

Nonparametric Empirical Bayes Confidence Intervals*

Zhen Xie

Department of Economics

Northwestern University

zhenxie@u.northwestern.edu

May 8, 2026

Abstract

Empirical Bayes methods can improve inference on unobservable individual effects by borrowing strength across units. This paper proposes nonparametric empirical Bayes confidence intervals (NP-EBCIs) for unobservable individual effects in a normal means model. The oracle intervals are constructed from posterior quantiles under a point-identified, fully nonparametric prior; feasible intervals replace these quantiles with nonparametric estimates. The NP-EBCIs are asymptotically exact in the sense that both their conditional and marginal coverage probabilities converge to the nominal level. The flexibility of this nonparametric construction has an unavoidable statistical cost. We demonstrate that posterior quantiles, unlike posterior means, inherit the severe ill-posedness of nonparametric deconvolution: the minimax optimal estimation rate is logarithmic. This logarithmic rate is minimax optimal for errors in the conditional coverage probability, and the resulting errors in the marginal coverage probability also vanish at the same logarithmic rate. Despite these slow asymptotic rates, simulations show that the NP-EBCIs remain close to nominal coverage when the prior is non-Gaussian, and deliver substantial length reductions relative to intervals that treat each unit in isolation.

Keywords: Empirical Bayes, Individual heterogeneity, Uncertainty quantification, Nonparametric methods, Ill-posed inverse problems

JEL classification codes: C11, C14, C13

*I am deeply grateful to Ivan Canay, Joel Horowitz, Federico Bugni, and Eric Auerbach for their guidance and encouragement throughout this project. I also thank Kirill Borusyak, Jiafeng Chen, Denis Chetverikov, Federico Crippa, Jiaying Gu, Michal Kolesár, Soonwoo Kwon, Carlos Lamarche, Amilcar Velez, and Kaspar Wüthrich for helpful conversations and comments. ChatGPT provided assistance with writing and proofreading the manuscript. All remaining errors are my own.

1 Introduction

Characterizing individual heterogeneity has become increasingly prominent in applied economics. Much modern empirical work treats individual effects, rather than only their average, as the primary object of interest.¹ This heterogeneity is often substantively important in its own right, as it describes how these economically relevant effects vary across units. Yet individual effects are unobserved and must be inferred from noisy data, often with heteroskedastic variances. Credible documentation of such heterogeneity therefore requires valid and informative uncertainty quantification for the individual effects themselves.

We consider the problem of constructing confidence intervals for unit-specific effects in a heteroskedastic normal means model. For each unit i , let Y_i be a noisy estimate of the unobservable individual effect θ_i and satisfy $Y_i|\theta_i \sim N(\theta_i, \sigma_i^2)$ with known sampling variance σ_i^2 . The naive z -interval $Y_i \pm z_{1-\alpha/2}\sigma_i$ has exact frequentist coverage for θ_i , but it can be quite wide when σ_i^2 is large.² Motivated by this concern, Cox (1975) and Morris (1983) introduced parametric empirical Bayes confidence intervals (EBCIs) constructed from the posterior distribution of $\theta_i|Y_i$ under a common Gaussian prior for all θ_i . Under this normal-normal model, the resulting parametric EBCIs borrow strength across units and can be much shorter than the naive z -interval (Morris, 1983). However, when the true prior is non-Gaussian, they can substantially undercover.

Armstrong, Kolesár, and Plagborg-Møller (2022, hereafter AKP) are the first to propose EBCIs that are robust to the failure of the Gaussian prior assumption. Like Cox-Morris EBCIs, their intervals are centered at linear shrinkage estimators, but use worst-case critical values calibrated to guarantee average coverage over a class of priors characterized by certain moments of θ_i . Because the full prior distribution is nonparametrically identified in the normal means model, this moment-based calibration leaves some available information unused. As a result, their robust EBCIs can be conservative when the true prior is far from the least favorable one in that moment class.

This paper proposes nonparametric empirical Bayes confidence intervals (NP-EBCIs) that fully utilize the entire prior distribution. Following Cox (1975); Morris (1983), we construct these intervals from the lower and upper quantiles of the posterior distribution of $\theta_i|Y_i$. Rather than imposing a Gaussian prior, we assume $\theta_i \sim G$ and allow the distribution G to be fully nonparametric. We target the oracle posterior credible intervals under the prior G and construct their feasible counterparts by estimating the posterior quantiles nonparametrically. This contrasts with Armstrong et al. (2022), whose robust EBCIs are calibrated using only moment information about θ_i .

The NP-EBCI construction makes posterior quantiles central to our analysis, and our first

¹See Walters (2024); Bonhomme and Denis (2024) for recent overviews. A growing body of empirical work studies unit-specific parameters, including neighborhood effects (Chetty and Hendren, 2018), workplace heterogeneity (Card, Heining, and Kline, 2013), teacher value-added (Chetty, Friedman, and Rockoff, 2014), manager effects (Fenizia, 2022), geographic variation in health care (Finkelstein, Gentzkow, and Williams, 2016), bank-specific credit-supply effects (Amiti and Weinstein, 2018), employer-specific discrimination (Kline, Rose, and Walters, 2022, 2024), police-officer heterogeneity (Goncalves and Mello, 2021), among many others.

²Here z_α is the α -quantile of the standard normal distribution.

contribution is to characterize the minimax optimal rate for estimating them. We establish this rate by deriving a minimax lower bound for the estimation risk and constructing a one-step kernel-based estimator that achieves it. This optimal rate is inherently logarithmic, reflecting the fundamental limits of empirical Bayes posterior quantile estimation. This result highlights a sharp contrast with posterior mean estimation (Jiang and Zhang, 2009; Brown and Greenshtein, 2009). In the normal means model, recovering the unknown prior G is a severely ill-posed deconvolution problem with a logarithmic minimax rate (Carroll and Hall, 1988).³ Although posterior mean and posterior quantiles are both nonlinear functionals of G , the posterior mean can circumvent the ill-posedness and be estimated at a nearly parametric rate (Zhang, 1997, 2009).⁴ We show that posterior quantiles cannot: because they are defined through a *non-smooth* criterion function, their estimation inherits the severe ill-posedness of nonparametric deconvolution.

We next investigate the *conditional* coverage properties of the feasible NP-EBCI, which relies on estimated posterior quantiles. By conditional coverage, we mean the probability that the reported interval contains θ_i after Y_i has been observed. This conditional perspective has a long history in statistics, dating back at least to Cox (1958),⁵ who argued that conditioning is needed so that the uncertainty statement reflects what can be learned from the realized data. Our second contribution is to characterize the minimax optimal rate at which the error in conditional coverage probability vanishes, and to show that the feasible NP-EBCI attains this rate. This optimal rate is again logarithmic, reflecting the fundamental limits of nonparametric empirical Bayes inference when targeting exact conditional coverage. Ultimately, the feasible NP-EBCI achieves asymptotically exact conditional coverage, but its coverage probability converges to the nominal level at this minimax optimal logarithmic rate.

As our third contribution, we show that the feasible NP-EBCI has asymptotically exact *marginal* coverage, with coverage error that again vanishes at a logarithmic rate. By marginal coverage, we mean the ex-ante probability that the reported interval contains θ_i under repeated sampling of both θ_i and the observed data. This notion is commonly referred to as “empirical Bayes coverage” (Morris, 1983). Our result differs from the parametric EBCI literature, where the error in marginal coverage can vanish at a faster parametric rate under a correctly specified Gaussian prior.⁶ However, their theoretical guarantees are vulnerable to misspecification when the true prior is non-Gaussian. Our result also differs from Armstrong et al. (2022, Theorem 4.1), who show that the empirical Bayes coverage of their robust EBCI is asymptotically at least the nominal level, thus their actual coverage need not be exact even asymptotically. The feasible NP-EBCI, by contrast, achieves asymptotically exact empirical Bayes coverage, but with coverage error of logarithmic order.

³ Throughout the paper, we use “prior distribution” and “mixing distribution” interchangeably for G . In empirical Bayes language, G is the prior distribution for θ_i , while the term “mixing distribution” reflects that the marginal law of Y_i is obtained by mixing the normal distribution over G , see Equation (2.2).

⁴ In the nonparametric empirical Bayes literature, faster rate results for the posterior mean are typically stated in terms of average mean squared errors for estimating θ_i ; see e.g. Jiang and Zhang (2009); Chen (2026).

⁵ Related classic discussions of the conditional perspective include Buehler (1959); Wallace (1959); Birnbaum (1962); Robinson (1979); see also Rubin (1984); Robins and Wasserman (2000) for later discussions.

⁶ See e.g. Datta, Ghosh, Smith, and Lahiri (2002); Chatterjee, Lahiri, and Li (2008); Yoshimori and Lahiri (2014).

This theoretical contrast motivates a comparison of the finite-sample performance of these EBCIs. In our simulations, all three EBCI procedures deliver average length reductions relative to the naive z -interval. AKP robust EBCIs with the second moment often exceed nominal coverage and yield the least length reductions. Cox-Morris parametric EBCIs are typically the shortest, but this efficiency can come at the cost of substantial undercoverage when the true prior is non-Gaussian, especially in low signal-to-noise ratio regimes. The feasible NP-EBCI, despite its severe ill-posedness, remains much closer to the nominal coverage level, even in small samples, while still delivering substantial length reductions.

Related literature. This paper contributes to the nonparametric empirical Bayes literature on the normal means problem. The central idea in this literature is to treat many related decision problems jointly, learn prior information from related units, and then use it to construct Bayes posterior summaries or decision rules for each unit (Robbins, 1956, 1964; Efron, 2014, 2019, see also Koenker and Gu, 2024; Walters, 2024 and references therein). Much of this literature has focused on estimation of individual effects,⁷ while more recent work, including Jiang (2019) and Koenker (2020), has considered confidence intervals for individual effects. Both papers provide constructions and numerical evidence, but do not develop theoretical guarantees for their procedures. Our contribution is to establish a general theory for posterior quantile estimation and the coverage properties of posterior-quantile-based EBCIs. More broadly, because posterior quantiles are Bayes rules under asymmetric loss, these results are also relevant for other empirical Bayes decision problems that involve posterior quantiles (e.g. Gu and Koenker, 2023, Section 2.5 in Walters, 2024).

This paper is broadly related to, but different from Ignatiadis and Wager (2022) and Ignatiadis and Sen (2025). Specifically, Ignatiadis and Wager (2022) study confidence intervals for empirical Bayes estimands, such as the posterior mean, rather than confidence intervals for the individual effect θ_i itself. Ignatiadis and Sen (2025) study empirical partially Bayes multiple testing. As they note in Footnote 2, their p -values can be inverted to construct individual intervals for θ_i . A key distinction is that they place a prior on the variances rather than individual effects, pooling information across units through the variance distribution. Consequently, in the homoskedastic model, their interval reduces to the naive z -interval (see their Example 14), while our procedure continues to exploit the common prior on the individual effects.

Our objective differs from that of the deconvolution literature. Deconvolution methods aim to recover the latent distribution of unobserved heterogeneity from noisy measurements (see e.g. Horowitz, 2014; Schennach, 2020 for reviews). Our goal, by contrast, is individualized inference for θ_i itself. The role of deconvolution in our analysis is therefore technical rather than substantive: the posterior quantiles underlying our NP-EBCIs are functionals of the latent distribution, so the problem is built on the same Gaussian deconvolution structure.

⁷See Chen (2026) and references therein. For other settings beyond normal mean models, see e.g. Gu and Koenker (2017); Liu, Moon, and Schorfheide (2020); Gilraïne, Gu, and McMillan (2020); Gaillac (2024); Kwon (2026); Cheng, Ho, and Schorfheide (2025).

Our results on minimax optimal rates for posterior quantiles connect to the literature on functional estimation in ill-posed inverse problems. For linear functionals of the mixing distribution G , the minimax rate can depend sharply on the smoothness of the weighting function that defines the functional (Butucea and Comte, 2009; Pensky, 2017). Our setting is different, because both posterior mean and posterior quantiles are *nonlinear* functionals of the mixing distribution. But a similar insight emerges after local linearization: the two functionals differ sharply in their local dependence on the mixing distribution. The posterior mean involves a comparatively smooth weighting function and therefore can circumvent the severe ill-posedness. By contrast, the posterior quantile is locally governed by a discontinuous weighting function and therefore inherits the logarithmic difficulty of the underlying inverse problem.

Outline. Section 2 presents the set-up, defines the oracle NP-EBCI, and contrasts it with the Cox-Morris parametric EBCI and AKP robust EBCIs. Assuming a homoskedastic normal mean model, Section 3 derives the minimax optimal rate of estimating posterior quantiles, and establishes the conditional and marginal coverage properties of the feasible NP-EBCI. Section 4 returns to the heteroskedastic setting and presents a feasible implementation of the NP-EBCI, including bandwidth selection. Section 5 uses Monte Carlo simulations to compare the finite-sample performance of NP-EBCI against existing methods. Section 6 concludes. Proofs and extended discussions are collected in the Appendix.

2 Empirical Bayes confidence intervals for individual effects

2.1 Set-up

We consider the following normal means model:

$$Y_i \mid \theta_i \sim N(\theta_i, \sigma_i^2), \quad \theta_i \stackrel{\text{i.i.d.}}{\sim} G(\cdot), \quad \text{for } i = 1, \dots, n \quad (2.1)$$

where Y_i is a noisy estimate of the unobserved outcome of interest θ_i for each unit i . For example, θ_i may represent the value-added of teacher i , and Y_i is an estimate constructed from the associated students' test scores. The unit-specific parameter θ_i is thought to be random, and drawn from the population distribution $G(\cdot)$. The variances σ_i^2 are known, but G and $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_n\}$ are unknown.

The model (2.1) arises naturally from a hierarchical data structure. For instance, let $\tilde{Y}_{i,j}$ denote the test outcome of student $j = 1, \dots, J_i$ taught by teacher i , and suppose that these test scores are normally distributed with a teacher-specific mean and variance: $\tilde{Y}_{i,j} \mid \theta_i \sim N(\theta_i, \sigma_i^2)$ for $j = 1, \dots, J_i$ where θ_i represents the value-added of teacher i . A natural estimator for θ_i is the average test score of students taught by teacher i : $Y_i = J_i^{-1} \sum_{j=1}^{J_i} \tilde{Y}_{ij}$. Then we have that $Y_i \mid \theta_i \sim N(\theta_i, \sigma_i^2/J_i)$ which falls into the model (2.1).

The model (2.1) fits into the classical measurement errors model, with $Y_i = \theta_i + \varepsilon_i$ and $\varepsilon_i \sim$

$N(0, \sigma_i^2)$. The normal error assumption is restrictive, but standard in the empirical Bayes literature with cross-sectional data (Efron, 2012; Koenker and Gu, 2024; Walters, 2024). It can be replaced by any other known error distribution $F_{\varepsilon_i}(\cdot)$, or even fully relaxed to an unknown distribution when panel data or other repeated measurements are available.⁸

The mixing distribution $G(\cdot)$ is nonparametrically identified from the marginal law of Y_i . Assume that the mixing distribution $G(\cdot)$ has a continuous density $g(\cdot)$ with respect to Lebesgue measure. Given the known variance σ_i^2 , the marginal density of Y_i is given by

$$f_{Y_i}(y) = \int \frac{1}{\sigma_i} \phi\left(\frac{y - \theta}{\sigma_i}\right) dG(\theta) \quad (2.2)$$

where $\phi(\cdot)$ is the standard normal density. For an integrable function $f : \mathbb{R} \rightarrow \mathbb{C}$, write $f^*(t) = \int_{\mathbb{R}} \exp(itx) f(x) dx$ for its Fourier transform. Let $f_{Y_i}^*(t), g^*(t)$ be the Fourier transforms of the densities $f_{Y_i}(y), g(\theta)$ respectively. The convolution theorem (Section 3.3.2, Schennach, 2020) gives that $f_{Y_i}^*(t) = g^*(t) f_{\varepsilon_i}^*(t)$ where $f_{\varepsilon_i}^*(t) = \exp(-\frac{1}{2}\sigma_i^2 t^2)$ is the Fourier transform of the normal error distribution. Since $f_{\varepsilon_i}^*(t) \neq 0$ for every $t \in \mathbb{R}$, we recover $g^*(t) = f_{Y_i}^*(t)/f_{\varepsilon_i}^*(t)$. Because characteristic functions uniquely determine probability laws, this establishes the identification of G .⁹

2.2 Empirical Bayes confidence intervals

We aim to construct an individual interval CI_i for each unobservable individual effect θ_i with a pre-specified coverage probability. Following Cox (1975); Morris (1983), we take an empirical Bayesian approach: assume that $\theta_i \sim G$ for some prior distribution $G(\cdot)$ before any observation is taken. After observing $Y_i = y$, uncertainty about θ_i is summarized by the posterior distribution $\theta_i | Y_i$.

If the true prior G were known, we could construct an *oracle* Bayesian credible interval by taking the lower and upper $\alpha/2$ posterior quantiles.¹⁰ Let $\mathcal{D}_i = (Y_i, \sigma_i)$, we denote this infeasible oracle interval by

$$CI_i^{\text{NP*}} = \left[q_G(\alpha/2; \mathcal{D}_i), q_G(1 - \alpha/2; \mathcal{D}_i) \right] \quad (2.3)$$

where the posterior quantile is defined as

$$q_G(\tau; \mathcal{D}_i) = \inf \left\{ u : P_G(\theta_i \leq u \mid Y_i = y) \geq \tau \right\} \quad \text{for } \tau \in (0, 1). \quad (2.4)$$

The subscript G is used to emphasize that these posterior quantiles are functionals of the underlying prior distribution.

⁸See, for example, Horowitz and Markatou (1996) and Li and Vuong (1998), as well as the surveys by Schennach (2020) and Bonhomme and Denis (2024, Section 7), for further details.

⁹This simplified exposition is intended to align with the theoretical analysis in Section 3. However, we note that the identification argument applies to any Borel probability measure G (Folland, 1999, Proposition 8.50).

¹⁰Our interval is the “equal-tailed credible set” for θ_i . One could alternatively construct a highest posterior density (HPD) credible set, defined as the set where the posterior density exceeds some threshold (Berger, 1985, Definition 5). However, a practical limitation of the HPD credible set is that it may be disjoint, making it difficult to report.

Because G is unknown in practice, we must distinguish between the oracle target (2.3) and its feasible analogue. Our proposed nonparametric empirical Bayes confidence interval (NP-EBCI) replaces the oracle posterior quantiles with nonparametric estimates:

$$\text{CI}_i^{\text{NP}} = \left[\hat{q}_G(\alpha/2, \mathcal{D}_i), \hat{q}_G(1 - \alpha/2; \mathcal{D}_i) \right]. \quad (2.5)$$

The theoretical justification of the feasible procedure (2.5) relies on large-sample asymptotics as $n \rightarrow \infty$, in the spirit of Robbins (1964). Before turning to that analysis in Section 3, we discuss several oracle properties of the interval (2.3). First, the oracle interval (2.3) admits a simple decision-theoretic characterization.

Proposition 2.1. *The quantile-based posterior credible interval (2.3) minimizes*

$$\mathcal{R}_i(L, U) = \text{E} \left[U(\mathcal{D}_i) - L(\mathcal{D}_i) + \frac{2}{\alpha} (L(\mathcal{D}_i) - \theta_i)_+ + \frac{2}{\alpha} (\theta_i - U(\mathcal{D}_i))_+ \right]$$

over all measurable interval rules (L, U) satisfying $L(\mathcal{D}_i) \leq U(\mathcal{D}_i)$ almost surely, where the expectation is taken over (Y_i, θ_i) .

Proposition 2.1 gives a direct decision-theoretic justification for using posterior quantiles as interval endpoints. The interval (2.3) is not only an equal-tailed posterior credible interval, but also the Bayes rule for an interval-valued decision problem that trades off interval length against linear penalties for undercoverage and overcoverage. In this sense, the lower and upper posterior quantiles arise as optimal endpoints under a particular loss function, in line with the decision-theoretic treatment of Bayesian interval estimation studied by Winkler (1972); Berger (1985). The criterion in Proposition 2.1 also coincides with the interval score used to evaluate prediction intervals (Gneiting and Raftery, 2007).

The oracle interval (2.3) has the usual conditional coverage property of an empirical Bayes confidence interval (EBCI) for θ_i (Carlin and Louis, 2000, Definition 3.2):

$$\text{P}(\theta_i \in \text{CI}_i^{\text{NP}*} \mid Y_i) = 1 - \alpha \quad \text{for each } i = 1, \dots, n \quad (2.6)$$

that summarizes the uncertainty about θ_i after the data are observed. It is interpreted as a data-specific measure of uncertainty: it evaluates inference procedure based on the realized sample rather than averaging over hypothetical, unobserved outcomes. This conditional perspective goes back at least to Cox (1958) who wrote that inference should reflect what can be learned from the data that we have, and it is central to Bayesian literature which treats the observed data as known and concerns the remaining uncertainty through the conditional distribution of unknowns given knowns (Rubin, 1984, see also Berger, 1985, Section 1.6).

The oracle interval (2.3) also satisfies the marginal coverage property, often called empirical Bayes coverage (Morris, 1983, Carlin and Louis, 2000, Definition 3.1). To see this, take averages

over the observed data distribution:

$$\begin{aligned} P(\theta_i \in \text{CI}_i^{\text{NP}*}) &= E_{\mathbf{Y}} [P(\theta_i \in \text{CI}_i^{\text{NP}*} \mid Y_i)] = 1 - \alpha \\ &= E_{\boldsymbol{\theta}} [P(\theta_i \in \text{CI}_i^{\text{NP}*} \mid \boldsymbol{\theta})] = 1 - \alpha \end{aligned}$$

the probability in $P(\theta_i \in \text{CI}_i)$ is taken over both $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ and the data $\mathbf{Y} = (Y_1, \dots, Y_n)$.¹¹ Thus marginal coverage is an *ex ante* statement: it characterizes the uncertainty prior to observing any data. It can also be interpreted as the population average of frequentist coverage $P(\theta_i \in \text{CI}_i^{\text{NP}*} \mid \boldsymbol{\theta})$ that considers the uncertainty from the data. This marginal coverage statement is weaker than the conditional coverage property (2.6) because it need only hold on average and therefore may be poorly calibrated for the sample data at hand. Even so, the notion of marginal coverage is standard in the empirical Bayes literature.

We now compare the oracle NP-EBCI with the existing alternatives in the literature. The natural starting point is the naive z -interval $Y_i \pm z_{1-\alpha/2}\sigma_i$, which treats each unit in isolation and has exact frequentist coverage for θ_i without imposing any prior structure. The limitation is that when σ_i^2 is large, the interval can be unnecessarily wide. This concern leads naturally to the empirical Bayes confidence interval (EBCI) literature, which aims to improve inference for each unit by combining the noisy observation Y_i with prior information learned from the distribution of individual effects.

Cox (1975) and Morris (1983) initiated the parametric EBCI literature by constructing individual intervals for θ_i from the posterior distribution under a Gaussian prior. In the homoskedastic case $\sigma_i^2 \equiv \sigma^2$, they assume $\theta_i \stackrel{\text{i.i.d.}}{\sim} N(0, A)$. Under this normal-normal specification, the posterior distribution takes a simple form: $\theta_i \mid Y_i = y \sim N(\frac{A}{A+\sigma^2}y, \frac{A\sigma^2}{A+\sigma^2})$ which yields the oracle interval

$$\text{CI}_i^{\text{Cox}*} = w_{\text{EB}}Y_i \pm w_{\text{EB}}^{1/2}\sigma \times z_{1-\alpha/2} \quad \text{where } w_{\text{EB}} = \frac{A}{A+\sigma^2} \quad (2.7)$$

A feasible version of this interval replaces the unknown variance component A with an estimator. This construction can produce substantially shorter intervals than the naive z -interval. This gain, however, comes from the parametric prior restriction.

Like Cox (1975) and Morris (1983), we construct intervals from the posterior distribution of $\theta_i \mid Y_i$. The difference is that we make no parametric assumptions about the prior distribution. Instead we treat the prior G as fully nonparametric and take the oracle posterior interval in (2.3) as the target. The interval proposed in practice is the feasible analog in (2.5), obtained by replacing the oracle posterior quantiles with nonparametric estimates.

A different alternative is the robust EBCI of Armstrong et al. (2022), which targets worst-case marginal coverage over a moment class rather than the posterior interval under a point-identified

¹¹This notion of marginal coverage also appears in the conformal prediction literature, see e.g. Angelopoulos and Bates (2023).

prior. In the homoskedastic case $\sigma_i^2 \equiv \sigma^2$, their oracle interval takes the form

$$\text{CI}_i^{\text{AKP}*} = w_{\text{EB}} Y_i \pm w_{\text{EB}} \sigma \times \chi(\alpha) \quad \text{where } w_{\text{EB}} = \frac{A}{A + \sigma^2}$$

where $A = E[Y_i^2] - \sigma^2$. The critical value $\chi(\alpha)$ is chosen such that the optimal value of the following optimization problem is at least $1 - \alpha$:

$$\inf_G \mathbb{P}_G [\theta_i \in \text{CI}_i^{\text{AKP}}(\chi)] \quad \text{s.t. } E_G[\theta_i^2] = A. \quad (2.8)$$

By construction, the marginal coverage probability of their oracle robust EBCI is at least $1 - \alpha$ for all prior distributions that satisfy the second moment constraint. However, the actual coverage probability may exceed the nominal level when the true prior is not least-favorable distribution that solves (2.8). The oracle NP-EBCI, by contrast, relies on nonparametric identification of the prior distribution and fully exploits the entire distribution of θ_i . Consequently, the interval (2.3) attains exact marginal coverage at the nominal level $1 - \alpha$. Moreover, the AKP robust EBCI does not target the exact posterior conditional coverage property in (2.6) which conditions on the observed data. By contrast, the oracle NP-EBCI (2.3) satisfies this property by construction.

We make two additional remarks to clarify how certain terminology in empirical Bayes inference are understood in our setting.

Remark 2.1. Morris (1983) assumed that the prior belongs to a known class $\mathcal{G}$, and evaluated the performance of procedures uniformly over all $G \in \mathcal{G}$. He also suggested placing a “second-stage” prior on $\mathcal{G}$ to account for uncertainty about the prior itself. In our setting, G is point identified. The remaining uncertainty is sampling uncertainty, because G must still be estimated from the data. One could, in principle, place a prior on $\mathcal{G}$ to account for this estimation uncertainty, see also Remark 3.6. ■

Remark 2.2. Morris (1983) discussed the robust Bayes literature; see also Berger (1985, Chapter 4.7) for a broader treatment that considers sensitivity analysis as the “subjective” prior varies over an ε -contamination class. In our setting, the prior is point identified, so we do not represent prior uncertainty through such a class. Robust Bayes methods would nevertheless be relevant if one wished to study sensitivity to prior misspecification or contamination, which we do not pursue in this paper. ■

3 Theoretical results under homoskedasticity

In this section, we analyze the theoretical properties of the feasible NP-EBCI. First we consider nonparametric estimation of the posterior quantile: we establish its optimal rates of convergence in Section 3.1 and propose a rate-optimal estimator in Section 3.2. Second, in Section 3.3, we establish the optimal rates of the errors in conditional coverage probability and show that the feasible NP-

EBCI based on plug-in estimate of posterior quantile achieves this rate. Finally, we examine its marginal coverage probability and connects it to the empirical Bayes literature.

Notation. For $1 \leq p < \infty$, the space $L^p(\mathbb{R})$ is defined as $L^p(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{C} \mid \int_{\mathbb{R}} |f(x)|^p dx < \infty\}$, with associated norm $\|f\|_p = (\int_{\mathbb{R}} |f(x)|^p dx)^{1/p}$. For $p = \infty$, we define the norm as $\|f\|_{\infty} = \text{ess sup}_{x \in \mathbb{R}} |f(x)|$. For any function $f : \mathbb{R} \rightarrow \mathbb{C}$ such that $f \in L^1(\mathbb{R})$, we define its Fourier transform by $f^*(t) = \int_{\mathbb{R}} \exp(itx) f(x) dx$ for $t \in \mathbb{R}$.

Throughout Section 3, we assume that $\sigma_i^2 \equiv \sigma^2$ for technical convenience, which implies that $(Y_i, \theta_i) \stackrel{\text{i.i.d.}}{\sim} \mathcal{F}$. In Sections 3.1 and 3.2, we consider the posterior quantile associated with a hypothetical random variable $(Y_{n+1}, \theta_{n+1}) \sim \mathcal{F}$, evaluated at $Y_{n+1} = y$, and estimate it using the data $\{Y_i\}_{i=1}^n$. In Sections 3.3 and 3.4, since our goal is to construct individual intervals CI_i for each θ_i , the relevant posterior quantile corresponds to (Y_i, θ_i) and is estimated using the leave-one-out data $\{Y_j\}_{j \neq i}$. We focus on a fixed quantile level $\tau \in (0, 1)$, tailored to the objective of the application.

3.1 Fundamental limits of estimating posterior quantiles

Both posterior mean¹² and posterior quantiles are nonlinear integral functionals of the unknown mixing distribution G . In the normal means model, the observed distribution of Y is the convolution of G with the normal error distribution, so recovering G from the data is a deconvolution problem. With normal errors, this inverse problem is severely ill-posed: the optimal rate for estimating the mixing density g is only logarithmic (Carroll and Hall, 1988, see also Horowitz (2014) for a review).¹³ In contrast to this severe ill-posedness, the posterior mean can be estimated at a nearly *parametric* rate (Zhang, 1997, 2009). Since the posterior quantile is also an integral functional of the unknown mixing density, it is natural to ask whether such “integration” can likewise lead to faster rates of convergence.

We will show that the answer is negative. Unlike the posterior mean, the posterior quantile is defined through a non-smooth criterion function. This non-smoothness prevents the associated Fourier weighting term from offsetting the exponential decay induced by the Gaussian error density. Thus, posterior quantile estimation inherits the severe ill-posedness of Gaussian deconvolution. To show that this difficulty is intrinsic, rather than a deficiency of a particular estimator, we derive a minimax lower bound on the risk of posterior quantile estimation over a functional class. This minimax result characterizes the fundamental limits of the estimation problem and provides a benchmark for evaluating the quality of an estimation procedure.

We begin by rewriting the posterior quantile in a form that exposes its non-smoothness. Al-

¹²The posterior mean, defined as $m(y) = \text{E}_G[\theta | Y = y]$, is one of the most widely studied objects in empirical Bayes literature. It is the optimal predictor of the individual parameter θ_i under the mean-square error, conditional on the observed data $Y = y$.

¹³ The normal density is a canonical example of a “super-smooth” density, which leads to the worst-case rates of convergence in nonparametric deconvolution. When the error density is “ordinarily smooth”, the optimal rate is polynomial (Fan, 1991); see Schennach (2020, page 511) for further discussion.

though defined in (2.4) as the solution to a moment condition weighted by the normal error density, it is more useful to reframe it as the minimizer of a convex optimization problem.

Proposition 3.1. *The posterior quantile (2.4) can be characterized as the minimizer of an ℓ_1 loss function:*

$$q_G(\tau; y) = \arg \min_q \int \rho_\tau(\theta - q) \phi\left(\frac{y - \theta}{\sigma}\right) dG(\theta) \quad (3.1)$$

where $\rho_\tau(u) = u(\tau - \mathbf{1}\{u \leq 0\})$ is the check function. We denote the objective function by $W(q)$.

We transform the objective function (3.1) into the frequency domain using Fourier basis functions. Define the weighting function $M(q, \theta; y) = \rho_\tau(\theta - q) \phi(\frac{y - \theta}{\sigma})$ that appears in the integral of the objective function (3.1). By the Parseval-Plancherel theorem (Theorem 9.13 in Rudin (1987)), the objective function $W(q)$ in Equation (3.1) can be expressed as

$$W(q) = \int M(q, \theta; y) g(\theta) d\theta = (2\pi)^{-1} \int \overline{M^*(q, t; y)} g^*(t) dt = (2\pi)^{-1} \int \frac{\overline{M^*(q, t; y)}}{f_\varepsilon^*(t)} f_Y^*(t) dt \quad (3.2)$$

where $\overline{M^*(q, t; y)} = \int \exp(-it\theta) M(q, \theta; y) d\theta$ is the conjugate Fourier transform of $M(q, \theta; y)$.

