TESTING FOR STOCHASTIC MONOTONICITY

Sokbae Lee, Oliver Linton and Yoon-Jae Whang

October 10, 2006

Abstract

We propose a test of the hypothesis of stochastic monotonicity. This hypothesis is of interest in many applications. Our test is based on the supremum of a rescaled U-statistic. We show that its asymptotic distribution is Gumbel. The proof is difficult because the approximating Gaussian stochastic process contains both a stationary and a nonstationary part and so we have to extend existing results that only apply to either one or the other case.

Key words: Distribution function; Extreme Value Theory; Gaussian Process; Monotonicity.

Journal of Economic Literature Classification: C14, C15.

AMS 2000 subject classifications: Primary 62G08

*Department of Economics, University College London, Gower Street, London, WC1E 6BT, UK. Email: l.simon@ucl.ac.uk. Thanks to the Leverhulme Trust through the funding of the Centre for Microdata Methods and Practice and of the research programme Evidence, Inference and Inquiry.

†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. Email: o.linton@lse.ac.uk; web page: http://econ.lse.ac.uk/staff/olinton/. Thanks to the ESRC for financial support.

‡Department of Economics, Seoul National University, Seoul 151-742, Korea. Email: whang@snu.ac.kr. Thanks to the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2005-041-B00074).
1 Introduction

Let $Y$ and $X$ denote two random variables whose joint distribution is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^2$. Let $F_{Y|X}(.|x)$ denote the distribution of $Y$ conditional on $X = x$. This paper is concerned with testing the stochastic monotonicity of $F_{Y|X}$. Specifically, we consider the hypothesis

$\quad \text{(1)} \quad H_0: \text{For each } y \in \mathcal{Y}, \quad F_{Y|X}(y|x) \leq F_{Y|X}(y|x') \text{ whenever } x \geq x' \text{ for } x, x' \in \mathcal{X},$

where $\mathcal{Y}$ and $\mathcal{X}$, respectively, are the supports of $Y$ and $X$. We propose a test statistic and obtain asymptotically valid critical values.

The hypothesis implies that the regression function $E(Y|X = x)$, when it exists, is monotonic increasing. It also implies that all conditional quantile functions are increasing. It is a strong hypothesis but can be reduced in strength by limiting the set of $X$ and $Y$ for which this property holds. Note that the transformation regression model structure considered in Ghosal, Sen and van der Vaart (2000) i.e., $Y = m(X) + \varepsilon$, where $\varepsilon$ is independent of $X$ and both $\phi, m$ are monotonic functions, actually implies stochastic monotonicity.

This hypothesis can be of interest in a number of applied settings. If $X$ is some policy, dosage, or other input variable, one might be interested in testing whether its effect on the distribution of $Y$ is increasing in this sense. Blundell, Gosling, Ichimura, and Meghir (2006) have recently used this hypothesis to obtain tight bounds on an unobservable cross-sectional wage distribution thus allowing them to characterize the evolution of its inequality over time. Specifically, they assumed that the distribution of wages $W$ for employed given observed characteristics $X$ and an instrument $Z$ is increasing in $Z$. Their instrument was the out of work income. They derived a bound on the implied distribution of wages given characteristics under this assumption of stochastic monotonicity. They also suggested a test of this hypothesis based on the implied bounds, using bootstrap to calculated critical values. They found that the hypothesis was not rejected on their data at standard significance levels, indeed the p-values were very high. They did not provide any theory to justify their critical values, and moreover did not test the monotonicity hypothesis itself but an implication of it.

A leading case is in time series when $Y = Y_{t+1}$ and $X = Y_t$ and $Y_t$ is a Markov process so that $F_{Y|X} = F_{t+1|t}$ is the transition measure of the process $Y_t$. In that case the property, along with mild technical conditions, implies that the process has a stationary distribution. The influential monograph of Lucas and Stokey (1989) uses the stochastic monotonicity
property frequently in solving dynamic optimization problems of the Markov type and characterizing the properties of the solution. It is particularly important in problems where nonconvexities give rise to discontinuous stochastic behaviour and it provides a route to proving the existence of stationary equilibria not requiring smoothness. Hopenhayn and Prescott (1992) argue that it arises ‘in economic models from the monotonicity of decision rules or equilibrium mappings that results from the optimizing behaviour of agents’. Pakes (1986) assumed that the distribution of the return to holding a patent conditional on current returns was nonincreasing in current returns. Consequently he showed that the optimal renewal policy took a very simple form based on the realization of current returns compared with the cost of renewing. Ericson and Pakes (1995), Olley and Pakes (1996), and Buettner (2003) have all used a similar property in various dynamic models of market structures.

We propose a simple test of hypothesis (1) for observed i.i.d. data. Our statistic is based on the supremum of a rescaled second order U-process indexed by two parameters $x$ and $y$. Nolan and Pollard (1987). It generalizes the corresponding statistic introduced by Ghosal, Sen and van der Vaart (2000) for testing the related hypothesis of monotonicity of a regression function. We prove that the asymptotic distribution of our test statistic is a Gumbel with certain nonstandard norming constants, thereby facilitating inference using critical values obtained from the limiting distribution. We also show that the test is consistent against all alternatives. The proof technique is quite complicated and novel because the approximating Gaussian stochastic process contains both a stationary part (corresponding to $x$) and a nonstationary part (corresponding to $y$) and so we have to extend existing results that only apply to either one or the other case. For example, Stute (1986) establishes the weak convergence to a Brownian Bridge of a conditional empirical process (effectively holding $x$ constant in our problem). In the other direction, using the local strong invariance principle of Rio (1994), Ghosal, Sen and van der Vaart (2000) establish local (in $x$ in our notation) weak convergence of an empirical process to a stationary limit, generalizing the seminal work of Bickel and Rosenblatt (1973). The most closely related work to ours is Beirlant and Einmahl (1996) who consider the asymptotics of some functional of a conditional empirical process except that they only consider a finite maximum over the covariate.

The remainder of the paper is organized as follows. Section 2 defines our test statistic and Section 3 outlines the general proof strategy and states the asymptotic results. Section 4 contains results of some Monte Carlo experiments. All the proofs are in the Appendix.
2 The Test Statistic

This section describes our test statistic. Let \( \{(Y_i, X_i) : i = 1, \ldots, n\} \) denote a random sample from \((Y, X)\). We suppose throughout that the data are i.i.d. but the main result also holds for the Markov case where \( Y_i = Y_{i+1} \) and \( X_i = Y_i \). Let \( 1(\cdot) \) denote the usual indicator function and let \( K(\cdot) \) denote a one-dimensional kernel function with a bandwidth \( h_n \). Consider the following \( U \)-process:

\[
U_n(y, x) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} [1(Y_i \leq y) - 1(Y_j \leq y)] \text{sgn}(X_i - X_j) 
\times K_{h_n}(X_i - x)K_{h_n}(X_j - x),
\]

where \( K_{h_n}(\cdot) = h_n^{-1}K(\cdot/h_n) \) and \( \text{sgn}(x) = 1(x > 0) - 1(x < 0) \). Note that the \( U \)-process \( U_n(y, x) \) can be viewed as a locally weighted version of Kendall’s tau statistic, applied to \( 1(Y_i \leq y) \) and that \( U_n(y, x) \) is related to the \( U \)-process considered in Ghosal, Sen, and van der Vaart (2000, equation (2.1)).

First, notice that under regularity conditions including smoothness of \( F_{Y|X}(y|x) \), as \( n \to \infty \),

\[
EU_n(y, x) \to h_n F_x(y|x) \left( \int \int |u_1 - u_2| K(u_1)K(u_2) du_1 du_2 \right) |f_X(x)|^2,
\]

where \( F_x(y|x) \) is a partial derivative of \( F_{Y|X}(y|x) \) with respect to \( x \). Therefore, under the null hypothesis such that \( F_x(y|x) \leq 0 \) for all \( (y, x) \in Y \times X \), \( U_n(y, x) \) is less than or equal to zero on average for large \( n \). Under the alternative hypothesis such that \( F_x(y|x) > 0 \) for some \( (y, x) \in Y \times X \), a suitably normalized version of \( U_n(y, x) \) can be very large. In view of this, we define our test statistic as a supremum statistic

\begin{equation}
S_n = \sup_{(y, x) \in Y \times X} \frac{U_n(y, x)}{c_n(x)}
\end{equation}

with some suitably defined \( c_n(x) \), which may depend on \( (X_1, \ldots, X_n) \) but not on \( (Y_1, \ldots, Y_n) \).

