Massachusetts Institute of Technology
Department of Economics
Working Paper Series

LIKELIHOOD ESTIMATION & INFERENCE IN A CLASS OF NONREGULAR ECONOMETRIC MODELS

Victor Chernozhukov
Han Hong

Working Paper 03-11
January 2003
Revised: May 9, 2003

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

LIKELIHOOD ESTIMATION AND INFERENCE IN A CLASS OF NONREGULAR ECONOMETRIC MODELS

VICTOR CHERNOZHUKOV AND HAN HONG

ABSTRACT. In this paper we study estimation and inference in structural models with a jump in the conditional density, where the location and size of the jump are described by regression lines. Two prominent examples are auction models, where the density jumps from zero to a positive value, and the equilibrium job search model, where the density jumps from one level to another, inducing kinks in the cumulative distribution function. An early model of this kind was introduced by Aigner, Amemiya, and Poirier (1976), but the estimation and inference in such models remained an unresolved problem, with the important exception of the specific cases studied by Donald and Paarsch (1993a) and the univariate case in Ibragimov and Has'minskii (1981a). The main difficulty is the statistical non-regularity of the problem caused by discontinuities in the likelihood function. This difficulty also makes the problem computationally challenging.

This paper develops estimation and inference theory and methods for such models based on likelihood procedures, focusing on the optimal (Bayes) procedures, including the MLEs. We obtain results on convergence rates and distribution theory, and develop Wald and Bayes type inference and confidence intervals. The Bayes procedures are attractive both theoretically and computationally. The Bayes confidence intervals, based on the posterior quantiles, are shown to provide a valid large sample inference method with good small sample properties. This inference result is of independent practical and theoretical interest due to the highly non-regular nature of the likelihood in these models, in which the maximum likelihood statistic or any finite dimensional statistic is not asymptotically sufficient.

Key Words: Likelihood Principle, Point Process, Frequentist Validity, Posterior, Structural Econometric Model, Auctions, Equilibrium Search, Production Frontier

The original draft of this paper was presented at the December 2000 EC^2 meetings in Dublin. This version is March, 2003. We would like to thank Joe Altonji, Stephen Donald, Christian Hansen, Jerry Hausman, Joel Horowitz, Ivan Fernandez, Shakeeb Khan, Yuichi Kitamura, Rosa Matzkin, Whitney Newey, George Neumann, Harry Paarsch, Frank Schorfheide, Robin Sickles, Richard Spady, Max Stinchcombe as well as participants in the CEME conference, the EC2 meeting, MIT, Northwestern, Princeton, Rice, University of Texas at Austin, and Duke for valuable suggestions. We thank three referees and the co-editor for numerous useful suggestions that improved the paper considerably. We are grateful to our our advisor Takeshi Amemiya for suggesting this problem to us.

1. Introduction

This paper develops theory for estimation and inference methods in structural models with jumps in the conditional density, where the locations of the jumps are described by parametric regression curves. The jumps in the density are very informative about the parameters of these curves, and result in non-regular and difficult inference theory, implying highly discontinuous likelihoods, non-standard 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 analysis. Many recent econometric models also share this interesting structure. For example, in structural procurement auction models, cf. Donald and Paarsch (1993a), the conditional density jumps from zero to a positive value at the lowest cost; in equilibrium job search models (Bowlus, Neumann, and Kiefer (2001)), the density jumps from one positive level to another at the wage reservation, inducing kinks in the wage distribution function. 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 locations of the jumps are linked to the parameters of the underlying structural economic model. Learning the parameters of the location of the jumps is thus crucial for learning the parameters of the underlying economic model.

Several early fundamental papers have developed inference methods for several cases of such models, including Aigner, Amemiya, and Poirier (1976), Ibragimov and Has'minskii (1981a), Flinn and Heckman (1982), Christensen and Kiefer (1991), Donald and Paarsch (1993a, 1993b, 1996, 2002), and Bowlus, Neumann, and Kiefer (2001). Ibragimov and Has'minskii (1981a) (Chapter V) obtained the limit theory of the likelihood-based optimal (Bayes) estimators in the general univariate non-regression case, and obtained the properties of MLE in the case of one-dimensional parameter. van der Vaart (1999) (Chapters 9.4-9.5) discussed the limit theory for the likelihood in the univariate Uniform and Pareto models, including Pareto models with parameter-dependent support and additional shape parameters. Paarsch (1992) and Donald and Paarsch (1993a, 1993b, 1996, 2002) introduced and developed the theory of likelihood (MLE) and related procedures in the one-sided regression models with discrete regressors, demonstrated the wide prevalence of such models in structural econometric modeling, and stimulated further research in this area.

Nevertheless, the general inference problem posed by Aigner, Amemiya, and Poirier (1976) has remained unsolved previously. Very little is known about likelihood-based estimation and inference in the general two-sided regression model. In the general one-sided regression model, the problem of likelihood-based estimation and inference also remains an important unresolved question, an important exception being the MLE theory for discrete regressors developed by Donald and Paarsch (1993a).[1] The general theory of such regression models is more involved and has a substantively different structure than the corresponding theory for the univariate (non-regression) or dummy

[1]There is also a literature on the ad hoc "linear programming" estimators of linear boundary functions covering continuous covariate case, see e.g. Smith (1994) (Chernozhukv (2001) provides a detailed review and other related

regressor case.[2] Moreover, there is a considerable implementation problem caused by the inherent computational difficulty of the classical (maximum likelihood) estimates.

This paper offers solutions to these open questions by providing theory for estimation and inference methods in both one and two-sided models with general regressors. These methods rely on the likelihood-based optimal[3] Bayes and also the MLE procedures. This paper demonstrates that these are tractable, computationally and theoretically attractive ways to obtain parameter estimates, construct confidence intervals, and carry out statistical inference. These results cover Bayes type inference as well as Wald type inference.

We show that Bayes inference methods, based on the posterior quantiles, are valid in large samples and also perform well in small samples. Moreover, these inference methods are tractable and require no knowledge of asymptotic theory on the practitioner's part. These estimation methods are also attractive due to their well-known finite-sample and large sample average risk optimality. They are computationally attractive when carried out through the Markov Chain Monte Carlo procedure (MCMC), see e.g. Robert and Casella (1998), which helps avoid the inherent curse of dimensionality in the computation of the MLE.

All of these results are preceded by a complete large sample theory of likelihood for these models, which is useful not only for the present analysis but also for any kind of inference based on the likelihood principle. Importantly, we show that the MLE is generally not an asymptotically sufficient statistic in these models (in contrast to the non-regression case or dummy regressor case). Therefore, the likelihood contains more information than the MLE does, and the totality of likelihood-based procedures are generally not functions of the MLE asymptotically, as they are in the non-regression or dummy regressor case (or regular models). This motivates the study of the entire likelihood and the wide class of the likelihood-based procedures.

results). In some special cases, such as homoscedastic exponential linear regression models, these estimators coincide with the MLE asymptotically.

[2]In fact, we show in this paper that unlike in the univariate models, such as the Uniform and Pareto models discussed in details by van der Vaart (1999) or dummy regression case, there are no finite-dimensional sufficient statistics. MLE is not asymptotically sufficient either, making inference theory difficult to analyze. We show that the limit likelihoods depend on multivariate Poisson point processes with complex correlation structure.

[3]This terminology follows that of Berger (1993), p. 17.

Our work is also related to a recent important contribution by Hirano and Porter (2002). They provide a detailed analysis of asymptotic minimax efficiency in a class of boundary models.[4] They employ an exponential-shift experiment framework along with group analysis to generate new results and insights on the efficiency structure of Bayesian estimators (which also motivate the present research) and prove the inefficiency (sub-optimality) of the MLE under the common mean squared and absolute deviation criteria. We study a different set of questions - focusing on the estimation and inference problem in the general two-sided and boundary models.

We briefly summarize the contributions of this paper as follows. First, we derive the large sample behavior of the likelihood ratio process and show that it approaches a simple, explicit function of a Poisson process that tracks the extreme (near-to-jump) events and depends on regressors in an interesting way. This limit result is useful for any inference that relies on the likelihood principle. The limit is useful since it can be easily simulated in order to evaluate the limit distributions of derived estimators and various likelihood-based statistics. To our knowledge, these results are new.

Second, we prove the consistency, derive the rates of convergence, and provide the limit distribution of the likelihood-based optimal estimators (BE) and MLE. The results are basic prerequisites for using these estimators in empirical work. More importantly, these results justify general Wald type inference based on limit distribution, subsampling, and parametric bootstrap.

Third, we show that posterior τ-quantiles are asymptotically $(1-\tau)$-quantile unbiased estimators of the true parameters. This property implies the validity of Bayes type confidence intervals based on the posterior quantiles. These confidence intervals provide valuable practical inference methods since they are simple to implement and require no detailed knowledge of asymptotic theory. This frequentist validity result is also of general theoretical interest because it covers models with complicated likelihoods where no finite dimensional sufficient statistics exist asymptotically, and its proof applies more generally to other problems. We further generalize this result to cover Bayes inference about general smooth functions of parameters, and show that it provides valid inference asymptotically.

Fourth, we briefly discuss how the well-known finite-sample (average-risk) optimality of Bayes procedures carries over to the limit.[5] The discussion is auxiliary and given here to prove some

[4]Hirano and Porter (2002) also derive the limit distributions of BE's for continuous covariate boundary models in a different form. (As Hirano and Porter (2002) note, the present treatment of continuous covariates appears to precede theirs somewhat.) An important difference is that our limit likelihood is stated in terms of a simple transform of a Poisson process, hence it is quite simple to simulate for inference purposes (which is our focus). Hirano and Porter (2002)'s limit is implicit, given as a process indexed by continuous covariate values. Consequently, their result is more suited for efficiency analysis (which is their focus), and its use for classical inference in practice may be infeasible without the (equivalent) Poisson type representations obtained in this paper. Also, the present paper focuses on inference in both one- and two-sided models.

[5]The optimality of Bayes estimates is treated in considerable details elsewhere in the literature. The recent contribution by Hirano and Porter (2002) provides a detailed limit-of-experiments analysis for a class of boundary models. Lehmann and Casella (1998) provides a basic discussion. van der Vaart (1999), Chapter 9.3-9.4, treats

of the previously stated results and for the justification of BE's, including the approximate and exact MLEs. The exact MLE, even when bias-corrected, generally does not coincide with optimal procedures asymptotically. But it is close to the approximate MLE defined as a BE under any loss function that approximates the delta function such as the 0-1 loss $I[|u| > \epsilon]/\epsilon$ or (truncated) p-th power losses.[6] Such loss functions penalize mistakes differently than the squared loss function. In that sense, the exact MLEs is approximately optimal under any loss function that approximate the delta function, and may perform better under the alternative loss functions than other likelihood procedures such as posterior means or posterior medians. Importantly, this implies that the MLE generally can not be dominated by any other *given* BE when the 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 her result. Thus, the MLE method, advocated by Donald and Paarsch (1993a) in the context of the discrete-covariate boundary models, provides a valuable method for estimation and inference in both the two-sided and one-sided regression models.[7]

Fifth, we show through simulation examples based on an empirical auction model from Paarsch (1992) that(1) particular BE's and MLEs work quite well and their relative performance depends critically on the measure of risk, and (2) the Bayes confidence intervals and the Wald confidence intervals based on the limit distributions perform as accurately as the Wald confidence intervals based on the parametric bootstrap, but are much less expensive computationally. The Bayes confidence intervals also produce the shortest confidence intervals among other methods. Thus, this paper justifies a whole array of useful and practical inference techniques, ranging from Wald type to Bayes type inference methods.

2. The Model, Examples, Assumptions, Procedures

This section describes the model and provides an informal discussion of the assumptions, results, and inference procedures developed in the later sections of the paper.

2.1. The Model. It is convenient to describe the class of models we consider in terms of a regression model where the errors have a discontinuous density. Let $(Y_i, X_i), i = 1, \ldots, n$, denote the random iid sample of size n generated by the model

$$Y_i = g(X_i, \beta) + \epsilon_i, \tag{2.1}$$

---

efficiency in the Uniform and Pareto models in the non-regression case. Ibragimov and Has'minskii (1981b), p.93 prove a fundamental result on the generic asymptotic efficiency of Bayes procedures under general, non-primitive (hard-to-verify) conditions.

[6]The approximations are Bayes procedures and are optimal in that regard, and hence can be a good substitute for exact MLE.

[7]Additional important and distinct properties of the MLE include (i) invariance to reparameterization and (ii) independence from prior information. Arguably, both properties are very useful.

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

$$
\begin{aligned}
&\lim_{\epsilon \uparrow 0} f(\epsilon|x, \beta, \alpha) = q(x, \beta, \alpha), \\
&\lim_{\epsilon \downarrow 0} f(\epsilon|x, \beta, \alpha) = p(x, \beta, \alpha), \\
&p(x, \beta, \alpha) > q(x, \beta, \alpha) + \delta, \quad \delta > 0, \quad \forall x \in \mathbf{X} = \text{support}(X), \quad \forall (\beta, \alpha) \in \mathcal{B} \times \mathcal{A}.
\end{aligned}
\tag{2.2}
$$

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 error distribution and α affects the shape of the error distribution only. We assume that $\beta \in \mathcal{B} \subset \mathbb{R}^{d_\beta}$ and $\alpha \in \mathcal{A} \subset \mathbb{R}^{d_\alpha}$. We also assume that the parameter set $\mathcal{G} = \mathcal{B} \times \mathcal{A}$ is compact and convex, and that the true parameter belongs to the interior of this set.

We consider two models: the one-sided model and the two-sided model. In the one-sided model, the conditional density jumps from zero to a positive constant. In the two-sided model, the conditional density jumps from one positive value to another. The one-sided model is a special 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. 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 parameter estimates in such models is approximated by that in the model with a density jump. The two-sided models also naturally arise in equilibrium search models, see e.g. Bowlus, Neumann, and Kiefer (2001).

The key feature of the regression model 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 non-regularities 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 is estimated at the rate n). On the other hand, inference about α is standard in many regards.

Note that classification of the model's parameters into α and β is motivated statistically, as in Donald and Paarsch (1993a) and van der Vaart (1999) (who considered univariate Pareto models). The boundary parameters β usually coincide with the main economic parameters, as indicated earlier. If they do not and the Wald type inference is to be used, then one needs to reparameterize them into α and β, see e.g. Donald and Paarsch (1993a).[8] However, the practical use of Bayes type inference or parametric bootstrap methods do not require such reparameterization.

In the following, we briefly review a structural example, which will serve to illustrate the plausibility of our regularity conditions and explain the results. It also provides an example for the Monte Carlo work.

Example: Independent Private Value Procurement Auction. Consider the following econometric model of an independent private value procurement auction, formulated in Paarsch (1992) and Donald and Paarsch (2002). Here, Y_i is the winning bid for auction i and the covariates $X_i = (Z_i, m_i)$ describes variation across auctions, where m_i denotes the number of bidders in the i-th auction minus 1, and Z_i denotes other observed characteristics of auctions.

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

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

where θ_2 and θ_1 are parameterized as functions of X and β (but this dependence is suppressed for notation convenience). E.g. $\theta_1(X,\beta) = \exp(\beta_1' Z)$ and $\theta_2(X,\beta) = \exp(\beta_2' Z)$.

Assuming the Bayesian 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 density function of the winning bid Y, which is the first order statistic generated by the specified bidding rule, conditional on covariates X:

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

[8] A referee pointed out the following example. Suppose $g(\theta) = \theta_1\theta_2$, where θ_2 also affects the shape of the error distribution. Although this example does not correspond to the economic model we used in the simulations, it highlights the important issue of reparameterization. In this case the asymptotic theory requires reparameterization into $\beta = \theta_1\theta_2$ and $\alpha = \theta_1$, and then the estimates of θ_1 and θ_2 are deduced from the estimates of α and β, and Wald type inference may be carried out using the Delta method (preferably based on the second order expansion, so that finite-sample estimation uncertainty about β is not neglected). E.g. for $\hat\theta_2 = \hat\beta/\hat\alpha$ $(\hat\theta_2 - \theta_2) \overset{d}{\sim} (\hat\beta - \beta)/\alpha - (\hat\alpha - \alpha)/\alpha^2 + 2(\hat\alpha - \alpha)^2/\alpha^3 \overset{d}{\sim} n^{-1}Z^\beta/\alpha + n^{-1/2}Z^\alpha/\alpha^2 + 2n^{-1}(Z^\alpha)^2/\alpha^3$, where Z^β and Z^α are the limit distributions of $\hat\beta$ and $\hat\alpha$. This expansion can be important because in finite samples, variability of estimates of $\hat\beta$ may be of comparable or larger order than that of $\hat\alpha$, motivating this expansion. Of course, one could use the first order Taylor expansion too, $(\hat\theta_2 - \theta_2) \overset{d}{\sim} -(\hat\alpha - \alpha)/\alpha^2 + o_p(1)$ but this approximation is less accurate.

Therefore, this is an example of a one-sided regression model (2.1) where

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

with

$$g(X, \beta) = \theta_1(X, \beta) \cdot \theta_2(X, \beta) \cdot (m-1) / (\theta_2(X, \beta)(m-1) - 1),$$

and ϵ_i has density $f(\epsilon|X, \beta) = f_Y(g(X, \beta) + \epsilon|X, \beta)$ conditional on X.

2.2. Regularity Conditions. The main regularity conditions **C0-C5** are collected in Appendix A. They serve to impose five basic types of assumptions:

(a) identification and compact, convex parameter space (with true parameters in the interior),
(b) continuous differentiability of the regression function $g(x; \beta)$ in β,
(c) nondegeneracy and boundedness of the vector $\partial g(X, \beta) / \partial \beta$,
(d) continuous differentiability and boundedness of the density function $f(\epsilon|x, \gamma)$, of its partial first derivatives in γ and ϵ, and of the second partial derivatives in γ(except at $\epsilon = 0$).
(e) continuous differentiability and integrability of the first and the second partial derivatives of $\ln f(\epsilon|x, \gamma)$ in γ.

Conditions of types (a) - (c) are standard in nonlinear likelihood analysis. Smoothness conditions of type (d) represent a generalization of the conditions of Ibragimov and Has'minskii (1981a). Conditions of type (e) are the standard conditions for regular smooth likelihood models, e.g. as in van der Vaart (1999), Chapter 7. Conditions of type (e) reflect that inference about α is standard if β is known.

These conditions are flexible enough to cover various auction models, frontier production function models, and equilibrium search models.[9]

2.3. Definitions of Estimation Procedures and Informal Overview of Results. Define the likelihood function as [10]

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

The optimal Bayes estimators are the likelihood-based estimators that minimize the average expected risk, where the risk is computed under different parameter values and then averaged over

---

[9]A technical addendum gives 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 $\{X_i, i \le n\}$. This term is omitted because this additional term does not affect the definition of the likelihood ratio or can be otherwise canceled out.

these parameter values. The procedures are generally of the following form:

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

where $\rho_n(\gamma) \equiv \rho(n\beta, \sqrt{n}\alpha)$ is a loss function, $\mu(\cdot)$ is the weight density function (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 the Bayes procedures carry over to the limit.[11]

The loss function ρ_n is made explicitly dependent on the sample size for purposes of asymptotic analysis, as in Ibragimov and Has'minskii (1981b), 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,
(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 deviation loss function,

Solutions of (2.4) with loss functions (A), (B), (C) generate BEs $\hat{\gamma}$ that are, respectively, (a) a vector of posterior means, (b) a vector of posterior medians (for each parameter component), (c) a vector of posterior τ-th quantiles.

Since BEs become very difficult to compute when ρ is not convex, we focus on convex loss functions for pragmatic reasons. However, proofs of the main results apply more generally to other loss functions specified in Ibragimov and Has'minskii (1981b). In practice, $\hat{\gamma}$ can be computed using Markov Chain Monte Carlo methods, which produce a sequence of draws

$$\left(\gamma^{(1)}, ..., \gamma^{(b)}\right), \tag{2.5}$$

whose marginal distribution is given by the posterior.Appropriate statistics of that sequence can be taken depending on the choice of ρ. (E.g. the means or component-wise medians for cases (A) and (B) above.) More generally, estimators $\hat{\gamma}$ are solutions of well defined globally convex differentiable optimization problems.[12]

The computational attractiveness of estimation and inference based on the Bayes procedures stems from the use of Markov Chain Monte Carlo (MCMC) and the statistical motivation of definition of the Bayes procedures. Since the Bayes estimates and the interval estimates 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,

[11] Furthermore, another motivation for (2.3) is that any optimal (admissible) estimation procedure is a Bayes procedure or a Bayes procedure with improper priors, cf. Wald (1950).

[12] Given the MCMC series (2.5) $\hat{\gamma}$ solves $\arg\inf_{\gamma \in \mathcal{G}} \frac{1}{b} \sum_{t=1}^{b} \rho_n\left(\gamma - \gamma^{(t)}\right)$, which is a globally convex and smooth (if ρ_n is smooth) optimization problem.

the computation of exact MLE requires optimization of a highly non-convex, discontinuous and otherwise highly nonlinear likelihood. MLE can be estimated by grid-based algorithms or MCMC only with an accuracy that worsens exponentially in the parameter dimension.

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

$$\widehat{\beta} - \beta = O_p\left(n^{-1}\right) \text{ and } \widehat{\alpha} - \alpha = O_p(n^{-1/2}). \tag{2.6}$$

The BE's are shown to converge in distribution to Pitman[13] functionals of the limit likelihood ratio process. We first develop a complete large sample theory of 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 type inference through either (a) simulation of the limit likelihood process, or (b) resampling techniques including subsampling and parametric bootstrap. Subsampling may be more robust than other methods under local misspecification of the parametric assumptions. However, the resampling methods are much more computationally expensive and require much more computational time than Bayes type inference. Simulating the limit distribution is comparable in terms of the computational expense to Bayes inference due to the linearity of the limit process.

An attractive practical alternative is the Bayes inference based on the posterior quantiles. Our results establish its large sample frequentist validity. Consider constructing a $\tau \times 100\%$ confidence intervals for $r_n(\gamma)$, where r_n is a smooth real function that possibly depends on n. Define the τ-th posterior quantile of the posterior distribution as

$$\widehat{c}(\tau) \equiv \underset{\tilde{r} \in \mathcal{R}_n}{\arg\inf} \int_{\mathcal{G}} \rho\left(\tilde{r} - r_n(\gamma); \tau\right) \frac{L_n(\gamma)\mu(\gamma)}{\int_{\mathcal{G}} L_n(\gamma')\mu(\gamma')d\gamma'} d\gamma, \tag{2.7}$$

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 taking the τth-quantile of the MCMC sequence evaluated at r_n

$$\left(r_n(\gamma^{(1)}), ..., r_n(\gamma^{(b)})\right). \tag{2.8}$$

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

$$\left[\widehat{c}(\tau/2), \widehat{c}(1-\tau/2)\right], \text{ where } \lim_{n\to\infty} P_{\gamma_0}\left\{\widehat{c}(\tau/2) \leq r_n(\gamma_0) \leq \widehat{c}(1-\tau/2)\right\} = \tau, \tag{2.9}$$

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

A pragmatic motivation for Bayesian intervals is that the empirical researcher does not need to have detailed knowledge of complex asymptotic limit theory to apply them. She 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.

