ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 641 September 2007

On the Consistency of Approximate Maximizing Estimator Sequences in the Case of Quasiconcave Functions

Gordon Kemp

On the Consistency of Approximate Maximizing Estimator Sequences in the Case of Quasiconcave Functions

Gordon C. R. Kemp*

University of Essex

Final Version: January 2007

Abstract

This paper demonstrates consistency for estimators obtained by approximately maximizing a sequence of stochastic quasiconcave functions on $\mathbb{R}^P$ that converges in probability pointwise to a non-stochastic function. In the scalar parameter case all that is necessary for consistency is that the parameter value of interest is a unique maximizer of the limiting function. However, in the vector parameter case certain further conditions on the limiting function are necessary to establish consistency. The paper also discusses the relation of these results to existing results on the consistency of estimators obtained by approximately maximizing concave functions and to the concepts of hypoconvergence and epiconvergence.

KEYWORDS: Consistency; approximate maximizing estimator sequence; quasiconcavity; concavity; hypoconvergence; epiconvergence.

1 Introduction

Many estimators in classical econometrics are obtained as solutions, or approximate solutions, to optimization problems. One of the main tasks of classical econometric theory has been to determine conditions under which such estimators are consistent and many results of this type have been obtained to date; see, for example, Newey and McFadden (1994). One such result is as follows. Suppose that a sequence of stochastic concave functions defined on a convex subset of $\mathbb{R}^p$ with a non-empty interior converges pointwise in probability to a limiting non-stochastic function

*I would like to thank seminar participants at the University of Essex for helpful comments. Address for correspondence: Dr. Gordon C.R. Kemp, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, U.K; email: kempgcr@essex.ac.uk.

and that the limiting function achieves a global maximum at a unique point in the interior of its domain. Then a maximization estimator will exist with probability approaching one and will converge in probability to that unique point (Newey and McFadden, 1994, Theorem 2.7).[1] A nice feature of this result is that it avoids imposing compactness on the domain of the function, i.e. on the parameter space.

The intuition which Newey and McFadden give for this result is that concavity prevents the objective function from "turning up" as the parameter moves a long way from the true parameter value, i.e. from the maximizer of the limiting objective function (Newey and McFadden, 1994, page 2133). Such a "no turning up" property is not unique to concave functions and, in fact, a little reflection immediately indicates that concavity is an overly strong requirement for this result. In particular, this result will still hold if the objective function undergoes a continuous monotonically strictly increasing transformation. It is clear that pointwise convergence is preserved under such a transformation as are the existence and location of maximizers. However, concavity is not in general preserved under such transformations.

One property which is preserved under such transformations is quasiconcavity, i.e. the property that all the upper level sets of the function are convex. As is well known, all monotonically non-decreasing transformations of concave functions are necessarily quasiconcave: hence concave functions are themselves necessarily quasiconcave. However, not all quasiconcave functions can be expressed as monotonically non-decreasing transformations of concave functions (Arrow and Enthoven, 1961). The purpose of the present paper is to examine the extent to which the "no turning up" intuition can provide the basis for consistency results in the case of sequences quasiconcave functions.

As in recent treatments of convergence of sequences of stochastic convex functions, such as that of Geyer (1996), the present paper considers stochastic sequences of extended real-valued functions defined on $\mathbb{R}^p$ rather than stochastic sequences of real-valued functions defined on a fixed subset of $\mathbb{R}^p$ as in Newey and McFadden (1994). By setting the objective function to $-\infty$ at points which fail to satisfy the constraints, this permits consideration of constrained maximization estimators when the constraints themselves can be stochastic and vary with the sample size. Furthermore, since it is often impossible to compute an exact maximizing estimator in a finite number of steps, the present paper considers approximate maximizing estimator sequences as defined in Section 2 below.

The first result of the paper is that in the case of a scalar parameter, quasiconcavity of each of the functions in the sequence combined with pointwise convergence and the existence of a unique maximizer of the limiting function is sufficient to ensure the consistency of approximate maximizing estimator sequences. Thus in the scalar case the "no turning up" intuition is powerful enough on its own to drive consistency results for quasiconcave functions.

The second result of the paper is that these conditions by themselves are not sufficient to ensure the consistency of approximate maximizing estimator sequences in the case of a vector parameter. This second result is not surprising since it is possible in the vector parameter case to construct sequences of concave functions which satisfy these conditions but where approximate maximizing estimator sequences are inconsistent. However, the standard supplementary conditions used to ensure consistency in the case of concave functions, namely upper semi-continuity of the limiting function and existence of an open set on which the limiting function is finite, are not sufficient to ensure consistency in the case of quasiconcave functions.

The third result of the paper is that in the case of a vector parameter, the conditions used in the scalar parameter case in conjunction with two other conditions on the limiting function are sufficient to ensure consistency for approximate maximizing estimator sequences. These other conditions are implied by the standard supplementary conditions for sequences of concave functions when the functions are indeed concave and the limiting function possesses a unique maximizer but not necessarily otherwise. This result can be viewed as a generalization of the result for the concavity case since the standard assumptions used in that context imply the set of the conditions used here.

An important point to note is that the proofs provided here for the consistency results do not involve demonstrating that the sequence of objective functions exhibits some such form of variational convergence such as hypoconvergence or epiconvergence. Such concepts of variational convergence were popularized by Attouch (1984) and have subsequently been used in the operations research literature on stochastic optimization (Zervos, 1999) and in the statistics literature on Monte Carlo methods (Geyer, 1994). In order to establish consistency for maximizers of concave functions, Newey and McFadden (1994) apply an intermediate result to the effect that if a sequence of concave functions on $\mathbb{R}^p$ converges pointwise to a function which is finite on some open set then the sequence converges uniformly on any compact subset of that open set. Uniform convergence then implies hypoconvergence. The analysis in Geyer (1996) involves showing that if a sequence of lower semi-continuous stochastic convex functions on $\mathbb{R}^p$ converges pointwise in law on a dense subset of $\mathbb{R}^p$ to a lower semi-continuous stochastic convex function then they epiconverge in law to that function. If the functions in such a sequence are all multiplied by minus one then resulting functions are upper semi-continuous stochastic concave and hypoconverge in law to an upper semi-continuous stochastic concave function. Thus in both of these cases, hypoconvergence is an intrinsic feature of the analysis. In contrast, it is not necessary in the present paper that the sequence of quasiconcave functions hypoconverges. Indeed, Section 4 below presents an example of a sequence of continuous quasiconcave functions which converges pointwise to a limiting continuous quasiconcave function with a unique maximizer but fails to hypoconverge.

The layout of the paper is as follows. Section 2 outlines the framework used in the paper and discusses the assumptions which are made. Section 3 presents the consistency results, first for the scalar parameter case and then for the vector parameter case. Section 4 discusses the relation of pointwise convergence of quasiconcave functions to hypoconvergence. Section 5 contains concluding remarks. All proofs are given in the Appendix.

2 Framework

In what follows all probability statements are made with respect to an underlying probability space $(\Omega, \mathcal{F}, P)$. Before making any assumptions we need to define quasiconcavity. Suppose that $g(\cdot)$ is a function from $\mathbb{R}^p$ (for some $p < \infty$) to the extended real line $\overline{\mathbb{R}} = [-\infty, +\infty]$. Then $g(\cdot)$ is *quasiconcave* if:

$$g(\lambda x + (1-\lambda)x') \geq \min\{g(x), g(x')\}, \tag{1}$$

whenever $0 < \lambda < 1$ and neither $g(x)$ nor $g(x')$ is equal to $-\infty$. This is equivalent to requiring that all the upper level sets of $g(\cdot)$ are convex, where for any reference value $\alpha \in \overline{\mathbb{R}}$ of $g(\cdot)$ the corresponding upper level set of $g(\cdot)$ is defined as:

$$\mathrm{ulev}_\alpha(g) \equiv \{x \in \mathbb{R}^p : g(x) \geq \alpha\}. \tag{2}$$

Note that allowing $-\infty$ as a value for $g(\cdot)$ permits constrained problems to be treated in the same fashion as unconstrained problems; allowing $+\infty$ as a value is then mathematically convenient. Maximizing $g(\cdot)$ over $\mathbb{R}^p$ is equivalent to maximizing $g(\cdot)$ over its *effective domain* given by:

$$\mathrm{dom}(g) \equiv \{x \in \mathbb{R}^p : g(x) > -\infty\} = \bigcup_{\alpha \in (-\infty, +\infty]} \mathrm{ulev}_\alpha(g). \tag{3}$$

Also note that allowing functions to take values anywhere in $\overline{\mathbb{R}}$ means that we need to consider extended real-valued random variables on $(\Omega, \mathcal{F}, P)$, i.e. functions from Ω to $\overline{\mathbb{R}}$ such that the inverse image under the function of any element of the Borel σ-algebra on $\overline{\mathbb{R}}$ is an element of $\mathcal{F}$.

