1 Applications to Threshold Estimation Models

1.1 Linear Regression

Consider the model \( Y = \alpha_0 + \beta_0 X + \epsilon \) and \( n \) i.i.d. observations from this model. For simplicity, let \( \epsilon \) be independent of \( X \) with mean 0 and variance \( \sigma^2 \). You can also assume the errors to be normal, even though this is not required for the subsequent development.

Estimates of \((\alpha_0, \beta_0)\) are obtained by minimizing

\[
M_n(\alpha, \beta) \equiv \mathbb{P}_n [(Y - \alpha - \beta X)^2] = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2.
\]

over all \((\alpha, \beta)\). This gives us our standard least squares estimates \((\hat{\alpha}, \hat{\beta})\), whose consistency we will take for granted in the ensuing discussion. Thus:

\[
(\hat{\alpha}, \hat{\beta}) = \text{argmin}_{(\alpha, \beta)} M_n(\alpha, \beta).
\]

It is not difficult to check that:

\[
(\alpha_0, \beta_0) = \text{argmin}_{(\alpha, \beta)} M(\alpha, \beta),
\]

where \( M(\alpha, \beta) = \mathbb{P} [(Y - \alpha - \beta X)^2] \) where \( \mathbb{P} \) is the distribution of \((X, Y)\). Check that, up to a constant, \( M \) is a second order polynomial in \((\alpha, \beta)\) with a constant non–singular Hessian \( 2H \), with \( h_{11} = 1, \) \( h_{22} = E(X^2) \) and \( h_{12} = h_{21} = EX \).

We will apply the rate theorem, Theorem 3.2.5, to deduce the rate of convergence of \((\hat{\alpha}, \hat{\beta})\). Firstly, as

\[
\mathbb{M}(\alpha, \beta) - \mathbb{M}(\alpha_0, \beta_0) = (\alpha - \alpha_0, \beta - \beta_0)^T H (\alpha - \alpha_0, \beta - \beta_0)^T
\]

where \( H \) is p.d., it is easy to see that the first condition of the theorem is satisfied, i.e.

\[
(\alpha - \alpha_0, \beta - \beta_0)^T H (\alpha - \alpha_0, \beta - \beta_0)^T \geq \text{constant} \times d((\alpha, \beta), (\alpha_0, \beta_0))
\]
where \( d((\alpha, \beta), (\alpha_0, \beta_0)) = \max \{ |\alpha - \alpha_0|, |\beta - \beta_0| \} \). Check that the constant can be chosen to be a multiple of the smallest eigen–value of the p.d. matrix \( H \).

In what follows, we generically denote \((\alpha, \beta)\) by \(\theta\) and \((\alpha_0, \beta_0)\) by \(\theta_0\). We seek a bound on:

\[
E^*[\sup_{d(\theta, \theta_0) < \delta} \left| \sqrt{n}(\mathbb{P}_n - P)[(Y - \alpha X - \beta X)^2 - (Y - \alpha_0 X - \beta_0 X)^2] \right|] .
\]

This simplifies easily to

\[
E^* \left[ \sup_{d(\theta, \theta_0) < \delta} \left| \mathcal{G}_n(m_\theta - m_{\theta_0}) \right| \right] \equiv E^* \| \mathcal{G}_n \|_{\mathcal{M}_\delta} (\ast)
\]

where \( \mathcal{G}_n = \sqrt{n}(\mathbb{P}_n - P), m_\theta(X, Y) = (Y - \alpha - \beta X)^2 \), so that:

\[
(m_\theta - m_{\theta_0})(X, Y) = 2 \left[ (\alpha_0 - \alpha) + (\beta_0 - \beta) \right] \left[ Y - (\alpha_0 + \alpha)/2 - ((\beta_0 + \beta)/2) X \right].
\]

Letting

\[
\mathcal{M}_\delta = \{(m_\theta - m_{\theta_0})(X, Y) : d(\theta, \theta_0) < \delta \}
\]

we can find a bound on (\ast) in terms of an envelope function for the class \(\mathcal{M}_\delta\) and the uniform entropy integral \(J(1, \mathcal{M}_\delta)\). For any fixed \(\delta\), the class \(\mathcal{M}_\delta\) is contained in the class of polynomials on \(\mathbb{R}^2\) of degree less than or equal to 2 which is a VC class of functions (since any finite dimensional vector space of measurable functions is VC by Lemma 2.6.15 of Van der Vaart and Wellner (1996)) and therefore itself VC. Also, since \(\delta \leq \delta_0\), for some fixed \(\delta_0\), we can find an envelope function \(M_\delta\) for the class \(\mathcal{M}_\delta\) of the form:

\[
M_\delta(X, Y) = 2 \delta (1 + |X|)(|Y| + C_1 + C_2 |X|)
\]

for positive constants \(C_1, C_2\) (which depend on \(\delta_0\)). By Theorem 2.6.7. of Van der Vaart and Wellner (1996), we have a bound on the covering numbers with respect to the \(L_2\) norm for any probability measure:

\[
N(\epsilon \|M_\delta\|_{Q, 2}, \mathcal{M}_\delta, L_2(Q)) \lesssim \left( \frac{1}{\epsilon} \right)^{2m}
\]

for some \(m \geq 1\). This inequality implies that for any \(\delta\):

\[
J(1, \mathcal{M}_\delta) = \sup_Q \int_0^1 \sqrt{1 + \log N(\epsilon \|M_\delta\|_{Q, 2}, \mathcal{M}_\delta, L_2(Q))} \, d\epsilon
\]

is finite, this number NOT depending on \(\delta\). By the inequalities on Page 291 of Van der Vaart and Wellner (1996) (see also the MAXIMAL INEQUALITIES on Page 199 of Kim and Pollard (1990)), we conclude that:

\[
E^* \| \mathcal{G}_n \|_{\mathcal{M}_\delta} \lesssim J(1, \mathcal{M}_\delta) (P^* M_\delta^2)^{1/2} \leq \text{constant} \times \delta,
\]

since \(P M_\delta^2 = O(\delta^2)\) by the square integrability of \(X\) and \(Y\). Hence \(\phi_n(\delta) = \delta\) works; indeed \(\phi_n(\delta)/\delta^\alpha\) is decreasing for some \(\delta < 2\). Solving \(r_n^2 \phi_n(1/r_n) \leq \sqrt{n}\), yields \(r_n = \sqrt{n}\) easily, showing that:

