Likelihood Inference in a Class of Nonregular Econometric Models *

Victor Chernozhukov

Department of Economics

Massachusetts Institute of Technology

Han Hong

Department of Economics

Princeton University

First Version: August 2000
This version: February 11, 2002

Abstract

In this paper we study inference for a conditional model with a jump in the conditional density, where the location and size of the jump are described by regression lines. This interesting structure is shared by several structural econometric models. Two prominent examples are the standard auction model where 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. This paper develops the asymptotic inference theory for likelihood based estimators of these models– the Bayes and maximum likelihood estimators. Bayes and ML estimators are useful classical procedures. While MLE is transformation invariant, Bayes estimators offer some theoretic and computational advantages. They also have desirable efficiency properties. We characterize the limit likelihood as a function of a Poisson process that tracks the near-to-jump events and depends on regressors. The approach is applied to an empirical model of a highway procurement auction. We estimated a pareto model of Paarsch (1992) and an alternative flexible parametric model.

Key Words: Structural Econometric Model, Auctions, Job Search, Highway Procurement Auction, Likelihood, Point Process, Stochastic Equisemicontinuity

[For convenience of refereeing only. To be removed.]

Contents

1 Introduction

In this paper we consider a conditional model with a jump in the conditional density, whose location and size are described by regression lines. This model was first proposed by Aigner, Amemiya, and Poirier (1976) in the context of production analysis. Many recent econometric models also share this interesting structure. For example, in standard auction models, cf. Donald and Paarsch (1993a), the conditional density jumps from zero to a positive value; in equilibrium job search models (Bowlus, Neumann, and Kiefer (2001)), the density jumps from one level to another, inducing kinks in the distribution function. In what follows, the former model is referred to as the *one-sided* or *boundary* model, while the latter model is the *two-sided*. It is typical in these models that the location of jump is indispensably related to the parameters of the underlying structural economic model. Learning the parameters of location is thus crucial for learning the parameters of the underlying economic model.

Several important, fundamental papers developed inference methods for such models, including Aigner, Amemiya, and Poirier (1976), Ibragimov and Has'minskii (1981), Flinn and Heckman (1982), Christensen and Kiefer (1991), Donald and Paarsch (1996), Donald and Paarsch (1993b), Donald and Paarsch (1993a), Bowlus, Neumann, and Kiefer (2001). Ibragimov and Has'minskii (1981)(IH afterwards) obtained the limit distributions of Bayes and maximum likelihood estimators(MLE) without covariates. Donald and Paarsch (1996) dealt with MLE in the one-sided (boundary) models with discrete covariates.

Nevertheless, the general inference problem posed by Aigner, Amemiya, and Poirier (1976) has remained unresolved. The basic asymptotic properties of Bayes and ML estimators in the general two-sided regression model are still unknown. The properties of Bayes estimators in the one-sided model and the properties of MLE in the one-sided model with general regressors are also open questions. Without understanding these basic properties, using classical estimation principles in these econometric applications may be questionable.

In this paper, we develop the asymptotic theory of Bayes and Maximum Likelihood estimators for a general conditional model of a density jump, including one-sided and two-sided models with arbitrary covariates. Bayes estimators and MLE are attractive estimation procedures. While MLE is transformation invariant, the Bayes estimators offer some theoretic and computational advantages, and are convenient in practice. They also have desirable efficiency properties.

Further details may be summarized as follows. We will show that the limit of the likelihood process is a stochastic integral of a Poisson point process that tracks the conditional near-to-jump events. The result is analogous in spirit to that of Chernozhukov (2000), obtained for the extremal quantile regression.

It will be shown that Bayes estimators behave asymptotically as functions of the likelihood limit. Unlike the usual case of regular parametric models, Bayes estimators are not asymptotically equivalent to ML. In fact, Bayes estimators are efficient in terms of finite-sample average risk optimality and asymptotic average-risk optimality, which strongly justifies their use.[1] We do not study the minimax criteria. Recent contribution by Hirano

---

[1] MLE is not optimal for loss functions in conventional sense, but it may be shown to be optimal for a

and Porter (2001) offer a substantive analysis in this interesting direction in the context of one-sided discrete covariate models.

We will also demonstrate that the MLE behaves asymptotically as a function of the likelihood limit. Our proof uses the concept of stochastic equi-semicontinuity of Knight (1999). In our opinion, the result makes a convincing case for its further applications in econometrics and statistics.

Finally, we will study these methods in simulations and apply them to estimate models of a highway procurement auction. The first model we estimate is a stylized pareto model of Paarsch (1992). The second one is a flexible parametric alternative to the non-parametric model of Guerre, Perrigne, and Vuong (2000). We also implemented computer programs with Monte Carlo Markov Chain methods for the estimators. These programs are available from the authors.

The paper is organized as follows. Section 2 describes a basic linear model. Section 3 develops the asymptotic theory for this model. Section 4 considers a more general nonlinear model with nuisance parameters. Section 5 discusses efficiency issues. Section 6 discusses practical aspects of inference and estimation. Throughout the paper, c and C denote generic positive constants; $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and distribution, respectively; and $|\cdot|$ denotes the *supremum* norm of a vector.

2 The Basic Model

This section begins with a basic linear model, which helps establish the main results clearly. Extensions to general nonlinear models are given in section 4.

2.1 Assumptions

The basic model, denoted **R**, takes the following form

$$Y_i \equiv X_i'\beta + \epsilon_i, \tag{1}$$

where the error ϵ_i has the conditional density $f(\epsilon|X_i, \beta)$, parameterized by β belonging to the set $\mathcal{B}$, a compact, convex subset of $\mathbb{R}^d$. We denote the reference parameter as β_0, and assume $\beta_0 \in \text{interior}\mathcal{B}$. The conditional density has a jump at zero:

$$\lim_{\epsilon \uparrow 0} f(\epsilon|x, \beta) = q(x, \beta),$$
$$\lim_{\epsilon \downarrow 0} f(\epsilon|x, \beta) = p(x, \beta), \tag{2}$$
$$p(x, \beta) > q(x, \beta) + \delta, \delta > 0, \quad \forall x \in \mathbf{X}, \forall \beta \in \mathcal{B}.$$

In other words, the conditional density of Y given X jumps at the location $X'\beta$, which depends on the parameter β and covariate X. The shape of the density may also depend on the parameter β. In section 4, it will be made dependent on other parameters as well, and $X'\beta$ will be generalized to a nonlinear function.

---

generalized Dirac loss function.

We have two models to consider: the one-sided model and the two-sided model. The one-sided model has its conditional density jumping from zero to a positive constant. The two-sided model has its conditional density jumping from one positive value to another positive value. Figure 1 illustrates the two models. The one-sided model is a special case of the two-sided model. In addition, as suggested by Aigner, Amemiya, and Poirier (1976), the two-sided model may be applied to one-sided models, using the lower density region to account for outliers. More generally, the two-sided model approximates models with a sharp increase in the density, whose location depends on parameters and regressors. The finite sample distribution of parameter estimates in such a model are approximated by that in the model with density jump.

It is typical in these models that the location of jump is indispensably related to the parameters of the underlying structural economic model. Learning the parameters of location is thus crucial for learning the parameters of the underlying economic model.

We maintain the following additional assumptions for model $\mathbf{R}$.

Assumption 1   The following statements apply to x in $\mathbf{X}$ and β in $\mathcal{B}$:

(C.1) (Y_i, X_i) is an i.i.d. sequence of vectors in $\mathbb{R} \times \mathbb{R}^d$, defined on $(\Omega, \mathcal{F}, P_\beta)$. X_t has c.d.f F_X, with compact support $\mathbf{X}$, that does not depend on β, $Var(X) > 0$.

(C.2) in addition to (1)-(2), uniformly in β and x
  i. $q(x, \beta) > c > 0$ or
  ii. $f(u|x, \beta) = q(x, \beta) = 0$, for $u < 0$.

(C.3) Except at $\epsilon = 0$, $f(\epsilon|x, \beta)$ has continuous derivatives in ϵ and β, that are bounded uniformly in u, x, β. W.l.o.g. $f(\epsilon|x, \beta)$ is upper-semicontinuous at 0; its derivative is dominated: $\sup_{\beta \in \mathcal{B}} E_X \int |\frac{\partial}{\partial \beta} f(y - X'\beta|X; \beta)| \, dy < \infty$.

(C.4) There exist $K > 0, C > 0, \delta \geq 0$, such that uniformly in x and β: in case C.2.**i**, for all $c, \epsilon \in \mathbb{R}$, $|c| \leq K$,

$$|\tfrac{\partial}{\partial \epsilon} \ln f(\epsilon + c|x, \beta)| \leq C(\epsilon, x) \equiv C|\tfrac{\partial}{\partial \epsilon} \ln f(\epsilon|x, \beta)|^{1+\delta};$$

in case C.2.**ii** this apply only to $c, \epsilon : \epsilon + c > 0$. Moreover, $\sup_\beta E_{P_\beta} C(\epsilon_t, X_t) < \infty$.

Assumption C.2 allows for the boundary case, where density is zero to the left side of the jump and is positive on the right side. It also allows for the two-sided case, where density is positive on both sides. We distinguish these two cases to organize the proofs better. C.3 and C.4 are needed for uniform convergence of a continuous part of the likelihood ratio. It will be satisfied as long as the derivative of the density is not ill-behaved in the tails. Finally, the linearity of the regression function eases the exposition. Section 4 will consider a more general non-linear model with nuisance parameters.

2.2   Definition and Motivation of Bayes and ML Estimators

The likelihood function for the model is given by

$$L_n(\beta) \equiv \prod_{i \leq n} f(Y_i - X_i'\beta|X_i; \beta) dF_X(X_i).$$

A. density of a two-sided model

B. density of a one-sided model

A. cdf of a two-sided model

B. cdf of a one-sided model
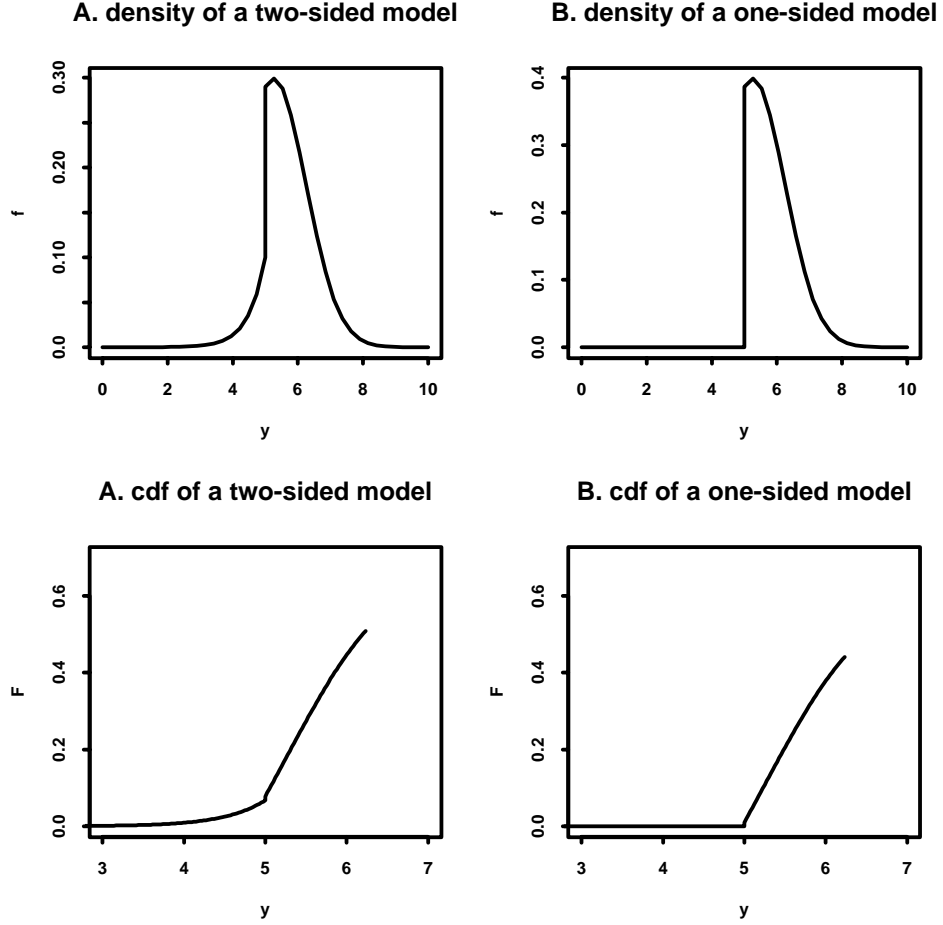
Figure 1: Panels B. correspond to the one-sided model. Panel A. corresponds to the two-sided model, which arises in equilibrium job search models and translates into kinks of cumulative distribution functions, see Bowlus et al (2001). The two-sided model also arises from one-sided models, when with a low probability we draw an outlier which ends up below the boundary of the support. This model was initially proposed by Aigner et al (1976). See also Bajari (1999) for a robustness critique of the one-sided model in the auction context. Note that the location of the density jump depends on parameters and regressors. Additional shape parameters will be introduced in section 4.

and the ML estimator[2] is defined as

$$\hat{\beta}_{ML} \equiv \arg\min_{\beta \in \mathcal{B}} -L_n(\beta).$$

On the other hand, the Bayes estimator minimizes the posterior expected loss

$$\hat{\beta} \equiv \arg\inf_{b \in \mathcal{B}} \int_{\mathcal{B}} \rho_n(b - \beta) \frac{L_n(\beta)q(\beta)d\beta}{\int_{\mathcal{B}} L_n(\beta)q(\beta)} d\beta,$$

whhere $\rho_n(x) \equiv \rho(nx)$ is a loss function, and $q(\cdot)$ is the prior density or weight function on $\mathcal{B}$. In the above expression, $L_n(\beta)q(\beta) / \int_{\mathcal{B}} L_n(\beta)q(\beta)d\beta$ is the posterior density conditional on the data $(Y_t, X_t, t \leq n)$. It does not depend on $dF_X(X_i)$. We impose the standard assumptions on $\rho(\cdot)$ and $q(\cdot)$, cf. IH(1982).

Assumption 2
(D.1) $q(\cdot) > 0$ is continous on $\mathcal{B}$,
(D.2) $\rho(\cdot) > 0$ is convex, and is majorized by a polynomial of $|u|$ as $|u| \longrightarrow \infty$.

Examples of conventional loss functions, satisfying condition (D.2), include the quadratic loss $\rho(z) = z'Wz$ for positive definite W and the absolute deviation loss $\rho(z) = \lambda' \mathrm{abs}(z)$, where $\lambda > 0$ and $\mathrm{abs}(z) \equiv \{|z|\}$.

For symmetric and bowl-shaped loss functions, the ML and Bayes estimators are efficient and asymptotically equivalent under asymptotic normality. If normality does not apply, as in our case, the Bayes estimators and MLE are typically not asymptotically equivalent. Therefore, MLE does not inherit the efficiency of Bayes estimators; Bayes estimators are average-risk efficient under conventional loss functions, while MLE is not.[3]

3 Asymptotic Theory for The Basic Model

We begin with the asymptotic behavior of the likelihood process, and then proceed to asymptotic distributions of Bayes and ML estimators.

3.1 Likelihood Limit

In modern asymptotic analysis, a common first step is to find the finite-dimensional or marginal limit of the likelihood ratio process. The limit eventually serves to describe the asymptotic distribution of Bayes and ML estimators. Such an initial step is called the convergence of experiments, see van der Vaart and Wellner (1996).

Consider the local likelihood ratio function

$$\ell_n(z) \equiv L_n(\beta_n + z/n)/L_n(\beta_n),$$

---

[2]$dF_X(X_i)$ does not depend on β and is hence irrelevant.
[3]See for example van der Vaart (1999). Efficiency for MLE could be claimed for the "delta loss" but not for quadratic or absolute deviation loss.

where $\beta_n = \beta_0 + \delta/n$, $\delta \in \mathbb{R}^d$ denotes the true parameter sequence. This is needed to study asymptotic efficiency later. Finite-dimensional (fi-di) weak convergence means that for any finite J

$$(\ell_n(z_j), \quad j \leq J) \overset{d}{\longrightarrow} (\ell_\infty(z_j), \quad j \leq J), \tag{3}$$

and $\ell_\infty(\cdot)$ is called a fi-di limit. In this section, $\overset{d}{\longrightarrow}$ denotes convergence under P_{β_n}. Define $p(X) = p(X, \beta_0)$ and $q(X) = q(X, \beta_0)$.

