation details are given, for example, in Robert and Casella (1998, p. 245). The Metropolis–Hastings algorithm is valid under general conditions that do not depend on the likelihood having discontinuities; see, e.g., Tierney (1994) and Corollary 6.2.6 in Robert and Casella (1998). Given an MCMC sequence $(\gamma^{(1)}, \ldots, \gamma^{(b)})$, the estimators $\hat{\gamma}$ can be taken as the appropriate statistics of the sequence depending on ρ, for example, the component-wise means, medians, and quantiles of $(\gamma^{(1)}, \ldots, \gamma^{(b)})$ in the cases (A), (B), and (C), respectively. More generally, the estimators $\widehat{\gamma}$ are solutions of well defined globally convex optimization problems. Indeed, in practice the estimators $\widehat{\gamma}$ solve $\arg\min_{\gamma \in \mathcal{G}}(1/b)\sum_{t=1}^{b} \rho_n(\gamma - \gamma^{(t)})$, which is a globally convex optimization problem.

---

[11]For example, a vector of posterior means is given by

$$\int_{\mathcal{G}} \bar{\gamma} \frac{L_n(\bar{\gamma})\mu(\bar{\gamma})}{\int_{\mathcal{G}} L_n(\gamma')\mu(\gamma') \, d\gamma'} \, d\bar{\gamma}.$$

Other estimators do not admit such closed form expressions.

The computational attractiveness of estimation and inference based on Bayes procedures stems from the use of MCMC and the statistical motivation entering the definition of Bayes procedures. Since Bayes estimators and confidence intervals are typically means, medians, or quantiles of the posterior distribution, by drawing the MCMC sample of size b from the posterior distribution, we can compute these quantities with an accuracy of order $1/\sqrt{b}$. In contrast, computation of the MLE requires optimization of a highly nonconvex and discontinuous likelihood function. The exact MLE can be estimated by grid-based algorithms or MCMC only with an accuracy that worsens exponentially in the parameter dimension.

BEs and the MLE are consistent, and it is shown in this paper that

$$(2.5) \qquad \widehat{\beta} - \beta = O_p(n^{-1}) \quad \text{and} \quad \widehat{\alpha} - \alpha = O_p(n^{-1/2}).$$

BEs are shown to converge in distribution to Pitman functionals of the limit likelihood ratio process. MLEs are shown to converge in distribution to arg max functionals of the limit likelihood ratio process. We first develop a complete large sample theory of the likelihood for these models, which is a prerequisite for any inference based on the likelihood principle. In particular, we obtain an explicit form of the limit likelihood ratio process as a function of a Poisson process that can be easily simulated.

This result implies that the limit distributions of the estimators can be simulated for purposes of Wald inference through either (a) simulating the limit likelihood process, or (b) subsampling and other resampling methods such as parametric bootstrap. Implementation protocols for subsampling are standard and can be found in Politis, Romano, and Wolf (1999). Subsampling is more robust than other methods under local misspecification of the parametric assumptions because it only relies on the existence, and not on a specific form, of the limit distribution. However, subsampling and other resampling methods are more computationally expensive than Bayes inference. Simulating the limit distribution is comparable in terms of computational expense to Bayes inference due to linearity of the limit process.

An attractive practical alternative is Bayes inference based on posterior quantiles. Our results establish its large sample (frequentist) validity. Following van der Vaart (2000), consider constructing a $\tau \times 100\%$ confidence interval for $r_n(\gamma)$, where r_n is a smooth real function that possibly depends on n. Define the τth posterior quantile of the posterior distribution as

$$(2.6) \qquad \widehat{c}(\tau) \equiv \underset{\tilde{r} \in \mathcal{R}_n}{\arg\inf} \int_{\mathcal{G}} \rho\big(\tilde{r} - r_n(\gamma); \tau\big) \frac{L_n(\gamma)\mu(\gamma)}{\int_{\mathcal{G}} L_n(\gamma')\mu(\gamma')\,d\gamma'}\,d\gamma,$$

where $\rho(z; \tau)$ is the check function defined above, and $\mathcal{R}_n \equiv \{r_n(\gamma), \gamma \in \mathcal{G}\}$. In practice, $\widehat{c}(\tau)$ is computed simply by taking the τth quantile of the MCMC sequence evaluated at r_n,

$$(2.7) \qquad \big(r_n(\gamma^{(1)}), \ldots, r_n(\gamma^{(b)})\big).$$

The resulting $\tau \times 100\%$-confidence intervals are given by

(2.8) $[\widehat{c}(\tau/2), \widehat{c}(1 - \tau/2)]$,   where

$$\lim_{n \to \infty} P_\gamma \{\widehat{c}(\tau/2) \le r_n(\gamma) \le \widehat{c}(1 - \tau/2)\} = 1 - \tau,$$

under mild conditions on r_n, which is one of the main results of this paper.

A pragmatic motivation for Bayesian intervals is that, in order to apply them, the empirical researcher does not need to know the estimation rate of the γ components, the form of the limit distributions, or other details of the complex asymptotic limit theory. One can simply compute the intervals through generic MCMC methods, and then rely upon the present results that establish the large sample (frequentist) validity of these intervals.

Another classical procedure is the MLE, which is defined by maximizing the likelihood function: $\widehat{\gamma} = (\widehat{\beta}', \widehat{\alpha}')' \equiv \arg\sup_{\gamma \in \mathcal{G}} L_n(\gamma)$. The MLE is the limit of BEs under any sequence of loss functions that approximates the delta function. The MLE converges in distribution to a random variable that maximizes the limit likelihood ratio. We will only briefly discuss the limit distribution of the MLE due to page constraints. A detailed analysis of the MLE is given in Chernozhukov and Hong (2001a).

3. LARGE SAMPLE THEORY

This section contains the formal results of the paper. It begins with the analysis of the large sample properties of the likelihood ratio function, followed by the analysis of Bayes estimators and derived inference procedures, and concludes with a brief discussion of the limit theory for maximum likelihood estimation.

3.1. *Large Sample Theory for the Likelihood*

A common first step in modern asymptotic analysis is to find the finite-dimensional marginal limit of the likelihood ratio process or other criterion functions; see, e.g., van der Vaart (1999) and Knight (2000). After appropriate strengthening, this limit serves to describe the asymptotic distribution of all likelihood based estimators. In the likelihood analysis, such an initial step is also called the convergence of experiments.

Consider the local likelihood ratio function $\ell_n(z) \equiv L_n(\gamma_n + H_n z)/L_n(\gamma_n)$, where $\gamma_n \equiv \gamma_0 + H_n \delta$ denotes the true local parameter sequence, $\delta \in \mathbb{R}^d$, and H_n is a diagonal matrix with $1/n$ in the first $d_\beta = \dim(\beta)$ diagonal entries and $1/\sqrt{n}$ in the remaining $d_\alpha = \dim(\alpha)$ diagonal entries. Consideration of the local parameter sequence is necessary for subsequent analysis and for showing

that the limit likelihood does not depend on δ. The scaling by H_n corresponds to convergence rates $\sqrt{n}$ for α and n for β.[12]

The function $\ell_n(z)$ is said to converge in distribution to $\ell_\infty(z)$ in the finite-dimensional sense, and $\ell_\infty(z)$ is called the finite-dimensional limit, if, for any finite k,

$$(3.1) \qquad (\ell_n(z_j), j \le k) \overset{d}{\to} (\ell_\infty(z_j), j \le k),$$

where $\overset{d}{\to}$ denotes convergence in distribution under P_{γ_n}. We partition the localized parameter z accordingly into $z = (u', v')'$, where $u \in \mathbb{R}^{d_\beta}$ corresponds to the localized β parameters and $v \in \mathbb{R}^{d_\alpha}$ corresponds to the localized α parameters.

THEOREM 3.1 (Limits of the likelihood function): *Given Conditions* C0–C5 *collected in Appendix* A, *the finite-dimensional weak limit of the likelihood ratio process* $\ell_n(z)$ *takes the following form*: *For* $\Delta(x) \equiv \partial g(x, \beta_0)/\partial\beta$, $p(X) \equiv p(X, \gamma_0)$, $q(X) \equiv q(X, \gamma_0)$, *and* $l_i(\gamma) = \ln f(\epsilon_i | X_i, \gamma)$,

$$(3.2) \qquad \begin{aligned} &\ell_\infty(z) \equiv \ell_{1\infty}(v) \times \ell_{2\infty}(u), \quad \ell_{1\infty}(v) \equiv \exp(\mathbf{W}'v - v'\mathcal{J}v/2), \\ &\ell_{2\infty}(u) \equiv \exp\left(u'\mathbf{m} + \int_{\mathbb{R} \times \mathbf{X}} l_u(j, x) \, d\mathbf{N}(j, x) \right), \end{aligned}$$

where

$$\mathcal{J} \equiv E_{\gamma_0}\left(\frac{\partial}{\partial\alpha} l_i(\gamma_0) \frac{\partial}{\partial\alpha} l_i(\gamma_0)' \right), \quad \mathbf{m} \equiv E_{\gamma_0} \Delta(X)[p(X) - q(X)],$$

$$\mathbf{W} \overset{d}{=} \mathcal{N}(0, \mathcal{J}),$$

and

$$l_u(j, x) \equiv \ln \frac{q(x)}{p(x)} \mathbf{1}[0 < j < \Delta(x)'u] + \ln \frac{p(x)}{q(x)} \mathbf{1}[0 > j > \Delta(x)'u],$$

where in the one-sided case, i.e., when $q(x) = 0$, *we use the convention*: $\ln 0 = -\infty, \ln\infty = \infty, 1/0 = \infty$, *and* $\infty \cdot 0 = 0$. $\mathbf{N}$ *is a Poisson random measure* $\mathbf{N}(\cdot) \equiv \sum_{i=1}^\infty \mathbf{1}[(J_i, \mathcal{X}_i) \in \cdot] + \sum_{i=1}^\infty \mathbf{1}[(J_i', \mathcal{X}_i') \in \cdot]$, *where*

$$(3.3) \qquad J_i \equiv \Gamma_i/p(\mathcal{X}_i), \qquad \Gamma_i \equiv \mathcal{E}_1 + \cdots + \mathcal{E}_i, \quad i \ge 1,$$

$$(3.4) \qquad J_i' \equiv \Gamma_i'/q(\mathcal{X}_i'), \qquad \Gamma_i' \equiv -(\mathcal{E}_1' + \cdots + \mathcal{E}_i'), \quad i \ge 1.$$

$\{\mathcal{X}_i, \mathcal{E}_i, i \ge 1\}$ *is an i.i.d. sequence of variables where* $\mathcal{X}_i$ *follows the law* F_X, *and* $\mathcal{E}_i$ *is a unit exponential variable. The sequence* $\{\mathcal{E}_i, i \ge 1\}$ *is independent of the sequence* $\{\mathcal{X}_i, i \ge 1\}$. $\{\mathcal{X}_i', \mathcal{E}_i', i \ge 1\}$ *is an independent copy of* $\{\mathcal{X}_i, \mathcal{E}_i, i \ge 1\}$, *and*

---

[12]Convergence rates are established as parts of the proof of subsequent theorems, and follow from the exponential decay of the likelihood tail $E\ell_n^{1/2}(z) \sim \text{const} \cdot e^{-c|z|}$ as $|z| \to \infty$; see the proof of Theorem 3.2.

both sequences are independent of **W**.

This theorem extends the results of Ibragimov and Has'minskii (1981) on the likelihood limits in univariate one-sided and two-sided models. The main term in the limit likelihood is $\ell_{2\infty}(u)$, which is nonstandard. The discontinuities in the density are highly informative about β, and most of the information about β is contained in those Y_i's that are near the location of the discontinuity $g(X_i, \beta)$, that is, in those Y_i for which $\epsilon_i = Y_i - g(X_i, \beta)$ is close to zero. Thus, the behavior of extreme (closest to zero) ϵ_i's determines the behavior of $\ell_{2\infty}(u)$, as further explained in Section 3.2. Consequently, one expects that the rate of convergence of likelihood-based estimators will be n for β (in contrast to $\sqrt{n}$ for α), and their behavior will be determined by $\ell_{2\infty}(u)$.

In contrast to $\ell_{2\infty}(v)$, the term $\ell_{1\infty}(v)$ is a standard expression for the limit likelihood ratio in regular models, and inference about the shape parameter α is thus asymptotically regular. The limit log-likelihood has a standard linear-quadratic expression: $v'\mathbf{W} - v'\mathcal{J}v/2$. This limit contains a normal vector $\mathbf{W} \stackrel{d}{=} \mathcal{N}(0, \mathcal{J})$ and the information matrix $\mathcal{J}$. This implies, for example, that conventional estimators of α, such as the posterior mean and the MLE, have the standard limit distribution $\mathcal{J}^{-1}\mathbf{W} \stackrel{d}{=} \mathcal{N}(0, \mathcal{J}^{-1})$.