\[
\sqrt{n}(\hat{\alpha} - \alpha_0, \hat{\beta} - \beta_0) = O_p(1).
\]
Our next step is to obtain the asymptotic distributions of these normalized estimators. To this end, we introduce the local variables \((h_1, h_2)\), where \(\alpha = \alpha_0 + h_1/\sqrt{n}\) and \(\beta = \beta_0 + h_2/\sqrt{n}\). Set \(\hat{h}_1 = \sqrt{n}(\alpha - \alpha_0)\) and \(\hat{h}_2 = \sqrt{n}(\beta - \beta_0)\). We have:

\[
(h_1, h_2) = \text{argmin}_{(h_1, h_2)} \left[ M_n(\alpha_0 + h_1/\sqrt{n}, \beta_0 + h_2/\sqrt{n}) - M_n(\alpha_0, \beta_0) \right]
= \text{argmin}_{(h_1, h_2)} \left[ (P_n - P)[(m_{(\alpha_0+h_1/\sqrt{n}, \beta_0+h_2/\sqrt{n})} - m_{(\alpha_0, \beta_0)})(X, Y)]
+ P[(m_{(\alpha_0+h_1/\sqrt{n}, \beta_0+h_2/\sqrt{n})} - m_{(\alpha_0, \beta_0)})(X, Y)] \right]
= \text{argmin}_{(h_1, h_2)} \left[ n (P_n - P)[(m_{(\alpha_0+h_1/\sqrt{n}, \beta_0+h_2/\sqrt{n})} - m_{(\alpha_0, \beta_0)})(X, Y)]
+ n P[(m_{(\alpha_0+h_1/\sqrt{n}, \beta_0+h_2/\sqrt{n})} - m_{(\alpha_0, \beta_0)})(X, Y)] \right]
\]

Check that the second term inside the argmin converges to \((h_1, h_2) H (h_1, h_2)^T \), where \(H\) is the matrix referred to above and that this convergence is uniform over compact sets. This is a simple consequence of the analytical form of \(M(\alpha, \beta)\):

\[
M(\alpha, \beta) = E(\epsilon^2) + (\alpha - \alpha_0)^2 + (\beta - \beta_0)^2 E(X^2) + 2(\alpha - \alpha_0)(\beta - \beta_0) EX.
\]

The first term inside the argmin simplifies to:

\[
\sqrt{n}(P_n - P)[-(h_1 + h_2 X)] [2(Y - \alpha_0 - \beta_0 X) - n^{-1/2}(h_1 + h_2 X)]
\]

which in terms of the empirical measure and underlying distribution of \((X, \epsilon)\) (which we still continue to denote by \(P_n\) and \(P\) respectively) is simply:

\[
\sqrt{n}(P_n - P)[-h_1 + h_2 X] 2\epsilon + (P_n - P)(h_1 + h_2 X)^2.
\]

The second term in this display converges to 0 in probability, uniformly on compact subsets of \(\mathbb{R}^2\). The first term is simply:

\[
-2(h_1, h_2)(\sqrt{n}(P_n - P)\epsilon, \sqrt{n}(P_n - P)X \epsilon)^T
\]

and converges in distribution, under the topology of uniform convergence on compact sets, to:

\[
2\sigma^2(h_1, h_2)(W_1, W_2)^T
\]

where \(\sigma^2 = E(\epsilon^2)\) and \(W = (W_1, W_2)^T \sim N(0, H)\). It follows that

\[
(\hat{h}_1, \hat{h}_2) \quad \xrightarrow{d} \quad (h_1^*, h_2^*) \equiv \text{argmin}_{h_1, h_2}[2\sigma^2(h_1, h_2)(W_1, W_2)^T + (h_1, h_2) H (h_1, h_2)^T]
\equiv \text{argmin}_{h \in \mathbb{R}^2}[2\sigma^2 h^T W + h^T H h].
\]

Since \(W\) follows \(N(0, H)\), \(V \equiv H^{-1/2} W \sim N(0, I_2)\); letting \(H^{1/2} h = \xi\), the expression inside the argmin in the above display can be written as

\[
2\sigma^2[\xi_1 V_1 + \xi_2 V_2] + \xi_1^2 + \xi_2^2
\]
and this is minimized over all \((\xi_1, \xi_2)\) at \((\hat{\xi}_1 \equiv -\sigma^2 V_1, \hat{\xi}_2 \equiv -\sigma^2 V_2)\), whence
\[
\hat{h} = H^{-1/2} \hat{\xi} \sim N(0, \sigma^2 H^{-1}).
\]
This provides the asymptotic distribution.

**Exercise:** Work out the asymptotics using standard procedures (i.e. arguing from first principles) and show that the results match.

**Exercise:** Prove Corollary 3.2.3 (ii) on Page 288 of Van der Vaart and Wellner (1996) from first principles.

### 1.2 Change Point Estimation

We consider a simple change point estimation problem. A more general treatment is available in Chapter 14 of Michael Kosorok’s notes on Empirical Processes that can be downloaded off his website.

Consider i.i.d. data \(\{X_i, Y_i\}_{i=1}^n\) where \(Y_i = \mu(X_i) + \epsilon_i\). Assume that \(X_i\) is independent of (the error) \(\epsilon_i\) and follows the uniform distribution on \([0, 1]\) and that \(\mu(x) = \alpha_0 1(x \leq d^0) + \beta_0 1(x > d^0)\). Our goal is to estimate the change–point \(d^0\).

The three parameters can be estimated using least squares methods. For now, we will make the (unrealistic) assumption that the two levels of the regression function, \(\alpha_0\) and \(\beta_0\) are known. Then, writing \(\mathbb{P}_n\) as the empirical measure of the data-points \(\{Y_i, X_i\}_{i=1}^n\) we can write down our estimate of \(\hat{d}_n\) as:
\[
\hat{d}_n = \arg\min \mathbb{P}_n [(y - \alpha_0)^2 1(x \leq d) + (y - \beta_0)^2 1(x > d)].
\]

To simplify things, assume that \(\alpha_0 < \beta_0\). A little algebra shows that:
\[
\hat{d}_n = \arg\min \mathbb{P}_n \left[ (y - \alpha_0)^2 - (y - \beta_0)^2 \right] 1(x \leq d) - 1(x \leq d^0)
\]
\[
= \arg\min \mathbb{P}_n \left[ \left( y - \frac{\alpha_0 + \beta_0}{2} \right) 1(x \leq d) - 1(x \leq d^0) \right]
\]
\[
\equiv M_n(d).
\]

If the sequence of stochastic processes \(M_n\) converges to anything, that candidate has to be
\[
\mathbb{M}(d) \equiv P \left[ \left( y - \frac{\alpha_0 + \beta_0}{2} \right) 1(x \leq d) - 1(x \leq d^0) \right].
\]