Assumption 1 *$\{Q_n(\cdot) : \mathbb{R}^p \times \Omega \to \overline{\mathbb{R}}\}_{n=1}^{\infty}$ is a sequence of functions such that:*

1. *$Q_n(\theta, \cdot)$ is an extended real-valued random variable for all $\theta \in \mathbb{R}^p$ and $n \in \mathbb{N}$;*
2. *$Q_n(\cdot, \omega)$ is quasiconcave on $\mathbb{R}^p$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$.*

Note that we could relax this assumption somewhat. In particular, suppose that the set of realizations for which $Q_n(\cdot, \omega)$ is quasiconcave on $\mathbb{R}^p$ is denoted K_n for each $n \in \mathbb{N}$. Assumption 1 requires that $K_n = \Omega$ for all $n \in \mathbb{N}$ but all the results established later in the paper would still go through if there exists a sequence of sets $\{K_n^* \in \mathcal{F}\}_{n=1}^{\infty}$ such that $K_n^* \subseteq K_n$ for each $n \in \mathbb{N}$ and that $\lim_{n\to\infty} P(K_n^*) = 1$, although the proofs would be somewhat more complex.

In this framework, there is no guarantee that the supremum of $Q_n(\cdot)$ is achieved at some $\theta \in \mathbb{R}^p$ for given $\omega \in \Omega$, i.e. that $Q_n(\cdot)$ possess a maximizer. Furthermore, and even if such a maximizer exists it may be impossible to locate within a finite number of computational steps. Therefore we use the concept of an approximate maximizing estimator sequence defined as follows. $\{\hat{\theta}_n(\cdot) : \Omega \to \mathbb{R}^p\}_{n=1}^{\infty}$ is an *approximate maximizing estimator sequence* based on $\{Q_n(\cdot) : \mathbb{R}^p \times \Omega \to \overline{\mathbb{R}}\}_{n=1}^{\infty}$ if there exist sequences of extended real-valued random variables $\{\hat{\zeta}_n\}_{n=1}^{\infty}$ and $\{\hat{\alpha}_n\}_{n=1}^{\infty}$ such that:[2]

1. $\hat{\zeta}_n$ converges in probability to 0; and
2. $\hat{\alpha}_n$ converges in probability to $+\infty$;
3. $Q_n(\hat{\theta}_n) \geq \sup_{\theta \in \mathbb{R}^p} Q_n(\theta) - \hat{\zeta}_n$ if $\sup_{\theta \in \mathbb{R}^p} Q_n(\theta) < +\infty$; and
4. $Q_n(\hat{\theta}_n) \geq \hat{\alpha}_n$ if $\sup_{\theta \in \mathbb{R}^p} Q_n(\theta) = +\infty$.

Note that allowing for extended real-valued random variables means that we need to slightly extend the usual definition of convergence in probability. More generally, suppose that Ω is a sample space equipped with a metric $\rho(\cdot, \cdot)$, $\mathcal{F}$ is the Borel σ-algebra on (Ω, ρ), and that P is a probability measure on (Ω, ρ). A sequence of extended real-valued random variables $\{X_n(\cdot)\}_{n=1}^{\infty}$ on $(\Omega, \mathcal{F}, P)$ *converges in probability* to the extended real-valued constant c if for every $\varepsilon > 0$ there exists $n_\varepsilon < \infty$ such that:

1. $P(X_n < \varepsilon^{-1}) < \varepsilon$ for all $n > n_\varepsilon$ if $c = +\infty$;
2. $P(|X_n - c| > \varepsilon) < \varepsilon$ for all $n > n_\varepsilon$ if $c \in \mathbb{R}$; and
3. $P(X_n > -\varepsilon^{-1}) < \varepsilon$ for all $n > n_\varepsilon$ if $c = -\infty$.

This notion of convergence in probability for extended real-valued random variables can be viewed as a special case of convergence in probability for random sequences taking values in metric spaces. In this particular case, the metric space consists of all real numbers together with the additional elements $+\infty$ and $-\infty$ and equipped with the metric $d(x, y) = |\Phi(x) - \Phi(y)|$, where $\Phi(\cdot)$ is a fixed, bounded, strictly increasing continuous function, e.g. the normal cumulative distribution function, as in Example 18.4 from van der Vaart (1998).

Observe that approximate maximizing estimator sequences based on $\{Q_n(\cdot) : \mathbb{R}^p \times \Omega \to \mathbb{R}\}_{n=1}^{\infty}$ will always exist since for any sequence of strictly positive constants $\{\zeta_n\}_{n=1}^{\infty}$ tending to zero and any sequence of constants $\{\alpha_n\}_{n=1}^{\infty}$ tending to $+\infty$ we can always find a sequence of functions $\{\hat{\theta}_n^*(\cdot) : \Omega \to \mathbb{R}^p\}_{n=1}^{\infty}$ such that $Q_n(\hat{\theta}_n^*) \geq \sup_{\theta \in \mathbb{R}^p} Q_n(\theta) - \zeta_n$ if $\sup_{\theta \in \mathbb{R}^p} Q_n(\theta) < +\infty$ and that $Q_n(\hat{\theta}_n^*) \geq \alpha_n$ if $\sup_{\theta \in \mathbb{R}^p} Q_n(\theta) = +\infty$. However, nothing in the definition imposes a requirement that such estimators need be random variables.

Assumption 2 *There exists a non-random function $Q(\cdot) : \mathbb{R}^p \to \overline{\mathbb{R}}$ such that for all $\theta \in \mathbb{R}^p$, $\{Q_n(\theta, \cdot)\}_{n=1}^{\infty}$ converges in probability to $Q(\theta)$.*

Assumptions 1 and 2 are sufficient to ensure the following useful result.

Lemma 1 *Under Assumptions 1 and 2, $Q(\cdot)$ is quasiconcave on $\mathbb{R}^p$.*

Thus quasiconcavity (like concavity) is preserved under pointwise convergence in probability.

The third assumption specifies that the parameter value of interest is characterized as the unique maximizer of the limiting objective function.

Assumption 3 *There exists $\theta_0 \in \mathbb{R}^p$ such that $Q(\theta) \leq Q(\theta_0)$ for all $\theta \in \mathbb{R}^p$ and $Q(\theta) = Q(\theta_0)$ if and only if $\theta = \theta_0$.*

This clearly implies that θ_0 belongs to the effective domain of of $Q(\cdot)$.

Since, as noted above, the estimators being considered are not necessarily random variables the consistency results presented in Section 3 require a generalized notion of weak consistency. This is given as follows. A sequence of not necessarily measurable estimators $\{\tilde{\theta}_n(\cdot) : \Omega \to \mathbb{R}^p\}_{n=1}^{\infty}$ is *weakly consistent* in the *generalized sense* if for each $\delta > 0$ there exists a sequence of sets $\{C_n(\delta) \in \mathcal{F}\}_{n=1}^{\infty}$ such that:

1. $\lim_{n\to\infty} P(C_n(\delta)) = 1$; and

2. $\|\tilde{\theta}_n(\omega) - \theta_0\| \leq \delta$ for all $\omega \in C_n(\delta)$ and $n \in \mathbb{N}$.

This definition is equivalent to requiring that for any $\delta > 0$ the outer measure of $\{\omega \in \Omega : \|\tilde{\theta}_n(\omega) - \theta_0\| > \delta\}$ converges to zero as n tends to infinity. Such a generalized concept of weak consistency is widely used in asymptotic theory to circumvent measurability issues; see for example Footnote 9 of Newey and McFadden (1994) and Section 18.2 of van der Vaart (1998). The formulation given here parallels the concept of convergence in probability in the generalized sense given by Amemiya (1985, p. 340). Observe that if the elements of $\{\tilde{\theta}_n(\cdot) : \Omega \to \mathbb{R}^p\}_{n=1}^{\infty}$ were random variables then $\{\omega \in \Omega : \|\tilde{\theta}_n(\omega) - \theta_0\| \leq \delta\}$ would belong to $\mathcal{F}$ and hence $\tilde{\theta}_n$ would be weakly consistent in the usual sense.

