Quantile Mixture Models: Estimation and Inference *

Luis A. F. Alvarez[†] Victor M. Orestes[‡]

July, 2023

Abstract

Nonparametric density mixture models have received increasing attention from Statistics and Econometrics. The standard approach to estimation in these models is the nonparametric MLE of Kiefer and Wolfowitz (1956). In spite of its attractive statistical properties, this method suffers from computational and inferential hurdles. Motivated by these difficulties, this paper introduces nonparametric quantile mixture models as a convenient counterpart to density mixture models. We show that, similarly to nonparametric density mixtures, nonparametric mixtures of quantiles enjoy interesting approximation properties.

We introduce a computationally attractive sieve approach to estimate quantile mixture models. Our method relies on *L-moments*, robust alternatives to standard moments that characterize any distribution with finite first moment. We develop a full inferential theory for our proposed estimator, by working with the concept of strong approximation. In doing so, we make two contributions to statistical theory. First, we provide a lemma that bounds the strong approximation of a quadratic minimizer by another optimizer in terms of the constituent elements of each program. Second, we develop an approach to extending anticoncentration inequalities available in the literature to a constrained estimation setting. We believe these tools may be of independent interest.

We consider two applications of our proposed methodology. First, we show how one can use quantile mixtures to recover estimates and conduct inference on the mixture weights of a density mixture model. Second, we show that, as a direct byproduct of our theory, we are able to provide an inference method for the distributional synthetic controls of Gunsilius (2023), a novel approach to counterfactual analysis for which formal inference methods were not yet available. As an empirical application of the latter, we apply our proposed approach to inference in assessing the effects of the Brumadinho barrage rupture on the local wage distribution. Our results uncover a range of effects across percentiles, which we argue are consistent with *displacement effects*, whereby median-earning jobs are replaced by low-paying contracts.

Keywords: mixture models; L-moments; strong approximation; anti-concentration; synthetic controls.

---

*We would like to thank Alberto Abadie, Bruno Ferman, Ricardo Masini, Eduardo Mendes, Anna Mikusheva, Stephen Morris, Whitney Newey, Christopher Palmer, Vitor Possebom, Robert Townsend, Jaume Vives, Henry Zhang, and seminar participants at MIT and FGV-EESP for their useful comments and suggestions. Any errors are our sole responsibility. Alvarez gratefully acknowledges financial support from Fapesp grant 2022/16691-1.

[†]Post-doctoral researcher, Fundação Getúlio Vargas. *E-mail address:* luisfantozzialvarez@gmail.com

[‡]PhD candidate, Massachussets Institute of Technology. *E-mail address:* orestes@mit.edu.

1 Introduction

Consider the model:

$$Q_\mu(u) := \int Q_\theta(u)\mu(d\theta)\,, u \in [0,1]\,, \tag{1}$$

where $\{Q_\theta : \theta \in \Theta\}$ is a known family of *quantile functions* indexed by a a measurable space (Θ, Σ), and μ is an unknown (signed) measure on (Θ, Σ) such that the resulting Q_μ is a quantile function. Equation (1) defines a *nonparametric quantile mixture model.* Such models have a wide range of statistical applications. In causal inference, they can be used to assess the distributional effects of aggregate shocks, by constructing a counterfactual quantile function for an exposed unit based on the quantile function of non-treated units (*distributional synthetic controls*; Gunsilius, 2023). In financial applications, a mixture of suitable quantile basis functions can be used to extrapolate the tails of an asset return distribution (Karvanen, 2006; Gourieroux and Jasiak, 2008). Finally, and as we further argue below, nonparametric quantile mixture models can be seen as a convenient tool in the estimation of *nonparametric density mixtures*, a class of models which has received increasing attention in Econometrics and Statistics (Ignatiadis and Wager, 2022; Gu and Koenker, 2023; Armstrong et al., 2022; Kline et al., 2022).

The main tool in estimating nonparametric *density* mixture models is the nonparametric maximum likelihood estimator (NPMLE) of Kiefer and Wolfowitz (1956). In spite of its attractive statistical properties (e.g. Polyanskiy and Wu, 2020), this approach suffers from inferential and computational hurdles. The NPMLE still lacks a formal frequentist inferential theory (Ignatiadis and Wager, 2022). Computationally, the nonconvexity of the optimization program imposes challenges, which have estimulated several attempts at computing approximate solutions (Train, 2008; Koenker and Mizera, 2014; Feng and Dicker, 2018; Jagabathula et al., 2020). Motivated by these concerns, this paper aims to introduce nonparametric quantile mixture models as an attractive counterpart to density mixtures.

We begin by defining nonparametric quantile mixtures. We show that, similarly to nonparametric density mixtures, nonparametric quantile mixtures enjoy interesting approximation properties, being able to approximate sufficiently well-behaved quantile functions. We then develop a framework for conducting estimation and inference on nonparametric quantile mixture models. Our proposed estimator is a sieve-like version of the generalized method of L-moments estimator (GMLM) of Alvarez et al. (2023). Introduced by Hosking (1990), L-moments are robust alternatives to standard moments that characterize distributions with finite first moment. In our setting, the proposed estimator amounts to finding mixture weights that minimize a weighted distance between sample and theoretical L-moments. When these weights are constrained to belong to a convex set, this amounts to solving a quadratic program with convex constraints, which can be performed efficiently in standard statistical software. Another interesting feature of the GMLM is that, in parametric and some semiparametric settings, this approach to estimation has been shown to perform well in finite samples, whilst still retaining some notion of asymptotic efficiency. (Alvarez et al., 2023; Alvarez and Biderman, 2022).

Building upon our proposed estimator, we establish an inferential procedure for mixture weights and functionals thereof based on a novel bootstrap for quadratic minimizers, which may be of independent interest. Our sieve-like approach allows for the number of basis functions to be a function of the sample size. For valid inference, we rely an undersmoothing condition, whereby the number of basis functions used in the mixture is sufficiently large so as to control approximation bias. Alternatively, if one is willing to place restrictions on the true quantile process, we may replace the undersmoothing condition with bias-aware inference, whereby bounds on the approximation bias are computed to conduct conservative inference (Armstrong and Kolesár, 2021; Noack and Rothe, 2021; Ignatiadis and Wager, 2022). We also note that our approach to inference explicitly allows for regularization, which may be preferable to *ad-hoc* selection methods (Masini, 2022).

Inference in our setting is challenging, as it is generally not possible to find explicit limit distributions for estimators when the number of parameters diverges (Chernozhukov et al., 2017a) and existing bootstrap methods are not applicable when we might have solutions at the boundary of the parameter set Θ (see Chernozhukov et al. (2023) for a recent survey). Our analysis addresses these challenges by relying upon a novel strategy that combines the convexity arguments originally applied to (low-dimensional) M-estimators with convex objective functions (Pollard, 1991; Kato, 2009), with an approach to inference based on the concept of *strong approximation*. Strong approximations have been increasingly used to conduct inference in nonstandard or high-dimensional settings (see Chernozhukov et al., 2014; Armstrong and Kolesár, 2017; Cattaneo et al., 2020; Fang et al., 2023; Chernozhukov et al., 2023, for some recent examples).

Key to our strategy is a novel result that bounds the strong approximation of quadratic minimizers in terms of the constituent elements of the program, along with a restricted eigenvalue condition. Building upon a lemma in Fan et al. (2022), we show that, when combined with anticoncentration inequalities, our results enable us to derive high-level sufficient conditions that ensure asymptotic validity of our proposed inferential approach. Whilst we apply these tools to inference in quantile mixture models, we highlight that they may have potential applications on other problems of interest such as LASSO and Ridge regression.

We then show how the analysis can be specialized under additional assumptions on the smoothness of the quantile functions, thus providing sufficient rates for inference in specific applications. For example, in an unconstrained estimation setting, and under some assumptions, our results show that the number of mixture weights can grow much faster than the sample size. When considering the case of ridge regularization, we show that the number of mixtures used (p) and the sample size (n) must satisfy, up to logarithmic factors, $p/n \to 0$ a rate similar to the one obtained by Belloni et al. (2015) in an (unconstrained) sieve-regression setting. In order to demonstrate the latter result, we develop an extension of the anticoncentration inequalities for Gaussian Random variables available in the literature to the projection of Gaussian random variables onto Euclidean balls, another result of independent interest.

We then consider two applications of our quantile mixtures. First, we show how they can be used to recover estimates (and conduct inference) in nonparametric **density** mixture models. Key to this approach is the relation between the derivative of a quantile function and the density of a random variable, which we use to "invert" our quantile mixtures onto density mixtures.

As a second application, we show that, as a direct byproduct of our theory, we are able to provide a valid inferential method in the distributional synthetic control setup of Gunsilius (2023). The crucial condition to the validity of our approach in this environment is a weight "dilution" condition, which requires the oracle weights attached to each control to spread sufficiently quickly. A similar condition exists in the literature on synthetic controls (Ferman, 2021). To the best of our knowledge, formal inference methods were not yet available in the distributional synthetic control setting.[1]

Finally, to illustrate the usefulness of our method, we provide an empirical application on the distributive effects of a large-scale environmental disaster, the Brumadinho Barrage rupture in Minas Gerais, Brazil, on the local wage distribution. The impact of this the disarter is an open question as its effects remain uncertain, with potential negative effects mitigated by reparations, investments, and economic recovery efforts. The main empirical challenge lies in the fact that the rupture affected both the city of Brumadinho and neighboring municipalities simultaneously. This simultaneous shock hinders the availability of a clear control group, making direct comparisons difficult and motivating the use of the distributional synthetic control method, especially since there is a large pool of potential control units (other non-affected municipalities). Our results uncover a range of effects across percentiles of the wage distribution, which we argue are consistent with *displacement effects*, whereby median-earning jobs are replaced by low-paying contracts.

[1] Gunsilius (2023) suggests using placebos to assess uncertainty in his setup, though he does not provide a formal justification to it.

Overview The remainder of the paper is organized as follows. In Section 2 we define nonparametric quantile mixtures and discuss its approximation properties, as well as its connection with the distributional synthetic controls of Gunsilius (2023). Section 3 introduces our estimation procedure. The asymptotic inferential theory for mixture weights and linear functionals thereof is presented in Section 4. Section 5 presents the algorithm to implement our bootstrap procedure. In Section 6, we discuss applications to the empirical Bayes and distributional synthetic control problems. Section 7 presents the results of our empirical application. Section 8 concludes. The Appendices contain the proofs of the main results, as well as our lemma on the strong approximation of a quadratic minimizer and the anti-concentration inequality of the projection of Gaussian random variables onto Euclidean balls. We also discuss the choice of tuning parameters in our bootstrap procedure, present additional results on the approximation properties of quantile mixtures and discuss further details on the empirical application.

2 Quantile mixture models

In this section, we define a quantile mixture and discuss its approximation properties. We begin by recalling the definition of a quantile function.

Definition 1. A function $Q : [0, 1] \mapsto \mathbb{R} \cup \{-\infty, \infty\}$ is a *quantile function* if it is nondecreasing and continuous on the left.

As it is well known, given a distribution function $F : \mathbb{R} \mapsto [0, 1]$, the *generalized inverse*

$$Q_F(u) \coloneqq \inf\{x \in \mathbb{R} : F(x) \geq u\}\,, \quad u \in [0, 1]\,.$$

defines a quantile function.

We are now ready to define a quantile mixture.

Definition 2. Given a family of *quantile functions* $\{Q_\theta : \theta \in \Theta\}$ indexed by a a measurable space (Θ, Σ), and a signed measure μ on (Θ, Σ), the map:

$$Q_\mu(u) \coloneqq \int Q_\theta(u)\mu(d\theta)\,, u \in [0,1]\,,$$

defines a *quantile mixture* if $Q_\mu(u)$ exists as an extended real number for every $u \in [0,1]$ and the resulting Q_μ is a quantile function.

A quantile mixture combines a family of quantile functions through a signed measure μ. Notice that our definition imposes that the resulting mixture is itself a quantile function, which is a desirable feature in our main applications. We also remark that, if μ is a *measure*, then any well-defined Q_μ is necessarily a quantile mixture.

Definition 3. A quantile mixture model is a pair $(\mathcal{G}, \mathcal{M})$, where $\mathcal{G}$ is a family of quantile functions indexed by a measurable space (Θ, Σ), and $\mathcal{M}$ is a subset of

$$\{\mu \text{ is a signed measure on } (\Theta, \Sigma) : Q_\mu \text{ is a quantile function}\}\,.$$

It is well known that some density mixture models enjoy great approximation properties, being able to approximate quite general classes of densities with arbitrary error under specific norms (see Nguyen and McLachlan (2019) and Nguyen et al. (2020) for results on mixtures of densities from location-scale families). We conclude this section by providing examples of quantile mixture models that similarly exhibit interesting approximation properties.

Example 1 (Polynomial quantiles)**.** For $n \in \mathbb{N}$, define the vector of polynomials:

$$J_n(u) \coloneqq \begin{pmatrix} 1 \\ u \\ u^2 \\ \vdots \\ u^n \end{pmatrix}.$$

It follows from monotone approximation theory (Shvedov, 1981) that, for any $p \in [1, \infty]$ and quantile function $Q \in L^p[0,1]$, and for every $n \in \mathbb{N}$, there exists $\theta_n^* \in \mathbb{R}^n$ such that $\theta_n^{*\prime} J_n$ is a quantile function and:

$$\|Q(\cdot) - \theta_n^{*\prime} J_n(\cdot)\|_{L_p[0,1]} \leq C\omega_{2,p}(Q, n^{-1}) \,,$$

where $C > 0$ is an absolute constant and $\omega_{2,p}(Q,\delta) = \sup_{0<h\leq\delta} \|Q(\cdot + 2\delta) - Q(\cdot + \delta) - (Q(\cdot + \delta) - Q(\cdot))\|_{L_p[0,1]}$ is the second modulus of continuity. If Q is absolutely continuous, with density $q \in L^p[0,1]$, then we can further show that (DeVore and Lorentz, 1993, page 44):

$$\|Q(\cdot) - \theta_n^{*\prime} J_n(\cdot)\|_{L_p[0,1]} \leq Cn^{-1}\omega_p(q, n^{-1}) \,,$$

where $\omega_p(q,\delta) = \sup_{0<h\leq\delta} \|q(\cdot + \delta) - q(\cdot)\|_{L_p[0,1]}$ is the (first) modulus of continuity.

Example 2 (Pareto mixtures)**.** Consider the class of generalized Pareto distributions with shape parameter $k \in \mathbb{R}$. In this case, the associated quantile functions are given by (Hosking and Wallis, 1987):

$$Q_k(u) = \begin{cases} \frac{(1-(1-u)^k)}{k}, & \text{if } k \neq 0 \\ -\log(1-u), & \text{if } k = 0 \end{cases} .$$

Since $\log(1-x)$ is analytic on $[0,1]$, we are able to show that the class of quantile mixtures on the Pareto family is able to reproduce any polynomial on $[0,1]$. It then follows from Example 1 that, for a given $p \in [1, \infty]$ and every quantile function $Q \in L^p[0,1]$, and for every $\epsilon > 0$, there exists a quantile mixture Q_{μ^*} on the family of Pareto distributions such that:

$$\|Q(\cdot) - Q_{\mu^*}(\cdot)\|_{L^p[0,1]} \leq C\omega_{2,p}(Q, \epsilon) \,.$$

Example 3 (Extreme Value mixtures)**.** Consider the class of generalized extreme value distributions with shape parameter $k \in \mathbb{R}$. In this case, the associated quantile functions are given by (Hosking et al., 1985):

$$Q_k(u) = \begin{cases} \frac{(1-(-\log u)^k)}{k}, & \text{if } k \neq 0 \\ -\log(-\log u), & \text{if } k = 0 \end{cases}.$$

Notice that mixtures on this class are able to reproduce any polynomial of $-\log(u)$. Consequently, for a given $p \in [1, \infty]$, every $\underline{u} \in (0,1)$ and quantile $Q \in L^p[0,1]$, we have that, for every $\epsilon > 0$, there exists a quantile mixture Q_{μ^*} such that:

$$\|Q(\cdot) - Q_{\mu^*}(\cdot)\|_{L^p[\underline{u},1]} \leq C\omega_{2,p}(Q, \epsilon)\,.$$

By assuming that the lower tail of Q asymptotically behaves as a member of the extreme value family, we can then extend this approximation to the entire interval $[0,1]$.

Example 4 (Approximation through nonnegative measures)**.** The previous examples rely on possibly signed measures to construct quantile mixtures that arbitrarily approximate a target quantile function. One question is whether similar approximations can be constructed by relying solely on nonnegative measures. In Appendix D, we construct an example of a quantile mixture model that achieves general approximation properties under nonnegative weighting under additional assumptions on the quantile function.

Example 5 (Distributional synthetic controls)**.** Consider a population of $J+1$ units. Let $Q_{j,t}$ denote the quantile function of the distribution of an outcome of interest in unit j at period t. For example, $Q_{j,t}$ may denote the quantile function of the wage distribution in state j at year t, with $(J+1)$ being the number of states in the country of analysis. Suppose that a policy is implemented at state $j=0$ beginning at period t^*. Gunsilius (2023) provides conditions under which, if there exists a set of weights $(w_j)_{j=1}^J \in \Delta^{J-1}$ such that:

$$Q_{0,t^*-1} = \sum_{j=1}^{J} w_j Q_{j,t^*-1} \, ,$$

then the quantity,

$$\tilde{Q}_{0t} \coloneqq \sum_{j=1}^{J} w_j Q_{jt}, \quad t \geq t^* \, ,$$

yields a valid counterfactual for the quantile function at $j = 0$ in the absence of treatment. Such *distributional synthetic controls* may be used to assess the effect of policies on the entire distribution of an outcome of interest.

3 Proposed estimator

Observation: In what follows, all random variables are defined in the same probability space $(S, \mathcal{S}, \mathbb{P})$.

