Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Inference for Local Parameters in Convexity Constrained Models

Hang Deng, Qiyang Han & Bodhisattva Sen

To cite this article: Hang Deng, Qiyang Han & Bodhisattva Sen (2023) Inference for Local Parameters in Convexity Constrained Models, Journal of the American Statistical Association, 118:544, 2721-2735, DOI: 10.1080/01621459.2022.2071721

To link to this article: https://doi.org/10.1080/01621459.2022.2071721

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2023, VOL. 118, NO. 544, 2721–2735: Theory and Methods
https://doi.org/10.1080/01621459.2022.2071721

Taylor & Francis
Taylor & Francis Group

Check for updates

Inference for Local Parameters in Convexity Constrained Models

Hang Deng[a], Qiyang Han[a], and Bodhisattva Sen[b]

[a]Department of Statistics, Rutgers University, New Brunswick, NJ; [b]Department of Statistics, Columbia University, New York, NY

ABSTRACT
In this article, we develop automated inference methods for "local" parameters in a collection of convexity constrained models based on the natural constrained tuning-free estimators. A canonical example is given by the univariate convex regression model, in which automated inference is drawn for the function value, the function derivative at a fixed interior point, and the anti-mode of the convex regression function, based on the widely used tuning-free, piecewise linear convex least squares estimator (LSE). The key to our inference proposal in this model is a pivotal joint limit distribution theory for the LS estimates of the local parameters, normalized appropriately by the length of certain data-driven linear piece of the convex LSE. Such a pivotal limiting distribution instantly gives rise to confidence intervals for these local parameters, whose construction requires almost no more effort than computing the convex LSE itself. This inference method in the convex regression model is a special case of a general inference machinery that covers a number of convexity constrained models in which a limit distribution theory is available for model-specific estimators. Concrete models include: (i) log-concave density estimation, (ii) *s*-concave density estimation, (iii) convex nonincreasing density estimation, (iv) concave bathtub-shaped hazard function estimation, and (v) concave distribution function estimation from corrupted data. The proposed confidence intervals for all these models are proved to have asymptotically exact coverage and oracle length, and require no further information than the estimator itself. We provide extensive simulation evidence that validates our theoretical results. Real data applications and comparisons with competing methods are given to illustrate the usefulness of our inference proposals. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received August 2021
Accepted April 2022

KEYWORDS
Convex least squares estimator; Convex regression; Log-concave density estimation; Pivotal limit distribution theory; Shape constraints

1. Introduction

1.1. Overview of the Inference Problems

In this article, we shall consider the problem of inference for "local" parameters in a collection of convexity constrained models, the precise meaning of which will be clear below. Here are two prototypical examples we have in mind.

Example 1.1 (Convex regression). Consider the standard nonparametric regression model:

$$Y_i = f_0(X_i) + \xi_i, \quad 1 \le i \le n, \tag{1}$$

where $f_0 : [0,1] \to \mathbb{R}$ is an unknown convex function, $X_1, \ldots, X_n$ are fixed or random design points, and ξ_i's are iid mean 0 (unobserved) errors with variance $\sigma^2 > 0$. The convex/concave regression model has been studied for more than 60 years in statistics. It was first proposed by Hildreth (1954) to solve real problems particularly in economics where, for example, demand and supply relationship is often assumed to satisfy a concavity constraint; also see Varian (1984), Matzkin (1991), and Ait-Sahalia and Duarte (2003). Here we are interested in inference for local parameters of this model, including the function value $f_0(x_0)$ and its derivative $f_0'(x_0)$ at an interior point $x_0 \in (0,1)$, and the anti-mode of f_0, that is, the smallest minimizer of f_0.

Example 1.2 (Log-concave density estimation). Suppose that we observe iid data $X_1, \ldots, X_n$ from a log-concave density $f_0 \equiv \exp(\varphi_0)$ where φ_0 is a proper concave function on $\mathbb{R}$. The class of log-concave densities has become a popular nonparametric alternative to standard parametric models due to its many desirable statistical properties; see for example, Walther (2002), Cule, Samworth, and Stewart (2010), Cule and Samworth (2010), Dümbgen and Rufibach (2009), Dümbgen, Samworth, and Schuhmacher (2011), Pal, Woodroofe, and Meyer (2007), Seregin and Wellner (2010), Kim and Samworth (2016), Kim, Guntuboyina, and Samworth (2018), Feng et al. (2021), Doss and Wellner (2016), Barber and Samworth (2020), and Han (2021) for extensive study from theoretical, methodological and application points of view. The local parameters in this model we are interested in will be the density function value $f_0(x_0)$, its derivative $f_0'(x_0)$, for $x_0 \in \mathbb{R}$, and the mode of f_0.

The common theme in these two examples is the existence of natural tuning free estimators in both problems that exhibit common distributional properties in the large sample limit. To fix ideas, let us focus on the convex regression model (1), and the canonical tuning free estimator in this model is the convex least squares estimator (LSE) $\widehat{f}_n$, defined as the convex function that minimizes the mean squared error:

CONTACT Qiyang Han qh85@stat.rutgers.edu Department of Statistics, Rutgers University, New Brunswick, NJ.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

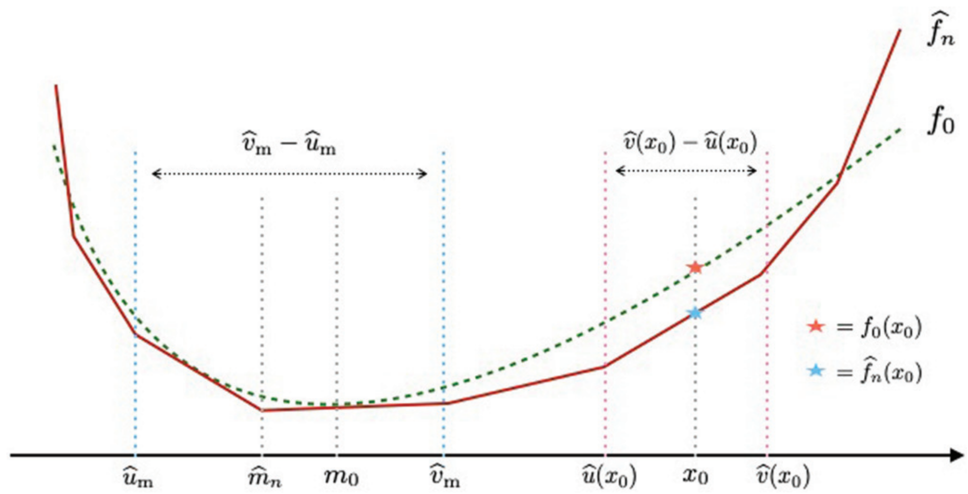


Figure 1. Figure illustration of the quantities in (2) and (3).

$$\widehat{f}_n \in \underset{f:\,\text{convex}}{\arg\min}\ \frac{1}{n}\sum_{i=1}^{n}\big(Y_i - f(X_i)\big)^2,$$

where the minimization is over all convex functions $f : \mathbb{R} \to \mathbb{R}$. Although not unique, the convex LSE $\widehat{f}_n$ has unique specification at the design points, that is, $(\widehat{f}_n(X_1), \ldots, \widehat{f}_n(X_n))^\top$ is unique. If we linearly interpolate this unique specification, the resulting piecewise linear function with kinks (i.e., points of change of slope) at design points is also unique and we treat this $\widehat{f}_n$ as the (unique) convex LSE without loss of generality. Consistency of the convex LSE $\widehat{f}_n$ is proved in Hanson and Pledger (1976). Mammen (1991) derives the pointwise convergence rate and Dümbgen, Freitag, and Jongbloed (2004) gives the uniform convergence rate of $\widehat{f}_n$. For global risk and the adaptation behavior of the convex LSE, results can be found in Chatterjee, Guntuboyina, and Sen (2015), Guntuboyina and Sen (2015), and Bellec (2018).

The hope of making inference using the convex LSE is spurred by the seminal work Groeneboom et al. (2001a), Groeneboom, Jongbloed, and Wellner (2001b) that established a joint limit distribution theory for $\big(n^{2/5}(\widehat{f}_n(x_0) - f_0(x_0)), n^{1/5}(\widehat{f}_n'(x_0) - f_0'(x_0))\big)$, when f_0 is twice continuously differentiable in a neighborhood of x_0 with $f_0''(x_0) > 0$, and the noise $\{\xi_i\}$ and the design points $\{X_i\}$ satisfy certain conditions; see Section 2.1 for a detailed technical review. One particularly unfortunate fact of this result is that it *cannot* be directly used in practice to form tuning-free confidence intervals for $f_0(x_0)$ and $f_0'(x_0)$, primarily due to the existence of the unknown second derivative $f_0''(x_0)$ in the limiting distribution theory. In fact, as the convex LSE $\widehat{f}_n$ is piecewise linear, that is, $\widehat{f}_n'' = 0$ a.e., so $f_0''(x_0)$ is not directly estimable by the LSE. Alternatively, it is tempting to look for a sample proxy of $f_0''(x_0)$ by considering, for example, kernel smoothing methods to estimate $f_0''(x_0)$; or we might consider bootstrap methods such as the m-out-of-n bootstrap and bootstrap with smoothing (Sen, Banerjee, and Woodroofe (2010), and Seijo and Sen (2011)) so that such a sample proxy can be bypassed. However, these inference approaches require careful tuning (bandwidth for smoothing and m for m-out-of-n bootstrap) that can be delicate and hard to evaluate, making them not very appealing in shape restricted problems.

1.2. Tuning-Free Inference with the Convex LSE

In this article, we introduce the first method for tuning-free inference using the convex LSE directly. The key idea is to make effective use of the length of the data-driven linear piece in the convex LSE around x_0 to "cancel out" the otherwise difficult-to-estimate second derivative. More concretely, let $[\widehat{u}(x_0), \widehat{v}(x_0)]$ be the maximal interval containing x_0 where $\widehat{f}_n$ is linear (see Figure 1). We rigorously establish a pivotal limit distribution theory (see Theorem 2.3): Under the same conditions for the limit distribution theory as in Groeneboom, Jongbloed, and Wellner (2001b),

$$\begin{pmatrix} \sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))}(\widehat{f}_n(x_0) - f_0(x_0)) \\ \sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))^3}(\widehat{f}_n'(x_0) - f_0'(x_0)) \end{pmatrix} \rightsquigarrow \sigma \cdot \begin{pmatrix} \mathbb{L}_2^{(0)} \\ \mathbb{L}_2^{(1)} \end{pmatrix}, \quad (2)$$

where $(\mathbb{L}_2^{(0)}, \mathbb{L}_2^{(1)})$ is a universal bivariate random vector, whose distribution does not depend on f_0, n, or σ. We also show that $\mathbb{L}_2^{(0)}, \mathbb{L}_2^{(1)}$ have exponentially decaying tails (see Corollary 2.11). As we may treat $\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))}$ and $\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))^3}$ in (2) as local normalizing factors for the magnitude of the standard deviation of $\widehat{f}_n(x_0) - f_0(x_0)$ and $\widehat{f}_n'(x_0) - f_0'(x_0)$, respectively, we call the normalized errors in (2) and other errors of this type the *locally normalized errors* (LNEs).

The asymptotically pivotal LNE theory in (2) can be used for inference immediately: For instance, a $1-\delta$ confidence interval (CI) for $f_0(x_0)$, based on (2), is

$$\Big[\widehat{f}_n(x_0) - \widehat{\sigma} \cdot c_\delta^{(0)} \big/ \sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))},$$
$$\widehat{f}_n(x_0) + \widehat{\sigma} \cdot c_\delta^{(0)} \big/ \sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))}\Big],$$

where $c_\delta^{(0)}$ is the $(1-\delta)$-quantile of $|\mathbb{L}_2^{(0)}|$, and $\widehat{\sigma}$ is a consistent estimator of σ. The above CI is easily seen to have asymptotically exact coverage by our theory (2). What makes this simple tuning-free CI even more attractive is that:

- Theoretically, the above CI shrinks at the oracle length suggested by the limit theory developed in Groeneboom, Jongbloed, and Wellner (2001b) which is known to be locally minimax optimal; so no other CIs can substantially beat the above CI in terms of length, at least in theory (see Theorem 2.5 for a formal statement).
- Numerically, the length of the above CI is typically very close to the oracle CI constructed using the limit distribution theory in Groeneboom, Jongbloed, and Wellner (2001b) with the knowledge of $f_0''(x_0)$. When $f_0''(x_0)$ is estimated using standard nonparametric methods, for example, local polynomial estimators with fixed or data-driven bandwidths, the resulting CI using the limit distribution theory in Groeneboom, Jongbloed, and Wellner (2001b) can suffer from severe under-coverage issues that makes it unreliable in practice unless the sample size is extremely large. This is so even in settings where estimation of $f_0''(x_0)$ is deemed relatively easy; see Section 4.3 for simulation evidence.
- Practically, our confidence interval proposal can easily handle general covariate designs, and can be extended to tackle heteroscedastic errors. See Section 2.4 for technical details, and Appendix B.5 of the Supplementary Materials for simulation results.

Another important problem in convex regression is inference for the anti-mode, defined as the smallest minimizer of f_0. It turns out that the above approach of constructing an asymptotically pivotal LNE is still applicable for this location parameter. We establish a pivotal limit distribution theory for the anti-mode as follows: Let m_0 and $\widehat{m}_n$ be the anti-mode of f_0 and $\widehat{f}_n$, respectively. Under regularity conditions on the noise variables and design points, it holds, when f_0 is twice continuously differentiable in a neighborhood of m_0 with $f_0''(m_0) > 0$, that (see Theorem 2.8)

$$\frac{1}{\widehat{v}_{\mathrm{m}} - \widehat{u}_{\mathrm{m}}}\left(\widehat{m}_n - m_0\right) \rightsquigarrow \mathbb{M}_2, \tag{3}$$

where $\widehat{u}_{\mathrm{m}}$ and $\widehat{v}_{\mathrm{m}}$ are the nearest kink points of $\widehat{f}_n$ to the left and right of $\widehat{m}_n$ (see Figure 1), and $\mathbb{M}_2$ has a pivotal distribution. What is even more striking in (3), compared to the pivotal limiting distribution in (2), is that the LNE for the anti-mode is scale-free and therefore it is not necessary to estimate σ.