As will be demonstrated in Section 3 below, when $p = 1$ then Assumptions 1–3 are sufficient to ensure that any approximate maximizing estimator sequence based on $Q_n(\cdot)$ is weakly consistent for θ_0 in the generalized sense. However, when $p > 1$ these assumptions are not sufficient to establish consistency of such estimator sequences. Two additional assumptions which have been used in the literature to establish consistency in the case of sequences of stochastic concave functions are as follows. First, there is a non-empty open set on which the pointwise limiting function is finite. Second, the realizations of the stochastic objective functions and the pointwise limiting function are all upper semi-continuous: recall that a function $g(\cdot) : \mathbb{R}^p \to \overline{\mathbb{R}}$ is *upper semi-continuous* if for any $\theta \in \mathbb{R}^p$:

$$\limsup_{m\to\infty} g(\theta_m) \leq g(\theta) \quad \text{whenever} \quad \lim_{m\to\infty} \theta_m = \theta. \tag{4}$$

Note that this is equivalent to requiring that the *hypograph* of $g(\cdot)$, given as:

$$\text{hyp}(g) = \{(x, \alpha) \in \mathbb{R}^p \times \mathbb{R} : g(x) \geq \alpha\}, \tag{5}$$

is closed. Note that if a concave functions is finite on some open set then it is continuous and hence upper semi-continuous on that open set. Thus a lack of upper semi-continuity is only important for a concave function on the boundary of its effective domain. These two assumptions combined with Assumptions 1–3 and the condition that the $Q_n(\cdot)$ are all concave are sufficient to ensure, first, that the $Q_n(\cdot)$ sequence hypoconverges in law to $Q(\cdot)$ and, second, that approximate maximizing estimator sequences are consistent; see, for example, Lemma 3.1 and Theorem 3.2 of Geyer (1996).

However, as will be demonstrated in Section 3 below, in the quasiconcavity case it is possible to construct examples of sequences of functions which satisfy Assumptions 1–3 together these two additional assumptions but where approximate maximizing sequences are inconsistent. Consequently, instead of the upper semi-continuity and finiteness on an open set assumptions we will make the two other assumptions. The first of these is a supplement to Assumption 3.

Assumption 4 *For every $\delta > 0$ there exists $\eta > 0$ such that:*

$$Q(\theta) \leq \sup_{\theta' \neq \theta_0} Q(\theta') - \eta, \quad \forall \ \|\theta - \theta_0\| \geq \delta. \tag{6}$$

This is closely related to the concept of identifiable uniqueness of maximizers: here θ_0 is an *identifiably unique maximizer* of $Q(\cdot)$ if for every $\delta > 0$ there exists $\eta > 0$ such that:

$$Q(\theta) \leq Q(\theta_0) - \eta, \quad \forall \ \|\theta - \theta_0\| \geq \delta. \tag{7}$$

Under Assumption 3 it is clear that $\sup_{\theta' \neq \theta_0} Q(\theta') \leq Q(\theta_0)$ and hence Assumptions 3 and 4 jointly imply identifiable uniqueness. However, if $\sup_{\theta' \neq \theta_0} Q(\theta') < Q(\theta_0)$ then θ_0 is necessarily an identifiably unique maximizer of $Q(\cdot)$ regardless of whether or not Assumption 4 holds.

Assumption 5 *Define:*

$$A \equiv \{\alpha \in \mathbb{R} : \exists \ \theta \in \text{ulev}_\alpha(Q) \ \ s.t. \ \ \theta \neq \theta_0\}; \tag{8}$$

then A is non-empty and for all $\alpha \in A$, $\text{ulev}_\alpha(Q)$ has a non-empty interior.

In the case of concavity these two additional assumptions are implied by the upper semi-continuity and finiteness on an open set assumptions in conjunction with the existence of a unique maximizer, as can be seen from the following lemma.

Lemma 2 *Suppose that $g(\cdot) : \mathbb{R}^p \to \overline{\mathbb{R}}$ is an upper semi-continuous concave function which satisfies Assumption 3 and is finite on some non-empty open subset of $\mathbb{R}^p$; then $g(\cdot)$ satisfies Assumptions 4 and 5.*

It should be noted that the assumptions made above do not impose any requirement that the maximum of the limiting function be achieved in the interior of the effective domain of the limiting function though Assumption 5 does require that the interior of the effective domain of the limiting function is non-empty. Thus the framework used here does permit cases in which the parameter value of interest is characterized as solution to an inequality constrained optimization problems in which the constraints bind. Furthermore, the constraints may vary depending on the sample size and realized outcome.

3 Results

3.1 The Scalar Parameter Case ($p = 1$)

The scalar parameter case is considerably easier to deal with than the vector parameter case for reasons discussed below.

Theorem 1 (Consistency I) *If $p = 1$ and Assumptions 1–3 are satisfied then any sequence of approximate maximizing estimators of θ_0 based on $\{Q_n(\cdot) : \mathbb{R}^p \times \Omega \to \mathbb{R}\}_{n=1}^{\infty}$ is weakly consistent in the generalized sense.*

The essence of the proof of consistency for the scalar parameter case ($p = 1$) is that if one takes any δ-neighborhood of θ_0 then for every element θ' of $\mathbb{R}^p$ which lies outside that δ-neighborhood there exists an element of the boundary of that δ-neighborhood which is a convex combination of θ_0 and θ'. Hence if $Q_n(\theta_0)$ is at least greater by some positive amount than the supremum of $Q_n(\theta)$ over the boundary of that δ-neighborhood then quasiconcavity implies that $Q_n(\theta_0)$ must be at least greater by that same amount than the supremum of $Q_n(\theta)$ over the entire subset of $\mathbb{R}^p$ in which $\|\theta - \theta_0\| \geq \delta$. It is this property which captures the "no turning up" intuition. The convenient feature of the scalar parameter case is that the boundary of any δ-neighborhood only has two elements and thus pointwise convergence in probability of $Q_n(\theta)$ to $Q(\theta)$ immediately implies convergence in probability of the supremum of $Q_n(\theta)$ over the boundary of a δ-neighborhood to the supremum of $Q(\theta)$ over the boundary of that δ-neighborhood.

3.2 The Vector Parameter Case ($p > 1$)

The main result in the vector parameter case is the following theorem.

Theorem 2 (Consistency II) *If $p > 1$ and Assumptions 1–5 are satisfied then any sequence of approximate maximizing estimators of θ_0 based on $\{Q_n(\cdot) : \mathbb{R}^p \times \Omega \to \overline{\mathbb{R}}\}_{n=1}^{\infty}$ is weakly consistent in the generalized sense.*

It is important to observe that in the vector parameter case, Assumptions 1–3 are not sufficient to establish consistency for sequences of approximate maximizing estimators of θ_0 as can be seen from the following examples.

Example 1 *Suppose $p = 2$ and that for each $n \in \mathbb{N}$, $k_n(\cdot) : \mathbb{R}^2 \to \mathbb{R}$ is given by:*

$$k_n(\theta_1, \theta_2) = \begin{cases} \theta_1, & \text{if } 0 \leq \theta_1 = n\theta_2 \leq 1; \\ -1, & \text{otherwise.} \end{cases} \tag{9}$$

For each $n \in \mathbb{N}$, the upper levels set of $k_n(\cdot)$ are then given by

$$\text{ulev}_\alpha(k_n) = \begin{cases} \mathbb{R}^2, & -\infty \leq \alpha \leq -1; \\ \{(\theta_1, \theta_2) \in \mathbb{R}^2 : 0 \leq \theta_1 = n\theta_2 \leq 1\}, & -1 < \alpha \leq 0; \\ \{(\theta_1, \theta_2) \in \mathbb{R}^2 : \alpha \leq \theta_1 = n\theta_2 \leq 1\}, & 0 < \alpha \leq 1; \\ \emptyset, & 1 < \alpha \leq \infty, \end{cases} \tag{10}$$

and the effective domain of $k_n(\cdot)$ is equal to $\mathbb{R}^2$. It is straightforward to verify $\text{ulev}_\alpha(k_n)$ is convex for each $n \in \mathbb{N}$ and $\alpha \in \overline{\mathbb{R}}$ which is equivalent to each of the functions $k_n(\cdot)$ being quasiconcave. It is also easy to verify that each $k_n(\cdot)$ satisfies Equation (4) for each $\theta \in \mathbb{R}^2$ and thus that the $k_n(\cdot)$ functions are all upper semi-continuous. Furthermore, they are all finite on $\mathbb{R}^2$ which is an open set.