In this section, we introduce an estimation procedure for quantile mixture models. Specifically, given a sample estimator $\hat{Q}_n$ of a target quantile function Q, we propose to estimate mixture weights by solving:

$$\hat{\theta}_n \in \operatorname{arginf}_{\theta \in \Theta_p} \left(\int_{\underline{p}_n}^{\overline{p}_n} (\hat{Q}_n(u) - \theta' \boldsymbol{J}_{p,n}(u)) \boldsymbol{P}_L(u) du \right)' W_{L,n} \left(\int_{\underline{p}_n}^{\overline{p}_n} (\hat{Q}_n(u) - \theta' \boldsymbol{J}_{p,n}(u)) \boldsymbol{P}_L(u) du \right) , \tag{2}$$

where $\boldsymbol{J}_{p,n}(u) = (J_{1,n}(u), J_{2,n}(u), \ldots J_{p,n}(u))'$ is a vector of p quantile functions; Θ_p is a convex subset of $\mathbb{R}^p$; $\boldsymbol{P}_L(u) = (P_1(u), P_2(u), \ldots, P_L(u))'$ is a vector of L quantile weighting functions, with $\{P_l\}_{l \in \mathbb{N}}$ forming an orthonormal basis on $L^2[0,1]$; $W_{L,n}$ is an $L \times L$ symmetric positive semidefinite weighting matrix; and $0 \leq \underline{p}_n < \overline{p}_n \leq 1$ are trimming constants. Estimator (2) is a sieve-like version of the generalized method of L-moments (GMLM) estimator in Alvarez et al. (2023). Introduced by Hosking (1990), the r-th L-moment of a distribution function F is defined as $\lambda_r := \int_0^1 Q_F(u) P^*_{r-1}(u) du$, with P^*_l being the l-th shifted Legendre polynomial on $[0,1]$. L-moments provide robust alternatives to standard moments; they also characterize any distribution function with finite first moment. In a parametric setting, Alvarez et al. (2023) show that, under some conditions, an optimally-weighted GMLM estimator is asymptotically efficient as the sample size and the number of L-moments used in estimation (L) diverge. Moreover, by properly choosing L as a function of the sample size, they show that it is possible to improve (in a mean-squared error sense) over maximum likelihood estimation in finite samples.[2]

In light of its attractive statistical properties, we propose to use a GMLM approach in the estimation of quantile mixture models. We also note that, in our setting, such approach is highly advantageous from the computational viewpoint. Indeed, (2) is a quadratic program with convex constraints, which can be solved efficiently. Possible choices of Θ_p include:

1. **unconstrained mixtures:** $\Theta_p = \mathbb{R}^p$;
2. **ridge regularization:** $\Theta_p = B_{(\mathbb{R}^p, \|\cdot\|_2)}(0, M)$;
3. **lasso regularization:** $\Theta_p = B_{(\mathbb{R}^p, \|\cdot\|_1)}(0, M)$;
4. **nonnegative weights:** $\Theta_p = \mathbb{R}^p_+$; and
5. **simplex weights:** $\Theta_p = \Delta^{p-1}$.

In our setup, the quantile "series" functions $\boldsymbol{J}_{p,n}$ can be either nonstochastic, as in usual series estimation, or stochastic, as in distributional-synthetic-control-type applications. The weighting matrix $W_{L,n}$ is possibly estimated – we discuss possible choices of weights later on. Finally, we allow for the possibility of trimming, by specifying constants $0 \leq \underline{p}_n < \overline{p}_n \leq 1$, as this may be useful in applications with heavy-tailed data and in some extrapolation exercises.

In the next sections, we provide a valid inferential theory on the estimands $\theta_{0,n}$ of $\hat{\theta}_n$, and functionals thereof, as n and (possibly) p and L diverge. Our theory is purposefully generic,

[2]This approach to estimation has also been shown to produce efficient estimators in a class of semiparametric models of treatment effects (see Alvarez and Biderman, 2022).

in order to accommodate the different types of applications we have in mind. Conceptually, we rely on strong approximations of the quantile function $\hat{Q}_n$ to a Gaussian process in order to approximate the distribution function of $\hat{\theta}_n$ with the distribution of a minimiser of (2) where $\hat{Q}_n$ is replaced by a Gaussian random variable. Strong approximations of the empirical quantiles of a scalar sample with common marginal distribution have been derived in the iid setting by Csorgo and Revesz (1978), and extended to the stationary mixing setting by Fotopoulos and Ahn (1994) and Yoshihara (1995). We speculate that these results may be combined with statistical learning techniques, such as sample splitting and debiasing (Chernozhukov et al., 2018), to construct valid strong approximations to more complex quantile estimators, such as the empirical quantiles obtained from the prediction errors of a first-step algorithm, which may be useful in risk management applications (see Section 8 for further discussion). While our theory is ample enough to accomodate any such approximation, we do not pursue the construction of this type of coupling in our paper, as we do not require it in our main applications.

4 Inferential approximation

4.1 Inference on mixture weights

We begin by presenting general inferential results on the distributional approximation of our GMLM estimator. We then specialize to specific bases and choice sets. We implicitly index p and L by n, thus allowing these quantities to diverge with n, and consider limits as $n \to \infty$. In what follows, we work under the high-level assumption:

Assumption 1 (Existence of tight Gaussian Approximation)**.** *There exists a sequence of zero-mean Gaussian processes, $B_{0,n}$, $n \in \mathbb{N}$, defined on $(S, \mathcal{S}, \mathbb{P})$ and indexed by $[0,1]$, such that, as $n \to \infty$:*

$$\|\sqrt{n}(\hat{Q}_n - Q) - B_{0,n}\|_{L^2[\underline{p}_n, \bar{p}_n]} = O_{\mathbb{P}}(r_n)\,. \tag{3}$$

Moreover, this sequence of Gaussian processes is tight in $L^2_{[\underline{p}_n,\overline{p}_n]}$, in the sense that:

$$\|B_{0,n}\|_{L^2_{[\underline{p},\overline{p}]}} = O_{\mathbb{P}}(1)\,.$$

As briefly mentioned in the previous section, Csorgo and Revesz (1978) derived strong approximation results in the case where $\hat{Q}_n$ are empirical quantiles from a random sample with quantile function Q. In this case, under the assumptions of their Theorem 6, one may take $0 = \underline{p}_n < \overline{p}_n = 1$ and have $r_n = n^{-1/2}\log(n)^{\alpha}$ for some $\alpha > 0$.[3] In this case, the Gaussian processes $B_{0,n}$ are zero-mean with common covariance kernel $\Gamma_0(i,j) = Q'(i)Q'(j)(i \wedge j - ij)$, which shows the sequence is tight in $L^2_{[\underline{p}_n,\overline{p}_n]}$.

We also need to control the possible estimation error of the weighting matrix $W_{n,L}$.

Assumption 2 (Estimation error of weighting matrices)**.** *There exists a sequence of non-stochastic positive semidefinite (psd) matrices $\Omega_{n,L}$, $n \in \mathbb{N}$, such that, as $n \to \infty$, $\|\Omega_{n,L}\|_2 = O(1)$ and $\|W_{n,L} - \Omega_{n,L}\| = O_{\mathbb{P}}(s_n)$.*

We split our analysis into two cases: (i) nonstochastic basis; and (ii) stochastic basis.

4.1.1 Nonstochastic basis

In this section, the basis functions $\boldsymbol{J}_{n,p}$ are taken to be nonstochastic. Our target estimand is given by:

$$\theta_{0,n} \in \operatorname{arginf}_{\theta\in\Theta_n} \|Q(\cdot) - \theta'\boldsymbol{J}_{n,p}(\cdot)\|_{L^2[\underline{p}_n,\overline{p}_n]}\,, \tag{4}$$

and we note that we can write:

$$\sqrt{n}(\hat{\theta}_n - \theta_n) \in \operatorname{arginf}_{x\in\mathcal{X}_n} \left\| \int_{\underline{p}_n}^{\overline{p}_n} (\sqrt{n}(\hat{Q}_n(u) - Q(u)) - x'\boldsymbol{J}_{p,n}(u) + D_n(u))\boldsymbol{P}_L(u)du \right\|^2_{2,W_{L,n}}\,, \tag{5}$$

[3] The authors also have weaker results for trimming constants $\underline{p}_n$ and $\overline{p}_n$ that converge, respectively, to 0 and 1 at a rate. These may be used in the sample-size-dependent trimming case.

where, for a symmetric psd matrix A, $\|x\|_{A,2} = \sqrt{x'Ax}$, $\mathcal{X}_n = \sqrt{n}(\Theta_n - \theta_{0,n})$, and $D_n(u) = \sqrt{n}(Q(u) - \theta'_{0,n}\boldsymbol{J}_{n,p}(u))$ is the approximation bias. In light of representation (5) and Assumptions 1 and 2, we are led to consider the following approximation:

$$\sqrt{n}(\theta^*_n - \theta_{0,n}) \in \text{arginf}_{x\in\mathcal{X}_n} \left\| \int_{\underline{p}_n}^{\overline{p}_n} (B_{0,n}(u) - x'\boldsymbol{J}_{p,n}(u) + D_n(u))\boldsymbol{P}_L(u)du \right\|^2_{2,\Omega_{L,n}} . \tag{6}$$

As we show later on, representation (6) may be used as a basis for inference, under some assumptions. The bias term $D_n(u)$, even though unknown, may be bounded by using approximation results such as Example 1 or identification assumptions such as in Example 5.[4] Alternatively, one may consider an "undersmoothed" approximation:

$$\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n}) \in \text{arginf}_{x\in\mathcal{X}_n} \left\| \int_{\underline{p}_n}^{\overline{p}_n} (B_{0,n}(u) - x'\boldsymbol{J}_{p,n}(u))\boldsymbol{P}_L(u)du \right\|^2_{2,\Omega_{L,n}} . \tag{7}$$

Proposition 1. *Suppose that $d_n := \|D_n\|_{L^2[\underline{p}_n,\overline{p}_n]}$ is bounded. For $z \in \mathbb{R}^p$, define the restricted eigenvalue around z as :*

$$\lambda_{z,n} := \inf_{s\in\mathcal{X}_n} \frac{(s-z)'\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)'\Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)(s-z)}{\|s-z\|_2^2} . \tag{8}$$

Finally, let:

$$\rho_n := \lambda_{\max}\left(\int_{\underline{p}_n}^{\overline{p}_n} (\boldsymbol{J}_{n,p}(u)\boldsymbol{J}_{n,p}(u)')du\right) .$$

We then have that:

1. *Smoothened case: for any sequences $(\delta_n)_{n\in\mathbb{N}}$, $(M_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ such that:*

[4] See Remark 4 below.

$$
\begin{aligned}
M_n \frac{\lambda_{0,n}}{\sqrt{\rho_n}} &\to \infty\,, \\
\mathbb{P}[\lambda_{\sqrt{n}(\theta_n^* - \theta_{0,n}),n} \leq c_n] &\to 0\,, \\
\frac{c_n \delta_n^2}{(\delta_n + M_n)\sqrt{\rho_n}(r_n \vee s_n)} &\to \infty\,, \\
\frac{c_n \delta_n^2}{(\delta_n + M_n)^2 \rho_n s_n} &\to \infty\,,
\end{aligned}
\tag{9}
$$

we have that

$$
\mathbb{P}[\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\theta_n^* - \theta_{0,n})\| \geq \delta_n] \to 0\,.
$$

2. *Undersmoothened case:for any sequences* $(\delta_n)_{n\in\mathbb{N}}$*,* $(M_n)_{n\in\mathbb{N}}$ *and* $(c_n)_{n\in\mathbb{N}}$ *such that:*

$$
\begin{aligned}
M_n \frac{\lambda_{0,n}}{\sqrt{\rho_n}} &\to \infty\,, \\
\mathbb{P}[\lambda_{\sqrt{n}(\theta_n^* - \theta_{0,n}),n} \leq c_n] &\to 0\,, \\
\frac{c_n \delta_n^2}{(\delta_n + M_n)\sqrt{\rho_n}(r_n \vee d_n \vee s_n)} &\to \infty\,, \\
\frac{c_n \delta_n^2}{(\delta_n + M_n)^2 \rho_n s_n} &\to \infty\,,
\end{aligned}
\tag{10}
$$

we have that

$$
\mathbb{P}[\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n})\| \geq \delta_n] \to 0\,.
$$

Proof. See Appendix A.2. □

Proposition 1 provides rates of the approximation of (5) in terms of (6) (or (7)). It expresses these rates in terms of control of the norm of the approximating solution, as well as control of the restricted eigenvalues of the approximating program.[5] The proof of Proposition 1 is deferred to Appendix A, where it follows from a lemma on the strong approximation of a minimizer of a quadratic program, which may be of independent interest.

In the next subsections, we will combine Proposition 1 with Lemma 1 below to provide

[5]Restricted eigenvalue conditions are common in the literature on high dimensional linear models (e.g. Bickel et al., 2009).

valid inferential tools on mixture weights. This lemma bounds the Kolmogorov distance (in a subset of $\mathcal{B}(\mathbb{R}^p)$) between two random variables taking values in $\mathbb{R}^p$, in terms of two components: *strong approximation* and *anticoncentration*.

Lemma 1. *Let X and Y be two random variables on $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), \mathbb{P})$. Then, for any $\mathcal{C} \subseteq \mathcal{B}(\mathbb{R}^p)$, we have that:*

$$\sup_{A\in\mathcal{C}} |\mathbb{P}[X \in A] - \mathbb{P}[Y \in A]| \leq \inf_{s\in[2,\infty]} \inf_{\delta\geq 0} \{\mathbb{P}[\|X - Y\|_2 \geq \delta] + \Psi_s(\mathcal{C};\delta)\} \,,$$

where $\Psi_s(\mathcal{C};\delta) := \sup_{A\in\mathcal{C}} \mathbb{P}[Y \in A_s^\delta \setminus A_s^{-\delta}]$, where $A_s^\delta = \{x \in \mathbb{R}^p : \inf_{a\in A}\|x - a\|_s \leq \delta\}$ and $A_s^{-\delta} = \mathbb{R}^p \setminus (\mathbb{R}^p \setminus A)_s^\delta$.

Proof. The result follows from Lemma S.14 in Fan et al. (2022), by observing that, for $x \in \mathbb{R}^p$, $\|x\|_s \leq \|x\|_2$ for any $s \in [2, \infty]$, and optimizing. □

By combining the above lemma with anticoncentration results on Gaussian processes and Proposition 1, we will be able to find sequences δ_n such that the Kolmogorov distance between the statistic and its approximation converges to zero. This is precisely what we need to achieve a valid distributional approximation.

Unconstrained case We begin by considering the unconstrained case $\Theta_p = \mathbb{R}^p$. In this case, the estimator admits a closed form solution given by:

$$\hat{\theta}_n = \begin{array}{c}\left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right)^{-1} \times \\ \left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L} \int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\hat{Q}_n(u)du\right)\end{array}, \tag{11}$$

provided that the inverse exists. Similarly, for the undersmoothed approximation:

$$\tilde{\theta}^*_n = \begin{array}{l}\left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right)^{-1} \times \\ \left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\frac{B_{n,0}(u)}{\sqrt{n}}du\right)\end{array}, \tag{12}$$

In this case, the restricted eigenvalue collapses to the smallest eigenvalue, and we are able to show that:

Proposition 2. *Suppose that for constants $C > 0$, $\gamma \in [0,1)$, the smallest eigenvalue of $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)$, $\lambda_{n,\min}$ satisfies, $\lambda_{n,\min} > C(\rho_n s_n)^\gamma$. Then, if $\rho_n s_n \to 0$, $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)$ is invertible with probability approaching one and:*

$$\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n} - \boldsymbol{D}_n) - \sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})\| = O_{\mathbb{P}}((\rho_n s_n)^{1-2\gamma} \vee ((\rho_n s_n)^{-\gamma}\sqrt{\rho_n} r_n)),$$

where the bias term is given by:

$$\boldsymbol{D}_n = \begin{array}{l}\left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right)^{-1} \times \\ \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)(Q(u) - \theta'_{0,n}\boldsymbol{J}_{n,p}(u))du\right)\end{array}.$$

Proof. See Appendix A.3. ☐

The previous proposition can be combined with anticoncentration results available in the literature to provide valid approximation in specific classes of subsets of $\mathcal{B}(\mathbb{R}^n)$. To illustrate, we consider the class of hyperrectangles, $\mathcal{C}_p = \{[\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a}, \boldsymbol{b} \in \overline{\mathbb{R}}^p, \boldsymbol{a} \leq \boldsymbol{b}\}$. In this case, combining the previous proposition with Nazarov's inequality (Chernozhukov et al., 2017b, Theorem 1), we obtain the following conclusion.

Corollary 1. *Suppose that the conditions in Proposition 2 hold. Suppose that the variance* *of* $\sqrt{n}(\tilde{\theta}^*_{j,n} - \theta_{j,0,n})$, $j = 1, \ldots, p$ *is bounded below uniformly in* j *and* n*. If, for some* $\nu > \frac{1}{2}$*:*

$$\log(p)^{\nu}\left((\rho_n s_n)^{1-2\gamma} \vee ((\rho_n s_n)^{-\gamma}\sqrt{\rho_n} r_n)\right) \to 0$$

then

$$\sup_{A \in \mathcal{C}_p} |\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n} - \boldsymbol{D}_n) \in A] - \mathbb{P}[\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n}) \in A]| \to 0\,.$$

Proof. Nazarov's inequality applied to the random vector $\boldsymbol{v} = (\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})', -\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})')'$ implies that $\Psi_\infty(\mathcal{C}_p; \delta) \leq \frac{\delta}{\underline{\sigma}}(\sqrt{2\log(2p)} + 2)$, where $\underline{\sigma}^2$ is the lower bound on the variance of the approximation. The conclusion then follows from Proposition 2 and Lemma 1.