The exact form of \( c_n(x) \) will be defined below.

Define \( \mathcal{H}_1 = \{(y, x) \in Y \times X : F_x(y|x) = 0\} \) and \( \mathcal{H}_2 = \{(y, x) \in Y \times X : F_x(y|x) < 0\} \). Note that for any \( s \),

\begin{equation}
\Pr(S_n > s) \leq \Pr \left( \sup_{(y, x) \in \mathcal{H}_1} \frac{U_n(y, x)}{c_n(x)} > s \right) + \Pr \left( \sup_{(y, x) \in \mathcal{H}_2} \frac{U_n(y, x)}{c_n(x)} > s \right).
\end{equation}
The first probability on the right-hand side of (3) is maximized when $H_1 \equiv Y \times X$. Also, under regularity conditions for the consistency of the test, which will be stated in Section 3, the second probability on the right-hand side of (3) can be arbitrarily small for any fixed $s$. Thus, for all sufficiently large $n$, the type I error probability is maximized in $H_0$ when $F_x(y|x) \equiv 0$, equivalently $F_{Y|X}(y|x) = F_Y(y)$ for any $(y, x) \in Y \times X$. Therefore, we consider the case that $F_x(y|x) \equiv 0$ to derive the limiting distribution under the null hypothesis.

When $F_x(y|x) \equiv 0$, the projection of $U_n(y, x)$ is given by

\[ \hat{U}_n(y, x) = 2n^{-1} \sum_{i=1}^{n} [1(Y_i \leq y) - F(y)] \int \text{sgn}(X_i - \bar{x})K_{h_n}(\bar{x} - x)dF_X(\bar{x})K_{h_n}(X_i - x). \]

Since $E\hat{U}_n(y, x) = 0$, the variance of $\hat{U}_n(y, x)$ is given by $F_Y(y)[1 - F_Y(y)]\sigma_n^2(x)/n$, where

\[ \sigma_n^2(x) = 4 \int \left[ \int \text{sgn}(\bar{x} - \bar{x})K_{h_n}(\bar{x} - x)dF_X(\bar{x})K_{h_n}(\bar{x} - x) \right]^2 dF_X(\bar{x}). \]

As in Ghosal, Sen, and van der Vaart (2000), $\sigma_n^2(x)$ can be estimated by a $U$-statistic

\[ \hat{\sigma}_n^2(x) = \frac{4}{n(n-1)(n-2)} \sum_{i \neq j \neq k \leq n} \text{sgn}(X_i - X_j)\text{sgn}(X_i - X_k) \times K_{h_n}(X_j - x)K_{h_n}(X_k - x)[K_{h_n}(X_i - x)]^2. \]

This suggests that we use the scaling factor $c_n(x) = \hat{\sigma}_n(x)/\sqrt{n}$.

An alternative class of test statistics is based on explicit estimation of conditional c.d.f.’s thus, consider $T_n = \sup_{y \in Y, x' \in X} \int [\hat{F}_{Y|X}(y|x) - \hat{F}_{Y|X}(y|x')]$, where $\hat{F}_{Y|X}(y|x)$ is some e.g., kernel estimate of the conditional c.d.f., see Hall, Wolff, and Yao (1999). Under the null hypothesis $T_n \leq 0$ with probability tending to one, while under the alternative hypothesis $T_n > 0$. The advantage that $T_n$ has is that it does not require smoothness of $F_{Y|X}(y|x)$. The disadvantage is that its limiting distribution is not pivotal and it is not known how to make it so. One might also be interested in testing second or higher order dominance, Levy (2006), of the conditional distribution functions, which can be achieved by straightforward modification of either $S_n$ or $T_n$.

3 Asymptotic Theory

This section provides the asymptotic behaviour of the test statistic when the null hypothesis is true and when it is false. In particular, we determine the asymptotic critical region of the test and show that the test is consistent against general fixed alternatives at any level.
Although the test is easy to implement, the asymptotic theory for the test involves several lengthy steps:

1. The asymptotic approximation of \( U_n(y, x)/c_n(x) \) by a Gaussian process;

2. The asymptotic approximation of the excursion probability of the maximum of the Gaussian process on a fixed set;

3. The asymptotic approximation of the excursion probability of the maximum of the Gaussian process on an increasing set.

We carry out step 1 by mimicking arguments of Ghosal, Sen, and van der Vaart (2000); step 2 by drawing on a monograph by Piterbarg (1996); step 3 by using arguments similar to those used in Leadbetter, Lindgren, and Rootzén (1983) and Piterbarg (1996). None of these steps is trivial, and, in particular, step 2 requires that we develop a new result for the excursion probability of the maximum of the Gaussian process that has a non-standard covariance function. To be specific, the approximating Gaussian process contains both a stationary and a nonstationary part and therefore we need to extend existing results that only apply to either one or the other case. For example, see Section 7 of Piterbarg (1996) for the stationary case and Sections 8 and 9 of Piterbarg (1996) for the nonstationary case, but to our best knowledge, there is no known result regarding our case in the literature.

### 3.1 Gaussian Process Approximation

As noted in the previous section, it suffices to consider the case that \( F_x(y|x) \equiv 0 \) to derive the limiting distribution under the null hypothesis. In this section, we assume that \( F_x(y|x) \equiv 0 \), equivalently \( F_Y|X(y|x) = F_Y(y) \) for any \((y, x)\). That is, \( Y \) and \( X \) are independent. Further, assume that without the loss of generality, the support of \( X \) is \( \mathcal{X} = [0, 1] \).

Let \( f_X(\cdot) \), \( F_X(\cdot) \), and \( F_Y(\cdot) \), respectively, denote the p.d.f. and c.d.f. of \( X \) and c.d.f. of \( Y \). Define \( \mathcal{X}_n = [0, 1/h_n] \), \( q(u) = \int \text{sgn}(u - w)K(w)dw \), and

\[
\rho(s) = \frac{\int q(z)q(z - s)K(z)K(z - s)dz}{\int q^2(z)K^2(z)dz}.
\]

**Theorem 1.** Assume that (a) \( F_x(y|x) \equiv 0 \); (b) without the loss of generality, the set \( \mathcal{X} = [0, 1] \); (c) the distribution of \( X \) is absolutely continuous with respect to Lebesgue measure and the probability density function of \( X \) is continuous and strictly positive in \( \mathcal{X} \); (d) the
distribution of $Y$ is absolutely continuous with respect to Lebesgue measure; (e) $K$ is a kernel function with support $[-1, 1]$, and is twice continuously differentiable.

Let $h_n$ satisfy

$$h_n(\log n)^{1/2} \to 0 \quad \text{and} \quad nh_n^2/(\log n)^2 \to \infty.$$  

Then there exists a sequence of Gaussian processes \{\xi_n(u, s) : (u, s) \in [0, 1] \times \mathcal{X}_n\} with continuous sample paths such that

$$E[\xi_n(u, s)] = 0, \quad E[\xi_n(u_1, s_1)\xi_n(u_2, s_2)] = [\min(u_1, u_2) - u_1u_2]p(s_1 - s_2),$$

for $u, u_1, u_2 \in [0, 1]$ and $s, s_1, s_2 \in \mathcal{X}_n$, and that

$$\sup_{(y, x) \in \mathcal{Y} \times \mathcal{X}} \left| n^{1/2} \frac{U_n(y, x)}{\sigma_n(x)} - \xi_n[F_Y(y), h_n^{-1}x] \right| = O_p\left(n^{-1/4}h_n^{-1/2}(\log n)^{1/2} + h_n(\log n)^{1/2}\right).$$

The covariance function of $\xi_n$ is the product of a Brownian Bridge covariance function and a stationary covariance function.

