



HAL
open science

Conditional Bias Robust Estimation of the Total of Curve Data by Sampling in a Finite Population: An Illustration on Electricity Load Curves

Hervé Cardot, Anne de Moliner, Camelia Goga

► **To cite this version:**

Hervé Cardot, Anne de Moliner, Camelia Goga. Conditional Bias Robust Estimation of the Total of Curve Data by Sampling in a Finite Population: An Illustration on Electricity Load Curves. *Journal of Survey Statistics and Methodology*, 2020, 8 (3), pp.453-482. 10.1093/jssam/smz009 . hal-03035958

HAL Id: hal-03035958

<https://hal.science/hal-03035958>

Submitted on 30 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Conditional bias robust estimation of the total of curve data by sampling in a finite population: an illustration on electricity load curves

Hervé Cardot⁽¹⁾, Anne De Moliner^(1,2) and Camelia Goga⁽³⁾

⁽¹⁾ Institut de Mathématiques de Bourgogne, UMR 5584 CNRS
Université de Bourgogne Franche-Comté, France

⁽²⁾ EDF LAB, Palaiseau, France

⁽³⁾ Laboratoire de Mathématiques de Besançon, UMR 6623 CNRS
Université de Bourgogne Franche-Comté, France

herve.cardot@u-bourgogne.fr, camelia.goga@univ-fcomte.fr

November 5, 2018

Abstract

For marketing or power grid management purposes, many studies based on the analysis of the total electricity consumption curves of groups of customers are now carried out by electricity companies. Aggregated total or mean load curves are estimated using individual curves measured at fine time grid and collected according to some sampling design. Due to the skewness of the distribution of electricity consumptions, these samples often contain outlying curves which may have an important impact on the usual estimation procedures. We introduce several robust estimators of the total consumption curve which are not sensitive to such outlying curves. These estimators are based on the conditional bias approach and robust functional methods. We also derive mean square error estimators of these robust estimators and finally, we evaluate and compare the performance of the suggested estimators on Irish electricity data.

Keywords: bootstrap, conditional bias, functional data, modified band depth, spherical principal component analysis, survey sampling, wavelets.

1 Introduction and context

Many studies carried out by electricity companies are based on the analysis of total electricity consumption curves measured at fine time scales (often half-hourly) for one or several groups of clients sharing some common characteristics (e.g. customers from the same electricity provider, having a particular electric equipment or living in a given geographic area). The aim of these studies can be for example to assist the power grid manager in maintaining the balance between electricity consumption and production at every instant on the

power grid. The total consumption curves can also be used to help the Sales Division to quantify the impact of a specific electric use or equipment on the electricity consumption, to build new innovative pricing strategies or to create new services based on customers consumption analysis.

In order to avoid technical and budgetary constraints due to limited bandwidth or storage cost of huge databases, or in order to preserve privacy, the strategy of selecting a sample of individual curves from the whole datasets is often employed. The total consumption curve or the load curve of each population of interest is then estimated by using the curves of the customers belonging to the sample. The estimation with survey sampling techniques of parameters of interest such as the total or the mean, the median or the principal components when the data are curves has been developed over the last years: Cardot et al. (2010), Cardot and Josserand (2011), Cardot et al. (2013a) and Chaouch and Goga (2012). Several sampling designs and estimators have been compared by means of simulation on real electricity data set in Cardot et al. (2013b) and some asymptotic properties have been established in Cardot et al. (2013c) and Cardot et al. (2014). We cite also Degras (2014) for the Horvitz-Thompson estimation with optimal rotation of samples. A recent review of research works in this area is given in Lardin-Puech et al. (2014).

We address here the estimation of the total consumption curve in presence of outlying curves. Following Chambers (1986), we consider only representative outlying curves, namely curves which are representative for some non-sampled units and that do not come from measurement errors. With electricity data, it is not unusual to have units with consumption electricity much higher than the rest of the population (see Figure 1). Such outlying curves may have a huge impact on the estimation and it is very important to detect and treat them correctly. In order to detect such outlying curves, we use the notion of depth of a curve introduced by López-Pintado and Romo (2009).

In a finite population setting, stratification is a good method to reduce the potential impact of outlying curves. More exactly, the population is divided into disjointed subpopulations called strata and units from the same stratum are as similar as possible according to several criteria. Unfortunately, due to wrong classifications or sudden changes some units may be very different from the other units belonging to the same stratum. These units are influential and deteriorate the stratum homogeneity and the variance of the usual estimators for the total or the mean will be large. More generally, a unit is considered influential if, in a given configuration: study population and variable, parameter, estimator and sampling design, its value has a great impact on the variance of the estimator (Favre-Martinoz (2015)).

Several robust estimators not sensitive to influential units have been suggested in the survey sampling setting for real data, that are not curves. We can cite for example Chambers (1986), Gwet and Rivest (1992), Rivest (1994), Kokic and Bell (1994), Welsh and Ronchetti (1998). Broadly speaking, these estimators are based on winsorization techniques which consist in down-weighting the influence of outlying units. This is performed by considering a thresholding function depending on a tuning constant whose value must

be chosen carefully. The reader is referred to Chapter 11 of Pfeiffermann and Rao (2009) for a detailed presentation of the main methods dealing with outliers in survey data. Recently, Beaumont et al. (2013) considered a new robust estimator for finite population totals. This new approach is based on the notion of conditional bias introduced by Muñoz-Pichardo et al. (1995) to measure the influence of a unit and is closely related to the estimator of Chambers (1986). Besides, the conditional bias approach does not require to introduce a superpopulation model. Another popular approach for building robust estimators for survey data is the one suggested by Kokic and Bell (1994). The use of Kokic and Bell's method would require the knowledge of a model for the probability distribution for functional data. Such superpopulation models are generally very complex in our curve data framework and cannot generally be reduced to parametric models with a small number of parameters. A recent comparison of robust estimation strategies in a finite population by Deroyon and Martinoz (2018) has also shown that a misspecification of the superpopulation model can deteriorate much the accuracy of the robust estimator based on Kokic and Bell's approach. For these two reasons, we did not consider further the Kokic and Bell's approach in the present work.