□

Similar results can be obtained for other subclasses of sets, by relying on different anti-concentration results available in the literature. For example, if one wishes to provide valid approximations on the class of Euclidean balls of arbitrary center or convex subsets of $\mathbb{R}^p$; one could rely, respectively, on the anticoncentration results of Götze et al. (2019, Theorem 2.7) or Chernozhukov et al. (2017a, Lemma A.2).

Remark 1 (Restrictions on L)**.** Notice that our rates never depend on the number of L-moments L. This is similar to the results in Alvarez et al. (2023) in the parametric setting, where L may increase arbitrarily with n. It should be noted, however, that the assumption of the smallest eigenvalue of the matrix in Proposition 2 being positive requires at least $L \geq p$.

Remark 2 (Specialization to polynomial quantiles)**.** It is instructive to consider polynomial basis functions, i.e. $\boldsymbol{J}_{n,p}(u) = (1, u, \ldots, u^{p-1})$. In this case, results on the Hilbert matrix (Taussky, 1949) reveal that $\rho_n = \pi(1 + O(1/\log(n)))$. Moreover, if we consider an identity weighting matrix, i.e. $W_{n,L} = \mathbb{I}_L$, then we may take $s_n = 0$. In addition, if we assume, similarly to the least squares series regression setup (Newey, 1997; Belloni et al., 2015), that the smallest eigenvalue of the population matrix in the statement of Proposition 2 is bounded away from zero uniformly, then $\gamma = 0$. Finally, if $\hat{Q}_n(u)$ are the empirical quantiles from a random sample of size n from a distribution satisfying the assumptions in Theorem 6 of Csorgo and Revesz (1978), then $r_n = n^{-1/2}\log(n)^{\alpha}$ for some $\alpha > 0$ and we obtain that the approximation on the class of semi-intervals in Corollary 1 holds as soon as:

$$\frac{\log(p)^{\nu}\log(n)^{\alpha}}{n^{1/2}} \to 0\,.$$

Remark 3 (Undersmoothing)**.** The inferential approximation (12) does not account for the bias term $\boldsymbol{D}_n$. Observe that, in the setting of Corollary 1, this bias may be ignored in inference under the undersmoothing condition.

$$\log(p)^{\nu}\sqrt{n}\|\boldsymbol{D}_n\| \to 0\,.$$

To understand this condition, consider, again, the polynomial basis setup of the previous remark. In this case, Example 1 and Bessel's inequality (Kreyszig, 1978, Theorem 3.2.8) reveal that:

$$\|\boldsymbol{D}_n\| \leq \tilde{C}\omega_{2,2}(Q, p^{-1})\,.$$

If we assume that Q belongs to a generalized Lipschitz space (see Section 2.9 of DeVore and Lorentz (1993) for a definition), then $\omega_{2,2}(Q,t) \leq Dt^b$ for $D, b > 0$. In this case, the undersmoothing condition subsumes to:

$$\log(p)^{\nu}\frac{\sqrt{n}}{p^b} \to 0\,.$$

Remark 4 (Bias-aware inference)**.** If one is not willing to impose the undersmoothing conditions in the previous remark, it is still possible to conduct conservative inference on $\theta_{0,n}$ by bounding the bias. Specifically, Approximation Theory provides the value for the absolute constant in Example 1. Then, by placing a plausible upper bound on the generalized Lipschitz norm of Q, we may use Example 1 along with representation (12) to adjust the critical values for coverage even in the worst-case setting. Such a procedure is known as "bias-aware" inference and has received increasing attention from the Econometrics and Statistics literature (Armstrong and Kolesár, 2021; Noack and Rothe, 2021; Ignatiadis and Wager, 2022). Though we recognize the possibility of conducting inference this way, we do not directly pursue it in this paper.

Remark 5 (Optimal weighting-matrix)**.** From the closed form solution of the approximation, it is clear that the optimal choice of weights is given by:

$$\Omega_{n,L} = \mathbb{V}\left[\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u) B_{n,0}(u) du\right)\right]^{-},$$

where A^{-} denotes the generalized inverse of a matrix. As in two-step GMM, this matrix may be estimated in a first-step by first running the GMLM with identity weights and then using it to estimate $\Omega_{n,L}$. Under mild assumptions, this should not affect the overall rate of estimation.

Ridge regularization We now consider the case of ridge regularization, i.e. $\Theta_p = B_{\mathbb{R}^p, R_n}(0)$. In this case, by relying on Proposition 1, we are able to show that, under the assumptions of Proposition 2, the L-moment estimator with $\Theta_p = B_{\mathbb{R}^p, R_n}(0)$ achieves the same rates as the unconstrained estimator.

To provide valid inference in this setting, one should then combine these rates with anticoncentration results on the approximation (7). Notice that, under ridge regularization, (7) is the projection of a Gaussian random vector in the $\|\cdot\|_{\Omega_n}$-norm onto an Eucliden ball. As far as we are aware, anticoncentration results on projection of Gaussian vectors onto Euclidean balls are not available in the literature. In Appendix B, we provide an approach to extend results on Gaussian anticoncentration available in the literature to the ball projection setting. We believe this approach may be of independent interest.

Combining these results, we have the following proposition.

Proposition 3. *Consider the setting with ridge constraint. Suppose that for constants* $C > 0$, $\gamma \in [0,1)$, *the smallest eigenvalue of* $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)$, $\lambda_{n,\min}$, *satisfies* $\lambda_{n,\min} > C(\rho_n s_n)^{\gamma}$.

Then, if $\rho_n s_n \to 0$, $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)$ *is invertible with probability approaching one and:*

$$\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\theta_n^* - \theta_{0,n})\| = O_{\mathbb{P}}((\rho_n s_n)^{1-2\gamma} \vee ((\rho_n s_n)^{-\gamma}\sqrt{\rho_n} r_n))\,,$$

$$\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n})\| = O_{\mathbb{P}}((\rho_n s_n)^{1-2\gamma} \vee ((\rho_n s_n)^{-\gamma}\sqrt{\rho_n}(r_n \vee d_n)))\,.$$

Moreover, suppose that the variance of the Gaussian approximation in the unconstrained setting is bounded below uniformly. Suppose that $R_n = Gp^{-l}$ *for some* $l \in [0, 1/4)$. *If, for some* $\nu > 1/2$

$$\log(p)^{\nu}(\rho_n s_n)^{-\gamma} p^{1/2+l}\left((\rho_n s_n)^{1-2\gamma} \vee ((\rho_n s_n)^{-\gamma}\sqrt{\rho_n} r_n)\right) \to 0\,,$$

$$n(\rho_n s_n)^{\gamma} \to \infty,$$

then:

$$\sup_{A \in \mathcal{C}_p} |\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) \in A] - \mathbb{P}[\sqrt{n}(\theta_n^* - \theta_{0,n}) \in A]| \to 0\,.$$

In addition, if the undersmoothing condition $\log(p)^{\nu}(\rho_n s_n)^{-2\gamma} p^{1/2+l} d_n \to 0$ *holds, then the approximation is also valid for the undersmoothened approximation.*

Proof. See Appendix A.4. □

Remark 6 (Polynomial basis, cont.)**.** Consider the setting of polynomial basis with identity weights, population matrix with eigenvalues bounded away uniformly from zero, and empirical quantiles from a random sample satisfying the Csorgo and Revesz (1978) assumptions. Suppose the radius R_n is kept fixed at $\bar{R}$. In this case, the rate requirement in the statement of the proposition subsumes to:

$$\log(p)^{\nu}\log(n)^{\alpha}\sqrt{\frac{p}{n}} \to 0\,,$$

which, up to logs, is the same rate obtained by Belloni et al. (2015) in an (unconstrained) sieve-regression setting. For Q in a generalized Lipschitz class, the undersmoothing condition is given by:

$$\log(p)^{\nu}\sqrt{n}p^{1/2-b} \to 0\,,$$

thus requiring a degree of smoothness $b > 1$ for proper bias control.

Remark 7 (Relaxing regularization)**.** In the statement of Proposition 3, we have considered either a constant or decreasing radius R_n. It is also possible to consider increasing radii, though in this case one should conduct a careful analysis of the terms in Lemma 3 in the Appendix to understand the "region" where the anticoncentration "inhabits". Since constant radius are a typical choice in constrained least squares estimation (Chernozhukov et al., 2021); and decreasing radii may be required in some applications, we focus on these cases.

4.1.2 Stochastic basis

In this section, we consider the case of stochastic basis, i.e. $\boldsymbol{J}_{p,n}$ is a random variable. In this case, our target estimand is given by:

$$\theta_{0,n} \in \mathrm{arginf}_{\theta\in\Theta_n}\|Q(\cdot) - \theta'\boldsymbol{J}^*_{n,p}(\cdot)\|_{L^2[\underline{p}_n,\overline{p}_n]}\,, \tag{13}$$

where $\boldsymbol{J}^*_{n,p}$ is the estimand of $\boldsymbol{J}_{n,p}$. We propose to conduct inference conditionally on the controls $\boldsymbol{J}_{n,p}$. For that, we will require $\boldsymbol{J}_{n,p}$ to be independent of $\hat{Q}_n$, and we also need to control the estimation error in $\boldsymbol{J}_{n,p}$. These conditions are subsumed in the following assumption.

Assumption 3 (Estimation error of basis functions)**.** *$\boldsymbol{J}_{n,p}$ is a random vector, with $\hat{Q}_n$ independent of $\boldsymbol{J}_{n,p}$. Moreover, there exists a sequence of nonstochastic quantile functions $\boldsymbol{J}^*_{p,n}$, $n \in \mathbb{N}$, such that, as $n \to \infty$,*

1. $\left\|\theta'_{0,n}\sqrt{n}\left(\boldsymbol{J}_{p,n}(\cdot) - \boldsymbol{J}^*_{p,n}(\cdot)\right)\right\|_{L^2[\underline{p}_n,\overline{p}_n]} = O_{\mathbb{P}}(\xi_n)$.
2. $\lambda_{\max}\left(\int_{\underline{p}_n}^{\overline{p}_n}\left(\boldsymbol{J}_{n,p}(u) - \boldsymbol{J}^*_{p,n}(u)\right)\left(\boldsymbol{J}_{n,p}(u) - \boldsymbol{J}^*_{p,n}(u)\right)' du\right) = O_{\mathbb{P}}(\varepsilon_n)$

Since we propose an inferential procedure conditionally on the data $\boldsymbol{J}_{n,p}$, estimation error enters the distributional approximation similarly to the bias term $D_n(u)$ in the nonstochastic setting. As it is generally difficult to bound the bias when $\boldsymbol{J}_{n,p}$ does not belong to a known family, we thus focus on an inferential approach that ignores the bias. Specifically, we consider the approximation:

$$\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n}) \in \operatorname{arginf}_{x\in\sqrt{n}(\Theta_n-\theta_{0,n})}\left\|\int_{\underline{p}_n}^{\overline{p}_n}(B_{0,n}(u) - x'\boldsymbol{J}_{p,n}(u))\boldsymbol{P}_L(u)du\right\|^2_{2,\Omega_{L,n}}, \tag{14}$$

and provide conditions that ensure asymptotic validity of this approximation over a class of sets $\mathcal{C}_p$, conditionally on the data, i.e.:

$$\sup_{C\in\mathcal{C}_p}|\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) \in C|\boldsymbol{J}_{n,p}] - \mathbb{P}[\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n}) \in C|\boldsymbol{J}_{n,p}]| \xrightarrow{p} 0\,.$$

To verify the above, we first present an analog of Proposition 1 in the stochastic setting.

Proposition 4. *Suppose that Assumptions 1, 2 and 3 are satisfied. Let $d_n := \|\sqrt{n}(Q(\cdot) - \theta'_{0,n}\boldsymbol{J}^*_{n,p}(\cdot)\|_{L^2[\underline{p}_n,\overline{p}_n]}$. For $z \in \mathbb{R}^p$, define the restricted eigenvalue around z as :*

$$\lambda_{z,n} := \inf_{s\in\mathcal{X}_n}\frac{(s-z)'\left(\int_{\underline{p}_n}^{\overline{p}_n}\boldsymbol{P}_L(u)\boldsymbol{J}^*_{n,p}(u)'du\right)'\Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n}\boldsymbol{P}_L(u)\boldsymbol{J}^*_{n,p}(u)'du\right)(s-z)}{\|s-z\|_2^2}\,. \tag{15}$$

Moreover, let:

$$\rho_n := \lambda_{\max}\left(\int_{\underline{p}_n}^{\overline{p}_n} (\boldsymbol{J}^*_{n,p}(u)\boldsymbol{J}^*_{n,p}(u)')du\right).$$

We then have that, for any sequences $(\delta_n)_{n\in\mathbb{N}}$, $(M_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ such that:

$$\begin{aligned}
M_n \frac{\lambda_{0,n}}{\sqrt{\rho_n \vee \varepsilon_n}} &\to \infty\,,\\
\mathbb{P}[\lambda_{\sqrt{n}(\theta^*_n-\theta_{0,n}),n} \le c_n] &\to 0\,,\\
\frac{c_n\delta_n^2}{(\delta_n+M_n)\sqrt{(\rho_n\vee\varepsilon_n)}(r_n\vee d_n\vee s_n\vee\xi_n)} &\to \infty\,,\\
\frac{c_n\delta_n^2}{(\delta_n+M_n)^2\,(\rho_n\vee\varepsilon_n)\,s_n} &\to \infty\,,
\end{aligned} \tag{16}$$

we have:

$$\mathbb{P}[\|\sqrt{n}(\hat{\theta}_n-\theta_{0,n})-\sqrt{n}(\tilde{\theta}^*_n-\theta_{0,n})\| \ge \delta_n] \to 0\,.$$

In addition, the convergence holds conditionally on $\boldsymbol{J}_{n,p}$, i.e.

$$\mathbb{P}[\|\sqrt{n}(\hat{\theta}_n-\theta_{0,n})-\sqrt{n}(\tilde{\theta}^*_n-\theta_{0,n})\| \ge \delta_n|\boldsymbol{J}_{n,p}] \xrightarrow{p} 0\,.$$

Proof. See Appendix A.5. □

The previous proposition provides rates for the proposed approximation. These rates depend, crucially, on the estimation error of the basis functions and its interaction with the oracle $\theta_{0,n}$, as reflected by the constants ξ_n and ϵ_n. In Section 6.2, we provide conditions that bound these rates when $\boldsymbol{J}_{n,p}$ are empirical quantiles from random samples of p populations. In this setup, proper control of the estimation error is achieved under a condition that ensures the weights $\theta_{0,n}$ are diluted across basis functions. As we remark in Section 6.2, a similar condition appears in the Synthetic Control literature (Ferman, 2021).

We conclude this section by providing a version of Proposition 3 to the case of stochastic basis.

Corollary 2. *Consider the setting with ridge constraint. Suppose that for constants $C > 0$,*

$\gamma \in [0,1)$, the smallest eigenvalue of $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}^*_{n,p}(u)'du\right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}^*_{n,p}(u)'du\right)$, $\lambda_{n,\min}$, satisfies $\lambda_{n,\min} > C((\rho_n \vee \varepsilon_n)s_n)^\gamma$.

Then, if $(\rho_n \vee \varepsilon_n)s_n \to 0$, $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}^*_{n,p}(u)'du\right)' W_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p^*}(u)'du\right)$ is invertible with probability approaching one and:

$$\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})\| = O_{\mathbb{P}}(((\rho_n \vee \varepsilon_n)s_n)^{1-2\gamma} \vee ((\rho_n \vee \varepsilon_n)s_n)^{-\gamma}\sqrt{(\rho_n \vee \varepsilon_n)}(r_n \vee d_n \vee \xi_n)))\,.$$

Moreover, suppose that the variance of the Gaussian approximation in the unconstrained setting is bounded below uniformly. Suppose that $R_n = Gp^{-l}$ for some $l \in [0, 1/4)$. If, for some $\nu > 1/2$

$$\log(p)^\nu((\rho_n \vee \varepsilon_n)s_n)^{-\gamma}p^{1/2+l}\left(((\rho_n \vee \varepsilon_n)s_n)^{1-2\gamma} \vee ((\rho_n \vee \varepsilon_n)s_n)^{-\gamma}\sqrt{(\rho_n \vee \varepsilon_n)}(r_n \vee d_n \vee \xi_n))\right) \to 0\,,$$

$$n((\rho_n \vee \varepsilon_n)s_n)^\gamma \to \infty,$$

then:

$$\sup_{A \in \mathcal{C}_p} |\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) \in A] - \mathbb{P}[\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n}) \in A | \boldsymbol{J}_{n,p}]| \xrightarrow{p} 0\,.$$

Proof. The proof is analogous to that of Proposition 3, but applied conditionally on $\boldsymbol{J}_{n,p}$. □

4.2 Inference on linear functionals of mixture weights

In several settings, direct interest is not on the mixture weights, but on linear functionals thereof. Specifically, one may be interested in conducting inference on the functional

$$T(x) := \omega(x)'\theta_{0,n}, \quad x \in \mathcal{X}\,, \tag{17}$$

where $(\mathcal{X}, \mathcal{L}, \Pi)$ is a measure space. In this case, confidence sets for T may be constructed by approximating the distribution of a suitable L^p-norm. Specifically, for $p \in [1, \infty]$, we consider the scalar random variable:

$$\mathcal{T}_p \coloneqq \sqrt{n}\|\omega(\cdot)'(\hat{\theta}_n - \theta_{0,n})\|_{L^p(\Pi)}\,. \tag{18}$$

If the quantiles of $\mathcal{T}_p$ were known, a valid $(1-\alpha)$ uniform confidence band for T could be constructed as:

$$C_{T,1-\alpha} = \left[\omega(\cdot)'\hat{\theta}_n - \frac{1}{\sqrt{n}}Q_{\mathcal{T}_p}(1-\alpha), \omega(\cdot)'\hat{\theta}_n + \frac{1}{\sqrt{n}}Q_{\mathcal{T}_p}(1-\alpha)\right]\,.$$

In practice, the distribution of $\mathcal{T}_p$ is unknown. One is thus tempted to use the distributional approximations discussed in the previous subsection to compute the quantiles. For example, if one considers an undersmoothened approximation, one could consider approximating the distribution of $\mathcal{T}_p$ with:

$$\tilde{\mathcal{T}}_p^* = \sqrt{n}\|\omega(\cdot)'(\tilde{\theta}_n^* - \theta_{0,n})\|_{L^p(\Pi)}\,.$$

Can we ensure this approximation provides valid inference? Application of Lemma 1, along with the Cauchy-Schwarz inequality, reveals that:

$$\sup_{c\in\mathbb{R}}|\mathbb{P}[\mathcal{T}_p \leq c]-\mathbb{P}[\tilde{\mathcal{T}}_p^* \leq c]| \leq \inf_{\delta\geq 0}\left\{\mathbb{P}\left[\|\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n^*)\|_2 \geq \frac{\delta}{\left(\int_{\mathcal{X}}\|\omega(x)\|_2^p\Pi(dx)\right)^{1/p}}\right] + \sup_{d\in\mathbb{R}}\mathbb{P}[d \leq \tilde{\mathcal{T}}_p^* \leq d+\delta]\right\}\,.$$

The above inequality shows that to approximate the quantiles of $\mathcal{T}_p$, one has to balance the strong approximation between $\sqrt{n}(\hat{\theta}_n - \theta_{0,n})$ and $\sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n})$ with the anticoncentration of $\mathcal{T}_p^*$. We have provided rates for the former in the previous subsection. As for the latter point, bounds on the anticoncentration of $\mathcal{T}_p^*$ may be obtained, in the unconstrained setting, by applying Theorem 2.1 of Chernozhukov et al. (2014).[6] These bounds may be extended to

[6]In the case $p = 2$, one may alternatively use the results in Götze et al. (2019).

the ridge setting by a similar approach as the one presented in Appendix B.