### 3.2 Distribution of the Test Statistic

Since the distribution of $\xi_n(u, s)$ does not depend on $n$, for the purpose of deriving the distribution of the supremum statistic $S_n$, it suffices to consider the asymptotic behaviour of the excursion probability of the maximum of a zero-mean Gaussian process that has the same covariance function as $\xi_n(u, s)$. To do so, let $\xi(u, s)$ denote a Gaussian process \{\xi(u, s) : (u, s) \in [0, 1] \times \mathcal{X}\} with continuous sample paths such that

$$E[\xi(u, s)] = 0, \quad E[\xi(u_1, s_1)\xi(u_2, s_2)] = [\min(u_1, u_2) - u_1u_2]p(s_1 - s_2),$$

for $u, u_1, u_2 \in [0, 1]$ and $s, s_1, s_2 \in \mathcal{X}$. Then $\xi_n$ is the restriction of $\xi$ to $[0, 1] \times \mathcal{X}_n$.

**Theorem 2.** For any $x$,

$$\Pr \left(4\beta_n \left\{ \max_{(u, s) \in [0, 1] \times \mathcal{X}_n} \xi(u, s) - \beta_n \right\} < x \right) = \exp \left\{ - \exp \left( -x - \frac{x^2}{8\beta_n^2} \right) \left[ 1 + \frac{x}{4\beta_n^2} \right] \right\} + o(1),$$

where $\beta_n$ is the largest solution to the following equation:

$$h_n^{-1} \left( \frac{8\lambda}{\pi} \right)^{1/2} \beta_n \exp(-2\beta_n^2) = 1$$

and

$$\lambda = -\frac{6 \int q(x)K^2(x)K''(x)K(x)K''(x)dx}{\int q^2(x)K^2(x)dx}.$$
To use the conclusion of Theorem 2, it is necessary to compute \( \beta_n \). It is straightforward to show that

\[
\beta_n = \left( \frac{1}{2} \log \left[ h_n^{-1} c^* \right] \right)^{1/2} + \frac{\log \left[ \frac{1}{2} \log \left[ h_n^{-1} c^* \right] \right]}{8 \left( \frac{1}{2} \log \left[ h_n^{-1} c^* \right] \right)^{1/2}} + o \left( \frac{1}{\left( \log \left[ h_n^{-1} c^* \right] \right)^{1/2}} \right),
\]

where \( c^* = \left( \frac{8}{\pi} \right)^{1/2} \). Then one can use an approximation to \( \beta_n \) by the first two terms on the right side of (5).

Assume that \( n^{-1/4} n^{-1/2} (\log n)^{1/2} + h_n (\log n)^{1/2} \to 0 \). Then the following theorem is an immediate consequence of Theorems 1 and 2.

**Theorem 3.1.** Assume that \( h_n \log n \to 0 \) and \( nh_n^2 / (\log n)^4 \to \infty \). Then for any \( x \),

\[
\Pr \left( 4 \beta_n (S_n - \beta_n) < x \right) = \exp \left\{ - \exp \left( -x - \frac{x^2}{8 \beta_n^2} \right) \left[ 1 + \frac{x}{4 \beta_n^2} \right] \right\} + o(1).
\]

In particular,

\[
\lim_{n \to \infty} \Pr \left( 4 \beta_n (S_n - \beta_n) < x \right) = \exp \left( - e^{-x} \right).
\]

As in Theorem 4.2 of Ghosal, Sen, and van der Vaart (2000), the theorem suggests that one can construct a test with an asymptotic level \( \alpha \):

\[
\text{Reject } H_0 \text{ if } S_n \geq \beta_n + \frac{1}{4 \beta_n} \log \left( \frac{1}{\log(1 - \alpha)^{-1}} \right)
\]

for any \( 0 < \alpha < 1 \). Alternatively, one can construct an \( \alpha \)-level test with (6):

\[
\text{Reject } H_0 \text{ if } S_n \geq \beta_n + \frac{1}{4 \beta_n} z_n, \alpha - 1.
\]

where for each \( n \), \( z_n, 1-\alpha \) is the \( (1 - \alpha) \) quantile of the ‘distribution function’ \( F_n(x) \) of the form

\[
F_n(x) = \exp \left\{ - \exp \left( -x - \frac{x^2}{8 \beta_n^2} \right) \left[ 1 + \frac{x}{4 \beta_n^2} \right] \right\}.
\]

In the next section, we carry out Monte Carlo experiments using both critical regions (7) and (8). It turns out that in our experiments, a test based on (8) performs better in finite samples. We expect that (6) provides a higher order approximation with polynomial error as in Piterbarg (1996, Theorem G1), which would justify its use.

We now turn to the consistency of the test. It is straightforward to show that the test specified by (7) or (8) is consistent again general alternatives.

**Theorem 3.2.** Assume that \( nh_n^3 / \log h_n^{-1} \to \infty \). If \( F_x(y|x) > 0 \) for some \( (y, x) \in \mathcal{Y} \times \mathcal{X} \), then the test specified by (7) or (8) is consistent at any level.
We end this section by mentioning that the test and its asymptotic properties obtained in this section can be extended easily to the case when the null hypothesis in (1) holds only for $Y$ and $X_1$, where $X_1$ is a compact interval and a strict subset of $X$. In this case, $F_x(y|x) \equiv 0$ does not imply that $Y$ and $X$ are independent; however, this would not matter since our test statistic depends only on observations inside an open interval containing $X_1$. Thus, the asymptotic properties of the supremum test statistic would be the same with $X_1$.

4 A Monte Carlo Experiment

This section presents the results of some Monte Carlo experiments that illustrate the finite-sample performance of the test. For each Monte Carlo experiment, $X$ was independently drawn from a uniform distribution on $[0,1]$. To evaluate the performance of the test under the correct null hypothesis, $Y$ was generated independently from $U \sim \mathcal{N}(0,0.1^2)$. In addition, to see the power of the test, $Y$ was also generated from $Y = m(X) + U$, where $m(x) = x(1-x)$. The simulation design considered here is similar to that of Ghosal et al. (2000). To save computing time the test statistic was computed by the maximum of $\sqrt{n}U_n(y,x)/\tilde{\sigma}_n(x)$ over $Y \times X$, where $Y = \{Y_1,Y_2,\ldots,Y_n\}$, $X = \{0.05,0.10,\ldots,0.90,0.95\}$, and $\tilde{\sigma}_n(x) = 4h_n^{-1}\left[\int q^2(u)K^2(u)du\right] \times \hat{f}_X(x)$. Here, $\hat{f}_X(x)$ denotes the kernel density estimator of $f_X(x)$. Notice that $\tilde{\sigma}_n(x)$ is asymptotically equivalent to $\sigma_n(x)$ but it is easier to compute. The kernel function was $K(u) = 0.75(1-u^2)$ for $-1 \leq u \leq 1$. The simulations used sample sizes of $n = 50, 100, 200$ and $500$, and all the simulations were carried out in GAUSS using GAUSS pseudo-random number generators. For each simulation, the number of replications was 1500.

Table 1 reports results of Monte Carlo experiments using critical values obtained from the asymptotic expansion $F_n$ of the limiting distribution (see (8)) and also using those from the type I extreme value distribution (see (7)). The nominal level was 5%. First, consider the first panel of the table that shows results with the critical values from $F_n$. When the null hypothesis is true, each rejection proportion is below the nominal level for all the bandwidths and is maximized at $n = 500$ and $h_n = 0.5$. It can be seen that the best $h_n$ is decreasing with the sample size and the performance of the test is less sensitive with $h_n$ as $n$ gets large. When the null hypothesis is false, for all values of $h_n$, the powers of the test are high for $n = 50$, almost one for $n = 100$, and one for $n = 200$. The performance of the test with critical values from the type I extreme value distribution is uniformly worse, as
seen from the second panel of the table.