Next, consider any fixed point (θ_1, θ_2). If $\theta_1 = \theta_2 = 0$ then $k_n(\theta_1, \theta_2) = 0$ for all $n \in \mathbb{N}$ in which case $\lim_{n\to\infty} k_n(\theta_1, \theta_2) = 0$. If there exists $m \in \mathbb{N}$ such that $0 < \theta_1 = m\theta_2 \leq 1$ then $k_n(\theta_1, \theta_2) = \theta_1$ for $n = m$ and is equal to -1 otherwise in which case $\lim_{n\to\infty} k_n(\theta_1, \theta_2) = -1$. Finally, if $\theta_1 \neq 0$ and $\theta_2 \neq 0$ but there is no $m \in \mathbb{N}$ such that $0 < \theta_1 = m\theta_2 \leq 1$ then $k_n(\theta_1, \theta_2) = -1$ for all $n \in \mathbb{N}$ in which case $\lim_{n\to\infty} k_n(\theta_1, \theta_2) = -1$.

This establishes that the sequence of functions $\{k_n(\cdot) : \mathbb{R}^2 \to \mathbb{R}\}_{n=1}^{\infty}$ converges pointwise to the function $k(\cdot) : \mathbb{R}^2 \to \mathbb{R}$ given by:

$$k(\theta_1, \theta_2) = \begin{cases} 0, & \text{if } \theta_1 = \theta_2 = 0; \\ -1, & \text{otherwise;} \end{cases} \tag{11}$$

which achieves a unique global maximum at the point $(0, 0)$. The upper level sets of $k(\cdot)$ are given by:

$$\text{ulev}_\alpha(k) = \begin{cases} \emptyset, & 0 < \alpha \leq +\infty \\ \{(0,0)\}, & -1 < \alpha \leq 0; \\ \mathbb{R}^2, & -\infty \leq \alpha \leq -1; \end{cases} \tag{12}$$

and the effective domain of $k(\cdot)$ is $\mathbb{R}^2$. Since these are all convex sets it follows that $k(\cdot)$ is quasiconcave. In addition, it is easy to verify that $k(\cdot)$ satisfies Equation (4) for each $\theta \in \mathbb{R}^2$, so that $k(\cdot)$ is upper semi-continuous, and that $k(\cdot)$ is finite on $\mathbb{R}^2$ which is an open set.

However, for each $n \in \mathbb{N}$, $k_n(\cdot)$ achieves a unique global maximum at the point $\theta^n = (1, n^{-1})$ and clearly $\{\theta^n\}_{n=1}^{\infty}$ converges to $(1,0)$ as n tends to infinity. Thus in this example, the limit of the sequence of maximizers and the maximizer of the limiting function both exist and are unique but are clearly different from each other even though all the functions in the sequence are quasiconcave and the maximizer of the limiting function is unique.

Note that although this is a deterministic example it would be trivial to create a stochastic version from it. Careful examination of Example 1 reveals that the limiting function $k(\cdot)$ fails to satisfy Assumption 4 since although $k(\cdot)$ is uniquely maximized at $\theta = (0,0)$ nevertheless $k(\theta) = \sup_{\theta' \neq (0,0)} k(\theta') = -1$ for all $\theta \neq (0,0)$. However, $k(\cdot)$ does satisfy Assumption 5: here $A = (-\infty, -1]$ and it is clear that $\text{ulev}_\alpha(k) = \mathbb{R}^2$ for all $-\infty < \alpha \leq -1$. Further examination of Example 1 reveals that the problem which arises here is closely related to an issue which can arise when dealing with sequences of concave functions. For each $n \in \mathbb{N}$, define $k_n^*(\theta) = k_n(\theta)$ if $k_n(\theta) \geq 0$ and $k_n^*(\theta) = -\infty$ if $k_n(\theta) < 0$. It is easy to see that $k_n^*(\cdot)$ is both concave, since it is linear on a convex set and equal to $-\infty$ everywhere else, and that the maximizer of $k_n^*(\cdot)$ is the same as that of $k_n(\cdot)$. Furthermore, $k_n^*(\cdot)$ converges pointwise to $k^*(\cdot)$ given by $k^*(\theta) = k(\theta)$ if $k(\theta) \geq 0$ and $k^*(\theta) = -\infty$ if $k(\theta) < 0$, and it is easy to see that $k^*(\cdot)$ is concave and that the maximizer of $k^*(\cdot)$ is the same as that of $k(\cdot)$. However, although $k^*(\cdot)$ and all the $k_n^*(\cdot)$ are upper-semi continuous, there is no open set on which $k^*(\cdot)$ is finite.

However, Assumptions 1–4 are also not sufficient to ensure consistency of approximating maximizing estimator sequences as can be seen from this second example.

Example 2 *Suppose $p = 2$ and that for each $n \in \mathbb{N}$, $k_n(\cdot) : \mathbb{R}^2 \to \mathbb{R}$ is given by:*

$$k_n(\theta_1, \theta_2) = \begin{cases} n\theta_2 - \theta_1, & \text{if } 0 \leq \theta_1 \leq 1 \ \& \ 0 \leq \theta_2 \leq 2n^{-1}\theta_1; \\ -1, & \text{otherwise.} \end{cases} \tag{13}$$

For each $n \in \mathbb{N}$ it is clear that $\{(\theta_1, \theta_2) \in \mathbb{R}^2 : \ 0 \leq \theta_1 = 1 \ \& \ 0 \leq \theta_2 \leq 2n^{-1}\theta_1\}$ is a compact convex set and that $k_n(\cdot)$ is linear on this set with a supremum over this set equal to 1 which is achieved at $(0,1)$ and an infimum over this set equal to -1 which is achieved at $(1,0)$. Outside of this set $k_n(\cdot)$ is constant with a value of -1 which is thus the global infimum of $k_n(\cdot)$. It follows that all of the upper level sets of $k_n(\cdot)$ are convex and hence $k_n(\cdot)$ is quasiconcave. Clearly each $k_n(\cdot)$ is finite on $\mathbb{R}^2$ which is an open set and it is easy to verify that each $k_n(\cdot)$ satisfies Equation (4) for each $\theta \in \mathbb{R}^2$ and hence is upper semi-continuous.

Now consider any fixed point (θ_1, θ_2). If $0 \leq \theta_1 \leq 1$ and $\theta_2 = 0$ then $k_n(\theta_1, \theta_2)$ is equal to $-\theta_1$ for all n and hence converges to $-\theta_1$. If $\theta_1 < 0$ or $\theta_1 > 1$ or $\theta_2 \neq 0$ then for all n sufficiently large $k_n(\theta_1, \theta_2)$ is equal to -1 and hence converges to -1.

This establishes that the sequence of functions $\{k_n(\cdot) : \mathbb{R}^2 \to \mathbb{R}\}_{n=1}^{\infty}$ *converges pointwise to the function* $k(\cdot) : \mathbb{R}^2 \to \mathbb{R}$ *given by:*

$$k(\theta_1, \theta_2) = \begin{cases} -\theta_1, & \text{if } 0 \leq \theta_1 \leq 1 \ \& \ \theta_2 = 0; \\ -1, & \text{otherwise;} \end{cases} \tag{14}$$

which achieves a unique global maximum at the point $(0, 0)$*. The upper level sets of* $k(\cdot)$ *are given by:*

$$\text{ulev}_\alpha(k) = \begin{cases} \emptyset, & 0 < \alpha \leq +\infty \\ \{(\theta_1, 0) : 0 \leq \theta_1 \leq -\alpha\}, & -1 < \alpha \leq 0; \\ \mathbb{R}^2, & -\infty \leq \alpha \leq -1; \end{cases} \tag{15}$$

and the effective domain of $k(\cdot)$ *is* $\mathbb{R}^2$*. Since these are all convex sets it follows that* $k(\cdot)$ *is quasiconcave. In addition, it is easy to verify that* $k(\cdot)$ *satisfies Equation* (4) *for each* $\theta \in \mathbb{R}^2$*, so that* $k(\cdot)$ *is upper semi-continuous, and that* $k(\cdot)$ *is finite on* $\mathbb{R}^2$ *which is an open set.*

However, $k(\cdot)$ *is uniquely maximized at* $(0, 0)$ *while each of the* $k_n(\cdot)$ *is uniquely maximized at* $(1, 2n^{-1})$ *which converges to* $(1, 0)$*.*