Theorem 1 *Given assumption 1, the fi-di weak limit of likelihood ratio $\ell_n(z)$ equals*

$$\ell_\infty(z) \equiv \exp\{z' EX[p(X) - q(X)]\}$$
$$\times \exp\left[\int_E l_z(j, x) d\mathbf{N}(j, x)\right], \quad where$$

$$l_z(j, x) \equiv \ln \frac{q(x)}{p(x)} 1 [0 < j < x'z] + \ln \frac{p(x)}{q(x)} 1 [0 > j > x'z],$$

$\mathbf{N}$ *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* $\{\mathcal{X}_i\}$ *are i.i.d. with d.f. F_X, and $\{\mathcal{X}_i'\}$ is an independent copy of $\{\mathcal{X}_i\}$,*

$$\begin{aligned} J_i &\equiv \Gamma_i/p(\mathcal{X}_i), & \Gamma_i &\equiv \quad \mathcal{E}_1 + ... + \mathcal{E}_i, \\ J_i' &\equiv \Gamma_i'/q(\mathcal{X}_i'), & \Gamma_i' &\equiv -(\mathcal{E}_1' + ... + \mathcal{E}_i'), \end{aligned} \tag{4}$$

$\{\mathcal{E}_i\}$ *and* $\{\mathcal{E}_i'\}$ *are two i.i.d., mutually independent sequences of standard exponential random variables that are also independent of $\{\mathcal{X}_i\}$ and $\{\mathcal{X}_i'\}$.*

The result has an intuitive appeal. The limit likelihood $\ell_\infty(z)$ has two informative parts. The deterministic "outer" part,

$$\ell_1(z) \equiv \exp\{z' EX[p(X) - q(X)]\},$$

can be regarded as information created by the far-from-jump data. The "inner" stochastic part,

$$\ell_2(z) \equiv \exp\left[\int_E l_z(j, x) d\mathbf{N}(j, x)\right] \equiv \exp\left[\sum_{t=1}^\infty l_z(J_i, \mathcal{X}_i) + \sum_{t=1}^\infty l_z(J_i', \mathcal{X}_i')\right],$$

can be interpreted as information created by near-to-jump data.

$\mathbf{N}$ is an asymptotic model of the near-to-jump data. As equation (4) shows, the points of $\mathbf{N}$, $(J_i, \mathcal{X}_i)$ and $(J_i', \mathcal{X}_i')$, depend on regressors $\mathcal{X}_i$ in a complex way. $\mathbf{N}$ is the limit of the point process

$$\widehat{\mathbf{N}}(\cdot) \equiv \sum_{i:\epsilon_i > 0} 1[(n\epsilon_i, X_i) \in \cdot] + \sum_{i:\epsilon_i < 0} 1[n\epsilon_i, X_i) \in \cdot].$$

The measure $\widehat{\mathbf{N}}(A)$ counts the number of points in any given set A. For any bounded set A, the limit behavior of $\widehat{\mathbf{N}}(A)$ depends only on near-to-jump errors $n\epsilon_i$ and the corresponding

6

covariate values. The smallest $|\epsilon_i|$'s are the ones that matter and they converge in law to mutually dependent gamma variables. Furthermore, in large samples the likelihood is driven mainly by near-to-jump data, revealing β at $O_p(n^{-1})$ rate. The fast convergence rate is not surprising. In a simplest one-sided case with no covariates, MLE is the minimal order statistic that converges to the end-point of the support at $O_p(n^{-1})$ rate.

Note also an important simplification of the formulae in the one-sided case. Since $q(x) = 0$

$$\ell_1(z) \equiv \exp\{z'EXp(X)\},$$

$$\ell_2(z) \equiv \left\{ \begin{array}{ll} 1 & \text{if } J_i \geq \mathcal{X}'_i z, \forall i \\ 0 & \text{otherwise.} \end{array} \right.$$

The inner part ℓ_2 is very informative in assigning the zero likelihood to certain values of z. Otherwise, $\ell_2(z)$ is flat. Once $\ell_2(z)$ equals 1, the outer part $\ell_1(z)$ further shapes the likelihood. In the two-sided model, when $q(x) > 0$, no z is assigned a zero inner likelihood. Both the ℓ_1 and ℓ_2 shape the limit likelihood.

3.2 Large Sample Properties of Bayes Estimators

The normalized Bayes estimator $Z_n = n(\hat{\beta} - \beta_n)$ is related to the likelihood ratio process by minimizing the posterior loss:

$$\Gamma_n(z) = \int_{U_n} \rho(z - u) \pi_n(u) du.$$

where $\pi_n(u)$ is the posterior density on the rescaled parameter space $U_n = n(\mathcal{B} - \beta_n)$.

$$\pi_n(u) = \ell_n(u) q(\beta_n + u/n) / \int_{U_n} \ell_n(u) q(\beta_n + u/n) du.$$

$\rho(z)$ is the loss function. $\ell_n(z)$ is the likelihood ratio process defined in the previous section and $q(\beta)$ is the prior density function.

As $n \to \infty$, U_n approaches $\mathbb{R}^d$ and the posterior $\pi_n(z)$ approaches the limit $\pi_\infty(u) \equiv \ell_\infty(u) / \int_{\mathbb{R}^d} \ell_\infty(u) du$. The limit posterior π_∞ is a function of the likelihood limit only; it is free from prior information. The result is thus simple to conjecture.

Theorem 2 *Suppose assumptions 1 and 2, also define for $\ell_\infty(\cdot)$ specified in Theorem 1:*

$$\Gamma_\infty(z) \equiv \int_{\mathbb{R}^d} \rho(z - u) \frac{\ell_\infty(u)}{\int_{\mathbb{R}^d} \ell_\infty(u) du} du.$$

Suppose that $Z_\infty \equiv \operatorname{argmin}_{z \in \mathbb{R}^d} \Gamma_\infty(z)$ is uniquely defined in $\mathbb{R}^d$ a.s. (*), then

$$Z_n \xrightarrow{d} Z_\infty.$$

Remark 3.1 The condition (*) is automatic for strictly convex functions $\rho(z)$ with unique minimum at $z = 0$, since $\ell_\infty(z)$ is positive a.s. on a subset of $\mathbb{R}^d$ with positive Lebesgue measure by the assumed non-degeneracy of X.

3.3 Maximum Likelihood

MLE $Z_n = n(\hat{\beta}_{ML} - \beta_n)$ maximizes the local likelihood ratio process:[4]

$$Z_n \equiv \operatorname{argmin}_{z \in U_n} - \ell_n(z)$$

Because $\ell_n(\cdot)$ is a highly non-regular function, the standard uniform convergence arguments are not applicable. One approach, taken by IH(1982), treats $\ell_n(\cdot)$ as an element of a Skorohod space. There are substantive difficulties with this approach in the regression case where there is more than one parameter. Instead, we employ Knight's stochastic equisemicontinuity, which converts the finite-dimensional convergence of discontinuous objective functions into convergence of argmins.[5] Appendix A provides a brief discussion of this new concept.

Theorem 3 *Suppose assumption 1, and that* $-\ell_\infty(z)$ *attains a unique minimum in* $\mathbb{R}^d$ *a.s., then*

$$Z_n \overset{d}{\longrightarrow} Z_\infty \equiv \operatorname{argmin}_{z \in \mathbb{R}^d} - \ell_\infty(z).$$

Remark 3.2 The condition that $-\ell_\infty(z)$ attains a unique minimum a.s. is needed, otherwise the limit distribution may fail to exist.

The important special case of MLE with discrete covariates in the one-sided model has been studied in the remarkable pioneering work of Donald and Paarsch (1996) and Donald and Paarsch (1993a). The results obtained here extend to continuous covariates and, importantly, also two-sided cases.

4 Nonlinear Model with Nuisance Parameters

In this section we consider a more general nonlinear model and also introduce nuisance parameters. While the linear model conveys the basic flavor and allows for a better explanation of the proofs, the nonlinear setup conforms with the economic models described in the introduction. The generalized model, denoted $\bar{\mathbf{R}}$, is given by

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

where the error ϵ_i has conditional density $f(\epsilon|X_i, \beta, \alpha)$, parameterized by $\beta \in \mathcal{B} \subset R^{d_1}$ and $\alpha \in \mathcal{A} \subset R^{d_2}$. We assume that the set $\mathcal{G} = \mathcal{B} \times \mathcal{A}$ is compact and convex and that the reference parameter $\gamma_0 = (\beta_0, \alpha_0)$ belongs to the interior of this set.

---

[4] If Z_n is set-valued, we may choose any measurable solution. Alternatively, define Z_n as any measurable ϵ_n-approximate argmin:Z_n is s.t. $-\ell_n(Z_n) \leq \inf_{z \in \mathbb{R}^d} -\ell_n(z) + \epsilon_n$, $\epsilon_n \searrow 0$. Allowing approximate solutions is useful for situations in which it may be difficult to find the exact optimum.

[5] MLE is a special Bayes estimator minimizing the posterior loss $\Gamma_n(z) \equiv \int \delta_z(u)\ell_n(u)\ du \equiv \ell_n(z)$, where $\delta_z(\cdot)$ is the delta function, which is too irregular to be a subject of the previous section.

The size of the jump of the conditional density of ϵ_i at 0, given X_i, may depend on both β and α:

$$\lim_{\epsilon \uparrow 0} f\left(\epsilon | x, \beta, \alpha\right) = q\left(x, \beta, \alpha\right),$$

$$\lim_{\epsilon \downarrow 0} f\left(\epsilon | x, \beta, \alpha\right) = p\left(x, \beta, \alpha\right), \tag{5}$$

$$p\left(x, \beta, \alpha\right) > q(x, \beta, \alpha) + \delta, \delta > 0; \quad \forall x \in \mathbf{X}, \quad (\beta, \alpha) \in \mathbb{R}^d, d = d_1 + d_2.$$

In other words, the conditional density of Y given X jumps at the location $g(X, \beta)$, which depends nonlinearly on the parameter β and covariate X. The shape of the density depends on the parameters β, α and covariates X. The additional shape parameter α is not related to the parameter of the location function. This model is therefore considerably more flexible than the basic linear model.

The assumptions and results for the non-linear model are very similar to the linear model. However, the presence of nuisance parameters adds to the complexity of exposition. We make the following additional technical assumptions:

Assumption 3 The following statements apply to x in $\mathbf{X}$ and $\gamma = (\beta, \alpha)$ in $\mathcal{G}$:

(E.1) (Y_i, X_i) is an i.i.d. sequence of vectors in $\mathbb{R} \times \mathbb{R}^d$, defined on $(\Omega, \mathcal{F}, P_\gamma)$. X_i has c.d.f F_X, with compact support $\mathbf{X}$. (5) holds, and uniformly in β, α, and x
  i. $q\left(x, \beta, \alpha\right) > c > 0$ or **ii.** $f\left(\epsilon | x, \beta\right) = q\left(x, \beta, \alpha\right) = 0$, for $\epsilon < 0$.

(E.2) Density $f(\epsilon | x, \gamma)$ has continuous derivatives in ϵ, γ for each ϵ, x and γ, except at $\epsilon = 0$, and is bounded uniformly in ϵ, x, γ; has continuous and bounded second derivative in α, uniformly in ϵ, x, γ. W.l.o.g. $f(\epsilon | x, \gamma)$ is upper-semicontinuous at $\epsilon = 0$ for each x and γ; and $\sup_{\gamma \in \mathcal{G}} E_X \int |\frac{\partial f}{\partial \gamma}\left(y - g(X, \beta) | X; \gamma\right)| dy < \infty$.

(E.3) $g\left(x, \beta\right)$ has two continuous and bounded derivatives w.r.t. β, uniformly in x and β. $Var\left[\frac{\partial g(X; \beta)}{\partial \beta}\right]$ is positive definite uniformly in β.

(E.4) Let $l_i\left(\gamma'\right) \equiv \ln f\left(Y_i - g\left(X_i, t\right); X_i, \gamma'\right)$. For $\gamma' = (t, s)$ in an open ball at $\gamma = (\beta, \alpha)$, either (a) $E_{P_\gamma}\left[\frac{\partial}{\partial \gamma} l_i\left(\gamma'\right)\right]\left[\frac{\partial}{\partial \gamma} l_i\left(\gamma'\right)\right]'$ is uniformly nonsingular and bounded, or (b) $\frac{\partial f}{\partial t}\left(y - g\left(x, t\right); x, \gamma\right) \equiv 0$ and $E_{P_\gamma}\left[\frac{\partial}{\partial \alpha} l_i\left(\gamma'\right)\right]\left[\frac{\partial}{\partial \alpha} l_i\left(\gamma'\right)\right]'$ is uniformly nonsingular and bounded; (a) and (b) hold uniformly in γ.

(E.5) There exist $K > 0, C > 0, \delta \geq 0$, such that for any x and γ: in case E.1.**i**, for all $c, \epsilon \in \mathbb{R}$, $|c| \leq K$, $|\frac{\partial}{\partial \epsilon} \ln f(\epsilon + c | x, \gamma)| \leq C(\epsilon, x)$ and in case E.1.**ii** this only needs to hold for all $c, \epsilon : \epsilon + c > 0$. Moreover, $\sup_\gamma E_{P_\gamma} C\left(\epsilon_i, X_i\right) < \infty$.

(E.6) Under the same conditions, second derivatives are dominated: $|\frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f\left(\epsilon + c | x, \gamma\right)| \leq C''\left(\epsilon, x\right)$ and $\sup_\gamma E_{P_\gamma} C''(\epsilon_i, X_i) < \infty$

The parameters include location parameters β and shape parameters α. If β is known, the inference about α is regular. Thus, the assumptions E.1-E.6 reflect a mixture of non-regular assumptions like in section 3 and in IH(1982) and regular ones like in van der Vaart and Wellner (1996), chapter 7 (mean-square differentiability). Conditions E.2-E.3

impose reasonable smoothness on the location function and the density function. E.4 imposes a standard mean-square differentiability and finite information matrix for the shape parameter α. E.5 and E.6 impose a standard domination on the score function and its derivative.

We next derive the limit likelihood process, followed by the asymptotics of Bayes and ML estimators.

4.1 Limits of Likelihood

The likelihood for this model is of the form:

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

where $dF_X(X_i)$ does not depend on γ and factors out. ML and Bayes estimators, $\hat{\gamma}_{ML}$ and $\hat{\gamma}_{Bayes}$, are defined as in section 2. The convergence rates needed to obtain a nondegenerate limiting likelihood ratio process are given by n and $\sqrt{n}$, for β and α respectively. Define H_n as a diagonal matrix with $1/n$ in the first $\dim(\beta)$ diagonal entries and $1/\sqrt{n}$ in the remaining diagonal entries. Let $\gamma_n(z) \equiv \gamma_n + H_n z$ for $z \equiv (u, v) \in \mathbb{R}^d$. The local likelihood ratio process is given by $\ell_n(z) = L_n(\gamma_n(z))/L_n(\gamma_n)$.

In this section, $\xrightarrow{d}$ denotes convergence in distribution under P_{γ_n}, where $\gamma_n \equiv (\beta_n, \alpha_n) = \gamma_0 + H_n \delta$ for given $\delta \in \mathbb{R}^d$.

Theorem 4 *In model $\bar{R}$, given assumption 3, the finite-dimensional weak limit of the localized likelihood ratio $\ell_n(z)$ takes the following form: for $\Delta(x) \equiv \partial g(x, \beta_0)/\partial \beta$*

$$\ell_\infty(z) \equiv \ell_{1\infty}(v) \times \ell_{2\infty}(u),$$

$$\ell_{1\infty}(v) \equiv \exp\left(W'v - \frac{1}{2}v'\mathcal{J}_{\gamma_0}v\right)$$

$$\ell_{2\infty}(u) \equiv \exp\left[u'E\Delta(X)[p(X) - q(X)]\right]$$

$$\times \exp\left[\int_E l_u(j, x)d\mathbf{N}(j, x)\right]$$

$$\mathcal{J}_{\gamma_0} \equiv E_{P_{\gamma_0}}\left[\frac{\partial}{\partial \alpha}l_i(\gamma_0)\right]\left[\frac{\partial}{\partial \alpha}l_i(\gamma_0)\right]'$$

where $l_u(j, x) \equiv \ln\frac{q(x)}{p(x)}1\left[0 < j < \Delta(x)'u\right] + \ln\frac{p(x)}{q(x)}1\left[0 > j > \Delta(x)'u\right]$, $\mathbf{N}$ is the Poisson process in Theorem 1. W is normally distributed $N(0, \mathcal{J}_{\gamma_0})$ and independent of $\mathbf{N}$.

The result differs from that in section 3. First, $\Delta(X) = \partial g(x, \beta_0)/\partial \beta$ replaces X, as expected. Second, we have a new term $\ell_{1\infty}(v)$, the log of which is a normal random variable with its variance inversely related to the information matrix. Thus, $\ell_{1\infty}(v)$ is a standard term for regular likelihood inference, e.g. van der Vaart and Wellner (1996), ch. 7. Note that if the parameter β were known, we would end up only with the standard term $\ell_{1\infty}(v)$. Since β needs to be estimated, we have the mixture of "regular" information about the shape parameters α and the "non-regular" information about the location parameters β. Moreover, these information components are asymptotically independent.

10

4.2 Asymptotic Behavior of Bayesian and ML Estimators

Next consider the normalized Bayes estimator $Z_n \equiv (Z_n^b, Z_n^s) \equiv (n(\hat{\beta} - \beta_n), \sqrt{n}(\hat{\alpha} - \alpha_n))$.

Theorem 5 (Bayesian Asymptotics for Nonlinear Models) *Assume model $\bar{R}$, assumptions 2 and 3, and that $\rho(z) = \rho_1(u) + \rho_2(v)$. For $\ell_\infty(\cdot)$ in Theorem 4 define:*

$$\Gamma_\infty(z) \equiv \int_{\mathbb{R}^d} \rho(z - z') \frac{\ell_\infty(z')}{\int_{\mathbb{R}^d} \ell_\infty(z') \, dz'} dz'$$

1. Suppose also that $Z_\infty \equiv \mathrm{argmin}_{z \in \mathbb{R}^d} \Gamma_\infty(z)$ is uniquely defined in $\mathbb{R}^d$ a.s. (*), then

$$Z_n \xrightarrow{d} Z_\infty$$

2. $Z_n^b \xrightarrow{d} Z_\infty^b = \mathrm{argmin}_u \int_{\mathbb{R}^{d_1}} \rho_1(u - u') \ell_{1\infty}(u') \, du'$ and $Z_n^s \xrightarrow{d} Z_\infty^s \equiv \mathrm{argmin}_v \int_{\mathbb{R}^{d_2}} \rho_2(v - v') \ell_{2\infty}(v') \, dv'$. Z_∞^b and Z_∞^s are independent.