REMARK 3.1 (An alternative form): To analyze the limit $\ell_{2\infty}(u)$ further, write the Lebesgue integral $\int_{\mathbb{R}\times\mathbf{X}} l_u(j, x)\, d\mathbf{N}(j, x)$ appearing in the statement of Theorem 3.1 as

$$(3.5) \qquad \sum_{i=1}^{\infty} l_u(J_i, \mathcal{X}_i) + \sum_{i=1}^{\infty} l_u(J'_i, \mathcal{X}'_i) \equiv \sum_{i=1}^{\infty} \ln \frac{q(\mathcal{X}_i)}{p(\mathcal{X}_i)} \mathbf{1}[0 < J_i < \Delta(\mathcal{X}_i)'u]$$

$$+ \sum_{i=1}^{\infty} \ln \frac{p(\mathcal{X}'_i)}{q(\mathcal{X}'_i)} \mathbf{1}[0 > J'_i > \Delta(\mathcal{X}'_i)'u],$$

which is a simple function of the variables $\{\mathcal{X}_i, \mathcal{X}'_i, J_i, J'_i\}$. This suggests that the limit likelihood function can be simulated simply by generating sequences of $\{\mathcal{X}_i, \mathcal{X}'_i, J_i, J'_i, i \leq b\}$ for some large b, say $b \propto n$, according to the distributions specified in Theorem 3.1, and then evaluating the corresponding expressions. In practice, the quantities $p(\mathcal{X}_i)$ and $q(\mathcal{X}_i)$ are replaced by their estimates, and F_X is replaced by the empirical distribution function.

REMARK 3.2 (The boundary or one-sided case): There is a drastic simplification of $\ell_{2\infty}(u)$ in the one-sided (boundary) model. Since $q(\mathcal{X}) = 0$ a.s., using the rules stated in Theorem 3.1,

$$(3.6) \qquad \sum_{i=1}^{\infty} l_u(J_i, \mathcal{X}_i) + \underbrace{\sum_{i=1}^{\infty} l_u(J'_i, \mathcal{X}'_i)}_{\equiv 0} \equiv \begin{cases} 0 & \text{if } J_i \geq \Delta(\mathcal{X}_i)'u, \text{ for all } i \geq 1, \\ -\infty & \text{otherwise.} \end{cases}$$

Hence for $\mathbf{m} = E\Delta(X)p(X)$,

$$(3.7) \qquad \ell_{2\infty}(u) \equiv \begin{cases} \exp(u'\mathbf{m}) & \text{if } J_i \geq \Delta(\mathcal{X}_i)'u, \text{ for all } i \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, in one-sided models, the limit depends only on the set of variables in (3.3) and does not depend on the set of variables in (3.4). Another way to represent the finite-dimensional distribution of $\ell_{2\infty}(u)$ is to observe that for each $(u_1, \ldots, u_k)$, $(\ell_{2\infty}(u_1), \ldots, \ell_{2\infty}(u_k))$ is $\exp(u'\mathbf{m})$ times a vector of correlated Bernoulli variables with success probabilities that depend on u; see Hirano and Porter (2003) who obtain this alternative form and use it for efficiency analysis.

REMARK 3.3 (Misspecification): It can be conjectured from the proof of Theorem 3.1 that the limit theory for β is robust under local $o(1/n)$ misspecification of the regression function $g(x, \beta)$, and local $o(1/\sqrt{n})$ misspecification of the heights of the densities, $p(x, \gamma)$ and $q(x, \gamma)$, at the jump points. It also appears that the qualitative nature of the limit theory would be preserved under local $O(1/n)$ misspecification of the regression function $g(x, \beta)$ and possibly under $O(1)$ misspecification of $p(x, \gamma)$ and $q(x, \gamma)$ as long as $p(x, \gamma) > q(x, \gamma)$ for all x. A formal development of these results is beyond the scope of this paper.

3.2. *An Example for Intuition*

It is useful to pause here and highlight the intuition behind Theorem 3.1 by considering the following simple example. Suppose

$$(3.8) \qquad Y_i = X_i'\beta_0 + \epsilon_i, \quad \epsilon_i \overset{d}{=} \mathcal{E},$$

where $\mathcal{E}$ is a standard unit exponential variable. This is a boundary model with the density at the boundary equal to $p(X) = 1$. Assume that there are no shape parameters α (we do not discuss inference about α since it is regular as stated earlier). This model is a linearized, homoscedastic version of more realistic nonlinear models.

Intuitively, the smallest values of ϵ_i will be the most informative about β, as the likelihood function will be positive only if $Y_i - X_i'\beta \geq 0$, for all i, that is, when $n\epsilon_i \geq X_i'n(\beta - \beta_0)$, for all i. Letting $u = n(\beta - \beta_0)$, this constraint takes the form $n\epsilon_i \geq X_i'u$, for all i. What we can learn about the parameter β_0 will depend on these constraints.

The likelihood for this example is $L_n(\beta) = \prod_{i \leq n} e^{-\epsilon_i + X_i'(\beta - \beta_0)}\mathbf{1}(n\epsilon_i \geq X_i'n(\beta - \beta_0))$. Hence the likelihood ratio $L_n(\beta)/L_n(\beta_0)$ as a function of $u = n(\beta - \beta_0)$ takes the form $\ell_n(u) = \prod_{i \leq n}(e^{-\epsilon_i + X_i'u/n}/e^{-\epsilon_i})\mathbf{1}(n\epsilon_i \geq X_i'u)$, which further reduces to

$$(3.9) \qquad \ell_n(u) = e^{\bar{X}'u} \cdot \mathbf{1}(n\epsilon_i \geq X_i'u, \text{ for all } i).$$

Since $\bar{X} \to_p EX$, the behavior of $\ell_n(u)$ for fixed u is determined by the lowest order statistics $n\epsilon_{(1)}, n\epsilon_{(2)}, n\epsilon_{(3)}, \ldots$. The Reny representation allows these rescaled order statistics to be represented almost surely as

$$\mathcal{E}_1, \mathcal{E}_1 + \frac{n}{n-1}\mathcal{E}_2, \mathcal{E}_1 + \frac{n}{n-1}\mathcal{E}_2 + \frac{n}{n-2}\mathcal{E}_3, \ldots,$$

where $\{\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n\}$ is an i.i.d. sequence of unit-exponential variables; see, e.g., Embrechts, Klüppelberg, and Mikosch (1997, p. 189). For a given u, essentially only a stochastically bounded number of order statistics, say k, matters in the constraints (3.9). Hence as $n \to \infty$, for any finite k: $(n\epsilon_{(1)}, n\epsilon_{(2)}, \ldots, n\epsilon_{(k)}) \overset{d}{\to} (\mathcal{E}_1, \mathcal{E}_1 + \mathcal{E}_2, \ldots, \sum_{j=1}^k \mathcal{E}_j) \equiv (\Gamma_1, \Gamma_2, \ldots, \Gamma_k)$. Hence the marginal limit of $\ell_n(u)$ may be seen as

$$\ell_\infty(u) \equiv e^{E(X)'u} \cdot \mathbf{1}(\Gamma_i \geq \mathcal{X}_i'u, \text{ for all } i \geq 1),$$

where $\{\Gamma_i\}$ is the sequence of gamma variables defined above, and $\{\mathcal{X}_i\}$ is an i.i.d. sequence of regressors with distribution F_X, which is independent of the $\{\Gamma_i\}$ sequence. Note that this is just a special case of the limit stated in (3.3), where $p(X) = 1$. (Also there are no α parameters in this example, so that $\ell_\infty(u) = \ell_{2\infty}(u)$.) The use of point process methods in Theorem 3.1 formalizes the intuition described above and extends it to more general heteroscedastic disturbances.

The result stated in Theorem 3.1 is more complicated due to the following reasons: First, in more general two-sided models, there is also an additional negative disturbance in equations like (3.8). The information about β is then largely deduced from the ϵ_i's closest to 0 from above and the ϵ_i's closest to zero from below. This explains the presence of the additional set of gamma variables and associated regressors in equation (3.4) as the limit distributions of "extremes from below." Second, the density of the ϵ_i's may vary near zero, which changes the hazard rates of the limit gamma variables Γ_i and Γ_i', resulting in their division by varying hazard rates $p(\mathcal{X}_i)$ or $q(\mathcal{X}_i')$. Third, the uncertainty about the additional shape parameter α leads to the presence of an additional term $\ell_{1\infty}(v)$. The form of this term reflects that the inference about α is standard. The limit information about α is given by the limit average score $\mathbf{W}$ and the information matrix $\mathcal{J}$. Since information about β comes from a small portion of the entire sample and is based on extreme type statistics, the average score $\mathbf{W}$ is independent of those statistics asymptotically, which follows by a standard argument, e.g., Lemma 21.19 in van der Vaart (1999) and Section 4.3 in Resnick (1986).

3.3. *Basic Large Sample Properties of Bayes Estimation*

Given the above discussion, the following Theorem 3.2 can be easily conjectured. The Bayes estimator

$$Z_n = \left(n(\widehat{\beta} - \beta_n)', \sqrt{n}(\widehat{\alpha} - \alpha_n)'\right)',$$

centered at the true parameter $\gamma_n = (\beta_n', \alpha_n')'$ and normalized by the convergence rates, is related to the localized likelihood ratio $\ell_n(z)$. Z_n minimizes the posterior loss redefined in terms of the local deviation from the true parameter: $\Gamma_n(z) = \int_{\mathbb{R}^d} \rho(z - z')\pi_n(z')\,dz'$, where $\pi_n(z)$ is the posterior density for the local deviation z from the true parameter: $\pi_n(z) = \ell_n(z)\mu(\gamma_n + H_n z)/\int_{\mathbb{R}^d} \ell_n(\bar{z})\mu(\gamma_n + H_n \bar{z})\,d\bar{z}$; $\ell_n(z)$ is the local likelihood ratio process; and μ is the prior density. As $n \to \infty$, it can be conjectured that the posterior $\pi_n(z)$ approaches $\pi_\infty(z) \equiv \ell_\infty(z)/\int_{\mathbb{R}^d} \ell_\infty(z)\,dz$, and that the following other results take place.

THEOREM 3.2 (Basic properties of BEs): *Suppose that* C0–C5 *and* D1–D3 *stated in Appendix* A *hold. Then*:
  (i) *The convergence rate is n for estimating β and $\sqrt{n}$ for estimating α, i.e.,* $Z_n = O_p(1)$.
  (ii) $Z_n \xrightarrow{d} Z$, *where*

$$Z \equiv \underset{z \in \mathbb{R}^d}{\arg\inf} \int_{\mathbb{R}^d} \rho(z - z')\frac{\ell_\infty(z')}{\int_{\mathbb{R}^d} \ell_\infty(\bar{z})\,d\bar{z}}\,dz'.$$

  (iii) *If $\rho(z) = \rho_\beta(u) + \rho_\alpha(v)$, then*

$$n(\widehat{\beta} - \beta_n) \xrightarrow{d} Z^\beta \equiv \underset{u}{\arg\inf} \int_{\mathbb{R}^{d_\beta}} \rho_\beta(u - u')\ell_{2\infty}(u')\,du',$$

$$\sqrt{n}(\widehat{\alpha} - \alpha_n) \xrightarrow{d} Z^\alpha \equiv \underset{v}{\arg\inf} \int_{\mathbb{R}^{d_\alpha}} \rho_\alpha(v - v')\ell_{1\infty}(v')\,dv',$$

and Z^β and Z^α are independent.

Theorem 3.2 establishes consistency, rates of convergence, and limit distributions of BEs. Theorem 3.2 also enables one to perform Wald inference by simulating the limit distribution of BEs according to Remark 3.1 or by estimating it with subsampling according to Remark 3.5. In either case it is desirable to use MCMC methods for recomputing the BEs. A useful alternative to Wald inference is Bayes inference, as discussed below.

REMARK 3.4: In the stated result, Z^β and Z^α are independent due to the factorization of $\ell_\infty(z)$ into independent terms $\ell_{1\infty}(v)$ and $\ell_{2\infty}(u)$. If $\rho(z) = \rho_\beta(u) + \rho_\alpha(v)$ does not hold, part (iii) of Theorem 3.2 does not apply. Also, the limit distribution of the Bayes estimator of the shape parameter α coincides with that of the MLE if the loss function ρ_α is symmetric, i.e., the limit distribution of $\widehat{\alpha}$ is given by $\mathcal{N}(0, \mathcal{J}^{-1})$. This does not take place in the case of the location parameter β. Furthermore, Bayes estimators of β generally are not transformations of the MLE of β asymptotically, as shown below, unlike in the nonregression or dummy regression cases.

REMARK 3.5: Theorem 3.2 justifies the validity of subsampling for Wald inference. Subsampling approximates the distribution of an estimator in the full sample based on values of this estimator in many smaller subsets of the sample. Politis, Romano, and Wolf (1999, Chapter 2) state implementation protocols for subsampling. Theorem 2.2.1 in Politis, Romano, and Wolf (1999) on validity of subsampling applies here provided (i) the estimates are consistent at polynomial in n rates, and (ii) the estimates have a limit distribution. As both (i) and (ii) are proven in Theorem 3.2, it follows immediately that subsampling is valid for large sample inference. Subsampling is useful in practice because it is less demanding than, say, parametric bootstrap in terms of computation and more attractive than other methods in terms of robustness to local misspecification of the parametric model. The latter property follows because we only need conditions (i) and (ii) for subsampling to be valid.

REMARK 3.6: The parametric bootstrap is an alternative to subsampling. Similarly to Ibragimov and Has'minskii (1981), the main convergence results of this paper can be stated uniformly in the parameter γ, and conditional on almost every realization of the covariate sequence $\{X_i, i \leq n\}$ as $n \to \infty$. This would formally establish that the parametric bootstrap is valid in the usual sense. Although Bayes estimates are not difficult to recompute, we found the parametric bootstrap to be extremely cumbersome and expensive computationally. In addition, it appears to be nonrobust against local misspecification of the parametric model.

