Journal of Econometrics 252 (2025) 106101

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Inference on model parameters with many L-moments

Luis A.F. Alvarez [a,*], Chang Chiann [b], Pedro A. Morettin [b]

[a] Department of Economics, University of São Paulo, Brazil
[b] Department of Statistics, University of São Paulo, Brazil

ARTICLE INFO

Keywords:
L-statistics
Quantiles
Generalised method of moments
Tuning parameter selection methods
Higher-order expansions

ABSTRACT

This paper studies parameter estimation using L-moments, an alternative to traditional moments with attractive statistical properties. The estimation of model parameters by matching sample L-moments is known to outperform maximum likelihood estimation (MLE) in small samples from popular distributions. The choice of the number of L-moments used in estimation remains *ad-hoc*, though: researchers typically set the number of L-moments equal to the number of parameters, which is inefficient in larger samples. In this paper, we show that, by properly choosing the number of L-moments and weighting these accordingly, one is able to construct an estimator that outperforms MLE in finite samples, and yet retains asymptotic efficiency. We do so by introducing a generalised method of L-moments estimator and deriving its properties in an asymptotic framework where the number of L-moments varies with sample size. We then propose methods to automatically select the number of L-moments in a sample. Monte Carlo evidence shows our approach can provide mean-squared-error improvements over MLE in smaller samples, whilst working as well as it in larger samples. We consider extensions of our approach to the estimation of conditional models and a class semiparametric models. We apply the latter to study expenditure patterns in a ridesharing platform in Brazil.

1. Introduction

L-moments, expected values of linear combinations of order statistics, were introduced by Hosking (1990) and have been successfully applied in areas as diverse as computer science (Hosking, 2007; Yang et al., 2021), hydrology (Wang, 1997; Sankarasubramanian and Srinivasan, 1999; Das, 2021; Boulange et al., 2021), meteorology (Wang and Hutson, 2013; Šimková, 2017; Li et al., 2021b) and finance (Gourieroux and Jasiak, 2008; Kerstens et al., 2011). By appropriately combining order statistics, L-moments offer robust alternatives to traditional measures of dispersion, skewness and kurtosis. Models fit by matching sample L-moments (a procedure labelled "method of L-moments" by Hosking, 1990) have been shown to outperform maximum likelihood estimators in small samples from flexible distributions such as generalised extreme value (Hosking et al., 1985; Hosking, 1990), generalised Pareto (Hosking and Wallis, 1987; Broniatowski and Decurninge, 2016), generalised exponential (Gupta and Kundu, 2001) and Kumaraswamy (Dey et al., 2018).

Statistical analyses of L-moment-based parameter estimators rely on a framework where the number of moments is fixed (Hosking, 1990; Broniatowski and Decurninge, 2016). Practitioners often choose the number of L-moments equal to the number of parameters in the model, so as to achieve the order condition for identification. This approach is generally inefficient.[1] It also raises

* Corresponding author.
E-mail addresses: luis.alvarez@usp.br (L.A.F. Alvarez), chang@ime.usp.br (C. Chiann), pam@ime.usp.br (P.A. Morettin).

[1] In the generalised extreme value distribution, there can be asymptotic root mean-squared error losses of 30% with respect to the MLE when the target estimand are the distribution parameters (Hosking et al., 1985; Hosking, 1990). In our Monte Carlo exercise, we verify root mean squared error losses of over 10% when the goal is tail quantile estimation.

https://doi.org/10.1016/j.jeconom.2025.106101
Received 16 July 2024; Received in revised form 3 May 2025; Accepted 7 September 2025
Available online 23 September 2025

the question of whether overidentifying restrictions, together with the optimal weighting of L-moment conditions, could improve the efficiency of "method of L-moments" estimators, as in the framework of generalised-method-of-moment (GMM) estimation (Hansen, 1982). Another natural question would be how to choose the number of L-moments in finite samples, as it is well-known from GMM theory that increasing the number of moments with a fixed sample size can lead to substantial biases (Koenker and Machado, 1999; Newey and Smith, 2004). In the end, one can only ask if, by correctly choosing the number of L-moments and under an appropriate weighting scheme, it may not be possible to construct an estimator that outperforms maximum likelihood estimation of some target quantities in small samples from popular distributions and yet achieves the Cramér–Rao bound asymptotically. Intuitively, the answer appears to be positive, especially if one takes into account that Hosking (1990) shows L-moments characterise distributions with finite first moments.

The goal of this paper lies in answering the questions outlined in the previous paragraph. Specifically, we propose to study L-moment-based estimation in a context where: (i) the number of L-moments varies with sample size; and (ii) weighting is used in order to optimally account for overidentifying conditions. In this framework, we introduce a "generalised" method of L-moments estimator and analyse its properties. We provide sufficient conditions under which our estimator is consistent and asymptotically normal; we also derive the optimal choice of weights and introduce a test of overidentifying restrictions. We then show that, under independent and identically distributed (iid) data and the optimal weighting scheme, the proposed generalised L-moment estimator achieves the Cramér–Rao lower bound. We provide simulation evidence that our L-moment approach outperforms (in a mean-squared error sense) MLE estimation of some target quantities in smaller samples from popular distributions; while working as well as the MLE in larger sample sizes. We then construct methods to automatically select the number of L-moments used in estimation. For that, we rely on higher order expansions of the method-of-L-moment estimator, similarly to the procedure of Donald and Newey (2001) and Donald et al. (2009) in the context of GMM. We use these expansions to find a rule for choosing the number of L-moments so as to minimise the estimated (higher-order) mean-squared incurred when targeting a given transformation of the model parameters of interest. We also consider an approach based on ℓ_1-regularisation (Luo et al., 2015). We provide computational code to implement both methods,[2] and evaluate their performance through Monte Carlo simulations. With these tools, we aim to introduce a fully automated procedure for estimating parametric models that is able to improve upon maximum likelihood estimation in small samples from popular distributions, and yet has the *guarantee* of not underperforming in larger datasets. As our examples and simulations throughout the paper indicate, our approach seems especially useful in tail quantile estimation of heavy-tailed distributions.[3]

We also consider two extensions of our main approach. First, we show how the generalised method-of-L-moment approach introduced in this paper can be extended to the estimation of conditional models. Second, we show how our approach may be used in the analysis of the "error term" in semiparametric models, an important task in specification testing and the construction of prediction bands. We apply the latter extension to study the tail behaviour of expenditure patterns in a ridesharing platform in São Paulo, Brazil. We provide evidence that the heavy-tailedness in consumption patterns persists even after partialing out the effect of unobserved time-invariant heterogeneity and observable heterogeneity in consumption trends. We also show how our estimators can be used to construct prediction bands for individual treatment effects when one is interested in causal inference on individualised interventions. With these extensions, we hope more generally to illustrate how the generalised-method-of-L-moment approach to estimation may be a convenient tool in a variety of settings, e.g. when a model's quantile function is easier to evaluate than its likelihood. The latter feature has been explored in followup work by one of the authors in semi- and nonparametric settings (Alvarez and Orestes, 2023; Alvarez and Biderman, 2024).

Related literature. This paper contributes to two main literatures. First, we relate to a couple of papers that, building on Hosking's original approach, propose new L-moment-based estimators. Gourieroux and Jasiak (2008) introduce a notion of L-moment for conditional moments, which is then used to construct a GMM estimator for a class of dynamic quantile models. As we argue in more detail in Section 6, while conceptually attractive, their estimator is not asymptotically efficient (vis-à-vis the conditional MLE), as it focuses on a finite number of moment conditions and does not optimally explore the set of overidentifying restrictions available in the parametric model. In contrast, our proposed extension of the generalised method-of-L-moment estimator to conditional models is able to restore asymptotic efficiency. In an unconditional setting, Broniatowski and Decurninge (2016) propose estimating distribution functions by relying on a *fixed* number of L-moments and a minimum divergence estimator that nests the empirical likelihood and generalised empirical likelihood estimators as particular cases. Even though these estimators are expected to perform better than (generalised) method-of-L-moment estimators in terms of higher-order properties (Newey and Smith, 2004), both would be *first-order* inefficient (vis-à-vis the MLE) when the number of L-moments is held fixed. In this paper, we thus focus on improving L-moment-based estimation in terms of first-order asymptotic efficiency, by suitably increasing the number of L-moments with sample size and optimally weighting the overidentifying restrictions, while retaining its known good finite-sample behaviour. We do note, however, that one of our suggested approaches to select the number of L-moments aims at minimising an estimate of the higher-order mean-squared error, which may be useful in improving the higher-order behaviour of estimators even when a bounded (as a function of sample sizes) number of L-moments is used in estimation.

[2] The repository https://github.com/luisfantozzialvarez/lmoments_redux contains R script that implements our main methods, as well as replication code for our Monte Carlo exercise and empirical application.

[3] More generally, the question of whether our approach will generate significant small sample root mean squared error gains over the MLE for a given family of distributions and transformation of the model parameters of interest must be answered on a case-by-case basis. We provide tools to automatically select the number of L-moments used in estimation based on higher-order expansions of the mean-squared error that may be applied to any estimation target that is a smooth transformation of the model parameters. Moreover, our asymptotic efficiency result ensures that, asymptotically, there will be no losses in adopting this approach vis-à-vis the MLE.

Secondly, we contribute to a literature that seeks to construct estimators that, while retaining asymptotic (first-order) unbiasedness and efficiency, improve upon maximum likelihood estimation in finite samples. The classical method to achieve finite-sample improvements over the MLE is through (higher-order) bias correction (Pfanzagl and Wefelmeyer, 1978). However, analytical bias corrections may be difficult to implement in practice, which has led the literature to consider jackknife and bootstrap corrections (Hahn et al., 2002). More recently, Ferrari and Yang (2010) introduced a maximum Lq-likelihood estimator for parametric models that replaces the log-density in the objective function of the MLE with $\frac{f(x)^{1-q}-1}{1-q}$, where $q > 0$ is a tuning parameter. They show that, by suitably choosing q in finite samples, one is able to trade-off bias and variance, thus enabling MSE improvements over the MLE. Moreover, if $q \to 1$ asymptotically at a rate, the estimator is asymptotically unbiased and achieves the Crámer–Rao lower bound. There are some important differences between our approach and maximum Lq-likelihood estimation, though. First, we note that the theoretical justification for our construction is distinct from their method. Indeed, for a fixed number of L-moments, our proposed estimator is first-order asymptotically unbiased, whereas the maximum Lq-likelihood estimator is inconsistent in an asymptotic regime with q fixed and consistent but first-order biased if $q \to 1$ slowly enough. Therefore, whereas the choice of the tuning parameter q is justified as capturing a tradeoff between first-order bias and variance; the MSE-optimal choice of L-moments in our setting concerns a tradeoff between the first-order variance of the estimator and its higher-order terms. This is precisely what we capture in our proposal to select the number of L-moments by minimising an estimator of the higher-order MSE; whereas presently no general rule for choosing the tuning parameter $q > 0$ in maximum Lq-likelihood estimation exists (Li et al., 2021a).

Structure of paper. The remainder of this paper is organised as follows. In the next section, we briefly review L-moments and parameter estimation based on these quantities. Section 3 works out the asymptotic properties of our proposed estimator. In Section 4 we conduct a small Monte Carlo exercise which showcases the gains associated with our approach. Section 5 proposes methods to select the number of L-moments and assesses their properties in the context of the Monte Carlo exercise of Section 4. Section 6 presents the extensions of our main approach, as well as the empirical application. Section 7 concludes. The Supplemental Appendix presents the proofs of the main results in the paper, as well as additional details on the methods of selection of L-moments, and the extensions to "residual analysis" and conditional models.

2. L-moments: definition and estimation

Consider a scalar random variable Y with distribution function F and finite first moment. For $r \in \mathbb{N}$, Hosking (1990) defines the rth L-moment as:

$$\lambda_r := \int_0^1 Q_Y(u) P^*_{r-1}(u) du\,, \tag{1}$$

where $Q_Y(u) := \inf\{y \in \mathbb{R} : F(y) \geq u\}$ is the quantile function of Y, and the functions $P^*_r(u) = \sum_{k=0}^r (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} u^k$, $r \in \{0\} \cup \mathbb{N}$, are shifted Legendre polynomials.[4] Expanding the polynomials and using the quantile representation of a random variable (Billingsley, 2012, Theorem 14.1), we arrive at the equivalent expression:

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathbb{E}[\tilde{Y}_{(r-k):r}]\,, \tag{2}$$

where, $\tilde{Y}_{j:l}$ is the jth order statistic of a random sample from F with l observations. Eq. (2) motivates our description of L-moments as the expected value of linear combinations of order statistics. Notice that the first L-moment corresponds to the expected value of Y.

To see how L-moments may offer "robust" alternatives to conventional moments, it is instructive to consider, as in Hosking (1990), the second L-moment. In this case, we have:

$$\lambda_2 = \frac{1}{2}\mathbb{E}[\tilde{Y}_{2:2} - \tilde{Y}_{1:2}] = \frac{1}{2} \iint \left(\max\{y_1, y_2\} - \min\{y_1, y_2\}\right) F(dy_1) F(dy_2) = \frac{1}{2}\mathbb{E}|\tilde{Y}_1 - \tilde{Y}_2|\,,$$

where $\tilde{Y}_1$ and $\tilde{Y}_2$ are independent copies of Y. This is a measure of dispersion. Indeed, comparing it with the variance, we have:

$$\mathbb{V}[Y] = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \frac{1}{2}\mathbb{E}[(\tilde{Y}_1 - \tilde{Y}_2)^2]\,,$$

from which we note that the variance puts more weight to larger differences.

Next, we discuss sample estimators of L-moments. Let $Y_1, Y_2 \ldots Y_T$ be an identically distributed sample of T observations, where each Y_t, $t = 1, \ldots, T$, is distributed according to F. A natural estimator of the rth L-moment is the sample analog of (1), i.e.

$$\hat{\lambda}_r = \int_0^1 \hat{Q}_Y(u) P^*_{r-1}(u) du\,, \tag{3}$$

where $\hat{Q}_Y$ is the left-continuous (*càglàd*) empirical quantile process:

$$\hat{Q}_Y(u) = Y_{i:T}, \quad \text{if} \quad \frac{i-1}{T} < u \leq \frac{i}{T}\,, \tag{4}$$

[4] Legendre polynomials are defined by applying the Gram–Schmidt orthogonalisation process to the polynomials $1, x, x^2, x^3 \ldots$ defined on $[-1, 1]$ (Kreyszig, 1989, p. 176–180). If P_r denotes the rth Legendre polynomial, shifted Legendre polynomials are related to the standard ones through the affine transformation $P^*_r(u) = P_r(2u - 1)$ (Hosking, 1990).

with $Y_{i:T}$ being the ith sample order statistic. The estimator given by (3) is generally biased (Hosking, 1990; Broniatowski and Decurninge, 2016). When observations $Y_1, Y_2, \dots, Y_T$ may be assumed to be independent, researchers thus often resort to an unbiased estimator of λ_r, which is given by an empirical analog of (2):

$$\tilde{\lambda}_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \binom{T}{r}^{-1} \sum_{1 \le i_1, i_2 \le \cdots \le i_r \le T} Y_{i_{r-k}:T} \,. \tag{5}$$

In practice, it is not necessary to iterate over all size r subsamples of $Y_1, \dots, Y_T$ to compute the sample rth L-moment through (5). Hosking (1990) provides a direct formula that avoids such computation.

