UNIQUENESS AND GLOBAL OPTIMALITY OF THE MAXIMUM LIKELIHOOD ESTIMATOR FOR THE GENERALIZED EXTREME VALUE DISTRIBUTION*

BY LIKUN ZHANG1,* BENJAMIN SHABY2

1Department of Statistics, Pennsylvania State University, *likunz@psu.edu
2Department of Statistics, Colorado State University, bshaby@colostate.edu

The three-parameter generalized extreme value distribution arises from classical univariate extreme value theory and is in common use for analyzing the far tail of observed phenomena. Curiously, important asymptotic properties of likelihood-based estimation under this standard model have yet to be established. In this paper, we formally prove that the maximum likelihood estimator is global and unique. An interesting secondary result entails the uniform consistency of a class of limit relations in a tight neighborhood of the shape parameter.

1. Introduction. Classical extreme value theory was introduced almost a century ago (Fisher and Tippett, 1928) and is in wide practical use, yet a basic theoretical elucidation of likelihood-based inference under its central distributional construct remains incomplete. Here, we fill in some of the important gaps. The generalized extreme value (GEV) distribution arises as the only limit of suitably renormalized maxima over independent and identically distributed (iid) random variables, and has therefore routinely been used in modeling the tail behavior of observed phenomena. However, since the support of the distribution depends on its parameters, standard regularity conditions of classic asymptotic theory are not satisfied. It is only recently that consistency and asymptotic normality of the maximum likelihood estimator (MLE), found locally on a restricted compact set, have been established. In this paper, we show that the local MLE uniquely and globally maximizes the GEV log-likelihood function. In addition, we establish a number of convergence properties related to the GEV, including uniform consistency of a class of limit relations, revealing a much richer understanding of the GEV likelihood than has previously appeared.

The family of GEV distributions forms a continuous parametric family with respect to $\theta = (\tau, \mu, \xi)$ on some measurable space $(\mathcal{X}, \mathcal{A})$,

$$P_\theta(y) = \exp \left\{- \left[ 1 + \xi \left( \frac{y - \mu}{\tau} \right) \right]^{-1/\xi} \right\} \quad \xi \neq 0$$

$$= \exp \left\{ - \exp \left[ - \frac{y - \mu}{\tau} \right] \right\} \quad \xi = 0,$$

where $1 + \xi \left( \frac{y - \mu}{\tau} \right) > 0$ for $\xi \neq 0$, and the scale parameter $\tau > 0$, location parameter $\mu \in \mathbb{R}$, shape parameter $\xi \in \mathbb{R}$. The GEV distribution unites the Gumbel, Fréchet and Weibull distributions into a single family to allow various shapes.

The estimation of GEV parameters, especially the shape parameter $\xi$, is pivotal in studying tail behavior. The Picklands estimators (Pickands, 1975), probability weighted moments...
estimators (Hosking, Wallis and Wood, 1985), and method of moments quantile estimators (Madsen, Rasmussen and Rosbjerg, 1997) are among many estimators available in the literature. Beirlant et al. (2004) provides a detailed review of the aforementioned estimators. In this paper, we focus on the asymptotic properties of estimators obtained from the maximum likelihood method. Denote $p_\theta$ as the density function of $P_\theta$ with respect to some dominating measure $\mathcal{P}$. Since the support of the GEV density function is not independent of $\theta$, the regularity conditions for standard likelihood inference do not hold, which gives rise to fundamental difficulties when studying the existence, consistency and asymptotic normality of the MLE.

Suppose $\theta_0 = (\tau_0, \mu_0, \xi_0)$ is the true parameter, and $Y_1, \ldots, Y_n$ are independent samples from $P_{\theta_0}$. Smith (1985) was the first to consider the MLE of a large class of irregular parametric families, whose formulation includes the GEV distribution when $-1 < \xi_0 < 0$. Treating the samples as coming from a distribution in the domain of attraction of a GEV, Dombry (2015) derived the existence of local MLE, which is implicitly defined as solutions of the score functions, under the setting of triangular arrays of block maxima when $\xi_0 > -1$. He proved that for any fixed compact set $K \subset \{ \theta : \tau > 0, \mu \in \mathbb{R}, \xi > -1 \}$ that contains $\theta_0$, the maximum of the likelihood function in $K$ is confined in an arbitrarily smaller neighborhood of $\theta_0$, $\tilde{K}$, for all $n$ large enough. The corresponding MLE $\hat{\theta}_n = (\tilde{\xi}_n, \tilde{\mu}_n, \tilde{\tau}_n)$ solves the score functions, and it converges almost surely to $\theta_0$.

Bücher and Segers (2017) extended this result, in the simpler setting where $Y_1, \ldots, Y_n$ are assumed to be independent samples from univariate GEV distribution, establishing a $O_p(1/\sqrt{n})$ rate of convergence for the local MLE, and refining the incomplete proof of Smith (1985) to establish the asymptotic normality of $\tilde{\theta}_n$ for $\xi_0 > -1/2$.

However, it is not guaranteed that the local MLE $\tilde{\theta}_n$ studied by Dombry (2015) and Bücher and Segers (2017) attains a unique, global maximum of the log-likelihood

$$L_n(\theta) = \sum_{i=1}^{n} l_\theta(Y_i),$$

in which $l_\theta : \theta \mapsto \log p_\theta(y)$, and $\theta \in \Omega_n = \{ \theta : p_\theta(Y_i) > 0, i = 1, \ldots, n \}$. Among other things, the uniform and global properties of $L_n$ in $\Omega_n$ are needed in Bayesian theory to develop optimal decision rules and perform posterior-based inference (Hartigan, 1983), to establish asymptotic posterior normality (Von Mises, 1931; Chen, 1985), and to construct rule-based noninformative priors (Bernardo, 2005). If $L_n$ is highly peaked and concentrates in a small neighborhood of $\theta_0$, the information contained in the observations $Y_1, \ldots, Y_n$ will dominate any prior knowledge as $n$ approaches infinity, and hence posterior distribution will behave like a normal kernel.

In this paper, we only consider $\theta_0 \in \Theta = (0, \infty) \times \mathbb{R} \times (-1/2, \infty)$. We will prove that the local MLE gives a unique, global maximum point for the log-likelihood function by following a two-step strategy:

(I) We first construct a small compact set $\tilde{K}$ containing $\theta_0$ in its interior, and prove that for all large $n$, $L_n$ in $\tilde{K}$ is strictly concave and attains a unique maximum point;

(II) We then define a larger compact set $K$, explicitly defined in terms of $\theta_0$, such that $\tilde{K} \subset K$. We prove for all large $n$, global maximum must be attained in $K$; that is, $\arg \max_{\theta \in \Theta} L_n(\theta) \in \tilde{K}$.

By Proposition 2 in Dombry (2015), $\tilde{\theta}_n = \arg \max_{\theta \in \tilde{K}} L_n(\theta) \in \tilde{K}$ for all large $n$. We will therefore conclude that $L_n(\tilde{\theta}_n)$ is indeed the unique and global maximum $L_n$—the global optimality is ensured by (II), while the uniqueness is ensured by (I). This main result is stated in the following theorem.
Theorem 1.1 (Global optimality and uniqueness). Suppose $Y_1, Y_2, \ldots \iid P_{\theta_0}$, and $\hat{\theta}_n$ is the sequence of local maxima of $L_n$ that is found on a fixed compact neighborhood of $\theta_0$. Define $\Theta_n = \{ \theta \in \Theta : \xi < n - 1 \}$. Then

$$ \mathcal{P} \left( \exists N > 0 \text{ such that for all } n > N, \arg \max_{\theta \in \Theta_n} L_n(\theta) \text{ is unique and equals } \hat{\theta}_n \right) = 1. $$

Remark 1. Later we will show that for any $n$ and a sequence of numbers $Y_1, \ldots, Y_n$ that are not equal, $\max_{\theta \in \{ \xi \geq n - 1 \}} L_n(\theta) = \infty$. One may object that the optimality result is not truly global because of the restriction $\xi < n - 1$. Considering the shape parameters are less than 1 for most observed data-generating processes, the ever-expanding $\Theta_n$ is hardly a restriction, and does not interfere with the derivation of asymptotic posterior properties.

Remark 2. Our asymptotic results are probabilistic in nature—that is, we do not treat the observed $Y_1, Y_2, \ldots$ as a deterministic sequence (as opposed to Chen, 1985, e.g.). Under the dominating measure $\mathcal{P}$, the desired properties hold almost surely.

To prove (I) and (II), we first examine the support $\Omega_n$ of the log-likelihood $L_n$ in Section 2.1. In Section 2.2, we study the properties of $L_n$ for a fixed $n$ and $Y_1, \ldots, Y_n$. To circumvent the non-convexity of the support $\Omega_n$, we slice $\Omega_n$ at different levels of $\xi$, and work with the maximum profile likelihood $PL_n(\xi)$ (Murphy and van der Vaart, 2000), which is defined as the maximum value of $L_n$ on the slice of $\xi$. With the help of the classic Seitz inequalities (Seitz, 1936), we show that $L_n$ is uniquely maximized on each slice of $\xi$. To locate the global maximum of $L_n$, we only need to compare the profile likelihood $PL_n(\xi)$ across different $\xi$ values. In Section 3, we note that the boundary of $\Omega_n$ becomes infinitely close to $\theta_0$ as $n$ approaches infinity. Although the convergence rate of the boundary to $\theta_0$ is slower than the $1/\sqrt{n}$ rate of local MLE, the close proximity of the two poses fundamental challenges on deriving asymptotic properties. In Section 4, we overcome these challenges, and obtain the local concavity condition (I) in section 4.1 via establishing pointwise and uniform consistencies for a class of limit relations. We then prove the result (II) in Section 4.2. Finally, we conclude that $\theta_n$ is indeed the unique global maximum point for $L_n$.

2. Preliminaries.

2.1. The joint likelihood function and its support. First we define

$$ \beta = \beta(\theta) = \mu - \frac{\tau}{\xi}. $$

This one-to-one mapping from $(\tau, \mu, \xi)$ to $(\tau, \beta, \xi)$ will be used in the subsequent analysis to simplify notation. In addition, define

$$ w_i(\theta) := 1 + \frac{\xi}{\tau} \left( \frac{Y_i - \mu}{\tau} \right) = \frac{\xi}{\tau} (Y_i - \beta), $$

which helps alleviate the complexity of the log-likelihood function:

$$ L_n(\theta) = -n \log \tau - \frac{\xi + 1}{\xi} \sum_{i=1}^{n} \log w_i(\theta) - \sum_{i=1}^{n} w_i^{-1/\xi}(\theta), \text{ when } \xi \neq 0. $$

The form of $\beta$ also helps to concisely delineate the support of the joint density function, i.e., the domain of the log-likelihood $L_n(\theta)$, which can be written out as follows, given the observations $Y_1, \ldots, Y_n$:

$$ \Omega_n = \{ \theta \in \Theta : w_i(\theta) > 0, i = 1, \ldots, n \} $$

$$ = \{ \theta \in \Theta : \xi(Y_i - \beta) > 0, i = 1, \ldots, n \}. $$

The joint likelihood function and its support. First we define

$$ \beta = \beta(\theta) = \mu - \frac{\tau}{\xi}. $$

This one-to-one mapping from $(\tau, \mu, \xi)$ to $(\tau, \beta, \xi)$ will be used in the subsequent analysis to simplify notation. In addition, define

$$ w_i(\theta) := 1 + \frac{\xi}{\tau} \left( \frac{Y_i - \mu}{\tau} \right) = \frac{\xi}{\tau} (Y_i - \beta), $$

which helps alleviate the complexity of the log-likelihood function:

$$ L_n(\theta) = -n \log \tau - \frac{\xi + 1}{\xi} \sum_{i=1}^{n} \log w_i(\theta) - \sum_{i=1}^{n} w_i^{-1/\xi}(\theta), \text{ when } \xi \neq 0. $$

The form of $\beta$ also helps to concisely delineate the support of the joint density function, i.e., the domain of the log-likelihood $L_n(\theta)$, which can be written out as follows, given the observations $Y_1, \ldots, Y_n$:

$$ \Omega_n = \{ \theta \in \Theta : w_i(\theta) > 0, i = 1, \ldots, n \} $$

$$ = \{ \theta \in \Theta : \xi(Y_i - \beta) > 0, i = 1, \ldots, n \}. $$
It can be easily verified that \( \Omega_n \) is not a convex set, which means Taylor expansion will not be helpful for studying \( L_n(\theta) \), which precludes the use of routine techniques such as the mean-value theorem, and makes it difficult to approximate the difference of the function on a certain intervals. Nonetheless, if we slice \( \Omega_n \) at different levels of \( \xi \), every cross section is convex; see Figure 1 for illustration. On a cross section at a fixed \( \xi \), the value of \( \beta = \mu - \tau / \xi \) can be construed as the intercept of the line which has a slope of \( 1/\xi \) and passes through \((\tau, \mu)\). When \( \xi > 0 \), the condition in (3) requires this intercept \( \beta < Y(1) \), and when \( \xi < 0 \), the intercept \( \beta > Y(n) \), where \( Y(1) \) and \( Y(n) \) are the minimum and maximum values of the observations, respectively. Therefore, for any \( \theta \in \Theta \), we can immediately tell whether \( \theta \in \Omega_n \) using only \( Y(1) \) and \( Y(n) \).

2.2. Profile likelihood. Denote the cross section of \( \Omega_n \) at a certain \( \xi \) as \( \Omega_n(\xi) \). The convexity of \( \Omega_n(\xi) \) suggests examining the log-likelihood via profiling out \((\tau, \mu)\):

\[
PL_n(\xi) := \sup_{(\tau, \mu) \in \Omega_n(\xi)} L_n(\theta).
\]

We show in the following proposition that on each cross section \( \Omega_n(\xi) \), \( L_n(\theta) \) is uniquely maximized.

**Proposition 2.1.** Suppose the sequence \( Y_1, \ldots, Y_n \) are not all equal to each other. For any \( \xi \neq 0 \) and \(-1 < \xi < n - 1\), there exists a maximizer \((\tau_n(\xi), \mu_n(\xi))\) of the log-likelihood \( L_n \) that is unique and global on the cross section \( \Omega_n(\xi) \). The values of \( \tau_n(\xi) \) and \( \mu_n(\xi) \) are implicitly defined by \( \xi \) and the observations through

\[
\begin{align*}
\tau &= \left\{ \frac{1}{n} \sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1/\xi} \right\}^{-\xi}, \\
(\xi + 1) \sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1} &= \frac{n \sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1-1/\xi}}{\sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1/\xi}}.
\end{align*}
\]

**Proof.** See Appendix A.

**Remark 3.** As defined in (4), \( PL_n(\xi) = L_n(\tau_n(\xi), \mu_n(\xi), \xi) \). Inserting (5) into (2),

\[
PL_n(\xi) = -n \log \left\{ \frac{1}{n} \sum_{i=1}^{n} [\xi(Y_i - \beta_n(\xi))]^{-1/\xi} \right\} - \frac{\xi + 1}{\xi} \sum_{i=1}^{n} \log[\xi(Y_i - \beta_n(\xi))] - n,
\]

where \( \beta_n(\xi) = \mu_n(\xi) - \tau_n(\xi)/\xi \), \(-1 < \xi < n - 1\) and \( \xi \neq 0 \).
REMARK 4. We show within the proof of Proposition 2.1 in Appendix A that $PL_n(\xi) = \infty$, if $\xi > n - 1$. For $\xi = 0$, there also exists an unique maximizer $(\tau_n(0), \mu_n(0))$ on the cross section $\Omega_n(0)$, which is implicitly defined by

$$
\begin{align*}
\{ & n\tau = \sum_{i=1}^{n} \left[ 1 - \exp \left( -\frac{Y_i - \mu}{\tau} \right) \right] Y_i, \\
& n = \sum_{i=1}^{n} \exp \left( -\frac{Y_i - \mu}{\tau} \right).
\end{align*}
$$

By the continuity of $L_n$ at $\xi = 0$, we know that

$$
\lim_{\xi \to 0} \mu_n(\xi) = \mu_n(0), \quad \lim_{\xi \to 0} \tau_n(\xi) = \tau_n(0) \quad \text{and} \quad \lim_{\xi \to 0} PL_n(\xi) = PL_n(0).
$$

REMARK 5. Notice that for this proposition, we are not assuming that $Y_1, \ldots, Y_n$ are drawn from $P_{\theta_0}$. Rather they can be independent copies of any unitary distribution on $X$. The fact that $PL_n$ becomes senseless when $\xi > n - 1$ is a result of the mathematical form of the GEV log-likelihood $L_n$.

Since this result does not rely on the asymptotics of the GEV distribution, the lower bound of $\xi$ is extended to $-1$. The reason we restrict the lower bound to $-1/2$ in Theorem 1.1 is that the propositions in the forthcoming sections require second-order consistencies. The existence and first-order consistency arguments in Dombry (2015) also pertain to $\xi > -1$.

To find the global maximum, we now only need to compare the 'representative' maximum likelihood from the each cross section. If the profile likelihood $PL_n$ as a function of $\xi$ is strictly concave in $(-1, n - 1)$, it has a unique maximum at a $\xi$ value such that $PL_n(0) = 0$, and then $(\tau_n(0), \mu_n(0))$ is the unique global maximizer for $L_n$. Unfortunately, $PL_n$ is not a strictly concave function of $\xi$. As demonstrated in the following proposition, the first derivative $PL'_n$ is not monotonically decreasing, and it behaves irregularly when $\xi$ approaches the bounds of $(-1, n - 1)$.

PROPOSITION 2.2. Under the assumptions of Proposition 2.1, the first derivative $PL'_n$ is well-defined and continuous in $\xi \in (-1, n - 1)$. When $\xi \searrow -1$, $PL'_n(\xi) \to -\infty$. When $\xi \nearrow n - 1$, $PL'_n(\xi) \to \infty$. By the intermediate zero theorem, there must exist a $\xi \in (-1, n - 1)$ such that $PL'_n(\xi) = 0$.

PROOF. See Appendix A. \hfill \Box

REMARK 6. For $\xi \neq 0$,

$$
PL'_n(\xi) = -\frac{n}{\xi} - \frac{n}{\xi^2} \sum_{i=1}^{n} [\xi(Y_i - \beta_n(\xi))]^{-1/\xi} \log [\xi(Y_i - \beta_n(\xi))]^{-1/\xi} + \frac{1}{\xi^2} \sum_{i=1}^{n} \log [\xi(Y_i - \beta_n(\xi))].
$$

For $\xi = 0$, the first derivative can be defined as the limit:

$$
\lim_{\xi \to 0} PL'_n(\xi) = \frac{n\mu'_n(0) - \sum_{i=1}^{n} [Y_i - \mu_n(0) + \tau'_n(0)]}{\tau_n(0)} + \frac{\sum_{i=1}^{n} [Y_i - \mu_n(0) + \tau'_n(0)]^2 - n\tau'_n(0)^2}{2\tau_n(0)^2}.
$$

It is easy to verify that if a $\xi$ value solves $PL'_n(\xi) = 0$, (5) and (7) together ensure that $(\tau_n(\xi), \mu_n(\xi), \xi)$ solves the score equations of $L_n$. Hence this result provides an alternative approach to proving the existence of the local MLE for $L_n$. However, acquiring the strong consistency of the local MLE requires the assumption that $Y_1, \ldots, Y_n \overset{\text{ind}}{\sim} P_{\theta_0}$ and the limiting behavior as $n \to \infty$. 