As an application of our proposed CIs via (2)–(3), we consider two real datasets on mean weakly wage (Bierens and Ginther (2001)) and Belgian firm production data (Verbeek (2000)) that serve as canonical examples of concave regression. Both datasets exhibit natural heteroscedasticity which can be handled particularly easily using our inference proposals; see Section 5 for details. As will be reported therein, our CIs not only give simultaneous inference results for the "local parameters" of interest in these applications, but also provide substantial gain compared to some classical methods.

1.3. General Inference Machinery in Convexity Constrained Models

The asymptotically pivotal LNE theories developed in (2)–(3) in the convex regression model (see Example 1.1) naturally suggest a similar inference method in the log-concave density estimation model (see Example 1.2). This is so as the limiting distributional behavior of the log-concave maximum likelihood estimator (MLE) can be described in a similar way as the convex LSE. We describe in Section 3 a general inference machinery in convexity constrained models, by establishing pivotal LNE theories analogous to (2) and (3) whenever a natural tuning-free estimator exhibits similar limit behavior as that of the convex LSE. In additional to the log-concave density estimation model, other models that fall in this general machinery include

- s-concave density estimation (Dharmadhikari and Joag-Dev (1988), Koenker and Mizera (2010), and Han and Wellner (2016)),
- convex nonincreasing density estimation (Groeneboom, Jongbloed, and Wellner (2001b)),
- convex bathtub-shaped hazard function estimation (Jankowski and Wellner (2009)),
- concave distribution function estimation from corrupted data (Jongbloed and van der Meulen (2009)).

We will present detailed inference procedures in the important log-concave density estimation model in Section 3. Due to space constraints, inference details for other convexity constrained models mentioned above are relegated to Appendix A of the Supplementary Materials.

To the best of our knowledge, inference procedures with theoretical guarantees in the above models are limited to the problem of inference for the mode of log-concave densities, for which Doss and Wellner (2019) developed the likelihood ratio test (LRT). We discuss this LRT based method in Section 3 and provide a numerical performance comparison with the proposed CIs in Appendix B.6 of the Supplementary Materials.

1.4. Related Literature

The idea of constructing such an asymptotically pivotal LNE for inference in shape constrained problems was first employed in isotonic regression. Deng, Han, and Zhang (2020) establishes a pivotal limit distribution theory similar to the first line of (2), with the convex LSE $\widehat{f}_n$ replaced by the isotonic LSE $\widehat{f}_n^{(\mathrm{iso})}$, and $[\widehat{u}(x_0), \widehat{v}(x_0))$ replaced by $[\widehat{u}^{(\mathrm{iso})}(x_0), \widehat{v}^{(\mathrm{iso})}(x_0))$, the maximal interval containing x_0 where $\widehat{f}_n^{(\mathrm{iso})}$ remains constant. Compared to the result in Deng, Han, and Zhang (2020), the asymptotically pivotal LNE theory (2)–(3) demonstrates the additional advantage of convexity/concavity constraints in providing simultaneous inference for all local parameters $f_0(x_0)$, $f_0'(x_0)$, m_0. This is possible as the convexity/concavity constraints induce a natural second-order curvature condition under which sufficient information is available for all these local parameters, whereas it is not possible to infer more than $f_0(x_0)$ from the first-order monotonicity constraint as in Deng, Han, and Zhang (2020).

In a different line of work, Cai, Low, and Xia (2013) established the existence of an adaptive CI in the convex regression setting for the function value $f_0(x_0)$. Their proposal is however, purely theoretical and tailored to the specific function value setting in regression. To the best of our knowledge, the confidence intervals for local parameters of a convex regression function proposed in this article, based on the pivotal distribution theories of the LSE in (2) and (3), are the first practical tuning-free inference procedures in the literature.

1.5. Organization

The rest of the article is organized as follows. We study the local inference problem mainly through (2)–(3) and its extensions to heteroscedastic errors for convex regression in Section 2. Section 3 builds a framework for constructing the LNEs for general models under convexity/concavity constraints and applies it to the models mentioned above. Section 4 reports (approximate) critical values of $\mathbb{L}_2^{(0)}$, $\mathbb{L}_2^{(1)}$ and $\mathbb{M}_2$, investigates the numerical performance of the proposed CIs in convex regression, and makes comparisons to some competing methods. Section 5 details the application of the proposed CIs to two real datasets mentioned above. Due to space constraints, inference procedures for the remaining models not included in Section 3 are detailed in Appendix A of the Supplementary Materials. Appendix B contains a number of additional simulation studies, including: (i) numerical performance of the proposed CIs in the log-concave density estimation model, (ii) simulation results for the (modified) CIs with interval trimming, or with random design and heteroscedastic errors in convex regression, and (iii) comparison with the LRT-based CIs for mode of log-concave densities. Technical proofs are collected in the remaining appendices of the Supplementary Materials. All appendices can be found in the Supplementary Materials.

Notation. For simplicity of presentation, we write the CI $[\widehat{\theta} - c_0, \widehat{\theta} + c_0]$ which is symmetric around $\widehat{\theta}$ as $\mathcal{I} = [\widehat{\theta} \pm c_0]$. The anti-mode, or the smallest minimizer, of a convex function f is denoted by $[f]_{\mathrm{m}} = [f]_{\mathrm{m}^+}$, and the mode, or the smallest maximizer of a concave function g is denoted by $[g]_{\mathrm{m}^-}$ which equals $[-g]_{\mathrm{m}}$; see (8) for a formal definition. Let $f_0^{(k)}(\cdot)$, with $k = 1, 2, \ldots$, denote the kth derivative of $f_0(\cdot)$. We may also use $f_0^{(0)}(x_0) \equiv f_0(x_0)$ and $f_0^{(1)}(x_0) \equiv f_0'(x_0)$ interchangeably. For two real numbers a, b, $a \vee b \equiv \max\{a, b\}$, $a \wedge b \equiv \min\{a, b\}$, and $a_+ \equiv a \vee 0$, $a_- \equiv (-a) \vee 0$. The indicator function $\mathbf{1}_A(x) = \mathbf{1}_{\{x \in A\}}$ outputs 1 if $x \in A$ and 0 otherwise. We use C_x or K_x to denote a generic constant that depends only on x, whose numeric value may change from line to line unless otherwise specified. $a \lesssim_x b$ and $a \gtrsim_x b$ mean $a \leq C_x b$ and $a \geq C_x b$, respectively, and $a \asymp_x b$ means $a \lesssim_x b$ and $a \gtrsim_x b$ ($a \lesssim b$ means $a \leq Cb$ for some absolute constant C). $\mathcal{O}_{\mathbf{P}}$ and $\mathfrak{o}_{\mathbf{P}}$ denote the usual big and small O notation in probability. $\rightsquigarrow$ is reserved for weak convergence for general metric-space valued random variables. In this article we will consider weak convergence of stochastic processes in the topology induced by uniform convergence on compacta (that is, compact sets). A function f is locally C^α at x_0 if it has a continuous αth derivative in a neighborhood of x_0. Lastly, $C([a, b])$ is the class of real-valued continuous functions defined on $[a, b] \subset \mathbb{R}$.

2. Pivotal LNE Theory: Convex Regression

2.1. Review of the Limit Distribution Theory

First we state the assumptions we make about the regression model (1).

Assumption A. Suppose that $f_0 : [0, 1] \to \mathbb{R}$ is a convex function and there exists some integer $\alpha \geq 2$ such that f_0 is locally C^α at $x_0 \in (0, 1)$ with $f_0^{(\beta)}(x_0) = 0$, $\beta = 2, \ldots, \alpha - 1$, and $f_0^{(\alpha)}(x_0) \neq 0$.

Assumption A is standard in the literature, see Balabdaoui, Rufibach, and Wellner (2009). For example, if $f_0(x) = (x - 1/2)^4$, then $\alpha = 2$ for all $x \in (0, 1) \setminus \{1/2\}$ and $\alpha = 4$ for $x_0 = 1/2$. A simple Taylor's expansion of degree $\alpha - 2$ of $f_0^{(2)}(\cdot)$ at x_0 yields that α must be even and $f_0^{(\alpha)}(x_0) > 0$ (see Balabdaoui, Rufibach, and Wellner 2009, pp. 1305). The canonical and most interesting case is $\alpha = 2$.

Assumption B. Suppose the design points $\{X_i\}$ are either: (i) Equally spaced fixed points on $[0, 1]$, or (ii) iid sampled from a distribution on $[0, 1]$ with Lebesgue density $\pi(\cdot)$ that is locally continuous at x_0 with $\pi(x_0) > 0$. In the fixed design case we write $\pi(x_0) = 1$.

Assumption C. Suppose the errors $\{\xi_i\}$ are iid mean-zero with variance σ^2 and sub-Gaussian, that is, $\mathbb{E}\exp(t\xi_1^2) < \infty$ for t in a neighborhood of 0, and are independent of $\{X_i\}$ in the case of a random design.

Here we have used a perhaps stronger-than-necessary sub-Gaussianity assumption, as in Groeneboom, Jongbloed, and Wellner (2001b), on the errors to avoid unnecessary detours. The reader should however, keep in mind that our main pivotal limit distribution theory (see Theorem 2.3) will work under the same conditions that validate the proof of Theorem 2.1.

Now we state the limit distribution theory for the convex LSE $\widehat{f}_n$ due to Groeneboom, Jongbloed, and Wellner (2001b) and Ghosal and Sen (2017).

Theorem 2.1. Suppose Assumptions A–C hold. Then with $\bar{\sigma}^2(x_0) \equiv \sigma^2/\pi(x_0)$,

$$\begin{pmatrix} \big(n/\bar{\sigma}^2(x_0)\big)^{\alpha/(2\alpha+1)}\big(\widehat{f}_n(x_0) - f_0(x_0)\big) \\ \big(n/\bar{\sigma}^2(x_0)\big)^{(\alpha-1)/(2\alpha+1)}\big(\widehat{f}_n'(x_0) - f_0'(x_0)\big) \end{pmatrix} \rightsquigarrow \begin{pmatrix} d_\alpha^{(0)}(f_0, x_0) \cdot \mathbb{H}_\alpha^{(2)}(0) \\ d_\alpha^{(1)}(f_0, x_0) \cdot \mathbb{H}_\alpha^{(3)}(0) \end{pmatrix}.$$

Here $d_\alpha^{(0)}(f_0, x_0) \equiv \big(f_0^{(\alpha)}(x_0)/(\alpha + 2)!\big)^{1/(2\alpha+1)}$, $d_\alpha^{(1)}(f_0, x_0) \equiv \big(f_0^{(\alpha)}(x_0)/(\alpha + 2)!\big)^{3/(2\alpha+1)}$, and $\mathbb{H}_\alpha$ is an a.s. uniquely well-defined random continuous function satisfying the following conditions: (i) For all $t \in \mathbb{R}$, $\mathbb{H}_\alpha(t) \geq \mathbb{Y}_\alpha(t) \equiv \int_0^t \mathbb{B}(s)\, \mathrm{d}s + t^{\alpha+2}$, where $\mathbb{B}$ is the standard two-sided Brownian motion starting from 0; (ii) $\mathbb{H}_\alpha$ has a convex second derivative $\mathbb{H}_\alpha^{(2)}$; (iii) $\mathbb{H}_\alpha$ satisfies $\int_{-\infty}^{\infty} \big(\mathbb{H}_\alpha(t) - \mathbb{Y}_\alpha(t)\big)\, \mathrm{d}\mathbb{H}_\alpha^{(3)}(t) = 0$.

In words, $\mathbb{H}_\alpha$ is a.s. determined as a random continuous function with piecewise linear convex second derivative, that majorizes $\mathbb{Y}_\alpha$ with equality (touch points) taken at jumps of its third derivative. The process $\mathbb{H}_\alpha$ is called the "invelope" function of $\mathbb{Y}_\alpha$.

2.2. Pivotal LNE Theory I: Pointwise Inference for the Function and its Derivative

In this section, we consider the inference problem for the parameters $f_0(x_0)$ and $f_0'(x_0)$. We propose the following construction of CIs: Let $[\widehat{u}(x_0), \widehat{v}(x_0)]$ be the "maximal interval" containing x_0 on which $\widehat{f}_n$ is linear, and

$$\mathcal{I}_n^{(0)}(c_\delta^{(0)}) \equiv \left[\widehat{f}_n(x_0) \pm \frac{c_\delta^{(0)} \cdot \widehat{\sigma}}{\sqrt{\sum_i \mathbf{1}_{\{\widehat{u}(x_0)\le X_i \le \widehat{v}(x_0)\}}}}\right],$$

$$\mathcal{I}_n^{(1)}(c_\delta^{(1)}) \equiv \left[\widehat{f}_n'(x_0) \pm \frac{c_\delta^{(1)} \cdot \widehat{\sigma}}{\sqrt{\left(\sum_i \mathbf{1}_{\{\widehat{u}(x_0)\le X_i \le \widehat{v}(x_0)\}}\right) \cdot (\widehat{v}(x_0) - \widehat{u}(x_0))^2}}\right], \tag{4}$$

where $c_\delta^{(i)}$ $(i = 0, 1)$ are universal critical values determined only by the confidence level $1 - \delta$, and will be detailed (see Theorem 2.5). Here $\widehat{\sigma}$ is the square root of a consistent estimator of σ^2. In the fixed design case, we may also use

$$\mathcal{I}_n^{(0)}(c_\delta^{(0)}) \equiv \left[\widehat{f}_n(x_0) \pm \frac{c_\delta^{(0)} \cdot \widehat{\sigma}}{\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))}}\right],$$

$$\mathcal{I}_n^{(1)}(c_\delta^{(1)}) \equiv \left[\widehat{f}_n'(x_0) \pm \frac{c_\delta^{(1)} \cdot \widehat{\sigma}}{\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))^3}}\right]. \tag{5}$$

Remark 2.2. We require $[\widehat{u}(x_0), \widehat{v}(x_0)]$ to be the "maximal interval" which means: (i) The only interval containing x_0 if x_0 is not a kink of $\widehat{f}_n$, and (ii) the longer one (either one for equal length) if x_0 is a kink. This definition is primarily for practical concerns, as in theory any fixed point x_0 is a kink of $\widehat{f}_n$ with vanishing probability in the large sample limit.