It is readily checked that \(M_n = |d - d^0| (\beta_0 - \alpha_0)/2\). To show that \(\hat{d}_n\) is consistent for \(d^0\), we can check the conditions of Corollary 3.2.3 (i) of Van der Vaart and Wellner (1996) appropriately.
modified to cater to the fact that we are dealing with minimization instead of maximization. Firstly, note that $\|M_n - M\|_{[0,1]}$ converges to 0 in probability. We have:

$$\|M_n - M\|_{[0,1]} = \sup_{d \in [0,1]} |(P_n - P)f_d(x, y)|$$

where $\{f_d(x, y) \equiv (y - (\alpha_0 + \beta_0)/2)(1(x \leq d) - 1(x \leq d^0)): d \in [0,1]\}$ is a Glivenko-Cantelli class of functions. Furthermore, $M(d^0) = 0 < \inf_{d \in B(\delta', \epsilon)}: M(d) = \epsilon (\beta_0 - \alpha_0)/2$. It follows that $\hat{d}_n$ must converge in probability to $d^0$.

In the discussion above we have not specified the choice of $\hat{d}_n$. However, the minimizer is not unique – there is an entire left–closed right–open interval of minimizers. For this problem it turns out that the limit distribution of $\hat{d}_n$ appropriately normalized, of course, depends on which minimizer is chosen. We will talk about this issue later. The key difference between this example and the others that we will encounter lies in the fact that the limit distribution of the normalized minimizer converges to an appropriate minimizer of a compound Poisson process. We next attempt to determine the rate of convergence.

We apply Theorem 3.2.5 of Van der Vaart and Wellner (1996). The first condition reduces to:

$$M(d) - M(d^0) \geq K \rho^2(d, d^0)$$

for some “distance function” $\rho$. Choosing $\rho(d, d^0) = |d - d^0|^{1/2}$ works. We, next, obtain a bound on the expected modulus of continuity of the empirical process, for all sufficiently small $\delta$, say $\delta < \delta_0$. Thus, we seek to find functions $\phi_n(\delta)$, such that

$$E^* \sup_{|d - d^0|^{1/2} < \delta} |\sqrt{n}(M_n - M)(d) - \sqrt{n}(M_n - M)(d^0)| \lesssim \phi_n(\delta).$$

The left side is simply $E^*[\sup_{d \in [d^0 - \delta^2, d^0 + \delta^2]} |\mathbb{G}_n f_d(x, y)|]$ where $\mathbb{G}_n \equiv \sqrt{n}(P_n - P)$. Set

$$\mathcal{M}_\delta = \{f_d(x, y) : d \in [d^0 - \delta^2, d^0 + \delta^2]\}.$$

A natural envelope function for this class is given by

$$M_\delta = |y - (\alpha_0 + \beta_0)/2| 1(x \in [d^0 - \delta^2, d^0 + \delta^2]).$$

With $P$ denoting the probability distribution of $(X_1, Y_1)$, we have:

$$E^* \|\mathbb{G}_n\|_{\mathcal{M}_\delta} \lesssim J(1, \mathcal{M}_\delta)(P^* M_\delta^2)^{1/2}.$$

It is easily seen that $\mathcal{M}_\delta$ is a VC class of functions and hence satisfies the uniform entropy bound on Page 141 of Van der Vaart and Wellner (1996); consequently, the quantity $J(1, \mathcal{M}_\delta)$ is uniformly bounded for all sufficiently small $\delta$. Check that $P^* M_\delta^2$ is $O(\delta^2)$, showing that the choice $\phi_n(\delta) = \delta$ works.

This yields $r_n = \sqrt{n}$ from the requirement that $r_n^2 \phi_n(r_n^{-1}) \leq \sqrt{n}$. Thus, $\sqrt{n}\rho(\hat{d}_n, d^0) = \sqrt{n}|\hat{d}_n - d^0| = O_p(1)$. It follows that $n(\hat{d}_n - d^0) = O_p(1)$.
1.3 Split Point Estimation

In this section we study the problem of estimating the split point in a binary decision tree for nonparametric regression.

Let $X, Y$ denote the (one-dimensional) predictor and response variables, respectively. The nonparametric regression function $f(x) = E(Y|X = x)$ is to be approximated using a decision tree with a single (terminal) node, i.e., a piecewise constant function with a single jump. The predictor $X$ is assumed to vary in a compact interval $[0, K]$ and to have a density $p_X(\cdot)$. For convenience, we adopt the usual representation $Y = f(X) + \epsilon$, with the error $\epsilon = Y - E(Y|X)$ having zero conditional mean given $X$. The conditional density of $\epsilon$ given $X = x$ is denoted by $p_{\epsilon}(\cdot | x)$, and the conditional variance of $\epsilon$ given $X = x$ is denoted $\sigma^2(x)$.

Suppose we have $n$ i.i.d. observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ of $(X, Y)$. Consider the working model in which $f$ is treated as a piecewise constant function with a single jump, with parameters $(\beta_l, \beta_u, d)$, where $d$ is the point at which the function jumps, $\beta_l$ is the value to the left of the jump and $\beta_u$ is the value to the right of the jump. The best projected values are defined by

$$(\beta_l^0, \beta_u^0, d^0) = \arg\min_{\beta_l, \beta_u, d} E \left[ Y - \beta_l 1(X < d) - \beta_u 1(X \geq d) \right]^2$$

and satisfy the normal equations

$$\beta_l^0 = E(Y|X < d^0), \quad \beta_u^0 = E(Y|X \geq d^0), \quad f(d^0) = \frac{\beta_l^0 + \beta_u^0}{2}.$$

Estimates of these quantities are obtained using least squares as

$$(\hat{\beta}_l, \hat{\beta}_u, \hat{d}_n) = \arg\min_{\beta_l, \beta_u, d} \sum_{i=1}^n \left( Y_i - \beta_l 1(X_i < d) - \beta_u 1(X_i \geq d) \right)^2.$$

The goal is to estimate the split point $d^0$. Before proceeding further, we list some mild conditions.

Conditions

(A1) $d^0$ is the unique solution of the normal equations and $0 < d^0 < K$. Also, $f(d) \neq f(d^0)$ for $d \neq d^0$.

(A2) $f(x)$ is continuous and its first and second derivatives exist and are uniformly bounded in a neighborhood $N$ of $d^0$. Also, $f'(d^0) \neq 0$.

(A3) $p_X(x)$ does not vanish and is continuously differentiable on $(0, K)$.

(A4) $\sigma^2(x)$ is continuous and $0 < \inf_{x \in N} \sigma^2(x) \leq \sup_{x \in N} \sigma^2(x) < \infty$.