[13] We follow the terminology of Ibragimov and Has'minskii (1981b) p. 21.

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 a limit of BEs under any sequence of loss functions that approximates the delta functions. We shall only briefly discuss the limit distribution of exact MLE for editorial reasons. A detailed analysis of MLE is given in the technical report, cf. Chernozhukov and Hong (2003). The MLE converges in distribution to a random variable that maximizes the limit likelihood ratio.

3. Large Sample Theory

This section contains the main formal results of the paper. Section 3.1 examines the large sample properties of the likelihood ratio function. Characterization of the limiting behavior of the likelihood is necessary for obtaining all of the main results and is useful for any likelihood based inference methojds. Section 3.2 provides an intuitive discussion of this result and subsequent results through an example. Section 3.3 describes the large sample properties of optimal Bayes estimators and both Wald and Bayes type inference procedures. Section 3.5 briefly discusses the limit theory of exact MLE.

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, e.g. van der Vaart (1999) and Knight (2000). After appropriate strengthening, the limit serves to describe the asymptotic distribution of all likelihood based estimators. Such an initial step is sometimes called the convergence of experiments, see van der Vaart (1999).

Consider the local likelihood ratio function

$$\ell_n(z) \equiv L_n(\gamma_n(\delta) + H_n z)/L_n(\gamma_n(\delta)),$$

where $\gamma_n(\delta) = \gamma_0 + H_n \delta$ denotes the true 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 results. The scaling by H_n corresponds to the convergence rates $\sqrt{n}$ for α and n for β.[14]

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

$$\left(\ell_n(z_j),\ \ j \leq k\right) \rightarrow_d \left(\ell_\infty(z_j),\ \ j \leq k\right), \tag{3.1}$$

and $\ell_\infty(\cdot)$ is called a finite-dimensional limit. In this section, $\rightarrow_d$ denotes convergence in distribution under $P_{\gamma_n(\delta)}$. We partition the localized parameter z accordingly into $z = (u', v')'$, where $u \in \mathbb{R}^{d_\beta}$

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

corresponds to the localized location parameters and $v \in \mathbb{R}^{d_\alpha}$ corresponds to the localized shape 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 $l_i(\gamma) = \ln f(\epsilon_i | X_i, \gamma)$,*

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

where $\mathcal{J} \equiv E_{P_{\gamma_0}}(\frac{\partial}{\partial\alpha} l_i(\gamma_0) \frac{\partial}{\partial\alpha} l_i(\gamma_0)')$, $\mathbf{m} \equiv E_{P_{\gamma_0}} \Delta(X)[p(X) - q(X)]$, $\mathbf{W} \stackrel{d}{=} N(0, \mathcal{J})$, and

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

[where Ibragimov and Has'minskii (1981b)'s convention applies to the case when $q(x) = 0$: $\ln 0 = -\infty$, $\ln \infty = \infty$ and $1/0 = \infty$, $\infty \cdot 0 = 0$, see equation (3.6) below].

$\mathbf{N}$ is a Poisson random measure $\mathbf{N}(\cdot) \equiv \sum_{i=1}^{\infty} 1[(J_i, \mathcal{X}_i) \in \cdot] + \sum_{i=1}^{\infty} 1[(J_i', \mathcal{X}_i') \in \cdot]$, where

$$J_i \equiv \Gamma_i / p(\mathcal{X}_i), \qquad \Gamma_i \equiv \mathcal{E}_1 + ... + \mathcal{E}_i, \qquad i \geq 1 \tag{3.3}$$

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

$\{\mathcal{X}_i, \mathcal{E}_i, i \geq 1\}$ is an iid sequence of variables where $\mathcal{X}_i$ follows law F_X, and $\mathcal{E}_i$ is a unit exponential variable. $\{\mathcal{X}_i', \mathcal{E}_i', i \geq 1\}$ is an independent copy of $\{\mathcal{X}_i, \mathcal{E}_i, i \geq 1\}$, and both sequences are independent of $\mathbf{W}$.

Remark 3.1 (***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

$$\begin{aligned} \sum_{i=1}^{\infty} l_u(J_i, \mathcal{X}_i) + \sum_{i=1}^{\infty} l_u(J_i', \mathcal{X}_i') \equiv & \sum_{i=1}^{\infty} \ln \frac{q(\mathcal{X}_i)}{p(\mathcal{X}_i)} 1\left[0 < J_i < \Delta(\mathcal{X}_i)' u\right] \\ & + \sum_{i=1}^{\infty} \ln \frac{p(\mathcal{X}_i')}{q(\mathcal{X}_i')} 1\left[0 > J_i' > \Delta(\mathcal{X}_i')' u\right], \end{aligned} \tag{3.5}$$

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 $\{\mathcal{X}_i, \mathcal{X}_i', J_i, J_i', i \leq b\}$ according to the distributions specified in Theorem 3.1 for large b, 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. This replacement is permissible for purposes of large sample inference, see e.g. Chernozhukov (2001).

Remark 3.2 (***Boundary 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

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

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

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

Thus in the one-sided models, the limit depends only on the set of variables in (3.3) and does not depend on the variables in (3.4).

Remark 3.3 (***Robustness to Misspecification***). It is not difficult to observe from the proof of Theorem 3.1 (hence Theorems 3.2–3.4) that the limit theory for β is robust under local misspecification of the regression function $g(x, \beta)$ of order $o(1/n)$, and local misspecification of the height of densities at the jump points $p(x, \gamma)$ and $q(x, \gamma)$ of order $o(1/n)$. 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 on $p(x, \gamma)$ and $q(x, \gamma)$ as long as $p(x, \gamma) > q(x, \gamma)$ for all x. Given these mentioned conditions hold, the inference about α appears to be robust up to local $o(1)$- violations of the information matrix equality for α. A formal development of these results is beyond the scope of this paper.

Theorem 3.1 extends the results of Donald and Paarsch (1993a) on the boundary models with discrete regressors and the results of Ibragimov and Has'minskii (1981a) on the univariate models. Despite its unusual form, the limit likelihood has a simple structure. 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 $\mathbf{W} = \mathcal{N}(0, \mathcal{J})$ 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

$$\mathcal{J}^{-1}\mathbf{W} = \mathcal{N}\left(0, \mathcal{J}^{-1}\right).$$

Because β is unknown, the limit likelihood also includes a nonstandard term $\ell_{2\infty}(u)$. The discontinuities in the density are highly informative about β and are of a local nature. A lot of information about β is contained in the observations Y_i that are near the location of the discontinuity $g(X_i, \beta)$, that is, for those Y_i such that

$$\epsilon_i = Y_i - g(X_i, \beta)$$

is close to zero. Thus the behavior of extreme (closest to zero) $\epsilon_i' s$ 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 that the behavior of likelihood estimators of β will be determined by $\ell_{2\infty}(u)$.

3.2. Informal Explanation Through an Example. Consider a simple model[15]

$$Y_i = X_i'\beta_0 + \epsilon_i, \quad \epsilon_i \stackrel{d}{=} \mathcal{E}, \tag{3.8}$$

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 the inference about α as it is regular as stated earlier). The model is a linearized, homoscedastic version of more realistic nonlinear models.

Intuitively, the smallest values of ϵ_i will be 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 $z = n(\beta - \beta_0)$, this constraint takes the form

$$n\epsilon_i \geq X_i'z, \text{ for all } i.$$

What we can learn about the parameter β_0 will depend on these constraints.

The likelihood for this example is $L_n(\beta) = \prod_i e^{-\epsilon_i + X_i'(\beta-\beta_0)} 1(n\epsilon_i \geq X_i'n(\beta - \beta_0))$. Hence the likelihood ratio $L_n(\beta)/L_n(\beta_0)$ as a function of $z = n(\beta - \beta_0)$ takes the form

$$\ell_n(z) = \prod_{i \leq n} \left(e^{-\epsilon_i + X_i'z/n} / e^{-\epsilon_i} \right) 1\left(n\epsilon_i \geq X_i'z\right),$$

which further reduces to

$$\ell_n(z) = e^{\bar{X}'z} 1(n\epsilon_i \quad \geq X_i'z, \text{ for all } i). \tag{3.9}$$

Since $\bar{X} \to_p EX$, the behavior of $\ell_n(z)$ for fixed z is determined by the lowest order statistics

$$n\epsilon_{(1)},\ n\epsilon_{(2)},\ n\epsilon_{(3)}, \ldots.$$

The Reny representation, see e.g. Embrechts, Klüppelberg, and Mikosch (1997) p. 189, allows these re-scaled order statistics to be represented almost surely as

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

---

[15] We thank a referee for suggesting using a similar example.

where $\{\mathcal{E}_1, \mathcal{E}_2, ..., \mathcal{E}_n\}$ is an iid sequence of unit-exponential variables. For given z, 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

$$\begin{aligned}(n\epsilon_{(1)}, n\epsilon_{(2)}, ..., n\epsilon_{(k)}) \to_d & (\mathcal{E}_1, \mathcal{E}_1 + \mathcal{E}_2, ..., \sum_{j=1}^{k} \mathcal{E}_j) \\ \equiv & (\Gamma_1, \Gamma_2, ..., \Gamma_k).\end{aligned}$$

Hence the marginal limit of $\ell_n(z)$ may be seen as

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

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

The result stated in Theorem 3.1 is more complicated for the following reasons:

1. In more general two-sided models, there is also an additional negative error 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".

2. The density of 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 functions $p(\mathcal{X}_i)$ or $q(\mathcal{X}_i')$.

3. 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 fully regular. 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. This follows from the standard proof of asymptotic independence of sample averages of general form and sample minimal order statistics, see e.g. Resnick (1986) for a general treatment and van der Vaart (1999) Lemma 21.19 for a simple example.

3.3. **Large Sample Properties of Bayes Procedures.** Given the above discussion, the following Theorem 3.2 can be easily conjectured. The Bayes estimator $Z_n = (n(\widehat{\beta} - \beta_n(\delta))', \sqrt{n}(\widehat{\alpha} - \alpha_n(\delta))')'$ centered at the true parameter $\gamma_n(\delta) = (\beta_n(\delta)', \alpha_n(\delta)')'$ and normalized by the convergence rates, is related to the localized likelihood ratio $\ell_n(z)$ as follows – it minimizes the posterior loss redefined in terms of the local deviation from the true parameter:

$$\Gamma_n(z) = \int_{\mathbb{R}^d} \rho(z - z') \pi_n(z') \, dz'.$$

Here $\pi_n(z)$ is the posterior density for the local deviation z from the true parameter:

$$\pi_n(z) = \ell_n(z) \mu(\gamma_n(\delta) + H_n z) / \int_{\mathbb{R}^d} \ell_n(\bar{z}) \mu(\gamma_n(\delta) + H_n \bar{z}) \, d\bar{z},$$

where $\ell_n(z)$ is the local likelihood ratio process and μ is the prior density. As $n \to \infty$, it can be conjectured from the discussion in the previous section that the posterior $\pi_n(z)$ approaches $\pi_\infty(z) \equiv \ell_\infty(z) / \int_{\mathbb{R}^d} \ell_\infty(z) \, dz$. The limit local posterior density π_∞ is a function of the likelihood only and does not depend on prior information.

Theorem 3.2 (***Properties of BEs***). *Suppose that the conditions of Theorem 3.1 and* **D1-D3** *hold. Then*

1. The convergence rate is n for estimating β and $\sqrt{n}$ for estimating α, i.e. $Z_n = O_p(1)$.

2. $Z_n \xrightarrow{d} Z$, where

$$Z \equiv \arg \inf_{z \in \mathbb{R}^d} \int_{\mathbb{R}^d} \rho(z - z') \frac{\ell_\infty(z')}{\int_{\mathbb{R}^d} \ell_\infty(\bar{z}) \, d\bar{z}} dz'. \tag{3.10}$$

3. If $\rho(z) = \rho_\beta(u) + \rho_\alpha(v)$, then $n(\widehat{\beta}_n - \beta_n(\delta)) \to_d Z^\beta \equiv \arg\inf_u \int_{\mathbb{R}^{d_\beta}} \rho_\beta(u - u') \ell_{2\infty}(u') \, du'$ and $\sqrt{n}(\widehat{\alpha}_n - \alpha_n(\delta)) \to_d Z^\alpha \equiv \arg\inf_v \int_{\mathbb{R}^{d_\alpha}} \rho_\alpha(v - v') \ell_{1\infty}(v') \, dv'$, and Z^β and Z^α are independent.

Theorem 3.2 obtains the consistency and establishes the rates of convergence and the limit distributions of the BEs. The limit is given in the form of a *Pitman functional* of a limit likelihood and is not difficult to simulate using MCMC methods according to Remark 3.1. The result also justifies the use of the parametric bootstrap, cf. Remark 3.5.

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 3 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 (by Anderson's lemma, see van der Vaart (1999)), i.e. the limit distribution of $\widehat{\alpha}$ is given by

$$\mathcal{N}(0, \mathcal{J}^{-1}).$$

This is not the case for the estimators of the location parameter β. Furthermore, as shown below the optimal estimators generally are not transformations of the MLE asymptotically, contrary to the non-regression or dummy regression cases.

Remark 3.4 (***Wald Inference with Subsampling***). Theorem 3.2 immediately justifies the validity of subsampling for Wald type inference. Subsampling approximates the distribution of the estimator in the full sample based on values of this estimator in many smaller subsets of data. Implementation protocols are standard and can be found in Politis, Romano, and Wolf (1999). Theorem 2.2.1 in Politis, Romano, and Wolf (1999) applies provided (i) the estimates are consistent at polynomial in n rates, (ii) the estimates posses a limit distribution. Both of these conditions are proven in Theorem 3.2. Thus, Theorem 3.2 immediately implies the validity of inference based on subsampling. Subsampling may not be as high quality as the parametric bootstrap or simulation of the limit. However, subsampling is (a) computationally less demanding than the parametric bootstrap and (b) is likely to be more robust than other methods to local misspecification of parametric models that change the parameters of the limit distribution but do not affect the rates of convergence.

Remark 3.5 (***Wald Inference with Parametric Bootstrap***). As in Ibragimov and Has'minskii (1981a), the weak convergence results and the proof can be stated uniformly in the parameter γ, and conditional on almost every realization of the covariate sequence $\{X_i, i \leq n\}$ $(n \to \infty)$. In order to do so, the notation must be made more complicated in a manner similar to Ibragimov and Has'minskii (1981a) to denote the dependence of the limit on the parameter γ and on the realization of the covariate sequence. The uniform convergence in distribution is defined as the convergence of distributions under the Levy metric uniformly in the parameter γ, and conditional on covariate sequences. This immediately implies that the parametric bootstrap is valid in the usual sense that the bootstrap distribution converges to the limit distribution in probability under the Levy metric as long as the preliminary estimate $\widehat{\gamma} \to_p \gamma$, conditional on covariate samples. Any initial consistent estimator $\widehat{\gamma}$ may be used. Hence the parametric bootstrap can be used for Wald type inference based on the point estimates (See for example Horowitz (2000)). Although the Bayes estimates are not difficult to recompute (especially with a good starting value such as the initial Bayes estimate $\widehat{\gamma}$), the parametric bootstrap appears to be very expensive computationally. As discussed in Section 4, it is much more computationally demanding than any other method. The parametric bootstrap may not be robust against the local misspecification of the parametric models.

Next consider the posterior mean $\bar{\gamma}$ and posterior quantile $\widehat{\gamma}(\tau)$ as the solutions to the problem (2.4) under the squared loss and the check loss functions, respectively (each defined in Section 2). Also define $\bar{Z}$ and $Z(\tau)$ as the solutions of the limit problem (3.10) under the squared and the check functions, respectively.

Theorem 3.3 (***Mean-Unbiasedness, Quantile-Unbiasedness, Posterior Confidence Intervals***). *Under the conditions of Theorem 3.2*
1. Posterior mean estimators are asymptotically mean-unbiased:

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

2. Consider any $0 < \tau' < \tau'' < 1$. *If* $Z(\tau)$ *has positive density in the neighborhood of* 0, *for* $\tau = \tau'$ *and* $\tau = \tau''$, *then posterior* τ*-quantiles are* $1 - \tau$*-quantile unbiased:*

$$\lim_{n\to\infty} P_{\gamma_n(\delta)}\left\{(\widehat{\gamma}(\tau))_j \le (\gamma_n(\delta))_j\right\} = P_{\gamma_0}\left\{(Z(\tau))_j \le 0\right\} = 1 - \tau, \tag{3.11}$$

where $(a)_j$ *denotes the* j*-th components of vector* a. *Hence*

$$\lim_{n\to\infty} P_{\gamma_n(\delta)}\left\{(\widehat{\gamma}(\tau'))_j \le (\gamma_n(\delta))_j \le (\widehat{\gamma}(\tau''))_j\right\} = \tau'' - \tau'. \tag{3.12}$$

A very useful implication of the quantile unbiasedness result is the validity of confidence intervals $[(\widehat{\gamma}(\tau'))_j, (\widehat{\gamma}(\tau''))_j]$ for large sample inference on parameter components $(\gamma)_j$.

Results 1 and 2 follow from the asymptotic optimality of posterior means and quantiles respectively under the squared and check function losses, which are defined and established in section 3.4. For example, if the limit posterior mean Z had a mean $EZ = c \ne 0$, then the estimator $\bar{\gamma} + H_n c$ would have a strictly lower asymptotic risk regardless of the local parameter sequence. Hence it must be that $E\bar{Z} = 0$. A similar argument applies to the τth-posterior quantile. The τ-posterior quantile is $1 - \tau$-quantile unbiased because it is asymptotically optimal under the τ-check loss function. The requirement that $Z_j(\tau)$ has positive density around 0 is technical.

The next result concerns the asymptotic validity of the posterior quantiles $\widehat{c}(\tau)$ for inference about smooth functions of the parameters. Consider inference about the function $r_n(\gamma)$ where $r_n : \mathbb{R}^{d_\alpha + d_\beta} \to \mathbb{R}$ is such that for $a > 1$ and $R \equiv [\dot{R}', \ddot{R}']'$ with rank $R = 1$:

$$r_n(\gamma) - r_n(\gamma_0) = \dot{R}'(\alpha - \alpha_0)\sqrt{n} + \ddot{R}'(\beta - \beta_0)n + O\left(n|\beta - \beta_0|^a + \sqrt{n}|\alpha - \alpha_0|^a\right). \tag{3.13}$$

For purposes of theoretical analysis, the function is made dependent on n specifically to have a better finite-sample approximation through the avoidance of the trivial case where all of the asymptotic inference is determined by either parameter α or β due to the difference in rates of convergence. If a smooth function $m(\beta)$ is of prime interest, taking $r_n(\gamma) = n \cdot m(\beta)$ fulfills condition (3.13). If a smooth function $m(\alpha)$ is of interest, then taking $r_n(\gamma) = \sqrt{n} \cdot m(\alpha)$ also fulfills condition (3.13). Note that these transformations by $\sqrt{n}$ or n do not affect the practical formulations (2.7) - (2.9) in Section 2.3 by the linearity of transformations and equivariance of quantiles to monotone transformations.

Theorem 3.4 (***Inference with Posterior Quantiles***). *Under the conditions of Theorem 3.2*
1. For any $0 < \tau < 1$, $(\widehat{c}(\tau) - r_n(\gamma_n(\delta))) \to_d \widetilde{Z}(\tau)$, *where*

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

2. Provided $\widetilde{Z}(\tau)$ *has positive density over a neighborhood of* 0 *for* $\tau = \tau'$ *and* $\tau = \tau''$

$$\lim_{n\to\infty} P_{\gamma_n(\delta)}\left\{\widehat{c}(\tau) \le r_n(\gamma_n(\delta))\right\} = P_{\gamma_0}\left\{\widetilde{Z}(\tau) \le 0\right\} = 1 - \tau, \tag{3.14}$$

and

$$\lim_{n\to\infty} P_{\gamma_n(\delta)}\left\{\widehat{c}(\tau') \le r_n(\gamma_n(\delta)) \le \widehat{c}(\tau'')\right\} = \tau'' - \tau'. \tag{3.15}$$

Theorem 3.4 generalizes Theorem 3.3 to more inference about more general functions than construction of parameter confidence intervals. For example, a $1-\tau$-level asymptotic test of the null hypothesis $r_n(\gamma_0) = r_n$ is given by the decision rule that rejects the null if $r_n \notin [\widehat{c}(\tau/2), \widehat{c}(1-\tau/2)]$, where (3.15) can be used to deduce the local power and consistency of the test.

3.4. **Optimality.** Lemma 3.1 below briefly records the finite-sample and asymptotic average risk optimality properties of BE's, which is needed only for the auxiliary purposes of proving Theorems 3.3 and 3.4. A detailed valuable analysis of (minimax) optimality in non-regular models can be found in Hirano and Porter (2002).

Define the normalization matrix H_n as in section 3.1, and let $\gamma_n(\delta) \equiv \gamma_0 + H_n\delta$, $\delta \in \mathbb{R}^d$, denote the local parameter sequence. Consider the set Υ_n of all statistics (measurable mappings of data) $\widehat{\gamma}_n$. Define the expected risk associated with a loss function ρ and estimator $\widehat{\gamma}_n$ as $E_{P_{\gamma_n(\delta)}}\rho(\widehat{Z}_n)$, 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 $\widehat{\gamma}$ is given by:

$$\frac{1}{\lambda(K)}\int_K E_{P_{\gamma_n(\delta)}}\rho(\widehat{Z}_n)\mu(\gamma_n(\delta))d\delta, \tag{3.16}$$

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

$$\limsup_{K\uparrow\mathbb{R}^d}\limsup_{n\to\infty}\frac{1}{\lambda(K)}\int_K E_{P_{\gamma_n(\delta)}}\rho(\widehat{Z}_n)d\delta, \tag{3.17}$$

where $K\uparrow\mathbb{R}^d$ denotes an increasing sequence of cubes centered at the origin and converging to $\mathbb{R}^d$. Compared to the previous formula, the weight μ is replaced by the objective (uninformative) weight over $\mathbb{R}^d$.

Lemma 3.1. *Suppose the conditions of Theorem 3.2 hold. For $\widehat{\gamma}_{\rho,\mu,n} \in \Upsilon_n$ denoting the Bayes estimator under loss ρ and prior weight μ, $Z_n \equiv H_n^{-1}[\widehat{\gamma}_{\rho,\mu,n} - \gamma_n(\delta)]$, $U_n \equiv n(\mathcal{B}-\beta_0)\times\sqrt{n}(\mathcal{A}-\alpha_0)$*

1. *For each $n \ge 1$ the infimum of finite sample average risk for $K = U_n$ is achieved over Υ_n by the Bayes estimator $\widehat{\gamma}_{\rho,\mu,n}$, i.e. at $\widehat{Z}_n = Z_n$ in (3.16).*
2. *The infimum of asymptotic average risk over estimator sequences in Υ_n equals $E_{P_{\gamma_0}}\rho(Z) < \infty$ and is attained by the sequence of the Bayes estimators $\widehat{\gamma}_{\rho,\mu,n}$, i.e. $\{\widehat{Z}_n\} = \{Z_n\}$ in (3.17). (Z denotes the weak limit of Z_n).*

Statement 1 is a basic result of statistics, that the optimal estimator under loss ρ is a Bayes estimator defined by the risk-weighting function μ and a loss function ρ, cf. Wald (1950) and

Lehmann and Casella (1998), Chapter 5. Statement **1** is often simply used as an alternate definition of the Bayes (optimal) procedures. Statement **2** translates finite-sample efficiency into asymptotic average risk efficiency (this result essentially follows from Ibragimov and Has'minskii (1981b) p.93).