3.4. *Further Properties of Bayes Procedures and Bayes Inference*

Next consider the posterior mean $\bar{\gamma}$ and the posterior quantile $\widehat{\gamma}(\tau)$ as the solutions of the problem (2.4) under squared loss and check loss functions, respectively (each defined in Section 2). Also define $\bar{Z}$ and $Z(\tau)$ as the solutions of the limit problem in Theorem 3.2 under squared and check functions, respectively.

THEOREM 3.3 (Further properties of the some Bayesian procedures): *Suppose that* C0–C5 *and* D1–D3 *stated in Appendix* A *hold. Then*:
 (i) *Posterior mean estimators are asymptotically mean-unbiased*:

$$\lim_{n\to\infty} E_{\gamma_n}[H_n^{-1}(\bar{\gamma} - \gamma_n)] = E_{\gamma_0}[\bar{Z}] = 0.$$

(ii) *For any* $0 < \tau' < \tau'' < 1$, *if* $Z(\tau)$ *has positive density in an open neighborhood of* 0 *for* $\tau = \tau'$ *and* $\tau = \tau''$, *then posterior* τ-*quantiles are asymptotically* $1 - \tau$-*quantile unbiased*:

$$(3.10) \quad \lim_{n\to\infty} P_{\gamma_n}\{(\widehat{\gamma}(\tau))_j \leq (\gamma_n)_j\} = P_{\gamma_0}\{(Z(\tau))_j \leq 0\} = 1 - \tau,$$

where $(\gamma)_j$ denotes the jth component of a vector γ. Hence

$$(3.11) \quad \lim_{n \to \infty} P_{\gamma_n}\{(\widehat{\gamma}(\tau'))_j \le (\gamma_n)_j \le (\widehat{\gamma}(\tau''))_j\} = \tau'' - \tau'.$$

Theorem 3.3 has a number of important implications. The asymptotic mean-unbiasedness of posterior means and the median-unbiasedness of posterior medians ($\tau = 1/2$) are very useful in practice. The asymptotic quantile-unbiasedness of posterior quantiles demonstrates that the posterior confidence intervals $[(\widehat{\gamma}(\tau'))_j, (\widehat{\gamma}(\tau''))_j]$ are valid for inference on the parameter components $(\gamma)_j$. The practical usefulness from this result is substantial since the intervals are easy to implement and require no knowledge of asymptotic theory. The theoretical usefulness is also substantial, since the argument used in the proof is generic and can be applied to other models where analytical demonstration of posterior validity is intractable.

The results (i) and (ii) in Theorem 3.3 follow from the asymptotic average risk optimality of posterior means and quantiles under squared and check loss functions, respectively; cf. Section 3.5. For example, if the limit posterior mean $\bar{Z}$ had a mean $E\bar{Z} = c \ne 0$, then the estimator $\bar{\gamma} - H_n c$ would have strictly lower asymptotic risk regardless of the local parameter sequence. Hence it must be that $E\bar{Z} = 0$. A similar argument extends to posterior quantiles. The posterior τ-quantile is $1 - \tau$-quantile unbiased because it is asymptotically optimal under the τ-check loss function. That $Z_j(\tau)$ has positive density around 0 is a technical requirement.

The next result extends the asymptotic validity of posterior quantiles to inference about smooth real functions of the parameter.[13] Consider inference about the function $r_n(\gamma)$, where $r_n : \mathcal{G} \to \mathbb{R}$ is such that for $a > 1$ and $R \equiv [\dot{R}', \ddot{R}']'$ with rank $R = 1$:

$$(3.12) \quad r_n(\gamma) - r_n(\gamma') = \dot{R}'(\alpha - \alpha')\sqrt{n} + \ddot{R}'(\beta - \beta')n$$
$$+ O\big(n|\beta - \beta'|^a + \sqrt{n}|\alpha - \alpha'|^a\big),$$

for all γ and γ' in an open neighborhood of γ_0. For purposes of theoretical analysis, the function is made dependent on n specifically to have a better finite-sample approximation by avoiding the trivial cases where all of the asymptotic inference is determined by either α or β due to the difference in rates of convergence. However, if a smooth function $m(\beta)$ is of prime interest, taking $r_n(\gamma) = n \cdot m(\beta)$ fulfills the condition (3.12). If a smooth function $m(\alpha)$ is of interest, then taking $r_n(\gamma) = \sqrt{n} \cdot m(\alpha)$ also fulfills the condition (3.12). Note that these transformations by $\sqrt{n}$ or n do not affect the practical formulations (2.6)–(2.8) in Section 2.2, by linearity of the transformations and equivariance of the quantiles to monotone transformations.

---

[13]Following, e.g., van der Vaart (2000) and for clarity of presentation, we consider only real functions r_n, but the results can be potentially generalized to multivariate functions by considering various multivariate analogs of the check function.

THEOREM 3.4 (Inference validity of posterior quantiles): *Suppose that* C0–C5 *and* D1–D3 *stated in Appendix* A *hold. Then*:

(i) *For any* $0 < \tau < 1$, $\widehat{c}(\tau) - r_n(\gamma_n) \overset{d}{\to} \widetilde{Z}(\tau)$, *where*

$$\widetilde{Z}(\tau) \equiv \underset{\tilde{z} \in \mathbb{R}}{\arg\inf} \int_{\mathbb{R}^d} \rho(\tilde{z} - R'z; \tau) \frac{\ell_\infty(z)}{\int_{\mathbb{R}^d} \ell_\infty(z')\, dz'}\, dz.$$

(ii) *Provided* $\widetilde{Z}(\tau)$ *has a positive density over an open neighborhood of* 0 *for* $\tau = \tau'$ *and* $\tau = \tau''$,

(3.13) $$\lim_{n\to\infty} P_{\gamma_n}\{\widehat{c}(\tau) \le r_n(\gamma_n)\} = P_{\gamma_0}\{\widetilde{Z}(\tau) \le 0\} = 1 - \tau,$$

$$\lim_{n\to\infty} P_{\gamma_n}\{\widehat{c}(\tau') \le r_n(\gamma_n) \le \widehat{c}(\tau'')\} = \tau'' - \tau'.$$

3.5. *On the Optimality of Bayes Procedures and Maximum Likelihood*

Lemma 3.1 given below states that BEs are asymptotically average risk optimal, which is most essential for establishing the inference results of Theorems 3.3 and 3.4. Lemma 3.1 is also of independent interest since it complements the results of Hirano and Porter (2003) on asymptotic (minimax) optimality of BEs in boundary models, by demonstrating asymptotic (average) optimality in both two-sided and one-sided models. Lemma 3.1 and the discussion that follows are also motivated as a response to the critique of maximum likelihood estimation by Hirano and Porter (2003). In fact, we argue that optimality analysis is useful not only for motivating BEs but also the MLE.

Define the normalization matrix H_n as in Section 3.1, and let $\gamma_{n\delta} \equiv \gamma_0 + H_n\delta$, for $\delta \in \mathbb{R}^d$, denote a local parameter sequence, where the dependence on δ needs to be emphasized here. Consider the set Y_n of all statistics (measurable mappings of data) $\widehat{\gamma}_n$. Define the expected risk associated with the loss function ρ and the estimator $\widehat{\gamma}_n$ as $E_{\gamma_{n\delta}}\rho(\widehat{Z}_n)$, where $\widehat{Z}_n = H_n^{-1}[\widehat{\gamma}_n - \gamma_{n\delta}]$ and the expectation is computed under $\gamma_{n\delta}$. Consider the following measures of risk.

The finite sample average risk (AR) of $\widehat{\gamma}_n$ is given by

(3.14) $$\frac{1}{\lambda(K)} \int_K E_{\gamma_{n\delta}}\rho(\widehat{Z}_n)\mu(\gamma_{n\delta})\, d\delta,$$

where μ is the weight or prior measure over K, ρ is the loss function, and λ is the Lebesgue measure. The asymptotic average risk (AAR) of the sequence $\{\widehat{\gamma}_n\}$ is given by

(3.15) $$\limsup_{K\uparrow\mathbb{R}^d} \limsup_{n\to\infty} \frac{1}{\lambda(K)} \int_K E_{\gamma_{n\delta}}\rho(\widehat{Z}_n)\, d\delta,$$

where $K \uparrow \mathbb{R}^d$ denotes an increasing sequence of cubes centered at the origin and converging to $\mathbb{R}^d$. Letting $K \uparrow \mathbb{R}^d$ is needed to rule out nonregular

(superefficient) estimators. Compared to the previous formula, the weight μ is replaced by the objective (uninformative) weight.

LEMMA 3.1: *Suppose that* C0–C5 *and* D1–D3 *hold. For* $\widehat{\gamma}_{\rho,\mu,n}$ *denoting the Bayes estimator under loss* ρ *and prior weight* μ, $Z_n \equiv H_n^{-1}[\widehat{\gamma}_{\rho,\mu,n} - \gamma_{n\delta}]$, $U_n \equiv n(\mathcal{B} - \beta_0) \times \sqrt{n}(\mathcal{A} - \alpha_0)$, *the following is true.*
 (i) *For each* n, *the infimum of the finite sample average risk for* $K = U_n$ *is achieved over* Y_n *by the Bayes estimator* $\widehat{\gamma}_{\rho,\mu,n}$, *i.e., by setting* $\widehat{Z}_n = Z_n$ *in* (3.14).
 (ii) *The infimum of the asymptotic average risk over estimator sequences in* Y_n *equals* $E_{\gamma_0}\rho(Z) < \infty$, *where* Z *denotes a weak limit of the normalized Bayes estimator* Z_n. *The infimum is attained by the sequence of the Bayes estimators* $\widehat{\gamma}_{\rho,\mu,n}$, *i.e., by setting* $\{\widehat{Z}_n\} = \{Z_n\}$ *in* (3.15).

Finite-sample average risk efficiency of Bayes estimators, appearing in part (i), is a basic result of mathematical statistics; cf. Wald (1950). Part (i) is often simply used as an alternative definition of Bayes procedures. Part (ii) translates part (i) into a statement of asymptotic average risk efficiency of Bayes estimators (essentially following the fundamental result of Ibragimov and Has'minskii (1981, p. 93)).

A very important point to be emphasized here is that, in contrast to the regular case, the asymptotic efficiency rankings critically rely on the loss function ρ. For example, the MLE is worse than the posterior mean under the squared loss, a point well emphasized in Hirano and Porter (2003), but performs better than the posterior mean under other loss functions, a point emphasized here and supported further in Section 4. A possible explanation for this is that the MLE can be approximated by a Bayes estimator under any loss function that approximates the delta function, e.g., $0 - 1$ loss $I[|u| > \epsilon]/\epsilon$, and penalizes the estimation mistakes differently than the squared loss does. In that sense the MLE is approximately optimal under these loss functions and thus may perform better under these loss functions than, say, the posterior mean. Thus, the MLE generally cannot be dominated by any other *given* BE when risk comparisons are made across different loss functions. Such comparisons are relevant when the empirical investigator does not know the loss function of the end-user of the estimation results. Moreover, the MLE is invariant to reparameterization, which makes it a valuable estimation approach, given its approximate optimality.

3.6. *Basic Large Sample Properties of Maximum Likelihood Procedures*

Consider the MLE $\widehat{\gamma} = (\widehat{\beta}', \widehat{\alpha}')' \equiv \arg\sup_{\gamma \in \mathcal{G}} L_n(\gamma)$ and define $Z_n \equiv (Z_n^{\beta'}, Z_n^{\alpha'})' \equiv (n(\widehat{\beta} - \beta_n)', \sqrt{n}(\widehat{\alpha} - \alpha_n)')'$ as the MLE centered at the true parameter and normalized by the convergence rates.