The aim of this paper is to build robust design-based estimators of the total consumption curves which are less sensitive to influential curves. Since generally the curve data are observed at a finite number of time instants, the easiest and most intuitive way to construct such a robust estimator is to apply the method suggested by Beaumont et al. (2013) at each instant of time. Unfortunately, this method does not take into account possible temporal correlations. In order to deal with this issue, we can transform the data by using dimension reduction methods such as functional principal component analysis or projection of the data onto basis functions. We suggest in this paper to perform a robust principal component analysis (PCA) as introduced by Locantore et al. (1999) in order to obtain uncorrelated real principal components. The total consumption curve may be then approximated in a smaller dimensional space spanned by robust eigenfunctions. Then, the coordinates in this new robust basis can be robustified by using the method of Beaumont et al. (2013) and a second robust estimator for the total consumption is then obtained.

Instead of using robust PCA, one may also project the data onto a basis functions, such as wavelets, which are known to be effective to deal with irregular temporal signals such as individual electricity load curves (see *e.g.* Mallat (1998)). A third robust estimator for the total consumption is then obtained by robustifying the coordinates in the wavelet basis.

The choice of a positive cut-off constant c is required to build these estimators. Choosing an adequate value is crucial since a trade-off between bias and variance must be made. We suggest in this paper a new criterion for choosing this tuning constant based on the q th power of the conditional bias. We also introduce a functional truncation method based on the concept of depth of curves (López-Pintado and Romo (2009)) as a functional measure of outlyingness. This method consists in finding a zone which entirely contains the conditional biases considered as “inliers” and to use the upper and lower bounds of this zone as

truncation limits. A fourth estimator may then be constructed.

This paper is organized as follows: in Section 2, we describe the estimation of totals with sampling designs from a finite population of curves and we extend the definition of the conditional bias for functional data. In Section 3, we apply point-wisely the approach of Beaumont et al. (2013) for building a robust estimator for the total curve and we use their minimax criterion for choosing the tuning constant as well as a new one based on the q th power of the conditional bias. We introduce in Section 4 two robust estimators based on dimension reduction techniques and in Section 5, a robust estimator built for the functional truncation method based on the modified band depth as suggested by López-Pintado and Romo (2009). In Section 6, we address the question of the estimation of pointwise mean square error. Due to confidentiality reasons the electricity data from EDF can not be used for publication. We illustrate, in Section 7, the performances of the different robust approaches on the estimation of the total curve on Irish electricity consumption curves. Concluding remarks are given in brief Section 8 and some proofs are postponed in an Appendix.

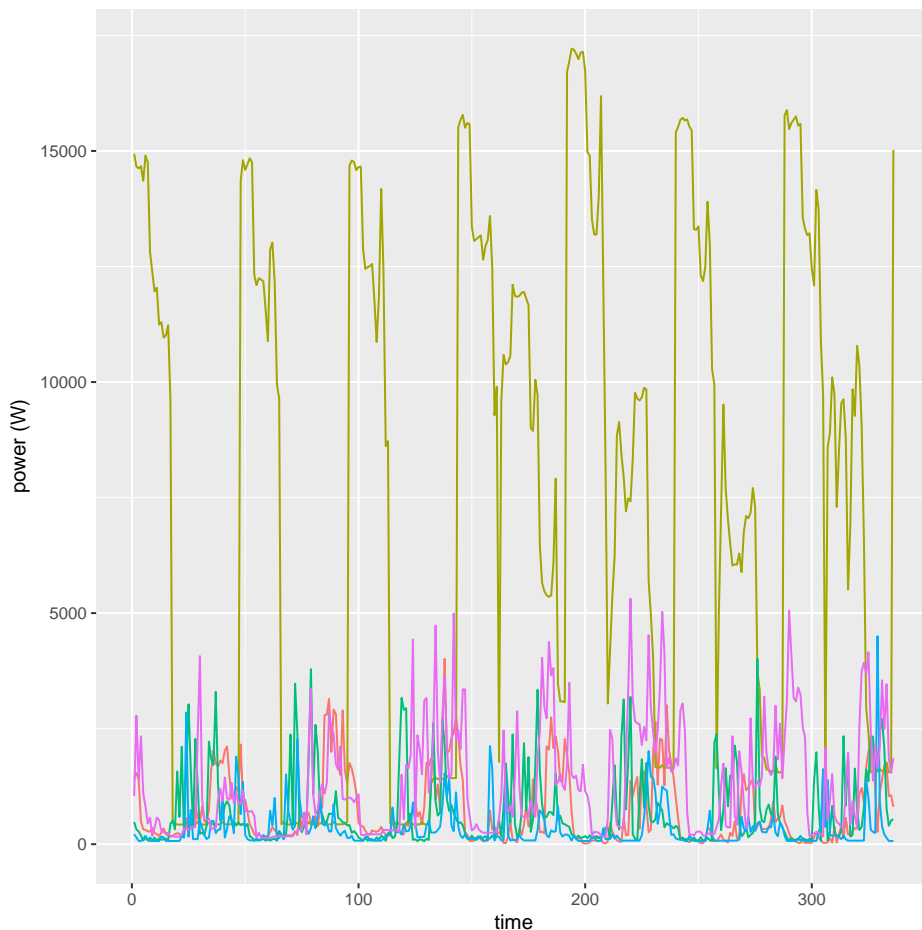


Figure 1: A sample of five load electricity curves measured every half an hour over a period of one week.

2 Robust estimation in a finite population of curves

2.1 Notations and framework

Let U be a population of interest of known size N . To each unit i of the population we associate a (load) curve defined over a time interval $[0, T]$: for each unit i we have a function of time $Y_i(t), t \in [0, T]$, where the continuous index t represents time.