We are now ready to discuss the estimation of parametric models based on matching L-moments. Suppose that F belongs to a parametric family of distribution functions $\{F_\theta : \theta \in \Theta\}$, where $\Theta \subseteq \mathbb{R}^d$ and $F = F_{\theta_0}$ for some $\theta_0 \in \Theta$. Let $l_r(\theta) := \int_0^1 P^*_{r-1}(u) Q(u|\theta) du$ denote the theoretical rth L-moment, where $Q(\cdot|\theta)$ is the quantile function associated with F_θ. Let $H^R(\theta) := (\lambda_1(\theta), \lambda_2(\theta), \dots, \lambda_R(\theta))'$, and $\hat{H}^R$ be the vector stacking estimators for the first R L-moments (e.g. (3) or (5)). Researchers then usually estimate θ_0 by solving:

$$H^d(\theta) - \hat{H}^d = 0 \,.$$

As discussed in Section 1, this procedure has been shown to lead to efficiency gains over maximum likelihood estimation in small samples from several distributions. Nonetheless, the choice of L-moments $R = d$ appears rather *ad-hoc*, as it is based on an order condition for identification. One may then wonder whether increasing the number of L-moments used in estimation – and weighting these properly –, might lead to a more efficient estimator in finite samples. Moreover, if one correctly varies the number of L-moments with sample size, it may be possible to construct an estimator that does not underperform MLE even asymptotically. The latter appears especially plausible if one considers the result in Hosking (1990), who shows that L-moments characterise a distribution with finite first moment.

In light of the preceding discussion, we propose to analyse the behaviour of the "generalised" method of L-moments estimator:

$$\hat{\theta} \in \arg\inf_{\theta \in \Theta} (H^R(\theta) - \hat{H}^R)' W^R (H^R(\theta) - \hat{H}^R) \,, \tag{6}$$

where R may vary with sample size; and W^R is a (possibly estimated) weighting matrix. In Section 3, we work out the asymptotic properties of this estimator in a framework where both T and R diverge, i.e. we index R by the sample size and consider the asymptotic behaviour of the estimator in the large-sample limit, for sequences $(R_T)_{T \in \mathbb{N}}$ where $\lim_{T \to \infty} R_T = \infty$.[5,6] We derive sufficient conditions for an asymptotic linear representation of the estimator to hold. We also show that the estimator is asymptotically efficient, in the sense that, under iid data and when optimal weights are used, its asymptotic variance coincides with the inverse of the Fisher information matrix. In Section 4, we conduct a small Monte Carlo exercise which showcases the gains associated with our approach. Specifically, we show that, for the task of tail quantile estimation, our L-moment approach entails mean-squared error gains over MLE in smaller samples, and performs as well as it in larger samples.[7] In light of these results, in Section 5 we propose to construct a semiautomatic method of selection of the number of L-moments by working with higher-order expansions of the mean-squared error of the estimator – in a similar fashion to what has already been done in the GMM literature (Donald and Newey, 2001; Donald et al., 2009; Okui, 2009; Abadie et al., 2024). We also consider an approach based on ℓ_1-regularisation borrowed from the GMM literature (Luo et al., 2015). We then return to the Monte Carlo setting of Section 4 in order to assess the properties of the proposed selection methods. In Section 6, we consider extensions of our main approach to the estimation of conditional models and a class of semiparametric models.

In this paper, we will focus on the case where estimated L-moments are given by (3). As shown in Hosking (1990), under random sampling and finite second moments, for each $r \in \mathbb{N}$, $\hat{\lambda}_r - \tilde{\lambda}_r = O_p(T^{-1})$, which implies that the estimator (6) using either (3) or (5) as $\hat{H}^R$ are first-order asymptotically equivalent *when R is fixed*. However, in an asymptotic framework where R increases with the sample size, this need not be the case. Indeed, note that, for $r > T$, $\tilde{\lambda}_r$ is not even defined, whereas $\hat{\lambda}_r$ is. Relatedly, the simulations in Section 4 show that, for values of R close to T, the generalised-method-of L-moment estimator (6) based on $\tilde{\lambda}_r$ breaks down, whereas the estimator based on $\hat{\lambda}_r$ does not.[8] We thus focus on the properties of the estimator that relies on $\hat{\lambda}_r$, as it is especially well-suited for settings where one may wish to make R large.

Remark 1 (*On Computation of Integrals*)**.** Parameter estimation through L-moments hinges crucially on computation of integrals $\int_0^1 Q(u) u^r du$, for a given quantile function Q and $r \in \{0, 1, \dots, R-1\}$. These quantities, known in the literature as probability-weighted moments (Landwehr et al., 1979), are directly available in closed form when Q is stepwise-constant (as when Q are the empirical quantiles), though this need not be the case for the theoretical moments $\int_0^1 Q(u|\theta) u^r du$ from parametric families of distributions. Hosking (1986) provides closed-form formulae for the probability-weighted moments of several families of distributions, including those from the Generalised Pareto (GPD) and Generalised Extreme Value (GEV) families that are popular in

[5] We maintain the indexing of R by T implicit to keep the notation concise. In this setting, the phrase "as $T, R \to \infty$" should be read as meaning that a property holds in the limit $T \to \infty$, for any sequence $(R_T)_{T \in \mathbb{N}}$ with $\lim_{T \to \infty} R_T = \infty$. The phrase "as $T, R \to \infty$ with $\phi(R, T) \to c$", where $\phi : \mathbb{N}^2 \mapsto \mathbb{R}$ and $c \in \mathbb{R}$, should be read as meaning that a property holds in the limit $T \to \infty$, for any sequence $(R_T)_{T \in \mathbb{N}}$ with $\lim_{T \to \infty} R_T = \infty$ and $\lim_{T \to \infty} \phi(R_T, T) = c$.

[6] As it will become clear in Section 3, our framework nests the setting with fixed R as a special case by properly filling the weighting matrix W^R with zeros.

[7] In the Supplemental Appendix, we consider an alternative setting where the estimation target consists of linear combinations of the model parameters, and also verify gains in adopting our approach.

[8] The latter phenomenon is corroborated by our theoretical results in Section 3 for the estimator based on $\hat{\lambda}_r$, which allow R to be much larger that T.

the modelling of extreme events, and which we consider in the Monte Carlo exercise of Section 4.[9] For those distributions where the integral does not admit a closed-form expression, the R package `lmom` (Hosking, 2024) provides access to Fortran routines that compute probability-weighted moments using numerical integration.

Remark 2 (*L-Moment Estimation as a Computationally Attractive Alternative to the MLE*)**.** There are instances where L-moment-based parameter estimates are easily computable, whereas maximum likelihood estimation can be computationally complicated. One example is that given by *quantile mixtures*, e.g. when the quantile function of the distribution of interest is given by $\theta_1 Q_1(u)+\theta_2 Q_2(u)$, with θ_1 and θ_2 unknown constants and Q_1 and Q_2 known quantile functions whose L-moments are easily computable. In this case, the conventional method-of L-moment estimator collapses to solving a linear system, and finding our generalised method-of-L-moment estimator amounts to solving a quadratic program. In contrast, estimation of quantile mixtures through maximum likelihood can be much more complicated computationally, as it involves, for each candidate parameter value, differentiation of the inverse of the quantile function, which is generally unavailable in closed-form. The estimation of quantile mixtures through the method-of-L-moments has been studied by Karvanen (2006) and Gourieroux and Jasiak (2008) in the context of modelling asset returns, while Alvarez and Orestes (2023), building on our proposed generalised-method-of-L-moment estimator, study quantile mixture models as a general tool for approximating a distribution of interest, with a particular focus in causal inference on distributional outcomes in observational settings.

3. Asymptotic properties of the generalised method of L-moments estimator with many moments

3.1. Setup

As in the previous section, we consider a setting where we have a sample with T identically distributed observations, $Y_1, Y_2 \dots Y_T$, $Y_t \sim F$ for $t = 1, 2 \dots T$, where F belongs to a parametric family $\{F_\theta : \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^d$; and $F = F_{\theta_0}$ for some $\theta_0 \in \Theta$. We will analyse the behaviour of the estimator:

$$\hat{\theta} \in \underset{\theta\in\Theta}{\arg\inf} \sum_{k=1}^{R}\sum_{l=1}^{R}\left(\int_{\underline{p}}^{\bar{p}} \left[\hat{Q}_Y(u) - Q_Y(u|\theta)\right] P_k(u)du\right) w^R_{k,l} \left(\int_{\underline{p}}^{\bar{p}} \left[\hat{Q}_Y(u) - Q_Y(u|\theta)\right] P_l(u)du\right), \tag{7}$$

where $\hat{Q}_Y(\cdot)$ is the empirical quantile process given by (4); $Q_Y(\cdot|\theta)$ is the quantile function associated with F_θ; $\{w^R_{k,l}\}_{1\le k,l\le R}$ are a set of (possibly estimated) weights; $\{P_k\}_{1\le k\le R}$ are a set of quantile "weighting" functions with $\int_0^1 P_k(u)^2 du = 1$; and $0 \le \underline{p} < \bar{p} \le 1$. This setting encompasses the generalised-method-of-L-moment estimator discussed in the previous section. Indeed, by choosing $P_k(u) = \sqrt{2(2k-1)} \cdot P^*_{k-1}(u)$, where $P^*_k(u)$ are the shifted Legendre polynomials on $[0, 1]$, and $0 = \underline{p} < \bar{p} = 1$, we have the generalised L-moment-based estimator in (6) using (3) as an estimator for the L-moments.[10] We leave $\underline{p} < \bar{p}$ fixed throughout.[11] All limits are taken **jointly** with respect to T **and** R.

To facilitate analysis, we let $\mathbf{P}^R(u) := (P_1(u), P_2(u) \dots P_R(u))'$; and write W^R for the $R \times R$ matrix with entry $W^R_{i,j} = w^R_{i,j}$. We may then rewrite our estimator in matrix form as:

$$\hat{\theta} \in \underset{\theta\in\Theta}{\arg\inf} \left[\int_{\underline{p}}^{\bar{p}} \left(\hat{Q}_Y(u) - Q_Y(u|\theta)\right) \mathbf{P}^R(u)' du\right] W^R \left[\int_{\underline{p}}^{\bar{p}} \left(\hat{Q}_Y(u) - Q_Y(u|\theta)\right) \mathbf{P}^R(u) du\right].$$

3.2. Consistency

In this section, we present conditions under which our estimator is consistent.

We impose the following assumptions on our environment. In what follows, we write $Q_Y(\cdot) = Q_Y(\cdot|\theta_0)$.

Assumption 1 (*Consistency of Empirical Quantile Process*)**.** The empirical quantile process is uniformly consistent on $(\underline{p}, \bar{p})$, i.e.

$$\sup_{u\in(\underline{p},\bar{p})} |\hat{Q}_Y(u) - Q_Y(u)| \xrightarrow{P} 0. \tag{8}$$

Assumption 1 is satisfied in a variety of settings. For example, if $Y_1, Y_2 \dots Y_T$ are iid and the family $\{F_\theta : \theta \in \Theta\}$ is continuous with a (common) compact support; then (8) follows with $\underline{p} = 0$ and $\bar{p} = 1$ (Ahidar-Coutrix and Berthet, 2016, Proposition 2.1.). Yoshihara (1995) and Portnoy (1991) provide sufficient conditions for uniform consistency (8) to hold when observations are dependent. We also note that, for the result in this section, it would have been sufficient to assume convergence in probability in the $L^2(\underline{p}, \bar{p})$ norm.[12] We only state results in the sup-norm because convergence statements regarding the empirical quantile process available in the literature are usually proved in $L^\infty(\underline{p}, \bar{p})$.

[9] For completeness, we reproduce the closed-form expressions of these integrals in the GEV and GPD families in Supplemental Appendix M.

[10] The rescaling by $\sqrt{2(2k-1)}$ is adopted so the polynomials have unit $L^2[0,1]$-norm.

[11] In Remark 8 later on, we briefly discuss an extension to sample-size-dependent trimming.

[12] Under random sampling from a distribution with Lebesgue density f such that $u \mapsto f(Q_Y(u))$ is continuous on $(0, 1)$, empirical quantiles are consistent in $L^2(0, 1)$ if one of the two conditions hold: the distribution has finite $(2+\delta)$-moment, or there exist real constants C, k_1, k_2 such that $f(Q_Y(u))^{-1} \le Cu^{k_1}(1-u)^{k_2}, \forall u \in (0, 1)$. See Supplemental Appendix N for a proof.

Assumption 2 (*Quantile Weighting Functions*). The functions $\{P_l : l \in \mathbb{N}\}$ constitute an orthonormal sequence on $L^2[0,1]$.

Assumption 2 is satisfied by (rescaled) shifted Legendre polynomials, shifted Jacobi polynomials and other weighting functions. Next, we impose restrictions on the estimated weights. In what follows, we write, for a $c \times d$ matrix A, $\|A\|_2 = \sqrt{\lambda_{\max}(A'A)}$.

Assumption 3 (*Estimated Weights*). There exists a sequence of nonstochastic symmetric positive semidefinite matrices Ω^R such that, as $T, R \to \infty$, $\|W^R - \Omega^R\|_2 = o_{P^*}(1)$[13]; $\|\Omega^R\|_2 = O(1)$.

Assumption 3 restricts the range of admissible weight matrices. Notice that $W^R = \Omega^R = \mathbb{I}_R$ trivially satisfies these assumptions. By the triangle inequality, Assumption 3 implies that $\|W^R\|_2 = O_{P^*}(1)$.

Finally, we introduce our identifiability assumption. For some $X \in L^2[0,1]$, let $\|X\|_{L^2[0,1]} = \left(\int_0^1 X(u)^2 du\right)^{\frac{1}{2}}$:

Assumption 4 (*Strong Identifiability and Suprema of L^2 Norm of Parametric Quantiles*). For each $\epsilon > 0$:

$$\liminf_{R\to\infty} \inf_{\theta\in\Theta:\|\theta-\theta_0\|_2\geq\epsilon} \left[\int_{\underline{p}}^{\bar{p}} \left(Q_Y(u|\theta) - Q_Y(u|\theta_0)\right) \mathbf{P}^R(u)' du\right] \Omega^R \left[\int_{\underline{p}}^{\bar{p}} \left(Q_Y(u|\theta) - Q_Y(u|\theta_0)\right) \mathbf{P}^R(u) du\right] > 0\,.$$

Moreover, we require that $\sup_{\theta\in\Theta} \|Q_Y(\cdot|\theta)\mathbb{1}_{[\underline{p},\bar{p}]}\|_{L^2[0,1]} < \infty$.

The first part of this assumption is closely related to the usual notion of identifiability in parametric distribution models. Indeed, if $0 = \underline{p} < \bar{p} = 1$, Θ is compact, $\theta \mapsto \|Q(\cdot|\theta)\|_{L^2[0,1]}$ is bounded and $(\theta', \theta'') \mapsto \|Q(\cdot|\theta') - Q(\cdot|\theta'')\|_{L^2[0,1]}$ is continuous, the $\{P_l\}_l$ constitute an orthonormal *basis* in $L^2[0,1]$ (this is the case for rescaled shifted Legendre polynomials), and if the smallest eigenvalue of Ω_R is bounded away from zero uniformly in R (for example, if we take $W^R = \mathbb{I}_R$), then the first part is equivalent to identifiability of the parametric family $\{F_\theta\}_\theta$ (see Supplemental Appendix C.1 for a proof).