5 A practical algorithm for conducting inference

Based on the discussion in Section 4, we propose the following algorithm for conducting inference. Suppose that we wish to approximate the quantity $\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) \in B]$, where $B \in \mathcal{B}(\mathbb{R}^p)$. We can then proceed as follows:

Algorithm 1 Approximating the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$

1: Fix $S \in \mathbb{N}$ and $\gamma_n > 0$.
2: Estimate $\hat{\theta}_n$ by solving (2).
3: Estimate the variance matrix $\mathbb{V}_n \coloneqq \mathbb{V}\left[\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u) B_{n,0}(u) du\right)\right]$ using an estimator $\hat{V}_n$.
4: **for** s=1 to S **do**
5: Draw $Z_s \sim \mathcal{N}\left(0, \hat{V}_n\right)$.
6: Find and store $\mu_s^* \in \operatorname{argmin}_{x \in \gamma_n(\Theta_n - \hat{\theta}_{0,n})} \|Z_s - \int_{\underline{p}_n}^{\overline{p}_n} x' \boldsymbol{J}_{p,n}(u) \boldsymbol{P}_L(u) du\|_{W_{L,n}}^2$
7: **end for**
8: For $B \in \mathcal{B}(\mathbb{R}^p)$, estimate $\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \in B]$ by $\frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\{\mu_s^* \in B\}$.

Algorithm 1 provides a simulation method for approximating the quantity $\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \in B]$. For simplicity, the algorithm abstracts from bias-aware considerations, thus relying on an undersmoothened approximation. The algorithm requires two hyperparameters: (i) the number of simulations S; and (ii) a tuning parameter γ_n for setting the choice set, whose choice we discuss further below.

In Section 4, we have provided conditions such that $\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \in B]$ is asymptotically approximated by $\mathbb{P}[\sqrt{n}(\tilde{\theta}_n^* - \theta_0) \in B]$, uniformly over a class $B \in \mathcal{C}$. Note, however, that the quantity $\mathbb{P}[\sqrt{n}(\tilde{\theta}_n^* - \theta_0) \in B]$ depends on three unknowns, namely the variance matrix $\mathbb{V}_n = \mathbb{V}\left[\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u) B_{n,0}(u) du\right)\right]$, the weighting matrix $\Omega_{L,n}$ and the centering point $\theta_{0,n}$ of the choice set $\mathcal{X}_n$. Algorithm 1 suggests a "plug-in" approach, whereby these quantities are replaced by estimators $\hat{V}_n$, $W_{L,n}$ and $\hat{\theta}_n$. It also replaces $\sqrt{n}$ in the definition of the choice set $\mathcal{X}_n$ with a tuning parameter $\gamma_n > 0$. Such modification is needed in constrained settings to control how estimation error of the centering point of $\mathcal{X}_n$ affects the quality of the approximation of the true choice set, especially with regard to the "bindingness" of the constraints (see Hong and Li, 2020; Li, 2021; Chernozhukov et al., 2023, for related bootstrap strategies in different constrained settings).

The choice of γ_n depends crucially on the local geometry of $\mathcal{X}_n$ around $\theta_{0,n}$. For example, in the unconstrained setting ($\Theta_n = \mathbb{R}^p$), $\mathcal{X}_n = \mathbb{R}^p$ irrespectively of the value of $\theta_{0,n}$ or γ_n. As a consequence, in this case, validity of Algorithm 1 hinges solely on proper estimation of $\mathbb{V}_n$ and $\Omega_{L,n}$. More generally, the value of $\theta_{0,n}$ affects the geometry of $\mathcal{X}_n$. In this case, Appendix C outlines a general approach which, building upon Lemma 2 in Appendix A, provides conditions which can be used to find a sequence γ_n in specific settings. We also refer the reader to Cattaneo et al. (2021) for a discussion on the role of properly accounting for the local geometry of $\mathcal{X}_n$ when bootstrapping in a constrained setting.

6 Theoretical Applications

6.1 Empirical Bayes

We consider the Empirical Bayes problem, which we briefly outlined in the introduction. Suppose the researcher has access to a sample of identically distributed real random variables, which admit a marginal Lebesgue density f satisfying:

$$f(y) = \int_{\Xi} \phi(y;\xi) G(d\xi)\,, \quad y \in \mathbb{R}, \tag{19}$$

where $\Phi \coloneqq \{\phi(\cdot;\xi) : \xi \in \Xi\}$ is a known parametric family of densities on $\mathbb{R}$, $(\Xi, \mathcal{J})$ is a measurable space, and G is an unknown probability measure on $(\Xi, \mathcal{J})$. Our interest lies in estimating G. For example, one may have access to a sample of noisy standardized measurements of school quality ξ_i for n schools, $Y_i|\xi_i \sim N(\xi_i, 1)$, $i = 1, \ldots, n$; and would like to estimate the population distribution of school quality G, so as to select the schools above 95th percentile of school quality. Alternatively, the researcher may have access to noisy measurements on treatment effects of a policy across different sites, and would like to estimate G to understand the population distribution of these effects. Finally, the researcher may need to estimate G to perform shrinkage on some preliminary noisy measure (Armstrong et al., 2022).

In estimating model (19), Efron (2014) distinguishes between two possible approaches. In f-modelling, one estimates a model for f, and then inverts the map $G \mapsto f_G$ given by (19) to find an estimate of G. When the model Φ corresponds to a location family, this procedure subsumes to the well known *density deconvolution problem*. In contrast, in G-modelling, one uses the structure given by (19) to directly estimate G, e.g. by applying the nonparametric MLE of Kiefer and Wolfowitz (1956).

Efron (2016) argues that fully nonparametric f- or G-modelling approaches are often undesirable, as rates of convergence can be poor. In the f-modelling setting, this corresponds to known problems of nonparametric density deconvolutions (Fan, 1991; Meister, 2009). In the G-modelling setup, the nonparametric MLE, in spite of enjoying "optimality" properties (e.g. Polyanskiy and Wu, 2020), can also perform poorly. Perhaps equally or more importantly, it is not yet clear how to conduct frequentist uncertainty quantification based on the nonparametric MLE (Ignatiadis and Wager, 2022).

To circumvent the problems with a fully nonparametric approach, Efron (2016) proposes a sieve G-modelling approach. Motivated by his discussion, and given the general approximation properties of quantile mixture models discussed in Section 2, as well as the attractive statistical properties of L-moments, we propose a sieve f-modelling approach based on our estimation method. For that, we rely on the observation that, for a quantile function Q_F obtained upon inversion of a strictly increasing and differentiable distribution function F, we have that:

$$Q'_F(u) = \frac{1}{F'(Q_F(u))} \,.$$

In light of this observation, we propose the following steps to estimate G.

1. Choose nonstochastic differentiable basis functions $\boldsymbol{J}_{n,p}$ and estimate a quantile mixture model by (2).

2. Estimate $f(Q_F(u))$ by $f(\widehat{Q_F(u)}) = \frac{1}{\hat{\theta}_n' \partial_u \boldsymbol{J}_{n,p}(u)}$.

3. Invert (19) to obtain $\hat{G}$.

To conduct frequentist uncertainty quantification on G or functionals thereof, one can rely on Algorithm 1. Specifically, for a given confidence level $(1-\alpha)$, if the resulting confidence set for (a functional of) G can be written as $\{\theta \in \Theta_n : \sqrt{n}(\hat{\theta}_n - \theta) \in D\}$, where D is a convex set such that $\mathbb{P}[\{\sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n}) \in D\}] \geq 1-\alpha$, then the validity of the approach in conducting inference can be ascertained by proceeding similarly to Section 4.2. In this case, one must rely, in the unconstrained setting, on an anticoncentration inequality for convex sets (Chernozhukov et al., 2017a, Lemma A.2). In the ridge setting, this inequality may be extended using the arguments in Appendix B. The assumption that D is convex is not restrictive: note that the inversion step consists on the application of a linear operator on $\hat{f}$. Combined with the fact that $x \mapsto 1/x$ is convex (thus exhibiting convex lower contour sets), we can construct convex confidence sets for G.

Remark 8. Bias-aware inference has been recently advocated in the Empirical Bayes setup by Ignatiadis and Wager (2022). We note that, by placing bounds on the approximation error of the sieve-quantile model on the true Q, it may be possible to conduct bias-aware inference on G. It should be noted, however, that strong restrictions may be needed for these bounds to be informative. Indeed, since the differentiation operator is unbounded in general function spaces, the construction of informative bounds may require strong restrictions on the candidate function space for Q. As we remarked earlier on, we do not pursue bias-aware inference in this paper.

6.2 Distributional synthetic controls

In this section, we show how our methodology can be applied to the distributional synthetic control setting of Gunsilius (2023), which we briefly overviewed in Example 5.[7] We note that Gunsilius (2023) does not introduce formal inference methods in his distributional setup: in the article, there is a brief suggestion of using placebo to assess the uncertainty in estimates, though no formal justification is given to it.

We start by recasting the problem in the notation of Section 4. We consider a setting where we only have one pre-intervention period, which we denote by t^*-1. The stochastic basis functions $\mathbf{J}_{t^*-1,n,p}$ represent the empirical quantile functions of the p control units at time $t^* - 1$. These empirical quantiles are estimators of the true population quantile functions $\mathbf{J}^*_{t^*-1,n,p}$, which are unknown. The empirical quantile function of the treated unit in the pre-treatment period is $\hat{Q}_{t^*-1}$, and Q_{t^*-1} is its population counterpart. We begin with the identification assumption in our distributional setting.

Assumption 4 (Identification Assumption of Distributional Synthetic Control)**.** *There exists $\theta_{0,n} \in B_{\mathbb{R}^p,1}(0)$ such that $Q_{t^*-1} = \theta_{0,n}' \boldsymbol{J}^*_{t^*-1,n,p}$.*

Assumption 4 requires that, if the population counterparts were known, it would be possible to perfectly replicate the quantile function of the treated unit in the pre-treatment period as a mixture of the controls. We assume that the weights belong to the unit Euclidean ball, which motivates the use of ridge regularization in the analysis.

The next assumption constrains the estimation error of the empirical quantile functions.

Assumption 5 (Sampling)**.** *$\hat{Q}_n$ is the empirical quantile function from a random sample of size n from a continuous distribution F_0. Similarly, for each $i = 1, \ldots, p$. $\mathbf{J}_{t^*-1,n,i}$ is the empirical quantile function from a random sample of size n_i from a continuous distribution function F_i. Finally, for each $k \in \{0, 1, \ldots, p\}$, we assume that:*

[7] Its worth noting that, when there is only one pre-intervention period, the estimator proposed by Gunsilius (2023) is equivalent to (2) when $L = \infty$, $W_{n,L} = \mathbb{I}_L$, $\underline{p}_n = 0$, $\overline{p}_n = 1$ and $\Theta_n = \Delta^{p-1}$. More generally, by using the optimal weighting scheme, we expect improvements over Gunsilius (2023) estimator.

1. *F_k is twice differentiable on (a_k, b_k), where $a_k = \sup\{x : F_k(x) = 0\}$, $b_k = \inf\{x : F_k(x) = 1\}$, and $F_k' \neq 0$ on (a_k, b_k).*

2. $\sup_{a_k < x < b_k} F_k(x)(1 - F_k(x)) \left| \frac{F''(x)}{F'^2(x)} \right| \leq \alpha_k$*, for some* $\alpha_k > 0$.

Assumption 5 requires that the population distributions satisfy the assumptions of Theorem 6 of Csorgo and Revesz (1978), which provides rates for the strong approximation of empirical quantile functions.

Assumption 6 ("Dilution condition")**.** *We require that:*

1. $\min_{i \in 1,\ldots,p} n_i \to \infty$.

2. *There exists* $\underline{f} > 0$ *such that,* $\inf_{i \in \mathbb{N}} \inf_{x \in \mathbb{R}} F_i'(x) \geq \underline{f}$.

3. *For some* $\mu > 1$ *we have*
$$\sum_{i=1}^{p} \theta_{0,i,n}^2 \frac{n}{n_i} = O_{\mathbb{P}} \left(\frac{1}{p^{\mu}} \right)$$

Assumption 6 provides conditions that allow us to properly control the estimation error of the basis functions. Specifically, under the Csorgo and Revesz conditions in Assumption 5 and Assumption 6, we are able to show that:

$$\left\| \theta_{0,n}' \sqrt{n} \left(\boldsymbol{J}_{p,n}(\cdot) - \boldsymbol{J}_{p,n}^{*}(\cdot) \right) \right\|_{L^2[\underline{p}_n, \overline{p}_n]}^2 = O_{\mathbb{P}} \left(\frac{1}{p^{\mu - 1}} \right),$$

ensuring that estimation error of the basis functions may be safely ignored when the number of controls is large. Crucially, Assumption 6 requires the oracle weights to be sufficiently diluted across controls, so the sampling error of any control unit becomes asymptotically negligible. A similar condition appears, either explicitly or implicitly, in the literature on inference in Synthetic Controls, where randomness of the outcomes in the donor pool is abstracted from when conducting inference (Ferman, 2021).

Under the previous assumptions, we can derive explicit convergence rates by directly appealing to Corollary 2.

Corollary 3. *Suppose that for constants $C > 0$, $\gamma \in [0, 1)$, the smallest eigenvalue of $\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}^*(u)'du\right)' \Omega_{n,L} \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}^*(u)'du\right)$, $\lambda_{n,\min}$ satisfies, $\lambda_{n,\min} > C(\rho_n s_n)^\gamma$. Suppose that the assumptions in Proposition 4 and Assumptions 4, 5 and 6 hold. Moreover, suppose the variance of the Gaussian Approximation in the unconstrained setting is bounded below uniformly, $\varepsilon_n \leq \rho_n$, and $R_n = 1$ for all n. If*

$$\log(p)^\nu (\rho_n s_n)^{-\gamma} p^{1/2} \left((\rho_n s_n)^{1-2\gamma} \vee (\rho_n s_n)^{-\gamma} \sqrt{\rho_n} \frac{\log(n)^\alpha}{n^{1/2}} \right) \to 0\,,$$

$$n(\rho_n s_n)^\gamma \to \infty\,,$$

$$\log(p)^\nu \frac{\sqrt{n}}{p^b} \to 0\,,$$

then:

$$\sup_{A \in \mathcal{C}_p} |\mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) \in A | \boldsymbol{J}_{n,p}] - \mathbb{P}[\sqrt{n}(\tilde{\theta}_n^* - \theta_{0,n}) \in A | \boldsymbol{J}_{n,p}]| \xrightarrow{p} 0\,.$$

7 Empirical Application

In this section, we apply our methodology to assess the effects of an environmental catastrophe on the local wage distribution of the affected municipality.

7.1 Brumadinho Dam Disaster

7.1.1 Context

On January 25, 2019, a devastating tailings dam collapse occurred near the city of Brumadinho, located in the state of Minas Gerais, Brazil. This incident resulted in the loss of 270 lives and caused extensive contamination of the neighboring Paraopeba River. The victims primarily consisted of employees of Vale, the largest mining conglomerate in Latin America, which operated both the dam and the mines providing materials for its construction.

The release of mining waste resulted in significant environmental and socioeconomic damage. Brumadinho suffered the most severe impact as the closest city to the collapsed dam. However, according to official authorities, an additional 25 cities were also affected. Damages resulted in the loss of hectares of forest and vegetation cover, as well as the contamination of the Paraopeba River with elevated levels of copper, negatively impacting its ecosystem and local agriculture. The city of Brumadinho experienced a decline in revenue due to the halting of mining operations, adversely affecting the local economy and government finances.

Vale faced legal and financial repercussions, including penalties and the obligation to provide compensation to affected individuals and communities.[8] According to the judicial agreement reached in 2021, Vale is required to allocate approximately 7.58 billion USD towards reparations, of which 4.4 billion USD has already been disbursed, with a substantial portion allocated to economic restitution for the affected cities.[9] Amidst all the repercussions, the labor market consequences of such an event still remain an important economic and policy question.