Note that the independence is due to multiplicative separability of $\ell_\infty(z)$ in $\ell_{1\infty}(u)$ and $\ell_{2\infty}(v)$ and additive separability of $\rho(z)$. If the additive separability does not hold, part 1 of the Theorem 4 is still applies, while part 2 does not. Consider next the normalized MLE $Z_n \equiv (Z_n^b, Z_n^s) = (n(\hat{\beta} - \beta), \sqrt{n}(\hat{\alpha} - \alpha))$.

Theorem 6 (MLE Asymptotics for Nonlinear Models) *Under model $\bar{R}$, assumption 3, and assuming that $\ell_\infty(z)$ attains a unique minimum a.s,*

$$Z_n \xrightarrow{d} Z_\infty \equiv \mathrm{argmin}_{z \in \mathbb{R}^d} - \ell_\infty(z)$$

By multiplicative additive separability of $\ell_\infty(z)$, we have $Z_n^s \xrightarrow{d} Z_\infty^s = \mathcal{J}^{-1}W = N(0, \mathcal{J}^{-1})$ and $Z_n^b \xrightarrow{d} Z_\infty^b = \mathrm{argmin}_{z \in \mathbb{R}^d} - \ell_{2\infty}(z)$. Z_∞^b and Z_∞^s are independent.

These results generalize Theorems 2 and 3. In view of asymptotic independence between the shape information and location information, the estimators for these parameters are asymptotically independent. Also, the limit distribution of the Bayes estimator of shape parameter α coincides with that of MLE, if the loss function ρ_2 is symmetric (by Anderson's lemma). This is not the case for the estimators of the location parameter β.

5 Efficiency

The Bayes estimators are *exactly* finite-sample average-risk efficient(ARE) under particular loss functions. This is an instance of a well known result, formally stated in Theorem 7. Theorem 8 makes this statement an asymptotic one. These results justify one of the main efforts of this paper – the study of Bayes estimators. The ML estimators of location parameters are not equivalent to Bayes estimators even asymptotically and, unlike the usual case, do not share the optimality of Bayes estimators in large samples.

Average risk efficiency is one of the classic efficiency concepts developed by Wald, Lehmann, and others. Before writing it down formally for our case, it is helpful to review

the basic idea. Given a parameter γ, an estimator $\hat{\gamma}$, and a loss function $\rho_n(x) = \rho\left(H_n^{-1}x\right)$, we can compute the expected risk as $E_{P_\gamma}\rho_n(\hat{\gamma} - \gamma)$. The average risk takes the form $\int_{\mathcal{G}} E_{P_\gamma}\rho_n(\hat{\gamma} - \gamma)q(\gamma)d\gamma$ where q is a weight function (e.g. $q(\gamma) = 1$).

To address the asymptotic results, consider the following notation, define H_n as in section 4, and let $\gamma_n(\delta) \equiv \gamma_0 + H_n\delta$. Consider all statistics (measurable mappings of data) $f_n \equiv f_n((Y_i, X_i)_{i=1}^n)$, and denote the set of all such mappings as $\mathbf{F}_n$.

Define the (exact) *average risk criterion* (ARC) as

$$R_{\rho,q}(f, K) = \left[\int_K E_{P_{\gamma_n(\delta)}}\left[\rho\left(H_n^{-1}\left[f_n - \gamma_n(\delta)\right]\right)\right]q(\gamma_n(\delta))d\delta\right] / \mathrm{Leb}(K),$$

where q is the *weight* or *prior* measure (e.g. uniform) and ρ is the loss function, defined earlier. Division by $\mathrm{Leb}(K)$ is immaterial at this point.

Theorem 7 (Finite-Sample ARE) *Suppose model $\bar{\mathbf{R}}$ and conditions E.1-E.6 hold. For $f_{Bayes} \in \mathbf{F}_n$, defined by loss function ρ and prior weight q:*

$$f_{Bayes} \in \arg\inf_{f \in \mathbf{F}_n} R_{\rho,q}(f, U_n).$$

We next define asymptotic average risk (AARC) as

$$\overline{R}_\rho(\{f_n\}, K) = \limsup_{n\to\infty} R_{\rho,\mathrm{Leb}}(f_n, K),$$

for a compact cube K of $\mathbb{R}^d$ with center 0, and sequence of estimators $\{f_n\}$ in $\{\mathbf{F}_n\}$. To extend this definition to entire $\mathbb{R}^d$, define

$$\overline{R}_\rho(\{f_n\}, \mathbb{R}^d) = \limsup_{K\uparrow\mathbb{R}^d}\left[\overline{R}_\rho(\{f_n\}, K)\right],$$

where $K \uparrow \mathbb{R}^d$ denotes an increasing sequence of cubes converging to $\mathbb{R}^d$.

Theorem 8 (Asymptotic ARE) *Suppose model $\bar{\mathbf{R}}$ and conditions E.1-E.6 hold. For $\{f_{Bayes}\} \in \{\mathbf{F}_n\}$, defined by loss ρ and prior (weight) q:*

$$\inf_{\{f_n\}\in\{\mathbf{F}_n\}} \overline{R}_\rho(f_n, \mathbb{R}^d) = \overline{R}_\rho(\{f_{Bayes}\}, \mathbb{R}^d)$$

Because Bayes and ML estimators of location parameters are not asymptotically equivalent (equivalence holds for the shape parameter α under symmetric loss function ρ_2), the ML estimators are not optimal under the convex loss functions considered here. This is in contrast to the usual case where, for a large class of loss functions, MLE is asymptotically equivalent to the Bayes estimator and shares its optimality.

We next examine whether theoretical efficiency translates into actual efficiency gains with a simple monte carlo example. We consider simple one and two sided models with two covariates. In the one-sided case we generate data as

$$Y_0 \sim \mathrm{Uniform}\left[\bar{c}/m + (1 - 1/m)\underline{c}, \bar{c}\right]$$
$$\underline{c} = \bar{c} - \beta_1 X_1 - \beta_2 X_2, \quad X_j \sim \mathrm{Uniform}(a, b), j = 1, 2$$

12

Table 1: Mean Squared Error For Two Simulation Experiments

	MLE		Bayes	
	β_1	β_2	β_1	β_2
One-sided Model	0.8769	2.2030	0.0473	0.0395
Two-sided Model	0.1761	0.1244	0.0587	0.0502

The distribution in this example can be rationalized by a simple procurement auction model, in which there are n auctions. In each auction there are m bidders. Each bidder draws a random cost C from the uniform distribution on $(\underline{c}, \bar{c})$. We only observe the submitted bids, which depend on C through the Bayesian Nash equilibrium bid function: $Y = \bar{c}/m + (1 - 1/m) C$, resulting in the model above.

The two-sided example is a contaminated version of the first example. In particular, the data is generated from:

$$Y = Y_0 \text{ with prob. } \lambda \text{ and } Y = \text{Uniform} \left(L, \bar{c}/m + (1 - 1/m) \underline{c} \right) \text{ with prob. } 1 - \lambda$$

where we chose $\lambda = 0.9$ and $L = 2$. We simulate the above two models, using $\beta_1 = \beta_2 = 0.5$ for $n = 200$. Table 1 reports the sum of mean square errors across 300 simulations. The Bayes estimator has a substantively smaller mean square error than MLE.

6 Confidence Intervals and Some Practical Questions

The obtained results enable the construction of confidence sets.

Confidence Intervals. We must distinguish between the estimates of the shape parameter α and the estimates of the location parameters β.

Inference about α parameters is fully regular. The limit distribution of either MLE and Bayes (for symmetric loss function ρ_1) is given by $N(0, \mathcal{J}_{\gamma_0}^{-1})$. To facilitate inference we need to estimate $\mathcal{J}_{\gamma_0}^{-1}$. This can be done by conventional methods, taking the parameter estimate $\hat{\beta}$ as given.

Corollary 1 *Under assumption 3, for Bayes or MLE $\hat{\alpha}, \hat{\beta}$*

$$\hat{J} \equiv -\frac{1}{n} \sum_{i=1}^{n} \tfrac{\partial^2}{\partial \alpha \alpha'} \ln f(Y_i - g(X_i, \hat{\beta}) | X_i, \hat{\beta}, \hat{\alpha}) \xrightarrow{p} J_{\gamma_0}$$

An alternative is the familiar outer product of scores in α, which we do not state for brevity. The resampling methods available for inference about $\hat{\alpha}$ include subsampling and bootstrap. Although the bootstrap is not investigated formally, we can conjecture it works due to Mammen's theorem and asymptotic normality (see Horowitz (2000)).

Inference about the parameter β poses more difficulties. Neither Bayes nor ML estimators have a standard limit distribution. The nonparametric bootstrap is not consistent in the present setting. A simple counterexample is the boundary model without covariates,

in which case both Bayes and MLE are functions of the minimum order statistics. The nonparametric bootstrap is known to fail in this case (e.g. Horowitz (2000)).

We discuss a subsampling and an analytical approach to confidence intervals. We believe subsampling is a more practical and computationally simpler method. Following Politis, Romano, and Wolf (1999), let $W_1, ..., W_{N_n}$ be equal to the $N_n = \binom{n}{b}$ subsets of size b of $\{(Y_i, X_i), i \leq n\}$, ordered in any fashion. Let $I_1, ... I_B$ be chosen randomly with or without replacement from $\{1, 2, ... N_n\}$. Now, let $\hat{\theta}_{n,b,i}$ be equal to the statistic of interest $\hat{\theta}_b$ evaluated at the data set W_i. The approximation to the limit distribution function of $\tau_n(\hat{\theta} - \theta)$, where $\tau_n = n$ for the location parameters or $\tau_n = \sqrt{n}$ for the shape parameters, is given by

$$L_{n,b}(x) = \frac{1}{B} \sum_{i=1}^{B} 1\{\tau_b(\hat{\theta}_{n,b,I_i} - \hat{\theta}_n) \leq x\}.$$

By inverting $L_{n,b}(x)$, we obtain various α-quantiles $c_{n,b}(\alpha) = L_{n,b}^{-1}(\alpha)$. The level $1 - \alpha$ two-sided confidence interval is obtained as $[\hat{\theta}_n - \tau_n^{-1} c_{n,b}(1 - \alpha/2), \hat{\theta}_n - \tau_n^{-1} c_{n,b}(\alpha/2)]$. Similarly, the empirical distribution of $\tau_b|\hat{\theta}_{n,b,I_i} - \hat{\theta}_n|$ can be used to construct symmetric confidence intervals.

Corollary 2 *The subsampling method of estimating the limit distribution of $\tau_n(\hat{\theta} - \theta)$, where $\hat{\theta}$ is Bayes or MLE, is consistent in the sense of Politis, Romano, and Wolf (1999) (Theorem 2.2.1(i)-(iii)), and the asymptotic coverage probability of the confidence intervals achieves the correct nominal value, as long as $b \to \infty$, $b/n \to 0$, and $B \to \infty$, as $n \to \infty$.*

The choice of the block size b is discussed in detailed in chapter 9 of Politis, Romano, and Wolf (1999). They provide the calibration and the minimum volatility methods. In the empirical section, we use $1/10$ of the sample size. The confidence intervals are not sensitive to block size variation. This is probably due to the fast rate of convergence to the limiting distribution. The insensitivity principle underlies the minimum volatility method.

An alternative is an analytical method, based on simulating the distribution of a poisson process **N**, taking the estimated parameters as given, then obtaining ℓ_∞ and computing the solutions Z_∞. This method is detailed in Chernozhukov (1999). In the present context, subsampling is preferable on computational grounds.

Computational Methods. Modern computational methods are important for making the inference and estimation methods available to practitioners. It used to be that the Bayes computations were cumbersome and hampered the applicability for many years. Since approximately 1990, this problem has been overcome by Markov Chain Monte-Carlo (MCMC). This technique allows the simulation of a markov chain $Z_1, ... Z_J$ whose marginal distributions are approximately the posterior distribution. The method allows efficient and numerically stable computation of the Bayes estimates. Detailed discussion can be found e.g. in Robert and Casella (1998).

We implement these computational methods for the models considered in this paper. The programs are available from the authors. Our implementation is fast since the main subroutines are coded in C. Our program uses uninformative (flat over $\mathbb{R}^d$) prior. To compute the MLE, we use simulated annealing, an algorithm that handles general nonsmooth

Table 2: Summary Statistics

Worktype	#auctions	Avg. winning bid (1989$, mil)	Stdev., winning bid	Avg #. bidders
2	141	1.006	1.149	5.91
3	181	1.500	1.870	8.59
4	405	5.015	9.497	7.46

objective functions. Therefore, we provide not only the theory, but also the tools and computer programs needed for implementation.

7 Empirical Illustration

We consider a data set of bids submitted in a procurement contract auctions conducted by the New Jersey department of transportation (NJDOT) in the years 1989-1997. Over this period, the NJDOT conducted 1025 low-price, sealed-bid auctions of contracts to procure various types of services such as highway work, bridge construction and maintenance, and road paving. Most of the services procured had few auctions conducted. In the following, we consider only three types of services. See table (2) for the summary statistics. Hong and Shum (2000) give a detailed description of the data.

We focus on the independent private value model formulated in Paarsch (1992). In particular we assume that the construction cost follows independent pareto distributions, as studied by Paarsch (1992) and Donald and Paarsch (2000). Precisely, the cost distribution for construction companies is given by, for $\theta = (\theta_1, \theta_2)$:

$$h(c) = \frac{\theta_2 \theta_1^{\theta_2}}{c^{\theta_2+1}} \quad 0 < \theta_1 \le c, \ 0 < \theta_2.$$

Paarsch (1992) and Donald and Paarsch (2000) showed that this implies the following density function of the winning bids, conditional on the number of bidders m and other covariates that affect the distribution parameters θ:

$$f(y|m,\theta) = \begin{cases} \frac{\theta_2 m \{\theta_1 \theta_2 (m-1)/[\theta_2(m-1)-1]\}^{\theta_2 m}}{y^{\theta_2 m+1}} & \text{if } y > \frac{\theta_1 \theta_2 (m-1)}{\theta_2(m-1)-1} \\ 0 & \text{otherwise} \end{cases}$$

Table 3 reports the ML estimates and the Bayesian posterior mean estimates for this model. The variation in the summary statistics of the winning bid across types of contracts indicate that the jobs defined in these contracts are very different. Hence we present separate parameter estimates for each type of contract. Also, we give the 95% equal tailed and 95% symmetric confidence intervals constructed using subsampling method described earlier.

Table 3: One sided Pareto Estimates

worktype	MLE			Bayes	
	θ_1	θ_2	$\log L$	θ_1	θ_2
2	0.0160483	0.57543	-1341.05	0.0789063	0.706494
equal tail	-0.0140822	0.562586		0.0455028	0.541959
	0.0154688	0.598813		0.0839134	0.73919
symmetric	-0.00578062	0.552359		0.0620981	0.669373
	0.0378773	0.598501		0.0957144	0.743614
3	0.0593045	0.568535	-1955.05	0.0465869	0.356914
equal tail	0.0377357	0.56294		0.017097	0.268854
	0.060384	0.589204		0.0493621	0.357122
symmetric	0.0396208	0.55182		0.0264918	0.291758
	0.0789883	0.58525		0.066682	0.42207
4	0.0222681	0.646935	-7975.9	0.146176	0.555795
equal tail	0.00208421	0.627295		0.131181	0.329598
	0.0223695	0.666575		0.157989	0.566001
symmetric	0.00392716	0.627295		0.13406	0.368623
	0.040609	0.666575		0.158291	0.742967

Table 4 reports the parameter estimates from an alternative two sided model:

$$f\left(y|m,\theta\right) = \begin{cases} (1-\lambda)\,\dfrac{\theta_2 m\{\theta_1\theta_2(m-1)/[\theta_2(m-1)-1]\}^{\theta_2 m}}{y^{\theta_2 m+1}} & \text{if } y > \frac{\theta_1\theta_2(m-1)}{\theta_2(m-1)-1} \\ \lambda\left(1/\frac{\theta_1\theta_2(m-1)}{\theta_2(m-1)-1}\right) & \text{if } 0 < y < \frac{\theta_1\theta_2(m-1)}{\theta_2(m-1)-1} \end{cases},$$

and 0 otherwise. We chose $\lambda = 0.02$, which accommodates outliers that do not conform to the theoretical model.

In table 5 we introduce a continuous covariate for the traffic volume. This covariate is only available for work type 4. We parameterize $\theta_1 = \exp\left(\alpha_1 + \alpha_3 \times X\right)$ and $\theta_2 = \exp\left(\alpha_2\right)$, where X denotes traffic volume. The coefficient appears to be significant, although there are large discrepancies depending on the estimation method and the model. This is indicative of misspecification.