The Fourier representation (3.2) shows that the problem has two ingredients. First, Gaussian deconvolution is severely ill-posed because $f_\varepsilon^*(t)$ decays exponentially as $|t| \rightarrow \infty$, so division by $f_\varepsilon^*(t)$ amplifies high-frequency noise. Second, whether this difficulty can be mitigated depends on the decay of the Fourier weighting term. For smooth functionals, such as the posterior mean, that weighting term decays fast enough to downweigh the exponential decay at the tail, leading to a nearly parametric rate.¹⁴ For posterior quantiles, however, the check function in (3.1) is non-smooth at zero, so $\overline{M^*(q, t; y)}$ decays too slowly to offset it. The posterior quantile therefore inherits the severe ill-posedness of Gaussian deconvolution.

To state the formal results, we introduce the following regularity assumptions.

Assumption 3.1. *The mixing density $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.*

Assumption 3.2. *We assume that the mixing density g belongs to the Sobolev class*

$$\mathcal{G}(s, L) = \left\{ g : \int |g^*(t)|^2 (t^2 + 1)^s dt < L \right\}$$

where $s \geq \frac{1}{2}$ is an integer smoothness parameter.

We denote the posterior density evaluated at $\theta = u$ by

$$\pi(u|Y = y) = \frac{\phi\left(\frac{y-u}{\sigma}\right) g(u)}{\int \phi\left(\frac{y-\theta}{\sigma}\right) g(\theta) d\theta} = \frac{\phi\left(\frac{y-u}{\sigma}\right) g(u)}{f_Y(y)}. \quad (3.3)$$

¹⁴ A similar insight appears in the semiparametric models with measurement errors that achieve the parametric $\sqrt{n}$ rate, see Schennach (2020, Page 512) for related discussion.

Assumption 3.3. *The normalized posterior density evaluated at the target posterior quantile $q_0 := q_G(\tau; y)$ is strictly positive, $|\phi(\frac{y-q_0}{\sigma})g(q_0)| > 0$.*

Assumption 3.4. *The Fourier transformation of the mixing density satisfies that $g^*(t) \in L^1(\mathbb{R})$. It implies the density g is continuous and uniformly bounded as $\|g\|_\infty \leq (2\pi)^{-1}\|g^*\|_1 < \infty$.*

Assumption 3.5. *The target parameter $q_G(\tau; y)$ lies in a compact parameter space.*

Assumption 3.1 is a regularity condition that facilitates the application of Fourier-based methods. Assumption 3.2 is equivalent to requiring that the sum of the squared $L^2(\mathbb{R})$ -norms of the first s derivatives of the mixing density $g(\cdot)$ be bounded, i.e. $\sum_{l=0}^s \|g^{(l)}\|_2^2 < L$, see e.g. Meister (2009, Appendix A.2). This Sobolev condition provides global control of the smoothness and is well suited for analyzing posterior quantile, a nonlinear integral functional of mixing density, while Hölder condition imposes smoothness only locally around some specific point.

Assumption 3.3 is the identification condition for the posterior quantile by requiring that the criterion function (3.1) is non-flat at its solution, ensuring a unique minimizer. Assumption 3.4 together with Assumption 3.1, ensures that the inverse Fourier transform of $g^*(t)$ is well-defined and yields a continuous mixing density $g(\cdot)$. Consequently, the posterior distribution has no atoms at the target quantile. Finally, Assumption 3.5 ensures that the posterior quantile is well-defined and lies in a compact subset of $\mathbb{R}$.

Theorem 3.1 (Minimax lower bounds). *Let Assumptions 3.1 to 3.5 hold. Then for any estimator $\hat{q}$ based on the i.i.d. data $\{Y_i\}_{i=1}^n$, we have that*

$$\inf_{\hat{q}} \sup_{g \in \mathcal{G}(s, L)} \mathbb{E} \left[(\hat{q}_G(\tau; y) - q_G(\tau; y))^2 \right] \geq \text{const} \cdot (\log n)^{-(2s+1)/2}$$

for n sufficiently large.

Theorem 3.1 establishes a minimax lower bound for estimating posterior quantiles uniformly over the Sobolev space $\mathcal{G}(s, L)$. In particular, it shows that no estimator can attain a faster rate of convergence than logarithmic uniformly over this function class. This lower bound reflects that posterior quantile estimation, due to the *non-smoothness* of its criterion function, inherits the intrinsic difficulty of nonparametric deconvolution. The next subsection shows that this lower bound is attained by a feasible estimator. Hence the logarithmic rate is the minimax optimal rate for estimating posterior quantiles in the homoskedastic normal-means model.

Theorem 3.1, together with the nearly parametric rate result for posterior mean estimation (Zhang, 1997, 2009), highlights a broader point in functional estimation for ill-posed inverse problems: the minimax optimal rate is determined *jointly* by the degree of ill-posedness of the inverse operator and the regularity of the target functional. Posterior quantiles are too irregular to circumvent the difficulty of the underlying severely ill-posed deconvolution problem, and their minimax rate therefore remains logarithmic.

Remark 3.1. It is worthwhile to mention that as with the monotonicity property of the posterior mean (Houwelingen and Stijnen, 1983; Koenker and Mizera, 2014), the posterior quantile $q_G(\tau; y)$ is non-decreasing in y for a broader class of error distributions F_ε that includes the normal (Proposition C.1). However, this monotonicity property does not improve the rate of convergence of the posterior quantile in the normal means model. The reason is that the monotonicity holds automatically for any valid densities, it does not impose additional restrictions on the risk when deriving the lower bound. Moreover, the posterior quantile $q_G(\tau; y)$ is non-decreasing in τ ,¹⁵ but this likewise does not improve the rate of convergence for the same reason above. ■

Remark 3.2. Escanciano (2023) shows that in many structural models, quantile functionals of nonparametric unobserved heterogeneity have *infinite* efficiency bounds. This occurs because the non-smoothness of the influence function violates the necessary condition in van der Vaart (1991). Our result in Theorem 3.1 also illustrates the implications of this non-smoothness for statistical performance but focuses on characterizing minimax optimal rates. ■

3.2 A rate-optimal estimator of posterior quantiles

In this subsection, we propose a one-step estimator for the posterior quantile that achieves the optimal rate established in Theorem 3.1.

Observe that the representation (3.2) expresses the posterior quantile directly as a function of the data, and naturally suggests an estimator for the posterior quantile, defined as the minimizer of the empirical counterpart of the objective function. In this representation, the only unknown component is $f_Y^*(t)$. Therefore, to construct the empirical analog $W_n(q)$, we replace $f_Y^*(t)$ with $\hat{f}_Y^*(t)K^*(h_nt)$, where $\hat{f}_Y^*(t) = n^{-1} \sum_{i=1}^n \exp(itY_i)$ is the empirical characteristic function, and $K^*(t)$ denotes the Fourier transformation of a kernel function $K(\cdot)$ with bandwidth h_n . The kernel serves as a regularization device to ensure the existence of the integral when $f_Y^*(t)$ is replaced by its empirical analog. The resulting kernel estimator of the posterior quantile is then

$$\hat{q}_G(\tau; y) = \arg \min_q (2\pi)^{-1} \int \frac{\overline{M^*(q, t; y)}}{f_\varepsilon^*(t)} \left(\hat{f}_Y^*(t) K^*(h_nt) \right) dt \quad (3.4)$$

where the sample objective function is denoted by $W_n(q)$.

Our proposed estimator (3.4) is a one-step procedure that bypasses the need to estimate the mixing distribution $G(\cdot)$ as an intermediate step. This estimation strategy is in the spirit of f -modeling for posterior mean estimation, which involves estimating only the marginal density $f_Y(y)$ and its derivative (Robbins, 1956; Efron, 2011, 2014). Another advantage of this direct estimation is that the regularization parameter can be tuned directly for the target functional, rather than for an intermediate estimate of the mixing distribution.

Our estimator for the posterior quantile is, by construction, similar to the classical deconvolu-

¹⁵This is known as the “quantile non-crossing” property, see Koenker (2005, Chapter 2.5)

tion kernel density estimator (Carroll and Hall, 1988; Stefanski and Carroll, 1990). However, as discussed in Section 3.1, the posterior quantile is a nonlinear integral functional of the deconvolution density. This leads to two key distinctions and challenges of the theoretical analysis. First, our estimator is defined as the solution to a Fourier integral, rather than an integral itself. Second, the integrand involves a weighting function $\overline{M^*(q, t; y)}$ whose (non-)smoothness crucially affects the convergence rate of the estimator.

We then formally establish the theoretical properties of the proposed estimator $\hat{q}_G(\tau; y)$.

Assumption 3.6. (i) *The kernel satisfies that $K \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ with $K_{h_n} = h_n^{-1}K(\cdot/h_n)$. Its Fourier transformation $K^*(t)$ is supported on $[-1, 1]$.*

(ii) *Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a bounded function that satisfies $\int v^d K(v) dv = 0$ for $d = 1, \dots, s$, $\int |v|^{s+1} |K(v)| dv < \infty$.*

Assumption 3.6 (i) requires that the Fourier transform of the kernel function be bounded, integrable and have compact support. Assumption 3.6 (ii) requires K to be higher-order kernel, that is a device designed to increase the rate of convergence of the asymptotic bias. Here we use a kernel of the order $s + 1$, which is one degree higher than the smoothness of the mixing density.

Theorem 3.2 (Consistency and rates of convergence). *Let Assumptions 3.1 to 3.6 hold. Setting the bandwidth $h_n \asymp (\log n)^{-1/2}$, we have that*

$$\left| \hat{q}_G(\tau; y) - q_G(\tau; y) \right| = O_p \left((\log n)^{-(2s+1)/4} \right).$$

Theorem 3.2 establishes that our proposed kernel estimator for the posterior quantile is consistent and achieves the optimal rate specified in Theorem 3.1. The slow rate of convergence reflects the difficulty of the estimation problem, and is not an indicator that our estimator is deficient.

The key step in the proof of Theorem 3.2 is to derive the following asymptotic expansion

$$r_n (\hat{q} - q_0) = -r_n \left[\phi \left(\frac{y - q_0}{\sigma} \right) g(q_0) \right]^{-1} \frac{\partial W_n(q_0)}{\partial q} + o_p(1) \quad (3.5)$$

where r_n denotes the rate of convergence, $\hat{q} = \hat{q}_G(\tau; y)$ is the estimated posterior quantile, and $q_0 = q_G(\tau; y)$ its population counterpart. This expansion is analogous to the classical Bahadur-Kiefer representation of the sample quantiles (Bahadur, 1966; Kiefer, 1967), and more generally to the linearization of M-estimators (see e.g. Section 3.2.4 in van der Vaart and Wellner, 1996).

However, establishing the asymptotic expansion in (3.5) is delicate when the functional of interest is possibly irregular, as is often the case in ill-posed inverse problems. One can invoke stochastic equicontinuity and empirical process arguments to control the nonlinear remainder term,¹⁶ but this

¹⁶See Assumption 3.5 in Chen and Pouzo (2015) for a formal statement of the required high-level conditions.

approach has a direct implication on the rates of the functional estimation.¹⁷ We instead exploit the convexity of the sample criterion in (3.4) and apply the convexity lemma of Pollard (1991); Hjort and Pollard (2011), thereby avoiding a separate stochastic equicontinuity argument. The key step is to establish a uniform quadratic approximation of the localized criterion on compact sets; the convexity lemma then implies that the sample argmin is close to the argmin of this approximation. Thus, the proof relies on the local uniform control already established for the deconvolution kernel density estimator, rather than additional high-level empirical process conditions.

Having established the asymptotic expansion (3.5), the rate of convergence is governed by the leading term $\partial W_n(q_0)/\partial q$. The same Fourier-space heuristic described after (3.2) applies here: once the criterion function is linearized, this term behaves like a functional with a non-smooth weighting function. Its Fourier coefficient therefore decays too slowly to offset the exponential decay of $f_\varepsilon^*(t)$ as $|t| \rightarrow \infty$, resulting in a logarithmic convergence rate. For comparison, Appendix C.2 studies the analogous kernel-based estimator for the posterior mean. There the leading term is associated with a smooth weighting function whose Fourier coefficient does offset this decay, so with a suitable bandwidth choice the estimator achieves a nearly parametric rate.

Remark 3.3 (Nonsmooth criterion function). Our setting presents an ill-posed inverse problem characterized by a non-smooth criterion function. Instead of tackling the non-smoothness componentwise, we linearize $W_n(q)$ around the true parameter. This builds on the insight that the non-smoothness may “average out” in the limit (Huber, 1967; Pollard, 1985; Newey and McFadden, 1994, Chapter 7; Chen, Linton, and Van Keilegom, 2003). However, unlike regular models, the (non)-smoothness of the criterion function crucially affects the rate of convergence of our estimator.¹⁸ Moreover, since our estimator already attains its optimal rate, smoothing the criterion – as in Horowitz (1998) – cannot improve convergence rates without additional assumptions. ■

Remark 3.4 (Alternative two-step estimators). Another widely used estimation strategy in the empirical Bayes literature is called “ g -modeling” (Koenker and Gu, 2019). It proceeds in two steps. First, one obtains an estimator of the mixing distribution $\hat{G}$ by maximizing generalized or sieve maximum likelihood (Jiang and Zhang, 2009; Koenker and Mizera, 2014; Efron, 2016). Second, the posterior quantile $q_{\hat{G}}(\tau; y)$ is computed based on plug-in $\hat{G}$. We do not expect these two-step procedures to improve the rate of convergence due to Theorem 3.1. ■

¹⁷The issue arises for regular functionals in severely ill-posed inverse problems. Consider a two-step procedure for estimating the posterior mean, where a deconvolution kernel density estimator is used as a plug-in. In this context, stochastic equicontinuity arguments typically bound the nonlinear remainder term by requiring uniform control over the first-stage estimate, i.e. $\|\hat{g} - g\|_\infty$, which results in a suboptimal convergence rate for the posterior mean itself.

¹⁸The reason is that the second-order expansion of our estimator includes the difference between deconvolution density and its sample counterpart. To ensure this term converges, one must choose the bandwidth at a logarithmic rate, which renders the leading term and thus the estimator converges at a logarithmic rate. In a regular model such as quantile regression, the leading term in its asymptotic expansion converges at a parameter rate. Although the second-order term involves the difference between regression error density and its sample counterpart, it does not affect the parameter rate of the leading term. As a result, quantile regression estimator achieves the parametric rate.

3.3 Conditional coverage of the feasible NP-EBCI

We now examine the conditional coverage of the feasible NP-EBCI. Because the target is the conditional coverage probability $P_G(\theta_i \in \text{CI}_i \mid Y_i = y)$, we evaluate CI_i at the realized observation Y_i , while estimating its endpoints from the remaining sample $\mathbf{Y}_{-i} := \{Y_j\}_{j \neq i}$. This leave-one-out construction separates the observation being conditioned on from the data used to learn the prior.

Let $\hat{q}_G(\tau; Y_i)$ denote the leave-one-out analogue of the posterior quantile estimator in Equation (3.4). We define the feasible NP-EBCI as

$$\text{CI}_i^{\text{NP}} = \left[\hat{q}_G(\alpha/2; Y_i), \hat{q}_G(1 - \alpha/2; Y_i) \right]. \quad (3.6)$$

To evaluate accuracy, we consider the error in conditional coverage probability, that is the difference between the actual and nominal coverage probabilities. The next theorem characterizes the best possible rate at which this error can vanish uniformly over the function class $\mathcal{G}(s, L)$.

Theorem 3.3. *Suppose Assumptions 3.1 to 3.5 hold. Let $\mathcal{C}$ denote the class of all two-sided intervals constructed based on the data $\mathbf{Y}_{-i} = \{Y_j\}_{j \neq i}$, that is the collection of $[\ell(\mathbf{Y}_{-i}), U(\mathbf{Y}_{-i})]$ with $\ell(\mathbf{Y}_{-i}) \leq U(\mathbf{Y}_{-i})$ almost surely, then we have that*

$$\inf_{\text{CI}_i \in \mathcal{C}} \sup_{g \in \mathcal{G}(s, L)} \mathbb{E}_{\mathbf{Y}_{-i}} \left[(P_G(\theta_i \in \text{CI}_i(\mathbf{Y}_{-i}) \mid Y_i = y) - (1 - \alpha))^2 \right] \geq \text{const} \cdot (\log n)^{-(2s+1)/2}$$

for n sufficiently large.

Theorem 3.3 provides a minimax lower bound on the errors in conditional coverage probabilities. The infimum is taken over all two-sided intervals constructed from the data, and the supremum is over all mixing densities satisfying Assumption 3.2. Thus, if conditional coverage is the target criterion, no leave-one-out two-sided interval can have a coverage error that vanishes faster than a logarithmic rate. In this sense, Theorem 3.3 reveals the fundamental limits on the accuracy of nonparametric empirical Bayes inference if the conditional coverage is the target criterion.

The proof of Theorem 3.3 relies on Fano's inequality (Scarlett and Cevher, 2019). Unlike the lower bound construction for mean square error, where alternative data distributions are separated by the induced target parameters, the separation here is based on the conditional coverage probabilities. We construct many alternative mixing densities around a baseline such that the induced observed laws of $\mathbf{Y}_{-i}$ are statistically indistinguishable, while the conditional coverage errors are bounded away from zero at the desired rate. Hence, if an interval had conditional coverage error uniformly smaller than this rate, it could decode which alternative generated the data. But Fano's inequality precludes this possibility because the alternatives are statistically close.

Remark 3.5. The minimax rates of errors in coverage probabilities of *frequentist* confidence intervals have been considered in Hall and Jing (1995); Calonico, Cattaneo, and Farrell (2022). Their notion of coverage is different from ours. In their setting, the parameter is fixed and coverage

refers to the sampling probability that the interval contains this fixed parameter. In our setting, the object of inference is the random, unobservable individual effect θ_i , and Theorem 3.3 concerns about the conditional coverage probability. ■

We next show that the feasible NP-EBCI (3.6) achieves the rate in the lower bound established in Theorem 3.3. Because we construct the feasible interval by replacing infeasible posterior quantiles with their estimators, our task is to translate the theoretical results of posterior quantile estimation (Theorem 3.2) into a statement about conditional coverage. To do so, we impose one additional regularity condition on the posterior density at the relevant quantiles.

Assumption 3.7. *We assume that the posterior density (3.3) evaluated at the posterior quantile is finite, i.e. $\pi(q(\tau; y)|Y = y) < \infty$.*

Theorem 3.4. *Let Assumptions 3.1 to 3.7 hold. Setting the bandwidth $h_n \asymp (\log n)^{-1/2}$, the proposed individual interval (3.6) is a $100(1 - \alpha)\%$ asymptotic conditional empirical Bayes confidence interval:*

$$P\left(\theta_i \in \text{CI}_i^{\text{NP}} \mid Y_i\right) = 1 - \alpha + O_p\left((\log n)^{-(2s+1)/4}\right)$$

that holds uniformly over $g \in \mathcal{G}(s, L)$.

Theorem 3.4 shows that the conditional coverage probability of the feasible NP-EBCI (3.6) asymptotically fully adapts to the nominal $1 - \alpha$, but differs at a logarithmic rate. The logarithmic conditional coverage error comes from that the endpoints themselves can only be estimated at a logarithmic rate. Nevertheless, Theorems 3.3 and 3.4 together establish the optimal rate at which the worst-case conditional coverage probability converges to the nominal level, and the feasible NP-EBCI (3.6) attains this rate.

Remark 3.6. The feasible NP-EBCI (3.6) replaces the infeasible endpoints with their empirical analogs. A limitation is that it does not take into account the sampling uncertainty from estimating posterior quantile. Laird and Louis (1987); Carlin and Gelfand (1990) propose adjustments to incorporate this source of uncertainty in parametric EBCIs within the normal-normal model. Their methods can be adapted to our setting where the prior is fully nonparametric and may improve the finite-sample performance of coverage probabilities. However, such adjustments cannot improve the rate of errors in conditional coverage probabilities due to Theorem 3.3. ■

3.4 Marginal coverage of the feasible NP-EBCI

We turn to the analysis of the marginal coverage probability of the feasible NP-EBCI (3.6). As this coverage statement is defined with respect to the joint distribution of both θ and $\{Y_i\}_{i=1}^n$, our analysis requires additional regularity conditions that hold uniformly for $y \in \mathcal{Y} \subset \mathbb{R}$.

Assumption 3.8. *The normalized posterior density evaluated at $q(\tau; y)$, i.e. $\phi(y - q(\tau; y)/\sigma)g(q(\tau; y))$ is bounded below uniformly over $y \in \mathcal{Y}$. That is, there exists $\underline{\pi} > 0$ such that $\inf_{y \in \mathcal{Y}} |\phi(y - q(\tau; y)/\sigma)g(q(\tau; y))| \geq \underline{\pi}$.*

Assumption 3.9. *The expected posterior density evaluated at posterior quantile $q(\tau; y)$ is finite, that is $\int \pi(q(\tau; y)|Y = y) dF_Y(y) < \infty$.*

Assumption 3.8 is a standard condition used to establish L_2 or uniform convergence in the quantile regression literature (see, e.g., Assumption 3 in Horowitz and Lee (2005), Condition D1 in Belloni and Chernozhukov (2011)). Assumption 3.9 is weaker than the uniform version of Assumption 3.7, requiring only that the expectation be finite.

Theorem 3.5. *Let Assumptions 3.1, 3.2, 3.4 to 3.6, 3.8 and 3.9 hold. Setting the bandwidth $h_n \asymp (\log n)^{-1/2}$, the marginal coverage probability of individual confidence interval (3.6) satisfies that*

$$\mathbb{P}(\theta_i \in \text{CI}_i^{\text{NP}}) = 1 - \alpha + O((\log n)^{-(2s+1)/4})$$

that holds uniformly over $g \in \mathcal{G}(s, L)$, and the probability measure $\mathbb{P}$ is with respect to the joint distribution of $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ and the data $\mathbf{Y} = (Y_1, \dots, Y_n)$.

For the oracle NP-EBCI, marginal coverage follows immediately from conditional coverage by iterated expectations. Theorem 3.5, however, concerns the feasible NP-EBCI whose endpoints are estimated. To obtain the marginal coverage result for this feasible interval, we need to control the errors in conditional coverage uniformly over $y \in \mathcal{Y}$ so they can be averaged over the distribution of Y_i . The averaging does not improve the convergence rate, since the dominant error still comes from estimating posterior quantiles, which suffers from the severe ill-posedness. Hence the errors in the coverage probability of the feasible NP-EBCI (3.6) vanish at the same logarithmic rate under both marginal and conditional coverage.

Theorem 3.5 gives the accuracy assessment about the feasible NP-EBCI under a fully non-parametric prior. The resulting slow logarithmic rate again arises from the intrinsic difficulty of nonparametric estimation in ill-posed inverse problems. It is useful to compare this result with the two alternatives discussed above: Cox-Morris parametric EBCIs and robust EBCIs of Armstrong et al. (2022).

We first compare Theorem 3.5 with the marginal coverage result for Cox-Morris parametric EBCIs. Under a correctly specified Gaussian prior, Cox-Morris EBCIs can achieve a parametric rate for errors in marginal coverage probabilities; see e.g. Theorem 1 in Yoshimori and Lahiri (2014). This advantage, however, relies on correct specification of the normal-normal model. When the true prior is non-Gaussian, especially in low signal-to-noise ratio regimes, Cox-Morris EBCIs can substantially undercover. The feasible NP-EBCI avoids this source of parametric misspecification by leaving the prior fully nonparametric. The price is that its marginal coverage error vanishes only at a logarithmic rate.

We then compare our result with AKP robust EBCIs. The oracle NP-EBCI targets exact marginal coverage under the true, point-identified prior. AKP instead calibrate their robust EBCI to guarantee marginal coverage over a moment class of priors, and their Theorem 4.1 implies that,

under exchangeability, the feasible interval satisfies $\liminf_{n \rightarrow \infty} \mathbb{P}(\theta_i \in \text{CI}_i^{\text{AKP}}) \geq 1 - \alpha$ (Armstrong et al., 2022). Because this calibration relies only on finite-dimensional moments, rather than on the full prior distribution, the statistical problem is simpler than posterior quantile estimation. Furthermore, their Theorem 4.1 requires only the convergence of the relevant moments, while our Theorem 3.5 imposes a stronger continuity condition on the prior distribution. The gain of AKP is robustness over the moment class; the cost is conservatism, since their EBCIs need not be asymptotically exact. The NP-EBCI instead targets exact marginal coverage under the true point-identified prior, but pays the logarithmic price of the severely ill-posed inverse problem.

Taken together, Theorem 3.5 clarifies the trade-offs behind the feasible NP-EBCI. Relative to the Cox-Morris parametric EBCI, it sacrifices the faster coverage error rates achievable under a correctly specified Gaussian prior in exchange for robustness to prior misspecification. Relative to the AKP robust EBCI, it gives up worst-case moment-class robustness in exchange for asymptotically exact marginal coverage under a point-identified prior. Section 5 examines how these trade-offs play out in finite samples.

4 Practical Implementation under heteroskedasticity

This section returns to the baseline heteroskedastic normal mean model in (2.1) and describes how to implement the proposed NP-EBCI when units have known but unequal variances σ_i^2 . Section 3 imposed the homoskedastic restriction $\sigma_i^2 \equiv \sigma^2$ to focus on the optimal rate results. In empirical applications, however, heteroskedasticity is the natural default. Accordingly, in this section, we outline the practical implementation of our approach and discuss bandwidth selection.

4.1 Baseline implementation

We now describe the baseline implementation of the proposed NP-EBCI under heteroskedastic Gaussian noise. Let $\mathcal{D}_i = (Y_i, \sigma_i)$. The oracle posterior quantile can be written as

$$q_G(\tau; \mathcal{D}_i) = \arg \min_q \int \overline{M^*(q, t; \mathcal{D}_i)} g^*(t) dt$$

where $\overline{M^*(q, t; \mathcal{D}_i)} = \int \exp(-it\theta) \rho_\tau(\theta - q) \phi((Y_i - \theta)/\sigma_i) d\theta$. The Fourier transform of the mixing density $g^*(t)$ can be identified from the following population identity

$$\frac{1}{n-1} \sum_{j \neq i} \mathbb{E}[\exp(itY_j)] = g^*(t) \left[\frac{1}{n-1} \sum_{j \neq i} \exp\left(-\frac{1}{2}\sigma_j^2 t^2\right) \right]. \quad (4.1)$$

Here, the leave-one-out form is used for estimation rather than identification: it allows us to estimate the posterior quantile from the other units, while $\mathcal{D}_i$ enters only through the evaluation step.

The representation (4.1) leads to a natural plug-in estimator. Define

$$\hat{f}_Y^*(t) = \frac{1}{n-1} \sum_{j \neq i} \exp(itY_j), \quad \bar{f}_\varepsilon^*(t) = \frac{1}{n-1} \sum_{j \neq i} \exp\left(-\frac{1}{2}\sigma_j^2 t^2\right).$$

For a given bandwidth h_n , we estimate the posterior quantile by

$$\hat{q}_G(\tau; \mathcal{D}_i) = \arg \min_q \int \frac{\overline{M^*(q, t; \mathcal{D}_i)}}{\bar{f}_\varepsilon^*(t)} \left(\hat{f}_Y^*(t) K^*(h_n t) \right) dt$$

As in the homoskedastic case, the kernel $K^*(h_n t)$ regularizes the Fourier integral. In practice, we use the compactly supported flat-top spectral kernel (Politis and Romano, 1999) but incorporate a cosine taper (Harris, 1978) to smooth the transition from the flat-top region to zero and avoid a hard spectral cutoff. We discuss the choice of the bandwidth parameter h_n in Section 4.2.

Using the estimated posterior quantiles, we construct the NP-EBCI

$$\text{CI}_i = \left[\hat{q}_G(\alpha/2; \mathcal{D}_i), \hat{q}_G(1 - \alpha/2; \mathcal{D}_i) \right], \quad i = 1, \dots, n. \quad (4.2)$$

So far, we have presented the baseline implementation for heteroskedastic variances treated as fixed and known. The same construction also covers the case when σ_i^2 is random across units and independent with θ_i . To see this, under $\theta_i \perp \sigma_i$, we have $E[\exp(itY_i)|\sigma_i] = g^*(t) \exp(-\frac{1}{2}\sigma_i^2 t^2)$, so averaging over the realized units gives the same population identity in (4.1). In this sense, the baseline implementation can also be viewed as conditioning on observed heteroskedastic variances. Appendix C.3 discusses the extension to the case in which θ_i and σ_i^2 may be correlated.

4.2 Empirical choice of the bandwidth parameter

We choose a common bandwidth h_n to minimize average interval length subject to an average coverage requirement.¹⁹ Since this criterion depends on the unknown mixing distribution G , we implement it by a plug-in rule that replaces G with its nonparametric maximum likelihood estimator (NPMLE) $\hat{G}(\cdot)$ (Soloff, Guntuboyina, and Sen, 2025). In this respect, our bandwidth selection method is close in spirit to plug-in or rule-of-thumb bandwidth choice in nonparametric kernel density estimation (see e.g. Li and Racine, 2007, Section 2.7). However, unlike these methods relying on preliminary pilot values, our procedure uses the plug-in NPMLE $\hat{G}$, which itself is a consistent estimator of the mixing distribution G (Kiefer and Wolfowitz, 1956; Pfanzagl, 1988) and requires *no* additional smoothing tuning parameter (Polyanskiy and Wu, 2020).

It is important to distinguish the role of the NPMLE in our implementation from the two-step plug-in approach discussed in Remark 3.4. Our feasible NP-EBCI is based on the one-step kernel

¹⁹One could, in principle, choose a separate bandwidth for each CI_i by minimizing the error of conditional coverage probability of that interval. This selection method is natural but computationally impractical because it requires solving n distinct bandwidth selection problems.

estimator for posterior quantiles. The NPMLE is only used for bandwidth selection. We do not suggest estimating posterior quantiles with the NPMLE plug-in. In simulations not reported here, the NP-EBCI based on these plug-in estimates can substantially undercover, which aligns with the simulation evidence in [Koenker \(2020\)](#). One plausible explanation is that the NPMLE has discrete support ([Lindsay, 1983](#)), so plugging it directly into the posterior quantile can yield an overly concentrated posterior distribution and hence intervals that are too short.

The bandwidth selection procedure is detailed as follows:

Step 1. Partition the units $\{1, \dots, n\}$ into V disjoint folds $\mathcal{I}_1, \dots, \mathcal{I}_V$, and let $\mathcal{I}_v^c = \{1, \dots, n\} \setminus \mathcal{I}_v$. For each fold $v = 1, \dots, V$, we compute the leave- v -out nonparametric maximum likelihood estimator for the mixing distribution $\hat{G}^{(-v)}(\cdot)$.

Step 2. Let $\mathcal{H}_n$ be a grid of candidate bandwidths. For each $h \in \mathcal{H}_n$, each fold $v = 1, \dots, V$ and each held-out unit $i \in \mathcal{I}_v$, we estimate the lower and upper endpoints $\hat{q}_h^{(-v)}(\alpha/2; \mathcal{D}_i)$ and $\hat{q}_h^{(-v)}(1 - \alpha/2; \mathcal{D}_i)$ on the fold $\mathcal{I}_v^c$, but evaluate at the held-out unit $\mathcal{D}_i$.

Step 3. For each $h \in \mathcal{H}_n$, we compute the average conditional coverage and average length of intervals under the out-of-fold prior estimate $\hat{G}^{(-v)}(\cdot)$:

$$\begin{aligned}\widehat{C}(h) &= \frac{1}{n} \sum_{v=1}^V \sum_{i \in \mathcal{I}_v} \mathbb{P}_{\hat{G}^{(-v)}} \left(\theta_i \in \left[\hat{q}_h^{(-v)}(\alpha/2; \mathcal{D}_i), \hat{q}_h^{(-v)}(1 - \alpha/2; \mathcal{D}_i) \right] \mid \mathcal{D}_i \right) \\ \widehat{L}(h) &= \frac{1}{n} \sum_{v=1}^V \sum_{i \in \mathcal{I}_v} \left[\hat{q}_h^{(-v)}(1 - \alpha/2; \mathcal{D}_i) - \hat{q}_h^{(-v)}(\alpha/2; \mathcal{D}_i) \right]\end{aligned}$$

Step 4. We set the bandwidth that yields the shortest average-length interval among those that satisfy the nominal coverage constraint:

$$\hat{h} = \arg \min_{h \in \mathcal{H}_n} \widehat{L}(h) \quad \text{s.t.} \quad \widehat{C}(h) \geq 1 - \alpha.$$

Appendix [B](#) provides further strategies that can substantially reduce the computational cost of this bandwidth selection procedure.