7.1.2 Empirical question: distributive consequence on wage earners

The distributive impact of the events in Brumadinho on the wages of formal sector workers is an open empirical question. The economic disruptions resulting from the dam collapse would negatively affect job opportunities but it is unclear which workers are more exposed to the shock. However, the allocation of reparations and investments in the affected areas as well as income support programs would go in the other direction, helping mitigate the impacts. The overall economic recovery and reconstruction efforts may generate new jobs and attract investments, benefiting workers across different skill levels.

The main empirical challenge that motivates the choice of the distributional synthetic

[8]Upon the initial news of the disaster, the stock price of Vale on the B3 (São Paulo Stock Exchange) fell over 10%. It then fell another 24% on the next trading day, January 28, corresponding to $19 billion in lost market capitalization (Source: Reuters). At the same time, bond prices fell, reflecting an increased risk of non-payment by Vale, as one of the consequences included a court order that froze 11.8 billion reais (USD 3.1 billion at the time) in Vale's accounts, comprising roughly half of Vale's cash on hand.

[9]For comprehensive and detailed information regarding affected cities and the reparations, please refer to the official website of the State Government of Minas Gerais. For the expenditure report on reparations by Vale, please refer to the official website of the company. Both sources are in Portuguese.

control method is the nature of the shock's impact, affecting at once both the city of Brumadinho and neighboring municipalities. The lack of a clear control group hinders a direct comparison and motivates the need to construct the synthetic control unit. The approach discussed in Section 6.2 is particularly suitable as a large number of potential control units (other municipalities) favors regularization and correct inference is needed to assess the effects on the entire distribution and validate the identification Assumption 4 by evaluating the fit on pre-shock period data not used in the analysis.

7.2 Data

Our primary outcome of interest is the distribution of average monthly wages for private sector workers in each municipality. We use the publicly accessible Brazilian employer-employee matched data (RAIS) to construct the empirical quantile function for each municipality $\times$ year pair. RAIS database, also known as Relação Anual de Informações Sociais (Annual Social Information Report) is a comprehensive dataset maintained by the Brazilian the Ministry of Labor and Employment and contains information on formal sector employment in the country, including both public and private sector data.[10] We use data from 2017-2021, details are on Appendix E.

7.3 Empirical Strategy

We set the treatment year to 2019, $t^* = 2019$, and use as potential control units other municipalities within the same state as Brumadinho (Minas Gerais) that were not affected by the barrage rupture ($p = 827$). We solve for the synthetic control estimator as in (2):

$$\hat{\theta}_n \in \text{arginf}_{\theta \in B_{\mathbb{R}^p,1}(0)} \left\| \left(\int_0^1 (\hat{Q}_{2018}(u) - \theta' \boldsymbol{J}_{2018,p}(u)) \boldsymbol{P}_L(u) du \right) \right\|^2_{2,\mathbb{I}_L}, \tag{20}$$

[10]The dataset is collected annually from employers through mandatory reporting requirements. Employers are required to provide detailed data about their employees, including their employment status, occupation, earnings, and other relevant indicators and demographic characteristics. Within the publicly accessible version of the dataset, we can observe for a given municipality the entire wage distribution of formal sector workers (but not individual IDs).

where $\hat{Q}_{2018}(\cdot)$ is Brumadinho's empirical average monthly wage quantile function in the year 2018, $\boldsymbol{J}_{2018,p}(\cdot)$ are the empirical wage quantile functions of the control units. We set $L = 1000$ and use the identity weighting matrix.

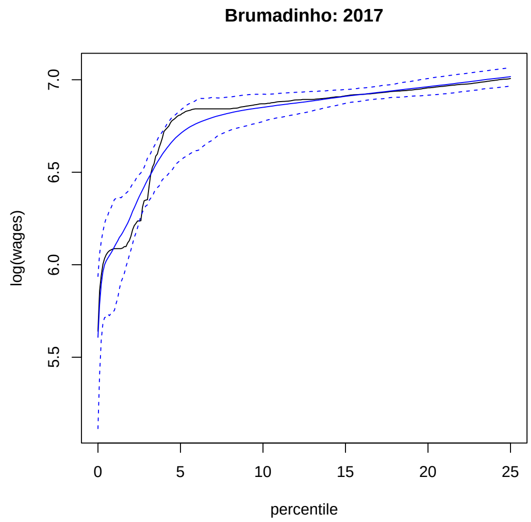
To construct confidence sets for $\boldsymbol{J}^*_{t,p}(\cdot)'\theta_{0,n}$, the counterfactual distribution at year t, we rely on Algorithm 1. Specifically, we set $\gamma_n = \sqrt{n}$, where $n = 9,632$ is the number of observations in Brumadinho in 2018; and consider $S = 500$ simulations. We then use these simulations to construct pointwise confidence intervals for $\boldsymbol{J}^*_{t,p}(u)'\theta_{0,n}$ at different years and quantile levels u.

7.4 Results

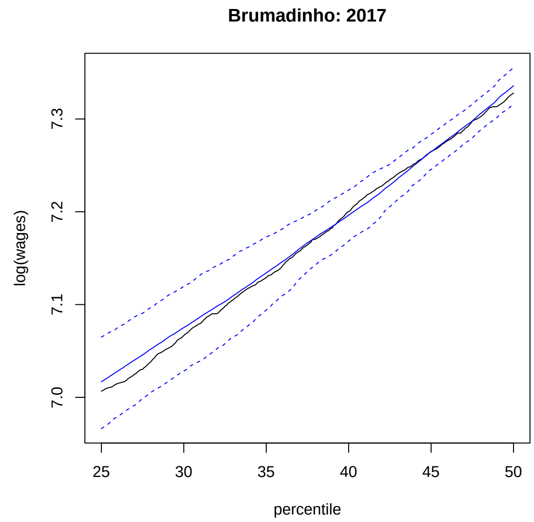
We conduct two types of analysis. First, as a check on the plausibility of Assumption 4, we evaluate how well the weights in 2018 replicate the wage distribution of Brumadinho in 2017. Next, we evaluate the distributional effects at years $t \in \{2019, 2020, 2021\}$. The observed empirical quantile functions are reported in black solid lines, whereas the estimated counterfactuals are reported in blue, along with 95% pointwise confidence intervals for the counterfactual in blue dashed lines.

7.4.1 Pretraining Fit

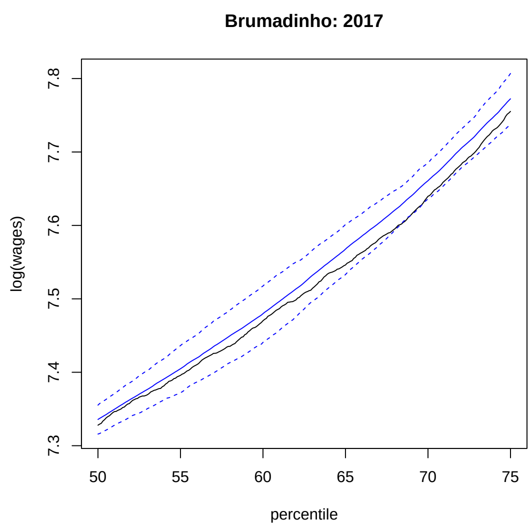
Figure 1 reports the observed quantile function of Brumadinho in 2017 (black line), along with the counterfactual that uses the weights estimated in 2018 (blue solid line) and the associated 95% pointwise confidence intervals (blue dashed line). Overall, pretreatment fit is good, though 2018 weights have some difficulty in reproducing the distribution at the lowermost percentiles.



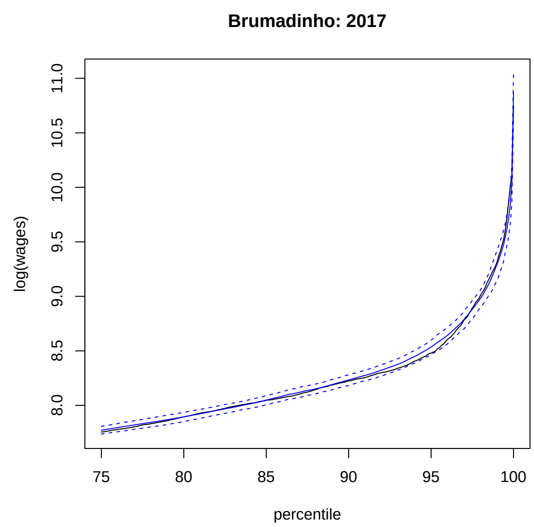
(a) Percentiles 0-25



(b) Percentiles 25-50



(c) Percentiles 50-75



(d) Percentiles 75-100

Figure 1: Pretraining fit (2017).

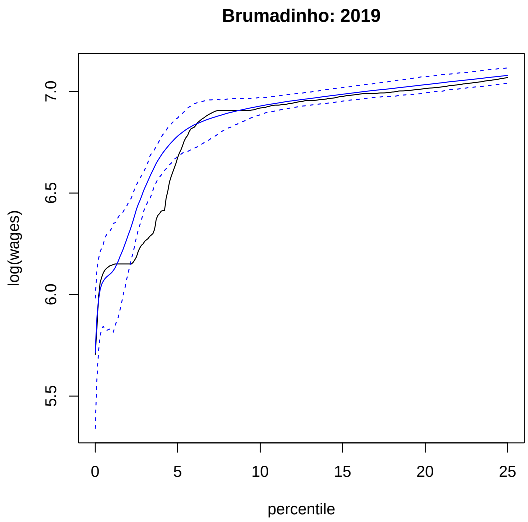
7.4.2 Treatment Effects

Figure 2 reports distributional treatment effects in 2019. We note that the observed quantiles remain below the counterfactual at the smallest percentiles, with this difference being statistically significant around percentiles 3-5. In the range 6-30, observed and counterfactual quantiles remain close to each other, after which the observed quantiles *cross* the counterfactual, remaining above it in the range 35-65, with the difference being (slightly) significant in the range 50-60. After the 65th quantile, observed and counterfactual quantiles return to being close to each other.

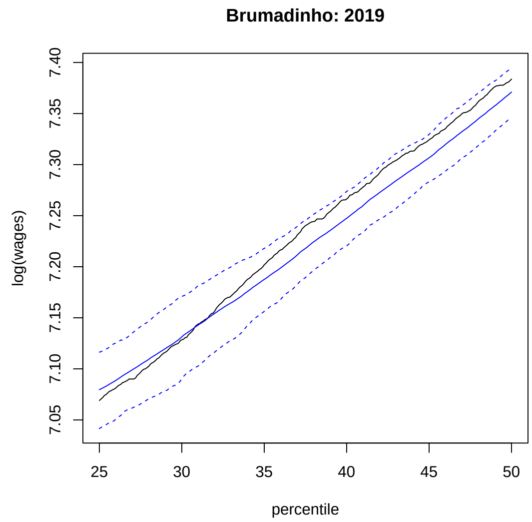
Distributional effects in 2020 (Figure 3) and 2021 (Figure 4) exhibit quite different patterns. In 2020, there is still a slight, though barely significant, decrease in the observed quantiles, vis-às-vis the counterfactual, around the 5th percentile. In contrast, after the median, the observed quantile is consistently (and significantly), above the counterfactual. In 2021, effects at the lower tail mostly disappear, with the observed counterfactual being above the counterfactual around the median and, especially, after the 95th percentile.

Overall, our results uncover a range of distributional effects on the wage distribution. The crossing of distributions, which is observed in 2019 and, to a lesser degree, in 2020, is consistent with a *displacement effect*, whereby intermediate-paying wages are replaced by low-payment contracts. Such a phenomenon leads simultaneously to a decrease in the lowest quantiles, and an increase from the median on. The counterfactual and wage distributions in 2021 exhibit a pattern which more closely resembles first-order stochastic dominance, with quantiles being above the counterfactual around the median and especially in the upper tail. Such pattern is compatible with two possible explanations. First, an increase in the median may be explained by a loss of intermediate-paying jobs, with *no subsequent replacement* by low-earning contracts. The increase in the upper tail could also be driven by a similar phenomenon, though an alternative explanation is also plausible. In Brumadinho, most private high-earning jobs are offered by Vale. Inasmuch as Vale workers are able to extract higher concessions from the company after the rupture, this could lead to an increase in the upper tail.[11]

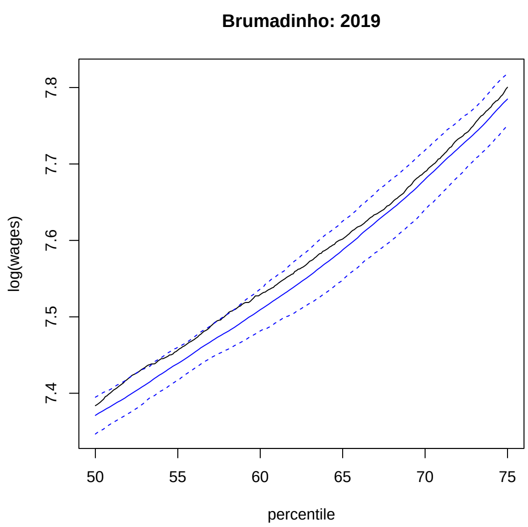
[11] In a similar vein, the increase in the median could be alternatively driven by an increase in median-paying jobs three years into the barrage rupture, e.g. due to sectoral changes in the local economy driven by the disaster.



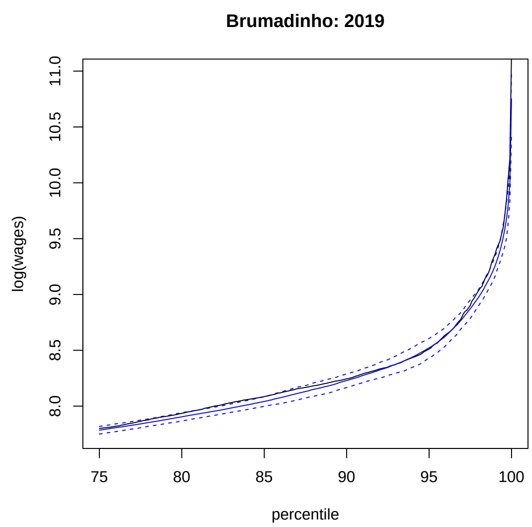
(a) Percentiles 0-25



(b) Percentiles 25-50

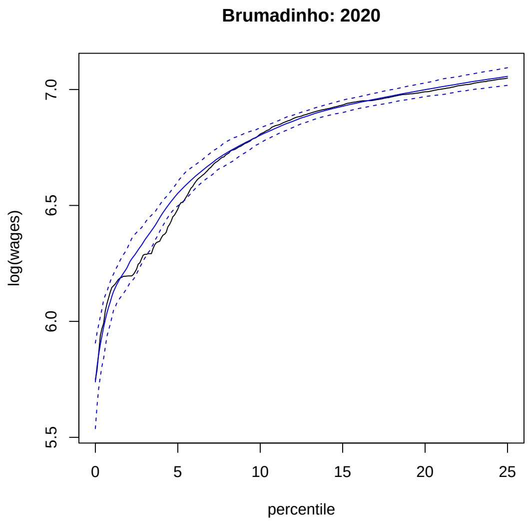


(c) Percentiles 50-75

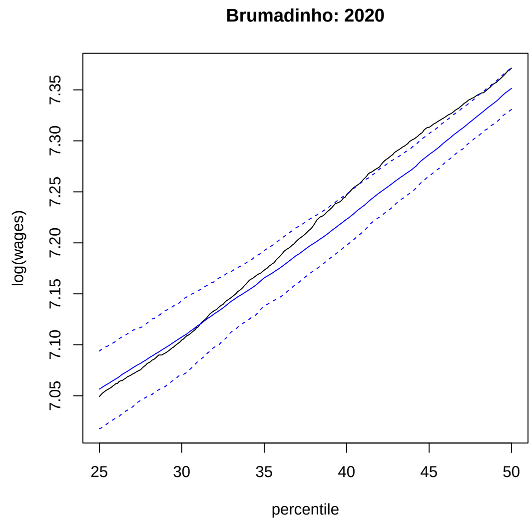


(d) Percentiles 75-100

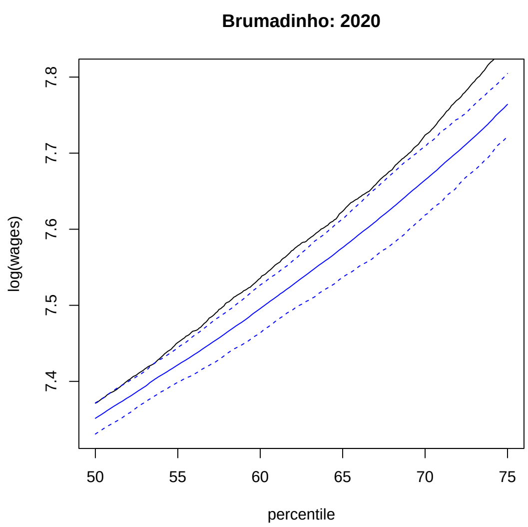
Figure 2: Treatment effects (2019).