The asymptotic critical values from the asymptotic expansion $F_{n}$ of the limiting distribution are easy to compute and appear to work satisfactorily in the simple numerical example we examined. Instead one could employ a standard bootstrap resample applied to a recentered statistic (or a bootstrap resample imposing independence between $Y$ and $X$) to improve the size of the test, motivated by the reasoning of Hall (1993). This method should work better in finite samples but is much more time consuming than the asymptotic critical values. See Einmahl and Van Keilegom (2006) for recent work in a related problem.

A Appendix: Technical Proofs

Following the notation of Ghosal, Sen, and van der Vaart (2000), for a metric space $M$ with a norm on it, let $N(\varepsilon, S, M)$, $\varepsilon > 0$, denote the $\varepsilon$-covering number of $S \in M$. Throughout the Appendix, $\lesssim$ will stand for an inequality up to a constant multiple.

A.1 Proofs for Section 3.1

Proof of Theorem 1. The proof of Theorem 1 follows closely Theorem 3.1 of Ghosal, Sen, and van der Vaart (2000). In particular, the theorem can be proved by combining arguments identical to those used in the proof of Theorem 3.1 of Ghosal, Sen, and van der Vaart (2000) with Lemmas A.1-A.4 proved below.

Lemma A.1.

$$\sup_{(y,x) \in Y \times X} \left| U_{n}(y, x) - \hat{U}_{n}(y, x) \right| = O_{P} \left( n^{-1}h_{n}^{-3/2} \right).$$

Proof. The proof is similar to that of Lemma 3.1 of Ghosal, Sen, and van der Vaart (2000). Hence, we will only indicate the differences. Consider a class of functions $\mathcal{M} = \{m_{(y, x)} : (y, x) \in Y \times X\}$, where

$$m_{(y, x)}((y_{1}, x_{1}), (y_{2}, x_{2})) = [1(y_{1} \leq y) - 1(y_{2} \leq y)]\text{sgn}(x_{1} - x_{2})$$

$$\times K_{h_{n}}(x_{1} - x)K_{h_{n}}(x_{2} - x).$$
This class is contained in the product of the classes

\[ M_1 = \{ 1(y_1 \leq y) - 1(y_2 \leq y) : y \in \mathcal{Y} \} \]
\[ M_2 = \left\{ K \left( \frac{x_1 - x}{h_n} \right) : x \in \mathcal{X} \right\} \]
\[ M_3 = \left\{ K \left( \frac{x_2 - x}{h_n} \right) : x \in \mathcal{X} \right\} \]
\[ M_4 = \{ h_n^{-2} \text{sgn}(x_1 - x_2)1\{|x_1 - x_2| \leq 2h_n\} \}. \]

Note that the class \( M_1 \) is the difference of two classes of functions \( y_1 \mapsto 1(y_1 \leq y) \) and \( y_2 \mapsto 1(y_2 \leq y) \), respectively. Hence, by Example 2.5.4 of van der Vaart and Wellner (1996) and the fact that the \( 2\varepsilon \)-covering number of the sum of the two classes is bounded by the product of the \( \varepsilon \)-covering numbers of the two classes,

\[ N(2\varepsilon, M_1, L_2(Q)) \leq \frac{4}{\varepsilon^4} \]
for any probability measure \( Q \) and \( \varepsilon \leq 1 \). Following the proof of Lemma 3.1 of Ghosal, Sen, and van der Vaart (2000), we have

\[
\text{nE} \left[ \sup_{(y,x) \in \mathcal{Y} \times \mathcal{X}} U_n(y, x) - \hat{U}_n(y, x) \right] \\
\leq \left( \text{E} \left[ h_n^{-4}1\{|X_1 - X_2| \leq 2h_n\} \right] \right)^{1/2} \\
\leq h_n^{-3/2},
\]

which gives the conclusion of the lemma.

Define

\[
\phi_{n,y,x}(Y, X) = 2[1(Y \leq y) - F_Y(y)] \int \text{sgn}(X - \hat{x})K_{h_n}(\hat{x} - x)dF_X(\hat{x})K_{h_n}(X - x).
\]

**Lemma A.2.** There exists a sequence of Gaussian processes \( G_n(\cdot) \), indexed by \( \mathcal{Y} \times \mathcal{X} \), with continuous sample paths and with

\[
\text{E}[G_n(y, x)] = 0, \quad \text{for } (y, x) \in \mathcal{Y} \times \mathcal{X},
\]

\[
\text{E}[G_n(y_1, x_1)G_n(y_2, x_2)] = \text{E}[\phi_{n,y_1,x_1}(Y, X)\phi_{n,y_2,x_2}(Y, X)],
\]

for \( (y_1, x_1) \) and \( (y_2, x_2) \in \mathcal{Y} \times \mathcal{X} \), such that

\[
\sup_{(y,x) \in \mathcal{Y} \times \mathcal{X}} \left| n^{1/2} \hat{U}_n(y, x) - G_n(y, x) \right| = O\left( n^{-1/4}h_n^{-1}(\log n)^{1/2} \right) \quad \text{a.s.}
\]
Proof. As in the proof of Lemma 3.2 of Ghosal, Sen, and van der Vaart (2000), we use Theorem 1.1 of Rio (1994). Since it can be proved using arguments identical to those used to prove Lemma 3.2 of Ghosal, Sen, and van der Vaart (2000), we will only highlight the differences.

To apply Rio’s theorem, we rewrite \( \phi_{n,y,x}(Y,X) \) as

\[
\phi_{n,y,x}(Y,X) = \varphi_{n,u,x}(U,X) = 2[1(U \leq u) - u] \int \text{sgn}(X - \hat{x})K_{h_n}(\hat{x} - x)dF_X(\hat{x})K_{h_n}(X - x),
\]

where \( U = F_Y(Y) \) and \( u = F_Y(y) \). Then \( U \) is uniformly distributed in \([0,1]\) \( \equiv \mathcal{U} \). Thus, Theorem 1.1 of Rio (1994) can be applied to a normalized empirical process associated with \( \varphi_{n,u,x}(U,X) \).

First, we verify that the class of functions \( (v,t) \mapsto h_n \varphi_{n,u,x}(v,t) \), indexed by \((u,x) \in \mathcal{U} \times \mathcal{X}\), is uniformly of bounded variation (UBV). By the definition of Rio (1994), it suffices to show that

\[
\sup_{(u,x) \in \mathcal{U} \times \mathcal{X}} \sup_{g \in \mathcal{D}_2([0,1]^2)} \left( \int_{\mathbb{R}^2} h_n \varphi_{n,u,x}(v,t) \text{ div } g(v,t) \, dv \, dt / \|g\|_\infty \right) < \infty,
\]

where \( \mathcal{D}_2([0,1]^2) \) denotes the space of \( C^\infty \) functions with values in \( \mathbb{R}^2 \) and with compact support included in \([0,1]^2\), \text{ div} denotes the divergence, and \( \|g\|_\infty = \sup_{(v,t) \in \mathbb{R}^2} \|g(v,t)\| \) with \( \|\cdot\| \) being the usual Euclidean norm.