An alternative approach to direct parametric inference for independent private value auction model is the indirect inference approach of Guerre, Perrigne, and Vuong (2000). Their insight is based on examing the first order condition of the optimization problem of a representative bidder i in the equilibrium:

$$\max_b \int_b^\infty g_{-i}\left(x\right)\left(b - c\right)dx \quad \Rightarrow \quad -g_{-i}\left(b\right)\left(b - c\right) + \bar{G}_{-i}\left(b\right) = 0$$

where $\bar{G}_{-i}$ and g_{-i} denotes the survival function and the density function of the distribution of the minimum bid among bidder i's competitors. Therefore, a two step procedure can be used. In the first step, $\bar{G}_{-i}$ and g_{-i} are estimated using the bid data. In the second

Table 4: Two sided Pareto Estimates

worktype	MLE			Bayes	
	θ_1	θ_2	$\log L$	θ_1	θ_2
2	0.259243	0.55482	-372.445	0.22572	0.810755
equal tail	0.180359	0.498768		0.149349	0.776126
	0.272495	0.569737		0.236047	0.85172
symmetric	0.188709	0.498771		0.154353	0.297086
	0.329778	0.61087		0.770756	0.850753
3	0.418519	0.512854	-656.18	0.3074	0.361401
equal tail	0.316776	0.453844		0.203662	0.336442
	0.439423	0.529476		0.31702	0.36304
symmetric	0.320138	0.491592		0.208184	0.344999
	0.516895	0.534106		0.406617	0.377803
4	2.87942	0.535876	-2375.17	1.84626	0.549167
equal tail	2.46807	0.529865		1.35817	0.53917
	3.1199	0.553112		1.98644	0.569452
symmetric	2.55317	0.518958		1.42778	0.530891
	3.20566	0.552793		2.26474	0.567443

Table 5: Model with Traffic Volume (Type 4)

	MLE				Bayes		
	α_1	α_2	α_3	$\log L$	α_1	α_2	α_3
one-side	-3.812	-0.436	-0.000	-6996.4	-1.096	-0.488	-0.697
equal tail	-3.833	-0.440	-0.001		-1.637	-1.092	-0.941
	-3.810	-0.409	0.006		0.137	-0.453	-0.557
symmetric	-3.830	-0.453	-0.006		-2.160	-0.615	-0.911
	-3.793	-0.419	0.005		-0.032	-0.362	-0.484
two-side	-1.131	-0.652	0.547	-1984	-0.540	-0.639	0.091
equal tail	-2.150	-0.655	0.544		-0.842	-0.670	0.058
	-1.114	-0.634	0.645		-0.461	-0.588	0.138
symmetric	-1.820	-0.670	0.451		-0.812	-0.681	0.049
	-0.441	-0.634	0.642		-0.269	-0.596	0.134

step, the pseudo-cost values

$$c = b - \frac{\bar{G}_{-i}(b)}{g_{-i}(b)}$$

are constructed for each bid in each auction. Then the distribution of these pseudo-values can be used to infer the latent distribution of the cost parameter. $\bar{G}_{-i}$ and g_{-i} can be easily inferred from the data for any symmetric affiliated private value model. It suffices to observe the winning bid, since

$$\bar{G}_{-i}(x) = \left(\bar{F}(x)\right)^{\frac{m-1}{m}} \quad \text{and} \quad g_{-i}(x) = \left(\frac{m-1}{m}\right) \bar{F}(x)^{-\frac{1}{m}} f(x)$$

where $\bar{F}(\cdot)$ and $f(\cdot)$ are the survival and density function of the whining bid, respectively.

The indirect inference approach of Guerre, Perrigne, and Vuong (2000) uses a non-parametric method to estimate $\bar{F}(x)$ and $f(x)$. Here we consider a flexible parametric approach. We used the truncated normal distribution parameterized as:

$$f(x) = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Big/ \bar{\Phi}\left(\frac{b-\mu}{\sigma}\right)$$

where we take $\ln b = \alpha_1 + \alpha_4 \times m$, $\ln \mu = \alpha_2 + \alpha_5 \times m$, $\ln \sigma = \alpha_3 + \alpha_6 \times m$. Table 6 reports the results for worktype 4.

First, we notice that the new model is an improvement over the pareto model in terms of the log likelihood value, suggesting a significant improvement in the fit. Although we did not develop a formal testing procedure for the nonstandard likelihood, the principle of Vuong (1989) suggests that the model with the higher likelihood is closer to the data in the information-theoretic sense.

Second, we note that the Bayes posterior mean estimates differ from the maximum likelihood estimates, although none of the slope coefficient reverse its sign. This is not surprising, since the likelihood surface has many more modes when the number of parameters is greater. Based on the computational experiments, theoretical efficiency properties, and the fact that the Bayes estimators (posterior means) minimize the globally convex objective function, they may be preferred.

We did not report the results for the case with the traffic volume covariate. Adding this covariate produced a tiny improvement of the log likelihood and did not change the original coefficients. The estimates for the traffic volume coefficients were highly insignificant. Both methods yielded agreeable results concerning that covariate.

Overall, the two-sided models fit data better that one-sided ones, indicating that controlling outliers that do not conform the model is important. We also observe that the parametric variant of the indirect inference approach of Guerre, Perrigne, and Vuong (2000) is quite valuable and allows to fit the particular data better than the Pareto model.

Conclusion

We studied a general model in which the conditional density of the dependent variable jumps at a location that is parameter dependent. This includes a variety of the boundary-dependent model discussed in the recent literature of structural estimation. We derive

Table 6: Truncated Normal Model for Worktype 4

worktype 4	MLE					
	α_1	α_2	α_3	α_4	α_5	α_6
	$\log L = -694.854$					
	-2.38883	0.404125	0.628959	-0.0702671	-2.58986	0.207458
equal tail	-2.56234	-2.39153	-0.23598	-0.14270	-2.61471	0.167639
	-2.3802	0.54492	0.80567	-0.04021	-2.5829	0.23593
symmetric	-2.55203	-2.07808	0.15484	-0.12891	-2.60658	0.168674
	-2.22562	2.88633	1.10308	-0.01162	-2.57313	0.24624
	Bayes					
	α_1	α_2	α_3	α_4	α_5	α_6
	4.40077	-2.41695	0.643259	-2.38707	-2.26628	0.211352
equal tail	3.84919	-2.93896	0.462919	-2.60164	-2.56957	0.172165
	4.87133	-1.88786	1.047	-1.8852	-1.77164	0.23446
symmetric	3.88809	-2.94462	0.347846	-2.8686	-2.72846	0.182684
	4.91346	-1.88929	0.938672	-1.90554	-1.80411	0.24002

asymptotic distributions of Bayes and ML estimators under general conditions, and offer practical computation and inference methods. The results provide a solution to a long-standing econometric problem.

Our results extend previous work in several directions: (1) handling general regression models; (2) inclusion of Bayes estimators, which enjoy the small and large sample efficiency; (3) considering the two sided model as a robust alternative to the one-sided model; and (4) using the point process methods to give a precise characterization of large sample distributions. Bayes estimators are important alternatives to MLE due to efficiency properties which the MLE does not share. The methodology in this paper also provides new insights into the analysis of asymptotic distributions for models that cannot be studied using the conventional tools of locally asymptotically normal models.

The empirical application presented illustrates the usefulness of the results. We estimated some key auction models. The first model was a stylized pareto model of Paarsch (1992). The second auction model represented a flexible parametric alternative to the nonparametric approach of Guerre, Perrigne, and Vuong (2000). We find that the two-sided models fit data better than one-sided ones, indicating that controlling for outliers that do not conform the model is important. We also find that a parametric variant of the indirect inference approach of Guerre, Perrigne, and Vuong (2000) is quite valuable and may allow to fit data better than the direct parametric approach.

A Useful Background Definitions

A.1 Point Processes

Definition 1 (Point Measures, $M_p(E)$, cf. Resnick (1987)) 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] (cf. Leadbetter, Lindgren, and Rootzen (1983)). Vague convergence induces *vague topology* on $M_p(E)$. 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 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 $N_n(w)$ is some point measure in $M_p(E)$. *Weak convergence* of the point process $\mathbf{N}_n$ 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 $\int_E f(x) d\mathbf{N}_n(x) \xrightarrow{d} \int_E f(x) d\mathbf{N}(x)$ for any $f \in C_K(E)$ by continuous mapping theorem.

Definition 3 (Poisson Point Process or Random Measure (PRM)) Point process $\mathbf{N}$ is a PRM with *mean intensity measure* m (defined 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.

A.2 Convex Objectives

The result can be found in Knight (1999). It is a generalization of earlier convexity lemmas by Knight and Pollard. It allows discontinuities and $\bar{\mathbb{R}}$ - valued objective functions.

Lemma 1 (Guyer) *Suppose $\{Q_T\}$ is a sequence of lower-semi-continuous (lsc) convex $\bar{\mathbb{R}}$-valued random functions, defined on $\mathbb{R}^d$, and let $\mathcal{D}$ be a countable dense subset of $\mathbb{R}^d$. If Q_T fidi-converges to Q_∞ in $\bar{\mathbb{R}}$ on $\mathcal{D}$ where Q_∞ is lsc convex and finite on an open non-empty set a.s., then*

$$\underset{z \in \mathbb{R}^d}{argmin} \, Q_T(z) \xrightarrow{d} \underset{z \in \mathbb{R}^d}{argmin} \, Q_\infty(z),$$

provided the latter is uniquely defined a.s. in $\mathbb{R}^d$.

A.3 Stochastic Equisemicontinuity

The remarkable concepts of this section were recently developed by Knight (1999). The following summarizes some essential elements we need.

Epi-Convergence. Suppose the sequence of objectives $\{Q_n\}$ are random lower semi-continuous (l-sc) functions (i.e. $Q_n(x) \leq \liminf_{x_j \to x} Q_n(x_j), \forall x, \forall x_j \to x$). Let $\mathcal{L}$ be the space of l-sc functions $f : \mathbb{R}^d \to \bar{\mathbb{R}}$, s.t. $f \not\equiv \infty$. $\mathcal{L}$ can be made into a complete separable metric space by considering a special metric, convergence in which is equivalent to epi-convergence (cf. Knight(2000), Rockafellar and Wets (1998)). Hence one can metrize the weak convergence in $\mathcal{L}$: Q_n is said to *epi-converge in distribution* to Q 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 $r_1, ..., r_k$:

$$P\big(\cap_{j=1}^k \{ \inf_{x \in R_j} Q(x) > r_j \} \big) \leq \liminf_n P\big(\cap_{j=1}^k \{ \inf_{x \in R_j} Q_n(x) > r_j \} \big)$$

$$\leq \limsup_n P\big(\cap_{j=1}^k \{ \inf_{x \in R_j^o} Q_n(x) \geq r_j \} \big) \leq P\big(\cap_{j=1}^k \{ \inf_{x \in R_j^o} Q(x) \geq r_j \} \big).$$

Epi-convergence is a weak condition that leads to the convergence of argmins.

Lemma 2 (Knight, Theorem 1) *Suppose that*

   i. Z_n is s.t. $Q_n(Z_n) \leq \inf_{z \in \mathbb{R}^d} Q_n(z) + \epsilon_n$, $\epsilon_n \searrow 0$; $Z_n = O_p(1)$

   ii. $Z_\infty \equiv argmin_{z \in \mathbb{R}^d} Q_\infty(z)$ is uniquely defined in $\mathbb{R}^d$ a.s.

   iii. $Q_n(\cdot)$ epi-converges in distribution to $Q_\infty(\cdot)$, then

$$Z_n \overset{d}{\longrightarrow} Z_\infty$$

Epi-convergence is more general than uniform convergence, because it allows for rather general discontinuities. In our case, (lots of) non-vanishing discontinuities make the uniform convergence of the likelihood function *impossible*.

Provided the finite dimensional distributional (fidi) limit exists, the necessary and sufficient condition for epi-convergence in distribution is stochastic equi-lower-semi-continuity (s. e-l-sc.), developed by Knight (1999).

Stochastic equi-semi-continuity. Sequence $\{Q_n\} \in \mathcal{L}$ is s. e-sc. if for each bounded set B, $\epsilon > 0$, and $\delta > 0$, there exist $u_1, ..., u_k \in B$ and some open sets $V(u_1), ..., V(u_k)$ covering B and containing $u_1, ..., u_k$ s.t.

$$\limsup_n P\big(\cup_{j=1}^k \{ \inf_{x \in V(u_j)} Q_n(x) \leq \min(\epsilon^{-1}, Q_n(u_j) - \epsilon) \} \big) < \delta.$$

Lemma 3 (Knight, Thm 2) *Suppose Q_n is s.e-lsc. Then $\{Q_n\}$ converges to Q_∞ in distribution in finite-dimensional sense if and only if $\{Q_n\}$ epi-converges in distribution to $Q_\infty(\cdot)$.*

The *s-esc condition* amounts to the possibility of *approximating* the *distribution* of the infimum (and hence of argmin) of Q_n over bounded set B by *an approximate minimum* of Q_n over a *carefully chosen grid* $\{u_i, ...u_k\}$, with given precision (ϵ, δ). It is worth emphasizing that the naive uniform grids will not do the job in our case. The application of stochastic equisemi-continuity leads to a simple proof even in the no-covariate case, which could improve (in terms of length) over the arguments of Ibragimov and Hasminskii.

B   Proofs for the Linear Model

In the proof we set the local parameter sequence $\beta_n = \beta_0$. Putting through the general local sequence β_n does not change the arguments but introduces a lot of notational complexity.

B.1 Proof of Theorem 1.

Consider the local log likelihood ratio process $Q_n(z) = \ln L_n(\beta_0 + z/n)/L_n(\beta_0)$

$$Q_n(z) = \sum_{i=1}^{n} q_{in}(z) \times \left[1(\epsilon_i > X_i'z/n \vee 0) + 1(\epsilon_i < X_i'z/n \wedge 0)\right]$$

$$+ \sum_{i=1}^{n} q_{in}(z) \times \left[1(0 < \epsilon_i \leq X_i'z/n) + 1(0 > \epsilon_i \geq X_i'z/n)\right]$$

$$\equiv Q_{1n}(z) + Q_{2n}(z), \text{ where}$$

$$q_{in}(z) \equiv \ln\left[f(Y_i - X_i'(\beta_0 + z/n))|X_i, \beta_0 + z/n)/f(Y_i - X_i'\beta_0|X_i, \beta_0)\right].$$

$Q_{1n}(z)$ and $Q_{2n}(z)$ behave very differently.

I. Limit of $Q_{1n}(z)$ is analyzed by techniques similar to those in IH (1982). $\{Q_{1n}(z)\}$ is an average like statistic. By C1, C4 and LLN, it converges in probability for each z:

$$Q_{1n}(z) \overset{p}{\longrightarrow} Q_{1\infty}(z) \equiv -z'EX\frac{f'(\epsilon_t|X, \beta_0)}{f(\epsilon_t|X, \beta_0)} \tag{6}$$

For a density function f that has a dominated derivative everywhere except at 0: $\int_{\mathbb{R}} f'(u)du = -f(0^+) + f(0^-)$. Thus applying this to the conditional density $f(\epsilon|X, \beta_0)$, we have:

$$Q_{1\infty}(z) \equiv z'EX[p(X) - q(X)].$$

Next we use stochastic equi-continuity to convert pointwise convergence to uniform convergence in z. It suffices to show that for any $|z_1 - z_2| \overset{p}{\longrightarrow} 0$, the term

$$\left|\sum_{i=1}^{n} q_{in}(z_1) \times 1(\epsilon_i > X_i'z_1/n \vee 0) - \sum_{i=1}^{n} q_{in}(z_2) \times 1(\epsilon_i > X_i'z_2/n \vee 0)\right| \overset{p}{\longrightarrow} 0, \tag{7}$$

as well as other similar terms in $Q_n(z)$. The left hand side of (7) is bounded by

$$\sum_{i=1}^{n} 1\left(\epsilon_i > \frac{X_i'z_1}{n} \vee \frac{X_i'z_2}{n} \vee 0\right) \times \left|\ln f\left(\epsilon_i - \frac{X_i'z_1}{n}|\beta_0 + z_1/n\right) - \ln f\left(\epsilon_i - \frac{X_i'z_2}{n}|\beta_0 + z_2/n\right)\right|$$

$$+ \sum_{i=1}^{n} 1\left(\epsilon_i \in (0, \frac{X_i'z_1}{n} \vee \frac{X_i'z_2}{n}]\right) \times \max_{j=1,2}\left|\ln\frac{f\left(\epsilon_i - \frac{X_i'z_j}{n}|\beta_0 + z_j/n\right)}{f(\epsilon_i|\beta_0)}\right|. \tag{8}$$

By C.4, the first sum in (8) is bounded by, for z^* in the convex hull of z_1 and z_2

$$\sum_{i=1}^{n} 1(\epsilon_i > X_i'z_1/n \vee X_i'z_2/n \vee 0) \times \left|\frac{f'(\epsilon_i - X_i'z^*/n|\beta_0 + z^*/n)}{f(\epsilon_i - X_i'z^*/n|\beta_0 + z^*/n)}\right| \times \left|\frac{X_i'(z_1 - z_2)}{n}\right|$$

$$\leq \frac{1}{n}\sum_{i=1}^{n} 1(\epsilon_i > 0) \times \text{const} \times \left|\frac{f'(\epsilon_i|\beta_0)}{f(\epsilon_i|\beta_0)}\right|^{1+\delta} \times |z_1 - z_2| = O_p(1) \times |z_1 - z_2| \tag{9}$$

The second sum in (8) is bounded by, for some $C, C', C'' > 0$

$$\sup_{z_1, z_2 \in \mathbf{Z}} \sum_{i=1}^{n} 1\left(\epsilon_i \in (0, \frac{X_i'z_1}{n} \vee \frac{X_i'z_2}{n}]\right) \times C \times \left|\frac{f'(\epsilon_i|\beta_0)}{f(\epsilon_i|\beta_0)}\right|^{1+\delta} \times \frac{|X_i'z_1| \vee |X_i'z_2|}{n}$$

$$\leq \sup_{z_1, z_2 \in \mathbf{Z}} \frac{1}{n}\sum_{i=1}^{n} 1\left(\epsilon_i \in (0, C'/n]\right) \times C'' = o_p(1)$$

22

since $f'(\epsilon|x)$ is uniformly bounded above and $f(u|x)$ uniformly bounded below when $\epsilon > u > 0$ by C.2-C.3, and X has a compact support. Thus (7) follows. Other terms are similarly checked.