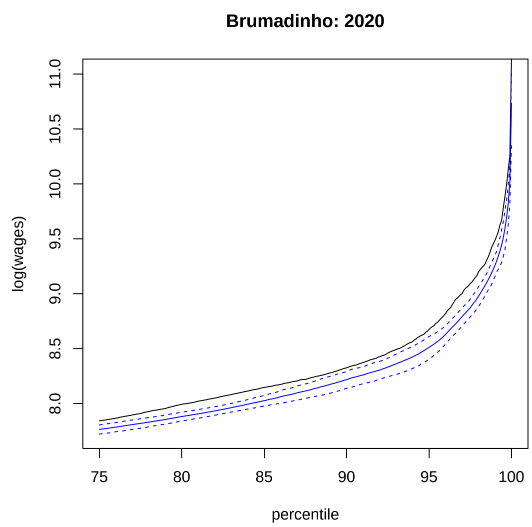
(a) Percentiles 0-25



(b) Percentiles 25-50

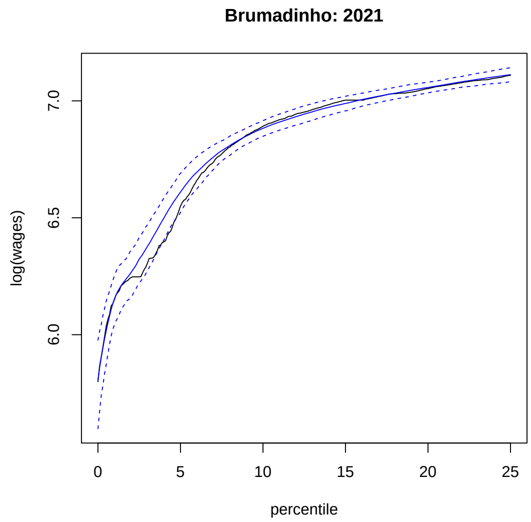


(c) Percentiles 50-75

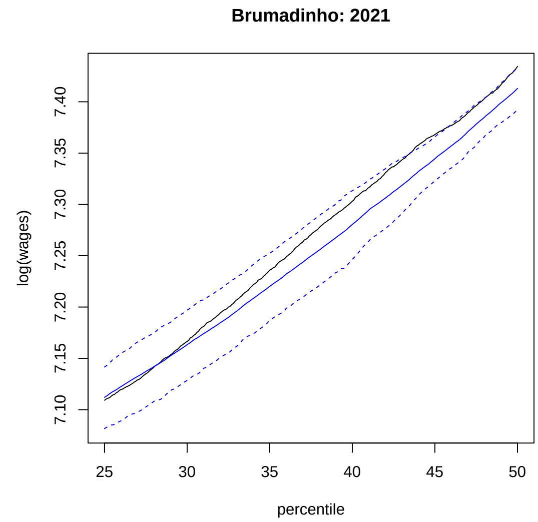


(d) Percentiles 75-100

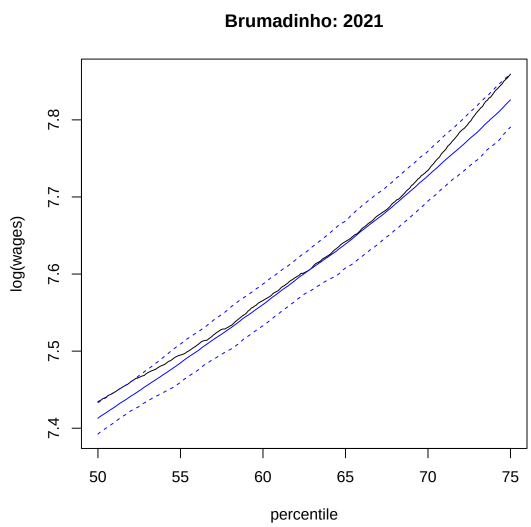
Figure 3: Treatment effects (2020).



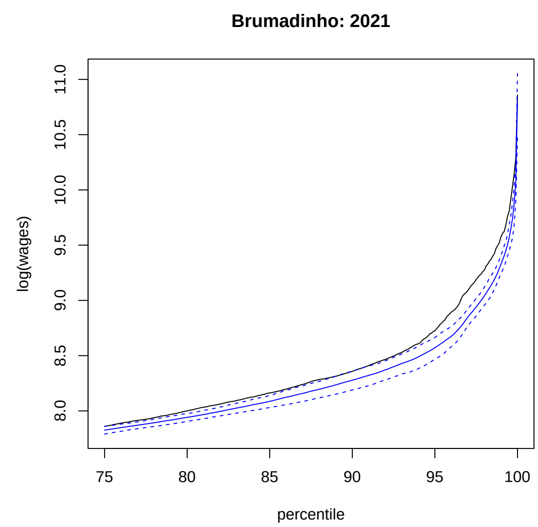
(a) Percentiles 0-25



(b) Percentiles 25-50



(c) Percentiles 50-75



(d) Percentiles 75-100

Figure 4: Treatment effects (2021).

8 Conclusion

In this paper, we have introduced nonparametric quantile mixture models as an attractive counterpart to nonparametric density mixture models. We have shown that, similarly to density mixture models, nonparametric quantile mixtures exhibit interesting approximation properties. They are also closely connected to the distributional synthetic controls recently proposed by Gunsilius (2023).

We have introduced estimation and inference tools for nonparametric mixtures by relying on the concept of *L-moments*. Introduced by Hosking (1990), L-moments are robust alternative alternative to standard moments that characterise distributions with finite first moments. The estimation of models by matching a weighted distance between sample and theoretical L-moments has been shown to produce statistically efficient estimators in parametric (Alvarez et al., 2023) and semiparametric (Alvarez and Biderman, 2022) settings. In this paper, we have extended this estimation approach to a sieve setting, which we show leads to a computationally convenient estimator with tractable statistical properties.

We develop a full inferential theory for our proposed estimator – including a general approach to constructing confidence sets on mixture weights –, by relying on the concept of strong approximation. In so doing, we make two contributions to statistical theory, which we believe may be of independent interest. First, we introduce a lemma that bounds the strong approximation of a quadratic optimizer by an approximating program in terms of its constituent elements. Inasmuch as strong approximations are an ever more prevalent tool in devising inferential procedures in nonstandard or high-dimensional settings (see Chernozhukov et al., 2014; Armstrong and Kolesár, 2017; Cattaneo et al., 2020; Fang et al., 2023; Chernozhukov et al., 2023, for some recent examples), this result may be useful in other contexts. Second, we develop a strategy that enables extending anticoncentration inequalities for Gaussian random variables available in the literature to the projection of Gaussian random variables onto Euclidean balls. Such procedure is useful for providing inferential guarantees in constrained estimation settings.

As theoretical applications of our proposed methodology, we show how quantile mixtures may be used to recover estimates (and confidence sets) of a nonparametric *density* model. We also show that, as a direct byproduct of our theory, one is able to provide a valid inferential procedure in the distributional synthetic control setup of Gunsilius (2023), where formal inference procedures were previously unavailable. As an empiricl application, we apply our inferential procedure in order to assess the distributional impacts of the Brumadinho barrage rupture. Our approach enables us to uncover a rage of effects across the wage distribution.

We believe there are several additional applications to our proposed methodology. For example, in a risk manamagement setting, one may be tempted to model the quantile function of the prediction error of an algorithm as a mixture of extreme value distributions in order to better understand the likelihood of extreme returns. Even though our theory is general enough to accomodate this setting, our results require a strong approximation to the adopted quantile estimator, which is generally unavailable when these are empirical quantiles from the residuals of a first-step. As we discuss in Section 3, it appears possible to extend the strong approximation of Csorgo and Revesz (1978) to this setting by relying on sample-splitting and cross-fitting. We intend to pursue such extensions in future research.

References

Alvarez, L. and C. Biderman (2022). Semiparametric analysis of randomized experiments using l-moments.

Alvarez, L., C. Chiann, and P. Morettin (2023). Inference in parametric models with many l-moments.

Armstrong, T. B. and M. Kolesár (2017, 08). A Simple Adjustment for Bandwidth Snooping. *The Review of Economic Studies 85*(2), 732–765.

Armstrong, T. B. and M. Kolesár (2021). Finite-sample optimal estimation and inference on average treatment effects under unconfoundedness. *Econometrica 89*(3), 1141–1177.

Armstrong, T. B., M. Kolesár, and M. Plagborg-Møller (2022). Robust empirical bayes confidence intervals. *Econometrica 90*(6), 2567–2602.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics 186*(2), 345–366. High Dimensional Problems in Econometrics.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics 37*(4), 1705 – 1732.

Cattaneo, M. D., M. H. Farrell, and Y. Feng (2020). Large sample properties of partitioning-based series estimators. *The Annals of Statistics 48*(3), 1718 – 1741.

Cattaneo, M. D., Y. Feng, and R. Titiunik (2021). Prediction intervals for synthetic control methods. *Journal of the American Statistical Association 116*(536), 1865–1880. PMID: 35756161.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, 01). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics 42*(5), 1787–1818.

Chernozhukov, V., D. Chetverikov, and K. Kato (2017a). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability 45*(4), 2309–2352.

Chernozhukov, V., D. Chetverikov, and K. Kato (2017b). Detailed proof of nazarov's inequality.

Chernozhukov, V., D. Chetverikov, K. Kato, and Y. Koike (2023). High-dimensional data bootstrap. *Annual Review of Statistics and Its Application 10*, 427–449.

Chernozhukov, V., W. K. Newey, and A. Santos (2023). Constrained conditional moment restriction models. *Econometrica 91*(2), 709–736.

Chernozhukov, V., K. Wüthrich, and Y. Zhu (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association 116*(536), 1849–1864.

Csorgo, M. and P. Revesz (1978). Strong Approximations of the Quantile Process. *The Annals of Statistics 6*(4), 882 – 894.

DeVore, R. A. and G. G. Lorentz (1993). *Constructive approximation*, Volume 303. Springer Science & Business Media.

Efron, B. (2014). Two modeling strategies for empirical bayes estimation. *Statistical Science 29*(2), 285–301.

Efron, B. (2016, 02). Empirical Bayes deconvolution estimates. *Biometrika 103*(1), 1–20.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 1257–1272.

Fan, J., R. Masini, and M. C. Medeiros (2022). Bridging factor and sparse models.

Fang, Z., A. Santos, A. M. Shaikh, and A. Torgovitsky (2023). Inference for large-scale linear systems with known coefficients. *Econometrica 91*(1), 299–327.

Feng, L. and L. H. Dicker (2018). Approximate nonparametric maximum likelihood for mixture models: A convex optimization approach to fitting arbitrary multivariate mixing distributions. *Computational Statistics & Data Analysis 122*, 80–91.

Ferman, B. (2021). On the properties of the synthetic control estimator with many periods and many controls. *Journal of the American Statistical Association 116*(536), 1764–1772.

Fotopoulos, S. and S. Ahn (1994). Strong approximation of the quantile processes and its applications under strong mixing properties. *Journal of Multivariate Analysis 51*(1), 17–45.

Götze, F., A. Naumov, V. Spokoiny, and V. Ulyanov (2019). Large ball probabilities, Gaussian comparison and anti-concentration. *Bernoulli 25*(4A), 2538 – 2563.

Gourieroux, C. and J. Jasiak (2008). Dynamic quantile models. *Journal of Econometrics 147*(1), 198–205. Econometric modelling in finance and risk management: An overview.

Gu, J. and R. Koenker (2023). Invidious comparisons: Ranking and selection as compound decisions. *Econometrica 91*(1), 1–41.

Gunsilius, F. F. (2023). Distributional synthetic controls. *Econometrica 91*(3), 1105–1117.

Hong, H. and J. Li (2020). The numerical bootstrap. *The Annals of Statistics 48*(1), 397 – 412.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological) 52*(1), 105–124.

Hosking, J. R. M. and J. R. Wallis (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics 29*(3), 339–349.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics 27*(3), 251–261.

Hsu, D., S. Kakade, and T. Zhang (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability 17*(none), 1 – 6.

Ignatiadis, N. and S. Wager (2022). Confidence intervals for nonparametric empirical bayes analysis. *Journal of the American Statistical Association 117*(539), 1149–1166.

Jagabathula, S., L. Subramanian, and A. Venkataraman (2020). A conditional gradient approach for nonparametric estimation of mixing distributions. *Management Science 66*(8), 3635–3656.

Karvanen, J. (2006). Estimation of quantile mixtures via l-moments and trimmed l-moments. *Computational Statistics & Data Analysis 51*(2), 947–959.

Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis 100*(8), 1816–1829.

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 27*(4), 887 – 906.

Kline, P., E. K. Rose, and C. R. Walters (2022, 06). Systemic Discrimination Among Large U.S. Employers*. *The Quarterly Journal of Economics 137*(4), 1963–2036.

Koenker, R. and I. Mizera (2014). Convex optimization, shape constraints, compound decisions, and empirical bayes rules. *Journal of the American Statistical Association 109*(506), 674–685.

Kreyszig, E. (1978). Introductory functional analysis with applications. jonh wiley & sons. *Inc., New York*.

Li, J. (2021, May). The proximal bootstrap for finite-dimensional regularized estimators. *AEA Papers and Proceedings 111*, 616–20.

Masini, R. (2022). Distributional counterfactual analysis in high-dimensional setup. *arXiv preprint arXiv:2202.11671*.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics.* Springer Berlin Heidelberg.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168.

Nguyen, H. D. and G. McLachlan (2019). On approximations via convolution-defined mixture models. *Communications in Statistics - Theory and Methods 48*(16), 3945–3955.

Nguyen, T. T., H. D. Nguyen, F. Chamroukhi, and G. J. McLachlan (2020). Approximation by finite mixtures of continuous density functions that vanish at infinity. *Cogent Mathematics & Statistics 7*(1), 1750861.

Noack, C. and C. Rothe (2021). Bias-aware inference in fuzzy regression discontinuity designs.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory 7*(2), 186–199.

Polyanskiy, Y. and Y. Wu (2020). Self-regularizing property of nonparametric maximum likelihood estimator in mixture models.

Rudelson, M. and R. Vershynin (2013). Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability 18*(none), 1 – 9.

Shvedov, A. S. (1981). Orders of coapproximation of functions by algebraic polynomials. *Mathematical notes of the Academy of Sciences of the USSR 29*, 63–70.

Taussky, O. (1949, 01). A remark concerning the characteristic roots of the finite segments of the Hilbert Matrix. *The Quarterly Journal of Mathematics os-20*(1), 80–83.

Train, K. E. (2008). Em algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling 1*(1), 40–69.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer New York.

Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Yoshihara, K. (1995). The bahadur representation of sample quantiles for sequences of strongly mixing random variables. *Statistics & Probability Letters 24*(4), 299 – 304.

Appendices

A Proof of main results in Section 4

A.1 A lemma on the approximation of quadratic minimizers

Before proving the results in Section 4, we state an auxiliary lemma, which concerns the strong approximation of quadratic minimizers. This lemma may be of independent interest:

Lemma 2 (Strong approximation of quadratic minimizers)**.** *Consider the program:*

$$\psi \in \min_{\xi\in\mathcal{X}} -\xi' S + \frac{1}{2}\xi' W \xi \,,$$

where $\mathcal{X}$ is a convex subset of $\mathbb{R}^d$, S is a $d\times 1$ random vector, and W is a $d \times d$ symmetric random matrix. Consider the alternative program:

$$\psi^* \in \min_{\xi\in\mathcal{X}} -\xi' Z + \frac{1}{2}\xi' \Omega \xi \,,$$

where Z is a $d\times 1$ random vector and Ω is a $d\times d$ symmetric random matrix. Define the restricted eigenvalue:

$$\lambda_{\psi^*}(\Omega) = \inf_{\xi\in\mathcal{X}} \frac{(\xi-\psi^*)'\Omega(\xi-\psi^*)}{\|\xi-\psi^*\|_2^2} \,.$$

We then have that, for any $\delta > 0$:

$$\mathbb{P}[\|\psi-\psi^*\|_2>\delta] \leq \inf_{c>0,M>0,\kappa\in(0,1)} \Big\{\mathbb{P}\left[\|S-Z\|_2 \geq \kappa\frac{c}{2}\frac{\delta^2}{(\delta+M)}\right]+ \mathbb{P}\left[\|W-\Omega\|_2 \geq (1-\kappa)\frac{c}{2}\frac{\delta^2}{(\delta+M)^2}\right]+\mathbb{P}\left[\|\psi^*\|\geq M\right]+\mathbb{P}[\lambda_{\psi^*}(\Omega)\leq c]\Big\} \quad (21)$$

Proof. Let $g^*(\xi) := -\xi' Z + \frac{1}{2}\xi'\Omega\xi$. We begin by noticing that, by convexity, for any $\xi \in \mathcal{X}$, we must have that:

$$(-Z+\Omega\psi^*)'(\xi-\psi^*) \geq 0.$$

The latter implies that:

$$g^*(\xi) - g^*(\psi^*) \geq -{\psi^*}'\Omega(\xi-\psi^*) + \frac{1}{2}\xi'\Omega\xi - \frac{1}{2}\psi^{*\prime}\Omega\psi^* = \frac{1}{2}(\xi-{\psi^*}')'\Omega(\xi-\psi^*).$$

It then follows, by the definition of the restricted eigenvalue, that:

$$g^*(\xi) - g^*(\psi^*) \geq \frac{\lambda_{\theta_n^*}(\Omega_n)}{2}\|\hat{\theta}_n-\theta_n^*\|_2^2.$$

Next, proceeding similarly to the proof of Theorem 2 of Kato (2009), we obtain the following bound, for each $\delta>0$:

$$\mathbb{P}[\|\psi-\psi^*\|_2>\delta] \leq \mathbb{P}[\Delta(\delta)>\lambda_{\psi^*}(\Omega)\delta^2/2], \quad (22)$$

where $\Delta(\delta) = \sup_{\xi\in\mathcal{X}:\|\xi-\psi^*\|_2\leq\delta}|-\xi'(S-Z)+(1/2)\xi'(W-\Omega)\xi|$. Note that, for any $c>0$ and $M>0$:

$$\mathbb{P}[\|\psi-\psi^*\|_2>\delta] \leq \mathbb{P}[(M+\delta)\|S-Z\|_2+(M+\delta)^2\|\Omega-W\|_2/2 > c\delta^2/2]+\mathbb{P}[\|\psi^*\|\geq M]+\mathbb{P}[\lambda_{\psi^*}(\Omega)\leq c],$$

and, then, by the union bound, for any $\kappa \in (0,1)$

$$\mathbb{P}[\|\psi - \psi^*\|_2 > \delta] \leq \mathbb{P}\left[\|S - Z\|_2 \geq \kappa \frac{c}{2}\frac{\delta^2}{(\delta + M)}\right] + \mathbb{P}\left[\|W - \Omega\|_2 \geq (1-\kappa)\frac{c}{2}\frac{\delta^2}{(\delta + M)^2}\right] + \\ \mathbb{P}\left[\|\psi^*\| \geq M\right] + \mathbb{P}[\lambda_{\psi^*}(\Omega) \leq c].$$

Optimizing leads to the desired result. □

Lemma 2 decomposes the approximation of a quadratic minimizer by another quadratic minimizer onto four terms: the first two terms concern the approximation of the linear and quadratic terms of the target program; the third term restricts the norm of the approximating solution; and the final term limits the "restricted eigenvalue" of the approximating program. We note that the third term, which concerns the norm of the approximating solution, may be further bounded by applying Lemma 2 recursively. Specifically, one may bound the norm of the approximation ψ^* in terms of a nonstochastic approximation (where random variables are replaced by their expected values). In the next subsection, we explore this idea to prove a general result on the rates of our L-moment estimator.