Figure 2 illustrates some key features of the profile likelihood function. We simulate $Y_1, \ldots, Y_n$ from $P_{\theta_0}$ and calculate the profile likelihood $PL_n$ at a grid of $\xi$ values ranging from $-1$ to $1$. For both positive shape parameter $\xi_0 = 0.2$ and negative shape parameter $\xi_0 = -0.2$, $PL_n$ appears to be uniquely maximized by the local MLE which is close to $\xi_0$. Although it is not a concave function globally, we still observe local concavity around $\xi_0$, which suggests adoption of the two-step strategy introduced in Section 1. Roughly speaking, these two steps are established in Section 4 via proving (I) $PL_n$ is strictly concave in a small neighborhood of $\hat{\xi}_n$, and (II) $PL_n(\xi) < PL_n(\hat{\xi}_n)$ for $\xi$ that is far from $\hat{\xi}_n$.

3. Convergence rate of the support boundary. To prove (I) and (II), we will need to study the distance between the true parameter $\theta_0$ and the boundary of the support $\Omega_n$. It is true from the definition of $\Omega_n$ that if $Y_1, \ldots, Y_n$ are drawn from $P_{\theta_0}$,

$$\theta_0 \in \Omega_n$$

for any $n \geq 1$.

It is clear that $\Omega_n$ is an open set for any $n$, and hence the true parameter $\theta_0$ is always an interior point of $\Omega_n$. This raises the question: can we always find a neighborhood of $\theta_0$ that is large enough to allow us to examine the log-likelihood in the vicinity of $\theta_0$? Unfortunately, this is not possible because $\theta_0$ becomes infinitely close to the boundary as $n$ approaches infinity.

To quantify the distance between $\theta_0$ and the boundary of $\Omega_n$, we first assume $\xi_0 > 0$ and examine the cross section $\Omega_n(\xi_0)$. This is illustrated in Figure 3, where $\theta_0 = (\tau_0, \mu_0, \xi_0)$ is shown as a red point, and $\beta_0 = \mu_0 - \tau_0/\xi_0$ is the intercept of the line that passes through $(\tau_0, \xi_0)$ with a slope of $1/\xi_0$. Figure 3 illustrates that the difference of intercepts, $Y_{(1)} - \beta_0$, is a good measure of the distance. By analogy, if true shape parameter $\xi_0 < 0$, the distance can be well-measured by $\beta_0 - Y_{(n)}$.

Note that when $\xi_0 > 0$, the distribution of $P_{\theta_0}$ is lower bounded by $\beta_0$, which guarantees $Y_{(1)} \overset{a.s.}{\rightarrow} \beta_0$. When $\xi_0 < 0$, the distribution of $P_{\theta_0}$ is upper bounded by $\beta_0$, which guarantees $Y_{(n)} \overset{a.s.}{\rightarrow} \beta_0$. Thus in both cases, the distance between $\theta_0$ and the boundary of $\Omega_n$ converges almost surely to zero. Also, Bücher and Segers (2017) showed that

$$\hat{\theta}_n - \theta_0 = O_p(1/\sqrt{n}),$$

which means $\hat{\theta}_n$ is also infinitely close to $\theta_0$ as $n$ grows, and thus close to the boundary of $\Omega_n$. This is concerning for the purpose of proving global optimity of $\hat{\theta}_n$ because it would be rather challenging to handle the log-likelihood near the boundary of the support.
If \(B > A\) and \(\theta \in \tilde{\Omega}(\xi_0)\) if true \(\xi_0 > 0\). The two parallel dashed lines have a slope of \(1/\xi_0\). The red point is \(\theta_0 = (\tau_0, \mu_0, \xi_0)\). Here we also compare the convergence rates of \(\hat{\Theta}_n\) and \(Y_{(1)}\), which are \(1/\sqrt{n}\) and \(1/\log \xi_0 n\). The red circle marks the neighborhood of \(\theta_0\) with radius \(1/\sqrt{n}\).

Therefore, it is imperative that we compare the convergence rate of the distance between \(\theta_0\) and the boundary with \(1/\sqrt{n}\) in order to get a clearer picture of \(L_n(\theta)\) near the boundary.

**PROPOSITION 3.1.** Suppose \(Y_1, \ldots, Y_n \overset{iid}{\sim} P_{\theta_0}\), and \(\gamma > 0\) is an arbitrary constant.

(A) If \(\xi_0 > 0\), \(Y_{(1)} \xrightarrow{a.s.} \beta_0\) and \(Y_{(n)} \xrightarrow{a.s.} \infty\), as \(n \to \infty\). Moreover,
\[
(\log n)^{(1+\gamma)\xi_0} (Y_{(1)} - \beta_0) \xrightarrow{a.s.} \infty, \quad (\log n)^{(1-\gamma)\xi_0} (Y_{(1)} - \beta_0) \xrightarrow{a.s.} 0,
\]
\[
n^{-2(1+\gamma)\xi_0} Y_{(n)} \xrightarrow{a.s.} 0, \quad \text{and} \quad n^{-2(1-\gamma)\xi_0} Y_{(n)} \xrightarrow{a.s.} \infty.
\]

(B) If \(\xi_0 < 0\), \(Y_{(1)} \xrightarrow{a.s.} -\infty\) and \(Y_{(n)} \xrightarrow{a.s.} \beta_0\), as \(n \to \infty\). Moreover,
\[
(\log n)^{(1+\gamma)\xi_0} Y_{(1)} \xrightarrow{a.s.} 0, (\log n)^{(1-\gamma)\xi_0} Y_{(1)} \xrightarrow{a.s.} -\infty,
\]
\[
n^{-2(1+\gamma)\xi_0} (\beta_0 - Y_{(n)}) \xrightarrow{a.s.} \infty, \quad \text{and} \quad n^{-2(1-\gamma)\xi_0} (\beta_0 - Y_{(n)}) \xrightarrow{a.s.} 0.
\]

**PROOF.** See Appendix B. \(\square\)

When \(\xi_0 > 0\), this theorem demonstrates that the convergence rate of \(Y_{(1)}\) to \(\beta_0\) is roughly \(1/\log \xi_0 n\). The convergence rate of \(\hat{\Theta}_n\) to \(\theta_0\), \(1/\sqrt{n}\), is much faster than the rate of \(Y_{(1)}\) to \(\beta_0\). These two rates are compared schematically in Figure 3. If \(\xi_0 < 0\), the convergence rate of \(Y_{(n)}\) to \(\beta_0\) is \(n^{1/\xi_0}\), which is still slower than \(1/\sqrt{n}\) because of the restriction \(\xi_0 > -1/2\). Thus for a ball neighborhood of \(\hat{\Theta}_n\) to be contained in \(\Omega_n\), its radius can be up to \(1/n^{\gamma}\) for some \(\gamma \in (0, 1/2)\). This property will be of vital importance in the proof of (I) and (II).

4. **Proof of Theorem 1.1.**

4.1. **Step (I) and its proof.** Construct the following compact set
\[
\tilde{K} = \{ \theta \in \Theta : |\tau - \tau_0| \leq r, |\beta - \beta_0| \leq r, |\xi - \xi_0| \leq r \},
\]
where \(r\) is a small constant to be determined by \(\theta_0\) such that the local concavity holds in \(\tilde{K}\). Slicing \(\tilde{K}\) at different levels of \(\xi\) will produce parallelograms; see Figure 4 for illustration. In this section, we will prove that for all large \(n\), the Hessian matrix of \(L_n\) is negative definite in \(\tilde{K} \cap \Omega_n\), and hence \(L_n\) is strictly concave.
the sums in \( L_\theta \) gives strong consistency for a class of limit relations Gamma function. Since this is an iid sequence, the strong law of large numbers immediately \( k, b \).

\[ \xi_0 < 0 \quad \xi_0 > 0 \]

\( \xi_0 < 0 \) (left) and \( \xi_0 > 0 \) (right). In both cases, the set \( \tilde{K} \) sliced at \( \xi = \xi_0 \) is shown in yellow, with \( \Omega_\xi(\xi_0) \) shown in blue. Since \( |r - \tau_0| < r \) and \( |\beta - \beta_0| < r \), the slice is a parallelogram. Similarly, \( \tilde{K} \) sliced at another \( \xi \) in \( (\xi_0 - r, \xi_0 + r) \) will also yield a parallelogram.

4.1.1. Smoothness of Hessian matrix. First we study \( L''_n(\hat{\theta}_n) \), the Hessian at the local MLE \( \hat{\theta}_n \). The log-likelihood \( L_n(\theta) \) in (2) and elements of its Hessian matrix \( L''_n(\theta) \) can all be written as linear combinations of sums of the form

\[ \sum_{i=1}^{n} w_i \frac{-k - \frac{a}{\tau}}{\xi} w_i(\theta), \]

where \( k, b = 0, 1, 2, a = 0, 1 \); see Appendix D for the explicit expressions of the Hessian.

For \( \xi_0 \neq 0 \) and constants \( k \) and \( a \) such that \( k\xi_0 + a + 1 > 0 \), it is straightforward to calculate

\[ E_{\theta_0} \left[ w \frac{-k - \frac{a}{\tau}}{\xi} w(\theta_0) \log^b w(\theta_0) \right] = (-\xi_0)^b \Gamma^{(b)}(k\xi_0 + a + 1), \]

where \( w(\theta_0) = \xi_0(Y - \beta_0)/\tau_0 \) with \( Y \sim P_{\theta_0} \), and \( \Gamma^{(b)} \) is the \( b \)th-order derivative of the Gamma function. Since this is an iid sequence, the strong law of large numbers immediately gives strong consistency for a class of limit relations

\[ \frac{1}{n} \sum_{i=1}^{n} w_i \frac{-k - \frac{a}{\tau}}{\xi} w_i(\theta_0) \log^b w_i(\theta_0) \xrightarrow{a.s.} (-\xi_0)^b \Gamma^{(b)}(k\xi_0 + a + 1). \]

To examine \( L''_n(\hat{\theta}_n), \theta_0 \) in the preceding averages needs to be changed to \( \hat{\theta}_n \). Since \( \hat{\theta}_n \xrightarrow{a.s.} \theta_0 \), the continuity of the sums with respect to \( \theta \) permits a pseudo large law of numbers for the sums in \( L''_n(\hat{\theta}_n) \).

**PROPOSITION 4.1.** Suppose \( Y_1, Y_2, \ldots, i.i.d \sim P_{\theta_0} \) and \( \xi_0 \neq 0 \), and \( \hat{\theta}_n \) is the local MLE of \( L_n(\theta) \) that is strongly consistent. Then for constants \( k \) and \( a \) such that \( k\xi_0 + a + 1 > 0 \),

\[ \frac{1}{n} \sum_{i=1}^{n} w_i \frac{-k - \frac{a}{\tau}}{\xi} w_i(\hat{\theta}_n) \log^b w_i(\hat{\theta}_n) \xrightarrow{a.s.} (-\xi_0)^b \Gamma^{(b)}(k\xi_0 + a + 1), \]

where \( b \) is a non-negative integer.

**PROOF.** The proof of this result depends on Proposition 3.1. For details see Appendix C. \( \square \)

Proposition 4.1 ensures that \( L''_n(\hat{\theta}_n) \) behaves like \( L''_n(\theta_0) \) for large \( n \). The following result states that if we select \( r \) carefully, \( L''_n(\theta) \) can be approximated by \( L''_n(\hat{\theta}_n) \) in the neighborhood \( \tilde{K} \cap \Omega_n \); hence the negative-definiteness of \( L''_n(\theta) \) in this neighborhood.
Therefore, I, n > N previously produces a continuous path of the limit. If we fix the non-negative integer \( b \) of the \( \hat{\theta}_n \) that is strongly consistent. Then we can find a \( \Phi \) tends to zero as \( t \to 0 \), where \( (\Phi(\alpha|n) = (n) \log^b w_i(\hat{\theta}_n) \theta \) is an unique maximum point in \( n \). Namely, \( \exists n > N, L_n(\hat{\theta}_n) < 0 \) in \( K \cap \Omega \). Namely, \( \exists N > 0 \) such that for all \( n > N, L_n''(\theta) < 0 \) in \( K \cap \Omega \). Namely, 

\[
\mathbb{P}\left( \exists N > 0 \text{ such that for all } n > N, L_n''(\theta) < 0 \text{ in } K \cap \Omega \right) = 1.
\]

Equivalently, \( \hat{\theta}_n \) is an unique maximum point in \( K \).

PROOF. Proposition 4.1 ensures that

\[
\frac{1}{n} L_n''(\hat{\theta}_n) \xrightarrow{a.s.} -I(\theta_0),
\]

where \( I(\theta_0) \) is the Fisher information of \( P_{\theta_0} \), and we know \( |I(\theta_0)| > 0 \) for all \( \xi_0 > -1/2 \). Therefore, \( I(\theta_0) \) is positive definite, and there almost surely exists \( N > 0 \) such that for all \( n > N, L_n''(\hat{\theta}_n) < 0 \).
By Proposition 4.2, \( A_0(r) \) only depends on \( \theta_0 \) and \( r \). We now fix \( r \) small enough such that the smallest eigenvalue of \( I - A_0(r) \) is positive. By (9),

\[
L''_n(\theta) - L''(\hat{\theta}_n)(I - A_0(r)) < 0.
\]

Note that the choice of \( r \) only depends on \( \theta_0 \).

4.2. Step (II) and its proof. Step (II) confines the global MLE to a fixed compact set \( K \) which is constructed using the values of \( \theta_0 \). Since \( \hat{\theta}_n = \arg\max_{\theta \in K} L_n(\theta) \), we can deduce the global optimality of \( \hat{\theta}_n \). When \( \xi_0 > 0 \), \( K \) is defined as

\[
K = \{ \theta : \beta_0 - 1 \leq \mu \leq \beta_0 + 1 + M_0, 0 \leq \tau \leq \xi_0 M_0, c_0 \xi_0 \leq \xi \leq C_0 \xi_0 \},
\]

in which

\[
M_0 = \max \left\{ \frac{\tau_0 C_0}{\xi_0} 2 \cdot 5 \cdot 2 \cdot 0 \cdot c_0 + 1, \frac{2^{1/\tau_0} \Gamma(1 / \tau_0)}{\xi_0} \left( 1 - \kappa_0 \right) + \frac{2^{1/\xi_0} - 1}{\xi_0} \right\},
\]

\[
\kappa_0 = \min \left\{ \frac{1}{2}, \frac{1}{2 \xi_0} \right\}, \kappa_1 = \min \{1, c_0 \xi_0 \},
\]

and \( c_0 \in (0, 1) \) is a fixed constant such that when \( 0 < x < c_0 \),

\[
\frac{1}{2} \log \Gamma \left( \frac{1}{x} \right) + \frac{1}{2} \log x + \frac{1}{20x} - \frac{21}{10} > 0,
\]

and \( C_0 > 1 \) is a fixed constant such that when \( x > C_0 \),

\[
\frac{1}{x} - 1 \log \frac{\tau_0}{\xi_0} + \log \Gamma \left( \frac{1}{x} \right) > 0.
\]

The construction of \( K \) when \( \xi_0 < 0 \) is done similarly in the supplementary material. Note \( K \) is only dependent on the value of \( \theta_0 \), and for small \( r \) defined in Section 4.1, \( K \subset K \).

**Proposition 4.5 (Step (II)).** Let \( Y_1, Y_2, \ldots \) \( \sim \) \( P_{\theta_0} \) and \( (\mu_n(\xi), \tau_n(\xi)) \) be the maximizer of \( L_n \) on the cross section \( \Omega_n(\xi) \). Then the global maximum must be in the compact set \( K \) as defined in (10) for large \( n \) - that is,

\[
\mathcal{P} \left( \exists N > 0 \text{ such that for all } n > N, \arg\max_{\theta \in \Theta_n} L_n(\theta) \in K \right) = 1.
\]

**Proof.** Applying Proposition 2.1 and 2.2 from Section 2, we will show in Appendix E that

\[
PL_n(\xi) < PL_n(\hat{\xi}_n) \text{ if } \xi \notin [c_0 \xi_0, C_0 \xi_0],
\]

and

\[
(\mu_n(\xi), \tau_n(\xi), \xi) \in K \text{ if } \xi \in [c_0 \xi_0, C_0 \xi_0],
\]

hold for all large \( n \).

Since \( \hat{\xi}_n \) converges quickly to \( \xi_0 \), we have \( \hat{\xi}_n \in [c_0 \xi_0, C_0 \xi_0] \) for sufficiently large \( n \). Denote \( K_1 = \{ \theta \in \Theta : c_0 \xi_0 \leq \xi \leq C_0 \xi_0 \} \). Clearly, \( K \subset K_1 \) and (11) implies

\[
\arg\max_{\theta \in \Theta_n} L_n(\theta) \in K_1.
\]
When $\xi \in [c_0 \xi_0, C_0 \xi_0]$, (12) encloses the unique maximizer $(\mu_n(\xi), \tau_n(\xi))$ on the cross section $\Omega_n(\xi)$ in $K$. Equivalently, 

$$\arg\max_{\theta \in K} L_n(\theta) \in K.$$ 

Combining (11) and (12), 

$$\arg\max_{\theta \in \Theta_n} L_n(\theta) \in K.$$ 

4.3. Completing the Proof of Theorem 1.1. Proposition 2 in Dombry (2015) ascertained that for all large $n$, the argmax point on the set $K$ defined in Proposition 4.5 is confined in any smaller neighborhood $\tilde{K}$. Although his result was developed within the framework of triangular arrays of block maxima, the proof can be adapted to work on iid GEV samples.

**Lemma 4.6 (Consistency).** Let $K \subset \Theta$ be a compact set that contains $\theta_0$ as an interior point, and $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed random variables with common distribution $P_{\theta_0}$. Then a sequence of estimators $\hat{\theta}_n$ can be found to maximize the log-likelihood $L_n$ over $K$, and for any smaller neighborhood $\tilde{K}$ of $\theta_0$ such that $\tilde{K} \subset K$,

$$P \left( \exists N > 0 \text{ such that for all } n > N, \hat{\theta}_n \in \tilde{K} \right) = 1,$$

and hence $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ as $n \to \infty$.

**Proof.** Bücher and Segers (2017) noted that Proposition 2 in Dombry (2015) is applicable for unitary GEV distributions. Noticing that a GEV distribution is in its own domain of attraction, the block size sequence $m(n)$ is set to be 1 with $a_m = \tau_0$ and $b_m = \mu_0$.

Following the lines of the proof in Dombry (2015), $K$ is limited to be a ball neighborhood of $\theta_0$ with an arbitrarily small radius. It is straightforward to generalize the proof to any small neighborhood of $\theta_0$ such that $\tilde{K} \subset K$. Because the closure of the set $\Delta = K \setminus \tilde{K}$ is still compact, any open cover of $\Delta$ has a finite subcover, whence the remaining proof can be passed through without modification.

Combining Proposition 4.5 and Lemma 4.6, we obtain

$$\arg\max_{\theta \in \Theta} L_n(\theta) \in \tilde{K} \cap \Omega_n,$$

and by the local strict concavity in $\tilde{K} \cap \Omega_n$ ensured by Proposition 4.4,

$$\hat{\theta}_n = \arg\max_{\tilde{K} \cap \Omega_n} L_n(\theta),$$

whence we conclude that $\hat{\theta}_n$ attains the unique and global maximum of $L_n$.