Our goal is to estimate the total curve t_Y over the population:

$$t_Y = \sum_{i \in U} Y_i, \quad (1)$$

with value $t_Y(t) = \sum_{i \in U} Y_i(t)$ for each instant $t \in [0, T]$. In practice, the curves are not observed continuously for $t \in [0, T]$ but only for a set of D measurement instants $0 = t_1 < t_2 < \dots < t_D = T$ which are generally supposed to be equi-spaced and the same for all units. Under weak assumptions on the number of discretization points, the regularity of the trajectories and the sampling design, it can be shown that the approximation error due to linear interpolation or kernel smoothing is negligible compared to the sampling error (see Cardot and Josserand (2011) and Cardot et al. (2013a)).

To evaluate and compare the different approaches, we consider in this work a test population composed of $N = 3994$ electricity consumption curves extracted from the Irish Commission for Energy Regulation (CER) Smart Metering Project that was conducted in 2009-2010 (CER, 2011)¹. The electricity consumptions are recorded during one week, from the 18th to the 24th of January 2010, we have $D = 336$ points in time (see Section 7 for more details). We display in Figure 1, the electricity consumption curves for five smart meters selected from that population.

A sample s of size n is selected from U according to a random sampling design $p(\cdot)$. We denote by I_i the sample membership indicator of unit i which is equal to 1 if the unit i belongs to the sample s and zero otherwise. The probability that unit i will be included in a sample is denoted by $\pi_i = P(I_i = 1) = \sum_{s, i \in s} p(s)$ and the probability that both of the units i and j will be included is denoted by $\pi_{ij} = P(I_i I_j = 1) = \sum_{s, (i, j) \in s} p(s)$. The first-order inclusion probabilities π_i and the second-order inclusion probabilities π_{ij} are assumed to be known and strictly positive. We also assume that π_i and π_{ij} do not depend on time t .

We will particularly be interested by two simple sampling designs, simple random sampling without replacement (SRS) and stratified sampling with simple random sampling within strata (STR). In STR, units with similar characteristics (according to some auxiliary information) are grouped into disjointed strata U_h of size N_h for $h = 1, \dots, H$. A simple random sampling without replacement s_h of size n_h is selected from U_h and the selection in one stratum is independent of the selection in all other strata. Note also that in the following, inference is made under the design-based approach in a finite population setting. This means that the sample membership indicators $I_i, i \in U$ are binary random variables

¹The data are available on request at the address:

<http://www.ucd.ie/issda/data/commissionforenergyregulation/>

and the values of the variable of interest Y_i are treated as being deterministic. In this context, the total curve t_Y can be estimated by the Horvitz-Thompson estimator,

$$\hat{t}_Y(t) = \sum_{i \in s} d_i Y_i(t), \quad t \in [0, T], \quad (2)$$

where $d_i = 1/\pi_i$, $i \in U$ are the sampling weights. The Horvitz-Thompson estimator \hat{t}_Y is a random curve, with covariance function given by

$$\text{Cov}(\hat{t}_Y(r), \hat{t}_Y(t)) = \sum_{i \in U} \sum_{j \in U} (\pi_{ij} - \pi_i \pi_j) \frac{Y_i(r)}{\pi_i} \frac{Y_j(t)}{\pi_j}, \quad \text{for all } r, t \in [0, T].$$

A unit i with a large sampling weight d_i and a large value of Y_i at some time instant t is influent for the Horvitz-Thompson estimator given in (2) since it increases considerably the covariance of the Horvitz-Thompson estimator given above.

2.2 Conditional bias when the data are curves

In order to construct robust estimators, Beaumont et al. (2013) have used the conditional bias as a tool for quantifying the influence of sampled and non sampled units on an estimator. The conditional bias, as defined by Beaumont et al. (2013) is, in a design-based approach, the expectation of the estimator conditionally to the inclusion indicator I_i of the unit i . In our context, the conditional bias of a sampled unit is a function of time t ,

$$B_{1i}^{HT}(t) = E_p(\hat{t}_Y(t) | I_i = 1) - t_Y(t) = \sum_{j \in U} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) Y_j(t), \quad t \in [0, T], \quad (3)$$

and for a non-sampled unit:

$$B_{0i}^{HT}(t) = E_p(\hat{t}_Y(t) | I_i = 0) - t_Y(t) = -\frac{1}{d_i - 1} B_{1i}^{HT}(t), \quad t \in [0, T], \quad (4)$$

where E_p is the expectation with respect to the sampling design p . For simple random sampling without replacement (SRS), the conditional bias have the following expression,

$$B_{1i}^{HT}(t) = \frac{N}{N-1} \left(\frac{N}{n} - 1 \right) (Y_i(t) - \bar{Y}_U(t)), \quad i \in U, \quad t \in [0, T],$$

where $\bar{Y}_U(t) = \sum_{i \in U} Y_i(t)/N$, and for stratified sampling with SRS within each stratum (STR), the conditional bias of a sampled unit i belonging to the stratum U_h is

$$B_{1i}^{HT}(t) = \frac{N_h}{N_h - 1} \left(\frac{N_h}{n_h} - 1 \right) (Y_i(t) - \bar{Y}_{U_h}(t)), \quad i \in U_h, \quad t \in [0, T],$$

where $\bar{Y}_{U_h}(t) = \sum_{i \in U_h} Y_i(t)/N_h$ is the mean curve within stratum h . We can see that for stratified sampling, a unit $i \in U_h$ has a large influence if its value $Y_i(t)$ is far from the mean stratum $\bar{Y}_{U_h}(t)$ and its influence is even larger if it is associated with a large sampling weight N_h/n_h .

We can see from (3) that the conditional bias $B_{1i}^{HT}(t)$ is unknown and must be estimated. A conditionally design-unbiased estimator of $B_{1i}^{HT}(t)$, given $I_i = 1$, is:

$$\hat{B}_{1i}^{HT}(t) = \sum_{j \in s} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_j \pi_{ij}} \right) Y_j(t), \quad \text{for all } t \in [0, T]. \quad (5)$$

In the case of SRS sampling, the conditional bias can be estimated by

$$\hat{B}_{1i}^{HT}(t) = \frac{n}{n-1} \left(\frac{N}{n} - 1 \right) (Y_i(t) - \bar{Y}_s(t)), \quad i \in U, \quad t \in [0, T],$$

where $\bar{Y}_s(t) = \sum_{i \in s} Y_i(t)/n$ and for STR sampling, it can be estimated by

$$\hat{B}_{1i}^{HT}(t) = \frac{n_h}{n_h-1} \left(\frac{N_h}{n_h} - 1 \right) (Y_i(t) - \bar{Y}_{s_h}), \quad i \in U_h, \quad t \in [0, T], \quad (6)$$

where $\bar{Y}_{s_h} = \sum_{i \in s_h} Y_i/n_h$ is the sample mean of Y -values within the stratum h .