(A5) For some $\delta > 0$, $|u|^{3+\delta} \sup_{x \in N} p_\epsilon(u | x) \to 0$ as $|u| \to \infty$. 

6
We now introduce the basic idea used to construct our proposed confidence interval for $d^0$. Consider testing the null hypothesis that the best projected value $d^0$ is located at some point $d$. The best piecewise constant approximation to $f$ under this null hypothesis is estimated by $(\hat{\beta}^d_t, \hat{\beta}^d_u, d)$, where

$$(\hat{\beta}^d_t, \hat{\beta}^d_u) = \text{argmin}_{\beta_t, \beta_u} \sum_{i=1}^n \left( Y_i - \beta_t 1(X_i < d) - \beta_u 1(X_i \geq d) \right)^2.$$ 

A reasonable test statistic is the (centered) residual sum of squares defined by

$$\text{RSS}_n(d) \equiv \sum_{i=1}^n \left( Y_i - \hat{\beta}^d_t 1(X_i < d) - \hat{\beta}^d_u 1(X_i \geq d) \right)^2 - \sum_{i=1}^n \left( Y_i - \beta_t 1(X_i < \hat{d}_n) - \beta_u 1(X_i \geq \hat{d}_n) \right)^2.$$ 

If the working model for $f$ is true, and the errors are Gaussian with constant variance, then $\text{RSS}_n(d)$ is a likelihood ratio statistic, but in general it has no such interpretation.

The key result below provides the joint asymptotic distribution of $\hat{d}_n$ and $\text{RSS}_n(d^0)$ for the “toy version”.

**Theorem 1.1** Suppose conditions (A1)–(A5) hold. Then

$$\left(n^{1/3}(\hat{d}_n - d^0), n^{-1/3} \text{RSS}_n(d^0)\right) \rightarrow_d (\text{argmax}_t Q(t), 2 \mid \beta^0_t - \beta^0_u \mid \max_t Q(t)),$$

where

$$Q(t) = a W(t) - b t^2.$$ 

Here $W(t)$ is standard two-sided Brownian motion started from 0 and $a, b$ are positive constants given by

$$a^2 = p_X(d^0) \sigma^2(d^0), \quad b = \frac{1}{2} \left| p_X(d^0) f'(d^0) \right|.$$ 

**Proof of Theorem:** In what follows we will assume that $\beta^0_t > \beta^0_u$. The derivation for the other case is analogous. Now, elementary algebra shows that

$$(Y_i - \beta^0_t 1(X_i < d) - \beta^0_u 1(X_i \geq d))^2 - Y_i^2 = (\beta^0_t - \beta^0_u) (\beta^0_t + \beta^0_u - 2 Y_i) 1(X_i < d) + \beta^0_u (\beta^0_u - 2 Y_i) .$$
Letting $\mathbb{P}_n$ denote the empirical measure of the pairs $\{(X_i, Y_i)\}_{i=1}^n$, we have

\[
{n^{-1/3} \text{RSS}_n(d^0) = n^{-1/3} \sum_{i=1}^n (\beta_i^0 - \beta_u^0) (\beta_i^0 + \beta_u^0 - 2Y_i) \left(1(X_i < d^0) - 1(X_i < \hat{d}_n)\right)}
\]

\[
= 2n^{-1/3} (\beta_i^0 - \beta_u^0) \sum_{i=1}^n \left( Y_i - \frac{\beta_i^0 + \beta_u^0}{2} \right) \left(1(X_i < \hat{d}_n) - 1(X_i < d^0)\right)
\]

\[
\equiv 2n^{2/3} (\beta_i^0 - \beta_u^0) \mathbb{P}_n g(\cdot, \hat{d}_n),
\]

where

\[
g((X, Y), d) = \left(Y - \frac{\beta_i^0 + \beta_u^0}{2}\right) \left[1(X < d) - 1(X < d^0)\right].
\]

It is easily seen that

\[
\hat{d}_n = \arg\max_d \left\{ \sum_{i=1}^n \left( Y_i - \beta_i^0 1(X_i < d^0) - \beta_u^0 1(X_i \geq d^0)\right)^2 \right. \\
\left. - \sum_{i=1}^n (Y_i - \beta_i^0 1(X_i < d) - \beta_u 1(X_i \geq d))^2 \right\}
\]

\[
= \arg\max_d 2n (\beta_i^0 - \beta_u^0) \mathbb{P}_n g(\cdot, d)
\]

\[
= \arg\max_d n^{2/3} \mathbb{P}_n g(\cdot, d).
\]

Now, define the process

\[
Q_n(t) \equiv n^{2/3} \mathbb{P}_n g(\cdot, d^0 + t n^{-1/3}).
\]

Letting $\hat{t} = n^{1/3}(\hat{d}_n - d^0)$, so that $\hat{d}_n = d^0 + \hat{t} n^{-1/3}$, we have $\hat{t} = \arg\max_{\hat{t}} Q_n(t)$ and $n^{-1/3} \text{RSS}_n(d^0) = 2(\beta_i^0 - \beta_u^0)Q_n(\hat{t})$. It therefore suffices to find the joint limit distribution of $(\hat{t}, Q_n(\hat{t}))$. Lemma 1.1 below shows that the random processes $Q_n(t)$ converge in distribution in the space $B_{loc}(\mathbb{R})$ (the space of locally bounded functions on $\mathbb{R}$ equipped with the topology of uniform convergence on compacta) to the Gaussian process $Q(t)$, whose distribution is a tight Borel measure concentrated on $C_{max}(\mathbb{R})$ (the separable subspace of $B_{loc}(\mathbb{R})$ of all continuous functions on $\mathbb{R}$ that converge to $-\infty$ as the argument runs off to $\infty$ or $-\infty$ and that have a unique maximum). Furthermore, the sequence $\{\hat{t}_n\}$ of maximizers of $\{Q_n(t)\}$ is $O_P(1)$. It then follows by Theorem 1.2 below that:

\[
(\arg\max_{t \in \mathbb{R}} Q_n(t), \max_{t \in \mathbb{R}} Q_n(t)) \equiv (\hat{t}, Q_n(\hat{t})) \rightarrow_d (\arg\max_{t \in \mathbb{R}} Q(t), \max_{t \in \mathbb{R}} Q(t)).
\]