5 Monte Carlo Simulations

Both NP-EBCI and the AKP robust EBCI move beyond the parametric Gaussian prior, but use the information about the prior distribution in different ways. NP-EBCI targets the oracle posterior interval under a point-identified nonparametric prior and thus exploits the entire distribution, but suffers from the severe ill-posedness. The AKP robust EBCI, by contrast, calibrate intervals to guarantee worst-case coverage uniform over a moment class of priors, so the resulting intervals may be conservative when the true prior is not least favorable. This section uses Monte Carlo simulations to examine how these differences appear in finite samples. We compare the feasible NP-EBCI with

AKP (using the second moment) and the Cox-Morris parametric EBCI, and summarize performance by average marginal coverage and average length reduction relative to the naive z -interval.

5.1 Monte Carlo Design

We generate simulated data from the heteroskedastic normal means model (2.1) with known variances:

$$Y_i|\boldsymbol{\theta} \sim N(\theta_i, \sigma_i^2), \quad \theta_i \stackrel{\text{i.i.d.}}{\sim} G(\cdot), \quad \text{for } i = 1, \dots, n.$$

We consider sample sizes $n \in \{100, 200, 500, 1000\}$ and target signal-to-noise ratios

$$\text{SNR} = \frac{\text{Var}(\theta_i)}{\frac{1}{n} \sum_{i=1}^n \sigma_i^2} \in \{0.1, 0.2, 0.4, 1\}.$$

For each design cell (G, n, SNR) , we first draw $\sigma_i^2 \stackrel{\text{i.i.d.}}{\sim} \text{Lognormal}(0, 1)$, $i = 1, \dots, n$, and rescale it to attain the target SNR, and then keep the resulting variance vector fixed across all Monte Carlo replications. We then generate 500 independent replications by redrawing only $\{Y_i, \theta_i\}_{i=1}^n$.

We consider eight designs for the mixing distribution G :

Design 1. Gaussian: $\theta_i \sim N(0, 1)$

Design 2. Laplace: $\theta_i \sim \text{Laplace}(0, 1/\sqrt{2})$

Design 3. Student- t : $\theta_i \sim \sqrt{0.2}t_{2.5}$

Design 4. Bimodal Gaussian mixture: $\theta_i \sim 0.5N(-5/\sqrt{29}, 4/29) + 0.5N(5/\sqrt{29}, 4/29)$

Design 5. Spike-and-slab: $\theta_i \sim 0.9N(0, 0.05^2) + 0.1N(0, 3.159^2)$

Design 6. 3-point distribution that places masses $(0.25, 0.5, 0.25)$ on $(-\sqrt{2}, 0, \sqrt{2})$

Design 7. The least favorable distribution for AKP robust EBCI

Design 8. The least favorable distribution for Cox-Morris parametric EBCI

All designs for the mixing distribution G are normalized to have unit variance, so differences across designs are driven by the shape of G . Design 1 is the Gaussian benchmark under which the Cox-Morris normal-normal model is correctly specified. Designs 2 and 3 introduce heavier tails and hence represent progressively larger departures from the Gaussian benchmark. Designs 1 and 3 have supersmooth mixing densities in the Fourier sense, so they are relatively favorable cases for nonparametric deconvolution (Lacour, 2006), and thus for feasible NP-EBCI. Design 4 is included to examine how prior multimodality affects the posterior quantiles that form NP-EBCI. Design 5 is a nearly sparse prior with most mass concentrated near zero and a small fraction of large effects.

Designs 6-8 are discrete priors, thus violate the absolute continuity condition on G (Assumptions 3.1, 3.2 and 3.4), which underlies our theoretical analysis of NP-EBCI. Design 7 is the least favorable prior for the AKP robust EBCI (Design 5, Armstrong et al., 2022). Among priors with the

same second moment, it attains the smallest marginal coverage and therefore comes closest to the worst-case coverage bound. Design 8 is the least favorable prior for Cox-Morris parametric EBCI (Design 6, [Armstrong et al., 2022](#)), under which the interval attains its lowest marginal coverage.

For each EBCI method, we focus on the average marginal coverage and the average length reduction relative to the naive z -interval, both averaged across units $i = 1, \dots, n$ and across all Monte Carlo replications. The nominal coverage level is $1 - \alpha = 95\%$ throughout.

5.2 Results and Discussion

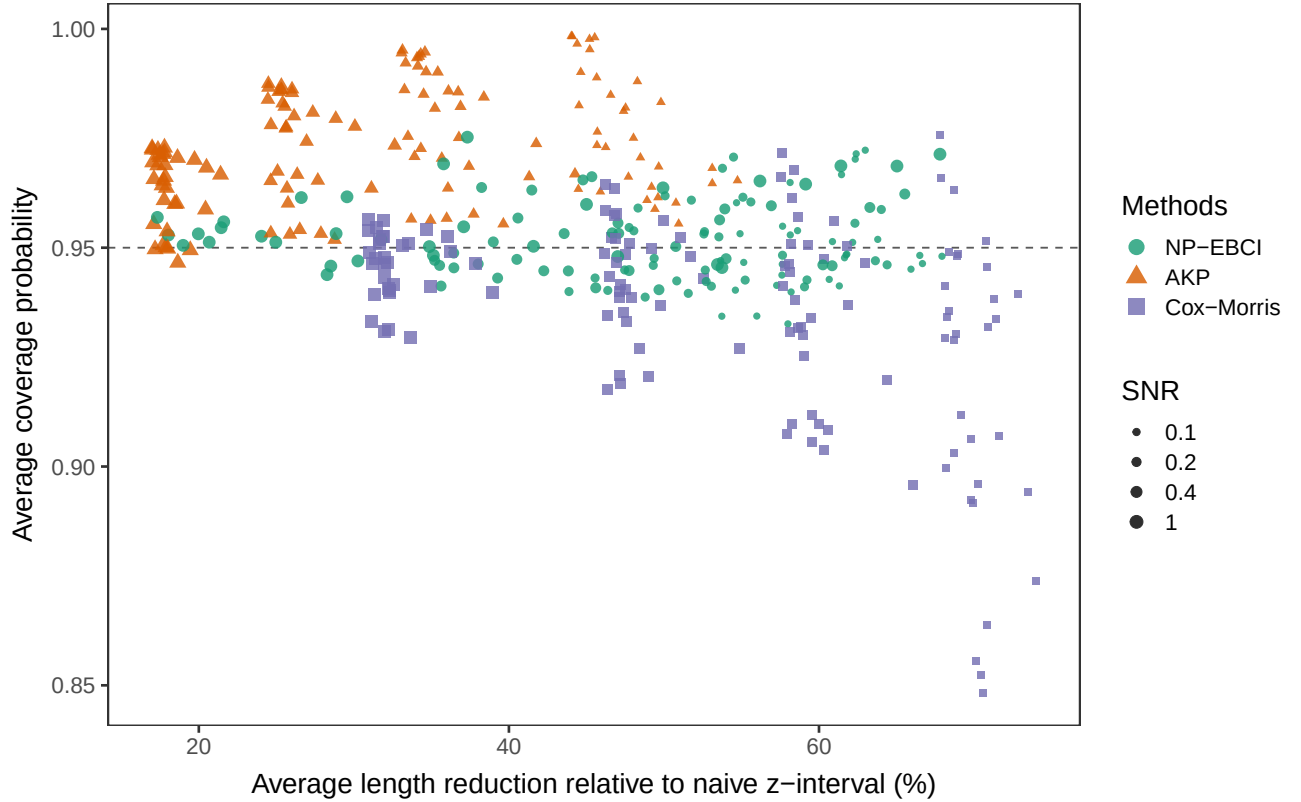


Figure 1: *Average coverage probability and length reduction across simulation designs.*

Note. Each point corresponds to one DGP $\times$ sample size $\times$ SNR cell. The dashed horizontal line is the nominal coverage level $1 - \alpha = 95\%$.

Figure 1 summarizes the full simulation grid; each point corresponds to one DGP $\times$ sample size $\times$ SNR cell.²⁰ First, all three EBCI procedures deliver substantial average length reductions relative to the naive z -interval, reflecting the classical empirical Bayes gain from borrowing strength across related units ([Morris, 1983](#)). Second, Figure 1 reveals a clear ordering among the three EBCI

²⁰Monte Carlo standard errors are small: across the three EBCI procedures, they are at most 0.17 percentage points for average coverage and 0.45 percentage points for average length reduction.

procedures in terms of average coverage and length. AKP (second moment) is conservative: many of its points lie above the nominal 95% coverage line, but its length reductions are comparatively modest. Cox-Morris parametric EBCI often delivers the largest length reductions, but many of these gains come with substantial undercoverage. NP-EBCI achieves large length reductions while keeping empirical marginal coverage much closer to the nominal level.

Figure 1 further shows that the average coverage and length of EBCIs depend strongly on the signal-to-noise ratio (SNR). Lower-SNR cells tend to exhibit larger average length reductions for all three EBCIs, which is consistent with the fact that empirical Bayes pooling is most valuable when noise is large relative to signal (Morris, 1983). These same cells also generate the sharpest differences in coverage across methods. At SNR=0.1, AKP tends to lie above the nominal 95% line, Cox-Morris falls below it in many cells, and NP-EBCI remains much closer to the nominal 95% line. This pattern motivates a closer examination of the low-SNR regime in Table 1.

Table 1: Average coverage probabilities and average length reductions for SNR = 0.1

DGP	Average coverage			Average length reduction (%)		
	NP-EBCI	AKP	Cox-Morris	NP-EBCI	AKP	Cox-Morris
<i>Panel A. $n = 100$, SNR = 0.1</i>						
Gaussian	0.947	0.964	0.848	53.7	49.2	70.6
Laplace	0.940	0.961	0.864	54.6	49.6	70.9
$t_{2.5}$	0.953	0.965	0.874	58.1	54.7	74.0
Bimodal	0.960	0.971	0.856	54.7	48.5	70.2
Spike-slab	0.954	0.965	0.939	58.6	53.1	72.9
3-point	0.953	0.961	0.852	54.9	48.9	70.4
Least favorable for AKP	0.947	0.960	0.934	55.1	50.8	71.4
Least favorable for Cox-Morris	0.934	0.955	0.907	53.7	51.0	71.6
<i>Panel B. $n = 1000$, SNR = 0.1</i>						
Gaussian	0.949	0.997	0.941	61.8	44.4	68.1
Laplace	0.941	0.990	0.934	61.3	44.6	68.3
$t_{2.5}$	0.946	0.988	0.952	66.7	48.3	70.8
Bimodal	0.972	0.998	0.976	63.0	44.0	67.8
Spike-slab	0.948	0.973	0.949	67.9	45.7	69.0
3-point	0.970	0.998	0.966	62.3	44.1	67.9
Least favorable for AKP	0.952	0.963	0.929	63.8	44.5	68.1
Least favorable for Cox-Morris	0.941	0.983	0.900	57.3	44.5	68.2

Note. Nominal average confidence level is $1 - \alpha = 95\%$. Average CI length reduction is measured relative to the naive z -interval. For each DGP, average coverage and average length reduction refer to averages across units $i = 1, \dots, n$ and across 500 Monte Carlo repetitions. The cells shaded with gray colors highlight the undercoverage.

Table 1 focuses on the regime SNR = 0.1. Panel A reports the results for small samples $n = 100$. NP-EBCI maintains average coverage near the nominal 95% level across all designs (0.934 to 0.960), while reducing the average length by 53% to 58.6% relative to the naive z -interval. Although Cox-

Morris EBCIs achieve larger average length reductions (70.2% to 74.0%), these gains come with substantial undercoverage in most cells, even in the Gaussian design, coverage falls to 0.848. One possible explanation is that when the prior variance A is small relative to σ_i^2 , the linear shrinkage factor $w_i = A/(A + \sigma_i^2)$ is also correspondingly small. Consequently, if the estimate $\hat{A}$ is biased downward in small samples, both the estimated posterior means and variances shrink aggressively toward zero, producing artificially narrow intervals (see e.g. [Li and Lahiri \(2010\)](#)).

Panel B of Table 1 reports the results for large samples $n = 1000$. NP-EBCI continues to maintain average coverage close to the nominal 95% level (0.941 to 0.972), while reducing average length by 57.3% to 67.9%. By contrast, AKP (second moment) substantially overcovers in nearly every design (0.973 to 0.998) except its own least-favorable design, and yields modest length reductions (44.0%-48.3%). While Cox-Morris EBCI produces the shortest intervals (reductions of 67.8% to 70.8%) and improves coverage compared to Panel A, it still undercovers markedly in some specific least-favorable designs. Furthermore, as [Armstrong et al. \(2022\)](#) emphasize, the Cox-Morris EBCI’s empirical Bayes coverage can drop to 0.74 for the nominal 95% level in the low-SNR limit.

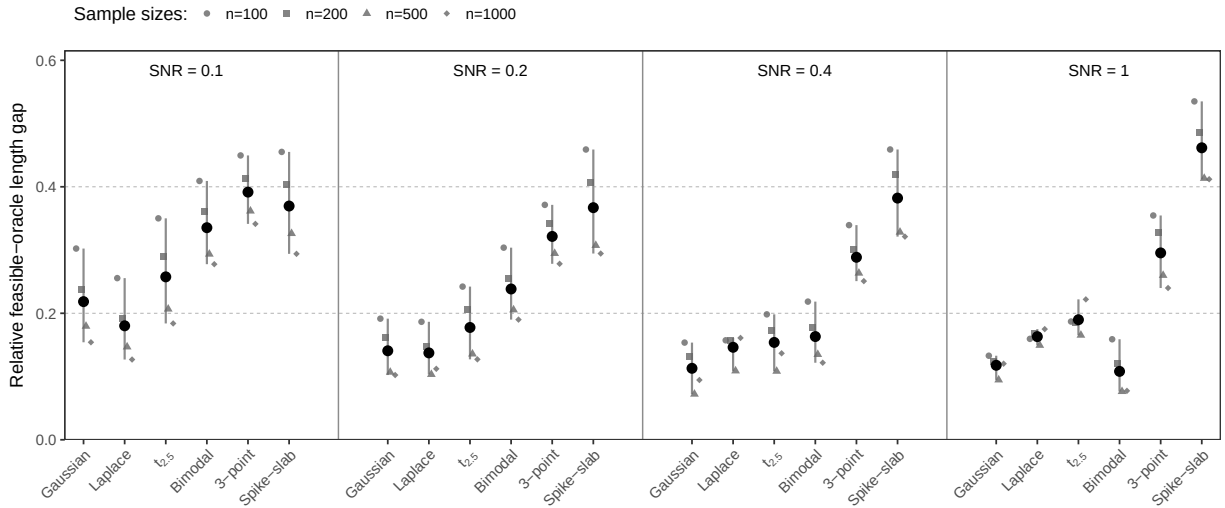


Figure 2: *Relative feasible-oracle length gap of NP-EBCI across prior designs and SNRs.*

Note. The y -axis “oracle length gap” is defined in (5.1). Gray markers correspond to $n = 100, 200, 500, 1000$. The black circle is the average across these sample sizes, and the vertical segment gives the range across them.

The strong finite-sample performance of feasible NP-EBCI in average marginal coverage may seem at odds with the logarithmic ill-posedness established in Theorem 3.5. It is useful to notice that the bandwidth choice in Section 4.2 is designed to minimize average interval length subject to an average coverage requirement, rather than to minimize the mean squared errors of estimating posterior quantiles. Such coverage-oriented regularization may keep average marginal coverage close to the nominal level, with the finite-sample cost of ill-posedness showing up primarily as excess length relative to oracle NP-EBCI rather than as severe undercoverage. For this reason, we

compare the average length of feasible NP-EBCI with that of its oracle counterpart in Figure 2.

In Figure 2, we define “relative feasible-oracle length gap” as

$$\left[\overline{\text{Length}}(\text{CI}_i^{\text{NP}}) - \overline{\text{Length}}(\text{CI}_i^{\text{NP}^*}) \right] / \overline{\text{Length}}(\text{CI}_i^{\text{NP}}) \quad (5.1)$$

where $\overline{\text{Length}}(\text{CI}_i^{\text{NP}})$ and $\overline{\text{Length}}(\text{CI}_i^{\text{NP}^*})$ denote, respectively, the average lengths of the feasible NP-EBCI in Equation (4.2) and its oracle counterpart in Equation (2.3).

Figure 2 first shows that the relative feasible-oracle length gap of NP-EBCI changes only marginally with sample size n . This aligns with the logarithmic ill-posedness of posterior quantile estimation being driven primarily by the structure of the inverse problem rather than by the sample size. Notably, this feasible-oracle gap varies significantly more across prior designs than across SNRs. In every SNR panel, the gap is small for the smooth priors but largest for the discrete and spike-and-slab priors. Although the gap varies slightly with SNR within a given design, the ordering across designs is quite stable. This pattern suggests that the relevant ill-posedness is governed mainly by the shape of G : smooth priors are comparatively favorable, while priors with sharp local structure, such as point masses or narrow spikes, make posterior quantiles harder to estimate.

6 Conclusion

This paper proposes empirical Bayes confidence intervals based on posterior quantiles for unobservable individual effects under a fully nonparametric prior, and establishes their theoretical properties. On the one hand, the feasible NP-EBCI achieves asymptotically exact conditional and marginal coverage. On the other hand, this flexibility comes at an unavoidable statistical cost: because posterior quantiles inherit the severe ill-posedness of nonparametric deconvolution, the relevant errors in quantile estimation and in coverage probabilities vanish at logarithmic rates. Thus the accuracy of nonparametric empirical Bayes inference is fundamentally limited by the difficulty of underlying inverse problem. Despite these slow asymptotic rates, the simulations show that the feasible NP-EBCI remains much closer to nominal coverage, even in small samples, is less conservative than the AKP robust EBCI, and delivers substantial length reductions relative to the naive z -interval.

While this paper focuses on empirical Bayes confidence intervals, an important direction for future work is to study other empirical Bayes decision problems that involve posterior quantiles. As Walters (2024) emphasizes, posterior quantiles can be more appropriate when the objective is to identify units in the tails of the distribution while controlling the risk of costly selection errors. This concern arises naturally in empirical Bayes compound decision problems such as ranking and selection; see e.g. Gu and Koenker (2023). An interesting question for future research is whether our minimax results can be extended to characterize the statistical difficulty of such tail-oriented empirical Bayes decision problems.

A useful feature of the NP-EBCI construction is that it relies on point identification of the

distribution of individual effect G , rather than on the normal means structure alone. This suggests another direction for future work: extending posterior-quantile-based empirical Bayes inference to a broader class of latent-variable models in which G remains point identified (Schennach, 2020). Such extensions would help clarify which parts of our theory are specific to Gaussian deconvolution and which reflect a broader difficulty of posterior-quantile-based empirical Bayes inference.

References

- AMITI, M. AND D. E. WEINSTEIN (2018): “How much do idiosyncratic bank shocks affect investment? Evidence from matched bank-firm loan data,” *Journal of Political Economy*, 126, 525–587.
- ANGELOPOULOS, A. N. AND S. BATES (2023): “Conformal Prediction: A Gentle Introduction,” *Foundations and Trends in Machine Learning*, 16, 494–591.
- ARMSTRONG, T. B., M. KOLESÁR, AND M. PLAGBORG-MØLLER (2022): “Robust empirical bayes confidence intervals,” *Econometrica*, 90, 2567–2602.
- BAHADUR, R. R. (1966): “A note on quantiles in large samples,” *The Annals of Mathematical Statistics*, 37, 577–580.
- BELLONI, A. AND V. CHERNOZHUKOV (2011): “ ℓ_1 -penalized quantile regression in high-dimensional sparse models,” *The Annals of Statistics*, 39, 82 – 130.
- BERGER, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics, New York: Springer, 2nd ed.
- BIRNBAUM, A. (1962): “On the foundations of statistical inference,” *Journal of the American Statistical Association*, 57, 269–306.
- BONHOMME, S. AND A. DENIS (2024): “Estimating heterogeneous effects: applications to labor economics,” *Labour Economics*, 91, 102638.
- BROWN, L. D. AND E. GREENSHTEIN (2009): “Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means,” *The Annals of Statistics*, 1685–1704.
- BUEHLER, R. J. (1959): “Some validity criteria for statistical inferences,” *The Annals of Mathematical Statistics*, 30, 845–863.
- BUTUCEA, C. AND F. COMTE (2009): “Adaptive estimation of linear functionals in the convolution model and applications,” *Bernoulli*, 15, 69–98.
- CALONICO, S., M. D. CATTANEO, AND M. H. FARRELL (2022): “Coverage error optimal confidence intervals for local polynomial regression,” *Bernoulli*, 28, 2998–3022.
- CARD, D., J. HEINING, AND P. KLINE (2013): “Workplace heterogeneity and the rise of West

- German wage inequality,” *The Quarterly journal of economics*, 128, 967–1015.
- CARLIN, B. P. AND A. E. GELFAND (1990): “Approaches for empirical Bayes confidence intervals,” *Journal of the American Statistical Association*, 85, 105–114.
- CARLIN, B. P. AND T. A. LOUIS (2000): *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall/CRC Texts in Statistical Science, Boca Raton, FL: Chapman & Hall/CRC, 2nd ed.
- CARROLL, R. J. AND P. HALL (1988): “Optimal rates of convergence for deconvolving a density,” *Journal of the American Statistical Association*, 83, 1184–1186.
- CHATTERJEE, S., P. LAHIRI, AND H. LI (2008): “Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models,” *The Annals of Statistics*, 36, 1221.
- CHEN, J. (2026): “Empirical Bayes when estimation precision predicts parameters,” *Econometrica*, 94, 305–340.
- CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): “Estimation of semiparametric models when the criterion function is not smooth,” *Econometrica*, 71, 1591–1608.
- CHEN, X. AND D. POUZO (2015): “Sieve Wald and QLR inferences on semi/nonparametric conditional moment models,” *Econometrica*, 83, 1013–1079.
- CHENG, X., S. C. HO, AND F. SCHORFHEIDE (2025): “Optimal estimation of two-way effects under limited mobility,” Tech. rep., National Bureau of Economic Research.
- CHETTY, R., J. N. FRIEDMAN, AND J. E. ROCKOFF (2014): “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates,” *American economic review*, 104, 2593–2632.
- CHETTY, R. AND N. HENDREN (2018): “The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects,” *The quarterly journal of economics*, 133, 1107–1162.
- COX, D. R. (1958): “Some problems connected with statistical inference,” *The Annals of Mathematical Statistics*, 29, 357–372.
- (1975): “Prediction intervals and empirical Bayes confidence intervals,” *Journal of Applied Probability*, 12, 47–55.
- CSÖRGŐ, S. AND V. TOTIK (1983): “On how long interval is the empirical characteristic function uniformly consistent,” *Acta Sci. Math.(Szeged)*, 45, 141–149.
- DATTA, G. S., M. GHOSH, D. D. SMITH, AND P. LAHIRI (2002): “On an asymptotic theory of conditional and unconditional coverage probabilities of empirical Bayes confidence intervals,” *Scandinavian Journal of Statistics*, 29, 139–152.
- EFRON, B. (2011): “Tweedie’s formula and selection bias,” *Journal of the American Statistical Association*, 106, 1602–1614.

- (2012): *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, vol. 1, Cambridge University Press.
- (2014): “Two modeling strategies for empirical Bayes estimation,” *Statistical science: a review journal of the Institute of Mathematical Statistics*, 29, 285.
- (2016): “Empirical Bayes deconvolution estimates,” *Biometrika*, 103, 1–20.
- (2019): “Bayes, oracle Bayes and empirical Bayes,” *Statistical science*, 34, 177–201.
- ESCANCIANO, J. C. (2023): “Irregular identification of structural models with nonparametric unobserved heterogeneity,” *Journal of Econometrics*, 234, 106–127.
- FAN, J. (1991): “On the optimal rates of convergence for nonparametric deconvolution problems,” *The Annals of Statistics*, 1257–1272.
- FENIZIA, A. (2022): “Managers and productivity in the public sector,” *Econometrica*, 90, 1063–1084.
- FINKELSTEIN, A., M. GENTZKOW, AND H. WILLIAMS (2016): “Sources of geographic variation in health care: Evidence from patient migration,” *The quarterly journal of economics*, 131, 1681–1726.
- FOLLAND, G. B. (1999): *Real analysis: modern techniques and their applications*, John Wiley & Sons.
- GAILLAC, C. (2024): “Predicting Unobserved Individual-level Causal Effects,” Tech. rep., Working paper.
- GILRAINE, M., J. GU, AND R. MCMILLAN (2020): “A new method for estimating teacher value-added,” Tech. rep., National Bureau of Economic Research.
- GNEITING, T. AND A. E. RAFTERY (2007): “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American statistical Association*, 102, 359–378.
- GONCALVES, F. AND S. MELLO (2021): “A few bad apples? Racial bias in policing,” *American Economic Review*, 111, 1406–1441.
- GU, J. AND R. KOENKER (2017): “Unobserved heterogeneity in income dynamics: An empirical Bayes perspective,” *Journal of Business & Economic Statistics*, 35, 1–16.
- (2023): “Invidious comparisons: Ranking and selection as compound decisions,” *Econometrica*, 91, 1–41.
- HALL, P. AND B.-Y. JING (1995): “Uniform coverage bounds for confidence intervals and Berry-Esseen theorems for Edgeworth expansion,” *The Annals of Statistics*, 23, 363–375.
- HARRIS, F. J. (1978): “On the use of windows for harmonic analysis with the discrete Fourier transform,” *Proceedings of the IEEE*, 66, 51–83.
- HJORT, N. L. AND D. POLLARD (2011): “Asymptotics for minimisers of convex processes,” *arXiv*

preprint *arXiv:1107.3806*.

- HOROWITZ, J. L. (1998): “Bootstrap methods for median regression models,” *Econometrica*, 1327–1351.
- (2009): *Semiparametric and nonparametric methods in econometrics*, vol. 12, Springer.
- (2014): “Ill-posed inverse problems in economics,” *Annu. Rev. Econ.*, 6, 21–51.
- HOROWITZ, J. L. AND S. LEE (2005): “Nonparametric estimation of an additive quantile regression model,” *Journal of the American Statistical Association*, 100, 1238–1249.
- HOROWITZ, J. L. AND M. MARKATOU (1996): “Semiparametric estimation of regression models for panel data,” *The Review of Economic Studies*, 63, 145–168.
- HOUWELINGEN, J. V. AND T. STIJNEN (1983): “Monotone empirical Bayes estimators for the continuous one-parameter exponential family,” *Statistica Neerlandica*, 37, 29–43.
- HUBER, P. J. (1967): “The behavior of maximum likelihood estimates under nonstandard conditions,” in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1: statistics*, University of California Press, vol. 5, 221–234.
- IGNATIADIS, N. AND B. SEN (2025): “Empirical partially Bayes multiple testing and compound χ^2 decisions,” *The Annals of Statistics*, 53, 1–36.
- IGNATIADIS, N. AND S. WAGER (2022): “Confidence intervals for nonparametric empirical Bayes analysis,” *Journal of the American Statistical Association*, 117, 1149–1166.
- JIANG, W. (2019): “Comment: Empirical Bayes Interval Estimation,” *Statistical science*, 34, 219–223.
- JIANG, W. AND C.-H. ZHANG (2009): “General maximum likelihood empirical Bayes estimation of normal means,” *The Annals of Statistics*, 37, 1647–1684.
- KIEFER, J. (1967): “On Bahadur’s representation of sample quantiles,” *The Annals of Mathematical Statistics*, 38, 1323–1342.
- KIEFER, J. AND J. WOLFOWITZ (1956): “Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters,” *The Annals of Mathematical Statistics*, 887–906.
- KLINE, P., E. K. ROSE, AND C. R. WALTERS (2022): “Systemic discrimination among large US employers,” *The Quarterly Journal of Economics*, 137, 1963–2036.
- (2024): “A discrimination report card,” *American Economic Review*, 114, 2472–2525.
- KNIGHT, K. (1998): “Limiting distributions for L_1 regression estimators under general conditions,” *Annals of statistics*, 755–770.
- KOENKER, R. (2005): *Quantile regression*, vol. 38, Cambridge university press.
- (2020): “Empirical Bayes confidence intervals: an R vinaigrette,” .

- KOENKER, R. AND J. GU (2019): “Comment: Minimalist g-Modeling,” *Statistical Science*, 34, pp. 209–213.
- (2024): “Empirical Bayes for the reluctant frequentist,” *arXiv preprint arXiv:2404.03422*.
- KOENKER, R. AND I. MIZERA (2014): “Convex optimization, shape constraints, compound decisions, and empirical Bayes rules,” *Journal of the American Statistical Association*, 109, 674–685.
- KWON, S. (2026): “Optimal shrinkage estimation of fixed effects in linear panel data models,” *Econometrica*, 94, 663–677.
- LACOUR, C. (2006): “Rates of convergence for nonparametric deconvolution,” *Comptes Rendus. Mathématique*, 342, 877–882.
- LAIRD, N. M. AND T. A. LOUIS (1987): “Empirical Bayes confidence intervals based on bootstrap samples,” *Journal of the American Statistical Association*, 82, 739–750.
- LE CAM, L. (1973): “Convergence of estimates under dimensionality restrictions,” *The Annals of Statistics*, 38–53.
- LI, H. AND P. LAHIRI (2010): “An adjusted maximum likelihood method for solving small area estimation problems,” *Journal of multivariate analysis*, 101, 882–892.
- LI, Q. AND J. S. RACINE (2007): *Nonparametric econometrics: theory and practice*, Princeton University Press.
- LI, T. AND Q. VUONG (1998): “Nonparametric estimation of the measurement error model using multiple indicators,” *Journal of Multivariate Analysis*, 65, 139–165.
- LINDSAY, B. G. (1983): “The geometry of mixture likelihoods: a general theory,” *The annals of statistics*, 86–94.
- LIU, L., H. R. MOON, AND F. SCHORFHEIDE (2020): “Forecasting with dynamic panel data models,” *Econometrica*, 88, 171–201.
- MEISTER, A. (2009): *Deconvolution Problems in Nonparametric Statistics*, Lecture Notes in Statistics, Springer Berlin Heidelberg.
- MORRIS, C. N. (1983): “Parametric empirical Bayes inference: theory and applications,” *Journal of the American statistical Association*, 78, 47–55.
- NEWAY, W. K. AND D. MCFADDEN (1994): “Large sample estimation and hypothesis testing,” *Handbook of econometrics*, 4, 2111–2245.
- PAGAN, A. AND A. ULLAH (1999): *Nonparametric Econometrics*, Themes in Modern Econometrics, Cambridge: Cambridge University Press.
- PENSKY, M. (2017): “Minimax theory of estimation of linear functionals of the deconvolution density with or without sparsity,” *The Annals of Statistics*, 45, 1516–1541.