To do so, note that

\[
\int_{\mathbb{R}^2} \varphi_{n,u,x}(v,t) \text{ div } g(v,t) \, dv \, dt \\
= \int_{\mathbb{R}^2} [2[1(v \leq u) - u] \int \text{sgn}(t - \hat{x})K_{h_n}(\hat{x} - x)dF_X(\hat{x})K_{h_n}(t - x) \left[ \frac{\partial g(v,t)}{\partial v} + \frac{\partial g(v,t)}{\partial t} \right] \, dv \, dt \\
= \int_{\mathbb{R}} \int_{\mathbb{R}} [2[1(v \leq u) - u] \frac{\partial g(v,t)}{\partial v} \, dv \int \text{sgn}(t - \hat{x})K_{h_n}(\hat{x} - x)dF_X(\hat{x})K_{h_n}(t - x) \, dt \\
+ \int_{\mathbb{R}^2} [2[1(v \leq u) - u] \int \text{sgn}(t - \hat{x})K_{h_n}(\hat{x} - x)dF_X(\hat{x})K_{h_n}(t - x) \frac{\partial g(v,t)}{\partial t} \, dv \, dt.
\]

Then it is straightforward to verify that

\[
\sup_{g \in \mathcal{D}_2([0,1]^2)} \left( \int_{\mathbb{R}^2} h_n \varphi_{n,u,x}(v,t) \text{ div } g(v,t) \, dv \, dt / \|g\|_\infty \right) = O\left(h_n^{-1}\right)
\]

uniformly over \((u,x) \in \mathcal{U} \times \mathcal{X}\). This implies that the class of functions \( \{h_n \varphi_{n,u,x} : (u,x) \in \mathcal{U} \times \mathcal{X}\} \) satisfies the UBV condition of Rio (1994).
Furthermore, it is also straightforward to verify that

$$
\sup_{g \in D_2([a,b]^2)} \left( \int_{\mathbb{R}^2} h_n \varphi_{n,u,x}(v,t) \ \text{div} \ g(v,t) \ dv \ dt / \|g\|_{\infty} \right) = O \left( h_n^{-1}[b-a] \right)
$$

uniformly over \((u,x) \in \mathcal{U} \times \mathcal{X}\). This implies that the class of functions \(\{h_n \varphi_{n,u,x} : (u,x) \in \mathcal{U} \times \mathcal{X}\}\) also satisfies the LUBV condition of Rio (1994).

We now verify that the class of functions \(\{h_n \varphi_{n,u,x} : (u,x) \in \mathcal{U} \times \mathcal{X}\}\) is a VC class. The function \(h_n \varphi_{n,u,x}\) is bounded by a constant uniformly in \((u,z) \in \mathcal{U} \times \mathcal{X}\) and is obtained by taking an average of

$$
2h_n[1(v \leq u) - 1(\bar{u} \leq u)]\text{sgn}(\bar{x} - t)K_{h_n}(\bar{x} - t)K_{h_n}(t - x)
$$

over \((\bar{u}, \bar{x})\).

Then it is easy to show that \(\{h_n \varphi_{n,u,x} : (u,x) \in \mathcal{U} \times \mathcal{X}\}\) is a VC class by using arguments similar to those used in the proof of Lemma 3.2 of Ghosal, Sen, and van der Vaart (2000, in particular equation 8.5). Finally, by applying Theorem 1.1 of Rio (1994), there exists a sequence of centered Gaussian processes \(G_n(u,x)\) with covariance

$$
E[G_n(u_1,x_1)G_n(u_2,x_2)] = E[\varphi_{n,u_1,x_1}(U,X)\varphi_{n,u_2,x_2}(U,X)].
$$

By switching back to the original variable \(Y\) and its corresponding index \(y\), we obtain the desired result.

Define

$$
\sigma^2(x) = 4 \left[ \int q^2(u)K^2(u)du \right] [f_X(x)]^3.
$$

**Lemma A.3.**

\((a)\) \(\sup_{x \in \mathcal{X}} \left| h_n \sigma_n^2(x) - \sigma^2(x) \right| = o(1)\).

\((b)\) \(\lim_{n \to \infty} \inf_{x \in \mathcal{X}} h_n \inf \sigma_n^2(x) > 0\).

\((c)\) \(\sup_{x \in \mathcal{X}} \left| \hat{\sigma}_n^2(x) - \sigma_n^2(x) \right| = O_p \left( n^{-1/2}h_n^{-2} \right)\).

**Proof.** Since our \(\hat{\sigma}_n^2(x)\) is just three times \(\hat{\sigma}_n^2(t)\) defined in the equation (2.13) of Ghosal, Sen, and van der Vaart (2000), the lemma follows directly from Lemma 3.3 of Ghosal, Sen, and van der Vaart (2000).
Lemma A.4. For the sequence of Gaussian processes \( \{G_n(y, x) : (y, x) \in \mathcal{Y} \times \mathcal{X}\} \) obtained in Lemma A.2, there corresponds a sequence of Gaussian processes \( \{\xi_n(u, s) : (u, s) \in [0, 1] \times \mathcal{X}_n\} \) with continuous sample paths such that

\[
E[\xi_n(u, s)] = 0, \quad E[\xi_n(u_1, s_1)\xi_n(u_2, s_2)] = [\min(u_1, u_2) - u_1u_2] \rho(s_1 - s_2),
\]

for \( u, u_1, u_2 \in [0, 1] \) and \( s, s_1, s_2 \in \mathcal{X}_n \), where \( \rho(\cdot) \) is defined by

\[
\rho(s) = \frac{\int q(z)q(z-s)K(z)K(z-s)dz}{\int q^2(z)K^2(z)dz}
\]

and

\[
\sup_{(y,x) \in \mathcal{Y} \times \mathcal{X}} \left| \frac{G_n(y, x)}{\sigma_n(x)} - \xi_n[F_Y(y), h_n^{-1}(x - x_0)] \right| = O_p\left(h_n\sqrt{\log h_n^{-1}}\right)
\]

Proof. Let \( \mathcal{G}_n \) denote the class of functions \( \{g_{n,u,x} : (u, x) \in \mathcal{U} \times \mathcal{X}\} \), where \( g_{n,u,x}(U, X) = \varphi_{n,u,x}(U, X)/\sigma_n(x) \). Also, let \( \tilde{\mathcal{G}}_n \) denote the class of functions \( \{\tilde{g}_{n,u,x} : (u, x) \in \mathcal{U} \times \mathcal{X}\} \), where

\[
\tilde{g}_{n,u,x}(U, X) = \varphi_{n,u,x}(U, X)/\tilde{\sigma}_n(x),
\]

\[
\varphi_{n,u,x}(U, X) = \begin{cases} 1(U \leq u) - u & \int \text{sgn}(X - \tilde{x}) K_{h_n}(\tilde{x} - x) d\tilde{x} K_{h_n}(X - x), \\
\tilde{\sigma}_n(x) = \left[ \int \left( \int \text{sgn}(\tilde{x} - x) K_{h_n}(\tilde{x} - x) d\tilde{x} \right)^2 [K_{h_n}(\tilde{x} - x)]^2 d\tilde{x} \right]^{1/2} \end{cases}
\]

As explained in Remark 8.3 of Ghosal, Sen, and van der Vaart (2000), it is possible to extend Lemma A.2 in that there exists a sequence of Gaussian processes, say \( \{B_n(g) : g \in \mathcal{G}_n \cup \tilde{\mathcal{G}}_n\} \), with

\[
E[B_n(g)] = 0, \quad E[B_n(g_1)B_n(g_2)] = \text{cov}(g_1, g_2)
\]

for all \( g, g_1, g_2 \in \mathcal{G}_n \cup \tilde{\mathcal{G}}_n \) and with continuous sample paths with respect to the \( L_2 \)-metric such that

\[
G_n(u, x) = \sigma_n(x)B_n(\varphi_{n,u,x}),
\]

where \( G_n(u, x) \) is defined in the proof of Lemma A.2.

Now let \( \tilde{\xi}_n(u, x) = B_n(\tilde{g}_{n,u,x}) \) and \( \gamma_n(u, x) = G_n(u, x)/\sigma_n(x) - \tilde{\xi}_n(u, x) \). As in the proof of Lemma 3.4 of Ghosal, Sen, and van der Vaart (2000), note that \( \gamma_n(u, x) \) is a mean zero Gaussian process with

\[
E[\gamma_n(u_1, x_1)\gamma_n(u_2, x_2)] = E[(g_{n,u_1,x_1} - \tilde{g}_{n,u_1,x_1})(g_{n,u_2,x_2} - \tilde{g}_{n,u_2,x_2})].
\]

Then the lemma can be proved using identical arguments used in the proof of Lemma 3.4 of Ghosal, Sen, and van der Vaart (2000). \( \square \)
A.2 Proofs for Section 3.2

A.2.1 Asymptotic Behaviour of the Excursion Probability on the Fixed Set

We first consider the asymptotic behaviour of the tail probability of the maximum of \( \xi(u, s) \) on a fixed set \([0, 1] \times I\), where \( I \equiv [0, L] \) is an interval with a fixed length \( L \). Define

\[
\Psi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2\right) dx.
\]

**Theorem A.1.** Let \( \lambda \) denote the quantity defined in Theorem 2. In addition, let \( I = [0, L] \).

Then

\[
\Pr\left( \max_{(u, s) \in [0,1] \times I} \xi(u, s) > a \right) = L \left( \frac{8\lambda}{\pi} \right)^{1/2} a \exp(-2a^2)[1 + o(1)]
\]

as \( a \to \infty \).