As for the second part of the assumption, we note that boundedness of the L^2 norm of parametric quantiles uniformly in θ is satisfied in several settings. If the parametric family $\{F_\theta : \theta \in \Theta\}$ has common compact support, then the assumption is trivially satisfied. Alternatively, if we assume Θ is compact and $Q_Y(u|\theta)$ is jointly continuous and bounded on $[\underline{p}, \bar{p}] \times \Theta$, then the condition follows from Weierstrass' theorem, as in this case: $\sup_{\theta\in\Theta} \|Q_Y(\cdot|\theta)\mathbb{1}_{[\underline{p},\bar{p}]}\|_{L^2[0,1]} \leq \sqrt{\bar{p} - \underline{p}} \cdot \sup_{(\theta,u)\in\Theta\times[\underline{p},\bar{p}]} |Q_Y(u|\theta)| < \infty$. More generally, if the support of the family under consideration is unbounded, then we may ensure that the assumption is satisfied by considering compact parameter spaces, or by a proper choice of trimming constants $\underline{p}$ and $\bar{p}$. For example, if we assume that Θ is compact, and that $\theta \mapsto \int_{\underline{p}}^{\bar{p}} Q_Y(u|\theta)^2 du$ is continuous and bounded on Θ, then the condition is satisfied. Supplemental Appendix B shows that, for the GEV and GPD families of distributions mentioned in Remark 1 and considered in the Monte Carlo Exercise, the uniform boundedness assumption is satisfied with $0 = \underline{p} < \bar{p} = 1$ and compact parameter spaces in the region where the distributions have finite variance. Moreover, we show that, by taking $0 < \underline{p}$ and $\bar{p} < 1$ in the GEV family and $\bar{p} < 1$ in the GPD family, it is possible to extend these parameter spaces to the region where the distributions have infinite variance.[14]

Under the previous assumptions, the estimator is consistent.

Proposition 1. *Suppose Assumptions 1 to 4 hold. Then $\hat{\theta} \xrightarrow{P^*} \theta_0$ as $R, T \to \infty$.*

Proof. See Supplemental Appendix A.1. □

Remark 3. Note that Proposition 1 does not impose any restrictions on the rate of growth of L-moments. This stands in contrast with consistency results in the literature exploring the behaviour of GMM in asymptotic sequences with an increasing number of moments. For example, when estimating a finite-dimensional parameter identified by a conditional moment restriction through many unconditional moments that span the available restrictions, the series-IV estimator proposed by Donald et al. (2003) is consistent in an asymptotic regime where the number of moments R satisfies $R/T \to 0$. Intuitively, one needs to impose this growth restriction in order to control the variance of an increasing number of moments, even in this particular case where moments are derived from series regressors. In contrast, the special structure of L-moments in our setting, being written as the projection coefficients of the *same* quantile function on an orthonormal sequence in $L^2[0,1]$, enables us to properly control the variance even when R is arbitrarily large, for Bessel's inequality (Kreyszig, 1989, page 157) implies that, for every R, $\left\|\int_{\underline{p}}^{\bar{p}} \left(\hat{Q}_Y(u) - Q_Y(u)\right) \mathbf{P}^R(u) du\right\|_2 \leq \|(\hat{Q}_Y(\cdot) - Q_Y(\cdot))\mathbb{1}_{[\underline{p},\bar{p}]}\|_{L^2[0,1]}$, with the upper bound crucially not depending on R. See Supplemental Appendix D for a detailed comparison between the consistency arguments underlying our Proposition 1 and Theorem 5.1 of Donald et al. (2003).

[13] The notation $o_{P^*}(1)$ expresses convergence in outer probability to zero. We state our main assumptions and results in outer probability in order to abstract from measurability concerns. We note these results are equivalent to convergence in probability when the appropriate measurability assumptions hold.

[14] As we show in the Appendix, it is actually possible to extend the parameter space to regions where even the first moment does not exist, since, in this case, even though untrimmed L-moments are not defined, trimmed L-moments are. We discuss a data-driven method to select the trimming constants in Remark 8.

3.3. Asymptotic linear representation

In this section, we provide conditions under which the estimator admits an asymptotic linear representation. In what, follows, define $h^R(\theta) := \int_{\underline{p}}^{\bar{p}} \left(\hat{Q}_Y(u) - Q_Y(u|\theta)\right) \mathbf{P}^R(u)du$; and write $\nabla_{\theta'} h^R(\tilde{\theta})$ for the Jacobian of h^R with respect to θ, evaluated at $\tilde{\theta}$. We assume that:

Assumption 5. There exists an open ball $\mathcal{O}$ in $\mathbb{R}^d$ containing θ_0 such that $\mathcal{O} \subseteq \Theta$ and $Q_Y(u|\theta)$ is differentiable on $\mathcal{O}$, uniformly in $u \in [\underline{p}, \bar{p}]$. Moreover, $\theta \mapsto Q_Y(u|\theta)$ is **continuously** differentiable on $\mathcal{O}$ for each u; and, for each $\theta \in \mathcal{O}$, $\nabla_{\theta'} Q_Y(\cdot|\theta)$ is square integrable on $[\underline{p}, \bar{p}]$.

Assumption 6. $\sqrt{T}(\hat{Q}_Y(\cdot) - Q_Y(\cdot))$ converges weakly in $L^\infty(\underline{p}, \bar{p})$ to a zero-mean Gaussian process B with continuous sample paths and covariance kernel Γ.

Assumption 7. $Q_Y(u|\theta)$ is **twice** continuously differentiable on $\mathcal{O}$, for each $u \in [\underline{p}, \bar{p}]$. Moreover, $\sup_{\theta\in\mathcal{O}} \sup_{u\in[\underline{p},\bar{p}]} \|\nabla_{\theta\theta'} Q_Y(u|\theta)\|_2 < \infty$.

Assumption 8. The smallest eigenvalue of $\nabla_{\theta'} h^R(\theta_0)' \Omega^R \nabla_{\theta'} h^R(\theta_0)$ is bounded away from 0, uniformly in R.

Assumption 5 requires θ_0 to be an interior point of Θ. It also implies the objective function is continuously differentiable on a neighbourhood of θ_0, which enables us to linearise the first order condition satisfied with high probability by $\hat{\theta}$.

Weak convergence of the empirical quantile process (Assumption 6) has been derived in a variety of settings, ranging from iid data (van der Vaart, 1998, Corollary 21.5) to nonstationary and weakly dependent observations (Portnoy, 1991). In the iid setting, if the family $\{F_\theta : \theta \in \Theta\}$ is continuously differentiable with strictly positive density f_θ over a (common) compact support; then weak-convergence holds with $\underline{p} = 0$ and $\bar{p} = 1$. In this case, the covariance kernel is $\Gamma(i,j) = \frac{(i\wedge j - ij)}{f_Y(Q_Y(i)) f_Y(Q_Y(j))}$. Similarly to the discussion of Assumptions 1, 6 is stronger than necessary: it would have been sufficient to assume $\|\sqrt{T}(Q_Y(\cdot) - \hat{Q}_Y(\cdot))\mathbb{1}_{[\underline{p},\bar{p}]}\|^2_{L^2[0,1]} = O_{P^*}(1)$, which is implied by weak convergence in $L^2(\underline{p}, \bar{p})$ (see Mason, 1984 and Barrio et al., 2005 for results in this direction).

Assumption 7 is a technical condition which enables us to provide an upper bound to the linearisation error of the first order condition satisfied by $\hat{\theta}$.

Assumption 8 is similar to the rank condition used in the proof of asymptotic normality of M-estimators (Newey and McFadden, 1994), which is known to be equivalent to a local identification condition under rank-regularity assumptions (Rothenberg, 1971). In our setting, where R varies with sample size, we show in Supplemental Appendix C.2 that a stronger version of Assumption 4 implies Assumption 8.

Under Assumptions 1–8, we have that:

Proposition 2. *Suppose Assumptions 1–8 hold. Then, as $T, R \to \infty$, the estimator admits the asymptotic linear representation:*

$$\sqrt{T}(\hat{\theta} - \theta_0) = -(\nabla_{\theta'} h^R(\theta_0)' \Omega^R \nabla_{\theta'} h^R(\theta_0))^{-1} \nabla_{\theta'} h^R(\theta_0)' \Omega^R (\sqrt{T} h^R(\theta_0)) + o_{P^*}(1). \tag{9}$$

Proof. See Supplemental Appendix A.2. □

In the next subsection, we work out an asymptotic approximation to the distribution of the leading term in (9).

Remark 4. Note that our linearisation result in Proposition 2 does not impose any restrictions on the rate of growth of L-moments, which again contrasts with existing results in the GMM literature (Koenker and Machado, 1999; Donald et al., 2003; Han and Phillips, 2006), where rate restrictions are typically required in order to establish an asymptotic linear representation. This difference may again be attributed to the special structure of L-moments in our setting. Indeed, whereas the asymptotic normality result on the series-IV estimator discussed in Donald et al. (2003) assumes the rate restriction $R/T^2 \to 0$ in order to control a crucial bias term in the asymptotic linear representation stemming from correlation between the gradient of the empirical moment condition at the true parameter and sample moments at the truth, the fact that, in our setting, the gradient of the difference between empirical and theoretical L-moments at the truth is not affected by estimation error of the empirical quantiles $\hat{Q}_Y$, coupled with Bessel's inequality, enables us to control the linearisation error without such restriction. See Supplemental Appendix D for further discussion.

3.4. Asymptotic distribution

Finally, to work out the asymptotic distribution of the proposed estimator, we rely on *a strong approximation concept*. The idea is to construct, in the *same* underlying probability space, a sequence of Brownian bridges that approximates, in the supremum norm, the empirical quantile process. This can then be used to conduct inference based on a Gaussian distribution. In Supplemental Appendix H, we alternatively show how a *Bahadur–Kiefer representation* of the quantile process can be used to conduct inference in the iid case. In this alternative, one approximates the distribution of the leading term of (9) by a transformation of independent Bernoulli random variables.

We first consider a strong approximation to a Gaussian process in the iid setting. We state below a classical result, due to Csorgo and Revesz (1978):

Theorem 1 (*Csorgo and Revesz, 1978*). *Let $Y_1, Y_2 \dots Y_T$ be an iid sequence of random variables with a continuous distribution function F which is also twice differentiable on (a,b), where $-\infty \leq a = \sup\{z : F(z) = 0\}$ and $b = \inf\{z : F(z) = 1\} \leq \infty$. Suppose that $F'(z) = f(z) > 0$ for $z \in (a,b)$. Assume that, for $\gamma > 0$:*

$$\sup_{a<x<b} F(x)(1-F(x))\left|\frac{f'(x)}{f^2(x)}\right| \leq \gamma\,,$$

where f denotes the density of F. Moreover, assume that f is nondecreasing (nonincreasing) on an interval to the right of a (to the left of b). Then, if the underlying probability space is rich enough, one can define, for each $t \in \mathbb{N}$, a Brownian bridge $\{B_t(u) : u \in [0,1]\}$ such that, if $\gamma < 2$:

$$\sup_{0<u<1} |\sqrt{T} f(Q_Y(u))(\hat{Q}_Y(u) - Q_Y(u)) - B_T(u)| \overset{a.s.}{=} O(T^{-1/2}\log(T))\,, \tag{10}$$

and, if $\gamma \geq 2$

$$\sup_{0<u<1} |\sqrt{T} f(Q_Y(u))(\hat{Q}_Y(u) - Q_Y(u)) - B_T(u)| \overset{a.s.}{=} O(T^{-1/2}(\log\log T)^{\gamma}(\log T)^{\frac{(1+\epsilon)}{(\gamma-1)}})\,, \tag{11}$$

for arbitrary $\epsilon > 0$.

The above theorem is stronger than the weak convergence of Assumption 6. Indeed, Theorem 1 requires variables to be defined in the same probability space and yields explicit bounds in the sup norm; whereas weak convergence is solely a statement on the convergence of integrals (van der Vaart and Wellner, 1996). Suppose the approximation (10)/(11) holds in our context. Let B_T be as in the statement of the theorem, and assume in addition that $\int_{\underline{p}}^{\bar{p}} \frac{1}{f_Y(Q_Y(u))^2} du < \infty$. A simple application of Bessel's inequality then shows that:

$$\sqrt{T}(\hat{\theta} - \theta_0) = -(\nabla_{\theta'} h^R(\theta_0)' \Omega^R \nabla_{\theta'} h^R(\theta_0))^{-1} \nabla_{\theta'} h^R(\theta_0)' \Omega^R \left[\int_{\underline{p}}^{\bar{p}} \frac{B_T(u)}{f_Y(Q_Y(u))} \mathbf{P}^R(u) du\right] + o_{P^*}(1)\,. \tag{12}$$

Note that the distribution of the leading term in the right-hand side is known (by Riemann integration, it is Gaussian) up to θ_0. This representation could thus be used as a basis for inference. The validity of such approach can be justified by verifying that the Kolmogorov distance between the distribution of $\sqrt{T}(\hat{\theta} - \theta_0)$ and that of the leading term of the representation goes to zero as T and R increase. We show that this indeed is true later on, where convergence in the Kolmogorov distance is obtained as a byproduct of weak convergence.

Next, we reproduce a strong approximation result in the context of dependent observations. The result is due to Fotopoulos and Ahn (1994) and Yu (1996).

Theorem 2 (*Fotopoulos and Ahn, 1994; Yu, 1996*). *Let $Y_1, Y_2 \dots Y_T$ be a strictly stationary, α-mixing sequence of random variables, with mixing coefficient satisfying $\alpha(t) = O(t^{-8})$. Let F denote the distribution function of Y_1. Suppose the following* ***Csorgo and Revesz conditions*** *hold:*

a. F is twice differentiable on (a,b), where $-\infty \leq a = \sup\{z : F(z) = 0\}$ and $b = \inf\{z : F(z) = 1\} \leq \infty$;
b. $\sup_{0<s<1} |f'(Q_Y(s))| < \infty$;

as well as the condition:

c. $\inf_{0<s<1} f(Q_Y(s)) > 0$.

Let $\Gamma(s,t) := \mathbb{E}[g_1(s)g_1(t)] + \sum_{n=2}^{\infty}\{\mathbb{E}[g_1(s)g_n(t)] + \mathbb{E}[g_1(t)g_n(s)]\}$, where $g_n(u) := \mathbb{1}\{U_n \leq u\} - u$ and $U_n := F(Y_n)$. Then, if the probability space is rich enough, there exists a sequence of Brownian bridges $\{\tilde{B}_n : n \in \mathbb{N}\}$ with covariance kernel Γ and a positive constant $\lambda > 0$ such that:

$$\sup_{0<u<1} |\sqrt{T}(\hat{Q}_Y(u) - Q_Y(u)) - f(Q_Y(u))^{-1}\tilde{B}_T(u)| \overset{a.s.}{=} O((\log T)^{-\lambda})\,. \tag{13}$$

A similar argument as the previous one then shows that, under the conditions of the theorem above:

$$\sqrt{T}(\hat{\theta} - \theta_0) = -(\nabla_{\theta'} h^R(\theta_0)' \Omega^R \nabla_{\theta'} h^R(\theta_0))^{-1} \nabla_{\theta'} h^R(\theta_0)' \Omega^R \left[\int_{\underline{p}}^{\bar{p}} \frac{\tilde{B}_T(u)}{f_Y(Q_Y(u))} \mathbf{P}^R(u) du\right] + o_{P^*}(1)\,. \tag{14}$$

Differently from the iid case, the distribution of the leading term on the right-hand side is now known up to θ_0 **and the covariance kernel** Γ. The latter could be estimated with a Newey and West (1987) style estimator.