5. Discussion. In this paper, we proved the uniqueness and global optimity of the MLE for the generalized extreme value distribution. This improves on previous results where the existence and asymptotic normality of the local MLE in Dombry (2015) and Bücher and Segers (2017) were established on a predetermined compact set. Our proof uses a two-step strategy. We first prove that the log-likelihood is strictly concave in a fixed small compact neighborhood when the sample size is large, and then enclose the global maximizer in a larger
neighborhood. Finally, the two main results are integrated using Proposition 2 in Dombry (2015) to ensure that the local MLE truly maximizes the log-likelihood globally.

Intermediate results necessary for the the proofs of local strict concavity and boundedness of the global MLE unveiled additional strong and interesting characteristics of the GEV likelihood function that may be of independent interest. We found that the profile likelihood attains a unique maximum at each slice of the support, identified the convergence rate of the support boundary to the local MLE, and demonstrated the uniform consistency of a class of limit relations that are the building blocks of the Hessian matrix. These results enhance our understanding of the GEV distribution, and in future work will become instrumental in establishing asymptotic properties of the GEV posterior distribution that are key in formulating rule-based reference analysis.

In applications, however, observations are never generated exactly from a GEV distribution; rather, they come from a distribution which we typically assume to be in the domain of attraction of a GEV. Dividing the observations into non-overlapping blocks, we make the approximating assumption that the maxima extracted from each block are GEV distributed. Thus, the asymptotic setup of Dombry (2015) should be viewed as the more realistic, and our work offers theoretical foundations for maximum likelihood estimation using the GEV when the block size is large.

Finally, the number of block maxima in any observational record is limited. For future research, it is important to examine the minimum sample size required for the observations to manifest large-sample behavior, as had been done for previous asymptotic results in extreme value statistics. Relatelly, small-sample estimators for the GEV tend to be unstable, so taking advantage of the profile likelihood might provide an effective—and to our knowledge unexplored—approach to estimating the shape parameter. That is, one could first calculate the maximum likelihood on the cross sections of the support at different levels of \( \xi \), and then find the \( \xi \) that maximizes the profile likelihood. Doing so is guaranteed (at least asymptotically) to find the global MLE, and might improve numerical stability in small samples.

Acknowledgments. The authors gratefully acknowledge support from the US National Science Foundation NSF DMS-2001433.

APPENDIX A: PROOFS CONCERNING THE PROFILE LIKELIHOOD

For the proof of Proposition 2.1 and 2.2, we focus on the mathematical form of the log-likelihood \( L_n \) while fixing \( n \) and a sequence of numbers \( Y_1, \ldots, Y_n \). We begin by recalling the Seitz inequality, which generalizes both Cauchy and Chebyshev’s sum inequalities; see Seitz (1936). The following lemma is a special case of the original Seitz inequality.

**Lemma A.1** (Seitz inequality). Let \( x = (x_1, \ldots, x_n) \), \( y = (y_1, \ldots, y_n) \), \( z = (z_1, \ldots, z_n) \) and \( u = (u_1, \ldots, u_n) \) be given sequences of real numbers. If for every pair of indices \( i, j \) \((i < j)\) and for every pair \( r, s \) \((r < s)\)

\[
\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} \times \begin{vmatrix} z_r & z_s \\ u_r & u_s \end{vmatrix} \geq 0,
\]

then the scalar products between the vectors satisfy

\[
(x \cdot z) \times (y \cdot u) \geq (y \cdot z) \times (x \cdot u).
\]

**Lemma A.2.** Suppose \( Y_1, \ldots, Y_n \in \mathbb{R} \) are not all equal to each other, and fix any non-zero \( \xi \) such that \(-1 < \xi < n - 1\). Define

\[
H_n(\beta) = \frac{\sum_{i=1}^{n} [\xi (Y_i - \beta)]^{-1} - 1}{\sum_{i=1}^{n} [\xi (Y_i - \beta)]^{-1} - 1 - \xi} - \frac{\xi + 1}{n},
\]
the domain of which is $\{\beta < Y_{(1)}\}$ if $\xi > 0$, and $\{\beta > Y_{(n)}\}$ if $\xi < 0$. Then $H_n$ is a strictly increasing function of $\beta$, and there exists a unique $\tilde{\beta}$ in the domain such that $H_n(\tilde{\beta}) = 0$, and $\text{sgn}(H_n(\beta)) = \text{sgn}(\beta - \tilde{\beta})$ for any $\beta$ in the domain.

**PROOF.** A. $\xi > 0$. We first show that $H'_n(\beta) > 0$ for $\beta \in (-\infty, Y_{(1)})$, where

$$H'_n(\beta) = \frac{\frac{1}{\xi} \sum_{i=1}^{n} (Y_i - \beta)^{-1} \left[ \sum_{i=1}^{n} (Y_i - \beta)^{-2} \sum_{i=1}^{n} (Y_i - \beta)^{-1} - \left( \sum_{i=1}^{n} (Y_i - \beta)^{-1} - \frac{1}{\xi} \right)^2 \right]}{\sum_{i=1}^{n} (Y_i - \beta)^{-2} \sum_{i=1}^{n} (Y_i - \beta)^{-1} - \sum_{i=1}^{n} (Y_i - \beta)^{-1} \left( \sum_{i=1}^{n} (Y_i - \beta)^{-2} \right)}.$$

By Cauchy inequality,

$$\sum_{i=1}^{n} (Y_i - \beta)^{-2} \sum_{i=1}^{n} (Y_i - \beta)^{-1} > \left( \sum_{i=1}^{n} (Y_i - \beta)^{-1} \right)^2.$$

Note the equality does not hold because $Y_1, \ldots, Y_n$ are not all equal to each other. Therefore, it suffices to prove

$$(A.1) \quad \sum_{i=1}^{n} (Y_i - \beta)^{-2} \sum_{i=1}^{n} (Y_i - \beta)^{-1} \geq \sum_{i=1}^{n} (Y_i - \beta)^{-1} \sum_{i=1}^{n} (Y_i - \beta)^{-2}.$$

Denote $x_i = (Y_{(i)} - \beta)^{-\frac{1}{\xi}}$, $y_i = 1$, $z_i = (Y_{(i)} - \beta)^{-2}$ and $u_i = (Y_{(i)} - \beta)^{-1}$. Since for any pairs of indices $i < j$ and $r < s$,

$$\frac{|x_i x_j|}{|y_i y_j|} \times \frac{|z_r z_s|}{|u_r u_s|} = \frac{(Y_{(r)} - \beta)^{-1} - (Y_{(s)} - \beta)^{-1}}{(Y_{(r)} - \beta)(Y_{(s)} - \beta)} \left[ (Y_{(i)} - \beta)^{-1} - (Y_{(j)} - \beta)^{-1} \right] > 0,$$

we can apply the Seitz inequality to obtain $\text{(A.1)}$, and hence $H'_n(\beta) > 0$ for $\beta \in (-\infty, Y_{(1)})$.

On the other hand, we can easily check the limits of $H_n$ on the bounds of the domain:

$$\lim_{\beta \searrow Y_{(1)}} H_n(\beta) = 1 - \frac{\xi + 1}{n} > 0.$$ 

By the intermediate zero theorem, there exists an unique $\tilde{\beta} \in (-\infty, Y_{(1)})$ that is determined by $\xi$ and the $Y_i$’s such that $H_n(\tilde{\beta}) = 0$, and $H_n(\beta) < 0$ when $\beta < \tilde{\beta}$, or vice versa.

Furthermore, when the fixed $\xi$ diminishes to 0, $\lim_{\beta \searrow -\infty} H_n(\beta)$ is close to 0. This means the root can only be found close to the boundary because $H_n$ is increasing and thus positive almost everywhere on $(-\infty, Y_{(1)})$. Treat $\beta$ as a function of fixed $\xi$, and we conclude

$$\tilde{\beta} \to -\infty, \text{ as } \xi \searrow 0.$$ 

Similarly, when the fixed $\xi$ approaches $n - 1$ from below, $\lim_{\beta \nearrow Y_{(1)}} H_n(\beta)$ is close to 0, and $H_n$ is negative almost everywhere on $(-\infty, Y_{(1)})$. Therefore,

$$\tilde{\beta} \to Y_{(1)}, \text{ as } \xi \nearrow n - 1.$$
Similarly, we examine the derivative:

\[ \lim_{\beta \to Y(n)} H_n(\beta) = -\frac{\xi + 1}{n} < 0, \]

\[ \lim_{\beta \to \infty} H_n(\beta) = \frac{1}{n} - \frac{\xi + 1}{n} = -\frac{\xi}{n} > 0. \]

Therefore, there exists an unique \( \tilde{\beta} \in (Y(n), \infty) \) that is determined by the fixed \( \xi \) and \( Y_i \)'s such that \( H_n(\tilde{\beta}) = 0 \), and \( H_n(\beta) < 0 \) when \( \beta < \tilde{\beta} \), or vice versa.

As we move \( \xi \) to \(-1\) from the right and \( 0 \) from the left, it can be shown accordingly that

\[ \tilde{\beta} \to Y(n), \text{ as } \xi \searrow -1, \text{ and } \tilde{\beta} \to \infty, \text{ as } \xi \nearrow 0. \]

\[ \square \]

A.1. Proof of Proposition 2.1.

PROOF. By (1) and (2), the log-likelihood function on one cross section \( \Omega_n(\xi) \) can be expressed in terms of \( (\tau, \beta) \):

\[ h(\tau, \beta) := L_n(\tau, \beta, \xi) = \frac{n}{\xi} \log \tau - \frac{\xi + 1}{\xi} \sum_{i=1}^{n} \log[\xi(Y_i - \beta)] - \tau^{1/\xi} \sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1/\xi}. \]

Now we profile out \( \tau \) via fixing \( \beta \). Taking the partial derivative with respect to \( \tau \), we obtain

\[ \frac{\partial h}{\partial \tau} = \frac{n}{\xi \tau} - \tau^{1/\xi} \frac{1}{\xi} \sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1/\xi}. \]

The root of the above partial derivative is

\[ \tau(\beta) = \left\{ \frac{1}{n} \sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1/\xi} \right\}^{-\frac{\xi}{\xi+1}}. \]

For both \( \xi > 0 \) and \( \xi < 0 \), it holds that \( \frac{\partial h}{\partial \tau} > 0 \) when \( \tau < \tau(\beta) \), and \( \frac{\partial h}{\partial \tau} < 0 \) when \( \tau > \tau(\beta) \). Fixing \( \beta \), \( h \) as a function of \( \tau \) monotonically increases and then decreases when \( \tau \) grows from \( 0 \) to \( \infty \). Thus, the maximizer of \( L_n \) on \( \Omega_n(\xi) \) must be on the parametric graph \( \{ (\beta, \tau) : \tau = \tau(\beta) \} \). To locate the maximizer, we simply need to find \( \beta \) such that \( h(\beta, \tau(\beta)) \) is maximized.

Simple calculations lead to

\[ h_1(\beta) := h(\tau(\beta), \beta) = -n \log \left\{ \frac{1}{n} \sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1/\xi} \right\} - \frac{\xi + 1}{\xi} \sum_{i=1}^{n} \log[\xi(Y_i - \beta)] - n. \]

Similarly, we examine the derivative:

\[ h_1'(\beta) = \sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1} \left\{ (\xi + 1) - \frac{\xi}{n} \sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1} \sum_{i=1}^{n} [\xi(Y_i - \beta)]^{-1} \right\}. \]

By Lemma A.2, there exists an unique \( \tilde{\beta} \) in the domain such that \( h_1'(\tilde{\beta}) = 0 \). Meanwhile, \( h_1'(\beta) > 0 \) when \( \beta < \tilde{\beta} \) and \( h_1'(\beta) < 0 \) when \( \beta > \tilde{\beta} \). That is, \( h_1(\beta) \) attains the global unique maximum at \( \tilde{\beta} \).
To sum up, there exists a maximizer \( (\tau(\hat{\beta}), \hat{\beta}) \) of \( L_n \) on the cross section \( \Omega_n(\xi) \) that is unique and global with respect to \( \Omega_n(\xi) \). The sufficient and necessary condition for \( (\hat{\tau}, \hat{\beta}) \) to be the maximizer is (5). The proof for maximizing \( L_n \) on \( \Omega_n(0) \) is deferred to the supplementary material.

Since \( (\tau, \mu) \mapsto (\tau, \beta) \) is a one-to-one mapping, it’s clear that \( (\hat{\tau}, \hat{\mu}) = (\hat{\tau}, \hat{\beta} + \hat{\tau}/\xi) \) is the unique maximizer on \( \Omega_n(\xi) \) under the parametrization \( (\tau, \mu) \). We write \( (\hat{\tau}, \hat{\mu}) \) as \( (\tau_n(\xi), \mu_n(\xi)) \) to highlight that \( (\hat{\tau}, \hat{\mu}) \) depends on \( Y_1, \ldots, Y_n \) and \( \xi \).

A.2. Proof of Proposition 2.2.

PROOF. For notational simplicity, we will denote \( \hat{\delta}_i = \xi_n(Y_i - \hat{\beta}_n) \) and \( \delta_i := \delta_i(\xi, n) = \xi(Y_i - \beta_n(\xi)) \), where \( \beta_n(\xi) = \mu_n(\xi) - \tau_n(\xi)/\xi \). We calculate

\[
\left\{ \frac{n}{\sum_{i=1}^{n} \delta_i^{-1/\xi}} \right\}' = \left\{ \frac{\sum_{i=1}^{n} [\xi(Y_i - \beta_n(\xi))]^{-1/\xi}}{\sum_{i=1}^{n} \delta_i^{-1/\xi}} \right\}'
\]

\[
= \frac{1}{\xi^2} \sum_{i=1}^{n} \delta_i^{-1/\xi} \log \delta_i + \beta_n'(\xi) \sum_{i=1}^{n} \delta_i^{-1-1/\xi} - \frac{1}{\xi^2} \sum_{i=1}^{n} \delta_i^{-1/\xi},
\]

and by (6),

\[
PL_n' = \frac{-n \left\{ \sum_{i=1}^{n} \delta_i^{-1/\xi} \right\}'}{\sum_{i=1}^{n} \delta_i^{-1/\xi}} + \sum_{i=1}^{n} \log \delta_i + \frac{\xi}{\xi} \sum_{i=1}^{n} \{ \log \delta_i \}'
\]

\[
= -\frac{n}{\xi^2} \sum_{i=1}^{n} \delta_i^{-1/\xi} \log \delta_i + \frac{\xi}{\xi^2} \sum_{i=1}^{n} \log \delta_i
\]

\[
+ \beta_n'(\xi) \left\{ (\xi + 1) \sum_{i=1}^{n} \delta_i^{\xi} - \frac{n \sum_{i=1}^{n} \delta_i^{-1-1/\xi}}{\sum_{i=1}^{n} \delta_i^{-1/\xi}} \right\},
\]

in which the term in the curly brackets is zero for \( \xi \neq 0 \) due to the second equation of (5).

Next we examine the limits of \( PL_n' \) as \( \xi \searrow n - 1, \xi \to 0 \) and \( \xi \nearrow -1 \).

A. \( \xi \not\to n - 1 \). From Lemma A.2, we know that

\[
\beta_n(\xi) \not\to Y_{(1)}, \quad \text{as } \xi \not\to n - 1,
\]

and thus

\[
\sum_{i=2}^{n} (Y_i - \beta_n(\xi))^{-1/\xi} \log(Y_i - \beta_n(\xi)) \to \sum_{i=2}^{n} (Y_i - Y_{(1)})^{-1/(n-1)} \log(Y_i - Y_{(1)}) =: C_1,
\]

\[
\sum_{i=2}^{n} (Y_i - \beta_n(\xi))^{-1/\xi} \to \sum_{i=2}^{n} (Y_i - Y_{(1)})^{-1/(n-1)} =: C_2,
\]

\[
\sum_{i=2}^{n} \log(Y_i - \beta_n(\xi)) \to \sum_{i=2}^{n} \log(Y_i - Y_{(1)}) =: C_3.
\]
As a result,
\[
\lim_{\xi \to n^{-1}} PL'_n(\xi) = - \lim_{\xi \to n^{-1}} \frac{1}{\xi^2} \left[ \frac{nC_1 + n(Y_{(1)} - \beta_n(\xi))^{-1/\xi} \log(Y_{(1)} - \beta_n(\xi))}{C_2 + (Y_{(1)} - \beta_n(\xi))^{-1/\xi}} - \log(Y_{(1)} - \beta_n(\xi)) \right] + \frac{C_3}{(n - 1)^2} \frac{n}{n - 1} = -\log(\xi) - \frac{n}{n - 1} = \infty.
\]

B. $\xi \to 0$. In this case, we need to be more cautious about $\beta'_n(\xi)$ and the term in the curly brackets in (A.2). It holds that as $\xi \to 0$,
\[
PL'_n(\xi) = \frac{n\mu_n(0) - \sum_{i=1}^n [Y_i - \mu_n(0) + \tau'_n(0)]}{\tau_n(0)} + \frac{\sum_{i=1}^n [Y_i - \mu_n(0) + \tau'_n(0)]^2 - n\tau'_n(0)^2}{2\tau_n(0)^2} + o(1),
\]
the proof of which is deferred to the supplementary material.

C. $\xi \to -1$. Denote $\delta(i) = \xi(Y_{(i)} - \beta_n(\xi))$. From Lemma A.2, we know that
\[\beta_n(\xi) \to Y_{(n)}, \text{ as } \xi \to -1.\]
Thus, $\delta(n) = \xi(Y_{(n)} - \beta_n(\xi)) \to 0$ and
\[
\sum_{i=1}^{n-1} \delta_{(i)}^{-1/\xi} \log \delta_{(i)} \to \sum_{i=1}^{n-1} (Y_{(n)} - Y_{(i)}) \log (Y_{(n)} - Y_{(i)}) =: C'_1,
\]
\[
\sum_{i=1}^{n-1} \delta_{(i)}^{-1/\xi} \to \sum_{i=1}^{n-1} (Y_{(n)} - Y_{(i)}) =: C'_2,
\]
\[
\sum_{i=1}^{n-1} \log \delta_{(i)} \to \sum_{i=1}^{n-1} \log (Y_{(n)} - Y_{(i)}) =: C'_3.
\]
As a result,
\[
\lim_{\xi \to n^{-1}} PL'_n(\xi) = n + C'_3 - \lim_{\xi \to n^{-1}} \left\{ \frac{nC'_1 + n\delta_{(n)}^{-1/\xi} \log \delta_{(n)}}{\xi^2 C'_2 + \xi^2 \delta_{(n)}^{-1/\xi}} - \frac{\log \delta_{(n)}}{\xi^2} \right\}
\]
\[
= n + C'_3 - \frac{C'_1 C'_3}{C'_2} + \lim_{\xi \to n^{-1}} \left\{ \frac{\log \delta_{(n)}}{\xi^2} \right\} = -\infty.
\]

Appendix B: Convergence Rate of the Support Boundary

Proposition B.1. Suppose $Y_1, \ldots, Y_n$ are i.i.d. samples from some parametric model $\mathcal{M} = \{p(y|\theta) : y \in Y, \theta \in \Theta\}$, where $Y$ may either depend on $\theta$ or not. If $\beta_n = \beta(Y_1, \ldots, Y_n)$ is a strongly consistent estimator of $\beta = \beta(\theta)$ that satisfies $\sqrt{n} (\hat{\beta}_n - \beta) = V + o_p(1)$ and $\hat{\beta}_n \xrightarrow{a.s.} \beta$, where $V$ is a continuous random variable. Then
\[n^\gamma (\hat{\beta}_n - \beta) \xrightarrow{a.s.} 0, \forall \gamma \in (0, 1/2).\]
Define \( n \) (1/2 - \( \gamma \)) > 1. Since \( \sqrt{n}(\hat{\beta}_n - \beta)/V - 1 = o_p(1) \), we know that \( E \left( |\sqrt{k}(\hat{\beta}_k - \beta)/V - 1|^{\gamma} \right) \to 0 \) as \( k \to \infty \), that is, there exists \( N_q \), such that \( E \left( |\sqrt{k}(\hat{\beta}_k - \beta)/V - 1|^{\gamma} \right) < 1 \) for all \( k > N_q \).