Consider again the test population of Irish electricity consumption curves. Two estimated conditional bias curves, with simple random sampling of size $n = 200$, are drawn in Figure 2. We can remark on this small example how different, in shape and values, the conditional bias can be from one individual to another and also, from one instant of time to another.

Following the lines of Beaumont et al. (2013), we obtain in our functional setting:

$$\hat{t}_Y(t) - t_Y(t) = \sum_{i \in s} B_{1i}^{HT}(t) + \sum_{i \in U-s} B_{0i}^{HT}(t) + \left(\sum_{i \in s} d_i A_i(t) - \sum_{i \in U} A_i(t) \right), \quad t \in [0, T], \quad (7)$$

where

$$A_i(t) = \frac{-1}{1 - \pi_i} \sum_{j \in U, j \neq i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_j} Y_j(t).$$

The inclusion probabilities not varying with time, it is straightforward to see, as in Beaumont et al. (2013), that the term in parentheses at the right-hand side of (7) is zero for Poisson sampling. Moreover, as shown in the Appendix, under broad assumptions upon the inclusion probabilities which are satisfied by the simple random sampling without replacement and fixed-size high-entropy designs, the term in parentheses at the right-hand side of (7) is negligible in the sense that

$$\sup_{t \in [0, T]} \left| \sum_{i \in s} d_i A_i(t) - \sum_{i \in U} A_i(t) \right| = O_p(n^{-1/2}).$$

Thus, we can consider that

$$\hat{t}_Y(t) \simeq t_Y(t) + \sum_{i \in s} B_{1i}^{HT}(t) + \sum_{i \in U-s} B_{0i}^{HT}(t), \quad t \in [0, T]. \quad (8)$$

The first term at the right-hand side of previous approximation is not random. Consequently, the precision of the estimator will be influenced only by the two other terms in (8).

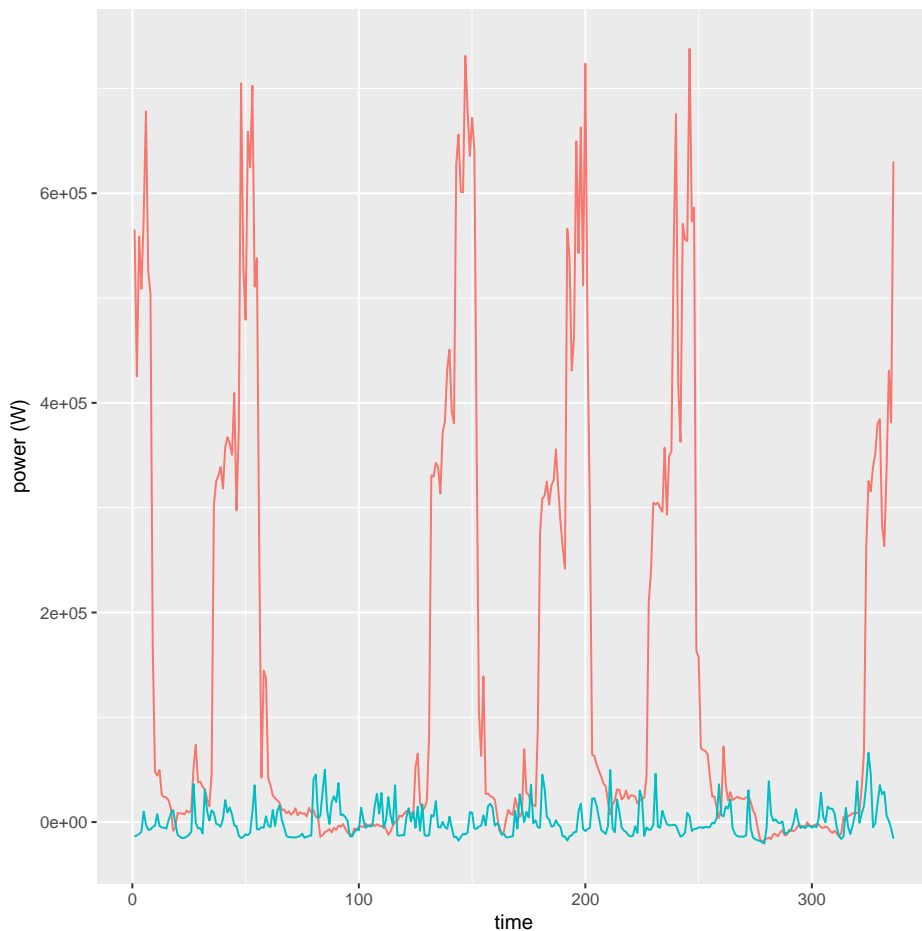


Figure 2: Two conditional bias curves estimated with simple random sampling of size $n = 200$.

The conditional bias of a particular unit can thus be interpreted as the contribution of this unit to the sampling error. An influential unit is defined as a unit with a large conditional bias and the idea is to downplay the impact of such units at the right-hand side of (7). A new challenge, compared to the univariate framework studied in Beaumont et al. (2013) comes from the fact that the conditional bias is now a function of time and as we can note in Figure 2, the shape and the values of the conditional bias can be very different from one individual, or time instant, to another. Different ways of dealing with this issue are developed in the following.