Our proposal (4) (or (5) specific to the fixed design case) for the CIs of $f_0(x_0)$ and $f_0'(x_0)$ is based on the following asymptotically pivotal LNE theory.

Theorem 2.3. Suppose Assumptions A–C hold. Then with $\bar{\sigma}^2(x_0) = \sigma^2/\pi(x_0)$,

$$\begin{pmatrix} \sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))}\left(\widehat{f}_n(x_0) - f_0(x_0)\right) \\ \sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))^3}\left(\widehat{f}_n'(x_0) - f_0'(x_0)\right) \end{pmatrix} \rightsquigarrow \bar{\sigma}(x_0) \cdot \begin{pmatrix} \mathbb{L}_\alpha^{(0)} \\ \mathbb{L}_\alpha^{(1)} \end{pmatrix}.$$

Here $\mathbb{L}_\alpha^{(0)}$ and $\mathbb{L}_\alpha^{(1)}$ are a.s. finite random variables defined by $\mathbb{L}_\alpha^{(0)} \equiv \sqrt{h_{\alpha;-}^* + h_{\alpha;+}^*} \cdot \mathbb{H}_\alpha^{(2)}(0)$ and $\mathbb{L}_\alpha^{(1)} \equiv \sqrt{\left(h_{\alpha;-}^* + h_{\alpha;+}^*\right)^3} \cdot \mathbb{H}_\alpha^{(3)}(0)$, where $h_{\alpha;-}^*$ (resp. $h_{\alpha;+}^*$) is the absolute value of the location of the first touch point of the pair $(\mathbb{H}_\alpha, \mathbb{Y}_\alpha)$ to the left (resp. right) of 0.

In Section 2.5, we will derive an exponential tail bound for $\mathbb{H}_\alpha^{(2)}(0)$, $\mathbb{H}_\alpha^{(3)}(0)$ and $h_{\alpha;\pm}^*$ that holds uniformly in α, which immediately implies a uniform exponential tail bound for $\mathbb{L}_\alpha^{(0)}$ and $\mathbb{L}_\alpha^{(1)}$ over all α.

Remark 2.4. Theorem 2.3 can be understood via the bias-variance heuristics in connection to Theorem 2.1. For simplicity we focus on the canonical case $\alpha = 2$ and the fixed design setting. As the convex LSE $\widehat{f}_n$ around x_0 can be roughly identified as the best linear fit for the observations over the interval $[\widehat{u}(x_0), \widehat{v}(x_0)]$, the squared bias of $\widehat{f}_n$ can be "calculated" as follows: With $n_{\widehat{u},\widehat{v}} \equiv n(\widehat{v}(x_0) - \widehat{u}(x_0))$ denoting the number of samples in $[\widehat{u}(x_0), \widehat{v}(x_0)]$ and $\widehat{\Delta} \equiv \widehat{v}(x_0) - \widehat{u}(x_0)$, we have approximations (omitting x_0 in the notation) $\text{bias}^2 \approx \min_{a,b}(n_{\widehat{u},\widehat{v}})^{-1}\sum_{X_i\in[\widehat{u},\widehat{v}]}\left(f_0(X_i) - a - bX_i\right)^2 \approx \min_{a,b} \widehat{\Delta}^{-1}\int_0^{\widehat{\Delta}}\left(f_0''(\widehat{u})x^2/2 - (a - f_0(\widehat{u})) - (b - f_0'(\widehat{u}))x\right)^2 \mathrm{d}x \asymp (f_0''(\widehat{u}))^2\widehat{\Delta}^4 \approx (f_0''(x_0))^2(\widehat{v} - \widehat{u})^4$. On the other hand, the variance is roughly σ^2 normalized by the number of sample in $[\widehat{u}(x_0), \widehat{v}(x_0)]$ so is roughly $\sigma^2/n_{\widehat{u},\widehat{v}}$. Now bias-variance balance leads to the heuristic

$$(\text{bias}) \quad f_0''(x_0)(\widehat{v}(x_0) - \widehat{u}(x_0))^2 \asymp \frac{\sigma}{\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))}} \quad (\text{s.d.})$$

$$\Rightarrow \quad f_0''(x_0) \asymp \sigma/\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))^5}. \tag{6}$$

Therefore, it is reasonable to expect that, by plugging (6) into $d_2^{(i)}(f_0, x_0) (i = 0, 1)$ in Theorem 2.1, the resulting quantities will be asymptotically pivotal. Theorem 2.3 formalizes this intuition. It is however, important to note that the distribution of $(\mathbb{L}_\alpha^{(0)}, \mathbb{L}_\alpha^{(1)})$ in Theorem 2.3 is different from $(\mathbb{H}_\alpha^{(2)}(0), \mathbb{H}_\alpha^{(3)}(0))$ in Theorem 2.1, as the sample proxy $\sigma/\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))^5}$ in (6) is actually *not* a consistent estimator of $f_0''(x_0)$; it has the same order of magnitude as $f_0''(x_0)$ up to a multiplicative, universal random variable.

Although Theorem 2.3 bears certain resemblance to its "isotonic" analogue in Deng, Han, and Zhang (2020), its proof uses a completely different method. Fundamentally, this is caused by the lack of an explicit representation of the convex LSE (or more generally, convexity constrained estimators), whereas the isotonic LSE has a closed-form min–max formula. Technical complications due to the lack of such explicit formulas are well documented in convexity constrained problems (Groeneboom et al. 2001a; Groeneboom, Jongbloed, and Wellner 2001b; and Doss and Wellner 2019).

More concretely, the proof of Theorem 2.3, at a high level, proceeds via a careful application of the continuous mapping theorem, by combining the proof of Theorem 2.1 and a suitable implicit characterization of $\widehat{u}(x_0)$ and $\widehat{v}(x_0)$. Intuitively, one may wish to do so by considering $\widehat{u}(x_0)$ and $\widehat{v}(x_0)$ as two functionals $\mathcal{H}_\pm$ of the underlying process $(\mathbb{H}_n^{\text{loc}})^{(2)}$, the finite sample version of $\mathbb{H}_\alpha^{(2)}$ defined in Theorem 2.1, whose realizations are piecewise linear convex functions (see Appendix C.1 for a precise definition). However, it turns out that $\mathcal{H}_\pm$ are not continuous with respect to the topology induced by uniform convergence on compacta in which $(\mathbb{H}_n^{\text{loc}})^{(2)}$ converges weakly to $\mathbb{H}_\alpha^{(2)}$; see Equation (C.5) in the Supplementary Materials for a counterexample. To overcome this difficulty, we employ a dual characterization of $\widehat{u}(x_0), \widehat{v}(x_0)$ using both $(\mathbb{H}_n^{\text{loc}})^{(2)}$ and $\mathbb{H}_n^{\text{loc}}$ (see Equations (C.6)–(C.7) in the Supplementary Materials) that maintains suitable topological openness and closedness properties. In essence, the additional information on $\mathbb{H}_n^{\text{loc}}$ shows that the convergence of the underlying process $(\mathbb{H}_n^{\text{loc}})^{(2)}$ to its limit must occur inside the "continuity set" of the functionals $\mathcal{H}_\pm$ in the prescribed topology, and therefore $\widehat{u}(x_0)$ and $\widehat{v}(x_0)$, after proper scaling, converge in distribution to their white noise analogues. The universality of the limit then follows from Brownian scaling arguments. We provide a detailed proof for the canonical case $\alpha = 2$ in Appendix C.2 in the Supplementary Materials, while the general case follows from minor modifications (indicated in the beginning of the proof therein).

One particularly important and the canonical case is $\alpha = 2$, where the CIs in (4) have asymptotically exact coverage and shrink at the optimal rate, as detailed below. See Appendix C.3 in the Supplementary Materials for a proof of the following result.

Theorem 2.5. Suppose Assumptions A–C hold with $\alpha = 2$. Let $c_\delta^{(0)}, c_\delta^{(1)}$ be chosen such that $\mathbb{P}(|\mathbb{L}_2^{(i)}| > c_\delta^{(i)}) = \delta, i = 0, 1$. Then for any consistent variance estimator $\widehat{\sigma}$, the CIs in (4) (or (5) specific to the fixed design case) satisfy the following:

1. (exact asymptotic coverage) $\lim_n \mathbb{P}_{f_0}\big(f_0^{(i)}(x_0) \in \mathcal{I}_n^{(i)}(c_\delta^{(i)})\big) = 1 - \delta, i = 0, 1$;
2. (oracle length) For any $\varepsilon > 0$, $\liminf_n \mathbb{P}_{f_0}\big(\big|\mathcal{I}_n^{(i)}(c_\delta^{(i)})\big| < 2c_\delta^{(i)} \mathfrak{g}_\varepsilon^{(i)} \cdot \bar{\sigma}^2(x_0)/n)^{(2-i)/5} d_2^{(i)}(f_0, x_0)\big) \geq 1 - \varepsilon, i = 0, 1$. Here $\mathfrak{g}_\varepsilon^{(i)}$ $(i = 0, 1)$'s are constants that depend only on ε.

In the above theorem, $\mathfrak{g}_\varepsilon^{(i)}$ can be chosen to be a constant no smaller than the $(1-\varepsilon)$-quantile of $(h^*_{2;+} + h^*_{2;-})^{-i-1/2}$ $(i = 0, 1)$ (defined in Theorem 2.3). Note that although the above theorem focuses only on separate CIs for $f_0(x_0), f_0'(x_0)$, the joint pivotal limit theory in Theorem 2.3 allows the construction of joint confidence regions for any given functional of $f_0(x_0), f_0'(x_0)$.

Remark 2.6 (Oracle length of the proposed CIs). The lengths of the proposed CIs shrink at the optimal rates in the sense that they adapt to the oracle rates which are locally asymptotically minimax optimal as shown in Groeneboom, Jongbloed, and Wellner 2001b, Theorem 5.1. In the oracle case where $f_0''(x_0)$ and $\bar{\sigma}(x_0)$ are both known, Theorem 2.1 implies an oracle CI for $f_0^{(i)}(x_0)$ $(i = 0, 1)$ as

$$\begin{aligned}&\mathcal{I}_{n,\mathrm{ora}}^{(i)}(c_\delta(|\mathbb{H}_{(i+2)}^{(2)}(0)|))\\&\equiv \Big[\widehat{f}_n^{(i)}(x_0) \pm (\bar{\sigma}^2(x_0)/n)^{(2-i)/5} d_2^{(i)}(f_0, x_0) c_\delta\big(|\mathbb{H}_2^{(i+2)}(0)|\big)\Big],\end{aligned}$$

where $c_\delta\big(|\mathbb{H}_2^{(i+2)}(0)|\big)$ is the $(1-\delta)$-quantile of $|\mathbb{H}_2^{(i+2)}(0)|$. The length of this oracle CI shrinks at the rate $(\bar{\sigma}^2(x_0)/n)^{(2-i)/5} d_2^{(i)}(f_0, x_0)$. The proposed CI $\mathcal{I}_n^{(i)}(c_\delta^{(i)})$ in (4) has an oracle length in the sense that $|\mathcal{I}_n^{(i)}(c_\delta^{(i)})| \overset{d}{\approx} U \cdot |\mathcal{I}_{n,\mathrm{ora}}^{(i)}(c_\delta(|\mathbb{H}_{(i+2)}^{(2)}(0)|))|$ for some universal, nondegenerate and nonnegative random variable U. Our simulation results in Section 4.2 indicate that the proposed CIs actually have almost the same length as the oracle CIs above for moderate sample size in an averaged sense, so it seems reasonable to conjecture that the mean or median of U is close to 1.

Let us now consider the case when $\alpha \neq 2$. Let $c_\delta^{(0)}, c_\delta^{(1)}$ be chosen such that

$$\sup_\alpha \Big\{\mathbb{P}\big(|\mathbb{L}_\alpha^{(0)}| > c_\delta^{(0)}\big) \vee \mathbb{P}\big(|\mathbb{L}_\alpha^{(1)}| > c_\delta^{(1)}\big)\Big\} \leq \delta. \tag{7}$$

Then we may construct adaptive CIs for both $f_0(x_0)$ and $f_0'(x_0)$. We formalize this result in the following theorem; the proof is essentially the same as that of Theorem 2.5 and is thus, omitted.

Theorem 2.7. Suppose Assumptions A–C hold. Let $c_\delta^{(0)}, c_\delta^{(1)}$ be chosen as above. Then:

1. (asymptotic coverage) $\liminf_n \mathbb{P}_{f_0}\big(f_0^{(i)}(x_0) \in \mathcal{I}_n^{(i)}(c_\delta^{(i)})\big) \geq 1 - \delta, i = 0, 1$;
2. (oracle length) For any $\varepsilon > 0$, $\liminf_n \mathbb{P}_{f_0}\big(\big|\mathcal{I}_n^{(i)}(c_\delta^{(i)})\big| < 2c_\delta^{(i)} \mathfrak{g}_{\varepsilon,\alpha}^{(i)} \cdot (\bar{\sigma}^2(x_0)/n)^{(\alpha-i)/(2\alpha+1)} d_\alpha^{(i)}(f_0, x_0)\big) \geq 1-\varepsilon, i = 0, 1$. Here $\mathfrak{g}_{\varepsilon,\alpha}^{(i)}$'s (for $i = 0, 1$) are constants that depend only on ε, α, and $d_\alpha^{(i)}(f_0, x_0)$'s are defined in Theorem 2.1.

