[redacted]

Massachusetts Institute of Technology
Department of Economics
Working Paper Series

AN MCMC APPROACH TO CLASSICAL ESTIMATION

Victor Chernozhukov
Han Hong

Working Paper 03-21
December 2002

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

An MCMC Approach to Classical Estimation

Victor Chernozhukov[a], Han Hong[b]

[a] Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

[b] Department of Economics, Princeton University, Princeton, NJ 08544, USA

First Version: October 2000 This Version: December 2002 [1]

Project Funded by the National Science Foundation.

Abstract

This paper studies computationally and theoretically attractive estimators referred here as to the Laplace type estimators (LTE). The LTE include means and quantiles of Quasi-posterior distributions defined as transformations of general (non-likelihood-based) statistical criterion functions, such as those in GMM, nonlinear IV, empirical likelihood, and minimum distance methods. The approach generates an alternative to classical extremum estimation and also falls outside the parametric Bayesian approach. For example, it offers a new attractive estimation method for such important semi-parametric problems as censored and instrumental quantile regression, nonlinear IV, GMM, and value-at-risk models. The LTE's are computed using Markov Chain Monte Carlo methods, which help circumvent the computational curse of dimensionality. A large sample theory is obtained and illustrated for regular cases.

JEL Classification: C10, C11, C13, C15

Keywords: Laplace, Bayes, Markov Chain Monte Carlo, GMM, instrumental regression, censored quantile regression, instrumental quantile regression, empirical likelihood, value-at-risk

1 Introduction

A variety of important econometric problems pose not only a theoretical but a serious computational challenge, cf. Andrews (1997). A small (and by no means exhaustive) set of such examples include (1) Powell's censored median regression for linear and nonlinear problems, (2) nonlinear IV estimation, e.g in the Berry et al. (1995) model, (3) the instrumental quantile regression, (4) the continuous-updating GMM estimator of Hansen et al. (1996), and related empirical likelihood problems. These problems represent a formidable practical challenge as the extremum estimators are known to be difficult to compute due to highly nonconvex criterion functions with many local optima (but well pronounced global optimum). Despite extensive efforts, see notably Andrews (1997), the problem of extremum computation remains a formidable impediment in these applications.

[1] A shorter version of this paper is forthcoming in Journal of Econometrics 115 (August 2003), p. 293-346

This paper develops a class of estimators, which we call the Laplace type estimators (LTE) or Quasi-Bayesian estimators (QBE),[2] which are defined similarly to Bayesian estimators but use general statistical criterion functions in place of the parametric likelihood function. This formulation circumvents the curse of dimensionality inherent in the computation of the classical extremum estimators by instead focusing on LTE which are functions of integral transformations of the criterion functions and can be computed using Markov Chain Monte Carlo methods (MCMC), a class of simulation techniques from Bayesian statistics. This formulation will be shown to yield both computable and theoretically attractive new estimators to such important problems as (1)-(4) listed above. Although the aforementioned applications are mostly microeconometric, the obtained results extend to many other models, including GMM and quasi-likelihoods in the nonlinear dynamic framework of Gallant and White (1988).

The class of LTE's or QBE's aim to explore the use of the Laplace approximation (developed by Laplace to study large sample approximations of Bayesian estimators and for use in other non-statistical problems) outside of the canonical Bayesian framework – that is, outside of parametric likelihood settings when the likelihood function is not known. Instead, the approach relies upon other statistical criterion functions of interest in place of the likelihood, transforms them into proper distributions – Quasi-posteriors – over a parameter of interest, and defines various moments and quantiles of that distribution as the point estimates and confidence intervals, respectively. It is important to emphasize that the underlying criterion functions are mainly motivated by the analogy principle in place of the likelihood principle, are not the likelihoods (densities) of the data, and are most often semi-parametric.[3]

The resulting estimators and inference procedures possess a number of good theoretical and computational properties and yield new, alternative approaches for the important problems mentioned earlier. The estimates are as efficient as the extremum estimates; and, in many cases, the inference procedures based on the quantiles of the Quasi-posterior distribution or other posterior quantities yield asymptotically valid confidence intervals, which also perform notably well in finite samples. For example, in the quantile regression setting, those intervals provide valid large sample and excellent small sample inference without requiring nonparametric estimation of the conditional density function (needed in the standard approach). The obtained results are general and useful – they cover the examples listed above under general, non-likelihood based conditions that allow discontinuous, non-smooth semi-parametric criterion functions, and data generating processes that range from iid settings to the nonlinear dynamic framework of Gallant and White (1988). The results thus extend the theoretical work on large sample theory of Bayesian procedures in econometrics and statistics, e.g. Bickel and Yahav (1969), Ibragimov and Has'minskii (1981), Andrews (1994b), Kim (1998).

The LTE's are computed using MCMC, which simulates a series of parameter draws such that

[2] A preferred terminology is taken here to be the 'Laplace Type Estimators', since the term 'Quasi-Bayesian Estimators' is already used to name Bayesian procedures that use either 'vague' or 'data-dependent' prior or multiple priors, cf. Berger (2002).

[3] In this paper, the term 'semi-parametric' refers to the cases where the parameters of interest are finite-dimensional but there are nonparametric nuisance parameters such as unspecified distributions.

the marginal distribution of the series is (approximately) the Quasi-posterior distribution of the parameters. The estimator is therefore a function of this series, and may be given explicitly as the mean or a quantile of the series, or implicitly as the minimizer of a smooth globally convex function.

As stated above, the LTE approach is motivated by the estimation and inference efficiency as well as computational attractiveness. Indeed, the LTE approach is as efficient as the extremum approach, but generally may not suffer from the computational curse of dimensionality (through the use of MCMC). The reason is that the computation of LTE's is itself statistically motivated. LTE's are typically means or quantiles of a quasi-posterior distribution, hence can be estimated (computed) at the parametric rate $1/\sqrt{B}$, where B is the number of draws from that distribution (functional evaluations). In contrast, the mode (extremum estimator) is estimated (computed) by the MCMC and similar grid-based algorithms at the nonparametric rate $(1/B)^{\frac{p}{d+2p}}$, where d is the parameter dimension and p is the smoothness order of the objective function.

Another useful feature of LT estimation is that, by using information about the overall shape of the objective function, point estimates and confidence intervals may be calculated simultaneously. It also allows incorporation of prior information, and allows for a simple imposition of constraints in the estimation procedure.

The remainder of the paper proceeds as follows. Section 2 formally defines and further motivates the Laplace type estimators with several examples, reviews the literature, and explains other connections. The motivating examples, which are all semi-parametric and involve no parametric likelihoods, will justify the pursuit of a more general theory than is currently available. Section 3 develops the large sample theory, and Sections 3 and 4 further explore it within the context of the econometric examples mentioned earlier. Section 4 briefly reviews important computational aspects and illustrates the use of the estimator through simulation examples. Section 5 contains a brief empirical example, and Section 6 concludes.

Notation. Standard notation is used throughout. Given probability measure P, $\rightarrow_p$ denotes the convergence in (outer) probability with respect to the outer probability P^*; $\rightarrow_d$ denotes the convergence in distribution under P^*, etc. See e.g. van der Vaart and Wellner (1996) for definitions. $|x|$ denotes the Euclidean norm $\sqrt{x'x}$; $B_\delta(x)$ denotes the ball of radius δ centered at x. A notation table is given in the appendix.

2 Laplacian or Quasi-Bayesian Estimation: Definition and Motivation

2.1 Motivation

Extremum estimators are usually motivated by the analogy principle and defined as maximizers of random average-like criterion functions $L_n(\theta)$, where n denotes the sample size. $n^{-1}L_n(\theta)$ are typically viewed as transformations of sample averages that converge to criterion functions $M(\theta)$ that are maximized uniquely at some θ_0. Extremum estimators are usually consistent and asymptotically normal, cf. Amemiya (1985), Gallant and White (1988), Newey and McFadden (1994), Pötscher and Prucha (1997). However, in many important cases, actually computing the extremum estimates remains a large problem, as discussed by Andrews (1997).

Example 1: ***Censored and Nonlinear Quantile Regression.*** A prominent model in econometrics is the censored median regression model of Powell (1984). Powell's censored quantile regression estimator is defined to maximize the following nonlinear objective function

$$L_n(\theta) = -\sum_{i=1}^{n} \omega_i \cdot \rho_\tau(Y_i - q(X_i, \theta)), \quad q(X_i, \theta) = \max(0, g(X_i, \theta)),$$

where $\rho_\tau(u) = (\tau - 1(u < 0))\, u$ is the check function of Koenker and Bassett (1978), ω_i is a weight, and Y_i is either positive or zero. Its conditional quantile $q(X_i, \theta)$ is specified as $\max(0, g(X_i, \theta))$. The censored quantile regression model was first formulated by Powell (1984) as a way to provide valid inference in Tobin-Amemiya models without distributional assumptions and with heteroscedasticity of unknown form. The extremum estimator based on the Powell's criterion function, while theoretically elegant, has a well-known computational difficulty. The objective function is similar to that plotted in Figure 1 – it is nonsmooth and highly nonconvex, with numerous local optima, posing a formidable obstacle to the practical use of this extremum estimator; see Buchinsky and Hahn (1998), Buchinsky (1991), Fitzenberger (1997), and Khan and Powell (2001) for related discussions. In this paper, we shall explore the use of LT estimators based on Powell's criterion function and show that this alternative is attractive both theoretically and computationally.

Example 2: ***Nonlinear IV and GMM.*** Amemiya (1977), Hansen (1982), Hansen et al. (1996) introduced nonlinear IV and GMM estimators that maximize

$$L_n(\theta) = -\frac{1}{2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_i(\theta)\right)' W_n(\theta) \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_i(\theta)\right) + o_p(1),$$

where $m_i(\theta)$ is a moment function defined such that the economic parameter of interest solves

$$Em_i(\theta_0) = 0.$$

The weighting matrix may be given by $W_n(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n} m_i(\theta)\, m_i(\theta)'\right]^{-1} + o_p(1)$ or other sensible choices. Note that the term "$o_p(1)$" in L_n implicitly incorporates generalized empirical likelihood estimators, which will be discussed in section 4. Up to the first order, objective functions of empirical likelihood estimators for θ (with the Lagrange multiplier concentrated out) locally coincide with L_n. Applications of these estimators are numerous and important (e.g. Berry et al. (1995), Hansen et al. (1996), Imbens (1997)), but while global maxima are typically well-defined, it is also typical to see many local optima in applications. This leads to serious difficulties with applying the extremum approach in applications where the parameter dimension is high. As in the previous example, LTE's provide a computable and theoretically attractive alternative to extremum estimators. Furthermore, Quasi-posterior quantiles provide a valid and effective way to construct confidence intervals and explore the shape of the objective function.

Example 3: ***Instrumental and Robust Quantile Regression.*** Instrumental quantile regression may be defined by maximizing a standard nonlinear IV or GMM objective function[4]

$$L_n(\theta) = -\frac{1}{2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_i(\theta)\right)' W_n(\theta)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_i(\theta)\right),$$

where

$$m_i(\theta) = (\tau - 1(Y_i \leq q(D_i, X_i, \theta))\, Z_i,$$

Y_i is the dependent variable, D_i is a vector of possibly endogeneous variables, X_i is a vector of regressors, Z_i is a vector of instruments, and $W_n(\theta)$ is a positive definite weighting matrix, e.g.

$$W_n(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n} m_i(\theta)\, m_i(\theta)'\right]^{-1} + o_p(1) \quad \text{or} \quad W_n(\theta) = \frac{1}{\tau(1-\tau)}\left[\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'\right]^{-1},$$

or other sensible versions. Motivations for estimating equations of this sort arise from traditional separable simultaneous equations, cf. Amemiya (1985), and also more general nonseparable simultaneous equation models and heterogeneous treatment effect models.[5]

Clearly, a variety of Huber (1973) type robust estimators can be defined in this way. For example, suppose in the absence of endogeneity

$$q(X, \theta) = X'\beta(\tau),$$

then $Z = f(X)$ can be constructed to preclude the influence of outliers in X on the inference. For example, choosing $Z_{ij} = 1(X_{ij} < x_j)$, $j = 1, ..., \dim(X)$, where x_j denotes the median of $\{X_{ij}, i \leq n\}$ produces an approach that is similar in spirit to the maximal regression depth estimator of Rousseeuw and Hubert (1999), whose computational difficulty is well known, as discussed in van Aelst et al. (2002). The resulting objective function $L_n(\beta)$ is highly robust to both outliers in X_{ij} and Y_i. In fact, it appears that the breakdown properties of this objective function are similar to those of the objective function of Rousseeuw and Hubert (1999).

Despite a clear appeal, the computational problem is daunting. The function L_n is highly non-convex, almost everywhere flat, and has numerous discontinuities and local optima.[6] (Note that the global optimum is well pronounced.) Figure 1 illustrates the situation. Again, in this case the LTE approach will yield a computable and theoretically attractive alternative to the extremum-based estimation and inference.[7] Furthermore, we will show that the Quasi-posterior confidence intervals provide a valid and effective way to construct confidence intervals for parameters and their smooth

[4]Early variants based on the Wald instruments go back to Mood (1950) and Hogg (1975), cf. Koenker (1998).

[5]See Chernozhukov and Hansen (2001) for the development of this direction.

[6]Macurdy and Timmins (2001) propose to smooth out the edges using kernels, however this does not eliminate non-convexities and local optima; see also Abadie (1995).

[7]Another computationally attractive approach, based on an extension of Koenker and Bassett (1978) quantile regression estimator to instrumental problems like these, is given in Chernozhukov and Hansen(2001).

functions without non-parametric estimation of the conditional density function evaluated at quantiles (needed in standard approach).

The LTE's studied in this paper can be easily computed through Markov Chain Monte Carlo and other posterior simulation methods. To describe these estimators, note that although the objective function $L_n(\theta)$ is generally not a log-likelihood function, the transformation

$$p_n(\theta) = \frac{e^{L_n(\theta)}\pi(\theta)}{\int_\Theta e^{L_n(\theta)}\pi(\theta)d\theta} \tag{2.1}$$

is a proper distribution density over the parameter of interest, called here the *Quasi-posterior.* Here $\pi(\theta)$ is a weight or prior probability density that is strictly positive and continuous over Θ, for example, it can be constant over the parameter space. Note that p_n is generally not a true posterior in the Bayesian sense, since it may not involve the conditional data density or likelihood, and is thus generally created through non-Bayesian statistical learning.

The Quasi-posterior mean is then defined as

$$\hat{\theta} = \int_\Theta \theta p_n(\theta)d\theta = \int_\Theta \theta\left(\frac{e^{L_n(\theta)}\pi(\theta)}{\int_\Theta e^{L_n(\theta)}\pi(\theta)d\theta}\right)d\theta, \tag{2.2}$$

where Θ is the parameter space. Other quantities such as medians and quantiles will also be considered. A formal definition of LTE's is given in Definition 1.

In order to compute these estimators, using Markov Chain Monte Carlo methods, we can draw a Markov chain (see Figure 1),

$$S = \left(\theta^{(1)}, \theta^{(2)}, ..., \theta^{(B)}\right),$$

whose marginal density is approximately given by $p_n(\theta)$, the Quasi-posterior distribution. Then the estimate $\widehat{\theta}$, e.g. the Quasi-posterior mean, is computed as

$$\widehat{\theta} = \frac{1}{B}\sum_{i=1}^{B}\theta^{(i)}.$$

Analogously, for a given continuously differentiable function $g : \Theta \to \mathbb{R}$, the 90%-confidence intervals are constructed simply by taking the .05-th and .95-th quantiles of the sequence

$$g(S) = \left(g(\theta^{(1)}), ..., g(\theta^{(B)})\right),$$

see Figure 1. Under the information equality restrictions discussed later, such confidence regions are asymptotically valid. Under other conditions, it is possible to use other Quasi-posterior quantities such as the variance-covariance matrix of the series S to define asymptotically valid confidence regions, see Section 3. It shall be emphasized repeatedly that the validity of this approach does not depend on the likelihood formulation.

2.2 Formal Definitions

Let $\rho_n(u)$ be a *penalty or loss function* associated with making an incorrect decision. Examples of $\rho_n(u)$ include

i. $\rho_n(u) = |\sqrt{n}u|^2$, the squared loss function,

ii. $\rho_n(u) = \sqrt{n}\sum_{j=1}^{d} |u_j|$, the absolute deviation loss function,

iii. $\rho_n(u) = \sqrt{n}\sum_{j=1}^{d} (\tau_j - 1(u_j \leq 0))\, u_j$, for $\tau_j \in (0,1)$ for each j, the check loss function of Koenker and Bassett (1978).

The parameter is assumed to belong to the subset Θ of Euclidean space. Using the Quasi-posterior p_n density in (2.1), define the Quasi-posterior risk function as:

$$Q_n(\zeta) = \int_\Theta \rho_n(\theta - \zeta)\, p_n(\theta)\, d\theta = \int_\Theta \rho_n(\theta - \zeta) \left(\frac{e^{L_n(\theta)}\pi(\theta)}{\int_\Theta e^{L_n(\theta)}\pi(\theta) d\theta} \right) d\theta. \tag{2.3}$$

Definition 1 *The class of LTE minimize the function $Q_n(\zeta)$ in (2.3) for various choices of ρ_n:*

$$\hat{\theta} = \arg \inf_{\zeta \in \Theta} \Big[Q_n(\zeta) \Big]. \tag{2.4}$$

The estimator $\widehat{\theta}$ is a decision rule that is least unfavorable given the statistical (non-likelihood) information provided by the probability measure p_n, using the loss function ρ_n. In particular, the loss function ρ_n may asymmetrically penalize deviations from the truth, and π may give differential weights to different values of θ. The solutions to the problem (2.4) for loss functions **i-iii** include the Quasi-posterior means, medians, and marginal τ_j-th quantiles, respectively.[8]

2.3 Related Literature

Our analysis will rely heavily on the previous work on Bayesian estimators in the likelihood setting. The initial large sample work on Bayesian estimators was done by Laplace (see Stigler (1975) for a detailed review). Further early work of Bernstein (1917) and von Mises (1931) has been considerably extended in both econometric and statistical research, cf. Ibragimov and Has'minskii (1981), Bickel and Yahav (1969), Andrews (1994b), Phillips and Ploberger (1996), and Kim (1998), among others.

In general, Bayesian asymptotics require very delicate control of the tail of the posterior distribution and were developed in useful generality much later than the asymptotics of extremum estimators. The treatments of Bickel and Yahav (1969) and Ibragimov and Has'minskii (1981) are most useful for the present setting, but are inevitably tied down to the likelihood setting. For example, the

[8]This formulation implies that conditional on the data, the decision $\widehat{\theta}$ satisfies Savage's axioms of choice under uncertainty with subjective probabilities given by p_n (these include the usual asymmetry and negative transitivity of strict preference relationship, independence, and some other standard axioms).

latter treatment relies heavily on Hellinger bounds that are firmly rooted in the objective function being a likelihood of iid data. However, the general flavor of the approach is suited for the present purposes. The treatment of Bickel and Yahav (1969) can be easily extended to smooth, possibly incorrect iid likelihoods,[9] but does not apply to censored median regression or any of the GMM type settings. Andrews (1994b) and Phillips and Ploberger (1996) study the large sample approximation of posteriors and posterior odds ratio tests in relation to the classical Wald tests in the context of smooth, correctly specified likelihoods. Kim (1998) derives the limit behavior of posteriors in likelihood models over shrinking neighborhood systems. Kim's approach and related approaches have been important in describing the essence of posterior behavior, but the limit behavior of point estimates like ours does not follow from it.[10]

Formally and substantively, none of the above treatments apply to our motivating examples and the estimators given in Definition 1. These examples do not involve likelihoods, deal mostly with GMM type objective functions, and often involve discontinuous and non-smooth criterion functions to which the above mentioned results do not apply. In order to develop the theory of LTE's for such examples, we extend the previous arguments. The results obtained here enable the use of Bayesian tools outside of the Bayesian framework – covering models with non-likelihood-based criterion functions, such as examples listed earlier and other semi-parametric objective functions that may, for example, depend on preliminary estimates of infinite-dimensional nuisance parameters. Moreover, our results apply to general forms of data generating processes – from the cross-sectional framework of Amemiya (1985) to the nonlinear dynamic framework of Gallant and White (1988) and Pötscher and Prucha (1997).

Our motivating problems are all semi-parametric, and there are several pure Bayesian approaches to such problems, see notably Doksum and Lo (1990), Diaconis and Freedman (1986), Hahn (1997), Chamberlain and Imbens (1997), Kottas and Gelfand (2001). Semi-parametric models have some parametric and nonparametric components, e.g. the unspecified nonparametric distribution of data in Examples 1-3. The mentioned papers proceed with the pure Bayesian approach to such problems, which involves Bayesian learning about these two components via a two-step process. In the first step, Bayesian non-parametric learning with Dirichlet priors is used to form beliefs about the joint non-parametric density of data, and then draws of the non-parametric density ("Bayesian bootstrap") are made repeatedly to compute the extremum parameter of interest. This approach is purely Bayesian, as it fully conforms to the Bayes learning model. It is clear that this approach is generally quite different from LTE's or QBE's studied in this paper, and in applications, it still requires numerous re-computations of the extremum estimates in order to construct the posterior distribution over the parameter of interest. In sharp contrast, the LT estimation takes a "shortcut" by essentially using the common criterion functions as posteriors, and thus entirely avoids both the estimation of the nonparametric distribution of the data and the repeated computation of extremum estimates.

[9]See Bunke and Milhaud (1998) for an extension to the more than three times differentiable smooth misspecified iid likelihood case. The conditions do not apply to GMM or even Example 1.

[10]E.g. to describe the behavior of posterior median one needs to know $\int_{-\infty}^{x} p_n(\theta)d\theta$ which requires the study of $L_n(\theta)$ beyond the compact $1/\sqrt{n}$ neighborhoods of θ_0. Similarly, the posterior mean is $\int_{-\infty}^{\infty} \theta p_n(\theta)d\theta$, which also requires the study of the complete $L_n(\theta)$.

Finally note that the LTE approach has a limited-information or semi-parametric nature in the sense that we do not know or are not willing to specify the complete data density. The limited-information principle is powerfully elaborated in the recent work of Zellner (1998), who starts with a set of moment conditions, calculates the maximum entropy densities consistent with the moment equations, and uses those as formal (misspecified) likelihoods. While in the present framework, calculation of the maximum entropy densities is not needed, the large sample theory obtained here does cover Zellner's (1998) estimators as one fundamental case. Related work by Kim (2002) derives a limited information likelihood interpretation for certain smooth GMM settings.[11] In addition, the LTE's based on the empirical likelihood are introduced in Section 4 and motivated there as respecting the limited information principle.

3 Large Sample Properties

This section shows that under general regularity conditions the Quasi-posterior distribution concentrates at the speed $1/\sqrt{n}$ around the true parameter θ_0 as measured by the "total variation of moments" norm (and total variation norm as a special case), that the LT estimators are consistent and asymptotically normal, and that Quasi-posterior quantiles and other relevant quantities provide asymptotically valid confidence intervals.

3.1 Assumptions

We begin by stating the main assumptions. In addition, it is assumed without further notice that the criterion functions $L_n(\theta)$ and other primitive objects have the standard measurability properties. For example, given the underlying probability space $(\Omega, \mathcal{F}, P)$, for any $\omega \in \Omega$, $L_n(\theta)$ is a measurable function of θ, and for any $\theta \in \Theta$, $L_n(\theta)$ is a random variable, that is a measurable function of ω.

ASSUMPTION 1 (***Parameter***) *The true parameter θ_0 belongs to the interior of a compact convex subset Θ of Euclidean space $\mathbb{R}^d$.*

ASSUMPTION 2 (***Penalty Function***) *The loss function $\rho_n : \mathbb{R}^d \to \mathbb{R}_+$ satisfies:*

i. $\rho_n(u) = \rho(\sqrt{n}u)$, *where* $\rho(u) \geq 0$ *and* $\rho(u) = 0$ *iff* $u = 0$,

ii. ρ *is convex and* $\rho(h) \leq 1 + |h|^p$ *for some* $p \geq 1$,

iii. $\varphi(\xi) = \int_{\mathbb{R}^d} \rho(u - \xi) e^{-u' a u} du$ *is minimized uniquely at some* $\xi^* \in \mathbb{R}^d$ *for any finite* $a > 0$,

iv. *the weighting function* $\pi : \Theta \to \mathbb{R}_+$ *is a continuous, uniformly positive density function.*

[11] Kim (2002) also provided some useful asymptotic results for $\exp(L_n(\theta))$ using the shrinking neighborhood approach. However, Kim's (2002) approach does not cover the estimators and procedures considered here, see previous footnote.

ASSUMPTION 3 (***Identifiability***) *For any $\delta > 0$, there exists $\epsilon > 0$, such that*

$$\liminf_{n\to\infty} P_* \left\{ \sup_{|\theta-\theta_0|\geq\delta} \frac{1}{n}\Big(L_n(\theta) - L_n(\theta_0)\Big) \leq -\epsilon \right\} = 1.$$

ASSUMPTION 4 (***Expansion***) *For θ in an open neighborhood of θ_0,*

i. $L_n(\theta) - L_n(\theta_0) = (\theta-\theta_0)'\Delta_n(\theta_0) - \frac{1}{2}(\theta-\theta_0)'[nJ_n(\theta_0)](\theta-\theta_0) + R_n(\theta)$,

ii. $\Omega_n^{-1/2}(\theta_0)\Delta_n(\theta_0)/\sqrt{n} \to_d \mathcal{N}(0, I)$,

iii. $J_n(\theta_0) = O(1)$ *and* $\Omega_n(\theta_0) = O(1)$ *are uniformly in n positive-definite constant matrices,*

iv. for each $\epsilon > 0$ there is a sufficiently small $\delta > 0$ and large $M > 0$ such that

$$(a)\quad \limsup_{n\to\infty} P^* \left\{ \sup_{M/\sqrt{n}\leq|\theta-\theta_0|\leq\delta} \frac{|R_n(\theta)|}{n|\theta-\theta_0|^2} > \epsilon \right\} < \epsilon,$$

$$(b)\quad \limsup_{n\to\infty} P^* \left\{ \sup_{|\theta-\theta_0|\leq M/\sqrt{n}} |R_n(\theta)| > \epsilon \right\} = 0.$$

3.2 Discussion of Assumptions

In the following we discuss the stated assumptions under which Theorem 1-4 stated below will be true. We argue that these assumptions are simple but encompass a wide variety of econometric models – from cross-sectional models to nonlinear dynamic models. This means that Theorems 1-4 are of wide interest and applicability.

In general, Assumptions 1-4 are related to but different from those in Bickel and Yahav (1969) and Ibragimov and Has'minskii (1981). The most substantial differences appear in Assumption 4, and are due to the general non-likelihood setting. Also in Assumption 4 we introduce Huber type conditions to handle the tail behavior of discontinuous and non-smooth criterion functions. In general, the early approaches are inevitably tied to the iid likelihood formulation, which is not suited for the present purposes.

The compactness Assumption 1 is conventional. It is shown in the proof of Theorem 1 that it is not difficult to drop compactness. For example, in the case of Quasi-posterior quantiles in Theorem 3, it is only required that π is a proper density; in the case of Quasi-posterior variances in Theorem 4, it is only required that $\int_\Theta |\theta|^2\pi(\theta)d\theta < \infty$; and for the general loss functions considered in Theorem 2 it is required that $\int_\Theta |\theta|^p\pi(\theta)d\theta < \infty$. Of course, compactness guarantees all of the above given Assumption 2. Also, that the parameter is on the interior of the parameter space rules out some non-regular cases; see for example Andrews (1999).

Assumption 2 imposes convexity on the penalty function. We do not consider non-convex penalty functions for pragmatic reasons. One of the main motivations of this paper is the generic computability of the estimates, given that they solve well-defined convex optimization problems. The domination condition, $\rho(h) \leq 1 + |h|^p$ for some $1 \leq p < \infty$, is conventional and is satisfied in all examples of ρ we gave.

The assumption that $\varphi(\xi) = \int \rho(u - \xi)e^{-u'au}du \propto E\rho(\mathcal{N}(0, a^{-1}) - \xi)$ attains a unique minimum at some finite ξ^* for any positive definite a is required, and it clearly holds for all of examples of ρ we mentioned. In fact, when ρ is symmetric, $\xi^* = 0$ by Anderson (1955)'s lemma.

Assumption 3 is implied by the usual uniform convergence and unique identification conditions as in Amemiya (1985). The proof of Lemma 1 can be found in Amemiya (1985) and White (1994).

LEMMA 1 *Given Assumption 1, Assumption 3 holds if there is a function $M_n(\theta)$ that*

i. is nonstochastic, continuous on Θ, for any $\delta > 0$, $\limsup_n(\sup_{|\theta-\theta_0|>\delta} M_n(\theta) - M_n(\theta_0)) < 0$,

ii. $L_n(\theta)/n - M_n(\theta)$ converges to zero in (outer) probability uniformly over Θ.

Assumption 4 is satisfied under the conditions of Lemma 2, which are known to be mild in nonlinear models. Assumption 4.ii requires asymptotic normality to hold, and is generally a weak assumption for cross-sectional and many time-series applications. Assumption 4.iii rules out the cases of mixed asymptotic normality for some non-stationary time series models (which can be incorporated at a notational cost with different scaling rates).

Assumption 4.iv easily holds when there is enough smoothness.

LEMMA 2 *Given Assumptions 1 and 3, Assumption 4 holds with*

$$\Delta_n(\theta_0) = \nabla_\theta L_n(\theta_0) \text{ and } J_n(\theta_0) = -\nabla_{\theta\theta'} M_n(\theta_0) = O(1), \text{ if}$$

i. for some $\delta > 0$, $L_n(\theta)$ and $M_n(\theta)$ are twice continuously differentiable in θ when $|\theta - \theta_0| < \delta$,

ii. there is $\Omega_n(\theta_0)$ such that $\Omega_n^{-1/2}(\theta_0)\nabla_\theta L_n(\theta_0)/\sqrt{n} \to_d \mathcal{N}(0, I)$, $J_n(\theta_0) = O(1)$ and $\Omega_n(\theta_0) = O(1)$ are uniformly positive definite, and

iii. for some $\delta > 0$ and each $\epsilon > 0$

$$\limsup_{n\to\infty} P^* \left\{ \sup_{|\theta-\theta_0|<\delta} |\nabla_{\theta\theta'} L_n(\theta)/n - \nabla_{\theta\theta'} M_n(\theta)| > \epsilon \right\} = 0.$$

Lemma 2 is immediate, hence its proof is omitted. Both Lemmas 1 and 2 are simple but useful conditions that can be easily verified using standard uniform laws of large numbers and central limit theorems. In particular, they have been proven to hold for criterion functions corresponding to

1. Most smooth cross-sectional models described in Amemiya (1985);

2. The smooth nonlinear stationary and dynamic GMM and Quasi-likelihood models of Hansen (1982), Gallant and White (1988) and Pötscher and Prucha (1997), covering Gordin(mixingale type) conditions and near-epoch dependent processes such as ARMA, GARCH, ARCH, and other models alike;

3. General empirical likelihood models for smooth moment equation models studied by Imbens (1997), Kitamura and Stutzer (1997), Newey and Smith (2001), Owen (1989,1990,1991, 2001), Qin and Lawless (1994), and the recent extensions to the conditional moment equations.