THEOREM 3.5 (Properties of MLE): *Under* C0–C5, *and supposing that* $-\ell_\infty(z)$ *attains a unique minimum in* $\mathbb{R}^d$ *a.s., then* $Z_n = O_p(1)$ *and* $Z_n \xrightarrow{d} Z \equiv$ $\arg\inf_{z \in \mathbb{R}^d} -\ell_\infty(z)$. *In particular,* $Z_n^\alpha \xrightarrow{d} Z^\alpha = \mathcal{J}^{-1}\mathbf{W} \overset{d}{=} \mathcal{N}(0, \mathcal{J}^{-1})$, $Z_n^\beta \xrightarrow{d} Z^\beta =$ $\arg\inf_{u \in \mathbb{R}^{d_\beta}} -\ell_{2\infty}(u)$, *and* Z^β *and* Z^α *are independent.*

Theorem 3.5 generalizes the results of Ibragimov and Has'minskii (1981) and Donald and Paarsch (1993a). To meet space constraints, the proof, which uses the remarkable epi-convergence framework of Knight (2000), and a detailed discussion are given in Chernozhukov and Hong (2001a). The limit result allows one to conduct Wald inference by either simulating the limit distribution or estimating it with subsampling; see previous Remarks 3.1 and 3.5 that apply here equally well. The limit variable Z is the argmin of the limit likelihood, which inherits the discontinuities of the finite sample likelihood. Due to asymptotic independence of the information about the shape parameter α from the information about the location parameter β, the MLEs for these parameters are asymptotically independent. Moreover, in boundary models, the limit result can be stated more explicitly for β as follows:

$$n(\widehat{\beta} - \beta_n) \xrightarrow{d} Z^\beta$$
$$\equiv \arg\inf_u \left(-\exp(u'\mathbf{m})\right) \text{ such that } J_i \geq \Delta(\mathcal{X}_i)'u, \text{ for all } i \geq 1).$$

REMARK 3.7 (Asymptotic nonsufficiency of MLE): It is important to note here that posterior means and medians are generally not equal to the bias corrected MLE. Consider the example of Section 3.2 where $\ell_\infty(z) \equiv e^{E(X)'z}\mathbf{1}(\Gamma_i \geq \mathcal{X}_i'z, \text{ for all } i \geq 1)$. The limit maximum likelihood variable $\widehat{Z}$ maximizes $\ell_\infty(z)$, which is equivalent to maximizing $E(X)'z$ subject to the constraint $\Gamma_i \geq \mathcal{X}_i'z$, for all $i \geq 1$. In the no covariate case, the limit MLE $\widehat{Z}$ maximizes e^z over z such that $\Gamma_i \geq z$ for all i; thus $\widehat{Z} = \min\{\Gamma_i, i \geq n\}$ and $\ell_\infty(z) = e^z\mathbf{1}(z \leq \widehat{Z})$, implying the sufficiency of $\widehat{Z}$. If $\widehat{Z}$ is sufficient, then the limit optimal (Bayes) estimators Z are all some shift transformations of $\widehat{Z}$ by the well-known Rao–Blackwell argument. This raises the question of whether $\widehat{Z}$ is a sufficient statistic for $\ell_\infty(z)$ in the general regression case. Taking the example with $X = (1, \tilde{X})$ where $\tilde{X}$ is continuous, it is easy to see that $\ell_\infty(z)$ is not a function of the MLE $\hat{Z}$ only, in particular $\ell_\infty(z) \neq e^{E(X)'z}\mathbf{1}(\mathcal{X}_i'z \leq \mathcal{X}_i'\widehat{Z}, \text{ for all } i \geq 1)$ with strictly positive probability. This implies that $\widehat{Z}$ is not sufficient for $\ell_\infty(z)$ even conditional on covariates. In fact, $\ell_\infty(z)$ is determined by the infinite-dimensional sufficient statistic $\{\mathcal{X}_i, \mathcal{X}_i', J_i, J_i'\}$ specified in Theorem 3.1. Thus, the limit Bayes estimators Z generally are functions of the entire likelihood $\ell_\infty(z)$ and are not nonrandom functions of the MLE $\widehat{Z}$.

4. COMPUTATIONAL EXPERIMENTS

4.1. *Monte Carlo Design and Computation*

We used a procurement auction model similar to that in Section 2.1, where we set $\theta_2 = 1$ and $\theta_1 = \exp(\beta_0 + \beta_1 X)$, with $X \sim U(0, 1)$, $\beta_0 = 1$, $\beta_1 = 1$, $m = 3$. We take $n = 100$ and $n = 400$, which are close to sample sizes encountered in empirical work on auctions. We used the parameter space $\mathcal{B} = [\beta_0 \pm 5] \times [\beta_1 \pm 5]$ and a flat prior to compute Bayes estimates. The starting value was set to be 0 in the computation of the estimates. The computations were performed using the canonical random-walk MCMC algorithm described in Robert and Casella (1998, p. 245).[14] The MLE was computed by taking the argmax over the sequence of draws generated by the MCMC algorithm.

4.2. *Quality of Estimation Procedures*

We compare the performance of (i) the posterior mean, (ii) the posterior median, and (iii) the posterior mode (MLE) across different risk measures. Results given in Tables I and II demonstrate that: (a) the posterior mean is the best under the mean squared loss, (b) the posterior median is the best under the mean absolute deviation loss, (c) the MLE does better under the mean 10th power loss function. Thus, all of the listed likelihood-based estimators perform well relative to some risk measure and none dominate, supporting the theoretical discussion given previously.

TABLE I

ESTIMATOR PERFORMANCE FOR INTERCEPT β_0 (BASED ON 500 REPETITIONS)

Estimator	RMSE	MAD	Median AD	pth loss ($p = 10$)
$n = 100$				
Posterior Mean	.0114	.0081	.0059	.0401
Posterior Median	.0115	.0078	.0054	.0433
MLE	.0127	.0088	.0063	.0369
$n = 400$				
Posterior Mean	.0029	.0021	.0015	.0104
Posterior Median	.0030	.0020	.0014	.0106
MLE	.0034	.0023	.0015	.0103

[14]In our implementation, the first 20,000 draws are made for a "burn-in stage," with adjustments made to the variance of transition kernel every 200 draws in order to keep the rejection rate near .5. Then, additional 20,000 draws were made with a fixed variance, and used in the computation of the estimates. The C++ implementation is available from the authors.

TABLE II

ESTIMATOR PERFORMANCE FOR SLOPE β_1 (BASED ON 500 REPETITIONS)

Estimator	RMSE	MAD	Median AD	pth loss ($p = 10$)
$n = 100$				
Posterior Mean	.0212	.0147	.0105	.0983
Posterior Median	.0214	.0144	.0096	.1053
MLE	.0216	.0145	.0096	.0693
$n = 400$				
Posterior Mean	.0053	.0037	.0026	.0227
Posterior Median	.0054	.0037	.0025	.0228
MLE	.0057	.0038	.0024	.0221

4.3. *Quality of Inferential Procedures*

In the next step we compare several methods for constructing confidence intervals. We compare confidence intervals based on posterior quantiles, which is the primary inference method advocated in the paper with a number of other procedures. In Table III we record the performance of (i) confidence intervals based on posterior quantiles, (ii) percentile confidence intervals based on parametric bootstrap of the posterior mean, (iii) percentile confidence intervals based on subsampling the posterior mean (using $1/4 \times n$ as the subsample size), and (iv)–(vi) percentile confidence intervals based on simulating the limit distribution of the MLE and other estimators, as described in Remark 3.1 (using $b = n$).

Results reported in Table III indicate that confidence intervals based on posterior quantiles and confidence intervals based on simulating the limit distribution perform better than other methods considered. Confidence intervals based on posterior quantiles also appear to be the shortest on average. Given that posterior quantiles are the least expensive to compute, they could be preferred in practice along with subsampling.[15] Subsampling is also an inexpensive method, in fact much less expensive than the parametric bootstrap, and is probably the most robust inference method in the presence of local misspecification.

---

[15]In terms of computational expense, computation of posterior quantiles takes less than 1 minute on a Pentium III PC. Simulation of the limit distribution is roughly twice as expensive (because limit expressions are simple transformations of linear functions and do not contain nonlinear expressions). We used 200 bootstrap draws and full sample estimates as starting values in the MCMC algorithm, which reduces the number of MCMC draws needed in the re-computation of the estimates. Using this implementation, 200 bootstrap draws take between 7 and 30 minutes for samples $n = 100$ and $n = 400$. Thus, 1000 bootstrap draws take up to 150 minutes for $n = 400$. The subsampling with 200 draws takes about 1/5th of the time of the parametric bootstrap.

TABLE III

COMPARISON OF INFERENCE METHODS: COVERAGE AND AVERAGE LENGTH OF THE NOMINAL
90% CONFIDENCE INTERVALS (BASED ON 500 REPETITIONS)

Confidence Interval	Coverage: Intercept	Length: Intercept	Coverage: Slope	Length: Slope
$n = 100$				
Posterior Interval	.87	.0298	.85	.0555
Bootstrap *P*-mean	.88	.0392	.86	.0720
Subsampling: *P*-mean	.83	.0416	.82	.0770
Limit Process: *P*-mean	.86	.0346	.84	.0587
Limit Process: *P*-median	.85	.0352	.86	.0610
Limit Process: MLE	.93	.0347	.95	.0653
$n = 400$				
Posterior Intervals	.87	.0075	.84	.0140
Bootstrap: *P*-mean	.84	.0089	.88	.0167
Subsampling: *P*-mean	.82	.0085	.83	.0158
Limit Process: *P*-mean	.89	.0085	.82	.0145
Limit Process: *P*-median	.86	.0087	.86	.0150
Limit Process: MLE	.89	.0084	.92	.0157

5. CONCLUSION

We study estimation and inference in a general model in which the conditional density of the dependent variable jumps at a location that is parameter dependent. This model includes a number of important economic models discussed in the recent literature on structural estimation. We derive large sample theory for a variety of likelihood-based procedures and offer an array of useful and practical inference techniques, including Bayes and Wald inference methods. The results provide a theoretical and practical solution to the important econometric problem posed by Aigner, Amemiya, and Poirier (1976).

Dept. of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, E52-262F, Cambridge, MA 02142, U.S.A.; vchern@mit.edu
and
Dept. of Economics, Duke University, P.O. Box 90097, Durham, NC 27708-0097, U.S.A.; hanhong@duke.edu.

APPENDIX A: REGULARITY CONDITIONS C0–C5 AND D1–D3

NOTATION: Throughout the paper, c and const denote generic positive constants; $\|x\|$ is the Euclidian norm $\sqrt{x'x}$, and $|x|$ is the supremum norm, i.e., $|x| = \sup_{j \leq k} |x_j|$, where $x = (x_1, \dots, x_k)$. The densities $f(\epsilon|x, \gamma)$ are discontinuous at $\epsilon = 0$ and are not differentiable in ϵ at $\epsilon = 0$. To simplify notation, we use $\partial f(\epsilon|x, \gamma)/\partial \epsilon$ to denote both the usual partial derivative when $\epsilon \neq 0$ and the directional partial derivative $\partial f(0^+|x, \gamma)/\partial \epsilon$ when $\epsilon = 0$, etc. Also, $B_\delta(\gamma)$ denotes a closed ball at γ of radius δ as measured by $|\cdot|$.

Conditions C0–C5

The following conditions apply to $x \in \mathbf{X}$ and $\epsilon \in \mathbb{R}$. Conditions C0–C3 apply to any $\gamma = (\beta, \alpha) \in \mathcal{G}$. Conditions C4 and C5 apply to any $\gamma = (\beta, \alpha) \in \mathcal{G}$ and $\bar{\gamma} = (\bar{\beta}, \bar{\alpha}) \in B_\delta(\gamma)$ for some $\delta > 0$.

CONDITION C0: For each γ, (Y_i, X_i) is an i.i.d. sequence of vectors in $\mathbb{R} \times \mathbb{R}^k$, defined on a probability space $(\Omega, \mathcal{F}, P_\gamma)$. $\mathcal{G} \subset \mathbb{R}^d$ is a compact convex set such that $\gamma_0 \in$ interior $\mathcal{G}$; for any γ and $\bar{\gamma} \neq \gamma$, $P_\gamma\{f(Y_i - g(X_i, \bar{\beta})|X_i, \bar{\gamma}) \neq f(Y_i - g(X_i, \beta)|X_i, \gamma)\} > 0$.

CONDITION C1: X_i has a distribution function F_X, which does not depend on γ, and has a compact support $\mathbf{X}$. In addition to (2.2), uniformly in γ and x, we have either:
  (i) the two-sided model: $p(x, \gamma) > q(x, \gamma) > \underline{f} > 0$, or
  (ii) the one-sided model: $p(x, \gamma) > \underline{f} > 0$ and $f(\epsilon|x, \gamma) = q(x, \gamma) = 0$, for all $\epsilon < 0$.

CONDITION C2: Without loss of generality, the density $f(\epsilon|x, \gamma)$ is upper-semicontinuous at $\epsilon = 0$ for each (x, γ). $f(\epsilon|x, \gamma)$ has continuous first partial derivative in ϵ (except at $\epsilon = 0$) for each (x, γ); $f(\epsilon|x, \gamma)$ has continuous first and second partial derivative in γ, for each (x, ϵ, γ). The density and the said derivatives are continuous in x on $\mathbf{X}$ for each (ϵ, γ). Moreover,

$$\sup_{\epsilon \in \mathbb{R}, x \in \mathbf{X}, \gamma \in \mathcal{G}} f(\epsilon|x, \gamma) < \bar{f},$$

$$\sup_{\epsilon \in \mathbb{R}, x \in \mathbf{X}, \gamma \in \mathcal{G}} \left| \frac{\partial}{\partial(\epsilon, \gamma)} f(\epsilon|x, \gamma) \right| < \bar{f}',$$

$$\sup_{\epsilon \in \mathbb{R}, x \in \mathbf{X}, \gamma \in \mathcal{G}} \left| \frac{\partial^2}{\partial \gamma \, \partial \gamma'} f(\epsilon|x, \gamma) \right| < \bar{f}'',$$

where $\bar{f}, \bar{f}', \bar{f}'' < \infty$. Lastly, $\sup_\gamma E_X \int |\frac{\partial}{\partial \gamma} f(y - g(X, \beta)|X; \gamma)| \, dy < \infty$.