The existence of critical values $c_\delta^{(i)}$ $(i = 0, 1)$ satisfying (7) is verified in Corollary 2.11 ahead, so indeed adaptive CIs for both $f_0(x_0)$ and $f_0'(x_0)$ can be constructed by calibrating the critical values alone. Theorem 2.7 should however be primarily viewed as purely theoretical as it seems difficult to obtain a practically useful value for $c_\delta^{(i)}$ $(i = 0, 1)$ either via simulation or analytic theory.

2.3. Pivotal LNE Theory II: Inference for the Anti-mode

The above idea of constructing CIs for $f_0(x_0)$ and $f_0'(x_0)$ can be taken further to other "local parameters" for which a limit distribution theory is available. In this section we consider the inference problem for the anti-mode of the convex regression function f_0. More precisely, we define the *anti-mode* of a convex function f on $[0, 1]$ as its smallest minimizer

$$[f]_{\mathrm{m}} = [f]_{\mathrm{m}^+} \equiv \min\Big\{t : f(t) = \min_{u\in[0,1]} f(u)\Big\}. \tag{8}$$

For a concave function g, the *mode* is defined as its smallest maximizer $[g]_{\mathrm{m}^-} \equiv [-g]_{\mathrm{m}}$. We continue to use this notion of the mode for densities not necessarily convex or concave.

Let $m_0 \equiv [f_0]_{\mathrm{m}} \in (0, 1)$ be the anti-mode of f_0 and $\widehat{m}_n \equiv [\widehat{f}_n]_{\mathrm{m}}$ be the anti-mode of the convex LSE $\widehat{f}_n$. Note that $\widehat{m}_n$ is a kink point of $\widehat{f}_n$. Let $\widehat{u}_{\mathrm{m}}$ (resp. $\widehat{v}_{\mathrm{m}}$) be the first kink of $\widehat{f}_n$ to the left (resp. right) of $\widehat{m}_n$. We propose the following CI for m_0:

$$\mathcal{I}_n^{\mathrm{m}}(c_\delta^{\mathrm{m}}) \equiv \Big[\widehat{m}_n \pm c_\delta^{\mathrm{m}}\big(\widehat{v}_{\mathrm{m}} - \widehat{u}_{\mathrm{m}}\big)\Big] \cap [0, 1]. \tag{9}$$

Here c_δ^{m} is a universal critical value determined only by the confidence level $1-\delta$, to be described below (see Theorem 2.9). For finite samples, when $\widehat{m}_n$ has no kink to its left (resp. right), we simply let $\widehat{u}_{\mathrm{m}} = \widehat{m}_n$ (resp. $\widehat{v}_{\mathrm{m}} = \widehat{m}_n$). It does not affect the limit theory as either case happens with vanishing probability for $m_0 \in (0, 1)$. Note that $\widehat{v}_{\mathrm{m}} - \widehat{u}_{\mathrm{m}} > 0$ always holds unless $n = 1$.

The above proposal (9) for a CI of m_0 is based on the following asymptotically pivotal LNE theory (see Appendix C.4 in the Supplementary Materials for a proof of the following result). We will focus on the canonical case $\alpha = 2$ for simplicity of exposition.

Theorem 2.8. Suppose f_0 is locally C^2 at $m_0 \in (0, 1)$ with $f_0''(m_0) > 0$, and that Assumptions B–C hold. Then

$$(n/\sigma^2)^{1/5}\big(\widehat{m}_n - m_0\big) \rightsquigarrow d_2^{\mathrm{m}}(f_0) \cdot \big[\mathbb{H}_2^{(2)}\big]_{\mathrm{m}}, \tag{10}$$

where $d_2^{\mathrm{m}}(f_0) \equiv \big(4!/f_0''(m_0)\big)^{2/5}$. Furthermore,

$$\frac{1}{\widehat{v}_{\mathrm{m}} - \widehat{u}_{\mathrm{m}}}\big(\widehat{m}_n - m_0\big) \rightsquigarrow \mathbb{M}_2. \tag{11}$$

Here $\mathbb{M}_2$ is an a.s. finite random variable defined by $\mathbb{M}_2 \equiv [\mathbb{H}_2^{(2)}]_{\mathrm{m}}/(h^*_{2,\mathrm{m};-} + h^*_{2,\mathrm{m};+})$, where $h^*_{2,\mathrm{m};-}$ (resp. $h^*_{2,\mathrm{m};+}$) is the first kink of the random convex function $\mathbb{H}_2^{(2)}$ (defined in Theorem 2.1) to the left (resp. right) of its anti-mode $\big[\mathbb{H}_2^{(2)}\big]_{\mathrm{m}}$.

As we will mention later in Section 3, Balabdaoui, Rufibach, and Wellner (2009) proved a limit distribution theory for the mode of the MLE of log-concave densities analogous to (10). Although our proof strategy is similar to that in Balabdaoui, Rufibach, and Wellner (2009), (10) is new in convex regression. The proof of the more significant result (11) is more difficult than the proofs of Theorem 2.3 and (10). As $\widehat{u}_{\mathrm{m}}$ and $\widehat{v}_{\mathrm{m}}$ have to be characterized by processes with center $\widehat{m}_n$ that is *random*, the continuous mapping argument in the proof of Theorem 2.3 and the argmax continuous mapping argument in the proof of (10) (originally developed in Balabdaoui, Rufibach, and Wellner 2009) cannot be applied, at least directly. As a result, the weak convergence on compacta must be argued for randomly centered processes. Details of the resulting technical complications and the proof can be found in Appendix C.4 in the Supplementary Materials.

One striking difference of the CI (9) compared to (4) is the *complete elimination of the need to estimate the effective noise level* $\bar{\sigma}^2(x_0) = \sigma^2/\pi(x_0)$. This is clearly reflected in the pivotal limiting distribution for m_0 in the above theorem. The intuition is that both the quantities $\widehat{m}_n - m_0$ and $\widehat{v}_{\mathrm{m}} - \widehat{u}_{\mathrm{m}}$ have roughly the same order of magnitudes, so their ratio becomes pivotal in the limit.

As a straightforward consequence of Theorem 2.8, the CI (9) has asymptotically exact coverage and shrinks at the oracle length.

Theorem 2.9. Let c_δ^{m} be chosen such that $\mathbb{P}\big(|\mathbb{M}_2| > c_\delta^{\mathrm{m}}\big) = \delta$. Then the CI in (9) satisfies:

1. (exact asymptotic coverage) $\lim_n \mathbb{P}_{m_0}\big(m_0 \in \mathcal{I}_n^{\mathrm{m}}(c_\delta^{\mathrm{m}})\big) = 1 - \delta$;
2. (oracle length) For any $\varepsilon > 0$, $\liminf_n \mathbb{P}_{m_0}\big(\big|\mathcal{I}_n^{\mathrm{m}}(c_\delta^{\mathrm{m}})\big| < 2c_\delta^{\mathrm{m}}\mathfrak{g}_\varepsilon^{\mathrm{m}} \cdot (\bar{\sigma}^2(x_0)/n)^{1/5} d_2^{\mathrm{m}}(f_0)\big) \geq 1 - \varepsilon$. Here $\mathfrak{g}_\varepsilon^{\mathrm{m}}$ is a constant depending only on ε, and $d_2^{\mathrm{m}}(f_0)$ is defined in Theorem 2.8.

The proof of the above theorem can be found in Appendix C.5 in the Supplementary Materials.

2.4. Inference with Heteroscedastic Errors

Our results in Theorems 2.5 only require a consistent variance estimator $\widehat{\sigma}^2$. Typically, it can be very well approximated by, say, the difference-based estimators Rice (1984) and Munk et al. (2005). In a fixed design case, we can naturally extend the conclusions to heteroscedastic errors where the variance function $\sigma^2 : [0,1] \to \mathbb{R}_{>0}$ is strictly positive and continuous over $[0,1]$. Both the residual-based estimator

$$\widehat{\sigma}^2_{\mathrm{res}}(x_0) = \frac{1}{\sum_{1\leq i\leq n} \mathbf{1}_{\{\widehat{u}(x_0)\leq X_i \leq \widehat{v}(x_0)\}}} \sum_{\widehat{u}(x_0)\leq X_i \leq \widehat{v}(x_0)} (Y_i - \widehat{f}_n(x_0))^2$$

and the difference estimators (von Neumann 1941 and Munk et al. 2005), for example,

$$\widehat{\sigma}^2_{\mathrm{dif}}(x_0) = \frac{1}{2(\sum_{1\leq i\leq n} \mathbf{1}_{\{\widehat{u}(x_0)\leq X_i \leq \widehat{v}(x_0)\}} - 1)} \sum_{\widehat{u}(x_0)\leq X_{i-1}, X_i \leq \widehat{v}(x_0)} (Y_i - Y_{i-1})^2 \tag{12}$$

over the region $[\widehat{u}(x_0), \widehat{v}(x_0)]$ would be consistent for $\sigma^2(x_0)$. See, for example, Wang et al. (2008) and Shen et al. (2020) for more detailed treatments when stronger smoothness conditions are imposed on the variance function. Some simulations validating the use of (12) in convex regression with heteroscedastic errors are reported in Appendix B.5.2 in the Supplementary Materials.

2.5. A Uniform Tail Bound for the Limit Distributions

We present a result on an exponential tail bound of the limit processes in Theorem 2.1 that holds uniformly in α; see Appendix C.6 in the Supplementary Materials for its proof.

Theorem 2.10. There exist universal constants $L > 0, b > 0$ such that

$$\sup_\alpha \Big\{\mathbb{P}\big(|\mathbb{H}_\alpha^{(2)}(0)| > t\big) \vee \mathbb{P}\big(|\mathbb{H}_\alpha^{(3)}(0)| > t\big) \vee \mathbb{P}\big(h^*_{\alpha;\pm} > t\big)\Big\} \leq Le^{-t^b/L}. \tag{13}$$

Here $h^*_{\alpha;-}$ (resp. $h^*_{\alpha;+}$) is the absolute value of the location of the first touch point of the pair $(\mathbb{H}_\alpha, \mathbb{Y}_\alpha)$ to the left (resp. right) of 0.

The above result resolves an open question posed in Groeneboom et al. (2001a) concerning the existence of moments of $\mathbb{H}_2^{(2)}(0)$ (see pp. 1648 therein). In fact, the theorem above shows that all moments of $\mathbb{H}_\alpha^{(2)}(0)$ and $\mathbb{H}_\alpha^{(3)}(0)$ can be controlled uniformly in α.

As a direct consequence of Theorem 2.10, we have the following exponential tail for the limit distributions in Theorem 2.3; see Appendix C.7 in the Supplementary Materials for its proof.

Corollary 2.11. There exist universal constants $L > 0, b > 0$ such that

$$\sup_\alpha \Big\{\mathbb{P}\big(|\mathbb{L}_\alpha^{(0)}| > t\big) \vee \mathbb{P}\big(|\mathbb{L}_\alpha^{(1)}| > t\big)\Big\} \leq L\exp(-t^b/L).$$

The above corollary verifies the existence of $c_\delta^{(i)} (i = 0, 1)$ in (7) and hence the existence of adaptive CIs in Theorem 2.7.

3. Inference in Other Convex/Concave Models

3.1. General Inference Machinery

The inference methods developed in Section 2 in regression can be generalized to other convexity/concavity constrained models, in which a natural estimator (not necessarily the LSE/MLE) exhibits a nonstandard limiting distribution similar to Theorem 2.1. Concrete models include:

- log-concave density estimation (Balabdaoui, Rufibach, and Wellner 2009),
- s-concave density estimation (Han and Wellner 2016),
- convex nonincreasing density estimation (Groeneboom, Jongbloed, and Wellner 2001b),
- convex bathtub-shaped hazard function estimation (Jankowski and Wellner 2009),
- concave distribution function estimation from corrupted data (Jongbloed and van der Meulen 2009).

We develop below a general inference machinery for these convexity/concavity constrained models, and provide the specific inference procedures for the log-concave density estimation model. The detailed procedures and the theoretical results for the other models we mention above are relegated to Appendix A in the Supplementary Materials.

Before stating our most general (and abstract) result, we shall first present some heuristics. Suppose that a piecewise linear convex estimator $\widehat{g}_n$ for a convex function g_0, where g_0 is locally C^2 at x_0 with $g_0''(x_0) > 0$, satisfies the following nonstandard limit distribution theory with $(a,b) \in \mathbb{R}^2_{>0}$:

$$\begin{pmatrix} n^{2/5}\big(\widehat{g}_n(x_0) - g_0(x_0)\big) \\ n^{1/5}\big(\widehat{g}_n'(x_0) - g_0'(x_0)\big) \end{pmatrix} \rightsquigarrow \begin{pmatrix} H^{(2)}_{a,b}(0) \\ H^{(3)}_{a,b}(0) \end{pmatrix}.$$

Here $H_{a,b}$ is a.s. uniquely determined as a piecewise cubic function that majorizes a drifted integrated Brownian motion $Y_{a,b}(t) \equiv a\int_0^t \mathbb{B}(s)\,\mathrm{d}s + bt^4$, with equality taken at jumps of the piecewise constant nondecreasing function $H^{(3)}_{a,b}$, see Groeneboom et al. (2001a). For convex regression in Section 2, $a = \bar{\sigma}(x_0)$ and $b = f_0^{(2)}(x_0)/4!$. So although two nuisance parameters a, b are present in $Y_{a,b}(t)$, the really difficult nuisance parameter to estimate is b, which is typically related to the second derivative of the underlying unknown convex/concave function. This parameter cannot be estimated directly from a piecewise linear estimator $\widehat{g}_n$ as its second derivative is a.e. 0, and hence its elimination constitutes the main hurdle in the construction of valid CIs.