The following Lemmas are useful to prove Theorem A.1.

**Lemma A.5.** Let \( \Pi_\delta = [1/2 - \delta(a), 1/2 + \delta(a)] \), where \( \delta(a) = a^{-1} \log a \). Then

\[
\Pr\left( \max_{(u, s) \in [0,1] \times I} \xi(u, s) > a \right) = \Pr\left( \max_{(u, s) \in \Pi_\delta \times I} \xi(u, s) > a \right) [1 + o(1)]
\]

as \( a \to \infty \).

**Proof.** For all sufficiently large \( a \),

\[
\Pr\left( \max_{(u, s) \in \Pi_\delta \times I} \xi(u, s) > a \right) \leq \Pr\left( \max_{(u, s) \in [0,1] \times I} \xi(u, s) > a \right)
\]

\[
\leq \Pr\left( \max_{(u, s) \in \Pi_\delta \times I} \xi(u, s) > a \right) + \Pr\left( \max_{(u, s) \in ([0,1] \setminus \Pi_\delta) \times I} \xi(u, s) > a \right).
\]

Note that

\[
E[\xi(u_1, s_1) - \xi(u_2, s_2)]^2 = u_1(1 - u_1) + u_2(1 - u_2) - 2[\min(u_1, u_2) - u_1u_2]\rho(s_1 - s_2).
\]

Furthermore, by some straightforward manipulation,

\[
E[\xi(u_1, s_1) - \xi(u_2, s_2)]^2 \lesssim |u_1 - u_2| + |s_1 - s_2|.
\]

Thus, Assumption E3 of Piterbarg (1996, p.118) is satisfied.

Then since

\[
\max_{(u, s) \in ([0,1] \setminus \Pi_\delta) \times I} \sigma^2(u, s) \leq 1/4 - \delta(a)^2,
\]


\[
\text{Pr} \left( \frac{\max_{(u, s) \in [0,1] \times I} \xi(u, s)}{\sigma(u, s)} > a \right) \sim L \left( \frac{8\lambda}{\pi} \right)^{1/2} a \exp(-2a^2)[1 + o(1)]
\]

as \( a \to \infty \).

The following Lemmas are useful to prove Theorem A.1. **Lemma A.5.** Let \( \Pi_\delta = [1/2 - \delta(a), 1/2 + \delta(a)] \), where \( \delta(a) = a^{-1} \log a \). Then

\[
\Pr\left( \max_{(u, s) \in [0,1] \times I} \xi(u, s) > a \right) = \Pr\left( \max_{(u, s) \in \Pi_\delta \times I} \xi(u, s) > a \right) [1 + o(1)]
\]

as \( a \to \infty \).

**Proof.** For all sufficiently large \( a \),

\[
\Pr\left( \max_{(u, s) \in \Pi_\delta \times I} \xi(u, s) > a \right) \leq \Pr\left( \max_{(u, s) \in [0,1] \times I} \xi(u, s) > a \right)
\]

\[
\leq \Pr\left( \max_{(u, s) \in \Pi_\delta \times I} \xi(u, s) > a \right) + \Pr\left( \max_{(u, s) \in ([0,1] \setminus \Pi_\delta) \times I} \xi(u, s) > a \right).
\]

Note that

\[
E[\xi(u_1, s_1) - \xi(u_2, s_2)]^2 = u_1(1 - u_1) + u_2(1 - u_2) - 2[\min(u_1, u_2) - u_1u_2]\rho(s_1 - s_2).
\]

Furthermore, by some straightforward manipulation,

\[
E[\xi(u_1, s_1) - \xi(u_2, s_2)]^2 \lesssim |u_1 - u_2| + |s_1 - s_2|.
\]

Thus, Assumption E3 of Piterbarg (1996, p.118) is satisfied.

Then since

\[
\max_{(u, s) \in ([0,1] \setminus \Pi_\delta) \times I} \sigma^2(u, s) \leq 1/4 - \delta(a)^2,
\]


\[
\text{Pr} \left( \frac{\max_{(u, s) \in [0,1] \times I} \xi(u, s)}{\sigma(u, s)} > a \right) \sim L \left( \frac{8\lambda}{\pi} \right)^{1/2} a \exp(-2a^2)[1 + o(1)]
\]

as \( a \to \infty \).
by Theorem 8.1 of Piterbarg (1996, p.119), there exists a constant C such that

$$\Pr \left( \max_{(u,s) \in \{0,1\} \times I} \xi(u,s) > a \right) \leq C \text{mes} \left( \{0,1\} \times I \right) a^4 \Psi \left( \frac{a}{\sqrt{1/4 - \delta(a)^2}} \right).$$

Note that by (D.8) of Piterbarg (1996, p.15), as $a \to \infty$,

$$a^4 \Psi \left( \frac{a}{\sqrt{1/4 - \delta(a)^2}} \right) \sim \frac{1}{\sqrt{2\pi}} a^3 \exp \left( \frac{-a^2/2}{1/4 - \delta(a)^2} \right),$$

where $A \sim B$ stands for $A/B \to 1$. Also, for some fixed interior point $\bar{s} \in I$, we have

$$\Pr (\xi(1/2, \bar{s}) > a) = \Psi(a/2) \sim \frac{2}{\sqrt{2\pi}} a^{-1} \exp \left( \frac{-a^2/2}{4} \right),$$

Then it is easy to show that as $a \to \infty$, the probability on the left-hand side of (10) converges to zero at a rate of $\exp \left[ -a^2/8 - O(\log a) \right]$ and $\Pr (\xi(1/2, \bar{s}) > a)$ converges to zero at a rate of $\exp \left[ -a^2/8 - O(\log a) \right]$. Thus, the probability on the left-hand side of (10) converges to zero faster than $\Pr (\xi(1/2, \bar{s}) > a)$. Since $\Pr (\xi(1/2, \bar{s}) > a) \leq \Pr \left( \max_{(u,s) \in \Pi_d \times I} \xi(u,s) > a \right)$, then the lemma follows immediately from (9).

Let $\sigma^2(u,s) = u(1 - u)$ and $r[(u_1, s_1), (u_2, s_2)] = \min(u_1, u_2) - u_1 u_2 \rho(s_1 - s_2)$, respectively, denote the variance and covariance functions of $\xi(u,s)$.