CONDITION C3: The function $g(x, \beta)$ has two continuous derivatives in β, for each (x, β), and

$$\sup_{\beta \in \mathcal{B}, x \in \mathbf{X}} |g(x, \beta)| < \bar{g},$$

$$\sup_{\beta \in \mathcal{B}, x \in \mathbf{X}} \left| \frac{\partial}{\partial \beta} g(x, \beta) \right| < \bar{g}',$$

$$\sup_{\beta \in \mathcal{B}, x \in \mathbf{X}} \left| \frac{\partial^2}{\partial \beta \, \partial \beta'} g(x, \beta) \right| < \bar{g}'',$$

where $\bar{g}, \bar{g}', \bar{g}'' < \infty$, and

$$E_X \left[\frac{\partial g(X; \beta)}{\partial \beta} \frac{\partial g(X; \beta)'}{\partial \beta} \right]$$

is positive definite uniformly in β. The function $g(x, \beta)$ and its derivatives

$$\frac{\partial}{\partial \beta} g(x, \beta) \quad \text{and} \quad \frac{\partial^2}{\partial \beta \, \partial \beta'} g(x, \beta)$$

are continuous in x on $\mathbf{X}$ for each β.

CONDITION C4: When the parameter α is present, for $l_i(\bar{\gamma}) \equiv \ln f(Y_i - g(X_i, \bar{\beta})|X_i, \bar{\gamma})$, where $\bar{\gamma} = (\bar{\beta}, \bar{\alpha})$, uniformly in $\bar{\gamma}$ and γ, either (a)

$$E_\gamma\left[\frac{\partial}{\partial\bar{\gamma}}l_i(\bar{\gamma})\frac{\partial}{\partial\bar{\gamma}}l_i(\bar{\gamma})'\right]$$

is positive definite and bounded or (b) if $\frac{\partial}{\partial\beta}l_i(\gamma) = 0$ P_γ-a.s., then

$$E_\gamma\left[\frac{\partial}{\partial\bar{\alpha}}l_i(\bar{\gamma})\frac{\partial}{\partial\bar{\alpha}}l_i(\bar{\gamma})'\right]$$

is positive definite and bounded.

CONDITION C5: In the two-sided model C1(i), the terms

$$\left|\frac{\partial}{\partial\bar{\beta}}\ln f(Y_i - g(X_i, \bar{\beta})|X_i, \bar{\gamma})\right|,$$

(A.1) $\qquad \left|\frac{\partial}{\partial\bar{\gamma}}\ln f(Y_i - g(X_i, \bar{\beta})|X_i, \bar{\gamma})\right|^2,$

$$\left|\frac{\partial^2}{\partial\bar{\gamma}\partial\bar{\gamma}'}\ln f(Y_i - g(X_i, \bar{\beta})|X_i, \bar{\gamma})\right|$$

are bounded respectively by $C_j(\epsilon_i, X_i)$, $j = 1, 2, 3$, for all $Y_i - g(X_i, \bar{\beta}) \in \mathbb{R} \setminus \{0\}$, uniformly in $\bar{\gamma} \in B_\delta(\gamma)$, where $\sup_\gamma E_\gamma C_j(\epsilon_i, X_i) < \infty$ for $j = 1, 2, 3$. Similarly, for the one-sided model C1(ii), the terms in (A.1) are bounded respectively by $C_j(\epsilon_i, X_i)$, $j = 1, 2, 3$, for all $Y_i - g(X_i, \bar{\beta}) > 0$, uniformly in $\bar{\gamma} \in B_\delta(\gamma)$, where $\sup_\gamma E_\gamma C_j(\epsilon_i, X_i) < \infty$ for $j = 1, 2, 3$.

Conditions D1–D3

The prior $\mu: \mathcal{G} \to \mathbb{R}_+$ and the loss function $\rho: \mathbb{R}^{d_\alpha + d_\beta} \to \mathbb{R}_+$ have the following properties:

PROPERTY D1: $\mu(\cdot) > 0$, is continuous on $\mathcal{G}$, and vanishes outside $\mathcal{G}$.

PROPERTY D2: $\rho(\cdot) \geq 0$, $\rho(z) = 0$ if and only if $z = 0$, and ρ is convex.

PROPERTY D3: $\rho(z)$ is dominated by a polynomial of $|z|$ as $|z| \longrightarrow \infty$.

REMARK A.1 (Constants): The constants $\bar{f}$, $\bar{f}'$, $\bar{f}''$, $\bar{g}$, $\bar{g}'$, and $\bar{g}''$ specified above as well as the constant $\underline{f}$ will be extensively used in the proofs. In view of C1–C2, for some $\delta > 0$ and $V(0) = [-\delta, \delta] \setminus \{0\}$ in the case C1(i) and $V(0) = (0, \delta]$ in the case C1(ii), $\underline{f}$ is any constant that satisfies $\inf_{\epsilon \in V(0), x \in \mathbf{X}, \gamma \in \mathcal{G}} f(\epsilon|x, \gamma) > \underline{f} > 0$.

REMARK A.2: Conditions C0–C5 and D1–D3 represent a generalization of the assumptions in Ibragimov and Has'minskii (1981, Chapter V) and Smith (1985). Condition C1(ii) allows for the boundary model, where the density is zero to the left side of the jump and is positive on the right side. Condition C1(i) allows for the two-sided model where the density is positive on both sides. Conditions of type C3–C5 are common in nonlinear analysis and conditions of type D1–D3 in Bayesian analysis.

APPENDIX B: PROOFS

PROOF OF THEOREM 3.1: In the proof we set the local parameter sequence $\gamma_n = \gamma_0$. Considering a general sequence does not change the argument but complicates notation.

Following Ibragimov and Has'minskii (1981), we split the log-likelihood ratio process $Q_n(z) \equiv \ln \ell_n(z) \equiv \ln L_n(\gamma_0 + H_n z)/L_n(\gamma_0)$ into the continuous part $Q_n^c(z)$ and the piecewise constant part $Q_n^d(z)$, and analyze each part separately. Our goal is to show that $Q_n(z)$ converges in distribution in the finite-dimensional sense to $Q_\infty(z) \equiv Q_\infty^c(z) + Q_\infty^d(z)$, where $Q_\infty^c(z) = \mathbf{W}'v - \frac{1}{2}v'\mathcal{J}v + \mathbf{m}'u$, $Q_\infty^d(z) = \int_{\mathbb{R} \times \mathbf{X}} l_u(j, x) \, d\mathbf{N}(j, x)$, and each term is defined in Theorem 3.1. Given this result, the finite-dimensional limit of $\ell_n(z)$ is $\ell_\infty(z) = \exp(Q_\infty(z))$. For $z = (u', v')'$, using that $\epsilon_i = Y_i - g(X_i, \beta_0)$,

$$Q_n(z) \equiv \underbrace{\sum_{i=1}^n \widehat{r}_{in}(z) \times \left[\mathbf{1}(\epsilon_i > \{\Delta_n(X_i, u)/n\} \vee 0) + \mathbf{1}(\epsilon_i < \{\Delta_n(X_i, u)/n\} \wedge 0) \right]}_{Q_{1n}^c(z)}$$

$$+ \underbrace{\sum_{i=1}^n \left(\widehat{r}_{in}(z) - r_{in}(z) \right) \times \left[\mathbf{1}(0 < \epsilon_i \leq \Delta_n(X_i, u)/n) \right.}_{}$$

$$\left. + \mathbf{1}(0 > \epsilon_i \geq \Delta_n(X_i, u)/n) \right]$$

$$\underbrace{\phantom{+ \sum_{i=1}^n}}_{Q_{2n}^c(z)}$$

$$+ \underbrace{\sum_{i=1}^n r_{in}(z) \times \left[\mathbf{1}(0 < \epsilon_i \leq \Delta_n(X_i, u)/n) + \mathbf{1}(0 > \epsilon_i \geq \Delta_n(X_i, u)/n) \right]}_{Q_n^d(z)}$$

$$\equiv Q_n^c(z) + Q_n^d(z), \quad \text{where}$$

$$\widehat{r}_{in}(z) \equiv \ln \left[\frac{f(Y_i - g(X_i, \beta_0 + u/n)|X_i, \beta_0 + u/n, \alpha_0 + v/\sqrt{n})}{f(Y_i - g(X_i, \beta_0)|X_i, \gamma_0)} \right]$$

$$\equiv \ln \left[\frac{f(\epsilon_i - \Delta_n(X_i, u)/n|X_i, \beta_0 + u/n, \alpha_0 + v/\sqrt{n})}{f(\epsilon_i|X_i, \gamma_0)} \right],$$

$$r_{in}(z) \equiv \ln \left[\frac{q(X_i)}{p(X_i)} \right] \mathbf{1}(0 < \epsilon_i) + \ln \left[\frac{p(X_i)}{q(X_i)} \right] \mathbf{1}(0 > \epsilon_i),$$

$$\Delta_n(x, u) \equiv n \big(g(X_i, \beta_0 + u/n) - g(X_i, \beta_0) \big).$$

The convergence analysis of the continuous part $Q_n^c(z)$ is standard. In sharp contrast, the behavior of the discontinuous part $Q_n^d(z)$ differs from that of $Q_n^c(z)$, and is analyzed using the point process methods. Also, in the expressions above and all the proofs we use the algebraic rules of Ibragimov and Has'minskii (1981) defined in Theorem 3.1 for working with ∞'s . This is done to include the proof for the boundary model as a special case. In particular, the expressions involving $\mathbf{1}(0 > \epsilon_i > \cdots)$ cancel, since in the boundary model $\epsilon_i > 0$. Also in the boundary model we define $\widehat{r}_{in}(z) - r_{in}(z) \equiv 0$ when $0 < \epsilon_i \leq \Delta_n(X_i, u)/n$, so that $Q_{2n}^c(z) \equiv 0$. Thus $Q_{2n}^c(z) \not\equiv 0$ only for the two-sided model. Further details follow in Parts I and II.

Part I obtains the finite-dimensional limit of $Q_n^c(z)$. The proof method is standard for the smooth likelihood analysis. Application of Taylor expansion to each $\widehat{r}_{in}(z)$, $i = 1, \ldots, n$, so that the

expanded terms are i.i.d., followed by application of the Markov LLN and Chebyshev inequality, yields for a given z

$$Q_{1n}^c(z) \equiv \underbrace{-u'E\Delta(X_i)\frac{f'(\epsilon_i|X, \gamma_0)}{f(\epsilon_i|X_i, \gamma_0)}}_{\equiv u'\mathbf{m}} + v'\underbrace{\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\ln f(\epsilon_i|X_i, \gamma_0)\right]}_{\equiv \mathbf{W}_n}$$

$$+ \frac{1}{2}v'\underbrace{\left[E\frac{\partial^2}{\partial\alpha\,\partial\alpha'}\log f(\epsilon|X, \gamma_0)\right]}_{\equiv -\mathcal{J}}v + o_p(1),$$

where $\Delta(X_i) = \partial g(X_i, \beta_0)/\partial\beta$. The information matrix equality for α implies

$$-\mathcal{J} = E\frac{\partial^2\log f(\epsilon|X, \gamma_0)}{\partial\alpha\,\partial\alpha'}$$

$$= -E\frac{\partial\log f(\epsilon|X, \gamma_0)}{\partial\alpha}\frac{\partial\log f(\epsilon|X, \gamma_0)}{\partial\alpha}',$$

and the CLT gives $\mathbf{W}_n \xrightarrow{d} \mathbf{W} = \mathcal{N}(0, \mathcal{J})$. Also, it follows by C2 that $\mathbf{m} = E\Delta(X_i)(p(X_i) - q(X_i))$. Therefore, the finite-dimensional limit of $Q_{1n}^c(z)$ is given by $Q_{\infty}^c(z) \equiv u'\mathbf{m} + \mathbf{W}'v - \frac{1}{2}v'\mathcal{J}v$.