Inspired by the bias calculation in (6), with $[\widehat{u}(x_0), \widehat{v}(x_0)]$ denoting the maximal interval containing x_0 on which $\widehat{g}_n$ is linear, we expect $\widehat{v}(x_0) - \widehat{u}(x_0)$ to be of the same order as $(a/b)^{2/5} \cdot n^{-1/5}$ up to a universal random variable. By a Brownian scaling argument, it is easy to show that $H^{(2)}_{a,b}(0) \overset{d}{=} (a^4 b)^{1/5} \cdot \mathbb{H}_2^{(2)}(0)$, where the distribution of $\mathbb{H}_2$, as defined in Theorem 2.1, is pivotal. Consequently, we may expect that

$$\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))}\big(\widehat{g}_n(x_0) - g_0(x_0)\big) \approx a \times \text{universal random variable},$$

and similar conclusions would hold for estimators of $g_0'(x_0)$ and the mode m_0 using $\widehat{g}_n$ with appropriate normalization.

The following theorem makes the above heuristics rigorous. Its formulation essentially requires an identification of a pair of processes (H_n, Y_n) so that: (i) H_n can be viewed as a version of the double (indefinite) integral of the (properly rescaled) convex function estimator $\widehat{g}_n$ under study, and (ii) (H_n, Y_n) can be viewed as a "finite-sample" version of the (rescaled) limiting process $(\mathbb{H}_2, \mathbb{Y}_2)$ as defined in Theorem 2.1.

Theorem 3.1. Let $\widehat{g}_n$ be a piecewise linear convex (resp. concave) function estimator for an underlying convex (resp. concave) function g_0, where g_0 is locally C^2 at $x_0 \in \mathbb{R}$ with $g_0''(x_0) > 0$ (resp. $g_0''(x_0) < 0$). Suppose that there exists a pair of stochastic processes $(\widetilde{H}_n, \widetilde{Y}_n)$ on $\mathbb{R}$ such that the following hold with the convex case taking $+$ and concave case taking $-$:

(C1) $\widetilde{H}_n$ is approximately the (scaled) double integral of $\widehat{g}_n$ up to a linear indeterminacy in the following sense: Let $\bar{G}_n(t) \equiv \pm n^{2/5}\big(\widehat{g}_n(x_0 + n^{-1/5}t) - g_0(x_0) - n^{-1/5}g_0'(x_0)t\big)$. There exist tight sequences of random variables $\{A_n\}, \{B_n\}$ and a process $\{\varepsilon_n(\cdot)\}$ asymptotically vanishing on any compacta (i.e., $\sup_{|t|\le K}|\varepsilon_n(t)| \to_p 0$ for any $K > 0$), such that with $\bar{H}_n(t) \equiv (1 + \varepsilon_n(t))\int_0^t\int_0^v \bar{G}_n(u)\,\mathrm{d}u\mathrm{d}v + A_n + B_n t$, the processes $\{\bar{H}_n^{(\ell)}\}_{\ell=0}^3$ are tight on compacta and $\sup_{|t|\le K}|\widetilde{H}_n(t) - \bar{H}_n(t)| \to_p 0$ for any $K > 0$.

(C2) $(\widetilde{H}_n, \widetilde{Y}_n)$ is a "finite-sample version" of $(H_{a,b}, Y_{a,b})$, where $H_{a,b}$ is a.s. uniquely determined as a piecewise cubic function that majorizes a drifted integrated Brownian motion $Y_{a,b}(t) \equiv a\int_0^t \mathbb{B}(s)\,\mathrm{d}s + bt^4$, with equality taken at jumps of the piecewise constant nondecreasing function $H^{(3)}_{a,b}$, in the following sense: (i) $\widetilde{H}_n''$ is piecewise linearly convex, and $\widetilde{H}_n(t) \ge \widetilde{Y}_n(t)$ with equality taken at the jumps of $\widetilde{H}_n^{(3)}$; (ii) $\widetilde{Y}_n \to Y_{a,b}$ in $C([-K,K])$ for any $K > 0$.

Whenever anti-mode (resp. mode) $m_0 \in \mathbb{R}$ is concerned, suppose that g_0 is locally C^2 at m_0 with $g_0''(m_0) > 0$ (resp. $g_0''(m_0) < 0$), (C1) holds with x_0 replaced by m_0, and let $\widehat{m}_n \equiv [\widehat{g}_n]_{\mathrm{m}^{\pm}}$. Then the following hold:

1. For any $\psi_i(\cdot)$ continuously differentiable at $g_0^{(i)}(x_0)$ $(i = 0, 1)$,

$$\begin{pmatrix} n^{2/5}\big\{\psi_0\big(\widehat{g}_n(x_0)\big) - \psi_0\big(g_0(x_0)\big)\big\} \\ n^{1/5}\big\{\psi_1\big(\widehat{g}_n'(x_0)\big) - \psi_1\big(g_0'(x_0)\big)\big\} \\ n^{1/5}\big(\widehat{m}_n - m_0\big) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \pm(a^4b)^{1/5}\psi_0'\big(g_0(x_0)\big)\cdot \mathbb{H}_2^{(2)}(0) \\ \pm(a^2b^3)^{1/5}\psi_1'\big(g_0'(x_0)\big)\cdot \mathbb{H}_2^{(3)}(0) \\ (a/b)^{2/5}[\mathbb{H}^{(2)}_{2,\mathrm{m}}]_{\mathrm{m}} \end{pmatrix}. \tag{14}$$

   Here $\mathbb{H}_{2,\mathrm{m}} = \mathbb{H}_2$ if $x_0 = m_0$, and $\mathbb{H}_{2,\mathrm{m}}$ is an independent copy of $\mathbb{H}_2$ if $x_0 \neq m_0$.
2. Let $[\widehat{u}(x_0), \widehat{v}(x_0)]$ be the maximal interval containing x_0 on which $\widehat{g}_n$ is linear, and let $\widehat{u}_{\mathrm{m}}$ (resp. $\widehat{v}_{\mathrm{m}}$) be the first kink of $\widehat{g}_n$ to the left (resp. right) of $\widehat{m}_n$. Then the pivotal limit distribution theory holds:

$$\begin{pmatrix} \sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))}\big\{\psi_0\big(\widehat{g}_n(x_0)\big) - \psi_0\big(g_0(x_0)\big)\big\} \\ \sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))^3}\big\{\psi_1\big(\widehat{g}_n'(x_0)\big) - \psi_1\big(g_0'(x_0)\big)\big\} \\ \big(\widehat{m}_n - m_0\big)/\big(\widehat{v}_{\mathrm{m}} - \widehat{u}_{\mathrm{m}}\big) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \pm a\cdot\psi_0'\big(g_0(x_0)\big)\cdot\mathbb{L}_2^{(0)} \\ \pm a\cdot\psi_1'\big(g_0'(x_0)\big)\cdot\mathbb{L}_2^{(1)} \\ \mathbb{M}_2 \end{pmatrix}. \tag{15}$$

   Here $\mathbb{M}_2$ is independent of $\mathbb{L}_2^{(i)}$ if and only if $x_0 \neq m_0$. If $x_0 = m_0$, then $(\mathbb{L}_2^{(0)}, \mathbb{L}_2^{(1)}, \mathbb{M}_2)$ can be realized as a vector-valued functional of one underlying process $\mathbb{H}_2$.

3. For any consistent estimator $\widehat{a}_n$ of a, let the CIs for $\psi_0(g_0(x_0)), \psi_1(g_0'(x_0))$ and m_0 be

$$\mathcal{I}_{n,*}^{(0)}(c_\delta^{(0)}) \equiv \left[\psi_0\big(\widehat{g}_n(x_0)\big) \pm \frac{\widehat{a}_n \cdot \psi_0'\big(\widehat{g}_n(x_0)\big) \cdot c_\delta^{(0)}}{\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))}}\right],$$

$$\mathcal{I}_{n,*}^{(1)}(c_\delta^{(1)}) \equiv \left[\psi_1\big(\widehat{g}_n'(x_0)\big) \pm \frac{\widehat{a}_n \cdot \psi_1'\big(\widehat{g}_n'(x_0)\big) \cdot c_\delta^{(1)}}{\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))^3}}\right], \quad (16)$$

$$\mathcal{I}_{n,*}^{\mathrm{m}}(c_\delta^{\mathrm{m}}) \equiv \left[\widehat{m}_n \pm c_\delta^{\mathrm{m}}\big(\widehat{v}_{\mathrm{m}} - \widehat{u}_{\mathrm{m}}\big)\right], \quad (17)$$

where the critical values $c_\delta^{(i)}, c_\delta^{\mathrm{m}}$ are chosen to be the corresponding quantiles for the universal random variables $\mathbb{L}_2^{(i)}$ in Theorem 2.3 and $\mathbb{M}_2$ in Theorem 2.8. Then

- (exact asymptotic coverage) Both $\lim_n \mathbb{P}\big(\psi_i(g_0^{(i)}(x_0)) \in \mathcal{I}_{n,*}^{(i)}(c_\delta^{(i)})\big)$ for $i = 0, 1$, and $\lim_n \mathbb{P}\big(m_0 \in \mathcal{I}_{n,*}^{\mathrm{m}}(c_\delta^{\mathrm{m}})\big)$ equal $1 - \delta$.
- (oracle length) For any $\varepsilon > 0$, both $\liminf_n \mathbb{P}\big(|\mathcal{I}_{n,*}^{(i)}(c_\delta^{(i)})| < 2c_\delta^{(i)}\mathfrak{g}_\varepsilon^{(i)} \cdot n^{-(2-i)/5}(a^{4-2i}b^{1+2i})^{1/5}\psi_i'(g_0^{(i)}(x_0))\big)$ for $i = 0, 1$, and $\liminf_n \mathbb{P}\big(|\mathcal{I}_{n,*}^{\mathrm{m}}(c_\delta^{\mathrm{m}})| < 2c_\delta^{\mathrm{m}}\mathfrak{g}_\varepsilon^{\mathrm{m}} n^{-1/5}(a/b)^{2/5}\big)$ are bounded from below by $1 - \varepsilon$, where $\mathfrak{g}_\varepsilon^{(i)} (i = 0, 1)$ and $\mathfrak{g}_\varepsilon^{\mathrm{m}}$ are constants depending only on ε.

The above theorem is essentially an abstraction of the underlying proof mechanism used in Theorems 2.3 and 2.8, and its proof can be found in Appendix D.1. in the Supplementary Materials. As has been clear from (15), the crux of (1) and (2) in Theorem 3.1 is to eliminate the nuisance parameter b that is present in (14). The reason that b is the essential difficulty in inference for convexity constrained models is due to the nature of the theory (14) that holds in the regime where *the second derivative is possibly not consistently estimable*.

The conditions (C1)–(C2) are abstract and intrinsically tied to the model structure and the estimator under study. As already mentioned before Theorem 3.1, one should aim to find a process H_n such that its second derivative recovers exactly the estimator $\widehat{g}_n$, while being a finite sample version of the limit process appearing in (14). It is sometimes not possible for one process to simultaneously satisfy these two requirement, so it is often useful to find two "close enough" processes $\bar{H}_n, \widetilde{H}_n$ that serve these two purposes separately. A common way to identify the pair of processes $(\widetilde{H}_n, \widetilde{Y}_n)$ is to look at the precise KKT conditions (or "characterizations" as in, for example, (Groeneboom, Jongbloed, and Wellner 2001b, Lemma 2.2) for the specific convex function estimator. This step depends crucially on the model/sampling structures as well as how convexity enters as a constraint, so will necessarily be problem specific in nature. Once these processes are identified, tightness can typically be satisfied by establishing apriori rates of convergence for $\widehat{g}_n$ and $\widehat{g}_n'$ (e.g., Groeneboom, Jongbloed, and Wellner 2001b, Lemma 4.4).

Example 3.2. In the (relatively simple) convex regression model, the processes $(\widetilde{H}_n, \widetilde{Y}_n, \bar{H}_n)$ in Theorem 3.1 can be identified as $(\mathbb{H}_n^{\mathrm{loc}}, \mathbb{Y}_n^{\mathrm{loc}}, \widetilde{\mathbb{H}}_n^{\mathrm{loc}})$ defined in Eq. (C.2) of Appendix C.1 in the Supplementary Materials, where the slight difference between $\mathbb{H}_n^{\mathrm{loc}}$ and $\widetilde{\mathbb{H}}_n^{\mathrm{loc}}$ is due to the discretization of $[0, 1]$ by the design points $X_i = i/n$.

Remark 3.3. The choice of the functions ψ_0 and ψ_1 is geared toward the inference target. For instance, in the log concave density estimation model to be studied below, g_0 is the log density, so we will take $\psi_0 = \psi_1 = e^{(\cdot)}$ when interests lie in the density and its derivative.

3.2. Log-Concave Density Estimation

Recall the setting in Example 1.2: We observe iid data $X_1, \ldots, X_n$ from a log-concave density $f_0 \equiv \exp(\varphi_0)$ where φ_0 is a proper concave function on $\mathbb{R}$. Let $\widehat{f}_n = \exp(\widehat{\varphi}_n)$ be the log-concave MLE based on $X_1, \ldots, X_n$, that is,

$$\widehat{\varphi}_n \equiv \mathop{\arg\max}_{\varphi:\,\text{concave}, \int_{\mathbb{R}} e^{\varphi} = 1} \int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}\mathbb{F}_n(x) = \mathop{\arg\max}_{\varphi:\,\text{concave}} \left\{ \int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}\mathbb{F}_n(x) - \int_{-\infty}^{\infty} e^{\varphi(x)}\,\mathrm{d}x \right\}. \quad (18)$$

Here $\mathbb{F}_n$ is the empirical distribution function of the sample $X_1, \ldots, X_n$. It can be shown that $\widehat{\varphi}_n$ is a piecewise linear concave function with possible kinks at the data points.