We prove \( \beta \) for any \( \epsilon > 0 \), which completes the proof of (B.1).

First, we prove

\[
P \left( \{ \omega \in \mathcal{X} : \exists N \text{ such that, } \forall n \geq N, n^\gamma|\hat{\beta}_n(\omega) - \beta| < |V(\omega)| \} \right) = 1.
\]

Define

\[
A_n = \bigcap_{k=n}^{\infty} \left\{ \omega \in \mathcal{X} : k^\gamma|\hat{\beta}_k(\omega) - \beta| < |V(\omega)| \right\}.
\]

Since \( A_n \subseteq A_{n+1} \), it is clear that (B.1) is equivalent to \( P(A_n) \to 1 \) as \( n \to \infty \). However,

\[
P \left( \{ \omega \in \mathcal{X} : k^\gamma|\hat{\beta}_k(\omega) - \beta| \geq |V(\omega)| \} \right) = P(\{ k^{1/2}|\hat{\beta}_k - \beta| - |V| \geq (k^{1/2-\gamma} - 1)|V| \}) \leq P \left( \left| \frac{\sqrt{k}(\hat{\beta}_k - \beta)}{V} - 1 \right| \geq k^{1/2-\gamma} - 1 \right) + P(V = 0) \leq \frac{1}{(k^{1/2-\gamma} - 1)^q}, \forall k > N_q,
\]

in which we utilized Markov’s inequality and the fact that \( V \) is continuous.

Note \( \sum_{k=2}^{\infty} \frac{1}{(k^{1/2-\gamma} - 1)^q} \) is convergent for \( q(1/2 - \gamma) > 1 \). Countable subadditivity implies

\[
P(A_n^c) = P \left( \bigcup_{k=n}^{\infty} \left\{ \omega \in \mathcal{X} : k^\gamma|\hat{\beta}_k(\omega) - \beta| \geq |V(\omega)| \right\} \right) \leq \sum_{k=n}^{\infty} \frac{1}{(k^{1/2-\gamma} - 1)^q} \to 0,
\]

which completes the proof of (B.1).

Since \( \gamma < 1/2 \), there exists \( \eta > 0 \) such that \( 1 - 2\gamma - \eta > 0 \) and \( \frac{1}{2\gamma+\eta} < 1/2 \). Define

\[
\gamma = \left\{ \omega \in \mathcal{X} : \lim_{n \to \infty} \hat{\beta}_n(\omega) = \beta \right\}, \quad C = \left\{ \omega \in \mathcal{X} : V(\omega) \neq 0 \right\},
\]

\[
D = \left\{ \omega \in \mathcal{X} : \exists N \text{ such that, } \forall n \geq N, n^{\gamma/(2\gamma+\eta)}|\hat{\beta}_n(\omega) - \beta| < |V(\omega)| \right\}.
\]

By definition of strong consistency and the preceding discussion, \( P(B) = P(C) = P(D) = 1 \), which implies \( P(B \cap C \cap D) = 1 - P(B^c \cup C^c \cup D^c) \geq 1 - P(B^c) - P(C^c) - P(D^c) = 1 \), i.e. \( P(B \cap C \cap D) = 1 \). Focusing on a particular element \( \omega \in B \cap C \cap D \), we have for \( \epsilon > 0 \), there exists \( N_1 \) such that \( |\hat{\beta}_n(\omega) - \beta| < \left( \epsilon/|V(\omega)|^{2\gamma+\eta} \right)^{1/(1-2\gamma-\eta)} \) for any \( n > N_1 \); there also exists \( N_2 \) such that \( n^{\gamma/(2\gamma+\eta)}|\hat{\beta}_n(\omega) - \beta| < |V(\omega)| \) for any \( n > N_2 \).

Therefore, for any \( n > \max\{ N_1, N_2 \} \),

\[
n^\gamma|\hat{\beta}_n(\omega) - \beta| = \left[ n^{\gamma/(2\gamma+\eta)}|\hat{\beta}_n(\omega) - \beta| \right]^{2\gamma+\eta} \cdot |\hat{\beta}_n(\omega) - \beta|^{1-2\gamma-\eta} \leq |V(\omega)|^{2\gamma+\eta} \cdot \frac{\epsilon}{|V(\omega)|^{2\gamma+\eta}} = \epsilon.
\]

Equivalently, if \( \omega \in B \cap C \cap D \), then \( n^\gamma(\hat{\beta}_n(\omega) - \beta) \to 0 \). Since \( P(B \cap C \cap D) = 1, n^\gamma(\hat{\beta}_n - \beta) \xrightarrow{a.s.} 0. \)
B.1. Proof of Proposition 3.1.

PROOF. Here we only provide the proof of (A), the positive shape $\xi_0$. The proof of the case $\xi_0 < 0$ is analogous.

(i) Prove $(\log n)^{(1+\gamma)\xi_0}(Y_1 - \beta_0) \xrightarrow{a.s.} \infty$. Note that $Y_i > \beta_0$, $i = 1, \ldots, n$, and $Y_1 > \beta_0$. Fix $M > 0$, and define

$$A_n = \left\{ \omega : \forall k \geq n, (\log k)^{(1+\gamma)\xi_0} \left( \min_{1 \leq i \leq k} Y_i(\omega) - \beta_0 \right) \geq M \right\}$$

$$= \left\{ \omega : Y_k \geq \beta_0 + \frac{M}{(\log n)^{(1+\gamma)\xi_0}}, 1 \leq k \leq n, \text{ and } Y_k \geq \beta_0 + \frac{M}{(\log k)^{(1+\gamma)\xi_0}}, k \geq n + 1 \right\}.$$  

Then $(\log n)^{(1+\gamma)\xi_0}(Y_1 - \beta_0) \xrightarrow{a.s.} \infty$ if and only if $P(A_n) \to 1$ for any $M > 0$.

We can easily calculate

$$P(A_n) = \left[ P \left( Y_1 \geq \beta_0 + \frac{M}{(\log n)^{(1+\gamma)\xi_0}} \right) \right]^n \prod_{k=n+1}^{\infty} P \left( Y_k \geq \beta_0 + \frac{M}{(\log k)^{(1+\gamma)\xi_0}} \right)$$

$$= \left[ 1 - \exp \left( -M_1 (\log n)^{1+\gamma} \right) \right]^n \prod_{k=n+1}^{\infty} \left[ 1 - \exp \left( -M_1 (\log k)^{1+\gamma} \right) \right],$$

where constant $M_1 = (M\xi_0/\tau_0)^{-1/\xi_0} > 0$. By L’Hospital rule,

$$(B.2) \lim_{x \to \infty} x \log \left\{ 1 - \exp \left( -M_1 (\log x)^{1+\gamma} \right) \right\} = -M_1 (1 + \gamma) \lim_{x \to \infty} \frac{x \log x \cdot \exp \left( -M_1 (\log x)^{1+\gamma} \right)}{1 - \exp \left( -M_1 (\log x)^{1+\gamma} \right)} = 0.$$

Thus, $\left[ 1 - \exp \left( -M_1 (\log n)^{1+\gamma} \right) \right]^n \to \exp(0) = 1$ as $n \to \infty$.

On the other hand, consider $\sum_{k=1}^{\infty} \log \left\{ 1 - \exp \left( -M_1 (\log k)^{1+\gamma} \right) \right\}$. Apply L’Hospital rule again, and we know

$$\lim_{x \to \infty} x^2 \log \left\{ 1 - \exp \left( -M_1 (\log x)^{1+\gamma} \right) \right\} = -M_1 (1 + \gamma) \lim_{x \to \infty} \frac{x^2 \log x \cdot \exp \left( -M_1 (\log x)^{1+\gamma} \right)}{1 - \exp \left( -M_1 (\log x)^{1+\gamma} \right)} = 0.$$

Thus, $k^2 \log \left\{ 1 - \exp \left( -M_1 (\log k)^{1+\gamma} \right) \right\} \to 0$ as $k \to \infty$. By limit comparison test, we know that $\sum_{k=1}^{\infty} \log \left\{ 1 - \exp \left( -M_1 (\log k)^{1+\gamma} \right) \right\} < \infty$. Hence, as $n \to \infty$,

$$\prod_{k=n+1}^{\infty} \left[ 1 - \exp \left( -M_1 (\log k)^{1+\gamma} \right) \right] = \exp \left\{ \sum_{k=n+1}^{\infty} \log \left\{ 1 - \exp \left( -M_1 (\log k)^{1+\gamma} \right) \right\} \right\} \to 1,$$

which completes proving $P(A_n) \to 1$ for any $M > 0$.

(ii) Prove $(\log n)^{(1-\gamma)\xi_0}(Y_1 - \beta_0) \xrightarrow{a.s.} 0$. If $\gamma \geq 1$, the second strong convergence holds because $Y_1 \xrightarrow{a.s.} \beta_0$. The more interesting case is when $1 - \gamma > 0$.

We fix $\epsilon > 0$, and define

$$B_n = \left\{ \omega : \exists k \geq n, \log^{(1-\gamma)\xi_0} k \cdot \left( \min_{1 \leq i \leq k} Y_i(\omega) - \beta_0 \right) > \epsilon \right\} = \bigcup_{k=n}^{\infty} C_k,$$

where $C_k = \left\{ \omega : \log^{(1-\gamma)\xi_0} k \cdot \left( \min_{1 \leq i \leq k} Y_i(\omega) - \beta_0 \right) > \epsilon \right\}$, and

$$P(C_k) = \prod_{i=1}^{k} P \left( Y_i \geq \beta_0 + \frac{\epsilon}{\log^{(1-\gamma)\xi_0} (k)} \right) = [1 - \exp \left( -M_2 \log^{1-\gamma} (k) \right)]^k,$$
where $M_2 = (\xi_0/\tau_0)^{-1/\xi_0} > 0$.

Now we want to prove $k^2 P(C_k) \to 0$ as $k \to \infty$. Application of L'Hospital rule similar to that leading to (B.2) yields

$$
\lim_{x \to \infty} \frac{x \log[1 - \exp(-M_2 x)]}{\log x} = \lim_{x \to \infty} \frac{M_2 (1 - \gamma) x \exp(-M_2 x)}{1 - \exp(-M_2 x)} \cdot x \log^{1-\gamma}(x)
$$

$$
= -M_2 (1 - \gamma) \lim_{x \to \infty} \frac{x \log(x) - 1}{x^{1-\gamma}(x)} \cdot x \log^{1-\gamma}(x)
$$

which in turn results in

$$
\log[x^2(1 - \exp(-M_2 x))^\gamma] = \log x \cdot \left\{ \frac{x \log[1 - \exp(-M_2 x)]}{\log x} + 2 \right\} \to -\infty, \text{ as } x \to \infty.
$$

Therefore, we have $P(C_k) = o(1/k^2)$.

By the limit comparison test, the series $\sum_{k=1}^\infty P(C_k)$ is convergent, and its tail satisfies

$$
\forall \epsilon > 0, \quad P(B_n) = P\left( \bigcup_{k=n}^\infty C_k \right) \leq \sum_{k=n}^\infty P(C_k) \to 0, \text{ as } n \to \infty.
$$

This proves $(\log n)^{(1-\gamma)\xi_0} (\min_n Y_i - \beta_0) \overset{a.s.}{\to} 0, \forall \gamma \in (0, 1)$.

(iii) **Proof** $n^{-(1+\gamma)\xi_0} Y_{(n)} \overset{a.s.}{\to} 0$. Similarly to (i), we fix $\epsilon > 0$, and define

$$
D_n = \left\{ \omega : \forall k \geq n, k^{-(1+\gamma)\xi_0} \left( \max_{1 \leq i \leq k} Y_i(\omega) - \beta_0 \right) \leq \epsilon \right\}
$$

$$
= \left\{ \omega : Y_k \leq \beta_0 + \epsilon n^{(1+\gamma)\xi_0}, 1 \leq k \leq n, \text{ and } Y_k \leq \beta_0 + \epsilon k^{(1+\gamma)\xi_0}, k \geq n + 1 \right\}.
$$

To prove $n^{-(1+\gamma)\xi_0} Y_{(n)} \overset{a.s.}{\to} 0$, it suffices to show $n^{-(1+\gamma)\xi_0} (Y_{(n)} - \beta_0) \overset{a.s.}{\to} 0$, which is true if and only if $P(D_n) \to 1$ for any $\epsilon > 0$.

The independence of $Y_k, k \geq 1$ gives

$$
P(D_n) = \exp\left( -\frac{M_2}{n^{1+\gamma}} \right) \exp\left( -M_2 \sum_{k=n+1}^\infty \frac{1}{k^{1+\gamma}} \right),
$$

where $M_2 = (\xi_0/\tau_0)^{-1/\xi_0} > 0$. Since $1 + \gamma > 1$, we know by limit comparison test that $\sum_{k=n+1}^\infty 1/k^{1+\gamma} \to 0$, whence $P(D_n) \to 1$ as $n \to \infty$.

(iv) **Proof** $n^{-(1-\gamma)\xi_0} Y_{(n)} \overset{a.s.}{\to} \infty$. Similar to (ii), we only consider $0 < \gamma < 1$, and for $M > 0$ define

$$
E_n = \left\{ \omega : \exists k \geq n, k^{-(1-\gamma)\xi_0} \left( \max_{1 \leq i \leq k} Y_i(\omega) - \beta_0 \right) < M \right\} = \bigcup_{k=n}^\infty F_k,
$$

where $F_k = \{ \omega : k^{-(1-\gamma)\xi_0} (\max_i Y_i(\omega) - \beta_0) < M \}$, and

$$
P(F_k) = \prod_{i=1}^k P\left( Y_i \leq \beta_0 + Mk^{(1-\gamma)\xi_0} \right) = \exp(-M_k \gamma),
$$

where $M_1 = (M\xi_0/\tau_0)^{-1/\xi_0} > 0$. 

Clearly the series $\sum_{k=1}^{\infty} P(F_k)$ is convergent, and its tail satisfies
\[
\forall \epsilon > 0, \quad P(E_n) = P\left( \bigcup_{k=n}^{\infty} F_k \right) \leq \sum_{k=n}^{\infty} P(F_k) \to 0, \text{ as } n \to \infty.
\]
This proves $n^{-(1-\gamma)\xi_0} (Y(n) - \beta_0) \overset{a.s.}{\to} \infty$, which implies $n^{-(1-\gamma)\xi_0} Y(n) \overset{a.s.}{\to} \infty$.

\begin{corollary}
Suppose $Y_1, \ldots, Y_n \overset{iid}{\sim} P_{\theta_0}$. If $\xi_0 > 0$, for any $\gamma \in [0, 1/2)$,
\[
n^\gamma (\hat{\beta}_n - \beta_0) / Y(1) \overset{a.s.}{\to} 0.
\]
If $-1/2 < \xi_0 < 0$, for any $\gamma \in [0, \xi_0 + 1/2)$,
\[
n^\gamma (\hat{\beta}_n - \beta_0) / \beta_0 - Y(n) \overset{a.s.}{\to} 0.
\]
\end{corollary}

\begin{proof}
When $\xi_0 > 0$, choose $\eta > 0$ such that $\gamma + \eta < 1/2$. By Proposition B.1, $n^{\gamma + \eta}(\hat{\beta}_n - \beta_0) \overset{a.s.}{\to} 0$, and by Proposition 3.1,
\[
n^\gamma (\hat{\beta}_n - \beta_0) / Y(1) = n^{\gamma + \eta}(\hat{\beta}_n - \beta_0) / n^\eta(Y(1) - \beta_0) \overset{a.s.}{\to} 0.
\]
When $-1/2 < \xi_0 < 0$, since $-\xi_0 < 1/2 - \gamma$, choose $\eta > 0$ such that $-\xi_0 < \eta < 1/2 - \gamma$. By Proposition 3.1, $n^\eta(\beta_0 - Y(n)) \overset{a.s.}{\to} \infty$, and by Proposition B.1,
\[
n^\gamma (\hat{\beta}_n - \beta_0) / \beta_0 - Y(n) = n^{\gamma + \eta}(\hat{\beta}_n - \beta_0) / n^\eta(\beta_0 - Y(n)) \overset{a.s.}{\to} 0.
\]
\end{proof}

\section*{Appendix C: Proofs of Pseudo-Law of Large Numbers}

For the proof of Proposition 4.1, we begin by proving a few useful results.

\begin{lemma}
Suppose $\xi_0 \neq 0$, and $Y \sim P_{\theta_0}$. Denote $w(\theta_0) = \xi_0(Y - \beta_0)/\tau \theta_0$. Then for any constant $\alpha$ such that $\alpha \xi_0 + 1 > 0$ and a positive integer $b$, we have
\[
E_{\theta_0} \left[ w^{-\alpha}(\theta_0) | \log w(\theta_0) |^b \right] < |\xi_0|^b \Gamma(b) (\alpha \xi_0 + 1) + \frac{3|\xi_0|^b}{(\alpha \xi_0 + 1)^{b+1}} \Gamma(b + 1).
\]
\end{lemma}

\begin{proof}
It is easy to calculate
\[
E_{\theta_0} \left[ w^{-\alpha}(\theta_0) | \log w(\theta_0) |^b \right] = |\xi_0|^b \int_0^{\infty} s^{\alpha \xi_0} \exp(-s) | \log s |^b ds
\]
\[
= |\xi_0|^b \int_0^{\infty} s^{\alpha \xi_0} \exp(-s) \log^b s ds + \left(-1\right)^b - 1 |\xi_0|^b \int_0^{1} s^{\alpha \xi_0} \exp(-s) \log^b s ds
\]
\[
= |\xi_0|^b \Gamma(b) (\alpha \xi_0 + 1) + \left(-1\right)^b - 1 |\xi_0|^b \int_0^{1} s^{\alpha \xi_0} \exp(-s) \log^b s ds.
\]
\end{proof}
Furthermore,
\[
\left| \left( -1 \right)^b - 1 \right| \xi_0^b \int_0^1 s^\alpha \xi_0 \exp\left( -s \right) \log^b s ds \leq 2|\xi_0|^b \int_0^1 s^\alpha \xi_0 \exp\left( -s \right) \left( -\log s \right)^b ds
\]
\[
< 3|\xi_0|^b \int_0^1 s^\alpha \xi_0 \left( -\log s \right)^b ds = \frac{3|\xi_0|^b}{(\alpha \xi_0 + 1)^{b+1}} \Gamma(b+1),
\]
which proves (C.1).