It is critical that unlike in the regular case, the efficiency rankings are largely determined by the loss function ρ. For example, MLE may be worse than the posterior mean under the squared loss, but performs better under other loss functions, cf. Section 4.

3.5. **Large Sample Theory of Maximum Likelihood Procedures.** We provide only a brief discussion of the MLE. Consider the MLE $Z_n \equiv \left(Z_n^{\beta\prime}, Z_n^{\alpha\prime}\right)' \equiv \left(n(\widehat{\beta} - \beta_n(\delta))', \sqrt{n}(\widehat{\alpha} - \alpha_n(\delta))'\right)'$, which is 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*

$$Z_n \to_d Z \equiv \operatorname{arginf}_{z\in\mathbb{R}^d} - \ell_\infty(z).$$

In particular, $Z_n^\alpha \to_d Z^\alpha = \mathcal{J}^{-1}\mathbf{W} \overset{d}{=} \mathcal{N}(0, \mathcal{J}^{-1})$, $Z_n^\beta \to_d Z^\beta = \operatorname{argmin}_{u\in\mathbb{R}^{d_\beta}} - \ell_{2\infty}(u)$, and Z^β and Z^α are independent.

The proof is given in the technical report (Chernozhukov and Hong (2003)). The limit variable is an argmin of a limit likelihood, which inherits the discontinuities of the finite sample likelihood. Due to asymptotic independence of the information about the shape parameter from the information about the location parameter, the MLEs for these parameters are asymptotically independent. In the boundary models, the limit result can be stated more explicitly for β as follows:

$$\begin{aligned} n(\widehat{\beta} - \beta_n(\delta)) \to_d Z^\beta &\equiv \arg\inf_u \Big(-\exp(u'\mathbf{m}) \quad \text{such that } J_i \geq \Delta(\mathcal{X}_i)' u, \text{for all } i \geq 1\Big), \\ &= \arg\sup_u \Big(\quad u'\mathbf{m} \qquad \text{such that } J_i \geq \Delta(\mathcal{X}_i)' u, \text{for all } i \geq 1\Big). \end{aligned}$$

This result generalizes the results of Donald and Paarsch (1993a) and Smith (1994). Note that the solutions of the linear programs like these are unique under fairly weak conditions. [16]

Remark 3.6 (***Asymptotic Non-Sufficiency of MLE's***). It is important to note here that the posterior means and medians are generally not equal to the bias corrected MLE. Consider the example of Section 3.2 where

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

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

[16]For example, a very simple sufficient condition for almost sure uniqueness is that there is one continuously distributed element in $\Delta(\mathcal{X}_i)$ and that $\Delta(\mathcal{X}_i)$ has a nondegenerate distribution, cf. Portnoy (1991) for a related problem. When $\Delta(\mathcal{X}_i)$ has discrete support, the stated limit result coincides with the result of Donald and Paarsch (1993a) who show that uniqueness holds if $\Delta(\mathcal{X}_i)$ has nondegenerate distribution (assumed in **C3**).

MLE $\widehat{Z}$ maximizes over z such that $\Gamma_i \geq z$, thus $\widehat{Z} = \min\{\Gamma_i, i \leq n\}$, $\ell_\infty(z) = e^z 1(z \leq \widehat{Z})$, implying sufficiency of $\widehat{Z}$. If $\widehat{Z}$ is sufficient then the limit optimal Bayes estimators 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

$$\ell_\infty(z) \neq e^{E(X)'z} 1(\mathcal{X}_i'z \leq \mathcal{X}_i'\widehat{Z}, \text{for all } i \geq 1) \text{ with strictly positive probability,}$$

implying that $\widehat{Z}$ is not sufficient for $\ell_\infty(z)$ even conditional on covariates. Thus, the limit likelihood-based Bayes estimators Z are generally not nonrandom functions of the limit MLE $\widehat{Z}$.

4. Computational Experiments

4.1. Monte Carlo Design. We used a simple 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 practical 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 the 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 on p. 245 in Robert and Casella (1998).[17]

4.2. Quality of Estimation Procedures. We compare the performance of

(1) the posterior median,
(2) the posterior mean,
(3) the posterior mode (MLE),

across different risk measures. Here the MLE is computed by taking the argmax over the grid generated by the MCMC sequence.[18]

The results given in Table 1 and Table 2 show that: (a) the posterior median is the best under the mean absolute deviation loss, (b) the posterior mean is the best under the mean squared loss, (c) the MLEs do better under the mean 10-th power loss function. Thus, it appears that all of the likelihood procedures perform quite well relative to some risk measure.

[17]In the 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 estimates. The C++ implementation is available from the authors.

[18]We also tried the approximate MLE defined as a BE under (truncated) 10-th power loss function but the performance of the approximate and exact MLEs coincided up to many digits and is not reported separately. Other loss functions that approximate delta function can also be used to approximate MLEs by some BE.

4.3. **Quality of Inferential Procedures.** In the next step, we compare the performance of several inference methods, focusing on the coverage properties of the confidence intervals. We compare

(1) the confidence intervals based on the posterior quantiles,
(2) the percentile confidence intervals based on the parametric bootstrap of the posterior mean,
(3) the percentile confidence intervals based in the parametric bootstrap of the MLE,
(4) the percentile confidence intervals based on the subsampling of the posterior mean (using $1/4 \times n$ as the subsample size),
(5) the simulation of the limit distribution of MLE and other estimators as described in Remark 3.1 (using $b = n$).

Results reported in Table 3 indicate that the intervals based on the posterior quantiles and simulation of the limit distribution perform nearly as well as the parametric bootstrap, while subsampling performs worse than any of them. The confidence intervals based on the posterior quantiles also appear to be the shortest on average. Given that the posterior intervals are the least expensive to compute, they should be preferred. The subsampling is less expensive than the parametric bootstrap and is probably more robust for inference purposes under local misspecification.

In terms of computational expense, computation of the posterior quantiles takes less than 1 minute on a Pentium III PC. Simulation of the limit distribution is roughly twice as expensive (because the limit expressions are simple transformations of linear functions and do not contain nonlinear expressions). We used 200 bootstrap draws and the full sample estimate as the starting value 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 takes about 1/5-th of the time of the parametric bootstrap, using the same number of draws. The entire Monte Carlo work took several weeks of computer time.

5. Conclusion

We studied estimation and inference 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 interesting economic models discussed in the recent literature of structural estimation. We derived the large sample theory of a variety of likelihood based procedures, and offered an array of useful and practical inference techniques, including Wald type and Bayes type inference methods. The results provide a theoretical and practical solution to an important econometric problem.

Appendix A. Regularity Conditions C0-C5 and D1-D3

Notation. Throughout the paper, c and const denote generic positive constants unless stated otherwise; $\to_p$ and $\to_d$ denote convergence in probability and distribution, respectively; $\|x\|$ is the usual Euclidian norm $\sqrt{x'x}$, and $|x|$ is used to denote the supremum norm , i.e. $|x| = \sup_{j\leq k}|x_j|$ where $x = (x_1, ..., x_k)$. Note the densities of interest $f(\epsilon|x,\gamma)$ 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 the usual partial derivative when $\epsilon \neq 0$, and also use it to denote 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 with radius δ as measured by $|\cdot|$.

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

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

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

C2 Without loss of generality, the density $f(\epsilon|x,\gamma)$ is upper-semicontinuous at $\epsilon = 0$ for each x and γ. The density $f(\epsilon|x,\gamma)$ is bounded from above uniformly in (ϵ, x, γ). $f(\epsilon|x,\gamma)$ has continuous first partial derivative in ϵ (except at $\epsilon = 0$) that is bounded uniformly in (ϵ, x, γ); $f(\epsilon|x,\gamma)$ has continuous first and second partial derivative in γ that is bounded uniformly in (ϵ, x, γ). The density and the derivatives specified above are continuous in x on $\mathbf{X}$ for each ϵ and γ. Lastly, $\sup_\gamma E_X \int |\frac{\partial}{\partial\bar{\gamma}} f(y - g(X,\beta)|X;\gamma)|dy < \infty$.

C3 The function $g(x,\beta)$ has two continuous and bounded derivatives in β, uniformly in x and β, and $E[\frac{\partial g(X;\beta)}{\partial\beta}\frac{\partial g(X;\beta)'}{\partial\beta}]$ is positive definite uniformly in β. The function and the specified above derivatives are continuous in x on $\mathbf{X}$ for each β.

C4 When the nuisance parameter α is present, for $l_i(\bar{\gamma}) \equiv \ln f\left(Y_i - g\left(X_i, \bar{\beta}\right)|X_i, \bar{\gamma}\right)$ where $\bar{\gamma} = (\bar{\beta}, \bar{\alpha})$, uniformly in $\bar{\gamma}$ and γ either (a) $E_{P_\gamma}[\frac{\partial}{\partial\bar{\gamma}} l_i(\bar{\gamma}) \frac{\partial}{\partial\bar{\gamma}} l_i(\bar{\gamma})']$ is positive definite and bounded or (b) if $\frac{\partial}{\partial\beta} l_i(\gamma) = 0$ P_γ-a.s., then $E_{P_\gamma}[\frac{\partial}{\partial\bar{\alpha}} l_i(\bar{\gamma}) \frac{\partial}{\partial\bar{\alpha}} l_i(\bar{\gamma})']$ is positive definite and bounded.

C5 In the two-sided model C1.i, the terms

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

are bounded respectively by $C_j(\epsilon_i, X_i)$, $j = 1,2,3$, for all $Y_i - g(X_i,\bar{\beta}) \in \mathbb{R}\setminus\{0\}$, uniformly in $\bar{\gamma} \in B_\delta(\gamma_0)$, where $\sup_\gamma E_{P_\gamma} C_j(\epsilon_i, X_i) < \infty$ 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 $\bar{\gamma} \in B_\delta(\gamma_0)$, where $\sup_\gamma E_{P_\gamma} C_j(\epsilon_i, X_i) < \infty$ for $j = 1,2,3$.

Lemma A.1 (*Important Constants*). *The conditions C0-C3 imply that there are finite constants $\bar{f}$, $\bar{f}'$, $\bar{f}''$, $\underline{f}$, $\bar{g}$, $\bar{g}'$, $\bar{g}''$ such that*

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

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

and also that for some $\delta > 0$ and $V(0) = [-\delta, \delta] \setminus \{0\}$ in case of C1(i) and $V(0) = (0, \delta]$ in case of C1(ii)

$$\inf_{\epsilon \in V(0), x \in \mathbf{X}, \gamma \in \mathcal{G}} f(\epsilon | x, \gamma) > \underline{f} > 0.$$

Remark A.1. The regularity conditions are summarized in Section 2.2. The parameters γ include the location parameters β and the shape parameters α. If β is known, the inference about α is regular. Thus, the conditions **C0-C5** are a mixture of non-regular and regular assumptions, as in Ibragimov and Has'minskii (1981a) and van der Vaart (1999), Ch. 7. 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 **C3-C5** are common in nonlinear analysis.

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:

D1 $\mu(\cdot) > 0$ is continuous on $\mathcal{G}$,
D2 $\rho(\cdot) \geq 0$ and $\rho(z) = 0$ iff $z = 0$, ρ is convex,
D3 $\rho(z)$ is dominated by a polynomial of $|z|$ as $|z| \longrightarrow \infty$.

Remark A.2. These are standard assumptions on the loss function ρ and the prior μ, see for example Ibragimov and Has'minskii (1981b). Since BE's become essentially uncomputable when ρ is not convex, we do not consider the non-convex loss functions for pragmatic reasons. However, the proof of Theorems 3.1-3.4 do not rely upon the convexity assumption and the results apply more generally to other loss functions specified in Ibragimov and Has'minskii (1981b).

Appendix B. Proofs for Section 3

[N.B. In the proofs we extensively use the constants defined in Lemma A.1]

B.1. **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 (1981a), 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 piece-wise 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

$$Q_\infty^c(z) = \mathbf{W}'v - \frac{1}{2} v' \mathcal{J} v + \mathrm{m}' u, \quad Q_\infty^d(z) = \int_{\mathbf{R} \times \mathbf{X}} l_u(j, x) d\mathbf{N}(j, x),$$

where each term is defined in Theorem 3.1. Given this result, the finite-dimensional limit of $\ell_n(z)$ is

$$\ell_\infty(z) = \exp\left(Q_\infty(z)\right).$$

For $z = (u', v')'$, using that $\epsilon_i = Y_i - g(X_i, \beta_0)$,

$$
\begin{aligned}
Q_n(z) \equiv & \underbrace{\sum_{i=1}^{n} \widehat{r}_{in}(z) \ \times [1(\epsilon_i > \{\Delta_n(X_i, u)/n\} \vee 0) + 1(\epsilon_i < \{\Delta_n(X_i, u)/n\} \wedge 0)]}_{Q^c_{1n}(z)} \\
& \underbrace{+ \sum_{i=1}^{n} (\widehat{r}_{in}(z) - r_{in}(z)) \times [1(0 < \epsilon_i \leq \Delta_n(X_i, u)/n) + 1(0 > \epsilon_i \geq \Delta_n(X_i, u)/n)]}_{Q^c_{2n}(z)} \\
& \underbrace{+ \sum_{i=1}^{n} r_{in}(z) \ \times [1(0 < \epsilon_i \leq \Delta_n(X_i, u)/n) + 1(0 > \epsilon_i \geq \Delta_n(X_i, u)/n)]}_{Q^d_n(z)} \\
& \equiv Q^c_n(z) + Q^d_n(z), \text{ where} \\
& \widehat{r}_{in}(z) \equiv \ln \left[\frac{f(Y_i - g(X_i, \beta_0 + u/n) \,|X_i, \beta_0 + u/n, \alpha_0 + v/\sqrt{n})}{f(Y_i - g(X_i, \beta_0) \,|X_i, \gamma_0)} \right] \\
& \equiv \ln \left[\frac{f(\epsilon_i - \Delta_n(X_i, u)/n)|X_i, \beta_0 + u/n, \alpha_0 + v/\sqrt{n})}{f(\epsilon_i | X_i, \gamma_0)} \right], \\
& r_{in}(z) \equiv \ln \left[\frac{q(X_i)}{p(X_i)} \right] 1(0 < \epsilon_i) + \ln \left[\frac{p(X_i)}{q(X_i)} \right] 1(0 > \epsilon_i), \\
& \Delta_n(x, u) \equiv n(g(X_i, \beta_0 + u/n) - g(X_i, \beta_0)).
\end{aligned}
$$

The convergence analysis of the continuous part $Q^c_n(z)$ is standard. In sharp contrast, behavior of discontinuous part $Q^d_n(z)$ differs from that of $Q^c_n(z)$, and is analyzed using the point process methods.

Also, in above expressions and all proofs we use the algebraic rules of Ibragimov and Has'minskii (1981a) for working with ∞'s defined in Theorem 3.1. This is done to include the proof for the boundary model as a special case. In particular, the expressions involving $1(0 > \epsilon_i > \ldots)$ cancel, since in the boundary models $\epsilon_i > 0$. Also in the boundary model $\widehat{r}_{in}(z) \equiv r_{in}(z) = -\infty$ when $0 < \epsilon_i \leq \Delta_n(X_i, u)/n$, so that

$$
Q^c_{2n}(z) \equiv 0.
$$

Thus, the term $Q^c_{2n}(z)$ is only non-zero for the two-sided model. Further details follow.

Part I obtains the finite-dimensional limit of $Q^c_n(z)$. 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 iid, followed by application of the Markov LLN and Chebyshev inequality, yields for a given z [see Addendum]

$$
Q^c_n(z) \equiv \underbrace{-u' E\Delta(X_i) \frac{f'(\epsilon_i|X, \gamma_0)}{f(\epsilon_i|X_i, \gamma_0)}}_{\equiv u' \, \mathbf{m}} + v' \underbrace{\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \alpha} \ln f(\epsilon_i|X_i, \gamma_0) \right]}_{\equiv \mathbf{W}_n} + \frac{1}{2} v' \underbrace{\left[E \frac{\partial^2}{\partial \alpha \partial \alpha'} \log f(\epsilon|X, \gamma_0) \right]}_{\equiv -\mathcal{J}} v + o_p(1),
$$

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 $\mathbf{W}_n \to_d \mathbf{W} = \mathcal{N}(0, \mathcal{J})$. Also it follows by **C2**[19]

$$\mathbf{m} = E\Delta(X_i)(p(X_i) - q(X_i)).$$

Therefore, the finite-dimensional limit of $Q^c_{1n}(z)$ is given by

$$Q^c_\infty(z) \equiv u'\mathbf{m} + \mathbf{W}'v - \frac{1}{2}v'\mathcal{J}v.$$

It remains to show $Q^c_{2n}(z) = o_p(1)$. In the one-sided case $Q^c_{2n}(z) \equiv 0$, hence consider the two-sided case. Note that by assumptions **C1-C3**, for any compact set $\mathbf{Z}$, as $n \to \infty$,

$$\left|\ln\left[\frac{f(\epsilon - \Delta_n(x,u)/n|x, \beta_0 + u/n, \alpha_0 + v/\sqrt{n})}{f(\epsilon|x,\gamma_0)}\right] - \ln\left[\frac{q(x)}{p(x)}\right]\right| \le 2 \times (\bar{f}'/\underline{f}) \times \bar{g}'\|z\|/\sqrt{n} \tag{B.1}$$

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

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

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

$$\sup_{z\in\mathbf{Z}} |Q^c_{2n}(z)| \le 2 \times (\bar{f}'/\underline{f}) \times \bar{g}' \times \|\mathbf{Z}\|/\sqrt{n} \times \sum_{i=1}^n 1(|\epsilon_i| < K/n) = O_p(1/\sqrt{n}) \tag{B.3}$$

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

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

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

$$Q^d_n(z) \equiv \sum_{i=1}^n \left[\ln\frac{q(X_i)}{p(X_i)}1(0 < n\epsilon_i \le \Delta_n(X_i,u)) + \ln\frac{p(X_i)}{q(X_i)}1(0 > n\epsilon_i \ge \Delta_n(X_i,u)\right].$$

By **C2** and **C3**

$$\begin{aligned} E\sum_{i=1}^n &\Big|1(0 < n\epsilon_i \le \Delta_n(X_i,u)) - 1(0 < n\epsilon_i \le \Delta(X_i)'u)\Big| \\ &+ \Big|1(0 > n\epsilon_i \ge \Delta_n(X_i,u)) - 1(0 > n\epsilon_i \ge \Delta(X_i)'u)\Big| \le 2\bar{f}\bar{g}''\|u\|^2/n = o(1), \end{aligned}$$

where $\Delta(X_i) \equiv \frac{\partial g(X_i,\beta_0)}{\partial\beta}$, which implies that for given z

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

Now note that $(Q^d_n(z_j), j \le l)$ and $(Q^c_n(z_j), j \le l)$, for any finite l, are asymptotically independent. This follows by applying a standard argument concerning the independence of minimal order statistics and sample averages of general form, see for example Resnick (1986) or Lemma 21.19 in van der Vaart (1999) [Addendum provides the proof for completeness].

[19]Recall that for a density function f, which is everywhere continuously differentiable except at 0 with an integrable derivative, $\int_{\mathbb{R}} f'(u)du = -f(0^+) + f(0^-)$.

The next step is to obtain the finite-dimensional limit of Q_n^d. The behavior of Q_n^d is determined by the near-to-jump observations, whose behavior is described using a point process. 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: 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

$$\widehat{\mathbf{N}}(A) = \sum_{i=1}^{n} 1\left[(n\epsilon_i, X_i) \in A\right],$$

$\widehat{\mathbf{N}}$ is a random element of $M_p(E)$, the metric space of nonnegative point measures on E, with the metric generated by the usual topology of vague convergence, Resnick (1987) Ch. 3. [Technical Addendum provides a review]. We show that

$$\widehat{\mathbf{N}} \Rightarrow \mathbf{N} \text{ in } M_p(E),$$

for $\mathbf{N}$ given in Theorem 3.1. This is done in 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), $\lim_{n\to\infty} E\widehat{\mathbf{N}}(F) \equiv \lim_{n\to\infty} nP((n\epsilon_i, X_i) \in F)$

$$= \int_F [p(x)1(u > 0)du + q(x)1(u < 0)du] \times dF_X(x) = m(F) < \infty, \tag{B.5}$$

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

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

Statements (B.5) and (B.6) imply by Kallenberg's Theorem – 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 $m_2(du, dx) = [1(u > 0)du + 1(u < 0)du] \times F_X(dx)$ 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 $T : (u, x) \mapsto (1(u > 0)u/p(x) + 1(u < 0)u/q(x), x)$ is a Poisson process with the desired mean measure $m(du, dx) = m_2 \circ T^{-1}(du, dx) = [p(x)1(u > 0) + q(x)1(u < 0)]du \times F_X(dx)$, by Proposition 3.7 in Resnick (1987).

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

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

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

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

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 two-sided model: By conditions **C1-3**, 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 $N \mapsto (\int_E l_{u_k}(j,x) dN(j,x), k \le l)$ for $l < \infty$. Hence by Proposition 3.13 in Resnick (1987) T is discontinuous at $\mathcal{D}(T) \equiv \{N \in M_p(E) : j_i^N = 0 \text{ or } j_i^N = u_k'\Delta(x_i^N) \text{ for some } i \ge 1, k \le l\}$ where $(j_i^N, x_i^N, i \ge 1)$ denote the points of N. Since ϵ_i's are absolutely continuous $P[\widehat{\mathbf{N}} \in \mathcal{D}(T), \text{ for some } n \ge 1] = 0$, 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 $T(\widehat{\mathbf{N}}) \rightarrow_d T(\mathbf{N})$ by the continuous mapping theorem, Resnick (1987), p. 153. It follows $(Q_n^d(u_k), k \le l) \rightarrow_d (Q_\infty^d(u_k), k \le l)$, where

$$Q_\infty^d(u) \equiv \int_E l_u(j,x) d\mathbf{N}(j,x).$$

(b) In one-sided model: Using the Ibragimov and Has'minskii (1981a) rules for algebraic operations with ∞'s stated in Theorem 3.1, note $Q_n^d(u) = Q_\infty^d(u) = \int_E l_u(j,x) d\widehat{\mathbf{N}}(j,x)$ as a binomial random variable: $Q_n^d(u) = -\infty$ if $\widehat{\mathbf{N}}(A(u)) > 0$, $Q_n^d(u) = 0$ if $\widehat{\mathbf{N}}(A(u)) = 0$, where $A(u) \equiv \{(j,x) \in \mathbb{R}_+ \times \mathbf{X} : j \le \Delta(x)'u\}$. Also define $Q_\infty^d(u) \equiv Q_\infty^d(u) = \int_E l_u(j,x) d\mathbf{N}(j,x) \equiv -\infty$ if $\mathbf{N}(A(u)) > 0$, $Q_\infty^d(u) \equiv 0$ if $\mathbf{N}(A(u)) = 0$. Thus, to show finite-dimensional convergence (for $\gamma_k = -\infty$ or 0):

$$\lim_{n \to \infty} P(Q_n^d(u_k) = \gamma_k, k \le l) = P(Q_\infty^d(u_k) = \gamma_k, k \le l),$$

it suffices to show $(\widehat{\mathbf{N}}(A(u_k)), k \le l) \rightarrow_d (\mathbf{N}(A(u_k)), k \le l)$ 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. ∎

B.2. **Proof of Theorem 3.2.** The proof applies Theorem I.10.2, p. 107 of Ibragimov and Has'minskii (1981b) that states the limit distribution of BE's provided some general conditions hold.