**Theorem 1.2** Suppose that the process $Q_n(t)$ converges in distribution in the space $B_{loc}(\mathbb{R})$ (the space of locally bounded functions on $\mathbb{R}$ equipped with the topology of uniform convergence on compacta) to the Gaussian process $Q(t)$, whose distribution is a tight Borel measure concentrated on $C_{max}(\mathbb{R})$ (the separable subspace of $B_{loc}(\mathbb{R})$ of all continuous functions on $\mathbb{R}$ that converge to $-\infty$ as the argument runs off to $\infty$ or $-\infty$ and that have a unique maximum). Furthermore, suppose that the sequence $\{\hat{t}_n\}$ of maximizers of $\{Q_n(t)\}$ is $O_P(1)$ and converges to $\arg\max_{t \in \mathbb{R}} Q(t)$. Then,

\[
(\hat{t}_n, Q_n(\hat{t}_n)) \rightarrow_d (\arg\max_{t \in \mathbb{R}} Q(t), Q(\arg\max_{t \in \mathbb{R}} Q(t)) \equiv \max_{t \in \mathbb{R}} Q(t)).
\]
Proof: By invoking Dudley’s representation theorem (Theorem 2.2 of Kim and Pollard (1990)), for the processes $Q_n$, we can construct a sequence of processes $\tilde{Q}_n$ and a process $\tilde{Q}$ defined on a common probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ with (a) $\tilde{Q}_n$ being distributed like $Q_n$, (b) $\tilde{Q}$ being distributed like $Q$ and (c) $\tilde{Q}_n$ converging to $\tilde{Q}$ almost surely (with respect to $\tilde{P}$) under the topology of uniform convergence on compact sets. Thus (i) $\tilde{t}_n$, the maximizer of $\tilde{Q}_n$, has the same distribution as $\tilde{t}_n$, (ii) $\tilde{t}$, the maximizer of $\tilde{Q}(t)$, has the same distribution as $\arg\max Q(t)$ and (iii) $\tilde{Q}_n(\tilde{t}_n)$ and $\tilde{Q}(\tilde{t})$ have the same distribution as $Q_n(\tilde{t}_n)$ and $\max Q(t)$ respectively. So to prove the theorem it suffices to show that $\tilde{t}_n$ converges in $\tilde{P}^*$ (outer) probability to $\tilde{t}$ and $\tilde{Q}_n(\tilde{t}_n)$ converges in $\tilde{P}^*$ (outer) probability to $\tilde{Q}(\tilde{t})$. The convergence of $\tilde{t}_n$ to $\tilde{t}$ in outer probability is shown in Theorem 2.7 of Kim and Pollard (1990).

To show that $\tilde{Q}_n(\tilde{t}_n)$ converges in probability to $\tilde{Q}(\tilde{t})$, let $\epsilon > 0, \delta > 0$ be given. We need to show that, eventually,

$$P^*\left(\left|\tilde{Q}_n(\tilde{t}_n) - \tilde{Q}(\tilde{t})\right| > \delta\right) < \epsilon.$$ 

Since $\tilde{t}_n$ and $\tilde{t}$ are $O_P(1)$, given $\epsilon > 0$, we can find $M_\epsilon > 0$ such that, with

$$A_n^c \equiv \{\tilde{t}_n \notin [-M_\epsilon, M_\epsilon]\}, \quad B_n^c \equiv \{\tilde{t} \notin [-M_\epsilon, M_\epsilon]\},$$

$P^*(A_n^c) < \epsilon/4$ and $P^*(B_n^c) < \epsilon/4$, eventually. Furthermore, as $\tilde{Q}_n$ converges almost surely and therefore in probability, uniformly to $\tilde{Q}$ on every compact set, with

$$C_n^c \equiv \{\sup_{s \in [-M_\epsilon, M_\epsilon]} |\tilde{Q}_n(s) - \tilde{Q}(s)| > \delta\},$$

$P^*(C_n^c) < \epsilon/2$, eventually. Hence, eventually, $P^*(A_n^c \cup B_n^c \cup C_n^c) < \epsilon$, so that $P_*(A_n \cap B_n \cap C_n) > 1 - \epsilon$.

But

$$A_n \cap B_n \cap C_n \subset \{|\tilde{Q}_n(\tilde{t}_n) - \tilde{Q}(\tilde{t})| \leq \delta,\}$$ (1.1)

and consequently

$$P_*(|\tilde{Q}_n(\tilde{t}_n) - \tilde{Q}(\tilde{t})| \leq \delta) \geq P_*(A_n \cap B_n \cap C_n) > 1 - \epsilon$$

for all sufficiently large $n$. This implies immediately that for all sufficiently large $n$

$$P^*(|\tilde{Q}_n(\tilde{t}_n) - \tilde{Q}(\tilde{t})| > \delta) < \epsilon.$$ 

It remains to show (1.1). To see this, note that for any $\omega \in A_n \cap B_n \cap C_n$ and $s \in [-M_\epsilon, M_\epsilon]$,

$$\tilde{Q}_n(s) = \tilde{Q}(s) + \tilde{Q}_n(s) - \tilde{Q}(s) \leq \tilde{Q}(\tilde{t}) + |\tilde{Q}_n(s) - \tilde{Q}(s)|.$$ 

Taking the supremum over $s \in [-M_\epsilon, M_\epsilon]$ and noting that $\tilde{t}_n \in [-M_\epsilon, M_\epsilon]$ on the set $A_n \cap B_n \cap C_n$ we have

$$\tilde{Q}_n(\tilde{t}_n) \leq \tilde{Q}(\tilde{t}) + \sup_{s \in [-M_\epsilon, M_\epsilon]} |\tilde{Q}_n(s) - \tilde{Q}(s)|,$$

or equivalently

$$\tilde{Q}_n(\tilde{t}_n) - \tilde{Q}(\tilde{t}) \leq \sup_{s \in [-M_\epsilon, M_\epsilon]} |\tilde{Q}_n(s) - \tilde{Q}(s)|.$$ 


An analogous derivation (replacing \( \tilde{Q}_n \) everywhere by \( Q \), and \( \tilde{t}_n \) by \( t \), and vice-versa) yields
\[
\tilde{Q}(\tilde{t}) - \tilde{Q}_n(\tilde{t}_n) \leq \sup_{s \in [-M_c,M_c]} |\tilde{Q}(s) - \tilde{Q}_n(s)|.
\]
Thus
\[
|\tilde{Q}_n(\tilde{t}_n) - \tilde{Q}(\tilde{t})| \leq \sup_{s \in [-M_c,M_c]} |\tilde{Q}_n(s) - \tilde{Q}(s)| \leq \delta,
\]
which completes the proof.