**Lemma C.2.** If \( \hat{\theta}_n \in \Omega_n \) solves the likelihood equations of \( L_n(\theta), \forall L_n(\theta) = 0 \), then
\[
\sum_{i=1}^n (1 + \xi_n) w_i^{-1} (\hat{\theta}_n) = \sum_{i=1}^n w_i^{-1/\xi_n} (\hat{\theta}_n),
\]
\[
\sum_{i=1}^n w_i^{-1/\xi_n} (\hat{\theta}_n) = n,
\]
\[
\sum_{i=1}^n \log w_i (\hat{\theta}_n) - n \xi_n = \sum_{i=1}^n w_i^{-1/\xi_n} (\hat{\theta}_n) \log w_i (\hat{\theta}_n).
\]

**Remark 7.** This lemma is an immediate result of the definition of local MLE. It dovetails with Proposition 4.1, which says that \( \frac{1}{n} \sum_{i=1}^n w_i^{-1/\xi_n} (\hat{\theta}_n) \rightarrow \hat{\theta}_n \). However, for the power of \( 1/\xi_n \), \( \frac{1}{n} \sum_{i=1}^n w_i^{-1/\xi_n} (\hat{\theta}_n) \equiv 1 \). Similarly for the other two equations, the asymptotic relations turned out to be always true.

To prove Proposition 4.1, let us first only change the power with functions of the local MLE before we replace the \( \theta_0 \) entirely with \( \hat{\theta}_n \).

**Lemma C.3.** Suppose \( Y_1, \ldots, Y_n \overset{iid}{\sim} P_{\theta_0} \), where \( \xi_0 \neq 0 \), and \( \hat{\theta}_n \) is the local MLE of \( L_n(\theta) \) that is strongly consistent. Then for constants \( k \) and \( a \) such that \( k \xi_0 + a + 1 > 0 \),
\[
\frac{1}{n} \sum_{i=1}^n w_i^{-k-\xi_n} (\theta_0) \log^b w_i (\theta_0) \rightarrow \frac{a}{\xi_0} \Gamma(b)(k \xi_0 + a + 1),
\]
where \( b \) is a non-negative integer.

**Proof.** Firstly, we prove that
\[
1 + x \log t \leq t^x \leq 1 + x \log t + (t^\eta + t^{-\eta}) (\log t)^2 x^2
\]
for \( t > 0 \), and \(|x| < \eta \), where \( \eta > 0 \). The first inequality holds immediately on account of \( 1 + x \leq e^x, \forall x \in \mathbb{R} \). For the second inequality, we denote \( f(x) = t^x - 1 - x \log t - (t^\eta + t^{-\eta}) (\log t)^2 x^2 \). Since \(|x| < \eta \), it’s obvious that \( f''(x) = t^x (\log t)^2 - 2(t^\eta + t^{-\eta}) (\log t)^2 < 0 \), and \( f \) is a strictly concave function. The maximum of \( f \) in \((-\eta, \eta)\) is \( x = 0 \) since \( f'(0) = 0 \). Therefore, \( f(x) \leq f(0) = 0 \).

Secondly, we examine the difference between \( w_i^{-k-\xi_n} (\theta_0) \) and \( w_i^{-k-\xi_n} (\theta_0) \).

Fix \( \eta > 0 \) such that \( |\eta \xi_0| < k \xi_0 + a + 1 \). Since \( \frac{1}{\xi_n} \rightarrow \frac{1}{\xi_0} \), there almost surely exists \( N \) such that
\[
\left| \frac{a}{\xi_n} - \frac{a}{\xi_0} \right| < \eta \text{ for all } n > N. \text{ Apply (C.3) to get}
\]
\[
\left| w_i^{-\xi_n} (\theta_0) - 1 - \left( \frac{a}{\xi_0} - \frac{a}{\xi_n} \right) \log w_i (\theta_0) \right| \leq \left| w_i^{-\eta} (\theta_0) + w_i^{-\eta} (\theta_0) \right| \log^2 w_i (\theta_0) \left( \frac{a}{\xi_0} - \frac{a}{\xi_n} \right)^2,
\]
in which \( w_i(\theta_0) > 0 \) for \( i = 1, \ldots, n \). Multiplying both sides by \( w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log w_i(\theta_0) \),
\[
\left| w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log^b w_i(\theta_0) - w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log^b w_i(\theta_0) - \left( \frac{a}{\xi_0} - \frac{a}{\xi_n} \right) w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log^{b+1} w_i(\theta_0) \right|
\]
\[
\leq \left| w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) + w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \right| \log w_i(\theta_0) |^{b+2} \left( \frac{a}{\xi_0} - \frac{a}{\xi_n} \right)^2.
\]

Summing over \( i \), we obtain
\[
\frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log^b w_i(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log^b w_i(\theta_0) + \frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log^{b+1} w_i(\theta_0) + R(\hat{\theta}_n),
\]
where
\[
|R(\hat{\theta}_n)| \leq \left( \frac{a}{\xi_0} - \frac{a}{\xi_n} \right)^2 \left[ \frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log w_i(\theta_0) |^{b+2} + \frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log w_i(\theta_0) |^{b+2} \right].
\]

Lemma C.1 ensures that \( E_{\theta_0} \{ w(\theta_0)^{-\alpha} \log w(\theta_0)^b \} < \infty \) for any non-negative integer \( b \)
and \( \alpha \) such that \( \alpha \xi_0 + \xi_0 + 1 > 0 \). Since \( |\eta\xi_0| < k \xi_0 + a + 1 \), we can assure \( (k + \eta)\xi_0 + a + 1 > 0 \)
and \( (k + \eta)\xi_0 + a + 1 > 0 \). By law of large numbers, \( \frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log w_i(\theta_0) |^{b+2} \)
and \( \frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log w_i(\theta_0) |^{b+2} \) are bounded almost surely. Meanwhile, \( \left( \frac{a}{\xi_0} - \frac{a}{\xi_n} \right)^2 \rightarrow 0 \),
and hence \( R(\hat{\theta}_n) \rightarrow 0 \).

Also, \( \left( \frac{a}{\xi_0} - \frac{a}{\xi_n} \right) \cdot \frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log^b w_i(\theta_0) \rightarrow 0 \), and
\[
\frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log^b w_i(\theta_0) \rightarrow (\xi_0)^b \Gamma(b) (k \xi_0 + a + 1).
\]

Consequently we can conclude from (C.4) that (C.2) is true.

\[ \square \]

**C.1. Proof of Proposition 4.1.** Next we replace the \( \theta_0 \) with \( \hat{\theta}_n \) in \( w_i(\theta_0) \) to prove Proposition 4.1.

**Proof. (A) Case \( \xi_0 > 0 \).** Firstly, we notice that for \( \alpha < 0 \) and \( |x - 1| < \eta \),
\[
\log^2 x \leq \frac{2}{(1 - \eta)^2} (x - 1)^2,
\]
and
\[
1 + t(x - 1) \leq x^t \leq 1 + t(x - 1) + \frac{t(t - 1)}{2} (1 - \eta)^{t-2} (x - 1)^2,
\]
where \( \eta \in (0, 1) \) is a small fixed number. Concavity argument similar to that yielding (C.3) will prove (C.5) and (C.6).

Secondly, we prove
\[
\frac{1}{n} w_i^{-k - \frac{\alpha}{\xi_0}}(\theta_0) \log^b w_i(\hat{\theta}_n) \rightarrow (\xi_0)^b \Gamma(b)(k \xi_0 + a + 1)
\]
for \( b \geq 1 \) (the case \( b = 0 \) is covered by Lemma C.3).

Since \( \frac{1}{n} \xrightarrow{a.s.} 0 \) and by Corollary B.1.1 \( \frac{\tilde{\beta}_n - \hat{\beta}_n}{\hat{\beta}_n - \beta_0} \xrightarrow{a.s.} 0 \), there almost surely exists \( N \) such that for all \( n > N \),

\[
1/\hat{\xi}_n > 0, \quad \text{and } \frac{\xi_0(\hat{\beta}_0 - \tilde{\beta}_n)}{\tau_0w_i(\theta_0)} = \frac{\beta_0 - \hat{\beta}_n}{Y_i - \beta_0} \leq \frac{\beta_0 - \tilde{\beta}_n}{Y(1) - \beta_0} < \eta.
\]

We expand the terms using multinomial theorem

\[
\log^b w_i(\tilde{\theta}_n) = \left[ \log \left( \frac{\xi_n \tau_0}{\tau_n \xi_0} \right) + \log w_i(\theta_0) + \log \left( 1 + \frac{\xi_0(\hat{\beta}_0 - \tilde{\beta}_n)}{\tau_0w_i(\theta_0)} \right) \right]^b \leq \log^b \left( \frac{\xi_n \tau_0}{\tau_n \xi_0} \right) + \log^b w_i(\theta_0) + \log^b \left( 1 + \frac{\xi_0(\hat{\beta}_0 - \tilde{\beta}_n)}{\tau_0w_i(\theta_0)} \right).
\]

Thus,

\[
\left| \log^b w_i(\tilde{\theta}_n) - \log^b \left( \frac{\xi_n \tau_0}{\tau_n \xi_0} \right) - \log^b w_i(\theta_0) \right| \leq \sum_{b_1 + b_2 + b_3 = b} \left( \frac{b}{b_1, b_2, b_3} \right) \log \left( \frac{\xi_n \tau_0}{\tau_n \xi_0} \right)^{b_1} \cdot \log w_i(\theta_0)^{b_2} \cdot \log \left( 1 + \frac{\xi_0(\hat{\beta}_0 - \tilde{\beta}_n)}{\tau_0w_i(\theta_0)} \right)^{b_3}.
\]

The last inequality holds due to (C.5) with \( x = 1 + \frac{\xi_0(\hat{\beta}_0 - \tilde{\beta}_n)}{\tau_0w_i(\theta_0)} \). Multiplying both sides of the inequality by \( w_i^{−k−\frac{a}{c}}(\theta_0) \) and summing over \( i \),

\[
\frac{1}{n} \sum_{i=1}^n w_i^{−k−\frac{a}{c}}(\theta_0) \log^b w_i(\tilde{\theta}_n) = \frac{1}{n} \sum_{i=1}^n w_i^{−k−\frac{a}{c}}(\theta_0) \log^b w_i(\theta_0) + \frac{1}{n} \log^b \left( \frac{\xi_n \tau_0}{\tau_n \xi_0} \right) \cdot \sum_{i=1}^n w_i^{−k−\frac{a}{c}}(\theta_0) + R_1(\tilde{\theta}_n),
\]

where

\[
|R_1(\tilde{\theta}_n)| \leq \sum_{b_1 + b_2 + b_3 = b} \left( \frac{b}{b_1, b_2, b_3} \right) \frac{2^{b_3}}{(1 - \eta)^{b_3}} \left| \xi_0(\hat{\beta}_0 - \tilde{\beta}_n) \right|^{b_3} \left| \log \left( \frac{\xi_n \tau_0}{\tau_n \xi_0} \right) \right|^{b_1} \times \frac{1}{n} \sum_{i=1}^n w_i^{−k−\frac{a}{c}−b_3}(\theta_0) \log w_i(\theta_0)^{b_2}.
\]

Lemma C.3 ensures the averages in the right side of last inequality are bounded almost surely. Since \( b_3 \geq 1 \) and \( \beta_0 - \tilde{\beta}_n \xrightarrow{a.s.} 0 \), we know \( R_1(\tilde{\theta}_n) \xrightarrow{a.s.} 0 \).

In the meantime, Lemma C.3 entails \( \frac{1}{n} \log^b \left( \frac{\xi_n \tau_0}{\tau_n \xi_0} \right) \cdot \sum_{i=1}^n w_i^{−k−\frac{a}{c}}(\theta_0) \xrightarrow{a.s.} 0 \), and thus (C.7) is proved.

Lastly, we prove \( \frac{1}{n} w_i^{−k−\frac{a}{c}}(\tilde{\theta}_n) \log^b w_i(\tilde{\theta}_n) \xrightarrow{a.s.} (-\xi_0)^b \tau^b(k\xi_0 + a + 1) \).
Apply (C.6) with $x = 1 + \frac{\xi_0(\beta_0 - \hat{\beta}_n)}{\tau_0 w_i(\theta_0)}$, and

$$\left| \left(1 + \frac{\xi_0(\beta_0 - \hat{\beta}_n)}{\tau_0 w_i(\theta_0)} \right)^{-k - \frac{a}{\xi_n}} - 1 - \left(k + \frac{a}{\xi_n}\right) \frac{\xi_0(\beta_0 - \hat{\beta}_n)}{\tau_0 w_i(\theta_0)} \right| \leq \frac{(k + \frac{a}{\xi_n})(k + \frac{a}{\xi_n} + 1)(1 - \eta)^{-k - \frac{a}{\xi_n} - 2}}{2} \frac{\xi_0^2(\beta_0 - \hat{\beta}_n)^2}{\tau_0^2 w_i^2(\theta_0)}.$$  

Multiplying every term by $w_i^{-k - \frac{a}{\xi_n}(\theta_0)} \log(\tilde{\theta}_n)^b$ and summing over $i$,

$$\frac{1}{n} \sum_{i=1}^{n} w_i(\theta_0) + \frac{\xi_0(\beta_0 - \hat{\beta}_n)}{\tau_0} \log^b(\tilde{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{a}{\xi_n}(\theta_0)} \log^b(\tilde{\theta}_n) - \frac{(k + \frac{a}{\xi_n})(k + 1 + \frac{1}{\xi_n})(1 - \eta)^{-k - \frac{a}{\xi_n} - 2}(\beta_0 - \hat{\beta}_n)^2}{2\tau_0^2} \times \frac{1}{n} \sum_{i=1}^{n} w_i^{-k + \frac{a}{\xi_n}(\theta_0)} \log(\tilde{\theta}_n)^b.$$  

(C.8)

in which the left-hand side is also equal to $(\xi_n \tau_0)^{k + \frac{a}{\xi_n}} \cdot \frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{a}{\xi_n}(\theta_0)} \log^b(\tilde{\theta}_n)$, and

$$|R_2(\tilde{\theta}_n)| \leq \frac{\xi_0^2(\beta_0 - \hat{\beta}_n)^2}{\tau_0^2} \times \frac{1}{n} \sum_{i=1}^{n} w_i^{-k + \frac{a}{\xi_n}(\theta_0)} \log(\tilde{\theta}_n)^b.$$  

Using arguments similar to that leading to $R_1(\hat{\theta}_n) \overset{a.s.}{\to} 0$ yields $R_2(\tilde{\theta}_n) \overset{a.s.}{\to} 0$. Simplifying terms in (C.8), we conclude

$$\frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{a}{\xi_n}(\theta_0)} \log^b(\tilde{\theta}_n) \sim \left(\xi_n \tau_0\xi_0\right)^{-k - \frac{a}{\xi_n}} \cdot \frac{1}{n} \sum_{i=1}^{n} w_i^{-(k + \frac{a}{\xi_n})(\theta_0)} \log^b(\tilde{\theta}_n) \quad a.s.,$$

the right-hand side of which converges almost surely to $(-\xi_0)^b \Gamma^{(b)} (k \xi_0 + a + 1)$ by (C.7). This ends the proof of Proposition 4.1 for $\xi_0 > 0$.

**B) Case $\xi_0 < 0$.** Since $k \xi_0 + a + 1 > 0$, $k + \frac{a}{\xi_n} < -\frac{1}{\xi_n}$, which means $k + \frac{a}{\xi_n}$ can be positive. Therefore we need to approximate $x^t$ for $t > 0$ in $(1 - \eta, 1 + \eta)$ by the polynomial terms of $(x - 1)$—that is, to establish different versions of (C.6). When $1 \leq t < 2$, (C.6) holds. When $0 < t < 1$,

$$1 + t(x - 1) + \frac{t(t - 1)}{2} (1 - \eta)^{t - 2} (x - 1)^2 \leq x^t \leq 1 + t(x - 1),$$

and when $t \geq 2$,

$$1 + t(x - 1) \leq x^t \leq 1 + t(x - 1) + \frac{t(t - 1)}{2} (1 + \eta)^{t - 2} (x - 1)^2.$$

On the other hand, we know by Corollary B.1.1 that $\frac{\beta_0 - \hat{\beta}_n}{\beta_0 - Y_n \overset{a.s.}{\to} 0}$. Hence there almost surely exists $N$ such that for all $n > N$,

$$1/\xi_n < 0, \quad \text{and} \quad \left| \frac{\xi_0(\beta_0 - \hat{\beta}_n)}{\tau_0 w_i(\theta_0)} \right| = \left| \frac{\beta_0 - \hat{\beta}_n}{Y_i - \beta_0} \right| \leq \frac{\beta_0 - \hat{\beta}_n}{\beta_0 - Y_n} < \eta,$$

where $\eta \in (0, 1)$ is fixed.

To prove Proposition 4.1 for $\xi_0 < 0$, apply (C.5) and the varieties of (C.6) with $x = 1 + \xi_0(\beta_0 - \hat{\beta}_n)/\tau_0 w_i(\theta_0)$, and follow the same steps as proving the case $\xi_0 > 0$. \qed
C.2. Proof of Proposition 4.3.