To conclude the discussion, we note that the strong representation (12) (resp. (14)) allows us to establish asymptotic normality of our estimator. Indeed, let L_T be the leading term of the representation on the right-hand side of (12) (resp. (14)), and $V_{T,R}$ be its variance. Observe that $V_{T,R}^{-1/2} L_T$ is distributed according to a multivariate standard normal. It then follows by Slutsky's theorem that $V_{T,R}^{-1/2}\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathbb{I}_d)$. Since pointwise convergence of cumulative distribution functions to a continuous distribution function implies uniform convergence (Parzen, 1960, page 438), and given that $V_{T,R}^{-1/2}$ is positive definite, we obtain that:

$$\lim_{T,R\to\infty} \sup_{c\in\mathbb{R}^d} |P[\sqrt{T}(\hat{\theta} - \theta_0) \leq c] - P[L_T \leq c]| = 0\,, \tag{15}$$

which justifies our approach to inference based on the distribution of the leading term on the right-hand side of (12) (resp. (14)).

We collect the main results in this subsection under the corollary below.

Corollary 1. *Suppose Assumptions 1–8 hold. Moreover, suppose a strong approximation condition such as (10)/(11) or (13) is valid; and, in addition, that* $\int_{\underline{p}}^{\bar{p}} \frac{1}{f_Y(Q_Y(u))^2} du < \infty$. *Then, as* $T, R \to \infty$, *the approximation (12) (resp. (14)) holds. Moreover, we have that, as* $T, R \to \infty$, $V_{T,R}^{-1/2} \sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathbb{I}_d)$ *and that (15) holds.*

Remark 5 (*Optimal Choice of Weighting Matrix Under Gaussian Approximation*)**.** Under (12), the optimal choice of weights that minimises the variance of the leading term is:

$$\Omega_R^* = \mathbb{E}\left[\left(\int_{\underline{p}}^{\bar{p}} \frac{B_T(u)}{f_Y(Q_y(u))} \mathbf{P}^R(u) du\right)\left(\int_{\underline{p}}^{\bar{p}} \frac{B_T(u)}{f_Y(Q_y(U))} \mathbf{P}^R(u) du\right)'\right]^{-}, \tag{16}$$

where A^- denotes the generalised inverse of a matrix A. This weight can be estimated using a preliminary estimator for θ_0. An analogous result holds under (14), though in this case one also needs an estimator for the covariance kernel Γ. In Supplemental Appendix E, we provide an estimator for Ω_R in the iid case when the $\{P_l\}$ are shifted Legendre Polynomials.

Remark 6 (*A Test Statistic for Overidentifying Restrictions*)**.** The strong approximation discussed in this subsection motivates a test statistic for overidentifying restrictions. Suppose $R > d$. Denoting by $M(\cdot)$ the objective function of (7), we consider the test-statistic:

$$J := T \cdot M(\hat{\theta}_T).$$

An analogous statistic exists in the overidentified GMM setting (Newey and McFadden, 1994; Wooldridge, 2010). Under the null that the model is correctly specified (i.e. that there exists $\theta \in \Theta$ such that $Q_Y(\cdot) = Q_Y(\cdot|\theta)$), we can use the results in this section to compute the distribution of this test statistic. Specifically, if the optimal weighting scheme is adopted, the distribution of the test statistic under the null may be approximated by a chi-squared distribution with $R - d$ degrees of freedom. To establish this fact, we rely on an anticoncentration inequality due to Götze et al. (2019). See Supplemental Appendix F for details.

Remark 7 (*Sample-Size-Dependent Trimming*)**.** It is possible to adapt our assumptions and results to the case where the trimming constants $\underline{p}$, $\bar{p}$ are functions of the sample size. In particular, Theorem 6 of Csorgo and Revesz (1978) provides uniform strong approximation results for sample quantiles ranging from $[1-\delta_T, \delta_T]$, where $\delta_T = 25T^{-1}\log\log T$. This result imposes fewer restrictions on the distribution, and could be used as the basis for inference on a variable-trimming estimator.

Remark 8 (*Data-Driven Method to Select Trimming Proportions*)**.** If one wishes to adopt trimming, then, for a given R, a data-driven method for selecting $\underline{p}$ and $\bar{p}$ can be obtained by choosing these constants so as to minimise an estimate of the variance of the leading term in (9). See Athey et al. (2023) for a discussion of this approach in estimating the mean of a symmetric distribution; and Crump et al. (2009) for a related approach when choosing trimming constants for the estimated propensity score in observational studies.

Remark 9 (*Inference Based on the Weighted Bootstrap*)**.** In Supplemental Appendix G, we show how one can leverage the strong approximations discussed in this section to conduct inference on the model parameters using the weighted bootstrap.

Finally, we observe that, in some settings, we are not interested in conducting inference on θ_0, but rather on a sequence of scalar functions $g_T(\theta_0)$. Typical examples include the estimation of tail probabilities and quantiles. The following result, which is an immediate consequence of Corollary 1, provides conditions for inference based on the Delta Method to be valid in this setting.

Corollary 2. *Let* $g_n : \mathbb{R}^d \mapsto \mathbb{R}$, $n \in \mathbb{N}$, *be a sequence of functions such that there exists an open ball* $\mathcal{B} \subseteq \mathbb{R}^d$ *containing* θ_0 *with: (1) each* g_n *is continuously differentiable on* $\mathcal{B}$, *and (ii) the gradient functions* $\{\nabla g_n : n \in \mathbb{N}\}$ *are equicontinuous on* $\mathcal{B}$, *with* $\nabla g_n(\theta_0) \neq 0$ *for every* $n \in \mathbb{N}$. *If the conditions of Corollary 1 are satisfied, then, as* $T, R \to \infty$, $\frac{\sqrt{T}(g_T(\hat{\theta}_T) - g_T(\theta_0))}{\sqrt{\nabla g_T(\theta_0)' V_{T,R} \nabla g_T(\theta_0)}} \xrightarrow{d} N(0,1)$.

Proof. The conditions in the statement of the corollary imply that, as $T, R \to \infty$: $\sqrt{T}(g_T(\hat{\theta}_T) - g_T(\theta_0)) = \nabla g_T(\theta_0)' \sqrt{T}(\hat{\theta}_T - \theta_0) + o_{P^*}(1)$. The conclusion then follows from Corollary 1. □

3.5. Asymptotic efficiency

Supplemental Appendix I discusses efficiency of our proposed L-moment estimator. Specifically, we show that, when no trimming is adopted ($0 = \underline{p} < \bar{p} = 1$), the optimal weighting scheme (16) is used, the $\{P_l\}_l$ constitute an orthonormal *basis* in $L^2[0,1]$ (recall this is satisfied by shifted Legendre polynomials), and the data is iid, the generalised method of L-moments estimator is asymptotically efficient, in the sense that its asymptotic variance coincides with the inverse of the Fisher information matrix of the parametric model.[15] We leave details to the Supplemental Appendix, though we briefly outline the argument here. The idea is to consider the

[15] In the dependent case, "efficiency" should be defined as achieving the efficiency bound of the semiparametric model that parametrises the marginal distribution of the Y_t, but leaves the time series dependence unrestricted up to regularity conditions (Newey, 1990; Komunjer and Vuong, 2010). Indeed, in general, our L-moment estimator will be inefficient with respect to the MLE that models the dependency structure between observations. See Carrasco and Florens (2014) for further discussion.

alternative estimator:

$$\tilde{\theta}_T \in \operatorname{argmin}_{\theta\in\Theta} \sum_{i\in\mathcal{G}_T}\sum_{j\in\mathcal{G}_T} (\hat{Q}_Y(i) - Q_Y(i|\theta))\kappa_{i,j}(\hat{Q}_Y(j) - Q_Y(j|\theta)), \tag{17}$$

for a grid of G_T points $\mathcal{G}_T = \{g_1, g_2, \dots, g_{G_T}\} \subseteq (0,1)$ and weights $\kappa_{i,j}$, $i,j \in \mathcal{G}_T$. This is a weighted version of a "percentile-based estimator", which is used in contexts where it is difficult to maximise the likelihood (Gupta and Kundu, 2001). It amounts to choosing θ so as to match a weighted combination of the order statistics in the sample. In the Supplemental Appendix, we show that, under a suitable sequence of gridpoints and optimal weights, this estimator is asymptotically efficient. We then show, by using the fact that the $\{P_l\}_l$ form an orthonormal basis, that estimator (17) can be seen as a generalised L-moment estimator that uses infinitely many L-moments. The final step of the argument then consists in observing that a generalised L-moment estimator that uses a finite but increasing number of L-moments is asymptotically equivalent to estimator (17), which implies that a generalised L-moment estimator under optimal weights is no less efficient than the (efficient) percentile estimator.

4. Monte Carlo exercise

In our experiments, we draw random samples $Y_1, Y_2, \dots, Y_T$ from a distribution function $F = F_{\theta_0}$ belonging to a parametric family $\{F_\theta : \theta \in \Theta\}$. Following Hosking (1990), we consider the goal of the researcher to be estimating quantiles $Q_Y(\tau)$ of the distribution F_{θ_0} by using a plug-in approach: first, the researcher estimates θ_0; then she estimates $Q_Y(\tau)$ by setting $\widehat{Q_Y(\tau)} = Q_Y(\tau|\hat{\theta})$. As in Hosking (1990), we consider $\tau \in \{0.9, 0.99, 0.999\}$. In order to compare the behaviour of alternative procedures in estimating more central quantiles, we also consider the median $\tau = 0.5$. We analyse sample sizes $T \in \{50, 100, 500\}$. The number of Monte Carlo draws is set to 5000.

We compare the root mean squared error of four types of generalised method of L-moment estimators under varying choices of R with the root mean squared error obtained were θ_0 to be estimated via MLE. We consider the following estimators: (i) the generalised method of L-moments estimator that uses the càglàd L-moment estimates (3) and identity weights (**Càglàd FS**)[16]; (ii) a two step-estimator which first estimates (i) and then uses this preliminary estimator[17] to estimate the optimal weighting matrix (16), which is then used to reestimate θ_0 (**Càglàd TS**); (iii) the generalised method of L-moments estimator that uses the unbiased L-moment estimates (5) and identity weights (**Unbiased FS**); and (iv) the two-step estimator that uses the unbiased L-moment estimator in the first and second steps (**Unbiased TS**). The estimator of the optimal-weighting matrix we use is given in Supplemental Appendix E.

4.1. Generalised Extreme value distribution (GEV)

Following Hosking et al. (1985) and Hosking (1990), we consider the family of distributions

$$F_\theta(z) = \begin{cases} \exp\{-[1-\theta_2(x-\theta_1)/\theta_3]^{1/\theta_3}\}, & \theta_3 \neq 0 \\ \exp\{-\exp(-(x-\theta_1)/\theta_2)\}, & \theta_3 = 0, \end{cases}$$

and $\theta_0 = (0, 1, -0.2)'$.

Table 1 reports the RMSE of each procedure, divided by the RMSE of the MLE, under the choice of R that achieves the smallest RMSE. Values above 1 indicate the MLE outperforms the estimator in consideration; and values below 1 indicate the estimator outperforms MLE. The value of R that minimises the RMSE is presented under parentheses. Some patterns are worth highlighting. Firstly, the L-moment estimator, under a proper choice of R and (estimated) optimal weights (two-step estimators) is able to outperform MLE in most settings, especially at the tail of the distribution function. Reductions in these settings can be as large as 31.9%. At the median, two-step L-moment estimators behave similarly to the MLE. The performance of two-step càglàd and unbiased estimators is also quite similar. Secondly, the power of overidentifying restrictions is evident: except in three out of twenty-four cases, two-step L-moment estimators never achieve a minimum RMSE at $R = 3$, the number of parameters. Two of these three exceptions are found at the smallest sample size ($T = 50$), where the benefit of overidentifying restrictions may be outweighed by noisy estimation of the weighting matrix.[18] Thirdly, the relationship between T and the MSE-minimising choice of R in the two-step Càglàd estimator is monotonic when we move from the smallest ($T = 50$) to the largest ($T = 500$) sample size.[19] This is

[16] To be precise, our choice of weights does not coincide with actual identity weights. Given that the coefficients of Legendre polynomials rapidly scale with R – and that this increase generates convergence problems in the numerical optimisation – we work directly with the underlying estimators of the probability-weighted moments $\int_0^1 Q_Y(u)u^r du$ (Landwehr et al., 1979), of which L-moments are linear combinations. When (estimated) optimal weights are used, such approach is without loss, since the optimal weights for L-moments constitute a mere rotation of the optimal weights for probability-weighted moments, in such a way that the optimally-weighted objective function for L-moments and probability-weighted moments coincide. In other cases, however, this is not the case: a choice of identity weights when probability-weighted moments are directly targeted coincides with using $\boldsymbol{D}^{-1'}\boldsymbol{D}^{-1}$ as a weighting matrix for L-moments, where $\boldsymbol{D}$ is a matrix which translates the first R probability-weighted moments onto the first R L-moments. For small R, we have experimented with using the "true" L-moment estimator with identity weights, and have obtained the same patterns presented in the text.

[17] This preliminary estimator is computed with $R = d$.

[18] Indeed, as we discuss in Supplemental Appendix K, a higher-order expansion of our proposed estimator shows that correlation of the estimator of the optimal weighting matrix with sample L-moments plays a key role in the higher-order bias and variance of the two-step estimator.

[19] The optimal choice of R also increases in three out of four quantiles when we move from $T = 50$ to $T = 100$. The exception occurs at the median, where the optimal choice decreases slightly from $R = 12$ to $R = 11$. At $T = 100$, the difference between $R = 11$ and $R = 12$ is negligible, though: choosing $R = 11$ leads to a relative RMSE of 1.002787, whereas the choice $R = 12$ leads to a relative RMSE of 1.002800.

Table 1
GEV : relative RMSE under MSE-minimising choice of R.

	$T = 50$				$T = 100$				$T = 500$			
	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
Càglàd FS	1.026 (3)	0.962 (3)	0.821 (3)	0.737 (3)	1.031 (3)	0.982 (4)	0.950 (3)	0.928 (3)	1.028 (3)	1.000 (8)	1.061 (3)	1.095 (3)
Càglàd TS	1.005 (12)	0.960 (3)	0.818 (5)	0.692 (5)	1.003 (11)	0.981 (3)	0.910 (5)	0.840 (5)	1.004 (30)	0.998 (4)	0.990 (90)	0.979 (90)
Unbiased FS	1.016 (3)	0.951 (5)	0.853 (3)	0.811 (3)	1.027 (3)	0.975 (6)	0.972 (3)	0.979 (3)	1.027 (3)	0.998 (9)	1.065 (3)	1.106 (3)
Unbiased TS	1.000 (21)	0.950 (3)	0.815 (5)	0.681 (9)	1.000 (8)	0.976 (4)	0.904 (5)	0.834 (5)	1.003 (29)	0.994 (7)	0.985 (21)	0.974 (21)

consistent with our theoretical results: given $\sqrt{T}$-consistency of the estimators, as T increases, one expects the contribution of the bias component in the RMSE to decrease, and, given asymptotic efficiency of the two-step estimator as R diverges, a larger choice of R may lead to variance reduction. Finally, the role of optimal weights is clear: first-step estimators tend to underperform the MLE as the sample size increases. In larger samples, and when optimal weights are not used, the best choice tends to be setting R close to or equal to 3, which reinforces the importance of weighting when overidentifying restrictions are included.