**Lemma A.6.** As $u \to 1/2$,

$$\sigma^2(u,s) = \frac{1}{4} - \left( u - \frac{1}{2} \right)^2 [1 + o(1)]$$

Furthermore, as $(u_1, u_2) \to (1/2, 1/2)$ and $|s_1 - s_2| \to 0$,

$$r[(u_1, s_1), (u_2, s_2)] = \frac{1}{4} - \frac{1}{2} |u_1 - u_2| [1 + o(1)] - \frac{\lambda}{8} (s_1 - s_2)^2 [1 + o(1)] - \frac{1}{2} \left( u_1 - \frac{1}{2} \right)^2 [1 + o(1)] - \frac{1}{2} \left( u_2 - \frac{1}{2} \right)^2 [1 + o(1)].$$

**Proof.** The first result (11) follows easily from a second-order Taylor series expansion of the variance of $\xi(u,s)$ with respect to $u$. We now consider the second result (12). In view of the proof of Theorem 9.2 of Piterbarg (1996, p.138), note that as $(u_1, u_2) \to (1/2, 1/2)$,

$$\frac{\min(u_1, u_2) - u_1 u_2}{\sqrt{u_1(1-u_1)u_2(1-u_2)}} = 1 - \frac{1}{2} \frac{|u_1 - u_2|}{\sqrt{u_1(1-u_1)u_2(1-u_2)}} + o(|u_1 - u_2|).$$
Note that by (4.9) of Ghosal, Sen, and van der Vaart (2000),

\[ \rho(s_1 - s_2) = 1 - \frac{\lambda(s_1 - s_2)^2}{2} + o\left(|s_1 - s_2|^2\right), \]

as \(|s_1 - s_2| \to 0\). As in (11), a Taylor series expansion of \(\sigma(u, s)\) around \(u = 1/2\) gives

\[ \sigma(u, s) = \frac{1}{2} - \left(u - \frac{1}{2}\right)^2 [1 + o(1)], \quad \text{as } u \to \frac{1}{2} \]

for any \(s \in I\). Thus, we have

\[ \sqrt{u_1(1-u_1)u_2(1-u_2)} = \frac{1}{4} - \frac{1}{2} \left(u_1 - \frac{1}{2}\right)^2 [1 + o(1)] - \frac{1}{2} \left(u_2 - \frac{1}{2}\right)^2 [1 + o(1)], \]

as \((u_1, u_2) \to (1/2, 1/2)\). Then the lemma follows from combining (13) and (14) with (15).

Let \(\varepsilon > 0\) be a fixed constant. Define Gaussian processes \(\psi_1^-(u)\) and \(\psi_1^+(u)\) such that

\[ \psi_1^-(u) = \frac{\zeta_1^-(u)}{2^{3/2}[1 + 4(1 + \varepsilon)(u - 0.5)^2]} \quad \text{and} \quad \psi_1^+(u) = \frac{\zeta_1^+(u)}{2^{3/2}[1 + 4(1 - \varepsilon)(u - 0.5)^2]} \]

where \(\zeta_1^-(u)\) and \(\zeta_1^+(u)\) are Gaussian stationary processes with zero means and the covariance functions \(r_1^-(u) = \exp[-4(1 - \varepsilon)|u|]\) and \(r_1^+(u) = \exp[-4(1 + \varepsilon)|u|]\). In addition, define mean-zero stationary Gaussian processes \(\psi_2^-(s)\) and \(\psi_2^+(s)\) such that they are independent of \(\psi_1^-(u)\) and \(\psi_1^+(u)\) and have the the covariance functions of the form

\[ r_2^-(s) = \frac{1}{8} \left[1 - \lambda(1 - \varepsilon)s^2 + o(s^2)\right], \]

\[ r_2^+(s) = \frac{1}{8} \left[1 - \lambda(1 + \varepsilon)s^2 + o(s^2)\right], \]

respectively. Finally, define

\[ \psi^-(u, s) = \psi_1^-(u) + \psi_2^-(s) \quad \text{and} \quad \psi^+(u, s) = \psi_1^+(u) + \psi_2^+(s). \]

**Lemma A.7.** Let \(\varepsilon > 0\) be any fixed, arbitrarily small, constant. Then for all sufficiently large \(a\),

\[ \Pr\left(\max_{(u, s) \in H_4 	imes I} \psi^-(u, s) > a\right) \leq \Pr\left(\max_{(u, s) \in H_4 	imes I} \xi(u, s) > a\right) \leq \Pr\left(\max_{(u, s) \in H_4 	imes I} \psi^+(u, s) > a\right). \]

**Proof.** As noted in the proofs of Theorems D.4 and 8.2 of Piterbarg (1996, p.23 and p.133), the lemma follows from the fact that the distribution of the maximum is monotone with respect to the variance and the Slepian inequality (see, for example, Theorem C.1 of Piterbarg (1996, p.6)).

\[ \square \]
Lemma A.8. Let $\varepsilon > 0$ be any fixed, arbitrarily small, constant. As $a \to \infty$,

\begin{align}
&\Pr \left( \max_{u \in \Pi_s} 2^{3/2} \psi_1^-(u) > a \right) = 2^{1/2} \frac{(1 - \varepsilon)}{(1 + \varepsilon)^{1/2}} \exp(-a^2/2)[1 + o(1)], \\
&\Pr \left( \max_{u \in \Pi_s} 2^{3/2} \psi_1^+(u) > a \right) = 2^{1/2} \frac{(1 + \varepsilon)}{(1 - \varepsilon)^{1/2}} \exp(-a^2/2)[1 + o(1)].
\end{align}

Proof. This lemma can be proved by one of results given in the proof of Theorem D.4 of Piterbarg (1996, p.21). In particular, using the notation used in the proof of Theorem D.4 of Piterbarg (1996), the excursion probability of $2^{3/2} \psi_1^-(u)$ can be obtained by the result of Case 1 with $\alpha = 1, \beta = 2, b = 4(1 + \varepsilon), \text{ and } d = 4(1 - \varepsilon)$. It follows from the second display on page 22 of Piterbarg (1996) that as $a \to \infty$,

\begin{equation}
\Pr \left( \max_{u \in \Pi_s} 2^{3/2} \psi_1^-(u) > a \right) = \frac{H_1 \Gamma(1/2)[4(1 - \varepsilon)]}{[4(1 + \varepsilon)]^{1/2}} a \Psi(a)[1 + o(1)],
\end{equation}

where $H_1$ is the Pickands’ constant with $\alpha = 1$ (defined on pages 13 and 16 of Piterbarg (1996)) and $\Gamma(\cdot)$ is the Gamma function. Note that $\Gamma(1/2) = \sqrt{\pi}$. Furthermore, by (9.6) of Piterbarg (1996, p.138), $H_1 = 1$ and by (D.8) of Piterbarg (1996, p.15),

\begin{equation}
a \Psi(a) \sim (2\pi)^{-1/2} \exp(-a^2/2)
\end{equation}
as $a \to \infty$. Therefore, (16) follows immediately. The excursion probability of $2^{3/2} \psi_1^+(u)$ can be obtained analogously.

Lemma A.9. Let $\varepsilon > 0$ be any fixed, arbitrarily small, constant. As $a \to \infty$,

\begin{align}
&\Pr \left( \max_{s \in I} 2^{3/2} \psi_2^-(s) > a \right) = \frac{[(\lambda/2)(1 - \varepsilon)]^{1/2}L}{\pi} \exp(-a^2/2)[1 + o(1)], \\
&\Pr \left( \max_{s \in I} 2^{3/2} \psi_2^+(s) > a \right) = \frac{[(\lambda/2)(1 + \varepsilon)]^{1/2}L}{\pi} \exp(-a^2/2)[1 + o(1)].
\end{align}

Proof. Recall that $I = [0, L]$. By Theorem D.2 of Piterbarg (1996, p.16) and a simple scaling of $\psi_2^-(u)$,

\begin{equation}
\Pr \left( \max_{s \in [0, L]} 2^{3/2} \psi_2^-(s) > a \right) = H_2 L^2 a \Psi(a)[1 + o(1)]
\end{equation}

where $H_2$ is the Pickands’ constant with $\alpha = 1$ and $L^* = |\lambda(1-\varepsilon)|^{1/2}L$. By (F.4) of Piterbarg (1996, p.31), $H_2 = 1/\sqrt{\pi}$. Then (18) follows immediately. The excursion probability of $2^{3/2} \psi_2^+(u)$ can be obtained similarly. \qed
Proof of Theorem A.1. Let $\varepsilon > 0$ be any fixed, arbitrarily small, constant. Note that $\psi^- (u, s)$ and $\psi^+ (u, s)$ are convolutions of $\psi_1^- (u)$ and $\psi_2^- (s)$ and of $\psi_1^+ (u)$ and $\psi_2^+ (s)$, respectively. Then an application of Lemma 8.6 of Piterbarg (1996, p.128) with Lemmas A.8 and A.9 gives

\[
\Pr \left( \max_{(u,s) \in \Pi_3 \times I} 2^{3/2} \psi^- (u, s) > a \right) = L \left( \frac{(1 - \varepsilon)^{3/2}}{(1 + \varepsilon)^{1/2}} \frac{\lambda}{\pi} \right)^{1/2} a \exp(-a^2/4)[1 + o(1)],
\]