The class of log-concave densities is statistically appealing due to its several nice closure properties with respect to marginalization, conditioning and convolution operations (see e.g., Saumard and Wellner 2014). The estimation of log-concave densities can be carried out using the method of maximum likelihood, and has been investigated by many authors; see Walther (2002), Cule, Samworth, and Stewart (2010), Cule and Samworth (2010), Dümbgen and Rufibach (2009), Dümbgen, Samworth, and Schuhmacher (2011), Pal, Woodroofe, and Meyer (2007), Seregin and Wellner (2010), Kim and Samworth (2016), Kim, Guntuboyina, and Samworth (2018), Feng et al. (2021), Doss and Wellner (2016), Barber and Samworth (2020), and Han (2021), just to name a few. The log-concave shape constraint also has applications in other settings; see, for example, Müller and Rufibach (2009), Samworth and Yuan (2012), Chen and Samworth (2013), and Balabdaoui and Doss (2018). We refer the reader to Saumard and Wellner (2014) and Samworth (2018) for comprehensive reviews.

Balabdaoui, Rufibach, and Wellner (2009) established the pointwise limit distribution theory, as in (14) for $\widehat{\varphi}_n$ with $a = 1/\sqrt{f_0(x_0)}, b = -\varphi_0''(x_0)/4!$ and then, by the delta method, the limit distribution theory for the log-concave MLE $\widehat{f}_n$ can be established. Now we consider the inference problem. Let $[\widehat{u}(x_0), \widehat{v}(x_0)]$ be the maximal interval containing x_0 on which $\widehat{\varphi}_n$ is linear, and let $\widehat{u}_{\mathrm{m}}$ (resp. $\widehat{v}_{\mathrm{m}}$) be the first kink of $\widehat{\varphi}_n$ to the left (resp. right) of $\widehat{m}_n$. The CIs for $f_0(x_0), f_0'(x_0)$ and the mode m_0 are given by

$$\mathcal{I}_{n,\mathrm{lc}}^{(i)}(c_\delta^{(i)}) \equiv \left[\widehat{f}_n^{(i)}(x_0) \pm \frac{\sqrt{\widehat{f}_n(x_0)} \cdot c_\delta^{(i)}}{\sqrt{n(\widehat{v}(x_0) - \widehat{u}(x_0))^{2i+1}}}\right], \qquad i = 0, 1, \quad (19)$$

$$\mathcal{I}_{n,\mathrm{lc}}^{\mathrm{m}}(c_\delta^{\mathrm{m}}) \equiv \left[\widehat{m}_n \pm c_\delta^{\mathrm{m}}\big(\widehat{v}_{\mathrm{m}} - \widehat{u}_{\mathrm{m}}\big)\right], \quad (20)$$

which are justified by the following theorem.

Theorem 3.4. Suppose f_0 is a log-concave density with $f_0 = e^{\varphi_0}$ for some concave function φ_0, $f_0(x_0) > 0$ and φ_0 is locally C^2 at x_0 and m_0 with $\varphi_0''(x_0) \vee \varphi_0''(m_0) < 0$, where $m_0 \equiv [\varphi_0]_{\mathrm{m}^-}$ is the mode of f_0. Then the conclusions in Theorem 3.1-(1)(2)

hold for concave $\widehat{g}_n = \widehat{\varphi}_n$, and the CIs in (19)–(20) satisfy the conclusions in Theorem 3.1-(3) with $\psi_0 \equiv e^{(\cdot)}$, $\widehat{a}_n = 1/\sqrt{\widehat{f}_n(x_0)}$.

Doss and Wellner (2019) developed a different procedure for inference of the mode m_0 based on the likelihood ratio test (LRT). More specifically, consider the hypothesis testing problem:

$$H_0 : [\varphi_0]_{m^-} = m_0 \qquad \text{versus} \qquad H_1 : [\varphi_0]_{m^-} \neq m_0.$$

Let $\widehat{f}_{n,0}$ be the mode-constrained log-concave MLE, that is, $\widehat{f}_{n,0} = e^{\widehat{\varphi}_{n,0}}$, where $\widehat{\varphi}_{n,0}$ is defined by $\widehat{\varphi}_{n,0} \equiv \arg\max\{\int_{-\infty}^{\infty} \varphi(x)\, d\mathbb{F}_n(x) - \int_{-\infty}^{\infty} e^{\varphi(x)}\, dx\}$ with the argmax running over all concave functions φ with $\varphi(m_0) \geq \varphi(x)$ for all $x \in \mathbb{R}$. The LRT statistic is now defined as

$$2\log\lambda_n(m_0) \equiv 2n\mathbb{P}_n\big(\log\widehat{f}_n - \log\widehat{f}_{n,0}\big) = 2n\mathbb{P}_n\big(\widehat{\varphi}_n - \widehat{\varphi}_{n,0}\big), \tag{21}$$

where $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$ is the empirical measure based on iid observations $X_1, \ldots, X_n$. Doss and Wellner (2019) proved the following result: Under the same conditions as in Theorem 3.4,

$$2\log\lambda_n(m_0) \rightsquigarrow \mathbb{K}, \tag{22}$$

where $\mathbb{K}$ has a universal limiting distribution. A CI for m_0 can then be obtained by inverting the above LRT statistic: Let

$$\mathcal{I}_{n,\mathrm{lc}}^{(m),\mathrm{DW}}(d_\delta) \equiv \{m_0 : 2\log\lambda_n(m_0) \leq d_\delta\}, \tag{23}$$

where d_δ is chosen such that $\mathbb{P}(\mathbb{K} > d_\delta) = \delta$. Then $\lim_n \mathbb{P}_{m_0}\big(m_0 \in \mathcal{I}_{n,\mathrm{lc}}^{(m),\mathrm{DW}}(d_\delta)\big) = \mathbb{P}\big(\mathbb{K} \leq d_\delta\big) = 1-\delta$. It is easy to see that the implementation of (23) requires the computation of many mode-constrained log-concave MLEs, whereas our proposed CI (20) only requires the computation of the log-concave MLE once. A detailed numerical comparison between the Doss-Wellner CI (23) and our proposed CI (20) is conducted in Appendix B.6 in the Supplementary Materials.

4. Numerical Experiments

4.1. Approximated Limiting Distributions

We report in Table 1 some important quantiles of these empirical distributions as the approximate corresponding critical values $c_\delta(\mathbb{T})$, defined by $\mathbb{P}\{\mathbb{T} > c_\delta(\mathbb{T})\} = \delta$ for $\mathbb{T} \in \{\mathbb{L}_2^{(0)}, \mathbb{L}_2^{(1)}, \mathbb{M}_2\}$.

We give in Table 2 some absolute sample quantiles which approximate the corresponding critical values $c_\delta(|\mathbb{T}|)$ for $\mathbb{T} \in \{\mathbb{L}_2^{(0)}, \mathbb{L}_2^{(1)}, \mathbb{M}_2\}$. They are used to construct the symmetric CIs (e.g., in (4) and (9)).

Finally, as we will later compute the lengths of the oracle CIs (with derivatives of f_0 known), we shall simulate the quantiles of $\mathbb{H}_2^{(2)}(0)$, $\mathbb{H}_2^{(3)}(0)$ and $[\mathbb{H}_2^{(2)}]_{\mathrm{m}}$ in Theorems 2.1 and 2.8 (10). They can be conveniently obtained as byproducts when we simulate the critical values of $\mathbb{L}_2^{(0)}$, $\mathbb{L}_2^{(1)}$ and $\mathbb{M}_2$; see Table 3 for the approximate quantiles.

Details of the simulation methods used in the above tables can be found in Appendix B.1 of the Supplementary Materials.

Table 1. Approximate quantiles of $\mathbb{L}_2^{(0)}$, $\mathbb{L}_2^{(1)}$ and $\mathbb{M}_2$.

δ	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010
$c_\delta(\mathbb{L}_2^{(0)})$	−2.60	−2.04	−1.62	−1.19	0.04	1.40	1.82	2.20	2.65
$c_\delta(\mathbb{L}_2^{(1)})$	−11.88	−9.03	−6.78	−4.54	0.00	4.55	6.77	8.96	11.87
$c_\delta(\mathbb{M}_2)$	−0.83	−0.60	−0.47	−0.35	0.00	0.35	0.47	0.60	0.82

Table 2. Approximate critical values of $|\mathbb{L}_2^{(0)}|$, $|\mathbb{L}_2^{(1)}|$ and $|\mathbb{M}_2|$.

δ	0.50	0.20	0.10	0.05	0.02	0.01
$c_\delta(\lvert\mathbb{L}_2^{(0)}\rvert)$	0.65	1.30	1.73	2.13	2.63	2.98
$c_\delta(\lvert\mathbb{L}_2^{(1)}\rvert)$	1.72	4.55	6.77	9.00	11.87	14.03
$c_\delta(\lvert\mathbb{M}_2\rvert)$	0.19	0.35	0.47	0.60	0.82	1.06

Table 3. Approximate critical values of $|\mathbb{H}_2^{(2)}(0)|$, $|\mathbb{H}_2^{(3)}(0)|$ and $\big|[\mathbb{H}_2^{(2)}]_{\mathrm{m}}\big|$.

δ	0.50	0.20	0.10	0.05	0.02	0.01
$c_\delta(\lvert\mathbb{H}_2^{(2)}(0)\rvert)$	0.89	1.69	2.16	2.58	3.08	3.43
$c_\delta(\lvert\mathbb{H}_2^{(3)}(0)\rvert)$	4.27	7.77	9.64	11.13	12.71	13.69
$c_\delta(\lvert[\mathbb{H}_2^{(2)}]_{\mathrm{m}}\rvert)$	0.18	0.32	0.40	0.46	0.53	0.57

4.2. Numerical Performance for the CIs in Convex Regression

We report in this section the numerical performance of the proposed CIs in convex regression; simulation results of a similar flavor in the log-concave density estimation model is deferred to Appendix B.3 of the Supplementary Materials. The following results mainly serve as numerical support of Theorems 2.5 and 2.9, showing that: (i) the corresponding proposed CIs have asymptotically accurate coverage, and (ii) their lengths adapt to oracle rates (see Remark 2.6).

Suppose we observe in convex regression data $\{(X_i, Y_i), 0 \leq i \leq n\}$ with $f_0(x) = 20 - 20\sqrt{1-(x-0.5)^2}$, $X_i = i/n$ and $\xi_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 = 1)$. The goal is to construct 95% CIs for the function value $f_0(x_0)$, derivative value $f_0'(x_0)$ at $x_0 = 0.5$ and the anti-mode $m_0 = x_0 = 0.5$. The noise level $\sigma^2 > 0$ can be very well estimated by the difference estimators Rice (1984) and Munk et al. (2005); here we use $\widehat{\sigma}^2 = \sum_{i=1}^{n-1}(Y_{i+1} + Y_{i-1} - 2Y_i)^2/(6(n-1))$ whenever variance estimation is needed.

We apply the support reduction algorithm implemented in the R function `conreg` from package `cobs` to compute the convex LSE and construct the 95% CIs defined in (5) and (9) with approximate critical values in Table 2. Simulated coverage probabilities are reported in Figure 2(a) with 10^4 repetitions. Boxplots of the lengths of these 10^4 CIs for each of $\{f_0(x_0), f_0'(x_0), m_0\}$ are reported in Figures 2(b)–(d), along with the oracle CI lengths in red dashed lines. By Theorem 2.1 and (10) in Theorem 2.8, the symmetric oracle CIs are: $\big[\widehat{f}_n^{(i)}(x_0) \pm (f_0^{(2)}(x_0)/24)^{1/5}(n/\sigma^2)^{-(2-i)/5}c_\delta(|\mathbb{H}_2^{(2+i)}(0)|)\big]$ for $f_0^{(i)}(x_0)$ with $i = 0, 1$ and $\big[\widehat{m}_n \pm (24/f_0^{(2)}(m_0))^{2/5}(n/\sigma^2)^{-1/5}c_\delta(|[\mathbb{H}_2^{(2)}]_{\mathrm{m}}|)\big]$ for m_0.