**Lemma 1.1** The process \( Q_n(t) \) defined in the proof of Theorem 1.1 converges in distribution in the space \( B_{loc}(\mathbb{R}) \) (the space of locally bounded functions on \( \mathbb{R} \) equipped with the topology of uniform convergence on compacta) to the Gaussian process \( Q(t) \equiv aW(t) - b^2 \) whose distribution is a tight Borel measure concentrated on \( C_{max}(\mathbb{R}) \). Here \( a \) and \( b \) are as defined in Theorem 1.1. Furthermore, the sequence \( \{\tilde{t}_n\} \) of maximizers of \( \{Q_n(t)\} \) is \( O_P(1) \) (and hence converges to \( \arg\max_{t} Q(t) \) by Theorem 1.2).

**Proof:** We apply the general approach outlined on page 288 of VdV and Wellner (1996). Letting \( M_n(d) = \mathbb{P}_n [g(\cdot,d)] \) and \( M(d) = \mathbb{P}[g(\cdot,d)] \), we have \( \bar{d}_n = \arg\max_{0 \leq d \leq K} M_n(d) \) and \( d^0 = \arg\max_{0 \leq d \leq K} M(d) \) and, in fact, \( d^0 \) is the unique maximizer of \( M \) under the stipulated conditions. The last assertion needs proof, which will be supplied later. We establish the consistency of \( \tilde{d}_n \) for \( d^0 \) and then find the rate of convergence \( r_n \) of \( \tilde{d}_n \); in other words that \( r_n(\tilde{d}_n - d^0) \) is \( O_P(1) \). For consistency of \( \tilde{d}_n \), note that \( \{g(\cdot,d) : 0 \leq d \leq K\} \) is a VC class of functions and hence universally Glivenko–Cantelli in probability. Therefore
\[
\sup_{0 \leq d \leq K} |M_n(d) - M(d)| = \sup_{0 \leq d \leq K} |(\mathbb{P}_n - \mathbb{P}) g(\cdot,d)| \to 0
\]
in outer probability. Also \( d \mapsto M(d) \) is continuous and therefore upper semicontinuous with unique maximizer \( d^0 \). It follows from Corollary 3.2.3 of VdV and Wellner (1996) that \( \bar{d}_n = \sup_{0 \leq d \leq K} M_n(d) \) converges in probability to \( d^0 \).

Next, to derive the rate \( r_n \), we invoke Theorem 3.2.5 of VdV and Wellner (1996), with \( d \) playing the role of \( \theta \), \( d^0 = \theta_0 \) and \( \Theta = [0,K] \). We have
\[
M(\theta) - M(\theta_0) = M(d) - M(d^0) \leq -C (d - d^0)^2
\]
(for some positive constant \( C \)) for all \( d \) in a neighborhood of \( d^0 \), on using the continuity of \( M''(d) \) in a neighborhood of \( d^0 \) and the fact that \( M''(d^0) < 0 \) (which follows from arguments at the end of this proof). We now need to find functions \( \phi_n(\delta) \) such that
\[
\sqrt{n} \mathbb{E}^* \left[ \sup_{|d - d^0| < \delta} \left| (\mathbb{P}_n - \mathbb{P}) [g(\cdot,d)] - (\mathbb{P}_n - \mathbb{P}) [g(\cdot,d^0)] \right| \right] \leq K \phi_n(\delta)
\]
for some universal positive constant \( K \). Letting \( \mathbb{G}_n = \sqrt{n} (\mathbb{P}_n - \mathbb{P}) \) denote the empirical process and
\[
\mathcal{M}_\delta = \left\{ \left( y - \frac{\beta^0_l + \beta^0_u}{2} \right) \left( 1 \{ x < d \} - 1 \{ x < d^0 \} \right) : |d - d^0| < \delta \right\},
\]
where
\[
\beta^0_l = \sup_{0 \leq d \leq K} \mathbb{P}_n [g(\cdot,d)] - \mathbb{P}[g(\cdot,d)], \quad \beta^0_u = \sup_{0 \leq d \leq K} \mathbb{P}_n [g(\cdot,d)] - \mathbb{P}[g(\cdot,d)],
\]
what we need to show can be restated as

$$E_P \left( \|G_n\|_{M_\delta} \right) \leq K \phi_n(\delta).$$

From page 291 of VdV and Wellner (1996)

$$E_P \left( \|G_n\|_{M_\delta} \right) \leq K' J(1, M_\delta) (P^* M_{\delta}^d)^{1/2},$$

where $M_\delta$ is an envelope function for $M_\delta$ and $K'$ is some universal constant. Since $M_\delta$ is a VC class of functions, it is easy to show that (in the notation of VdV and Wellner (1996))

$$J(1, M_\delta) = \sup_{Q \in Q} \int_0^1 \sqrt{1 + \log N(\epsilon \|M_\delta\|_{Q,2}, M_\delta, L_2(Q))} \, d\epsilon < R,$$

where $R$ is a positive number not depending on $\delta$. We take

$$M_\delta(x, y) = |y - f(d^0)| 1(x \in [d^0 - \delta, d^0 + \delta]).$$

Next, we have

$$E(M_\delta(X,Y)^2) = E \left[ (Y - f(d^0))^2 1 \{ X \in [d^0 - \delta, d^0 + \delta] \} \right]$$

$$= \int_{d^0 - \delta}^{d^0 + \delta} E \left[ (Y - f(d^0))^2 \mid X = x \right] \, p_X(x) \, dx$$

$$= \int_{d^0 - \delta}^{d^0 + \delta} (\sigma^2(x) + (f(x) - f(d^0))^2) \, p_X(x) \, dx$$

$$\leq K \delta,$$

whenever $\delta < \delta^0$ for some constant $K$ (depending on $\delta^0$) by the continuity of all functions involved in the integrand in a neighborhood of $d^0$. Therefore

$$(E M_{\delta}^2)^{1/2} = K^{1/2} \delta^{1/2}$$

and it follows that $E^* (\|G_n\|_{M_\delta}) \leq \tilde{K} \delta^{1/2}$ for some constant $\tilde{K}$ depending neither on $n$, nor on $\delta$, whenever $\delta < \delta^0$. Hence $\phi_n(\delta) = \sqrt{\delta}$ works. Indeed $\phi_n(\delta)/\delta^\alpha$ is decreasing for $\alpha = 1$. Solving

$$r_n^2 \phi_n(1/r_n) \leq \sqrt{n}$$

we find $r_n = n^{1/3}$ works. Since $\hat{d}_n$ maximizes $M_n(d)$, it follows that $n^{1/3} (\hat{d}_n - d^0) = O_P(1)$.