It remains to show $Q_{2n}^c(z) = o_p(1)$. In the one-sided case $Q_{2n}^c(z) \equiv 0$; hence we only need to consider the two-sided case. Note that by C1–C3, for any compact set $\mathbf{Z}$, as $n \to \infty$, for some constant c that depends only on $\bar{f}/\underline{f}$ and $\bar{g}'$,

(B.1)
$$\left|\ln\left[\frac{f(\epsilon - \Delta_n(x, u)/n|x, \beta_0 + u/n, \alpha_0 + v/\sqrt{n})}{f(\epsilon|x, \gamma_0)}\right] - \ln\left[\frac{q(x)}{p(x)}\right]\right|$$

$$\leq c \times \|z\|/\sqrt{n}$$

uniformly in $\{\epsilon, z, x \in \mathbb{R}_+ \times \mathbf{Z} \times \mathbf{X}: \Delta_n(x, u) > 0, 0 < \epsilon \leq \Delta_n(x, u)/n\}$. Likewise

(B.2)
$$\left|\ln\left[\frac{f(\epsilon - \Delta_n(x, u)/n|x, \beta_0 + u/n, \alpha_0 + v/\sqrt{n})}{f(\epsilon|x, \gamma_0)}\right] - \ln\left[\frac{p(x)}{q(x)}\right]\right|$$

$$\leq c \times \|z\|/\sqrt{n}$$

uniformly in $\{\epsilon, z, x \in \mathbb{R}_- \times \mathbf{Z} \times \mathbf{X}: \Delta_n(x, u) < 0, 0 > \epsilon \geq \Delta_n(x, u)/n\}$. Thus

(B.3)
$$\sup_{z\in\mathbf{Z}}|Q_{2n}^c(z)| \leq c \times \|\mathbf{Z}\|/\sqrt{n} \times \sum_{i=1}^{n}\mathbf{1}(|\epsilon_i| < K/n) = O_p(1/\sqrt{n})$$

for $K = \|\mathbf{Z}\| \times \bar{g}' < \infty$, where $\|\mathbf{Z}\| = \sup\{\|z\|: z \in \mathbf{Z}\}$, where $K < \infty$ by C3. The $O_p(1/\sqrt{n})$ conclusion is by C2:

(B.4)
$$E\sum_{i=1}^{n}\mathbf{1}(|\epsilon_i| < K/n) \leq 2\bar{f}K < \infty.$$

Part II obtains the finite-dimensional limit of $Q_n^d(z)$. Recall

$$Q_n^d(z) \equiv \sum_{i=1}^{n}\left[\ln\frac{q(X_i)}{p(X_i)}\mathbf{1}(0 < n\epsilon_i \leq \Delta_n(X_i, u)) + \ln\frac{p(X_i)}{q(X_i)}\mathbf{1}(0 > n\epsilon_i \geq \Delta_n(X_i, u))\right].$$

By C2 and C3

$$E \sum_{i=1}^{n} \left| \mathbf{1}(0 < n\epsilon_i \leq \Delta_n(X_i, u)) - \mathbf{1}(0 < n\epsilon_i \leq \Delta(X_i)'u) \right|$$

$$+ \left| \mathbf{1}(0 > n\epsilon_i \geq \Delta_n(X_i, u)) - \mathbf{1}(0 > n\epsilon_i \geq \Delta(X_i)'u) \right|$$

$$\leq 2\bar{f}\bar{g}'' \|u\|^2 / n = o(1),$$

where $\Delta(X_i) \equiv \partial g(X_i, \beta_0)/\partial \beta$, which implies that for a given z

$$Q_n^d(z) = \sum_{i=1}^{n} \left[\ln \frac{q(X_i)}{p(X_i)} \mathbf{1}(0 < n\epsilon_i < \Delta(X_i)'u) + \ln \frac{p(X_i)}{q(X_i)} \mathbf{1}(0 > n\epsilon_i > \Delta(X_i)'u) \right]$$

$$+ o_p(1).$$

Now note that $(Q_n^d(z_j), j \leq l)$ and $(Q_n^c(z_j), j \leq l)$, for any finite l, are asymptotically independent. This follows by applying a standard argument concerning the independence of extreme order statistics and sample averages; see, e.g., Lemma 21.19 in van der Vaart (1999). Details are omitted for brevity, but can be found in Chernozhukov and Hong (2001a).

The next step is to obtain the finite-dimensional limit of Q_n^d. The behavior of Q_n^d is determined by near-to-jump observations, which behavior is described using point processes. We split the argument in two steps. Step 1 constructs the required point process and derives its limit. Step 2 applies Step 1 to obtain the finite-dimensional limit of Q_n^d.

Step 1. The intuition for Step 1 is provided in Section 3.2 of the main text.

Define $E \equiv \mathbb{R} \times \mathbf{X}$. The topology on E is standard; e.g., $[a, b] \times \mathbf{X}$ is a compact subset relative to E. The point process of interest is a random measure taking the following form: for any Borel subset A, $\widehat{\mathbf{N}}(A) = \sum_{i=1}^{n} \mathbf{1}[(n\epsilon_i, X_i) \in A]$. We take $\widehat{\mathbf{N}}$ to be a random element of $M_p(E)$, the metric space of nonnegative point measures on E, with the metric generated by the topology of vague convergence; cf. Resnick (1987, Chapter 3). We show that $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E)$, for $\mathbf{N}$ given in Theorem 3.1. This is done in the steps (a) and (b).

(a) By C1 and C2, for any $F \in \mathcal{T}$, the basis of relatively compact open sets in E (finite unions and intersections of open bounded rectangles in E),

$$
\begin{aligned}
\lim_{n \to \infty} E\widehat{\mathbf{N}}(F) &\equiv \lim_{n \to \infty} nP((n\epsilon_i, X_i) \in F) \\
&= \int_F [p(x)\mathbf{1}(u > 0)\, du + q(x)\mathbf{1}(u < 0)\, du]\, dF_X(x) \\
&= m(F) < \infty,
\end{aligned}
$$
(B.5)

where the measure m is defined as $dm(u, x) = [p(x)\mathbf{1}(u > 0)\, du + q(x)\mathbf{1}(u < 0)\, du]\, dF_X(x)$. Since $\{(n\epsilon_i, X_i) \in F\}$ are independent across i by C0, by Meyer's Theorem (cf. Meyer (1973))

$$\lim_{n \to \infty} P(\widehat{\mathbf{N}}(F) = 0) = e^{-m(F)}.$$
(B.6)

Statements (B.5) and (B.6) imply by Kallenberg's theorem (cf. Resnick (1987, Proposition 3.22)) that $\widehat{\mathbf{N}} \Rightarrow N$ in $M_p(E)$, where N is a Poisson point process with the mean intensity measure $m(\cdot)$.

(b) Next we show that N has the same distribution as $\mathbf{N}$ given in Theorem 3.1. First, consider the canonical Poisson processes N_0 and N_0' with points $\{\Gamma_i\}$ and $\{\Gamma_i'\}$ defined in Theorem 3.1. N_0 has the mean measure $m_0(du) = du$ on $(0, \infty)$, and N_0' has the mean measure $m_0'(du) = du$ on $(-\infty, 0)$; see Resnick (1987, p. 138). Because N_0 and N_0' are independent, $N_1(\cdot) \equiv N_0(\cdot) + N_0(\cdot)'$ is a Poisson point process with mean measure $m_1(du) = du$ on $\mathbb{R}$, by definition of the Poisson process; see Resnick (1987, p. 130). Because $\{\mathcal{X}_i, \mathcal{X}_i'\}$ are i.i.d. and independent of $\{\Gamma_i, \Gamma_i'\}$, by Proposition 3.8 in Resnick (1987), the composed process N_2 with points $(\{\Gamma_i, \mathcal{X}_i\}, \{\Gamma_i', \mathcal{X}_i'\}, i \geq 1)$ is a Poisson process with the mean measure $m_2(du, dx) = [\mathbf{1}(u >$

0) $du + \mathbf{1}(u < 0)\, du] \times F_X(dx)$ on $\mathbb{R} \times \mathbf{X}$. Finally, $\mathbf{N}$ with the points $\{T(\Gamma_i, \mathcal{X}_i), T(\Gamma_i', \mathcal{X}_i')\}$, where $T : (u, x) \mapsto (\mathbf{1}(u > 0)u/p(x) + \mathbf{1}(u < 0)u/q(x), x)$, is a Poisson process with the desired mean measure $m(du, dx) = m_2 \circ T^{-1}(du, dx) = [p(x)\mathbf{1}(u > 0) + q(x)\mathbf{1}(u < 0)]\, du\, F_X(dx)$, by Proposition 3.7 in Resnick (1987).

Step 2. We have for $z = (u', v')'$,

$$
Q_n^d(z) = Q_n^d(u) = \left[\sum_{i=1}^{n} \ln \frac{q(X_i)}{p(X_i)} \mathbf{1}[0 < n\epsilon_i \leq \Delta(X_i)'u] \right.
$$

$$
\left. + \sum_{i=1}^{n} \ln \frac{p(X_i)}{q(X_i)} \mathbf{1}[0 > n\epsilon_i \geq \Delta(X_i)'u] \right] + o_p(1).
$$

Ignoring the $o_p(1)$ term, write $Q_n^d(u)$ as a Lebesgue integral with respect to $\widehat{\mathbf{N}}$:

$$
Q_n^d(u) = \int_E l_u(j, x)\, d\widehat{\mathbf{N}}(j, x),
$$

where $l_u(j, x)$ is defined in Theorem 3.1. The convergence of this integral is implied by $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in both the two-sided and one-sided model:

(a) In the two-sided model: By conditions C1–C3, the function $(j, x) \mapsto l_u(j, x)$ is bounded and vanishes outside the compact set $\mathcal{K}_u \equiv [-\eta, +\eta] \times \mathbf{X}$, $\eta = \sup_{x \in \mathbf{X}} |\Delta(x)'u|$, where $\eta < \infty$ by C3. Thus $(j, x) \mapsto l_u(j, x)$ has compact support but is discontinuous when $j = 0$ and $j = \Delta(x)'u$. Define the map $T : M_p(E) \mapsto \mathbb{R}^l$ as $N \mapsto (\int_E l_{u_k}(j, x)\, dN(j, x), k \leq l)$ for $l < \infty$. Hence by Proposition 3.13 in Resnick (1987), T is discontinuous at $\mathcal{D}(T) \equiv \{N \in M_p(E) : j_i^N = 0 \text{ or } j_i^N = u_k'\Delta(x_i^N)$ for some $i \geq 1, k \leq l\}$ where $(j_i^N, x_i^N, i \geq 1)$ denote the points of N. Since ϵ_i's are absolutely continuous, $P[\widehat{\mathbf{N}} \in \mathcal{D}(T)$, for some $n \geq 1] = 0$, and by definition of $\mathbf{N}$, $P[\mathbf{N} \in \mathcal{D}(T)] = 0$. Therefore $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E)$ implies $T(\widehat{\mathbf{N}}) \xrightarrow{d} T(\mathbf{N})$ by the continuous mapping theorem; cf. Resnick (1987, p. 153). It follows that $(Q_n^d(u_k), k \leq l) \xrightarrow{d} (Q_\infty^d(u_k), k \leq l)$, where $Q_\infty^d(u) \equiv \int_E l_u(j, x)\, d\mathbf{N}(j, x)$.

(b) In the one-sided model: Using the Ibragimov and Has'minskii (1981) rules for algebraic operations with ∞'s stated in Theorem 3.1, note that $Q_n^d(u) = \int_E l_u(j, x)\, d\widehat{\mathbf{N}}(j, x)$ is a binomial random variable: $Q_n^d(u) = -\infty$ if $\widehat{\mathbf{N}}(A(u)) > 0$ and $Q_n^d(u) = 0$ if $\widehat{\mathbf{N}}(A(u)) = 0$, where $A(u) \equiv \{(j, x) \in \mathbb{R}_+ \times \mathbf{X} : j \leq \Delta(x)'u\}$. Also define $Q_\infty^d(u) \equiv \int_E l_u(j, x)\, d\mathbf{N}(j, x)$, so that $Q_\infty^d(u) = -\infty$ if $\mathbf{N}(A(u)) > 0$ and $Q_\infty^d(u) \equiv 0$ if $\mathbf{N}(A(u)) = 0$. Thus, to show the finite-dimensional convergence (for $\gamma_k = -\infty$ or 0): $\lim_{n \to \infty} P(Q_n^d(u_k) = \gamma_k, k \leq l) = P(Q_\infty^d(u_k) = \gamma_k, k \leq l)$, it suffices to show $(\widehat{\mathbf{N}}(A(u_k)), k \leq l) \xrightarrow{d} (\mathbf{N}(A(u_k)), k \leq l)$ for $l < \infty$. By a definition of weak convergence of point processes (cf. Embrechts, Klüppelberg, and Mikosch (1997, p. 232)), this is immediate from $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$, since by C2 and construction of $\mathbf{N}$, $\widehat{\mathbf{N}}(\partial A(u_k)) = 0$ and $\mathbf{N}(\partial A(u_k)) = 0$ a.s. *Q.E.D.*

PROOF OF THEOREM 3.2: The proof applies Theorem I.10.2 of Ibragimov and Has'minskii (1981, p. 107), which allows one to obtain the limit distribution of BEs provided some conditions on the likelihood ratio process are satisfied.

First, BEs are measurable by the Jennrich's measurability theorem since they minimize objective functions that are continuous in data and parameters. Second, we shall make use of the following important lemmas proved in Appendix C. Let $\gamma = (\beta, \alpha)$ and $h = (h_\beta, h_\alpha)$. Define the Hellinger distance

$$
r_2(\gamma; \gamma + h)^2
$$

$$
= \int \int \left| f^{1/2}(y - g(x, \beta + h_\beta); x, \gamma + h) - f^{1/2}(y - g(x, \beta); x, \gamma) \right|^2 dy\, F_X(dx).
$$

Note that $F_X^{1/2}(dx)$ is taken outside the $|\cdot|^2$-brackets, since it does not depend on the parameters.

LEMMA B.1 (Hellinger distance properties): *Under* C0–C5, *there are* $a > 0$ *and* $A > 0$ *such that for all* h *such that* $\gamma + h \in \mathcal{G}$, *uniformly in* $\gamma \in \mathcal{G}$

$$(B.7) \qquad (a) \quad r_2^2(\gamma; \gamma + h) \geq 2 \frac{a \max(|h_\beta|, |h_\alpha|^2)}{1 + \max(|h_\beta|, |h_\alpha|^2)} \quad and$$

$$(b) \quad r_2^2(\gamma; \gamma + h) \leq A(|h_\beta| + |h_\alpha|^2).$$

LEMMA B.2 (Exponential tails and Holder continuity): *Given* (B.7), *for all* $n > n_0$, (z, z') *with* $\gamma + H_n z \in \mathcal{G}$ *and* $\gamma + H_n z' \in \mathcal{G}$, *and some* $a' > 0$ *and* n_0, *uniformly in* $\gamma \in \mathcal{G}$

$$(B.8) \qquad E_\gamma \ell_n(z)^{1/2} \leq e^{-a'(|z|-1)}, \qquad E_\gamma |\ell_n(z)^{1/2} - \ell_n(z')^{1/2}|^2 \leq A(|z - z'|)(1 + 2 \cdot |z'| \vee |z|).$$

The following conditions 1–4 verify the conditions of Theorem I.10.2 of Ibragimov and Has'minskii (1981, p. 107):

1. Holder continuity of $\ell_n^{1/2}(z)$ in the mean square, and the exponential bound on the expected likelihood tail, both proved in Lemma B.2. In the latter case, we have that for $a' > 0$, $E_\gamma \ell_n(z)^{1/2} \leq e^{-a'(|z|-1)}$, where $z \mapsto a'(|z|-1)$ falls into the function class **G** of Ibragimov and Has'minskii (1981, p. 41), i.e., $a'(|z|-1)$ is increasing in $|z|$ on $[0, \infty)$ and $\lim_{|z| \to \infty} |z|^N e^{-a'(|z|-1)} = 0$ for any $N > 0$.