- PFANZAGL, J. (1988): “Consistency of maximum likelihood estimators for certain nonparametric families, in particular: mixtures,” *Journal of Statistical Planning and Inference*, 19, 137–158.
- POLITIS, D. N. AND J. P. ROMANO (1999): “Multivariate density estimation with general flat-top kernels of infinite order,” *Journal of Multivariate Analysis*, 68, 1–25.
- POLLARD, D. (1985): “New ways to prove central limit theorems,” *Econometric Theory*, 1, 295–313.
- (1991): “Asymptotics for least absolute deviation regression estimators,” *Econometric Theory*, 7, 186–199.
- POLYANSKIY, Y. AND Y. WU (2020): “Self-regularizing property of nonparametric maximum likelihood estimator in mixture models,” *arXiv preprint arXiv:2008.08244*.
- ROBBINS, H. (1956): “An Empirical Bayes Approach to Statistics,” in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume I*, ed. by J. Neyman, Berkeley, CA: University of California Press, 157–163.
- (1964): “The empirical Bayes approach to statistical decision problems,” *The Annals of Mathematical Statistics*, 35, 1–20.
- ROBINS, J. AND L. WASSERMAN (2000): “Conditioning, likelihood, and coherence: A review of some foundational concepts,” *Journal of the American Statistical Association*, 95, 1340–1346.
- ROBINSON, G. (1979): “Conditional properties of statistical procedures,” *The Annals of Statistics*, 742–755.
- RUBIN, D. B. (1984): “Bayesianly justifiable and relevant frequency calculations for the applied statistician,” *The Annals of Statistics*, 1151–1172.
- RUDIN, W. (1987): *Real and complex analysis*, 3rd ed., USA: McGraw-Hill, Inc.
- SAUMARD, A. AND J. A. WELLNER (2014): “Log-concavity and strong log-concavity: a review,” *Statistics surveys*, 8, 45.
- SCARLETT, J. AND V. CEVHER (2019): “An introductory guide to Fano’s inequality with applications in statistical estimation,” *arXiv preprint arXiv:1901.00555*.
- SCHENNACH, S. M. (2020): “Mismeasured and unobserved variables,” in *Handbook of econometrics*, Elsevier, vol. 7, 487–565.
- SOLOFF, J. A., A. GUNTUBOYINA, AND B. SEN (2025): “Multivariate, heteroscedastic empirical Bayes via nonparametric maximum likelihood,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87, 1–32.
- STEFANSKI, L. A. AND R. J. CARROLL (1990): “Deconvolving kernel density estimators,” *Statistics*, 21, 169–184.
- STONE, C. J. (1980): “Optimal rates of convergence for nonparametric estimators,” *The annals of*

Statistics, 1348–1360.

- TAYLOR, R. AND H. ZHANG (1990): “On a strongly consistent nonparametric density estimator for the deconvolution problem,” *Communications in Statistics-Theory and Methods*, 19, 3325–3342.
- TSYBAKOV, A. B. (2008): *Introduction to Nonparametric Estimation*, Springer Publishing Company, Incorporated, 1st ed.
- VAN DER VAART, A. (1991): “On differentiable functionals,” *The Annals of Statistics*, 178–204.
- VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Series in Statistics, New York: Springer.
- WALLACE, D. L. (1959): “Conditional confidence level properties,” *The Annals of Mathematical Statistics*, 30, 864–876.
- WALTERS, C. (2024): “Empirical Bayes methods in labor economics,” in *Handbook of Labor Economics*, Elsevier, vol. 5, 183–260.
- WINKLER, R. L. (1972): “A decision-theoretic approach to interval estimation,” *Journal of the American Statistical Association*, 67, 187–191.
- YOSHIMORI, M. AND P. LAHIRI (2014): “A second-order efficient empirical Bayes confidence interval,” *The Annals of Statistics*, 42, 1233–1261.
- ZHANG, C.-H. (1997): “Empirical Bayes and compound estimation of normal means,” *Statistica Sinica*, 7, 181–193.
- (2009): “Generalized maximum likelihood estimation of normal mixture densities,” *Statistica Sinica*, 1297–1318.

Appendix A Proofs of main results

The additional notation will be used throughout the Appendix: the symbol $*$ denotes the convolution of two functions, defined by $(f * g)(x) = \int_{\mathbb{R}} f(x - y)g(y) dy$, provided that $f, g \in L^1(\mathbb{R})$. For sequences a_n and b_n , we write $a_n \lesssim b_n$ if there exists a constant $C > 0$ such that $a_n \leq Cb_n$ for all sufficiently large n , and $a_n \gtrsim b_n$ if $a_n \geq C^{-1}b_n$ for all sufficiently large n . We write $a_n \asymp b_n$ when both $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold, meaning that a_n and b_n are of the same order up to constants.

A.1 Analytical properties of the weighting function for posterior quantile

The posterior quantile $q_0 := q(\tau; y)$ is the solution to the moment condition

$$\int (\tau - \mathbf{1}\{\theta \leq q_0\}) \phi\left(\frac{y - \theta}{\sigma}\right) dG(\theta) = 0.$$

In this subsection, we characterize the decay behavior of the Fourier coefficients of the weighting function $\psi(q_0, \theta; y) = (\tau - \mathbf{1}\{\theta \leq q_0\}) \phi\left(\frac{y - \theta}{\sigma}\right)$:

$$\overline{\psi^*(q_0, t; y)} := \int \exp(-it\theta) (\tau - \mathbf{1}\{\theta \leq q_0\}) \phi\left(\frac{y - \theta}{\sigma}\right) d\theta \quad (\text{A.1})$$

Its decay behavior plays a crucial role in our analysis of convergence rates of the posterior quantile.

Proposition A.1. *The Fourier coefficients of the weighting function $\overline{\psi^*(q_0(y), t; y)}$ has the following asymptotic expansion as $|t| \rightarrow \infty$.*

$$\overline{\psi^*(q_0(y), t; y)} = \sigma\tau \exp(-ity) \exp\left(-\frac{1}{2}t^2\sigma^2\right) + \frac{\exp(-itq_0(y))}{it} \phi\left(\frac{y - q_0(y)}{\sigma}\right) + o(t^{-1})$$

where the leading term is of order t^{-1} , and the remainder term $o(t^{-1})$ holds uniformly over $y \in \mathbb{R}$.

Proof. We carry out infinite integration by parts for $\overline{\psi^*(q_0, t; y)}$ since $\phi(\cdot)$ is infinitely smooth:

$$\overline{\psi^*(q_0, t; y)} = \sigma\tau \exp(-ity) \exp\left(-\frac{1}{2}t^2\sigma^2\right) + \exp(-itq_0(y)) \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{(it)^k \sigma^{k-1}} H_{k-1}\left(\frac{y - q_0(y)}{\sigma}\right) \phi\left(\frac{y - q_0(y)}{\sigma}\right).$$

where $H_k(\cdot)$ is the k -th Hermite polynomial, e.g. $H_0(z) = 1, H_1(z) = z, H_2(z) = z^2 - 1$.

An important property is that $\sup_{z \in \mathbb{R}} |H_k(z)| \phi(z) < C_k$ for any $k \in \mathbb{N}_+$. Thus for each $k \geq 2$:

$$\left| \exp(-itq_0(y)) \frac{(-1)^{k-1}}{(it)^k \sigma^{k-1}} H_{k-1}\left(\frac{y - q_0(y)}{\sigma}\right) \phi\left(\frac{y - q_0(y)}{\sigma}\right) \right| \leq \frac{C_{k-1}}{(it)^k \sigma^{k-1}} = O(t^{-k})$$

that holds for all $y \in \mathbb{R}$. Then we obtain the desired results. ■

Remark A.1. The leading t^{-1} decay rate arises from the discontinuity at $\theta = q_0$, which governs the asymptotic behavior of the Fourier coefficients $\overline{\psi^*(q_0, t; y)}$. ■

Remark A.2. The asymptotic expansion in Proposition A.1 reflects the key features of Fourier coefficients for functions with discontinuities. The second term decays at the slower rate of t^{-1} , reflecting the impact of the discontinuity introduced by the indicator function $\mathbf{1}\{\theta \leq q_0\}$. This slower decay rate reflects the Gibbs phenomenon which arises in the Fourier analysis of functions with jump discontinuities. ■

A.2 Proof of Theorem 3.1

We follow the classical two-point method of Le Cam (1973), see also Fan (1991); Butucea and Comte (2009); Pensky (2017) for related lower bound arguments in deconvolution problems. The proof has three parts. First, we construct two mixing densities g_0 and $g_1 = g_0 + H_T$ in $\mathcal{G}(s, L)$, where H_T is a local perturbation concentrated at frequencies of order T . Second, we show that the corresponding posterior quantiles are separated by order $T^{-s-1/2}$. Third, we show that the induced data distributions $((F_Y)_0, (F_Y)_1)$ are statistically indistinguishable in the sense that their χ^2 distance is of order $1/n$. Proposition A.2 then yields the desired minimax lower bound.

The non-standard step in our proof is to translate the perturbation in the mixing densities into a separation of their posterior quantiles. This step handles the nonlinearity of the functional by relying on the local behavior of the objective function around its minimizer and is where the non-smoothness enters.

Proposition A.2. *Assume two densities $g_0, g_1 \in \mathcal{G}(s, L)$. Then, with respect to an arbitrary estimator $\hat{q}$ of $q_G(\tau; y)$ based on i.i.d. data $\{Y_i\}_{i=1}^n$ having the density $f_Y = g * f_\varepsilon$ that satisfies that*

$$\sup_{g \in \mathcal{G}(s, L)} \mathbb{E} \left[\hat{q}_G(\tau; y) - q_G(\tau; y) \right]^2 \geq \text{const} \cdot \left| q_{G_1}(\tau; y) - q_{G_0}(\tau; y) \right|^2$$

for sufficiently large n , if we have that

$$\chi^2((f_Y)_1, (f_Y)_0) = \int \frac{((f_Y)_1(y) - (f_Y)_0(y))^2}{(f_Y)_0(y)} dy = O(1/n) \quad (\text{A.2})$$

where $(f_Y)_1 = g_1 * f_\varepsilon, (f_Y)_0 = g_0 * f_\varepsilon$.

Proof of Proposition A.2. The proof follows similar arguments to those used in Proposition 2.12 of Meister (2009) to derive the lower bound for the deconvolution density $g(\cdot)$. ■

We now consider two candidate deconvolution densities

$$g_0(\theta) = \frac{\lambda}{\pi(\lambda^2 + \theta^2)} \quad \text{and} \quad g_1(\theta) = g_0(\theta) + H_T(\theta) \quad (\text{A.3})$$

where $g_0(\theta)$ is a fixed baseline density and $H_T(\theta)$ is a local perturbation that will be specified later.

The baseline density g_0 is Cauchy with Fourier transformation $g_0^*(t) = \exp(-\lambda|t|)$. Moreover

$$\int |g_0^*(t)|^2 (1+t^2)^s dt = \int \exp(-2\lambda|t|) (1+t^2)^s dt \rightarrow 0$$

as $\lambda \rightarrow \infty$. Therefore, we may choose λ large enough so that $\int |g_0^*(t)|^2 (1+t^2)^s dt \leq L/4$ so that $g_0 \in \mathcal{G}(s, L)$. Let $q_0 = q_{G_0}(\tau; y)$. Since g_0 is continuous and strictly positive, the posterior quantile q_0 is well defined and unique.

To construct a localized least-favorable perturbation, we use the standard notation $C_c^\infty(E)$ for the space of infinitely differentiable functions on $\mathbb{R}$ whose support is compact and contained in E ; see [Folland \(1999, Chapter 9.1, p. 282\)](#). Choose a nonnegative function $k^* \in C_c^\infty((1, 2))$ that is not identically zero, and define

$$\xi(u) := -i \operatorname{sgn}(u) k^*(|u|), \quad h(x) := \frac{1}{2\pi} \int e^{-iux} \xi(u) du.$$

Since $\operatorname{supp}(k^*) \subset (1, 2)$, there exists some $\varepsilon > 0$ such that $k^*(u) = 0$ for all $u \notin [1 + \varepsilon, 2 - \varepsilon]$. Since $k^*(|u|) = 0$ for $|u| < 1 + \varepsilon$, we have that $\xi(u) = 0$ for $|u| < 1 + \varepsilon$. So near $u = 0$, ξ is just the zero function. Away from 0, both $\operatorname{sgn}(u)$ and $k^*(|u|)$ are smooth. Therefore $\xi \in C_c^\infty(\mathbb{R})$. In particular, $\operatorname{supp}(\xi) \subset [-2, -1] \cup [1, 2]$.

Moreover, since k^* is real-valued, it follows that $\xi(-u) = \overline{\xi(u)}$ for all $u \in \mathbb{R}$, and therefore the inverse Fourier transform h is real-valued. Recall that the Schwartz class ([Folland, 1999, p.237](#)) consists of smooth functions²¹ whose derivative decay faster than any polynomial grows at infinity:

$$\mathcal{S}(\mathbb{R}) = \left\{ f \in C^\infty(\mathbb{R}) : \sup_{x \in \mathbb{R}^n} (1 + |x|)^N |\partial^\alpha f(x)| < \infty \text{ for all } N, \alpha \right\} \quad (\text{A.4})$$

clearly $C_c^\infty \subset \mathcal{S}$. Since $\xi \in C_c^\infty(\mathbb{R})$, it follows that $\xi \in \mathcal{S}(\mathbb{R})$ as well. By [Folland \(1999, Corollary 8.23, Corollary 8.28\)](#), the Fourier transform maps $\mathcal{S}(\mathbb{R})$ into itself, and the inverse Fourier transform does as well. In particular, taking $N = 2$ and $\alpha = 0$ in [\(A.4\)](#), there exists a constant $C_h > 0$ such that

$$|h(x)| \leq C_h (1 + x^2)^{-1} \quad \text{for all } x \in \mathbb{R}.$$

Finally, Fourier inversion gives $h^* = \xi$, so $\int h(x) dx = h^*(0) = \xi(0) = 0$. For $T \geq 1$ and a small constant $\nu > 0$, we define $H_T(\theta) := \nu T^{-s+1/2} h(T(\theta - q_0))$. Then its Fourier transform is

$$H_T^*(t) = \nu T^{-s-1/2} (-i \operatorname{sgn}(t)) \exp(itq_0) k^*(|t|/T). \quad (\text{A.5})$$

²¹Here $C^\infty(E)$ means the smooth functions on E : all partial derivatives of every finite order exist and are continuous, see [Folland \(1999, p.235\)](#).

Step 1. Verifying that $g_1(\cdot) \in \mathcal{G}(s, L)$ for sufficiently large T .

First, g_1 integrates to one. Indeed, using the change of variable $u = T(\theta - q_0)$,

$$\int_{\mathbb{R}} H_T(\theta) d\theta = \nu T^{-s+1/2} \int_{\mathbb{R}} h(T(\theta - q_0)) d\theta = \nu T^{-s-1/2} \int_{\mathbb{R}} h(u) du = 0.$$

Therefore $\int_{\mathbb{R}} g_1(\theta) d\theta = \int_{\mathbb{R}} g_0(\theta) d\theta + \int_{\mathbb{R}} H_T(\theta) d\theta = 1$.

Second, $g_1(\theta)$ is nonnegative if ν is chosen small enough. Since q_0 is fixed, there exists a constant C_0 such that $1 + \theta^2 \leq C_0(1 + (\theta - q_0)^2)$ for all $\theta \in \mathbb{R}$. Hence

$$(1 + \theta^2)|H_T(\theta)| \leq C_0 C_h \nu T^{-s+1/2} \frac{1 + (\theta - q_0)^2}{1 + T^2(\theta - q_0)^2} \leq C \nu T^{-s+1/2} \leq C \nu$$

because $T \geq 1$ and $s \geq 1/2$. Therefore

$$|H_T(\theta)| \leq \frac{C \nu}{1 + \theta^2} \quad \text{for all } \theta \in \mathbb{R}.$$

On the other hand, $g_0(\theta) \geq [\pi \lambda (1 + \theta^2)]^{-1}$ for $\lambda \geq 1$, thus $|H_T(\theta)| \leq C \nu \pi \lambda g_0(\theta)$. If we choose $\nu > 0$ so small that $C \nu \pi \lambda \leq 1/2$, then $|H_T(\theta)| \leq \frac{1}{2} g_0(\theta)$ for all $\theta \in \mathbb{R}$. Consequently, $g_1(\theta) = g_0(\theta) + H_T(\theta) \geq g_0(\theta) - |H_T(\theta)| \geq \frac{1}{2} g_0(\theta) \geq 0$ for all $\theta \in \mathbb{R}$. Thus g_1 is a density.

Third, we verify that $g_1 \in \mathcal{G}(s, L)$. Since $g_1^* = g_0^* + H_T^*$, we have that $|g_1^*(t)|^2 \leq 2|g_0^*(t)|^2 + 2|H_T^*(t)|^2$ and hence

$$\int_{\mathbb{R}} |g_1^*(t)|^2 (1 + t^2)^s dt \leq 2 \int_{\mathbb{R}} |g_0^*(t)|^2 (1 + t^2)^s dt + 2 \int_{\mathbb{R}} |H_T^*(t)|^2 (1 + t^2)^s dt.$$

By construction, the first term is at most $L/2$. For the second term,

$$\begin{aligned} \int_{\mathbb{R}} |H_T^*(t)|^2 (1 + t^2)^s dt &= \nu^2 T^{-2s-1} \int_{\mathbb{R}} |\xi(t/T)|^2 (1 + t^2)^s dt = \nu^2 T^{-2s} \int_{\mathbb{R}} |\xi(u)|^2 (1 + T^2 u^2)^s du \\ &= \nu^2 \int_{\mathbb{R}} |\xi(u)|^2 (T^{-2} + u^2)^s du \leq \int_{\mathbb{R}} |\xi(u)|^2 (1 + u^2)^s du. \end{aligned}$$

Because $\xi \in C_c^\infty(\mathbb{R})$, the integral is finite. Hence, after possibly decreasing ν , we may ensure that $\int |H_T^*(t)|^2 (1 + t^2)^s dt \leq L/4$. Therefore $g_1 \in \mathcal{G}(s, L)$.

Step 2. Lower bound for the distance $|q_{G_1}(\tau; y) - q_{G_0}(\tau; y)|$.

Here $q_{G_1}(\tau; y), q_{G_0}(\tau; y)$ are posterior quantiles that correspond to densities g_1, g_0 defined in Equation (A.3). For notational convenience, write $q_j = q_{G_j}(\tau; y)$ for $j = 0, 1$. For $j = 0, 1$, define

$$\Psi_j(q) := \int \left(\mathbf{1}\{\theta \leq q\} - \tau \right) \phi\left(\frac{y - \theta}{\sigma}\right) g_j(\theta) d\theta.$$

Then the first-order conditions give that $\Psi_j(q_j) = 0$ for $j = 0, 1$.

We first show that q_1 lies in a small neighborhood of q_0 . Since $\Psi'_0(q) = \phi\left(\frac{y-q}{\sigma}\right)g_0(q)$, the positivity assumption at q_0 implies that $\Psi'_0(q_0) > 0$. Hence there exists $\delta > 0$ such that $m_\delta := \min\{-\Psi_0(q_0 - \delta), \Psi_0(q_0 + \delta)\} > 0$. Moreover, because $g_1 - g_0 = H_T$,

$$\begin{aligned} \sup_{q \in \mathbb{R}} |\Psi_1(q) - \Psi_0(q)| &\leq \sup_{q \in \mathbb{R}} \int \phi\left(\frac{y-\theta}{\sigma}\right) |H_T(\theta)| d\theta \\ &\leq \left\| \phi\left(\frac{y-\cdot}{\sigma}\right) \right\|_\infty \|H_T\|_1 \leq C\nu T^{-s-1/2}. \end{aligned}$$

Therefore, for all sufficiently large T , $\Psi_1(q_0 - \delta) < 0 < \Psi_1(q_0 + \delta)$. Since $q \mapsto \Psi_1(q)$ is increasing, it follows that $q_1 \in (q_0 - \delta, q_0 + \delta)$.

Next, subtracting $\Psi_0(q_0) = 0$ from $\Psi_1(q_1) = 0$ gives

$$\int (\mathbf{1}\{\theta \leq q_0\} - \tau) \phi\left(\frac{y-\theta}{\sigma}\right) (g_1(\theta) - g_0(\theta)) d\theta = \int (\mathbf{1}\{\theta \leq q_0\} - \mathbf{1}\{\theta \leq q_1\}) \phi\left(\frac{y-\theta}{\sigma}\right) g_1(\theta) d\theta.$$

By the mean-value theorem, there exists q^* between q_0 and q_1 such that

$$\int (\mathbf{1}\{\theta \leq q_0\} - \mathbf{1}\{\theta \leq q_1\}) \phi\left(\frac{y-\theta}{\sigma}\right) g_1(\theta) d\theta = (q_0 - q_1) \phi\left(\frac{y-q^*}{\sigma}\right) g_1(q^*).$$

Hence

$$q_1 - q_0 = - \left[\phi\left(\frac{y-q^*}{\sigma}\right) g_1(q^*) \right]^{-1} \left[\int (\mathbf{1}\{\theta \leq q_0\} - \tau) \phi\left(\frac{y-\theta}{\sigma}\right) (g_1(\theta) - g_0(\theta)) d\theta \right]. \quad (\text{A.6})$$

The denominator in (A.6) is bounded away from zero. Indeed, because $q^* \in [q_0 - \delta, q_0 + \delta]$, $g_1 \geq g_0/2$, and since $g_1 \geq g_0/2$ on this compact interval for all sufficiently large T , while g_0 is continuous and strictly positive there, there exists $c_1 > 0$ such that

$$\phi\left(\frac{y-q^*}{\sigma}\right) g_1(q^*) \geq c_1. \quad (\text{A.7})$$

It remains to lower bound the numerator in (A.6). Define $\psi(q_0, \theta; y) = (\mathbf{1}\{\theta \leq q_0\} - \tau) \phi\left(\frac{y-\theta}{\sigma}\right)$ and write

$$I_T := \int \psi(q_0, \theta; y) H_T(\theta) d\theta = \frac{1}{2\pi} \int \overline{\psi^*(q_0, t; y)} H_T^*(t) dt.$$

By Proposition A.1, we have that $\overline{\psi^*(q_0, t; y)} = \phi\left(\frac{y-q_0}{\sigma}\right) \frac{\exp(-itq_0)}{it} + r(t; y)$ where $|r(t; y)| \leq Ct^{-2}$.

Using the expression of $H_T^\star(t)$ from (A.5), we obtain that $I_T = I_{T1} + I_{T2}$ where

$$I_{T1} = \frac{\nu}{2\pi} T^{-s-1/2} \phi\left(\frac{y-q_0}{\sigma}\right) \int \frac{\exp(-itq_0)}{it} (-i \operatorname{sgn}(t)) \exp(itq_0) k^\star(|t|/T) dt$$

$$I_{T2} = \frac{\nu}{2\pi} T^{-s-1/2} \int r(t; y) (-i \operatorname{sgn}(t)) \exp(itq_0) k^\star(|t|/T) dt$$

We bound these two terms separately. For the first term,

$$|I_{T1}| = \frac{\nu}{2\pi} T^{-s-1/2} \phi\left(\frac{y-q_0}{\sigma}\right) \int \frac{k^\star(|t|/T)}{|t|} dt \stackrel{(a)}{=} \frac{\nu}{\pi} T^{-s-1/2} \phi\left(\frac{y-q_0}{\sigma}\right) \int_T^{2T} \frac{k^\star(t/T)}{|t|} dt$$

$$\stackrel{(b)}{=} \frac{\nu}{\pi} T^{-s-1/2} \phi\left(\frac{y-q_0}{\sigma}\right) \int_1^2 \frac{k^\star(u)}{u} du.$$

where (a) uses $\operatorname{supp}(k^\star) \subset (1, 2)$, and (b) follows from the change of variable $u = t/T$. Since $k^\star$ is nonnegative and not identically zero, $\int_1^2 \frac{k^\star(u)}{u} du > 0$. Therefore, $|I_{T1}| \geq c_2 \nu T^{-s-1/2}$ for some constant $c_2 > 0$.

For the second term,

$$|I_{T2}| = C \nu T^{-s-1/2} \int |r(t; y)| k^\star(|t|/T) dt \leq C \nu T^{-s-1/2} \int_T^{2T} t^{-2} dt \leq C \nu T^{-s-3/2},$$

thus I_{T2} is smaller than I_{T1} by an additional factor T^{-1} . Hence, by the triangle inequality, $|I_T| \geq |I_{T1}| - |I_{T2}| \geq c_3 \nu T^{-s-1/2}$ for all sufficiently large T where $c_3 > 0$.

Combining this bound with (A.6) and (A.7), we conclude that $|q_1 - q_0| \geq C \nu T^{-s-1/2}$.

Step 3. Upper bound for chi-divergence $\chi^2((f_Y)_1, (f_Y)_0)$.

Let $u_T(y) = (f_Y)_1(y) - (f_Y)_0(y) = H_T * f_\varepsilon(y)$, where f_ε is the $N(0, \sigma^2)$ density.

We first show that $(f_Y)_0$ is bounded below by a Cauchy tail. Choose $\eta > 0$ such that $p_\eta := \int_{|z| \leq \eta} f_\varepsilon(z) dz > 0$. Then

$$(f_Y)_0(y) = \int_{\mathbb{R}} g_0(y-z) f_\varepsilon(z) dz \geq \int_{|z| \leq \eta} g_0(y-z) f_\varepsilon(z) dz \geq p_\eta \inf_{|z| \leq \eta} g_0(y-z).$$

Moreover, for $|z| \leq \eta$, $1 + (y-z)^2 \leq C_\eta(1 + y^2)$, so $g_0(y-z) = \frac{\lambda}{\pi(\lambda^2 + (y-z)^2)} \geq \frac{c_\eta}{1+y^2}$. Hence $(f_Y)_0(y) \geq \frac{c}{1+y^2}$ for all $y \in \mathbb{R}$. Therefore

$$\chi^2((f_Y)_1, (f_Y)_0) = \int_{\mathbb{R}} \frac{u_T(y)^2}{(f_Y)_0(y)} dy \leq C \int_{\mathbb{R}} (1 + y^2) u_T(y)^2 dy.$$

Then it suffices to show that $S_T := \int_{\mathbb{R}} (1 + y^2) u_T(y)^2 dy = O(n^{-1})$.

By Parseval's identity, we have that

$$\int u_T(y)^2 dy = \frac{1}{2\pi} \int |H_T^*(t)|^2 \exp(-\sigma^2 t^2) dt \leq C\nu^2 T^{-2s} \exp(-\sigma^2 T^2). \quad (\text{A.8})$$

Also, since the Fourier transformation of $yu_T(y)$ is $i\partial_t(u_T^*(t))$,

$$\int_{\mathbb{R}} y^2 u_T(y)^2 dy = \frac{1}{2\pi} \int \left| \partial_t \left(H_T^*(t) \exp(-\sigma^2 t^2/2) \right) \right|^2 dt.$$

On the support of H_T^* , namely $T \leq |t| \leq 2T$, we have that $|H_T^*(t)| \leq C\nu T^{-s-1/2}$, and $|\partial_t H_T^*(t)| \leq C\nu T^{-s-1/2}$. Hence

$$\left| \partial_t \left(H_T^*(t) \exp(-\sigma^2 t^2/2) \right) \right| \leq C\nu T^{-s+1/2} \exp(-\sigma^2 T^2/2).$$

Since the support has length $O(T)$, it follows that

$$\int_{\mathbb{R}} y^2 u_T(y)^2 dy \leq C\nu^2 T^{-2s+2} \exp(-\sigma^2 T^2). \quad (\text{A.9})$$

Consequently, $S_T \leq C\nu^2 T^{-2s+2} \exp(-\sigma^2 T^2)$. Now set $T = T_n := \sigma^{-1} \sqrt{(1+\eta) \log n}$ for any fixed $\eta > 0$, then we have that $S_{T_n} \lesssim Cn^{-(1+\eta)} (\log n)^{-s+1}$, and therefore $S_T = O(n^{-1})$. Hence $\chi^2((f_Y)_1, (f_Y)_0) = O(n^{-1})$.

By Proposition A.2,

$$\sup_{g \in \mathcal{G}(s, L)} \mathbb{E} \left[(\hat{q} - q_G(\tau; y))^2 \right] \geq c |q_1 - q_0|^2 \geq c T_n^{-2s-1} \asymp (\log n)^{-(2s+1)/2}.$$

This proves Theorem 3.1.

A.3 Proof of Theorem 3.2

For notational convenience, let $\hat{q} = \hat{q}_G(\tau; y)$ denote the estimated posterior quantile, and $q_0 = q_{G_0}(\tau; y)$ its population counterpart. The proof proceeds in two main steps. First, we establish the following asymptotic expansion in Section A.3.1

$$r_n (\hat{q} - q_0) = r_n \left[\phi \left(\frac{y - q_0}{\sigma} \right) g(q_0) \right]^{-1} \Psi_n(q_0) + o_p(1) \quad (\text{A.10})$$

with a suitable sequence $r_n \rightarrow \infty$ as $n \rightarrow \infty$, and the term $\Psi_n(q_0)$ ²² is given by

$$\Psi_n(q_0) = (2\pi)^{-1} \int \frac{\overline{\psi^*(q_0, t; y)}}{f_\varepsilon^*(t)} \left(\widehat{f_Y^*}(t) K^*(h_n t) \right) dt \quad (\text{A.11})$$

where the Fourier coefficient $\overline{\psi^*(q_0, t; y)}$ is defined in Equation (A.1). Second we establish the properties of the leading term $\Psi_n(q_0)$, namely its consistency and rate of convergence, in Section A.3.2.

A.3.1 Asymptotic expansion

Define that $\widehat{\delta}_n = r_n(\widehat{q} - q_0)$ and the localized, rescaled function $\widetilde{W}_n(\delta) := r_n^2 \{W_n(q_0 + r_n^{-1}\delta) - W_n(q_0)\}$ for $\delta \in \mathbb{R}$. Observe that $\widehat{\delta}_n$ is the unique minimizer of $\widetilde{W}_n$. We now derive a quadratic approximation to $\widetilde{W}_n$ on compact sets, and then use Lemma 2 of Hjort and Pollard (2011) to pass from the approximation of the criterion to the approximation of its minimizer.

By Proposition D.1, we can write

$$\rho_\tau(\theta - (r_n^{-1}\delta + q_0)) - \rho_\tau(\theta - q_0) = -(r_n^{-1}\delta) \left(\tau - \mathbf{1}\{\theta - q_0 \leq 0\} \right) + \Delta$$

where

$$\Delta = \int_0^{r_n^{-1}\delta} \left(\mathbf{1}\{\theta - q_0 \leq s\} - \mathbf{1}\{\theta - q_0 \leq 0\} \right) ds.$$

Then we can decompose the objective function $\widetilde{W}_n(\delta) = -r_n \delta \Psi_n(q_0) + r_n^2 W_{n2}(\delta)$ where

$$W_{n2}(\delta) = (2\pi)^{-1} \int \frac{\exp(-it\theta) \Delta \phi\left(\frac{y-\theta}{\sigma}\right) d\theta}{f_\varepsilon^*(t)} \left(\widehat{f_Y^*}(t) K^*(h_n t) \right) dt. \quad (\text{A.12})$$

By Parseval-Plancherel theorem, we have that $W_{n2}(\delta) = \int \Delta \phi\left(\frac{y-\theta}{\sigma}\right) \widehat{g}(\theta) d\theta$ where

$$\widehat{g}(\theta) = (2\pi)^{-1} \int \frac{\exp(-it\theta)}{f_\varepsilon^*(t)} \left(\widehat{f_Y^*}(t) K^*(h_n t) \right) dt \quad (\text{A.13})$$

is algebraically the classical deconvolution kernel density estimator.