PROOF. To prove \( \sup_{\alpha \in I} |\Phi_n(\alpha) - \Phi(\alpha)| \to \epsilon \), it suffices to prove that for any fixed \( \epsilon > 0 \), there almost surely exists \( N > 0 \) such that for all \( n > N \),
\[
\sup_{\alpha \in I} |\Phi_n(\alpha) - \Phi(\alpha)| < \epsilon.
\]

We first consider the case where \( \xi_0 > 0 \). Since \( \Phi(\alpha) = (-\xi_0)^b \Gamma(\alpha \xi_0 + 1) \) is uniformly continuous in \( I = [m, M] \), we can find \( \eta_1 > 0 \) such that for any \( \alpha_1, \alpha_2 \in I \) with \( |\alpha_1 - \alpha_2| < \eta_1 \),
\[
|\Phi(\alpha_1) - \Phi(\alpha_2)| < \frac{\epsilon}{3}.
\]

Define the interval \( \tilde{I} = \left( \frac{m-1/\xi_0}{2}, \frac{2M+m+1/\xi_0}{2} \right) \) and
\[
\Pi_2 = \xi_0^{b+1} \sup_{\alpha \in I} \Gamma^{(b+1)}(\alpha \xi_0 + 1) + \frac{3\xi_0^{b+1}}{(m \xi_0 + 1)^{b+2}} \Gamma(b + 2),
\]
\[
\Pi_3 = \xi_0^{b+2} \sup_{\alpha \in I} \Gamma^{(b+2)}(\alpha \xi_0 + 1) + \frac{3\xi_0^{b+2}}{(m \xi_0 + 1)^{b+3}} \Gamma(b + 3).
\]

Fix \( \eta = \min \left\{ \eta_1, \frac{\epsilon}{3 \Pi_2}, \frac{\epsilon}{3 \Pi_3}, \frac{m+1/\xi_0}{2} \right\} \). Denote \( B(\alpha, \eta) = (\alpha - \eta, \alpha + \eta) \). Note the set \( I \) is compact and is covered by the intervals \( \{ B(\alpha, \eta) : \alpha \in I \} \). Let \( B_i = B(\alpha_i, \eta) \), \( 1 \leq i \leq p \), be a finite cover. Then
\[
\sup_{\alpha \in I} |\Phi_n(\alpha) - \Phi(\alpha)| \leq \max_{1 \leq i \leq p} \sup_{\alpha \in B_i} |\Phi_n(\alpha) - \Phi(\alpha)|
\]
\[
\leq \max_{1 \leq i \leq p} \sup_{\alpha \in B_i} |\Phi_n(\alpha) - \Phi_n(\alpha_i)| + \max_{1 \leq i \leq p} \sup_{\alpha \in B_i} |\Phi(\alpha) - \Phi(\alpha_i)| + \max_{1 \leq i \leq p} |\Phi_n(\alpha_i) - \Phi(\alpha_i)|.
\]

From (C.9), we know
\[
\max_{1 \leq i \leq p} \sup_{\alpha \in B_i} |\Phi(\alpha) - \Phi(\alpha_i)| < \frac{\epsilon}{3}.
\]

Since \( p \) is a finite number, the pointwise consistency in Proposition 4.1 will ensure that there almost surely exists \( N_1 > 0 \) such that for all \( n > N_1 \),
\[
\max_{1 \leq i \leq p} |\Phi_n(\alpha_i) - \Phi(\alpha_i)| < \frac{\epsilon}{3}.
\]

Now we examine the first term on the right-hand side of (C.10). Since \( \eta \leq \frac{m+1/\xi_0}{2} \), we have
\[
\alpha_i - \eta \geq \frac{m-1/\xi_0}{2}, \xi_0(\alpha_i - \eta) + 1 \geq \frac{m \xi_0 + 1}{2} > 0,
\]
\[
\alpha_i + \eta \leq \frac{2M + m + 1/\xi_0}{2},
\]
for all \( 1 \leq i \leq p \). From (C.3), we deduce for \( \alpha \in B_i \) that
\[
|w_i^{-\alpha}(\hat{\theta}_n) - w_i^{-\alpha_i}(\hat{\theta}_n)| \leq |\alpha - \alpha_i| \cdot |w_i^{-\alpha_i}(\hat{\theta}_n) \cdot \log w_i(\hat{\theta}_n)| + (\alpha - \alpha_i)^2 \cdot (w_i^{-\alpha_i + \eta}(\hat{\theta}_n) + w_i^{-\alpha_i - \eta}(\hat{\theta}_n)) \log^2 w_i(\hat{\theta}_n),
\]
Therefore,

$$|\Phi_n(\alpha) - \Phi_n(\alpha_i)| = \left| \frac{1}{n} \sum_{i=1}^{n} \{ w_i^{-\alpha}(\hat{\theta}_n) - w_i^{-\alpha_i}(\hat{\theta}_n) \} \log w_i(\hat{\theta}_n) \right|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \{ w_i^{-\alpha}(\hat{\theta}_n) - w_i^{-\alpha_i}(\hat{\theta}_n) \} \left| \log w_i(\hat{\theta}_n) \right|^{b+1} +$$

$$\frac{(\alpha - \alpha_i)^2}{n} \sum_{i=1}^{n} w_i^{-\alpha_i}(\hat{\theta}_n) \left| \log w_i(\hat{\theta}_n) \right|^{b+2} +$$

$$\frac{(\alpha - \alpha_i)^2}{n} \sum_{i=1}^{n} w_i^{-\alpha_i}(\hat{\theta}_n) \left| \log w_i(\hat{\theta}_n) \right|^{b+2}. \quad (C.14)$$

Recall that Lemma C.1 guarantees that there almost surely exists $N_2 > 0$ such that for all $n > N_2$ and $1 \leq i \leq p$,

$$\frac{1}{n} \sum_{i=1}^{n} w_i^{-\alpha_i}(\hat{\theta}_n) \left| \log w_i(\hat{\theta}_n) \right|^{b+1} \leq \eta \left( b+1 \right) (\alpha_i \xi_0 + 1) + \frac{3\xi_0^{b+1} \Gamma (b+2)}{(\alpha_i \xi_0 + 1)^{b+2}} < \Pi_2. \quad (C.13)$$

Similarly, there almost surely exists $N_3 > 0$ such that for all $n > N_3$ and $1 \leq i \leq p$,

$$\frac{1}{n} \sum_{i=1}^{n} w_i^{-\alpha_i+\eta}(\hat{\theta}_n) \left| \log w_i(\hat{\theta}_n) \right|^{b+2} \leq \eta \left( b+2 \right) (\xi_0 (\alpha_i - \eta) + 1) + \frac{3\xi_0^{b+2} \Gamma (b+3)}{(\xi_0 (\alpha_i - \eta) + 1)^{b+3}} < \Pi_3. \quad (C.13)$$

Therefore, the bound in (C.14) can be relaxed as

$$|\Phi_n(\alpha) - \Phi_n(\alpha_i)| < |\alpha - \alpha_i| \Pi_2 + 2(\alpha - \alpha_i)^2 \Pi_3,$$

and thus for all $1 \leq i \leq p$,

$$(C.15) \quad \sup_{\alpha \in B_i} |\Phi_n(\alpha) - \Phi_n(\alpha_i)| < \eta \Pi_2 + 2\eta^2 \Pi_3 < \eta + 2\eta = \frac{\epsilon}{3}. \quad (C.13)$$

The last inequality stems from $\eta < \frac{\epsilon}{3\Pi_2}$ and $\eta < \frac{\epsilon}{3\Pi_3}$.

To sum up, we plug (C.11), (C.12) and (C.15) back into (C.10). We conclude that there almost surely exists $N = \max \{ N_1, N_2, N_3 \}$ such that for all $n > N$,

$$\sup_{\alpha \in I} |\Phi_n(\alpha) - \Phi(\alpha)| < \epsilon,$$

which completes the proof of this proposition for $\xi_0 > 0$. For the case $\xi_0 < 0$, the proof is analogous.

APPENDIX D: PROOFS CONCERNING THE LOCAL CONCAVITY

D.1. Proof of Proposition 4.2.

PROOF. Let $B_r(\theta) = \{ \theta' \in \Theta : ||\theta' - \theta||_\infty < r \}$, where $|| \cdot ||_\infty$ is the maximum norm.

For the conciseness of the appendices, we only provide proof for the case when the shape parameter $\xi_0 < 0$. The proof of the case $\xi_0 > 0$ is much easier because $k + 1/\xi_0$ is always positive when $k$ is a positive integer.
If \( \xi_0 \in (-1/2, 0) \), the proof requires more careful treatment. We first find an integer \( K_0 \geq 2 \) such that \(-1/K_0 < \xi_0 \leq -1/(K_0 + 1)\). We further select \( \epsilon_0 > 0 \) to tighten the left bound such that \(-1/(K_0 + \epsilon_0) < \xi_0 \leq -1/(K_0 + 1)\). Find \( r > 0 \) small enough such that

\[
(D.1) \quad -\frac{1}{K_0 + \epsilon_0} < \xi_0 - r < \xi_0 + r < -\frac{1}{K_0 + 2},
\]

which leads to \( \xi_0 / (\xi_0 - r) > -\xi_0 (K_0 + \epsilon_0) \). Therefore, for any integer \( k \in \{0, 1, \ldots, K_0 + 1\} \),

\[
(D.2) \quad k\xi_0 + \frac{\xi_0}{\xi_0 - r} + 1 \geq (K_0 + 1)\xi_0 + \frac{\xi_0}{\xi_0 - r} + 1
\]

\[
> (K_0 + 1)\xi_0 - \xi_0 (K_0 + \epsilon_0) + 1 = \xi_0 (1 - \epsilon_0) + 1 > 0,
\]

where the last inequality holds because \( \xi_0 \in (-1, 0) \).

Meanwhile, for any \( \xi \in (\xi_0 - r, \xi_0 + r) \),

\[
(D.3) \quad K_0 + \epsilon_0 < -\frac{1}{\xi_0 - r} < -\frac{1}{\xi} < -\frac{1}{\xi_0 + r}
\]

To approximate the Hessian in \( B_r(\theta) \) and prove Proposition 4.2, we need to check a few things:

(i) We calculate the Hessian matrix of the log-likelihood function \( L_n(\theta) \), and show that each element of the matrix is a linear combination of a few different types of sums:

\[
\sum_{i=1}^{n} w_i^{-k-\frac{1}{b}}(\theta), \quad \sum_{i=1}^{n} w_i^{-k-\frac{1}{b}}(\theta) \log^b w_i(\theta), \quad \text{where } k \text{ and } b \text{ are integers. This is verified in Lemma D.1.}
\]

(ii) For small \( r > 0 \) that satisfies (D.1) and \( r < \min \left\{ \frac{\xi_0}{3}, \frac{4(\xi_0 - \tau_0)}{\xi_0 \tau_0} \right\} \), we prove that these sums divided by \( n \) are uniformly bounded for large sample size \( n \) in the neighborhood \( B_r(\hat{\theta}_n) \) due to the pseudo-law of large numbers (Proposition 4.1). The bounds only depend on \( \theta_0 \) and \( r \). This is established in Lemma D.2 and D.3.

(iii) Utilizing the convexity of the set \( B_r(\hat{\theta}_n) \cap \Omega_n \) and the mean value theorem, we prove the sums are Lipschitz continuous:

\[
\left| \sum_{i=1}^{n} w_i^{-k-\frac{1}{b}}(\theta) \log^b w_i(\theta) - \sum_{i=1}^{n} w_i^{-k-\frac{1}{b}}(\hat{\theta}_n) \log^b w_i(\hat{\theta}_n) \right| < nM||\theta - \hat{\theta}_n||_\infty,
\]

where \( M \) is positive constant that is defined by \( \theta_0 \) and the radius \( r \). This is assured by Proposition D.4.

With the Lipschitz continuity, we can easily show in the neighborhood \( B_r(\hat{\theta}_n) \cap \Omega_n \),

\[
I - A_0(r) \leq L_n''(\theta)\left\{L_n''(\hat{\theta}_n)\right\}^{-1} \leq I + A_0(r),
\]

where \( I \) is the \( 3 \times 3 \) identity matrix and \( A_0(r) \) is a symmetric positive-semidefinite matrix which only depends on \( \theta_0 \) and the radius \( r \), and whose largest eigenvalue tends to zero as \( r \to 0 \).

\[
(D.4) \quad (a + b)^r \leq C_r (a^r + b^r),
\]

where \( C_r = 1 \) if \( 0 < r \leq 1 \), and \( C_r = 2^{r-1} \) if \( r > 1 \).
Lemma D.1 (Hessian matrix). Given i.i.d random variables $Y_1, Y_2, \ldots$ with common distribution $P_{\theta_0}$, the elements in the Hessian matrix, $L''_n(\theta)$, can each be expressed as a linear combination of the following terms

$$ \sum_{i=1}^{n} w_i^{-k}(\theta), \quad \sum_{i=1}^{n} w_i^{-k - \frac{1}{\xi}}(\theta), \quad \sum_{i=1}^{n} \log w_i(\theta), $$

$$ \sum_{i=1}^{n} w_i^{-k'-\frac{1}{\xi}}(\theta) \log w_i(\theta), \quad \sum_{i=1}^{n} w_i^{-\frac{1}{\xi}}(\theta)(\log w_i(\theta))^2, $$

where $k = 0, 1, 2$, $k' = 0, 1$ and $\xi \neq 0$.

Proof. Given i.i.d observations $Y_1, \ldots, Y_n$, the elements in the Hessian matrix, $\partial^2 L_n / \partial \theta \partial \theta^T$, can be obtained as follows (the detailed calculations are omitted):

$$ \frac{\partial^2 L_n}{\partial \mu^2} |_{\theta} = \frac{(1 + \xi)\xi}{\tau^2} \sum_{i=1}^{n} w_i^{-2}(\theta) - \frac{(1 + \xi)}{\tau^2} \sum_{i=1}^{n} w_i^{-2 - \frac{1}{\xi}}(\theta), $$

$$ \frac{\partial^2 L_n}{\partial \mu \xi} |_{\theta} = -\frac{1}{\xi \tau} \sum_{i=1}^{n} w_i^{-1}(\theta) + \frac{\xi + 1}{\tau \xi} \sum_{i=1}^{n} w_i^{-1 - \frac{1}{\xi}}(\theta) + \frac{\xi + 1}{\tau \xi} \sum_{i=1}^{n} w_i^{-2}(\theta) $$

$$ - \left( \frac{1}{\tau \xi^2} + \frac{1}{\tau \xi} \right) \sum_{i=1}^{n} w_i^{-2 - \frac{1}{\xi}}(\theta) - \frac{1}{\tau \xi^2} \sum_{i=1}^{n} w_i^{-1 - \frac{1}{\xi}}(\theta) \log w_i(\theta), $$

$$ \frac{\partial^2 L_n}{\partial \mu \tau} |_{\theta} = -\frac{1}{\tau \xi^2} \sum_{i=1}^{n} w_i^{-1}(\theta) - \frac{(\xi + 1)}{\xi \tau^2} \sum_{i=1}^{n} w_i^{-2}(\theta) + \frac{\xi + 1}{\xi \tau^2} \sum_{i=1}^{n} w_i^{-2 - \frac{1}{\xi}}(\theta), $$

$$ \frac{\partial^2 L_n}{\partial \tau^2} |_{\theta} = -\frac{n \xi}{\xi \tau^2} + \frac{(\xi - 1)}{\xi \tau^2} \sum_{i=1}^{n} w_i^{-\frac{1}{\xi}}(\theta) + \frac{2 \xi}{\xi \tau^2} \sum_{i=1}^{n} w_i^{-1 - \frac{1}{\xi}}(\theta) + \frac{\xi (\xi + 1)}{\xi \tau^2} \sum_{i=1}^{n} w_i^{-2}(\theta) $$

$$ - \frac{\xi + 1}{\xi \tau^2} \sum_{i=1}^{n} w_i^{-2 - \frac{1}{\xi}}(\theta), $$

$$ \frac{\partial^2 L_n}{\partial \tau \xi} |_{\theta} = \frac{1}{\tau \xi} \left[ -(n + \xi + 1) \sum_{i=1}^{n} w_i^{-\frac{1}{\xi}}(\theta) + (2 + \xi) \sum_{i=1}^{n} w_i^{-1}(\theta) - \frac{2 (\xi + 1)}{\xi} \sum_{i=1}^{n} w_i^{-1 - \frac{1}{\xi}}(\theta) $$

$$ - (\xi + 1) \sum_{i=1}^{n} w_i^{-2}(\theta) + \frac{\xi + 1}{\xi} \sum_{i=1}^{n} w_i^{-2 - \frac{1}{\xi}}(\theta) - \frac{1}{\xi} \sum_{i=1}^{n} w_i^{-\frac{1}{\xi}}(\theta) \log w_i(\theta) $$

$$ + \frac{1}{\xi} \sum_{i=1}^{n} w_i^{-1 - \frac{1}{\xi}}(\theta) \log w_i(\theta) \right], $$

$$ \frac{\partial^2 L_n}{\partial \tau^2} |_{\theta} = \frac{n (\xi + 3)}{\xi^3} - \frac{3 \xi + 1}{\xi^4} \sum_{i=1}^{n} w_i^{-\frac{1}{\xi}}(\theta) - \frac{2 (\xi + 2)}{\xi^4} \sum_{i=1}^{n} w_i^{-1}(\theta) + \frac{2 (\xi + 1)}{\xi^4} \sum_{i=1}^{n} w_i^{-1 - \frac{1}{\xi}}(\theta) $$

$$ + \frac{\xi + 1}{\xi^3} \sum_{i=1}^{n} w_i^{-2}(\theta) - \frac{(\xi + 1)}{\xi^4} \sum_{i=1}^{n} w_i^{-2 - \frac{1}{\xi}}(\theta) - \frac{2}{\xi^3} \sum_{i=1}^{n} \log w_i(\theta) $$

$$ + \frac{2 (\xi + 1)}{\xi^3} \sum_{i=1}^{n} w_i^{-\frac{1}{\xi}}(\theta) \log w_i(\theta) - \frac{2}{\xi^3} \sum_{i=1}^{n} w_i^{-\frac{1}{\xi}}(\theta) (\log w_i(\theta))^2. $$

Lemma D.2 (Uniform bound). Suppose a set of true GEV parameters $\theta_0$ satisfies $-\frac{1}{K_0} < \xi_0 \leq -\frac{1}{(K_0 + 1)}$, where $K_0 \geq 1$ is some integer. Fix any $r > 0$ small enough such that both (D.1) and $r < \frac{3}{4}$ hold. Then there almost surely exists $N_1 > 0$ such that for...
any \( n > N_1 \),
\[
\sum_{i=1}^{n} w_i^{-k-\frac{1}{\xi}}(\theta) \leq n\phi_k(\theta_0, r), \text{ for } \theta \in B_r(\hat{\theta}_n) \cap \Omega_n,
\]
in which \( k = 0, \ldots, K_0 + 1 \), and \( \phi_k(\theta_0, r) > 0 \) is determined by \( k, \theta_0 \) and the radius \( r \).

**Proof. A. Case** \( k = 0, \ldots, K_0 \). When \( \theta \in B_r(\hat{\theta}_n) \), we know from (D.3) that
\[
0 > k - (K_0 + 1) > k + 1 > k + 1 + r.
\]
By the monotonicity of the power means,
\[
(D.5) \quad \frac{1}{n} \sum_{i=1}^{n} w_i^{-k-\frac{1}{\xi}}(\theta) \leq \left( \frac{1}{n} \sum_{i=1}^{n} w_i^{-k-\frac{1}{\xi}}(\theta) \right)^{\frac{k+1}{k+1/\xi}}.
\]
To prove the right-hand side of (D.5) is uniformly bounded, we first define three positive constants
\[
(D.6) \quad \eta_k = -k - \frac{1}{\xi_0 + r}, \quad \Psi_0 = -\frac{2\xi_0}{r_0}, \quad \text{and } \Lambda_0 = \frac{5(\xi_0 - \tau_0)}{\xi_0\tau_0}.
\]
Apply the classic \( c_\gamma \)-inequality (D.4) to \( \eta_k \), and we obtain
\[
\sum_{i=1}^{n} w_i^{\eta_k}(\theta) = \left( \frac{\xi \tau_n}{\tau_n} \right) \eta_k \sum_{i=1}^{n} \left( w_i(\hat{\theta}_n) + \frac{\hat{\xi}_n}{\tau_n}(\hat{\beta}_n - \beta) \right)^{\eta_k}
\]
\[
\leq C \eta_k \left( \frac{\xi \tau_n}{\tau_n} \right) \eta_k \left( \sum_{i=1}^{n} w_i^{\eta_k}(\theta) + n \left( \frac{\hat{\xi}_n}{\tau_n}(\hat{\beta}_n - \beta) \right)^{\eta_k} \right).
\]
By the strong consistency of \( \hat{\theta}_n \) and the fact that \( r < \tau_0/3 \), there almost surely exists \( N_{1,1} > 0 \) such that
\[
(D.8) \quad \hat{\xi}_n < 0, \quad r < \tau_n/2, \quad \left| \frac{\hat{\xi}_n}{\tau_n}(\hat{\beta}_n - \beta) \right| < -\frac{2\xi_0}{\tau_0}r = \Psi_0 r,
\]
for all \( n > N_{1,1} \) and \( \theta \in B_r(\hat{\theta}_n) \). Thereupon it can be shown via a convexity argument similar to that yielding Lemma 2.1 that
\[
\forall \theta \in B_r(\hat{\theta}_n), \quad 0 < \frac{\xi \tau_n}{\tau_n} \hat{\xi}_n \leq \left( \frac{\xi \tau_n}{\tau_n} \right) \hat{\xi}_n < 1 + \frac{4(\hat{\xi}_n - \tau_n)}{\xi_n \hat{\tau}_n} r.
\]
Then we can find \( N_{1,2} > 0 \) such that for all \( n > N_{1,2} \),
\[
(D.9) \quad \forall \theta \in B_r(\hat{\theta}_n), \quad 0 < \frac{\xi \tau_n}{\tau_n} < 1 + \Lambda_0 r.
\]
Also, (D.2) guarantees \( \frac{1}{n} \sum_{i=1}^{n} w_i^{-k-\frac{1}{\xi}}(\theta) \overset{a.s.}{\rightarrow} \Gamma(k\xi_0 + \frac{\xi_0}{\xi_0 - r} + 1) \), there almost surely exists \( N_{1,3} > 0 \) such that for any \( n > N_{1,3} \),
\[
\frac{1}{n} \sum_{i=1}^{n} w_i^{-k-\frac{1}{\xi}}(\theta) \leq \Gamma(k\xi_0 + \frac{\xi_0}{\xi_0 - r} + 1) + 1,
\]
where \( k = 1, \ldots, K_0 \).