Hence the main results of this paper, Theorems 1-4, apply to these fundamental econometric and statistical models. Moreover, Assumption 4 does not require differentiability of the criterion function and thus holds even more generally. Assumption **4.iv** is a Huber-like stochastic equicontinuity condition, which requires that the remainder term of the expansion can be controlled in a particular way over a neighborhood of θ_0. In addition to Lemma 2, many sufficient conditions for Assumption 4 are given in empirical process literature, e.g. Amemiya (1985), Andrews (1994a), Newey (1991), Pakes and Pollard (1989), and van der Vaart and Wellner (1996). Section 4 verifies Assumption 4 for the leading models with nonsmooth criterion functions, including the examples discussed in the previous section.

3.3 Convergence in the Total Variation of Moments Norm

Under Assumptions 1-4, we show that the Quasi-posterior density concentrates around θ_0 at the speed $1/\sqrt{n}$ as measured by the total variation of moments norm, and then use this preliminary result to prove all other main results.

Define the local parameter h as a normalized deviation from θ_0 and centered at the normalized random "score function":

$$h = \sqrt{n}\left(\theta - \theta_0\right) - J_n\left(\theta_0\right)^{-1} \Delta_n\left(\theta_0\right)/\sqrt{n}.$$

Define by the Jacobi rule the localized Quasi-posterior density for h as

$$p_n^*\left(h\right) = \frac{1}{\sqrt{n}} p_n\left(h/\sqrt{n} + \theta_0 + J_n\left(\theta_0\right)^{-1} \Delta_n\left(\theta_0\right)/n\right).$$

Define the total variation of moments norm for a real-valued measurable function f on S as

$$\|f\|_{TVM(\alpha)} \equiv \int_S (1 + |h|^\alpha)|f(h)|dh.$$

THEOREM 1 (*Convergence in Total Variation of Moments Norm*) *Under Assumptions 1 - 4, for any* $0 \le \alpha < \infty$,

$$\|p_n^*(h) - p_\infty^*(h)\|_{TVM(\alpha)} \equiv \int_{H_n} (1 + |h|^\alpha)|p_n^*(h) - p_\infty^*(h)|dh \to_p 0,$$

where $H_n = \{\sqrt{n}(\theta - \theta_0) - J_n(\theta_0)^{-1}\Delta_n(\theta_0)/\sqrt{n} : \theta \in \Theta\}$ *and*

$$p^*_\infty(h) = \sqrt{\tfrac{\det J_n(\theta_0)}{(2\pi)^d}} \cdot \exp\left(-\frac{1}{2}h' J_n(\theta_0) h\right).$$

Theorem 1 shows that $p_n(\theta)$ is concentrated at a $1/\sqrt{n}$ neighborhood of θ_0 as measured by the total variation of moments norm. For large n, $p_n(\theta)$ is approximately a random normal density with the random mean parameter $\theta_0 + J_n(\theta_0)^{-1}\Delta_n(\theta_0)/n$, and constant variance parameter $J_n(\theta_0)^{-1}/n$.

Theorem 1 applies to general statistical criterion functions $L_n(\theta)$, hence it covers the parametric likelihood setting as a fundamental case, in particular implying the Bernstein-Von Mises theorems, which state the convergence of the likelihood posterior to the limit random density in the total variation norm. Note also that the total variation norm results from setting $\alpha = 0$ in the total variation of moments norm. The use of the latter is needed to deduce the convergence of LTE's such as the posterior means or variances in Theorems 2-4.

3.4 Limit Results for Point Estimates and Confidence Intervals

As a consequence of Theorem 1, Theorem 2 establishes $\sqrt{n}$- consistency and asymptotic normality of LTE's. When the loss function $\rho(\cdot)$ is symmetric, LTE's are asymptotically equivalent to the extremum estimators.

Recall that the extremum estimator $\sqrt{n}(\widehat{\theta}_{ex} - \theta_0)$, where $\widehat{\theta}_{ex} = \arg\sup_{\theta\in\Theta} L_n(\theta)$, is first order-equivalent to

$$U_n \equiv \frac{1}{\sqrt{n}} J_n(\theta_0)^{-1}\Delta_n(\theta_0).$$

Given that the p^*_n approaches p^*_∞, it may be expected that the LTE $\sqrt{n}(\widehat{\theta} - \theta_0)$ is asymptotically equivalent to

$$Z_n = \arg\inf_{z\in\mathbb{R}^d}\left\{\int_{\mathbb{R}^d} \rho(z-u)\, p^*_\infty(u - U_n)\, du\right\}.$$

To see a relationship between Z_n and U_n, define

$$\xi_{J_n(\theta_0)} \equiv \arg\inf_{z\in\mathbb{R}^d}\left\{\int_{\mathbb{R}^d} \rho(z-u)\, p^*_\infty(u)\, du\right\},$$

which exists by Assumption 2.[12] If ρ is symmetric, i.e. $\rho(h) = \rho(-h)$, then by Anderson's lemma $\xi_{J_n(\theta_0)} = 0$. Hence

$$Z_n = \xi_{J_n(\theta_0)} + U_n,$$

and we are prepared to state the result.

[12] For example, in the scalar parameter case, if $\rho(h) = (\alpha - 1(h < 0))h$, the constant $\xi_{J(\theta_0)} = q_\alpha J_n(\theta_0)^{-1/2}$, where q_α is the α-quantile of $\mathcal{N}(0,1)$.

THEOREM 2 (***LTE in Large Samples***) *Under Assumptions 1-4,*

$$\sqrt{n}(\hat{\theta}-\theta_0) = \xi_{J_n(\theta_0)} + U_n + o_p(1), \quad \Omega_n^{-1/2}(\theta_0)J_n(\theta_0)U_n \to_d \mathcal{N}(0\ ,I).$$

Hence

$$\Omega_n^{-1/2}(\theta_0)J_n(\theta_0)\left(\sqrt{n}(\hat{\theta}-\theta_0) - \xi_{J_n(\theta_0)}\right) \to_d \mathcal{N}(0,I).$$

If loss function ρ_n *is symmetric, i.e.* $\rho_n(h) = \rho_n(-h)$ *for all* h, $\xi_{J_n(\theta_0)} = 0$ *for each* n.

In order for the Quasi-posterior distribution to provide valid large sample confidence intervals, the density of $W_n = \mathcal{N}\left(0, J_n(\theta_0)^{-1}\Omega_n(\theta_0)J_n(\theta_0)^{-1}\right)$ should coincide with that of $p_\infty^*(h)$. This requires

$$\int_{\mathbb{R}^d} hh'p_\infty^*(h)dh \equiv J_n(\theta_0)^{-1} \sim \mathrm{Var}(W_n) \equiv J_n(\theta_0)^{-1}\Omega_n(\theta_0)J_n(\theta_0)^{-1},$$

or equivalently

$$\Omega_n(\theta_0) \sim J_n(\theta_0),$$

which is a **generalized information equality.** The information equality is known to hold for regular, correctly specified likelihoods. It is also known to hold for appropriately constructed criterion functions of generalized method of moments, minimum distance estimators, generalized empirical likelihood estimators, and properly weighted extremum estimators; see Section 4.

Consider construction of the confidence intervals for the quantity $g(\theta_0)$, and suppose g is continuously differentiable. Define

$$F_{g,n}(x) = \int_{\theta\in\Theta: g(\theta)\leq x} p_n(\theta)d\theta, \quad \text{and} \quad c_{g,n}(\alpha) = \inf\{x : F_{g,n}(x) \geq \alpha\}.$$

Then a LT confidence interval is given by $[c_{g,n}(\alpha/2), c_{g,n}(1-\alpha/2)]$. As previously mentioned, these confidence intervals can be constructed by using the $\alpha/2$ and $1-\alpha/2$ quantiles of the MCMC sequence

$$\left(g(\theta^{(1)}), ..., g(\theta^{(B)})\right)$$

and thus are quite simple in practice. In order for the intervals to be valid in large samples, one needs to ensure the generalized information equality, which can be done easily through the use of optimal weighting in GMM and minimum-distance criterion functions or the use of generalized empirical likelihood functions; see Section 4.

Consider now the usual asymptotic intervals based on the Δ- method and any estimator with the property

$$\sqrt{n}(\widehat{\theta}-\theta) = J_n(\theta_0)^{-1}\Delta_n(\theta)/\sqrt{n} + o_p(1).$$

Such intervals are usually given by

$$\left[g(\widehat{\theta}) + q_{\alpha/2}\frac{\sqrt{\nabla_\theta g(\widehat{\theta})' J_n(\theta_0)^{-1}\nabla_\theta g(\widehat{\theta})}}{\sqrt{n}}, \quad g(\widehat{\theta}) + q_{1-\alpha/2}\frac{\sqrt{\nabla_\theta g(\widehat{\theta})' J_n(\theta_0)^{-1}\nabla_\theta g(\widehat{\theta})}}{\sqrt{n}}\right],$$

where q_α is the α-quantile of the standard normal distribution. The following theorem establishes the large sample correspondence of the Quasi-posterior confidence intervals to the above intervals.

THEOREM 3 (*Large Sample Inference I*) *Suppose Assumptions 1-4 hold. In addition suppose that the generalized information equality holds:*

$$\lim_{n\to\infty} J_n(\theta_0)\Omega_n(\theta_0)^{-1} = I.$$

Then for any $\alpha \in (0,1)$

$$c_{g,n}(\alpha) - g(\widehat{\theta}) - q_\alpha \frac{\sqrt{\nabla_\theta \theta(\theta_0)' J_n(\theta_0)^{-1} \nabla_\theta g(\theta_0)}}{\sqrt{n}} = o_p\left(\frac{1}{\sqrt{n}}\right),$$

and

$$\lim_{n\to\infty} P^*\Big\{c_{g,n}(\alpha/2) \le g(\theta_0) \le c_{g,n}(1-\alpha/2)\Big\} = 1-\alpha.$$

One practical limitation of this result arises in the case of regression criterion functions (M-estimators), where achieving the information equality may require nonparametric estimation of appropriate weights, e.g. as in censored quantile regression discussed in Section 4. This may entirely be avoided by using a different method for construction of confidence intervals. Instead of the Quasi-posterior quantiles, we can use the Quasi-posterior variance as an estimate of the inverse of the population Hessian matrix $J_n^{-1}(\theta_0)$, and combine it with any available estimate of $\Omega_n(\theta_0)$ (which typically is easier to obtain) in order to obtain the Δ-method style intervals. The usefulness of this methods is particularly evident in the censored quantile regression, where direct estimation of $J_n(\theta_0)$ requires use of nonparametric methods.

THEOREM 4 (*Large Sample Inference II*) *Suppose Assumptions 1-4 hold. Define for* $\widehat{\theta} = \int_\Theta \theta p_n(\theta) d\theta$,

$$\widehat{J}_n^{-1}(\theta_0) \equiv \int_\Theta n(\theta - \widehat{\theta})(\theta - \widehat{\theta})' p_n(\theta) d\theta,$$

and

$$c_{g,n}(\alpha) \equiv g(\widehat{\theta}) + q_\alpha \cdot \frac{\sqrt{\nabla_\theta g(\widehat{\theta})' \widehat{J}_n(\theta_0)^{-1} \widehat{\Omega}_n(\theta_0) \widehat{J}_n(\theta_0)^{-1} \nabla_\theta g(\widehat{\theta})}}{\sqrt{n}},$$

where $\widehat{\Omega}_n(\theta_0)\Omega_n^{-1}(\theta_0) \to_p I$. *Then* $\widehat{J}_n(\theta_0) J_n(\theta_0)^{-1} \to_p I$, *and*

$$\lim_{n\to\infty} P^*\Big\{c_{g,n}(\alpha/2) \le g(\theta_0) \le c_{g,n}(1-\alpha/2)\Big\} = 1-\alpha.$$

In practice $\widehat{J}_n(\theta_0)^{-1}$ is computed by multiplying by n the variance-covariance matrix of the MCMC sequence $S = \big(\theta^{(1)}, \theta^{(2)}, ..., \theta^{(B)}\big)$.

4 Applications to Selected Problems

This section further elaborates the approach through several examples. Assumptions 1-4 cover a wide variety of smooth econometric models (by virtue of Lemma 1 and Lemma 2). Thus, what follows next is mainly motivated by models with non-smooth moment equations, such as those occurring in Examples 1 - 3. Verification of the key Assumption 4 is not immediate in these examples, and Propositions 1-3 and the forthcoming examples show how to do this in a class of models that are of prime interest to us.

4.1 Generalized Method of Moments and Nonlinear Instrumental Variables

Going back to Example 2, recall that a typical model that underlies the applications of GMM is a set of population moment equations:

$$Em_i(\theta) = 0 \qquad \text{if and only if} \qquad \theta = \theta_0. \tag{4.1}$$

Method of moment estimators involve maximizing an objective function of the form

$$\begin{aligned}
L_n(\theta) &= -n\,(g_n(\theta))'\,W_n(\theta)\,(g_n(\theta))\,/2, && (4.2)\\
g_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n} m_i(\theta), && (4.3)\\
W_n(\theta) &= W(\theta) + o_p(1) \ \text{ uniformly in } \theta \in \Theta, && (4.4)\\
W(\theta) &> 0 \text{ and continuous uniformly in } \theta \in \Theta, && (4.5)\\
W(\theta_0) &= \left[\lim_{n\to\infty} \mathrm{Var}[\sqrt{n}g_n(\theta_0)]\right]^{-1}. && (4.6)
\end{aligned}$$

The choice (4.6) of the weighting matrix implies the generalized information equality under standard regularity conditions.

Generally, by a Central Limit Theorem $\sqrt{n}g_n(\theta_0) \to_d \mathcal{N}(0, W^{-1}(\theta_0))$, so that the objective function can be interpreted as the approximate log-likelihood for the sample moments of the data $g_n(\theta)$. Thus we can think of GMM as an approach that specifies an approximate likelihood for selected moments of the data without specifying the likelihood of the entire data.[13]

We may impose Assumptions 1-4 directly on the GMM objective function. However, to highlight the plausibility and elaborate on some examples that satisfy Assumption 4 consider the following proposition.

Proposition 1 (***Method-of-Moments and Nonlinear IV***) *Suppose that Assumptions 1-2 hold, and that for all θ in Θ, $m_i(\theta)$ is stationary and ergodic, and*

i. *conditions (4.1)-(4.5) hold,*

ii. $J(\theta) \equiv G(\theta)'W(\theta)G(\theta) > 0$ *and is continuous,* $G(\theta) = \nabla_\theta Em_i(\theta)$ *is continuous,*

iii. $\Delta_n(\theta_0)/\sqrt{n} = -\sqrt{n}g_n(\theta_0)'W(\theta_0)G(\theta_0) \to_d \mathcal{N}(0, \Omega(\theta_0))$, $\Omega(\theta_0) \equiv G(\theta_0)'W(\theta_0)G(\theta_0)$,

iv. *for any $\epsilon > 0$, there is $\delta > 0$ such that*

$$\limsup_{n\to\infty} P^* \left\{ \sup_{|\theta-\theta'|\leq\delta} \frac{\sqrt{n}|\,(g_n(\theta) - g_n(\theta')) - (Eg_n(\theta) - Eg_n(\theta'))\,|}{1 + \sqrt{n}|\theta - \theta'|} > \epsilon \right\} < \epsilon. \tag{4.7}$$

Then Assumption 4 holds. In addition the information equality holds by construction. Therefore the conclusions of Theorems 1-4 hold with $\Delta_n(\theta_0)$, $\Omega_n(\theta_0) \equiv \Omega(\theta_0)$ and $J_n(\theta_0) \equiv J(\theta_0)$ defined above, where the condition (4.6) is only needed for the conclusions of Theorem 3 to hold.

[13] This does not help much in terms of providing formal asymptotic results for the GMM model.

Therefore, for symmetric loss functions ρ_n, the LTE is asymptotically equivalent to the GMM extremum estimator. Furthermore, the generalized information equality holds by construction, hence Quasi-posterior quantiles provide a computationally attractive method of "inverting" the objective function for the confidence intervals.

For twice continuously differentiable smooth moment conditions, the smoothness conditions on $\nabla_\theta L_n(\theta)$ and $\nabla_{\theta\theta'} L_n(\theta)$ stated in Lemma 2 trivially imply condition **iv** in Proposition 1. More generally, Andrews (1994a), Pakes and Pollard (1989) and van der Vaart and Wellner (1996) provide many methods to verify that condition in a wide variety of method-of-moments models.

Example 3 Continued. Instrumental median regression falls outside of both the classical Bayesian approach and the classical smooth nonlinear IV approach of Amemiya (1977). Yet the conditions of Proposition 1 are satisfied under mild conditions:

i. (Y_i, D_i, X_i, Z_i) is an iid data sequence, $E[m_i(\theta_0)Z_i] = 0$, and θ_0 is identifiable,

ii. $\{m_i(\theta) = (\tau - 1(Y_i \leq q(D_i, X_i, \theta)))\, Z_i, \theta \in \Theta\}$ is a Donsker class,[14] $E \sup_\theta |m_i(\theta)|^2 < \infty$,

iii. $G(\theta) = \nabla_\theta E m_i(\theta) = -E f_{Y|D,X,Z}(q(D,X,\theta)) Z \nabla_\theta q(D,X,\theta)'$ is continuous,

iv. $J(\theta) = G(\theta)'W(\theta)G(\theta) > 0$ and is continuous in an open ball at θ_0.

In this case the weighting matrix can be taken as

$$W_n(\theta) = \frac{1}{\tau(1-\tau)} \left[\frac{1}{n} \sum_{i=1}^{n} Z_i Z_i' \right]^{-1},$$

so that the information equality holds. Indeed, in this case

$$\Omega(\theta_0) = G(\theta_0)'W(\theta_0)G(\theta_0) = J(\theta_0),$$

where

$$W(\theta_0) = \text{plim } W_n(\theta_0) = [\text{Var } m_i(\theta_0)]^{-1}, \quad \text{Var } m_i(\theta_0) = \tau(1-\tau) E Z_i Z_i'.$$

When the model q is linear and the dimension of D is small, there are computable and practical estimators in the literature.[15] In more general models, the extremum estimates are quite difficult to compute, and the inference faces the well-known difficulty of estimating sparsity parameters.

On the other hand, the Quasi-posterior median and quantiles are easy to compute and provide asymptotically valid confidence intervals. Note that the inference does not require the estimation of

[14]This is a very weak restriction on the function class, and is known to hold for all practically relevant functional forms, see van der Vaart (1999).

[15]These include e.g. the "inverse" quantile regression approach in Chernozhukov and Hansen (2001), which is an extension of Koenker and Bassett (1978)'s quantile regression to endogenous settings.

the density function. The simulation example given in Section 5 strongly supports this alternative approach.

Another important example which poses computational challenge is the estimation problem of Berry et al. (1995). This example is similar in nature to the instrumental quantile regression, and the application of the LT methods may be fruitful there.

4.2 Generalized Empirical Likelihood

A class of objective functions that are first-order equivalent to optimally weighted GMM (after recentering) can be formulated using the generalized empirical likelihood framework.

A class of generalized empirical likelihood functions(GEL) are studied in Imbens et al. (1998), Kitamura and Stutzer (1997), and Newey and Smith (2001). For a set of moment equations $Em_i(\theta_0) = 0$ that satisfy the conditions of section 4.1, define

$$\bar{L}_n(\theta,\gamma) \equiv \sum_{i=1}^{n} \left(s\left(m_i(\theta)'\gamma\right) - s(0)\right). \tag{4.8}$$

Then set

$$L_n(\theta) = \bar{L}_n(\theta, \hat{\gamma}(\theta)), \tag{4.9}$$

where $\widehat{\gamma}(\theta)$ solves

$$\widehat{\gamma}(\theta) \equiv \arg \inf_{\gamma \in \mathbb{R}^p} \bar{L}_n(\theta,\gamma), \tag{4.10}$$

and $p = \dim(m_i)$.

The scalar function $s(\cdot)$ is a strictly convex, finite, and three times differentiable function on an open interval of $\mathbb{R}$ containing 0, denoted $\mathcal{V}$, and is equal to $+\infty$ outside such an interval. $s(\cdot)$ is normalized so that both $\nabla s(0) = 1$ and $\nabla^2 s(0) = 1$. The choices of the function $s(v) = -\ln(1-v)$, $\exp(v)$, and $(1+v)^2/2$ lead to the well-known empirical likelihood, exponential tilting, and continuous-updating GMM criterion functions.

Simple and practical sufficient conditions for Lemma 2 are given in Qin and Lawless (1994), Imbens et al. (1998), Kitamura (1997), Kitamura and Stutzer (1997), including stationary weakly dependent data, Newey and Smith (2001), and Christoffersen et al. (1999). Thus, the application of LTE's to these problems is immediate.

To illustrate a further use of LTE's we state a set of simple conditions geared towards non-smooth microeconometric applications such as the instrumental quantile regression problem. These regularity conditions imply the first order equivalence of the GEL and GMM objective functions. The Donskerness condition below is a weak assumption that is known to hold for all reasonable linear and nonlinear functional forms encountered in practice, as discussed in van der Vaart (1999).

Proposition 2 (***Empirical Likelihood Problems***) *Suppose that Assumptions 1- 2 hold, and that the following conditions are satisfied: for some $\delta > 0$ and all $\theta \in \Theta$*

i. condition (4.1) holds and that $m_i(\theta)$ is iid,

*ii. $\partial P[m_i(\theta) < x]/\partial\theta$ is continuous in θ uniformly in $x : |x| \leq K$, for K in **iii**.*

iii. $\sup_{|\theta-\theta_0|<\delta} |m_i(\theta)| < K$ a.s., for some constant K

iv. $\{m_i(\theta), \theta \in \Theta\}$ is Donsker class, where

$$\sqrt{n}g_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_i(\theta_0) \to_d \mathcal{N}(0, V(\theta_0)), \quad V(\theta_0) = E\{m_i(\theta_0)m_i(\theta_0)'\} > 0,$$

then Assumptions 3 and 4 hold, and thus the conclusions of Theorems 1 - 4 are true with

$$\begin{aligned}
&\Delta_n(\theta_0)/\sqrt{n} = \sqrt{n}g_n(\theta_0)V(\theta_0)^{-1}G(\theta_0) \to_d \mathcal{N}(0, \Omega(\theta_0)),\\
&\Omega(\theta_0) = G(\theta_0)'V(\theta_0)^{-1}G(\theta_0),\\
&J(\theta_0) = G(\theta_0)'V(\theta_0)^{-1}G(\theta_0), \quad G(\theta_0) = \nabla_\theta E m_i(\theta_0).
\end{aligned}$$

The information equality holds in this case.

Another (equivalent) way to proceed is through the dual formulation. Consider the following criterion function

$$L_n(\theta) = \sup_{\pi_1,\ldots,\pi_n \in [0,1]} \sum_{i=1}^{n} h(\pi_i) \text{ subject to } \sum_{i=1}^{n} m_i(\theta)\pi_i = 0, \sum_{i=1}^{n} \pi_i = 1, \tag{4.11}$$

where h is the Cressie-Reid divergence criterion, cf. Newey and Smith (2001)

$$h(\pi) = \frac{1}{\gamma(\gamma+1)} \frac{(\frac{\pi}{1/n})^{\gamma+1} - 1}{n}.$$

The function $L_n(\theta)$ in (4.11) is the generalized empirical likelihood function for θ with the concentrated out probabilities. In fact, (4.11) corresponds to (4.9) by the argument given in Qin and Lawless (1994) p.303-304, so that Proposition 2 covers (4.11) as a special case up to renormalization.

Empirical probabilities $\widehat{\pi}_i(\theta)$'s are obtained in (4.11) using the extremum method. The case $\gamma = -1$ yields the empirical likelihood case, where $\widehat{\pi}_i(\theta)$'s are obtained through the maximum likelihood method. Taking $\gamma = 0$ yields the exponential tilting case, where $\widehat{\pi}_i(\theta)$'s are obtained through minimization of the Kullback-Leibler distance from the empirical distribution. Taking $\gamma = 1$ yields the continuous-updating case, where $\widehat{\pi}_i(\theta)$'s are obtained through the minimization of the Euclidean distance from the empirical distribution. Each approach generates the implied probabilities $\widehat{\pi}_i(\theta)$ given θ. Qin and Lawless (1994) and Newey and Smith (2001) provide the formulas:

$$\widehat{\pi}_i(\theta) = \nabla s(\widehat{\gamma}(\theta)' m_i(\theta)) / \sum_{i=1}^{n} \nabla s(\widehat{\gamma}(\theta)' m_i(\theta)).$$

The Quasi-posterior for θ and $\widehat{\pi}_i(\theta)$ can be used for predictive inference. Suppose $m_i(\theta) = m(X_i, \theta)$ for some random vector X_i. Then the Quasi-posterior predictive probability is given by

$$\widehat{P}\{X_i \in A\} = \int \underbrace{\sum_{i=1}^{n} \widehat{\pi}_i(\theta) 1\{X_i \in A\}}_{\equiv h_n(\theta)} p_n(\theta) d\theta = \int h_n(\theta) p_n(\theta) d\theta,$$

which can be computed by averaging over the MCMC sequence evaluated at h_n, $\left(h_n(\theta^{(1)}), ..., h_n(\theta^{(B)})\right)$. It follows similarly to the proof of Theorem 1 in Qin and Lawless (1994) that $\sqrt{n}(\widehat{P}\{X_i \in A\} - P\{X_i \in A\}) \rightarrow_d \mathcal{N}(0, \Pi_A)$, where $\Pi_A \equiv P\{X_i \in A\}(1 - P\{X_i \in A\}) - Em_i(\theta_0)' 1\{X_i \in A\} \cdot U \cdot Em_i(\theta_0) 1\{X_i \in A\}$, $U = V(\theta_0)^{-1}\{I - G(\theta_0) J(\theta_0)^{-1} G(\theta_0) V(\theta_0)^{-1}\}$.

4.3 M-estimation

M-estimators, which include many linear and nonlinear regressions as special cases, typically maximize objective functions of the form

$$L_n(\theta) = \sum_{i=1}^{n} m_i(\theta).$$

$m_i(\theta)$ need not be the log likelihood function of observation i, and may depend on preliminary non-parametric estimation. Assumptions 1–3 usually are satisfied by uniform laws of large numbers and by unique identification of the parameter; see for example Amemiya (1985) and Newey and McFadden (1994). The next proposition gives a simple set of sufficient conditions for Assumption 4.

Proposition 3 (***M-problems***) *Suppose Assumptions 1-3 hold for the criterion function specified above with the following additional conditions: Uniformly in θ in an open neighborhood of θ_0, $m_i(\theta)$ is stationary and ergodic, and for $\overline{m}_n(\theta) = \sum_{i=1}^{n} m_i(\theta)/n$,*

i. there exists $\dot{m}_i(\theta_0)$ such that $E\dot{m}_i(\theta_0) = 0$ for each i and, for some $\delta > 0$,

$$\left\{ \frac{m_i(\theta) - m_i(\theta_0) - \dot{m}_i(\theta_0)'(\theta - \theta_0)}{|\theta - \theta_0|}, \ \theta : |\theta - \theta_0| < \delta \right\} \text{ is a Donsker class,}$$

$$E[\overline{m}_n(\theta) - \overline{m}_n(\theta_0) - \overline{\dot{m}}_n(\theta_0)'(\theta - \theta_0)]^2 = o(|\theta - \theta_0|^2),$$

$$\Delta_n(\theta_0)/\sqrt{n} = \sum_{i=1}^{n} \dot{m}_i(\theta_0)/\sqrt{n} \rightarrow_d \mathcal{N}(0, \Omega(\theta_0)).$$

ii. $J(\theta) = -\nabla_{\theta\theta'} E[m_i(\theta)]$ is continuous and nonsingular in a ball at θ_0.

Then Assumption 4 holds. Therefore, the conclusions of Theorems 1, 2, and 4 hold. If in addition $J(\theta_0) = \Omega(\theta_0)$, then the conclusions of Theorem 3 also hold.

The above conditions apply to many well known examples such as LAD, see for example van der Vaart and Wellner (1996). Therefore, for many nonlinear regressions, Quasi-posterior means, modes, and medians are asymptotically equivalent, and Quasi-posterior quantiles provide asymptotically valid confidence statements if the generalized information equality holds. When the information equality fails to hold, the method of Theorem 4 provides valid confidence intervals.

Example 1 Continued. Under the conditions given in Powell (1984) or Newey and Powell (1990) for the censored quantile regression, the assumptions of Proposition 3 are satisfied. Furthermore, it is not difficult to show that when the weights ω_i^* are nonparametrically estimated, the conditions of Newey and Powell (1990) imply Assumption 4. Under iid sampling, the use of efficient weighting

$$\omega_i^* = \frac{1}{\tau(1-\tau)} f_i, \tag{4.12}$$

where $f_i = f_{Y_i|X_i}(q_i)$, $q_i = q(X_i;\theta_0)$, validates the generalized information equality, and the Quasi-posterior quantiles form asymptotically valid confidence intervals. Indeed, since

$$J(\theta_0) = \frac{1}{\tau(1-\tau)} E f_i^2 \nabla q_i \nabla q_i', \text{ for } \nabla q_i = \partial q(X_i,\theta_0)/\partial\theta, \tag{4.13}$$

and

$$\frac{1}{\sqrt{n}}\Delta_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \cdot (\tau - 1(Y_i \le q_i))\nabla q_i \to_d \mathcal{N}(0,\Omega(\theta_0)), \tag{4.14}$$

with

$$\Omega(\theta_0) = \frac{1}{\tau(1-\tau)} E f_i^2 \nabla q_i \nabla q_i', \tag{4.15}$$

we have

$$\Omega(\theta_0) = J(\theta_0). \tag{4.16}$$

For this class of problems, the Quasi-posterior means and medians are asymptotically equivalent to the extremum estimators. The Quasi-posterior quantiles provide asymptotically valid confidence intervals when the efficient weights are used. However, estimation of efficient weights requires preliminary estimation of parameter θ_0. When other weights are used, the method of Theorem 4 provides valid confidence intervals.