We calculate $W_{n2}(\delta)$ as follows

$$\begin{aligned} W_{n2}(\delta) &= \int \left[\int_0^{r_n^{-1}\delta} \left(\mathbf{1}\{\theta - q_0 \leq s\} - \mathbf{1}\{\theta - q_0 \leq 0\} \right) ds \right] \phi\left(\frac{y-\theta}{\sigma}\right) \widehat{g}(\theta) d\theta \\ &\stackrel{(a)}{=} r_n^{-1} \int \left[\int_0^\delta \left(\mathbf{1}\{\theta - q_0 \leq r_n^{-1}z\} - \mathbf{1}\{\theta - q_0 \leq 0\} \right) dz \right] \phi\left(\frac{y-\theta}{\sigma}\right) \widehat{g}(\theta) d\theta \end{aligned}$$

²²Note that its population counterpart $\Psi(q_0) = (2\pi)^{-1} \int \overline{\psi^*(q_0, t; y)} g^*(t) dt = \int (\tau - \mathbf{1}\{\theta \leq q_0\}) \phi\left(\frac{y-\theta}{\sigma}\right) g(\theta) d\theta \equiv 0$, which corresponds to the moment condition. In this sense, $\Psi_n(q_0)$ can be viewed as a sample moment function, and $\widehat{q}_n$ as the associated Z -estimator.

$$\begin{aligned}
&\stackrel{(b)}{=} r_n^{-1} \int_0^\delta \left[\int \left(\mathbf{1}\{\theta - q_0 \leq r_n^{-1}z\} - \mathbf{1}\{\theta - q_0 \leq 0\} \right) \phi\left(\frac{y-\theta}{\sigma}\right) \hat{g}(\theta) d\theta \right] dz \\
&= r_n^{-1} \int_0^\delta \int_{q_0}^{q_0+r_n^{-1}z} \phi\left(\frac{y-\theta}{\sigma}\right) \hat{g}(\theta) d\theta dz \stackrel{(c)}{=} r_n^{-1} \int_{q_0}^{q_0+r_n^{-1}\delta} \int_{r_n(\theta-q_0)}^\delta \phi\left(\frac{y-\theta}{\sigma}\right) \hat{g}(\theta) dz d\theta \\
&= r_n^{-1} \int_{q_0}^{q_0+r_n^{-1}\delta} [\delta - r_n(\theta - q_0)] \phi\left(\frac{y-\theta}{\sigma}\right) \hat{g}(\theta) d\theta \\
&\stackrel{(d)}{=} r_n^{-2} \int_0^\delta (\delta - u) \phi\left(\frac{y - q_0 - r_n^{-1}u}{\sigma}\right) \hat{g}(q_0 + r_n^{-1}u) du
\end{aligned}$$

where (a) uses changes of variable $s = r_n^{-1}z$, (b) and (c) switch the order of integration, (d) uses changes of variable again $u = r_n(\theta - q_0)$.

We now show that $r_n^2 W_{n2}(\delta)$ is asymptotically quadratic, uniformly on compact sets. By Assumptions 3.3 to 3.5 and continuity of g , there exists $\eta > 0$ such that for $I = [q_0 - \eta, q_0 + \eta]$, $c_I = \inf_{q \in I} \phi\left(\frac{y-q}{\sigma}\right)g(q) > 0$.

Fix $M > 0$. Since $r_n \rightarrow \infty$, for all sufficiently large n , $q_0 + r_n^{-1}u \in I$ whenever $|u| \leq M$. By the uniform consistency of $\hat{g}$ (Proposition D.2) and continuity of $q \mapsto \phi((y-q)/\sigma)g(q)$, we have that

$$\sup_{|u| \leq M} \left| \phi\left(\frac{y - q_0 - r_n^{-1}u}{\sigma}\right) \hat{g}(q_0 + r_n^{-1}u) - \phi\left(\frac{y - q_0}{\sigma}\right) g(q_0) \right| = o_p(1).$$

Therefore,

$$\begin{aligned}
&\sup_{|\delta| \leq M} \left| r_n^2 W_{n2}(\delta) - \frac{\delta^2}{2} \phi\left(\frac{y - q_0}{\sigma}\right) g(q_0) \right| \\
&= \sup_{|\delta| \leq M} \left| \int_0^\delta (\delta - u) \left[\phi\left(\frac{y - q_0 - r_n^{-1}u}{\sigma}\right) \hat{g}(q_0 + r_n^{-1}u) - \phi\left(\frac{y - q_0}{\sigma}\right) g(q_0) \right] du \right| \\
&\leq M^2 \sup_{|u| \leq M} \left| \phi\left(\frac{y - q_0 - r_n^{-1}u}{\sigma}\right) \hat{g}(q_0 + r_n^{-1}u) - \phi\left(\frac{y - q_0}{\sigma}\right) g(q_0) \right| = o_p(1).
\end{aligned}$$

Define the quadratic approximation $\widetilde{W}_n^0(\delta) = -r_n \delta \Psi_n(q_0) + \frac{\delta^2}{2} \phi\left(\frac{y-q_0}{\sigma}\right) g(q_0)$. Then for every fixed $M > 0$,

$$\sup_{|\delta| \leq M} \left| \widetilde{W}_n(\delta) - \widetilde{W}_n^0(\delta) \right| = o_p(1).$$

Recall that $\widehat{\delta}_n$ is the minimizer of $\widetilde{W}_n(\delta)$, and let

$$\widehat{\delta}_0 = \arg \min_{\delta \in \mathbb{R}} \widetilde{W}_n^0(\delta) = r_n \left[\phi\left(\frac{y - q_0}{\sigma}\right) g(q_0) \right]^{-1} \Psi_n(q_0)$$

where the minimizer is assumed to be unique. The next subsection shows that $r_n \Psi_n(q_0) = O_p(1)$,

hence $\widehat{\delta}_0 = O_p(1)$.

We can now compare the argmins of $\widetilde{W}_n$ and $\widetilde{W}_n^0$. First by Proposition D.2, $\sup_{q \in I} |\widehat{g}(q) - g(q)| = o_p(1)$. Hence $\inf_{q \in I} \phi\left(\frac{y-q}{\sigma}\right) \widehat{g}(q) \geq c_I/2 > 0$ with probability tending to one. On this event, we have that

$$\Psi'_n(q) = -\phi\left(\frac{y-q}{\sigma}\right) \widehat{g}(q) \leq -\frac{c_I}{2} < 0 \quad \text{for } q \in I$$

so Ψ_n is strictly decreasing on I , and therefore W_n is strictly convex on I .

Fix $\varepsilon > 0$ and $\eta > 0$. Since $\widehat{\delta}_0 = O_p(1)$, there exists $M > 0$ such that $P(|\widehat{\delta}_0| > M) \leq \varepsilon$ for all sufficiently large n . By Lemma 2 of Hjort and Pollard (2011),

$$P(|\widehat{\delta}_n - \widehat{\delta}_0| \geq \eta) \leq P\left(\sup_{|\delta - \widehat{\delta}_0| \leq \eta} \left| \widetilde{W}_n(\delta) - \widetilde{W}_n^0(\delta) \right| \geq \frac{1}{2} \inf_{|\delta - \widehat{\delta}_0| = \eta} \left[\widetilde{W}_n^0(\delta) - \widetilde{W}_n^0(\widehat{\delta}_0) \right]\right).$$

Because $\widetilde{W}_n^0$ is quadratic,

$$\inf_{|\delta - \widehat{\delta}_0| = \eta} \left[\widetilde{W}_n^0(\delta) - \widetilde{W}_n^0(\widehat{\delta}_0) \right] = \frac{\eta^2}{2} \phi\left(\frac{y - q_0}{\sigma}\right) g(q_0).$$

Therefore,

$$P(|\widehat{\delta}_n - \widehat{\delta}_0| \geq \eta) \leq P(|\widehat{\delta}_0| > M) + P\left(\sup_{|\delta| \leq M+\eta} \left| \widetilde{W}_n(\delta) - \widetilde{W}_n^0(\delta) \right| \geq \frac{\eta^2}{4} \phi\left(\frac{y - q_0}{\sigma}\right) g(q_0)\right).$$

The second probability tends to zero by the uniform approximation established above, while the first is at most ε . Since $\varepsilon > 0$ is arbitrary, we conclude that $\widehat{\delta}_n - \widehat{\delta}_0 = o_p(1)$.

Equivalently,

$$r_n(\widehat{q} - q_0) = \left[\phi\left(\frac{y - q_0}{\sigma}\right) g(q_0) \right]^{-1} r_n \Psi_n(q_0) + o_p(1), \quad (\text{A.14})$$

which is the desired asymptotic expansion. ■

A.3.2 Analysis of $\Psi_n(q_0)$

We decompose $\Psi_n(q_0) = \Psi_{n1} + \Psi_{n2}$ where

$$\begin{aligned} \Psi_{n1} &= (2\pi)^{-1} \int \frac{\overline{\psi^*(q_0, t; y)}}{f_\varepsilon^*(t)} \left(\widehat{f}_Y^*(t) - f_Y^*(t) \right) K^*(h_n t) dt, \\ \Psi_{n2} &= (2\pi)^{-1} \int \frac{\overline{\psi^*(q_0, t; y)}}{f_\varepsilon^*(t)} f_Y^*(t) (K^*(h_n t) - 1) dt. \end{aligned} \quad (\text{A.15})$$

The first term Ψ_{n1} captures the effects of random sampling error in estimating $f_Y^*(t)$. The second term Ψ_{n2} is a nonstochastic bias arising from the kernel regularization.

Analysis of Ψ_{n2} .

By Proposition A.1, there exists a constant $C > 0$ such that

$$|\psi^*(q_0, t; y)| \leq \frac{C}{1 + |t|} \quad \text{for all } t \in \mathbb{R}.$$

We split $\Psi_{n2}(q_0) = \Psi_{n2}^{\leq}(q_0) + \Psi_{n2}^{>}(q_0)$, where the first term integrates over $|t| \leq h_n^{-1}$ and the second over $|t| > h_n^{-1}$. Because K^* is supported on $[-1, 1]$, we have $K^*(h_n t) = 0$ whenever $|t| > h_n^{-1}$. Hence

$$|\Psi_{n2}^{>}(q_0)| \leq C \int_{|t| > h_n^{-1}} \frac{|g^*(t)|}{1 + |t|} dt.$$

By Cauchy-Schwarz inequality and Assumption 3.2,

$$|\Psi_{n2}^{>}(q_0)| \leq C \left(\int |g^*(t)|^2 (1 + t^2)^s dt \right)^{1/2} \left(\int_{|t| > h_n^{-1}} (1 + t^2)^{-(s+1)} dt \right)^{1/2} = O(h_n^{s+1/2}).$$

For the low-frequency part $\Psi_{n2}^{\leq}(q_0)$, first we aim to establish that $|K^*(h_n t) - 1| \lesssim |h_n t|^{s+1}$. Expanding $K^*(h_n t)$ around $t = 0$:

$$K^*(h_n t) = 1 + \frac{(h_n t)^{s+1}}{(s+1)!} \int (-iv)^{s+1} K(v) \exp(-i\xi h_n t v) dv$$

for some $\xi \in [0, 1]$. Then,

$$\begin{aligned} |K^*(h_n t) - 1|^2 &\lesssim |h_n t|^{2(s+1)} \left[\int |(-iv)^{s+1} K(v) \exp(-i\xi h_n t v)| dv \right]^2 \\ &\stackrel{(a)}{\lesssim} |h_n t|^{2(s+1)} \left[\int |v|^{s+1} |K(v)| dv \right]^2 \stackrel{(b)}{\lesssim} |h_n t|^{2(s+1)} \end{aligned}$$

where (a) is obtained from $|\exp(-it\xi h_n t v)| = 1$; (b) follows from Assumption 3.6 (ii). Therefore,

$$|\Psi_{n2}^{\leq}(q_0)| \leq C h_n^{s+1} \int_{|t| \leq h_n^{-1}} \frac{|t|^{s+1} |g^*(t)|}{1 + |t|} dt.$$

Applying Cauchy-Schwarz inequality again,

$$|\Psi_{n2}^{\leq}(q_0)| \leq C h_n^{s+1} \left(\int |g^*(t)|^2 (1 + t^2)^s dt \right)^{1/2} \left(\int_{|t| \leq h_n^{-1}} \frac{|t|^{2(s+1)}}{(1 + t^2)^{s+1}} dt \right)^{1/2}.$$

It remains to bound the last integral. On $|t| \leq 1$, the integrand is bounded. On $1 \leq |t| \leq h_n^{-1}$, $\frac{|t|^{2(s+1)}}{(1+t^2)^{s+1}} \lesssim 1$. Hence

$$\int_{|t| \leq h_n^{-1}} \frac{|t|^{2(s+1)}}{(1 + t^2)^{s+1}} dt = O(h_n^{-1}).$$

Therefore we have $|\Psi_{n2}^{\leq}(q_0)| = O(h_n^{s+1/2})$, and conclude that $\Psi_{n2}(q_0) = O(h_n^{s+1/2})$.

Analysis of Ψ_{n1} .

Observe that Ψ_{n1} is equal to an average of i.i.d. term:

$$\Psi_{n1} = n^{-1} \sum_{i=1}^n (2\pi)^{-1} \int \frac{\overline{\psi^*(q_0, t; y)}}{f_{\varepsilon}^*(t)} \left(\exp(itY_i) - f_Y^*(t) \right) K^*(h_n t) dt.$$

The variance term of Ψ_{n1} can be computed as

$$\text{Var}(\Psi_{n1}) \leq \frac{1}{4\pi^2 n} \mathbb{E} \left| \int \frac{\overline{\psi^*(q_0, t; y)}}{f_{\varepsilon}^*(t)} \exp(itY_i) K^*(h_n t) dt \right|^2 \quad (\text{A.16})$$

Letting $\omega(q_0, t; y) = \overline{\psi^*(q_0, t; y)} \exp(itY_i) / f_{\varepsilon}^*(t)$, the first term of the right-hand side of Equation (A.16) can be computed as

$$\begin{aligned} \text{First term of Equ. (A.16)} &= \frac{1}{4\pi^2 n} \mathbb{E} \left| \int \omega(q_0, t; y) \exp(-it(y - Y_i)) K^*(h_n t) dt \right|^2 \\ &= \frac{1}{4\pi^2 n} \int \left| \int \omega(q_0, t; y) \exp(-itz) K^*(h_n t) dt \right|^2 f_Y(y - z) dz \\ &\leq \frac{1}{4\pi^2 n} \|f_Y\|_{\infty} \int \left| \int \omega(q_0, t; y) \exp(-itz) K^*(h_n t) dt \right|^2 dz \\ &\stackrel{(a)}{=} \frac{1}{2\pi n} \|f_Y\|_{\infty} \int |\omega(q_0, t; y) K^*(h_n t)|^2 dt \end{aligned} \quad (\text{A.17})$$

where (a) is obtained from Parseval's identity, and $\|f_Y\|_{\infty}$ is finite since that

$$\sup_{y \in \mathbb{R}} |f_Y(y)| \leq \int \left(\sup_{y \in \mathbb{R}} |g(y - z)| \right) dF_{\varepsilon}(z) = \|g\|_{\infty} < \infty.$$

We calculate that $\omega(q_0, t; y) = \sigma\tau + \eta(t)$ where

$$\eta(t) = \exp\left(\frac{1}{2}t^2\sigma^2\right) \left[-\frac{\exp(-it(q_0 - y))}{it} \phi\left(\frac{y - q_0}{\sigma}\right) + o(t^{-1}) \right]. \quad (\text{A.18})$$

Then we have that

$$\int |\omega(q_0, t; y) K^*(h_n t)|^2 dt \leq 2\sigma^2 \int |K^*(h_n t)|^2 dt + 2 \int |\eta(t) K^*(h_n t)|^2 dt. \quad (\text{A.19})$$

For the first term in Equation (A.19), by Parseval's identity and Assumption 3.6, we have that $\int |K^*(h_n t)|^2 dt = h_n^{-1} \int |K^*(t)|^2 dt = 2\pi h_n^{-1} \int |K(z)|^2 dz = O(h_n^{-1})$.

It remains to bound the second term in Equation (A.19)

$$I_{n2}(h_n) := \int |\eta(t)K^\star(h_nt)|^2 dt.$$

Choose $T > 0$ large enough that, by (A.18),

$$|\eta(t)| \leq C \exp\left(\frac{1}{2}t^2\sigma^2\right) |t|^{-1} \quad \text{for all } |t| \geq T.$$

Since η is continuous on compact sets and $K^\star$ is bounded, we may split $I_{n2}(h_n)$ as

$$I_{n2}(h_n) = \int_{|t| \leq T} |\eta(t)K^\star(h_nt)|^2 dt + \int_{T < |t| \leq h_n^{-1}} |\eta(t)K^\star(h_nt)|^2 dt =: I_{n2}^{(1)}(h_n) + I_{n2}^{(2)}(h_n),$$

where the upper limit h_n^{-1} comes from the support condition on $K^\star(h_nt)$. The first term satisfies $I_{n2}^{(1)}(h_n) = O(1)$. For the second term, the above bound on $\eta(t)$ yields

$$\begin{aligned} I_{n2}^{(2)}(h_n) &\leq C \int_{T < |t| \leq h_n^{-1}} \frac{\exp(t^2\sigma^2)}{t^2} |K^\star(h_nt)|^2 dt \\ &\leq Ch_n \int_{Th_n \leq |u| \leq 1} \frac{\exp(u^2h_n^{-2}\sigma^2)}{u^2} |K^\star(u)|^2 du \end{aligned}$$

with the change of variable $u = h_nt$.

Fix any $\delta \in (0, 1)$. Split this integral into the interior region $Th_n \leq |u| \leq 1 - \delta$ and the boundary region $1 - \delta \leq |u| \leq 1$. On the interior region,

$$\begin{aligned} h_n \int_{Th_n \leq |u| \leq 1 - \delta} \frac{\exp(u^2h_n^{-2}\sigma^2)}{u^2} |K^\star(u)|^2 du &\leq Ch_n \exp(\sigma^2(1 - \delta)^2h_n^{-2}) \int_{Th_n \leq |u| \leq 1 - \delta} u^{-2} du \\ &= O(\exp(\sigma^2(1 - \delta)^2h_n^{-2})). \end{aligned}$$

Since $h_n \rightarrow 0$, it follows that $\exp(\sigma^2(1 - \delta)^2h_n^{-2}) = o(h_n^3 \exp(\sigma^2h_n^{-2}))$.

On the boundary region, the contributions from neighborhoods of $u = 1$ and $u = -1$ are the same, so it suffices to consider $u \in [1 - \delta, 1]$. By boundedness of $K^\star$,

$$\begin{aligned} h_n \int_{1 - \delta \leq |u| \leq 1} \frac{\exp(u^2h_n^{-2}\sigma^2)}{u^2} |K^\star(u)|^2 du &\leq Ch_n \int_{1 - \delta}^1 \exp(u^2h_n^{-2}\sigma^2) du \\ &= Ch_n \exp(\sigma^2h_n^{-2}) \int_0^\delta \exp(-\sigma^2(2v - v^2)h_n^{-2}) dv \\ &\leq Ch_n \exp(\sigma^2h_n^{-2}) \int_0^\delta \exp(-\sigma^2vh_n^{-2}) dv \\ &= O(h_n^3 \exp(\sigma^2h_n^{-2})) \end{aligned}$$

where $u = 1 - v$, and $2v - v^2 \geq v$ for $v \in [0, \delta] \subset [0, 1]$.

Combining the interior and boundary bounds, we obtain $I_{n2}^{(2)}(h_n) = O(h_n^3 \exp(\sigma^2 h_n^{-2}))$ and thus $I_{n2}(h_n) = O(1) + O(h_n^3 \exp(\sigma^2 h_n^{-2})) = O(h_n^3 \exp(\sigma^2 h_n^{-2}))$, since $h_n^3 \exp(\sigma^2 h_n^{-2}) \rightarrow \infty$ as $h_n \rightarrow 0$.

Substituting this and the bound for the first term into Equation (A.19), we get

$$\int |\omega(q_0, t; y) K^*(h_n t)|^2 dt = O(h_n^{-1}) + O(h_n^3 \exp(\sigma^2 h_n^{-2})).$$

Because $h_n^{-1} = o(h_n^3 \exp(\sigma^2 h_n^{-2}))$, it follows that

$$\int |\omega(q_0, t; y) K^*(h_n t)|^2 dt = O(h_n^3 \exp(\sigma^2 h_n^{-2})).$$

Hence, by Equation (A.17), $\text{Var}(\Psi_{n1}) = O(n^{-1} h_n^3 \exp(\sigma^2 h_n^{-2}))$. Combining this with $\Psi_{n2}^2 = O(h_n^{2s+1})$, we obtain

$$\mathbb{E}[\Psi_n(q_0) - \Psi(q_0)]^2 = \text{Var}(\Psi_{n1}) + \Psi_{n2}^2 \leq O(n^{-1} h_n^3 \exp(\sigma^2 h_n^{-2})) + O(h_n^{2s+1}).$$

Now choose $h_n = c(\log n)^{-1/2}$ for some $c > \sigma$. Then

$$n^{-1} h_n^3 \exp(\sigma^2 h_n^{-2}) = n^{-1+\sigma^2/c^2} (\log n)^{-3/2} = o((\log n)^{-(2s+1)/2}),$$

so

$$\mathbb{E}[\Psi_n(q_0) - \Psi(q_0)]^2 = O((\log n)^{-(2s+1)/2}).$$

This choice of h_n also guarantees the uniform consistency of $\hat{g}(\cdot)$, which is needed in the analysis of $W_{n2}(\delta)$ in (A.12), see Proposition D.2.

Finally, since $\Psi(q_0) = 0$, let $M_\varepsilon = \sqrt{C/\varepsilon}$. For every $\varepsilon > 0$, Chebyshev's inequality yields

$$\mathbb{P}\left((\log n)^{(2s+1)/4} |\Psi_n(q_0)| > M_\varepsilon\right) \leq \frac{(\log n)^{(2s+1)/2} \mathbb{E}[\Psi_n(q_0) - \Psi(q_0)]^2}{M_\varepsilon^2} \leq \frac{C}{M_\varepsilon^2} \leq \varepsilon.$$

Therefore $\Psi_n(q_0) = O_p((\log n)^{-(2s+1)/4})$. ■

A.3.3 Results for $\hat{q}_G(\tau; y)$

Let $r_n = (\log n)^{(2s+1)/4}$. From the asymptotic expansion established above, we obtain that

$$(\log n)^{(2s+1)/4} (\hat{q}_G(\tau; y) - q_G(\tau; y)) = \underbrace{-(\log n)^{(2s+1)/4} \left[\phi\left(\frac{y - q_0}{\sigma}\right) g(q_0) \right]^{-1}}_{O_p(1)} \Psi_n(q_0) + o_p(1)$$

which completes the proof, establishing both consistency and the convergence rate of $\widehat{q}_G(\tau; y)$. ■

A.4 Proof of Theorem 3.3

The proof reduces the lower bound problem for conditional coverage errors to a multiple testing problem. We construct a collection of alternative deconvolution densities $\{g_m\}_{m=1}^M$ around the baseline g_0 such that the induced conditional coverage error has a specific “step-function” structure. Under this structure, if an interval had conditional coverage error uniformly smaller than the claimed rate, it could decode which alternative generated the data. The proof then shows that such decoding is impossible because the alternatives are statistically indistinguishable.

To formalize this argument, we fix a baseline density $g_0 \in \mathcal{G}(s, L/2)$ introduced in Appendix A.2. Then define

$$I(\beta; y) := \left[q_{G_0}(\beta; y), q_{G_0}(\beta + 1 - \alpha; y) \right], \quad \text{for } \beta \in [0, \alpha]$$

so that $P_{G_0}(\theta \in I(\beta; y) \mid Y = y) = 1 - \alpha$. Thus every posterior interval under G_0 with conditional coverage $1 - \alpha$ can be indexed by its left-tail probability $\beta \in [0, \alpha]$. The restriction $\beta \in [0, \alpha]$ ensures that the right endpoint is a valid posterior quantile since $\beta + 1 - \alpha \in [0, 1]$. For a general mixing distribution G , write $f_G(y) = \int \phi(\frac{y-\theta}{\sigma}) dG(\theta)$ for the marginal density of Y evaluated at y .

The proof has four steps. Step 1 constructs the alternatives, Step 2 shows that their corresponding data distributions are statistically indistinguishable, Step 3 shows that a single interval can have conditional coverage error at the claimed rate only over a local window of alternatives, and Step 4 converts this into a multiple testing problem and applies Fano’s inequality.

A.4.1 Step 1. Constructing many alternatives

Step 1 constructs a collection of mixing densities $\{g_m\}_{m=1}^M \subset \mathcal{G}(s, L)$ that are well separated in the target criterion: under alternative m , the associated conditional coverage error changes sign across an m -specific transition window of width $O(T^{-1})$, and outside that window it is uniformly separated from zero by an order of $T^{-s-1/2}$. Because of this uniform separation, if an interval has conditional coverage error of the order $T^{-s-1/2}$, then under alternative m its left endpoint must lie inside the corresponding transition window.

To index these alternatives, we first choose a compact interval $\mathcal{J} \subset (0, \alpha)$ such that $\text{dist}(\mathcal{J}, \mathcal{J} + 1 - \alpha) = \inf_{\beta \in \mathcal{J}, \beta' \in \mathcal{J} + 1 - \alpha} |\beta - \beta'| > 0$. Next, choose $\Delta > 0$ so that Lemmas D.3, D.6 and D.7 apply on $\mathcal{J}$, and let β_1 be the left-endpoint of $\mathcal{J}$. We define a grid of points $\beta_m := \beta_1 + \Delta(m - 1)/T$ for $m = 1, \dots, M$, where M is the largest integer such that $\beta_M \in \mathcal{J}$. Because $\mathcal{J}$ has a fixed positive length, we have $M \asymp T$.

We now define the alternatives. For each $m = 1, \dots, M$, let $g_m := g_0 + \bar{H}_T + \tilde{H}_{m,T}$ where g_0 is the baseline Cauchy density, $\bar{H}_T$ is a common perturbation defined in Appendix D.2.1, and $\tilde{H}_{m,T}$ is an m -specific cumulative perturbation that determines the location of the sign change inside $\mathcal{J}$. Following Appendix A.2, we set $\theta_m := q_{G_0}(\beta_m; y)$, and $H_{m,T} := \nu T^{-s+1/2} h(T(\theta - \theta_m))$

whose Fourier transform is $H_{m,T}^*(t) = \nu T^{-s-1/2}(-i \operatorname{sgn}(t)) \exp(it\theta_m)k^*(|t|/T)$. We then define the cumulative perturbations

$$\tilde{H}_{m,T} := \varepsilon \left(\sum_{j < m} H_{j,T} - \sum_{j \geq m} H_{j,T} \right)$$

where $\varepsilon > 0$ is sufficiently small. Thus each alternative is obtained by adding to the baseline density a common perturbation $\bar{H}_T$ and an m -dependent signed cumulative perturbation $\tilde{H}_{m,T}$.

As verified in Appendix D.2.2, these alternatives are admissible densities within the target function class; that is, $g_m \in \mathcal{G}(s, L)$ for all $m = 1, \dots, M$ whenever T is sufficiently large.

We now define the conditional coverage errors under each alternative. For $\beta \in [0, \alpha]$, let $D_m(\beta; y) = P_{G_m}(\theta \in I(\beta; y) \mid Y = y) - (1 - \alpha)$. To characterize where the sign of $D_m(\beta; y)$ can change, partition $\mathcal{J}$ into cells

$$B_k = \left[\beta_k - \frac{\Delta}{2T}, \beta_k + \frac{\Delta}{2T} \right] \cap \mathcal{J}, \quad k = 1, \dots, M.$$

so that $\mathcal{J} = \bigcup_{k=1}^M B_k$.

The key output of Step 1 is a sign-separation property for $D_m(\beta; y)$: away from a uniformly bounded neighborhood of the index m , the sign of $D_m(\beta; y)$ is fixed and its magnitude is of order $T^{-s-1/2}$. Specifically, Appendix D.2.3 shows that there exist constants $c_1 > 0$ and $r \geq 1$, independent of m and T , such that for all sufficiently large T , the error $D_m(\beta; y)$ is uniformly positive on the left-hand cells:

$$D_m(\beta; y) \geq c_1 T^{-s-1/2} \quad \text{for } \beta \in \bigcup_{k \leq m-r} B_k, \quad (\text{A.20})$$

and uniformly negative on the right-hand cells:

$$D_m(\beta; y) \leq -c_1 T^{-s-1/2} \quad \text{for } \beta \in \bigcup_{k \geq m+r} B_k. \quad (\text{A.21})$$

Consequently, its sign changes only within the transition region:

$$\mathcal{T}_m := \bigcup_{m-r < k < m+r} B_k.$$

Then $\beta \in \mathcal{J} \setminus \mathcal{T}_m$ implies that $|D_m(\beta; y)| \geq c_1 T^{-s-1/2}$.

For each m , $D_m(\beta; y)$ changes sign across an m -specific transition region $\mathcal{T}_m$ of uniform width $O(T^{-1})$. Outside of $\mathcal{T}_m$, it is uniformly bounded away from zero by the order of $T^{-s-1/2}$. As a result, the family $\{D_m(\cdot; y)\}_{m=1}^M$ behaves like a collection of step functions with a transition region of uniformly bounded width.

A.4.2 Step 2. The alternatives are statistically close

We now show that the alternatives constructed in Step 1 are statistically close. First, we bound the Kullback-Leibler divergence between the laws of $\mathbf{Y}_{-i}$ under different alternatives. Second, because the target of the lower bound is the conditional coverage probability at a fixed y , we show that the corresponding posterior distributions are also uniformly close.

We use two notions of divergence: the Kullback-Leibler divergence, defined in Definition 1, and the χ^2 divergence, introduced in Proposition A.2.

Definition 1 (Kullback-Leibler divergence, Definition 2.5, [Tsybakov \(2008\)](#)). *The Kullback-Leibler (KL) divergence between two probability measures P and Q is defined by*

$$\text{KL}(P, Q) = \begin{cases} \int \log \frac{dP}{dQ} dP, & \text{if } P \ll Q \\ +\infty, & \text{otherwise} \end{cases}$$

where $P \ll Q$ means that P is absolutely continuous with respect to Q , and $\frac{dP}{dQ}$ is the Radon-Nikodym derivative.

Let P_m denote the law of $\mathbf{Y}_{-i} = \{Y_j\}_{j \neq i}$ under g_m and let $(f_Y)_m$ be the corresponding marginal density. Since $\mathbf{Y}_{-i}$ consists of $n-1$ i.i.d. observations, the joint divergence tensorizes over the marginals. Using the inequality (2.27) in [Tsybakov \(2008\)](#), we obtain

$$\begin{aligned} \text{KL}(P_m, P_{m'}) &= (n-1) \text{KL}\left((f_Y)_m, (f_Y)_{m'}\right) \leq (n-1) \chi^2\left((f_Y)_m, (f_Y)_{m'}\right) \\ &\leq 2(n-1) \int_{\mathbb{R}} \frac{\left((f_Y)_m(y) - (f_Y)_{m'}(y)\right)^2}{(f_Y)_0(y)} dy \end{aligned}$$

The last inequality holds because Step 1 gives $g_m \geq c'g_0$ so that $(f_Y)_m(y) \geq c'(f_Y)_0(y)$ for some constant $c' > 0$.

Define $u_{m,m'} := (f_Y)_m - (f_Y)_{m'} = (g_m - g_{m'}) * f_\varepsilon$. By Lemma D.4, we have that

$$\int_{\mathbb{R}} \frac{u_{m,m'}(y)^2}{(f_Y)_0(y)} dy \leq C\varepsilon^2 \nu^2 T^{-2s+2} \exp(-\sigma^2 T^2).$$

Choosing $T = T_n = \sigma^{-1} \sqrt{(1+\eta) \log n}$ as in Appendix A.2, we obtain

$$(n-1) \int_{\mathbb{R}} \frac{u_{m,m'}(y)^2}{(f_Y)_0(y)} dy \leq C(\log n)^{-s+1} n^{-\eta} \leq C \quad \text{for all } m, m'.$$

Therefore

$$\sup_{m, m'} \text{KL}(P_m, P_{m'}) \leq C. \tag{A.22}$$

Thus the full leave-one-out sample $\mathbf{Y}_{-i}$ cannot distinguish the alternatives well.

We next give an analogous argument for errors in conditional coverage probability at the fixed value y . Let $F_G(\cdot | y)$ denote the posterior distribution of θ_i given $Y_i = y$:

$$F_G(x | y) = \frac{\int \mathbf{1}\{\theta \leq x\} \phi\left(\frac{y-\theta}{\sigma}\right) dG(\theta)}{\int \phi\left(\frac{y-\theta}{\sigma}\right) dG(\theta)}.$$

By Lemma D.5, the posterior distributions induced by the alternatives are uniformly close to the baseline posterior:

$$\sup_m \sup_{x \in \mathbb{R}} |F_{G_m}(x | y) - F_{G_0}(x | y)| \leq CT^{-s-1/2}. \quad (\text{A.23})$$

Consequently, let $I = [\ell, u] \subset \mathbb{R}$ be any interval. Then $P_G(\theta \in I | Y = y) = F_G(u | y) - F_G(\ell | y)$, so we have that

$$\left| P_{G_m}(\theta \in I | Y = y) - P_{G_0}(\theta \in I | Y = y) \right| \leq 2CT^{-s-1/2}. \quad (\text{A.24})$$

Hence, at the fixed value y , the posterior mass assigned to any interval deviates from the baseline by at most $O(T^{-s-1/2})$ across the alternatives.