First, BE's are measurable by Jennrich's (1969) measurability theorem since they minimize the objective functions that are continuous in data and parameters.

Second, the following conditions (1)-(3) verify the conditions of Theorem I.10.2, p. 107 of Ibragimov and Has'minskii (1981b):

(1) (a) Holder continuity of $\ell_n^{1/2}(z)$ in expectation proved in Lemma C.2,
(b) Exponential bound on the expected likelihood tails proved in Lemma C.2: for $a' > 0$

$$E_{P_\gamma} \ell_n(z)^{1/2} \le e^{-a'(|z|-1)},$$

where function $a'(|z|-1)$ falls to the class of functions $\mathbf{G}$ defined on p. 41 in Ibragimov and Has'minskii (1981b), i.e. (i) $a'(|z|-1)$ is monotonically increasing to ∞ in $|z|$ on $[0,\infty)$, and (ii) for any $N > 0$, $\lim_{|z| \to \infty} |z|^N e^{-a'(|z|-1)} = 0$.

(2) Finite-dimensional convergence of $\ell_n(z) = \exp(Q_n(z))$ to $\ell_\infty(z) = \exp(Q_\infty(z))$: for any finite collection $(z_j, j \leq l)$

$$(\ell_n(z_j), j \leq l) \rightarrow_d (\ell_\infty(z_j), j \leq l),$$

(all of the terms are defined in the proof of 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. p.107 in Ibragimov and Has'minskii (1981b).)

and conditions **D1 -D3** on the loss functions ρ and prior μ.

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

Thus, conditions (1)-(3) imply by Theorem I.10.2 of Ibragimov and Has'minskii (1981b) our result:

$$Z_n \rightarrow_d Z.$$

Furthermore, conditions (1)-(3) imply by Theorem I.5.2 and Theorem I.10.2 of Ibragimov and Has'minskii (1981b) that for any local sequence $\gamma_n(\delta) = \gamma_0 + H_n\delta, \delta \in \mathbb{R}^d$ and any $N > 0$

$$\lim_{H \to \infty, n \to \infty} H^{-N} P_{\gamma_n(\delta)}\{|Z_n| > H\} = 0, \quad \text{and} \lim_{n \to \infty} E_{P_{\gamma_n(\delta)}} \rho(Z_n) = E_{P_{\gamma_0}} \rho(Z) < \infty. \tag{B.7}$$

The last result is not needed to prove Theorem 3.2, but will be used later. ■

B.3. **Proof of Theorem 3.3.** To show claim 1, note under the conditions of Theorem 3.2 by (B.7)

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

Consider the problem $\min_c E_{P_{\gamma_0}} \rho(\bar{Z} + c)$, where $\rho(z) = z'z$. The solution of this minimization problem is to set $c = -E\bar{Z}$. Suppose that $c \neq 0$, then

$$E_{P_{\gamma_0}} \rho(\bar{Z} + c) < E_{P_{\gamma_0}} \rho(\bar{Z}), \tag{B.8}$$

where by Lemma 3.1 the lhs of (B.8) is the asymptotic average risk of the sequence of estimators $\bar{\gamma} + H_n c$ and the rhs of (B.8) 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.

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

$$\lim_{n \to \infty} P_{\gamma_n(\delta)} \left\{ (\hat{\gamma}(\tau))_j \leq (\gamma_0)_j \right\} = \lim_{n \to \infty} P_{\gamma_n(\delta)} \left\{ (Z_n(\tau))_j \leq 0 \right\} = P_{\gamma_0} \left\{ (Z(\tau))_j \leq 0 \right\}.$$

since 0 is assumed to be a continuity point of the distribution of $(Z(\tau))_j$. Consider the problem

$$\min_c E_{P_{\gamma_0}} \rho\left((Z(\tau))_j - c; \tau \right),$$

where $\rho(z; \tau) = (1(z \geq 0) - \tau) z$. Note that the quantity $E_{P_{\gamma_0}} \rho\left((Z(\tau))_j - c; \tau \right)$ is finite for any c by (B.7) or by Lemma 3.1. A solution of this problem is given by the root of the first order condition

$$P_{\gamma_0} \left\{ (Z(\tau))_j \geq c \right\} = \tau \text{ or } P_{\gamma_0} \left\{ (Z(\tau))_j \leq c \right\} = 1 - \tau, \tag{B.9}$$

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

$$E_{P_{\gamma_0}} \rho\left((Z(\tau))_j - c; \tau\right) < E_{P_{\gamma_0}} \rho\left((Z(\tau))_j; \tau\right), \tag{B.10}$$

where by Lemma 3.1 the lhs of (B.10) is the asymptotic average risk of the sequence of estimators defined as $(\hat{\gamma}(\tau) - H_n c)_j$, and the rhs is the asymptotic average risk of the posterior τ-th quantile $(\hat{\gamma}(\tau))_j$. See section 3.4 and Lemma 3.1 for definitions. Then (B.10) contradicts the asymptotic average risk efficiency of the posterior quantiles under the check function loss established in Lemma 3.1.

Thus it must be that $c = 0$ in (B.9), so that the first part of claim 2, equation (3.11), is proven. The second part of claim 2, equation (3.12) is immediate from (B.9) with $c = 0$ for $\tau = \tau'$ and $\tau = \tau''$:

$$\begin{aligned}
\lim_{n\to\infty} P_{\gamma_n(\delta)}\left\{(\hat{\gamma}(\tau'))_j \leq (\gamma_n(\delta))_j \leq (\hat{\gamma}(\tau''))_j\right\} &= \lim_{n\to\infty} P_{\gamma_n(\delta)}\left\{(Z_n(\tau'))_j \leq 0 \leq (Z_n(\tau''))_j\right\} \\
&= 1 - \lim_{n\to\infty} P_{\gamma_n(\delta)}\left\{(Z_n(\tau''))_j \leq 0\right\} - \lim_{n\to\infty} P_{\gamma_n(\delta)}\left\{(Z_n(\tau'))_j \geq 0\right\} \\
&= 1 - P_{\gamma_0}\left\{(Z(\tau''))_j \leq 0\right\} - P_{\gamma_0}\left\{(Z(\tau'))_j \geq 0\right\} = \tau'' - \tau'. \quad \blacksquare
\end{aligned} \tag{B.11}$$

B.4. **Proof of Theorem 3.4.** Consider at first two useful lemmas.

Lemma B.1 (Integral Convergence of $\ell_n(z)$). *Suppose that (1) $\ell_n(z)$ has the properties specified in Lemma C.2 and (2) $\ell_n(z)$ 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 ball at zero a.s., (b) for any vector-valued continuous function $g(z)$ dominated by a polynomial as $z \to \infty$*

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

Here K_n is either the set $\{z : \gamma_n(\delta) + H_n z \in \mathcal{G}\}$, in which case $K = \mathbb{R}^d$; or $K_n = K$ is a fixed cube centered at the origin. [Convergence in distribution here is taken under any local sequence $\gamma_n(\delta)$.]

Lemma B.2 (Convexity Lemma). *Suppose $\{R_T\}$ is a sequence of $\mathbb{R}$-valued random functions, defined on $\mathbb{R}^d$. If R_T converges in distribution in finite-dimensional sense to R_∞, i.e. for any $l < \infty$ $(R_T(z_j), j \leq l) \to_d (R_\infty(z_j), j \leq l)$, where R_∞ is convex and finite on an open non-empty set a.s., then $\arg\inf_{z\in\mathbb{R}^d} R_T(z) \to_d \arg\inf_{z\in\mathbb{R}^d} R_\infty(z)$.*

Proof of Lemma B.1 Assertion (a) is a special case of Lemma I.5.1 in Ibragimov and Has'minskii (1981b). Assertion (b) is proven on p. 106-109 of Ibragimov and Has'minskii (1981b) under more general conditions than conditions (1) and (2). $\blacksquare$

Proof of Lemma B.2 See Davis, Knight, and Liu (1992) and Pollard (1991). $\blacksquare$

The first part of the proof of Theorem 3.4 is done by setting the true parameter sequence $\gamma_n(\delta) = \gamma_0$. Considering general sequence does not change the argument but significantly complicates notation.

Write $\tilde{Z}_n(\tau) = (\hat{c}(\tau) - r_n(\gamma_0))$. Note that

$$\tilde{Z}_n(\tau) \equiv \arg\inf_{\tilde{z}\in\mathbf{R}} \Gamma_n(\tilde{z}), \ \Gamma_n(\tilde{z}) \equiv \int_{\mathbf{R}^d} \rho(\tilde{z} - r_n(\gamma_0 + H_n z) + r_n(\gamma_0); \tau) \frac{\ell_n(z)\mu(\gamma_0 + H_n z)}{\int_{\mathbf{R}^d} \ell_n(z')\mu(\gamma_0 + H_n z')\,dz'} dz.$$

Since $|r_n(\gamma_0 + H_n z) - r_n(\gamma_0) - R'z| = O\left(z \cdot |H_n z|^{a'}\right)$ for $a' > 0$, it is the case by the properties of the check function that

$$|\rho(\tilde{z} - r_n(\gamma_0 + H_n z) + r_n(\gamma_0); \tau) - \rho(\tilde{z} - R'z; \tau)| \leq 2|O(z \cdot |H_n z|^{a'})|,$$

for any $\tilde{z}$. Hence by Lemma B.1

$$\begin{aligned} \Gamma_n(\tilde{z}) &= \int_{\mathbf{R}^d} \left(\rho(\tilde{z} - R'z; \tau) + O(z \cdot |H_n z|^{a})\right) \frac{\ell_n(z)\mu(\gamma_0 + H_n z)}{\int_{\mathbf{R}^d} \ell_n(z')\mu(\gamma_0 + H_n z')\,dz'} dz \\ &= \int_{\mathbf{R}^d} \rho(\tilde{z} - R'z; \tau) \frac{\ell_n(z)\mu(\gamma_0 + H_n z)}{\int_{\mathbf{R}^d} \ell_n(z')\mu(\gamma_0 + H_n z')\,dz'} dz + o_p(1). \end{aligned} \tag{B.12}$$

Applying Lemma B.1 again, it follow that the marginal limit of $\Gamma_n(\tilde{z})$ is given by

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

Recall that $\tilde{Z}(\tau)$ denotes the minimizer of $\Gamma_\infty(\tilde{z})$. By Lemma B.2

$$\tilde{Z}_n(\tau) \rightarrow_d \tilde{Z}(\tau).$$

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 for $\tilde{Z}_n(\tau)$. Consider linear transformation $\xi = M'z$ defined by the nonsingular matrix M such that R is a column of M. Then, the likelihood for ξ given by $\ell_n\left(M^{-1}\xi\right)$ has the properties (for some $c > 0$ and $c' > 0$):

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

by nonsingularity of M and Lemma C.2. By Theorem I.5.2 of Ibragimov and Has'minskii (1981b) for any local sequence of $\gamma_n(\delta) = \gamma_0 + H_n\delta, \delta \in \mathbb{R}^d$ and any $N > 0$

$$\begin{aligned} &\lim_{H\to\infty, n\to\infty} H^{-N} P_{\gamma_n(\delta)}\{|\tilde{Z}_n(\tau)| > H\} = 0, \\ &\text{hence } \lim_{n\to\infty} E_{P_{\gamma_n(\delta)}}\rho\left(\tilde{Z}_n(\tau); \tau\right) = E_{P_{\gamma_0}}\rho\left(\tilde{Z}(\tau); \tau\right) < \infty. \end{aligned} \tag{B.13}$$

Then identically to the steps in the proof of Lemma 3.1, it can be concluded that $\{\hat{c}(\tau)\}$ minimizes the asymptotic average risk in the sense of achieving the infimum of

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

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

B.5. **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 an argument similar to that given by Ibragimov and Has'minskii (1981b) p.93. Details are omitted for brevity. See Chernozhukov and Hong (2003). ■

Appendix C. Important Lemmas

Let $\gamma = (\beta, \alpha)$ and $h = (h_\beta, h_\alpha)$. Define the standard Hellinger distance $r_2(\gamma; \gamma + h)^2 =$

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

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

Lemma C.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 γ*

$$\text{(a) } r_2^2(\gamma; \gamma + h) \geq 2\frac{a \max(|h_\beta|, |h_\alpha|^2)}{1 + \max(|h_\beta|, |h_\alpha|^2)} \text{ and } \text{(b) } r_2^2(\gamma; \gamma + h) \leq A(|h_\beta| + |h_\alpha|^2). \tag{C.1}$$

Lemma C.2 (Exponential Tails and Holder Continuity). *Given (C.1), for some n_0 and all $n > n_0$ for any z, z' such that $\gamma + H_n z \in \mathcal{G}$ and $\gamma + H_n z' \in \mathcal{G}$, and some $a' > 0$ uniformly in γ*

$$E_{P_\gamma}\ell_n(z)^{1/2} \leq e^{-a'(|z|-1)}, \quad E_{P_\gamma}|\ell_n(z)^{1/2} - \ell_n(z')^{1/2}|^2 \leq A(|z - z'|)(1 + 2\cdot|z'| \vee |z|). \tag{C.2}$$

C.1. **Proof of Lemma C.2.** For some $B > 0$ and $|\cdot|$ denoting the sup norm,

$$E_{P_\gamma}\ell_n(z)^{1/2} \overset{(1)}{\leq} \left[1 - \frac{1}{2}r_2^2(\gamma; \beta + u/n, \alpha + v/\sqrt{n})\right]^n$$
$$\overset{(2)}{\leq} e^{-\frac{n}{2}r_2^2(\gamma;\beta+u/n,\alpha+v/\sqrt{n})} \overset{(3)}{\leq} e^{-a\frac{\max(|u|,|v|^2)}{1+\max(|u|,|v|^2)/n}} \overset{(4)}{\leq} e^{-a\frac{\max(|u|,|v|^2)}{1+K_{\mathcal{G}}}} \overset{(5)}{\leq} e^{\frac{-a|z|+a}{1+K_{\mathcal{G}}}},$$

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

$$E_{P_\gamma}|\ell_n(z)^{1/2} - \ell_n(z')^{1/2}|^2 \overset{(1)}{\leq} nr_2^2(\gamma + (u/n, v/\sqrt{n}); \gamma + (u'/n, v'/\sqrt{n})) \overset{(2)}{\leq} A(|u - u'| + |v - v'|^2)$$
$$\overset{(3)}{\leq} A(|z - z'| + |z - z'|^2) \overset{(4)}{\leq} A(|z - z'|)(1 + 2\cdot|z'| \vee |z|),$$

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

C.2. **Proof of Lemma C.1.** In order to establish (C.1)(b), let $\gamma = (\beta, \alpha)$ and $h = (h_\beta, h_\alpha)$,

$$
\begin{aligned}
& r_2^2(\gamma;\gamma+h) \le E_X \int \left(f^{1/2}\left(y - g\left(X, \beta + h_\beta\right)|X; \beta + h_\beta, \alpha + h_\alpha\right) - f^{1/2}\left(y - g\left(X,\beta\right)|X;\gamma\right)\right)^2 dy \\
& \overset{(1)}{\le} E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]} \left| f\left(y - g\left(X,\beta+h_\beta\right)|X;\beta+h_\beta,\alpha+h_\alpha\right) - f\left(y - g\left(X,\beta\right)|X;\gamma\right)\right| dy \\
& + E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]^c} \left(f^{1/2}\left(y - g\left(X,\beta+h_\beta\right)|X;\gamma+h\right) - f^{1/2}\left(y - g\left(X,\beta+h_\beta\right)|X;\beta+h_\beta,\alpha\right)\right)^2 dy \\
& + E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]^c} \left| f\left(y - g\left(X,\beta+h_\beta\right)|X;\beta+h_\beta,\alpha\right) - f\left(y - g\left(X,\beta\right)|X;\gamma\right)\right| dy \\
& \overset{(2)}{\le} 2E_X \left| g\left(X,\beta+h_\beta\right) - g\left(X,\beta\right)\right| \bar{f} \\
& + |h_\alpha|^2 \int_0^1 E_X \int \left| \frac{\partial f^{1/2}\left(y - g\left(X,\beta+h_\beta\right)|X,\beta+h_\beta,\alpha+\omega h_\alpha\right)}{\partial\alpha} \right|^2 dy d\omega \\
& + |h_\beta| \int_0^1 E_X \int \left| \frac{\partial f\left(y - g\left(X,\beta+\omega h_\beta\right)|X,\beta+\omega h_\beta,\alpha\right)}{\partial\beta} \right| dy d\omega \\
& \overset{(3)}{\le} 2\bar{f}|h_\beta| E_X \int_0^1 \left| \frac{\partial g\left(X,\beta+\omega h_\beta\right)}{\partial\beta} \right| d\omega + O\left(|h_\alpha|^2\right) + O\left(|h_\beta|\right) \\
& = O\left(|h_\beta|\right) + O\left(|h_\alpha|^2\right),
\end{aligned}
$$

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

The lower bound from below, the equation (C.1)(a), is established by considering separately $|h| < \delta$ for some sufficiently small δ and $|h| > \delta$.

Indeed, for sufficiently small δ and $|h| \le \delta$ it is shown below that

$$r_2^2(\gamma;\gamma+h) \ge \text{ const } \max(|h_\beta|, |h_\alpha|^2). \tag{C.3}$$

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

$$r_2^2(\gamma;\gamma+h) \ge \varepsilon_\delta > 0. \tag{C.4}$$

Hence for some $a > 0$ the bound in (C.1)(a) is immediate from (C.3) - (C.4).

It remains to prove (C.3) for $|h| < \delta$ for some sufficiently small δ. Write

$$
\begin{aligned}
r_2^2(\gamma;\gamma+h) = & \underbrace{E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]} \left(f^{1/2}\left(y - g\left(X,\beta+h_\beta\right)|X;\gamma+h\right) - f^{1/2}\left(y - g\left(X,\beta\right)|X;\gamma\right)\right)^2 dy}_{I} \\
& \underbrace{+ E_X \int_{[g(X,\beta), g(X,\beta+h_\beta)]^c} \left(f^{1/2}\left(y - g\left(X,\beta+h_\beta\right)|X;\gamma+h\right) - f^{1/2}\left(y - g\left(X,\beta\right)|X;\gamma\right)\right)^2 dy.}_{II}
\end{aligned}
$$

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

$$E_X \frac{1}{2}\left|g(X,\beta+h_\beta)-g(X,\beta)\right|\left|p^{1/2}(X,\gamma)-q^{1/2}(X,\gamma)\right|^2 \geq \text{const } E_X\left|\frac{\partial g(X;\beta)'}{\partial\beta}h_\beta\right| \geq \text{const } |h_\beta|,$$

using assumption **C3** and Taylor expansion. On the other hand, by **C1-C3**, bound II from below by:

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

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

Under **C2**, **C3** and **C4(a)**, a further lower bound is $|h|^2 \inf_{|u|=1} E_X \int_{[g(X,\beta),g(X,\beta+h_\beta)]^c}\left(\frac{f^{1/2}(y-g(X,\beta);\gamma)'}{\partial\gamma}u\right)^2 dy$

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

for 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 bounds in **C2** and **C3** to do so. On the other hand, if assumption **C4(b)** holds, the uniform lower bound is

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

Conclude $\inf_\gamma r_2^2(\gamma;\gamma+h) \geq$ const $\cdot \max\left(|h_\beta|, |h_\alpha^2|\right)$. ■

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.

Andrews, D. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle, and D. McFadden. North Holland.

Andrews, D. W. (1997): "A stopping rule for the computation of generalized method of moments estimators," *Econometrica*, 65(4), 913–931.

Artstein, Z., and R. J.-B. Wets (1995): "Consistency of minimizers and the SLLN for stochastic programs," *J. Convex Anal.*, 2(1-2), 1–17.

Berger, J. O. (1993): *Statistical decision theory and Bayesian analysis*, Springer Series in Statistics. Springer-Verlag, New York, Corrected reprint of the second (1985) edition.

Bowlus, A., G. Neumann, and A. Kiefer (2001): "Equilibrium Search Models and the Transition from School to Work," *International Economic Review*, 42(2), 317–343.

Chernozhukov, V. (2001): "Extremal Quantile Regression," MIT Working Paper.

Chernozhukov, V., and H. Hong (2003): "Likelihood Estimation and Inference in Some Nonregular Econometrics Models," Working Paper, Department of Economics MIT, to be published by the Social Science Reserach Network at www.ssrn.com.

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 Process. Appl.*, 40(1), 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," Typescript. London, Canada: Department of Economics, University of Western Ontario.

Donald, S. G., and H. J. Paarsch (1993b): "Piecewise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review*, 34, 121–148.

Donald, S. G., and H. J. Paarsch (1996): "Identifiction, 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," *J. Econometrics*, 109(2), 305–340.

Dupačová, J., and R. Wets (1988): "Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems," *Ann. Statist.*, 16(4), 1517–1549.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997): *Modelling extremal events*, vol. 33 of *Applications of Mathematics*. Springer-Verlag, Berlin, For insurance and finance.

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

Hirano, K., and J. Porter (2002): "Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support," October 2002, forthcoming in Econometrica.

Horowitz, J. (2000): "The Bootstrap," University of Iowa; to appear, *Handbook of Econometrics*, vol. 5.

Ibragimov, I., and R. Has'minskii (1981a): *Statistical Estimation: Asymptotic Theory*. Springer Verlag, Chapter V: Independent Identically Distributed Observations. Densities with Jumps.

——— (1981b): *Statistical Estimation: Asymptotic Theory*. Springer Verlag, Chapter I: The Problem of Statistical Estimation.

Jennrich, R. I. (1969): "Asymptotic properties of non-linear least squares estimators," *Ann. Math. Statist.*, 40, 633–643.

Knight, K. (2000): "Epi-convergence and Stochastic Equisemicontinuity," Preprint.

Koenker, R., and G. S. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Lehmann, E., and G. Casella (1998): *Theory of Point Estimation*. Springer.