II. Limit of $Q_{2n}(z)$. Q_{2n} is driven by the "rare" occurrences of the near-to-jump observations and can be modeled as an integral w.r.t. to the point process that measures these occurrences. We split the proof into two parts: step 1 constructs the key point process and derives its limit representation. Step 2 shows that $(Q_{2n}(z_j), j \leq J)$ is a continuous transformation of the point process.

Step 1: The Key Point Process. Define $E \equiv (-\infty, 0) \cup (0, +\infty) \times \mathbf{X}$. The topological space on E is taken to be a product of standard topologies of $\mathbb{R} \setminus \{0\}$ and $\mathbb{R}^d \cap \mathbf{X}$. So that, e.g., $[a, b] \times \mathbf{X}$ is a compact subset of E. $\mathcal{E}$ is the Borel σ-algebra of subsets of E.

The key point process is a random measure taking the following form:

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

for any set A in $\mathcal{E}$. $\widehat{\mathbf{N}}$ is a random element of $M_p(E)$, the metric space of nonnegative point measures on E, with the metric generated by the topology of vague convergence. Appendix A gives definitions. We will show that

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

where $\mathbf{N}$ is defined in the proposition 1. This is done in steps (a) and (b).

(a) By C.1 and C.3, 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), $\lim_{n \to \infty} E\widehat{\mathbf{N}}(F) \equiv \lim_{n \to \infty} nP(\{n\epsilon, X\} \in F) =$

$$\int_F [p(x)1(u > 0)du + q(x)1(u < 0)du] \times dF_X(x) = m(F) < \infty$$

where measure m is defined in the proposition. Since the events $\{(n\epsilon_i, X_i) \in F\}$ are independent across i by C.1, by the Meyer's Lemma (Meyer (1973)) we also have:

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

which by Kallenberg's Theorem[6] [$\widehat{\mathbf{N}}$ is clearly simple a.s.] and the definition of the Poisson process (Appendix A) implies that $\widehat{\mathbf{N}} \Rightarrow N$ in $M_p(E)$, where N is a Poisson point process with the mean measure $m(\cdot)$.

(b). Next we show that N has the same distribution as the process $\mathbf{N}$, stated in the proposition. First, define *canonical* homogeneous PRMs $\mathbf{N}_0$ and $\mathbf{N}_0'$ with points $\{\Gamma_i\}$ and $\{\Gamma_i'\}$, defined in the proposition. $\mathbf{N}_0$ has the mean measure $m_0(du) = du$ on $(0, \infty)$, and $\mathbf{N}_0'$ has the mean measure $m_0'(du) = du$ on $(-\infty, 0)$. Now because $\mathbf{N}_0$ and $\mathbf{N}_0'$ are independent,

$$\mathbf{N}_1(\cdot) \equiv \mathbf{N}_0(\cdot) + \mathbf{N}_0(\cdot)'$$

is the Poisson point process with the mean measure

$$m_1(du) = 1(u > 0)du + 1(u < 0)du \text{ on } \mathbb{R} \setminus \{0\}.$$

Because $\{\mathcal{X}_i, \mathcal{X}_i'\}$ are i.i.d. and independent of $\{\Gamma_i, \Gamma_i'\}$, by the Composition Lemma(cf. Proposition 3.8 in Resnick (1987)), the *composed* PRM $\mathbf{N}_2$ with points

$$\left(\{\Gamma_i, \mathcal{X}_i\}, \{\Gamma_j', \mathcal{X}_j'\}, i \geq 1, j \geq 1\right) \text{ in } \mathbb{R}^d$$

---

[6] For example, Resnick (1987), Prop. 3.22

has the *product* mean intensity measure given by

$$m_2(du, dx) = [1(u > 0)du + 1(u < 0)du] \times F_X(dx) \text{ on } \mathbb{R} \setminus \{0\} \times \mathbf{X}.$$

Finally, $\mathbf{N}$ with the *transformed* points $\{\mathbf{T}(\Gamma_i, \mathcal{X}_i), \mathbf{T}(\Gamma'_i, \mathcal{X}'_i)\}$, where

$$\mathbf{T} : (u, x) \mapsto \left(1(u > 0)\frac{u}{p(x)} + 1(u < 0)\frac{u}{q(x)}, \quad x \right)$$

has the desired mean measure on E

$$m(dj, dx) = m_2 \circ \mathbf{T}^{-1}(dj, dx) = [p(x)1(j > 0) + q(x)1(j < 0)]\, dj \times F_X(dx),$$

by the Transformation Theorem for Poisson Processes(Proposition 3.7 in Resnick (1987)).

Step 2: The Functional of the Key Point Process. Here we distinguish two cases (a) $\mathbf{X} \equiv \{x \in \mathbf{X} : q(x) > \delta > 0\}$ and (b) $\mathbf{X} \equiv \{x \in \mathbf{X} : q(x) = 0\}$. Note that by assumption C.3, for any compact set $\mathbf{Z}$, as $n \to \infty$,

$$\ln \left[\frac{f(\delta - x'z/n|x, \beta_0 + z/n)}{f(\delta|x, \beta_0)} \right] = \ln \left[\frac{q(x)}{p(x)} \right] (1 + O(n^{-1})) \tag{10}$$

uniformly in $\{\delta, z, x \in \mathbb{R}_+ \times \mathbf{Z} \times \mathbf{X} : x'z > 0, 0 < \delta < x'z/n\}$, and

$$\ln \left[\frac{f(\delta - x'z/n|x, \beta_0 + z/n)}{f(\delta|x, \beta_0)} \right] = \ln \left[\frac{p(x)}{q(x)} \right] (1 + O(n^{-1})) \tag{11}$$

uniformly in $\{\delta, z, x \in \mathbb{R}_- \times \mathbf{Z} \times \mathbf{X} : x'z < 0, 0 > \delta > x'z/n\}$. Note that in case (a) this holds by C.3, while in case (b) equation (10) holds *identically* equal to $-\infty$. Hence

$$Q_{2n}(z) = \Big[\sum_{i=1}^{n} \ln \frac{q(X_i)}{p(X_i)} 1\left[0 < n\epsilon_i < X'_i z \right] + \sum_{i=1}^{n} \ln \frac{p(X_i)}{q(X_i)} 1\left[0 > n\epsilon_i > X'_i z \right] \Big] \times (1 + o_p(1)) \tag{12}$$

$$\equiv \bar{Q}_{2n}(z) \times (1 + o_p(1))$$

uniformly in z over $\mathbf{Z}$. (Again this expression may equal $-\infty$ in case (b), but that does not create a problem for the proof.) Write $\bar{Q}_{2n}(z)$ as an integral w.r.t. $\widehat{\mathbf{N}}$:

$$\bar{Q}_{2n}(z) \equiv \int_E l_z(j, x) d\widehat{\mathbf{N}}(j, x), \tag{13}$$

where $l_z(j, x)$ is defined in the theorem. We consider cases (a) and (b) separately:

(a): By conditions C.1-2, the function $(j, x) \mapsto l_z(j, x)$ is bounded and vanishes outside the set $\mathbf{K}_z \equiv [-\eta, +\eta] \times \mathbf{X}$, $\eta = \sup_{x \in \mathbf{X}} |x'z|$, where $\eta < \infty$ by C.1. $\mathbf{K}_z$ is a compact set in E (by the standard compactification of E). The function $(j, x) \mapsto l_z(j, x)$ is discontinuous at $j = 0$ and at $j = x'z$. Next define the map $\mathbf{T} : M_p(E) \mapsto \mathbb{R}^l$ as

$$N \mapsto \left(\int_E l_{z_k}(j, x) dN(j, x), k \leq l \right).$$

The mapping $\mathbf{T}$ is discontinuous at the set

$$\mathcal{D}(\mathbf{T}) \equiv \{N \in M_p(E) : j_i^N = 0 \text{ or } j_i^N = z'_k x_i^N \text{ for some } i \geq 1, k \leq l\}$$

where $(j_i^N, x_i^N, i \geq 1)$ denote the points of N. By C.3 and the construction of $\mathbf{N}$

$$P[\widehat{\mathbf{N}} \in \mathcal{D}(\mathbf{T}), \text{ for some } n \geq 1] = 0, \quad P[\mathbf{N} \in \mathcal{D}(\mathbf{T})] = 0. \tag{14}$$

24

Therefore $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E)$ implies $\mathbf{T}(\widehat{\mathbf{N}}) \xrightarrow{d} \mathbf{T}(\mathbf{N})$. We conclude that

$$\left(Q_{2n}(z_k), k \leq l\right) \xrightarrow{d} \left(Q_{2\infty}(z_k), k \leq l\right), \tag{15}$$

where $Q_{2\infty}(z) \equiv \int_E l_z(j, x) d\mathbf{N}(j, x)$.

(b): Now consider the second case: $\ell_{2n}(z) \equiv \exp\{\bar{Q}_{2n}(z)\}$. Note that

$$\ell_{2n}(z) = 0 \text{ if } \widehat{\mathbf{N}}(A(z)) > 0, \quad \ell_{2n}(z) = 1 \text{ if } \widehat{\mathbf{N}}(A(z)) = 0,$$

where

$$A(z) \equiv \{(j, x) \in \mathbb{R}_+ \times \mathbf{X} : j < x'z\}.$$

Observe also that $\ell_{2\infty}(z) = 0$ if $\mathbf{N}(A(z)) > 0$, $\ell_{2\infty}(z) = 1$ if $\mathbf{N}(A(z)) = 0$. Thus to show finite-dimensional convergence (for $\Delta_k = 0$ or 1):

$$P\left(\ell_{2n}(z_k) = \Delta_k, k \leq K\right) \longrightarrow P\left(\ell_{2\infty}(z_k) = \Delta_k, k \leq K\right),$$

it suffices to show $\left(\widehat{\mathbf{N}}(A(z_k)), k \leq K\right) \xrightarrow{d} \left(\mathbf{N}(A(z_k)), k \leq K\right)$. By the continuous mapping theorem, this follows from $\widehat{\mathbf{N}} \Rightarrow \mathbf{N}$ in $M_p(E)$, since $\widehat{\mathbf{N}}(\partial A(z_k)) = 0$ and $\mathbf{N}(\partial A(z_k)) = 0$ a.s. ∎

B.2   Proof of Theorem 2

Using the convexity lemma (1), it suffices to show the finite-dimensional (fidi) convergence of $\Gamma_n(\cdot)$ to $\Gamma_\infty(\cdot)$,

$$\left(\Gamma_n(z_k), k \leq K\right) \xrightarrow{d} \left(\Gamma_\infty(z_k), k \leq K\right).$$

$\Gamma_n(z)$ is an integral with respect to ℓ_n over $\mathbb{R}^d$. There are two steps in the proof:

1. Approximate $\Gamma_n(z)$ over $\mathbb{R}^d$ by an integral over a compact subset of $\mathbb{R}^d$. For large $M \in \mathbb{R}_+$, approximate $\Gamma_n(z)$ by $\Gamma_n^M(z) \equiv \Gamma_{n1}^M(z) / \Gamma_{n2}^M$, where

$$\Gamma_{n1}^M(z) \equiv \int_{|u| \leq M} \rho(z - u) \ell_n(u) q(\beta_0 + u/n) du, \quad \Gamma_{n2}^M \equiv \int_{|u| \leq M} \ell_n(u) q(\beta_0 + u/n) du.$$

Define their corresponding limit process as

$$\Gamma_{1\infty}^M(z) \equiv \int_{|u| \leq M} \rho(z - u) \ell_\infty(u) q(\beta_0) du, \quad \Gamma_{2\infty}^M \equiv \int_{|u| \leq M} \ell_\infty(u) q(\beta_0) du.$$

Now suppose that for each z and each $\epsilon > 0, \delta > 0, \exists M \in \mathbb{R}_+$:

$$\limsup_n P\left(|\Gamma_n(z) - \Gamma_n^M(z)| > \epsilon\right) < \delta, \tag{16}$$

This is a tail-smallness property. It follows from the exponential smallness of the likelihood tails, which is demonstrated in the subsequent lemmas (4) and (5).

2. In view of (16), it suffices to show $\left(\Gamma_{n2}^M, \Gamma_{n1}^M(z_k), k \leq K\right) \xrightarrow{d} \left(\Gamma_{2\infty}^M, \Gamma_{1\infty}^M(z_k), k \leq K\right)$. This follows by two facts:

   i $E\ell_n(z) = 1 < \infty$

   ii $|E\ell_n^{1/2}(z') - \ell_n^{1/2}(z)|^2 \leq C|z - z'|$

Fact i. is the definition of likelihood ratio. Fact ii. is by Lemma 5-6. These facts check the conditions of a limit theorem for integrals of random functions, Theorem 22 in Appendix I of IH (1982). ∎

B.3 Proof of Theorem 3

By condition C.3, $-\ell_n$ is lower-semi-continuous. Two steps are needed:

- Show finite-dimensional convergence of ℓ_n to the stochastic limit ℓ_∞,
- Show the stochastic equi-lower-semi-continuity (e-lsc) of $\{-\ell_n\}$,

The first step was shown in Theorem 1. We demonstrate the second step next. Theorem I.5.1 in IH(1982) shows that the conclusion of lemma (6) implies that $Z_n = O_p(1)$. Combining these two steps, the conclusion follows by Lemma 2.

It remains to show stochastic equi-lower-semi-continuity (e-lsc) of $\{-\ell_n\}$, or equivalently of $\{-Q_{2n} \equiv -\log \ell_n\}$. We know from the proof of Theorem 1 that:

$$Q_{1n}(z) - Q_{1\infty}(z) \xrightarrow{p} 0, \quad Q_{2n}(z) - \bar{Q}_{2n}(z) \xrightarrow{p} 0,$$

uniformly in z over fixed compact sets, where $\bar{Q}_{2n}(z) \equiv \int_E l_z(j,x)d\widehat{\mathbf{N}}(j,x)$. $Q_{1\infty}(z)$ is a fixed linear function. It therefore suffices to show s-e-lsc of $\{-\bar{Q}_{2n}(z)\}$ only.

Because $\bar{Q}_{2n}(z)$ is a piece-wise constant function it suffices to show that or any bounded set $B \subset \mathbb{R}^d$ and $\delta > 0$, there are open neighborhoods $V(z_1), \dots V(z_m)$ of some $z_1, \dots, z_m$ s.t. $B \subset \cup_{k=1}^m V(z_k)$ and

$$P\left[\cup_{k=1}^m \{\inf_{z \in V(z_k)} -\bar{Q}_{2n}(z) < -\bar{Q}_{2n}(z_k)\}\right] < \delta.$$

This is done in several steps.

(a)[A seemingly strange point process] Construct the point process:

$$\tilde{N}(\cdot) \equiv \sum_{\epsilon_i > 0} 1\left[(n\epsilon_i p(X_i), X_i) \in \cdot\right] + \sum_{\epsilon_i < 0} 1\left[(n\epsilon_i q(X_i), X_i) \in \cdot\right].$$

Represent the points of $\tilde{N}$ equivalently in terms of the order statistics

$$\Gamma_{1n}, \Gamma_{2n}, \dots \quad \text{of } \{ n\epsilon_i p(X_i), \quad i : \epsilon_i > 0 \}, \quad \Gamma'_{1n}, \Gamma'_{2n}, \dots \quad \text{of } \{ n\epsilon_i q(X_i), \quad i : \epsilon_i < 0 \}.$$

so that $0 < \Gamma_{1n} \leq \Gamma_{2n}\dots$; $\quad 0 > \Gamma'_{1n} \geq \Gamma'_{2n}\dots$. Denote by X_{in}, X'_{in} the corresponding to $\Gamma_{in}, \Gamma'_{in}$ realizations of the covariate. Thus $\tilde{N}(\cdot) \equiv \sum_{i \geq 1} 1\left[(\Gamma_{in}, X_{in}) \in \cdot\right] + 1\left[(\Gamma'_{in}, X'_{in}) \in \cdot\right]$. $\tilde{N}$ is a continuous transform of $\widehat{\mathbf{N}}$, say $\mathbf{T}(\widehat{\mathbf{N}})$, from $M_p(E)$ to $M_p(E)$. Therefore, in $M_p(E)$ $\tilde{N}(\cdot) \Rightarrow N(\cdot) \equiv \mathbf{T}(\mathbf{N}) \equiv \sum_{i \geq 1} 1\left[(\Gamma_i, \mathcal{X}_i) \in \cdot\right] + 1\left[(\Gamma'_i, \mathcal{X}'_i) \in \cdot\right]$ where the distribution of the points is defined in the statement of Theorem 1. Also, by continuous mapping theorem, in $\mathbb{R}^k$

$$(\Gamma_{1n}, \dots, \Gamma_{kn}) \xrightarrow{d} (\Gamma_1, \dots, \Gamma_k).$$

We need all of this to characterize the equisemi-continuity of $\bar{Q}_{2n}$. Write

$$\bar{Q}_{2n}(z) \equiv \int_{E:j>0} l_z(j,x)d\widehat{\mathbf{N}}(j,x) + \int_{E:j<0} l_z(j,x)d\widehat{\mathbf{N}}(j,x) \equiv \bar{Q}_{2n}^+(z) + \bar{Q}_{2n}^-(z).$$

We examine *discontinuities* of $\bar{Q}_{2n}$ by examining that of $\bar{Q}_{2n}^+(z)$ and $\bar{Q}_{2n}^-(z)$. Because the arguments are identical for either part, consider only part $\bar{Q}_{2n}^+(z)$.