A.4.3 Step 3. Accuracy under an alternative forces localization

This step turns the sign pattern from Step 1 into a localization statement for any candidate interval. Fix an arbitrary interval $I = [\ell, u]$. We show that if I has a conditional coverage error of order at most $T^{-s-1/2}$ under alternative m , the corresponding baseline tail index β must lie in the m -specific transition region $\mathcal{T}_m$. Since these transition regions have uniformly bounded widths and are eventually disjoint when the indices are sufficiently separated, a single interval can have a conditional coverage error of $O(T^{-s-1/2})$ for only finitely many alternatives.

Let $\beta_0(I) := P_{G_0}(\theta \in (-\infty, \ell] | Y = y)$. Suppose that $|P_{G_m}(\theta \in I | Y = y) - (1 - \alpha)| \leq AT^{-s-1/2}$ for some m . Then by (A.24), setting $C_1(A) = A + 2C$, we have

$$|P_{G_0}(\theta \in I | Y = y) - (1 - \alpha)| \leq C_1(A)T^{-s-1/2}. \quad (\text{A.25})$$

Define $\beta(I) := \beta_0(I) \wedge \alpha$ and $\tilde{I} := I(\beta(I), y)$. We first claim that

$$0 \leq \beta_0(I) - \beta(I) \leq C_1(A)T^{-s-1/2}. \quad (\text{A.26})$$

Because $I = [\ell, u]$ is an interval, we have that $P_{G_0}(\theta \in I | Y = y) \leq P_{G_0}(\theta \geq \ell | Y = y) = 1 - \beta_0(I)$. Combining this with (A.25) gives $1 - \beta_0(I) \geq 1 - \alpha - C_1(A)T^{-s-1/2}$, which is equivalent to (A.26).

Let $F_{G_0}(\cdot | y)$ denote the posterior distribution under G_0 , and write $\tilde{I} = [\tilde{\ell}, \tilde{u}]$. By construction, $F_{G_0}(\ell | y) = \beta_0(I)$ and $F_{G_0}(\tilde{\ell} | y) = \beta(I)$. Furthermore, $F_{G_0}(u | y) = \beta_0(I) + P_{G_0}(\theta \in I | Y = y)$

and $F_{G_0}(\tilde{u} \mid y) = \beta(I) + (1 - \alpha)$. The continuity of $F_{G_0}(\cdot \mid y)$ yields

$$\begin{aligned} \mathbb{P}_{G_0}(\theta \in I \triangle \tilde{I} \mid Y = y) &\leq \left| F_{G_0}(\ell \mid y) - F_{G_0}(\tilde{\ell} \mid y) \right| + \left| F_{G_0}(u \mid y) - F_{G_0}(\tilde{u} \mid y) \right| \\ &\leq 2|\beta_0(I) - \beta(I)| + \left| \mathbb{P}_{G_0}(\theta \in I \mid Y = y) - (1 - \alpha) \right| \\ &\leq 3C_1(A)T^{-s-1/2}. \end{aligned}$$

Hence, by (A.23), $\mathbb{P}_{G_m}(\theta \in I \triangle \tilde{I} \mid Y = y) \leq (3C_1(A) + 4C)T^{-s-1/2}$. Combined with $|\mathbb{P}_{G_m}(\theta \in I \mid Y = y) - (1 - \alpha)| \leq AT^{-s-1/2}$, the triangle inequality implies $|\mathbb{P}_{G_m}(\theta \in \tilde{I} \mid Y = y) - (1 - \alpha)| \leq (4A + 10C)T^{-s-1/2}$. Since $\tilde{I} = I(\beta(I); y)$, setting $C_2(A) := 4A + 10C$, this becomes

$$|D_m(\beta(I); y)| \leq C_2(A)T^{-s-1/2}. \quad (\text{A.27})$$

We next show that $\beta(I)$ must lie in $\mathcal{J}$. If $\beta(I) \notin \mathcal{J}$, then Lemma D.1 (ii) together with the decomposition $D_m = \bar{D}_T + \Gamma_{\tilde{H}_{m,T}} + \tilde{R}_m$ implies that

$$\begin{aligned} |D_m(\beta(I); y)| &\geq |\bar{D}_T(\beta(I); y)| - |\Gamma_{\tilde{H}_{m,T}}(\beta(I); y)| - |\tilde{R}_m(\beta(I); y)| \\ &\geq (c\eta - C\varepsilon\nu - o(1))T^{-s-1/2}. \end{aligned}$$

Choose $\varepsilon > 0$ sufficiently small so that $C\varepsilon\nu \leq c\eta/4$. Then for all sufficiently large T , $|D_m(\beta(I); y)| \geq (c\eta/2)T^{-s-1/2}$ which contradicts (A.27) whenever $C_2(A) < c\eta/2$. Hence $\beta(I) \in \mathcal{J}$.

Once $\beta \in \mathcal{J}$, the sign-separation argument from Step 1 applies. In particular, if $\beta \in \mathcal{J} \setminus \mathcal{T}_m$, it follows that $|D_m(\beta; y)| \geq c_1T^{-s-1/2}$ where c_1 is the constant from (A.20)-(A.21). Now choose $A > 0$ such that $C_2(A) < c_1$, which ensures $\beta(I) \in \mathcal{T}_m$. Thus, whenever I has a conditional coverage error of order $AT^{-s-1/2}$ under alternative m , its corresponding baseline tail index $\beta(I)$ must lie in the m -specific transition region $\mathcal{T}_m$.

Finally, we show that any fixed interval I can satisfy this coverage error bound for at most $2r + 1$ alternatives. If $|m - m'| > 2r + 1$, the two transition regions are disjoint. To see this, assume without loss of generality that $m' > m + 2r + 1$. Then

$$\beta_{m'+r} - \frac{\Delta}{2T} - \left(\beta_{m+r} + \frac{\Delta}{2T} \right) = \frac{\Delta}{T} (m' - m - 2r - 1) > 0$$

so $\mathcal{T}_m \cap \mathcal{T}_{m'} = \emptyset$. Thus the right endpoint of $\mathcal{T}_m$ lies strictly to the left of the left endpoint of $\mathcal{T}_{m'}$. Since $\beta(I)$ is a single point, it can belong to at most $2r + 1$ transition regions.

A.4.4 Step 4. Reduction to multiple testing

This step completes the proof by using a local Fano argument. First, we construct a finite local packing of alternatives. Choose a subset $\mathcal{M}_T \subset \{1, \dots, M\}$ such that $|m - m'| > 2r + 1$ for all

distinct $m, m' \in \mathcal{M}_T$. Because r is fixed and $M \asymp T$, such a subset can be chosen with $|\mathcal{M}_T| \asymp T$.

For each $m \in \mathcal{M}_T$, we define the errors in conditional coverage probability of a realized interval I under G_m by $\text{ECP}_m(I; y) = \mathbb{P}_{G_m}(\theta_i \in I \mid Y_i = y) - (1 - \alpha)$. We define the $AT^{-s-1/2}$ -accuracy set under G_m by

$$\mathcal{A}_m = \left\{ I : |\text{ECP}_m(I; y)| \leq AT^{-s-1/2} \right\}.$$

Step 3 implies the following separation property of the finite packing:

$$\mathcal{A}_m \cap \mathcal{A}_{m'} = \emptyset, \quad \text{for all distinct } m, m' \in \mathcal{M}_T.$$

Thus the family $\{G_m : m \in \mathcal{M}_T\}$ is separated in the sense that no single realized interval can lie in the $AT^{-s-1/2}$ -accuracy sets of two distinct alternatives in $\mathcal{M}_T$.

We now associate each interval rule with a decoder $\hat{m} = \hat{m}(\text{CI}_i(\mathbf{Y}_{-i}))$: if $\text{CI}_i(\mathbf{Y}_{-i}) \in \mathcal{A}_m$ for some $m \in \mathcal{M}_T$, set $\hat{m} = m$; otherwise, set $\hat{m}$ to be an arbitrary fixed element of $\mathcal{M}_T$. This map is well-defined because the collection $\{\mathcal{A}_m : m \in \mathcal{M}_T\}$ is pairwise disjoint.

Now fix $m \in \mathcal{M}_T$. If the true alternative is m and $\hat{m} \neq m$, then the realized interval cannot be $AT^{-s-1/2}$ -accurate under G_m . Hence

$$\mathbb{P}_m(\hat{m} \neq m) \leq \mathbb{P}_m\left(|\text{ECP}_m(\text{CI}_i(\mathbf{Y}_{-i}); y)| > AT^{-s-1/2}\right).$$

Using the inequality $x^2 \geq a^2 \mathbf{1}\{|x| > a\}$, we obtain

$$\mathbb{E}_m \left[\text{ECP}_m(\text{CI}_i(\mathbf{Y}_{-i}); y)^2 \right] \geq A^2 T^{-2s-1} \mathbb{P}_m(\hat{m} \neq m).$$

Taking the supremum over $m \in \mathcal{M}_T$ yields

$$\sup_{m \in \mathcal{M}_T} \mathbb{E}_m \left[\text{ECP}_m(\text{CI}_i(\mathbf{Y}_{-i}); y)^2 \right] \geq A^2 T^{-2s-1} \sup_{m \in \mathcal{M}_T} \mathbb{P}_m(\hat{m} \neq m). \quad (\text{A.28})$$

Let $\hat{\mathcal{M}}_T$ denote the class of decoders induced by interval rules in $\mathcal{C}$, and let $\mathcal{M}_T$ denote the class of all measurable decoders mapping $\mathbf{Y}_{-i}$ to $\mathcal{M}_T$. Because $\hat{\mathcal{M}}_T \subseteq \mathcal{M}_T$, we have

$$\inf_{\hat{m} \in \hat{\mathcal{M}}_T} \sup_{m \in \mathcal{M}_T} \mathbb{P}_m(\hat{m} \neq m) \geq \inf_{\hat{m} \in \mathcal{M}_T} \sup_{m \in \mathcal{M}_T} \mathbb{P}_m(\hat{m} \neq m)$$

Taking the infimum in (A.28) over all interval rules in $\mathcal{C}$ produces the lower bound

$$\inf_{\text{CI}_i \in \mathcal{C}} \sup_{m \in \mathcal{M}_T} \mathbb{E}_m \left[\text{ECP}_m(\text{CI}_i(\mathbf{Y}_{-i}); y)^2 \right] \geq A^2 T^{-2s-1} \inf_{\hat{m} \in \hat{\mathcal{M}}_T} \sup_{m \in \mathcal{M}_T} \mathbb{P}_m(\hat{m} \neq m). \quad (\text{A.29})$$

It remains to lower bound the testing error probability on the right-hand side of (A.29). By

Fano's inequality (Theorem 1, [Scarlett and Cevher, 2019](#)), we have that

$$\sup_{m \in \mathcal{M}_T} \mathbb{P}_m(\hat{m} \neq m) \geq \frac{1}{|\mathcal{M}_T|} \sum_{m \in \mathcal{M}_T} \mathbb{P}_m(\hat{m} \neq m) \geq 1 - \frac{I(V; \mathbf{Y}_{-i}) + \log 2}{\log |\mathcal{M}_T|}$$

where $I(V; \mathbf{Y}_{-i})$ is the mutual information between the random index $V \sim \text{Unif}(\mathcal{M}_T)$ and the observed data $\mathbf{Y}_{-i}$, that is defined and upper bounded as follows:

$$\begin{aligned} I(V; \mathbf{Y}_{-i}) &:= \text{KL}(\mathbb{P}_{V, \mathbf{Y}_{-i}}, \mathbb{P}_V \otimes \mathbb{P}_{\mathbf{Y}_{-i}}) \stackrel{(a)}{=} \frac{1}{|\mathcal{M}_T|} \sum_{m \in \mathcal{M}_T} \text{KL}(\mathbb{P}_{\mathbf{Y}_{-i}|V=m}, \mathbb{P}_{\mathbf{Y}_{-i}}) \\ &\stackrel{(b)}{=} \frac{1}{|\mathcal{M}_T|} \sum_{m \in \mathcal{M}_T} \text{KL}(\mathbb{P}_m, \mathbb{P}_{\mathbf{Y}_{-i}}), \quad \text{here } \mathbb{P}_{\mathbf{Y}_{-i}} = \frac{1}{|\mathcal{M}_T|} \sum_{m' \in \mathcal{M}_T} \mathbb{P}_{m'} \\ &\stackrel{(c)}{\leq} \frac{1}{|\mathcal{M}_T|^2} \sum_{m, m' \in \mathcal{M}_T} \text{KL}(\mathbb{P}_m, \mathbb{P}_{m'}) \leq \sup_{m, m' \in \mathcal{M}_T} \text{KL}(\mathbb{P}_m, \mathbb{P}_{m'}) \stackrel{(d)}{\leq} C \end{aligned}$$

where (a) follows from Lemma 4, Eq. (23) of [Scarlett and Cevher \(2019\)](#), (b) follows because $\mathbf{Y}_{-i}|V=m \sim \mathbb{P}_m$, (c) is obtained from Lemma 4, Eq. (26) of [Scarlett and Cevher \(2019\)](#), (d) follows from the uniform bound on pairwise Kullback-Leibler divergences ([A.22](#)).

Then we have that

$$\inf_{\hat{m} \in \mathcal{M}_T} \sup_{m \in \mathcal{M}_T} \mathbb{P}_m(\hat{m} \neq m) \geq 1 - \frac{C + \log 2}{\log |\mathcal{M}_T|}. \quad (\text{A.30})$$

Because $|\mathcal{M}_T| \asymp T \rightarrow \infty$, the right-hand side of ([A.30](#)) is bounded away from zero; in particular, it is at least 1/2 for all sufficiently large T .

Combining ([A.28](#))-([A.30](#)), and using $\{g_m : m \in \mathcal{M}_T\} \subset \mathcal{G}(s, L)$, we conclude that

$$\inf_{\text{CI}_i \in \mathcal{C}} \sup_{g \in \mathcal{G}(s, L)} E \left[\left(\mathbb{P}_G(\theta_i \in \text{CI}_i(\mathbf{Y}_{-i}) \mid Y_i = y) - (1 - \alpha) \right)^2 \right] \geq cT^{-2s-1}$$

for some constant $c > 0$. Letting $T \asymp \sqrt{\log n}$ proves the theorem.

A.5 Proof of Theorem 3.4

Fix $Y_i = y$ and $\tau \in (0, 1)$, write $\hat{q} = \hat{q}_G(\tau; y)$ and $q_0 = q_G(\tau; y)$. First we consider the conditional coverage probability $\mathbb{P}(\theta \leq \hat{q} \mid Y_i = y)$ for a fixed $\tau \in (0, 1)$.

Because the posterior distribution $\mathbb{P}(\theta \leq u \mid Y = y)$ is differentiable in a neighborhood around q_0 with the posterior density ([3.3](#)), the mean-value theorem gives that

$$\mathbb{P}(\theta \leq \hat{q} \mid Y_i = y) = \mathbb{P}(\theta \leq q_0 \mid Y_i = y) + \pi(q^* \mid Y_i = y)(\hat{q} - q_0) \quad (\text{A.31})$$

where q^* lies between $\hat{q}$ and q_0 . Notice that since $\hat{q} \xrightarrow{P} q_0$ by Theorem 3.2 and $|q^* - q_0| \leq |\hat{q} - q_0|$,

it must be that $q^* \xrightarrow{P} q_0$. By Assumption 3.2 and $s > 1/2$, the mixing density $g(\cdot)$ and thus $\pi(\cdot|Y = y)$ is continuous, the continuous mapping theorem yields that $\pi(q^*|Y_i = y) \xrightarrow{P} \pi(q_0|Y_i = y)$. By $P(\theta \leq q_0 | Y_i = y) = \tau$ and Theorem 3.2, we obtain that

$$\begin{aligned} P(\theta_i \leq \hat{q}_G(\tau; Y_i) | Y_i = y) &= \tau + \left(\pi(q_0|Y_i = y) + o_p(1) \right) (\hat{q} - q_0) \\ &= \tau + O_p\left((\log n)^{-(2s+1)/4}\right). \end{aligned}$$

Applying this result with $\tau = \alpha/2$ and $\tau = 1 - \alpha/2$ obtains the desired results.

A.6 Proof of Theorem 3.5

Only in this subsection, for $\tau \in (0, 1)$, we write $q(y) := q_G(\tau; y)$, $\hat{q}_{-i}(y) := \hat{q}_G(\tau; y)$, where $\hat{q}_{-i}(y)$ is computed from the leave-one-out sample $\mathbf{Y}_{-i} = \{Y_j\}_{j \neq i}$. Also let $r_n := (\log n)^{(2s+1)/4}$. It suffices to prove that

$$P\left(\theta_i \leq \hat{q}_{-i}(Y_i)\right) = \tau + O\left((\log n)^{-(2s+1)/4}\right).$$

Applying this with $\tau = \alpha/2$ and $\tau = 1 - \alpha/2$ gives the desired result for

$$\text{CI}_i^{\text{NP}} = \left[\hat{q}_G(\alpha/2; Y_i), \hat{q}_G(1 - \alpha/2; Y_i) \right].$$

By iterated expectations and the independence of $\mathbf{Y}_{-i}$ and (θ_i, Y_i) ,

$$\begin{aligned} P(\theta_i \leq \hat{q}_{-i}(Y_i)) - \tau &= E[P(\theta_i \leq \hat{q}_{-i}(Y_i) | Y_i, \mathbf{Y}_{-i}) - P(\theta_i \leq q(Y_i) | Y_i)] \\ &= E[P(\theta_i \leq \hat{q}_{-i}(Y_i) | Y_i) - P(\theta_i \leq q(Y_i) | Y_i)]. \end{aligned} \tag{A.32}$$

Therefore, by the mean-value theorem, there exists $q^*(y)$ between $q(y)$ and $\hat{q}_{-i}(y)$ such that

$$P(\theta_i \leq \hat{q}_{-i}(Y_i)) - \tau = E\left[\pi(q^*(Y_i)|Y_i) (\hat{q}_{-i}(Y_i) - q(Y_i))\right] \tag{A.33}$$

Next we use the same argument as in Appendix A.3, but uniformly over $y \in \mathcal{Y}$. Specifically Assumption 3.8 gives that

$$\phi\left(\frac{y - q(y)}{\sigma}\right) g(q(y)) \geq \pi \quad \text{for all } y \in \mathcal{Y}$$

and Proposition D.2 gives the uniform consistency of $\hat{g}$. Hence the proof of the asymptotic expansion in Appendix A.3 yields

$$\sup_{y \in \mathcal{Y}} \left| r_n (\hat{q}_{-i}(y) - q(y)) + r_n \left[\phi\left(\frac{y - q(y)}{\sigma}\right) g(q(y)) \right]^{-1} (\Psi_{n1, -i}(y) + \Psi_{n2}(y)) \right| = o_p(1) \tag{A.34}$$

where $\Psi_{n1,-i}(y)$ and $\Psi_{n2}(y)$ are the stochastic and bias terms from the decomposition in Appendix A.3 with $\hat{f}_Y^*$ formed from Y_{-i} .

Moreover, the same argument also gives that $\sup_{y \in \mathcal{Y}} |\hat{q}_{-i}(y) - q(y)| = o_p(1)$. Since $|q^*(y) - q(y)| \leq |\hat{q}_{-i}(y) - q(y)|$, it follows that $\sup_{y \in \mathcal{Y}} |q^*(y) - q(y)| = o_p(1)$. Because g is continuous by Assumption 3.4, the posterior density $\pi(\cdot | Y = y)$ is continuous in its first argument. Hence we have that

$$\sup_{y \in \mathcal{Y}} \left| \pi(q^*(y) | Y = y) - \pi(q(y) | Y = y) \right| = o_p(1). \quad (\text{A.35})$$

Substituting (A.34) and (A.35) into (A.33), we obtain that

$$\mathbb{P} \left(\theta_i \leq \hat{q}_{-i}(\tau; Y_i) \right) - \tau = A_{n1} + A_{n2} + o(r_n^{-1}), \quad (\text{A.36})$$

where

$$A_{n1} := \mathbb{E} \left[\frac{\pi(q(Y_i) | Y_i)}{\phi\left(\frac{Y_i - q(Y_i)}{\sigma}\right) g(q(Y_i))} \Psi_{n1,-i}(Y_i) \right], \quad A_{n2} := \mathbb{E} \left[\frac{\pi(q(Y_i) | Y_i)}{\phi\left(\frac{Y_i - q(Y_i)}{\sigma}\right) g(q(Y_i))} \Psi_{n2}(Y_i) \right].$$

We first consider A_{n1} . For every fixed y , $\mathbb{E}[\Psi_{n1,-i}(y) | Y_i = y] = 0$, because $\Psi_{n1,-i}(y)$ is the centered empirical-process term built from $\mathbf{Y}_{-i}$, and $\mathbf{Y}_{-i}$ is independent of Y_i . Hence $A_{n1} = 0$.

We next consider A_{n2} . Assumption 3.8 implies that

$$|A_{n2}| \leq \pi^{-1} \mathbb{E} [\pi(q(Y_i) | Y_i) |\Psi_{n2}(Y_i)|]. \quad (\text{A.37})$$

Now, the same argument used to analyze $\Psi_{n2}(q_0)$ in Appendix A.3 applies uniformly over $y \in \mathcal{Y}$ because Proposition A.1 gives the t^{-1} bound on the Fourier coefficient of the weighting function uniformly over $y \in \mathcal{Y}$. Therefore

$$\sup_{y \in \mathcal{Y}} |\Psi_{n2}(y)| = O(h_n^{s+1/2}). \quad (\text{A.38})$$

Combining (A.37) and (A.38), we have

$$|A_{n2}| \leq \pi^{-1} \sup_{y \in \mathcal{Y}} |\Psi_{n2}(y)| \int \pi(q(\tau; y) | Y_i = y) dF_Y(y).$$

By Assumption 3.9, the integral on the right-hand side is finite. Therefore $A_{n2} = O(h_n^{s+1/2})$.

Since $h_n \asymp (\log n)^{-1/2}$, we have that $h_n^{s+1/2} = O((\log n)^{-(2s+1)/4})$. Substituting the results for A_{n1} and A_{n2} into (A.36) gives the desired result. ■

Appendix B Computational details of NP-EBCI

The practical implementation in Section 4, especially the bandwidth selection procedure in Section 4.2, can be computationally demanding. The procedure requires the computation of two posterior-quantile endpoints for each candidate bandwidth $h \in \mathcal{H}_n$, each fold v , and each held-out unit $i \in \mathcal{I}_v$. In large samples, this repeated evaluation can be costly.

The purpose of this section is therefore computational. We outline two strategies to speed up the computation. First, for each fold and each candidate bandwidth, the deconvolution objects are common to all held-out units and can therefore be computed once and reused throughout interval construction. Second, since the leave-fold-out NPMLE has a discrete structure, the coverage criterion in Section 4.2 reduces to finite weighted sums over its support points. Consequently, the bandwidth search avoids repeated simulation or additional high-dimensional numerical integration.

Fix a fold $v \in \{1, \dots, V\}$ and a candidate bandwidth $h \in \mathcal{H}_n$. Define

$$\hat{f}_Y^{\star(-v)}(t) = \frac{1}{|\mathcal{I}_v^c|} \sum_{j \in \mathcal{I}_v^c} \exp(itY_j), \quad \tilde{f}_\varepsilon^{\star(-v)}(t) = \frac{1}{|\mathcal{I}_v^c|} \sum_{j \in \mathcal{I}_v^c} \exp\left(-\frac{1}{2}\sigma_j^2 t^2\right),$$

and define the fold-specific deconvolution weight

$$\hat{g}_{v,h}^{\star}(t) := \frac{\hat{f}_Y^{\star(-v)}(t)}{\tilde{f}_\varepsilon^{\star(-v)}(t)} K^{\star}(ht).$$

Then, for each held-out unit $i \in \mathcal{I}_v$, the posterior quantile estimator in Step 2 of Section 4.2 can be written as

$$\hat{q}_h^{(-v)}(\tau; \mathcal{D}_i) = \arg \min_q \int \overline{M^{\star}(q, t; \mathcal{D}_i)} \hat{g}_{v,h}^{\star}(t) dt. \quad (\text{B.1})$$

A key computational feature is that $\hat{g}_{v,h}^{\star}(t)$ depends on the training sample $\mathcal{I}_v^c$ and on h , but not on the held-out unit i . Hence, within each fold v , the same deconvolution weight can be used for all units $i \in \mathcal{I}_v$. In implementation, we *pre-compute* $\hat{g}_{v,h}^{\star}(t)$ once for each (v, h) and then reuse it to evaluate the one-dimensional objective for all held-out units in that fold.

The remaining computational bottleneck is the repeated evaluation of $\overline{M^{\star}(q, t; \mathcal{D}_i)}$ over many candidate values of q , many held-out units i , and many bandwidths h . We compute these transforms in batches. Specifically, we stack the integrands in (B.1) for many held-out units into a single large matrix, rewrite the complex exponential terms in real arithmetic, and evaluate the resulting objective by matrix multiplication. This replaces repeated scalar quadrature problems with a small number of batched linear algebra operations. In practice, we evaluate the objective on a moderate grid of candidates q -values centered near Y_i , locate the minimizing grid point, and then apply a local quadratic refinement to obtain the posterior quantile estimates.

We next turn to Step 3 of Section 4.2. For each fold v , let the leave-fold-out NPMLE be

$$\hat{G}^{(-v)}(A) = \sum_{\ell=1}^{L_v} \hat{p}_{\ell v} \mathbf{1}\{\hat{\theta}_{\ell v} \in A\}, \quad \hat{p}_{\ell v} \geq 0, \quad \sum_{\ell=1}^{L_v} \hat{p}_{\ell v} = 1$$

for any Borel set $A \subset \mathbb{R}$. The discreteness of $\hat{G}^{(-v)}$ implies that, for each held-out unit $i \in \mathcal{I}_v$, the posterior distribution under $\hat{G}^{(-v)}$ is also discrete. Its posterior atom probabilities, given $\mathcal{D}_i = (Y_i, \sigma_i)$, are

$$\hat{\pi}_{\ell v}(\mathcal{D}_i) = \frac{\hat{p}_{\ell v} \phi((Y_i - \hat{\theta}_{\ell v})/\sigma_i)}{\sum_{m=1}^{L_v} \hat{p}_{mv} \phi((Y_i - \hat{\theta}_{mv})/\sigma_i)}, \quad \ell = 1, \dots, L_v.$$

Therefore the average conditional coverage under the out-of-fold prior estimate $\hat{G}^{(-v)}(\cdot)$ is

$$\hat{C}(h) = \frac{1}{n} \sum_{v=1}^V \sum_{i \in \mathcal{I}_v} \sum_{\ell=1}^{L_v} \hat{\pi}_{\ell v}(\mathcal{D}_i) \mathbf{1}\left\{\hat{q}_h^{(-v)}(\alpha/2; \mathcal{D}_i) \leq \hat{\theta}_{\ell v} \leq \hat{q}_h^{(-v)}(1 - \alpha/2; \mathcal{D}_i)\right\}$$

which is a finite weighted sum over the discrete support of the NPMLE. Hence, once the endpoints have been computed, evaluating $\hat{C}(h)$ requires no additional numerical integration.

Finally, for Step 4 of the bandwidth selection procedure in Section 4.2, one can replace the hard feasibility rule with a penalized criterion:

$$\hat{h} = \arg \min_{h \in \mathcal{H}_n} \left\{ \frac{\hat{L}(h)}{L_{\text{naive}}} + \lambda \left(\max\{(1 - \alpha) - \hat{C}(h), 0\} \right)^2 \right\},$$

where $L_{\text{naive}} = \frac{1}{n} \sum_{i=1}^n 2z_{1-\alpha/2}\sigma_i$ is the average length of the naive z -interval. The soft penalty continues to favor short intervals among candidates whose estimated coverage is close to the nominal level. But unlike the hard-feasibility rule, it remains well defined even when no candidate bandwidth satisfies the nominal coverage requirement, and it is often more numerically stable in finite samples.

Appendix C Extended discussions

C.1 Monotonicity of posterior quantiles

Proposition C.1. *Consider the model $Y_i = \theta_i + \varepsilon_i$ where $\theta_i \sim G(\cdot)$, $\varepsilon_i \sim F_\varepsilon(\cdot)$ and $\theta_i \perp\!\!\!\perp \varepsilon_i$. The posterior quantile $q_G(\tau; y)$ is non-decreasing in y if $f_\varepsilon(\cdot)$ is log-concave on $\mathbb{R}$.*

Proof. Given $Y = y$, the posterior density of θ is $\pi(\theta|y) \propto g(\theta)f_\varepsilon(y - \theta)$. Since f_ε is log-concave, the location family $\{f_\varepsilon(y - \theta)\}$ satisfies the monotone likelihood ratio property, i.e. for any $y_1 < y_2$, the ratio $\theta \mapsto f_\varepsilon(y_2 - \theta)/f_\varepsilon(y_1 - \theta)$ is non-decreasing in θ (Proposition 2.3, [Saumard and Wellner](#))

(2014)). Therefore

$$\frac{\pi(\theta|y_2)}{\pi(\theta|y_1)} = \frac{\int g(u)f_\varepsilon(y_1 - u)du}{\int g(u)f_\varepsilon(y_2 - u)du} \cdot \frac{f_\varepsilon(y_2 - \theta)}{f_\varepsilon(y_1 - \theta)}$$

is also non-decreasing in θ .

Since both $\pi(\theta|y_1)$ and $\pi(\theta|y_2)$ integrate to one, the function $\pi(\theta|y_2) - \pi(\theta|y_1)$ can change sign at most once, and if it changes sign, it does so from negative to positive. Hence there exists c such that $\pi(\theta|y_2) \leq \pi(\theta|y_1)$ for $\theta \leq c$ and $\pi(\theta|y_2) \geq \pi(\theta|y_1)$ for $\theta \geq c$. It follows that

$$P_G(\theta \leq u|Y = y_2) - P_G(\theta \leq u|Y = y_1) = \int_{-\infty}^u (\pi(\theta|y_2) - \pi(\theta|y_1)) d\theta \leq 0.$$

Thus, for each u , the posterior distribution $P_G(\theta \leq u|Y = y)$ is non-increasing in y , which in turn implies that the posterior quantile $q_G(\tau; y)$ is non-decreasing in y . ■

C.2 Fourier-based kernel estimator for posterior means

In this subsection, we consider the Fourier-based kernel estimator for the posterior mean (Zhang, 1997) analogous to the procedure (3.4) for the posterior quantile. Our purpose is not to propose an estimator for the posterior mean per se, but to facilitate a comparison of its rate of convergence with that of the posterior quantile.

Recall that the posterior mean $m_G(y) = E_G[\theta|Y = y] := m_0$. Let the estimator $\hat{m}_G(y)$ be Equation (3.4) but replace the check function with ℓ_2 loss. Following the strategy in Section A.3.1, one can establish the following asymptotic expansion $r_n(\hat{m} - m_0) = -r_n f_Y^{-1}(y) \Psi_n(m_0) + o_p(1)$ with a sequence $r_n \rightarrow \infty$ as $n \rightarrow \infty$, that will be specified later. The leading term $\Psi_n(m_0)$ is

$$\Psi_n(m_0) = (2\pi)^{-1} \int \frac{\overline{\psi^*(m_0, t; y)}}{f_\varepsilon^*(t)} \left(\hat{f}_Y^*(t) K^*(h_n t) \right) dt$$

where the weighting function for the posterior mean is $\overline{\psi^*(m_0, t; y)} := \int \exp(-it\theta) (\theta - m_0) \phi\left(\frac{y-\theta}{\sigma}\right) d\theta$.