Careful examination of Example 2 reveals that the limiting function $k(\cdot)$ satisfies Assumption 4: here $k(0,0) = \sup_{\theta \neq (0,0)} k(\theta) = 0$ and for any $\delta > 0$, $\sup_{\theta : \|\theta\| \geq \delta} k(\theta) = \max\{-1, -\delta\}$. However, $k(\cdot)$ does not satisfy Assumption 5: here $A = (-\infty, 0)$ and it is clear that $\text{ulev}_\alpha(k)$ is contained in a linear subspace for all $-1 < \alpha < 0$. Further examination of Example 2 reveals that again the problem which arises here is closely related to an issue which can arise when dealing with sequences of concave functions. For each $n \in \mathbb{N}$, define $k_n^*(\theta) = k_n(\theta)$ if $0 \leq \theta_1 \leq 1$ and $0 \leq \theta_2 \leq 2n^{-1}\theta_1$ and $k_n^*(\theta) = -\infty$ otherwise. It is easy to see that $k_n^*(\cdot)$ is both concave, since it is linear on a convex set and equal to $-\infty$ everywhere else, and that the maximizer of $k_n^*(\cdot)$ is the same as that of $k_n(\cdot)$. Furthermore, $k_n^*(\cdot)$ converges pointwise to $k^*(\cdot)$ given by $k^*(\theta) = k(\theta)$ if $0 \leq \theta_1 \leq 1$ and $\theta_2 = 0$ and $k^*(\theta) = -\infty$ everywhere else, and it is easy to see that $k^*(\cdot)$ is concave and that the maximizer of $k^*(\cdot)$ is the same as that of $k(\cdot)$. However, although $k^*(\cdot)$ and all the $k_n^*(\cdot)$ are upper-semi continuous, there is no open set on which $k^*(\cdot)$ is finite.

Examples 1 and 2 thus indicate that even exact maximizing sequences can fail to be consistent if either Assumption 4 or Assumption 5 fails to hold. However, if we assume both Assumption 4 and 5 in conjunction with the assumptions made in the scalar parameter case then we can ensure that all approximate maximizing estimator sequences are consistent as indicated above in Theorem 2.

4 Discussion

There are a number of points of interest to note about the results obtained in Section 3 above. First, as observed in the Introduction, concavity is a special case of quasiconcavity since any monotonically non-decreasing transformation of a concave function is quasiconcave. Since convergence in probability is preserved under continuous transformations it follows that if a sequence of stochastic concave functions converges pointwise in probability then applying a given continuous monotonically non-decreasing transformation to each member of that sequence will result in a sequence of stochastic quasiconcave functions which converge pointwise in probability. Furthermore, the location of any maximizer of the original function is preserved under such transformations. Thus the existing results on consistency of maximization estimators based on concave functions imply consistency of maximization estimators based on at least some sequences of quasiconcave functions. However, as is well known from Arrow and Enthoven (1961), not all quasiconcave functions can be obtained monotonically non-decreasing transformations of concave functions. Hence the results obtained in Section 3 cannot be reduced simply to the application of the consistency result for concave objective functions in combination with continuous increasing transformations.

A second point to note is that, unlike the usual practice for concave functions, we have not assumed that the functions of interest are upper semi-continuous. In maximization theory it is often convenient to impose that the function being maximized is upper semi-continuous as this ensures that the supremum of the function over any given compact set is actually achieved at some point belonging to that set. A concave function on $\mathbb{R}^p$ is necessarily continuous, and thus upper semi-continuous, on the interior of its effective domain. Hence imposing upper semi-continuity on a concave function on $\mathbb{R}^p$ only really affects the behavior of that function on the boundary of its effective domain. However, a quasiconcave function on $\mathbb{R}^p$ need not be upper semi-continuous everywhere in the interior of its effective domain. Hence imposing upper semi-continuity on a quasiconcave function on $\mathbb{R}^p$ may affect the behavior of that function on the interior as well as the boundary of its effective domain.

A third point to note is that the proofs of Theorems 1 and 2 do not involve showing that the sequence of stochastic quasiconcave objective functions exhibits any form of variational convergence such as continuous convergence, i.e. uniform convergence to a continuous function, or hypoconvergence. This is on contrast to the existing literature on convergence of optimization estimators based on concave (or convex) functions which makes heavy use of such forms of variational convergence. For example, the proof of Theorem 2.7 from Newey and McFadden (1994) involves using a result to the effect that if a sequence of concave functions on $\mathbb{R}^p$ converges pointwise on a dense subset of an open set on which the limiting function is finite then it converges uniformly to that limiting function on any compact subset of that open set (Rockafellar, 1970, Theorem 10.8).[3] Newey and McFadden combine this with the well-known results that the pointwise limit of a sequence of concave functions is itself concave and that concave functions over $\mathbb{R}^p$ are continuous on open sets on which they are finite to establish that the sequence of functions exhibits continuous convergence to the pointwise limit on compact subsets of the interior of the parameter space.

The arguments presented by Kall (1986) indicate that such continuous convergence only matters because it implies hypoconvergence. Recall that a sequence of functions $\{g_n(\cdot) : \mathbb{R}^p \to \overline{R}\}_{n=1}^{\infty}$ *hypoconverges* to the function $\{g(\cdot) : \mathbb{R}^p \to \overline{R}\}_{n=1}^{\infty}$ at the point $x \in \mathbb{R}^p$ if:

$$\limsup_{n\to\infty} g_n(x_n) \leq g(x) \text{ for all } \{x_n\}_{n=1}^{\infty} \text{ s.t. } \lim_{n\to\infty} x_n = x; \text{ and} \tag{16}$$

$$\liminf_{n\to\infty} g_n(y_n) \geq g(x) \text{ for some } \{y_n\}_{n=1}^{\infty} \text{ s.t. } \lim_{n\to\infty} y_n = x; \tag{17}$$

see Attouch (1984, p. 30). Hypoconvergence is useful in analyzing the behavior of maximizers because of the following result implied by Kall (1986, Corollary 3). Suppose that $\{g_n(\cdot) : \mathbb{R}^p \to \overline{R}\}_{n=1}^{\infty}$ hypoconverges to $g(\cdot) : \mathbb{R}^p \to \overline{R}$ at every point $x \in \mathbb{R}^p$, that the effective domain of $g(\cdot)$ has a non-empty interior, that the maximum of $g(\cdot)$ over $\mathbb{R}^p$ is finite and is achieved at least one point, and that $\{\hat{x}_n\}_{n=1}^{\infty}$ is a sequence of maximizers of $\{g_n(\cdot)\}_{n=1}^{\infty}$; then any point of accumulation of $\{\hat{x}_n\}_{n=1}^{\infty}$ must be a maximizer of $g(\cdot)$.

Hypoconvergence is closely related to upper semi-continuity. Thus if $\{g_n(\cdot)\}_{n=1}^{\infty}$ hypoconverges to $g(\cdot)$ at every point $x \in \mathbb{R}^p$ then $g(\cdot)$ must be upper semi-continuous on $\mathbb{R}^p$. Furthermore, if $g_n(\cdot) \equiv g(\cdot)$ for all $n \in \mathbb{N}$ then $\{g_n(\cdot)\}_{n=1}^{\infty}$ hypoconverges to $g(\cdot)$ at every point $x \in \mathbb{R}^p$ if and only if $g(\cdot)$ is upper semi-continuous on $\mathbb{R}^p$. However, in the context of quasiconcavity, there are two drawbacks to making use of hypoconvergence. First, even if a sequence of upper semi-continuous quaisconcave functions converges pointwise to a function whose effective domain has a non-empty interior, that limiting function need not be upper semi-continuous on the interior of its effective domain and hence the sequence of functions need not hypoconverge to the pointwise limiting function.[4] Second, even a sequence of continuous quasiconcave functions which converges pointwise to a continuous quasiconcave function with a unique argmax need not hypoconverge at every point.