2. Finite-dimensional convergence of $\ell_n(z) = \exp(Q_n(z))$ to $\ell_\infty(z) = \exp(Q_\infty(z))$, established in Theorem 3.1.

3. The limit Bayes problem,

$$Z = \arg\inf_{z' \in \mathbb{R}^d} \int_{\mathbb{R}^d} \rho(z' - z) \frac{\ell_\infty(z)}{\int_{\mathbb{R}^d} \ell_\infty(\bar{z}) \, d\bar{z}} \, dz,$$

is uniquely solved by a random vector Z, which is by D2 (since ρ is convex with a unique minimum; cf. Ibragimov and Has'minskii (1981, p. 107)).

4. Conditions D1–D3 on the loss functions ρ and prior μ. (It must be noted that Ibragimov and Has'minskii (1981) impose the symmetry of ρ throughout their book. However, the inspection of their proof of Theorems I.10.2 (and Theorem I.5.2) reveals that the proof does not require symmetry and applies to the loss functions that satisfy D1–D3.)

Thus, conditions 1–4 imply by Theorem I.10.2 of Ibragimov and Has'minskii (1981) that $Z_n \xrightarrow{d} Z$. Furthermore, conditions 1–4 imply by Theorem I.5.2 and Theorem I.10.2 of Ibragimov and Has'minskii (1981) that for any $\delta \in \mathbb{R}^d$, $\gamma_{n\delta} = \gamma_0 + H_n \delta$, $N > 0$,

$$(B.9) \qquad \lim_{L \to \infty, n \to \infty} L^{-N} P_{\gamma_{n\delta}}\{|Z_n| > L\} = 0, \quad and \quad \lim_{n \to \infty} E_{\gamma_{n\delta}} \rho(Z_n) = E_{\gamma_0} \rho(Z) < \infty.$$

The result (B.9) is not needed to prove Theorem 3.2 but will be used later. Q.E.D.

PROOF OF THEOREM 3.3: To show Claim 1, note that $\lim_{n \to \infty} E_{\gamma_{n\delta}}[H_n^{-1}(\bar{\gamma} - \gamma_n)] = E_{\gamma_0} \bar{Z}$ by (B.9). Consider the problem $\min_c E_{\gamma_0} \rho(\bar{Z} + c)$, where $\rho(z) = z'z$. The solution of this problem is $c = -E\bar{Z}$. Suppose that $c \neq 0$; then

$$(B.10) \qquad E_{\gamma_0} \rho(\bar{Z} + c) < E_{\gamma_0} \rho(\bar{Z}),$$

where by Lemma 3.1 stated in Section 3.5 the left-hand side of (B.10) is the asymptotic average risk of the sequence of estimators $\bar{\gamma} + H_n c$ and the right-hand side of (B.10) is the asymptotic average risk of the sequence of posterior means $\bar{\gamma}$, which contradicts the asymptotic average risk efficiency of the posterior mean established in Lemma 3.1. Thus it must be that $c = -E\bar{Z} = 0$.

To show Claim 2, note that by Theorem 3.2 and the definition of weak convergence,

$$\lim_{n \to \infty} P_{\gamma_{n\delta}}\{(\hat{\gamma}(\tau))_j \leq (\gamma_0)_j\} = \lim_{n \to \infty} P_{\gamma_{n\delta}}\{(Z_n(\tau))_j \leq 0\} = P_{\gamma_0}\{(Z(\tau))_j \leq 0\},$$

since 0 is assumed to be a continuity point of the distribution of $(Z(\tau))_j$. Consider the problem $\min_c E_{\gamma_0} \rho((Z(\tau))_j - c; \tau)$, where $\rho(z; \tau) = (\mathbf{1}(z \geq 0) - \tau)z$. Note that the quantity

$E_{\gamma_0}\rho((Z(\tau))_j - c; \tau)$ is finite for any c by (B.9). A solution of this problem is given by the root of the first-order condition

$$\text{(B.11)} \qquad P_{\gamma_0}\{(Z(\tau))_j \geq c\} = \tau \quad \text{or} \quad P_{\gamma_0}\{(Z(\tau))_j \leq c\} = 1 - \tau,$$

i.e., $c = (1 - \tau)$th quantile of $(Z(\tau))_j$ (under the condition that $(Z(\tau))_j$ has positive density in any small neighborhood of 0). Suppose $c \neq 0$; then

$$\text{(B.12)} \qquad E_{\gamma_0}\rho\big((Z(\tau))_j - c; \tau\big) < E_{\gamma_0}\rho\big((Z(\tau))_j; \tau\big),$$

where by Lemma 3.1 the left-hand side of (B.12) is the asymptotic average risk of the sequence of estimators defined as $(\hat{\gamma}(\tau) - H_n c)_j$, and the right-hand side is the asymptotic average risk of the posterior τth quantile $(\hat{\gamma}(\tau))_j$. See Section 3.5 and Lemma 3.1 for definitions. Then (B.12) contradicts the asymptotic average risk efficiency of the posterior quantiles under the check loss function established in Lemma 3.1.

Thus it must be that $c = 0$ in (B.11), so that the first part of Claim 2, (3.10), is proven. The second part of Claim 2, (3.11), is immediate from (B.11) with $c = 0$ for $\tau = \tau'$ and $\tau = \tau''$:

$$\text{(B.13)} \qquad \lim_{n \to \infty} P_{\gamma_{n\delta}}\{(\hat{\gamma}(\tau'))_j \leq (\gamma_n(\delta))_j \leq (\hat{\gamma}(\tau''))_j\}$$

$$= \lim_{n \to \infty} P_{\gamma_{n\delta}}\{(Z_n(\tau'))_j \leq 0 \leq (Z_n(\tau''))_j\}$$

$$= 1 - \lim_{n \to \infty} P_{\gamma_{n\delta}}\{(Z_n(\tau''))_j \leq 0\} - \lim_{n \to \infty} P_{\gamma_{n\delta}}\{(Z_n(\tau'))_j \geq 0\}$$

$$= 1 - P_{\gamma_0}\{(Z(\tau''))_j \leq 0\} - P_{\gamma_0}\{(Z(\tau'))_j \geq 0\} = \tau'' - \tau'. \qquad \textit{Q.E.D.}$$

PROOF OF THEOREM 3.4: We will use the following lemma.

LEMMA B.3 (Integral convergence of $\ell_n(z)$): *Suppose that* (i) $\ell_n(z)$ *has the properties specified in Lemma B.2 and* (ii) $\ell_n(z)$ *converges marginally to* $\ell_\infty(z)$ *under a parameter sequence* $\gamma_{n\delta} = \gamma_0 + H_n\delta$. *Then* (a) $\ell_\infty(z) > 0$ *in some open ball at zero a.s.,* (b) *for any vector-valued continuous function* $g(z)$ *dominated by a polynomial as* $z \to \infty$

$$\int_{K_n} g(z) \frac{\ell_n(z)\mu(\gamma_0 + H_n z)}{\int_{K_n} \ell_n(z')\mu(\gamma_0 + H_n z')\, dz'}\, dz \xrightarrow{d} \int_K g(z) \frac{\ell_\infty(z)}{\int_K \ell_\infty(z')\, dz'}\, dz.$$

Here either $K_n = U_n \equiv \{z : \gamma_{n\delta} + H_n z \in \mathcal{G}\}$ *and* $K = \mathbb{R}^d$ *or* $K_n = K = U$, *where* U *is a fixed cube centered at the origin; and* $\xrightarrow{d}$ *denotes weak convergence under* P_{γ_n}.

The first part of the proof of Theorem 3.4 is done by setting the true parameter sequence $\gamma_{n\delta} = \gamma_0$. Considering a general sequence does not change the argument but complicates notation. Write $\tilde{Z}_n(\tau) = (\hat{c}(\tau) - r_n(\gamma_0))$. Note that $\tilde{Z}_n(\tau) \equiv \arg\inf_{\tilde{z} \in \mathbb{R}} \Gamma_n(\tilde{z})$, where

$$\Gamma_n(\tilde{z}) \equiv \int_{\mathbb{R}^d} \rho\big(\tilde{z} - r_n(\gamma_0 + H_n z) + r_n(\gamma_0); \tau\big) \frac{\ell_n(z)\mu(\gamma_0 + H_n z)}{\int_{\mathbb{R}^d} \ell_n(z')\mu(\gamma_0 + H_n z')\, dz'}\, dz.$$

Since $|r_n(\gamma_0 + H_n z) - r_n(\gamma_0) - R'z| \leq \text{const } |z| \cdot |H_n z|^{a'}$ for $a' > 0$, by the piece-wise linearity of the check function we have that $|\rho(\tilde{z} - r_n(\gamma_0 + H_n z) + r_n(\gamma_0); \tau) - \rho(\tilde{z} - R'z; \tau)| \leq \text{const} \cdot |z| \cdot |H_n z|^{a'}$. Hence by Lemma B.3

$$\text{(B.14)} \qquad \Gamma_n(\tilde{z}) = \int_{\mathbb{R}^d} \big(\rho(\tilde{z} - R'z; \tau) + O(|z| \cdot |H_n z|^{a'})\big) \frac{\ell_n(z)\mu(\gamma_0 + H_n z)}{\int_{\mathbb{R}^d} \ell_n(z')\mu(\gamma_0 + H_n z')\, dz'}\, dz$$

$$= \int_{\mathbb{R}^d} \rho(\tilde{z} - R'z; \tau) \frac{\ell_n(z)\mu(\gamma_0 + H_n z)}{\int_{\mathbb{R}^d} \ell_n(z')\mu(\gamma_0 + H_n z')\, dz'}\, dz + o_p(1).$$

Applying Lemma B.3 again, it follows that the marginal limit of $\Gamma_n(\tilde{z})$ is given by

$$\Gamma_\infty(\tilde{z}) = \int_{\mathbb{R}^d} \rho(\tilde{z} - R'z; \tau) \frac{\ell_\infty(z)}{\int_{\mathbb{R}^d} \ell_\infty(z')\,dz'}\,dz.$$

Recall that $\tilde{Z}(\tau)$ denotes the minimizer of $\Gamma_\infty(\tilde{z})$. By the Convexity Lemma, e.g., Knight (2000) and Davis, Knight, and Liu (1992), the stated finite-dimensional convergence of convex objective functions implies $\tilde{Z}_n(\tau) \overset{d}{\to} \tilde{Z}(\tau)$. Thus, in what follows it suffices to consider only linear functions such that $r_n(\gamma_0 + H_n z) - r_n(\gamma_0) - R'z = 0$ for all z.

Next, we need to establish the uniform integrability of $\rho(\tilde{Z}_n(\tau); \tau)$. Consider the linear transformation $\xi = M'z$ defined by a nonsingular matrix M such that R is a column of M. Then, the likelihood for ξ given by $\ell_n(M^{-1}\xi)$ satisfies for some $c > 0$ and $c' > 0$:

(a) $E_\gamma |\ell_n^{1/2}(M^{-1}\xi') - \ell_n^{1/2}(M^{-1}\xi'')|^2 \leq c|\xi' - \xi''|(1 + 2|\xi'| \vee |\xi''|)$,

(b) $E_\gamma \ell_n^{1/2}(M^{-1}\xi) \leq e^{-c'(|\xi|-1)}$,

by nonsingularity of M and Lemma B.2. By Theorem I.5.2 of Ibragimov and Has'minskii (1981) for any local sequence of $\gamma_{n\delta} = \gamma_0 + H_n\delta$, $\delta \in \mathbb{R}^d$, and any $N > 0$, $\lim_{L\to\infty, n\to\infty} L^{-N} P_{\gamma_{n\delta}}\{|\tilde{Z}_n(\tau)| > L\} = 0$. Hence $\lim_{n\to\infty} E_{\gamma_{n\delta}} \rho(\tilde{Z}_n(\tau); \tau) = E_{\gamma_0} \rho(\tilde{Z}(\tau); \tau) < \infty$. Then similarly to Lemma 3.1, it can be concluded that $\{\hat{c}(\tau)\}$ minimizes the asymptotic average risk, i.e., it achieves the infimum of

$$\limsup_{K\uparrow\mathbb{R}^d} \limsup_{n\to\infty} \frac{1}{\lambda(K)} \int_K E_{\gamma_{n\delta}} \rho(Z_n; \tau)\,d\delta$$

over all statistic sequences $\{\hat{c}\}$ that are measurable functions of $(Y_i, X_i, i \leq n)$, where $Z_n = \hat{c} - r_n(\gamma_{n\delta})$. The rest of the argument, which establishes the $(1 - \tau)$-quantile unbiasedness of the posterior τ-quantiles and the resulting coverage properties, is identical to the proof of Theorem 3.3.   Q.E.D.