3 Point-wise robust estimators

A first possibility is to directly apply the method of Beaumont et al. (2013) at the D instants t_1, \dots, T_D . Considering the Huber function, $\psi_c(z) = \text{sgn}(z)\min(|z|, c)$ which depends on the tuning constant $c > 0$, with $\text{sgn}(z) = 1$ if $z \geq 0$ and -1 otherwise, we can construct

the following point-wise robust estimator of $t_Y(t)$:

$$\hat{t}_Y^{(R1)}(t) = \hat{t}_Y(t) + \sum_{i \in s} \psi_{c(t)}(B_{1i}^{HT}(t)) - \sum_{i \in s} B_{1i}^{HT}(t) \quad (9)$$

$$= \hat{t}_Y(t) + \Delta(c(t)), \quad \text{for all } t \in [0, T]. \quad (10)$$

So, for a given value $c(t)$ those conditional bias $B_{1i}^{HT}(t)$ larger than $c(t)$ will be cut-off at $c(t)$ in the second-term at the right-hand side of (9). Clearly, the efficiency of the robust estimator depends on the choice of the tuning constant $c(t)$. As $c(t)$ increases, the estimator becomes closer to the non robust estimator. The new estimator $\hat{t}_Y^{(R1)}(t)$ is biased but of smaller variance than that of the non robust one, so we hope to improve the global precision measured by the mean squared error. The trade-off between variance and bias is controlled again by the tuning constant $c(t)$: a large value for $c(t)$ implies small bias but large variance and a small value for $c(t)$ implies large bias and small variance.

3.1 Minimax approach for choosing the optimal tuning constant

We determine the optimal tuning constant in a pointwise manner, namely we determine for each t , the optimal value $c_{opt}(t)$ is chosen according to the minimax approach suggested by Beaumont et al. (2013). The value $c_{opt}(t)$, that is not necessarily unique, satisfies

$$c_{opt}(t) = \arg \min_{c \geq 0} \max_{i \in s} \left| \hat{B}_{1i}^{RHT}(c(t)) \right|, \quad (11)$$

where $\hat{B}_{1i}^{RHT}(c(t))$ is the estimator of the conditional bias of the robust estimator $\hat{t}_y^R(t)$. Using relation (10), the conditional bias of the robust estimator is $B_{1i}^{RHT}(c(t)) = B_{1i}^{HT}(t) + E_p(\Delta(c(t)) | I_i = 1)$ and can be estimated by

$$\hat{B}_{1i}^{RHT}(t) = \hat{B}_{1i}^{HT}(t) + \Delta(c(t)).$$

Following Beaumont et al. (2013), the optimal value of $\Delta(c(t))$ is

$$\Delta(c_{opt}(t)) = -\frac{1}{2} \left(\hat{B}_{min}^{HT}(t) + \hat{B}_{max}^{HT}(t) \right),$$

where $\hat{B}_{min}^{HT}(t) = \min_{i \in s} \hat{B}_i^{HT}(t)$ and $\hat{B}_{max}^{HT}(t) = \max_{i \in s} \hat{B}_i^{HT}(t)$ are the minimum and respectively, the maximum of the estimated absolute conditional biases $\hat{B}_i(t)$ over the sample. The optimal robust estimator is therefore, at each instant t ,

$$\begin{aligned} \hat{t}_{Y_{opt}}^R(t) &= \hat{t}_Y(t) + \Delta(c_{opt}(t)) \\ &= \hat{t}_Y(t) - \frac{1}{2} \left(\hat{B}_{min}^{HT}(t) + \hat{B}_{max}^{HT}(t) \right). \end{aligned} \quad (12)$$

Remark that the optimal tuning constant $c_{opt}(t)$ varies over time, but there is not need to compute it in order to construct the optimal robust estimator. Note also that this method is essentially univariate since it deals independently with the different points in time and thus does not necessarily preserve the correlations between instants in the estimated total curves. We can think that some information is lost by not making use of the strong temporal correlations between the values of Y at different times. A robust estimator which takes into account such possible correlations is presented in Section 4.1.