A.2 Proof of Proposition 1

We apply Lemma 2 to our problem. We consider the "undersmoothed" approximation $\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})$. The approximation $\sqrt{n}(\theta^*_n - \theta_{0,n})$ is similar and therefore omitted. To apply the lemma, we define:

$$
\begin{aligned}
S = \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} (\sqrt{n}(\hat{Q}_n(u) - Q(u)) + D_n(u))\boldsymbol{P}_L(u)du\right) \\
W = 2\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right) \\
Z = \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} B_{0,n}(u)\boldsymbol{P}_L(u)du\right) \\
\Omega = 2\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)
\end{aligned} \tag{23}
$$

We will provide rates for each term. We first note that, for any $\phi \in \mathbb{R}^p$, Bessel's inequality yields:

$$
\left\|\left(\int_{\underline{p}_n}^{\overline{p}} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\phi\right\|_2^2 \leq \int_{\underline{p}_n}^{\overline{p}_n} (\boldsymbol{J}_{n,p}(u)'\phi)^2 du = \phi'\left(\int_{\underline{p}_n}^{\overline{p}_n} (\boldsymbol{J}_{n,p}(u)\boldsymbol{J}_{n,p}(u)')du\right)\phi\,.
$$

Consequently, by the definition of the spectal norm, we obtain that:

$$
\left\|\int_{\underline{p}_n}^{\overline{p}} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right\|_2 \leq \sqrt{\rho_n}\,.
$$

We also note that Bessel's inequality yields

$$
\left\|\int_{\underline{p}_n}^{\overline{p}_n} (\sqrt{n}(\hat{Q}_n(u) - Q(u)) - B_{0,n}(u))\boldsymbol{P}_l(u)du\right\|_2 \leq \|\sqrt{n}(\hat{Q}_n - Q) - B_{0,n}\|_{L^2[\underline{p}_n,\overline{p}_n]} = O_{\mathbb{P}}(r_n)\,,
$$

and

$$
\left\|\int_{\underline{p}_n}^{\overline{p}_n} D_n(u)\boldsymbol{P}_l(u)du\right\|_2 \leq \|D_n\|_{L^2[\underline{p}_n,\overline{p}_n]} = d_n\,.
$$

Combining these facts, we obtain that:

$$\|Z - S\|_2 = O_{\mathbb{P}}(\sqrt{\rho}_n(r_n \vee d_n \vee s_n))\,,$$

$$\|W - \Omega\|_2 = O_{\mathbb{P}}(\rho_n s_n)\,.$$

Combining these facts with the fact that $A_n = O_{\mathbb{P}}(1) \implies \lim_{s\to\infty} \mathbb{P}[|A_n| > s] = 0$, and applying Lemma 2 twice: first, to approximate $\sqrt{n}(\hat{\theta}_n - \theta_{0,n})$ with $\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})$; then, to approximate $\sqrt{n}(\tilde{\theta}^*_n - \theta_{0,n})$ with 0 (the minimizer of $\min_{x\in\mathcal{X}_n} x'\Omega x$); we obtain the desired result.

A.3 Proof of Proposition 2

One may prove this proposition by relying on Proposition 1. Instead, we offer a direct proof. By Weyl's inequality (Wainwright, 2019, page 241):

$$\Bigg|\lambda_{\min}\left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right) \\ -\lambda_{\min}\left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right)\Bigg| = \\ O(\rho_n s_n)$$

Consequently, under our assumptions:

$$\lim_{n\to\infty} \mathbb{P}\left[\lambda_{\min}\left(\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)\right) > 0\right] = 1\,,$$

and we may work with representation (11).

Next, we observe that, for symmetric invertible matrices A and B:

$$\|A^{-1}-B^{-1}\| = \|A^{-1}(A-B)B^{-1}\| \leq \|A^{-1}\|\|A-B\|\|B^{-1}\| \leq \|A^{-1}\|^2\|A-B\|+\|A^{-1}\|\|A-B\|^2\,,$$

where the last inequality follows from $\|B^{-1}\| \leq \|A^{-1}\| + |\|A^{-1}\| - \|B^{-1}\||$ and Weyl's inequality.

Applying these results, we obtain the desired conclusion.

A.4 Proof of Proposition 3

The first part of the proposition follows by applying Proposition 1. The second part of the proof follows from the projection characterization of (7) and Example 6 in Appendix B, by taking $\epsilon_p = (\log(p)^\nu (\rho_n s_n)^{-\gamma} p^{1/2+l})^{-1}$ and noticing that $\underline{\sigma}\sqrt{n} \leq \mathrm{tr}(\Sigma) \leq \sqrt{\|\Sigma\|_2}\sqrt{n}$, where Σ is the variance matrix of the approximation in the unconstrained setting, which can be shown to satisfy, $\|\Sigma\| = O((\rho_n s_n)^{-2\gamma})$; and $\underline{\sigma}^2$ is the lower bound on the variance.

A.5 Proof of Proposition 4

The proof is analogous to the proof of Proposition 1, in that we apply Lemma 2 to our problem. We start by defining $\hat{D}_n(u) := \theta'_{n,0}\left(\boldsymbol{J}_{n,p}(u) - \boldsymbol{J}^*_{n,p}(u)\right)$ and $D_n(u) := Q(u) - \theta'_{n,0}\boldsymbol{J}^*_{n,p}(u)$. Next, define:

$$\begin{aligned}
S = \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} (\sqrt{n}(\hat{Q}_n(u) - Q(u)) + D_n(u) + \hat{D}_n(u))\boldsymbol{P}_L(u)du\right) \\
W = 2\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' W_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right) \\
Z = \left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} B_{0,n}(u)\boldsymbol{P}_L(u)du\right) \\
\Omega = 2\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)' \Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n} \boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)
\end{aligned} \tag{24}$$

Following the same steps as in Appendix A.2, we arrive at:

$$\|Z - S\|_2 = O_{\mathbb{P}}(\sqrt{(\rho_n \vee \varepsilon_n)}(r_n \vee d_n \vee s_n \vee \xi_n)),$$

$$\|W - \Omega\|_2 = O_{\mathbb{P}}((\rho_n \vee \varepsilon_n)\, s_n)\,,$$

and the first conclusion follows similarly as in Appendix A.2. The conclusion that convergence also holds conditionally follows from Markov inequality.

B Anti-concentration inequalities on the projection onto Euclidean balls

This appendix shows how to extend Gaussian anti-concentration inequalities available in the literature to the projection of Gaussian vectors onto an Euclidean ball.

We consider $Z \sim N(0, \Sigma)$ a $d \times 1$ normal random variable. Let $X := P_{B(0;M)}(Z)$ be the projection of Z onto the Euclidean ball of radius M and centered at the origin. We consider bounding the anticoncentration function, $\mathcal{L}_X^\infty(\epsilon; \mathcal{C})$, where:

$$\mathcal{L}_X^\infty(\epsilon; \mathcal{C}) \coloneqq \sup_{C \in \mathcal{C}} \mathbb{P}[X \in C_\infty^\epsilon \setminus C_\infty^{-\epsilon}]\,.$$

We begin by noting that:

$$X = Z \mathbf{1}_{\|Z\|_2 \leq M} + M \frac{Z}{\|Z\|_2} \mathbf{1}_{\|Z\|_2 \leq M}\,.$$

Let $Z^* = M \frac{Z}{\|Z\|_2}$. We note that the law of total probability yields :

$$\mathcal{L}_X^\infty(\epsilon; \mathcal{C}) \leq \mathcal{L}_Z^\infty(\epsilon; \mathcal{C}) \wedge \mathbb{P}[\|Z\|_2 \leq M] + \underbrace{\sup_{C \in \mathcal{C}} \mathbb{P}\left[Z^* \in C_\infty^\epsilon \setminus C_\infty^{-\epsilon}, \|Z\|_2 > M\right]}_{=L^*(\epsilon)}$$

The literature provides upper bounds on the first term for specific classes of sets. We seek to bound the second term. Let $\psi^2 = \mathbb{E}[\|Z\|_2^2]$. Applying Lemma 1 with $s = \infty$ yields:

$$L^*(\epsilon) \leq \inf_{\delta \geq 0} \left\{ \mathbb{P}\left[M\|Z_j\|_\infty \left| \frac{\psi - \|Z\|_2}{\psi\|Z\|_2} \right| \geq \delta \right] + \sup_{C \in \mathcal{C}} \mathbb{P}\left[\frac{M}{\psi} Z \in \left(C_\infty^\epsilon \setminus C_\infty^{-\epsilon}\right)^\delta \setminus \left(C_\infty^\epsilon \setminus C_\infty^{-\epsilon}\right)^{-\delta} \right] \right\} \wedge \mathbb{P}[\|Z\|_2 > M$$

and we also have that, for any s:

$$\mathbb{P}\left[M\|Z_j\|_\infty \left| \frac{\psi - \|Z\|_2}{\psi\|Z\|_2} \right| \geq \delta \right] \leq \mathbb{P}\left[\|Z_j\|_\infty \geq s\right] + \mathbb{P}\left[\left| \frac{\psi - \|Z\|_2}{\psi\|Z\|_2} \right| \geq \frac{\delta}{Ms} \right].$$

We will bound both terms on the right-hand-side. Let $\bar{\sigma}^2 = \max_i \mathbb{V}[Z_i]$. First, by Lemma 2.2.2 of van der Vaart and Wellner (1996):

$$\mathbb{P}\left[\|Z\|_\infty \geq s\right] \leq 2\exp\left(-\frac{s^2}{K\log(d+1)\bar{\sigma}^2} \right).$$

Next, we note that, by concavity of $x \mapsto \sqrt{x}$:[12]

$$\sqrt{|\psi^2 - \|Z\|^2|} \geq |\psi - \|Z\|_2|$$

Consequently, we have that:

$$\mathbb{P}\left[\left| \frac{\psi - \|Z\|_2}{\psi\|Z\|_2} \right| \geq \frac{\delta}{Ms} \right] \leq \mathbb{P}\left[|\psi^2 - \|Z\|^2| > \psi^2 \|Z\|_2^2 \frac{\delta^2}{M^2 s^2} \right] \leq \mathbb{P}\left[\left|\psi^2 - \|Z\|^2\right| > \frac{\psi^2}{2} \left(\frac{\psi^2\delta^2}{Ms^2} \wedge 1 \right) \right]$$

But then, the Hanson-Wright inequality (Theorem 1.1 of Rudelson and Vershynin (2013)), yields:

$$\mathbb{P}[|\psi^2 - \|Z\|^2| > t] \leq 2\exp\left(-c\min\left\{ \frac{t^2}{\|\Sigma\|_F^2}, \frac{t}{\|\Sigma\|_2} \right\} \right) \implies$$

$$\mathbb{P}\left[\left|\psi^2 - \|Z\|^2\right| > \frac{\psi^2}{2} \left(\frac{\psi^2\delta^2}{Ms^2} \wedge 1 \right) \right] \leq 2\exp\left(-c\min\left\{ \frac{\psi^4}{4\|\Sigma\|_F^2} \left(\frac{\psi^2\delta^2}{Ms^2} \wedge 1 \right)^2, \frac{\psi^2}{2\|\Sigma\|_2} \left(\frac{\psi^2\delta^2}{Ms^2} \wedge 1 \right) \right\} \right),$$

[12] Recall that $\sqrt{a} + \sqrt{b} \geq \sqrt{a+b}$, for any $a, b \geq 0$.

But, by Von-Neumann trace inequality,

$$\|\Sigma\|_F^2 = \operatorname{tr}(\Sigma^2) \leq \|\Sigma\|_2 \operatorname{tr}(\Sigma) = \|\Sigma\|_2 \psi^2 \,.$$

Equalling the upper bounds corresponding to the tail inequalities yields:

$$\mathbb{P}\left[M\|Z_j\|_\infty \left|\frac{\psi - \|Z\|_2}{\psi\|Z\|_2}\right| \geq \delta\right] \leq 4\exp\Bigg(-\min\Bigg\{\frac{c^{1/3}\psi^2\delta^{4/3}}{K^{2/3}\log(d+1)^{2/3}\bar{\sigma}^{4/3}(4M\|\Sigma\|_2)^{1/3}}, \frac{\sqrt{c}\psi^2\delta}{K^{1/2}\log(d+1)^{1/2}\bar{\sigma}(2M\|\Sigma\|_2)^{1/2}}, \frac{c\psi^2}{4\|\Sigma\|_2}\Bigg\}\Bigg), \tag{25}$$

Combining the above leads to the following lemma.

Lemma 3. *For any class of sets $\mathcal{C} \subseteq \mathcal{B}\left(\mathbb{R}^d\right)$,*

$$\mathcal{L}_X^\infty(\epsilon;\mathcal{C}) \leq \mathcal{L}_Z^\infty(\epsilon;\mathcal{C}) \wedge \mathbb{P}[\|Z\|_2 \leq M] + \inf_{\delta\geq 0}\left\{\overline{U}(\delta) + \sup_{C\in\mathcal{C}}\mathbb{P}\left[\frac{M}{\psi}Z \in \left(C_\infty^\epsilon \setminus C_\infty^{-\epsilon}\right)^\delta \setminus \left(C_\infty^\epsilon \setminus C_\infty^{-\epsilon}\right)^{-\delta}\right]\right\} \wedge \mathbb{P}[\|Z\|_2 > M]\,, \tag{26}$$

where

$$\overline{U}(\delta) = 4\exp\left(-\min\left\{\frac{c^{1/3}\psi^2\delta^{4/3}}{K^{2/3}\log(d+1)^{2/3}\bar{\sigma}^{4/3}(4M\|\Sigma\|_2)^{1/3}}, \frac{\sqrt{c}\psi^2\delta}{K^{1/2}\log(d+1)^{1/2}\bar{\sigma}(2M\|\Sigma\|_2)^{1/2}}, \frac{c\psi^2}{4\|\Sigma\|_2}\right\}\right).$$

The previous lemma offers an upper bound to anticoncentration in the sup-norm in terms of estimable quantities. The second term inside the infimum is itself an anticoncentration inequality of a rescaled Gaussian vector. In what follows, we illustrate our lemma by providing anticoncentration rates in the class of hyperrectangles.

Example 6 (Hyperrectangles)**.** Consider the class of hyperrectangles $\mathcal{C}_d$. In this case, Nazarov's inequality reveals that:

$$\sup_{C\in\mathcal{C}_d}\mathbb{P}\left[\frac{M}{\psi}Z\in\left(C_\infty^{\epsilon}\setminus C_\infty^{-\epsilon}\right)^{\delta}\setminus\left(C_\infty^{\epsilon}\setminus C_\infty^{-\epsilon}\right)^{-\delta}\right]\leq\frac{\psi(\epsilon+\delta)}{\underline{\sigma}M}(\sqrt{2\log(2d)}+2)\,.$$

Take $\epsilon=\delta$. We can then use Lemma 3 to find a sequence ϵ_d such that $\mathcal{L}_X^{\infty}(\epsilon_d;\mathcal{C})\to 0$.

Remark 9. It is possible to adapt the proof of Lemma 3 to bound $\mathcal{L}_X^{2}(\epsilon;\mathcal{C})$. In this case, one replaces the maximal inequality that bounds $\mathbb{P}[\|Z\|_\infty\geq\delta]$ with Proposition 1 of Hsu et al. (2012), which provides an upper bound to $\mathbb{P}[\|Z\|_2\geq\delta]$.