To better understand the patterns in the table, we report in Fig. 1, the relative RMSE curve for different sample sizes and choices of R. The role of optimal weights is especially striking: first-step estimators usually exhibit an increasing RMSE, as a function of R. In contrast, two-step estimators are able to better control the RMSE across R. It is also interesting to note that the two-step unbiased L-moment estimator behaves poorly when R is close to T. This suggests that, in settings where one may wish to make R large, the càglàd estimator is preferable.[20] Finally, we note that, for two-step estimators, the RMSE curve is relatively flat over several regions of R. This implies that, if R is chosen in these regions, then the RMSE of the resulting estimator is robust to (local) perturbations on the number of L-moments used in estimation. As we discuss in Section 5, this flatness will be convenient when designing methods to automatically select R, since any method that sets this tuning parameter to be in the correct *region* where RMSE is small should perform well. In contrast, if the RMSE curve were locally very sensitive to the choice of R, it could be unfeasible to obtain a sufficiently accurate assessment of the RMSE in finite samples that were to result in a good choice of R.

4.2. Generalised Pareto distribution (GPD)

Following Hosking and Wallis (1987), we consider the family of distributions:

$$F_\theta(z) = \begin{cases} 1 - (1 - \theta_2 x/\theta_1)^{-1/\theta_2}, & \theta_2 \neq 0 \\ 1 - \exp(-x/\theta_1), & \theta_2 = 0, \end{cases}$$

and $\theta_0 = (1, -0.2)'$.

Table 2 and Fig. 2 summarise the results of our simulation. Overall patterns are similar to the ones obtained in the GEV simulations. Importantly, though, estimation of the optimal weighting matrix impacts two-step estimators quite negatively in this setup. As a consequence, we verify that the choice of $R = 2$ (i.e. a just-identified estimator that effectively does not rely on the weights) is optimal for TS estimators at five out of the eight cases in sample size $T = 50$. This behaviour also leads to FS estimators, which do not use estimated weights, outperforming TS estimators at the 0.99 and 0.999 quantiles when $T = 50$, and underperforming TS estimators in larger sample sizes by much smaller margins than in the GEV design. Finally, we note that, in all cases, L-moment estimators compare favourably to the MLE.

Remark 10 (*Other Target Parameters*)**.** In this section, we have focused in a setting where the goal is quantile estimation. In Supplemental Appendix J.1, we consider instead a situation where the targets are linear combinations $\delta'\theta_0$ of the model parameters. Since we do not have any particular linear combination in mind, we consider choices of δ (directions) that lead to the most and least favourable relative RMSE vis-à-vis the MLE. In the GEV design, two-step estimators, under the optimal choice of R, are able to offer RMSE improvements of around 8% in the most favourable direction and smaller sample sizes, while strongly mitigating the underperformance of first-step estimators in the least-favourable directions and larger sample sizes. Indeed, in the latter scenario, first-step estimators, even under an optimal choice of R, incur in RMSE losses relatively to the MLE of over 25%; in contrast, this underperformance shrinks to only 1.3% when càglàd two-step estimators are adopted. In the GPD design, both first-step and two-step estimators perform well relatively to the MLE, even when considering the least favourable directions and largest sample sizes, with gains reaching over 16% in the smallest sample size and most favourable direction (and 4% in the smallest sample size and least favourable direction).

[20] This is also in accordance with our theoretical results for the càglàd-based estimator, which essentially place no restriction on the growth rate of R.

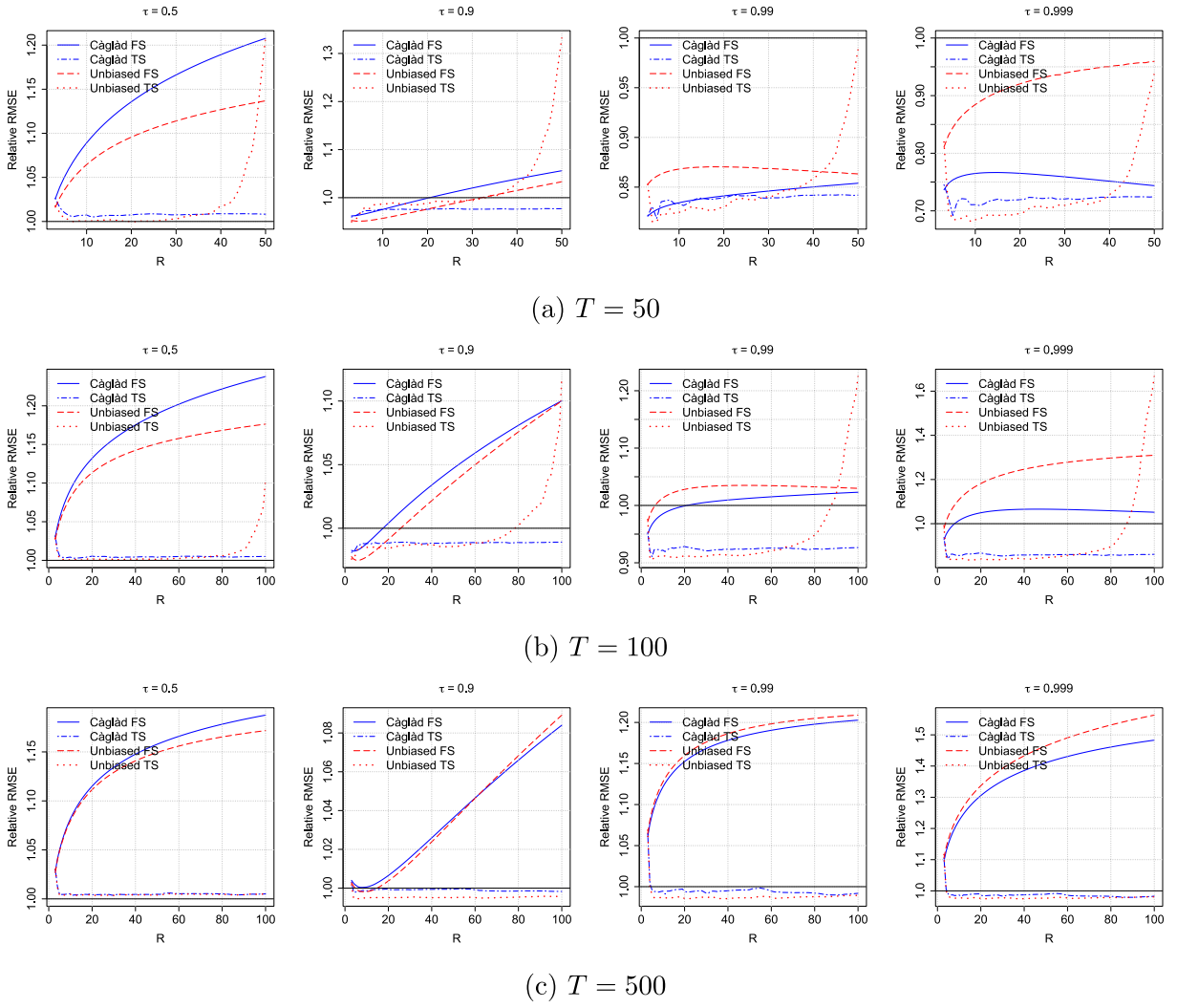

Fig. 1. GEV: relative RMSE for different choices of R.

Table 2
GPD : relative RMSE under MSE-minimising choice of R.

	$T = 50$				$T = 100$				$T = 500$			
	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
Càglàd FS	0.975 (4)	0.981 (3)	0.806 (34)	0.596 (50)	0.982 (4)	0.991 (7)	0.915 (4)	0.856 (3)	0.998 (4)	1.000 (23)	0.990 (3)	0.982 (2)
Càglàd TS	0.959 (3)	0.981 (2)	0.822 (3)	0.649 (2)	0.978 (3)	0.987 (3)	0.899 (3)	0.837 (3)	0.995 (5)	0.997 (100)	0.980 (3)	0.969 (100)
Unbiased FS	0.942 (4)	0.967 (5)	0.834 (10)	0.711 (4)	0.967 (4)	0.982 (12)	0.931 (3)	0.910 (2)	0.995 (4)	0.997 (28)	0.992 (2)	0.993 (2)
Unbiased TS	0.929 (5)	0.967 (2)	0.845 (2)	0.717 (2)	0.962 (5)	0.980 (3)	0.914 (3)	0.887 (3)	0.991 (39)	0.995 (27)	0.975 (27)	0.974 (27)

Remark 11 (*Comparison with Trimming and Tilting Approaches*)**.** In both of our Monte Carlo exercises, the distributions exhibit heavy tails. In these settings, a natural approach would be to consider maximum likelihood estimators that take additional steps to limit the influence of extreme observations. We compare the behaviour of these estimators with our L-moment-based approach in Supplemental Appendix J.2. Specifically, we contrast our L-moment-based estimators with a trimming approach that discards extreme observations and computes MLE estimates in a restricted dataset, and also with a "tilted" MLE that computes estimates in a reweighted dataset. In both the GEV and GDP designs, the Càglàd TS estimator under the RMSE-minimising choice of R consistently

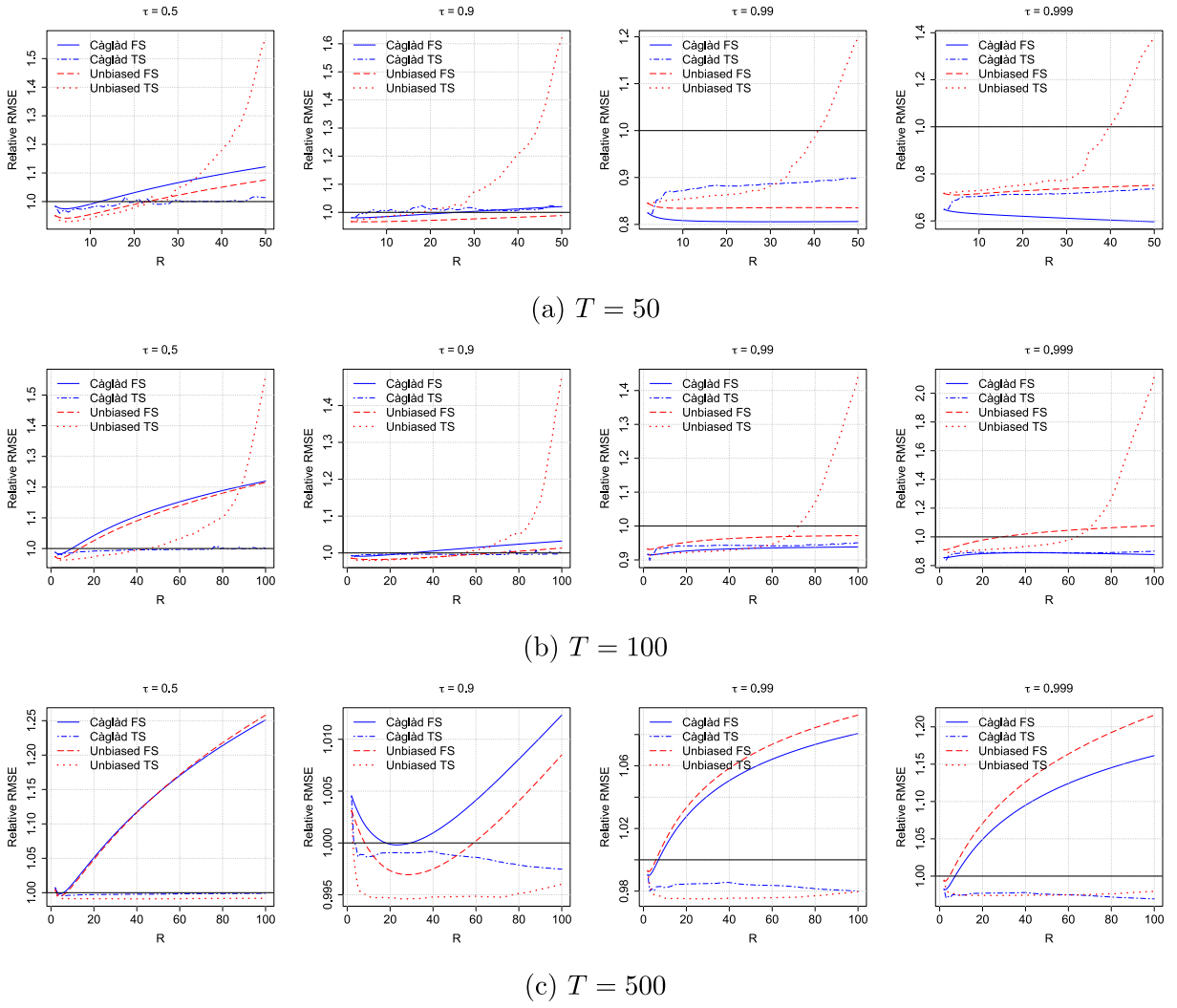

Fig. 2. GPD: relative RMSE for different choices of R.

outperforms the trimming approach. As for the tilted MLE, it is able to compete with the L-moment estimator at some combinations of tail quantiles and sample sizes, under a suitable amount of tilting. However, the competitiveness and overall performance of the tilted MLE estimator is extremely sensitive to the amount of tilting to which the data is subjected to, and, as we discuss in the Appendix, to the best of our knowledge, there currently do not exist methods to select the tilting proportion with an aim at MSE reductions.

Remark 12 (*Size and Coverage of Confidence Intervals Based on the Gaussian Approximation*)**.** Supplemental Appendix J.3 assesses the coverage and length of confidence intervals (CIs) derived from the normal approximation in Corollaries 1 and 2. We focus on the càglàd two-step estimator and provide two sets of results. First, to assess the quality of the normal approximations, we compute the coverage and length of confidence intervals for different target quantiles based on normal critical values and the true sampling variance of the estimators. We observe that these confidence intervals have coverage close to their nominal level even in the smallest sample size and more extreme quantiles, and that, consistent with our theoretical results that do not impose any growth restrictions on the number of L-moments in the derivation of the asymptotic normal approximation, coverage is approximately constant across choices of R. Confidence intervals based on the MLE estimator, normal critical values and the true sampling variance also exhibit correct coverage; however, their length is no less than the length of confidence intervals based on the generalised L-moment estimator under the length-minimising choice of R. Length reductions provided by the L-moment-based CIs can be substantial, especially at tail quantiles. We then compare the coverage and length of feasible versions of these confidence intervals that rely on estimators of the asymptotic variance. Both the MLE and L-moment-based confidence intervals display correct coverage at central quantiles in small sample sizes and at the tail when we consider larger samples. However, both methods display undercoverage at more extreme

quantiles in smaller sample sizes, with L-moment CIs in some cases undercovering more than the MLE in the GEV design (differences in undercoverage are insubstantial in the GPD design). As we argue in the Appendix, this is driven partly by correlation between the asymptotic variance estimator and the target quantile estimator, which generates distortions in the sampling distribution of the t-test that is inverted to construct the confidence interval. Motivated by our strong approximation results, we provide a simple correction to the critical values used in the L-moment confidence interval that improves coverage in smaller sample sizes and more extreme quantiles, while still preserving reduced length over (coverage-corrected) MLE-based CIs.