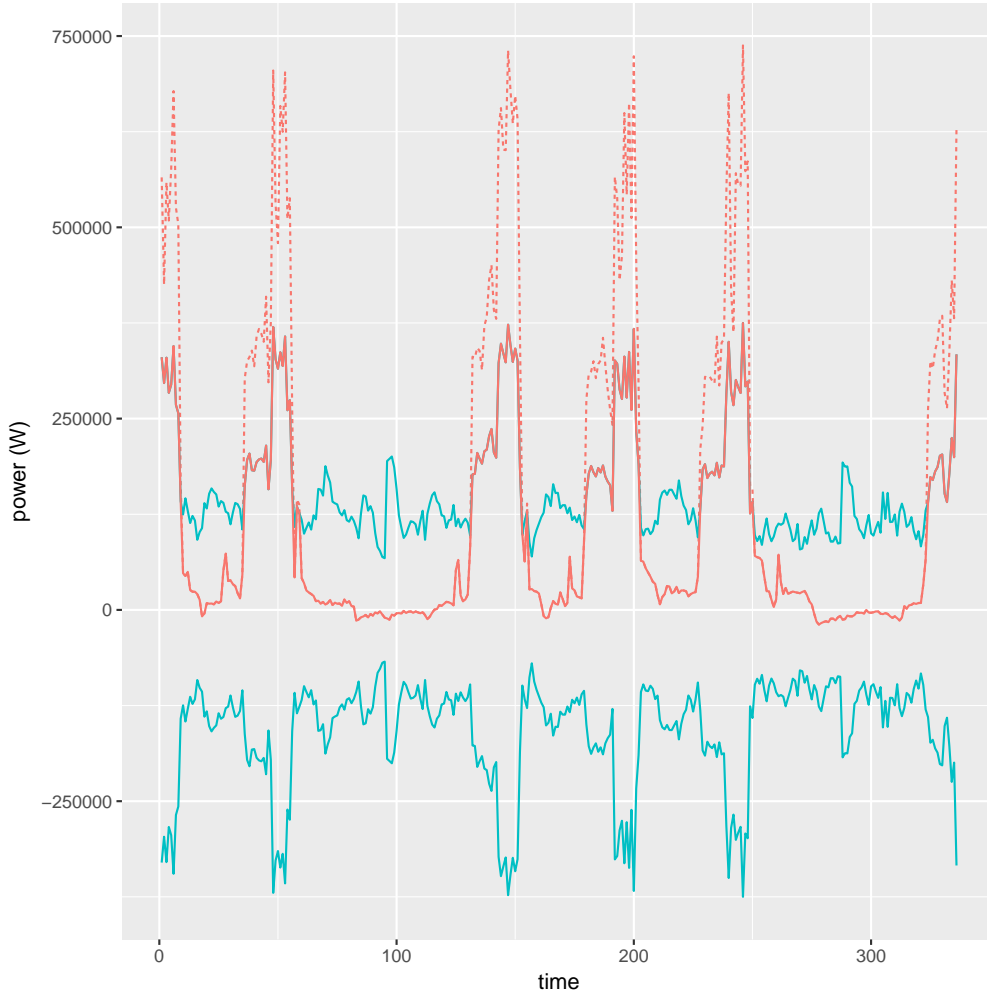


Figure 3: An estimated conditional bias (red solid line) and the same conditional bias truncated (red dotted line) for cut-off limits (blue lines) defined point-wisely.

3.2 A new criterion for choosing the optimal tuning constant

We suggest minimizing the sum over the sample of the q th power of the absolute value of $\widehat{B}_{1i}^{RHT}(c(t))$, the conditional bias of the robust estimator. This means that, for all t , we look for the optimal constant $c(t)$ satisfying the following criterion:

$$\begin{aligned} c_{opt}^{alt}(t) &= \arg \min_{c \geq 0} \sum_{i \in s} \left| \widehat{B}_{1i}^{RHT}(c(t)) \right|^q, \\ &= \arg \min_{c \geq 0} \sum_{i \in s} \left| \widehat{B}_{1i}^{HT}(c(t)) + \Delta(c(t)) \right|^q, \end{aligned} \quad (13)$$

where $\Delta(c(t))$ is given in (10) and q is a positive constant. The optimal solution may be found by numerical algorithms such as Newton-Raphson.

By using this criterion, we penalize the conditional bias computed for the whole sample of individuals not only the maximum and the minimum of $\widehat{B}_{1i}^{RHT}(c(t))$ as in Beaumont et al. (2013). In this way, each $\left| \widehat{B}_{1i}^{RHT}(c(t)) \right|$ for $i \in s$ will contribute to the optimisation

research and as a consequence, the presence in the sample at one instant t of a unit with very high influence will not cause a sudden change in the function $c(t)$ as it was the case with the minimax approach.

Large values of q will penalize large absolute values of $|\widehat{B}_{1i}^{RHT}(c(t))|$ while values of q between 0 and 1 will penalize small conditional bias $|\widehat{B}_{1i}^{RHT}(c(t))|$. So, for large q , this new criterion will be close to the minimax criterion but with better regularity properties allowing the use of bootstrap methods in order to estimate the variance. To obtain robust estimates we thus advise to consider large values of q ($q \geq 4$) that will ensure that high values of the conditional bias are sufficiently penalized. Note that for $q = 1$, we obtain the median conditional bias curve which is not of interest here because it is robust to outlying (extremely large or small) values. As a consequence, it will not be affected by these values and it can not be used to truncate outlying values.

4 Robust estimation based on dimension reduction

In a functional data setting, it is very common to use dimension reduction methods. In this paper, we use robust functional principal analysis and projection on basis functions such as wavelet function in order to transform the functional robust estimation issue into a series of univariate robust estimation issues.

4.1 Spherical principal components analysis

Principal components analysis is a popular tool to explore and to represent graphically the variations around their barycenter of multivariate and functional data (see Jolliffe (2002) and Ramsay and Silverman (2005) as well as Cardot et al. (2010) for a presentation in a finite population setting). The aim is to build new non-correlated variables, called principal components, that are linear combinations of the initial variables and of maximum variance. The principal components are obtained via the eigenfunctions of the covariance function γ of the data $Y_i, i = 1, \dots, N$,

$$\gamma(r, t) = \frac{1}{N} \sum_{i=1}^N (Y_i(r) - \bar{Y}_U(r))(Y_i(t) - \bar{Y}_U(t)), \quad r, t \in [0, T],$$

where $\bar{Y}_U = N^{-1} \sum_{i \in U} Y_i$ is the mean, or the center, of the data. However, it is well known that the mean and the covariance are highly sensitive to outlying units and consequently, principal components are also known to be highly non robust.

We consider now a robust version of PCA named spherical PCA (see Locantore et al. (1999)) that has nice properties (see Gervini (2008)) and is easy to compute. It consists in considering the eigenfunctions of the following sphericised ‘‘covariance’’ function

$$\mathbf{\Gamma}(r, t) = \frac{1}{N} \sum_{i=1}^N \frac{Y_i(r) - m_N(r)}{\|Y_i - m_N\|} \cdot \frac{Y_i(t) - m_N(t)}{\|Y_i - m_N\|}, \quad r, t \in [0, T] \quad (14)$$

where m_N is a robust indicator of location and $\|\cdot\|$ denotes the $L^2[0, T]$ -norm ($\|Y\|^2 = \int_0^T Y^2(t)dt$). Considering the unit norm functions $(Y_i(t) - m_N(t))/\|Y_i - m_N\|$ instead of $Y_i(t) - \bar{Y}_U(t)$, we perform a kind of winsorisation of the outlying curves Y_i . As in Locantore et al. (1999), we use the geometric median (see Kemperman (1987) or Small (1990)) as a robust location parameter of a set of points belonging to the space $L^2[0, T]$. With a finite population point of view, the median curve of the elements Y_1, \dots, Y_N , is defined by:

$$m_N = \arg \min_{y \in L^2[0, T]} \sum_{i=1}^N \|Y_i - y\|. \quad (15)$$

The relation (15) arises as a natural generalization of the well-known characterization of the univariate median. It is also called the *spatial median* (Brown (1983)) because, from a geometric point of view, the median is the point that minimizes the sum of distances to the points in the population. The names *L_1 -median* (Small (1990)) and *geometric median* (Chaudhuri (1996)) have been also employed for m_N .