Meyer, R. M. (1973): "A poisson-type limit theorem for mixing sequences of dependent 'rare' events," *Ann. Probability*, 1.

Paarsch, H. (1992): "Deciding between the common and private value paradigms in empirical models of auctions," *Journal of Econometrics*, 51, 191–215.

Pflug, G. C. (1995): "Asymptotic stochastic programs," *Math. Oper. Res.*, 20(4), 769–789.

Politis, D., J. Romano, and M. Wolf (1999): *Subsampling*. Springer Series in Statistics.

Pollard, D. (1991): "Asymptotics for Least Absolute Deviation Regression Estimator," *Econometric Theory*, 7, 186–199.

Portnoy, S. (1991): "Asymptotic behavior of the number of regression quantile breakpoints," *SIAM J. Sci. Statist. Comput.*, 12(4), 867–883.

Portnoy, S., and R. Koenker (1997): "The Gaussian Hare and the Laplacian Tortoise," *Statistical Science*, 12, 279–300.

Resnick, S. I. (1986): "Point processes, regular variation and weak convergence," *Adv. in Appl. Probab.*, 18(1), 66–138.

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

Robert, C. P., and G. Casella (1998): *Monte Carlo Statistical Methods*. Springer-Verlag.

Rockafellar, R. T., and R. B. Wets (1998): *Variational Analysis*. Springer-Verlag, Berlin.

Salinetti, G., and R. J.-B. Wets (1986): "On the convergence in distribution of measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima," *Math. Oper. Res.*, 11(3), 385–419.

SMITH, R. L. (1994): "Nonregular regression," *Biometrika*, 81(1), 173–183.
VAN DER VAART, A. (1999): *Asymptotic Statistics*. Cambridge University Press.
VAN DER VAART, A. (2000): "Lecture notes on Bayesian Estimation," preprint, Aarhus University.
VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak convergence and empirical processes*. Springer-Verlag, New York.
WALD, A. (1950): *Statistical Decision Functions*. John Wiley & Sons Inc., New York, N. Y.

Table 1. Estimator Performance for Intercept β_0. (Based on 500 repetitions).

Estimator	RMSE	MAD	Median AD	pthloss (p=10)
n=100				
Posterior mean	0.0114	0.0081	0.0059	0.0401
Posterior median	0.0115	0.0078	0.0054	0.0433
MLE	0.0127	0.0088	0.0063	0.0369
n=400				
Posterior mean	0.0029	0.0021	0.0015	0.0104
Posterior median	0.0030	0.0020	0.0014	0.0106
MLE	0.0034	0.0023	0.0015	0.0103

Table 2. Estimator Performance for Slope β_1. (Based on 500 repetitions)

Estimator	RMSE	MAD	Median AD	pthloss (p=10)
n=100				
Posterior mean	0.0212	0.0147	0.0105	0.0983
Posterior median	0.0214	0.0144	0.0096	0.1053
MLE	0.0216	0.0145	0.0096	0.0693
n=400				
Posterior mean	0.0053	0.0037	0.0026	0.0227
Posterior median	0.0054	0.0037	0.0025	0.0228
MLE	0.0057	0.0038	0.0024	0.0201

Table 3. 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	0.87	0.0298	0.85	0.0555
Bootstrap P-mean	0.88	0.0392	0.86	0.0720
Subsampling: P-mean	0.83	0.0416	0.82	0.0770
Limit process: P-mean	0.86	0.0346	0.84	0.0587
Limit process: P-median	0.85	0.0352	0.86	0.0610
Limit process: MLE	0.93	0.0347	0.95	0.0653
n=400				
Posterior Intervals	0.87	0.0075	0.86	0.0140
Bootstrap: P-mean	0.84	0.0089	0.88	0.0167
Subsampling: P-mean	0.82	0.0085	0.83	0.0158
Limit process: P-mean	0.89	0.0085	0.82	0.0145
Limit process: P-median	0.86	0.0087	0.86	0.0150
Limit process: MLE	0.89	0.0084	0.92	0.0157

Technical Addendum, Part I

This addendum includes the excluded material, such as omitted simpler calculations and simpler proofs, and some background material on point processes. The addendum will be made available as a part of an MIT Economics Department Technical Report published by the Social Science Research Network.

Appendix D. Point Processes

The following definitions are collected from Resnick (1987).

Definition D.1 ($M_p(E)$)**.** Let E be a locally compact topological space with a countable basis, and $\mathcal{E}$ to be the Borel σ-algebra of subsets of E. A *point measure* (p.m.) p on $(E, \mathcal{E})$ is a measure of the following form: for $\{x_i, i \geq 1\}$, a countable collection of points (called points of p), and any set $A \in \mathcal{E}$: $p(A) \equiv \sum_i 1(x_i \in A)$. If $p(K) < \infty$, for any $K \subset E$ compact, then p is said to be Radon. A p.m. p is simple if $p(x) \leq 1 \quad \forall x \in E$, and is compound otherwise. Let $M_p(E)$ be the collection of all Radon point measures. Sequence $\{p_n\} \subset M_p(E)$ converges vaguely to p, if $\int f dp_n \to \int f dp$ for all functions $f \in C_K(E)$ [continuous, real-valued, and vanishing outside a compact set]. Vague convergence induces *vague topology* on $M_p(E)$. The topological space $M_p(E)$ is metrizable as complete separable metric space. $M_p(E)$ denotes such metric space hereafter. Define $\mathcal{M}_p(E)$ to be σ-algebra generated by open sets.

Definition D.2 (Point Processes. Convergence in Distribution.). A *point process* in $M_p(E)$ is a measurable map $\mathbf{N} : (\Omega, \mathcal{F}, P) \to (M_p(E), \mathcal{M}_p(E))$, i.e. for every elementary event $w \in \Omega$, the realization of the point process $\mathbf{N}(w)$ is some point measure in $M_p(E)$. *Weak convergence* of the point processes $\mathbf{N}_n$ taking values in $M_p(E)$ is the same as for any metric space, cf. Resnick (1987): we shall write $\mathbf{N}_n \Rightarrow \mathbf{N}$ in $M_p(E)$ if $E_P h(\mathbf{N}_n) \to E_P h(\mathbf{N})$ for all continuous and bounded functions h mapping $M_p(E)$ to $\mathbb{R}$. Note that if $\mathbf{N}_n \Rightarrow \mathbf{N}$ in $M_p(E)$, then $\int_E f(x) d\mathbf{N}_n(x) \to_d \int_E f(x) d\mathbf{N}(x)$ for any $f \in C_K(E)$ by continuous mapping theorem.

Definition D.3. (Poisson Point Process) The point process $\mathbf{N}$ is a Poisson Point Process with *mean intensity measure* m on $(E, \mathcal{E})$, if

(a) for any $F \in \mathcal{E}$, and any non-negative integer k

$$P(\mathbf{N}(F) = k) = \begin{cases} e^{-m(F)} m(F)^k / k! & \text{if } m(F) < \infty \\ 0 & \text{if } m(F) = \infty, \end{cases}$$

(b) if $(F_i, i \leq k)$ are disjoint sets in $\mathcal{E}$, then $(\mathbf{N}(F_i), i \leq k)$ are independent random variables.

Appendix E. Plausibility of C0-C5

This section illustrates the plausibility of conditions **C0-C5** using the Paarsch's (1992) independent private value auction model in section 2.1. For clarity consider the example used in the Monte-Carlo section: $\theta_2 \equiv 1$ and $\theta_1 = \exp(\beta' Z)$, with the following assumptions. More general examples can be verified similarly. For example θ_2 can be made a function of regressors too, and verification proceeds similarly but is notationally much more burdensome.

Assumptions (in addition to those listed in Section 2.1)

A1 $Y_i, X_i = (Z_i, m_i)$ are iid, $\beta \in \mathcal{B}$, a compact convex subset of $\mathbb{R}^d$. $X \in \mathbf{X}$, a compact subset of $\mathbb{R}^{d_\beta + 1}$.

A2 $F_X(x)$ does not depend on β. EZZ' is positive definite.

A3 $3 \le m_i \le M$ for some $M < \infty$, where m_i is (non-degenerate) number of bidders (minus 1).

Verification of **C0**: **A1** implies iid sampling and compactness and convexity of parameter space stated in **C0**.

Under the stated parameterization

$$g(X,\beta) = \exp(\beta' Z)(m-1)/(m-2)$$

and

$$f(\epsilon|X,\beta) = m\frac{g(X,\beta)^m}{[\epsilon + g(X,\beta)]^{m+1}} 1(\epsilon \ge 0).$$

Note that $\gamma = \beta$ and **A2** and **A3** imply that for $\beta \neq \beta'$, $g(X,\beta) \neq g(X,\beta')$ for some X with positive probability. Since the $f(y - g(X,\beta)|X,\beta)$ density function is strictly monotone in $g(X,\beta)$, identification holds.

Verification of **C1**: Clearly the model is a one-sided model **C1**.(ii) in which $f(\epsilon|x,\beta) \equiv q(x,\beta) \equiv 0$ and

$$p(X,\beta) = m\frac{g(X,\beta)^m}{[g(X,\beta)]^{m+1}} = \frac{m}{g(X,\beta)}$$

is strictly bounded away from 0 and from above by **A1** and **A3**.

Verification of **C2**: $f(\epsilon|X,\beta)$, as defined, is obviously upper semi-continuous at $\epsilon = 0$, is maximized at $\epsilon = 0$ for each X and β, and is uniformly bounded by **A1**. Its first derivative in ϵ:

$$\frac{\partial}{\partial \epsilon} f(\epsilon|X,\beta) = -\frac{(m+1)}{[\epsilon + g(X,\beta)]} f(\epsilon|X,\beta),$$

is continuous in ϵ except at $\epsilon = 0$ and bounded uniformly in (ϵ, X, β) by **A1** and **A3**.

The first partial derivative of $f(\epsilon|X,\beta)$ in β:

$$\frac{\partial}{\partial \beta} f(\epsilon|X,\beta) = Z\left[m - (m+1)\frac{g(X,\beta)}{[\epsilon + g(X,\beta)]}\right] f(\epsilon|X,\beta)$$

is clearly continuous in X and β and bounded uniformly in (ϵ, X, β) by **A1** and **A3**. This holds similarly for $\frac{\partial^2}{\partial\beta\partial\beta'} f(\epsilon|X,\beta)$:

$$\frac{\partial^2}{\partial\beta\partial\beta'} f(\epsilon|X,\beta) = ZZ' f(\epsilon|X,\beta)\left[\left(m - (m+1)\frac{g(X,\beta)}{[\epsilon + g(X,\beta)]}\right)^2 - (m+1)\frac{g(X,\beta)}{[\epsilon + g(X,\beta)]} + (m+1)\frac{g(X,\beta)}{[\epsilon + g(X,\beta)]^2}\right].$$

All of the above quantities are continuous in X for each ϵ and β. Finally, for some constant C

$$\left|\frac{\partial}{\partial\beta} f(y - g(X,\beta)|X,\beta)\right| = 1(y \ge g(X,\beta))\left|\frac{m^2 g(X,\beta)^{m-1}}{y^{m+1}} g(X,\beta) Z\right| \le 1(y \ge -C)\frac{C}{y^{m+1}},$$

which implies

$$\sup_\beta E_X \int \left|\frac{\partial}{\partial\beta} f\left(y - g(X,\beta)\middle|X,\beta\right)\right| dy \le E_X \int_{-C}^{\infty} \left|\frac{C}{y^{m+1}}\right| dy < \infty.$$

Verification of **C3**: This is verified immediately by noting that:

$$\frac{\partial}{\partial\beta} g(X,\beta) = g(X,\beta) Z,$$

and

$$\frac{\partial^2}{\partial\beta\partial\beta'}g(X,\beta)=g(X,\beta)ZZ',$$

due to **A1**. Also note that

$$E\left[\frac{\partial g(X;\beta)}{\partial\beta}\frac{\partial g(X;\beta)'}{\partial\beta}\right]=E\left[g(X,\beta)^2ZZ'\right]\geq c^2EZZ'$$

for some $c>0$ due to **A1** and **A3**. By **A2** EZZ' is positive definite.

Verification of **C4**: Verification of **C4** is not needed since parameter α is not present.

Verification of **C5**: When the true parameter is γ so that $\epsilon_i=Y_i-g(X_i,\beta)$, we have that

$$f\left(Y_i-g(X_i,\bar\beta)|X,\bar\beta\right)=m\frac{g\left(X_i,\bar\beta\right)^m}{\left[\epsilon_i+g(X_i,\beta)\right]^{m+1}}1\left(Y_i-g(X_i,\bar\beta)\geq 0\right).$$

Therefore by **A1** and **A3**

$$\begin{aligned}\sup_{\bar\beta}\left|\tfrac{\partial}{\partial\beta}\ln f(Y_i-g(X_i,\bar\beta)|X_i,\bar\beta)\right|&=\sup_{\bar\beta}1(Y_i-g(X_i,\bar\beta)>0)\left|(m+1)\frac{g(X_i,\bar\beta)}{\epsilon_i+g(X_i,\beta)}Z_i\right.\\&\left.+\left[m-(m+1)\frac{g(X_i,\bar\beta)}{\left[\epsilon_i+g\left(X_i,\beta\right)\right]}\right]Z_i\right|\leq C_1<\infty,\end{aligned}$$

and

$$\begin{aligned}\sup_{\bar\beta}\left|\tfrac{\partial^2}{\partial\beta\partial\beta}\ln f(Y_i-g(X_i,\bar\beta)|X_i,\bar\beta)\right|&=\sup_{\bar\beta}1(Y_i-g(X_i,\bar\beta)>0)\left|(m+1)\frac{g(X_i,\bar\beta)}{\epsilon_i+g(X_i,\beta)}Z_iZ_i'\right.\\&\left.-\left[(m+1)\frac{g(X_i,\bar\beta)}{\left[\epsilon_i+g\left(X_i,\beta\right)\right]}\right]Z_iZ_i'\right|\leq C_3<\infty.\end{aligned}$$

Appendix F. Excluded Simple Derivations

F.1. **In Proof of Theorem 3.1.** These additional derivations are added at the request of a referee. Write

$$Q_{1n}^c(z)=\sum_{i=1}^n\widehat r_{in}(z)\quad\times\mathbf{1}_{in}(u)\tag{F.1}$$

where $\mathbf{1}_{in}(u)\equiv\left[1(\epsilon_i>\{\Delta_n(X_i,u)/n\vee 0\})+1(\epsilon_i<\{\Delta_n(X_i,u)/n\wedge 0\})\right]$.

At first, for a fixed z, we use a Taylor expansion of $\hat r_{in}(z)$ for each i and plugging this back into expression for $Q_{1n}^c(z)$

$$\begin{aligned}Q_{1n}^c(z)=&\underbrace{-u'\frac{1}{n}\sum_{i=1}^n\frac{\partial g\left(X_i,\beta_0+\frac{u_{in}}{n}\right)}{\partial\beta}\frac{\partial}{\partial\epsilon}\ln f\left(\epsilon_i-\frac{\Delta_n(X_i,u_{in})}{n}|X_i,\beta_0+u/n,\alpha_0+v/\sqrt{n}\right)\mathbf{1}_{in}(u)+}_{I}\\&+\underbrace{v'\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right)\mathbf{1}_{in}(u)}_{II}+\underbrace{u'\frac{1}{n}\sum_{i=1}^n\frac{\partial}{\partial\beta}\ln f\left(\epsilon_i-\frac{\Delta_n(X_i,u)}{n}|X_i,\beta_0+u_{in}/n,\alpha_0+v/\sqrt{n}\right)\mathbf{1}_{in}(u)}_{II'\to_p 0}\\&+\underbrace{\frac{1}{2}v'\left[\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\alpha\partial\alpha'}f\left(\epsilon_i-\frac{\Delta_n(X_i,u)}{n}|X_i,\beta_0+u/n,\alpha_0+v_{in}/\sqrt{n}\right)\mathbf{1}_{in}(u)\right]v,}_{III}\end{aligned}$$

where u_{in} is a point on the line between 0 and u that depend on u and (ϵ_i, X_i) (but not on other observations), and v_{in} is a point on the line between 0 and v that depends on v and (ϵ_i, X_i) (but not on other observations). Thus, the terms are iid in the above summations.

Application of Chebyshev inequality, the bounded density condition C1, and the bounded derivative condition C3 removes $\mathbf{1}_{in}(u)$ in the term II, and adds $o_p(1)$ to the above expression. For example, consider

$$\left\| Var\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right)\left(\mathbf{1}_{in}(u)-1\right)\right]\right\| \leq \text{const}\cdot(\bar{f}'/\underline{f})^2\cdot\frac{1}{n}\sum_{i=1}^{n}E|\mathbf{1}_{in}(u)-1|^2 \leq \text{const}\cdot(\bar{f}'/\underline{f})^2\cdot 2\cdot(\bar{f}\bar{g}')/n = o(1). \quad \text{(F.2)}$$

Similarly

$$E\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right)\left(\mathbf{1}_{in}(u)-1\right)=o(1).$$

Hence

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right)\mathbf{1}_{in}(u)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right)+o_p(1).$$

Also elementary calculations, as in (F.2), and application of Chebyshev inequality, show that by the bounded density conditions **C2** and bounded derivative condition **C3**, we can replace u_{in}, v_{in} by 0 and remove $\mathbf{1}_{in}(u)$ from the expression for I, II' and III, and add a term $o_p(1)$ to the whole expression. The application of Markov LLN along with **C5** allows to replace each of the terms with its limit expectation, and gives the required conclusion.

F.2. **Proof of Independence needed in the Proof of Theorem 3.1.** These additional derivations are added at the request of a referee. The proof was omitted from the main text, because it follows essentially from the standard proof concerning the independence of minimal order statistics (extremal processes) and sample averages of general form (partial sum processes), see e.g. Resnick (1986) and Lemma 21.19 in van der Vaart (1999). It is presented here for completeness.

Since up to $o_p(1)$ terms, the only stochastic element in $(Q_n^c(z_j), j\leq l)$ is $\mathbf{W}_n$, we need to show that $(Q_n^d(u_j), j\leq l)$ are asymptotically independent of $\mathbf{W}_n$, where ignoring $o_p(1)$ term

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

For clarity we first present the proof for the case when $l=1$, and discuss the changes needed to accommodate $l>1$. For notation sake, we do not index P by γ.

<u>Case I: $l=1$.</u> By an argument similar to that in (F.2), $\mathbf{W}_n-\bar{\mathbf{W}}_n=o_p(1)$, where

$$\bar{\mathbf{W}}_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right)1\left(\epsilon_i>\{\Delta(X_i)'u/n\}\vee 0\text{ or }\epsilon_i<\{\Delta(X_i)'u/n\}\wedge 0\right)$$

Therefore it suffices to show asymptotic independence between $\bar{\mathbf{W}}_n$ and $Q_n^d(u)$. Define

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

and

$$Q^p_\infty(u) \equiv \sum_{i=1}^{\infty}\left[1(0 < J_i < \Delta(\mathcal{X}_i)' u) + 1(0 > J_i' > \Delta(\mathcal{X}_i')' u)\right].$$

Then asymptotic independence between $\bar{\mathbf{W}}_n$ and $Q^d_n(u)$ follows by the Portmanteau Lemma and proving that for any real x, y and integer k:

$$\limsup_{n\to\infty} P\left\{Q^p_n(u) = k, Q^d_n(u) \le x, \bar{\mathbf{W}}_n \le y\right\} \le P\left\{Q^p_\infty(u) = k, Q^d_\infty(u) \le x\right\} P\{\mathbf{W} \le y\}. \quad \text{(F.3)}$$

To show (F.3), proceed in two steps. In Step 1 below, invoking iid sampling, it is shown that

$$P\left\{Q^p_n(u) = k, Q^d_n(u) \le x, \bar{\mathbf{W}}_n \le y\right\} = P\left\{Q^p_n(u) = k, Q^d_n(u) \le x\right\} \cdot P\left\{\sqrt{\tfrac{n-k}{n}}\widetilde{\mathbf{W}}_n \le y\right\} \quad \text{(F.4)}$$

where

$$\widetilde{\mathbf{W}}_n = \frac{1}{\sqrt{n-k}}\sum_{i=1}^{n-k}\frac{\partial}{\partial\alpha}\ln f\left(\bar{\epsilon}_i|\bar{X}_i,\gamma_0\right), \quad \text{(F.5)}$$

where $\bar{\epsilon}_i, \bar{X}_i$ for $i \le n-k$ are i.i.d. draws from the distribution of ϵ_i, X_i conditional on

$$A(\epsilon_i, X_i) \equiv \{\epsilon_i > \Delta(X_i)'u/n \vee 0 \text{ or } \epsilon_i < \Delta(X_i)'u/n \wedge 0\}.$$

Step 2 applies CLT to show that $\widetilde{\mathbf{W}}_n \to_d \mathbf{W}$. Finally, convergence $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ implies similarly to the proof in Theorem 3.1 and by the Portmanteau Lemma that

$$\limsup_{n\to\infty} P\left\{Q^p_n(u) = k, Q^d_n(u) \le x\right\} \le P\left\{Q^p_\infty(u) = k, Q^d_\infty(u) \le x\right\}.$$

Thus the proof is complete given Steps 1 and 2.

Step 1. Define $p_n = P\{A(\epsilon_i, X_i)^c\}$. By i.i.d. sampling in **C0**, the left hand side of (F.4) can be written as

$$\underbrace{\binom{n}{k}p_n^k(1-p_n)^{n-k}P\left\{\sum_{i=n-k+1}^{n} l_u(n\epsilon_i, X_i) \le x\middle| 1\left(0 < n\epsilon_i < \Delta(X_i)'u\right) + 1\left(0 > n\epsilon_i > \Delta(X_i)'u\right) = 1, \text{ for } i = n-k+1,\dots,n\right\}}_{\equiv P\{Q^p_n(u)=k, Q^d_n(u)\le x\}}$$

$$\underbrace{\times P\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n-k}\frac{\partial}{\partial\alpha}\ln f(\epsilon_i|X_i,\gamma_0) \le y\middle|\epsilon_i > \{\Delta(X_i)'u/n\} \vee 0 \text{ or } \epsilon_i < \{\Delta(X_i)'u/n\} \wedge 0, \text{ for all } i \le n-k\right\}}_{\equiv P\left\{\sqrt{\frac{n-k}{n}}\widetilde{\mathbf{W}}_n \le y\right\}},$$

where function l_u is defined in Theorem 3.1 and $\widetilde{\mathbf{W}}_n$ in (F.5).

Step 2. $\widetilde{\mathbf{W}}_n \to_d \mathbf{W}$ follows by checking three conditions:

(a) $E\widetilde{\mathbf{W}}_n \to 0$,
(b) $Var\left(\widetilde{\mathbf{W}}_n\right) \to Var(\mathbf{W})$, and
(c) Lindeberg's condition is satisfied.