5 Computation and Simulation Examples

In this section we briefly discuss the MCMC method and present simulation examples.

5.1 Markov Chain Monte Carlo

The Quasi-posterior density is proportional to

$$p_n(\theta) \propto e^{L_n(\theta)}\pi(\theta).$$

In most cases we can easily compute $e^{L_n(\theta)}\pi(\theta)$. However, computation of the point estimates and confidence intervals typically requires evaluation of integrals like

$$\frac{\int_\Theta g(\theta)e^{L_n(\theta)}\pi(\theta)d\theta}{\int_\Theta e^{L_n(\theta)}\pi(\theta)d\theta} \tag{5.1}$$

for various functions g. For problems for which no analytic solution exists for (5.1), especially in high dimensions, MCMC methods provide powerful tools for evaluating integrals like the one above. See for example Chib (2001), Geweke and Keane (2001), and Robert and Casella (1999) for excellent treatments.

MCMC is a collection of methods that produce an ergodic Markov chain with the stationary distribution p_n. Given a starting value $\theta^{(0)}$, a chain $(\theta^{(t)}, 1 \leq t \leq B)$ is generated using a transition kernel with stationary distribution p_n, which ensures the convergence of the marginal distribution of $\theta^{(B)}$ to p_n. For sufficiently large B, the MCMC methods produce a dependent sample $(\theta^{(0)}, \theta^{(1)}, ..., \theta^{(B)})$ whose empirical distribution approaches p_n. The ergodicity and construction of the chains usually imply that as $B \to \infty$,

$$\frac{1}{B}\sum_{t=1}^{B} g(\theta^{(t)}) \to_p \int_\Theta g(\theta)p_n(\theta)d\theta.$$

We stress that this technique does not rely on the likelihood principle and can be fruitfully used for computation of LTE's. (Appendix B provides the formal details.)

One of the most important MCMC methods is the Metropolis-Hastings algorithm.

Metropolis-Hastings (MH) algorithm with Quasi-Posteriors. Given the Quasi-posterior density $p_n(\theta) \propto e^{L_n(\theta)}\pi(\theta)$, known up to a constant, and a prespecified conditional density $q(\theta'|\theta)$, generate $\left(\theta^{(0)}, ..., \theta^{(B)}\right)$ in the following way,

1. Choose a starting value $\theta^{(0)}$.
2. Generate ξ from $q(\theta^{(j)}|\xi)$.
3. Update $\theta^{(j+1)}$ from $\theta^{(j)}$ for $j = 1, 2, ...,$ using

$$\theta^{(j+1)} = \begin{cases} \xi & \text{with probability} \quad \rho(\theta^{(j)}, \xi) \\ \theta^{(j)} & \text{with probability} \quad 1 - \rho(\theta^{(j)}, \xi) \end{cases},$$

where

$$\rho(x, y) = \inf\left(\frac{e^{L_n(y)}\pi(y)q(x|y)}{e^{L_n(x)}\pi(x)q(y|x)}, 1\right).$$

Note that the most important quantity in the algorithm is the probability $\rho(x, y)$ of the move from an "old" point x to the "new" point y, which depends on how much of an improvement in $e^{L_n(y)}\pi(y)$ a possible "new" value of y yields relative to $e^{L_n(x)}\pi(x)$ at the "old" value x. Thus, the generated chain of draws spends a relatively high proportion of time in the higher density regions and a lower proportion in the lower density regions. Because such proportions of times are balanced in the right way, the generated sequence of parameter draws has the requisite marginal distribution, which we then use for computation of means, medians, and quantiles. (How closely the sequence travels near the mode is not relevant.)

Another important choice is the transition kernel q, also called the instrumental density. It turns out that a wide variety of kernels yield Markov chains that converge to the distribution of interest. One canonical implementation of the MH algorithm is to take

$$q(x|y) = f(|x - y|),$$

where f is a density symmetric around 0, such as the Gaussian or the Cauchy density. This implies that the chain $(\theta^{(j)})$ is a random walk. This is the implementation we used in this paper. Chib (2001), Geweke and Keane (2001) and Robert and Casella (1999) can be consulted for important details concerning the implementation and convergence monitoring of the algorithm.

It is now worth repeating that the main motivation behind the LTE approach is based on its efficiency properties (stated in sections 3 and 4) as well as computational attractiveness. Indeed, the LTE approach is as efficient as the extremum approach, but may avoid the computational curse of dimensionality through the use of MCMC. LTE's are typically means or quantiles of a Quasi-posterior distribution, hence can be computed (estimated) at the parametric rate $1/\sqrt{B}$,[16] where B is the number of MCMC draws (functional evaluations). Indeed, under canonical implementations, the MCMC chains are geometrically mixing, so the rates of convergence are the same as under independent sampling. In contrast, the extremum estimator (mode) is computed (estimated) by the MCMC and similar grid-based algorithms at the nonparametric rate $(1/B)^{\frac{p}{d+2p}}$, where d is the parameter dimension and p is the smoothness order of the objective function.

We used an optimistic tone regarding the performance of MCMC. Indeed, in the problems we study, the objective functions have numerous local optima, but all exhibit a well pronounced global optimum. These problems are important, and therefore the good performance of MCMC and the derived estimators are encouraging. However, various pathological cases can be constructed, see Robert and Casella (1999). Functions may have multiple separated global modes (or approximate modes), in which case MCMC may require extended time for convergence. Another potential problem is that the initial draw $\theta^{(0)}$ may be very far in the tails of the posterior $p_n(\theta)$. In this case, MCMC may also take extended time to converge to the stationary distribution. In the problems we looked at, this may be avoided by choosing a starting value based on economic considerations or other simple considerations. For example, in the censored median regression example, we may use the starting values based on an initial Tobit regression. In the instrumental median regression, we may use the two stage least squares estimates as the starting values.

[16]Note that the rates are used for the informal motivation. We fix d in the discussion, but the rate may typically increase linearly or polynomially in d if d is allowed to grow.

5.2 Monte Carlo Example 1: Censored Median Regression

As discussed in Section 2, a large literature has been devoted to the computation of Powell's censored median regression estimator. In the simulation example reported below, we find that both in small and large samples with high degree censoring, the LT estimation may be a useful alternative to the popular iterated linear programming algorithm of Buchinsky (1991). The model we consider is

$$Y^* = \beta_0 + X'\beta + u,$$
$$X \overset{d}{=} \mathcal{N}(0, I_3), \quad u = X_1^2 \mathcal{N}(0,1), \quad Y = \max(0, Y^*).$$

The true parameter $(\beta_0, \beta_1, \beta_2, \beta_3)$ is $(-6, 3, 3, 3)$, which produces about 40% censoring.

The LTE is based on the Powell's objective function $L_n(\theta) = -\sum_{i=1}^n |Y_i - \max(0, \beta_0 + X_i'\beta)|$. The initial draw of the MCMC series is taken to be the ordinary least squares estimate, and other details are summarized in Appendix B.

Table 1 reports the results. The number in parentheses in the iterated linear programming (ILP) results indicates the number of times that this algorithm converges to a local minimum of 0. The first row for the ILP reports the performance of the algorithm among the subset of simulations for which the algorithm does not converge to the local minimum at 0. The second row reports the results for all simulation runs, including those for which the ILP algorithm does not move away from the local minimum. The LTE's (Quasi-posterior mean and median) never converge to the local minimum of 0, and they compare favorably to the ILP even when the local minima are excluded from the ILP results, as can be seen from Table 1. When the local minima are included in the ILP results, LTE's do markedly better.

[Table 1 goes here.]

5.3 Monte Carlo Example 2: Instrumental Quantile Regression

We consider a simulation example similar to that in Koenker (1994). The model is

$$Y = \alpha_0 + D'\beta_0 + u, \quad u = \sigma(D)\epsilon,$$
$$D = \exp \mathcal{N}(0, I_3), \quad \epsilon = \mathcal{N}(0,1), \quad \sigma(D) = (1 + \sum_{i=1}^{3} D_{(i)})/5.$$

The true parameter (α_0, β_0) equals 0, and we consider the instrumental moment conditions

$$g_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2} - 1(Y_i \le \alpha + D'\beta))Z, \quad Z = (1, D),$$
$$W_n(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2} - 1(Y_i \le \alpha + D'\beta))^2 Z_i Z_i'\right]^{-1}.$$

In simulations, the initial draw of the MCMC series is taken to be the ordinary least squares estimate, and other details are summarized in Appendix B.

While instrumental median regression is designed specifically for endogenous or nonlinear models, we use a classical exogenous example in order to provide a contrast with a clear undisputed benchmark – the standard linear quantile regression. The benchmark provides a reliable and high-quality estimation method for the exogenous model. In this regard, the performance of the LT estimation and inference, reported in Table 2 and Table 3, is encouraging.

Table 2 summarizes the performance of LTE's and the standard quantile regression estimator. Table 3 compares the performance of the LT confidence intervals to the standard inference method for quantile regression implemented in S-plus 4.0. The reported results are averaged across the slope parameters. The root mean square errors of the LTE's are no larger than those of quantile regression. Other criteria demonstrate similar performance of two methods, as predicted by the asymptotic theory. The coverage of Quasi-posterior quantile confidence intervals is also close to the nominal level of 90% in both small and large samples. It is also noteworthy that the intervals do not require nonparametric density estimation, as the standard method requires.

[Tables 2 and 3 go here.]

6 An Illustrative Empirical Application

The following illustrates the use of LT estimation in practice. We consider the problem of forecasting the conditional quantiles or value-at-risk (VaR) of the Occidental Petroleum (NYSE:OXY) security returns. The problem of forecasting quantiles of return distributions is not only important for economic analysis, but is fundamental to the real-life activities of financial firms. We offer an econometric analysis of a dynamic conditional quantile forecasting model, and show that the LTE approach provides a simple and effective method of estimating such models (despite the difficulties inherent in the estimation).

The dataset consists of 2527 daily observations (September, 1986 – November, 1998) on

Y_t, the one-day returns of the Occidental Petroleum (NYSE:OXY) security,

X_t, a vector of returns and prices of other securities that affect the distribution of Y_t: a constant, lagged one-day return of Dow Jones Industrials (DJI), the lagged return on the spot price of oil (NCL, front-month contract on crude oil on NYMEX), and the lagged return Y_{t-1}.

The choice of variables follows a general principle in which the relevant conditioning information for estimating value-at-risk of a stock return, X_t, may contain such variables as a market index of corresponding capitalization and type (for instance, the $S\&P500$ returns for a large-cap value stock), the industry index, a price of a commodity or some other traded risk that the firm is exposed to, and lagged values of its stock price.

Two functional forms of predictive τ-th quantile regressions were estimated:

$$\begin{array}{ll} \text{Linear Model :} & Q_{Y_{t+1}}(\tau|I_t,\theta(\tau)) = X_t'\,\theta(\tau), \\ \text{Dynamic Model:} & Q_{Y_{t+1}}(\tau|I_t,\theta(\tau),\varrho(\tau)) = X_t'\theta(\tau) + \varrho(\tau)\cdot Q_{Y_t}(\tau|I_{t-1},\theta(\tau),\varrho(\tau)), \end{array}$$

where $Q_{Y_{t+1}}(\tau|I_t,\theta(\tau))$ denotes the τ-th conditional quantile of Y_{t+1} conditional on the information I_t available at time t. In other words, $Q_{Y_{t+1}}(\tau|I_t,\theta(\tau))$ is the value-at-risk at the probability level τ. The idea behind the dynamic models is to better incorporate the entire past information and better predict risk clustering, as introduced by Engle and Manganelli (2001). The nonlinear dynamic models described by Engle and Manganelli (2001) are appealing, but appear to be difficult to estimate using conventional extremum methods, see Engle and Manganelli (2001) for discussion. An extended empirical analysis of the linear model is given in Chernozhukov and Umantsev (2001).

The LT estimation and inference strategy is based on the Koenker and Bassett (1978) criterion function,

$$L_n(\theta,\varrho) = -\sum_{t=s}^{n} w_t(\tau)\rho_\tau(Y_t - Q_{Y_t}(\tau|I_{t-1},\theta,\varrho)), \tag{6.1}$$

where $\rho_\tau(u) = (\tau - 1(u<0))u$. This criterion function is similar to that described in Example 1, with the exception that there is no censoring. The starting value $s = 100$ initializes the recursive specification so that the imputed initial conditions (taken to be the marginal quantiles) have a numerically negligible effect.

In the first step, we constructed the LT estimates using the flat weights

$$w_t(\tau) = 1/\tau(1-\tau) \text{ for each } \; t = s,...,T.$$

The results of the first step are not presented here, but they are very similar to those reported below. Because the weights are not optimal, the information equality does not hold, hence Quasi-posterior quantiles are not valid for confidence intervals. However, the confidence intervals suggested in Theorem 4 lead to asymptotically valid inference. Under the assumption of correct dynamic specification, stationary sampling, and the conditions specified in Proposition 3, the LTE's are consistent and asymptotically normal

$$(\widehat{\varrho}(\tau),\widehat{\theta}(\tau)')' \to_d \mathcal{N}\left(0, J(\theta_0)^{-1}\Omega(\theta_0)J(\theta_0)^{-1}\right), \tag{6.2}$$

where for $\nabla q_t(\tau) = \partial Q_{Y_t}(\tau|I_{t-1},\varrho(\tau),\theta(\tau))/\partial(\varrho,\theta')'$ and $q_t(\tau) = Q_{Y_t}(\tau|I_{t-1},\varrho(\tau),\theta(\tau))$,

$$J(\theta_0) = E f_{Y_t|I_{t-1}}(q_t(\tau))\nabla q_t(\tau)\nabla q_t(\tau)',$$

and for $\frac{\Delta_n(\theta_0)}{\sqrt{T-s}} = \frac{1}{\sqrt{T-s}}\sum_{t=s}^{T}\left[\tau - 1(Y_t < q_t(\tau))\right]\nabla q_t(\tau)$,

$$\Omega(\theta_0) = \lim_{T\to\infty}\frac{1}{T-s}E\Delta_n(\theta_0)\Delta_n(\theta_0)' = \tau(1-\tau)E\nabla q_t(\tau)\nabla q_t(\tau)'.$$

If the model is not correctly specified, then, for example, the Newey and West (1987) estimator provides a consistent and robust procedure for estimation of the limit variance $\Omega(\theta_0)$.

The estimation of the matrix $J(\theta_0)^{-1}$ can be done through the use of nonparametric methods as in Powell (1984). Alternatively, as suggested in Theorem 4, we can use the variance-covariance matrix of the MCMC sequence of parameter draws multiplied by $n = (T - s)$ as a consistent estimate of $J(\theta_0)^{-1}$. Plugging the estimates into the variance expression (6.2), we obtain the standard errors and confidence intervals that are qualitatively similar to those reported in Figures 4-7.

In order to illustrate the use of Quasi-posterior quantiles (Theorem 3) and improve estimation efficiency, we also carried out the second step estimation using the Koenker-Bassett criterion function (6.1) with the weights

$$\widehat{w}_t(\tau) = \frac{h}{[Q_{Y_t}(\tau + h/2|I_{t-1}, \widehat{\varrho}(\tau), \widehat{\theta}(\tau)) - Q_{Y_t}(\tau - h/2|I_{t-1}, \widehat{\varrho}(\tau), \widehat{\theta}(\tau))]} \cdot \frac{1}{\tau(1-\tau)},$$

where $h \propto Cn^{-1/3}$ and $C > 0$ is chosen using the rule given in Koenker (1994). Under the assumption of correct dynamic specification, these weights imply the generalized information equality, which validates Quasi-posterior quantiles for inference purposes, as in (4.12)-(4.16). The following analysis is based on the second step estimates. The .05-th, .5-th, and .95-th Quasi-posterior quantiles are computed for each coefficient $\theta_j(\tau)$ $(j = 1, ..., 4)$ and $\varrho(\tau)$, and then used to form the point estimates and the 90%-confidence intervals, which are reported Figures 4-7 for $\tau = .2, .4, ..., .8$.

Figures 2 and 3 present the estimated surfaces of the conditional VaR functions of the dynamic model and linear models, respectively, plotted in the time-probability level coordinates, (t, p), ($p = \tau$ is the quantile index.) We report VaR for many values of τ. The conventional VaR reporting typically involves the probability levels at a given τ. Clearly, the whole VaR surface formed by varying τ represents a more complete depiction of conditional risk.

The dynamics depicted in Figures 2 and 3 unambiguously indicate certain dates on which market risk tends to be much higher than its usual level. The difference between the linear and the recursive model is also striking. The risk surface generated by the recursive model is much smoother and is much more persistent. Furthermore, this difference is statistically significant, as Figure 7 shows.

Focusing on the recursive model, let us examine the economic and statistical interpretation of the slope coefficients $\widehat{\theta}_2(\cdot)$, $\widehat{\theta}_3(\cdot)$, $\widehat{\theta}_4(\cdot)$, $\widehat{\varrho}(\cdot)$, plotted in Figures 4-7.

The coefficient on the lagged oil price return, $\widehat{\theta}_2(\cdot)$, is insignificantly positive in the left and right tails of the conditional return distribution. It is insignificantly negative in the middle part. The coefficient on the lagged DJI return, $\widehat{\theta}_3(\cdot)$, in contrast, is significantly positive for all values of τ. We also notice a sharp increase in the middle range. Thus, in addition to the strong positive relation between the individual stock return and the market return (DJI) (dictated by the fact that $\theta_2(\cdot) > 0$ on $(0.2, 0.8)$) there is also additional sensitivity of the median of the security return to the market movements.

The coefficient on the own lagged return, $\widehat{\theta}_4(\cdot)$, on the other hand, is significantly negative, except for values of τ close to 0. This may be interpreted as a reversion effect in the central part of the distribution. However, the lagged return does not appear to significantly shift the quantile function in the tails. Thus, the lagged return is more important for the determination of intermediate risks.

Most importantly, the dynamic coefficient $\widehat{\varrho}(\cdot)$ on the lagged VaR is significantly negative in the low quantiles and in the high quantiles, but is insignificant in the middle range. The significance of $\widehat{\varrho}(\cdot)$ is a strong evidence in favor of the recursive specification. The magnitude and sign of $\widehat{\varrho}(\cdot)$ indicates both the reversion and significant risk clustering effects in the tails of the distribution (see Figure 7). As expected, there is zero effect over the middle range, which is consistent with the random walk properties of the stock price. Thus, the dynamic effect of lagged VaR is much more important for the tails of the quantile function, that is for risk management purposes.

7 Conclusion

In this paper, we study the Laplace-type Estimators or Quasi-Bayesian Estimators that we define using common statistical, non-likelihood based criterion functions. Under mild regularity conditions these estimators are $\sqrt{n}$-consistent and asymptotically normal, and Quasi-posterior quantiles provide asymptotically valid confidence intervals. A simulation study and an empirical example illustrate the properties of the proposed estimation and inference methods. These results show that in many important cases the Quasi-Bayesian estimators provide useful alternatives to the usual extremum estimators. In ongoing work, we are extending the results to models in which $\sqrt{n}$-convergence rate and asymptotic normality do not hold, including the maximum score problem.

Acknowledgments: We thank the editor for the invitation of this paper to Journal of Econometrics and an anonymous referee for prompt and highest quality feedback. We thank Xiahong Chen, Shawn Cole, Gary Chamberlain, Ivan Fernandez, Ronald Gallant, Jinyong Hahn, Bruce Hansen, Chris Hansen, Jerry Hausman, James Heckman, Bo Honore, Guido Imbens, Roger Koenker, Shakeeb Khan, Sergei Morozov, Whitney Newey, Ziad Nejmeldeen, Stavros Panaceas, Chris Sims, George Tauchen, and seminar participants at Brown University, Duke-UNC Triangle Seminar, MIT, MIT-Harvard, University of Chicago, Princeton University, University of Wisconsin at Madison, University of Michigan, Michigan State University, Texas-AM University, the Winter meeting of the Econometric Society, the 2002 European Econometric Society Meeting in Venice for insightful comments. We gratefully acknowledge the financial support provided by the U.S. National Science Foundation grants SES-0214047 and SES-0079495.

A Appendix of Proofs

A.1 Proof of Theorem 1

It suffices to show

$$\int_{H_n} |h|^{\alpha} \left| p_n^*(h) - p_\infty^*(h) \right| dh \rightarrow_p 0 \tag{A.1}$$

for all $\alpha \geq 0$. Our arguments follow those in Bickel and Yahav (1969) and Ibragimov and Has'minskii (1981), as presented by Lehmann and Casella (1998). As indicated in the text, the main difference are in part 2, and are due to (i) the non-likelihood setting, (ii) the use of Huber-like conditions in Assumption 4 to handle discontinuous criterion functions, (iii) allowing more general loss functions, which are needed for construction of confidence intervals.

Throughout this proof the range of integration for h is implicitly understood to be H_n. For clarity of the argument, we limit exposition only to the case where $J_n(\theta)$ and $\Omega_n(\theta)$ do not depend on n. The more general case follows similarly.

Part 1. Define

$$h \equiv \sqrt{n}\left(\theta - T_n\right), \; T_n = \theta_0 + \frac{1}{n} J\left(\theta_0\right)^{-1} \Delta_n\left(\theta_0\right), \; U_n \equiv \frac{1}{\sqrt{n}} J(\theta_0)^{-1}\Delta_n(\theta_0), \tag{A.2}$$

then

$$\begin{aligned} p_n^*\left(h\right) \equiv & \frac{1}{\sqrt{n}} p_n\left(h/\sqrt{n} + \theta_0 + U_n/\sqrt{n}\right) \\ = & \frac{\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp\left(L_n\left(\frac{h}{\sqrt{n}} + T_n\right)\right)}{\int_{H_n}\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp\left(L_n\left(\frac{h}{\sqrt{n}} + T_n\right)\right)dh} \\ = & \frac{\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp\left(\omega(h)\right)}{\int_{H_n}\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp\left(\omega(h)\right)dh} \\ = & \frac{\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp\left(\omega(h)\right)}{C_n}, \end{aligned}$$

where

$$\omega\left(h\right) \equiv L_n\left(T_n + \frac{h}{\sqrt{n}}\right) - L\left(\theta_0\right) - \frac{1}{2n}\Delta_n\left(\theta_0\right)' J\left(\theta_0\right)^{-1}\Delta_n\left(\theta_0\right), \tag{A.3}$$

and

$$C_n \equiv \int_{H_n}\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp\left(\omega(h)\right)dh.$$

Part 2 shows that for each $\alpha \geq 0$,

$$A_{1n} \equiv \int_{H_n}|h|^\alpha\left|\exp\left(w\left(h\right)\right)\pi\left(T_n + \frac{h}{\sqrt{n}}\right) - \exp\left(-\frac{1}{2}h' J\left(\theta_0\right)h\right)\pi\left(\theta_0\right)\right|dh \xrightarrow{p} 0. \tag{A.4}$$

Given (A.4), taking $\alpha = 0$ we have

$$C_n \rightarrow_p \int_{\mathbf{R}^d} e^{-\frac{1}{2}h' J(\theta_0)h}\pi\left(\theta_0\right)dh = \pi\left(\theta_0\right)\left(2\pi\right)^{\frac{d}{2}}\left|\det J\left(\theta_0\right)\right|^{-1/2}, \tag{A.5}$$

hence

$$C_n = O_p(1).$$

Next note

$$\text{left side of (A.1)} \; \equiv \int_{H_n}|h|^\alpha\left|p_n^*(h) - p_\infty^*(h)\right|dh = A_n \cdot C_n^{-1},$$

where

$$A_n \equiv \int_{H_n}|h|^\alpha\left|e^{w(h)}\pi\left(T_n + \frac{h}{\sqrt{n}}\right) - (2\pi)^{-d/2}\left|\det J\left(\theta_0\right)\right|^{1/2}\exp\left(-\frac{1}{2}h' J\left(\theta_0\right)h\right)\cdot C_n\right|dh.$$

Using (A.5), to show (A.1) it suffices to show that $A_n \xrightarrow{p} 0$. But

$$A_n \leq A_{1n} + A_{2n}$$

where

$$A_{2n} \equiv \int_{H_n} |h|^{\alpha} \left| C_n (2\pi)^{-d/2} |\det J(\theta_0)|^{1/2} \exp\left(-\frac{1}{2} h' J(\theta_0) h\right) - \pi(\theta_0) \exp\left(-\frac{1}{2} h' J(\theta_0) h\right) \right| dh.$$

Then by (A.4)

$$A_{1n} \xrightarrow{p} 0,$$

and

$$\begin{aligned} A_{2n} &= \left| C_n (2\pi)^{-d/2} |\det J(\theta_0)|^{1/2} - \pi(\theta_0) \right| \int_{H_n} |h|^{\alpha} \exp\left(-\frac{1}{2} h' J(\theta_0) h\right) dh \\ &\xrightarrow{p} 0. \end{aligned}$$

Part 2. It remains only to show (A.4). Given Assumption 4 and definitions in (A.2) and (A.3), write

$$\begin{aligned} \omega(h) &= \Delta_n(\theta_0)' \left(\frac{U_n + h}{\sqrt{n}}\right) - \frac{1}{2} n \left(\frac{U_n + h}{\sqrt{n}}\right)' J(\theta_0) \left(\frac{U_n + h}{\sqrt{n}}\right) \\ &\quad - \frac{1}{2n} \Delta_n(\theta_0)' J(\theta_0)^{-1} \Delta_n(\theta_0) + R_n\left(\frac{h}{\sqrt{n}} + T_n\right) \\ &= -\frac{1}{2} h' J(\theta_0) h + R_n\left(\frac{h}{\sqrt{n}} + T_n\right). \end{aligned}$$

Split the integral A_{1n} in (A.4) over three separate areas:

- Area (i) : $|h| \leq M$,
- Area (ii) : $M \leq |h| \leq \delta\sqrt{n}$,
- Area (iii) : $|h| \geq \delta\sqrt{n}$.

Each of these areas is implicitly understood to intersect with the range of integration for h, which is H_n.

Area (i): We will show that for each $0 < M < \infty$ and each $\epsilon > 0$

$$\begin{aligned} \liminf_n P_* \Bigg\{ \int_{|h| \leq M} |h|^{\alpha} \Bigg| \exp(w(h)) \pi\left(T_n + \frac{h}{\sqrt{n}}\right) \\ - \exp\left(-\frac{1}{2} h' J(\theta_0) h\right) \pi(\theta_0) \Bigg| dh < \epsilon \Bigg\} \geq 1 - \epsilon. \end{aligned} \tag{A.6}$$

This is proved by showing that

$$\sup_{|h| \leq M} |h|^{\alpha} \left| \exp(w(h)) \pi\left(T_n + \frac{h}{\sqrt{n}}\right) - \exp\left(-\frac{1}{2} h' J(\theta_0) h\right) \pi(\theta_0) \right| \xrightarrow{p} 0. \tag{A.7}$$

Using the definition of $\omega(h)$, (A.7) follows from:

$$(a) \sup_{|h| \leq M} \left| \pi\left(\frac{h}{\sqrt{n}} + T_n\right) - \pi(\theta_0) \right| \xrightarrow{p} 0, \quad (b) \sup_{|h| \leq M} \left| R_n\left(\frac{h}{\sqrt{n}} + T_n\right) \right| \xrightarrow{p} 0,$$

where (a) follows from the continuity of $\pi(\cdot)$ and because by Assumption 4.ii-4.iii:

$$\frac{1}{\sqrt{n}} J(\theta_0)^{-1} \Delta_n(\theta_0) = O_p(1). \tag{A.8}$$

Given (A.8), (b) follows from Assumption 4.iv, since

$$\sup_{|h|\le M}\left|T_n+\frac{h}{\sqrt{n}}-\theta_0\right|=O_p(1/\sqrt{n}).$$

Area (ii): We show that for each $\epsilon>0$ there exist large M and small $\delta>0$ such that

$$\begin{aligned}\liminf_n P_*\Bigg\{\int_{M<|h|<\delta\sqrt{n}}|h|^\alpha\Bigg|&\exp\left(w\left(h\right)\right)\pi\left(T_n+\frac{h}{\sqrt{n}}\right)\\&-\exp\left(-\frac{1}{2}h'J\left(\theta_0\right)h\right)\pi\left(\theta_0\right)\Bigg|dh<\epsilon\Bigg\}\ge 1-\epsilon.\end{aligned}\tag{A.9}$$

Since the integral of the second term is finite and can be made arbitrarily small by setting M large, it suffices to show that for each $\epsilon>0$ there exist large M and small $\delta>0$ such that

$$\liminf_n P_*\left\{\int_{M<|h|<\delta\sqrt{n}}|h|^\alpha\left|\exp\left(w\left(h\right)\right)\pi\left(T_n+\frac{h}{\sqrt{n}}\right)\right|dh<\epsilon\right\}\ge 1-\epsilon.\tag{A.10}$$

In order to do so, it suffices to show that for sufficiently large M as $n\to\infty$

$$\exp\left(w\left(h\right)\right)\pi\left(T_n+\frac{h}{\sqrt{n}}\right)\le C\exp\left(-\frac{1}{4}h'J\left(\theta_0\right)h\right),\ \text{for all } M<|h|<\delta\sqrt{n}.\tag{A.11}$$

By assumption $\pi(\cdot)<K$, so we can drop it from consideration.