Consequently, there almost surely exists \( N_1 = \max\{N_{1,1}, N_{1,2}, N_{1,3}\} \) such that for any \( n > N_1 \), (D.7) can be further bounded as follows:
\[
\sum_{i=1}^{n} w_i^{\eta_k}(\theta) \leq nC_{\eta_k} (1 + \Lambda_0 r)^{\eta_k} \left( \Gamma(k\xi_0 + \frac{\xi_0}{\xi_0 - r} + 1) + 1 + \Psi_0 r \right)^{\eta_k}.
\]
Plugging this result back into (D.5) while noticing that \(0 < \frac{k+1}{k+1/(\xi_0-r)} < 1\), we have
\[
\frac{1}{n} \sum_{i=1}^{n} w_i^{-k - \frac{1}{\xi}}(\theta) \leq C_{k_0} \frac{k+1/(\xi_0-r)}{k+1/(\xi_0-r)} \left( 1 + \Lambda_0 r \right)^{-k - \frac{1}{\xi}} \left( \Gamma(k\xi_0 + \frac{\xi_0}{\xi_0-r} + 1) + (\Psi_0 r)^{-k - \frac{1}{\xi}} \right) \frac{k+1}{k+1/(\xi_0-r)}.
\]

\[
\leq C_{k_0} (1 + \Lambda_0 r)^{-k - \frac{1}{\xi}} \left( \Gamma(k\xi_0 + \frac{\xi_0}{\xi_0-r} + 1) + (\Psi_0 r)^{-k - \frac{1}{\xi}} \right) \phi_k(\theta_0, r).
\]

\[\text{(D.3)}\]

**B. Case** \(k = K_0 + 1\). For \(\theta \in B_r(\hat{\theta}_n)\), (D.3) ensures
\[
K_0 + 1 + \frac{1}{\xi_0 - r} > K_0 + 1 + \frac{1}{\xi} > K_0 + 1 + \frac{1}{\xi_0 + r},
\]
and (D.1) ensures
\[
K_0 + 1 + \frac{1}{\xi_0 - r} > K_0 + 1 + \frac{1}{\xi_0} \geq 0.
\]

If \(K_0 + 1 + \frac{1}{\xi} > 0\), the monotonicity of the power means yields
\[
\frac{1}{n} \sum_{i=1}^{n} w_i^{-K_0 - 1 - \frac{1}{\xi}}(\theta) \leq \left( \frac{1}{n} \sum_{i=1}^{n} w_i^{-K_0 - 1 - \frac{1}{\xi}}(\theta) \right) \frac{K_0 + 1 + 1}{K_0 + 1 + 1/(\xi_0 - r)}.
\]

If \(K_0 + 1 + \frac{1}{\xi} < 0\), the monotonicity of the power means then yields
\[
\frac{1}{n} \sum_{i=1}^{n} w_i^{-K_0 - 1 - \frac{1}{\xi}}(\theta) \leq \left( \frac{1}{n} \sum_{i=1}^{n} w_i^{-K_0 - 1 - \frac{1}{\xi}}(\theta) \right) \frac{K_0 + 1 + 1}{K_0 + 1 + 1/(\xi_0 + r)}.
\]

For both cases, we follow a similar proof to the previous case with the help of the classic \(c_\theta\)-inequality (D.4) to obtain upper bounds that depends only on \(\theta_0\) and \(r\). Pick the greater of the two bounds, and we have \(\phi_{K_0+1}(\theta_0, r)\). \(\Box\)

**Lemma D.3** (Uniform bound). Under the assumptions of Lemma D.2, fix any \(r > 0\) small enough such that both (D.1) and \(r < \min\{r_0, \frac{0.8}{k_0}\}\) hold, where \(\Lambda_0\) is defined in (D.6).

Then there almost surely exists \(N_2\) such that for any \(n > N_2\) and \(\theta \in B_r(\hat{\theta}_n) \cap \Omega_n\),
\[
\sum_{i=1}^{n} w_i^{-k - \frac{1}{\xi}}(\theta) \left| \log w_i(\theta) \right|^b \leq n \varphi_k(\theta_0, r),
\]
where \(k \in \{0, \ldots, K_0 + 1 - b\}\), \(b \in \{1, 2, 3\}\), and \(\varphi_k(\theta_0, r)\) only depends on \(k\), \(\theta_0\) and \(r\).

**Proof.** We deduce by (1)
\[
\sum_{i=1}^{n} w_i^{-k - \frac{1}{\xi}}(\theta) \left| \log w_i(\theta) \right|^b = \left| \log \left( \frac{\xi \hat{\xi}_n}{\tau \hat{\xi}_n} \right) \right|^b \sum_{i=1}^{n} w_i^{-k - \frac{1}{\xi}}(\theta) + \sum_{i=1}^{n} w_i^{-k - \frac{1}{\xi}}(\theta) \left| \log w_i(\theta) + \frac{\hat{\xi}_n(\hat{\beta}_n - \beta)}{\tau_n} \right|^b.
\]

Then (D.9) implies for all \(n > N_1\),
\[
\left| \log \left( \frac{\xi \hat{\xi}_n}{\tau \hat{\xi}_n} \right) \right| \leq \frac{\sqrt{2}}{1-\Lambda_0 r} \Lambda_0 r < \frac{\sqrt{2}}{1-0.8} \Lambda_0 r = 5\sqrt{2} \Lambda_0 r,
\]
and Lemma D.2 guarantees
\begin{equation}
(D.11) \quad \left| \log \left( \frac{\xi \tau_n}{\tau \xi_n} \right) \right|^b \sum_{i=1}^n w_i^{-k-rac{1}{2}}(\theta) < n(5\sqrt{2}A_0r)^b \phi_k(\theta_0, r), \text{ for } \theta \in B_r(\hat{\theta}_n) \cap \Omega_n.
\end{equation}

Since $|\log x - \log y| \leq |x - y|/\min(x, y)$, and (D.8) ensures $\left| \frac{\xi_n(\beta_n - \beta)}{\tau_n} \right| < \Psi_0 r$ for $n > N_{1,1}$,
\[\left| \log \left( w_i(\hat{\theta}_n) + \frac{\xi_n(\beta_n - \beta)}{\tau_n} \right) - \log w_i(\hat{\theta}_n) \right| \leq \frac{\xi_n(\beta_n - \beta)}{\tau_n w_i(\hat{\theta}_n)} < \Psi_0 r w_i^{-1}(\hat{\theta}_n).\]
Therefore,
\begin{equation}
(D.12) \quad \left| \log \left( w_i(\hat{\theta}_n) + \frac{\xi_n(\beta_n - \beta)}{\tau_n} \right) \right|^b < \left| \log w_i(\hat{\theta}_n) \right|^b + \Psi_0 r w_i^{-b}(\hat{\theta}_n).
\end{equation}

Apply the $c_r$-inequality (D.4) on (D.12) to get
\begin{equation}
(D.13) \quad \left| \log \left( w_i(\hat{\theta}_n) + \frac{\xi_n(\beta_n - \beta)}{\tau_n} \right) \right|^b < C_b \left( \left| \log w_i(\hat{\theta}_n) \right|^b + (\Psi_0 r)^b w_i^{-b}(\hat{\theta}_n) \right).
\end{equation}

Plugging (D.11) and (D.13) back into (D.10), we have
\[\sum_{i=1}^n w_i^{-k-rac{1}{2}}(\theta) |\log w_i(\theta)|^b < C_b \sum_{i=1}^n w_i^{-k-rac{1}{2}}(\theta) |\log w_i(\hat{\theta}_n)|^b + C_b (\Psi_0 r)^b \sum_{i=1}^n w_i^{-k-rac{1}{2}}(\theta) w_i^{-b}(\hat{\theta}_n) + n(5\sqrt{2}A_0r)^b \phi_k(\theta_0, r).
\]
By Lemma D.2, the right-hand side of the previous inequation can be uniformly bounded by a constant $\varphi_k(\theta_0, r)$ that is determined by $k, \theta_0$ and $r$, $k = 0, \ldots, K_0 + 1 - b$.

\begin{proposition}[Lipschitz continuity] Under the assumptions of Lemma D.2 and D.3, there almost surely exists $N$ such that, for any $n > N$ and $\theta \in B_r(\hat{\theta}_n) \cap \Omega_n$,
\[\sum_{i=1}^n w_i^{-k-rac{1}{2}}(\theta) - \sum_{i=1}^n w_i^{-k-rac{1}{2}}(\hat{\theta}_n) \leq n \psi_k(\theta_0, r) \|\theta - \hat{\theta}_n\|_\infty,
\]
\[\sum_{i=1}^n w_i^{-k-rac{1}{2}}(\theta) \log w_i(\theta) - \sum_{i=1}^n w_i^{-k-rac{1}{2}}(\hat{\theta}_n) \log w_i(\hat{\theta}_n) \leq n \psi_{k'}(\theta_0, r) \|\theta - \hat{\theta}_n\|_\infty,
\]
\[\sum_{i=1}^n w_i^{-k-rac{1}{2}}(\theta) \log^2 w_i(\theta) - \sum_{i=1}^n w_i^{-k-rac{1}{2}}(\hat{\theta}_n) \log^2 w_i(\hat{\theta}_n) \leq n \psi_{k''}(\theta_0, r) \|\theta - \hat{\theta}_n\|_\infty,
\]
where $k = 0, \ldots, K_0$, $k' = 0, \ldots, K_0 - 1$, $k'' = 0, \ldots, K_0 - 2$, and $\psi_k(\theta_0, r), \psi_k(\theta_0, r), \psi_{k''}(\theta_0, r)$ are positive constants determined by $k$ (or $k', k''$), $\theta_0$ and the radius $r$.

\begin{proof}
Given that $\theta \in B_r(\hat{\theta}_n) \cap \Omega_n$ is a convex set under the parametrization $(\beta, \xi, \tau)$, the mean value theorem can be applied to assess
\[\sum_{i=1}^n w_i^{-k-rac{1}{2}}(\theta) - \sum_{i=1}^n w_i^{-k-rac{1}{2}}(\hat{\theta}_n) \leq \left\| \sum_{i=1}^n \nabla \left( w_i^{-k-rac{1}{2}} \right) \right\|_{\infty} \|\theta - \hat{\theta}_n\|_\infty,
\]
\end{proof}
where \( \mathbf{\theta}_n^+ = (\beta_n^+, \xi_n^+, \tau_n^+) \) lies between \( \mathbf{\theta} \) and \( \hat{\mathbf{\theta}}_n \), and the Jacobian
\[
\nabla (w_{i-1} - \frac{1}{\beta}) \bigg|_\theta = \frac{k \xi + 1}{\xi} w_i \bigg( \frac{\xi}{w_i - \frac{1}{\beta}} \bigg) \cdot \bigg( \frac{1}{\xi} \log w_i(\theta) - \frac{1}{\xi} \big(1 - \frac{1}{w_i - \frac{1}{\beta}}\big)\bigg).
\]
By Lemma D.2 and D.3, we can easily find \( \psi_k(\theta_0, r) > 0 \) which is determined by \( k, \theta_0 \) and the radius \( r \) such that
\[
\left\| \sum_{i=1}^n \nabla \left( w_{i-1} - \frac{1}{\beta} \right) \right\|_\infty \leq \psi_k(\theta_0, r).
\]
Therefore the first inequality holds.

Similarly, we derive
\[
\nabla \left( w_{i-1} - \frac{1}{\beta} \log w_i \right) \bigg|_\theta = \left( w_{i-1} - \frac{1}{\beta} \log w_i(\theta) \right) \cdot \left[ \frac{k \xi + 1}{\xi} \log w_i(\theta) - \frac{2}{\tau} \right],
\]
\[
\nabla \left( w_{i-1} - \frac{1}{\beta} \log^2 w_i \right) \bigg|_\theta = \left( w_{i-1} - \frac{1}{\beta} \log^2 w_i(\theta) \right) \cdot \left[ \frac{k \xi + 1}{\xi^2} \log^2 w_i(\theta) - \frac{1}{\xi} \big(1 - \frac{1}{w_i - \frac{1}{\beta}}\big)\bigg]\bigg(\frac{k \xi + 1}{\xi^2} \log w_i(\theta) - \frac{2}{\tau}\bigg),
\]
in which all the terms under expansion have been studied in Lemma D.2 and D.3. Hence, the second and third inequality also holds.

**APPENDIX E: PROOF OF PROPOSITION 4.5**

We first study the relationship between \( \beta_n(\xi) \) and \( \hat{\beta}_n(\xi) \) when \( \xi \neq \hat{\xi}_n \). This relationship is illustrated visually in Figure 5, where we simulate \( n = 1,000 \) samples using \((\tau_0, \mu_0, \xi_0) = (0.5, 20, 0.2)\), and then evaluate the log-likelihood on different cross sections. Through comparing the intercepts, the numerical results confirm Proposition 2.1 in that \( \beta_n(\xi) < \hat{\beta}_n \) when \( \xi \neq \hat{\xi}_n \), and \( \beta_n(\xi) > \hat{\beta}_n \) as \( \xi \geq \hat{\xi}_n \). This property stems from the monotonicity of \( \beta_n(\xi) \) as a function of \( \xi \).

**LEMMA E.1.** Under the same assumptions as Proposition 2.1, \( \beta_n(\xi) = \mu_n(\xi) - \tau_n(\xi)/\xi \) as function of \( \xi \) satisfies \( \beta'_n(\xi) > 0 \) for any \( \xi \neq 0 \) and \( -1 < \xi < n - 1 \). Furthermore, Lemma A.2 guarantees \( \beta_n(\xi) \nless Y_1 \text{ as } \xi \nless n - 1 \), and \( \beta_n(\xi) \nleq Y_n \text{ as } \xi \nleq n - 1 \); see Figure 6.

**PROOF.** See the supplementary material for the proof, in which we calculate the derivative via applying the implicit function theorem on (5).

**LEMMA E.2.** Suppose \( \xi_0 > 0 \), and \( Y_1, \ldots, Y_n \sim iid P_{\theta_0} \). Define
\[
F_n(\xi) = n \log \xi + \frac{1}{\xi} \sum_{i=1}^n \log w_i(\hat{\theta}_n) + n \log \sum_{i=1}^n w_i(\hat{\theta}_n)^{\frac{1}{n}}.
\]
Then for any \( \xi > 0 \), \( F_n(\xi) - F_n(\hat{\xi}_n) \geq n \left( \frac{k}{\xi} \log \hat{\xi}_n - 1 \right) \).
FIG 5. Contour plots of the log-likelihood $L_n(\theta)$ at cross sections of different levels of $\xi$, where $Y_i$, $i=1, \ldots, n$ that are sampled from true $\xi = 0.2$. Blue points mark the maximizer $(\tau_n(\xi), \mu_n(\xi))$ on each slice, and the black dashed lines showing the lines that has intercept $\hat{\beta}_n$ and slope $1/\xi$. Left: $\xi = 0.5\hat{\xi}_n$, and we see $\beta_n(\xi) < \hat{\beta}_n$. Middle: $\xi = \hat{\xi}_n$, and $\hat{\beta}_n = \beta_n(\hat{\xi}_n)$. Right: $\xi = 1.5\hat{\xi}_n$, and we see $\beta_n(\xi) > \hat{\beta}_n$.

FIG 6. Illustration of $\beta_n(\xi)$ when $\xi \in (-1, n-1)$.

PROOF. The first derivative of $F_n$ is

$$F'_n(\xi) = \frac{1}{\xi^2} \left[ n\xi - \sum_{i=1}^{n} \log w_i(\hat{\theta}_n) + \frac{n}{\xi} \sum_{i=1}^{n} w_i(\hat{\theta}_n)^{-\frac{1}{\xi}} \log w_i(\hat{\theta}_n) \right].$$

Since $\frac{\partial L_n}{\partial \xi}|_{\hat{\theta}_n} = 0$ by definition, we deduce

$$-\sum_{i=1}^{n} \log w_i(\hat{\theta}_n) = -n\hat{\xi}_n - \frac{n}{\xi} \sum_{i=1}^{n} w_i(\hat{\theta}_n)^{-\frac{1}{\xi}} \log w_i(\hat{\theta}_n).$$

Therefore,

$$F'_n(\xi) = \frac{n}{\xi^2} \left[ \xi - \hat{\xi}_n + \frac{1}{\xi} \sum_{i=1}^{n} w_i(\hat{\theta}_n)^{-\frac{1}{\xi}} \log w_i(\hat{\theta}_n) - \frac{1}{\xi} \sum_{i=1}^{n} \frac{w_i(\hat{\theta}_n)^{-\frac{1}{\xi}} \log w_i(\hat{\theta}_n)}{w_i(\hat{\theta}_n)} \right].$$

Denote $x_i = w_{(i)}(\hat{\theta}_n)^{-\frac{1}{\xi}}$, $y_i = w_{(i)}(\hat{\theta}_n)^{-\frac{1}{\xi}}$, $z_i = \log w_{(i)}(\hat{\theta}_n)$ and $u_i = 1$. When $\xi > \hat{\xi}_n$,

$$\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} \times \begin{vmatrix} z_r & z_s \\ u_r & u_s \end{vmatrix} = \frac{\left[ w_{(i)}(\hat{\theta}_n)^{\frac{1}{\xi} - \frac{1}{\xi}} - w_{(j)}(\hat{\theta}_n)^{\frac{1}{\xi} - \frac{1}{\xi}} \right] [\log w_{(r)}(\hat{\theta}_n) - \log w_{(s)}(\hat{\theta}_n)]}{\left[ w_{(i)}(\hat{\theta}_n) w_{(j)}(\hat{\theta}_n) \right]^{\frac{1}{\xi}}} > 0,$$

$$\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} \times \begin{vmatrix} z_r & z_s \\ u_r & u_s \end{vmatrix} = \frac{\left[ w_{(i)}(\hat{\theta}_n)^{\frac{1}{\xi} - \frac{1}{\xi}} - w_{(j)}(\hat{\theta}_n)^{\frac{1}{\xi} - \frac{1}{\xi}} \right] [\log w_{(r)}(\hat{\theta}_n) - \log w_{(s)}(\hat{\theta}_n)]}{\left[ w_{(i)}(\hat{\theta}_n) w_{(j)}(\hat{\theta}_n) \right]^{\frac{1}{\xi}}} > 0,$$
for any pairs of indices $i < j$ and $r < s$. Hence we can apply the Seitz inequality to acquire
\[
\sum_{i=1}^{n} w_i(\hat{\theta}_n)^{-\frac{1}{\xi}} \log w_i(\hat{\theta}_n) \sum_{i=1}^{n} w_i(\hat{\theta}_n)^{-\frac{1}{\xi}} \geq \sum_{i=1}^{n} w_i(\hat{\theta}_n)^{-\frac{1}{\xi}} \log w_i(\hat{\theta}_n) \sum_{i=1}^{n} w_i(\hat{\theta}_n)^{-\frac{1}{\xi}}.
\]
Similarly, we can prove the above inequality holds in an opposite direction when $\xi < \hat{\xi}_n$.
Consequently,
\[
F'_n(\xi) \geq \frac{n(\xi - \hat{\xi}_n)}{\xi^2} \text{ for } \xi > \hat{\xi}_n, \text{ and } F'_n(\xi) \leq \frac{n(\xi - \hat{\xi}_n)}{\xi^2} \text{ for } \xi < \hat{\xi}_n.
\]
By Newton-Leibniz formula,
\[
(F_n(\xi) - F_n(\hat{\xi}_n)) = \int_{\hat{\xi}_n}^{\xi} \frac{n(\xi - \hat{\xi}_n)}{\xi^2} d\xi = n \left( \frac{\hat{\xi}_n}{\xi} - \log \left( \frac{\xi}{\hat{\xi}_n} \right) - 1 \right) \text{ for all } \xi > 0.
\]

**Lemma E.3.** Suppose $\xi_0 > 0$, and $Y_1, \ldots, Y_n \overset{iid}{\sim} P_{\theta_0}$, Define
\[
G_n(\xi) = \frac{\xi + 1}{\xi} \sum_{i=1}^{n} \log \frac{Y_i - \hat{\beta}_n}{Y_i - \beta_n(\xi)} + n \log \frac{\sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{-1/\xi}}{\sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-1/\xi}}.
\]
Then for $0 < \xi < \hat{\xi}_n$, $G_n(\xi) \leq -n \left[ 2\xi_0 \gamma \left( \frac{1}{\xi_0} - \frac{1}{\xi} \right) + \frac{1}{2} \log \Gamma \left( \frac{\xi}{\xi_0} + 1 \right) \right]$ almost surely when $n$ is sufficiently large.