Example 3 *Suppose that:*

$$f_n(x) = n^{-1} - |x - n^{-1}|, \quad \textit{for all } x \in \mathbb{R} \textit{ and } n \in \mathbb{N}, \tag{18}$$

$$h_n(z) = \begin{cases} 2nz, & \textit{if } z \geq 0 \ \& \ n = 2k \textit{ for some } k \in \mathbb{N}; \\ nz, & \textit{if } z \geq 0 \ \& \ n = 2k+1 \textit{ for some } k \in \mathbb{N}; \\ z, & z \text{ ¡ 0.} \end{cases} \tag{19}$$

Then set $g_n(x) = h_n(f_n(x))$ *and observe that:*

$$g_n(x) = \begin{cases} 2n^{-1} - x, & \text{if } x \geq 2n^{-1}; \\ 4 - 2nx, & \text{if } n^{-1} \leq x < 2n^{-1} \ \& \ n = 2k \text{ for some } k \in \mathbb{N}; \\ 2 - nx, & \text{if } n^{-1} \leq x < 2n^{-1} \ \& \ n = 2k+1 \text{ for some } k \in \mathbb{N}; \\ 2nx, & \text{if } 0 \leq x < n^{-1} \ \& \ n = 2k \text{ for some } k \in \mathbb{N}; \\ nx, & \text{if } 0 \leq x < n^{-1} \ \& \ n = 2k+1 \text{ for some } k \in \mathbb{N}; \\ x, & \text{if } x < 0. \end{cases} \tag{20}$$

Since $f_n(\cdot)$ *and* $h_n(\cdot)$ *are continuous with respect to* $x \in \mathbb{R}$ *for each* $n \in \mathbb{N}$ *then so to is* $g_n(\cdot)$*. Furthermore, since each* $f_n(\cdot)$ *is concave and each* $h_n(\cdot)$ *is continuous and strictly increasing in* $x \in \mathbb{R}$ *for all* $n \in \mathbb{N}$ *then each* $g_n(\cdot)$ *is quasiconcave* $x \in \mathbb{R}$ *for all* $n \in \mathbb{N}$*. For any given* $\bar{x} \in \mathbb{R}$ *if* $\bar{x} \leq 0$ *then* $g_n(\bar{x}) = \bar{x}$ *for all* n *and hence* $\lim_{n\to\infty} g_n(\bar{x}) = \bar{x}$*. If* $\bar{x} > 0$ *then* $g_n(\bar{x}) = 2n^{-1} - \bar{x}$ *for all* $n \geq 2\bar{x}^{-1}$ *and hence* $\lim_{n\to\infty} g_n(\bar{x}) = -\bar{x}$ *as* $\lim_{n\to\infty} n^{-1} = 0$*. Putting these together implies that* $g_n(x)$ *converges pointwise to* $-|x|$*.*

However, this sequence of functions fails to hypoconverge at $x = 0$*. Observe that for each* $n \in \mathbb{N}$*,* $g_n(x)$ *possesses a unique maximizer at* $x = n^{-1}$ *since this is the value of* x *at which* $f_n(x)$ *is maximized and since* $h_n(z)$ *is monotonically strictly increasing in* z*. In addition, observe that* $g_n(n^{-1}) = 2$ *if* $n = 2k$ *for some* $k \in \mathbb{N}$ *while* $g_n(n^{-1}) = 1$ *if* $n = 2k - 1$ *for some* $k \in \mathbb{N}$*. Let* $x_n^* = n^{-1}$ *for each* $n \in \mathbb{N}$*; then* $\lim x_n^* = 0$ *and* $\limsup_{n\to\infty} g_n(x_n^*) = 2$*. However, for any sequence* y_n *such that* $\lim y_n = 0$ *it is easy to see that:*

$$\lim \inf_{n\to\infty} g_n(y_n) \leq \lim \inf_{n\to\infty} g_n(x_n) = 1 < 2 = \lim \sup_{n\to\infty} g_n(x_n^*), \tag{21}$$

since $g_n(y_n) \leq g_n(x_n^*)$ *for any* $y_n \in \mathbb{R}$ *and* $n \in \mathbb{N}$*, and hence* $g_n(x)$ *fails to hypoconverge at* $x = 0$*.*

This example is somewhat pathological in that, although the sequence of functions fails to hypoconverge, nevertheless the sequence of maximizers converges to the maximizer of the pointwise limiting function. However, one benefit of Theorem 2 is precisely that it does not require the imposition of assumptions which exclude such cases.

5 Conclusion

This paper has demonstrated the consistency of approximate maximizing estimator sequences for stochastic extended real-valued quasiconcave functions which converge pointwise and satisfy certain other conditions. In the scalar parameter case, the only such condition which is needed is that the pointwise limiting function has a unique maximizer which occurs at the parameter value of interest and is finite at this parameter value. In the vector parameter case, somewhat stronger conditions on the pointwise limiting function are needed. The paper demonstrates that these stronger conditions are automatically satisfied when the conditions used in the scalar parameter case are met and the limiting function is an upper semi-continuous concave function which is finite on some non-empty open set. However, these stronger conditions do not require that the limiting function be either concave or upper-semi continuous. In addition, the assumptions made in the present paper do not imply that the sequence of functions hypoconverges and thus hypoconvergence is not crucial to the theory of consistency presented here.

Appendix

Proof of Lemma 1. Select any $\theta, \theta' \in \mathrm{dom}(Q)$ and $0 < \lambda < 1$; then set $\theta_\lambda = \lambda\theta + (1-\lambda)\theta'$ and define:

$$\zeta_n(\omega) \equiv Q_n(\theta_\lambda, \omega) - \min\{Q_n(\theta, \omega), Q_n(\theta', \omega)\}, \tag{22}$$

$$\zeta \equiv Q(\theta_\lambda) - \min\{Q(\theta), Q(\theta')\}. \tag{23}$$

Assumption 1 implies that $P(\zeta_n \geq 0) = 1$ while Assumption 2 implies that ζ_n converges in probability to ζ since $\min\{\cdot,\cdot\}$ is a continuous function. It follows that $\zeta \geq 0$. But this implies that $Q(\cdot)$ is an extended real-valued function on $\mathbb{R}^p$ such that:

$$Q(\lambda\theta + (1-\lambda)\theta') \geq \min\{Q(\theta), Q(\theta')\} \tag{24}$$

for arbitrary $\theta, \theta' \in \mathbb{R}^p$ such that $Q(\theta), Q(\theta') > -\infty$ and $0 < \lambda < 1$ and hence is quasiconcave as desired. ■

Proof of Lemma 2. First, let θ_0 denote the unique value of $\theta \in \mathbb{R}^p$ at which $g(\theta)$ is maximized by Assumption 3. Since $g(\cdot)$ is finite on some open set then we can find $\theta', \theta'' \in \mathbb{R}^p$ and $0 < \lambda < 1$ such that $g(\theta'), g(\theta'') \in \mathbb{R}$ and $\theta'' = \lambda\theta' + (1-\lambda)\theta_0$. But since $g(\cdot)$ is concave it follows that $g(\theta'') \geq \lambda g(\theta') + (1-\lambda)g(\theta_0)$ which then implies that $g(\theta_0) \in \mathbb{R}$. Now define $\theta_m = m^{-1}\theta' + (1-m^{-1})\theta_0$ for $m \in \mathbb{N}$; it is clear that $g(\theta_0) \geq g(\theta_m) \geq m^{-1}g(\theta') + (1-m^{-1})g(\theta_0)$ for all $m \in \mathbb{N}$ which implies that $\lim_{m\to\infty} g(\theta_m) = g(\theta_0)$ and since $\theta_m \neq \theta_0$ for all $m \in \mathbb{N}$ it follows that $g(\theta_0) = \sup_{\theta\neq\theta_0} g(\theta)$.

Now, let $\{\theta'_m\}_{m=1}^\infty$ denote any sequence of elements of $\mathbb{R}^p$ such that $\lim_{m\to\infty} g(\theta'_m) = g(\theta_0)$, fix an arbitrary $\delta > 0$, and for each $m \in N$ define:

$$\theta''_m = \begin{cases} \theta'_m, & \text{if } \|\theta'_m - \theta_0\| \leq \delta; \\ \theta_0 + \frac{\delta}{\|\theta'_m - \theta_0\|}(\theta'_m - \theta_0), & \text{if } \|\theta'_m - \theta_0\| > \delta. \end{cases} \tag{25}$$

From concavity it follows that $g(\theta''_m) \geq g(\theta'_m)$ for all $m \in N$ and hence that $\lim_{m\to\infty} g(\theta''_m) = g(\theta_0)$. Since $\|\theta''_m - \theta_0\| \leq \delta$ for all $m \in \mathbb{N}$ it follows that $\{\theta''_m\}_{m=1}^\infty$ must possess at least one point of accumulation. Denote such a point of accumulation by θ''_∞. Upper semi-continuity then implies that $g(\theta''_\infty) \geq \lim_{m\to\infty} g(\theta'_m) = g(\theta_0)$ which is only possible if $\theta''_\infty = \theta_0$. But any bounded sequence in $\mathbb{R}^p$ which possesses only one point of accumulation must converge to that point and thus $\lim_{m\to\infty} \theta''_m = \theta_0$. This implies that $\|\theta'_m - \theta_0\| \leq \delta$ for all m sufficiently large and since $\delta > 0$ was arbitrary this implies that $\lim_{m\to\infty} \theta'_m = \theta_0$. But this in turn implies that θ_0 is an identifiably unique maximizer of $g(\theta)$ which combined with the earlier result that $g(\theta_0) = \sup_{\theta\neq\theta_0} g(\theta)$ implies that Assumption 4 holds.