By definition of $\omega(h)$

$$\exp\left(w\left(h\right)\right)\le\exp\left(-\frac{1}{2}h'J\left(\theta_0\right)h+\left|R_n\left(T_n+\frac{h}{\sqrt{n}}\right)\right|\right).$$

Since $|T_n-\theta_0|=o_p(1)$, for any $\delta>0$ wp $\to 1$

$$\left|T_n+\frac{h}{\sqrt{n}}-\theta_0\right|<2\delta,\quad\text{for all}\quad |h|\le\delta\sqrt{n}.$$

Thus, by Assumption 4.iv(a) there exists some small δ and large M such that

$$\liminf_n P_*\left\{\sup_{M\le|h|\le\delta\sqrt{n}}\frac{|R_n\left(T_n+\frac{h}{\sqrt{n}}\right)|}{|h+\frac{1}{\sqrt{n}}J\left(\theta_0\right)^{-1}\Delta_n\left(\theta_0\right)|^2}\le\frac{1}{4}\text{mineig}\left(J\left(\theta_0\right)\right)\right\}\ge 1-\epsilon.$$

Since $\frac{1}{n}|J\left(\theta_0\right)^{-1}\Delta_n\left(\theta_0\right)|^2=O_p\left(1\right)$, for some $C>0$

$$\begin{aligned}&\liminf_n P_*\left\{\exp\left(w\left(h\right)\right)\le C\exp\left(-\frac{1}{4}h'J\left(\theta_0\right)h\right)\right\}\\&\ge\liminf_n P_*\left\{e^{w(h)}\le C\exp\left(-\frac{1}{2}h'J\left(\theta_0\right)h+\frac{1}{4}\text{mineig}\left(J\left(\theta_0\right)\right)|h|^2\right)\right\}\\&\ge 1-\epsilon.\end{aligned}\tag{A.12}$$

(A.12) implies (A.11), which in turn implies (A.9).

Area (iii): We will show that for each $\epsilon > 0$ and each $\delta > 0$,

$$\liminf_n P_* \left\{ \int_{h \geq \delta\sqrt{n}} |h|^\alpha \left| \exp(w(h)) \pi \left(T_n + \frac{h}{\sqrt{n}} \right) - \exp\left(-\frac{1}{2} h' J(\theta_0) h \right) \pi(\theta_0) \right| dh < \epsilon \right\} \geq 1 - \epsilon. \tag{A.13}$$

The integral of the second term clearly goes to 0 as $n \to \infty$. Therefore we only need to show

$$\int_{|h| \geq \delta\sqrt{n}} |h|^\alpha e^{w(h)} \pi \left(T_n + \frac{h}{\sqrt{n}} \right) dh \to_p 0.$$

Recalling the definition of h, the term is bounded by

$$\sqrt{n}^{\alpha+1} \int_{|\theta - T_n| \geq \delta} \left| T_n - \theta \right|^\alpha \pi(\theta) \exp\left(L_n(\theta) - L_n(\theta_0) - \frac{1}{2n} \Delta_n(\theta_0)' J(\theta_0)^{-1} \Delta_n(\theta_0) \right) d\theta.$$

Since $T_n - \theta_0 \xrightarrow{p} 0$, wp $\to 1$ this is bounded by

$$K_n \cdot C \cdot \sqrt{n}^{\alpha+1} \int_{|\theta - \theta_0| \geq \delta/2} \left(1 + |\theta|^\alpha \right) \pi(\theta) \exp\left(L_n(\theta) - L_n(\theta_0) \right) d\theta,$$

where

$$K_n = \exp\left(-\frac{1}{2n} \Delta_n(\theta_0)' J(\theta_0)^{-1} \Delta_n(\theta_0) \right) = O_p(1).$$

By Assumption 3 there exists $\varepsilon > 0$ such that

$$\liminf_{n \to \infty} P_* \left\{ \sup_{|\theta - \theta_0| \geq \delta/2} e^{L_n(\theta) - L_n(\theta_0)} \leq e^{-n\epsilon} \right\} = 1$$

Thus, wp $\to 1$ the entire term is bounded by

$$K_n \cdot C \cdot \sqrt{n}^{\alpha+1} \cdot e^{-n\epsilon} \int_\Theta |\theta|^\alpha \pi(\theta) d\theta = o_p(1). \tag{A.14}$$

Here observe that compactness is only used to insure that

$$\int_\Theta |\theta|^\alpha \pi(\theta) d\theta < \infty. \tag{A.15}$$

Hence by replacing compactness with the condition (A.15), the conclusion (A.14) is not affected for the given α.

The entire proof is now completed by combining (A.6), (A.9), and (A.13). ■

A.2 Proof of Theorem 2

For clarity of the argument, we limit exposition only to the case where $J_n(\theta)$ and $\Omega_n(\theta)$ do not depend on n. The more general case follows similarly.

Recall that

$$h = \sqrt{n}(\theta - \theta_0) - J(\theta_0)^{-1} \Delta_n(\theta_0) / \sqrt{n}.$$

Define $U_n = J(\theta_0)^{-1}\Delta_n(\theta_0)/\sqrt{n}$. Consider the objective function

$$Q_n(z) = \int_{H_n} \rho(z - h - U_n)p_n^*(h)dh,$$

which is minimized at $\sqrt{n}(\widehat{\theta} - \theta_0)$. Also define

$$Q_\infty(z) = \int_{\mathbf{R}^d} \rho(z - h - U_n)p_\infty^*(h)dh.$$

which is minimized at a random vector denoted Z_n. Define

$$\xi = \arg \inf_{z \in \mathbf{R}^d} \left\{ \int_{\mathbf{R}^d} \rho(z - h)\, p_\infty^*(h)\, dh \right\}.$$

Note that solution is unique and finite by Assumption 2 parts (ii) and (iii) on the loss function ρ. When ρ is symmetric, $\xi = 0$ by Anderson's lemma.

Therefore, $Z_n = \arg\inf_{z \in \mathbf{R}^d} Q_\infty(z)$ equals

$$Z_n = \xi + U_n = O_p(1).$$

Next, we have for any fixed z

$$Q_n(z) - Q_\infty(z) \to_p 0$$

since by Assumption 2.ii $\rho(h) \leq 1 + |h|^p$ and by $|a + b|^p \leq 2^{p-1}|a|^p + 2^{p-1}|b|^p$ for $p \geq 1$:

$$\begin{aligned}
|Q_n(z) - Q_\infty(z)| &\leq \int_{H_n} (1 + |z - h - U_n|^p)\, |p_n^*(h) - p_\infty^*(h)| dh \\
&+ \int_{H_n^c} (1 + |z - h - U_n|^p)\, (p_\infty^*(h))\, dh \\
&\leq \int_{H_n} \left(1 + 2^{p-1}|h|^p + 2^{p-1}|z - U_n|^p\right) |p_n^*(h) - p_\infty^*(h)| dh \\
&+ \int_{H_n^c} \left(1 + 2^{p-1}|h|^p + 2^{p-1}|z - U_n|^p\right) (p_\infty^*(h))\, dh \\
&\leq \int_{H_n} \left(1 + 2^{p-1}|h|^p + O_p(1)\right) (p_n^*(h) - p_\infty^*(h))\, dh \\
&+ \int_{H_n^c} \left(1 + 2^{p-1}|h|^p + O_p(1)\right) (p_\infty^*(h))\, dh = o_p(1),
\end{aligned}$$

where $o_p(1)$-conclusion is by Theorem 1 and exponentially small tails of the normal density (Lebesgue measure of H_n^c converges to zero).

Now note that $Q_n(z)$ and $Q_\infty(z)$ are convex and finite, and $Z_n = \arg\inf_{z \in \mathbf{R}^d} Q_\infty(z) = O_p(1)$. By the convexity lemma of Pollard (1991), pointwise convergence entails the uniform convergence over compact sets K:

$$\sup_{z \in K} \left| Q_n(z) - Q_\infty(z) \right| \to_p 0.$$

Since $Z_n = O_p(1)$, uniform convergence and convexity arguments like those in Jureckova (1977) imply that $\sqrt{n}(\widehat{\theta} - \theta_0) - Z_n \to_p 0$, as shown below. ■

Proof of $Z_n - \sqrt{n}(\widehat{\theta} - \theta_0) = o_p(1)$. The proof follows by extending slightly the convexity argument of Jureckova (1977) and Pollard (1991) to the present context. Consider a ball $B_\delta(Z_n)$ with radius $\delta > 0$, centered at Z_n, and let $z = Z_n + dv$, where v is a unit direction vector such that $|v| = 1$ and $d > \delta$. Because $Z_n = O_p(1)$, for any $\delta > 0$ and $\epsilon > 0$, there exists $K > 0$ such that

$$\liminf_{n\to\infty} P_*\Big\{E_n = \{B_\delta(Z_n) \in B_K(0)\}\Big\} \geq 1-\epsilon.$$

By convexity, for any $z = Z_n + dv$ constructed so, it follows that .

$$\frac{\delta}{d}(Q_n(z) - Q_n(Z_n)) \geq Q_n(z^*) - Q_n(Z_n), \tag{A.16}$$

where z^* is a point of boundary of $B_\delta(Z_n)$ on the line connecting z and Z_n. By the uniform convergence of $Q_n(z)$ to $Q_\infty(z)$ over any compact set $B_K(0)$, whenever E_n occurs:

$$\begin{aligned}\frac{\delta}{d}\big(Q_n(z) - Q_n(Z_n)\big) &\geq Q_n(z^*) - Q_n(Z_n)\\ &\geq Q_\infty(z^*) - Q_\infty(Z_n) + o_p(1) \geq V_n + o_p(1),\end{aligned}$$

where $V_n > 0$ is a uniformly in n positive variable, because Z_n is the unique optimizer of Q_∞. That is, there exists an $\eta > 0$ such that $\liminf_n P(V_n > \eta) \geq 1-\epsilon$. Hence we have with probability at least as big as $1 - 3\epsilon$ for large n:

$$\frac{\delta}{d}\big(Q_n(z) - Q_n(Z_n)\big) > \eta.$$

Thus, $\sqrt{n}(\widehat{\theta} - \theta_0)$ eventually belongs to a complement of $B_\delta(Z_n)$ with probability at most 3ϵ. Since we can set ϵ as small as we like by picking (a) sufficiently large K, and (b) sufficiently large n, and (c) sufficiently small $\eta > 0$, it follows that

$$\limsup_{n\to\infty} P^*\Big\{|Z_n - \sqrt{n}(\widehat{\theta} - \theta_0)| > \delta\Big\} = 0.$$

Since this is true for any $\delta > 0$, it follows that

$$Z_n - \sqrt{n}(\widehat{\theta} - \theta_0) = o_p(1). \quad \blacksquare$$

A.3 Proof of Theorem 3

For clarity of the argument, we limit exposition only to the case where $J_n(\theta)$ and $\Omega_n(\theta)$ do not depend on n. The more general case follows similarly. We defined

$$F_{g,n}(x) = \int_{\theta\in\Theta: g(\theta)\leq x} p_n(\theta) d\theta.$$

Evaluate it at $x = g(\theta_0) + s/\sqrt{n}$ and change the variable of integration

$$H_{g,n}(s) = F_{g,n}(g(\theta_0) + s/\sqrt{n}) = \int_{h\in H_n: g(\theta_0 + h/\sqrt{n} + U_n/\sqrt{n}) \leq g(\theta_0) + s/\sqrt{n}} p_n^*(h) dh.$$

Define also

$$\widehat{H}_{g,n}(s) = \int_{h\in\mathbb{R}^d: g(\theta_0 + h/\sqrt{n} + U_n/\sqrt{n}) \leq g(\theta_0) + s/\sqrt{n}} p_\infty^*(h) dh$$

and

$$H_{g,\infty}(s) = \int_{h\in\mathbb{R}^d: \nabla g(\theta_0)'(h/\sqrt{n} + U_n/\sqrt{n}) \leq s/\sqrt{n}} p_\infty^*(h) dh.$$

By definition of total variation of moments norm and Theorem 1

$$\sup_s |H_{g,n}(s) - \widehat{H}_{g,n}(s)| \to_p 0,$$

where the sup is taken over the support of $H_{g,n}(s)$.

By the uniform continuity of the integral of the normal density with respect to the boundary of integration

$$\sup_s |\widehat{H}_{g,n}(s) - H_{g,\infty}(s)| \to_p 0,$$

which implies

$$\sup_s |H_{g,n}(s) - H_{g,\infty}(s)| \to_p 0.$$

where the sup is taken over the support of $H_{g,n}(s)$.

The convergence of distribution function implies the convergence of quantiles at continuity points of distribution functions, see e.g. Billingsley (1994), so

$$H_{g,n}^{-1}(\alpha) - H_{g,\infty}^{-1}(\alpha) \to_p 0.$$

Next observe

$$H_{g,\infty}(s) = P\Big\{\nabla g(\theta_0)'\mathcal{N}(U_n, J^{-1}(\theta_0)) < s \Big| U_n\Big\},$$

so

$$H_{g,\infty}^{-1}(\alpha) = \nabla g(\theta_0)'U_n + q_\alpha\sqrt{\nabla_\theta g(\theta_0)'J^{-1}(\theta_0)\nabla_\theta g(\theta_0)},$$

where q_α is the α-quantile of $\mathcal{N}(0,1)$.

Recalling that we defined $c_{g,n}(\alpha) = F_{g,n}^{-1}(\alpha)$, by quantile equivariance with respect to the monotone transformations

$$H_{g,n}^{-1}(\alpha) = \sqrt{n}\,(c_{g,n}(\alpha) - g(\theta_0))$$

so that

$$\sqrt{n}\,(c_{g,n}(\alpha) - g(\theta_0)) = \nabla g(\theta_0)'U_n + q_\alpha\sqrt{\nabla_\theta g(\theta_0)'J^{-1}(\theta_0)\nabla_\theta g(\theta_0)} + o_p(1).$$

The rest of the result follows by the Δ-method. ∎

A.4 Proof of Theorem 4

In view of Assumption 4, it suffices to show that

$$\widehat{J}_n^{-1}(\theta_0) - J_n^{-1}(\theta_0) \to_p 0, \qquad \text{(A.17)}$$

and then conclude by the Δ-method.

Recall that

$$h = \sqrt{n}\,(\theta - \theta_0) - \underbrace{J_n\,(\theta_0)^{-1}\,\Delta_n\,(\theta_0)\,/\sqrt{n}}_{U_n},$$

and the localized Quasi-posterior density for h is

$$p_n^*\,(h) = \frac{1}{\sqrt{n}}p_n\left(h/\sqrt{n} + \theta_0 + U_n/\sqrt{n}\right).$$

Note also

$$\widehat{J}_n^{-1}(\theta_0) \equiv \int_\Theta n(\theta - \widehat{\theta})(\theta - \widehat{\theta})' p_n(\theta) d\theta$$

$$= \int_{H_n} (h - \sqrt{n}(\widehat{\theta} - \theta_0) + U_n) \cdot (h - \sqrt{n}(\widehat{\theta} - \theta_0) + U_n)' p_n^*(h) dh,$$

and

$$J_n^{-1}(\theta_0) \equiv \int_{\mathbf{R}^d} hh' p_\infty^*(h) dh.$$

We have, denoting $h = (h_1, ..., d_d)$ and $\widetilde{T}_n = (\widetilde{T}_{n1}, ..., \widetilde{T}_{nd})$ where $\widetilde{T}_n = \sqrt{n}(\widehat{\theta} - \theta_0) - U_n$, for all $i, j \leq d$

(a) $\int_{H_n} h_i h_j \Big(p_n^*(h) - p_\infty^*(h)\Big) dh = o_p(1)$ by Theorem 1,

(b) $\int_{H_n^c} h_i h_j \Big(p_\infty^*(h)\Big) dh = o_p(1)$ by definition of p_∞^* and $J_n(\theta_0)$ being uniformly nonsingular,

(c) $\int_{H_n} \underbrace{|\widetilde{T}_n|^2}_{=o_p(1)} \Big(p_n^*(h) - p_\infty^*(h)\Big) dh = o_p(1)$ by Theorem 2,

(d) $\int_{H_n} \underbrace{|\widetilde{T}_n|^2}_{=o_p(1)} \Big(p_\infty^*(h)\Big) dh = o_p(1)$ by Theorem 2, definition of p_∞^*, and $J_n(\theta_0)$ being nonsingular,

(e) $\int_{H_n} h_j \underbrace{\widetilde{T}_{ni}}_{=o_p(1)} \Big(p_n^*(h) - p_\infty^*(h)\Big) dh = o_p(1)$ by Theorems 1 and 2,

(f) $\int_{H_n} h_j \underbrace{\widetilde{T}_{ni}}_{=o_p(1)} \Big(p_\infty^*(h)\Big) dh = o_p(1)$ by Theorems 1 and 2, definition of p_∞^*, and $J_n(\theta_0)$ being uniformly nonsingular, from which the required conclusion follows. ■

A.5 Proof of Proposition 1

Assumption 3 is directly implied by (4.1)-(4.4) and the uniform continuity of $Em_i(\theta)$, as shown in Lemma 1. It remains only to verify Assumption 4.

Define the identity

$$L_n(\theta) - L_n(\theta_0) \equiv - \underbrace{n g_n(\theta_0)' W(\theta_0) G(\theta_0)}_{\Delta_n(\theta_0)'} (\theta - \theta_0)$$
$$- \frac{1}{2}(\theta - \theta_0)' n \underbrace{G(\theta_0)' W(\theta_0) G(\theta_0)}_{J(\theta_0)} (\theta - \theta_0) \quad + \quad R_n(\theta). \tag{A.18}$$

Next, given the definition of $\Delta_n(\theta_0)$ and $J(\theta_0)$, conditions **i**, **ii**, **iii** of Assumption 4. are immediate from conditions **i-iii** of Proposition 1. Condition **iv** is verified as follows. Condition **iv** of Assumption 4 can be succinctly stated as:

$$\text{for each } \epsilon > 0 \text{ there exists a } \delta > 0 \text{ such that } \limsup_{n\to\infty} P^* \left\{ \sup_{|\theta - \theta_0| \leq \delta} \frac{|R_n(\theta)|}{1 + n|\theta - \theta_0|^2} > \epsilon \right\} < \epsilon.$$

This stochastic equicontinuity condition is equivalent to the following stochastic equicontinuity condition, see e.g. Andrews (1994a):

$$\text{for any } \delta_n \to 0 \qquad \sup_{|\theta-\theta_0|\le\delta_n} \frac{|R_n(\theta)|}{1+n|\theta-\theta_0|^2} = o_p(1). \tag{A.19}$$

This is weaker than condition (v) of Theorem 7.1 in Newey and McFadden (1994), which requires

$$\sup_{|\theta-\theta_0|\le\delta_n} \frac{R_n(\theta)}{\sqrt{n}|\theta-\theta_0|+n|\theta-\theta_0|^2} = o_p(1), \tag{A.20}$$

since

$$\frac{R_n(\theta)}{1+n|\theta-\theta_0|^2} = \frac{R_n(\theta)}{\sqrt{n}|\theta-\theta_0|+n|\theta-\theta_0|^2}\left[\frac{\sqrt{n}|\theta-\theta_0|+n|\theta-\theta_0|^2}{1+n|\theta-\theta_0|^2}\right],$$

where the term in brackets is bounded by

$$1+\frac{\sqrt{n}|\theta-\theta_0|}{1+n|\theta-\theta_0|^2} \le 2.$$

Hence the arguments of the proof, except for several important differences, follow those of Theorem 7.2 in Newey and McFadden (1994).

At first note that condition iv of Proposition 1 is implied by the condition (where we let $g(\theta) \equiv Eg_n(\theta)$):

$$\sup_{\theta\in B_{\delta_n}(\theta_0)} \epsilon(\theta) = o_p\left(\frac{1}{\sqrt{n}}\right), \text{ where } \epsilon(\theta) = \frac{g_n(\theta)-g_n(\theta_0)-g(\theta)}{1+\sqrt{n}|\theta-\theta_0|} \text{ for any } \delta_n \to 0. \tag{A.21}$$

From (A.18)

$$R_n(\theta) = R_{1n}(\theta) + R_{2n}(\theta) + R_{3n}(\theta),$$

where

$$\begin{aligned}
R_{1n}(\theta) \equiv& n\Big(g_n(\theta_0)' W_n(\theta) G(\theta_0)(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)' G(\theta_0)' W(\theta) G(\theta_0)(\theta-\theta_0) \\
&\qquad - \frac{1}{2}g_n(\theta)' W_n(\theta) g_n(\theta) + \frac{1}{2}g_n(\theta_0)' W_n(\theta) g_n(\theta_0)\Big), \\
R_{2n}(\theta) \equiv& n\Big(\frac{1}{2}g_n(\theta_0)'(W_n(\theta_0)-W_n(\theta))\, g_n(\theta_0)\Big), \\
R_{3n}(\theta) \equiv& n\Big(g_n(\theta_0)'(W(\theta_0)-W_n(\theta)))\, G(\theta_0)(\theta-\theta_0) \\
&\qquad + \frac{1}{2}(\theta-\theta_0)' G(\theta_0)'(W(\theta_0)-W(\theta))\, G(\theta_0)(\theta-\theta_0)\Big).
\end{aligned}$$

Verification of (A.19) for the terms $R_{2n}(\theta)$ and R_{3n} immediately follows from $\sqrt{n}g_n(\theta_0) = O_p(1)$ and the uniform consistency of $W_n(\theta)$ in θ as assumed in condition i of Proposition 1 and from the continuity of $W(\theta)$ in θ by condition i of Proposition 1, so that $W_n(\theta) - W(\theta) = o_p(1)$ uniformly in θ and $W(\theta) - W(\theta_0) = o(1)$ as $|\theta-\theta_0| \to 0$.

It remains to check condition (A.19) for the term $R_{1n}(\theta)$. Note that

$$g_n(\theta) = \left(1+\sqrt{n}|\theta-\theta_0|\right)\epsilon(\theta) + g(\theta) + g_n(\theta_0).$$

Substitute this into $R_{1n}(\theta)$ and decompose

$$-\frac{1}{n}R_{1n}(\theta) = \underbrace{\frac{1}{2}\left(1+\sqrt{n}|\theta-\theta_0|\right)^2 \epsilon(\theta)' W_n(\theta)\epsilon(\theta)}_{r_1(\theta)} + \underbrace{g_n(\theta_0)' W_n(\theta)(g(\theta) - G(\theta_0))(\theta-\theta_0)}_{r_2(\theta)}$$
$$+ \underbrace{\left(1+\sqrt{n}|\theta-\theta_0|\right)\epsilon(\theta)' W_n(\theta) g_n(\theta_0)}_{r_3(\theta)} + \underbrace{\left(1+\sqrt{n}|\theta-\theta_0|\right)\epsilon(\theta)' W_n(\theta) g(\theta)}_{r_4(\theta)}$$
$$+ \underbrace{\frac{1}{2}g(\theta)'(W_n(\theta) - W(\theta))\, g(\theta)}_{r_5(\theta)} + \underbrace{\frac{1}{2}g(\theta)' W(\theta) g(\theta) - \frac{1}{2}(\theta-\theta_0)' G(\theta_0)' W(\theta) G(\theta_0)(\theta-\theta_0)}_{r_6(\theta)}.$$

Using the inequalities, for $x \geq 0$:

$$\frac{\left(1+\sqrt{n}x\right)^2}{1+nx^2} \leq 2, \quad \frac{\sqrt{n}x}{1+nx^2} \leq 1, \quad \frac{1+\sqrt{n}x}{1+nx^2} \leq 2, \quad \frac{n\left(1+\sqrt{n}x\right)}{1+nx^2} \leq 2\frac{\sqrt{n}}{x}, \tag{A.22}$$

each of these terms can be dealt with separately, by applying the conditions **i**- **iii** and (A.21):

$$(a) \quad \sup_{\theta\in B_{\delta_n}(\theta_0)} \frac{n|r_1(\theta)|}{1+n|\theta-\theta_0|^2} \leq \sup_{\theta\in B_{\delta_n}(\theta_0)} n\epsilon(\theta)' W_n(\theta)\epsilon(\theta) = o_p(1),$$

$$(b) \quad \sup_{\theta\in B_{\delta_n}(\theta_0)} \frac{n|r_2(\theta)|}{1+n|\theta-\theta_0|^2} = \sup_{\theta\in B_{\delta_n}(\theta_0)} \frac{o\left(\sqrt{n}|\theta-\theta_0|\right)'}{1+n|\theta-\theta_0|^2} W_n(\theta)\sqrt{n} g_n(\theta_0) = o_p(1),$$

$$(c) \quad \sup_{\theta\in B_{\delta_n}(\theta_0)} \frac{n|r_3(\theta)|}{1+n|\theta-\theta_0|^2} \leq \sup_{\theta\in B_{\delta_n}(\theta_0)} 2n|\epsilon(\theta)' W_n(\theta) g_n(\theta_0)| = o_p(1),$$

$$(d) \quad \sup_{\theta\in B_{\delta_n}(\theta_0)} \frac{n|r_4(\theta)|}{1+n|\theta-\theta_0|^2} \leq \sup_{\theta\in B_{\delta_n}(\theta_0)} 2\sqrt{n}\left|\epsilon(\theta)' W_n(\theta)\frac{g(\theta)}{|\theta-\theta_0|}\right| = o_p(1),$$

$$(e) \quad \sup_{\theta\in B_{\delta_n}(\theta_0)} \frac{n|r_5(\theta)|}{1+n|\theta-\theta_0|^2} \leq \sup_{\theta\in B_{\delta_n}(\theta_0)} \frac{|g(\theta)|^2}{|\theta-\theta_0|^2}|W_n(\theta) - W(\theta)| = o_p(1),$$

$$(f) \quad \sup_{\theta\in B_{\delta_n}(\theta_0)} \frac{n|r_6(\theta)|}{1+n|\theta-\theta_0|^2} \leq \sup_{\theta\in B_{\delta_n}(\theta_0)} \frac{o\left(|\theta-\theta_0|^2|W(\theta)|\right)}{|\theta-\theta_0|^2} = o_p(1),$$

where (a) follows from (A.22), (A.21), and condition **i**, which states that $W_n(\theta) = W(\theta) + o_p(1)$ and $W(\theta) > 0$ is finite uniformly in θ; in (b) the first equality follows by Taylor expansion $g(\theta) = G(\theta_0)(\theta-\theta_0) + o(|\theta-\theta_0|)$ and the second conclusion follows from (A.22) and condition **iii**; (c) follows from (A.22), (A.21), condition **i** and **iii**; (d) follows by (A.22) and then replacing, by condition **ii**, $g(\theta)$ with $G(\theta_0)(\theta-\theta_0) + o(|\theta-\theta_0|)$, followed by applying (A.21) and condition **i**; (e) follows from replacing by condition **ii** $g(\theta)$ with $G(\theta_0)(\theta-\theta_0) + o(|\theta-\theta_0|)$, followed by applying condition **i**; and (f) follows from replacing $g(\theta)$ with $G(\theta_0)(\theta-\theta_0) + o(|\theta-\theta_0|)$, followed by applying condition **i**.

Verification of (A.20) for the term $R_{1n}(\theta)$ now follows by putting these terms together. ■

A.6 Proof of Proposition 2

Verification of Assumption 3 is standard given the stated conditions and is subsumed as a step in the consistency proofs of extremum estimators based on GEL in Kitamura and Stutzer (1997) for cases when s is finite and Kitamura (1997) for cases when s takes on infinite values. We shall not repeat it here. Next, we will verify Assumption 4. Define

$$\widehat{\gamma}(\theta) = \arg \inf_{\gamma \in \mathbb{R}^p} \bar{L}_n(\theta, \gamma).$$

It will suffice to show that uniformly in $\theta_n \in B_{\delta_n}(\theta_0)$ for any $\delta_n \to 0$, we have the GMM set-up:

$$\begin{aligned} L_n(\theta_n) &= \bar{L}_n(\theta_n, \widehat{\gamma}(\theta_n)) \\ &= -\frac{1}{2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n m_i(\theta_n)\right)' (V(\theta_0) + o_p(1))^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n m_i(\theta_n)\right), \end{aligned} \tag{A.23}$$

where

$$V(\theta_0) = E m_i(\theta_0) m_i(\theta_0)'.$$

The Assumptions 4.**i**-**iii** follow immediately from the conditions of Proposition 2, and Assumption 4.**iv** is verified exactly as in the proof of Proposition 1, given the reduction to the GMM case. Indeed, defining $g_n(\theta) = \frac{1}{n}\sum_{i=1}^n m_i(\theta)$, the Donsker property assumed in condition **iv** implies that for any $\epsilon > 0$, there is $\delta > 0$ such that

$$\limsup_{n\to\infty} P^*\left\{ \sup_{\theta \in B_\delta(\theta_0)} \sqrt{n}|g_n(\theta) - g_n(\theta_0) - (Eg_n(\theta) - Eg_n(\theta_0))| > \epsilon \right\} < \epsilon,$$

which implies

$$\limsup_{n\to\infty} P^*\left\{ \sup_{\theta \in B_\delta(\theta_0)} \frac{\sqrt{n}|g_n(\theta) - g_n(\theta_0) - (Eg_n(\theta) - Eg_n(\theta_0))|}{1 + \sqrt{n}|\theta - \theta_0|} > \epsilon \right\} < \epsilon,$$

which is condition **iv** in Proposition 1. The rest of the arguments follow that in the proof of Proposition 1.

It only remains to show the requisite expansion (A.23). We first show that

$$\widehat{\gamma}(\theta_n) \to_p 0.$$

For that purpose we use the convexity lemma, which was obtained by C. Geyer, and can be found in Knight (1999).

Convexity Lemma. Suppose Q_n is a sequence of lower-semi-continuous 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_n weakly converges to Q_∞ in $\bar{\mathbb{R}}$ marginally (in finite-dimensional sense) on $\mathcal{D}$ where Q_∞ is lower-semi-continuous convex and finite on an open non-empty set a.s., then

$$\arg\inf_{z \in \mathbb{R}^d} Q_n(z) \to_d \arg\inf_{z \in \mathbb{R}^d} Q_\infty(z),$$

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

Next, we show that $\widehat{\gamma}(\theta_n) \to_p 0$. Define $F = \{\gamma : Es[m_i(\theta_0)'\gamma] < \infty\}$ and $F^c = \{\gamma : Es[m_i(\theta_0)'\gamma] = \infty\}$. By convexity and lower-semicontinuity of s, F is convex, open, and its boundary is nowhere dense in $\mathbb{R}^p$. Thus for $\gamma \in F$, $Es[m_i(\theta)'\gamma] < \infty$ for all $\theta \in B_\delta(\theta_0)$ and some $\delta > 0$, which follows by continuity of $\theta \mapsto Es[m_i(\theta)'\gamma]$ over $B_\delta(\theta_0)$ implied by the condition **ii** and **iii**.

Thus, for a given $\gamma \in F$ and any $\theta_n \to_p \theta_0$

$$\frac{1}{n}\sum_{i=1}^n s[m_i(\theta_n)'\gamma] \to_p Es[m_i(\theta_0)'\gamma] < \infty.$$

This follows from the uniform law of large numbers implied by

1. $\{s[m_i(\theta)'\gamma], \theta \in B_\delta(\theta_0)\}$, where δ is sufficiently small, being a Donsker class wp $\to 1$, and

2. $Em_i(\theta) = \int x \cdot dP[m_i(\theta) \leq x]$ being continuously differentiable in θ by condition **ii** and **iii**.