We focus on the analysis of $\Psi_n(m_0)$ for the posterior mean and compare it with the corresponding term for the posterior quantile. First, we demonstrate that the weighting function $\overline{\psi^*(m_0, t; y)}$ is able to cancel out the exponential decay of $f_\varepsilon^*(t)$ in Proposition C.2.

Proposition C.2. *It calculates that $\overline{\psi^*(m_0, t; y)} = \sigma [(y - m_0) - i\sigma^2 t] \exp(-ity - \frac{1}{2}\sigma^2 t)$.*

Using the explicit expression in (C.2), we observe that the Fourier-based kernel estimator for the posterior mean is directly linked to the classical kernel density estimator for $f_Y(y)$ and its derivative. In particular, the leading term can be rewritten as follows:

$$\Psi_n(m_0) = (2\pi)^{-1} \int \sigma (y - m_0 - i\sigma^2 t) \exp(-ity) \left(\hat{f}_Y^*(t) K^*(h_n t) \right) dt$$

$$\begin{aligned}
&= \frac{\sigma}{2\pi n} \sum_{j=1}^n \int (y - m_0 - i\sigma^2 t) \exp(-it(y - Y_j)) K^*(h_n t) dt \\
&\stackrel{(a)}{=} \frac{\sigma}{2\pi n h_n} \sum_{j=1}^n \int \left(y - m_0 - i\sigma^2 \frac{u}{h_n}\right) \exp(-iu h_n^{-1}(y - Y_j)) K^*(u) du \\
&= \frac{\sigma}{2\pi n} \sum_{j=1}^n \left[(y - m_0) \int \exp\left(-iu \frac{y - Y_j}{h_n}\right) K^*(u) du - \frac{\sigma^2 i}{h_n^2} \int u \exp\left(-iu \frac{y - Y_j}{h_n}\right) K^*(u) du \right] \\
&\stackrel{(b)}{=} \frac{\sigma}{n} \sum_{j=1}^n \left[(y - m_0) K\left(\frac{y - Y_j}{h_n}\right) + \frac{\sigma^2}{h_n^2} K'\left(\frac{y - Y_j}{h_n}\right) \right]
\end{aligned}$$

where (a) uses the change-of-variable $t = h_n^{-1}u$; (b) obtains by inverse Fourier transform formulas $K(z) = (2\pi)^{-1} \int \exp(-itz) K^*(t) dt$ and $K'(z) = (2\pi)^{-1} \int \exp(-itz) (-it) K^*(t) dt$.

Similarly, we can obtain the expression of its popular counterpart:

$$\Psi(m_0) = (2\pi)^{-1} \sigma \int (y - m_0 - i\sigma^2 t) \exp(-ity) f_Y^*(t) dt = \sigma(y - m_0) f_Y(y) + \sigma^3 f_Y'(y) \equiv 0,$$

and $\Psi_n(m_0) - \Psi(m_0)$ can be represented as the following

$$\sigma(y - m_0) \left(\frac{1}{n} \sum_{i=1}^n K\left(\frac{y - Y_i}{h_n}\right) - f_Y(y) \right) + \sigma^3 \left(\frac{1}{n h_n^2} K'\left(\frac{y - Y_i}{h_n}\right) - f_Y'(y) \right).$$

The convergence rate analysis follows directly from standard results on kernel density estimation (see e.g. [Pagan and Ullah \(1999\)](#); [Li and Racine \(2007\)](#); [Horowitz \(2009, Appendix\)](#) for textbook treatments). We note that the optimal rates of convergence of $\Psi_n(m_0) - \Psi(m_0)$ coincides with that for estimating the derivative of a density, namely $n^{-(r-1)/(2r+1)}$, where r is the smoothness of $f_Y(y)$ ([Stone, 1980](#)). This smoothness r in our setting is characterized in [Proposition C.3](#).

Proposition C.3. *Under the normal location model (2.1) and the smoothness assumption for $g \in \mathcal{G}(s, L)$ (3.2), we have that $f_Y(y)$ is super-smooth in the sense that*

$$\int |f_Y^*(t)|^2 (t^2 + 1)^r dt < \infty, \quad \text{for all } r \geq 0.$$

Proof. Given that $f_Y^*(t) = g^*(t) \exp(-\frac{1}{2}\sigma^2 t^2)$. For any $r \geq 0$, we have that

$$\int |f_Y^*(t)|^2 (t^2 + 1)^r dt = \int |g^*(t)|^2 (t^2 + 1)^s \underbrace{\left[(t^2 + 1)^{r-s} \exp(-\sigma^2 t^2) \right]}_{w_r(t)} dt. \quad (\text{C.1})$$

The function $w_r(t)$ is continuous and vanishes as $|t| \rightarrow \infty$, thus it must achieve its supremum at some finite point. That is, $W_r = \sup_{t \in \mathbb{R}} w_r(t) < \infty$. Then the right-hand side of Equation (C.1) is not greater than $W_r \int |g^*(t)|^2 (1 + t^2)^s dt < W_r L =: L_r < \infty$. Hence $f_Y \in \mathcal{G}(r, L_r)$ for all $r \geq 0$ and

thus supersmooth. ■

By Proposition C.3, it's clear that the optimal rates of convergence of $\Psi_n(m_0) - \Psi(m_0)$ – and thus the posterior mean – is $O_p(n^{-1/2})$. To let the kernel estimator match this parametric rate, one should use an infinite-order kernel. For instance, the sinc kernel $K(z) = \sin(z)/\pi z$ has infinite order,²³ see e.g. Meister (2009, Page 42) for more details.

Remark C.1. The point of parametric rate of the posterior mean in the normal location model has been previously made in the empirical Bayes literature. Zhang (1997, Theorem 2) showed that the kernel estimator for the k -th derivative of marginal density $\hat{f}^{(k)}(\cdot)$ converges at parametric rate (up to a logarithm factor) in L_2 and L_∞ norm. By Tweedie's formula, one can obtain that the posterior mean shall also convergence at parametric rate. Zhang (2009, Theorem 1) showed that the generalized maximum likelihood estimator (GMLE) for the marginal density $f_Y(y)$ converges at parametric rate (up to a logarithm factor) in Hellinger distance given that the mixing distribution has compact support or has a heavier-than-normal exponential tail. ■

C.3 Extension to precision dependence

The baseline model in Section 4.1 treats the sampling variance σ_i^2 as observed and assumes precision independence, in the sense that $\theta_i \perp\!\!\!\perp \sigma_i$. Following Chen (2026), a tractable relaxation is to model the conditional distribution of θ_i given σ_i as a location-scale family.

Specifically, suppose

$$Y_i \mid \theta_i, \sigma_i \sim N(\theta_i, \sigma_i^2), \quad \theta_i = m(\sigma_i) + s(\sigma_i)\xi_i,$$

where $m(\cdot)$ and $s(\cdot) > 0$ are unknown functions, $\xi_i \sim G(\cdot)$, and $\xi_i \perp\!\!\!\perp \sigma_i$. Under this specification, precision dependence operates through the conditional location $m(\sigma_i)$ and conditional scale $s(\sigma_i)$, while the shape of condition distribution of $\theta_i \mid \sigma_i$, govern by G , is common across values of σ_i .

Define the normalized variables

$$Z_i = \frac{Y_i - m(\sigma_i)}{s(\sigma_i)}, \quad v_i = \frac{\sigma_i}{s(\sigma_i)}, \quad \tilde{\mathcal{D}}_i = (Z_i, v_i). \quad (\text{C.2})$$

Then

$$Z_i = \xi_i + \eta_i, \quad \eta_i \mid v_i \sim N(0, v_i^2), \quad (\text{C.3})$$

and, by construction, $\xi_i \perp\!\!\!\perp v_i$. Hence the transformed data satisfy the same heteroskedastic normal means model as in Section 4.1, but with latent effect ξ_i and known heteroskedastic variances v_i^2 .

Let $\mathcal{D}_i = (Y_i, \sigma_i)$. In the precision-dependent setting, the posterior quantile for θ_i is defined

²³The fourier transformation of the sinc kernel is $K^*(t) = \mathbf{1}_{[-1,1]}(t)$. It is flat at $t = 0$, that enables to ensure that all moments of the kernel vanish, i.e. $\int z^q K(z) dz = i^q \frac{d^q K^*(t)}{dt^q} \Big|_{t=0} = 0$ for any $q \geq 1$.

conditional on $\mathcal{D}_i$:

$$q_{\theta_i}(\tau; \mathcal{D}_i) = \inf \{u : P(\theta_i \leq u \mid \mathcal{D}_i) \geq \tau\}, \quad \tau \in (0, 1).$$

Because $\theta_i = m(\sigma_i) + s(\sigma_i)\xi_i$ is strictly increasing in ξ_i , posterior quantiles transform equivariantly:

$$q_{\theta_i}(\tau; \mathcal{D}_i) = m(\sigma_i) + s(\sigma_i) q_{\xi_i}(\tau; \tilde{\mathcal{D}}_i),$$

where $q_{\xi_i}(\tau; \tilde{\mathcal{D}}_i)$ denotes the posterior quantile of ξ_i under the normalized model (C.3).

Accordingly, the oracle NP-EBCI under precision dependence is

$$\text{CI}_i^{\text{NP*}, \text{CLOSE}} = \left[q_{\theta_i}(\alpha/2; \mathcal{D}_i), q_{\theta_i}(1 - \alpha/2; \mathcal{D}_i) \right]$$

or, equivalently,

$$\text{CI}_i^{\text{NP*}, \text{CLOSE}} = \left[m(\sigma_i) + s(\sigma_i) q_{\xi_i}(\alpha/2; \tilde{\mathcal{D}}_i), m(\sigma_i) + s(\sigma_i) q_{\xi_i}(1 - \alpha/2; \tilde{\mathcal{D}}_i) \right].$$

By construction, the oracle NP-EBCI achieves exact conditional coverage:

$$P\left(\theta_i \in \text{CI}_i^{\text{NP*}, \text{CLOSE}} \mid \mathcal{D}_i\right) = 1 - \alpha.$$

In practice, we replace the unknown functions $m(\cdot)$ and $s(\cdot)$ with their respective estimators, $\hat{m}(\cdot)$ and $\hat{s}(\cdot)$. We then construct the empirical normalized variables:

$$\hat{Z}_i = \frac{Y_i - \hat{m}(\sigma_i)}{\hat{s}(\sigma_i)}, \quad \hat{v}_i = \frac{\sigma_i}{\hat{s}(\sigma_i)}, \quad \hat{\tilde{\mathcal{D}}}_i = (\hat{Z}_i, \hat{v}_i).$$

Applying the baseline heteroskedastic posterior-quantile estimator in Section 4.1 to the normalized sample $\{(\hat{Z}_j, \hat{v}_j)\}_{j=1}^n$, we obtain $\hat{q}_{\xi_i}(\tau; \hat{\tilde{\mathcal{D}}}_i)$. The feasible NP-EBCI is then

$$\text{CI}_i^{\text{NP}, \text{CLOSE}} = \left[\hat{m}(\sigma_i) + \hat{s}(\sigma_i) \hat{q}_{\xi_i}(\alpha/2; \hat{\tilde{\mathcal{D}}}_i), \hat{m}(\sigma_i) + \hat{s}(\sigma_i) \hat{q}_{\xi_i}(1 - \alpha/2; \hat{\tilde{\mathcal{D}}}_i) \right].$$

We leave the comprehensive theoretical analysis of this NP-EBCI under precision dependence for future work.

Appendix D Auxiliary results

Proposition D.1. *Let $\rho_\tau(u) = u(\tau - \mathbf{1}\{u \leq 0\})$ be the check function. Then, for all $x, y \in \mathbb{R}$,*

$$\rho_\tau(x - y) - \rho_\tau(x) = -y(\tau - \mathbf{1}\{x \leq 0\}) + \int_0^y (\mathbf{1}\{x \leq s\} - \mathbf{1}\{x \leq 0\}) ds.$$

Proof. It follows from the identity from Knight (1998) and the reformulation of the check function $\rho_\tau(u) = (\tau - \frac{1}{2})u + \frac{1}{2}|u|$. ■

D.1 Uniform consistency of deconvolution kernel density estimator

Uniform consistency of the deconvolution kernel density estimator has been considered in the literature. In particular, Taylor and Zhang (1990) establish a stronger uniform strong-consistency result for a broader class of deconvolution estimators. Proposition D.2 is provided only as a self-contained specialization tailored to the deconvolution kernel density estimator used in our proof.

Proposition D.2. *Let $\hat{g}(\theta)$ be the deconvolution kernel density estimator defined in Equation (A.13). Let Assumptions 3.1, 3.2, 3.4 and 3.6 hold. Setting $h_n = c(\log n)^{-1/2}$ with any constant $c \geq \sigma$, we have that*

$$\sup_{\theta \in \mathbb{R}} |\hat{g}(\theta) - g(\theta)| \xrightarrow{P} 0.$$

Proof of Proposition D.2. We decompose $\hat{g}(\theta) - g(\theta) = S_{n1}(\theta) + S_{n2}(\theta)$ where

$$\begin{aligned} S_{n1}(\theta) &= (2\pi)^{-1} \int \frac{\exp(-it\theta)}{f_\varepsilon^\star(t)} \left(\hat{f}_Y^\star(t) - f_Y^\star(t) \right) K^\star(h_nt) dt \\ S_{n2}(\theta) &= (2\pi)^{-1} \int \frac{\exp(-it\theta)}{f_\varepsilon^\star(t)} f_Y^\star(t) (K^\star(h_nt) - 1) dt \end{aligned}$$

where $S_{n1}(\theta)$ is the stochastic deviation of the estimator from its expectation, $S_{n2}(\theta)$ is the bias term. Note that $\sup_{\theta \in \mathbb{R}} |\hat{g}(\theta) - g(\theta)| \leq \sup_{\theta \in \mathbb{R}} |S_{n1}(\theta)| + \sup_{\theta \in \mathbb{R}} |S_{n2}(\theta)|$.

To analyze the bias term $S_{n2}(\theta)$, note first that

$$\sup_{\theta \in \mathbb{R}} |S_{n2}(\theta)| \leq (2\pi)^{-1} \int |g^\star(t)| |K^\star(h_nt) - 1| dt$$

Since $K \in L^1(\mathbb{R})$, the function $K^\star$ is bounded and continuous, and $K^\star(0) = 1$. Thus $K^\star(h_nt) \rightarrow 1$ for every fixed t , while $|K^\star(h_nt) - 1| \leq \|K^\star\|_\infty + 1$. By Assumption 3.1, $g^\star \in L^1(\mathbb{R})$. Therefore the dominated convergence theorem yields that $\sup_{\theta \in \mathbb{R}} |S_{n2}(\theta)| = o(1)$.

For the first term $S_{n1}(\theta)$, since $K^\star$ is supported on $[-1, 1]$, we have that

$$\begin{aligned} \sup_{\theta \in \mathbb{R}} |S_{n1}(\theta)| &= \sup_{\theta \in \mathbb{R}} \left| (2\pi)^{-1} \int \frac{\exp(-it\theta)}{f_\varepsilon^\star(t)} \left(\hat{f}_Y^\star(t) - f_Y^\star(t) \right) K^\star(h_nt) dt \right| \\ &\leq \sup_{|t| \leq h_n^{-1}} \left| \hat{f}_Y^\star(t) - f_Y^\star(t) \right| \times (2\pi)^{-1} \int_{-h_n^{-1}}^{h_n^{-1}} \frac{|K^\star(h_nt)|}{|f_\varepsilon^\star(t)|} dt \end{aligned} \quad (\text{D.1})$$

where the last line obtains from Hölder's inequality. By Theorem 1 in Csörgő and Totik (1983), if

$\log(h_n^{-1})/n \rightarrow 0$, then

$$\sup_{|t| \leq h_n^{-1}} |\hat{f}_Y^*(t) - f_Y^*(t)| = O_p \left(\sqrt{\frac{\log(h_n^{-1})}{n}} \right).$$

Meanwhile, for the second product in Equation (D.1), we have that

$$\begin{aligned} (2\pi)^{-1} \int_{-h_n^{-1}}^{h_n^{-1}} \frac{|K^*(h_n t)|}{|f_\varepsilon^*(t)|} dt &\leq \frac{\|K\|_\infty}{2\pi} \int_{-h_n^{-1}}^{h_n^{-1}} \exp\left(\frac{\sigma^2 t^2}{2}\right) dt = \frac{\|K\|_\infty}{2\pi h_n} \int_{-1}^1 \exp\left(\frac{\sigma^2 u^2}{2h_n^2}\right) du \\ &\leq \frac{\|K\|_\infty}{\pi h_n} \int_0^1 \exp\left(\frac{\sigma^2 u}{2h_n^2}\right) du = O\left(h_n \exp\left(\frac{\sigma^2}{2h_n^2}\right)\right). \end{aligned}$$

Then if $h_n = c(\log n)^{-1/2}$ with $c \geq \sigma$, we obtain that

$$\sup_{\theta \in \mathbb{R}} |S_{n1}(\theta)| = O_p \left(h_n \exp\left(\frac{\sigma^2}{2h_n^2}\right) \sqrt{\frac{\log(h_n^{-1})}{n}} \right) = O_p \left(n^{\sigma^2/(2c^2)-1/2} \sqrt{\frac{\log \log n}{\log n}} \right) = o_p(1).$$

Combining the results for $S_{n1}(\theta)$ and $S_{n2}(\theta)$, we obtain that $\sup_{\theta \in \mathbb{R}} |\hat{g}(\theta) - g(\theta)| = o_p(1)$, as desired. $\blacksquare$

D.2 Technical Lemmas for the proof of Theorem 3.3

This subsection collects the auxiliary results used in Appendix A.4. The lemmas establish several ingredients of the lower-bound argument: the construction of the common perturbation $\bar{H}_T$, the verification that the alternatives g_m belong to $\mathcal{G}(s, L)$, the sign separation of the induced conditional coverage errors, and the bounds showing that the alternatives remain statistically close. Throughout this subsection, we maintain the notation of Appendix A.4.

D.2.1 Construction of common perturbation $\bar{H}_T$

Consider $\mathcal{J} = [\beta_L, \beta_U] \subset (0, \alpha)$ defined in Appendix A.4. Choose a constant $c_0 > 0$ such that for all sufficiently large T ,

$$\mathcal{J}_T := \left[\beta_L - \frac{c_0}{T}, \beta_U + \frac{c_0}{T} \right] \subset (0, \alpha) \quad \text{and} \quad \text{dist}(\mathcal{J}_T, \mathcal{J}_T + 1 - \alpha) > 0.$$

Let $\theta_L = q_{G_0}(\beta_L; y)$, $\theta_U = q_{G_0}(\beta_U + 1 - \alpha; y)$. We construct the common perturbation $\bar{H}_T$ as follows:

$$\bar{H}_T(\theta) := \eta T^{-s+1/2} \left[\kappa_-(T(\theta - \theta_L)) - \kappa_+(T(\theta - \theta_U)) \right]$$

where $\eta > 0$ is a sufficiently small constant, and $\kappa_-, \kappa_+ \in C_c^\infty(\mathbb{R})$ are nonnegative functions

satisfying

$$\int \kappa_-(u) du = \int \kappa_+(u) du = 1, \quad \text{supp}(\kappa_-) \subset [-2, -1], \quad \text{supp}(\kappa_+) \subset [1, 2].$$

Lemma D.1. (i) Define $\bar{g}_T := g_0 + \bar{H}_T$ where g_0 is the baseline Cauchy density, then $\int \bar{H}_T(\theta) d\theta = 0$, $\bar{g}_T(\theta) \geq c g_0(\theta) \geq 0$ for all $\theta \in \mathbb{R}$ for some constant $c > 0$ independent of T , and $\bar{g}_T \in G(s, 3L/4)$.

(ii) Define the conditional coverage error of the interval $I(\beta; y)$ under the common perturbation:

$$\bar{D}_T(\beta; y) := \mathbb{P}_{\bar{G}_T}(\theta \in I(\beta; y) \mid Y = y) - (1 - \alpha)$$

where $\bar{G}_T$ has density $\bar{g}_T$. Then there exists a constant $c > 0$ such that

$$\bar{D}_T(\beta; y) \geq c\eta T^{-s-1/2} \quad \text{for all } \beta \leq \beta_L - \frac{c_0}{T},$$

$$\bar{D}_T(\beta; y) \leq -c\eta T^{-s-1/2} \quad \text{for all } \beta \geq \beta_U + \frac{c_0}{T}.$$

Proof of Lemma D.1. First we check that

$$\int_{\mathbb{R}} \bar{H}_T(\theta) d\theta = \eta T^{-s+1/2} \left[\int \kappa_-(T(\theta - \theta_L)) d\theta - \int \kappa_+(T(\theta - \theta_U)) d\theta \right] = 0.$$

Its Fourier transform is

$$\bar{H}_T^*(t) = \eta T^{-s-1/2} \left[\exp(it\theta_L) \kappa_-^*(t/T) - \exp(it\theta_U) \kappa_+^*(t/T) \right],$$

so $\int_{\mathbb{R}} |\bar{H}_T^*(t)|^2 (1+t^2)^s dt \leq C\eta^2$ where C depends only on κ_- and κ_+ . Hence, for η sufficiently small,

$$\int_{\mathbb{R}} |\bar{H}_T^*(t)|^2 (1+t^2)^s dt \leq \frac{L}{4}.$$

Since $\bar{H}_T$ is supported on two $O(T^{-1})$ -neighborhoods of the fixed points θ_L and θ_U , and has size $O(\eta T^{-s+1/2})$, the same argument in Appendix A.4 implies $|\bar{H}_T(\theta)| \leq C\eta(1+\theta^2)^{-1}$ uniformly in θ and T . Because the baseline density g_0 satisfies $g_0(\theta) \geq c'_0(1+\theta^2)^{-1}$, choosing η sufficiently small yields

$$\bar{g}_T(\theta) = g_0(\theta) + \bar{H}_T(\theta) \geq c g_0(\theta) \geq 0$$

for all θ , and therefore $\bar{g}_T \in G(s, 3L/4)$, which proves part (i).

We now prove part (ii). We first consider the left-hand case $\beta \leq \beta_L - c_0/T$. We write $q(\beta) := q_{G_0}(\beta; y)$ and let $c_- = \inf_{\beta \in \mathcal{J}} q'(\beta) > 0$. By the mean-value theorem,

$$q(\beta_L) - q(\beta) = q'(\tilde{\beta})(\beta_L - \beta) \geq c_- \frac{c_0}{T} > \frac{2}{T}$$

for some $\tilde{\beta} \in [\beta, \beta_L] \subset \mathcal{J}$. Hence $q(\beta) \leq \theta_L - \frac{2}{T}$. Since $\text{supp}(\kappa_-(T(\cdot - \theta_L))) \subset [\theta_L - \frac{2}{T}, \theta_L - \frac{1}{T}]$, the

positive bump is contained in the interval $I(\beta; y)$. On the other hand, $\beta + 1 - \alpha \leq \beta_L + 1 - \alpha - \frac{c_0}{T} < \beta_U + 1 - \alpha$. So by monotonicity of q , $q(\beta + 1 - \alpha) < \theta_U$. Therefore $I(\beta; y)$ excludes the support of the negative bump, which lies in $[\theta_U + \frac{1}{T}, \theta_U + \frac{2}{T}]$.

The right-hand case $\beta \geq \beta_U + c_0/T$ is symmetric. Since $q(\beta + 1 - \alpha) - q(\beta_U + 1 - \alpha) \geq c_- \frac{c_0}{T} > \frac{2}{T}$, we have that $q(\beta + 1 - \alpha) \geq \theta_U + \frac{2}{T}$, which implies that $I(\beta; y)$ contains the support of the negative bump. Also $\beta > \beta_L$, hence $q(\beta) > \theta_L$, so $I(\beta; y)$ excludes the support of the positive bump.

A standard numerator-denominator expansion around G_0 therefore gives

$$\bar{D}_T(\beta; y) = f_{G_0}(y)^{-1} \int \phi\left(\frac{y - \theta}{\sigma}\right) (1\{\theta \in I(\beta; y)\} - (1 - \alpha)) \bar{H}_T(\theta) d\theta + o\left(T^{-s-1/2}\right),$$

uniformly over β . In the left-hand case, the coefficient $1\{\theta \in I(\beta; y)\} - (1 - \alpha)$ equals α on the support of the positive bump and equals $-(1 - \alpha)$ on the support of the negative bump; because the perturbation itself has signs $+$ and $-$ on these two supports, both contributions are positive and of order $\eta T^{-s-1/2}$. The right-hand case is symmetric and gives a negative quantity of the same order. This proves part (ii). ■

D.2.2 Verifying that alternatives $g_m \in \mathcal{G}(s, L)$

Recall from Section A.4.1 that the alternatives are defined by $g_m = g_0 + \bar{H}_T + \tilde{H}_{m,T}$. We now verify that each alternative g_m belongs to $\mathcal{G}(s, L)$. The argument has two parts. First, we control the Sobolev norm of $\tilde{H}_{m,T}$. We then show that this perturbation is uniformly small relative to the baseline density, so that g_m remains nonnegative.

Because each $H_{j,T}$ has an integral of zero, it follows that $\int \tilde{H}_{m,T}(\theta) d\theta = 0$. Moreover, the Fourier transform of the cumulative perturbation is

$$\tilde{H}_{m,T}^*(t) = \varepsilon \nu T^{-s-1/2} (-i \operatorname{sgn}(t)) k^*(|t|/T) S_{m,T}(t)$$

where

$$S_{m,T}(t) := \sum_{j < m} \exp(it\theta_j) - \sum_{j \geq m} \exp(it\theta_j).$$

By Lemma D.6, $|S_{m,T}(t)| \leq C$ uniformly in m and in t on the support of $k^*(|t|/T)$. Hence

$$\sup_{m,T} \int_{\mathbb{R}} |\tilde{H}_{m,T}^*(t)|^2 (1 + t^2)^s dt \leq C \varepsilon^2 \nu^2. \quad (\text{D.2})$$

It remains to verify the positivity of g_m . Since h belongs to the Schwartz space $S(\mathbb{R})$, taking $N = 4$ gives that

$$|h(u)| \leq C_4 (1 + |u|)^{-4} \quad \text{for all } u \in \mathbb{R}.$$

Therefore we can bound the cumulative perturbation by

$$|\tilde{H}_{m,T}(\theta)| \leq C\varepsilon\nu T^{-s+1/2} \sum_{j=1}^M (1 + T|\theta - \theta_j|)^{-4}. \quad (\text{D.3})$$

The centers $\{\theta_j\}_{j=1}^M$ lie within the compact set $q_{G_0}(\mathcal{J}; y)$ and have a spacing of order T^{-1} . Thus, for any θ inside a fixed compact set, the sum on the right-hand side of (D.3) is uniformly bounded. On the other hand, outside a fixed compact neighborhood of $q_{G_0}(\mathcal{J}; y)$, we have

$$\sum_{j=1}^M (1 + T|\theta - \theta_j|)^{-4} \leq CT^{-3}(1 + |\theta|)^{-4},$$

because $M \asymp T$. Combining these two bounds, and using $s \geq 1/2$, we obtain

$$|\tilde{H}_{m,T}(\theta)| \leq C\varepsilon\nu(1 + \theta^2)^{-1} \quad \text{for all } \theta, m,$$

for all sufficiently large T .

Now recall that g_0 is the baseline Cauchy density; thus, there exists $c > 0$ such that $g_0(\theta) \geq c(1 + \theta^2)^{-1}$ for all $\theta \in \mathbb{R}$. In addition, Lemma D.1 (i) gives that $\bar{g}_T(\theta) = g_0 + \bar{H}_T \geq c_1 g_0(\theta)$ for all θ and for all sufficiently large T . Then we may choose $\varepsilon > 0$ sufficiently small enough such that $g_m(\theta) = \bar{g}_T(\theta) + \tilde{H}_{m,T}(\theta) \geq c' g_0(\theta) \geq 0$ for all θ, m and all sufficiently large T .

Finally, $\bar{g}_T$ already satisfies the Sobolev bound by Lemma D.1 (i), and the display above gives the corresponding bound for $\tilde{H}_{m,T}$. We conclude that $g_m \in \mathcal{G}(s, L)$ for all m .

D.2.3 Sign separation of errors in conditional coverage probabilities

We consider the sign of the conditional coverage error $D_m(\beta; y)$. The argument has two steps. First, we decompose $D_m(\beta; y)$ into a leading linear term and a negligible common and remaining term. Second, we show that the leading term has a step-function sign pattern across the cells $\{B_k\}_{k=1}^M$. Since the other two terms are smaller than the separation scale $T^{-s-1/2}$, the actual conditional coverage error has the same sign pattern.

For each $m = 1, \dots, M$, recall that

$$\begin{aligned} D_m &= P_{G_m}(\theta \in I(\beta; y) \mid Y = y) - (1 - \alpha) \\ &= f_{G_m}^{-1}(y) \left[\int \phi\left(\frac{y - \theta}{\sigma}\right) (\mathbf{1}\{\theta \in I(\beta; y)\} - (1 - \alpha)) g_m(\theta) d\theta \right]. \end{aligned}$$

For any perturbation H , we also define

$$\Gamma_H(\beta; y) = f_{G_0}^{-1}(y) \left[\int \phi\left(\frac{y - \theta}{\sigma}\right) (\mathbf{1}\{\theta \in I(\beta; y)\} - (1 - \alpha)) H(\theta) d\theta \right].$$

This is the linearized effect of the perturbation H on the conditional coverage probability, evaluated around the baseline G_0 .

Lemma D.2. *For each $m = 1, \dots, M$, the error of conditional coverage probability has the decomposition*

$$D_m(\beta; y) = \bar{D}_T(\beta; y) + \Gamma_{\tilde{H}_{m,T}}(\beta; y) + \tilde{R}_m(\beta; y)$$

where $\bar{D}_T(\beta; y)$ is defined in Lemma D.1, and the remainder term is

$$\tilde{R}_m(\beta; y) = -\bar{D}_T(\beta; y) \frac{(\tilde{H}_{m,T} * f_\varepsilon)(y)}{f_{G_m}(y)} - \Gamma_{\tilde{H}_{m,T}}(\beta; y) \frac{((\bar{H}_T + \tilde{H}_{m,T}) * f_\varepsilon)(y)}{f_{G_m}(y)}.$$

Moreover,

$$\sup_m \sup_{\beta \in \mathcal{J}} |\tilde{R}_m(\beta; y)| = o\left(T^{-s-1/2}\right). \quad (\text{D.4})$$

Proof of Lemma D.2. By Lemma D.1 (ii), $\sup_{\beta \in \mathcal{J}} |\bar{D}_T(\beta; y)| \leq C\eta T^{-s-1/2}$.

Next, by linearity,

$$\Gamma_{\tilde{H}_{m,T}}(\beta; y) = \varepsilon \left(\sum_{j < m} \Gamma_{H_{j,T}}(\beta; y) - \sum_{j \geq m} \Gamma_{H_{j,T}}(\beta; y) \right).$$

Hence by Lemma D.7 (ii), $|\Gamma_{\tilde{H}_{m,T}}(\beta; y)| \leq \varepsilon C_0 T^{-s-1/2} \sum_{j=1}^M (1 + T|\beta - \beta_j|)^{-2}$. Because the grid points satisfy $\beta_{j+1} - \beta_j = \Delta/T$, the sum on the right-hand side is uniformly bounded over $\beta \in \mathcal{J}$. Therefore

$$\sup_m \sup_{\beta \in \mathcal{J}} |\Gamma_{\tilde{H}_{m,T}}(\beta; y)| \leq C\varepsilon T^{-s-1/2}.$$

We now bound the convolution terms. Since the support of $k^\star(|t|/T)$ is contained in the interval $[T, 2T]$, Lemma D.6 yields $|\tilde{H}_{m,T}^\star(t)| \leq C\varepsilon \nu T^{-s-1/2} |k^\star(|t|/T)|$. Hence

$$\left| (\tilde{H}_{m,T} * f_\varepsilon)(y) \right| \leq \frac{1}{2\pi} \int_{\mathbb{R}} |\tilde{H}_{m,T}^\star(t)| \exp(-\sigma^2 t^2 / 2) dt \leq C\varepsilon \nu T^{-s+1/2} \exp(-cT^2)$$

for some $c > 0$. For the common perturbation, we have that

$$|(\bar{H}_T * f_\varepsilon)(y)| \leq \left\| \phi\left(\frac{y - \cdot}{\sigma}\right) \right\|_\infty \|\bar{H}_T\|_1 \leq C\eta T^{-s-1/2}.$$

Therefore

$$|((\bar{H}_T + \tilde{H}_{m,T}) * f_\varepsilon)(y)| \leq C\eta T^{-s-1/2} + C\varepsilon \nu T^{-s+1/2} \exp(-cT^2).$$

Finally, by Lemma D.1(i), $f_{\tilde{G}_T}(y) \geq c f_{G_0}(y) > 0$. Since $(\tilde{H}_{m,T} * f_\varepsilon)(y) = o(1)$ uniformly in m , it follows that $\inf_m f_{G_m}(y) \geq c_y > 0$ for all sufficiently large T .