(b)[Picking the Cover] Fix a bounded set $B \subset \mathbb{R}^d$. Cover $\mathbf{X}$ by the minimal number of closed equal-sized cubes $\{X_\phi(x_j), j \leq J(\phi)\}$ with the side-length equal to ϕ and the centers of cubes denoted as x_j. Construct sets

$$\{V_{kj}, k = -m, \dots, m, j \leq J(\phi)\} \subset \mathbb{R}^d$$

as $V_{kj} \equiv \{z \in \mathbb{R}^d : \underline{v}_k - \varphi < p(x)x'z < \underline{v}_k + \varphi, \forall x \in X_\phi(x_j)\}$, where $\varphi > 0$ and

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

Since $\mathbf{X}$ is a nondegerate compact subset of $\mathbb{R}^d$, and $p(x)$ is bounded away from 0 and above by assumption, $\{\cup_{kj} V_{kj}\}$ can cover any given bounded set B by selecting sufficiently large m.

(c)[**Number of Break-points is** $O_P(1)$ **and separated for small** φ] Consider argument z in $\cup_j V_{kj}$, then a discontinuity in $\bar{Q}_{2n}^+(z)$ can *potentially* occur in $\cup_j V_{kj}$ only if there exist $z^* \in \cup_j V_{kj}$ and (Γ_{in}, X_{in}) s.t.

$$\Gamma_{in} = p(X_{in})X_{in}'z^*, \tag{17}$$

where it must be that

$$\underline{v}_k - \varphi \leq \Gamma_{in} \leq \underline{v}_k + \varphi.$$

If there is such (Γ_{in}, X_{in}), we say that $\bar{Q}_{2n}^+$ *has a breakpoint in* $\cup_j V_{kj}$. Note that $\bar{Q}_{2n}^+(z)$ can not have breakpoints in V_{kj} with $k < 0$ because $\Gamma_{in} > 0$. Define $\mathcal{N}_n \equiv \#\{i : \Gamma_{in} < \bar{k}\}$, $\mathcal{N} \equiv \#\{i : \Gamma_i < \bar{k}\}$, where $\bar{k} \equiv \sup_{x \in \mathbf{X}, z \in B} p(x)x'z$. $\mathcal{N}_n$ is the upper bound on the number of breakpoints of $\bar{Q}_{2n}^+(z)$ in set B. By continuous mapping theorem, $\mathcal{N}_n \xrightarrow{d} \mathcal{N}$ in $\mathbb{R}$. So the *number of breakpoints* $O(\mathcal{N}_n)$ is stochastically bounded $O_p(1)$. Furthermore, break-points are *separated*: *no more than one break point can happen in* $\cup_j V_{kj}$ with probability arbitrarily close to one if φ is sufficiently small. Define A_k to be the event that $\bar{Q}_{2n}^+(z)$ has more than two break-points in $\cup_j V_{kj}$.

$$\limsup_n \ P\left[\cup_k A_k\right] \leq \limsup_n \ P\left[\min_{1 \leq i \leq \mathcal{N}_n} |\Gamma_{in} - \Gamma_{(i-1)n}| < 2\varphi\right]$$

$$\leq \limsup_n \ P\left[\min_{1 \leq i \leq K} |\Gamma_{in} - \Gamma_{(i-1)n}| < 2\varphi\right] + P\left[\mathcal{N}_n > K\right]$$

$$\leq \qquad P\left[\min_{1 \leq i \leq K} |\Gamma_i - \Gamma_{(i-1)}| < 2\varphi\right] + P\left[\mathcal{N} > K\right]$$

$$\leq \delta/2$$

which is achieved by setting K sufficiently large so that $P\left[\mathcal{N} > K\right] < \delta/4$, followed by setting φ sufficiently small so that $P\left[\min_{1 \leq i \leq K} |\Gamma_i - \Gamma_{(i-1)}| < 2\varphi\right] < \delta/4$, which is possible since

$$\mathrm{E}\left[\min_{1 \leq i \leq K} |\Gamma_i - \Gamma_{(i-1)}|\right] = \mathrm{E}\left[\min_{1 \leq i \leq K} |\mathcal{E}_i|\right] > 0.$$

From now on, K and φ are fixed.

(c)[**"smart" grid-points by setting** ϕ **small**] Next construct "centers" z_{kj} in V_{kj} so that

$$\underline{v}_k - \varphi \ < \ x'z_{kj} \ < \ \underline{v}_k - \varphi + \eta, \quad \forall x \in X_\phi(x_j)$$

where $\eta : 0 < \eta << \varphi$ will be set sufficiently small in the next step. Depending on η, to satisfy the constraints, we will set ϕ sufficiently small as well.

(d) [**Stochastic Equi-semicontinuity**] Now follows the verification of stochastic equi-semicontinuty:

$$\limsup_n P\left[\cup_{kj} \{\inf_{z \in V_{kj}(z_{kj})} -\bar{Q}_{2n}^+(z) < -\bar{Q}_{2n}^+(z_{kj})\}\right] < \limsup_n P\left[B(\eta)\right] + \limsup_n P\left[\cup_k A_k\right]$$

where $B(\eta)$ is the event that $\{\Gamma_{in}, i \leq K\}$ are separated, and at least one of them falls into one of the "small" disjoint sets of the form:

$$[\underline{v}_{k_i} - \varphi, \underline{v}_{k_i} - \varphi + \eta], \quad i \leq K.$$

27

The bound is true because the last event contains the event that $(\Gamma_{in}, i \leq K)$ are separated, and one of them falls into a set of the form $[\underline{v}_{k_i} - \varphi, X'_{in} z_{kij}]$ which actually is the event

$$\cup_{kj} \{ \inf_{z \in V_{kj}(z_{kj})} -\bar{Q}^+_{2n}(z) < -\bar{Q}^+_{2n}(z_{kj}) \} \cap (\cup_k A_k)^c,$$

because $-\bar{Q}^+_{2n}(z)$ is *piecewise-constant and can only jump up if the index $X'_i z$ increases*. Now because $(\Gamma_{in}, i \leq K) \xrightarrow{d} (\Gamma_i, i \leq K)$, which have the bounded density, it follows that

$$\limsup_n P[B(\eta)] = O(K\eta) < \delta/2.$$

by picking sufficiently small η. (Note that K stays fixed and its choice does not depend on η). By step (b) $\limsup_n P[B(\eta)] + \limsup_n P[\cup_k A_k] < \delta.\blacksquare$

B.4   Lemmas 4-6

A critical ingredient in the proof of Theorem 2 is the requirement of tail smallness. In this section such properties of

$$\ell_n(z) = L_n(\beta + z/n)/L_n(\beta).$$

are established, where $z \in \mathbb{R}^d$.

Lemma 4 *Suppose $\exists b > 0, B > 0$, s.t. $\forall z \in \mathbb{R}^d, z' \in \mathbb{R}^d$, $\exists n_0$ s.t. $\forall n > n_0$:*

$$(i) E_{P_\beta} \ell_n(z)^{1/2} \leq e^{-b|z|}; \qquad (ii) E_{P_\beta} |\ell_n(z)^{1/2} - \ell_n(z')^{1/2}|^2 \leq B|z - z'|,$$

uniformly in β in an open ball at β_0. Then for

$$\Gamma^M_{n1} \equiv \int_{|u|>M} \frac{\ell_n(u) q(\beta_0 + u/n)}{\int_{U_n} \ell_n(u) q(\beta_0 + u/n) du} du, \quad \Gamma^M_{n2}(z) \equiv \int_{|u|>M} \rho(z-u) \frac{\ell_n(u) q(\beta_0 + u/n)}{\int_{U_n} \ell_n(u) q(\beta_0 + u/n) du} du$$

it is true that under $P = P_{\beta_n}$

$$\lim_{M \to \infty} \limsup_{n \to \infty} E \Gamma^M_{n1} = 0, \qquad \lim_{M \to \infty} \limsup_{n \to \infty} E \Gamma^{M'}_{n2}(z) = 0, \tag{18}$$

so that

$$\lim_{M \to \infty} \limsup_{n \to \infty} P \left(|\Gamma_n(z) - \Gamma^M_n(z)| > \epsilon \right) = 0. \tag{19}$$

IH(1982) studied the non-regression models and verified the conditions of this lemma by bounding a Hellinger distance. Our approach requires a modification: we bound a conditional Hellinger distance for the model **R**. The Conditional Hellinger Distance, denoted as $r_2(\beta; \beta + h)$, is defined as:

$$\left[\int \int \left| f^{1/2}(y - x'(\beta + h)|x; \beta + h) - f^{1/2}(y - x'\beta|x; \beta) \right|^2 dy F_X(dx) \right]^{1/2}$$

Upper and lower bounds on $r_2(\cdot)$ are used for to verify conditions of lemma 2:

Lemma 5 *If there are $a > 0, A > 0$ such that for each $h > 0$ small,*

$$\inf_\beta r_2^2(\beta; \beta + h) \geq \frac{a|h|}{1 + |h|} \quad and \quad \sup_\beta r_2^2(\beta; \beta + h) \leq A|h|$$

where sup/inf are computed over β in open ball at β_0. Then $\exists b > 0, B > 0$, s.t. $\forall z \in \mathbb{R}^d, z' \in \mathbb{R}^d$, $\exists n_0$ s.t. $\forall n > n_0$:

$$\sup_\beta E_{P_\beta} \ell_n(z)^{1/2} \leq e^{-ba|z|}; \qquad \sup_\beta E_{P_\beta} |\ell_n(z)^{1/2} - \ell_n(z')^{1/2}|^2 \leq B|z - z'|.$$

28

The following lemma verifies the conditions of lemma (5) and hence also verifies the bound on the conditional Hellinger distance in model **R**.

Lemma 6 *In model* **R**, *suppose (C.1)-(C.3) hold, then* $\exists a > 0, A > 0$, *such that for all* $h > 0$ *small enough, the following is true:*

$$\inf_\beta r_2^2 \left(\beta; \beta + h\right) \geq a|h| \quad and \quad \sup_\beta r_2^2 \left(\beta; \beta + h\right) \leq A|h|$$

Proofs of lemmas (4) to (6):

Lemma (4): (18) is a special case of lemma I.5.2 and Theorem I.10.2 of IH(1982). (19) follows immediately from (18). ∎

Lemma (5): It follows from the definition of the conditional Hellinger distance that $r_2^2 \left(\beta; \beta + z/n\right)$

$$= 2\left[1 - \int \int f^{1/2}\left(y - x'\left(\beta + z/n\right)|x; \beta + z/n\right) f^{1/2}\left(y - x'\beta|x; \beta\right) dy F_X\left(dx\right)\right].$$

Also $E\ell_n\left(z\right)^{1/2} \equiv \int \cdots \int \left\{ \prod_{i \leq n} \left[\frac{f^{1/2}\left(y_i - x_i'\left(\beta + z/n\right)|x_i; \beta + z/n\right)}{f^{1/2}\left(y_i - x_i'\beta|x_i; \beta\right)} \right]^{1/2} \right.$

$$\left. \prod_{i \leq n} \left[f(y_i - x_i'\beta|x; \beta) dy_i F_X\left(dx_i\right) \right] \right\}$$

$$\equiv \left[\int \int f^{1/2}(y - x'(\beta + z/n)|x_i; \beta + z/n) f^{1/2}\left(y - x'\beta|x; \beta\right) dy F_X\left(dx\right) \right]^n,$$

we can bound uniformly in $P = P_\beta$

$$E\ell_n\left(z\right)^{1/2} = \left[1 - \frac{1}{2} r_2^2\left(\beta; \beta + z/n\right) \right]^n \leq e^{-\frac{n}{2} r_2^2 \left(\beta; \beta + z/n\right)} \leq e^{-\frac{1}{2} \frac{a|z|}{1 + |z|/n}}.$$

Similarly, $E\left|\ell_n\left(z\right)^{1/2} - \ell_n\left(z'\right)^{1/2}\right|^2 = E\ell_n\left(z\right) + E\ell_n\left(z'\right) - 2E\ell_n\left(z\right)^{1/2}\ell_n\left(z'\right)^{1/2}$

$$\equiv 2\left[1 - \left(E_X \int f^{1/2}\left(y - X'\left(\beta + z/n\right)|X; \beta + z/n\right)\right.\right.$$

$$\left.\left. f^{1/2}\left(y - X'\left(\beta + z'/n\right)|X; \beta + z'/n\right) dy\right)^n\right]$$

$$\leq 2n\left[1 - E_X \int f^{1/2}\left(y - X'\left(\beta + z/n\right)|X; \beta + z/n\right)\right.$$

$$\left. f^{1/2}\left(y - X'\left(\beta + z'/n\right)|X; \beta + z'/n\right) dy\right]$$

$$\equiv nr_2^2\left(\beta + z/n; \beta + z'/n\right) \leq A|z - z'|. \ \blacksquare$$

Lemma (6): To obtain the uniform upper bound, for $\epsilon > 0$ small enough

$$r_2^2(\beta; \beta + h) \equiv E_X \int \left(f^{1/2}\left(y - X'\left(\beta + h\right)|X; \beta + h\right) - f^{1/2}\left(y - X'\beta|X; \beta\right) \right)^2 dy$$

$$\leq E_X \int \left| f\left(y - X'\left(\beta + h\right)|X; \beta + h\right) - f\left(y - X'\beta|X; \beta\right) \right| dy$$

$$\leq E_X \int_{[X'\beta, X'(\beta + h)]} \left| f\left(y - X'\left(\beta + h\right)|X; \beta + h\right) - f\left(y - X'\beta|X; \beta\right) \right| dy$$

$$+ E_X \int_{[X'\beta, X'(\beta + h)]^c} \left| f\left(y - X'\left(\beta + h\right)|X; \beta + h\right) - f\left(y - X'\beta|X; \beta\right) \right| dy,$$

where $[a, b] = [a, b]$ if $a \leq b$ and $= [b, a]$ if $b \leq a$. The first inequality follows since $|a - b|^2 \leq |(a + b)(a - b)| = |a^2 - b^2|$ for $a > 0$ and $b > 0$. This is further bounded by

$$2E_X |X'h| (p(X, \beta) + q(X, \beta)) + E_X \int \int_0^1 \left| h' \frac{\partial f}{\partial u} (y - X'(\beta + uh)|X; \beta + uh) \right| dudy,$$

which by compactness of $\mathbf{X}$, Cauchy-Schwartz inequality and changing order of integration by Fubini is bounded by,

$$\leq \text{const} \times |h| + \text{const} \times |h| \int_0^1 E_X \int \left| \frac{\partial f}{\partial \beta} (y - X'(\beta + uh)|X; \beta + uh) \right| dydu$$

The upper bound is now obtained by condition C.3.

Now bound uniformly in β the conditional Hellinger distance from below:

$$r_2^2(\beta; \beta + h) \geq E_X \int_{[X'\beta, X'(\beta+h)]} \left(f^{1/2} (y - X'(\beta + h)|X; \beta + h) - f^{1/2} (y - X'\beta|X; \beta) \right)^2 dy$$

$$\geq E_X \int_{[X'\beta, X'(\beta+h)]} \frac{1}{2} \times \left| p^{1/2}(X, \beta) - q^{1/2}(X, \beta) \right|^2 dy$$

$$\geq \frac{1}{2} \times \delta \times \int |X'h|\ ) = \frac{1}{2} \times \delta \times |h| \times E_X \left| \frac{X'h}{|h|} \right| \geq \text{const} \times |h|,$$

where the last line follows because $\inf_{c \in \mathbb{R}^d : |c| = 1} E|X'c| > 0$, since $Var(X) > 0$. ∎

C   Proofs for the Non-Linear Model

In the proofs we set the local parameter sequence $\gamma_n = \gamma_0$. Putting through the general local sequence γ_n does not change the arguments but introduces a lot of notational complexity.