The above function set is Donsker by

(a) $m_i(\theta)'\gamma \in M$ for some compact M and a given $\gamma \in F$, by condition **iii**,

(b) $\{m_i(\theta), \theta \in B_\delta(\theta_0)\}$ being Donsker class by condition **iv**,

(c) s being a uniform Lipschitz function over $\mathcal{V} \cap M$[17], by assumption on s,

(d) $m_i(\theta)'\gamma \in \mathcal{V}$ for all $\theta \in B_\delta(\theta_0)$, some $\delta > 0$, and a given $\gamma \in F$, by construction of F,

(e) Theorem 2.10.6 in van der Vaart and Wellner (1996) that says a uniform Lipschitz transform of a Donsker class is Donsker class itself.

Now take γ in $F^c \setminus \partial F$, where ∂F denotes the boundary of F. Then wp $\to 1$

$$\frac{1}{n}\sum_{i=1}^{n} s[m_i(\theta_n)'\gamma] = \infty \to_p Es[m_i(\theta_0)'\gamma] = \infty.$$

Now take all the rational numbers $\gamma \in \mathbb{R}^p \setminus \partial F$ as the set $\mathcal{D}$ appearing in the statement of the Convexity Lemma and conclude that

$$\widehat{\gamma}(\theta_n) \to_p 0 = \arg\inf_\gamma Es[m_i(\theta_0)'\gamma].$$

Given this result, we can expand the first order condition for $\widehat{\gamma}(\theta_n)$ in order to obtain the expression for its form. Note first

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \nabla s\left(\gamma(\theta_n)' m_i(\theta_n)\right) m_i(\theta_n) \\ &= \sum_{i=1}^{n} m_i(\theta_n) + \gamma(\theta_n) nV_n, \end{aligned} \tag{A.24}$$

where

$$V_n = \frac{1}{n}\sum_{i=1}^{n} \nabla^2 s\left(\bar{\gamma}(\theta_n)' m_i(\theta_n)\right) m_i(\theta_n) m_i(\theta_n)',$$

for some $\bar{\gamma}(\theta_n)$ between 0 and $\gamma(\theta_n)$, which is different from row to row of the matrix V_n.

Then

$$V_n \to_p V(\theta_0) = Em_i(\theta_0) m_i(\theta_0)'.$$

This follows from the uniform law of large numbers implied by

1. $\{\nabla^2 s(\gamma' m_i(\theta^*)) m_i(\theta) m_i(\theta)', (\theta^*, \gamma, \theta) \in B_{\delta_1}(\theta_0) \times B_{\delta_2}(0) \times B_{\delta_3}(\theta_0)\}$, where $\delta_j > 0$ are sufficiently small, being a Donsker class wp $\to 1$,

2. $Em_i(\theta)m_i(\theta)' = \int xx'dP[m_i(\theta) \leq x]$ being continuous function in θ by condition **i**,

[17] Recall that $\mathcal{V}$ is defined as the open convex set on which s is finite.

3. $E\nabla^2 s(\gamma' m_i(\theta^*)) m_i(\theta) m_i(\theta)' = \underbrace{E\nabla^2 s(0)}_{1} m_i(\theta) m_i(\theta)' + o(1)$ uniformly in $(\theta, \theta^*) \in B_\delta(\theta_0) \times B_\delta(\theta_0)$ for sufficiently small $\delta > 0$, for any $\gamma \to 0$, by assumptions on s and condition **iii**.

The claim **1** is verified by applying exactly the same logic as in the previously stated steps (a)-(e). For the sake of brevity, this will not be repeated.

Therefore, wp $\to 1$

$$\gamma(\theta_n) = -(V_n)^{-1} \frac{1}{n} \sum_{i=1}^{n} m_i(\theta_n) \equiv -(V(\theta_0)^{-1} + o_p(1)) \frac{1}{n} \sum_{i=1}^{n} m_i(\theta_n). \tag{A.25}$$

Consider the second order expansion,

$$\bar{L}_n(\theta_n, \gamma(\theta_n)) = \sqrt{n}\gamma(\theta_n)' \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_i(\theta_n) + \frac{1}{2} \sqrt{n}\gamma(\theta_n)' \tilde{V}_n \sqrt{n}\gamma(\theta_n), \tag{A.26}$$

where

$$\tilde{V}_n = \frac{1}{n} \sum_{i=1}^{n} \nabla^2 s\left(\tilde{\gamma}(\theta_n)' m_i(\theta_n)\right) m_i(\theta_n) m_i(\theta_n)',$$

for some $\tilde{\gamma}(\theta_n)$ between 0 and $\gamma(\theta_n)$, which is different from row to row of the matrix $\tilde{V}_n$. By a preceding argument,

$$\tilde{V}_n \to_p V(\theta_0).$$

Inserting (A.25) and $\tilde{V}_n = V(\theta_0) + o_p(1)$ into (A.26), we obtain the required expansion (A.23). ■

A.7 Proof of Proposition 3

Assumption 3 is assumed. We need to verify Assumption 4.

Define the identity

$$\begin{aligned} L_n(\theta) - L_n(\theta_0) \equiv & \underbrace{\sum_{i=1}^{n} \dot{m}_i(\theta_0)'}_{\Delta_n(\theta_0)'}(\theta - \theta_0) \\ & + \frac{1}{2}(\theta - \theta_0)' n \underbrace{\nabla_{\theta\theta'} E m_i(\theta_0)}_{-J(\theta_0)}(\theta - \theta_0) \\ & + R_n(\theta). \end{aligned} \tag{A.27}$$

Assumption 4.i-iii then follows immediately from conditions **i** and **ii**. Assumption 4.iv is verified as follows.

The remainder term $R_n(\theta)$ is given the following decomposition:

$$\begin{aligned} R_n(\theta) = & \underbrace{\sum_{i=1}^{n} \left\{ m_i(\theta) - m_i(\theta_0) - E m_i(\theta) + E m_i(\theta_0) - \dot{m}_i(\theta_0)'(\theta - \theta_0) \right\}}_{R_{1n}(\theta)} \\ & \underbrace{+ n(E m_i(\theta) - E m_i(\theta_0)) + \frac{1}{2}(\theta - \theta_0)' n J(\theta_0)(\theta - \theta_0)}_{R_{2n}(\theta)}. \end{aligned}$$

It suffices to verify Assumption 4.iv separately for $R_{1n}(\theta)$ and $R_{2n}(\theta)$. Since

$$R_{2n}(\theta) = -\frac{1}{2}n(\theta-\theta_0)'\left[J(\theta^*) - J(\theta_0)\right](\theta-\theta_0),$$

for some θ^* on the line connecting θ and θ_0, verification of Assumption 4 for $R_{2n}(\theta)$ follows immediately from continuity of $J(\theta)$ in θ over a ball at θ_0.

To show Assumption 4.iv-(b) for $R_{1n}(\theta)$, we note that for any given $M > 0$

$$\begin{aligned}
&\limsup_n P^*\left\{\sup_{|\theta-\theta_0|\le M/\sqrt{n}} |R_{1n}(\theta)| > \epsilon\right\} \\
&\le \limsup_n P\left\{\sup_{|\theta-\theta_0|\le M/\sqrt{n}} |\theta-\theta_0|\frac{|R_{1n}(\theta)|}{|\theta-\theta_0|} > \epsilon\right\} \\
&\le \limsup_n P\left\{\sup_{|\theta-\theta_0|\le M/\sqrt{n}} \frac{M}{\sqrt{n}}\frac{|R_{1n}(\theta)|}{|\theta-\theta_0|} > \epsilon\right\} = 0,
\end{aligned} \tag{A.28}$$

where the last conclusion follows from two observations.

First, note that

$$Z_n(\theta) \equiv \frac{R_{1n}(\theta)}{\sqrt{n}|\theta-\theta_0|} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{m_i(\theta) - m_i(\theta_0) - (Em_i(\theta) - Em_i(\theta_0)) - \dot{m}_i(\theta_0)'(\theta-\theta_0)}{|\theta-\theta_0|}\right)$$

is Donkser by assumption, that is it converges in $\ell^\infty(B_\delta(\theta_0))$ to a tight Gaussian process Z.

The process has uniformly continuous paths with respect to the semimetric ρ given by

$$\rho^2(\theta_1,\theta_2) = E(Z(\theta_1) - Z(\theta_2))^2,$$

so that $\rho(\theta,\theta) \to 0$ if $\theta \to \theta_0$. Thus almost all sample paths of Z are continuous at θ_0.

Second, since by assumption

$$E[\overline{m}_n(\theta) - \overline{m}_n(\theta_0) - \overline{\dot{m}_n}(\theta_0)'(\theta-\theta_0)]^2 = o(|\theta-\theta_0|^2),$$

we have for any $\theta_n \to \theta_0$

$$E^*\left[\frac{|R_{1n}(\theta_n)|}{\sqrt{n}|\theta_n-\theta_0|}\right]^2 = \frac{o\left(|\theta_n-\theta|^2\right)}{|\theta_n-\theta_0|^2} \to 0,$$

therefore

$$Z(\theta_0) = 0.$$

Therefore for any $\theta' \to_p \theta_0$, we have by the extended continuous mapping theorem

$$Z_n(\theta') \to_d Z(\theta_0) = 0, \text{ that is } Z_n(\theta') \to_p 0. \tag{A.29}$$

This shows (A.28).

To prove Assumption 4.iv-(a) for $R_{1n}(\theta)$, we need to show that for some $\delta > 0$ and constant M

$$\limsup_n P^*\left\{\sup_{M/\sqrt{n}\le|\theta-\theta_0|\le\delta} \frac{|R_{1n}(\theta)|}{n|\theta-\theta_0|^2} > \epsilon\right\} < \epsilon. \tag{A.30}$$

Using that $M/\sqrt{n} \leq |\theta - \theta_0|$, bound the left-hand-side by

$$\begin{aligned}
&\limsup_n P^* \left\{ \sup_{M/\sqrt{n} \leq |\theta - \theta_0| \leq \delta} \frac{|R_{1n}(\theta)|}{\sqrt{n}|\theta - \theta_0|} \cdot \frac{1}{\sqrt{n}|\theta - \theta_0|} > \epsilon \right\} \\
&\leq \limsup_n P^* \left\{ \sup_{M/\sqrt{n} \leq |\theta - \theta_0| \leq \delta} \frac{|R_{1n}(\theta)|}{\sqrt{n}|\theta - \theta_0|} \cdot \frac{1}{M} > \epsilon \right\} \\
&\leq \limsup_n P^* \left\{ \sup_{M/\sqrt{n} \leq |\theta - \theta_0| \leq \delta} |Z_n(\theta)| \cdot \frac{1}{M} > \epsilon \right\} \\
&< \epsilon,
\end{aligned} \tag{A.31}$$

where for any given $\epsilon > 0$ in order to make the last inequality true, we can make either δ sufficiently small by the property (A.29) of Z_n or make M sufficiently large by the property $Z_n = O_{P^*}(1)$. ■

B Appendix on Computation

B.1 A computational lemma

In this section we record some formal results on MCMC computation of the quasi-posterior quantities.

LEMMA 3 *Suppose the chain $(\theta^{(j)}, j \leq B)$ is produced by the Metropolis Hastings(MH) algorithm with q such that $q(\theta|\theta') > 0$ for each (θ, θ'). Suppose also that $P\{\rho(\theta^{(j)}, \xi) = 1\} > 0$ for all $j > t_0$. Then*

1. *$p_n(\cdot)$ is the stationary density of the chain,*
2. *the chain is ergodic with the limit marginal distribution given by $p_n(\cdot)$:*

$$\lim_{B \to \infty} \sup_A \left| P(\theta^{(B)} \in A|\theta_0) - \int_A p_n(\theta) d\theta \right| = 0,$$

   where the supremum is taken over the Borel sets,

3. *For any p_n- integrable function g:*

$$\frac{1}{B} \sum_{j=1}^{B} g(\theta^{(j)}) \to_p \int_\Theta g(\theta) p_n(\theta) d\theta.$$

Proof. The result is immediate from Theorem 6.2.5 in Robert and Casella (1999). ■

An immediate consequence of this lemma is the following result.

LEMMA 4 *Suppose Assumptions 1 and 2 hold. Suppose the chain $(\theta^{(j)}, j \leq B)$ satisfies the conditions of Lemma 3, then for any convex and p_n-integrable loss function $\rho_n(\cdot)$*

$$\arg \inf_{\theta \in \Theta} \left[\frac{1}{B} \sum_{j=1}^{B} \rho_n(\theta^{(j)} - \theta) \right] \to_p \widehat{\theta} = \arg \inf_{\theta \in \Theta} \left[\int_\Theta \rho_n(\tilde{\theta} - \theta) p_n(\tilde{\theta}) d\tilde{\theta} \right],$$

provided that $\widehat{\theta}$ is uniquely defined.

Proof. By Lemma 3 we have the pointwise convergence of the objective function: for any θ

$$\frac{1}{B}\sum_{j=1}^{B}\rho_n(\theta^{(j)}-\theta)\rightarrow_p \int_{\Theta}\rho_n(\tilde{\theta}-\theta)p_n(\tilde{\theta})d\tilde{\theta},$$

which implies the result by the Convexity Lemma, since $\theta \mapsto \int_{\Theta}\rho_n(\tilde{\theta}-\theta)p_n(\tilde{\theta})d\tilde{\theta}$ is convex by convexity of ρ_n. ∎

B.2 Quasi-Bayes Estimation and Simulated Annealing

The relation between drawing from the shape of a likelihood surface and optimizing to find the mode of the likelihood function is well known. It is well established that, e.g. Robert and Casella (1999),

$$\lim_{\lambda\rightarrow\infty}\frac{\int_{\Theta}\theta e^{\lambda L_n(\theta)}\pi(\theta)\,d\theta}{\int_{\Theta}e^{\lambda L_n(\theta)}\pi(\theta)\,d\theta}=\underset{\theta\in\Theta}{\operatorname{argmax}}L_n(\theta) \tag{B.1}$$

Essentially, as $\lambda \rightarrow \infty$, the sequence of probability measures

$$\frac{e^{\lambda L_n(\theta)}\pi(\theta)}{\int_{\Theta}e^{\lambda L_n(\theta)}\pi(\theta)\,d\theta} \tag{B.2}$$

converges to the generalized Dirac probability measure concentrated at $\underset{\theta\in\Theta}{\operatorname{argmax}}\, L_n(\theta)$.

The difficulty of nonlinear optimization has been an important issue in econometrics (Berndt et al. (1974), Sims (1999)). The simulated annealing algorithm (see e.g. Press et al. (1992), Goffe et al. (1994)) is usually considered a generic optimization method. It is an implementation of the simulation based optimization (B.1) with a uniform prior $\pi(\theta) \equiv c$ on the parameter space Θ. At each temperature level $1/\lambda$, the simulated annealing routine uses a large number of Metropolis-Hastings steps to draw from the quasi distribution (B.2). The temperature parameter is then decreased slowly while the Metropolis steps are repeated, until convergence criteria for the optimum are achieved.

Interestingly, the simulated annealing algorithm has been widely used in optimization of non-likelihood-based semiparametric objective functions. In principle, if the temperature parameter is decreased at an arbitrarily slow rate (that depends on the criterion function), simulated annealing can find the global optimum of non-smooth objective functions that may have many local extrema. Controlling the temperature reduction parameter is a very delicate matter and is certainly crucial to the performance of the algorithm with highly nonsmooth objective functions. On the other hand, as Theorems 1 and 2 apply equally to (B.2), the results of this paper show that we may fix the temperature parameter $1/\lambda$ at a positive constant and then compute the quasi-posterior medians or means for (B.2) using Metropolis steps. These estimates can be used in place of the exact maximum. They are consistent and asymptotically normal, and possess the same limiting distribution as the exact maximum. The interpretation of the simulated annealing algorithm as an implementation of (B.2) also suggests that for some problems with special structures, other MCMC methods, such as the Gibbs sampler, may be used to replace the Metropolis-Hasting step in the simulated annealing algorithm.

B.3 Details of Computation in Monte-Carlo Examples

The parameter space is taken to be $\Theta = [\theta_0 \pm 10]$. The transition kernel is a Normal density, and flat prior is truncated to Θ. Each parameter is updated via a Gibbs-Metropolis procedure, which modifies slightly the basic Metropolis-Hastings algorithm: for $k = 1, ..., d$, a draw of ξ_k from the univariate normal density $q(|\xi_k - \theta_k^{(j)}|, \phi)$ is made, then the candidate value ξ consisting of ξ_k and $\theta_{-k}^{(j)}$ replaces $\theta^{(j)}$ with probability $\rho(\theta^{(j)}, \xi)$ specified in the text. Variance parameter ϕ is adjusted every 100 draws (in the second simulation example and empirical example) or 200 draws (in the first simulation example) so that the rejection probability is roughly 50%.

The first $N \times d$ draws (the burn-in stage) are discarded, and the remaining $N \times d$ draws are used in computation of estimates and intervals. The starting value is the OLS estimate in all examples. We use $N = 5,000$ in the second simulation example and empirical example and $N = 10,000$ in the second simulation example. To give an idea of computational expense, computing one set of estimates takes 20-40 seconds depending on the example. All of the codes that we used to produce figures, simulation, and empirical results are available from the authors.

Notation and Terms

$\to_p$	convergence in (outer) probability P^*
$\to_d$	convergence in distribution under P^*
wp $\to 1$	with inner probability P_* converging to one
$\sim$	asymptotic equivalence denoted $A \sim B$ means $\lim AB^{-1} = I$
$B_\delta(x)$	ball centered at x of radius $\delta > 0$
I	identity matrix
$A > 0$	A is positive definite when A is matrix
$\mathcal{N}(0, a)$	normal random vector with mean 0 and variance matrix a
$\mathcal{F}$ Donsker class	here this means that empirical process $f \mapsto \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(W_i) - Ef(W_i))$ is asymptotically Gaussian in $\ell^\infty(\mathcal{F})$, see van der Vaart (1999)
$\ell^\infty(\mathcal{F})$	metric space of bounded over $\mathcal{F}$ functions, see van der Vaart (1999)
mineig(A)	minimum eigenvalue of matrix A

References

Abadie, A., 1995. Changes in spanish labor income structure during the 1980s: a quantile regression approach, CEMFI Working Paper.

Amemiya, T., 1977. The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. Econometrica 45 (4), 955–968.

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press.

Anderson, T. W., 1955. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proc. Amer. Math. Soc. 6, 170–176.

Andrews, D. W. K., 1994a. Empirical process methods in econometrics. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. North Holland, pp. 2248–2292.

Andrews, D. W. K., 1994b. The large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests. Econometrica 62 (5), 1207–1232.

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

Andrews, D. W. K., 1999. Estimation when a parameter is on a boundary. Econometrica 66, 1341–83.

Berger, J. O., 2002. Bayesian analysis: a look at today and thoughts of tomorrow. In: Statistics in the 21st Century. Chapman and Hall, New York, pp. 275–290.

Berndt, E., Hall, B., Hall, R., Hausman, J., 1974. Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement 3 (4), 653–665.

Bernstein, S., 1917. Theory of Probability. (Russian) Fourth Edition (1946) Gostekhizdat, Moscow-Leningrad.

Berry, S., Levinsohn, J., Pakes, A., July 1995. Automobile prices in market equilibrium. Econometrica 63, 841–890.

Bickel, P. J., Yahav, J. A., 1969. Some contributions to the asymptotic theory of Bayes solutions. Z. Wahrsch. Verw. Geb 11, 257–276.

Billingsley, P., 1994. Probability and Measure, 3rd Ed. John Wiley and Sons.

Buchinsky, M., 1991. Theory of and practice of quantile regression, Ph.D. dissertation, Department of Economics Harvard University.

Buchinsky, M., Hahn, J., 1998. An alternative estimator for the censored regression model. Econometrica 66, 653–671.

Bunke, O., Milhaud, X., 1998. Asymptotic behavior of Bayes estimates under possibly incorrect models. The Annals of Statistics 26 (2), 617–644.

Chamberlain, G., Imbens, G., 1997. Nonparametric appliations of Bayesian inference, NBER Working Paper.

Chernozhukov, V., Hansen, C., 2001. An IV model of quantile treatment effects, MIT Department of Economics Working Paper.

Chernozhukov, V., Umantsev, L., 2001. Conditional value-at-risk: Aspects of modeling and estimation. Empirical Economics 26, 271–92.

Chib, S., 2001. Markov chain monte carlo methods: Computation and inference. In: J.J.Heckman, Leamer, E. (Eds.), Handbook of Econometrics, Vol 5, Chapter 5. North Holland, pp. 3564–3634.

Christoffersen, P., Hahn, J., Inoue, A., 1999. Testing, comparing and combining value at risk measures, working Paper, Wharton School University of Pennsylvania.

Diaconis, P., Freedman, D., 1986. On the consistency of Bayes estimates. Annals of Statistics 14, 1–26.

Doksum, K. A., Lo, A. Y., 1990. Consistent and robust Bayes procedures for location based on partial information. Ann. Statist. 18 (1), 443–453.

Engle, R., Manganelli, S., 2001. Caviar: Conditional value at risk by regression quantiles, Working Paper, Department of Economics UC San Diego.

Fitzenberger, B., 1997. A guide to censored quantile regressions. In: Robust inference, Handbook of Statistics. Vol. 14. North-Holland, Amsterdam, pp. 405–437.

Gallant, A. R., White, H., 1988. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Oxford: Bail Blackwell.

Geweke, J., Keane, M., 2001. Computationally intensive methods for integration in econometrics. In: J.J.Heckman, Leamer, E. (Eds.), Handbook of Econometrics, Vol 5, Chapter 5. North Holland, pp. 3465–3564.

Goffe, W. L., Ferrier, G. D., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. Journal of Econometrics 60, 65–99.

Hahn, J., 1997. Bayesian bootstrap of the quantile regression estimator: a large sample study. Internat. Econom. Rev. 38 (4), 795–808.

Hansen, L., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.

Hogg, R. V., 1975. Estimates of percentile regression lines using salary data. Journal of American Statistical Association 70, 56–59.

Huber, P. J., 1973. Robust regression: Asymptotics, conjectures, and monte carlo. Annals of Statistics 1, 799–821.

Ibragimov, I., Has'minskii, R., 1981. Statistical Estimation: Asymptotic Theory. Springer Verlag.

Imbens, G., 1997. One-step estimators for over-identified generalized method of moments models. Rev. Econom. Stud. 64 (3), 359–383.

Imbens, G., Spady, R., Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66, 333–357.

Jureckova, J., 1977. Asymptotic relations of M-estimators and R-estimators in linear regression models. Annals of Statistics 5, 464–472.

Khan, S., Powell, J. L., 2001. Two step estimation of semiparametric censored regression models. Journal of Econometrics 103, 73–110.

Kim, J.-Y., 1998. Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models. Econometrica 66 (2), 359–380.

Kim, J.-Y., 2002. Limited information likelihood and Bayesian analysis. Journal of Econometrics , 175-193.

Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes. Ann. Statist. 25 (5), 2084–2102.

Kitamura, Y., Stutzer, M., 1997. An information-theoretic alternative to generalized method of moments estimation. Econometrica 65, 861–874.

Knight, K., 1999. Epi-convergence and stochastic equisemicontinuity, Working Paper, Department of Statistics University of Toronto.

Koenker, R., 1994. Confidence intevals for quantile regression. In: Proceedings of the 5th Prague Symposium on Asymptotic Statistics. Heidelberg: Physica-Verlag, pp. 10–20.

Koenker, R., 1998. Treating the treated, varieties of causal analysis, Lecture Note, Depratment of Economics University of Illinois.

Koenker, R., Bassett, G. S., 1978. Regression quantiles. Econometrica 46, 33–50.

Kottas, A., Gelfand, A., 2001. Bayesian semiparametric median regression modeling. Journal of the American Statistical Association 96, 1458–1468.

Lehmann, E., Casella, G., 1998. Theory of Point Estimation. Springer.

Macurdy, T., Timmins, C., 2001. Bounding the influence of attrition on the intertemporal wage variation in the NLSY, Working Paper, Department of Economics Yale University.

Mood, A. M., 1950. Introduction to the Theory of Statistics. McGraw-Hill Book Company, Inc.

Newey, W. K., 1991. Uniform convergence in probability and stochastic equicontinuity. Econometrica 59 (4), 1161–1167.

Newey, W. K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. North Holland, pp. 2113–2241.

Newey, W. K., Powell, J. L., 1990. Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. Econometric Theory 6, 295–317.

Newey, W. K., Smith, R., 2001. Higher order properties of GMM and generalized empirical likeliood estimators, Working Paper, Department of Economics MIT.

Newey, W. K., West, K. D., 1987. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.

Owen, A., 1989. Empirical likelihood ratio confidence regions. In: Proceedings of the 47th Session of the International Statistical Institute, Book 3 (Paris, 1989). Vol. 53. pp. 373–393.

Owen, A., 1990. Empirical likelihood ratio confidence regions. Ann. Statist. 18 (1), 90–120.

Owen, A., 1991. Empirical likelihood for linear models. Ann. Statist. 19 (4), 1725–1747.

Owen, A., 2001. Empirical Likelihood. Chapman and Hall/CRC.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57 (5), 1027–1057.

Phillips, P. C. B., Ploberger, W., 1996. An asymptotic theory of Bayesian inference for time series. Econometrica 64 (2), 381–412.

Pollard, D., 1991. Asymptotics for least absolute deviation regression estimator. Econometric Theory 7, 186–199.

Pötscher, B. M., Prucha, I. R., 1997. Dynamic Nonlinear Econometric Models. Springer-Verlag, Berlin.

Powell, J. L., 1984. Least absolute deviations estimation for the censored regression model. Journal of Econometrics , 303–325.

Press, W., Teukolsky, S. A., Vettering, W., Flannery, B., 1992. Numerical Recipes in C, The Art of Scientific Computing. Cambridge.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Annals of Statistics 22, 300–325.

Robert, C. P., Casella, G., 1999. Monte Carlo Statistical Methods. Springer.

Rousseeuw, P. J., Hubert, M., 1999. Regression depth. J. Amer. Statist. Assoc. 94 (446), 388–433.

Sims, C., 1999. Adaptive Metropolis-Hastings, or monte carlo kernel estimation, Working Paper, Department of Economics Princeton University.

Stigler, S. M., 1975. Studies in the history of probability and statistics. XXXIV. Napoleonic statistics: the work of Laplace. Biometrika 62 (2), 503–517.

van Aelst, S., Rousseeuw, P. J., Hubert, M., Struyf, A., 2002. The deepest regression method. J. Multivariate Anal. 81 (1), 138–166.

van der Vaart, A., 1999. Asymptotic Statistics. Cambridge University Press.

van der Vaart, A. W., Wellner, J. A., 1996. Weak Convergence and Empirical Processes. Springer-Verlag, New York.

von Mises, R., 1931. Wahrscheinlichkeitsrechnung. Berlin: Springer.

White, H., 1994. Estimation, Inference and Specification Analysis. Vol. 22 of Econometric Society Monographs. Cambridge University Press, Cambridge.

Zellner, A., 1998. Past and recent results on maximal data information priors. J. Statist. Res. 32 (1), 1–22.

Table 1: Monte Carlo Comparison of LTE's with Censored Quantile Regression Estimates Obtained using Iterated Linear Programming (Based on 100 repetitions).

Estimator	RMSE	MAD	Mean Bias	Median Bias	Median Abs. Dev.
n=400					
Q-posterior-mean	0.473	0.378	0.138	0.134	0.340
Q-posterior-median	0.465	0.372	0.131	0.137	0.344
Iterated LP(10)	0.518	0.284	0.040	0.016	0.171
	3.798	0.827	-0.568	-0.035	0.240
n=1600					
Q-posterior-mean	0.155	0.121	-0.018	0.009	0.089
Q-posterior-median	0.155	0.121	-0.020	0.002	0.092
Iterated LP(7)	0.134	0.106	0.040	0.067	0.085
	3.547	0.511	0.023	-0.384	0.087

Table 2: Monte Carlo Comparison of the LTE's with Standard Estimation for a Linear Quantile Regression Model (Based on 500 repetitions)

Estimator	RMSE	MAD	Mean Bias	Median Bias	Median AD
n=200					
Q-posterior-mean	.0747	.0587	.0174	.0204	.0478
Q-posterior-median	.0779	.0608	.0192	.0136	.0519
Standard Quantile Regression	.0787	.0628	.0067	.0092	.0510
n=800					
Q-posterior-mean	.0425	.0323	-.0018	-.0003	.0280
Q-posterior-median	.0445	.0339	-.0023	.0001	.0295
Standard Quantile Regression	.0498	.0398	.0007	.0025	.0356

Table 3: Monte Carlo Comparison of the LT Inference with Standard Inference for a Linear Quantile Regression Model (Based on 500 repetitions)

Inference Method	**coverage**	**length**
n=200		
Quasi-posterior confidence interval, equal tailed	.943	.377
Quasi-posterior confidence interval, symmetric (around mean)	.941	.375
Quantile Regression: Hall-Sheather Interval	.659	.177
n=800		
Quasi-posterior confidence interval, equal tailed	.920	.159
Quasi-posterior confidence interval, symmetric (around mean)	.917	.158
Quantile Regression: Hall-Sheather Interval	.602	.082

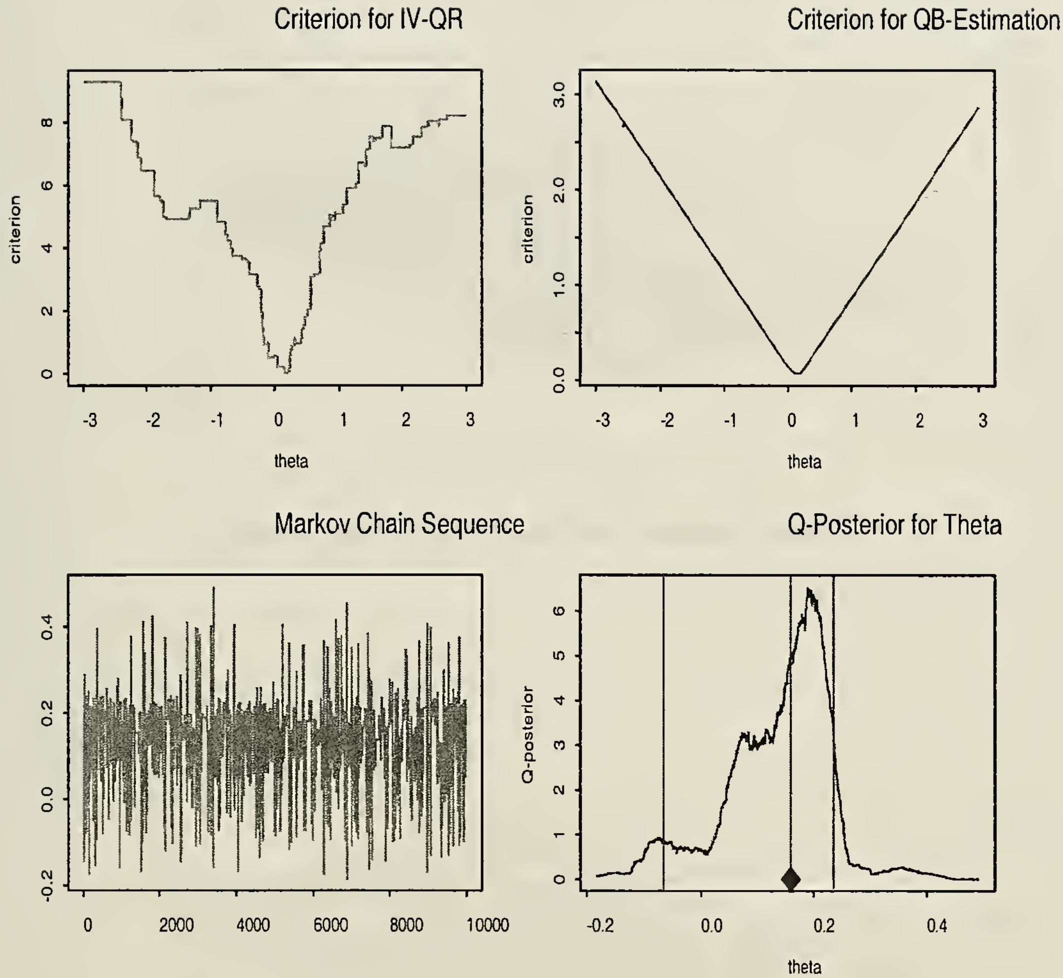

Figure 1: A Nonlinear IV Example involving Instrumental Quantile Regression. In the top-left panel the discontinuous objective function $L_n(\theta)$ is depicted (one-dimensional case). The true parameter $\theta_0 = 0$. In the bottom-left panel, a Markov Chain sequence of draws $(\theta^{(1)}, \ldots \theta^{(J)})$ is depicted. The marginal distribution of this sequence is $p_n(\theta) = e^{L_n(\theta)} / \int_\Theta e^{L_n(\theta)} d\theta$, see the bottom-right panel. The point estimate, the sample mean $\bar{\theta}$, is given by the vertical line with the romboid root. Two other vertical lines are the 10-th and the 90-th percentiles of quasi-posterior distribution. The upper-right panel depicts the expected loss function that the LTE minimize.