Second, it is well-established that if a concave function of $\mathbb{R}^p$ is finite on some open set than it is continuous on that open set. Let M denote a non-empty open set on which $g(\cdot)$ is finite and let K_1 a compact subset of M with a non-empty interior; then $g(\cdot)$ must achieve its minimum over K_1 at some point belonging to K_1. Let r_1 denote the maximum value achieved by $g(\cdot)$ on K_1. For any $0 \leq \lambda \leq 1$ define:

$$K(\lambda) = \{\theta \in R^p : \ \exists\theta' \in K_1 \ \text{s.t.}\, \theta = \lambda\theta' + (1-\lambda)\theta_0\}, \tag{26}$$

so that $K_1 = K(1)$ and $\{\theta_0\} = K(0)$. Clearly $K(\lambda)$ is compact for all $0 \leq \lambda \leq 1$ and has a non-empty interior for all $0 < \lambda \leq 1$. In addition, by concavity the maximum value of $g(\cdot)$ over $K(\lambda)$ must be greater than or equal to $\lambda r_1 + (1-\lambda)g(\theta_0)$ which can be made arbitrarily close to $g(\theta)$ by selecting λ sufficiently close to zero. Hence for any $\alpha \in A$, as defined in Assumption 5, there exists a value $0 < \lambda_\alpha \leq 1$ such that $K(\lambda_\alpha) \in \text{ulev}_\alpha(g)$ which implies that $\text{ulev}_\alpha(g)$ has a non-empty interior and hence that Assumption 5 is satisfied. ■

Proof of Theorem 1. From Assumption 4 it follows there exists $\delta_0 > 0$ such that $[\theta_0 - \delta, \theta_0 + \delta]$ belongs to the interior of the effective domain of $Q(\cdot)$ for all $0 < \delta \leq \delta_0$. Fix $0 < \delta \leq \delta_0$ and set $\eta = Q(\theta_0) - \max\{Q(\theta_0+\delta), Q(\theta_0-\delta)\}$; from Assumption 4 it is clear that $\eta > 0$. Next for each $n \in \mathbb{N}$ define:

$$D_n \equiv \{\omega \in \Omega : |Q_n(\theta,\omega) - Q(\theta)| \leq (\eta/4), \ \theta = (\theta_0-\delta), \theta_0, (\theta_0+\delta)\}, \tag{27}$$

which clearly depends on δ, and observe that Assumption 2 implies that $\lim_{n\to\infty} P(D_n) = 1$. Equation (27) then implies that:

$$\max\{Q_n(\theta_0-\delta,\omega), Q_n(\theta_0+\delta,\omega)\} \leq Q_n(\theta_0,\omega) - (\eta/2) \leq Q(\theta_0) - (\eta/4), \ \forall\, \omega \in D_n, \ n \in \mathbb{N}, \tag{28}$$

which combined with the quasiconcavity of $Q_n(\cdot,\omega)$ on $\mathbb{R}^p$ for all $\omega \in \Omega$ implies that:

$$\sup_{\theta\in\mathbb{R}^p:|\theta-\theta_0|\geq\delta} Q_n(\theta,\omega) \leq Q_n(\theta_0,\omega) - (\eta/2) \leq Q(\theta_0) - (\eta/4), \ \forall\, \omega \in D_n, \ n \in \mathbb{N}. \tag{29}$$

Now suppose that $\{\hat\theta_n(\cdot) : \Omega \to \mathbb{R}^p\}_{n=1}^\infty$ is a sequence of approximate maximizing estimators of θ_0 based on $\{Q_n(\cdot) : \mathbb{R}^P \to \overline{r}\}_{n=1}^\infty$. Fix $\varepsilon > 0$ such that $\varepsilon \leq \eta/4$ and $\varepsilon^{-1} \geq Q(\theta_0)$ and observe that there exists a sequence of measurable sets $\{A_n\}_{n=1}^\infty$ such that $\lim_{n\to\infty} P(A_n) = 1$ and that

$|\hat{\zeta}_n| \leq \varepsilon$ and $\hat{\alpha}_n \geq \varepsilon^{-1}$ for all $\omega \in A_n$ and $n \in \mathbb{N}$. Define $B_n = \{\omega \in \Omega : \sup_{\theta \in \mathbb{R}^p} Q_n(\theta, \omega) < +\infty\}$. If $\omega \in A_n \cap B_n \cap D_n$ then:

$$Q_n(\hat{\theta}_n(\omega), \omega) \geq Q_n(\theta_0, \omega) - (\eta/4) > \sup_{\theta \in \mathbb{R}^p : |\theta - \theta_0| \geq \delta} Q_n(\theta, \omega), \tag{30}$$

while if $\omega \in A_n \cap B_n^c \cap D_n$, where B_n^c denotes the complement of B_n, then:

$$Q_n(\hat{\theta}_n(\omega), \omega) \geq Q(\theta_0) > \sup_{\theta \in \mathbb{R}^p : |\theta - \theta_0| \geq \delta} Q_n(\theta, \omega), \tag{31}$$

and thus $|\hat{\theta}_n(\omega) - \theta_0| < \delta$ for all $\omega \in A_n \cap D_n$. But clearly $\lim_{n \to \infty} P(A_n \cap D_n) = 1$ which establishes that $\hat{\theta}_n$ is weakly consistent in the generalized sense as desired since $0 < \delta \leq \delta_0$ can be chosen arbitrarily close to zero. ■

Proof of Theorem 2. First, for any $\theta \in \mathbb{R}^p$ and $\gamma > 0$ let $S(\theta, \gamma)$ denote the open ball of radius γ centered on θ, given by:

$$S(\theta, \gamma) = \{\theta^* \in \mathbb{R}^p : \|\theta^* - \theta\| < \gamma\}, \tag{32}$$

let $\overline{S}(\theta, \gamma)$ denote the closure of $S(\theta, \gamma)$, and let $\partial S(\theta, \gamma)$ denote the boundary of $S(\theta, \gamma)$. Now fix any $\delta > 0$. Then Assumption 4 implies that there exists $\eta > 0$ such that:

$$\sup_{\theta \in S(\theta_0, \delta)^c} Q(\theta) = \sup_{\theta \in \mathbb{R}^p :\ \theta \neq \theta_0} Q(\theta) - \eta. \tag{33}$$

Now select $\alpha \in \mathbb{R}$ such that $\sup_{\theta \in S(\theta_0, \delta)^c} Q(\theta) < \alpha < \sup_{\theta \in \mathbb{R}^p :\ \theta \neq \theta_0} Q(\theta)$. Clearly $\text{ulev}_\alpha(Q) \subseteq \{\theta \in S(\theta_0, \delta)^c\}$ for all such α. Assumption 5 implies that there exist $\theta' \in \mathbb{R}^p$ and $0 < \delta' < \delta$ such that $\overline{S}(\theta', \delta') \subseteq \text{ulev}_\alpha(Q)$. Thus together Assumptions 4 and 5 imply that for any $\delta > 0$ there exist $\theta' \in \mathbb{R}^p$ and $\delta' > 0$ such that:

$$\overline{S}(\theta', \delta') \quad \subset \quad \overline{S}(\theta_0, \delta), \tag{34}$$

$$\sup_{\theta \in S(\theta_0, \delta)^c} Q(\theta) \quad < \quad \inf_{\theta \in \overline{S}(\theta', \delta')} Q(\theta). \tag{35}$$

Furthermore, it is clear that $\overline{S}(\theta_0, \delta) \subset \overline{S}(\theta', 2\delta)$ and hence that there exists $0 < \eta' < +\infty$ such that:

$$\sup_{\theta \in S(\theta', 2\delta)^c} Q(\theta) \leq \inf_{\theta \in \overline{S}(\theta', \delta')} Q(\theta) - \eta'. \tag{36}$$