It remains to find the limiting distribution of $\hat{t}_n = n^{1/3} (\hat{d}_n - d^0)$. Now, $\hat{t}_n = \arg\max_t Q_n(t)$ with $Q_n(t) = n^{2/3} \mathbb{P}_n g(\cdot, d^0 + t \, n^{-1/3})$. We show that $Q_n(t) \to_d Q(t)$, a Gaussian process in $C_{\max}(\mathbb{R})$ and then use the argmax continuous mapping theorem to deduce that $\hat{t}_n \to_d \hat{t}$, the unique maximizer of $Q(t)$. Write

$$n^{2/3} \mathbb{P}_n \left[ g(\cdot, d^0 + t \, n^{-1/3}) \right] = n^{2/3} (\mathbb{P}_n - P) \left[ g(\cdot, d^0 + t \, n^{-1/3}) \right] + n^{2/3} P \left[ g(\cdot, d^0 + t \, n^{-1/3}) \right]$$

$$= I_n(t) + II_n(t).$$
In terms of the empirical process \( G_n \), we have \( I_n = G_n(f_{n,t}) \) where

\[
 f_{n,t}(x, y) = n^{1/6}(y - f(d^0))(1(x \leq d^0 + t n^{-1/3}) - 1(x \leq d^0)).
\]

We will use Theorem 2.11.22 from VdV and Wellner (1996). On each compact set \([-K, K]\), \( G_n f_{n,t} \) converges as a process in \( l^\infty [-K, K] \) to the tight Gaussian process \( aW(t) \) with \( a^2 = \sigma^2(d^0) p_X(d^0) \). Hence \( G_n f_{n,t} \) converges to \( aW(t) \) on the line, in the topology of uniform convergence on compact sets. Also, \( II_n(t) \) converges on every \([-K, K]\) uniformly to the deterministic function \(-bt^2\), with \( b = |f'(d^0) p_X(d^0)|/2 > 0 \). Hence \( Q_n(t) \to_d Q(t) \equiv aW(t) - bt^2 \) in \( l^\infty [-K, K] \) for all \( K > 0 \). Consequently, \((t_n, Q_n(t_n)) \to_d (t, Q(t))\).

We now establish that \( I_n \) and \( II_n \) indeed converge to the claimed limits. As far as \( I_n \) is concerned, provided we can verify the other conditions of Theorem 2.11.22, the covariance kernel \( K(s, t) \) of the limit of \( G_n f_{n,t} \) is given by the limit of \( P(f_{n,s} f_{n,t}) - P f_{n,s} P f_{n,t} \) as \( n \to \infty \). We first compute \( P(f_{n,s} f_{n,t}) \). This vanishes if \( s \) and \( t \) are of opposite signs. For \( s, t > 0 \),

\[
 P f_{n,s} f_{n,t} = E[n^{1/3}(Y - f(d^0))^2 1\{X \in (d^0, d^0 + (s \land t) n^{-1/3})\}]
 = \int_{d^0}^{d^0+(s \land t) n^{-1/3}} n^{1/3} E[(f(X) + \epsilon - f(d^0))^2 | X = x] p_X(x) \, dx
 = n^{1/3} \int_{d^0}^{d^0+(s \land t) n^{-1/3}} (\sigma^2(x) + (f(x) - f(d^0))^2) p_X(x) \, dx
 \to \sigma^2(d^0) p_X(d^0) (s \land t)
 \equiv a^2 (s \land t).
\]

Also, it is easy to see that \( P f_{n,s} \) and \( P f_{n,t} \) converge to 0. Thus, when \( s, t > 0 \),

\[
 P(f_{n,s} f_{n,t}) - P f_{n,s} P f_{n,t} \to a^2 (s \land t) \equiv K(s, t).
\]

Similarly, it can be checked that for \( s, t < 0 \), \( K(s, t) = a^2 (-s \land -t) \). Thus \( K(s, t) \) is the covariance kernel of the Gaussian process \( aW(t) \).

Next we need to check

\[
\sup_Q \int_0^{\delta_n} \sqrt{\log N(\epsilon \|F_n\|_{Q,2}, \mathcal{F}_n, L_2(Q))} \, d\epsilon \to 0, \tag{1.2}
\]

for every \( \delta_n \to 0 \). Here

\[
\mathcal{F}_n = \left\{ n^{1/6}(y - f(d^0)) [1(x < d^0 + tn^{-1/3}) - 1(x < d^0)] : t \in [-K, K] \right\}
\]

and

\[
F_n(x, y) = n^{1/6} |y - f(d^0)| 1(x \in [d^0 - Kn^{-1/3}, d^0 + Kn^{-1/3}]).
\]
is an envelope for \( \mathcal{F}_n \). Now

\[
\log N(\epsilon \| F_n \|_{Q,2}, \mathcal{F}_n, L_2(Q)) \leq K V(\mathcal{F}_n) (16 \epsilon)^{V(\mathcal{F}_n)} \left( \frac{1}{\epsilon} \right)^{2 (V(\mathcal{F}_n) - 1)}
\]

for some universal constant \( K \). Here \( V(\mathcal{F}_n) \) is the VC-dimension of \( \mathcal{F}_n \). Using the fact that \( V(\mathcal{F}_n) \) is uniformly bounded, we see that the above inequality implies

\[
\log N(\epsilon \| F_n \|_{Q,2}, \mathcal{F}_n, L_2(Q)) \leq K^* \left( \frac{1}{\epsilon} \right)^s
\]

where \( s = \sup_n 2(V(\mathcal{F}_n) - 1) < \infty \) and \( K^* \) is a constant not depending upon \( n \) or \( Q \). To check (1.2) it therefore suffices to check that

\[
\int_0^{\delta_n} \sqrt{-\log \epsilon} d\epsilon \to 0
\]

as \( \delta_n \to 0 \). But this is trivial. We finally check the conditions (2.11.21) in VdV and Wellner (1996); these are:

\[
P^* F_n^2 = O(1), \quad P^* F_n^2 1\{F_n > \eta \sqrt{n}\} \to 0, \quad \forall \eta > 0,
\]

and

\[
\sup_{\rho(s,t) < \delta_n} P(f_{n,s} - f_{n,t})^2 \to 0, \quad \forall \delta_n \to 0.
\]

With \( F_n \) as defined above, an easy computation shows that

\[
P^* F_n^2 = K \frac{1}{K n^{-1/3}} \int_{d^0 - K n^{-1/3}}^{d^0 + K n^{-1/3}} (\sigma^2(x) + (f(x) - f(d^0))^2) p_X(x) dx = O(1).
\]