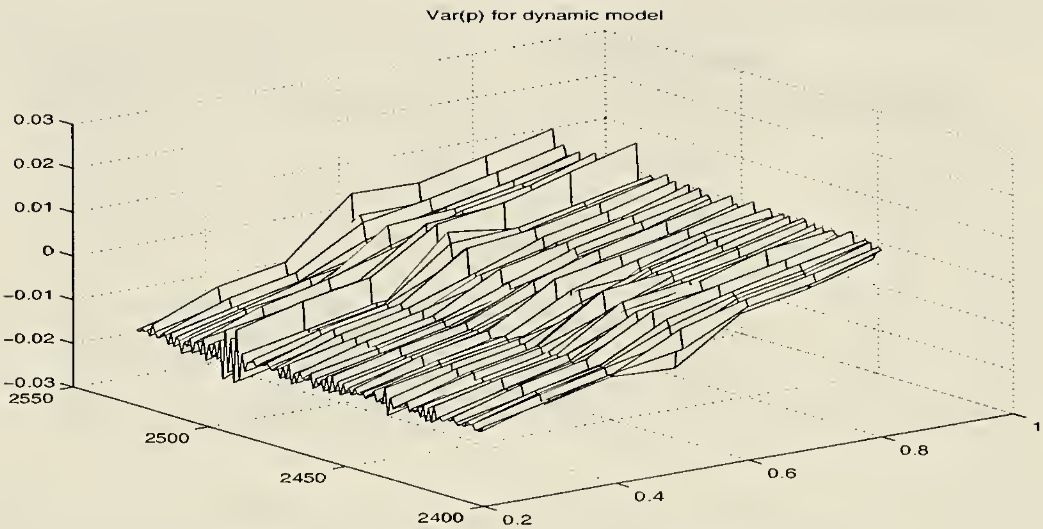

Figure 2: Recursive VaR Surface in time-probability space

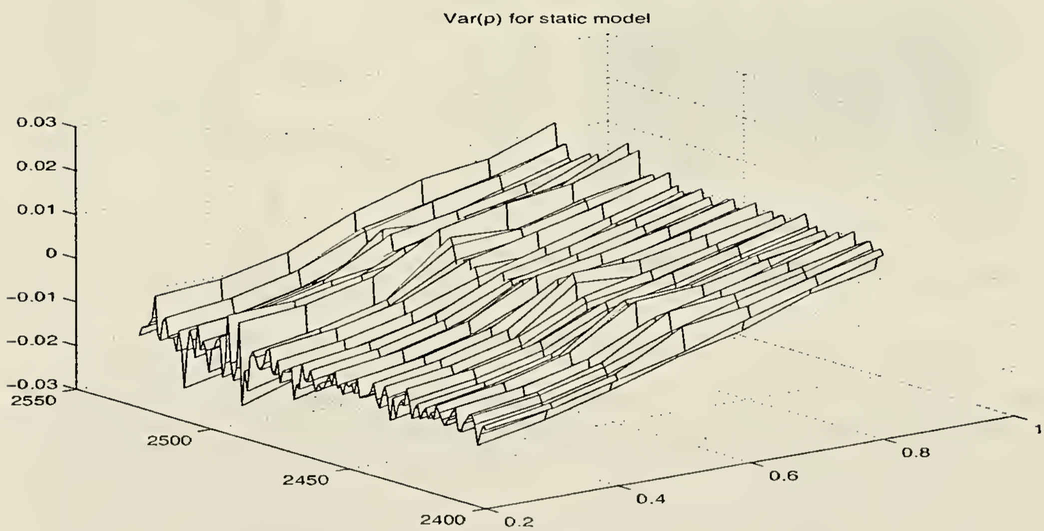

Figure 3: Non-recursive VaR Surface in time-probability space

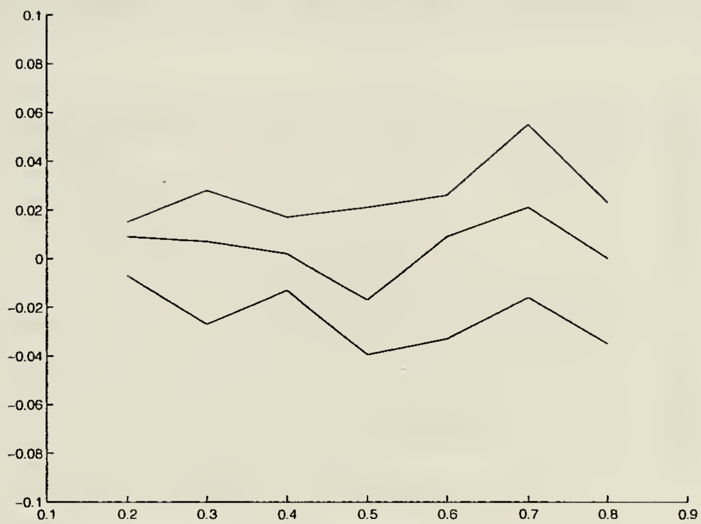

Figure 4: $\widehat{\theta}_2(\tau)$ for $\tau \in [.2, .8]$ and the 90% confidence intervals.

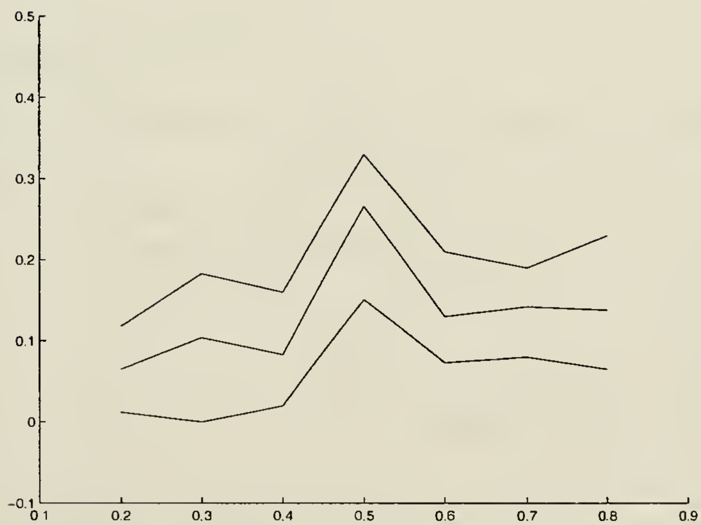

Figure 5: $\widehat{\theta}_3(\tau)$ for $\tau \in [.2, .8]$ and the 90% confidence intervals.

Figure 6: $\widehat{\theta}_4(\tau)$ for $\tau \in [.2, .8]$ and the 90% confidence intervals.

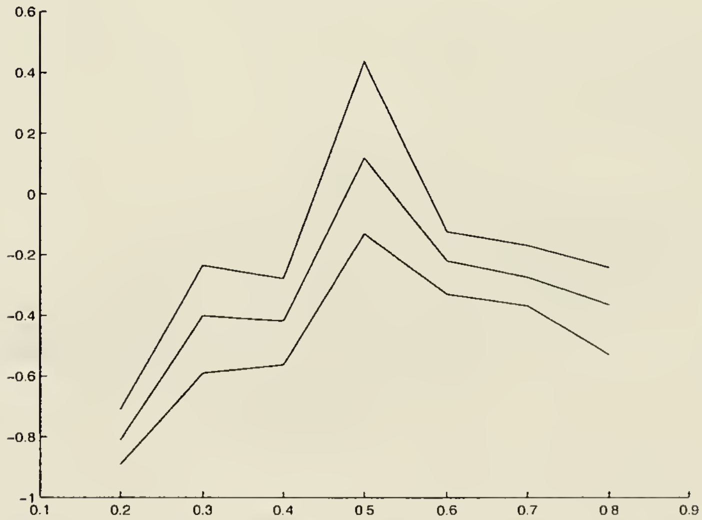

Figure 7: $\widehat{\varrho}(\tau)$ for $\tau \in [.2, .8]$ and the 90% confidence intervals.

Throughout this proof the range of integration for h is implicitly understood to be H_n. For clarity of the argument, we limit exposition only to the case where $J_n(\theta)$ and $\Omega_n(\theta)$ do not depend on n. The more general case follows similarly.

Part 1. Define

$$h \equiv \sqrt{n}(\theta - T_n)\,,\ T_n = \theta_0 + \frac{1}{n} J(\theta_0)^{-1}\Delta_n(\theta_0)\,,\ U_n \equiv \frac{1}{\sqrt{n}} J(\theta_0)^{-1}\Delta_n(\theta_0), \tag{A.2}$$

then

$$\begin{aligned} p_n^*(h) \equiv& \frac{1}{\sqrt{n}} p_n\left(h/\sqrt{n} + \theta_0 + U_n/\sqrt{n}\right) \\ =& \frac{\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp\left(L_n\left(\frac{h}{\sqrt{n}} + T_n\right)\right)}{\int_{H_n}\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp\left(L_n\left(\frac{h}{\sqrt{n}} + T_n\right)\right)dh} \\ =& \frac{\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp(\omega(h))}{\int_{H_n}\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp(\omega(h))\,dh} \\ =& \frac{\pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp(\omega(h))}{C_n}, \end{aligned}$$

where

$$\omega(h) \equiv L_n\left(T_n + \frac{h}{\sqrt{n}}\right) - L(\theta_0) - \frac{1}{2n}\Delta_n(\theta_0)' J(\theta_0)^{-1}\Delta_n(\theta_0)\,, \tag{A.3}$$

and

$$C_n \equiv \int_{H_n} \pi\left(\frac{h}{\sqrt{n}} + T_n\right)\exp(\omega(h))\,dh.$$

Part 2 shows that for each $\alpha \geq 0$,

$$A_{1n} \equiv \int_{H_n} |h|^{\alpha}\left|\exp(w(h))\pi\left(T_n + \frac{h}{\sqrt{n}}\right) - \exp\left(-\frac{1}{2}h' J(\theta_0)\,h\right)\pi(\theta_0)\right| dh \xrightarrow{p} 0. \tag{A.4}$$

Given (A.4), taking $\alpha = 0$ we have

$$C_n \to_p \int_{\mathbf{R}^d} e^{-\frac{1}{2}h' J(\theta_0) h}\pi(\theta_0)\,dh = \pi(\theta_0)\,(2\pi)^{\frac{d}{2}}|\det J(\theta_0)|^{-1/2}, \tag{A.5}$$

hence

$$C_n = O_p(1).$$

Next note

$$\text{left side of (A.1)} \equiv \int_{H_n} |h|^{\alpha}\left|p_n^*(h) - p_\infty^*(h)\right| dh = A_n \cdot C_n^{-1},$$

where

$$A_n \equiv \int_{H_n} |h|^{\alpha}\left|e^{w(h)}\pi\left(T_n + \frac{h}{\sqrt{n}}\right) - (2\pi)^{-d/2}|\det J(\theta_0)|^{1/2}\exp\left(-\frac{1}{2}h' J(\theta_0)\,h\right)\cdot C_n\right| dh.$$

Using (A.5), to show (A.1) it suffices to show that $A_n \xrightarrow{p} 0$. But

$$A_n \leq A_{1n} + A_{2n}$$

where

$$A_{2n} \equiv \int_{H_n} |h|^{\alpha} \left| C_n (2\pi)^{-d/2} |\det J(\theta_0)|^{1/2} \exp\left(-\frac{1}{2}h' J(\theta_0) h\right) - \pi(\theta_0) \exp\left(-\frac{1}{2}h' J(\theta_0) h\right) \right| dh.$$

Then by (A.4)

$$A_{1n} \xrightarrow{p} 0,$$

and

$$A_{2n} = \left| C_n (2\pi)^{-d/2} |\det J(\theta_0)|^{1/2} - \pi(\theta_0) \right| \int_{H_n} |h|^{\alpha} \exp\left(-\frac{1}{2}h' J(\theta_0) h\right) dh$$
$$\xrightarrow{p} 0.$$

Part 2. It remains only to show (A.4). Given Assumption 4 and definitions in (A.2) and (A.3), write

$$\begin{aligned} \omega(h) &= \Delta_n(\theta_0)' \left(\frac{U_n + h}{\sqrt{n}}\right) - \frac{1}{2} n \left(\frac{U_n + h}{\sqrt{n}}\right)' J(\theta_0) \left(\frac{U_n + h}{\sqrt{n}}\right) \\ &- \frac{1}{2n} \Delta_n(\theta_0)' J(\theta_0)^{-1} \Delta_n(\theta_0) + R_n\left(\frac{h}{\sqrt{n}} + T_n\right) \\ &= -\frac{1}{2} h' J(\theta_0) h + R_n\left(\frac{h}{\sqrt{n}} + T_n\right). \end{aligned}$$

Split the integral A_{1n} in (A.4) over three separate areas:

- Area (i) : $|h| \le M$,
- Area (ii) : $M \le |h| \le \delta\sqrt{n}$,
- Area (iii) : $|h| \ge \delta\sqrt{n}$.

Each of these areas is implicitly understood to intersect with the range of integration for h, which is H_n.

Area (i): We will show that for each $0 < M < \infty$ and each $\epsilon > 0$

$$\liminf_n P_* \left\{ \int_{|h| \le M} |h|^{\alpha} \left| \exp(\omega(h)) \pi\left(T_n + \frac{h}{\sqrt{n}}\right) - \exp\left(-\frac{1}{2}h' J(\theta_0) h\right) \pi(\theta_0) \right| dh < \epsilon \right\} \ge 1 - \epsilon. \tag{A.6}$$

This is proved by showing that

$$\sup_{|h| \le M} |h|^{\alpha} \left| \exp(\omega(h)) \pi\left(T_n + \frac{h}{\sqrt{n}}\right) - \exp\left(-\frac{1}{2}h' J(\theta_0) h\right) \pi(\theta_0) \right| \xrightarrow{p} 0. \tag{A.7}$$

Using the definition of $\omega(h)$, (A.7) follows from:

$$(a) \sup_{|h| \le M} \left| \pi\left(\frac{h}{\sqrt{n}} + T_n\right) - \pi(\theta_0) \right| \xrightarrow{p} 0, \quad (b) \sup_{|h| \le M} \left| R_n\left(\frac{h}{\sqrt{n}} + T_n\right) \right| \xrightarrow{p} 0,$$

where (a) follows from the continuity of $\pi(\cdot)$ and because by Assumption 4.ii-4.iii:

$$\frac{1}{\sqrt{n}} J(\theta_0)^{-1} \Delta_n(\theta_0) = O_p(1). \tag{A.8}$$

Given (A.8), (b) follows from Assumption 4.iv, since

$$\sup_{|h|\leq M}\left|T_n+\frac{h}{\sqrt{n}}-\theta_0\right|=O_p(1/\sqrt{n}).$$

Area (ii): We show that for each $\epsilon>0$ there exist large M and small $\delta>0$ such that

$$\begin{aligned}\liminf_n P_*\Bigg\{\int_{M<|h|<\delta\sqrt{n}}|h|^{\alpha}\Bigg|&\exp\left(w\left(h\right)\right)\pi\left(T_n+\frac{h}{\sqrt{n}}\right)\\&-\exp\left(-\frac{1}{2}h'J\left(\theta_0\right)h\right)\pi\left(\theta_0\right)\Bigg|dh<\epsilon\Bigg\}\geq 1-\epsilon.\end{aligned}\tag{A.9}$$

Since the integral of the second term is finite and can be made arbitrarily small by setting M large, it suffices to show that for each $\epsilon>0$ there exist large M and small $\delta>0$ such that

$$\liminf_n P_*\left\{\int_{M<|h|<\delta\sqrt{n}}|h|^{\alpha}\left|\exp\left(w\left(h\right)\right)\pi\left(T_n+\frac{h}{\sqrt{n}}\right)\right|dh<\epsilon\right\}\geq 1-\epsilon.\tag{A.10}$$

In order to do so, it suffices to show that for sufficiently large M as $n\to\infty$

$$\exp\left(w\left(h\right)\right)\pi\left(T_n+\frac{h}{\sqrt{n}}\right)\leq C\exp\left(-\frac{1}{4}h'J\left(\theta_0\right)h\right),\ \text{ for all } M<|h|<\delta\sqrt{n}.\tag{A.11}$$

By assumption $\pi\left(\cdot\right)<K$, so we can drop it from consideration.

By definition of $\omega(h)$

$$\exp\left(w\left(h\right)\right)\leq\exp\left(-\frac{1}{2}h'J\left(\theta_0\right)h+\left|R_n\left(T_n+\frac{h}{\sqrt{n}}\right)\right|\right).$$

Since $|T_n-\theta_0|=o_p(1)$, for any $\delta>0$ wp $\to 1$

$$\left|T_n+\frac{h}{\sqrt{n}}-\theta_0\right|<2\delta,\quad\text{for all}\quad |h|\leq\delta\sqrt{n}.$$

Thus, by Assumption 4.iv(a) there exists some small δ and large M such that

$$\liminf_n P_*\left\{\sup_{M\leq|h|\leq\delta\sqrt{n}}\frac{|R_n\left(T_n+\frac{h}{\sqrt{n}}\right)|}{|h+\frac{1}{\sqrt{n}}J\left(\theta_0\right)^{-1}\Delta_n\left(\theta_0\right)|^2}\leq\frac{1}{4}\text{mineig}\left(J\left(\theta_0\right)\right)\right\}\geq 1-\epsilon.$$

Since $\frac{1}{n}|J\left(\theta_0\right)^{-1}\Delta_n\left(\theta_0\right)|^2=O_p\left(1\right)$, for some $C>0$

$$\begin{aligned}&\liminf_n P_*\left\{\exp\left(w\left(h\right)\right)\leq C\exp\left(-\frac{1}{4}h'J\left(\theta_0\right)h\right)\right\}\\&\geq\liminf_n P_*\left\{e^{w(h)}\leq C\exp\left(-\frac{1}{2}h'J\left(\theta_0\right)h+\frac{1}{4}\text{mineig}\left(J\left(\theta_0\right)\right)|h|^2\right)\right\}\\&\geq 1-\epsilon.\end{aligned}\tag{A.12}$$

(A.12) implies (A.11), which in turn implies (A.9).

Area (iii): We will show that for each $\epsilon > 0$ and each $\delta > 0$,

$$\liminf_n P_* \left\{ \int_{h \geq \delta\sqrt{n}} |h|^\alpha \left| \exp(w(h))\, \pi\left(T_n + \frac{h}{\sqrt{n}}\right) - \exp\left(-\frac{1}{2}h' J(\theta_0)\, h\right) \pi(\theta_0) \right| dh < \epsilon \right\} \geq 1 - \epsilon. \tag{A.13}$$

The integral of the second term clearly goes to 0 as $n \to \infty$. Therefore we only need to show

$$\int_{|h| \geq \delta\sqrt{n}} |h|^\alpha e^{\omega(h)} \pi\left(T_n + \frac{h}{\sqrt{n}}\right) dh \to_p 0.$$

Recalling the definition of h, the term is bounded by

$$\sqrt{n}^{\alpha+1} \int_{|\theta - T_n| \geq \delta} \left| T_n - \theta \right|^\alpha \pi(\theta) \exp\left(L_n(\theta) - L_n(\theta_0) - \frac{1}{2n} \Delta_n(\theta_0)' J(\theta_0)^{-1} \Delta_n(\theta_0) \right) d\theta.$$

Since $T_n - \theta_0 \xrightarrow{p} 0$, wp $\to 1$ this is bounded by

$$K_n \cdot C \cdot \sqrt{n}^{\alpha+1} \int_{|\theta - \theta_0| \geq \delta/2} \left(1 + |\theta|^\alpha\right) \pi(\theta) \exp\left(L_n(\theta) - L_n(\theta_0)\right) d\theta,$$

where

$$K_n = \exp\left(-\frac{1}{2n} \Delta_n(\theta_0)' J(\theta_0)^{-1} \Delta_n(\theta_0) \right) = O_p(1).$$

By Assumption 3 there exists $\varepsilon > 0$ such that

$$\liminf_{n \to \infty} P_* \left\{ \sup_{|\theta - \theta_0| \geq \delta/2} e^{L_n(\theta) - L_n(\theta_0)} \leq e^{-n\varepsilon} \right\} = 1$$

Thus, wp $\to 1$ the entire term is bounded by

$$K_n \cdot C \cdot \sqrt{n}^{\alpha+1} \cdot e^{-n\epsilon} \int_\Theta |\theta|^\alpha \pi(\theta)\, d\theta = o_p(1). \tag{A.14}$$

Here observe that compactness is only used to insure that

$$\int_\Theta |\theta|^\alpha \pi(\theta)\, d\theta < \infty. \tag{A.15}$$

Hence by replacing compactness with the condition (A.15), the conclusion (A.14) is not affected for the given α.

The entire proof is now completed by combining (A.6), (A.9), and (A.13). ■

A.2 Proof of Theorem 2

For clarity of the argument, we limit exposition only to the case where $J_n(\theta)$ and $\Omega_n(\theta)$ do not depend on n. The more general case follows similarly.

Recall that

$$h = \sqrt{n}(\theta - \theta_0) - J(\theta_0)^{-1} \Delta_n(\theta_0) / \sqrt{n}.$$

Define $U_n = J(\theta_0)^{-1}\Delta_n(\theta_0)/\sqrt{n}$. Consider the objective function

$$Q_n(z) = \int_{H_n} \rho(z - h - U_n)p_n^*(h)dh,$$

which is minimized at $\sqrt{n}(\widehat{\theta} - \theta_0)$. Also define

$$Q_\infty(z) = \int_{\mathbf{R}^d} \rho(z - h - U_n)p_\infty^*(h)dh.$$

which is minimized at a random vector denoted Z_n. Define

$$\xi = \arg \inf_{z\in\mathbf{R}^d} \left\{ \int_{\mathbf{R}^d} \rho(z-h)\ p_\infty^*(h)\ dh \right\}.$$

Note that solution is unique and finite by Assumption 2 parts (ii) and (iii) on the loss function ρ. When ρ is symmetric, $\xi = 0$ by Anderson's lemma.

Therefore, $Z_n = \arg\inf_{z\in\mathbf{R}^d} Q_\infty(z)$ equals

$$Z_n = \xi + U_n = O_p(1).$$

Next, we have for any fixed z

$$Q_n(z) - Q_\infty(z) \rightarrow_p 0$$

since by Assumption 2.ii $\rho(h) \leq 1 + |h|^p$ and by $|a+b|^p \leq 2^{p-1}|a|^p + 2^{p-1}|b|^p$ for $p \geq 1$:

$$\begin{aligned}
|Q_n(z) - Q_\infty(z)| &\leq \int_{H_n} (1 + |z-h-U_n|^p)\,|p_n^*(h) - p_\infty^*(h)|dh \\
&+ \int_{H_n^c} (1 + |z-h-U_n|^p)\,(p_\infty^*(h))\,dh \\
&\leq \int_{H_n} \left(1 + 2^{p-1}|h|^p + 2^{p-1}|z-U_n|^p\right) |p_n^*(h) - p_\infty^*(h)|dh \\
&+ \int_{H_n^c} \left(1 + 2^{p-1}|h|^p + 2^{p-1}|z-U_n|^p\right)(p_\infty^*(h))\,dh \\
&\leq \int_{H_n} \left(1 + 2^{p-1}|h|^p + O_p(1)\right)(p_n^*(h) - p_\infty^*(h))\,dh \\
&+ \int_{H_n^c} \left(1 + 2^{p-1}|h|^p + O_p(1)\right)(p_\infty^*(h))\,dh = o_p(1),
\end{aligned}$$

where $o_p(1)$-conclusion is by Theorem 1 and exponentially small tails of the normal density (Lebesgue measure of H_n^c converges to zero).

Now note that $Q_n(z)$ and $Q_\infty(z)$ are convex and finite, and $Z_n = \arg\inf_{z\in\mathbf{R}^d} Q_\infty(z) = O_p(1)$. By the convexity lemma of Pollard (1991), pointwise convergence entails the uniform convergence over compact sets K:

$$\sup_{z\in K} \left| Q_n(z) - Q_\infty(z) \right| \rightarrow_p 0.$$

Since $Z_n = O_p(1)$, uniform convergence and convexity arguments like those in Jureckova (1977) imply that $\sqrt{n}(\widehat{\theta} - \theta_0) - Z_n \rightarrow_p 0$, as shown below. ■

Proof of $Z_n - \sqrt{n}(\widehat{\theta} - \theta_0) = o_p(1)$. The proof follows by extending slightly the convexity argument of Jureckova (1977) and Pollard (1991) to the present context. Consider a ball $B_\delta(Z_n)$ with radius $\delta > 0$, centered at Z_n, and let $z = Z_n + dv$, where v is a unit direction vector such that $|v| = 1$ and $d > \delta$. Because $Z_n = O_p(1)$, for any $\delta > 0$ and $\epsilon > 0$, there exists $K > 0$ such that

$$\liminf_{n\to\infty} P_*\Big\{E_n = \{B_\delta(Z_n) \in B_K(0)\}\Big\} \geq 1 - \epsilon.$$

By convexity, for any $z = Z_n + dv$ constructed so, it follows that

$$\frac{\delta}{d}\big(Q_n(z) - Q_n(Z_n)\big) \geq Q_n(z^*) - Q_n(Z_n), \tag{A.16}$$

where z^* is a point of boundary of $B_\delta(Z_n)$ on the line connecting z and Z_n. By the uniform convergence of $Q_n(z)$ to $Q_\infty(z)$ over any compact set $B_K(0)$, whenever E_n occurs:

$$\begin{aligned}\frac{\delta}{d}\big(Q_n(z) - Q_n(Z_n)\big) &\geq Q_n(z^*) - Q_n(Z_n)\\ &\geq Q_\infty(z^*) - Q_\infty(Z_n) + o_p(1) \geq V_n + o_p(1),\end{aligned}$$

where $V_n > 0$ is a uniformly in n positive variable, because Z_n is the unique optimizer of Q_∞. That is, there exists an $\eta > 0$ such that $\liminf_n P(V_n > \eta) \geq 1 - \epsilon$. Hence we have with probability at least as big as $1 - 3\epsilon$ for large n:

$$\frac{\delta}{d}\big(Q_n(z) - Q_n(Z_n)\big) > \eta.$$

Thus, $\sqrt{n}(\widehat{\theta} - \theta_0)$ eventually belongs to a complement of $B_\delta(Z_n)$ with probability at most 3ϵ. Since we can set ϵ as small as we like by picking (a) sufficiently large K, and (b) sufficiently large n, and (c) sufficiently small $\eta > 0$, it follows that

$$\limsup_{n\to\infty} P^*\Big\{|Z_n - \sqrt{n}(\widehat{\theta} - \theta_0)| > \delta\Big\} = 0.$$

Since this is true for any $\delta > 0$, it follows that

$$Z_n - \sqrt{n}(\widehat{\theta} - \theta_0) = o_p(1). \quad \blacksquare$$

A.3 Proof of Theorem 3

For clarity of the argument, we limit exposition only to the case where $J_n(\theta)$ and $\Omega_n(\theta)$ do not depend on n. The more general case follows similarly. We defined

$$F_{g,n}(x) = \int_{\theta\in\Theta: g(\theta)\leq x} p_n(\theta)d\theta.$$

Evaluate it at $x = g(\theta_0) + s/\sqrt{n}$ and change the variable of integration

$$H_{g,n}(s) = F_{g,n}(g(\theta_0) + s/\sqrt{n}) = \int_{h\in H_n: g(\theta_0 + h/\sqrt{n} + U_n/\sqrt{n}) \leq g(\theta_0) + s/\sqrt{n}} p_n^*(h)dh.$$

Define also

$$\widehat{H}_{g,n}(s) = \int_{h\in\mathbb{R}^d: g(\theta_0 + h/\sqrt{n} + U_n/\sqrt{n}) \leq g(\theta_0) + s/\sqrt{n}} p_\infty^*(h)dh$$

and

$$H_{g,\infty}(s) = \int_{h\in\mathbb{R}^d: \nabla g(\theta_0)'(h/\sqrt{n} + U_n/\sqrt{n}) \leq s/\sqrt{n}} p_\infty^*(h)dh.$$

By definition of total variation of moments norm and Theorem 1

$$\sup_s |H_{g,n}(s) - \widehat{H}_{g,n}(s)| \to_p 0,$$

where the sup is taken over the support of $H_{g,n}(s)$.