As in Figure 2(a), all CIs for the local parameters have rather accurate coverage and the convergence of coverage probabilities is approximately achieved for sample size as small as $n = 100$; most coverage errors deviate from the nominal coverage by less than 0.01. In terms of length, Figures 2(b)–(d) show that the

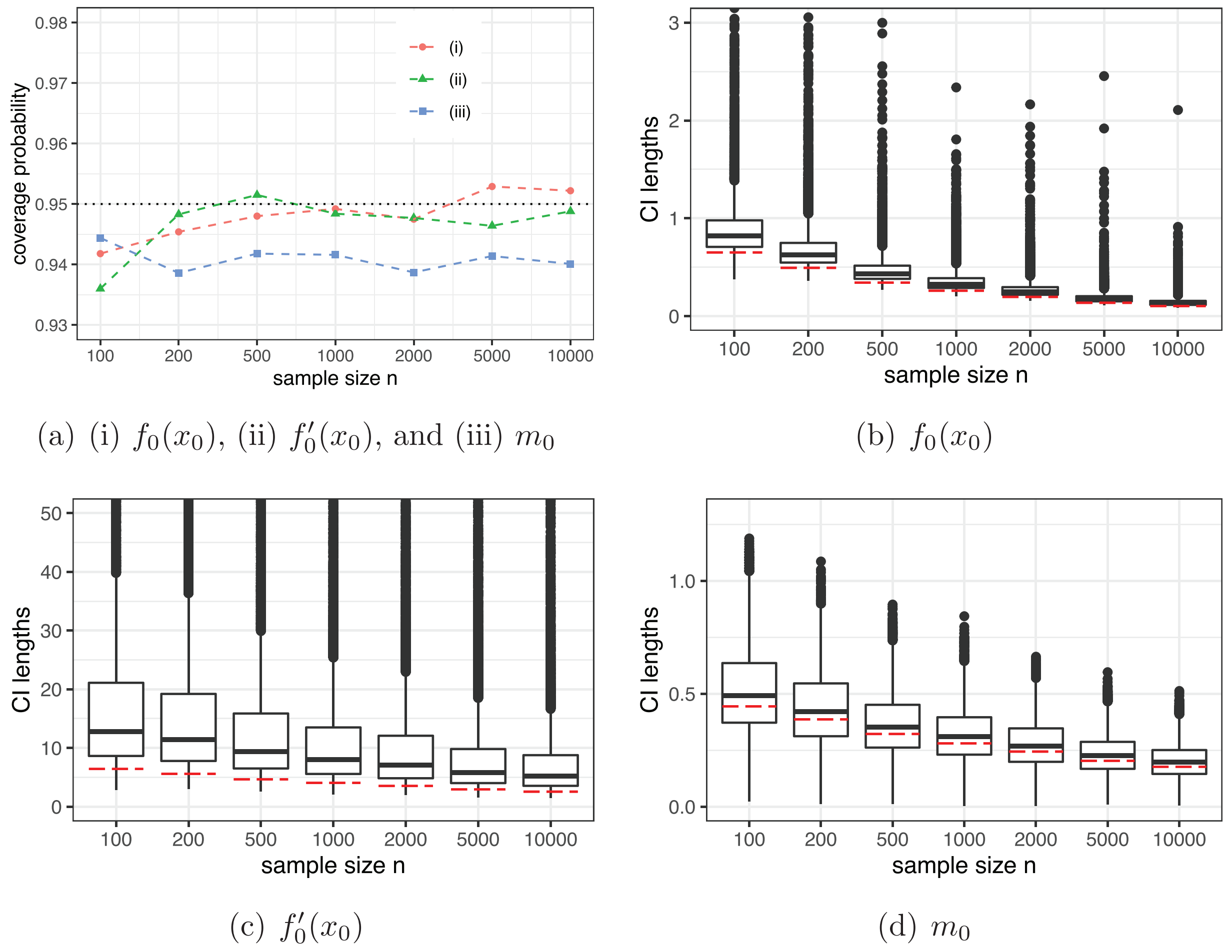


Figure 2. Plot of the simulated coverage probabilities and boxplots of the lengths of the proposed CIs for corresponding local parameters in convex regression. Here $f_0(x) = 20 - 20\sqrt{1-(x-0.5)^2}$, $x_0 = 0.5$ and anti-mode $m_0 = 0.5$. The red dashed lines in boxplots (b)–(d) represent the lengths of the oracle CIs.

lengths of the proposed CIs shrink at the same rate with those of the oracle CIs. Note that when $[\widehat{u}(x_0), \widehat{v}(x_0)]$ is short the proposed CI for $f_0(x_0), f_0'(x_0)$ may become quite wide; so we observe relatively more outliers on the CIs for $f_0(x_0)$ and $f_0'(x_0)$ than for m_0. A simple solution in practice would be forcing $\widehat{v}(x_0) - \widehat{u}(x_0)$ to be no smaller than a small value: As the interval length has asymptotic order $n^{-1/5}$, we find that replacing $\widehat{v}(x_0) - \widehat{u}(x_0)$ with $\max\{\widehat{v}(x_0) - \widehat{u}(x_0), n^{-3/10}\}$ removes extreme CIs significantly and maintain satisfactory coverage probabilities. See our simulation results in Appendix B.4 of the Supplementary Materials.

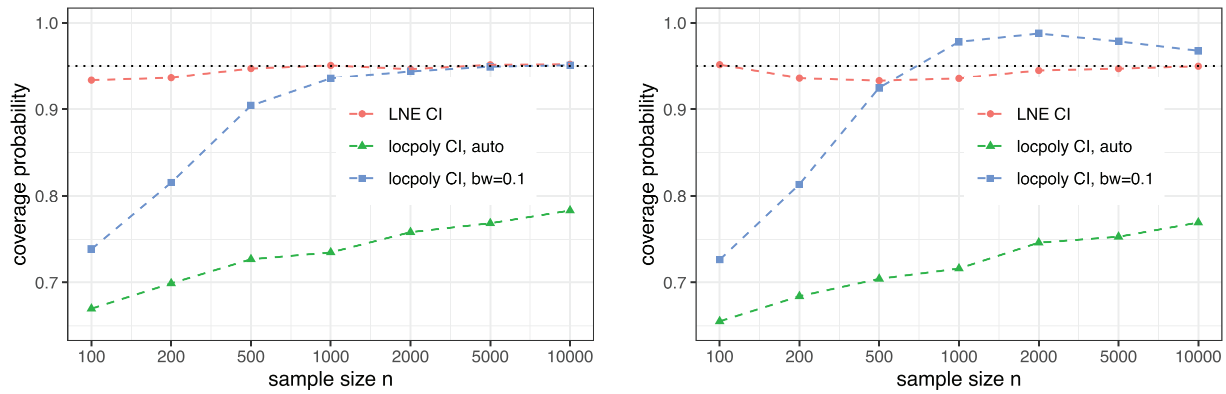
For the performance of the proposed CIs at different design points, rather than at one single point $x_0 = 0.5$, see Appendix B.2 of the Supplementary Materials. We observe that, other than at points near the boundary where the limit distribution theory is unlikely to hold, the coverage and length behaviors of the proposed CIs are similar to what we observe in this section for $x_0 = 0.5$.

4.3. Comparison to CIs with Estimated Second Derivative

While the goal of the article is to propose and study tuning-free CI construction procedures, it is natural to wonder how they compare with the CIs constructed with smoothing methods. Here we consider convex regression. Based on Theorem 2.1, we may use, for example, a local polynomial estimator $\widehat{f_0^{(2)}}(x_0)$ to estimate the second derivative $f_0^{(2)}(x_0)$ and construct CIs for $f_0(x_0)$ and $f_0'(x_0)$ as

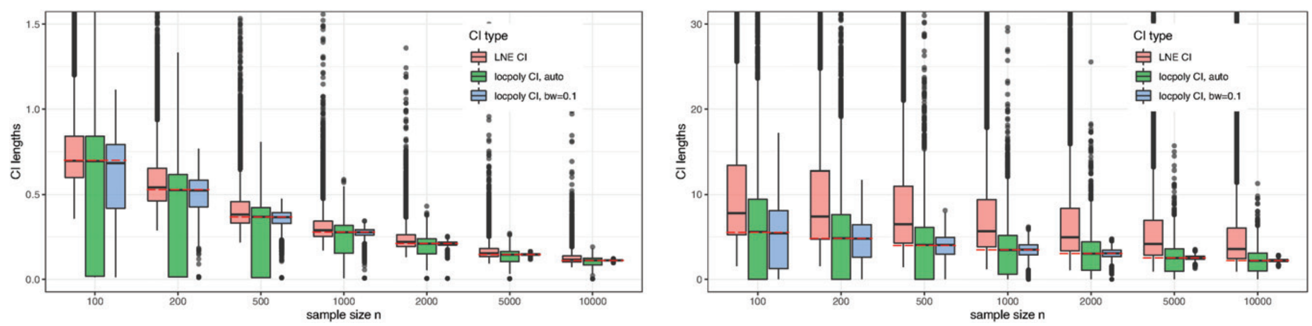
$$\mathcal{I}^{(i)}_{n,\text{locpoly}} \equiv \left[\widehat{f}^{(i)}_n(x_0) \pm c_{\delta,i}\left(\widehat{f_0^{(2)}(x_0)}/4!\right)^{(2i+1)/5} \Big/ \left(n/\widehat{\sigma}^2\right)^{(2-i)/5}\right], \quad i = 0, 1, \qquad (24)$$

where $c_{\delta,0}$ (resp. $c_{\delta,1}$) is the $(1-\delta)$-quantile of $\left|\mathbb{H}^{(2)}_\alpha(0)\right|$ (resp. $\left|\mathbb{H}^{(3)}_\alpha(0)\right|$). We let $f_0(x) = e^{2x}$ and continue to use the simulation setting from Section 4.2. Note that we make this choice of a smooth and highly regular f_0 in favor of estimation of its second derivative. To compute $\widehat{f_0^{(2)}}(x_0)$, we use the R function `locpoly` from package `KernSmooth`. For bandwidth selection, we consider a fixed bandwidth 0.1 and a data-driven bandwidth selector, proposed by Ruppert, Sheather, and Wand (1995) and implemented as function `dpill` in the same package. The simulated coverage probabilities of the proposed CIs at $x_0 = 0.5$, referred to as *LNE CIs*, and the CIs with estimated second derivative, referred to as *locpoly CIs, bw=0.1* for fixed bandwidth 0.1 and *locpoly CIs, auto* for bandwidth determined by the selector `dpill`, are all given in Figure 3. The boxplots of the CI lengths are reported in Figure 4.



(a) function value $f_0(x_0)$ (b) derivative value $f_0'(x_0)$

Figure 3. Simulated coverage probabilities of the proposed 95% CIs (*LNE CI*) and the 95% CIs with estimated second derivative (*locpoly CI*) for corresponding parameters in convex regression. Here $f_0(x) = 20 - 20\sqrt{1-(x-0.5)^2}$ and $x_0 = 0.5$.



(a) function value $f_0(x_0)$ (b) derivative value $f_0'(x_0)$

Figure 4. Boxplots of the lengths of the LNE and locpoly CIs for corresponding parameters in convex regression with $f_0(x) = e^{2x}$. Here "auto" means the bandwidth is chosen by the selector "dpill" (proposed by Ruppert, Sheather, and Wand 1995) and "bw=0.1" means the bandwidth is a constant 0.1 for all n. Red dashed lines represent the lengths of oracle CIs.

From these simulation results, it seems that the proposed LNE CIs produce the most stable coverage for all considered sample sizes. The bandwidth selector by Ruppert, Sheather, and Wand (1995) introduces a severe under-coverage and the resulting CIs seem not reliable enough for practical use for even large sample size n, say, $n = 10{,}000$. The locpoly CIs with fixed bandwidth 0.1 produce comparable coverage when sample size $n \geq 2000$ for function value, while they are less stable than the LNE CIs for derivative estimation. As such, even in settings where estimation of second derivatives is deemed easy, these simulation results still clearly show the advantage of our proposed LNE CIs in terms of stable coverage for all considered sample sizes while enjoying similar lengths, compared to either fixed and data-driven bandwidth selection methods.

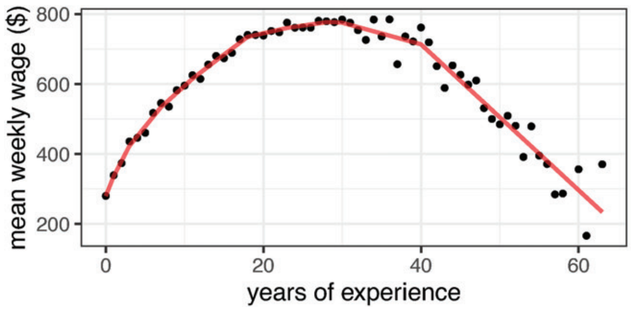
5. Real Data Analysis

5.1. Mean Weekly Wage Inference

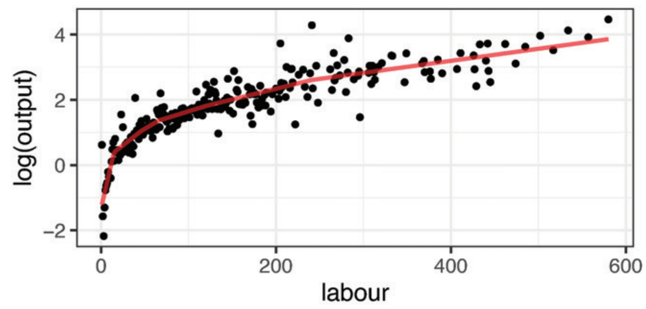
We apply the proposed inference procedures to facilitate understanding of the relationship between mean weekly wage and years of potential work experience based on the 1988 March Current Population Survey data. The dataset can be accessed as `ex1029` in R package `Sleuth2` and was first studied in Bierens and Ginther (2001) for different purposes. After cleaning, the dataset contains a sample of 25,437 men aged 18–70 with positive annual income greater than \$50 in 1992 after deflated by the deflator of Personal Consumption Expenditure for 1992.

We consider the relationship between mean weekly wage and years of experience. Let N_i be the number of males with x_i years of experience and $y_{i,j}$ with $1 \leq j \leq N_i$ the weekly wage of the jth male with x_i years of experience. Our model for the mean weekly wage and years of experience is $y_{i,j} = f(x_i) + \varepsilon_{i,j}$, where $x_i \in \{0, 1, \ldots, 63\} \setminus \{59, 62\}$. We assume $\varepsilon_{i,j}$ has the same variance for fixed i and $f(\cdot)$ is concave. The concavity constraint for $f(\cdot)$ is in line with our common sense: Years of experience is particularly important in early career but has decreasing positive impact until weekly wage hits the highest level; after that, the earning will decline as years of experience increases.

To fit into the standard nonparametric regression model as in (1), we let $y_i = \text{Ave}(y_{i,j},\ 1 \leq j \leq N_i)$ be the averaged weekly wage for x_i years of experience, so it follows that $y_i = f(x_i) + \xi_i$ where error $\xi_i = \text{Ave}(\varepsilon_{i,j},\ 1 \leq j \leq N_i)$ can vary for different i. In Figure 5, we plot the mean weekly wage data against years of experience and the fitted concave LSE, which clearly fits the data very well. General inference for heteroscedastic errors is discussed in Section 2.4. Essentially, inference procedures will not change as long as we have a good estimate for local variance. Here we have $\{\varepsilon_{i,j},\ \forall j\}$ to estimate the noise level of



(a) mean weekly wage data



(b) labor demand data

Figure 5. Plots of the real data and their concave LSEs.