C.1   Proof of Theorem 4

As in Theorem 1, we split the log likelihood ratio process $Q_n(z) = \ln L_n(\beta_0 + H_n z)/L_n(\beta_0)$ into the "jump" part and the "smooth" part, and analyze each part separately. For $z = (u, v)$,

$$Q_n(z) = \sum_{i=1}^n q_{in}(z) \times [1(\epsilon_i > \Delta_n(X_i, u)/n \vee 0) + 1(\epsilon_i < \Delta_n(X_i, u)/n \wedge 0)]$$

$$+ \sum_{i=1}^n q_{in}(z) \times [1(0 < \epsilon_i \leq \Delta_n(X_i, u)/n) + 1(0 > \epsilon_i \geq \Delta_n(X_i, u)/n)]$$

$$\equiv Q_{1n}(z) + Q_{2n}(z), \text{ where}$$

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

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

I. We first find the uniform limit of $Q_{1n}(z)$. In this part only, we apply the arguments similar to those of IH (1982). For each z, using assumptions (E.1) to (E.3) with a second order Taylor

expansion:

$$Q_{1n}(z) = -u'E\frac{\partial g(X,\beta_0)}{\partial\beta}\frac{f'(\epsilon|X,\gamma_0)}{f(\epsilon|X,\gamma_0)}$$

$$+ \frac{v'}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\ln f(\epsilon_i|X_i,\gamma_0) + \frac{1}{2}v'E\frac{\partial^2}{\partial\alpha\partial\alpha'}\log f(\epsilon|X,\gamma_0)v + o_p(1)$$

$$\equiv u'E\Delta(X)(p(X)-q(X)) + \bar{Q}_{1n}(z) \equiv u'\mu + \bar{Q}_{1n}(z)$$

where the definition of μ and $\bar{Q}_{1n}(z)$ is obvious. Information matrix equality for α gives $-\mathcal{J} = E\frac{\partial^2\log f(\epsilon|X,\gamma_0)}{\partial\alpha\partial\alpha'}$, and CLT gives $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}f(\epsilon_i|X_i,\gamma_0) \xrightarrow{d} N(0,\mathcal{J})$. Hence

$$\bar{Q}_{1n}(z) \Rightarrow \bar{Q}_{1\infty}(z) \equiv W'v - \frac{1}{2}v'\mathcal{J}v \text{ in } \ell^{\infty}(\mathbf{Z}),$$

a continuous Gaussian process, provided we can show that this convergence is uniform by demonstrating stochastic equicontinuity of $Q_{1n}(z)$. In particular, for any $|z_1 - z_2| \to 0$, we can show that terms like

$$\left|\sum_{i=1}^{n} q_{in}(z_1)1(\epsilon_i \geq \Delta_n(X_i,u_1)/n \vee 0) - q_{in}(z_2)1(\epsilon_i \geq \Delta_n(X_i,u_2)/n \vee 0)\right| \xrightarrow{p} 0$$

Split the term into

$$\sum_{i=1}^{n}1(\epsilon_i > \Delta_n(X_i,u_1)/n \vee \Delta_n(X_i,u_2)/n \vee 0)$$

$$\times |\ln f(\epsilon_i - \Delta_n(X_i,u_1)/n|X_i;\beta_0+u_1/n,\alpha_0+v_1/n)$$

$$-\ln f(\epsilon_i - \Delta_n(X_i,u_2)/n|X_i;\beta_0+u_2/n,\alpha_0+v_2/n)|$$

$$+\sum_{i=1}^{n}1(\epsilon \in (0, \Delta_n(X_i,u_1)/n \vee \Delta_n(X_i,u_2)/n \vee 0))$$

$$\times \max_{j=1,2}\left|\ln\frac{f(\epsilon_i - \Delta_n(X_i,u_j)/n|\beta_0+u_j/n,\alpha_0+v_j/\sqrt{n})}{f(\epsilon_i|\beta_0,\alpha_0)}\right|.$$

By the argument of the proof of Theorem 1, the second term is bounded by $\sum_{i=1}^{n}1(\epsilon_i \in (0,C'/n \vee 0)) \times \frac{C''}{\sqrt{n}} \xrightarrow{p} 0$. Each summand in the first summation can be bounded by, for some z^* between z_1 and z_2,

$$\left|\sum_{i=1}^{n}1(\epsilon_i > \Delta_n(X_i,u_1)/n \vee \Delta_n(X_i,u_2)/n \vee 0) \times \right.$$

$$\left(\left|\frac{\partial}{\partial\beta}\ln f(\epsilon_i - \Delta_n(X_i,u^*)/n|\beta_0+u^*/n,\alpha_0+v^*/\sqrt{n})\right|\left|u_1-u_2\right|/n\right.$$

$$\left.\left.+\frac{\partial}{\partial\alpha}\ln f(\epsilon_i - \Delta_n(X_i,u^*)/n|\beta_0+u^*/n,\alpha_0+v^*/\sqrt{n})\frac{v_1-v_2}{\sqrt{n}}\right)\right|$$

Using assumption (E.5), the first term in this summation can be bounded by

$$C\frac{1}{n}\sum_{i=1}^{n}|\frac{\partial}{\partial\beta}\ln f(\epsilon_i|\gamma_0)|^{1+\delta}|u_1-u_2| = O_p(1)|u_1-u_2|$$

31

The second term in the summation can be further expanded as, for z^{**} between z^* and 0,

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

$$+ \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\partial^2}{\partial \alpha \partial \alpha'} \ln f \left(\epsilon_i - \Delta_n \left(X_i, u^{**} \right) / n | X_i, \beta_0 + \frac{u^{**}}{n}, \alpha_0 + \frac{v^{**}}{\sqrt{n}} \right) \right| \left| v^{**} \left(v_1 - v_2 \right) \right|$$

Assumption (E.6) implies that the second term is bounded by

$$\text{const} \times \left| v^{**} \left(v_1 - v_2 \right) \right| \frac{1}{n} \sum_{i=1}^{n} C''(\epsilon_i, X_i) = \left| v_1 - v_2 \right| O_p(1) .$$

The first term by the CLT and assumption (E.4) is $O_p(1)$ and is linear in $v_1 - v_2$. (Note that the indicator in the summation can be replaced by $1(\epsilon_t > 0)$ by Chebyshev inequality.) Thus as $|z_1 - z_2| \xrightarrow{p} 0$ the entire term goes to 0 in probability. This completes the proofs for stochastic equicontinuity and uniform convergence of $Q_{1n}(z)$ for the nonlinear model.

II. Limit of $Q_{2n}(z)$. As in the proof of Theorem 1, in the expression for $Q_{2n}(z)$ we can replace $q_{in}(z)$ by either $\ln p(X_i)/q(X_i)$ or $\ln p(X_i)/q(X_i)$, uniformly in z in $\mathbf{Z}$, depending on the sign of the r.v. ϵ_i. Note that this expressions may equal $-\infty$. So uniformly in z over $\mathbf{Z}$, $Q_{2n}(z) = \bar{Q}_{2n}(z) + o_p(1)$, where

$$\bar{Q}_{2n}(z) \equiv \sum_{i=1}^{n} \left[\ln \frac{q(X_i)}{p(X_i)} 1(0 < n\epsilon_i \leq \Delta_n(X_i, u)) + \ln \frac{p(X_i)}{q(X_i)} 1(0 > n\epsilon_i \geq \Delta_n(X_i, u)) \right].$$

Next we note that by straightforward calculations

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

$$+ \left| 1(0 > n\epsilon_i \geq \Delta_n(X_i, u)) - 1(0 > n\epsilon_i > \Delta(X_i, u)) \right| = o(1),$$

where $\Delta(X_i, u) \equiv \frac{\partial g(X_i, \beta_0)}{\partial \beta}' u$, which yields that for any fixed z, $\bar{Q}_{2n}(z) =$

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

Having obtained the "linearized expression," the finite-dimensional convergence of $\bar{Q}_{2n}(z)$ now follows by the arguments identical to those in Theorem 1. It remains to show that $(\bar{Q}_{2n}(z_i), i \leq J)$ and $(\bar{Q}_{1n}(z_i), i \leq J)$, for any finite J, are asymptotically independent. This follows by noting that the dependence between these terms is realized through the sums that disappear in probability for any fixed $z = (u, v)$ $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \alpha} \ln f(\epsilon_i | X_i, \gamma_0) 1(\epsilon_i \in (0, \Delta_n(X_i, u))) = o_p(1)$, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \alpha} \ln f(\epsilon_i | X_i, \gamma_0) 1(\epsilon_i \in (\Delta_n(X_i, u), 0)) = o_p(1)$ ∎

C.2 Proof of Theorem 5

Similar to theorem 2, the proof is a straightforward application of convexity lemma (1), by which it suffices to show the finite-dimensional (fidi) convergence of $\Gamma_n(\cdot)$ to $\Gamma_\infty(\cdot)$,

$$\left(\Gamma_n(z_k), k \leq K \right) \xrightarrow{d} \left(\Gamma_\infty(z_k), k \leq K \right).$$

As before we follow two steps.

1. First we approximate the integral $\Gamma_n(z)$ over $\mathbb{R}^d$ by an integral over a compact subset of $\mathbb{R}^d$. For each $M \in \mathbb{R}_+$, the approximation of $\Gamma_n(z)$ is given by $\Gamma_n^M(z) \equiv \Gamma_{n1}^M(z) / \Gamma_{n2}^M$ where

$$\Gamma_{n1}^M(z) \equiv \int_{|z'| \leq M} \rho\left(z - z'\right) \ell_n\left(z'\right) q\left(\beta_0 + H_n z'\right) du, \quad \Gamma_{n2}^M \equiv \int_{|z'| \leq M} \ell_n\left(z'\right) q\left(\beta_0 + H_n z'\right) dz'.$$

Define

$$\Gamma_{1\infty}^M(z) \equiv \int_{|z'| \leq M} \rho\left(z - z'\right) \ell_\infty\left(z'\right) q(\beta_0) du, \quad \Gamma_{2\infty}^M \equiv \int_{|z'| \leq M} \ell_\infty\left(z'\right) q(\beta_0) dz'.$$

Similar to (19), lemmas (7) and (8) in section (C.5) demonstrate that the tail of the likelihood ratio process is exponentially small: for each $\epsilon > 0, \delta > 0, \exists M \in \mathbb{R}_+$:

$$\limsup_n P\left(|\Gamma_n(z) - \Gamma_n^M(z)| > \epsilon\right) < \delta, \tag{20}$$

2. In view of (20), it suffices to show $\left(\Gamma_{n2}^M, \Gamma_{n1}^M(z_k), k \leq K\right) \overset{d}{\longrightarrow} \left(\Gamma_{2\infty}^M, \Gamma_{1\infty}^M(z_k), k \leq K\right)$. This follows by two facts:

   i $E\ell_n(z) = 1 < \infty$

   ii $E|\ell_n^{1/2}(z') - \ell_n^{1/2}(z)|^2 \leq C|z - z'|$

Fact i. is be definition of likelihood ratio. Fact ii. is by Lemma 7. These facts check the conditions of a limit theorem for integrals of random functions, Theorem 22 in Appendix I of IH (1982). ∎

C.3 Proof of Theorem 6

By assumption $-\ell_n$ is lower-semi-continuous. Two steps are needed:

   i. Show finite-dimensional convergence of ℓ_n to the stochastic limit ℓ_∞,

   ii. Show the stochastic equi-lower-semi-continuity (e-lsc) of $\{-\ell_n\}$,

Theorem 4 handled step i. We demonstrate step ii next. After having these two steps the conclusion follows by Lemma 2, since $Z_n = O_p(1)$ by the tail smallness lemma 8 and Theorem I.5.1 in IH(1982).

It therefore remains to show the stochastic equi-lower-semi-continuity (e-lsc) of $\{-\ell_n\}$ or equivalently of $\{-Q_{2n} \equiv -\log \ell_n\}$. From the proof of Theorem 4:

$$Q_{1n}(z) \Rightarrow u'\mu + \bar{Q}_{1\infty}(v), \text{ in } \ell^\infty(\mathbf{Z}) \tag{21}$$

where $\bar{Q}_{1\infty}(\cdot)$ is the gaussian process defined in the proof of Theorem 4 and $\mathbf{Z}$ is any compact subset of $\mathbb{R}^d$. Also

$$Q_{2n}(z) - \bar{Q}_{2n}(z) \overset{p}{\longrightarrow} 0,$$

uniformly in z over $\mathbf{Z}$, where

$$\bar{Q}_{2n}(z) \equiv \int_E l_{un}(j, x) d\widehat{\mathbf{N}}(j, x),$$

$$l_{un}(j, x) \equiv \ln \frac{q(x)}{p(x)} 1(0 < j \leq \Delta_n(x, u)) + \ln \frac{p(x)}{q(x)} 1(0 > j \geq \Delta_n(x, u)).$$

Because of (21), it suffices to show s-e-lsc of $\{-\bar{Q}_{2n}(z)\}$ only. Because $\bar{Q}_{2n}(z)$ depends on z only through u, we may write $\bar{Q}_{2n}(u)$ instead. Because $\bar{Q}_{2n}(u)$ is a piece-wise constant function

it suffices to show that for any bounded set $B \subset \mathbb{R}^{d_1}$ and $\delta > 0$, there are open neighborhoods $V(u_1), ... V(u_m)$ of some $u_1, ..., u_m$ s.t. $B \subset \cup_{k=1}^{m} V(u_k)$ and

$$P\left[\cup_{k=1}^{m}\{\inf_{u \in V(u_k)} -\bar{Q}_{2n}(u) < -\bar{Q}_{2n}(u_k)\}\right] < \delta.$$

This is done in several steps. The proof is nearly identical to the proof of Theorem 3, and only the minor differences are highlighted below.

(a). This step is *identical* to step (a) in the proof of Theorem 3, where the point process $\tilde{N}$ is constructed and its limit is found. This process allows to demonstrate the equisemi-continuity of $\bar{Q}_{2n}$. Write

$$\bar{Q}_{2n}(u) \equiv \int_{E:j>0} l_{un}(j,x) d\widehat{\mathbf{N}}(j,x) + \int_{E:j<0} l_{un}(j,x) d\widehat{\mathbf{N}}(j,x) \equiv \bar{Q}_{2n}^{+}(u) + \bar{Q}_{2n}^{-}(u).$$

We wish to examine the nature of discontinuities of $\bar{Q}_{2n}$ by examining that of $\bar{Q}_{2n}^{+}$ and $\bar{Q}_{2n}^{-}$. Because the arguments are identical for either part, consider only part $\bar{Q}_{2n}^{+}$.

(b) This step is identical to step (b) in the proof of Theorem 3 except that the construction of $\{V_{kj}\}$. Here for for sufficiently large n, $V_{kj} \equiv \{u \in \mathbb{R}^{d_1} : \underline{v}_k - \varphi < p(x)\Delta_n(x,u) < \underline{v}_k + \varphi, \forall x \in X_{\phi}(x_j)\}$, where $\varphi > 0$ and $\underline{v}_k = k\varphi$, for $k \in \{-m, ..., 0, ...m\}$. Observe that $\Delta_n(X,u) = \frac{\partial g(X,\beta_0 + u*(X)/n)}{\partial \beta}$ has a positive definite variance matrix uniformly in u^* in any compact set, as $n \to \infty$, by assumption. Thus the support of this vector in non-degenerate in $\mathbb{R}^{d_1}$. Thus, as in Theorem 3, since also $\mathbf{X}$ is compact and $p(x)$ is bounded away from 0 and above by assumption, $\{\cup_{kj} V_{kj}\}$ can cover any given bounded set B by selecting sufficiently large m, for large n.

(c)[Number of Break-points is $O_P(1)$ and separated for small φ] Consider argument z in $\cup_j V_{kj}$, then a discontinuity in $\bar{Q}_{2n}^{+}(u)$ (since it only depends on u through the index $\Delta_n(x,u)$) can *potentially* occur in the set $\cup_j V_{kj}$ only if there exist $u^{\bullet} \in \cup_j V_{kj}$ and (Γ_{in}, X_{in}) s.t.

$$\Gamma_{in} = p(X_{in})\Delta_n(X_{in}, u^{\bullet}), \tag{22}$$

where it should be the case for large n that $\underline{v}_k - \varphi \le \Gamma_{in} \le \underline{v}_k + \varphi$. If there is such (Γ_{in}, X_{in}), we say that $\bar{Q}_{2n}^{+}$ *has a breakpoint in* $\cup_j V_{kj}$. Define $\mathcal{N}_n \equiv \#\{i : \Gamma_{in} < \bar{k}\}$, $\mathcal{N} \equiv \#\{i : \Gamma_i < \bar{k}\}$, where $\bar{k} \equiv \sup_{x \in \mathbf{X}, u \in B} p(x)\Delta(x,u) + 2\varphi$. For sufficiently large n, $\mathcal{N}_n$ is the upper bound on the number of breakpoints of $\bar{Q}_{2n}^{+}(z)$ in set B. By continuous mapping theorem, $\mathcal{N}_n \xrightarrow{d} \mathcal{N}$ in $\mathbb{R}$. So the *number of breakpoints* $O(\mathcal{N}_n)$ is stochastically bounded $O_p(1)$. Furthermore, break-points are *separated* in the same sense as in the proof of Theorem 3. Define A_k to be the event that $\bar{Q}_{2n}^{+}(z)$ has more than two break-points in $\cup_j V_{kj}$. Then by the arguments that are identical to those in the proof of Theorem 3, for any $\delta > 0$ φ can be picked small so that $\limsup_n \ P[\cup_k A_k] \le \delta/2$.

(c)[grid-points by setting ϕ small] Next step is to construct the "centers" of u_{kj} of V_{kj}. Pick $u_{kj} \in V_{kj}$ so that for large n $\underline{v}_k - \varphi < \Delta_n(x, u_{kj}) < \underline{v}_k - \varphi + \eta$, $\forall x \in X_{\phi}(x_j)$, where $\eta : 0 < \eta << \varphi$ will be set sufficiently small in the next step. Depending on η, to satisfy the constraints set ϕ sufficiently small as well.