By the uniform continuity of the integral of the normal density with respect to the boundary of integration

$$\sup_s |\widehat{H}_{g,n}(s) - H_{g,\infty}(s)| \to_p 0,$$

which implies

$$\sup_s |H_{g,n}(s) - H_{g,\infty}(s)| \to_p 0.$$

where the sup is taken over the support of $H_{g,n}(s)$.

The convergence of distribution function implies the convergence of quantiles at continuity points of distribution functions, see e.g. Billingsley (1994), so

$$H_{g,n}^{-1}(\alpha) - H_{g,\infty}^{-1}(\alpha) \to_p 0.$$

Next observe

$$H_{g,\infty}(s) = P\Big\{\nabla g(\theta_0)'\mathcal{N}(U_n, J^{-1}(\theta_0)) < s \Big| U_n\Big\},$$

so

$$H_{g,\infty}^{-1}(\alpha) = \nabla g(\theta_0)'U_n + q_\alpha \sqrt{\nabla_\theta g(\theta_0)' J^{-1}(\theta_0) \nabla_\theta g(\theta_0)},$$

where q_α is the α-quantile of $\mathcal{N}(0,1)$.

Recalling that we defined $c_{g,n}(\alpha) = F_{g,n}^{-1}(\alpha)$, by quantile equivariance with respect to the monotone transformations

$$H_{g,n}^{-1}(\alpha) = \sqrt{n}\,(c_{g,n}(\alpha) - g(\theta_0))$$

so that

$$\sqrt{n}\,(c_{g,n}(\alpha) - g(\theta_0)) = \nabla g(\theta_0)'U_n + q_\alpha \sqrt{\nabla_\theta g(\theta_0)' J^{-1}(\theta_0) \nabla_\theta g(\theta_0)} + o_p(1).$$

The rest of the result follows by the Δ-method. ■

A.4 Proof of Theorem 4

In view of Assumption 4, it suffices to show that

$$\widehat{J}_n^{-1}(\theta_0) - J_n^{-1}(\theta_0) \to_p 0, \tag{A.17}$$

and then conclude by the Δ-method.

Recall that

$$h = \sqrt{n}\,(\theta - \theta_0) - \underbrace{J_n(\theta_0)^{-1} \Delta_n(\theta_0)/\sqrt{n}}_{U_n},$$

and the localized Quasi-posterior density for h is

$$p_n^*(h) = \frac{1}{\sqrt{n}} p_n\left(h/\sqrt{n} + \theta_0 + U_n/\sqrt{n}\right).$$

Note also

$$\widehat{J}_n^{-1}(\theta_0) \equiv \int_\Theta n(\theta - \widehat{\theta})(\theta - \widehat{\theta})' p_n(\theta) d\theta$$
$$= \int_{H_n} (h - \sqrt{n}(\widehat{\theta} - \theta_0) + U_n) \cdot (h - \sqrt{n}(\widehat{\theta} - \theta_0) + U_n)' p_n^*(h) dh,$$

and

$$J_n^{-1}(\theta_0) \equiv \int_{\mathbf{R}^d} hh' p_\infty^*(h) dh.$$

We have, denoting $h = (h_1, ..., d_d)$ and $\widetilde{T}_n = (\widetilde{T}_{n1}, ..., \widetilde{T}_{nd})$ where $\widetilde{T}_n = \sqrt{n}(\widehat{\theta} - \theta_0) - U_n$, for all $i, j \leq d$

(a) $\int_{H_n} h_i h_j \Big(p_n^*(h) - p_\infty^*(h)\Big) dh = o_p(1)$ by Theorem 1,

(b) $\int_{H_n^c} h_i h_j \Big(p_\infty^*(h)\Big) dh = o_p(1)$ by definition of p_∞^* and $J_n(\theta_0)$ being uniformly nonsingular,

(c) $\int_{H_n} \underbrace{|\widetilde{T}_n|^2}_{=o_p(1)} \Big(p_n^*(h) - p_\infty^*(h)\Big) dh = o_p(1)$ by Theorem 2,

(d) $\int_{H_n} \underbrace{|\widetilde{T}_n|^2}_{=o_p(1)} \Big(p_\infty^*(h)\Big) dh = o_p(1)$ by Theorem 2, definition of p_∞^*, and $J_n(\theta_0)$ being nonsingular,

(e) $\int_{H_n} h_j \underbrace{\widetilde{T}_{ni}}_{=o_p(1)} \Big(p_n^*(h) - p_\infty^*(h)\Big) dh = o_p(1)$ by Theorems 1 and 2,

(f) $\int_{H_n} h_j \underbrace{\widetilde{T}_{ni}}_{=o_p(1)} \Big(p_\infty^*(h)\Big) dh = o_p(1)$ by Theorems 1 and 2, definition of p_∞^*, and $J_n(\theta_0)$ being uniformly nonsingular, from which the required conclusion follows. ■

A.5 Proof of Proposition 1

Assumption 3 is directly implied by (4.1)-(4.4) and the uniform continuity of $Em_i(\theta)$, as shown in Lemma 1. It remains only to verify Assumption 4.

Define the identity

$$\begin{aligned} L_n(\theta) - L_n(\theta_0) \equiv & -\underbrace{ng_n(\theta_0)'W(\theta_0)G(\theta_0)}_{\Delta_n(\theta_0)'}(\theta - \theta_0) \\ & - \frac{1}{2}(\theta - \theta_0)' n \underbrace{G(\theta_0)'W(\theta_0)G(\theta_0)}_{J(\theta_0)}(\theta - \theta_0) \quad + \quad R_n(\theta). \end{aligned} \tag{A.18}$$

Next, given the definition of $\Delta_n(\theta_0)$ and $J(\theta_0)$, conditions **i**, **ii**, **iii** of Assumption 4. are immediate from conditions **i-iii** of Proposition 1. Condition **iv** is verified as follows. Condition **iv** of Assumption 4 can be succinctly stated as:

$$\text{for each } \epsilon > 0 \text{ there exists a } \delta > 0 \text{ such that } \limsup_{n \to \infty} P^* \left\{ \sup_{|\theta - \theta_0| \leq \delta} \frac{|R_n(\theta)|}{1 + n|\theta - \theta_0|^2} > \epsilon \right\} < \epsilon.$$

This stochastic equicontinuity condition is equivalent to the following stochastic equicontinuity condition, see e.g. Andrews (1994a):

$$\text{for any } \delta_n \to 0 \qquad \sup_{|\theta-\theta_0|\le\delta_n} \frac{|R_n(\theta)|}{1+n|\theta-\theta_0|^2} = o_p(1). \tag{A.19}$$

This is weaker than condition (v) of Theorem 7.1 in Newey and McFadden (1994), which requires

$$\sup_{|\theta-\theta_0|\le\delta_n} \frac{R_n(\theta)}{\sqrt{n}|\theta-\theta_0|+n|\theta-\theta_0|^2} = o_p(1), \tag{A.20}$$

since

$$\frac{R_n(\theta)}{1+n|\theta-\theta_0|^2} = \frac{R_n(\theta)}{\sqrt{n}|\theta-\theta_0|+n|\theta-\theta_0|^2}\left[\frac{\sqrt{n}|\theta-\theta_0|+n|\theta-\theta_0|^2}{1+n|\theta-\theta_0|^2}\right],$$

where the term in brackets is bounded by

$$1+\frac{\sqrt{n}|\theta-\theta_0|}{1+n|\theta-\theta_0|^2} \le 2.$$

Hence the arguments of the proof, except for several important differences, follow those of Theorem 7.2 in Newey and McFadden (1994).

At first note that condition iv of Proposition 1 is implied by the condition (where we let $g(\theta) \equiv Eg_n(\theta)$):

$$\sup_{\theta\in B_{\delta_n}(\theta_0)} \epsilon(\theta) = o_p\left(\frac{1}{\sqrt{n}}\right), \text{ where } \epsilon(\theta) = \frac{g_n(\theta)-g_n(\theta_0)-g(\theta)}{1+\sqrt{n}|\theta-\theta_0|} \text{ for any } \delta_n \to 0. \tag{A.21}$$

From (A.18)

$$R_n(\theta) = R_{1n}(\theta) + R_{2n}(\theta) + R_{3n}(\theta),$$

where

$$\begin{aligned}
R_{1n}(\theta) \equiv & n\Big(g_n(\theta_0)' W_n(\theta) G(\theta_0)(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)' G(\theta_0)' W(\theta) G(\theta_0)(\theta-\theta_0) \\
& \qquad - \frac{1}{2} g_n(\theta)' W_n(\theta) g_n(\theta) + \frac{1}{2} g_n(\theta_0)' W_n(\theta) g_n(\theta_0)\Big), \\
R_{2n}(\theta) \equiv & n\Big(\frac{1}{2} g_n(\theta_0)' (W_n(\theta_0) - W_n(\theta)) g_n(\theta_0)\Big), \\
R_{3n}(\theta) \equiv & n\Big(g_n(\theta_0)' (W(\theta_0) - W_n(\theta))) G(\theta_0)(\theta-\theta_0) \\
& \qquad + \frac{1}{2}(\theta-\theta_0)' G(\theta_0)' (W(\theta_0) - W(\theta)) G(\theta_0)(\theta-\theta_0)\Big).
\end{aligned}$$

Verification of (A.19) for the terms $R_{2n}(\theta)$ and R_{3n} immediately follows from $\sqrt{n}g_n(\theta_0) = O_p(1)$ and the uniform consistency of $W_n(\theta)$ in θ as assumed in condition i of Proposition 1 and from the continuity of $W(\theta)$ in θ by condition i of Proposition 1, so that $W_n(\theta) - W(\theta) = o_p(1)$ uniformly in θ and $W(\theta) - W(\theta_0) = o(1)$ as $|\theta-\theta_0| \to 0$.

It remains to check condition (A.19) for the term $R_{1n}(\theta)$. Note that

$$g_n(\theta) = \left(1+\sqrt{n}|\theta-\theta_0|\right)\epsilon(\theta) + g(\theta) + g_n(\theta_0).$$

Substitute this into $R_{1n}(\theta)$ and decompose

$$-\frac{1}{n}R_{1n}(\theta) = \underbrace{\frac{1}{2}\left(1+\sqrt{n}|\theta-\theta_0|\right)^2 \epsilon(\theta)' W_n(\theta)\epsilon(\theta)}_{r_1(\theta)} + \underbrace{g_n(\theta_0)' W_n(\theta)(g(\theta) - G(\theta_0))(\theta-\theta_0)}_{r_2(\theta)}$$
$$+ \underbrace{\left(1+\sqrt{n}|\theta-\theta_0|\right)\epsilon(\theta)' W_n(\theta) g_n(\theta_0)}_{r_3(\theta)} + \underbrace{\left(1+\sqrt{n}|\theta-\theta_0|\right)\epsilon(\theta)' W_n(\theta) g(\theta)}_{r_4(\theta)}$$
$$+ \underbrace{\frac{1}{2}g(\theta)'(W_n(\theta) - W(\theta))\, g(\theta)}_{r_5(\theta)} + \underbrace{\frac{1}{2}g(\theta)' W(\theta) g(\theta) - \frac{1}{2}(\theta-\theta_0)' G(\theta_0)' W(\theta) G(\theta_0)(\theta-\theta_0)}_{r_6(\theta)}.$$

Using the inequalities, for $x \geq 0$:

$$\frac{\left(1+\sqrt{n}x\right)^2}{1+nx^2} \leq 2, \quad \frac{\sqrt{n}x}{1+nx^2} \leq 1, \quad \frac{1+\sqrt{n}x}{1+nx^2} \leq 2, \quad \frac{n\left(1+\sqrt{n}x\right)}{1+nx^2} \leq 2\frac{\sqrt{n}}{x}, \tag{A.22}$$

each of these terms can be dealt with separately, by applying the conditions **i**- **iii** and (A.21):

$$(a) \quad \sup_{\theta \in B_{\delta_n}(\theta_0)} \frac{n|r_1(\theta)|}{1+n|\theta-\theta_0|^2} \leq \sup_{\theta \in B_{\delta_n}(\theta_0)} n\epsilon(\theta)' W_n(\theta)\epsilon(\theta) = o_p(1),$$
$$(b) \quad \sup_{\theta \in B_{\delta_n}(\theta_0)} \frac{n|r_2(\theta)|}{1+n|\theta-\theta_0|^2} = \sup_{\theta \in B_{\delta_n}(\theta_0)} \frac{o\left(\sqrt{n}|\theta-\theta_0|\right)'}{1+n|\theta-\theta_0|^2} W_n(\theta)\sqrt{n}g_n(\theta_0) = o_p(1),$$
$$(c) \quad \sup_{\theta \in B_{\delta_n}(\theta_0)} \frac{n|r_3(\theta)|}{1+n|\theta-\theta_0|^2} \leq \sup_{\theta \in B_{\delta_n}(\theta_0)} 2n|\epsilon(\theta)' W_n(\theta) g_n(\theta_0)| = o_p(1),$$
$$(d) \quad \sup_{\theta \in B_{\delta_n}(\theta_0)} \frac{n|r_4(\theta)|}{1+n|\theta-\theta_0|^2} \leq \sup_{\theta \in B_{\delta_n}(\theta_0)} 2\sqrt{n}\left|\epsilon(\theta)' W_n(\theta)\frac{g(\theta)}{|\theta-\theta_0|}\right| = o_p(1),$$
$$(e) \quad \sup_{\theta \in B_{\delta_n}(\theta_0)} \frac{n|r_5(\theta)|}{1+n|\theta-\theta_0|^2} \leq \sup_{\theta \in B_{\delta_n}(\theta_0)} \frac{|g(\theta)|^2}{|\theta-\theta_0|^2}|W_n(\theta) - W(\theta)| = o_p(1),$$
$$(f) \quad \sup_{\theta \in B_{\delta_n}(\theta_0)} \frac{n|r_6(\theta)|}{1+n|\theta-\theta_0|^2} \leq \sup_{\theta \in B_{\delta_n}(\theta_0)} \frac{o\left(|\theta-\theta_0|^2|W(\theta)|\right)}{|\theta-\theta_0|^2} = o_p(1),$$

where (a) follows from (A.22), (A.21), and condition **i**, which states that $W_n(\theta) = W(\theta) + o_p(1)$ and $W(\theta) > 0$ is finite uniformly in θ; in (b) the first equality follows by Taylor expansion $g(\theta) = G(\theta_0)(\theta-\theta_0) + o(|\theta-\theta_0|)$ and the second conclusion follows from (A.22) and condition **iii**; (c) follows from (A.22), (A.21), condition **i** and **iii**; (d) follows by (A.22) and then replacing, by condition **ii**, $g(\theta)$ with $G(\theta_0)(\theta-\theta_0) + o(|\theta-\theta_0|)$, followed by applying (A.21) and condition **i**; (e) follows from replacing by condition **ii** $g(\theta)$ with $G(\theta_0)(\theta-\theta_0) + o(|\theta-\theta_0|)$, followed by applying condition **i**; and (f) follows from replacing $g(\theta)$ with $G(\theta_0)(\theta-\theta_0) + o(|\theta-\theta_0|)$, followed by applying condition **i**.

Verification of (A.20) for the term $R_{1n}(\theta)$ now follows by putting these terms together. ■

A.6 Proof of Proposition 2

Verification of Assumption 3 is standard given the stated conditions and is subsumed as a step in the consistency proofs of extremum estimators based on GEL in Kitamura and Stutzer (1997) for cases when s is finite and Kitamura (1997) for cases when s takes on infinite values. We shall not repeat it here. Next, we will verify Assumption 4. Define

$$\widehat{\gamma}(\theta) = \arg \inf_{\gamma \in \mathbf{R}^p} \bar{L}_n(\theta, \gamma).$$

It will suffice to show that uniformly in $\theta_n \in B_{\delta_n}(\theta_0)$ for any $\delta_n \to 0$, we have the GMM set-up:

$$\begin{aligned} L_n(\theta_n) &= \bar{L}_n(\theta_n, \widehat{\gamma}(\theta_n)) \\ &= -\frac{1}{2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n m_i(\theta_n)\right)'(V(\theta_0) + o_p(1))^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n m_i(\theta_n)\right), \end{aligned} \tag{A.23}$$

where

$$V(\theta_0) = E m_i(\theta_0) m_i(\theta_0)'.$$

The Assumptions 4.**i-iii** follow immediately from the conditions of Proposition 2, and Assumption 4.**iv** is verified exactly as in the proof of Proposition 1, given the reduction to the GMM case. Indeed, defining $g_n(\theta) = \frac{1}{n}\sum_{i=1}^n m_i(\theta)$, the Donsker property assumed in condition **iv** implies that for any $\epsilon > 0$, there is $\delta > 0$ such that

$$\limsup_{n\to\infty} P^*\left\{\sup_{\theta \in B_\delta(\theta_0)} \sqrt{n}|g_n(\theta) - g_n(\theta_0) - (Eg_n(\theta) - Eg_n(\theta_0))| > \epsilon\right\} < \epsilon,$$

which implies

$$\limsup_{n\to\infty} P^*\left\{\sup_{\theta \in B_\delta(\theta_0)} \frac{\sqrt{n}|g_n(\theta) - g_n(\theta_0) - (Eg_n(\theta) - Eg_n(\theta_0))|}{1 + \sqrt{n}|\theta - \theta_0|} > \epsilon\right\} < \epsilon,$$

which is condition **iv** in Proposition 1. The rest of the arguments follow that in the proof of Proposition 1.

It only remains to show the requisite expansion (A.23). We first show that

$$\widehat{\gamma}(\theta_n) \to_p 0.$$

For that purpose we use the convexity lemma, which was obtained by C. Geyer, and can be found in Knight (1999).

Convexity Lemma. Suppose Q_n is a sequence of lower-semi-continuous 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_n weakly converges to Q_∞ in $\bar{\mathbb{R}}$ marginally (in finite-dimensional sense) on $\mathcal{D}$ where Q_∞ is lower-semi-continuous convex and finite on an open non-empty set a.s., then

$$\arg\inf_{z \in \mathbf{R}^d} Q_n(z) \to_d \arg\inf_{z \in \mathbf{R}^d} Q_\infty(z),$$

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

Next, we show that $\widehat{\gamma}(\theta_n) \to_p 0$. Define $F = \{\gamma : Es[m_i(\theta_0)'\gamma] < \infty\}$ and $F^c = \{\gamma : Es[m_i(\theta_0)'\gamma] = \infty\}$. By convexity and lower-semicontinuity of s, F is convex, open, and its boundary is nowhere dense in $\mathbb{R}^p$. Thus for $\gamma \in F$, $Es[m_i(\theta)'\gamma] < \infty$ for all $\theta \in B_\delta(\theta_0)$ and some $\delta > 0$, which follows by continuity of $\theta \mapsto Es[m_i(\theta)'\gamma]$ over $B_\delta(\theta_0)$ implied by the condition **ii** and **iii**.

Thus, for a given $\gamma \in F$ and any $\theta_n \to_p \theta_0$

$$\frac{1}{n}\sum_{i=1}^n s[m_i(\theta_n)'\gamma] \to_p Es[m_i(\theta_0)'\gamma] < \infty.$$

This follows from the uniform law of large numbers implied by

1. $\{s[m_i(\theta)'\gamma], \theta \in B_\delta(\theta_0)\}$, where δ is sufficiently small, being a Donsker class wp $\to 1$, and

2. $Em_i(\theta) = \int x \cdot dP[m_i(\theta) \leq x]$ being continuously differentiable in θ by condition **ii** and **iii**.

The above function set is Donsker by

(a) $m_i(\theta)'\gamma \in M$ for some compact M and a given $\gamma \in F$, by condition **iii**,

(b) $\{m_i(\theta), \theta \in B_\delta(\theta_0)\}$ being Donsker class by condition **iv**,

(c) s being a uniform Lipschitz function over $\mathcal{V} \cap M$[17], by assumption on s,

(d) $m_i(\theta)'\gamma \in \mathcal{V}$ for all $\theta \in B_\delta(\theta_0)$, some $\delta > 0$, and a given $\gamma \in F$, by construction of F,

(e) Theorem 2.10.6 in van der Vaart and Wellner (1996) that says a uniform Lipschitz transform of a Donsker class is Donsker class itself.

Now take γ in $F^c \setminus \partial F$, where ∂F denotes the boundary of F. Then wp $\to 1$

$$\frac{1}{n}\sum_{i=1}^{n} s[m_i(\theta_n)'\gamma] = \infty \to_p Es[m_i(\theta_0)'\gamma] = \infty.$$

Now take all the rational numbers $\gamma \in \mathbb{R}^p \setminus \partial F$ as the set $\mathcal{D}$ appearing in the statement of the Convexity Lemma and conclude that

$$\widehat{\gamma}(\theta_n) \to_p 0 = \arg\inf_\gamma Es[m_i(\theta_0)'\gamma].$$

Given this result, we can expand the first order condition for $\widehat{\gamma}(\theta_n)$ in order to obtain the expression for its form. Note first

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \nabla s\left(\gamma(\theta_n)' m_i(\theta_n)\right) m_i(\theta_n) \\ &= \sum_{i=1}^{n} m_i(\theta_n) + \gamma(\theta_n)\, nV_n, \end{aligned} \tag{A.24}$$

where

$$V_n = \frac{1}{n}\sum_{i=1}^{n} \nabla^2 s\left(\bar{\gamma}(\theta_n)' m_i(\theta_n)\right) m_i(\theta_n) m_i(\theta_n)',$$

for some $\bar{\gamma}(\theta_n)$ between 0 and $\gamma(\theta_n)$, which is different from row to row of the matrix V_n.

Then

$$V_n \to_p V(\theta_0) = Em_i(\theta_0) m_i(\theta_0)'.$$

This follows from the uniform law of large numbers implied by

1. $\{\nabla^2 s(\gamma' m_i(\theta^*)) m_i(\theta) m_i(\theta)', (\theta^*, \gamma, \theta) \in B_{\delta_1}(\theta_0) \times B_{\delta_2}(0) \times B_{\delta_3}(\theta_0)\}$, where $\delta_j > 0$ are sufficiently small, being a Donsker class wp $\to 1$,

2. $Em_i(\theta)m_i(\theta)' = \int xx' dP[m_i(\theta) \leq x]$ being continuous function in θ by condition **i**,

[17] Recall that $\mathcal{V}$ is defined as the open convex set on which s is finite.

3. $E\nabla^2 s(\gamma' m_i(\theta^*)) m_i(\theta) m_i(\theta)' = \underbrace{E\nabla^2 s(0)}_{1} m_i(\theta) m_i(\theta)' + o(1)$ uniformly in $(\theta, \theta^*) \in B_\delta(\theta_0) \times B_\delta(\theta_0)$ for sufficiently small $\delta > 0$, for any $\gamma \to 0$, by assumptions on s and condition **iii**.

The claim **1** is verified by applying exactly the same logic as in the previously stated steps (a)-(e). For the sake of brevity, this will not be repeated.

Therefore, wp $\to 1$

$$\gamma(\theta_n) = -(V_n)^{-1}\frac{1}{n}\sum_{i=1}^{n} m_i(\theta_n) \equiv -(V(\theta_0)^{-1} + o_p(1))\frac{1}{n}\sum_{i=1}^{n} m_i(\theta_n). \tag{A.25}$$

Consider the second order expansion,

$$\bar{L}_n(\theta_n, \gamma(\theta_n)) = \sqrt{n}\gamma(\theta_n)'\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_i(\theta_n) + \frac{1}{2}\sqrt{n}\gamma(\theta_n)'\tilde{V}_n\sqrt{n}\gamma(\theta_n), \tag{A.26}$$

where

$$\tilde{V}_n = \frac{1}{n}\sum_{i=1}^{n}\nabla^2 s\left(\tilde{\gamma}(\theta_n)' m_i(\theta_n)\right) m_i(\theta_n) m_i(\theta_n)',$$

for some $\tilde{\gamma}(\theta_n)$ between 0 and $\gamma(\theta_n)$, which is different from row to row of the matrix $\tilde{V}_n$. By a preceding argument,

$$\tilde{V}_n \to_p V(\theta_0).$$

Inserting (A.25) and $\tilde{V}_n = V(\theta_0) + o_p(1)$ into (A.26), we obtain the required expansion (A.23). ■

A.7 Proof of Proposition 3

Assumption 3 is assumed. We need to verify Assumption 4.

Define the identity

$$\begin{aligned} L_n(\theta) - L_n(\theta_0) \equiv & \underbrace{\sum_{i=1}^{n} \dot{m}_i(\theta_0)'}_{\Delta_n(\theta_0)'}(\theta - \theta_0) \\ & + \frac{1}{2}(\theta - \theta_0)' n \underbrace{\nabla_{\theta\theta'} E m_i(\theta_0)}_{-J(\theta_0)}(\theta - \theta_0) \\ & + R_n(\theta). \end{aligned} \tag{A.27}$$

Assumption 4.**i-iii** then follows immediately from conditions **i** and **ii**. Assumption 4.**iv** is verified as follows.

The remainder term $R_n(\theta)$ is given the following decomposition:

$$\begin{aligned} R_n(\theta) = & \underbrace{\sum_{i=1}^{n}\left\{m_i(\theta) - m_i(\theta_0) - Em_i(\theta) + Em_i(\theta_0) - \dot{m}_i(\theta_0)'(\theta - \theta_0)\right\}}_{R_{1n}(\theta)} \\ & \underbrace{+ n(Em_i(\theta) - Em_i(\theta_0)) + \frac{1}{2}(\theta - \theta_0)' n J(\theta_0)(\theta - \theta_0)}_{R_{2n}(\theta)}. \end{aligned}$$

It suffices to verify Assumption 4.iv separately for $R_{1n}(\theta)$ and $R_{2n}(\theta)$. Since

$$R_{2n}(\theta) = -\frac{1}{2}n(\theta-\theta_0)'\left[J(\theta^*) - J(\theta_0)\right](\theta-\theta_0),$$

for some θ^* on the line connecting θ and θ_0, verification of Assumption 4 for $R_{2n}(\theta)$ follows immediately from continuity of $J(\theta)$ in θ over a ball at θ_0.

To show Assumption 4.iv-(b) for $R_{1n}(\theta)$, we note that for any given $M > 0$

$$\begin{aligned}
&\limsup_n P^*\left\{\sup_{|\theta-\theta_0|\le M/\sqrt{n}} |R_{1n}(\theta)| > \epsilon\right\} \\
&\le \limsup_n P\left\{\sup_{|\theta-\theta_0|\le M/\sqrt{n}} |\theta-\theta_0|\frac{|R_{1n}(\theta)|}{|\theta-\theta_0|} > \epsilon\right\} \\
&\le \limsup_n P\left\{\sup_{|\theta-\theta_0|\le M/\sqrt{n}} \frac{M}{\sqrt{n}}\frac{|R_{1n}(\theta)|}{|\theta-\theta_0|} > \epsilon\right\} = 0,
\end{aligned} \tag{A.28}$$

where the last conclusion follows from two observations.

First, note that

$$Z_n(\theta) \equiv \frac{R_{1n}(\theta)}{\sqrt{n}|\theta-\theta_0|} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{m_i(\theta) - m_i(\theta_0) - (Em_i(\theta) - Em_i(\theta_0)) - \dot{m}_i(\theta_0)'(\theta-\theta_0)}{|\theta-\theta_0|}\right)$$

is Donkser by assumption, that is it converges in $\ell^\infty(B_\delta(\theta_0))$ to a tight Gaussian process Z.

The process has uniformly continuous paths with respect to the semimetric ρ given by

$$\rho^2(\theta_1,\theta_2) = E(Z(\theta_1) - Z(\theta_2))^2,$$

so that $\rho(\theta,\theta) \to 0$ if $\theta \to \theta_0$. Thus almost all sample paths of Z are continuous at θ_0.

Second, since by assumption

$$E[\overline{m}_n(\theta) - \overline{m}_n(\theta_0) - \overline{\dot{m}_n}(\theta_0)'(\theta-\theta_0)]^2 = o(|\theta-\theta_0|^2),$$

we have for any $\theta_n \to \theta_0$

$$E^*\left[\frac{|R_{1n}(\theta_n)|}{\sqrt{n}|\theta_n-\theta_0|}\right]^2 = \frac{o\left(|\theta_n-\theta|^2\right)}{|\theta_n-\theta_0|^2} \to 0,$$

therefore

$$Z(\theta_0) = 0.$$

Therefore for any $\theta' \to_p \theta_0$, we have by the extended continuous mapping theorem

$$Z_n(\theta') \to_d Z(\theta_0) = 0, \text{ that is } Z_n(\theta') \to_p 0. \tag{A.29}$$

This shows (A.28).

To prove Assumption 4.iv-(a) for $R_{1n}(\theta)$, we need to show that for some $\delta > 0$ and constant M

$$\limsup_n P^*\left\{\sup_{M/\sqrt{n}\le|\theta-\theta_0|\le\delta} \frac{|R_{1n}(\theta)|}{n|\theta-\theta_0|^2} > \epsilon\right\} < \epsilon. \tag{A.30}$$

Using that $M/\sqrt{n} \leq |\theta - \theta_0|$, bound the left-hand-side by

$$\begin{aligned}
&\limsup_n P^*\left\{\sup_{M/\sqrt{n}\leq|\theta-\theta_0|\leq\delta} \frac{|R_{1n}(\theta)|}{\sqrt{n}|\theta-\theta_0|}\cdot\frac{1}{\sqrt{n}|\theta-\theta_0|} > \epsilon\right\} \\
&\leq \limsup_n P^*\left\{\sup_{M/\sqrt{n}\leq|\theta-\theta_0|\leq\delta} \frac{|R_{1n}(\theta)|}{\sqrt{n}|\theta-\theta_0|}\cdot\frac{1}{M} > \epsilon\right\} \\
&\leq \limsup_n P^*\left\{\sup_{M/\sqrt{n}\leq|\theta-\theta_0|\leq\delta} |Z_n(\theta)|\cdot\frac{1}{M} > \epsilon\right\} \\
&< \epsilon,
\end{aligned} \tag{A.31}$$

where for any given $\epsilon > 0$ in order to make the last inequality true, we can make either δ sufficiently small by the property (A.29) of Z_n or make M sufficiently large by the property $Z_n = O_{p^*}(1)$. ■

B Appendix on Computation

B.1 A computational lemma

In this section we record some formal results on MCMC computation of the quasi-posterior quantities.