C On the choice of tuning sequence γ_n

Define the following terms:

$$\begin{aligned}Z_n&=\left(\int_{\underline{p}_n}^{\overline{p}_n}\boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)'\Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n}B_{0,n}(u)\boldsymbol{P}_L(u)du\right)\\ \Omega_n&=2\left(\int_{\underline{p}_n}^{\overline{p}_n}\boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right)'\Omega_{n,L}\left(\int_{\underline{p}_n}^{\overline{p}_n}\boldsymbol{P}_L(u)\boldsymbol{J}_{n,p}(u)'du\right).\end{aligned}\tag{27}$$

For a given sequence $(\gamma_n)_{n\in\mathbb{N}}$, the bootstrap estimator is defined as,

$$\theta_n^B\in\operatorname{arginf}_{\theta\in\Theta_n}-\theta'Z_n+\frac{\gamma_n}{2}\left(\theta-\hat{\theta}\right)'\Omega_n\left(\theta-\hat{\theta}\right)\,.\tag{28}$$

We consider the problem of approximating $\sqrt{n}(\theta_n^*-\theta_{0,n})$ with $\gamma_n(\theta_n^B-\hat{\theta}_n)$. We start with the strong approximation, note that, by the triangle inequality

$$\|\gamma_n(\theta_n^B-\hat{\theta}_n)-\sqrt{n}(\theta_n^*-\theta_{0,n})\|\leq\left\|\frac{\gamma_n\sqrt{n}(\hat{\theta}_n-\theta_{0,n})}{\sqrt{n}}\right\|+\|\gamma_n(\theta_n^B-\theta_{0,n})-\sqrt{n}(\theta_n^*-\theta_{0,n})\|.$$

Following the proof of Proposition 1, we have

$$\left\|\frac{\gamma_n\sqrt{n}(\hat{\theta}_n-\theta_{0,n})}{\sqrt{n}}\right\|=O_{\mathbb{P}}\left(\frac{\gamma_n}{\sqrt{n}}\left(\frac{\sqrt{\rho_n}}{\lambda_{0,n}}\vee\xi_n\right)\right),$$

where $\|\sqrt{n}(\hat{\theta}_n - \theta_{0,n}) - \sqrt{n}(\theta_n^* - \theta_{0,n})\| = O_{\mathbb{P}}(\xi_n)$. Now define the event:

$$A_{\gamma_n} = \{\sqrt{n}(\theta_n^* - \theta_{0,n}) \in \gamma_n(\Theta_n - \theta_0)\} \cup \{\gamma_n(\theta_n^B - \theta_{0,n}) \in \sqrt{n}(\Theta_n - \theta_0)\},$$

we can decompose, for $\delta > 0$:

$$\mathbb{P}(\|\gamma_n(\theta_n^B - \theta_{0,n}) - \sqrt{n}(\theta_n^* - \theta_{0,n})\| > \delta) \leq \mathbb{P}(\|\gamma_n(\theta_n^B - \theta_{0,n}) - \sqrt{n}(\theta_n^* - \theta_{0,n})\| > \delta | A_{\gamma_n}) + (1 - \mathbb{P}(A_{\gamma_n}))$$

Conditional on the event A_{γ_n}, by Lemma 2 we have:

$$\|\gamma_n(\theta_n^B - \theta_{0,n}) - \sqrt{n}(\theta_n^* - \theta_{0,n})\| = O_{\mathbb{P}}\left(\frac{\gamma_n}{\sqrt{n}}\sqrt{\rho_n}\left(\frac{\sqrt{\rho_n}}{\lambda_{0,n}} \vee \xi_n\right)\right),$$

which implies that, conditional on A_{γ_n},

$$\|\gamma_n(\theta_n^B - \hat{\theta}_n) - \sqrt{n}(\theta_n^* - \theta_{0,n})\| = O_{\mathbb{P}}\left(\frac{\gamma_n}{\sqrt{n}}\sqrt{\rho_n}\left(\frac{\sqrt{\rho_n}}{\lambda_{0,n}} \vee \xi_n\right)\right).$$

And we conclude that for bootstrap validity we need the additional condition that we can find a sequence γ_n, such that

$$\frac{\gamma_n}{\sqrt{n}}\sqrt{\rho_n}\left(\frac{\sqrt{\rho_n}}{\lambda_{0,n}} \vee \xi_n\right) \to 0,$$

and

$$\mathbb{P}(A_{\gamma_n}) \to 1\,.$$

D Approximation through nonnegative weights

In this appendix, we show that any nonnegative random variable whose quantile function satisfies Assumption 7 can be arbitrarily well approximated (in the L^2-norm) by a mixture of "truncated" Extreme Value quantiles with an underlying nonnegative measure.

D.1 Setup

We consider the problem of approximating a quantile function Q by "truncations" of quantile functions given by

$$Q_\theta(u) = \frac{1}{\theta}(-\log u)^{-1}, \quad \theta \in (0, \infty)$$

The quantile functions presented above represent the quantiles of random variables that conform to a generalized extreme value distribution, with associated scale parameter given by $\sigma = \frac{1}{\theta}$, shape parameter given by $\xi = 1$ and scale given by $\mu = \frac{1}{\theta}$.

Assumption 7. *We make additional assumptions about* Q

1. $Q(0) \geq 0$.
2. $Q \in L^2[0,1]$.
3. Q *is differentiable and strictly increasing in* $(0,1)$.

D.2 Approximation Result

Lemma 4. *Let* Q *be a quantile function satisfying Assumption 7 and let,*

$$\alpha_{\theta,k}(u) := \frac{1}{\theta}(-\log u)^{-1}\mathbf{1}\left[\theta \geq \frac{u}{k}\right]\mathbf{1}\left[\theta \leq \frac{u^{\left(1-\frac{C}{k}\right)}}{k}\right],$$

for some $C > 0$. *There exists a sequence of nonnegative measure* $\{\mu_k^+\}_{k\in\mathbb{N}}$ *on* $(\mathbb{R}_+, \mathcal{B}[\mathbb{R}_+])$*, such that,*

$$\left\| Q - \int_{\mathbb{R}_{++}} \alpha_{\theta,k}\mu_k^+(d\theta) \right\|_{L^2[0,1]} \to 0,$$

as $k \to \infty$. *Moreover, these mixtures are themselves quantiles.*

Proof. Fix $C > 0$. Define the following functions:

$$\alpha(m) := \frac{1}{m}$$

$$h(m,x) := m^{-\log x}$$

$$g(m,k) := (em)^k$$

$$h^{-1}(z,x) := z^{-\frac{1}{\log x}}.$$

Here $h(e^{-1},x) = e^{\log x} = x$, then we define the identity element $\eta := e^{-1}$. Define the sequence $\{m_k\}_{k\in\mathbb{N}}$, $m_k = \exp\left(\frac{C}{k} - 1\right)$ and define a sequence of functions indexed by $k \in \mathbb{N}$,

$$\alpha_k(m) = \frac{dg(m,k)}{dm}\alpha(m) = \frac{k}{m}.$$

We claim that $\alpha_k(m)$ is an approximate identity. First,

$$\int_\eta^{m_k} \frac{k}{m} = k\left(\log\left(\exp\left(\frac{C}{k} - 1\right)\right) + 1\right) = C$$

. Moreover, for every $\delta > 0$,

$$\lim_{k\to\infty} \int_{\eta+\delta}^{\infty} k\mathbf{1}\left(e^{-1} + \delta < \exp\left(\frac{C}{k} - 1\right)\right) dm = 0$$

then we can apply Lemma 5 to conclude that $\alpha_k(m)$ is a C-approximate identity. Defining

$$\left[\alpha_k * \frac{Q}{C}\right](u) := \int_\eta^{m_k} \alpha_k(m)\frac{Q(h(m,u))}{C}dm,$$

we can apply Lemma 5 to conclude

$$\left\|\alpha_k * \frac{Q}{C} - Q\right\|_{L^2[0,1]} \to 0, \qquad \text{as } k \to \infty$$

Next, we show that the function $\left[\alpha_k * \frac{Q}{C}\right]$ is a quantile function. We can rewrite:

$$\left[\alpha_k * \frac{Q}{C}\right](u) = \int_\eta^{m_k} \alpha_k(m)\frac{Q(h(m,u))}{C}dm = \int_u^{u^{\left(1-\frac{C}{K}\right)}} \frac{dh^{-1}(z,u)}{dz}\alpha_k(h^{-1}(z,u))\frac{Q(z)}{C}dz =$$

$$= \int_u^{u^{\left(1-\frac{C}{K}\right)}} \frac{1}{-z\log u} z^{-\frac{1}{\log u}} \frac{k}{z^{-\frac{1}{\log u}}} \frac{Q(z)}{C} dz = \int_u^{u^{\left(1-\frac{C}{K}\right)}} \frac{k}{z}(-\log u)^{-1}\frac{Q(z)}{C}dz,$$

where the second equality follows from the change of variables formula of the Riemann Integral by letting $z := h(m,x)$. Since Q is nonnegative, $\left[\alpha_k * \frac{Q}{C}\right]$ is non negative as well. Taking the derivative with respect to u by applying the Leibnitz Rule and simplifying we get:

$$\frac{d}{du}\left[\alpha_k * \frac{Q}{C}\right](u) = \frac{k}{C}\frac{1}{u(\log u)^2}\left(\int_u^{u^{\left(1-\frac{C}{K}\right)}} \frac{Q(z)}{z}dz + (-\log u)\left(Q\left(u^{\left(1-\frac{C}{K}\right)}\right) - Q(u)\right)\right),$$

further integrating by parts the term $\int_u^{u^{\left(1-\frac{C}{K}\right)}} \frac{Q(z)}{z}dz$, we conclude

$$\frac{d}{du}\left[\alpha_k * \frac{Q}{C}\right](u) \geq 0 \iff \int_u^{u^{\left(1-\frac{C}{K}\right)}} (-\log z)Q'(z)dz \geq 0,$$

which is true by Assumption 7. Finally, we can get the desired result, by defining $\theta := \frac{z}{k}$ and applying the change of variables formula again to conclude:

$$\left[\alpha_k * \frac{Q}{C}\right](u) = \int_0^\infty \frac{1}{\theta}(-\log u)^{-1}\mathbf{1}\left[\theta \geq \frac{u}{k}\right]\mathbf{1}\left[\theta \leq \frac{u^{\left(1-\frac{C}{k}\right)}}{k}\right]\frac{kQ(\theta k)}{C}d\theta$$

□

Remark 10. The result goes through if we replace $[0,1]$ by any closed interval $\mathcal{X} \subseteq [0,1]$ and/or $Q(0) \geq 0$ by $Q(1) \leq 1$ in Assumption 7. In this case, the approximation is with respect to the norm of $L^2(\mathcal{X})$.

D.3 Auxiliary Definitions and Lemmas

We follow the idea from Nguyen and McLachlan (2019) and Nguyen et al. (2020), to construct convolution-based approximations. Because we are working with mixture of quantiles instead of mixture of densities we have to adapt the definitions and proofs.

D.3.1 Approximate Identity Definition

Let $\{\mathcal{M}_k\}_{k\in\mathbb{N}}$ be a sequence of sets, where for all k, $\mathcal{M}_k \subset \mathbb{R}_+$ and let $h_k : \mathcal{M}_k \times \mathbb{R}_+ \to \mathbb{R}_+$, be a family of functions indexed by k such that, there exists $\eta \in \mathbb{R}_+$ with $h_k(x,\eta) = x$, for all $x \in \mathbb{R}_+$. We define the modified convolution operator $*$, such that

$$[f * g]_k(x) = \int_{\mathcal{M}_k} f(m) g\left(h_k(m,x)\right) dm.$$

When there is no ambiguity We will drop the subscript k of the left-hand side.

Definition 4. Let $k \in \mathbb{N}$, and k^* be a limit point of $\mathbb{N}$. A family of functions $\{\alpha_k\}_{k\in\mathbb{N}}$ defined on $\mathbb{R}_+$ is called a C-approximate identity if

1. $\int_{\mathcal{M}_K} |\alpha_k(m)dm| \leq \tilde{C}$, for all $k \in \mathbb{R}_+$
2. $\int_{\mathcal{M}_k} \alpha_k(m)dm = C$, for all $k \in \mathbb{R}_+$, where $C \neq 0$ is a constant.
3. $\int_{\mathcal{M}_k/B_\eta(\delta)} |\alpha_k(m)|dm \underset{k\to k^*}{\longrightarrow} 0$ for every $\delta > 0$.
4. $\eta \in \mathcal{M}_k$, for all $k \in \mathbb{R}_+$

where $B_\eta(\delta)$ denotes the ball centered at η with radius δ.

Next, we provide a useful lemma for constructing approximate identities.

Lemma 5. *Let $g : \mathbb{R}_+ \times \mathbb{N} \to \mathbb{R}_+$, and $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$, if:*

1. *$\frac{dg(m,k)}{dm}$ exists and is positive almost everywhere,*
2. *There exists a sequence $\{m_k\}_{k\in\mathbb{N}}$, such that, for every k, $\mathcal{M}_k = [\eta, m_k]$, $m_{k+1} \leq m_k$, $\lim_{k\to\infty} m_k = \eta$,*

$$\int_\eta^{m_k} \frac{dg(m,k)}{dm} \alpha(g(m,k))dm = C$$

for some $C \neq 0$ and, for every $\delta > 0$,

$$\lim_{k\to\infty} \int_{\eta+\delta}^{\infty} \frac{dg(m,k)}{dm} \alpha(g(m,k)) \mathbf{1}\left(n + \delta < m_k\right) dm = 0$$

then

$$\alpha_k(x) := \frac{dg(m,k)}{dm}\alpha(g(m,k))$$

is a $C-$approximate identity with $k^* = \infty$.

Proof. The proof is direct by verifying that the definition holds.

$\square$

D.3.2 Generic Approximation

Lemma 6. *Let α_k be a C-approximate identity for some $k^* \in [0, \infty]$. If $\mathcal{X} \subseteq \mathbb{R}_+$, $f \in \mathcal{L}_2(\mathcal{X})$, and, for all $k \in \mathbb{N}$, $f(h_k(m,x))$ is continuous in m and uniformly continuous in a neighborhood of η, then $\left\|\alpha_k * \frac{f}{C} - f\right\|_2 \to 0$ as $k \to k^*$.*

Proof.

$$\begin{aligned}
&\left|\left[\alpha_k * \frac{f}{C}\right](x) - f(x)\right| = \left|\int_{\mathcal{M}_k} \alpha_k(m)\frac{f(h_k(m,x))}{C}dm - f(x)\right| \\
&= \frac{1}{|C|}\left|\int_{\mathcal{M}_k} \alpha_k(m)[f(h_k(m,x)) - f(x)]dm\right| \quad \text{(by 2.)} \\
&\leq \frac{1}{|C|}\int_{\mathcal{M}_k} |\alpha_k(m)[f(h_k(m,x)) - f(x)]|\, dm \\
&= \frac{1}{|C|}\int_{\mathcal{M}_k} |\alpha_k(m)|^{\frac{1}{2}}|[f(h_k(m,x)) - f(x)]||\alpha_k(m)|^{\frac{1}{2}}dm \\
&\leq \frac{1}{|C|}\left(\int_{\mathcal{M}_k} |\alpha_k(m)|[f(h_k(m,x)) - f(x)]^2 dm\right)^{\frac{1}{2}}\left(\int_{\mathcal{M}_k} |\alpha_k(m)|dm\right)^{\frac{1}{2}} \quad \text{(by C-S)} \\
&\leq \frac{1}{\sqrt{|C|}}\left(\int_{\mathcal{M}_k} |\alpha_k(m)|[f(h_k(m,x)) - f(x)]^2 dm\right)^{\frac{1}{2}} \quad \text{(by 1.).}
\end{aligned}$$

We can bound the L_2 norm:

$$\begin{aligned}
&\int_{\mathcal{X}}\left|\left[\alpha_k * \frac{f}{K}\right](x) - f(x)\right|^2 dx \leq \frac{1}{\sqrt{|C|}}\int_{\mathcal{X}}\int_{\mathcal{M}_k} |\alpha_k(m)|[f(h_k(m,x)) - f(x)]^2 dm dx \\
&= \frac{1}{\sqrt{|C|}}\int_{\mathcal{M}_k} |\alpha_k(m)|\left(\int_{\mathcal{X}}[f(h_k(m,x)) - f(x)]^2 dx\right) dm \quad \text{(Fubinis).}
\end{aligned}$$

Note that we we have that as $m \to \eta$,

$$\lim_{m\to\eta}\left(\int_{\mathcal{X}}[f(h_k(m,x)) - f(x)]^2 dx\right) = 0 \quad \text{(By continuity of } f(h_k(m,x)) \text{ with respect to } m\text{).}$$

Define $g_k(m) \equiv \left(\int_{\mathcal{X}}[f(h_k(m,x)) - f(x)]^2 dx\right)$, then $g(\eta) = 0$. Moreover,

$$\begin{aligned}
&\int_{\mathcal{M}_k}|\alpha_k(m)|g_k(m)dm \le \int_{\mathcal{M}_k}|\alpha_k(m)||g_k(m)|dm \\
&= \int_{B_\eta(\delta)}|\alpha_k(m)||g_k(m)|dm + \int_{\mathcal{M}_k/B_\eta(\delta)}|\alpha_k(m)||g_k(m)|dm \\
&\le \sup_{B_\eta(\delta)}|g_k(m)|\int_{B_\eta(\delta)}|\alpha_k(m)|dm + \sup_{\mathcal{M}_k/B_{\eta(\delta)}}|g_k(m)|\int_{\mathcal{M}_k/B_\eta(\delta)}|\alpha_k(m)|dm
\end{aligned}$$

Since we can bound $|g_k(m)| \le M$, taking the limit we get

$$\lim_{k\to k^*}\int_{\mathcal{M}_k}|\alpha_k(m)|g_k(m)dm \le C \lim_{k\to k^*}\sup_{B_\eta(\delta)}|g_k(m)| \quad \text{(By 1. and 2.)}$$

Since $g(\eta) = 0$, $f(h_k(m,x))$ is continuous in m we can pick delta such that the right-hand side is close enough to 0. Then we conclude that

$$\lim_{k\to k^*}\int_{\mathcal{X}}\left|\left[\alpha_k * \frac{f}{C}\right](x) - f(x)\right|^2 dx = 0$$

□

E Empirical Application

E.1 Data Source

Our data source is the publicly available version of the employer-employee matched data RAIS from 2017 to 2021. A standardized version is available to download at https://basedosdados.org/datas 96ba-448e-b053-d385a829ef00?table=dabe5ea8-3bb5-4a3e-9d5a-3c7003cd4a60.

E.2 Data Cleaning

The construction of the empirical quantile function of the average monthly wages for private sector workers in each municipality $\times$ year proceeds as follows:

1. We exclude entries with 0 average monthly wage (column valor_remuneracao_media equal 0).
2. We exclude entries where the employee is hired as a public servant (column tipo_vinculo equal 2, 30, 31 or 35).
3. We exclude entries where the employer sector is either public administration, defense, social security, or international organizations. (first two digits of column cnae_2 equal 84 or 99).