Denote the set \([d^0 - K n^{-1/3}, d^0 + K n^{-1/3}]\) by \( S_n \). Then

\[
P^* (F_n^2 1\{F_n > \eta \sqrt{n}\}) = E [n^{1/3} | Y - f(d^0) |^2 1\{X \in S_n\} 1\{|Y - f(d^0)| > \eta n^{1/3}\}] \\
\leq E \left[n^{1/3} | Y - f(d^0) |^2 1\{X \in S_n\} 1\{|\epsilon| > \eta n^{1/3}/2\}\right] \\
\leq E \left[2n^{1/3} (\epsilon^2 + (f(X) - f(d^0))^2) 1\{X \in S_n\} 1\{|\epsilon| > \eta n^{1/3}/2\}\right] \quad (1.3)
\]

eventually, since for all sufficiently large \( n \)

\[
\{|Y - f(d^0)| \geq \eta n^{1/3}\} \subset \{|\epsilon| > \eta n^{1/3}/2\}.
\]

Now, the right side of (1.3) can be written as \( T_1 + T_2 \) where

\[
T_1 = 2n^{1/3} E [\epsilon^2 1\{|\epsilon| > \eta n^{1/3}/2\} 1\{X \in S_n\}]
\]

and

\[
T_2 = 2n^{1/3} E [(f(X) - f(d^0))^2 1\{X \in S_n\} 1\{|\epsilon| > \eta n^{1/3}/2\}] .
\]
We will show that $T_1 = o(1)$. Similar arguments can be used to show that $T_2$ is also $o(1)$. Using condition (A5), eventually

$$T_1 = 2 n^{1/3} \int_{d^0-K n^{-1/3}}^{d^0+K n^{-1/3}} \left( \int_{|u|>\eta n^{1/3}/2} u^2 p_x(u | x) \, du \right) p_X(x) \, dx$$

$$\leq 2 n^{1/3} \int_{d^0-K n^{-1/3}}^{d^0+K n^{-1/3}} \left( \int_{|u|>\eta n^{1/3}/2} u^2 |u|^{-3(3+\delta)} \, du \right) p_X(x) \, dx$$

$$= 2 n^{1/3} \int_{d^0-K n^{-1/3}}^{d^0+K n^{-1/3}} \left( \int_{|u|>\eta n^{1/3}/2} |u|^{-(1+\delta)} \, du \right) p_X(x) \, dx$$

$$= \tilde{k}(\eta, \delta) n^{1/3} \int_{d^0-K n^{-1/3}}^{d^0+K n^{-1/3}} n^{-\delta/3} p_X(x) \, dx$$

$$= 2 \frac{\tilde{k}(\eta, \delta)}{n^{\delta/3}} n^{1/3} \int_{d^0-K n^{-1/3}}^{d^0+K n^{-1/3}} p_X(x) \, dx$$

$$= o(1).$$

In the above display $\tilde{k}(\eta, \delta)$ is a quantity depending only on $\eta$ and $\delta$. Finally, the fact that

$$\sup_{|s-t|<\delta_n} P(f_{n,s} - f_{n,t})^2 \to 0$$

as $\delta_n \to 0$ can be verified through analogous computations which are omitted.

We next deal with $II_n$. For convenience we sketch the uniformity of the convergence of $II_n(t)$ to the claimed limit on $0 \leq t \leq K$. We have

$$II_n(t) = n^{2/3} E \left[ (Y - f(d^0))(1(X < d^0 + t n^{-1/3}) - 1(X < d^0)) \right]$$

$$= n^{2/3} E \left[ (f(X) - f(d^0))(1(X \in [d^0, d^0 + t n^{-1/3}])) \right]$$

$$= n^{2/3} \int_{d^0}^{d^0+t n^{-1/3}} (f(x) - f(d^0)) p_X(x) \, dx$$

$$= n^{1/3} \int_0^t (f(d^0 + u n^{-1/3}) - f(d^0)) p_X(d^0 + u n^{-1/3}) \, du$$

$$= \int_0^t u \frac{f(d^0 + u n^{-1/3}) - f(d^0)}{u n^{-1/3}} p_X(d^0 + u n^{-1/3}) \, du$$

$$\to \int_0^t u f'(d^0) p_X(d^0) \, du \quad \text{(uniformly on } 0 \leq t \leq K)$$

$$= \frac{1}{2} f'(d^0) p_X(d^0) t^2 = -b t^2.$$
\[
\begin{align*}
E \left[ \left( Y - \frac{\beta_l^0 + \beta_u^0}{2} \right) (1(X < d) - 1(X < d^0)) \right].
\end{align*}
\]

Let,
\[
\xi(d) = E \left[ Y - \beta_l^0 1(X < d) - \beta_u^0 1(X \geq d) \right]^2.
\]

By condition (A1), it follows immediately that \( d^0 \in (0, K) \) is the unique minimizer of \( \xi(d) \). Consequently, \( d^0 \) is also the unique maximizer of the function \( \xi(d^0) - \xi(d) \). Straightforward algebra shows that
\[
\xi(d^0) - \xi(d) = 2 \left( \beta_l^0 - \beta_u^0 \right) M(d)
\]
and since \( \beta_l^0 - \beta_u^0 > 0 \), it follows that \( d^0 \) is also the unique maximizer of \( M(d) \). This shows (i).

Now,
\[
\begin{align*}
M(d) &= E \left[ \left( f(X) - \frac{\beta_l^0 + \beta_u^0}{2} \right) (1(X < d) - 1(X < d^0)) \right] + E \left[ \epsilon (1(X < d) - 1(X < d^0)) \right] \\
&= E \left[ \left( f(X) - \frac{\beta_l^0 + \beta_u^0}{2} \right) (1(X < d) - 1(X < d^0)) \right] + 0 \\
&= \int_0^K \left( f(x) - \frac{\beta_l^0 + \beta_u^0}{2} \right) \left( 1(x < d) - 1(x < d^0) \right) p_X(x) dx \\
&= \int_0^d \left( f(x) - \frac{\beta_l^0 + \beta_u^0}{2} \right) p_X(x) dx - \int_{d^0}^d \left( f(x) - \frac{\beta_l^0 + \beta_u^0}{2} \right) p_X(x) dx.
\end{align*}
\]

Thus, for \( 0 < d < K \)
\[
M'(d) = (f(d) - f(d^0)) p_X(d),
\]
on recalling that \( f(d^0) = (\beta_l^0 + \beta_u^0)/2 \) and
\[
M''(d) = f'(d) p_X(d) + (f(d) - f(d^0)) p_x'(d).
\]
Thus, \( M'(d^0) = 0 \) and \( M''(d^0) = f'(d^0) p_X(d^0) \). Since \( d^0 \) is the maximizer at an interior point, \( M''(d^0) \leq 0 \). This implies (ii), since by our assumptions \( f'(d^0) p_X(d^0) \neq 0 \).