If we assume that Y_i , for $i = 1, \dots, N$, are not concentrated on a line, the median exists and is unique (see Kemperman (1987)). If $m_N \neq Y_i$ for all $i = 1, \dots, N$, then it is the unique solution of the following estimating equation:

$$\sum_{i=1}^N \frac{Y_i(t) - m_N(t)}{\|Y_i - m_N\|} = 0, \quad t \in [0, T] \quad (16)$$

and it may be computed by using fast iterative algorithms such as Weiszfeld's algorithm (see Weiszfeld (1937) and Vardi and Zhang (2000)) for multivariate data or gradient algorithms (see Gervini (2008)) for sparse functional data.

Then, performing spherical PCA consists in computing the eigenvalues λ_j and the corresponding orthonormal eigenfunctions \mathbf{v}_j of the covariance $\mathbf{\Gamma}$ of these projected data instead of the initial data. As for the location estimate, the influence of the outlying observations can be greatly reduced. Furthermore Gervini (2008) shows that if the distribution of Y_i is symmetric, then the covariance γ and the spherical covariance $\mathbf{\Gamma}$ have the same orthonormal eigenfunctions \mathbf{v}_j , $j = 1, \dots, N$.

The curves Y_i in the population can also be approximated, in this new orthonormal basis, leading to a kind of robust Karhunen-Loeve expansion, that allows to get the best approximation of $Y_i(t) - m_N(t)$ in a finite K -dimensional space (see Ramsay and Silverman (2005)):

$$Y_i(t) = m_N(t) + \sum_{k=1}^K \langle Y_i - m_N, \mathbf{v}_k \rangle \mathbf{v}_k(t) + \epsilon_i(t), \quad \text{for } i \in U, \quad (17)$$

where $\epsilon_i(t) = Y_i(t) - m_N(t) - \sum_{k=1}^K \langle Y_i - m_N, \mathbf{v}_k \rangle \mathbf{v}_k(t)$ is a remainder term and $\langle \cdot, \cdot \rangle$ is the inner product in $L^2[0, T]$. Here, $\langle Y_i - m_N, \mathbf{v}_k \rangle \mathbf{v}_k(t)$ is the projection of the centered curve $Y_i - m_N$ onto the rank one space generated by function \mathbf{v}_k . For our purpose, we consider the same (large enough) value of K for all the curves Y_i to keep most of the variation in the data.

With these considerations, the approximation of the total curve in a finite K -dimensional space is given by

$$t_Y(t) \approx Nm_N(t) + \sum_{k=1}^K F_k \mathbf{v}_k(t), \quad (18)$$

where

$$F_k = \sum_{i \in U} \langle Y_i - m_N, \mathbf{v}_k \rangle, \quad \text{for } k = 1, \dots, K,$$

is the population total of the projections on \mathbf{v}_k of the "centered" data $Y_i - m_N$. So, we can write the finite population total t_Y as the sum of a robust location parameter, the median m_N , and the sum of K products between the robust eigenfunctions $\mathbf{v}_k(t)$ and the real coordinates F_k in this new basis. The interest of considering decomposition (18) is that the total of a function with a continuous time index is decomposed into a new multivariate problem in which robustification techniques can be applied to each real component.

4.1.1 Estimation of the robust principal components

In order to estimate t_Y , we need to estimate first the median and the eigenfunctions \mathbf{v}_k for all $k = 1, \dots, K$. A natural estimator of the geometric median m_N is given by the solution \hat{m} of the following non linear estimating equation (see Chaouch and Goga (2012)),

$$\sum_{i \in s} d_i \frac{Y_i(t) - \hat{m}(t)}{\|Y_i - \hat{m}\|} = 0, \quad t \in [0, T]. \quad (19)$$

Numerically, the solution is generally reached in a few iteration of a weighted version Weiszfeld's algorithm.

The spherical covariance function given in (14) is estimated as follows

$$\hat{\mathbf{\Gamma}}(r, t) = \frac{1}{N} \sum_{i \in s} d_i \frac{(Y_i(r) - \hat{m}(r)) \cdot (Y_i(t) - \hat{m}(t))}{\|Y_i - \hat{m}\|}, \quad \text{for all } r, t \in [0, T], \quad (20)$$

where \hat{m} is the estimator of the median m_N given in (19). Then, estimators of the eigenvalues λ_j of $\mathbf{\Gamma}$ with the associated eigenfunctions $\hat{\mathbf{v}}_j, j = 1 \dots, N$ are obtained by the spectral decomposition of the estimated covariance $\hat{\mathbf{\Gamma}}(r, t)$.

A natural estimator of the approximation of t_Y given in (18) is obtained by replacing the unknown quantities with their estimators:

$$\hat{t}_Y^{(2)}(t) = N\hat{m}(t) + \sum_{k=1}^K \hat{F}_k \hat{\mathbf{v}}_k(t), \quad (21)$$

where $\hat{F}_k = \sum_{i \in s} d_i \langle Y_i - \hat{m}, \hat{\mathbf{v}}_k \rangle$ is the substitution estimator for F_k . Note that even if $\hat{m}(t)$ and $\hat{\mathbf{v}}_k(t)$ are robust estimates, the estimator given in (21) is not robust because the coordinates $\hat{F}_k, k = 1, \dots, K$ are not robust.

4.1.2 Robustifying the coordinates in the spherical PCA basis

We suggest to build the following robust estimates of the coordinates

$$\hat{F}_k^R = \hat{F}_k - \sum_{i \in s} \hat{B}_{1i,k}^F + \sum_{i \in s} \psi_{c_k}(\hat{B}_{1i,k}^F), \quad k = 1, \dots, K, \quad (22)$$

where $\hat{B}_{1i,k}^F = \sum_{j \in s} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_j \pi_{ij}} \right) \langle Y_j - \hat{m}, \hat{\mathbf{v}}_k \rangle$ is the estimator of the conditional bias of \hat{F}_k , and ψ_{c_k} is the Huber function depending on the tuning constant c_k . An optimal value for c_k may be found by using the minimax criterion or the new criterion defined in (13). Finally, the second robust estimator of t_Y is defined as follows

$$\hat{t}_Y^{(R2)}(t) = N \hat{m}(t) + \sum_{k=1}^K \hat{F}_k^R \hat{\mathbf{v}}_k(t), \quad t \in [0, T]. \quad (23)$$

4.2 Projection on wavelet basis

Instead of using principal components, we may project data onto a basis of functions ϕ_1, \dots, ϕ_Q which do not depend on the data. Electricity load curves are known to be irregular, as seen in Figure 1, and natural candidates are wavelet basis (see Mallat (1998)).