Remark 13. We note that computation runtime of our L-moment estimators is quite fast in the GEV and GPD families. For example, the estimation of the parameters of a GEV distribution with the two-step càglàd L-moment estimator and a random sample with $T = 500$ observations takes around half a second in a 2017 i7 Macbook Pro with 16 GB RAM when $R = 100$; around three seconds when $R = 500$; and around ten seconds when $R = 1000$.

5. Choosing the number of L-moments in estimation

The simulation exercise in the previous section evidences that the number of L-moments R plays an important role in determining the relative behaviour of the generalised L-moment estimator. Indeed, Figs. 1 and 2 suggest that the RMSE of two-step estimators can be sensitive to the number of L-moments. For example, in the GEV design, at $T = 500$ and $\tau = 0.999$, the RMSE of the two-step unbiased L-moment estimator is around 11% *larger* than the MLE when $R = 3$, and around 2.3% *smaller* than the MLE when $R = 6$. This indicates that designing a proper method to select R is essential for competitiveness of the L-moment approach. Moreover, the pattern of the curves in Figs. 1 and 2 suggests that there is great hope that such methods will perform well in practice. Indeed, given that the RMSE curve is flat over several regions of R, one should expect any method that sets R to be in an appropriate *region* where RMSE is small to perform well.[21] In contrast, if the RMSE curve were locally very sensitive to the choice of the number of L-moments, then one would require a rather sharp assessment of the RMSE to select R, which could be unfeasible in smaller sample sizes.

In light of these points, in this section we introduce (semi)automatic methods to select R. We briefly outline two approaches, with the details being left to Supplemental Appendix K. We then contrast these approaches in the context of the Monte Carlo exercise of Section 4.

In Supplemental Appendix K.1, we derive a higher-order expansion of the "generalised" L-moment estimator (7). We then propose to choose R by minimising the resulting higher-order mean-squared error of a suitable linear combination of the parameters. Similar approaches were considered in the GMM literature by Donald and Newey (2001) – where the goal is to choose the number of instruments in linear instrumental variable models –, and Donald et al. (2009) – where one wishes to choose moment conditions in models defined by conditional moment restrictions (in which case infinitely many restrictions are available). Relatedly, Okui (2009) considers the choice of moments in dynamic panel data models; and, more recently, Abadie et al. (2024) use higher order expansions to develop a method of choosing subsamples in linear instrumental variables models with first-stage heterogeneity. Importantly, our higher-order expansions can be used to provide higher-order mean-squared error estimates of target estimands $g_T(\theta_0)$, where g_T is a function indexed by sample size. This can be useful when the parameter θ_0 is not of direct interest. So, for example, if our goal is quantile estimation, we can choose R so as to minimise the higher-order mean-squared error of estimating the target quantile.

In Supplemental Appendix K.2, we consider an alternative approach to selecting L-moments by employing ℓ_1-regularisation. Following Luo et al. (2015), we note that the first order condition of the estimator (7) may be written as:

$$A_R h^R(\hat{\theta}) = 0\,,$$

for a $d \times R$ matrix A_R which combines the L-moments linearly into d restrictions. The idea is to estimate A_R using a Lasso penalty. This approach implicitly performs moment selection, as the method yields exact zeros for several entries of A_R. In a GMM context, Luo et al. (2015) introduces an easy-to-implement quadratic program for estimating A_R with the Lasso regularisation. In the Supplemental Appendix, we show how this algorithm may be extended to our L-moment setting and provide conditions for its validity.

To conclude, we return to the Monte Carlo exercise of Section 4. We contrast the RMSE (relatively to the MLE) of the original L-moment estimator due to Hosking (1990) that sets $R = d$ (**FS**) with a two-step L-moment estimator where R is chosen so as to minimise a higher-order MSE of the target quantile (**TS RMSE**), and a "post-lasso" estimator that estimates θ_0 using only those L-moments selected by regularised estimation of A_R (**TS Post-Lasso**). For brevity, we focus on estimators based on the *càglàd* L-moments (3). Additional details on the implementation of each method can be found in Supplemental Appendix K.3.

Tables 3 and 4 present the results of the different methods in the GEV and GPD exercises. We report in parentheses the average number of L-moments used by each estimator. Overall, the TS RMSE estimator compares favourably to both Hosking's original estimator and the MLE. In the GEV exercise, the TS RMSE estimator improves upon both the FS estimator and the MLE when $T < 500$; and behaves similarly to the MLE and better than FS in the largest sample size. For example, at $T = 500$, Hosking's estimator has a 6.1% (9.5%) larger root-mean-squared error than then MLE at the 0.99 (0.999) quantile, whereas the relative performance of TS RMSE with respect to the MLE is 0.6% (0.9%). As for the GPD exercise, recall that this is a setting where estimation of the optimal weighting matrix impacts two-step estimators more negatively. Consequently, gains of TS RMSE over FS are more limited in this

[21] We thank two anonymous referees for pointing this out.

Table 3
GEV : relative RMSE under different selection procedures.

	$T = 50$				$T = 100$				$T = 500$			
	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
FS	1.026 (3)	0.962 (3)	0.821 (3)	0.737 (3)	1.031 (3)	0.983 (3)	0.950 (3)	0.928 (3)	1.028 (3)	1.004 (3)	1.061 (3)	1.095 (3)
TS RMSE	1.008 (16.81)	0.964 (3.66)	0.794 (3.3)	0.674 (3.41)	1.004 (33.02)	0.987 (4.17)	0.923 (4.32)	0.865 (4.57)	1.005 (20.29)	0.999 (35.51)	1.006 (40.91)	1.009 (43.17)
TS Post-Lasso	1.017 (7.99)	0.975 (7.99)	0.857 (7.99)	0.781 (7.99)	1.010 (9.24)	0.988 (9.24)	0.928 (9.24)	0.866 (9.24)	1.006 (9.95)	0.999 (9.95)	0.999 (9.95)	0.993 (9.95)

Table 4
GPD : relative RMSE under different selection procedures.

	$T = 50$				$T = 100$				$T = 500$			
	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
FS	0.984 (2)	0.981 (2)	0.824 (2)	0.648 (2)	0.988 (2)	0.993 (2)	0.917 (2)	0.856 (2)	1.007 (2)	1.005 (2)	0.990 (2)	0.982 (2)
TS RMSE	0.964 (2.86)	0.994 (4.43)	0.817 (2.61)	0.640 (2.96)	0.980 (3.59)	0.990 (4.31)	0.896 (3.02)	0.828 (3.09)	0.997 (5.34)	0.999 (48.42)	0.978 (31.11)	0.970 (29.71)
TS Post-Lasso	0.995 (3.52)	0.999 (3.52)	0.891 (3.52)	0.741 (3.52)	0.992 (3.7)	0.999 (3.7)	0.950 (3.7)	0.905 (3.7)	0.998 (3.78)	1.000 (3.78)	0.985 (3.78)	0.977 (3.78)

setting. Indeed, the average gain of TS RMSE over FS in the GPD exercise is 1.0 (relative) percentage points (pp), with the largest outperformance being 2.8 pp and an underperformance in $T = 50$ and $\tau = 0.9$ of 1.3 pp. In contrast, in the GEV exercise, the average gain of TS RMSE over FS is 3.2 pp, the largest gain is 8.6 pp and TS RMSE underperforms FS by only 0.4 pp at $T = 100$ and $\tau = 0.9$. More importantly, in both settings, our TS RMSE approach is able to *simultaneously* generate gains over MLE in smaller samples and mitigate inefficiencies of Hosking's original method in larger sample sizes. This phenomenon is especially pronounced at the tails of the distributions.

With regards to the Post-Lasso method, we note that it behaves similarly to TS RMSE in larger sample sizes,[22] though it can perform somewhat unfavourably vis-à-vis the other L-moment alternatives in the smallest sample size (the method still improves upon MLE at tail quantiles in this scenario). As we discuss in Supplemental Appendix K.3, this issue can be partly attributed to a "harsh" regularisation penalty being used in the selection step. There is room for improving this step by relying on an iterative procedure to select a less harsh penalty (see Belloni et al. (2012) and Luo et al. (2015) for examples). We also discuss in the Supplemental Appendix that it could be possible to improve the TS RMSE procedure by including additional higher-order terms in the estimated approximate RMSE. We leave exploration of these improvements as future topics of research.

Remark 14. We remark that comparisons between the RMSE and post-Lasso approaches should also take computational concerns into consideration. Indeed, as summarised in the pseudo-code in the Supplemental Appendix (Algorithm K.1), our numerical implementation of the RMSE approach requires evaluation of cross-products, for different test values of R, between the gradient and Hessian of the theoretical L-moment functions at different choices of R (which measure the sensitivity of estimates to the sample L-moments); the partial derivatives, with respect to θ, of the quantile density function $Q'(u|\theta)$ at different values of u (measuring higher-order terms pertaining to estimation of the optimal weighting matrix); and the gradient and Hessian, with respect to θ, of the quantile function $Q(\tau|\theta)$ at the quantiles τ of interest, when the goal is quantile estimation (pertaining to the expansion of the RMSE of the target quantile). Even though these derivatives are available in closed form for the GEV and GPD families (see Supplemental Appendix M), and while we do provide R code that leverages fast automatic differentiation tools to evaluate the derivatives when these expressions are not available in closed form, computation of the RMSE approach is generally slower than the Post-Lasso method. Indeed, while the post-Lasso also requires computation of derivatives to estimate the Lasso penalty (see Supplemental Appendix K.3 for details), it does not hinge on the evaluation of cross-products of these terms for different choices of R, which speeds up implementation considerably.[23] For comparison, in the GEV Monte Carlo exercise, with 500 observations, computing the higher-order RMSE estimate for the four target quantiles across test values $R \in \{3, \dots, 100\}$ takes around 42 s in a 2017 i7 Macbook Pro with 16 GB RAM (and around 10 s for test values $R \in \{3, \dots, 50\}$). In contrast, the Post-Lasso selection approach with a maximum allowed choice of $R_{\max} = 200$ takes around 6 s (and around 2 s with $R_{\max} = 100$). Given that the Post-Lasso approach compares favourably to TS RMSE in sample sizes $T > 50$, it may thus be preferable in these settings on computational grounds. One further advantage of this approach is its simplicity, as it delivers a single choice of R irrespective of the target parameter.

[22] In the largest sample size of the GEV distribution, the Post-Lasso performs especially well, incurring in a gain of 10.2 pp over the FS estimator at the $\tau = 0.999$ quantile.

[23] We remark that evaluation of the cross-products in the RMSE approach can be parallelised across different test values of R, an approach we adopt in our computational implementation. See the R script `selection.R` in the accompanying online repository for a generic implementation of our selection methods to any class of parametric distributions.

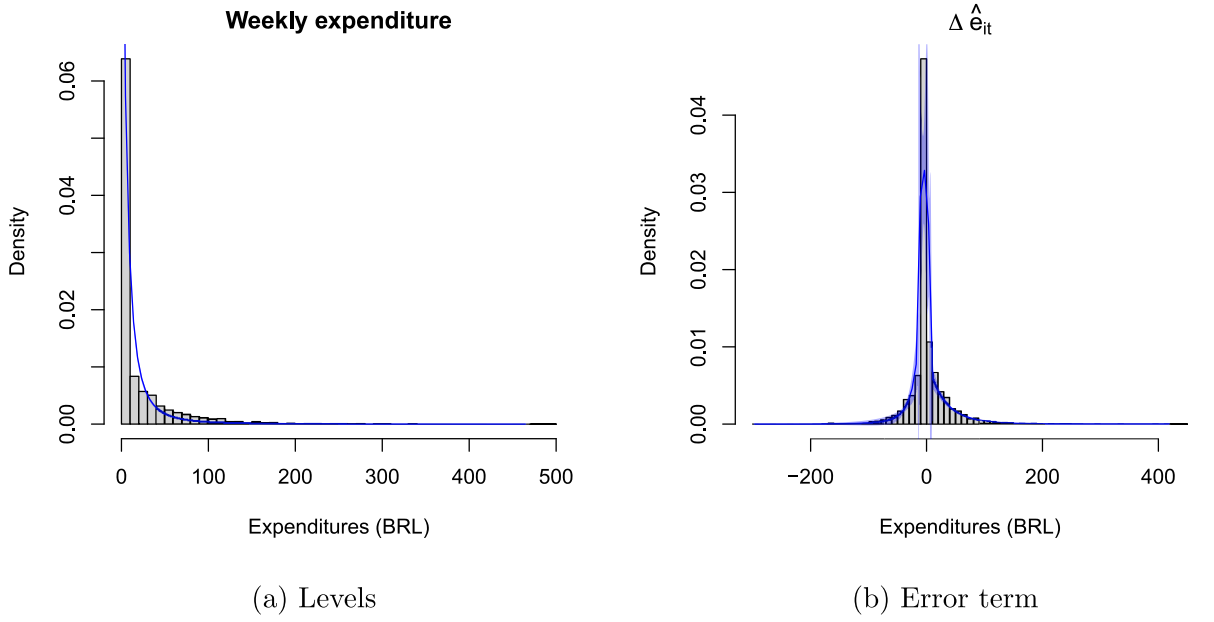

Fig. 3. Empirical application: expenditure patterns in mid-to-late September.

6. Extensions

6.1. "Residual" analysis in semi- and nonparametric models

In this subsection, we consider a setting where a researcher has postulated a model for a scalar real-valued outcome Y:

$$Y = h(\epsilon, X; \gamma_0), \quad \gamma \in \Gamma \subseteq \mathcal{B}, \tag{18}$$

where h is a known mapping, X is a vector of observable attributes taking values in $\mathcal{X}$, ϵ is an unobservable *scalar* real-valued disturbance, and γ_0 is a nuisance parameter that is known to belong to a subset Γ of a Banach space $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$. We assume that, for each possible value $(x, \gamma) \in \mathcal{X} \times \Gamma$, the map $e \mapsto h(e, x; \gamma)$ is invertible, and we denote its pointwise inverse by $h^{-1}(\cdot, x; \gamma)$.

We further assume that the researcher has access to an estimator of γ_0, and that her goal is to estimate a parametric model for the distribution of ϵ, i.e. she considers the model:

$$\epsilon \sim F_{\theta_0}, \quad \theta_0 \in \Theta \subseteq \mathbb{R}^d. \tag{19}$$

Interest in (19) nests different types of "residual" analyses, where one may wish to estimate (19) with an aim to (indirectly) assess the appropriateness of (18) – whenever theory imposes restrictions on the distribution of ϵ –, or as a means to construct unconditional prediction intervals for Y.

In Supplemental Appendix L.1, we show how our generalised L-moment approach may be adapted to estimate (19), while remaining agnostic about the first-step estimator of γ_0. We do so by borrowing insights from the double-machine learning literature (Chernozhukov et al., 2018, 2022; Kennedy, 2023). Specifically, we employ sample-splitting and debiasing to construct the generalised method-of-L-moment estimator:

$$\hat{\theta} \in \arg\inf_{\theta \in \Theta} \left[\int_{\underline{p}}^{\bar{p}} \left(\hat{Q}_{\hat{\epsilon}}(u) - Q_{\epsilon}(u|\theta) \right) \mathbf{P}^R(u)' du - \hat{\boldsymbol{A}} \right] W^R \left[\int_{\underline{p}}^{\bar{p}} \left(\hat{Q}_{\hat{\epsilon}}(u) - Q_Y(u|\theta) \right) \mathbf{P}^R(u) du - \hat{\boldsymbol{A}} \right], \tag{20}$$

where $\hat{Q}_{\hat{\epsilon}}(u)$ is the empirical quantile function of $\{h^{-1}(Y_i, X_i; \hat{\gamma}) : i = 1, \ldots, T\}$, with $\hat{\gamma}$ a first-step estimator of γ_0 computed from a sample independently from $\{(X_i, Y_i) : i = 1, \ldots, T\}$. The adjustment term $\hat{\boldsymbol{A}}$ is an estimator of the first-step influence function (Ichimura and Newey, 2022), which reflects the impact of estimating $\hat{\gamma}$ on $\gamma \mapsto \hat{Q}_{h^{-1}(Y,X;\gamma)}$. The Supplemental Appendix provides the form of this correction in three examples. We also show that the asymptotic distribution of $\hat{\theta}$ may be computed as in the previous sections, provided that we adjust it to account for first-step estimation error. This correction can also be used to compute the optimal-weighting scheme of L-moments.