Condition (a) requires $E\tilde{\mathbf{W}}_n = \sqrt{n-k}E\frac{\partial}{\partial\alpha}\ln f\left(\bar{\epsilon}_i|\bar{X}_i,\gamma_0\right) \to 0$. This is true because

$$E\widetilde{\mathbf{W}}_n = \sqrt{n-k}\frac{E_X\int_{A(\epsilon_i,X_i)}\frac{\partial}{\partial\alpha}\ln f(\epsilon|X_i,\gamma_0)f(\epsilon_i|X,\gamma_0)\,d\epsilon_i}{P\{A(\epsilon_i,X_i)\}} = -\sqrt{n-k}\frac{E_X\int_{A(\epsilon_i,X_i)^c}\frac{\partial}{\partial\alpha}\ln f(\epsilon_i|X,\gamma_0)f(\epsilon_i|X_i,\gamma_0)\,d\epsilon_i}{P\{A(\epsilon_i,X_i)\}}$$

By **C2** $\lim_{n\to\infty} P\{A(\epsilon_i, X_i)\} = 1$. In addition, for large enough n:

$$-E_X\int_{A(\epsilon_i,X_i)^c}\frac{\partial}{\partial\alpha}\ln f(\epsilon_i|X_i,\gamma_0)f(\epsilon_i|X_i,\gamma_0)\,d\epsilon_i < 2\cdot(\bar{f}'/\underline{f})\cdot(\bar{g}'\bar{f})\cdot(\frac{1}{n}).$$

Hence $E\widetilde{\mathbf{W}}_n \to 0$. A similar calculation verifies

$$\begin{aligned} Var\left(\widetilde{\mathbf{W}}_n\right) =& E\frac{\partial}{\partial\alpha}\ln f\left(\tilde{\epsilon}_i|\tilde{X}_i,\gamma_0\right)\frac{\partial}{\partial\alpha}\ln f\left(\tilde{\epsilon}_i|\tilde{X}_i,\gamma_0\right)' - \left[E\frac{\partial}{\partial\alpha}\ln f\left(\tilde{\epsilon}_i|\tilde{X}_i,\gamma_0\right)\right]^2 \\ \to& E\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right)\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right)' = Var\left(\mathbf{W}_n\right). \end{aligned}$$

The final step is to verify the Lindeberg condition, that for all $\lambda \neq 0$ and all $\xi > 0$,

$$\frac{1}{\sigma_{n-k}^2}\sum_{i=1}^{n-k} E\left[\left(\lambda'\frac{\partial}{\partial\alpha}\ln f\left(\tilde{\epsilon}_i|\tilde{X}_i,\gamma_0\right)^2\right)1\left(\left|\lambda'\frac{\partial}{\partial\alpha}\ln f\left(\tilde{\epsilon}_i|\tilde{X}_i,\gamma_0\right)\right| > \xi\sigma_{n-k}\sqrt{n}\right)\right] \to 0, \qquad \text{(F.6)}$$

where we define

$$\sigma_{n-k}^2 = \frac{1}{n-k}\sum_{i=1}^{n-k} Var\left[\lambda'\frac{\partial}{\partial\alpha}\ln f\left(\tilde{\epsilon}_i|\tilde{X}_i,\gamma_0\right)\right] = \lambda' Var\left[\frac{\partial}{\partial\alpha}\ln f\left(\tilde{\epsilon}_i|\tilde{X}_i,\gamma_0\right)\right]\lambda. \qquad \text{(F.7)}$$

Since σ_{n-k}^2 converges to a positive constant by (b), the conclusion (F.6) follows from

$$E\left(\lambda'\frac{\partial}{\partial\alpha}\ln f\left(\tilde{\epsilon}_i|\tilde{X}_i,\gamma_0\right)\right)^2 \le E\left(\lambda'\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right)\right)^2 / P\{A(\epsilon_i,X_i)\} < \infty;$$

which, in turn, follows from $\lim_{n\to\infty} P\{A(\epsilon_i,X_i)\} = 1$ and by **C4**.

Case II: $l > 1$. This involves more tedious notations but follows the same logic. Note that $\mathbf{W}_n - \bar{\mathbf{W}}_n = o_p(1)$, where

$$\bar{\mathbf{W}}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right)1\left(\epsilon_i > \{\Delta(X_i)'u_j/n\}\vee 0 \text{ or } \epsilon_i < \{\Delta(X_i)'u_j/n\}\wedge 0, \text{ for all } j \le l\right).$$

Hence it suffices to show the asymptotic independence between $\left(Q_n^d(u_j), j \le l\right)$ and $\bar{\mathbf{W}}_n$. By the Portmanteau Lemma it suffices to show that for any real x_j and y and any integers k_j and k for $j = 1, ..., l$

$$\begin{aligned} &\limsup_{n\to\infty} P\left\{Q_n^p(u_1) = k_1, Q_n^d(u_1) \le x_1, ..., Q_n^p(u_l) = k_l, Q_n^d(u_l) \le x_l, Q_n^t = k, \bar{\mathbf{W}}_n \le y\right\} \\ &\le P\left\{Q_\infty^p(u_1) = k_1, Q_\infty^d(u_1) \le x_1, ..., Q_\infty^p(u_l) = k_l, Q_\infty^d(u_l) \le x_l, Q_\infty^t = k\right\}\cdot P\{\mathbf{W} \le y\} \end{aligned} \qquad \text{(F.8)}$$

where

$$Q_n^t = \sum_{i=1}^{n}\left(1\left[0 < n\epsilon_i < \Delta(X_i)'u_j, \text{ for some } j \le l\right] + 1\left[0 > n\epsilon_i > \Delta(X_i)'u_j, \text{ for some } j \le l\right]\right)$$

and

$$Q_\infty^t = \sum_{i=1}^{\infty}\left(1\left[0 < J_i < \Delta(\mathcal{X}_i)'u_j, \text{ for some } j \le l\right] + 1\left[0 > J_i' > \Delta(\mathcal{X}_i')'u_j, \text{ for some } j \le l\right]\right).$$

By iid sampling in **C0**, the left hand side of (F.8) (without limsup) equals

$$\begin{aligned} &P\left\{Q_n^p(u_1) = k_1, Q_n^d(u_1) \le x_1, ..., Q_n^p(u_l) = k_l, Q_n^d(u_l) \le x_l, Q_n^t = k\right\} \\ &\times \underbrace{P\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n-k}\frac{\partial}{\partial\alpha}\ln f\left(\epsilon_i|X_i,\gamma_0\right) \le y \middle| \epsilon_i > \{\Delta(X_i)'u_j/n\}\vee 0 \text{ or } \epsilon_i < \{\Delta(X_i)'u_j/n\}\wedge 0, \text{ for all } j \le l, \text{ for all } i \le n-k\right\}}_{\equiv P\left\{\sqrt{\frac{n-k}{n}}\widetilde{\mathbf{W}}_n \le y\right\}}, \end{aligned}$$

where $\widetilde{\mathbf{W}}_n = \frac{1}{\sqrt{n-k}}\sum_{i=1}^{n-k}\frac{\partial}{\partial\alpha}\ln f\left(\tilde{\epsilon}_i|\tilde{X}_i,\gamma_0\right)$ and $\tilde{\epsilon}_i, \tilde{X}_i$ are i.i.d. draws from the distribution of ϵ_i, X_i conditional on

$$A(\epsilon_i, X_i) = \{\epsilon_i > \{\Delta(X_i)'u_j/n\}\vee 0 \text{ or } \epsilon_i < \{\Delta(X_i)'u_j/n\}\wedge 0, \text{ for all } j \le l\}.$$

Similarly to the proof of Theorem 3.1, convergence $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ implies that by the Portmanteau Lemma that

$$\limsup_{n\to\infty} P\left\{Q_n^p(u_1) = k_1, Q_n^d(u_1) \le x_1, ..., Q_n^p(u_l) = k_l, Q_n^d(u_l) \le x_l, Q_n^t = k\right\}$$
$$\le P\left\{Q_\infty^p(u_1) = k_1, Q_\infty^d(u_1) \le x_1, ..., Q_\infty^p(u_l) = k_l, Q_\infty^d(u_l) \le x_l, Q_\infty^t = k\right\}$$

Finally, that $\lim_{n\to\infty} P\{\sqrt{\frac{n-k}{n}}\widetilde{\mathbf{W}}_n \le y\} = P\{W \le y\}$ follows similarly to the proof of (a)-(c) in Step 2 in Case I (when $l = 1$). ■

F.3. **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 an argument similar to that given by Ibragimov and Has'minskii (1981b) p.93.

Let $Z_n^\delta \equiv H_n^{-1}(\widehat{\gamma}_{\rho,\mu,n} - \gamma_n(\delta))$, where index δ emphasizes the dependence of the distribution of Z_n^δ on the local parameter sequence $\gamma_n(\delta)$. Define

$$I(K) \equiv \limsup_{n\to\infty} I_n(K), \quad I_n(K) \equiv \frac{1}{\lambda(K)}\int_K E_{P_{\gamma_n(\delta)}}\rho(Z_n^\delta)d\delta.$$

It follows from Fatou's lemma and conclusion (B.7) that $I(K) = \frac{1}{\lambda(K)}\int_K E_{P_{\gamma_0}}\rho(Z_\infty)d\delta = E_{P_{\gamma_0}}\rho(Z)$. Thus $I \equiv \limsup_{K\uparrow\mathbb{R}^d} I(K) = E_{P_{\gamma_0}}\rho(Z)$.

Next let $Z_n^\delta(K) = H_n^{-1}(\widehat{\gamma}_{\rho,\lambda_K,n} - \gamma_n(\delta))$, where $\widehat{\gamma}_{\rho,\lambda_K,n}$ is the Bayes estimator defined with respect to the loss function ρ and prior $\lambda_K(\gamma) = 1\{H_n^{-1}(\gamma - \gamma_0) \in K\}$. Define

$$II(K) \equiv \limsup_{n\to\infty} II_n(K), \quad II_n(K) \equiv \frac{1}{\lambda(K)}\int_K E_{P_{\gamma_n(\delta)}}(Z_n^\delta(K))d\delta.$$

By Lemma B.2 and Lemma B.1 it follows that for any $\delta \in \mathbb{R}^d$

$$Z_n^\delta(K) \to_d Z^\delta(K) \equiv \arg\inf_{z\in\mathbb{R}^d}\int_K \rho(z - (\eta - \delta))\frac{\ell_\infty(\eta - \delta)}{\int_K \ell_\infty(\bar{\eta} - \delta)d\bar{\eta}}d\eta.$$

The property – $P_{\gamma_n(\delta)}\{Z_n^\delta(K) > |K| + \delta\} = 0$, for $|K| = \sup\{|z| : z \in K\}$ for any $\delta \in \mathbb{R}^d$ and $n \le \infty$, – provides the necessary uniform integrability to conclude $\lim_{n\to\infty} E_{P_{\gamma_n(\delta)}}\rho(Z_n^\delta(K)) = E_{P_{\gamma_0}}\rho(Z^\delta(K))$; which by Fatou's lemma implies that $II(K) = \frac{1}{\lambda(K)}\int_K E_{P_{\gamma_0}}\rho(Z^\delta(K))d\delta$.

By finite-sample average risk efficiency of the Bayes estimator $\widehat{\gamma}_{\rho,\lambda_K,n}$

$$II_n(K) \le I_n(K) \text{ for each } n, \text{ hence } II(K) \le I(K) = I.$$

Then $\limsup_{K\uparrow\mathbb{R}^d} II(K) = I$ follows from **(a)** $II(K) \le I$ for each K, **(b)** noting that for any $\delta \in \mathbb{R}^d$ as $K \uparrow \mathbb{R}^d$ $Z^\delta(K) \to_p Z$, and **(c)** dominated convergence theorem, as shown below.

The claim **(2)** now follows. Indeed, suppose there exists an estimator sequence $\{\widehat{\gamma}_n\}$ that achieves a strictly lower asymptotic average risk. Define $\widehat{Z}_n^\delta \equiv H_n^{-1}(\widehat{\gamma}_n - \gamma_n(\delta))$, then it must be that for some K, n_0, and infinitely many $n > n_0$, $\frac{1}{\lambda(K)}\int_K E_{P_{\gamma_n(\delta)}}\rho(\widehat{Z}_n^\delta)d\delta < II_n(K)$, which contradicts to finite-sample average risk-efficiency of the Bayes estimator $\gamma_{\rho,\lambda_K,n}$ for each such n. ■

Proof of $\limsup_{K\uparrow\mathbb{R}^d} II(K) = I$. Rewrite $II(K) \le I(K)$ as $\int_K E_{P_{\gamma_0}}\left[\rho(Z^\delta(K)) - \rho(Z)\right]d\delta/\lambda(K) \le 0$ or

$$\int_K E_{P_{\gamma_0}}\left[\rho(Z^\delta(K)) - \rho(Z)\right]^+ d\delta/\lambda(K) - \int_K E_{P_{\gamma_0}}\left[\rho(Z^\delta(K)) - \rho(Z)\right]^- d\delta/\lambda(K) \le 0. \quad \text{(F.9)}$$

Next as $r(K) \to \infty$

$$\int_K E_{P_{\gamma 0}} \left[\rho(Z^\delta(K)) - \rho(Z)\right]^- d\delta/\lambda(K) = \int_{(-1,1)^d} E_{P_{\gamma 0}} \left[\rho(Z^{\eta r(K)}(K)) - \rho(Z)\right]^- d\eta \to 0, \qquad \text{(F.10)}$$

where $r(K)$ denotes the width of the cube K (which is assumed to be centered at zero). Conclusion (F.10) follows by **(b)** , **(c)**, and the domination (uniform integrability) condition:

$$\text{for any } \eta \in (0,1)^d \text{ and any } K, \left[\rho(Z^{\eta r(K)}(K)) - \rho(Z)\right]^- \leq \rho(Z), \text{ where } E_{P_{\gamma_0}} \rho(Z) < \infty. \qquad \text{(F.11)}$$

But (F.9)- (F.11) imply that it must be that $\int_K E_{P_{\gamma 0}} \left[\rho(Z^\delta(K)) - \rho(Z)\right]^+ d\delta/\lambda(K) \to 0$ as $r(K) \to \infty$. Thus $II(K) - I \to 0$ as $K \uparrow \mathbb{R}^d$. ■

Technical Addendum, Part II: Maximum Likelihood Estimation

This addendum includes the material on the maximum likelihood estimation. The main text only contains the statement of the result. The addendum will be made available as part of a MIT Economics Department Working Paper published by the Social Science Research Network.

Appendix G. Maximum Likelihood Procedures.

The Maximum Likelihood Estimator (MLE) is given by

$$\hat{\gamma} = (\hat{\beta}', \hat{\alpha}')' \equiv \arg\inf_{\gamma \in \mathcal{G}} \; - L_n(\gamma).$$

We obtain various properties of the MLE such as consistency, rates of convergence (n for the parameter β and $\sqrt{n}$ for the parameter α), and its asymptotic distribution. The limit distribution for $\hat{\alpha}$ is the standard one for smooth likelihood analysis. The asymptotic distribution for $\hat{\beta}$ is an extreme type distribution, which may be used for inference in the same way as any standard distribution. For example, denoting the limit variable Z^β and the parameter sequence $\gamma_n(\delta) = \gamma_0 + H_n\delta$,

$$Z_n^\beta \equiv n(\hat{\beta} - \beta_n(\delta)) \to_d Z^\beta,$$

for any continuously differentiable functions r of β,

$$n(r(\hat{\beta}) - r(\beta_0)) \to_d \frac{\partial}{\partial \beta} r(\beta_0)' Z^\beta.$$

Then, quantiles of distribution of $\frac{\partial}{\partial\beta} r(\beta_0)' Z^\beta$ can be estimated by simulating a series of draws of $\frac{\partial}{\partial\beta} r(\hat{\beta})' Z^\beta$, according to the formulas given in this paper, and then used for classical inference and hypothesis testing. (Alternatively, parametric bootstrap may be used). For example, denoting by $\hat{c}(\tau)$ the α-quantile of $\frac{\partial}{\partial\beta} r(\hat{\beta})' Z^\beta$, an asymptotic 90% confidence interval is given by

$$I_{.90} = \left\{ r(\hat{\beta}) - \frac{\hat{c}(.95)}{n}, \;\; r(\hat{\beta}) - \frac{\hat{c}(.05)}{n} \right\}.$$

The limit distribution is also useful for bias correction. For example, to remove the (first-order) asymptotic median bias, simply take $r(\hat{\beta}) - \frac{\hat{c}(.5)}{n}$.

On the other hand, inference about α is standard. The limit distribution of either MLE or BE's (for symmetric loss functions) under the parameter sequence $\gamma_n(\delta) = \gamma_0 + H_n\delta$ is given by

$$Z_n^\alpha \equiv \sqrt{n}(\hat{\alpha} - \alpha_n(\delta)) \to_d Z^\alpha \overset{d}{=} \mathcal{N}(0, \mathcal{J}^{-1}),$$

where $\mathcal{J} \equiv -E \frac{\partial^2}{\partial\alpha\partial\alpha'} \ln f(Y_i - g(X_i, \beta) | X_i, \beta, \alpha)$ is the usual information matrix for α. The limit variable Z^α is in fact independent from Z^β, which follows from the information about β coming from a small fraction of the whole sample located near the jump points, and information about α coming from the whole sample and averaged over (so that impact of the small fraction is negligible).[20]

[20] This intuition is based on e.g. Resnick (1986) and van der Vaart (1999)'s Lemma 21.19 concerning the independence of minimal order statistics and sample averages.

The usual estimates of the information matrix can be used for inference. This result, combined with the one above, can also be used for inference about functions $r(\beta,\alpha)$ of both β and α using the delta-method:

$$\sqrt{n}(r(\hat\beta,\hat\alpha)-r(\beta_0,\alpha_0)) = \frac{\partial r(\beta_0,\alpha_0)}{\partial\beta}'\sqrt{n}(\hat\beta-\beta_0)+\frac{\partial r(\beta_0,\alpha_0)}{\partial\alpha}'\sqrt{n}(\hat\alpha-\alpha_0)+o_p(1/\sqrt{n})$$
$$\overset{d}{\approx}\frac{\partial r(\beta_0,\alpha_0)}{\partial\alpha}'Z^\alpha+o_p(1),$$

Also, in this case it is possible to use the second order expansion, from which $(\hat\beta-\beta_0)$ does not vanish, to better capture the estimation uncertainty, i.e.

$$\sqrt{n}(r(\hat\beta,\hat\alpha)-r(\beta_0,\alpha_0))\overset{d}{\approx}\frac{\partial r(\beta_0,\alpha_0)}{\partial\beta}'\frac{Z^\beta}{\sqrt{n}}+\frac{\partial r(\beta_0,\alpha_0)}{\partial\alpha}'Z^\alpha+\frac{Z^{\alpha\prime}}{\sqrt{n}}\frac{1}{2}\frac{\partial^2 r(\beta_0,\alpha_0)}{\partial\alpha\partial\alpha'}Z^\alpha+o_p(1/\sqrt{n}).$$

In many situations, such as the previous auction example, the functions of prime interest depend only on β, and the regular parameter α is not present. Quantiles of the above distributional approximation can be obtained by simulation, which consists of making draws of the variables Z^β and Z^α, independently of each other, evaluating the above expressions (with derivatives of function r evaluated at the estimates), and then taking the appropriate quantiles of the simulated series. The resulting quantiles can be used for classical Wald intervals and hypothesis testing.

Theorem G.1 (*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*

$$Z_n \to_d Z \equiv \text{arginf}_{z\in\mathbb{R}^d} - \ell_\infty(z).$$

Particularly, $Z_n^\alpha \to_d Z^\alpha = \mathcal{J}^{-1}\mathbf{W} \overset{d}{=} \mathcal{N}(0,\mathcal{J}^{-1})$, $Z_n^\beta \to_d Z^\beta = \text{argmin}_{u\in\mathbb{R}^{d_\beta}} - \ell_{2\infty}(u)$, and Z^β and Z^α are independent.

This result states the consistency, the rates of convergence, and the limit distribution of the MLE. The limit is given in form of argmin of a limit likelihood. Due to asymptotic independence of the information about the shape parameter from the information about the location parameter, the MLE's for these parameters are asymptotically independent.

Remark G.1 (*Boundary Models*). In the boundary models the limit result can be made more explicit:

$$\sqrt{n}(\hat\alpha-\alpha_n(\delta))\to_d \arg\sup_v\Big(\mathbf{W}'v-v'\mathcal{J}v/2\Big)=\mathcal{J}^{-1}\mathbf{W}\overset{d}{=}\mathcal{N}(0,\mathcal{J}^{-1}),$$

and by (3.7)

$$n(\hat\beta-\beta_n(\delta))\to_d \arg\inf_u\Big(-\exp(u'\mathrm{m}) \quad \text{such that } J_i\geq\Delta(\mathcal{X}_i)'u, \text{for all } i\geq 1\Big),$$
$$=\arg\sup_u\Big(\quad u'\mathrm{m} \quad \text{such that } J_i\geq\Delta(\mathcal{X}_i)'u, \text{for all } i\geq 1\Big).$$

The limit distribution of MLE is thus convenient in the boundary models and can be simulated by solving an L_1-linear programming problem, which can be done at the speeds of OLS through the use of interior point algorithms, cf. Portnoy and Koenker (1997). In contrast, bootstrapping requires repeated solutions of a nonlinear programming problem and is much less practical.

Remark G.2 (*Uniqueness*). The condition that $-\ell_\infty(z)$ attains a unique minimum a.s. is necessary and does not appear to be problematic. The invertibility of the information matrix $\mathcal{J}$ guarantees the uniqueness of Z^α, and the solutions of linear programs like those above are unique under mild conditions, which guarantees uniqueness of Z^{β}. E.g. when $\Delta(\mathcal{X}_i)$'s support does not concentrate on a proper linear subspace of lower dimension, has an absolutely continuous component and the variables J_i are absolutely continuous, see e.g. Portnoy (1991) for a related problem. Also when $\Delta(\mathcal{X}_i)$'s have discrete support the limit result corresponds to the result of Donald and Paarsch (1993a) who show that uniqueness holds in that case too.

G.1. **Epi-graphical Convergence.** Epi-convergence in distribution has been developed in the work on the stochastic approximation of optimization problems, cf. Knight (2000), Pflug (1995), Salinetti and Wets (1986), Rockafellar and Wets (1998), among others. Suppose that the sequence of objectives $\{Q_n\}$ are random lower semi-continuous (l-sc) functions, that is for each n, $Q_n(x) \leq \liminf_{x_j \to x} Q_n(x_j)$, $\forall x$, $\forall x_j \to x$. Let $\mathcal{L}$ be the space of l-sc functions $f : \mathbb{R}^d \to \bar{\mathbb{R}} = [-\infty, +\infty]$ such that $f \not\equiv \infty$.