Second, for any $\theta \in \mathbb{R}^p$ and $\gamma > 0$ let $C(\theta, \gamma)$ denote the closed cube centered on θ whose facets are perpendicular to the axes of $\mathbb{R}^p$ and whose edges have length 2γ, and let $V(\theta, \gamma)$ denote the set consisting of the vertices of $C(\theta, \gamma)$: clearly $V(\theta, \eta)$ then has 2^p elements. It is then easy to show that there exist $\rho_1, \rho_2 > 0$ such that:

$$\overline{S}(\theta, \rho_1 \rho_2 \gamma) \subseteq C(\theta, \rho_1 \gamma) \subseteq \overline{S}(\theta, \gamma), \quad \forall\ \gamma > 0\ \&\ \theta \in \mathbb{R}^p, \tag{37}$$

and thus that:

$$\left[\sup_{\theta\in S(\theta',2\delta)^c} Q(\theta)\right] \leq \left[\inf_{\theta\in C(\theta',\rho_1\delta')} Q(\theta)\right] - \eta' \leq \left[\inf_{\theta\in V(\theta',\rho_1\delta')} Q(\theta)\right] - \eta'. \tag{38}$$

Third, since $\partial S(\theta', 2\delta)$ is a compact subset of $S(\theta', 2\delta)^c$ it follows that there exists a finite collection of elements of $\partial S(\theta', 2\delta)^c$, denoted $H(\theta', 2\delta)$, such that for any $\theta \in \partial S(\theta', 2\delta)$ there exists $\theta^* \in H(\theta', 2\delta)$ satisfying $\|\theta^* - \theta\| \leq \rho_1\rho_2\delta'$. It thus follows that $Q(\cdot)$ satisfies:

$$\left[\sup_{\theta\in H(\theta',2\delta)} Q(\theta)\right] \leq \left[\inf_{\theta\in V(\theta',\rho_1\delta')} Q(\theta)\right] - \eta'. \tag{39}$$

Furthermore, since $Q_n(\cdot)$ converges pointwise to $Q(\cdot)$ and since $H(\theta', 2\delta)$ and $V(\theta', \rho_1\delta')$ only contain a finite number of points it follows that there exists a sequence of measurable sets D_n (which depend on δ) such that $\lim_{n\to\infty} P(D_n) = 1$ and that:

$$\left[\sup_{\theta\in H(\theta',2\delta)} Q_n(\theta,\omega)\right] \leq \left[\inf_{\theta\in V(\theta',\rho_1\delta')} Q_n(\theta,\omega)\right] - (\eta'/2), \quad |Q_n(\theta',\omega) - Q(\theta')| \leq (\eta'/4), \tag{40}$$

for all $n \in N$ and $\omega \in D_n$. Note that $Q(\theta') \in \mathbb{R}$ in view of Assumption 3 together with Equation (36).

Fourth, since $C(\theta, \gamma)$ is the convex hull of $V(\theta, \gamma)$ for any $\gamma > 0$ and $\theta \in \mathbb{R}^p$ it follows that all the elements of $\overline{S}(\theta', \rho_1\rho_2\delta')$ can be expressed as convex combinations of the elements of $V(\theta', \rho_1\delta')$ and hence that the value of $Q_n(\cdot, \omega)$ at any element of $\overline{S}(\theta_0, \rho_1\rho_2\delta')$, including therefore the value of $Q_n(\cdot, \omega)$ at $\theta = \theta'$, must be at least as large as the minimum value which $Q_n(\cdot, \omega)$ takes on the set $V(\theta', \rho_1\delta')$. This then implies that:

$$\left[\sup_{\theta\in H(\theta',2\delta)} Q_n(\theta,\omega)\right] \leq \left[\inf_{\theta\in \overline{S}(\theta',\rho_1\rho_2\delta')} Q_n(\theta,\omega)\right] - (\eta'/2) \leq Q_n(\theta',\omega) - (\eta'/2), \tag{41}$$

for all $n \in N$ and $\omega \in D_n$.

Fifth, take any $\theta \in S(\theta', 4\delta)^c$; then there exists $\theta'' \in \partial S(\theta', 2\delta)$ and $0 \leq \lambda \leq 1$ such that $\theta'' = \lambda\theta + (1-\lambda)\theta'$ and hence $\theta' = \theta + (1-\lambda)^{-1}(\theta'' - \theta)$. But then there exists $\theta''' \in H(\theta', 2\delta)$ such that $\|\theta''' - \theta''\| \leq \rho_1\rho_2\delta'$. Now let $\theta^* = \theta + (1-\lambda)^{-1}(\theta''' - \theta)$ and observe that $(\theta^* - \theta') = (1-\lambda)^{-1}(\theta''' - \theta'')$ and hence that $\|\theta^* - \theta'\| \leq \rho_1\rho_2\delta'$ since $0 \leq \lambda = \|\theta'' - \theta'\|/\|\theta - \theta'\| \leq (1/2)$. But this establishes that there exist $\theta''' \in H(\theta', 2\delta)$, $\theta^* \in \overline{S}(\theta', \rho_1\rho_2\delta')$ and $0 \leq \lambda \leq 1$ such that $\theta''' = \lambda\theta + (1-\lambda)\theta^*$ and hence that $Q_n(\theta''', \omega) \geq \min\{Q_n(\theta, \omega), Q_n(\theta^*, \omega)$ in view of the quasiconcavity of $Q_n(\cdot)$. This combined with Equation (41) then implies that:

$$Q_n(\theta,\omega) \leq Q_n(\theta',\omega) - (\eta'/2), \quad \forall\, n \in \mathbb{N}, \omega \in D_n, \theta \in S(\theta', 4\delta)^c, \tag{42}$$

and thus that:

$$\sup_{\theta\in S(\theta',4\delta)^c} Q_n(\theta,\omega) \leq Q_n(\theta',\omega) - (\eta'/2) \leq Q(\theta') - (\eta'/4), \quad \forall\, \omega \in D_n,\ n \in \mathbb{N}. \tag{43}$$

The remainder of the proof then follows the same lines as the proof of Theorem 1 from Equation (29) onwards but with 4δ in place of δ, θ' in place of θ_0 and η' in place of η. ■

Notes

[1] The earliest result of this type, namely on the consistency of optimization estimators based on stochastic convex (or concave) objective functions, seems to be Corollary II.2 from Andersen and Gill (1982) which establishes the consistency of such a maximization estimator but not its existence. Similar results can be found in Haberman (1989), Niemiro (1992), and Pollard (1991).

[2]This is a slight generalization of the concept of an approximate maximizing sequence, as used by Geyer (1996), in which $\hat{\zeta}_n$ and $\hat{\alpha}_n$ are non-random.

[3] The stochastic version of this uniform convergence result from Rockafellar (1970) appears to have been established first by Andersen and Gill (1982, Theorem II.1).

[4]Note that even in the case of a sequence of upper semi-continuous concave functions which converges pointwise to a function whose effective domain has a non-empty interior there is no guarantee that the pointwise limiting function is upper semi-continuous on the boundary of its effective domain though it will be upper semi-continuous on the interior of its effective domain.

References

Amemiya, T. (1985) *Advanced Econometrics.* Oxford: Basil Blackwell.

Andersen, P.K. & R.D. Gill (1982) Cox's regression model for counting processes: A large sample study. *The Annals of Statists* **10**, 1100–1120.

Arrow, K.J. & A.C. Enthoven (1961) Quasi-concave programming. *Econometrica* **29**, 779–800.

Attouch, H. (1984) *Variational Convergence for Functions and Operators.* London, UK: Pitman.

Geyer, C.J. (1994) On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society, Series B* **56**, 261–274.

Geyer, C.J. (1996) On the asymptotics of convex stochastic optimization. Mimeo, University of Minnesota.

Haberman, S.J. (1989) Concavity and estimation. *The Annals of Statists* **17**, 1631–1661.

Kall, P. (1986) Approximation to optimization problems: An elementary review. *Mathematics of Operations Research* **11**, 9–18.

Newey, W. & D.McFadden (1994) Large sample estimation and hypothesis testing. In *Handbook of Econometrics, Volume IV*, chap. 36. New York: North-Holland.

Niemiro, W. (1992) Asymptotics for M-estimators defined by convex minimization. *The Annals of Statists* **20**, 1514–1533.

Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**, 186–199.

Rockafellar, R.T. (1970) *Convex Analysis*. Princeton: Princeton University Press.

van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Zervos, M. (1999) On the epiconvergence of stochastic optimization problems. *Mathematics of Operations Research* **24**, 495–508.