PROOF OF LEMMA 3.1: Claim 1 is just a special case of Theorem 1.1 of Lehmann and Casella (1998, Chapter 5). Claim 2 follows by the argument given by Ibragimov and Has'minskii (1981, p. 93). Details can be found in Chernozhukov and Hong (2001a).   Q.E.D.

APPENDIX C: PROOFS OF LEMMAS B.1–B.4

PROOF OF LEMMA B.1: In order to establish (B.7)(b), let $\gamma = (\beta, \alpha)$ and $h = (h_\beta, h_\alpha)$,

$$
\begin{aligned}
&r_2^2(\gamma; \gamma + h) \\
&= E_X \int \left(f^{1/2}(y - g(X, \beta + h_\beta)|X; \beta + h_\beta, \alpha + h_\alpha) - f^{1/2}(y - g(X, \beta)|X; \gamma) \right)^2 dy \\
&\overset{(1)}{\leq} E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]} \left| f(y - g(X, \beta + h_\beta)|X; \beta + h_\beta, \alpha + h_\alpha) \right. \\
&\hspace{6cm} \left. - f(y - g(X, \beta)|X; \gamma) \right| dy \\
&\quad + E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]^c} \left(f^{1/2}(y - g(X, \beta + h_\beta)|X; \gamma + h) \right. \\
&\hspace{5cm} \left. - f^{1/2}(y - g(X, \beta + h_\beta)|X; \beta + h_\beta, \alpha) \right)^2 dy \\
&\quad + E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]^c} \left| f(y - g(X, \beta + h_\beta)|X; \beta + h_\beta, \alpha) \right. \\
&\hspace{6cm} \left. - f(y - g(X, \beta)|X; \gamma) \right| dy
\end{aligned}
$$

$$\overset{(2)}{\leq} 2E_X |g(X, \beta + h_\beta) - g(X, \beta)| \bar{f}$$

$$+ \text{const} |h_\alpha|^2 \int_0^1 E_X \int \left| \frac{\partial f^{1/2}(y - g(X, \beta + h_\beta)|X, \beta + h_\beta, \alpha + \omega h_\alpha)}{\partial \alpha} \right|^2 dy\, d\omega$$

$$+ \text{const} |h_\beta| \int_0^1 E_X \int \left| \frac{\partial f(y - g(X, \beta + \omega h_\beta)|X, \beta + \omega h_\beta, \alpha)}{\partial \beta} \right| dy\, d\omega$$

$$\overset{(3)}{\leq} \text{const}\, \bar{f} |h_\beta| E_X \int_0^1 \left| \frac{\partial g(X, \beta + \omega h_\beta)}{\partial \beta} \right| d\omega + \text{const}\, (|h_\alpha|^2) + \text{const}\, (|h_\beta|)$$

$$\overset{(4)}{=} \text{const}\, (|h_\beta|) + \text{const}\, (|h_\alpha|^2),$$

where $[a, b] = [a, b]$ if $a \leq b$ and $= [b, a]$ if $b \leq a$, and the bound is uniform in γ. (1) is by triangle inequality and $|a - b|^2 \leq |a^2 - b^2|$ for $a > 0$ and $b > 0$. The first term in (2) follows from $|f(\cdot|\cdot)| \leq \bar{f}$. The second and third terms in (2) are by Taylor expansion and Fubini. The first term in (3) follows from Taylor expansion and Fubini. The second term in (3) follows from C4, while the third term in (3) is by C2. (4) is by C3.

The lower bound from below, equation (B.7)(a), is established by considering separately $|h| \leq \delta$ for some sufficiently small δ and $|h| > \delta$. Indeed, for sufficiently small δ and $|h| \leq \delta$ it is shown below that

$$(C.1) \qquad r_2^2(\gamma; \gamma + h) \geq \text{const} \max(|h_\beta|, |h_\alpha|^2).$$

On the other hand, by the identification condition C0 for all $|h| > \delta$ such that $\gamma + h \in \mathcal{G}$,

$$(C.2) \qquad r_2^2(\gamma; \gamma + h) \geq \varepsilon_\delta > 0.$$

Hence for some $a > 0$ the bound in (B.7)(a) is immediate from (C.1)–(C.2).

It remains to prove (C.1) for $|h| < \delta$ for some sufficiently small δ. Write

$$r_2^2(\gamma; \gamma + h) = E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]} \big(f^{1/2}(y - g(X, \beta + h_\beta)|X; \gamma + h)$$

$$\underbrace{- f^{1/2}(y - g(X, \beta)|X; \gamma) \big)^2 dy}_{I}$$

$$+ E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]^c} \big(f^{1/2}(y - g(X, \beta + h_\beta)|X; \gamma + h)$$

$$\underbrace{- f^{1/2}(y - g(X, \beta)|X; \gamma) \big)^2 dy.}_{II}$$

For small h_β, we can bound I from below uniformly in γ by

$$E_X \frac{1}{2} |g(X, \beta + h_\beta) - g(X, \beta)| |p^{1/2}(X, \gamma) - q^{1/2}(X, \gamma)|^2$$

$$\geq \text{const}\, E_X \left| \frac{\partial g(X; \beta)}{\partial \beta}' h_\beta \right| \geq \text{const} |h_\beta|,$$

using C3 and Taylor expansion. On the other hand, by C1–C3, bound II from below by

$$E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]^c} \left(h' \frac{\partial f^{1/2}(y - g(X; \beta)|\gamma)}{\partial \gamma} \right)^2 dy - o(|h|^2).$$

By C2, C3, and C4(a), a further lower bound on II is

$$|h|^2 \inf_{|u|=1} E_X \int_{[g(X,\beta),g(X,\beta+h_\beta)]^c} \left(\frac{f^{1/2}(y-g(X,\beta);\gamma)'}{\partial\gamma} u \right)^2 dy$$

$$\geq |h|^2 \left(\inf_{|u|=1} E_X \int \left(\frac{\partial f^{1/2}(y-g(X,\beta);\gamma)'}{\partial\gamma} u \right)^2 dy + O(|h_\beta|) \right)$$

$$\geq \text{const} |h|^2 \geq \text{const} |h_\alpha|^2,$$

for all sufficiently small $|h|$, where the remainder term $O(|h_\beta|)$ arises from neglecting the integrand over the small area $[g(X,\beta), g(X,\beta+h_\beta)]$ and using the bounds in C2 and C3 to do so. On the other hand, if assumption C4(b) holds, the uniform lower bound on II is

$$E_X \int \left(h' \frac{\partial f^{1/2}(y-g(X;\beta)|\gamma)}{\partial\gamma} \right)^2 dy$$

$$\geq \text{const} |h_\alpha|^2 \inf_{|u|=1} E_X \int \left(\frac{f^{1/2}(y-g(X,\beta);\gamma)'}{\partial\alpha} u \right)^2 dy$$

$$\geq \text{const} |h_\alpha|^2,$$

for all sufficiently small $|h|$. Conclude $\inf_\gamma r_2^2(\gamma;\gamma+h) \geq \text{const} \max(|h_\beta|,|h_\alpha^2|)$.   Q.E.D.

PROOF OF LEMMA B.2: We have

$$E_\gamma \ell_n(z)^{1/2} \overset{(1)}{\leq} \left[1 - \frac{1}{2} r_2^2(\gamma; \beta+u/n, \alpha+v/\sqrt{n}) \right]^n$$

$$\overset{(2)}{\leq} \exp\left\{ -\frac{n}{2} r_2^2(\gamma; \beta+u/n, \alpha+v/\sqrt{n}) \right\}$$

$$\overset{(3)}{\leq} \exp\left\{ -a \frac{\max(|u|,|v|^2)}{1+\max(|u|,|v|^2)/n} \right\}$$

$$\overset{(4)}{\leq} \exp\left\{ -a \frac{\max(|u|,|v|^2)}{1+K_{\mathcal{G}}} \right\}$$

$$\overset{(5)}{\leq} \exp\left\{ \frac{-a|z|+a}{1+K_{\mathcal{G}}} \right\},$$

where the constant $K_{\mathcal{G}}$ depends only on the diameter of the parameter space $\mathcal{G}$; (1) follows by the standard manipulation of the Hellinger distance, as in Ibragimov and Has'minskii (1981, p. 260); (2) follows by the inequality $(1-r) \leq e^{-r}$ when $r > 0$; (3) is given by (B.7), and (4) and (5) are obvious. Also,

$$E_\gamma |\ell_n(z)^{1/2} - \ell_n(z')^{1/2}|^2 \overset{(1)}{\leq} n r_2^2\big(\gamma+(u/n,v/\sqrt{n}); \gamma+(u'/n,v'/\sqrt{n})\big)$$

$$\overset{(2)}{\leq} A(|u-u'|+|v-v'|^2)$$

$$\overset{(3)}{\leq} A(|z-z'|+|z-z'|^2)$$

$$\overset{(4)}{\leq} A(|z-z'|)(1+2\cdot|z'|\vee|z|),$$

where (1) follows by the standard manipulation of the Hellinger distance, as in Ibragimov and Has'minskii (1981, p. 260), (2) is given, and (3) and (4) are obvious. *Q.E.D.*

PROOF OF LEMMA B.3: Assertion (a) is a special case of Lemma I.5.1 in Ibragimov and Has'minskii (1981). Assertion (b) is proven of Ibragimov and Has'minskii (1981, pp. 106–109) under more general conditions than conditions (i) and (ii). *Q.E.D.*

REFERENCES

AIGNER, D., T. AMEMIYA, AND D. POIRIER (1976): "On the Estimation of Production Frontiers: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function," *International Economic Review*, 17, 377–396.

BOWLUS, A., G. NEUMANN, AND A. KIEFER (2001): "Equilibrium Search Models and the Transition from School to Work," *International Economic Review*, 42, 317–343.

CHERNOZHUKOV, V. (2000): "Conditional Extremes and Near-Extremes: Estimation, Inference, and Economic Applications," Ph.D. Dissertation, Department of Economics, Stanford.

CHERNOZHUKOV, V., AND H. HONG (2001a): "Likelihood Inference in a Class of Nonregular Econometrics Models," MIT Department of Economics Working Paper, Revised October 2003, distributed by the Social Science Research Network, http://www.ssrn.com.

——— (2001b): "Likelihood Inference with Density Jumps," MIT Department of Economics Working Paper, presented at CEME Conference.

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

DAVIS, R. A., K. KNIGHT, AND J. LIU (1992): "M-estimation for Autoregressions with Infinite Variance," *Stochastic Processes and Applications*, 40, 145–180.

DONALD, S. G., AND H. J. PAARSCH (1993a): "Maximum Likelihood Estimation when the Support of the Distribution Depends upon Some or All of the Unknown Parameters," Working Paper, Department of Economics, University of Western Ontario.

——— (1993b): "Piecewise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review*, 34, 121–148.

——— (1996): "Identification, Estimation, and Testing in Parametric Empirical Models of Auctions Within Independent Private Values Paradigm," *Econometric Theory*, 12, 517–567.

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

EMBRECHTS, P., C. KLÜPPELBERG, AND T. MIKOSCH (1997): *Modelling Extremal Events*. Berlin: Springer-Verlag.

FLINN, C., AND J. HECKMAN (1982): "New Methods for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, 18, 115–190.

GHOSAL, S. (1999): "Probability Matching Priors for Non-Regular Cases," *Biometrika*, 86, 956–964.

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 (2003): "Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support," *Econometrica*, 71, 1307–1338.

IBRAGIMOV, I., AND R. HAS'MINSKII (1981): *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag.

KNIGHT, K. (2000): "Epi-convergence and Stochastic Equisemicontinuity," Working Paper, University of Toronto.

——— (2002): "Limiting Distributions of Linear Programming Estimators," *Extremes*, 4, 87–103.

KOENKER, R., AND G. S. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

LEHMANN, E., AND G. CASELLA (1998): *Theory of Point Estimation*. New York: Springer-Verlag.

MEYER, R. M. (1973): "A Poisson-Type Limit Theorem for Mixing Sequences of Dependent Rare Events," *The Annals of Probability*, 1, 480–483.

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

POLITIS, D., J. ROMANO, AND M. WOLF (1999): *Subsampling*. New York: Springer-Verlag.

PORTNOY, S., AND J. JUREČKOVÁ (2000): "On Extreme Regression Quantiles," *Extremes*, 2, 227–243.

RESNICK, S. I. (1986): "Point Processes, Regular Variation and Weak Convergence," *Advances in Applied Probability*, 18, 66–138.

——— (1987): *Extreme Values, Regular Variation, and Point Processes*. New York: Springer-Verlag.

ROBERT, C. P., AND G. CASELLA (1998): *Monte Carlo Statistical Methods*. New York: Springer-Verlag.

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

——— (1994): "Nonregular Regression," *Biometrika*, 81, 173–183.

TIERNEY, L. (1994): "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701–1762.

VAN DER VAART, A. (1999): *Asymptotic Statistics*. New York: Cambridge University Press.

——— (2000): "Lectures on Bayesian Statistics," Unpublished Manuscript, Aarhus University.

WALD, A. (1950): *Statistical Decision Functions*. New York: John Wiley & Sons.