Q_n is said to *epi-converge in distribution* to Q in $\mathcal{L}$ if for any closed rectangles $R_1, ..., R_k$ in $\mathbb{R}^d$ with open interiors $R_1^o, ..., R_k^o$, and any real $r_1, ..., r_k$:

$$\begin{aligned} &P\Big\{ \inf_{x\in R_1} Q(x) > r_1, ..., \inf_{x \in R_k} Q(x) > r_k \Big\} \\ &\overset{(1)}{\leq} \liminf_n P\Big\{ \inf_{x\in R_1} Q_n(x) > r_1, ..., \inf_{x\in R_k} Q_n(x) > r_k \Big\} \\ &\overset{(2)}{\leq} \limsup_n P\Big\{ \inf_{x\in R_1^o} Q_n(x) \geq r_1, ..., \inf_{x\in R_k^o} Q_n(x) \geq r_k \Big\} \\ &\overset{(3)}{\leq} P\Big\{ \inf_{x\in R_1^o} Q(x) \geq r_1, ..., \inf_{x\in R_k^o} Q(x) \geq r_k \Big\}. \end{aligned} \tag{G.1}$$

Note that the inequality (2) is simply by lower-semi-continuity. Epi-convergence is a weak condition that leads to weak convergence of argmins. It is also an evident condition, since $P[\arg\inf_{x\in K} Q_n(x) \leq a] = P[\inf_{x\in K, x\leq a} Q_n(x) < \inf_{x\in K, x\nleq a} Q_n(x)]$. Thus if one can characterize the joint distribution of the terms $\inf_{x\in K, x\leq a} Q_n(x)$ and $\inf_{x\in K, x\nleq a} Q_n(x)$, one obtains the limit distribution of argmin.

The following lemma is given in Knight (2000), Pflug (1995) Salinetti and Wets (1986), among others.

Lemma G.1. *Let Z_n be s.t. $Q_n(Z_n) \leq \inf_{z\in\mathbf{R}^d} Q_n(z) + \epsilon_n$, $\epsilon_n \searrow 0$, and suppose*

- **i.** $Z_n = O_p(1)$,
- **ii.** $Z_\infty \equiv \arg\inf_{z\in\mathbf{R}^d} Q_\infty(z)$ *is uniquely defined in $\mathbb{R}^d$ a.s., and*
- **iii.** $Q_n(\cdot)$ *epi-converges in distribution to $Q_\infty(\cdot)$,*
  then $Z_n \to_d Z_\infty$.

Epi-convergence is more general than uniform convergence, because it allows for non-vanishing discontinuities. In our case, the non-vanishing discontinuities make the uniform convergence of the likelihood function impossible. The recent remarkable work of Knight (2000) provides convenient sufficient conditions for verifying epi-convergence, which amount to converting the finite-dimensional limits to epi-limits via a device called stochastic equisemicontinuity. The work extends Salinetti and Wets (1986) by defining an "in probability" version of stochastic equisemicontinuity a.s. We shall prove epi-convergence directly, albeit borrowing the general structure of the proof from Knight (2000). In fact, part of the proof replicates the proof of Theorem 2 of Knight (2000).

G.2. **Proof of Theorem G.1.** First, note that the MLE and other variables such as $\inf_{z\in K} -Q_n(z)$, are measurable by Proposition 3.2 in Dupačová and Wets (1988), given **C0-C3.**

Second, we use Lemma G.1 on epi-convergence to prove the result. By definition

$$Z_n = \arg\sup_{z\in U_n} \ell_n(z) = \arg\inf_{z\in U_n} -Q_n(z),$$

where U_n is the rescaled parameter space $\sqrt{n}(\mathcal{A} - \alpha_n(\delta)) \times n(\mathcal{B} - \beta_n(\delta))$, $Q_n(z)$ is defined in the proof of Theorem 3.1. It will be proved that

$$Z_n \rightarrow_d Z = \arg\sup_{z\in \mathbf{R}^d} \ell_\infty(z) = \arg\inf_{z\in \mathbf{R}^d} -Q_\infty(z),$$

where $Q_\infty(z)$ is defined in the proof of Theorem 3.1.

Lemma G.1 may be verified by checking three conditions:

(a) epi-convergence in distribution of $-Q_n$ to its finite-dimensional limit $-Q_\infty$,
(b) $Z_n = O_p(1)$, and
(c) uniqueness of Z.

Conditions (c) is assumed. Condition (b) is shown below. It is more difficult to prove (a). The general idea of the proof is borrowed from Knight (2000)'s proof of his Theorem 2. The specifics are based on bounding two types of modulus of continuity by a strategy that is similar to the one in Ibragimov and Has'minskii (1981a).

Definition of epi-convergence in (G.1) consists of parts (1)-(3). We verify part (1) only, part (3) follows almost identically, and part (2) holds trivially (by definition of lower-semi-continuity.) For notation sake, in what follows we do not index P by $\gamma_n(\delta)$.

Given a collection of rectangles R_1,..., R_k, write

$$P\Big\{\inf_{z\in R_1} -Q_n(z) > r_1, ..., \inf_{z\in R_k} -Q_n(z) > r_k\Big\} = 1 - P\Big\{\cup_{j\le k}\{\inf_{z\in R_j} -Q_n(z) \le r_j\}\Big\}.$$

Thus, to verify (1) in (G.1) it suffices to show

$$\limsup_{n\to\infty} P\Big\{\cup_{j\le k}\{\inf_{z\in R_j} -Q_n(z) \le r_j\}\Big\} \le P\Big\{\cup_{j\le k}\{\inf_{z\in R_j} -Q_\infty(z) \le r_j\}\Big\}. \tag{G.2}$$

To explain the result clearly, first bound the probability of the event

$$\Big\{\inf_{z\in R} -Q_n(z) \le r\Big\} = \Big\{\inf_{z\in R}(-Q_n^c(z) - Q_n^d(u)) \le r\Big\}.$$

Denote $R = R^\beta \times R^\alpha$. Define two sets of grid-points as follows.

Consider the grid of equidistant points $\{v_s\}$ and $\{u_m\}$ inside the rectangles R^α and R^β such that sup-distance between the adjacent points is at most φ. Also cover R^β by the sets V_{jk}, as defined in the proof of Lemma G.2 where and let u_{kj} denote a carefully chosen point inside the cover set V_{kj}, as defined in the proof of Lemma G.2.

Next, define collection of points $\{z_l\}$ as the Cartesian product $\{v_s\} \times (\{u_m\} \cup \{u_{kj}\})$. This collection of grid points has the property that the nearest grid-points are at most φ apart from each other. The collection of $\{z_l\}$ will be used to approximate the $\inf_{z\in R} -Q_n^c(z)$ by $\inf_{z\in\{z_l\}} -Q_n^c(z)$. The collection of $\{u_{kj}\}$ will be used to approximate the behavior of $\inf_{u\in R^\beta} -Q_n^d(u)$ by $\inf_{u\in\{u_{kj}\}} -Q_n^d(u)$.

Then

$$\Big\{\inf_{z\in R} -Q_n(z) \le r\Big\} \subset \Big\{\Big\{\inf_{z\in\{z_l\}} -Q_n^c(z) + \inf_{u\in\{u_{jk}\}} -Q_n^d(u) \le r+\varepsilon\Big\} \cap \Big\{A\Big\}\Big\} \cup A^c, \tag{G.3}$$

where A is the event that the finite-dimensional approximation "works" and A^c is its complement, that is

$$A \equiv \Big\{w_{Q_n^c}(R,\varphi) < \varepsilon,\ \xi_{Q_n^d}(R,\varphi) = 0\Big\},$$

where $w_{Q_n^c}(R,\varphi)$ and $\xi_{Q_n^d}(R,\varphi)$ are the moduli of continuity of the continuous part Q_n^c and discontinuous part Q_n^d, respectively:

$$\begin{aligned} w_{Q_n^c}(R,\varphi) &\equiv \sup_{z_1,z_2\in R:|z_1-z_2|<\varphi} |Q_n^c(z_1) - Q_n^c(z_2)|, \\ \xi_{Q_n^d}(R,\varphi) &\equiv 1\Big[\inf_{u\in R} -Q_n^d(u) < \inf_{u\in\{u_{kj}\}\cap R} -Q_n^d(u)\Big]. \end{aligned} \tag{G.4}$$

The modulus $w_{Q_n^c}(R,\varphi)$ is a standard measure of equicontinuity of Q_n^c. The modulus $\xi_{Q_n^d}(R,\varphi)$ is a Skorohod-type modulus, it tells whether the infimum of the step function $-Q_n^d(z)$ coincides with the minimum of $-Q_n^d(z)$ computed over a finite set of grid points $\{z_l\}$.

Lemma G.3 bounds the probability of A^c:

$$P\Big\{w_{Q_n^c}(R,\varphi) > \varepsilon\Big\} \le \text{const}\cdot|R|\cdot\varepsilon^{-1}\varphi, \quad P\Big\{\xi_{Q_n^d}(R,\varphi) = 1\Big\} \le \text{const}\cdot|R|\cdot\varphi, \tag{G.5}$$

where $|R| = \sup\{|z| : z \in R\}$. For any $\varepsilon > 0$ and given R, we can pick $\varphi(\varepsilon)$ small enough such that the rhs of (G.5) is smaller than $\varepsilon/2$.

Hence

$$P\Big\{\inf_{z\in R} -Q_n(z) \le r\Big\} \le P\Big\{\inf_{z\in\{z_l\}} -Q_n^c(z) + \inf_{u\in\{u_{kj}\}} -Q_n^d(u) \le r+\varepsilon\Big\} + \varepsilon,$$

and by Theorem 3.1 and the Portmanteu Lemma

$$\begin{aligned} \limsup_{n\to\infty} P\Big\{\inf_{z\in R} -Q_n(z) \le r\Big\} &\le P\Big\{\inf_{z\in\{z_l\}} -Q_n^c(z) + \inf_{z\in\{z_{jk}\}} -Q_n^d(z) \le r+\varepsilon\Big\} + \varepsilon \\ &\le P\Big\{\inf_{z\in R} -Q_\infty(z) \le r+\varepsilon\Big\} + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary

$$\limsup_{n\to\infty} P\Big\{\inf_{z\in R} -Q_n(z) \le r\Big\} \le P\Big\{\inf_{z\in R} -Q_\infty(z) \le r\Big\}.$$

Thus for $\cup_j R_j \subset R$, it follows that

$$\limsup_{n\to\infty} P\Big\{\cup_{j=1}^k\Big\{\inf_{z\in R_j} -Q_n(z) \le r_j\Big\}\Big\} \le P\Big\{\cup_{j=1}^k\Big\{\inf_{z\in R_j} -Q_\infty(z) \le r_j+\varepsilon\Big\}\Big\} + \varepsilon.$$

Since ε is arbitrary, the required conclusion (G.2) follows.

Finally, it remains to establish $Z_n = O_p(1)$. First, the MLE $\hat\gamma$ is consistent by a generalization of Wald's Theorem on Consistency of MLE – Theorem 3.3 of Artstein and Wets (1995) – which requires

(a) $\gamma \mapsto -\ln f(Y_i - g(X_i,\beta)|X_i,\gamma)$ is a.s. lower semi-continuous (which is true by **C2** and **C3**),
(b) the domination: $\sup_\gamma E_{P_\gamma} \sup_{\gamma'} \ln f(Y_i - g(X_i,\beta')|X_i,\gamma') \le \ln \bar f < +\infty$, which is true by **C2**,
(c) identification (by **C0**),
(d) compact parameter space (by **C0**).

Consistency implies wp $\to 1$ $Z_n \in S_n = \{z : |v|/\sqrt{n} \le \varepsilon_n, |u|/n \le \varepsilon_n\}$ for some $\varepsilon_n \to 0$; which allows the use of inequalities (G.6) and (G.7) proved in Lemma G.3, along with the exponential inequality for $E\ell_n^{1/2}(z)$ given in Lemma C.2. These inequalities imply, by a standard argument like that on p. 265 in Ibragimov and Has'minskii (1981a) [see Lemma G.4], that for sufficiently large $A > 0$ and any $N > 0$

$$P\Big\{ \sup_{z\in S_n:|z|>A} \ell_n(z) > A^{-N}\Big\} \le C_N A^{-N}, \text{ where } C_N > 0.$$

Hence $P\{|Z_n| > A\} \le C_N A^{-N}$, and it follows that $Z_n = O_p(1)$. ■

Lemma G.2 (Lipschitz Continuity of Q_n^c). *Under C0-C5, for a small $\delta > 0$, there is a random variable $C_n > 0$ such that for all $n > n_0$ and some large n_0*

$$|Q_n^c(z_1) - Q_n^c(z_2)| \le C_n|z_1 - z_2|(|z_1| + 1), \qquad \sup_{n>n_0,\gamma\in B_\delta(\gamma_0)} E_{P_\gamma} C_n < \infty,$$

uniformly over all $|z_1 - z_2| \le 1$ in the set $S_n = \{z : |v|/\sqrt{n} \le \varepsilon_n, |u|/n \le \varepsilon_n\}$, where $\varepsilon_n \to 0$.

Lemma G.3 (Bounding Moduli of Continuity). *Under C0-C5, for all $n : N \ge n_0$, where n_0 is sufficiently large, for some $\delta > 0$, and any bounded rectangle $R \subset S_n$:*

1. For all sufficiently small $\varphi > 0$ and $\varepsilon > 0$

$$\sup_{\gamma\in B_\delta(\gamma_0)} P_\gamma(w_{Q_n^c}(R,\varphi) > \epsilon) \le \text{const}\cdot|R|\cdot\varepsilon^{-1}\varphi, \tag{G.6}$$

2. For all sufficiently small $\varphi > 0$

$$\sup_{\gamma\in B_\delta(\gamma_0)} P_\gamma(\xi_{Q_n^d}(R,\varphi) = 1) \le \text{const}\cdot|R|\cdot\varphi, \tag{G.7}$$

where $|R| = \sup\{|z| : z \in R\}$, and $w_{Q_n^c}(\cdot)$ and $\xi_{Q_n^d}(\cdot)$ are defined in the proof of Theorem G.1.

Lemma G.4 (Tail Bound). *Under C0-C5, for sufficiently large $A > 0$ and any $N > 0$*

$$P_\gamma\Big\{ \sup_{z\in S_n:|z|>A} \ell_n(z) > A^{-N}\Big\} \le C_N A^{-N}, \text{ where } C_N > 0.$$

uniformly in $\gamma \in B_\delta(\gamma_0)$ for some $\delta > 0$.

G.3. **Proof of Lemma G.2(short proof).** The Lemma G.2 is needed for the MLE part only. The result follows by a standard empirical process argument, noting that the object of interest is a function that is an average and that is a spline type object. The result then follows by the Taylor-like expansion and obtaining expressions of the from $C_n(\epsilon_i, z_1, z_2)|z_1 - z_2||z_1| + 1|$, and finally applying a maximal moment inequality to the coefficients $C_n(\epsilon_i, z_1, z_2)$, specifically Lemma 19.34 in van der Vaart (1999). [Details are given in the last section of this document.] ■

G.4. **Proof of Lemma G.3.** The first part follows by the Markov inequality and Lemma G.2. The second part is proven below. In the one-dimensional case with no covariates, the argument essentially reduces to the proof given on p.262 in Ibragimov and Has'minskii (1981a).

(a) [Covering Sets.] For a hyper-cube $R = R^\alpha \times R^\beta$, where $R^\alpha \subset \mathbb{R}^{d_\alpha}$ and $R^\beta \subset \mathbb{R}^{d_\beta}$, construct a collection of (possibly overlapping) subsets $\{V_{kj}\}$ of R^β as follows. First cover the support of vector $\Delta(X)$ by the minimal number of closed equal-sized cubes $\{\mathbb{X}_{\phi,j}, j \le J(\phi)\}$ with the side-length of the cube equal to $\phi < 1$. There are $J(\phi) \le$ const $(1/\phi)^{d_\beta}$ such cubes.

Recall that

$$\frac{\partial \Delta_n(X,u)}{\partial u} = \frac{\partial g(x,\beta)}{\partial \beta}\Big|_{\beta=\beta_0+u/n}.$$

Note also that uniformly in u in R^β and uniformly in $R^\beta \subset \{u : \|u\|/n \le \varepsilon_n\}$ for some given sequence $\varepsilon_n \to 0$, we have

$$\frac{\partial \Delta_n(X,u)}{\partial u} = \Delta(X) + o(1).$$

In particular, choose n_0 such that for all $n > n_0$ (n_0 depends on ϕ)

$$\Big|\frac{\partial \Delta_n(X,u)}{\partial u} - \Delta(X)\Big| < \phi^2 \text{ a.s.}$$

Thus, for any given x and any R^β, and given that $\Delta(x) \in \mathbb{X}_{\phi,j}$, we have that

$$\cup_{u\in R^\beta} \frac{\partial \Delta_n(x,u)}{\partial u} \text{ belongs at most to } K^* = 2^{d_\beta} \text{ cubes of the form } \mathbb{X}_{\phi,j'} \text{ that are adjacent to } \mathbb{X}_{\phi,j}. \quad \text{(G.8)}$$

(We only need that K^* is finite and is independent of ϕ and R^β). Thus, in what follows it is helpful to think of $\frac{\partial \Delta_n(x,u)}{\partial u}$ as being equal $\Delta(x)$ and independent from u.

Construct the (overlapping) sets[21]

$$\Big\{V_{kj}, k = -m, ..., m, j = 1, ..., J(\phi)\Big\} \subset \mathbb{R}^{d_\beta}$$

such that

$$V_{kj} \equiv \Big\{u \in \mathbb{R}^{d_\beta} : \underline{v}_k - \varphi \le \Delta_n(x,u) \le \underline{v}_k + \varphi \text{ for all } n > n_0 \text{ and all } x \text{ s.t. } \Delta(x) \in \mathbb{X}_{\phi,j}\Big\},$$

where $\varphi > 0$ and

$$\underline{v}_k = k\varphi, \text{ for } k \in \{-m, ..., 0, ...m\}.$$

Since the range of $|\Delta_n(X,u)|$ is bounded a.s. by $\bar{g}\|R^\beta\|$ for all n, we can cover the range by $2m+1$ brackets of the form $[\underline{v}_k - \varphi, \underline{v}_k + \varphi]$ where $m \le$ const $|R^\beta|/\varphi$, $|R^\beta| = \sup\{|u| : u \in R^\beta\}$. Choose

$$\phi \propto \varphi^2/|R^\beta| \quad \text{(G.9)}$$

for all small φ. Hence the total number L of covering sets V_{kj} is bounded as $L \le (2m+1)\cdot J(\phi)$ and grows at most polynomially in $|R^\beta|$ and in $1/\varphi$.[22]

Next, construct the "centers" u_{kj} in $V_{kj} \cap R^\beta$ so that for all $n > n_0$ [23]

$$\underline{\delta}_{kj} \le \Delta_n(x, u_{kj}) < \underline{\delta}_{kj} + \eta, \quad \forall x : \Delta(x) \in \mathbb{X}_{\phi,j} \quad \text{(G.10)}$$

where

$$\underline{\delta}_{kj} = \inf_{n>n_0,x,u} \Delta_n(x,u) \text{ where inf is taken over } u \in V_{kj} \cap R^\beta \text{ and } x : \Delta(x) \in \mathbb{X}_{\phi,j}.$$

We will need that $\eta : 0 < \eta << \varphi$, i.e. that η is sufficiently small relative to φ. Moreover, in order to satisfy the constraint in (G.10), we need to have ϕ set sufficiently small as well. Setting ϕ small restricts the variation of $\Delta(x)$ and hence of $\partial\Delta_n(x, u_{kj})/\partial u$ at most to const $\cdot\, \phi$ when $x : \Delta(x) \in \mathbb{X}_{\phi,j}$. Thus, we choose η as $\eta \propto \varphi^2$ and ϕ as stated in (G.9).

[21]The covering sets V_{kj} can be thought of as "approximate linear subspaces" of $\mathbb{R}^{d_\beta}$.

[22]Note also that V_{kj} clearly cover R^β for $n > n_0$, because given u we have $\Delta_n(x,u)$ belongs to at least two different brackets of the form $[\underline{v}_k - \varphi, \underline{v}_k + \varphi]$ for all $n > n_0$, and $\Delta(x) \in \mathbb{X}_{\phi,j}$ for some $\mathbb{X}_{\phi,j}$ that is at most ϕ^2 away from $\partial\Delta_n(x,u)/\partial u$ for all $n > n_0$. Hence $u \in V_{kj}$ for some k and j.

[23]Note that for V_{kj} in the interior of R^β, it is the case that $\underline{\delta}_{kj} = \underline{v}_k - \varphi$; but otherwise, this is not the case.

(b)[**Characterization of Break-Points**] Recall that

$$Q_n^d(u) = \underbrace{\sum_{i=1}^{n}\left[\ln\frac{q(X_i)}{p(X_i)}1\left(0 < n\epsilon_i \le \Delta_n(X_i,u)\right)\right]}_{Q_n^{d+}(u)} + \underbrace{\sum_{i=1}^{n}\left[\ln\frac{p(X_i)}{q(X_i)}1\left(0 > n\epsilon_i \ge \Delta_n(X_i,u)\right)\right]}_{Q_n^{d-}(u)}.$$

We next examine the nature of the discontinuities of $Q_n^d(u)$ by first examining those of $Q_n^{d+}(u)$ and then those of $Q_n^{d-}(u)$.

Suppose we have $n\epsilon_i = \Delta_n(X_i,u)$ for some $u \in V_{kj}$ and $\Delta(X_i) \in \mathbb{X}_{\phi,j}$, then the pair $(n\epsilon_i, X_i)$ is said to induce a break-point in the set V_{kj} and in the bracket $[\underline{v}_k - \varphi, \underline{v}_k + \varphi]$ to which $\Delta_n(X_i,u)$ belongs.[24]

Given that this is the only pair that induces a break-point in V_{kj} it follows that

$$\inf_{u\in V_{kj}\cap R^B} -Q_n^{d+}(u) \neq -Q_n^{d+}(u_{kj}) \quad \text{only if} \quad n\epsilon_i \in [\underline{\delta}_{kj}, \underline{\delta}_{kj} + \eta]$$

since $-Q_n^{d+}(u)$ is piecewise-constant and can only jump *up* if the index $\Delta_n(X_i,u)$ increases.

Thus, what we need is as follows. First, we need to control the probability of the event that more than one break-point happens in any of the brackets of the form $[\underline{v}_k - \varphi, \underline{v}_k + \varphi]$ for $|k| \le m$. This is included in the event that the errors $n\epsilon_i$ are not separated in non-overlapping brackets, which is the event $A_1(R) \equiv \cup_{k\le m}\{$ there are $n\epsilon_i, n\epsilon_{i'} \in [\underline{v}_k - \varphi, \underline{v}_k + \varphi]\}$. Second, we need to control the probability that for all $n\epsilon_i$ that are separated into the brackets $[\underline{v}_k - \varphi, \underline{v}_k + \varphi]$, they do not fall into the "bad subset" $[\underline{\delta}_{kj}, \underline{\delta}_{kj} + \eta]$ of such brackets, given that $\Delta(X_i) \in \mathbb{X}_{\phi,j}$. Formally, conditionally on the complement of $A_1(R)$, i.e. on $A_1^c(R)$ define the event $A_2(R)$ as the union of

$$A_{2i,kj}(R) \equiv \left\{n\epsilon_i \in [\underline{\delta}_{kj}, \underline{\delta}_{kj} + \eta] \Big| n\epsilon_i \in [\underline{v}_k - \varphi, \underline{v}_k + \varphi], \Delta(X_i) \in \mathbb{X}_{\phi,j}, u \in V_{kj}\right\}$$

across $i \le n, |k| \le m, j \le J(\phi)$.