To illustrate this approach, we rely on expenditure data in a ridesharing platform collected by Biderman (2018). We observe weekly expenditures (in Brazilian reais) in the platform during eight weeks between August and September 2018, for a subset of 3961 users of the service in the municipality of São Paulo, Brazil. For these users, we also have access to survey data on sociodemographic traits and commuting motives. We denote by Y_{it} the amount spent by user i in week t, whereas X_i collects their survey information.

Panel 3(a) plots the histogram for the distribution of expenditures in the penultimate week of our sample (mid-to-late September). The data clearly exhibits heavy tails: the maximum observed expenditure is 464.34 Brazilian reais, whereas average expenditure amounts to 20.81 reais.[24] Moreover, 58% of individuals do not spend any money in rides during this period. The solid line presents the density of a GPD distribution fit to this data. The parameters of the distribution are estimated via the two-step generalised method-of L-moment estimator discussed in Section 4, with $R = 65$, which corresponds to the optimal choice for estimating several quantiles across the distribution, according to the RMSE criterion discussed in Section 5. Even though the overidentifying restrictions test clearly rejects the null (p-value ≈ 0), we take the plotted density as further confirmation of heavy-tailedness of expenditure patterns, since the estimated GPD density understates mass at larger support points.

We seek to understand whether individual time-invariant heterogeneity, along with persistence in expenditure patterns, is able to explain the observed heavy-tailedness. To accomplish this, we posit the following model for the evolution of expenditures:

$$Y_{it} = \alpha_i + a(X_i)t + b(X_i)Y_{i,t-1} + \epsilon_{i,t}\,,$$

where α_i is unobserved time-invariant heterogeneity (here treated as a fixed effect), and ϵ_{it} is time-varying idiosyncratic heterogeneity that is assumed to be, conditionally on X_i, independent across time. The coefficients $a(X_i)$ and $b(X_i)$ measure respectively deterministic trends and persistence in individual consumption patterns. We allow these coefficients to be nonparametric functions of survey information. Finally, we also assume that $\mathbb{E}[\epsilon_{it}|X_i] = 0$, meaning that $(a(X_i), b(X_i))$ correctly capture mean heterogeneity in consumption trends attributable to X_i.

To estimate the above model, we take first-differences to remove the fixed effect, i.e. we consider:

$$\Delta Y_{i,t} = a(X_i) + b(X_i)\Delta Y_{i,t-1} + \Delta\epsilon_{i,t}\,. \tag{21}$$

Under the assumption that the $\epsilon_{i,t}$ are (conditionally on X_i) independent across time, we may then use $Y_{i,t-2}$ as a valid instrument for the endogenous variable $\Delta Y_{i,t-1}$ (Anderson and Hsiao, 1982; Arellano and Bond, 1991). We estimate (21) using the instrumental forest estimator of Athey et al. (2019), which assumes a and b to be in the closure of the linear span of regression trees.

Panel 3(b) reports the histogram of the residuals $\Delta\hat{\epsilon}_{it}$ in mid-to-late September, where we adopt sample-splitting and estimate the functions $a(\cdot)$ and $b(\cdot)$ using data from the weeks prior to the penultimate week in the sample. The distribution is two-sided, with a large mass just below zero, suggesting that model (21) somewhat overpredicts expenditure variation in mid-to-late September. To assess whether the data exhibits heavy-tails, we estimate a GEV mixture model for the distribution of $\Delta\epsilon_{it}$, assuming that $\mathbb{P}[\Delta\epsilon_{it} \leq x] = \omega((1-\delta_1)/2+\delta_1 F_{\theta_1}(\delta_1 x))+(1-\omega)((1-\delta_2)/2+\delta_2 F_{\theta_1}(\delta_2 x))$, with $\omega \in [0,1]$, $\delta_1, \delta_2 \in \{-1, 1\}$, and F_{θ_1} and F_{θ_2} belonging to the GEV family described in Section 4. Our formulation allows for the left- and right-tails to exhibit different decay, e.g. if $\delta_1 = -1$ and $\delta_2 = -1$ and the shape parameter of F_{θ_1} is negative while F_{θ_1} is positive, then the left-tail behaves as a Fréchet, while the right-tail behaves as a Weibull distribution. Moreover, if the distributions being mixed by the weights ω exhibit disjoint supports, then the quantile function of the mixture admits a simple closed-form solution (Castellacci, 2012), which enables us to rely on closed-form expressions for the L-moments of the GEV family to compute theoretical L-moments. We estimate the parameters $(\omega, \delta_1, \delta_2, \theta_1, \theta_2)$ by relying on the adjusted estimator (20), with two-step optimal weights and the form of correction $\hat{\boldsymbol{A}}$ derived in Example 3 in Supplemental Appendix L.1.

The solid line in Panel 3(b) reports the fitted density of the GEV mixture, while the shaded area plots a 95% uniform confidence band for the density function over the support of $\Delta\hat{\epsilon}_{it}$. The band is computed using the delta-method and sup-t critical values (Freyberger and Rai, 2018). Overall, the fit appears adequate, as evidenced by the overidentifying restrictions test not rejecting the null at the usual significance levels (p-value ≈ 1). We then use our parameter estimates to test the null hypothesis that the left (right) tail exhibits exponentially light decay, against the alternative that it is heavy-tailed. This corresponds to testing whether the left-tail (right-tail) shape parameter is in the set $[0, 1]$, against the alternative that it is not.[25] Upon computation of a 95% confidence interval for the right- (left-) tail shape parameters, we verify that, while the confidence region for the right-tail shape parameter is entirely contained in the $(1, \infty)$ region, the confidence region for the left-tail shape parameter is $[-0.934, 0.124]$, which intersects with $[0, 1]$. Therefore, at the 5% significance level, we reject the null of exponentially light decay for the right-tail of the distribution, though we fail to do so for the left-tail. These results provide evidence that, even after accounting for heterogeneous persistence and trends, as well as time-invariant unobserved heterogeneity, weekly expenses still exhibit very positive idiosyncratic realisations. Such pattern is consistent with, in any given week, some individuals having to take very long trips (e.g. taking a ride to the airport), the demand for which may be hard to anticipate on the basis of observable traits.[26]

As a final application of our approach, we show how our estimator of the distribution of $\Delta\epsilon_{it}$ may be used to construct prediction intervals for individual treatment effects (Cattaneo et al., 2021; Chernozhukov et al., 2021a,b), a useful tool in assessing the impacts of personalised interventions (Kivaranovic et al., 2020; Lei and Candès, 2021). Specifically, suppose that, in some week, the ridesharing company implements a personalised policy in the platform, e.g. a change in a parameter of the pricing algorithm that may result in disparate fees being charged across users. In this causal inference setting, the model (21) may be seen as a model

[24] For completeness, in late September 2018, 1USD = 4 Brazilian reais.

[25] When the shape parameter equals zero, the GEV distribution collapses to a Gumbel distribution, which has exponentially light tails (Chernozhukov and Fernández-Val, 2011). When the shape parameter is strictly greater than zero but smaller than one, the distribution collapses to a Weibull distribution with shape parameter greater than one, a region for which the Weibull is known to have exponentially light tails (Foss et al., 2011). The region $(-\infty, 0)$ corresponds to a Fréchet distribution, whereas the region $(1, \infty)$ corresponds to a Weibull with shape parameter strictly greater than zero and strictly less than unity – both cases corresponding to heavy-tailed distributions.

[26] We thank a referee for providing this interpretation of the conclusion of the test.

for the untreated potential outcome, $\Delta Y_{i,t}(0)$, that is observed in the periods t prior to the intervention at t^*, where t^* is a random variable denoting the date when the intervention starts. Under the assumption that t^* is independent of $\Delta\epsilon_{i,t^*}$,[27] the interval:

$$\hat{I}_{i,1-\alpha} = [\Delta Y_{i,t^*} - \hat{a}(X_i) - \hat{b}(X_i)\Delta Y_{i,t^*-1} - Q_{\hat{\theta}}(1-\alpha), \infty),$$

is an asymptotically (in a regime where the number of users diverges) valid $(1-\alpha)$ prediction region for the individual treatment effect $Y_{it^*}(1) - Y_{it^*}(0)$, where $\hat{a}$, $\hat{b}$ and $\hat{\theta}$ are estimators computed using the pre-intervention sample, and asymptotic validity is meant as, when the number of users diverges:

$$\mathbb{P}[Y_{it^*}(1) - Y_{it^*}(0) \in \hat{I}_{i,1-\alpha}] \to 1-\alpha,$$

(see Supplemental Appendix L.2 for details). The lower-bound of our one-sided prediction region has a Value-at-Risk type interpretation, representing the largest loss c with the policy that cannot be rejected, in a test of the null $H_0 : Y_{it^*}(1) - Y_{it^*}(0) \leq c$ against the alternative $H_1 : Y_{it^*}(1) - Y_{it^*}(0) > c$, at the α significance level. Two-sided intervals can also be considered, as well as Bonferroni-style corrections to $\hat{I}_{i,1-\alpha}$ in order to account for estimation error of $\hat{a}$, $\hat{b}$ and $\hat{\theta}$ (see Cattaneo et al. (2021) and Supplemental Appendix L.2 for details).

As an illustration of our approach to constructing prediction intervals, we implement the intervals $\hat{I}_{i,0.95}$ in our data, assuming that the last week of September is the post-treatment period. Given that there was no known intervention in this period (i.e. $Y_{it^*}(1) - Y_{it^*}(0)$ is known to be zero for every unit), one would expect that these regions would contain 0 in approximately 95% of the cases. This is indeed what we observe in the data: out of the 3961 users in our sample, the corresponding individual prediction intervals do not contain zero in only 175 (4.4%) of the cases.

6.2. Conditional quantile models

Let $Q_{Y|X}(\cdot|X)$ be the quantile function of a conditional distribution function $F_{Y|X}(\cdot|X)$, where Y is a scalar outcome and X is a set of controls. Following Gourieroux and Jasiak (2008), we define the rth conditional L-moment as:

$$\lambda_r(X) := \int_{\underline{p}}^{\overline{p}} Q_{Y|X}(u|X) P_r(u) du. \tag{22}$$

When $0 = \underline{p} \leq \overline{p} = 1$, Gourieroux and Jasiak (2008) note that $\lambda_r(X) = \mathbb{E}[Y P_l(F_{Y|X}(Y|X))|X]$. They suggest estimating $F_{Y|X}$ nonparametrically, and, for a fixed number R of L-moments, to exploit the following RK unconditional moments in the estimation of conditional parametric models $\{Q_{Y|X}(\cdot|\cdot;\theta) : \theta \in \Theta\}$:

$$\mathbb{E}\left[w(X) \otimes \left(Y \boldsymbol{P}^R(F_{Y|X}(Y|X)) - \int_0^1 Q_{Y|X}(u|X;\theta_0) \boldsymbol{P}^R(u) du \right) \right] = 0, \tag{23}$$

where $w(X)$ is a $R \times 1$ vector of transformations of X.

In spite of its conceptual attractiveness – L-moment estimation is cast as method-of-moment estimation –, formulation (23) does not directly extend to settings with trimming.[28] Moreover, by working with fixed R and K, it does not fully exploit the identifying information in the parametric model. In light of these points, Supplemental Appendix L.3 proposes an alternative method-of-L-moment estimator for conditional models. We propose to estimate (22) by directly plugging into the representation a nonparametric conditional quantile estimator. Following Ai and Chen (2003), we then optimally exploit the conditional L-moment restrictions by weighting these using $R \times R$ weighting functionals $\Omega(X)$. In the Supplemental Appendix, we consider the case where we rely on the quantile series regression estimator of Belloni et al. (2019) for preliminary nonparametric estimation, though in principle any nonparametric estimator of conditional quantile functions for which an approximation theory is available could be used in this first step. Examples of such estimators include local polynomial quantile regression (Yu and Jones, 1998; Guerre and Sabbah, 2012) and quantile regression forests (Meinshausen, 2006; Athey et al., 2019). Under regularity conditions, our estimator admits an asymptotic linear representation that can be used as a basis for inference, and for finding the optimal choice of functional $\Omega(X)$. Moreover, by suitably taking $R \to \infty$, we expect our optimally-weighted estimator to achieve good finite-sample performance, whilst retaining asymptotic efficiency – indeed, as we argue in the Supplemental Appendix, the optimally-weighted estimator with no trimming and an orthonormal *basis* choice of $\{P_l\}_l$ is asymptotically efficient as $T, R \to \infty$.

[27] In our example, this assumption allows the decision of when to implement the policy to depend on the values of α_i, X_i and the $Y_{i,t}$ prior to the treatment, but essentially excludes the possibility of the ridesharing platform, in its decision of when to change the pricing parameter, to rely on better predictions of post-treatment values of $\Delta Y_{it}(0)$ than those obtained from the predictable part of (21), as this would introduce dependence between t^* and the post-treatment variation in idiosyncratic components driving demand absent the intervention ($\Delta\epsilon_{i,t^*}$). See Ferman and Pinto (2021) and Alvarez and Ferman (2024) for a discussion on the interpretation of similar assumptions in synthetic control designs.

[28] To implement trimming in this formulation would require nonparametric estimation of both the conditional distribution and conditional quantile functions, whereas our suggested approach solely relies on the latter.

7. Concluding remarks

This paper considered the estimation of parametric models using a "generalised" method of L-moments procedure, which extends the approach introduced in Hosking (1990) whereby a d-dimensional parametric model for a distribution function is fit by matching the first d L-moments. We have shown that, by appropriately choosing the number of L-moments and under an appropriate weighting scheme, we are able to construct an estimator that is able to outperform maximum likelihood estimation in small samples from popular distributions, and yet does not suffer from efficiency losses in larger samples. We have developed tools to automatically select the number of L-moments used in estimation, and have shown the usefulness of such approach in Monte Carlo simulations. We have also extended our L-moment approach to the estimation of conditional models, and to the "residual analysis" of semiparametric models. We then applied the latter to study expenditure patterns in a ridesharing platform in São Paulo, Brazil.