Combining the previous bounds, uniformly in m and $\beta \in \mathcal{J}$,

$$\begin{aligned} |\tilde{R}_m(\beta; y)| &\leq C|\bar{D}_T(\beta; y)| |(\tilde{H}_{m,T} * f_\varepsilon)(y)| + C|\Gamma_{\tilde{H}_{m,T}}(\beta; y)| |((\bar{H}_T + \tilde{H}_{m,T}) * f_\varepsilon)(y)| \\ &\leq C\eta\varepsilon\nu T^{-2s} \exp(-cT^2) + C\eta\varepsilon T^{-2s-1} + C\varepsilon^2\nu T^{-2s} \exp(-cT^2) = o\left(T^{-s-1/2}\right) \end{aligned}$$

as desired. ■

Recall from Section A.4.1 that

$$B_k = \left[\beta_k - \frac{\Delta}{2T}, \beta_k + \frac{\Delta}{2T} \right] \cap \mathcal{J}, \quad k = 1, \dots, M.$$

Suppose that Δ is chosen so that

$$2C_0 \sum_{\ell \geq 1} \left(1 + \Delta \left(\ell - \frac{1}{2} \right) \right)^{-2} \leq \frac{c_0}{4}, \quad (\text{D.5})$$

and that r is chosen so that

$$C_0 \sum_{\ell \geq r} \left(1 + \Delta \left(\ell - \frac{1}{2} \right) \right)^{-2} \leq \frac{c_0}{4}.$$

Lemma D.3. *For each $m = 1, \dots, M$,*

$$\Gamma_{\tilde{H}_{m,T}}(\beta; y) \geq \frac{\varepsilon c_0}{2} T^{-s-1/2}, \quad \text{for } \beta \in \bigcup_{k \leq m-r} B_k \quad (\text{D.6})$$

$$\Gamma_{\tilde{H}_{m,T}}(\beta; y) \leq -\frac{\varepsilon c_0}{2} T^{-s-1/2}, \quad \text{for } \beta \in \bigcup_{k \geq m+r} B_k. \quad (\text{D.7})$$

Proof of Lemma D.3. We prove the first claim; the second is symmetric. Fix $k \leq m - r$ and let $\beta \in B_k$. Isolating the $j = k$ term gives

$$\begin{aligned} \Gamma_{\tilde{H}_{m,T}}(\beta; y) &= \varepsilon \left(\Gamma_{H_{k,T}}(\beta; y) + \sum_{\substack{j < m \\ j \neq k}} \Gamma_{H_{j,T}}(\beta; y) - \sum_{j \geq m} \Gamma_{H_{j,T}}(\beta; y) \right) \\ &\geq \varepsilon \left(\Gamma_{H_{k,T}}(\beta; y) - \sum_{\substack{j < m \\ j \neq k}} |\Gamma_{H_{j,T}}(\beta; y)| - \sum_{j \geq m} |\Gamma_{H_{j,T}}(\beta; y)| \right). \end{aligned}$$

Since $\beta \in B_k$, Lemma D.7 (iii) yields $\Gamma_{H_{k,T}}(\beta; y) \geq c_0 T^{-s-1/2}$. For $j \neq k$, the same spacing argument gives that

$$T|\beta - \beta_j| \geq T|\beta_j - \beta_k| - T|\beta - \beta_k| \geq \Delta|j - k| - \frac{\Delta}{2} \geq \Delta \left(|j - k| - \frac{1}{2} \right). \quad (\text{D.8})$$

Then by Lemma D.7 (ii), we obtain that

$$\sum_{\substack{j \leq m \\ j \neq k}} |\Gamma_{H_{j,T}}(\beta; y)| \leq \frac{c_0}{4} T^{-s-1/2}, \quad \sum_{j \geq m} |\Gamma_{H_{j,T}}(\beta; y)| \leq \frac{c_0}{4} T^{-s-1/2}.$$

Combining these three bounds gives (D.6), as claimed. ■

We now turn to the actual errors in conditional coverage probability. By Lemma D.1 (ii), $\sup_{\beta \in \mathcal{J}} |\bar{D}_T(\beta; y)| \leq C\eta T^{-s-1/2}$. Choose η so that $C\eta \leq \varepsilon c_0/8$. Then combining with Lemma D.2 and Lemma D.3, we have (A.20)-(A.21) for all sufficiently large T . Thus, the actual conditional coverage error has the same step-function sign pattern as its linearization, up to the transition band.

D.2.4 Upper bounds on divergences

Recall from Appendix A.4.2 that P_m denotes the law of $\mathbf{Y}_{-i} = \{Y_j\}_{j \neq i}$ under g_m , and $(f_Y)_m$ is the corresponding marginal density, and $F_G(\cdot | y)$ denotes the posterior distribution of θ_i given $Y_i = y$ under the mixing distribution G .

Lemma D.4. *Recall that $u_{m,m'} := (f_Y)_m - (f_Y)_{m'} = (g_m - g_{m'}) * f_\varepsilon$, we have*

$$\int_{\mathbb{R}} \frac{u_{m,m'}(y)^2}{(f_Y)_0(y)} dy \leq C\varepsilon^2 \nu^2 T^{-2s+2} \exp(-\sigma^2 T^2).$$

Proof of Lemma D.4. First, as shown in Appendix A.2, we have $(f_Y)_0(y) \geq c(1+y^2)^{-1}$. Therefore, by Parseval's identity,

$$\begin{aligned} \int_{\mathbb{R}} \frac{u_{m,m'}(y)^2}{(f_Y)_0(y)} dy &\leq C \int_{\mathbb{R}} (1+y^2) u_{m,m'}(y)^2 dy \\ &= \frac{C}{2\pi} \int_{\mathbb{R}} (|u_{m,m'}^*(t)|^2 + |\partial_t u_{m,m'}^*(t)|^2) dt. \end{aligned}$$

Fix $m < m'$. Then $g_m - g_{m'} = -2\varepsilon \sum_{j=m}^{m'-1} H_{j,T}$. Taking the Fourier transformation yields that

$$(g_m - g_{m'})^*(t) = -2\varepsilon \nu T^{-s-1/2} (-i \operatorname{sgn}(t)) k^*(|t|/T) \sum_{j=m}^{m'-1} \exp(it\theta_j).$$

By Lemma D.6, the oscillatory sum and its derivative are bounded uniformly on $t \in [T, 2T]$:

$$\sup_{m,m'} \sup_{t \in [T, 2T]} \left| \sum_{j=m}^{m'-1} \exp(it\theta_j) \right| \leq C, \quad \text{and} \quad \sup_{m,m'} \sup_{t \in [T, 2T]} \left| \partial_t \left(\sum_{j=m}^{m'-1} \exp(it\theta_j) \right) \right| \leq CT.$$

Recall that $u_{m,m'}^*(t) = (g_m - g_{m'})^*(t) \exp(-\sigma^2 t^2/2)$.

On the support of $k^*(|t|/T)$, we have $|t| \asymp T$. Applying the oscillatory bounds gives:

$$|u_{m,m'}^*(t)| \leq C\varepsilon\nu T^{-s-1/2} |k^*(|t|/T)| \exp(-\sigma^2 t^2/2),$$

$$|\partial_t u_{m,m'}^*(t)| \leq C\varepsilon\nu T^{-s+1/2} |k^*(|t|/T)| \exp(-\sigma^2 t^2/2).$$

Because the support of $k^*(|t|/T)$ has length $O(T)$, integrating these point-wise bounds yields:

$$\int_{\mathbb{R}} |u_{m,m'}^*(t)|^2 dt \leq C\varepsilon^2\nu^2 T^{-2s} \exp(-\sigma^2 T^2),$$

$$\int_{\mathbb{R}} |\partial_t u_{m,m'}^*(t)|^2 dt \leq C\varepsilon^2\nu^2 T^{-2s+2} \exp(-\sigma^2 T^2).$$

The derivative term clearly dominates. Substituting this upper bound back into our Plancherel inequality gives the desired claim. ■

Lemma D.5. *For all sufficiently large T ,*

$$\sup_m \sup_{x \in \mathbb{R}} |F_{G_m}(x | y) - F_{G_0}(x | y)| \leq CT^{-s-1/2}.$$

Proof of Lemma D.5. We define the effect of a perturbation H on this posterior distribution:

$$\Lambda_H(x; y) := f_{G_0}^{-1}(y) \left[\int \phi\left(\frac{y - \theta}{\sigma}\right) (\mathbf{1}\{\theta \leq x\} - F_{G_0}(x | y)) H(\theta) d\theta \right].$$

By part (i) of Lemma D.7,

$$|\Lambda_{H_{j,T}}(x; y)| \leq C_0 T^{-s-1/2} (1 + T |F_{G_0}(x | y) - \beta_j|)^{-2}.$$

Using similar arguments as in the proof of Lemma D.2, we obtain that

$$\sup_m \sup_{x \in \mathbb{R}} |\Lambda_{\tilde{H}_{m,T}}(x; y)| \leq \sup_m \sup_{x \in \mathbb{R}} \varepsilon \sum_{j=1}^M |\Lambda_{H_{j,T}}(x; y)| \leq CT^{-s-1/2}$$

where we used Lemma D.7 (i), (D.5), and (D.8).

The same Fourier arguments used in Lemma D.7 (i), applied to the two compactly supported bumps in $\tilde{H}_T$, yields that $\sup_{x \in \mathbb{R}} |\Lambda_{\tilde{H}_T}(x; y)| \leq C\eta T^{-s-1/2}$, and then

$$\sup_m \sup_{x \in \mathbb{R}} |\Lambda_{\tilde{H}_T + \tilde{H}_{m,T}}(x; y)| \leq CT^{-s-1/2}.$$

Now we apply the numerator-denominator expansion

$$F_{G_0+H}(x | y) - F_{G_0}(x | y) = \Lambda_H(x; y) - \Lambda_H(x; y) \frac{(H * f_\varepsilon)(y)}{f_{G_0+H}(y)}.$$

Since the denominator remains bounded away from zero for the perturbations, the previous expansion gives the desired result. ■

D.2.5 Oscillatory-sum and localization lemmas

Lemma D.6 (Oscillatory exponential sums). *For the grid points $\{\theta_m\}_{m=1}^M$ defined in Appendix A.4, there exists a constant C , independent of m', m, t , and T , such that for all sufficiently large T ,*

$$\sup_{m, m'} \sup_{t \in [T, 2T]} \left| \sum_{j=m}^{m'-1} \exp(it\theta_j) \right| \leq C, \quad \text{and} \quad \sup_{m, m'} \sup_{t \in [T, 2T]} \left| \partial_t \left(\sum_{j=m}^{m'-1} \exp(it\theta_j) \right) \right| \leq CT.$$

Consequently,

$$|S_{m,T}(t)| = \left| \sum_{j < m} \exp(it\theta_j) - \sum_{j \geq m} \exp(it\theta_j) \right| \leq C$$

uniformly in m and in $t \in [T, 2T]$.

Proof of Lemma D.6. We write $q(\beta) := q_{G_0}(\beta; y)$ so that $\theta_j = q(\beta_j)$. Because the baseline density g_0 is Cauchy, the posterior distribution $F_{G_0}(\cdot | y)$ has a continuous density that is strictly positive on the compact set $q_{G_0}(\mathcal{J}; y)$. Consequently, the inverse map $\beta \mapsto q(\beta)$ is twice continuously differentiable on $\mathcal{J}$. Thus, there exist constants $0 < c_- \leq c_+ < \infty$ and $C_2 < \infty$ such that for all $\beta \in \mathcal{J}$

$$c_- \leq q'(\beta) \leq c_+, \quad |q''(\beta)| \leq C_2.$$

Let $\Delta_j(t) := t(\theta_{j+1} - \theta_j)$ for $j = 1, \dots, M-1$. The grid points are evenly spaced such that $\beta_{j+1} - \beta_j = \Delta T^{-1}$. By the mean-value theorem, $\theta_{j+1} - \theta_j = q(\beta_{j+1}) - q(\beta_j) = q'(\tilde{\beta}_j) \Delta T^{-1}$ for some $\tilde{\beta}_j \in (\beta_j, \beta_{j+1}) \subset \mathcal{J}$. Therefore, for $t \in [T, 2T]$, the phase difference is bounded by $\Delta c_- \leq \Delta_j(t) \leq 2\Delta c_+$. Then choose $\Delta > 0$ small enough that $2\Delta c_+ < \pi$, and set $\delta := \min\{\Delta c_-, \pi - 2\Delta c_+\} > 0$. Then we obtain that $0 < \delta \leq \Delta_j(t) \leq \pi - \delta$ for all $j \leq M-1, t \in [T, 2T]$.

Also, by the mean-value theorem for second differences, $\theta_{j+1} - 2\theta_j + \theta_{j-1} = q(\beta_{j+1}) - 2q(\beta_j) + q(\beta_{j-1}) = q''(\hat{\beta}_j) \Delta^2 T^{-2}$ for some $\hat{\beta}_j \in (\beta_{j-1}, \beta_{j+1}) \subset \mathcal{J}$. Hence $|\Delta_j(t) - \Delta_{j-1}(t)| \leq CtT^{-2} \leq CT^{-1}$ for $j = 2, \dots, M-1, t \in [T, 2T]$.

We now prove the partial-sum bound. Fix $1 \leq m \leq m' \leq M+1$. If $m = m'$, the sum is zero, so assume $m < m'$. If $m' = M+1$, we can split the sum as $\sum_{j=m}^{m'-1} \exp(it\theta_j) = \sum_{j=m}^{M-1} \exp(it\theta_j) + \exp(it\theta_M)$. Because the single tail term $\exp(it\theta_M)$ is trivially bounded by 1, it suffices to treat

the case $m < m' \leq M$. For such m' , define the multiplier $b_j(t) := 1/[1 - \exp(i\Delta_j(t))]$ for $j = m, \dots, m' - 1$. Because $\Delta_j(t) \in [\delta, \pi - \delta]$, the denominator is bounded away from zero, implying $|b_j(t)| \leq C$. Using $\exp(it\theta_j) = b_j(t) [\exp(it\theta_j) - \exp(it\theta_{j+1})]$, summation by parts yields

$$\sum_{j=m}^{m'-1} \exp(it\theta_j) = b_m(t) \exp(it\theta_m) - b_{m'-1}(t) \exp(it\theta_{m'}) + \sum_{j=m+1}^{m'-1} (b_j(t) - b_{j-1}(t)) \exp(it\theta_j).$$

Since $u \mapsto (1 - e^{iu})^{-1}$ is continuously differentiable on $[\delta, \pi - \delta]$, hence Lipschitz continuous on that interval. Therefore

$$|b_j(t) - b_{j-1}(t)| \leq C|\Delta_j(t) - \Delta_{j-1}(t)| \leq CT^{-1}.$$

Since $M \lesssim T$, the sum of the absolute differences is bounded: $\sum_{j=m+1}^{m'-1} |b_j(t) - b_{j-1}(t)| \leq C$. Applying the triangle inequality to the summation-by-parts expansion gives that $\left| \sum_{j=m}^{m'-1} \exp(it\theta_j) \right| \leq C$ uniformly in m, m', t , and T .

For the derivative bound, note that $|\theta_j| = |q(\beta_j)| \leq \sup_{\beta \in \mathcal{J}} |q(\beta)| \leq C$ uniformly in j . Hence

$$\left| \partial_t \sum_{j=m}^{m'-1} \exp(it\theta_j) \right| = \left| i \sum_{j=m}^{m'-1} \theta_j \exp(it\theta_j) \right| \leq C \sum_{j=m}^{m'-1} 1 \leq CM \leq CT.$$

The final claim for $S_{m,T}$ follows from the triangle inequality. ■

Lemma D.7 (Localized perturbations). *Assume the notation of Appendix A.4. Also recall that*

$$\Lambda_H(x; y) := f_{G_0}^{-1}(y) \left[\int \phi \left(\frac{y - \theta}{\sigma} \right) \left(\mathbf{1}_{\{\theta \leq x\}} - F_{G_0}(x | y) \right) H(\theta) d\theta \right].$$

Then there exist constants $c_0, C_0 > 0$, independent of m and T , s.t. for all sufficiently large T :

- (i) *for every $x \in \mathbb{R}$, $|\Lambda_{H_{m,T}}(x; y)| \leq C_0 T^{-s-1/2} (1 + T|F_{G_0}(x | y) - \beta_m|)^{-2}$;*
- (ii) *for every $\beta \in \mathcal{J}$, $|\Gamma_{H_{m,T}}(\beta; y)| \leq C_0 T^{-s-1/2} (1 + T|\beta - \beta_m|)^{-2}$;*
- (iii) *Let $B_m := [\beta_m - \frac{\Delta}{2T}, \beta_m + \frac{\Delta}{2T}] \cap \mathcal{J}$, then $\inf_{\beta \in B_m} \Gamma_{H_{m,T}}(\beta; y) \geq c_0 T^{-s-1/2}$.*

Proof of Lemma D.7. We choose a compact interval $\mathcal{J}^+ \subset (0, 1)$ such that $\mathcal{J} \cup (\mathcal{J} + 1 - \alpha) \subset \text{int}(\mathcal{J}^+)$, and the posterior quantile $q_{G_0}(\beta; y)$ is continuously differentiable on this $\mathcal{J}^+$. Therefore there exist constants $0 < c_- \leq c_+ < \infty$ such that, for all $\beta \in \mathcal{J}^+$ and all $m \leq M$,

$$c_- |\beta - \beta_m| \leq |q_{G_0}(\beta; y) - \theta_m| \leq c_+ |\beta - \beta_m|. \quad (\text{D.9})$$

Proof of part (i).

We write that $\Lambda_{H_{m,T}}(x; y) := f_{G_0}^{-1}(y) \int \psi_x(\theta; y) H_{m,T}(\theta) d\theta$ where

$$\psi_x(\theta; y) := \phi\left(\frac{y - \theta}{\sigma}\right) \left\{ \mathbf{1}\{\theta \leq x\} - F_{G_0}(x | y) \right\}.$$

As in Proposition A.1, integration by parts gives

$$\overline{\psi_x^*(t; y)} = \phi\left(\frac{y - x}{\sigma}\right) \frac{\exp(-itx)}{it} + r_x(t; y), \quad |r_x(t; y)| \leq Ct^{-2}, \quad (\text{D.10})$$

uniformly in x over compact sets. Then by Parseval's identity and the specific form of the perturbation $H_{m,T}^*(t) = \nu T^{-s-1/2}(-i \operatorname{sgn}(t)) \exp(it\theta_m) k^*(|t|/T)$, we decompose that $\Lambda_{H_{m,T}}(x; y) = \Lambda_{m,T,1} + \Lambda_{m,T,2}$ where

$$\Lambda_{m,T,1} = \frac{\nu T^{-s-1/2}}{2\pi f_{G_0}(y)} \phi\left(\frac{y - x}{\sigma}\right) \int \frac{\exp(-itx)}{it} (-i \operatorname{sgn}(t)) \exp(it\theta_m) k^*(|t|/T) dt \quad (\text{D.11})$$

and

$$\Lambda_{m,T,2} = \frac{1}{2\pi f_{G_0}(y)} \int r_x(t; y) H_{m,T}^*(t) dt. \quad (\text{D.12})$$

For the first term (D.11), since $(-i \operatorname{sgn}(t))/(it) = -1/|t|$, we have that

$$\begin{aligned} \Lambda_{m,T,1} &= -\frac{\nu T^{-s-1/2}}{2\pi f_{G_0}(y)} \phi\left(\frac{y - x}{\sigma}\right) \int \exp(it(\theta_m - x)) \frac{k^*(|t|/T)}{|t|} dt \\ &\stackrel{(a)}{=} -\frac{\nu T^{-s-1/2}}{2\pi f_{G_0}(y)} \phi\left(\frac{y - x}{\sigma}\right) \int_{|u| \in [1,2]} \exp(iTu(\theta_m - x)) \frac{k^*(|u|)}{|u|} du \\ &\stackrel{(b)}{=} -\frac{\nu T^{-s-1/2}}{2\pi f_{G_0}(y)} \phi\left(\frac{y - x}{\sigma}\right) \int_1^2 \frac{k^*(u)}{u} \left[\exp(iuT(\theta_m - x)) + \exp(-iuT(\theta_m - x)) \right] du \\ &= -\frac{\nu T^{-s-1/2}}{\pi f_{G_0}(y)} \phi\left(\frac{y - x}{\sigma}\right) \int_1^2 \frac{k^*(u)}{u} \cos(uT(\theta_m - x)) du \\ &= -\nu T^{-s-1/2} \frac{\phi((y - x)/\sigma)}{f_{G_0}(y)} \mathcal{K}(T(\theta_m - x)) \end{aligned}$$

where (a) is obtained by the change of variable $t = Tu$, (b) is obtain by splitting positive and negative u , and the kernel $\mathcal{K}$ in the last step is defined as

$$\mathcal{K}(v) := \frac{1}{\pi} \int_1^2 \frac{k^*(u)}{u} \cos(uv) du.$$

For the second term (D.12), we use that $|r_x(t; y)| \leq C|t|^{-2}$ and $|H_{m,T}^*(t)| \leq CT^{-s-1/2}|k^*(|t|/T)|$. And also $k^*(|t|/T)$ is supported where $|t| \asymp T$, we have that $|\Lambda_{m,T,2}| \leq CT^{-s-1/2} \int_{|t| \gtrsim T} |t|^{-2} dt =$

$CT^{-s-1/2}T^{-1} = CT^{-s-3/2}$. Combining with the first term, we obtain that

$$\Lambda_{H_{m,T}}(x; y) = -\nu T^{-s-1/2} \frac{\phi((y-x)/\sigma)}{f_{G_0}(y)} \mathcal{K}(T(\theta_m - x)) + O(T^{-s-3/2}).$$

Since $u \mapsto k^\star(u)/u$ belongs to $C_c^\infty((1, 2))$, we have $\mathcal{K} \in \mathcal{S}(\mathbb{R})$, and hence $|\mathcal{K}(v)| \leq C(1 + |v|)^{-2}$ and

$$|\Lambda_{H_{m,T}}(x; y)| \lesssim CT^{-s-1/2} (1 + T|x - \theta_m|)^{-2}$$

uniformly over $x \in q_{G_0}(\mathcal{J}^+; y)$.

To bound $\Lambda_{H_{m,T}}(x; y)$ uniformly over $x \in \mathbb{R}$, we consider two cases. If $F_{G_0}(x | y) \in \mathcal{J}^+$, then $x = q_{G_0}(\beta; y)$ for some $\beta \in \mathcal{J}^+$. By Equation (D.9), we have that

$$|\Lambda_{H_{m,T}}(x; y)| \lesssim CT^{-s-1/2} \left(1 + T|F_{G_0}(x | y) - \beta_m|\right)^{-2}.$$

If $F_{G_0}(x | y) \notin \mathcal{J}^+$, the asymptotic expansion (D.10) does not apply, so we need a different argument. First observe that there exists d_0 such that $|F_{G_0}(x | y) - \beta_m| \geq d_0$ for all m . Because $q_{G_0}(\cdot; y)$ is continuous and strictly increasing, we have that there exists $d_1 > 0$ such that $|x - \theta_m| = |q_{G_0}(F_{G_0}(x | y); y) - q_{G_0}(\beta_m; y)| \geq d_1 > 0$ uniformly in m .

Now define $U_m = \{|\theta - \theta_m| \leq d_1/2\}$, the function $\theta \mapsto \psi_x(\theta; y)$ is C^2 on a fixed neighborhood of θ_m , with derivatives bounded uniformly in x and m . We let $P_{x,m}(\theta) := \psi_x(\theta_m; y) + \psi'_x(\theta_m; y)(\theta - \theta_m)$ be the first-order Taylor polynomial of $\psi_x(\cdot; y)$ at θ_m .

Also, given that $H_{m,T}^\star(t) = \nu T^{-s-1/2}(-i \operatorname{sgn}(t)) \exp(it\theta_m) k^\star(|t|/T)$, we have

$$\begin{aligned} \int H_{m,T}(\theta) d\theta &= H_{m,T}^\star(0) = 0, \\ \int (\theta - \theta_m) H_{m,T}(\theta) d\theta &= -i \partial_t \left(\exp(-it\theta_m) H_{m,T}^\star(t) \right) \Big|_{t=0} = 0. \end{aligned}$$

Hence we obtain that

$$\begin{aligned} \Lambda_{H_{m,T}}(x; y) &= f_{G_0}^{-1}(y) \int (\psi_x(\theta; y) - P_{x,m}(\theta)) H_{m,T}(\theta) d\theta \\ &= f_{G_0}^{-1}(y) \left(\int_{U_m} + \int_{U_m^c} \right) (\psi_x(\theta; y) - P_{x,m}(\theta)) H_{m,T}(\theta) d\theta \end{aligned}$$

On U_m , Taylor's theorem gives that $|\psi_x(\theta; y) - P_{x,m}(\theta)| \leq C|\theta - \theta_m|^2$ uniformly in x, m . Therefore

$$\begin{aligned} \left| \int_{U_m} (\psi_x(\theta; y) - P_{x,m}(\theta)) H_{m,T}(\theta) d\theta \right| &\leq C \int_{U_m} |\theta - \theta_m|^2 |H_{m,T}(\theta)| d\theta \\ &\stackrel{(a)}{\leq} CT^{-s-5/2} \int |u|^2 |h(u)| du \stackrel{(b)}{\leq} CT^{-s-5/2} \end{aligned}$$

where (a) uses the changes of variable $u = T(\theta - \theta_m)$ and (b) is due to $h \in \mathcal{S}(\mathbb{R})$.

On U_m^c , since $|\psi_x(\theta; y) - P_{x,m}(\theta)| \leq C(1 + |\theta - \theta_m|)$, we have that

$$\begin{aligned} \left| \int_{U_m^c} \left(\psi_x(\theta; y) - P_{x,m}(\theta) \right) H_{m,T}(\theta) d\theta \right| &\leq C \int_{U_m^c} (1 + |\theta - \theta_m|) |H_{m,T}(\theta)| d\theta \\ &\stackrel{(a)}{=} \nu T^{-s-1/2} \int_{|u| > Td_1/2} \left(1 + \frac{|u|}{T} \right) |h(u)| du \\ &\stackrel{(b)}{\leq} \nu T^{-s-1/2} CT^{-3} = CT^{-s-7/2} \end{aligned}$$

where (a) uses again $u = T(\theta - \theta_m)$ and (b) uses $|h(u)| \leq C(1 + |u|)^{-5}$ as $h \in \mathcal{S}(\mathbb{R})$.

Combining these two bounds gives that $|\Lambda_{H_{m,T}}(x; y)| \leq CT^{-s-5/2}$ whenever $F_{G_0}(x | y) \notin \mathcal{J}^+$. Since $|F_{G_0}(x | y) - \beta_m| \geq d_0$, it follows that $T^{-s-5/2} \leq CT^{-s-1/2}(1 + T|F_{G_0}(x | y) - \beta_m|)^{-2}$ and thus again

$$|\Lambda_{H_{m,T}}(x; y)| \leq CT^{-s-5/2} \leq CT^{-s-1/2} \left(1 + T|F_{G_0}(x | y) - \beta_m| \right)^{-2}.$$

which completes the proof of part (i).

Proof of part (ii).

By the definition of $I(\beta; y)$, we can write

$$\Gamma_H(\beta; y) = \Lambda_H(q_{G_0}(\beta + 1 - \alpha; y); y) - \Lambda_H(q_{G_0}(\beta; y); y). \quad (\text{D.13})$$

Applying part (i) to the second term of (D.13) gives

$$\left| \Lambda_{H_{m,T}}(q_{G_0}(\beta; y); y) \right| \leq CT^{-s-1/2} (1 + T|\beta - \beta_m|)^{-2}.$$

For the first term of (D.13), we also apply part (i) and obtain that

$$\left| \Lambda_{H_{m,T}}(q_{G_0}(\beta + 1 - \alpha; y); y) \right| \leq CT^{-s-1/2} (1 + T|(\beta + 1 - \alpha) - \beta_m|)^{-2}.$$

Now with the assumption $\text{dist}(\mathcal{J}, \mathcal{J} + 1 - \alpha) > 0$ and $\beta_m \in \mathcal{J}$, we have that there exists $d_0 > 0$ such that $|(\beta + 1 - \alpha) - \beta_m| \geq d_0$ for all $\beta \in \mathcal{J}$ and all $m \leq M$. Therefore

$$\begin{aligned} \left| \Lambda_{H_{m,T}}(q_{G_0}(\beta + 1 - \alpha; y); y) \right| &\leq CT^{-s-1/2} (1 + Td_0)^{-2} \\ &\leq CT^{-s-5/2} \leq CT^{-s-1/2} (1 + T|\beta - \beta_m|)^{-2} \end{aligned}$$

where the final step follows because $\beta, \beta_m \in \mathcal{J}$ implies the uniform bound $|\beta - \beta_m| \leq C'$, which yields that $T^{-2} \leq C'(1 + T|\beta - \beta_m|)^{-2}$.

Combining these bounds via the triangle inequality directly yields

$$\left| \Gamma_{H_{m,T}}(\beta; y) \right| \leq C_0 T^{-s-1/2} (1 + T|\beta - \beta_m|)^{-2},$$

which proves part (ii).

Proof of part (iii).

First note that

$$\mathcal{K}(0) = \frac{1}{\pi} \int_1^2 \frac{k^*(u)}{u} du > 0.$$

By continuity of $\mathcal{K}$, there exists $\delta_0 > 0$ such that $\mathcal{K}(v) \geq \frac{1}{2}\mathcal{K}(0) > 0$ whenever $|v| \leq \delta_0$.

Choose $\Delta > 0$ so small that $c_+ \Delta/2 \leq \delta_0$. If $\beta \in B_m$, then (D.9) gives

$$|q_{G_0}(\beta; y) - \theta_m| \leq c_+ |\beta - \beta_m| \leq \frac{c_+ \Delta}{2T} \leq \frac{\delta_0}{T},$$

this ensures that the argument of $\mathcal{K}$ satisfies $T|\theta_m - q_{G_0}(\beta; y)| \leq \delta_0$, and thus $\mathcal{K}(T(\theta_m - q_{G_0}(\beta; y))) \geq \frac{1}{2}\mathcal{K}(0)$. Because $\beta \mapsto f_{G_0}^{-1}(y)\phi((y - q_{G_0}(\beta; y))/\sigma)$ is continuous and strictly positive on $\mathcal{J}$, we obtain the uniform lower bound $-\Lambda_{H_{m,T}}(q_{G_0}(\beta; y); y) \geq cT^{-s-1/2}$ for all $\beta \in B_m$. On the other hand, as shown in part (ii), $q_{G_0}(\beta + 1 - \alpha; y)$ remains uniformly separated from θ_m , so

$$|\Lambda_{H_{m,T}}(q_{G_0}(\beta + 1 - \alpha; y); y)| \leq CT^{-s-1/2}T^{-2}.$$

Therefore, for all sufficiently large T , the dominant term determines the lower bound:

$$\Gamma_{H_{m,T}}(\beta; y) = \Lambda_{H_{m,T}}(q_{G_0}(\beta + 1 - \alpha; y); y) - \Lambda_{H_{m,T}}(q_{G_0}(\beta; y); y) \geq c_0 T^{-s-1/2}$$

for all $\beta \in B_m$. This proves part (iii). ■