To begin,

$$P\left\{A_1(R)\right\} \le \sum_{|k|\le m}\sum_{i'=1:i'\neq i}^{n}\sum_{i=1}^{n} P\left\{n\epsilon_i, n\epsilon_i' \in [\underline{v}_k - \varphi, \underline{v}_k + \varphi]\right\} \le (2m+1)\cdot(2\bar{f}\varphi)^2 \le \text{ const } |R|\varphi.$$

Denote the total number of $n\epsilon_i$ that fall into brackets of the form $[\underline{v}_k - \varphi, \underline{v}_k + \varphi]$ by $\mathcal{N}_n$. Because (i) any bracket $[\underline{v}_k - \varphi, \underline{v}_k + \varphi]$ overlaps with at most two other brackets and (ii) (G.8) holds for $n > n_0$, there are at most $3 \cdot K^* \cdot \mathcal{N}_n$ neighborhoods of the form V_{kj} in which the break-point may occur (where K^* is defined in (G.8)). Then

$$P\left\{A_2(R)|\mathcal{N}_n, A_1^c(R)\right\} \le 3\cdot K^*\cdot\mathcal{N}_n\cdot \sup_{i\le n,|k|\le m,j\le J(\phi)} P\left\{A_{2i,j,k}(R)\right\} \le 3\cdot K^*\cdot\mathcal{N}_n\cdot(\bar{f}/\underline{f})\cdot(\eta/(2\varphi)).$$

Since $E\{\mathcal{N}_n\} \le nE1(|n\epsilon_i| \le g'\|R\|) \le 2\bar{g}'\|R\|\bar{f}$,

$$P\left\{A_2(R)|A_1^c(R)\right\} \le \text{ const } |R|(\eta/\varphi).$$

Hence (since $\eta \propto \varphi^2$)

$$P\left\{\cup_{kj}\{\inf_{u\in V_{kj}} -Q_n^{d+}(u) \neq -Q_n^{d+}(u_{jk})\}\right\} \le P\left\{A_2(R)|A_1^c(R)\right\} + P\left\{A_1(R)\right\}$$

$$\le \text{ const } |R|(\eta/\varphi + \varphi) \le \text{ const } |R|\varphi.$$

[24]The terminology "break-points" is borrowed from the linear programming literature.

Therefore, conclude that

$$I \equiv P\Big\{ \inf_{u\in R^\beta} -Q_n^{d+}(u) < \inf_{\{u_{kj}\}} -Q_n^{d+}(u_{kj})\Big\} \leq \text{ const } |R|\varphi.$$

Likewise, it follows that for a finite collection of grid points $\{\bar{u}_{kj}\}$

$$II \equiv P\Big\{ \inf_{u\in R^\beta} -Q_n^{d-}(u) < \inf_{\{\bar{u}_{kj}\}} -Q_n^{d}(u_{kj})\Big\} \leq \text{ const } |R|\varphi.$$

Finally,

$$P\Big\{ \inf_{u\in R^\beta} -Q_n^{d}(u) < \inf_{u\in\{u_{kj},\bar{u}_{kj}\}} -Q_n^{d}(u)\Big\} \leq I + II \leq \text{ const } |R|\varphi. \quad \blacksquare$$

G.5. **Proof of Lemma G.4.** This uses the method of the proof on p.265-266 of Ibragimov and Has'minskii (1981a). We need to establish for sufficiently large $A > 1$ and any $N > 0$

$$P_\gamma\Big\{ \sup_{z\in S_n:|z|>A} \ell_n(z) > A^{-N}\Big\} \leq C_N A^{-N}. \tag{G.11}$$

where C_N denotes a generic constant that only depends on N. Let $R(t) = \{z : t \leq |z| \leq t+1\}$. It will suffice to show that for sufficiently large $t \geq A$

$$P_\gamma\Big\{ \sup_{z\in R(t)\cap S_n} \ell_n(z) > t^{-N}\Big\} \leq C_N t^{-N}, \tag{G.12}$$

since then

$$P_\gamma\Big\{ \sup_{z\in S_n:|z|>A} \ell_n(z) > A^{-N}\Big\} \leq \sum_{t=0}^{\infty} P_\gamma\Big\{ \sup_{z\in R(A+t)\cap S_n} \ell_n(z) > (A+t)^{-N-1}\Big\} \leq C_N A^{-N}.$$

Next cover $R(t)$ by grid-points $\{z_l\}$ in the way defined in the proof of Theorem G.1. It follows that

$$P_\gamma\Big\{ \sup_{z\in R(t)\cap S_n} \ell_n(z) > t^{-N}\Big\} \leq \underbrace{P_\gamma\Big\{ \sup_{z\in\{z_l\}} \ell_n(z) > t^{-N}/2\Big\}}_{I} + \underbrace{P_\gamma\Big\{ w_{Q_n^c}(R(t),\varphi) > |\ln(t^{-N}/2)| \cup \xi_{Q_n^d}(R(t),\varphi) = 1\Big\}}_{II},$$

where $w_{Q_n^c}$ and $\xi_{Q_n^d}$ are the moduli of continuity defined in (G.4).

The number of partition points $\{z_l\}$ is bounded by $L \leq$ const $\cdot (|R(t)|/\varphi)^\kappa$, where $|R(t)| = \sup\{|z| : z \in R(t)\} = t+1$, that is $L \leq$ const $(t+1)^\kappa\varphi^{-\kappa}$, where $1 < \kappa < \infty$ (κ is given in the proof of Lemma G.2).

By Lemma C.2 and Markov inequality, $P_\gamma\{\ell_n(z_l) > t^{-N}/2\} \leq$ const $t^N \exp(-b|t|)$. Hence for $t \geq A$

$$I \leq \text{ const } \cdot (t+1)^\kappa \varphi^{-\kappa} \cdot t^N \cdot e^{-b|t|}. \quad b > 0 \tag{G.13}$$

By Lemma G.3 noting that $|R(t)| = t+1$

$$II \leq \text{ const } (t+1)|\ln(t^{-N}/2)|^{-1}\varphi + \text{ const } (t+1)\varphi. \tag{G.14}$$

Select $\varphi = t^{-2N-1}$, then (G.12) immediately follows from (G.13) and (G.14). $\blacksquare$

G.6. **Proof of Lemma G.2(Detailed proof).** We have

$$Q_n^c(z) \equiv \underbrace{\sum_{i=1}^n \hat{r}_{in}(z) \times [1(\epsilon_i > \{\Delta_n(X_i,u)/n\} \vee 0) + 1(\epsilon_i < \{\Delta_n(X_i,u)/n\} \wedge 0)]}_{Q_{1n}^c(z)}$$
$$+ \underbrace{\sum_{i=1}^n (\hat{r}_{in}(z) - r_{in}(z)) \times [1(0 < \epsilon_i \le \Delta_n(X_i,u)/n) + 1(0 > \epsilon_i \ge \Delta_n(X_i,u)/n)]}_{Q_{2n}^c(z)}.$$

Recall that $Q_{2n}^c(z) \equiv 0$ in the one-sided models.

Intuitively, note that the functions of interest are all Lipschitz-smooth (spline-type) objects by construction, given the differentiability assumptions in **C1- C3**. Thus, it is reasonable to expect the final result:

Under **C1-C5**, for a small $\delta > 0$, there is a random variable $C_n > 0$ such that for all $n > n_0$ and some large n_0

$$|Q_n^c(z_1) - Q_n^c(z_2)| \le C_n |z_1 - z_2|(|z_1| + 1), \qquad \sup_{n>n_0, \gamma \in B_\delta(\gamma_0)} E_{P_\gamma} C_n < \infty,$$

uniformly over all $|z_1 - z_2| \le 1$ in the set $S_n = \{z : |v|/\sqrt{n} \le \varepsilon_n, |u|/n \le \varepsilon_n\}$, where $\varepsilon_n \to 0$.

The proof is tedious but it does have a very simple structure. Given some careful Taylor-type expansions, the maximal inequalities will be applied to the coefficients of those expansions to obtain the required result.

Split $Q_{1n}^c(z_1) - Q_{1n}^c(z_2)$ into two terms

$$I = \sum_{i=1}^n \hat{r}_{in}(z_1) 1(\epsilon_i > \{\Delta_n(X_i,u_1)/n\} \vee 0) - \hat{r}_{in}(z_2) 1(\epsilon_i > \{\Delta_n(X_i,u_2)/n\} \vee 0),$$

$$II = \sum_{i=1}^n \hat{r}_{in}(z_1) 1(\epsilon_i < \{\Delta_n(X_i,u_1)/n\} \wedge 0) - \hat{r}_{in}(z_2) 1(\epsilon_i < \{\Delta_n(X_i,u_2)/n\} \wedge 0).$$

We focus on term I, and only briefly indicate the differences for term II. The term I can bounded as

$$I_1 - I_2 \le I \le I_1 + I_2,$$

where

$$I_1 \equiv \sum_{i=1}^n 1(\epsilon_i > \{\Delta_n(X_i,u_1)/n\} \vee \{\Delta_n(X_i,u_2)/n\} \vee 0)$$
$$\times (\ln f(\epsilon_i - \Delta_n(X_i,u_1)/n | X_i; \beta_0 + u_1/n, \alpha_0 + v_1/\sqrt{n})$$
$$- \ln f(\epsilon_i - \Delta_n(X_i,u_2)/n | X_i; \beta_0 + u_2/n, \alpha_0 + v_2/\sqrt{n}))$$

$$I_2 \equiv \sum_{i=1}^n 1(0 < \epsilon_i \in [\Delta_n(X_i,u_1)/n, \Delta_n(X_i,u_2)/n])$$
$$\times \max_{j=1,2} \left| \ln \frac{f(\epsilon_i - \Delta_n(X_i,u_j)/n | X_i; \beta_0 + u_j/n, \alpha_0 + v_j/\sqrt{n})}{f(\epsilon_i | X_i; \beta_0, \alpha_0)} \right|,$$

where $[a,b]$ denotes $\{x : a \le x \le b \text{ or } b \le x \le a\}$.

Analogously approximate the term II as follows:

$$II_1 - II_2 \le II \le II_1 + II_2,$$

where

$$II_1 \equiv \sum_{i=1}^{n} 1(\epsilon_i < \{\Delta_n (X_i, u_1)/n\} \wedge \{\Delta_n (X_i, u_2)/n\} \wedge 0)$$
$$\times (\ln f (\epsilon_i - \Delta_n (X_i, u_1)/n | X_i; \beta_0 + u_1/n, \alpha_0 + v_1/\sqrt{n})$$
$$- \ln f (\epsilon_i - \Delta_n (X_i, u_2)/n | X_i; \beta_0 + u_2/n, \alpha_0 + v_2/\sqrt{n}))$$
$$II_2 \equiv \sum_{i=1}^{n} 1(0 > \epsilon_i \in [\Delta_n (X_i, u_1)/n, \Delta_n (X_i, u_2)/n])$$
$$\times \max_{j=1,2} \left| \ln \frac{f (\epsilon_i - \Delta_n (X_i, u_j)/n | X_i; \beta_0 + u_j/n, \alpha_0 + v_j/\sqrt{n})}{f (\epsilon_i | X_i; \beta_0, \alpha_0)} \right|,$$

where $[a, b]$ denotes $\{x : a \leq x \leq b \text{ or } b \leq x \leq a\}$.

Part I. Bounds on Terms I_1 and II_1. Term I_1 equals by Taylor expansion

$$\underbrace{\sum_{i=1}^{n} 1 (\epsilon_i > \{\Delta_n (X_i, u_1)/n\} \vee \{\Delta_n (X_i, u_2)/n\} \vee 0) \times \frac{\partial}{\partial u} \ln f (\epsilon_i - \Delta_n (X_i, u^*)/n | X_i; \beta_0 + u^*/n, \alpha_0 + v^*/\sqrt{n}) (u_1 - u_2)/n}_{I_{11}} +$$

$$\underbrace{\sum_{i=1}^{n} 1 (\epsilon_i > \{\Delta_n (X_i, u_1)/n\} \vee \{\Delta_n (X_i, u_2)/n\} \vee 0) \frac{\partial}{\partial \alpha} \ln f (\epsilon_i | X_i, \gamma_0) (v_1 - v_2)/\sqrt{n}}_{I_{12}} +$$

$$\underbrace{\sum_{i=1}^{n} 1 (\epsilon_i > \{\Delta_n (X_i, u_1)/n\} \vee \{\Delta_n (X_i, u_2)/n\} \vee 0) \times (H_n z^*)' \frac{\partial^2}{\partial \gamma \partial \alpha'} \ln f (\epsilon_i - \Delta_n (X_i, u^{**})/n | X_i; \beta_0 + \frac{u^{**}}{n}, \alpha_0 + \frac{v^{**}}{\sqrt{n}}) (v_1 - v_2)/\sqrt{n}}_{I_{13}}.$$

Analogously decompose the term II_1 as $II_{11} + II_{12} + II_{13}$:

$$\underbrace{\sum_{i=1}^{n} 1 (\epsilon_i < \{\Delta_n (X_i, u_1)/n\} \wedge \{\Delta_n (X_i, u_2)/n\} \wedge 0) \times \frac{\partial}{\partial u} \ln f (\epsilon_i - \Delta_n (X_i, u^*)/n | X_i; \beta_0 + u^*/n, \alpha_0 + v^*/\sqrt{n}) (u_1 - u_2)/n}_{II_{11}} +$$

$$\underbrace{\sum_{i=1}^{n} 1 (\epsilon_i < \{\Delta_n (X_i, u_1)/n\} \wedge \{\Delta_n (X_i, u_2)/n\} \wedge 0) \frac{\partial}{\partial \alpha} \ln f (\epsilon_i | X_i, \gamma_0) (v_1 - v_2)/\sqrt{n}}_{II_{12}} +$$

$$\underbrace{\sum_{i=1}^{n} 1 (\epsilon_i < \{\Delta_n (X_i, u_1)/n\} \wedge \{\Delta_n (X_i, u_2)/n\} \wedge 0) \times (H_n z^*)' \frac{\partial^2}{\partial \gamma \partial \alpha'} \ln f (\epsilon_i - \Delta_n (X_i, u^{**})/n | X_i, \beta_0 + \frac{u^{**}}{n}, \alpha_0 + \frac{v^{**}}{\sqrt{n}}) (v_1 - v_2)/\sqrt{n}}_{II_{13}}.$$

By **C5**, the term $|I_{11} + II_{11}|$ is bounded by

$$\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{C_2(\epsilon_i, X_i)}\right)|u_1 - u_2|, \tag{G.15}$$

where the expectation of the random term is finite and constant across n by iid sampling.

By **C5**, the term $I_{13} + II_{13}$ is bounded by

$$|(|z_1| + 1)(z_1 - z_2)|\frac{1}{n}\sum_{i=1}^{n} C_3(\epsilon_i, X_i), \tag{G.16}$$

where the expectation of the random term is constant for all n.

Write

$$\begin{aligned} I_{12} + II_{22} \leq | & \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\ln f(\epsilon_i|X_i, \gamma_0)(v_1 - v_2)}_{III_1} \\ & \underbrace{-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} 1(0 < \epsilon_i \in [\Delta_n(X_i, u_1)/n, \Delta_n(X_i, u_2)/n])\frac{\partial}{\partial\alpha}\ln f(\epsilon_i|X_i, \gamma_0)(v_1 - v_2)}_{III_2} \\ & \underbrace{-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} 1(0 > \epsilon_i \in [\Delta_n(X_i, u_1)/n, \Delta_n(X_i, u_2)/n])\frac{\partial}{\partial\alpha}\ln f(\epsilon_i|X_i, \gamma_0)(v_1 - v_2)|}_{III_3} \end{aligned} \tag{G.17}$$

By **C5** III_1 has two finite moments, which remain constant for all n. Next we show a bound for III_2 and the same bound for III_3 follows identically. By Lemma 19.34 in van der Vaart (1999)

$$E \sup_{|u_1|/n+|u_2|/n\leq 2\epsilon_n} |III_2 - EIII_2| \leq J_{[]}(F, \mathcal{F}, L_2(P)) < \infty, \tag{G.18}$$

where $J_{[]}(F, \mathcal{F}, L_2(P))$ is the $L_2(P)$ bracketing entropy of the function class, which we rewrite in terms of original parameter

$$\mathcal{F} = \left\{1(g(X_i, \beta_0) < Y_i \in [g(X_i, \beta_1), g(X_i, \beta_2)])\frac{\partial}{\partial\alpha}\ln f(\epsilon_i|X_i, \gamma_0),\ |\beta_1 - \beta_0| + |\beta_2 - \beta_0| \leq 2\varepsilon_n\right\}$$

with the constant envelope $F = \bar{f}'/\underline{f}\mathbf{1}$ by **C3**, where $\mathbf{1}$ is the vector of ones. Note that the bound $|u_1 - u_2|/n \leq \epsilon_n$ eventually puts $\beta_0 + u_j/n$'s in a small fixed neighborhood of β_0 for $n > n_0$, and also puts $\Delta_n(X_i, u_j)/n$'s in any small fixed neighborhood of 0.

The entropy in (G.18) is finite uniformly in n by a standard argument, because $\mathcal{F}$ is formed as a product of a Donsker class

$$\left\{1(g(X_i, \beta_0) < Y_i \in [g(X_i, \beta_1), g(X_i, \beta_2)]),\ |\beta_1 - \beta_0| + |\beta_2 - \beta_0| \leq 2\varepsilon_n\right\}$$

(see type V functions in Andrews (1994)) and a bounded by F random variable, which by Theorem 2.10.6 in van der Vaart and Wellner (1996) implies that $\mathcal{F}$ is Donsker, and thus $J_{[]}(F, \mathcal{F}, L_2(P)) < \infty$.

Also note that by **C2** and **C3**:

$$|EIII_2| = E\sqrt{n}\int_{\Delta_n(X_i,u_1)/n}^{\Delta_n(X_i,u_2)/n}\left(\frac{|\frac{\partial}{\partial\alpha}f(\epsilon|X_i,\gamma_0)|}{f(\epsilon|X_i,\gamma_0)}\right)f(\epsilon|X_i,\gamma_0)|1(\epsilon>0)d\epsilon$$
$$\leq E\sqrt{n}\int_{\Delta_n(X_i,u_1)/n}^{\Delta_n(X_i,u_2)/n}\bar{f}'1(\epsilon>0)d\epsilon \leq \bar{g}\cdot\bar{f}'\cdot\|u_1-u_2\|/\sqrt{n},$$

where the constants are defined in Lemma A.1.

Thus, for some random variable C_n with bounded expectation uniformly in n

$$|I_{12}+II_{22}| \leq C_n|v_1-v_2|.$$

Now collecting all the bounds established so far, we have for some random variable C_n with bounded expectation uniformly in n

$$|I_1+II_1| \leq C_n|z_1-z_2|(|z_1|+1). \tag{G.19}$$

Part II. Bounds on Terms I_2 and II_2. Let's get back now to the terms I_2 and II_2. Recall that

$$I_2 \equiv \sum_{i=1}^{n} 1(0<\epsilon_i\in[\Delta_n(X_i,u_1)/n,\Delta_n(X_i,u_2)/n])$$
$$\times\max_{j=1,2}\left|\ln\frac{f(\epsilon_i-\Delta_n(X_i,u_j)/n|X_i;\beta_0+u_j/n,\alpha_0+v_j/\sqrt{n})}{f(\epsilon_i|X_i;\beta_0,\alpha_0)}\right|,$$

and that

$$II_2 \equiv \sum_{i=1}^{n} 1(0>\epsilon_i\in[\Delta_n(X_i,u_1)/n,\Delta_n(X_i,u_2)/n])$$
$$\times\max_{j=1,2}\left|\ln\frac{f(\epsilon_i-\Delta_n(X_i,u_j)/n|X_i;\beta_0+u_j/n,\alpha_0+v_j/\sqrt{n})}{f(\epsilon_i|X_i;\beta_0,\alpha_0)}\right|.$$

By **C2-C3** I_2 is bounded by

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}1(0<\epsilon_i\in[\Delta_n(X_i,u_1)/n,\Delta_n(X_i,u_2)/n])(\bar{f}'/\underline{f})}_{I_{21}}\times\|u_1-u_2\|,$$

where the constants are defined in Lemma A.1.

By **C2-C3** II_2 is bounded by

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}1(0>\epsilon_i\in[\Delta_n(X_i,u_1)/n,\Delta_n(X_i,u_2)/n])(\bar{f}'/\underline{f})}_{II_{21}}\times\|u_1-u_2\|,$$

where the constants are defined in Lemma A.1. Then, by an argument that is identical to the proof of inequality (G.18) we obtain that uniformly in n

$$E\sup_{u_1,u_2}|\sqrt{n}(I_{21}-EI_{21})|<\infty,$$

and identical bound follows for the term II_{21}:

$$E\sup_{u_1,u_2}|\sqrt{n}(II_{21}-EII_{21})|<\infty.$$

Furthermore by **C2- C3** $E(I_{21}+II_{21})\leq \text{const}\cdot|u_1-u_2|$.

Hence for a random variable C_n with uniformly bounded expectation (uniformly in n and in γ)

$$|I_2 + II_2| \leq |I_{21} + II_{21}| \leq \text{const}(1 + C_n/\sqrt{n})|u_1 - u_2|.$$

Combining this inequality with the one in (G.19), the bound in the statement of the lemma follows:

For a small $\delta > 0$, there is a random variable $C_n > 0$ such that for all $n > n_0$ and some large n_0

$$|Q_{1n}^c(z_1) - Q_{1n}^c(z_2)| \leq C_n|z_1 - z_2|(|z_1| + 1), \qquad \sup_{n>n_0, \gamma \in B_\delta(\gamma_0)} E_{P_\gamma} C_n < \infty,$$

uniformly over all $|z_1 - z_2| \leq 1$ in the set $S_n = \{z : |v|/\sqrt{n} \leq \varepsilon_n, |u|/n \leq \varepsilon_n\}$, where $\varepsilon_n \to 0$.

Similarly it follows that for a small $\delta > 0$, there is a random variable $C_n > 0$ such that for all $n > n_0$ and some large n_0

$$|Q_{2n}^c(z_1) - Q_{2n}^c(z_2)| \leq C_n|z_1 - z_2|(|z_1| + 1), \qquad \sup_{n>n_0, \gamma \in B_\delta(\gamma_0)} E_{P_\gamma} C_n < \infty,$$

uniformly over all $|z_1 - z_2| \leq 1$ in the set $S_n = \{z : |v|/\sqrt{n} \leq \varepsilon_n, |u|/n \leq \varepsilon_n\}$, where $\varepsilon_n \to 0$. Note that $C_n \equiv 0$ in the one-side models. ■