(20)

\[
\Pr \left( \max_{(u,s) \in \Pi_3 \times I} 2^{3/2} \psi^+ (u, s) > a \right) = L \left( \frac{(1 + \varepsilon)^{3/2}}{(1 - \varepsilon)^{1/2}} \frac{\lambda}{\pi} \right)^{1/2} a \exp(-a^2/4)[1 + o(1)].
\]

(21)

Then as $a \to \infty$,

\[
\Pr \left( \max_{(u,s) \in [0,1] \times I} 2^{3/2} \xi (u, s) > a \right) = L \left( \frac{\lambda}{\pi} \right)^{1/2} a \exp(-a^2/4)[1 + o(1)]
\]

since the choice of $\varepsilon$ can be made arbitrarily small and the constants on the right-hand sides of (20) and (21) are continuous at $\varepsilon = 0$. Therefore, the theorem follows immediately. \(\square\)

A.2.2 Asymptotic Behaviour of the Excursion Probability on the Increasing Set

Proof of Theorem 2. Note that the covariance function of $\xi (u, s)$, that is $r[(u_1, s_1), r(u_2, s_2)]$, has compact support and in particular it is zero when $|s_1 - s_2| > 2$. As in the proof of Theorem G.1 of Piterbarg (1996), define an increasing sequence $m_n$ such that $m_n \to \infty$ but $m_n h_n \to 0$ as $n \to \infty$. That is, $m_n$ converges to infinity slower than $h_n^{-1}$. Further, define sequences of sets

\[
I_k = \left[ k(m_n h_n)^{-1}, (k + 1)(m_n h_n)^{-1} - 2 \right],
\]

\[
J_k = \left[ (k + 1)(m_n h_n)^{-1} - 2, (k + 1)(m_n h_n)^{-1} \right],
\]

for $k = 0, 1, \ldots, m_n - 1$. Then as in (G.3) of Piterbarg (1996), we have

\[
\Pr \left( \max_{(u,s) \in [0,1] \times \mathbb{X}_n} \xi (u, s) < a \right) = \Pr \left( \max_{(u,s) \in [0,1] \times \bigcup_k I_k} \xi (u, s) < a \right)
\]

\[
- \Pr \left( \max_{(u,s) \in [0,1] \times \bigcup_k I_k} \xi (u, s) < a, \max_{(u,s) \in [0,1] \times \bigcup_k J_k} \xi (u, s) \geq a \right).
\]

(22)

We first consider the first probability on the right-hand side of (22). Let $c^* = \left( \frac{8A}{\pi} \right)^{1/2}$. Following the idea in the proof of Theorem G.1 of Piterbarg (1996), for each $x$, choose
\[ a_n = \beta_n + x/(4\beta_n), \text{ where } \beta_n \text{ is the largest solution to the following equation:} \]

\[ (23) \quad h_n^{-1} c^* \beta_n \exp(-2\beta_n^2) = 1. \]

Since \( I_k \)'s are separated by the diameter of the support and the distribution of \( \xi(u, s) \) is stationary in the direction of \( s \), it follows from Theorem A.1 that

\[
\Pr \left( \max_{(u,s)\in[0,1]x[\bigcup_k I_k]} \xi(u, s) < a_n \right)
= \left[ 1 - \Pr \left( \max_{(u,s)\in[0,1]xI_0} \xi(u, s) \geq a_n \right) \right]^{m_n}
= \exp \left( -m_n \log \left[ 1 - \Pr \left( \max_{(u,s)\in[0,1]xI_0} \xi(u, s) \geq a_n \right) \right] \right)
= \exp \left( -m_n \Pr \left( \max_{(u,s)\in[0,1]xI_0} \xi(u, s) \geq a_n \right) \right) + O \left( m_n \left[ \Pr \left( \max_{(u,s)\in[0,1]xI_0} \xi(u, s) \geq a_n \right) \right]^2 \right)
= \exp \left\{ -m_n \left[ (m_n h_n)^{-1} - 2c^* a_n \exp(-2a_n^2[1 + o(1)]) \right] + O \left( m_n^{-1} \right) \right\},
\]

so that

\[ (24) \quad \Pr \left( \max_{(u,s)\in[0,1]x[\bigcup_k I_k]} \xi(u, s) < a_n \right) = \exp \left\{ - \exp \left( -x - \frac{x^2}{8\beta_n^2} \right) \left[ 1 + \frac{x}{4\beta_n^2} \right] \right\} + o(1). \]

Now consider the second probability on the right-hand side of (22). As in the proof of Theorem G.1 of Piterbarg (1996), note that again using Theorem A.1 and the fact that the distribution of \( \xi(u, s) \) is stationary in the direction of \( s \),

\[
\Pr \left( \max_{(u,s)\in[0,1]x[\bigcup_k I_k]} \xi(u, s) < a_n, \max_{(u,s)\in[0,1]x[\bigcup_k J_k]} \xi(u, s) \geq a_n \right)
\leq \Pr \left( \max_{(u,s)\in[0,1]x[\bigcup_k J_k]} \xi(u, s) \geq a_n \right)
= m_n \Pr \left( \max_{(u,s)\in[0,1]xJ_1} \xi(u, s) \geq a_n \right)
= m_n \Pr \left( \max_{(u,s)\in[0,1]x[0,2]} \xi(u, s) \geq a_n \right)
= O \left( m_n h_n \right) = o(1).
\]

This and (24) together prove the theorem.

\[ \square \]

**Proof of Theorem 3.2.** The theorem can be proved by arguments identical to those used to prove Theorem 5.1 of Ghosal, Sen, and van der Vaart (2000). In fact, when \( F_x(y|x) > 0 \)
for some \((y,x)\), \(S_n\) is of order \(O_p(n^{1/2}h_n^{3/2})\) and the consistency follows from the restriction that \(nh_n^3/\log h_n^{-1} \to \infty\).

References


Table 1. *Simulation Results*

*Using critical values obtained from the asymptotic expansion $F_n$ of the limiting distribution*

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Bandwidth $h_n = 0.4$</th>
<th>$h = 0.5$</th>
<th>$h = 0.6$</th>
<th>$h = 0.7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rejection proportions when the null hypothesis is true:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 50$</td>
<td>0.014</td>
<td>0.021</td>
<td>0.025</td>
<td>0.030</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.028</td>
<td>0.033</td>
<td>0.034</td>
<td>0.034</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>0.025</td>
<td>0.031</td>
<td>0.036</td>
<td>0.033</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.032</td>
<td>0.039</td>
<td>0.033</td>
<td>0.037</td>
</tr>
<tr>
<td>Rejection proportions when the null hypothesis is false:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 50$</td>
<td>0.687</td>
<td>0.762</td>
<td>0.771</td>
<td>0.760</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.976</td>
<td>0.988</td>
<td>0.989</td>
<td>0.977</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*Using critical values obtained from the type I extreme value distribution*

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Bandwidth $h_n = 0.4$</th>
<th>$h = 0.5$</th>
<th>$h = 0.6$</th>
<th>$h = 0.7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rejection proportions when the null hypothesis is true:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 50$</td>
<td>0.009</td>
<td>0.017</td>
<td>0.013</td>
<td>0.017</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.022</td>
<td>0.024</td>
<td>0.022</td>
<td>0.021</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>0.015</td>
<td>0.021</td>
<td>0.022</td>
<td>0.021</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.021</td>
<td>0.021</td>
<td>0.022</td>
<td>0.023</td>
</tr>
<tr>
<td>Rejection proportions when the null hypothesis is false:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 50$</td>
<td>0.618</td>
<td>0.693</td>
<td>0.697</td>
<td>0.694</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.966</td>
<td>0.976</td>
<td>0.983</td>
<td>0.965</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>