ξ_i, that is, we use the variance estimator $\widehat{\sigma}_i^2 = \frac{1}{N_i-1}\sum_{j=1}^{N_i}(y_{i,j} - \frac{1}{N_i}\sum_{j=1}^{N_i} y_{i,j})^2/N_i$.

We first carry out point inference at $x = 8$, that is, we are interested in 95% CIs for the mean weekly wage and its growth rate for people with eight years of experience. The concave LSE for mean weekly wage and its growth rate are \$554.00 and \$22.15/year. Following the proposed procedures in (4), we can easily calculate the 95% CI for mean weekly wage as [\$542.70, \$565.32] and its growth rate as [\$12.74/year, \$31.56/year]. An interesting comparison can be made by also looking at the classical CI from t-test using sample $\{y_{i,j},\ 1 \le j \le N_i \text{ and } x_i = 8\}$, which is [\$511.97, \$558.60]. It is obvious that the convexity constraint helps us construct a much shorter CI, reducing the length by more than 50%, from \$46.62 to \$22.63.

Finally, we want to have a interval of high confidence (95%) for the number of years of experience to reach the highest mean weekly wage, which translates to a 95% CI for the mode/maximizer of the function $f(\cdot)$. A very simple calculation based on (9) will give us the 95% CI as [27.8 years, 30.2 years].

5.2. Labor Demand Inference

We apply the inference procedures to conduct labor demand inference using a dataset that contains information for 569 Belgian firms in 1996 on the total number of employees (labor) and a measure of output (value added in million euro). Here we intend to understand the relationship between output and the number of employees and thus the labor demand. We adopt the convention of taking the logarithm of output and assume y : log(output) $= f(x$: number of employees) satisfies a concavity constraint. It is a well-acknowledged assumption in practice as the extra value added to the company by hiring more employees will generally decrease as the company grows. See Figure 5(b) for the plot of the data and the fitted concave LSE.[1] To derive a CI for the log(output) and its growth rate at labor size 200, we estimate the local variance using difference estimator (12) and apply the procedure in (4). We obtain the 95% CI for log(output) as $[2.21, 2.44]$ and the 95% CI for the growth rate as $[0, 0.01]$.

[1]There were some ties in the covariate values; we added noise to break the ties.

6. Conclusion

In this article, we developed the first fully automated inference method for local parameters in the univariate convex regression model, based on the widely used tuning-free convex LSE. The key idea in our inference proposal is to make effective use of the length of certain data-driven linear piece in the convex LSE, to obtain a pivotal limiting distribution for the "locally normalized errors" (LNEs) that cancels out the otherwise difficult-to-estimate second derivative (impossible to estimate by the convex LSE directly). This inference method in convex regression using the convex LSE extends to other convexity constrained models, in which a natural tuning-free estimator exhibits a nonstandard limit distribution. Notably, inference for local parameters in the popular log-concave density estimation model can be carried out immediately using the standard log-concave MLE.

Finally we mention some open problems related to the topic of this article: (i) In the regression setting with a random design, it is would be interesting to extend our results to allow for the case that $\mathbb{E}[\xi_i|X_i] = 0$ and $\mathbb{E}[\xi_i^2|X_i] = \sigma^2(X_i)$; (ii) it would be interesting to develop fully automated inference method for the maximum/minimum of a concave/convex function in addition to its location considered in this article; (iii) the construction of global confidence regions/bands for the unknown convex regression functions/log-concave density, etc. is still open(we conjecture that an L_∞-aggregated version of Theorems 2.3/3.4 may lead to a tuning-free construction of confidence bands); (iv) developing tuning-free inference methods in multivariate convexity constrained models is another interesting problem. At this point, a limit distribution theory for convex function estimators in any of these models in multiple dimensions is still lacking.

Supplementary Materials

The supplementary material, which is available online, contains proofs of our main results and additional simulation results.

Funding

Hang Deng is supported by NSF grants DMS-1451817 and CCF-1934924. Qiyang Han is supported by NSF grants DMS-1916221 and DMS-2143468. Bodhisattva Sen is supported by NSF grant DMS-2015376.

References

Ait-Sahalia, Y., and Duarte, J. (2003), "Nonparametric Option Pricing Under Shape Restrictions," *Journal of Econometrics*, 116, 9–47. [2721]

Balabdaoui, F., and Doss, C. R. (2018), "Inference for a Two-Component Mixture of Symmetric Distributions Under Log-Concavity," *Bernoulli*, 24, 1053–1071. [2729]

Balabdaoui, F., Rufibach, K., and Wellner, J. A. (2009), "Limit Distribution Theory for Maximum Likelihood Estimation of a Log-Concave Density," *The Annals of Statistics*, 37, 1299–1331. [2724,2727,2728,2729]

Barber, R. F., and Samworth, R. J. (2020), "Local Continuity of Log-Concave Projection, with Applications to Estimation Under Model Misspecification," arXiv preprint arXiv:2002.06117. [2721,2729]

Bellec, P. C. (2018), "Sharp Oracle Inequalities for Least Squares Estimators in Shape Restricted Regression," *The Annals of Statistics*, 46, 745–780. [2722]

Bierens, H. J., and Ginther, D. K. (2001), "Integrated Conditional Moment Testing of Quantile Regression Models," *Empirical Economics*, 26, 307–324. [2723,2732]

Cai, T. T., Low, M. G., and Xia, Y. (2013), "Adaptive Confidence Intervals for Regression Functions Under Shape Constraints," *The Annals of Statistics*, 41, 722–750. [2723]

Chatterjee, S., Guntuboyina, A., and Sen, B. (2015), "On Risk Bounds in Isotonic and Other Shape Restricted Regression Problems," *The Annals of Statistics*, 43, 1774–1800. [2722]

Chen, Y., and Samworth, R. J. (2013), "Smoothed Log-Concave Maximum Likelihood Estimation with Applications," *Statistica Sinica*, 23, 1373–1398. [2729]

Cule, M., and Samworth, R. (2010), "Theoretical Properties of the Log-Concave Maximum Likelihood Estimator of a Multidimensional Density," *Electronic Journal of Statistics*, 4, 254–270. [2721,2729]

Cule, M., Samworth, R., and Stewart, M. (2010), "Maximum Likelihood Estimation of a Multi-Dimensional Log-Concave Density," *Journal of the Royal Statistical Society*, Series B, 72, 545–607. [2721,2729]

Deng, H., Han, Q., and Sen, B. (2022), "Supplement to 'Inference for Local Parameters in Convexity Constrained Models'," *Journal of the American Statistical Association*.

Deng, H., Han, Q., and Zhang, C.-H. (2020), "Confidence Intervals for Multiple Isotonic Regression and Other Monotone Models," *The Annals of Statistics* (to appear). Available at arXiv:2001.07064. [2723,2725]

Dharmadhikari, S., and Joag-Dev, K. (1988), *Unimodality, Convexity, and Applications*, Probability and Mathematical Statistics, Boston, MA: Academic Press. [2723]

Doss, C. R., and Wellner, J. A. (2016), "Global Rates of Convergence of the MLEs of Log-Concave and *s*-concave Densities," *The Annals of Statistics*, 44, 954–981. [2721,2729]

——— (2019), "Inference for the Mode of a Log-Concave Density," *The Annals of Statistics*, 47, 2950–2976. [2723,2725,2730]

Dümbgen, L., Freitag, S., and Jongbloed, G. (2004), "Consistency of Concave Regression with an Application to Current-Status Data," *Mathematical Methods of Statistics*, 13, 69–81. [2722]

Dümbgen, L., and Rufibach, K. (2009), "Maximum Likelihood Estimation of a Log-Concave Density and its Distribution Function: Basic Properties and Uniform Consistency," *Bernoulli*, 15, 40–68. [2721,2729]

Dümbgen, L., Samworth, R., and Schuhmacher, D. (2011), "Approximation by Log-Concave Distributions, with Applications to Regression," *The Annals of Statistics*, 39, 702–730. [2721,2729]

Feng, O. Y., Guntuboyina, A., Kim, A. K. H., and Samworth, R. J. (2021), "Adaptation in Multivariate Log-Concave Density Estimation," *The Annals of Statistics*, 49, 129–153. [2721,2729]

Ghosal, P., and Sen, B. (2017), "On Univariate Convex Regression," *Sankhya A*, 79, 215–253. [2724]

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001a), "A Canonical Process for Estimation of Convex Functions: The "invelope" of Integrated Brownian Motion $+t^4$," *The Annals of Statistics*, 29, 1620–1652. [2722,2725,2727,2728]

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001b), "Estimation of a Convex Function: Characterizations and Asymptotic Theory," *The Annals of Statistics*, 29, 1653–1698. [2722,2723,2724,2725,2726,2728,2729]

Guntuboyina, A., and Sen, B. (2015), "Global Risk Bounds and Adaptation in Univariate Convex Regression," *Probability Theory and Related Fields*, 163, 379–411. [2722]

Han, Q. (2021), "Set Structured Global Empirical Risk Minimizers are Rate Optimal in General Dimensions," *The Annals of Statistics* (to appear). Available at arXiv:1905.12823. [2721,2729]

Han, Q., and Wellner, J. A. (2016), "Approximation and Estimation of *s*-concave Densities via Rényi Divergences," *The Annals of Statistics*, 44, 1332–1359. [2723,2728]

Hanson, D. L., and Pledger, G. (1976), "Consistency in Concave Regression," *The Annals of Statistics*, 4, 1038–1050. [2722]

Hildreth, C. (1954), "Point Estimates of Ordinates of Concave Functions," *Journal of the American Statistical Association*, 49, 598–619. [2721]

Jankowski, H. K., and Wellner, J. A. (2009), "Nonparametric Estimation of a Convex Bathtub-Shaped Hazard Function," *Bernoulli*, 15, 1010–1035. [2723,2728]

Jongbloed, G., and van der Meulen, F. H. (2009), "Estimating a Concave Distribution Function from Data Corrupted with Additive Noise," *The Annals of Statistics*, 37, 782–815. [2723,2728]

Kim, A. K. H., Guntuboyina, A., and Samworth, R. J. (2018), "Adaptation in Log-Concave Density Estimation," *The Annals of Statistics*, 46, 2279–2306. [2721,2729]

Kim, A. K. H., and Samworth, R. J. (2016), "Global Rates of Convergence in Log-Concave Density Estimation," *The Annals of Statistics*, 44, 2756–2779. [2721,2729]

Koenker, R., and Mizera, I. (2010), "Quasi-Concave Density Estimation," *The Annals of Statistics*, 38, 2998–3027. [2723]

Mammen, E. (1991), "Nonparametric Regression Under Qualitative Smoothness Assumptions," *The Annals of Statistics*, 19, 741–759. [2722]

Matzkin, R. L. (1991), "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," *Econometrica*, 59, 1315–1327. [2721]

Müller, S., and Rufibach, K. (2009), "Smooth Tail-Index Estimation," *Journal of Statistical Computation and Simulation*, 79, 1155–1167. [2729]

Munk, A., Bissantz, N., Wagner, T., and Freitag, G. (2005), "On Difference-Based Variance Estimation in Nonparametric Regression When the Covariate is High Dimensional," *Journal of the Royal Statistical Society*, Series B, 67, 19–41. [2727,2730]

Pal, J. K., Woodroofe, M., and Meyer, M. (2007), "Estimating a Polya Frequency Function$_2$," in *Complex Datasets and Inverse Problems*, Volume 54 of IMS Lecture Notes—Monograph Series, eds. R. Liu, W. Strawderman, C.-H. Zhang, pp. 239–249, Beachwood, OH: Institute of Mathematical Statistics. [2721,2729]

Rice, J. (1984), "Bandwidth Choice for Nonparametric Regression," *The Annals of Statistics*, 12, 1215–1230. [2727,2730]

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995), "An Effective Bandwidth Selector for Local Least Squares Regression," *Journal of the American Statistical Association*, 90, 1257–1270. [2731,2732]

Samworth, R. J. (2018), "Recent Progress in Log-Concave Density Estimation," *Statistical Science*, 33, 493–509. [2729]

Samworth, R. J., and Yuan, M. (2012), "Independent Component Analysis via Nonparametric Maximum Likelihood Estimation," *The Annals of Statistics*, 40, 2973–3002. [2729]

Saumard, A., and Wellner, J. A. (2014), "Log-Concavity and Strong Log-Concavity: A Review," *Statistics Ssurveys*, 8, 45–114. [2729]

Seijo, E., and Sen, B. (2011), "Change-Point in Stochastic Design Regression and the Bootstrap," *The Annals of Statistics*, 39, 1580–1607. [2722]

Sen, B., Banerjee, M., and Woodroofe, M. (2010), "Inconsistency of Bootstrap: The Grenander Estimator," *The Annals of Statistics*, 38, 1953–1977. [2722]

Seregin, A., and Wellner, J. A. (2010), "Nonparametric Estimation of Multivariate Convex-Transformed Densities," *The Annals of Statistics*, 38, 3751–3781. [2721,2729]

Shen, Y., Gao, C., Witten, D., and Han, F. (2020), "Optimal Estimation of Variance in Nonparametric Regression with Random Design," *The Annals of Statistics*, 48, 3589–3618. [2727]

Varian, H. R. (1984), "The Nonparametric Approach to Production Analysis," *Econometrica*, 52, 579–597. [2721]

Verbeek, M. (2000), *A Guide to Modern Econometrics*, Chichester: Wiley. [2723]

von Neumann, J. (1941), "Distribution of the Ratio of the Mean Square Successive Difference to the Variance," *The Annals of Mathematical Statistics*, 12, 367–395. [2727]

Walther, G. (2002), "Detecting the Presence of Mixing with Multiscale Maximum Likelihood," *Journal of the American Statistical Association*, 97, 508–513. [2721,2729]

Wang, L., Brown, L. D., Cai, T. T., and Levine, M. (2008), "Effect of Mean on Variance Function Estimation in Nonparametric Regression," *The Annals of Statistics*, 36, 646–664. [2727]