**Proof.** We first calculate
\[
G'_n(\xi) = \frac{\beta'_n(\xi)}{\xi} \left[ \left( \xi + 1 \right) \sum_{i=1}^{n} \frac{(Y_i - \beta_n(\xi))^{-1}}{(Y_i - \beta_n(\xi))^{-1/\xi}} - \frac{n \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-1/\xi}}{\sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-1/\xi}} \right] + \frac{1}{\xi^2} \sum_{i=1}^{n} \log \frac{Y_i - \beta_n(\xi)}{Y_i - \beta_n} + \frac{n \sum_{i=1}^{n} (Y_i - \beta_n)^{-1/\xi} \log (Y_i - \beta_n)}{\xi^2 \sum_{i=1}^{n} (Y_i - \beta_n)^{-1/\xi}} - \frac{n \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-1/\xi} \log (Y_i - \beta_n(\xi))}{\xi^2 \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-1/\xi}},
\]
in which the term in the square brackets is 0 for $\xi > 0$ due to the second equation of (5).
For $\xi > 0$, we deduce from Chebyshev's sum inequality that
\[
\sum_{i=1}^{n} \log (Y_i - \beta_n(\xi)) \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-1/\xi} \geq n \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-1/\xi} \log (Y_i - \beta_n(\xi)),
\]
and thus
\[
(G'_n(\xi)) \geq -\frac{1}{\xi^2} \sum_{i=1}^{n} \log (Y_i - \hat{\beta}_n) + \frac{n \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{-1/\xi} \log (Y_i - \hat{\beta}_n)}{\xi^2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{-1/\xi}}.
\]
Apply the Newton-Leibniz formula to both sides of (E.2), and we obtain, for $\xi < \hat{\xi}_n$, that
\[
G_n(\hat{\xi}_n) - G_n(\xi) \geq \int_{\xi}^{\hat{\xi}_n} -\frac{1}{\xi^2} \sum_{i=1}^{n} \log (Y_i - \hat{\beta}_n) + \frac{n \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{-1/\xi} \log (Y_i - \hat{\beta}_n)}{\xi^2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{-1/\xi}} d\xi
\]
\[
= \left( \frac{1}{\hat{\xi}_n} - \frac{1}{\xi} \right) \sum_{i=1}^{n} \log (Y_i - \hat{\beta}_n) + n \log \frac{\sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{-1/\xi}}{\sum_{i=1}^{n} (Y_i - \beta_n)^{-1/\xi}}
\]
\[
= \left( \frac{1}{\hat{\xi}_n} - \frac{1}{\xi} \right) \sum_{i=1}^{n} \log w_i(\hat{\theta}_n) - n \log \frac{\sum_{i=1}^{n} w_i^{-1/\xi}(\theta_n)}{n}.
\]
Proposition 4.3 ensures \( \log \frac{1}{n} \sum_{i=1}^{n} w_i^{-\frac{1}{\xi}}(\hat{\theta}_n) \leq 2 \log \Gamma \left( \frac{\xi_0}{\xi} + 1 \right) \) uniformly, and \( \frac{1}{n} \sum_{i=1}^{n} \log w_i(\hat{\theta}_n) \leq 2\xi_0 \gamma \) for sufficiently large \( n \). Note that \( G_n(\xi_n) = 0 \), and (E.3) becomes

\[
-G_n(\xi) \geq 2n\xi_0 \gamma \left( \frac{1}{\xi_n} - \frac{1}{\xi} \right) + \frac{n}{2} \log \Gamma \left( \frac{\xi_0}{\xi} + 1 \right),
\]

which proves the lemma.

\[
\square
\]

E.1. Proof of relation (11).

Proof. When \( \xi > \xi_n, \beta_n(\xi) > \hat{\beta}_n \) and

\[
- \frac{n}{\xi} \sum_{i=1}^{\xi_n} \frac{\log \delta_i}{\delta_i^{-\frac{1}{\xi}}} + \sum_{i=1}^{n} \log \delta_i \leq - \frac{n}{\xi} \sum_{i=1}^{\xi_n} \frac{\log \hat{\delta}_i}{\hat{\delta}_i^{-\frac{1}{\xi}}},
\]

where \( \hat{\delta}_i = \hat{\xi}_n(Y_i - \hat{\beta}_n) \) and \( \delta_i = \xi(Y_i - \beta_n(\xi)) \). Plug this inequality in (A.2), and we obtain

\[
PL_n(\xi) \leq - \frac{n}{\xi} \sum_{i=1}^{\xi_n} \frac{\log \hat{\delta}_i}{\hat{\delta}_i^{-\frac{1}{\xi}}},
\]

Apply the Newton-Leibniz formula, and we get

\[
PL_n(\xi) - PL_n(\hat{\xi}_n) \leq - \int_{\hat{\xi}_n}^{\xi} \frac{n}{\xi} \frac{\log \hat{\delta}_i}{\hat{\delta}_i^{-\frac{1}{\xi}}} d\xi - \int_{\hat{\xi}_n}^{\xi_n} \frac{n}{\xi} \frac{\log \delta_i}{\delta_i^{-\frac{1}{\xi}}} d\xi
\]

(E.4)

\[
= -n \log \frac{\xi}{\xi_n} - n \log \frac{\sum_{i=1}^{\xi_n} \hat{\delta}_i^{-\frac{1}{\xi}}}{\sum_{i=1}^{\xi_n} \delta_i^{-\frac{1}{\xi}}}
\]

\[
= -n \log \frac{\xi}{\xi_n} - n \left( \frac{1}{\xi} - \frac{1}{\xi_n} \right) \log \tau_n - n \log \left[ \frac{1}{n} \sum_{i=1}^{n} w_i^{-\frac{1}{\xi}}(\hat{\theta}_n) \right].
\]

Proposition 4.3 ensures \( \frac{1}{n} \sum_{i=1}^{n} w_i^{-\frac{1}{\xi}}(\hat{\theta}_n) \xrightarrow{a.s.} \Gamma \left( \frac{\xi_0}{\xi} + 1 \right) \) uniformly. Therefore, for sufficiently large \( n \), the right-hand side of (E.4) can be further bounded as follows

\[
PL_n(\xi) - PL_n(\hat{\xi}_n) \leq - \frac{n}{2} \left[ \frac{1}{\xi} - \frac{1}{\xi_0} \right] \log \tau_n + \log \Gamma \left( \frac{\xi_0}{\xi} \right).
\]

We know from the definition of \( C_0 \) in (10) that the right-hand side of the above inequality is negative when \( \xi > C_0 \xi_n \).

For \( \xi < \xi_n \), we evaluate \( PL_n(\xi) - PL_n(\hat{\xi}_n) \) using (6):

\[
PL_n(\xi) - PL_n(\hat{\xi}_n) = -n \log \xi - n \log \left[ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-1/\xi} \right] - \frac{\xi + 1}{\xi} \sum_{i=1}^{n} \log(Y_i - \beta_n(\xi))
\]

\[
+ n \log \xi_n + n \log \left[ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{-1/\xi_n} \right] + \frac{\xi_n + 1}{\xi_n} \sum_{i=1}^{n} \log(Y_i - \hat{\beta}_n)
\]

\[
=F_n(\hat{\xi}_n) - F_n(\xi) + G_n(\xi),
\]
where $F_n$ and $G_n$ are defined in Lemma E.2 and E.3 respectively. Now combine the results from those two lemmas to get

\begin{equation}
PL_n(\xi) - PL_n(\hat{\xi}_n) \leq -n \left[ \frac{1}{2} \log \Gamma \left( \frac{\xi_0}{\xi} + 1 \right) - \log \frac{\hat{\xi}_n}{\xi} - \left( \frac{2\xi_0 \gamma}{\xi_n} - 1 \right) \left( 1 - \frac{\hat{\xi}_n}{\xi} \right) - 1 \right].
\end{equation}

Since $\hat{\xi}_n \to \xi_0$, we have $\log \frac{\xi_0}{\xi_n} > -1$, $\frac{2\xi_0 \gamma}{\xi_n} - 1 > \frac{1}{10}$ and

\begin{equation}
\left( \frac{2\xi_0 \gamma}{\xi_n} - 1 \right) \left( 1 - \frac{\hat{\xi}_n}{\xi} \right) \leq \frac{1}{10} \left( 1 - \frac{\xi_0}{2\xi} \right)
\end{equation}

for sufficiently large $n$. Therefore, the right-hand side of (E.3) can be further bounded by

\begin{equation}
PL_n(\xi) - PL_n(\hat{\xi}_n) \leq -n \left[ \frac{1}{2} \log \Gamma \left( \frac{\xi_0}{\xi} \right) - \frac{1}{2} \log \frac{\xi_0}{\xi} - \frac{1}{10} \left( 1 - \frac{\xi_0}{2\xi} \right) - 2 \right]
\end{equation}

Similarly, we know from the definition of $\hat{\xi}_n$ in (10) that the right-hand side of the above inequality is negative when $\xi < \xi_0$. \qed

\section*{E.2. Proof of relation (12)}

\textbf{Proof.} We need to show that the maximizer $(\mu_n(\xi), \tau_n(\xi))$ for $\xi \in [\xi_0, C_0\xi_0]$ is enfolded by the bounds in (10). For positive $\xi$, (5) can be expressed as,

\begin{equation}
\tau_n(\xi) = \xi \left[ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-\frac{1}{\xi}} \right]^{-\xi}.
\end{equation}

It follows immediately that

$$\tau_n(\xi) \geq \xi \left[ (Y_1(\xi) - \beta_n(\xi))^{-\frac{1}{\xi}} \right]^{-\xi} = \xi (Y_1(\xi) - \beta_n(\xi)),$$

which results in

$$\mu_n(\xi) = \beta_n(\xi) + \frac{\tau_n(\xi)}{\xi} > Y_1(\xi).$$

Since $Y_1(\xi) \to \beta_0$, we have $\mu_n(\xi) > \beta_0 - 1$, which is the lower bound in (10). In the following, we validate the upper bound for $\mu_n(\xi)$ and the bounds for $\tau_n(\xi)$ via splitting $[\xi_0, C_0\xi_0]$ into $[\xi_n, C_0\xi_0]$ and $[\xi_0, \xi_n]$.

\textbf{A. Case} $\xi \in [\xi_n, C_0\xi_0]$. Lemma E.1 implies $\hat{\beta}_n \leq \beta_0(\xi) < Y_1(\xi)$, and

$$\frac{1}{n} \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-\frac{1}{\xi}} \geq \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{-\frac{1}{\xi}},$$

which results in

$$\tau_n(\xi) \leq \xi \left[ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{-\frac{1}{\xi}} \right]^{-\xi} = \hat{\tau}_n \xi \left[ \frac{1}{n} \sum_{i=1}^{n} w_i^{-1/\xi(\hat{\theta}_n)} \right]^{-\xi}.$$

Since $\xi \in [\xi_n, C_0\xi_0]$, $\frac{1}{\xi}$ can be enclosed in $[\frac{1}{\xi_0}, \frac{2}{\xi_0}]$ for large $n$. By the uniform consistency in Proposition 4.3, there almost surely exists $N > 0$ such that for all $n > N$,

$$\frac{1}{n} \sum_{i=1}^{n} w_i^{-\frac{1}{\xi(\hat{\theta}_n)}} > \frac{1}{2} \Gamma \left( \frac{\xi_0}{\xi} + 1 \right) > 0.4$$

for $\xi \in [\xi_n, C_0\xi_0]$.  


where we utilize the fact that $\Gamma(x) > 0.8$ for all $x > 0$. Further, we have
\[
\tau_n(\xi) < \frac{\tau_n}{\xi_n} 0.4^{-\xi}.
\]
By the strong consistency of $\hat{\tau}_n$, $\frac{\tau_n}{\xi_n} < \frac{2.5\tau_n}{\xi_0} \leq 2.5\tau_0 C_0$ and $Y_{(1)} < \beta_0 + 1$ for large $n$. Thus,
\[
(\text{E.9}) \quad \tau_n(\xi) < \tau_0 C_0 2.5^{C_0 \xi_0 + 1}, \quad \text{and } \mu_n(\xi) = \beta_n(\xi) + \frac{\tau_n(\xi)}{\xi} < \beta_0 + 1 + \frac{\tau_0 C_0}{\xi_0} 2.5^{C_0 \xi_0 + 1},
\]
where we have used the relation $\beta_n(\xi) < Y_{(1)}$ when $0 < \xi < n - 1$.

**B. Case $\xi \in [c_0 \xi_0, \xi_n]$**. Denote $\kappa_0 = \min\{\frac{1}{2}, \frac{1}{2c_0}\}$, and we have $1 - \kappa_0 \xi_0 > 0$. We deduce by the power means inequality that
\[
(\text{E.10}) \quad \frac{\tau_n(\xi)}{\xi} = \left[ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{-\xi} \right]^{-\frac{1}{\xi}} \leq \left[ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{\kappa_0} \right]^{-\frac{1}{\xi_0}}.
\]
For $\xi < \xi_n$, $\beta_n(\xi) < \hat{\beta}_n$. Apply the classic $c_\nu$-inequality (D.4) twice to get
\[
\frac{1}{n} \sum_{i=1}^{n} (Y_i - \beta_n(\xi))^{\kappa_0} \leq \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{\kappa_0} + (\hat{\beta}_n - \beta_n(\xi))^{\kappa_0}
\]
and
\[
\left[ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{\kappa_0} + (\hat{\beta}_n - \beta_n(\xi))^{\kappa_0} \right]^{-\frac{1}{\xi_0}} \leq \frac{\beta_0}{\xi}\left[ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{\kappa_0} \right]^{-\frac{1}{\xi_0}} + 2\frac{1}{\xi_0} - 1 (\hat{\beta}_n - \beta_n(\xi)),
\]
where $2\frac{1}{\xi_0} - 1 > 1$. Plugging the results back in (E.10), we obtain
\[
\frac{\tau_n(\xi)}{\xi} \leq 2\frac{1}{\xi_0} - 1 \left[ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{\kappa_0} \right]^{-\frac{1}{\xi_0}} + 2\frac{1}{\xi_0} - 1 (\hat{\beta}_n - \beta_n(\xi)).
\]
In the supplementary material, we show that for $\xi \in [c_0 \xi_0, \xi_n]$, 
\[
\hat{\beta}_n - \beta_n(\xi) < \frac{4c_0 \xi_0 - \frac{1}{\xi_0}}{\tau_0 c_0^3},
\]
where $\kappa_1 = \min\{1, c_0 \xi_0\}$. In the meantime, for $n$ sufficiently large,
\[
2\frac{1}{\xi_0} - 1 \left[ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_n)^{\kappa_0} \right]^{-\frac{1}{\xi_0}} = 2\frac{1}{\xi_0} - 1 \frac{\tau_n}{\xi_n} \left[ \frac{1}{n} \sum_{i=1}^{n} w_i^{\kappa_0}(\theta_n) \right]^{-\frac{1}{\xi_0}} \leq \frac{2\frac{1}{\xi_0} \tau_0}{\xi_0} \Gamma^{\frac{1}{\xi_0}} (1 - \kappa_0 \xi_0).
\]
Consequently,
\[
\tau_n(\xi) < \frac{2\frac{1}{\xi_0} \tau_0}{\xi_0} \Gamma^{\frac{1}{\xi_0}} (1 - \kappa_0 \xi_0) + \frac{2\frac{1}{\xi_0} + \frac{1}{\xi_0} - 1 - \frac{1}{\xi_0}}{\tau_0 c_0^3},
\]
\[
\mu_n(\xi) = \beta_n(\xi) + \frac{\tau_n(\xi)}{\xi} < 2\beta_0 + \frac{2\frac{1}{\xi_0}}{\xi_0} \Gamma^{\frac{1}{\xi_0}} (1 - \kappa_0 \xi_0) + \frac{2\frac{1}{\xi_0} + 1 - \frac{1}{\xi_0}}{\tau_0 c_0^3}.
\]
Combining the previous two cases, we complete the proof of relation (12).
SUPPLEMENTARY MATERIAL

Supplement to “Uniqueness and global optimality of the maximum likelihood estimator for the generalized extreme value distribution”. The supplementary material contains additional technical results and proofs that complete the proofs in the appendices and extend the proof of relation (12) to the negative $\xi_0$.

REFERENCES

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). Statistics of extremes. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester Theory and applications, With contributions from Daniel De Waal and Chris Ferro. MR2108013

BERNARDO, J. M. (2005). Reference analysis. In Bayesian thinking: modeling and computation. Handbook of Statist. 25 17–90. Elsevier/North-Holland, Amsterdam. MR2490522

BÜCHER, A. and SEGERS, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. Extremes 20 839–872. MR3737387

CHEN, C. F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. J. Roy. Statist. Soc. Ser. B 47 540–546. MR844485

DOMBRY, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. Bernoulli 21 420–436. MR3322325

FISHER, R. A. and TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society 24 180–190.

HARTIGAN, J. A. (1983). Bayes theory. Springer Series in Statistics. Springer-Verlag, New York. MR715782

HOSKING, J. R. M., WALLIS, J. R. and WOOD, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27 251–261. MR797563

MASEND, H., RASMUSSEN, P. F. and ROSSBERG, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. Water resources research 33 747–757.

MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. J. Amer. Statist. Assoc. 95 449–485. With comments and a rejoinder by the authors. MR1803168

PICKANDS, J. III (1975). Statistical inference using extreme order statistics. Ann. Statist. 3 119–131. MR423667

SEITZ, G. (1936). Une remarque aux inégalités. Aktuárské vědy 6 167–171.

SMITH, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. Biometrika 72 67–90. MR790201

VON MISES, R. (1931). Wahrscheinlichkeitsrechnung. Springer-Verlag.