LEMMA 3 *Suppose the chain $(\theta^{(j)}, j \leq B)$ is produced by the Metropolis Hastings(MH) algorithm with q such that $q(\theta|\theta') > 0$ for each (θ, θ'). Suppose also that $P\{\rho(\theta^{(j)}, \xi) = 1\} > 0$ for all $j > t_0$. Then*

1. *$p_n(\cdot)$ is the stationary density of the chain,*
2. *the chain is ergodic with the limit marginal distribution given by $p_n(\cdot)$:*

$$\lim_{B\to\infty}\sup_A \left|P(\theta^{(B)} \in A|\theta_0) - \int_A p_n(\theta)d\theta\right| = 0,$$

where the supremum is taken over the Borel sets,

3. *For any p_n- integrable function g:*

$$\frac{1}{B}\sum_{j=1}^{B} g(\theta^{(j)}) \to_p \int_\Theta g(\theta)p_n(\theta)d\theta.$$

Proof. The result is immediate from Theorem 6.2.5 in Robert and Casella (1999). ■

An immediate consequence of this lemma is the following result.

LEMMA 4 *Suppose Assumptions 1 and 2 hold. Suppose the chain $(\theta^{(j)}, j \leq B)$ satisfies the conditions of Lemma 3, then for any convex and p_n-integrable loss function $\rho_n(\cdot)$*

$$\arg\inf_{\theta\in\Theta}\left[\frac{1}{B}\sum_{j=1}^{B}\rho_n(\theta^{(j)} - \theta)\right] \to_p \widehat{\theta} = \arg\inf_{\theta\in\Theta}\left[\int_\Theta \rho_n(\bar\theta - \theta)p_n(\bar\theta)d\bar\theta\right],$$

provided that $\widehat{\theta}$ is uniquely defined.

Proof. By Lemma 3 we have the pointwise convergence of the objective function: for any θ

$$\frac{1}{B}\sum_{j=1}^{B}\rho_n(\theta^{(j)}-\theta)\to_p \int_\Theta \rho_n(\tilde{\theta}-\theta)p_n(\tilde{\theta})d\tilde{\theta},$$

which implies the result by the Convexity Lemma, since $\theta \mapsto \int_\Theta \rho_n(\tilde{\theta}-\theta)p_n(\tilde{\theta})d\tilde{\theta}$ is convex by convexity of ρ_n. ■

B.2 Quasi-Bayes Estimation and Simulated Annealing

The relation between drawing from the shape of a likelihood surface and optimizing to find the mode of the likelihood function is well known. It is well established that, e.g. Robert and Casella (1999),

$$\lim_{\lambda\to\infty}\frac{\int_\Theta \theta e^{\lambda L_n(\theta)}\pi(\theta)\,d\theta}{\int_\Theta e^{\lambda L_n(\theta)}\pi(\theta)\,d\theta}=\underset{\theta\in\Theta}{\operatorname{argmax}} L_n(\theta) \tag{B.1}$$

Essentially, as $\lambda\to\infty$, the sequence of probability measures

$$\frac{e^{\lambda L_n(\theta)}\pi(\theta)}{\int_\Theta e^{\lambda L_n(\theta)}\pi(\theta)\,d\theta} \tag{B.2}$$

converges to the generalized Dirac probability measure concentrated at $\underset{\theta\in\Theta}{\operatorname{argmax}}\, L_n(\theta)$.

The difficulty of nonlinear optimization has been an important issue in econometrics (Berndt et al. (1974), Sims (1999)). The simulated annealing algorithm (see e.g. Press et al. (1992), Goffe et al. (1994)) is usually considered a generic optimization method. It is an implementation of the simulation based optimization (B.1) with a uniform prior $\pi(\theta)\equiv c$ on the parameter space Θ. At each temperature level $1/\lambda$, the simulated annealing routine uses a large number of Metropolis-Hastings steps to draw from the quasi distribution (B.2). The temperature parameter is then decreased slowly while the Metropolis steps are repeated, until convergence criteria for the optimum are achieved.

Interestingly, the simulated annealing algorithm has been widely used in optimization of non-likelihood-based semiparametric objective functions. In principle, if the temperature parameter is decreased at an arbitrarily slow rate (that depends on the criterion function), simulated annealing can find the global optimum of non-smooth objective functions that may have many local extrema. Controlling the temperature reduction parameter is a very delicate matter and is certainly crucial to the performance of the algorithm with highly nonsmooth objective functions. On the other hand, as Theorems 1 and 2 apply equally to (B.2), the results of this paper show that we may fix the temperature parameter $1/\lambda$ at a positive constant and then compute the quasi-posterior medians or means for (B.2) using Metropolis steps. These estimates can be used in place of the exact maximum. They are consistent and asymptotically normal, and possess the same limiting distribution as the exact maximum. The interpretation of the simulated annealing algorithm as an implementation of (B.2) also suggests that for some problems with special structures, other MCMC methods, such as the Gibbs sampler, may be used to replace the Metropolis-Hasting step in the simulated annealing algorithm.

B.3 Details of Computation in Monte-Carlo Examples

The parameter space is taken to be $\Theta=[\theta_0\pm 10]$. The transition kernel is a Normal density, and flat prior is truncated to Θ. Each parameter is updated via a Gibbs-Metropolis procedure, which modifies slightly the basic Metropolis-Hastings algorithm: for $k = 1, ..., d$, a draw of ξ_k from the univariate normal density $q(|\xi_k - \theta_k^{(j)}|, \phi)$ is made, then the candidate value ξ consisting of ξ_k and $\theta_{-k}^{(j)}$ replaces $\theta^{(j)}$ with probability $\rho(\theta^{(j)}, \xi)$ specified in the text. Variance parameter ϕ is adjusted every 100 draws (in the second simulation example and empirical example) or 200 draws (in the first simulation example) so that the rejection probability is roughly 50%.

The first $N \times d$ draws (the burn-in stage) are discarded, and the remaining $N \times d$ draws are used in computation of estimates and intervals. The starting value is the OLS estimate in all examples. We use $N = 5,000$ in the second simulation example and empirical example and $N = 10,000$ in the second simulation example. To give an idea of computational expense, computing one set of estimates takes 20-40 seconds depending on the example. All of the codes that we used to produce figures, simulation, and empirical results are available from the authors.

Notation and Terms

$\to_p$	convergence in (outer) probability P^*
$\to_d$	convergence in distribution under P^*
wp $\to 1$	with inner probability P_* converging to one
$\sim$	asymptotic equivalence denoted $A \sim B$ means $\lim AB^{-1} = I$
$B_\delta(x)$	ball centered at x of radius $\delta > 0$
I	identity matrix
$A > 0$	A is positive definite when A is matrix
$\mathcal{N}(0, a)$	normal random vector with mean 0 and variance matrix a
$\mathcal{F}$ Donsker class	here this means that empirical process $f \mapsto \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(W_i) - Ef(W_i))$ is asymptotically Gaussian in $\ell^\infty(\mathcal{F})$, see van der Vaart (1999)
$\ell^\infty(\mathcal{F})$	metric space of bounded over $\mathcal{F}$ functions, see van der Vaart (1999)
mineig(A)	minimum eigenvalue of matrix A

References

Abadie, A., 1995. Changes in spanish labor income structure during the 1980s: a quantile regression approach, CEMFI Working Paper.

Amemiya, T., 1977. The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. Econometrica 45 (4), 955–968.

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press.

Anderson, T. W., 1955. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proc. Amer. Math. Soc. 6, 170–176.

Andrews, D. W. K., 1994a. Empirical process methods in econometrics. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. North Holland, pp. 2248–2292.

Andrews, D. W. K., 1994b. The large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests. Econometrica 62 (5), 1207–1232.

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

Andrews, D. W. K., 1999. Estimation when a parameter is on a boundary. Econometrica 66, 1341–83.

Berger, J. O., 2002. Bayesian analysis: a look at today and thoughts of tomorrow. In: Statistics in the 21st Century. Chapman and Hall, New York, pp. 275–290.

Berndt, E., Hall, B., Hall, R., Hausman, J., 1974. Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement 3 (4), 653–665.

Bernstein, S., 1917. Theory of Probability. (Russian) Fourth Edition (1946) Gostekhizdat, Moscow-Leningrad.

Berry, S., Levinsohn, J., Pakes, A., July 1995. Automobile prices in market equilibrium. Econometrica 63, 841–890.

Bickel, P. J., Yahav, J. A., 1969. Some contributions to the asymptotic theory of Bayes solutions. Z. Wahrsch. Verw. Geb 11, 257–276.

Billingsley, P., 1994. Probability and Measure, 3rd Ed. John Wiley and Sons.

Buchinsky, M., 1991. Theory of and practice of quantile regression, Ph.D. dissertation, Department of Economics Harvard University.

Buchinsky, M., Hahn, J., 1998. An alternative estimator for the censored regression model. Econometrica 66, 653–671.

Bunke, O., Milhaud, X., 1998. Asymptotic behavior of Bayes estimates under possibly incorrect models. The Annals of Statistics 26 (2), 617–644.

Chamberlain, G., Imbens, G., 1997. Nonparametric appliations of Bayesian inference, NBER Working Paper.

Chernozhukov, V., Hansen, C., 2001. An IV model of quantile treatment effects, MIT Department of Economics Working Paper.

Chernozhukov, V., Umantsev, L., 2001. Conditional value-at-risk: Aspects of modeling and estimation. Empirical Economics 26, 271–92.

Chib, S., 2001. Markov chain monte carlo methods: Computation and inference. In: J.J.Heckman, Leamer, E. (Eds.), Handbook of Econometrics, Vol 5, Chapter 5. North Holland, pp. 3564–3634.

Christoffersen, P., Hahn, J., Inoue, A., 1999. Testing, comparing and combining value at risk measures, working Paper, Wharton School University of Pennsylvania.

Diaconis, P., Freedman, D., 1986. On the consistency of Bayes estimates. Annals of Statistics 14, 1–26.

Doksum, K. A., Lo, A. Y., 1990. Consistent and robust Bayes procedures for location based on partial information. Ann. Statist. 18 (1), 443–453.

Engle, R., Manganelli, S., 2001. Caviar: Conditional value at risk by regression quantiles, Working Paper, Department of Economics UC San Diego.

Fitzenberger, B., 1997. A guide to censored quantile regressions. In: Robust inference, Handbook of Statistics. Vol. 14. North-Holland, Amsterdam, pp. 405–437.

Gallant, A. R., White, H., 1988. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Oxford: Bail Blackwell.

Geweke, J., Keane, M., 2001. Computationally intensive methods for integration in econometrics. In: J.J.Heckman, Leamer, E. (Eds.), Handbook of Econometrics, Vol 5, Chapter 5. North Holland, pp. 3465–3564.

Goffe, W. L., Ferrier, G. D., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. Journal of Econometrics 60, 65–99.

Hahn, J., 1997. Bayesian bootstrap of the quantile regression estimator: a large sample study. Internat. Econom. Rev. 38 (4), 795–808.

Hansen, L., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.

Hogg, R. V., 1975. Estimates of percentile regression lines using salary data. Journal of American Statistical Association 70, 56–59.

Huber, P. J., 1973. Robust regression: Asymptotics, conjectures, and monte carlo. Annals of Statistics 1, 799–821.

Ibragimov, I., Has'minskii, R., 1981. Statistical Estimation: Asymptotic Theory. Springer Verlag.

Imbens, G., 1997. One-step estimators for over-identified generalized method of moments models. Rev. Econom. Stud. 64 (3), 359–383.

Imbens, G., Spady, R., Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66, 333–357.

Jureckova, J., 1977. Asymptotic relations of M-estimators and R-estimators in linear regression models. Annals of Statistics 5, 464–472.

Khan, S., Powell, J. L., 2001. Two step estimation of semiparametric censored regression models. Journal of Econometrics 103, 73–110.

Kim, J.-Y., 1998. Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models. Econometrica 66 (2), 359–380.

Kim, J.-Y., 2002. Limited information likelihood and Bayesian analysis. Journal of Econometrics , 175–193.

Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes. Ann. Statist. 25 (5), 2084–2102.

Kitamura, Y., Stutzer, M., 1997. An information-theoretic alternative to generalized method of moments estimation. Econometrica 65, 861–874.

Knight, K., 1999. Epi-convergence and stochastic equisemicontinuity, Working Paper, Department of Statistics University of Toronto.

Koenker, R., 1994. Confidence intevals for quantile regression. In: Proceedings of the 5th Prague Symposium on Asymptotic Statistics. Heidelberg: Physica-Verlag, pp. 10–20.

Koenker, R., 1998. Treating the treated, varieties of causal analysis, Lecture Note, Depratment of Economics University of Illinois.

Koenker, R., Bassett, G. S., 1978. Regression quantiles. Econometrica 46, 33–50.

Kottas, A., Gelfand, A., 2001. Bayesian semiparametric median regression modeling. Journal of the American Statistical Association 96, 1458–1468.

Lehmann, E., Casella, G., 1998. Theory of Point Estimation. Springer.

Macurdy, T., Timmins, C., 2001. Bounding the influence of attrition on the intertemporal wage variation in the NLSY, Working Paper, Department of Economics Yale University.

Mood, A. M., 1950. Introduction to the Theory of Statistics. McGraw-Hill Book Company, Inc.

Newey, W. K., 1991. Uniform convergence in probability and stochastic equicontinuity. Econometrica 59 (4), 1161–1167.

Newey, W. K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. North Holland, pp. 2113–2241.

Newey, W. K., Powell, J. L., 1990. Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. Econometric Theory 6, 295–317.

Newey, W. K., Smith, R., 2001. Higher order properties of GMM and generalized empirical likeliood estimators, Working Paper, Department of Economics MIT.

Newey, W. K., West, K. D., 1987. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.

Owen, A., 1989. Empirical likelihood ratio confidence regions. In: Proceedings of the 47th Session of the International Statistical Institute, Book 3 (Paris, 1989). Vol. 53. pp. 373–393.

Owen, A., 1990. Empirical likelihood ratio confidence regions. Ann. Statist. 18 (1), 90–120.

Owen, A., 1991. Empirical likelihood for linear models. Ann. Statist. 19 (4), 1725–1747.

Owen, A., 2001. Empirical Likelihood. Chapman and Hall/CRC.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57 (5), 1027–1057.

Phillips, P. C. B., Ploberger, W., 1996. An asymptotic theory of Bayesian inference for time series. Econometrica 64 (2), 381–412.

Pollard, D., 1991. Asymptotics for least absolute deviation regression estimator. Econometric Theory 7, 186–199.

Pötscher, B. M., Prucha, I. R., 1997. Dynamic Nonlinear Econometric Models. Springer-Verlag, Berlin.

Powell, J. L., 1984. Least absolute deviations estimation for the censored regression model. Journal of Econometrics , 303–325.

Press, W., Teukolsky, S. A., Vettering, W., Flannery, B., 1992. Numerical Recipes in C, The Art of Scientific Computing. Cambridge.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Annals of Statistics 22, 300–325.

Robert, C. P., Casella, G., 1999. Monte Carlo Statistical Methods. Springer.

Rousseeuw, P. J., Hubert, M., 1999. Regression depth. J. Amer. Statist. Assoc. 94 (446), 388–433.

Sims, C., 1999. Adaptive Metropolis-Hastings, or monte carlo kernel estimation, Working Paper, Department of Economics Princeton University.

Stigler, S. M., 1975. Studies in the history of probability and statistics. XXXIV. Napoleonic statistics: the work of Laplace. Biometrika 62 (2), 503–517.

van Aelst, S., Rousseeuw, P. J., Hubert, M., Struyf, A., 2002. The deepest regression method. J. Multivariate Anal. 81 (1), 138–166.

van der Vaart, A., 1999. Asymptotic Statistics. Cambridge University Press.

van der Vaart, A. W., Wellner, J. A., 1996. Weak Convergence and Empirical Processes. Springer-Verlag, New York.

von Mises, R., 1931. Wahrscheinlichkeitsrechnung. Berlin: Springer.

White, H., 1994. Estimation, Inference and Specification Analysis. Vol. 22 of Econometric Society Monographs. Cambridge University Press, Cambridge.

Zellner, A., 1998. Past and recent results on maximal data information priors. J. Statist. Res. 32 (1), 1–22.

Table 1: Monte Carlo Comparison of LTE's with Censored Quantile Regression Estimates Obtained using Iterated Linear Programming (Based on 100 repetitions).

Estimator	RMSE	MAD	Mean Bias	Median Bias	Median Abs. Dev.
n=400					
Q-posterior-mean	0.473	0.378	0.138	0.134	0.340
Q-posterior-median	0.465	0.372	0.131	0.137	0.344
Iterated LP(10)	0.518	0.284	0.040	0.016	0.171
	3.798	0.827	-0.568	-0.035	0.240
n=1600					
Q-posterior-mean	0.155	0.121	-0.018	0.009	0.089
Q-posterior-median	0.155	0.121	-0.020	0.002	0.092
Iterated LP(7)	0.134	0.106	0.040	0.067	0.085
	3.547	0.511	0.023	-0.384	0.087

Table 2: Monte Carlo Comparison of the LTE's with Standard Estimation for a Linear Quantile Regression Model (Based on 500 repetitions)

Estimator	RMSE	MAD	Mean Bias	Median Bias	Median AD
n=200					
Q-posterior-mean	.0747	.0587	.0174	.0204	.0478
Q-posterior-median	.0779	.0608	.0192	.0136	.0519
Standard Quantile Regression	.0787	.0628	.0067	.0092	.0510
n=800					
Q-posterior-mean	.0425	.0323	-.0018	-.0003	.0280
Q-posterior-median	.0445	.0339	-.0023	.0001	.0295
Standard Quantile Regression	.0498	.0398	.0007	.0025	.0356

Table 3: Monte Carlo Comparison of the LT Inference with Standard Inference for a Linear Quantile Regression Model (Based on 500 repetitions)

Inference Method	**coverage**	**length**
n=200		
Quasi-posterior confidence interval, equal tailed	.943	.377
Quasi-posterior confidence interval, symmetric (around mean)	.941	.375
Quantile Regression: Hall-Sheather Interval	.659	.177
n=800		
Quasi-posterior confidence interval, equal tailed	.920	.159
Quasi-posterior confidence interval, symmetric (around mean)	.917	.158
Quantile Regression: Hall-Sheather Interval	.602	.082

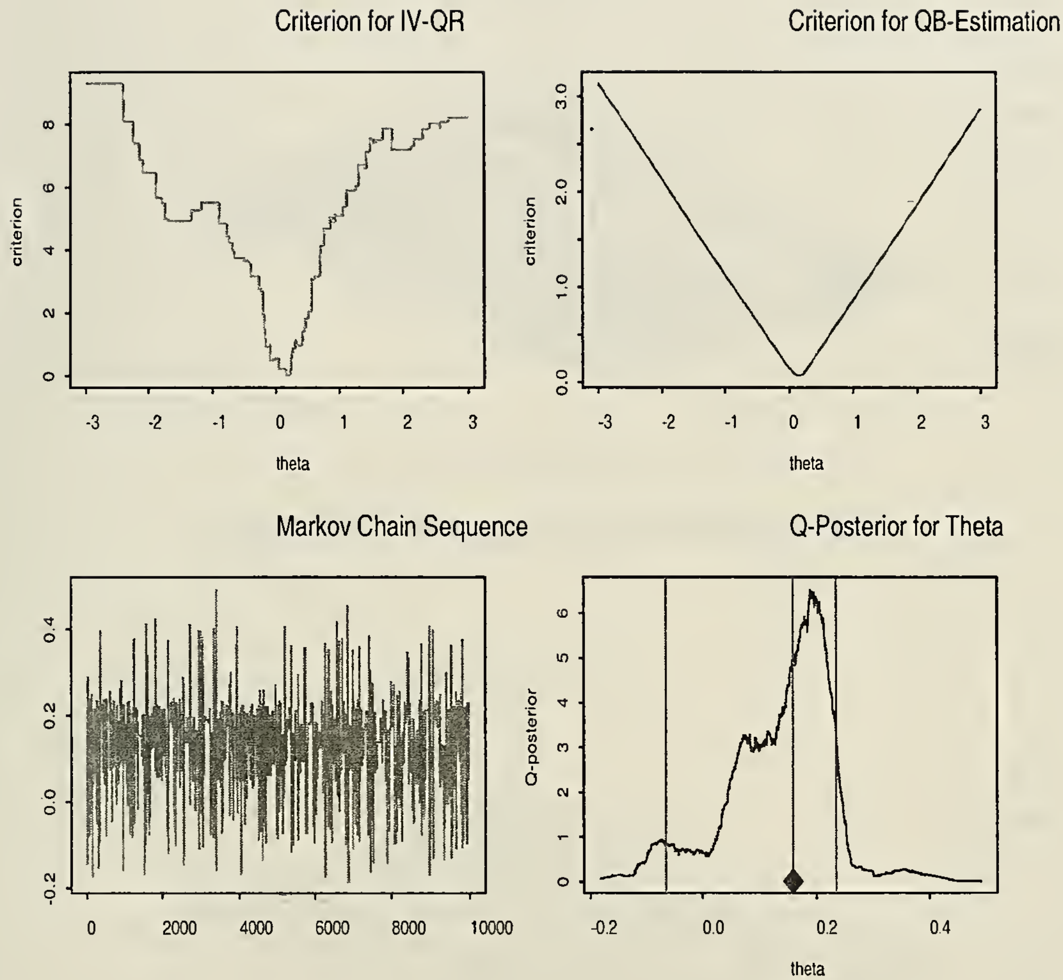

Figure 1: A Nonlinear IV Example involving Instrumental Quantile Regression. In the top-left panel the discontinuous objective function $L_n(\theta)$ is depicted (one-dimensional case). The true parameter $\theta_0 = 0$. In the bottom-left panel, a Markov Chain sequence of draws $\left(\theta^{(1)}, ...\theta^{(J)}\right)$ is depicted. The marginal distribution of this sequence is $p_n(\theta) = e^{L_n(\theta)} / \int_\Theta e^{L_n(\theta)} d\theta$, see the bottom-right panel. The point estimate, the sample mean $\bar{\theta}$, is given by the vertical line with the romboid root. Two other vertical lines are the 10-th and the 90-th percentiles of quasi-posterior distribution. The upper-right panel depicts the expected loss function that the LTE minimize.

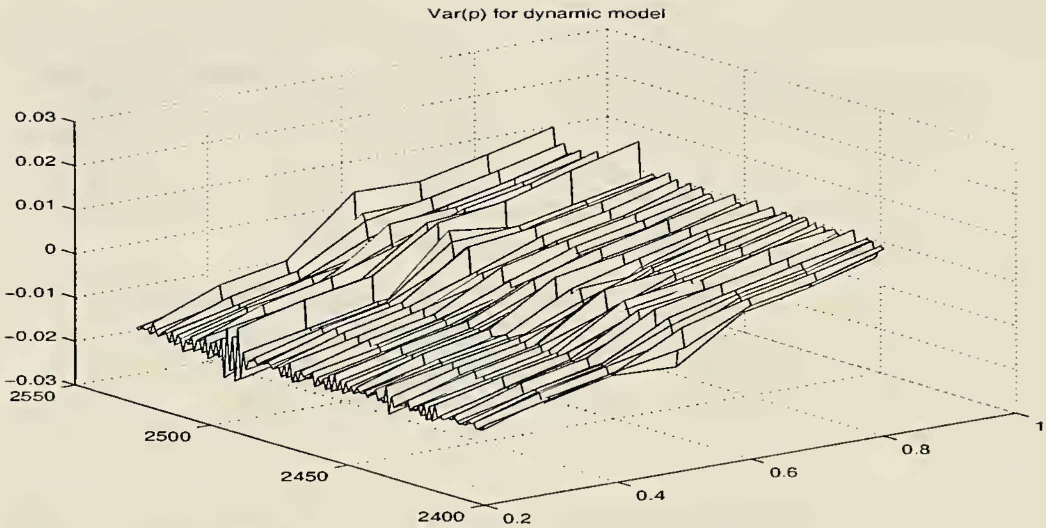

Figure 2: Recursive VaR Surface in time-probability space

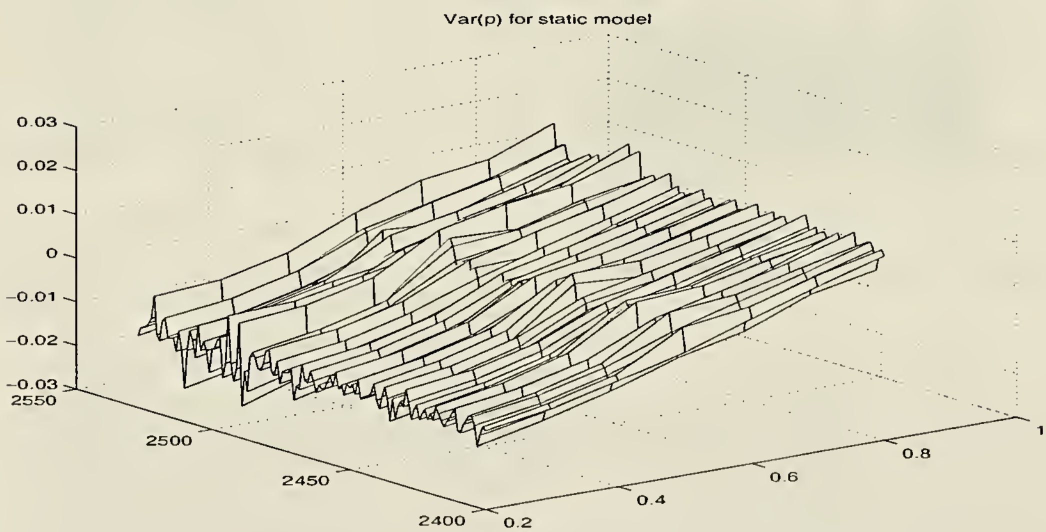

Figure 3: Non-recursive VaR Surface in time-probability space

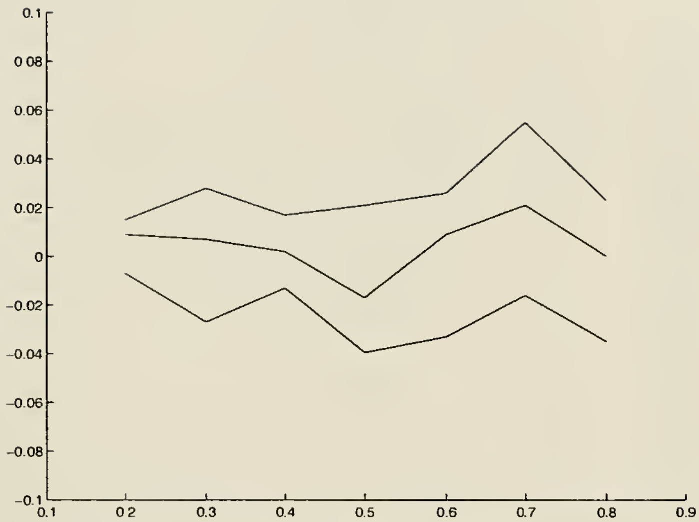

Figure 4: $\widehat{\theta}_2(\tau)$ for $\tau \in [.2, .8]$ and the 90% confidence intervals.

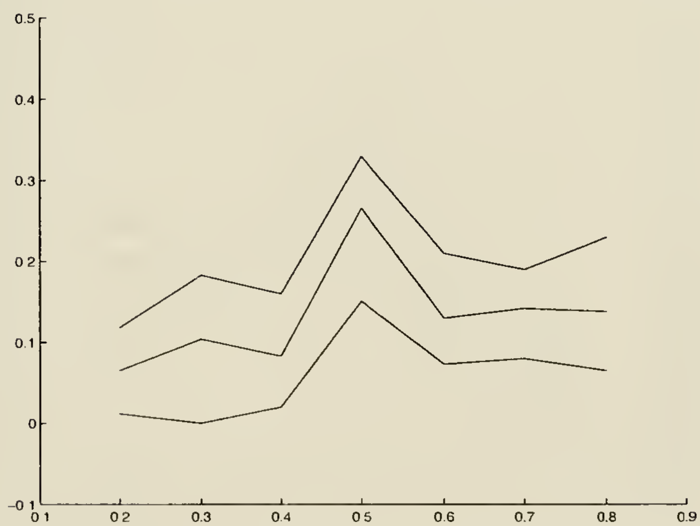

Figure 5: $\widehat{\theta}_3(\tau)$ for $\tau \in [.2, .8]$ and the 90% confidence intervals.

Figure 6: $\widehat{\theta}_4(\tau)$ for $\tau \in [.2, .8]$ and the 90% confidence intervals.

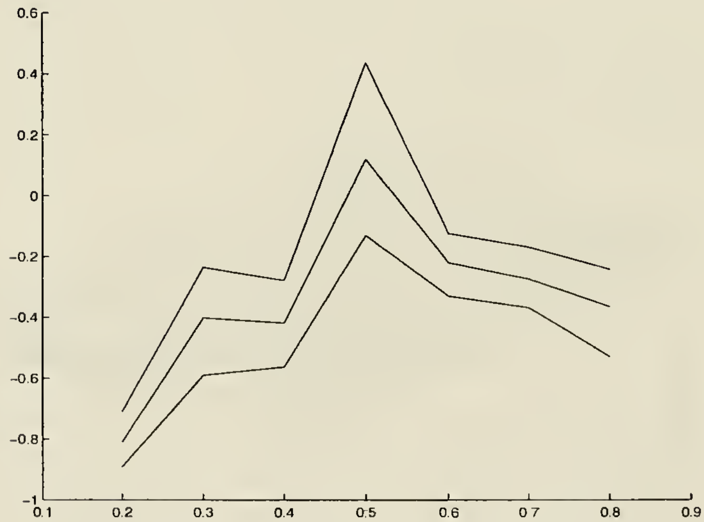

Figure 7: $\widehat{\varrho}(\tau)$ for $\tau \in [.2, .8]$ and the 90% confidence intervals.