The extension of the generalised L-moment approach to other semi- and nonparametric settings appears to be an interesting venue of future research. The L-moment approach appears especially well-suited to problems where semi- and nonparametric maximum likelihood estimation is computationally complicated, but evaluation of integrals of quantiles is not. In followup work, Alvarez and Orestes (2023) propose using the generalised method-of-L-moment approach to estimate nonparametric quantile mixture models, while Alvarez and Biderman (2024) introduce an efficient generalised L-moment estimator for the semiparametric models of treatment effects of Athey et al. (2023). The study of such extensions in more generality is a topic for future research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Earlier working paper versions of this article circulated as "Inference in parametric models with many L-moments". We thank Cristine Pinto, Marcelo Fernandes, Marcos Fernandes, Ricardo Masini, Silvia Ferrari and Victor Orestes and two anonymous referees for their useful comments and suggestions. We are also thankful to Ciro Biderman for sharing his dataset with us. All the remaining errors are ours. Luis Alvarez gratefully acknowledges financial support from Capes, Brazil grant 88887.353415/2019-00 and Cnpq, Brazil grant 141881/2020-8. Chang Chiann and Pedro Morettin gratefully acknowledge financial support from Fapesp, Brazil grant 2018/04654-9 and 2023/02538-0.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2025.106101.

References

Abadie, A., Gu, J., Shen, S., 2024. Instrumental variable estimation with first-stage heterogeneity. J. Econometrics 240 (2), 105425, URL https://www.sciencedirect.com/science/article/pii/S0304407623000702.

Ahidar-Coutrix, A., Berthet, P., 2016. Convergence of multivariate quantile surfaces. arXiv:1607.02604.

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71 (6), 1795–1843, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00470.

Alvarez, L., Biderman, C., 2024. The learning effects of subsidies to bundled goods: a semiparametric approach. arXiv:2311.01217.

Alvarez, L.A.F., Ferman, B., 2024. On "imputation of counterfactual outcomes when the errors are predictable": Discussions on misspecification and suggestions of sensitivity analyses. J. Bus. Econom. Statist. 42 (4), 1123–1127, URL https://doi.org/10.1080/07350015.2024.2359594.

Alvarez, L., Orestes, V., 2023. Quantile Mixture Models: Estimation and Inference. Working paper.

Anderson, T., Hsiao, C., 1982. Formulation and estimation of dynamic models using panel data. J. Econometrics 18 (1), 47–82, URL https://www.sciencedirect.com/science/article/pii/0304407682900951.

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econ. Stud. 58 (2), 277–297, URL https://academic.oup.com/restud/article-pdf/58/2/277/4454458/58-2-277.pdf.

Athey, S., Bickel, P.J., Chen, A., Imbens, G.W., Pollmann, M., 2023. Semi-parametric estimation of treatment effects in randomised experiments. J. R. Stat. Soc. Ser. B Stat. Methodol. 85 (5), 1615–1638, URL https://academic.oup.com/jrsssb/article-pdf/85/5/1615/56555154/qkad072.pdf.

Athey, S., Tibshirani, J., Wager, S., 2019. Generalized random forests. Ann. Statist. 47 (2), 1148–1178.

Barrio, E.D., Giné, E., Utzet, F., 2005. Asymptotics for L2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. Bernoulli 11 (1), 131–189.

Belloni, A., Chen, D., Chernozhukov, V., Hansen, C., 2012. Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica 80 (6), 2369–2429, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA9626. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA9626.

Belloni, A., Chernozhukov, V., Chetverikov, D., Fernández-Val, I., 2019. Conditional quantile processes based on series or many regressors. J. Econometrics 213 (1), 4–29, URL http://www.sciencedirect.com/science/article/pii/S0304407619300582. Annals: In Honor of Roger Koenker.

Biderman, C., 2018. E-hailing and last mile connectivity in Sao Paulo. Am. Econ. Assoc. Regist. Randomized Control. Trials AEARCTR-0003518.

Billingsley, P., 2012. Probability and Measure. Wiley.

Boulange, J., Hanasaki, N., Yamazaki, D., Pokhrel, Y., 2021. Role of dams in reducing global flood exposure under climate change. Nat. Commun. 12 (1), 1–7.

Broniatowski, M., Decurninge, A., 2016. Estimation for models defined by conditions on their L-moments. IEEE Trans. Inform. Theory 62 (9), 5181–5198.

Carrasco, M., Florens, J.-P., 2014. On the asymptotic efficiency of GMM. Econometric Theory 30 (2), 372–406.

Castellacci, G., 2012. A formula for the quantiles of mixtures of distributions with disjoint supports. Available at SSRN 2055022.

Cattaneo, M.D., Feng, Y., and, R.T., 2021. Prediction intervals for synthetic control methods. J. Amer. Statist. Assoc. 116 (536), 1865–1880, URL https://doi.org/10.1080/01621459.2021.1979561. PMID: 35756161.
Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J., 2018. Double/debiased machine learning for treatment and structural parameters. Econom. J. 21 (1), C1–C68, URL https://academic.oup.com/ectj/article-pdf/21/1/C1/27684918/ectj00c1.pdf.
Chernozhukov, V., Escanciano, J.C., Ichimura, H., Newey, W.K., Robins, J.M., 2022. Locally robust semiparametric estimation. Econometrica 90 (4), 1501–1535, URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA16294.
Chernozhukov, V., Fernández-Val, I., 2011. Inference for extremal conditional quantile models, with an application to market and birthweight risks. Rev. Econ. Stud. 78 (2), 559–589.
Chernozhukov, V., Wüthrich, K., Zhu, Y., 2021a. Distributional conformal prediction. Proc. Natl. Acad. Sci. 118 (48), e2107794118, URL https://www.pnas.org/doi/abs/10.1073/pnas.2107794118.
Chernozhukov, V., Wüthrich, K., Zhu, Y., 2021b. An exact and robust conformal inference method for counterfactual and synthetic controls. J. Amer. Statist. Assoc. 116 (536), 1849–1864, URL https://doi.org/10.1080/01621459.2021.1920957.
Crump, R.K., Hotz, V.J., Imbens, G.W., Mitnik, O.A., 2009. Dealing with limited overlap in estimation of average treatment effects. Biometrika 96 (1), 187–199, URL https://academic.oup.com/biomet/article-pdf/96/1/187/642537/asn055.pdf.
Csorgo, M., Revesz, P., 1978. Strong approximations of the quantile process. Ann. Stat. 6 (4), 882–894.
Das, S., 2021. Extreme rainfall estimation at ungauged locations: information that needs to be included in low-lying monsoon climate regions like Bangladesh. J. Hydrol. 601, 126616.
Dey, S., Mazucheli, J., Nadarajah, S., 2018. Kumaraswamy distribution: different methods of estimation. Comput. Appl. Math. 37 (2), 2094–2111, URL http://link.springer.com/10.1007/s40314-017-0441-1.
Donald, S.G., Imbens, G.W., Newey, W.K., 2003. Empirical likelihood estimation and consistent tests with conditional moment restrictions. J. Econometrics 117 (1), 55–93, URL https://www.sciencedirect.com/science/article/pii/S0304407603001180.
Donald, S.G., Imbens, G.W., Newey, W.K., 2009. Choosing instrumental variables in conditional moment restriction models. J. Econometrics 152 (1), 28–36, URL https://www.sciencedirect.com/science/article/pii/S0304407609000566.
Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. Econometrica 69 (5), 1161–1191, URL http://www.jstor.org/stable/2692218.
Ferman, B., Pinto, C., 2021. Synthetic controls with imperfect pretreatment fit. Quant. Econ. 12 (4), 1197–1221, URL https://onlinelibrary.wiley.com/doi/abs/10.3982/QE1596.
Ferrari, D., Yang, Y., 2010. Maximum L q -likelihood estimation. Ann. Stat. 38 (2), 753–783.
Foss, S., Korshunov, D., Zachary, S., et al., 2011. An Introduction to Heavy-Tailed and Subexponential Distributions. Vol. 6, Springer.
Fotopoulos, S., Ahn, S., 1994. Strong approximation of the quantile processes and its applications under strong mixing properties. J. Multivariate Anal. 51 (1), 17–45.
Freyberger, J., Rai, Y., 2018. Uniform confidence bands: Characterization and optimality. J. Econometrics 204 (1), 119–130, URL https://www.sciencedirect.com/science/article/pii/S0304407618300174.
Götze, F., Naumov, A., Spokoiny, V., Ulyanov, V., 2019. Large ball probabilities, Gaussian comparison and anti-concentration. Bernoulli 25 (4A), 2538–2563.
Gourieroux, C., Jasiak, J., 2008. Dynamic quantile models. J. Econometrics 147 (1), 198–205.
Guerre, E., Sabbah, C., 2012. Uniform bias study and bahadur representation for local polynomial estimators of the conditional quantile function. Econometric Theory 28 (1), 87–129, URL http://www.jstor.org/stable/41426508.
Gupta, R.D., Kundu, D., 2001. Generalized exponential distribution: Different method of estimations. J. Stat. Comput. Simul. 69 (4), 315–337.
Hahn, J., Kuersteiner, G., Newey, W., 2002. Higher order properties of bootstrap and jackknife bias corrected maximum likelihood estimators. Unpubl. Manuscr..
Han, C., Phillips, P.C.B., 2006. GMM with many moment conditions. Econometrica 74 (1), 147–192, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2006.00652.x.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054, URL http://www.jstor.org/stable/1912775.
Hosking, J.R., 1986. The Theory of Probability Weighted Moments. IBM Research Division, TJ Watson Research Center New York, USA.
Hosking, J.R.M., 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. J. R. Stat. Soc. Ser. B Stat. Methodol. 52 (1), 105–124, URL http://doi.wiley.com/10.1111/j.2517-6161.1990.tb01775.x.
Hosking, J.R., 2007. Some theory and practical uses of trimmed L-moments. J. Statist. Plann. Inference 137 (9), 3024–3039.
Hosking, J.R.M., 2024. L-moments. URL https://CRAN.R-project.org/package=lmom. R package, version 3.2.
Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29 (3), 339–349, URL http://www.jstor.org/stable/1269343.
Hosking, J.R.M., Wallis, J.R., Wood, E.F., 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27 (3), 251–261, URL https://www.tandfonline.com/doi/abs/10.1080/00401706.1985.10488049.
Ichimura, H., Newey, W.K., 2022. The influence function of semiparametric estimators. Quant. Econ. 13 (1), 29–61, URL https://onlinelibrary.wiley.com/doi/abs/10.3982/QE826.
Karvanen, J., 2006. Estimation of quantile mixtures via L-moments and trimmed L-moments. Comput. Statist. Data Anal. 51 (2), 947–959.
Kennedy, E.H., 2023. Semiparametric doubly robust targeted double machine learning: a review. arXiv:2203.06469.
Kerstens, K., Mounir, A., De Woestyne, I.V., 2011. Non-parametric frontier estimates of mutual fund performance using C- and L-moments: Some specification tests. J. Bank. Financ. 35 (5), 1190–1201.
Kivaranovic, D., Ristl, R., Posch, M., Leeb, H., 2020. Conformal prediction intervals for the individual treatment effect. arXiv:2006.01474.
Koenker, R., Machado, J.A., 1999. GMM inference when the number of moment conditions is large. J. Econometrics 93 (2), 327–344, URL https://www.sciencedirect.com/science/article/pii/S0304407699000147.
Komunjer, I., Vuong, Q., 2010. Semiparametric efficiency bound in time-series models for conditional quantiles. Econometric Theory 26 (2), 383–405, URL http://www.jstor.org/stable/40664470.
Kreyszig, E., 1989. Introductory Functional Analysis with Applications. Wiley.
Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1979. Probability weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles. Water Resour. Res. 15 (5), 1055–1064, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR015i005p01055.
Lei, L., Candès, E.J., 2021. Conformal inference of counterfactuals and individual treatment effects. J. R. Stat. Soc. Ser. B Stat. Methodol. 83 (5), 911–938, URL https://academic.oup.com/jrsssb/article-pdf/83/5/911/49322261/jrsssb_83_5_911.pdf.
Li, Y., Li, R., Qin, Y., Lin, C., Yang, Y., 2021a. Robust group variable screening based on maximum Lq-likelihood estimation. Stat. Med. 40 (30), 6818–6834, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.9212.
Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., Wehner, M., 2021b. Changes in annual extremes of daily temperature and precipitation in CMIP6 models. J. Clim. 34 (9), 3441–3460.
Luo, Y., et al., 2015. High-Dimensional Econometrics and Model Selection (Ph.D. thesis). Massachusetts Institute of Technology.
Mason, D.M., 1984. Weak convergence of the weighted empirical quantile process in $L^2(0,1)$. Ann. Probab. 12 (1), 243–255.
Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7 (35), 983–999, URL http://jmlr.org/papers/v7/meinshausen06a.html.

Newey, W.K., 1990. Semiparametric efficiency bounds. J. Appl. Econometrics 5 (2), 99–135, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.3950050202.

Newey, W.K., McFadden, D., 1994. Chapter 36 large sample estimation and hypothesis testing. In: Handbook of Econometrics. Vol. 4, Elsevier, pp. 2111–2245, URL http://www.sciencedirect.com/science/article/pii/S1573441205800054.

Newey, W.K., Smith, R.J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica 72 (1), 219–255, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2004.00482.x.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708, URL http://www.jstor.org/stable/1913610.

Okui, R., 2009. The optimal choice of moments in dynamic panel data models. J. Econometrics 151 (1), 1–16, URL https://EconPapers.repec.org/RePEc:eee:econom:v:151:y:2009:i:1:p:1-16.

Parzen, E., 1960. Modern Probability Theory and its Applications. Wiley.

Pfanzagl, J., Wefelmeyer, W., 1978. A third-order optimum property of the maximum likelihood estimator. J. Multivariate Anal. 8 (1), 1–29, URL https://www.sciencedirect.com/science/article/pii/0047259X78900167.

Portnoy, S., 1991. Asymptotic behavior of regression quantiles in non-stationary, dependent cases. J. Multivariate Anal. 38 (1), 100–113, URL http://www.sciencedirect.com/science/article/pii/0047259X9190034Y.

Rothenberg, T.J., 1971. Identification in parametric models. Econometrica 39 (3), 577–591, URL https://EconPapers.repec.org/RePEc:ecm:emetrp:v:39:y:1971:i:3:p:577-91.

Sankarasubramanian, A., Srinivasan, K., 1999. Investigation and comparison of sampling properties of L-moments and conventional moments. J. Hydrol. 218 (1–2), 13–34.

Šimková, T., 2017. Statistical inference based on L-moments. Statistika 97 (1), 44–58.

van der Vaart, A.W., 1998. Asymptotic Statistics. In: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer.

Wang, Q.J., 1997. LH moments for statistical analysis of extreme events. Water Resour. Res. 33 (12), 2841–2848.

Wang, D., Hutson, A.D., 2013. Joint confidence region estimation of L-moment ratios with an extension to right censored data. J. Appl. Stat. 40 (2), 368–379.

Wooldridge, J., 2010. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, Mass.

Yang, J., Bian, Z., Liu, J., Jiang, B., Lu, W., Gao, X., Song, H., 2021. No-reference quality assessment for screen content images using visual edge model and AdaBoosting neural network. IEEE Trans. Image Process. 30, 6801–6814.

Yoshihara, K., 1995. The Bahadur representation of sample quantiles for sequences of strongly mixing random variables. Statist. Probab. Lett. 24 (4), 299–304, URL http://www.sciencedirect.com/science/article/pii/016771529400187D.

Yu, H., 1996. A note on strong approximation for quantile processes of strong mixing sequences. Statist. Probab. Lett. 30 (1), 1–7, URL http://www.sciencedirect.com/science/article/pii/0167715295001948.

Yu, K., Jones, M.C., 1998. Local linear quantile regression. J. Amer. Statist. Assoc. 93 (441), 228–237, URL http://www.jstor.org/stable/2669619.