(d) [Stochastic Equi-semicontinuity] This step is identical to step (d) in the proof of Theorem 3, except we replace $X'z$ with $\Delta_n(X, u)$. ∎

C.4 Proof of Theorems 7 and 8

Proof of Theorem 7 The result is well known. Under the stated conditions, the posterior risk is finite for the Bayes estimator, so Theorem 1.1 in Ch.4 of Lehmann applies. ∎

Proof of Theorem 8. Z_n denotes the re-scaled Bayes estimator $H_n^{-1}(\hat{\gamma}_{Bayes} - \gamma)$

As a preliminary step we have for large n, for $c_1, c_2 > 0$ and any H

$$P^*\{|Z_n| > H\} \leq c_1 \exp\{-c_2|H|\}, \tag{23}$$

which follows by Theorem I.5.3 in Ibragimov and Hasminski that required the tails smallness conditions verified in Lemmas (7) and (8).

Z_n under P_γ depends on γ, which we emphasize by writing as Z_n^γ. (23) and majorzation of ρ by a polynomial imply that for some n_0,

$$\left\{\rho(Z_n^\gamma), n > n_0, \ \gamma \text{ in open ball at } \gamma_0 \right\} \text{ is uniformly integrable.} \tag{24}$$

Define $Z_n^\gamma(K)$ as $H_n^{-1}(f_{Bayes, \lambda_K} - \gamma)$, where f_{Bayes, λ_K} is the Bayes estimator defined with respect to the prior weight $\lambda_K(x) = 1\{H_n^{-1}(x - \gamma_0) \in K\}$. By construction for large H:

$$P^*\{|Z_n^\gamma(K)| > H\} = 0. \tag{25}$$

Thus for some n_0 and any compact sets K,

$$\left\{\rho(Z_n^\gamma(K) - \delta), n > n_0, \ \gamma \text{ in open ball at } \gamma_0 \right\} \text{ is uniformly integrable.} \tag{26}$$

Next we (a) evaluate

$$I \equiv \overline{R}_\rho(\{f_{Bayes}\}, \mathbb{R}^d) \text{ and } II(K) \equiv R_\rho(\{f_{Bayes, \lambda_K}\}, K)$$

and (b) show that $II(K)$ approaches I from below as $K \uparrow \mathbb{R}^d$.

Step (a). $\overline{R}_\rho(\{f_{Bayes}\}, K) = \lim_n \int_K E_{P_{\gamma_n(\delta)}} \rho(Z_n^{\gamma_n(\delta)}) d\delta / \text{Leb}(K)$

$$= \int_K E_{P_{\gamma_0}} \rho(Z_\infty) d\delta / \text{Leb}(K) = E_{P_{\gamma_0}} \rho(Z_\infty)$$

by Theorem 5, equations ((23)-(24)), and dominated convergence theorem. Thus

$$\overline{R}_\rho(\{f_{Bayes}\}, \mathbb{R}^d) = E_{P_{\gamma_0}} \rho(Z_\infty)$$

Analogously, $\overline{R}_\rho(\{f_{Bayes, \lambda_K}\}, K) = \lim_n \int_K E_{P_{\gamma_n(\delta)}} \rho(Z_n^{\gamma_n(\delta)}(K)) d\delta / \text{Leb}(K)$

$$= \int_K E_{P_{\gamma_0}} \rho(Z_\infty^\delta(K)) d\delta / \text{Leb}(K)$$

by Theorem 5(applies to $Z_n^{\gamma_n(\delta)}(K)$ by the same argument as it applies to $Z_n^{\gamma_n(\delta)}$) and equations (25)-(26), and dominated convergence, where

$$Z_\infty^\delta(K) \equiv \arg \inf_{z \in \mathbb{R}^d} \int_K \rho(z - \mu - \delta) \ell_\infty(\mu - \delta) d\mu.$$

Step (b). Next, for any δ, $\int_K E_{P_{\gamma_0}} \rho(Z_\infty^\delta(K)) d\delta / \text{Leb}(K) \leq E_{P_{\gamma_0}} \rho(Z_\infty)$. This follows because the lhs is the lower risk bound for $\overline{R}_\rho(\{f_n\}, K)$. Thus *no* other estimator sequence in $\{\mathbf{F}_n\}$ that differs from $\{f_{Bayes, \lambda_K}\}$ for i.o. n achieves lower risk value.

Rewrite the inequality as

$$\int_K E_{P_{\gamma_0}} \left[\rho(Z_\infty^\delta(K)) - \rho(Z_\infty)\right] d\delta / \text{Leb}(K) \equiv$$

$$\int_K E_{P_{\gamma_0}} \left[\rho(Z_\infty^\delta(K)) - \rho(Z_\infty)\right]^+ d\delta / \text{Leb}(K) - \int_K E_{P_{\gamma_0}} \left[\rho(Z_\infty^\delta(K)) - \rho(Z_\infty)\right]^- d\delta / \text{Leb}(K) \leq 0 \tag{27}$$

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

$$\int_K E_{P_{\gamma_0}} \left[\rho(Z_\infty^\delta(K)) - \rho(Z_\infty(K)) \right]^- d\delta/\text{Leb}(K) = \int_{(0,1)^d} E_{P_{\gamma_0}} \left[\rho(Z_\infty^{\mu r(K)}(K)) - \rho(Z_\infty) \right]^- d\mu \to 0.$$

(28)

where $r(K)$ denotes the width of the cube K (which is centered at zero by definition) . The last conclusion follows by **(a)** domination: for any $\mu \in (0,1)^d$, $r(K) \leq \infty$:

$$\left[\rho(Z_\infty^{\mu r(K)}(K)) - \rho(Z_\infty) \right]^- \leq \rho(Z_\infty)$$

(29)

so that $E_{P_{\gamma_0}} \left[\rho(Z_\infty^{\mu r(K)}(K)) - \rho(Z_\infty) \right]^- \leq E_{P_{\gamma_0}} \rho(Z_\infty)$ and **(b)** pointwise convergence: as $r(K) \to \infty$, for any $\mu \in (0,1)^d$

$$E_{P_{\gamma_0}} \left[\rho(Z_\infty^{\mu r(K)}(K)) - \rho(Z_\infty) \right]^- \to 0,$$

(30)

since **(i)** $Z_\infty^{\mu r(K)} \xrightarrow{p} Z_\infty$ for $\mu \in (0,1)^d$ by definition and convexity Lemma 1, due to fi-di convergence of the objective function that $Z_\infty^{\mu r(K)}$ minimizes to the objective function that Z_∞ minimizes, in probability; **(ii)** by (29) the collection of variables

$$\left\{ \left[\rho(Z_\infty^{\mu r(K)}(K)) - \rho(Z_\infty) \right]^-, \quad \text{s.t. } \mu \in (0,1)^d, r(K) \leq \infty \right\} \text{ is uniformly integrable.}$$

(27)- (30) also imply that it must be that $\int_K E_{P_{\gamma_0}} \left[\rho(Z_\infty^\delta(K)) - \rho(Z_\infty) \right]^+ d\delta/\text{Leb}(K) \to 0$ as $r(K) \to \infty$. Thus $II(K) \uparrow I$ as $K \uparrow \mathbb{R}^d$.

Note that this shows that $\{f_{Bayes}\}$ minimizes $\overline{R}_\rho(\{f\}, \mathbb{R}^d)$. Suppose that there is a sequence $\{f_n^*\}$ in $\{\mathbf{F}_n\}$ that achieves lower risk value than $E_{P_{\gamma_0}} \rho(Z_\infty)$. But then this sequence must achieve a *lower* value than $II(K)$ for some large K, which is impossible by the previous comment. ∎

C.5   Lemmas 7-8

In this section some important properties of

$$\ell_n(z) = L_n(\gamma(z))/L_n(\gamma).$$

are established, where $\gamma(z) \equiv \gamma + H_n z$ for $z \equiv (u,v) \in \mathbb{R}^d$.

First we note the conditions of lemma 4 that uniformly in γ in a ball at γ_0

$$E_{P_\gamma} \ell_n(z)^{1/2} \leq e^{-b|z|}, \qquad E_{P_\gamma} |\ell_n(z)^{1/2} - \ell_n(z')^{1/2}| \leq B|z - z'|$$

By the proof of lemma I.5.2 of IH(1982) the first inequality only needs to hold for z large, and the second inequality only needs to hold for $|z - z'| < 1$.

For the nonlinear model $\bar{\mathbf{R}}$, the conditional Hellinger distance $r_2(\gamma; \gamma + h)^2$ is defined as

$$\int \int |f^{1/2}(y - g(x, t + h_1); x, \gamma + h) - f^{1/2}(y - g(x, t); x, \gamma)|^2 dy F_X(dx)$$

Next we obtain the nonlinear versions of lemmas (5) and (6) for the model $\bar{\mathbf{R}}$.

Lemma 7 *If $\exists a > 0, A > 0$, such that for $h > 0$ small enough for γ in a ball at γ_0*

$$\inf_\gamma r_2^2(\gamma; \gamma + h) \geq a \max\left(|h_1|, |h_2|^2\right) \quad and \quad \sup_\gamma r_2^2(\gamma; \gamma + h) \leq A\left(|h_1| + |h_2|^2\right)$$

then $\exists b > 0, B > 0$, s.t. for all u large, $\forall z \in R^d$, $z' \in R^d$, $|z - z'| < 1$. $\exists n_0$, s.t. $\forall n > n_0$,

$$\sup_\gamma E_{P_\gamma} \ell_n(z)^{1/2} \leq e^{-b|z|}, \qquad \sup_\gamma E_{P_\gamma} |\ell_n(z)^{1/2} - \ell_n(z')^{1/2}| \leq B|z - z'|$$

Proof: For z large, the same calculation as in lemma 5 leads to

$$E_{P_\gamma} \ell_n (z)^{1/2} = \left[1 - \frac{1}{2} r_2^2 \left(\gamma_0; \beta_0 + u/n, \alpha_0 + v/\sqrt{n} \right) \right]^n$$

$$\leq e^{- \frac{n}{2} r_2^2 \left(\gamma_0; \beta_0 + u/n, \alpha + v/\sqrt{n} \right)} \leq e^{-b \max \left(|u|, |v|^2 \right)} \leq e^{-b \max(|u|, |v|)} \leq e^{-b|z|}$$

On the other hand, for $|z - z'|$ small,

$$E_{P_\gamma} |\ell_n (z)^{1/2} - \ell_n (z')^{1/2}| \leq n r_2^2 \left(\gamma_0 + \left(u/n, v/\sqrt{n} \right); \gamma_0 + \left(u'/n, v'/\sqrt{n} \right) \right) \leq A \left(|u - u'| + |v - v'|^2 \right)$$

$$\leq A \left(|z - z'| + |z - z'|^2 \right) \leq B \left(|z - z'| \right) \quad \blacksquare$$

Lemma 8 *In model* $\bar{R}$*, suppose condition (E.1)-(E.6) hold, then* $\exists a > 0$*,* $\exists A > 0$*, s.t.* $\forall h > 0$ *small enough, for* γ *in a ball at* γ_0

$$\inf_\gamma r_2^2 \left(\gamma; \gamma + h \right) \geq a \max \left(|h_1|, |h_2|^2 \right) \quad and \quad \sup_\gamma r_2^2 \left(\gamma; \gamma + h \right) \leq A \left(|h_1| + |h_2|^2 \right)$$

Proof of Lemma 8 For $\epsilon > 0$ and $\eta > 0$ small enough, let $\gamma = (t, s), h = (\epsilon, \eta)$,

$$r_2^2(\gamma; \gamma + h) \equiv E_X \int \left(f^{1/2} \left(y - g\left(X, t + \epsilon \right) | X; t + \epsilon, s + \eta \right) - f^{1/2} \left(y - g\left(X, t \right) | X; t, s \right) \right)^2 dy$$

$$\stackrel{(1)}{\leq} E_X \int_{[g(X,t), g(X,t+\epsilon)]} \left| f \left(y - g\left(X, t + \epsilon \right) | X; t + \epsilon, s + \eta \right) - f \left(y - g\left(X, t \right) | X; t, s \right) \right| dy$$

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

$$+ E_X \int_{[g(X,t), g(X,t+\epsilon)]^c} \left| f \left(y - g\left(X, t + \epsilon \right) | X; t + \epsilon, s \right) - f \left(y - g\left(X, t \right) | X; t, s \right) \right| dy$$

$$\stackrel{(2)}{\leq} 2 E_X \left| g\left(X, t + \epsilon \right) - g\left(X, t \right) \right| \left(p\left(X, \gamma \right) + q\left(X, \gamma \right) \right)$$

$$+ |\eta|^2 \int_0^1 E_X \int \left| \frac{\partial f^{1/2} \left(y - g\left(X, t + \epsilon \right) | X, t + \epsilon, s + \omega \eta \right)}{\partial s} \right|^2 dy d\omega$$

$$+ |\epsilon| \int_0^1 E_X \int \left| \frac{\partial f \left(y - g\left(X, t + \omega\epsilon \right) | X, t + \omega\epsilon, s \right)}{\partial t} \right| dy d\omega$$

$$\stackrel{(3)}{\leq} 2|\epsilon| E_X \left(p\left(X, \gamma \right) + q\left(X, \gamma \right) \right) \int_0^1 \left| \frac{\partial g\left(X, t + s\epsilon \right)}{\partial t} \right| ds F_X \left(dX \right) + O \left(|\eta|^2 \right) + O \left(|\epsilon| \right)$$

$$= O \left(|\epsilon| \right) + O \left(|\eta|^2 \right).$$

where $[a, b] = [a, b]$ if $a \leq b$ and $= [b, a]$ if $b \leq a$, and the bound is uniform in γ by (E.4). The first inequality follows by triangle inequality and from $|a - b|^2 \leq |a^2 - b^2|$ for $a > 0$ and $b > 0$. The first term in the second inequality follows from that fact that for $y \in [g\left(X, t \right), g\left(X, t + \epsilon \right)]$ and ϵ small enough, $|f \left(\cdot | \cdot \right)| \leq 2 \left(p\left(X, \gamma \right) + q\left(X, \gamma \right) \right)$, and then we integrate over y over that area. The second and third terms in the second inequality are usual multivariate first order Taylor expansions and Fubini. The first term in the third inequality follows from Taylor expansion and Fubini. The second term in the third inequality follows from assumption (E.4), while the third term in that inequality follows from assumption (E.2).

To obtain a bound from below, consider $r_2^2 \left(\gamma; \gamma + h \right)$:

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

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

37

We can bound the first term from below uniformly in γ by

$$E_X \frac{1}{2} \left| g\left(X, t+\epsilon\right) - g\left(X, t\right) \right| \left\| p^{1/2}\left(X, \gamma\right) - q^{1/2}\left(X, \gamma\right) \right\|^2 \geq c|\epsilon| E_X \left| \frac{\partial g\left(X; t\right)}{\partial t} \right| \geq c|\epsilon|$$

using assumption (E.3), Taylor expansion and Cauchy-Schwartz inequality. On the other hand, bound the second term of $r_2^2\left(\gamma; \gamma + h\right)$ uniformly below by:

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

$$= E_X \int_{[g(X,t),g(X,t+\epsilon)]^c} \left(h' \frac{\partial f^{1/2}\left(y - g\left(X; t\right) | \gamma\right)}{\partial \gamma} \right)^2 dy + o\left(|h|^2\right)$$

Under assumption (E.4) (a), a lower bound is $|h|^2 \inf_{|u|=1} E_X \int_{[g(X,t),g(X,t+\epsilon)]^c} \left(\frac{\partial f^{1/2}(y-g(X,t);\gamma)}{\partial \gamma}' u \right)^2 dy$

$$= |h|^2 \inf_{|u|=1} E_X \int \left(\frac{\partial f^{1/2}\left(y - g\left(X, t\right); \gamma\right)}{\partial \gamma}' u \right)^2 dy + o\left(|h|^2\right) \geq c|h|^2 \geq c|\eta|^2$$

On the other hand, if assumption (E.4) (b) holds, the uniform lower bound is

$$E_X \int \left(h' \frac{\partial f^{1/2}\left(y - g\left(X; t\right) | \gamma\right)}{\partial \gamma} \right)^2 dy \geq |\eta|^2 \inf_{|u|=1} E_X \int \left(\frac{\partial f^{1/2}\left(y - g\left(X, t\right); \gamma\right)}{\partial s}' u \right)^2 dy \geq c|\eta|^2$$

Thus, conclude that $\inf_\gamma r_2^2\left(\gamma; \gamma + h\right) \geq c \max\left(|\epsilon|, |\eta^2|\right)$. ∎

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.

BAJARI, P. (1999): "Auction Models Are Not Robust When Bidders Make Small Mistakes," Stanford University.

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. (2000): "Conditional Extremes and Near Extremes," mimeo. MIT and Stanford University.

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

DONALD, S., AND H. PAARSCH (2000): "Superconsistent Estimation and Inferences in Structural Econometric Models using Extreme Order Statistics," University of Texas and University of Iowa.

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.

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

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68(3), 525–574.

HIRANO, K., AND J. PORTER (2001): "Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support," UCLA and Harvard University.

HONG, H., AND M. SHUM (2000): "Increasing Competition and the Winner's Curse: Evidence from Procurement," manuscript.

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

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

KNIGHT, K. (1999): "Epi-convergence and Stochastic Equisemicontinuity," Preprint.

LEADBETTER, M., G. LINDGREN, AND H. ROOTZEN (1983): *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag.

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.

POLITIS, D., J. ROMANO, AND M. WOLF (1999): *Subsampling*. Springer Series in Statistics.

RESNICK, S. I. (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.

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

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak convergence and empirical processes*. Springer-Verlag, New York.

VUONG, Q. (1989): "Likelihood-ratio tests for model selection and non-nested hypotheses," *Econometrica*, pp. 307–333.