The curves $Y_i, i \in U$ may be expanded as follows

$$Y_i(t) = \sum_{q=1}^Q a_{iq} \phi_q(t) + \epsilon_i(t), \quad t \in [0, T],$$

where ϵ is an approximation residual. Note that, unlike the principal component analysis, the functions ϕ_1, \dots, ϕ_Q are known and do not need to be estimated. The coefficients a_{iq} , for $q = 1, \dots, Q$, depend on Y_i and are unknown for the non-sampled individuals. As in robust principal component analysis, the total curve may be approximated by

$$t_Y(t) \simeq \sum_{q=1}^Q t_{a_q} \phi_q(t), \quad t \in [0, T], \quad (24)$$

where $t_{a_q} = \sum_{i \in U} a_{iq}$ is the unknown real population total of the coefficients a_{iq} , for $q = 1, \dots, Q$. The Horvitz-Thompson estimator of this new approximation of the total t_Y is given by

$$\hat{t}_Y^{(3)}(t) = \sum_{q=1}^Q \hat{t}_{a_q}^{HT} \phi_q(t), \quad t \in [0, T],$$

where $\hat{t}_{a_q}^{HT} = \sum_{i \in s} d_i a_{iq}$. Robust estimators of $\hat{t}_{a_q}^{HT}$ may be built as above. Our third robust estimator of t_Y is defined as follows:

$$\hat{t}_Y^{(R3)}(t) = \sum_{q=1}^Q \hat{t}_{a_q}^{RHT} \phi_q(t), \quad t \in [0, T], \quad (25)$$

where $\hat{t}_{a_q}^{RHT} = \hat{t}_{a_q}^{HT} + \sum_{i \in s} (\psi_{c_q}(\hat{B}_{1i,q}) - \hat{B}_{1i,q})$ is the robust estimator of $\hat{t}_{a_q}^{HT}$, $q = 1, \dots, Q$, with $\hat{B}_{1i,q}$ the conditional bias of $\hat{t}_{a_q}^{HT}$ and ψ_{c_q} the Huber function depending on the tuning constant c_q whose value may be determined for each $\hat{t}_{a_q}^{HT}$.

5 Global functional truncation methods based on statistical depth

The aim of this section is to introduce a global way of truncating the conditional-bias curve. In order to do that, we use the notion of statistical depth which allows to define an order relation in a set of curves, from the most central curve to the most outlying one. In the context of functional data, the depth may be defined in many different ways: see for example Cuesta-Albertos et al. (2006), Gervini (2012), Fraiman and Muniz (2001) or Hyndman and Shang (2010). Many of these notions of depth are rather difficult to put into practice and are not considered here. In the following, we consider the modified band depth as defined by López-Pintado and Romo (2009) as well as a depth notion based on the $L^2[0, T]$ distance from the center of the projected data onto the axis obtained by spherical PCA.

5.1 Definition of the modified band depth (MBD)

The Modified Band Depth (MBD), studied by López-Pintado and Romo (2009), of a discretized curve is the number of times (or the proportion of time for continuous time observations) the curve, within a set of curves, is “lying between a couple of other curves”:

$$\begin{aligned} MBD_i &= \frac{1}{\binom{2}{n}} \sum_{j,k \in s, j \neq k} \frac{1}{D} \sum_{d=1}^D \mathbb{1}_{[\min(\hat{B}_{1j}^{HT}(t_d), \hat{B}_{1k}^{HT}(t_d)) \leq \hat{B}_{1i}^{HT}(t_d) \leq \max(\hat{B}_{1j}^{HT}(t_d), \hat{B}_{1k}^{HT}(t_d))]} \\ &\approx \frac{1}{\binom{2}{n}} \sum_{j,k \in s, j \neq k} \frac{1}{T} \int_0^T \mathbb{1}_{[\min(\hat{B}_{1j}^{HT}(t), \hat{B}_{1k}^{HT}(t)) \leq \hat{B}_{1i}^{HT}(t) \leq \max(\hat{B}_{1j}^{HT}(t), \hat{B}_{1k}^{HT}(t))]} dt. \end{aligned}$$

This indicator takes into account the length of the time interval during which the curve \hat{B}_{1i}^{HT} is not lying between each couple of other curves : a curve which is not included between others during a small time interval will be considered as “less outlying” than another one which is out during a longer period. So, the more often a curve is included entirely between others the more it is considered as central and by consequence, a curve with a high MBD will be considered as central.

5.2 Central area based on MBD and robust estimator

We compute the depth value MBD_i of the conditional bias curve $\hat{B}_{1i}^{HT}(t)$ for all units i belonging to the sample and let I be the central region containing the 50% of the deepest curves $\hat{B}_{1i}^{HT}(t), i \in s$. Let L be the lower functional bound and U the upper functional bound computed over I , for $t \in [0, T]$:

$$L(t) = \min_{i \in I} \hat{B}_{1i}^{HT}(t) \quad \text{and} \quad U(t) = \max_{i \in I} \hat{B}_{1i}^{HT}(t).$$

The idea of using a 50% central region has been suggested first in the functional bagplot introduced by Hyndman and Shang (2010) and in the functional boxplot by Sun and Genton (2011).

The conditional-bias curves entirely located inside these boundaries will not be modified whereas the curves taking values outside the central region, for some period of time, will be truncated by using a truncation function ψ as in the non-functional case. An obvious candidate is the Huber function $\psi_c(y) = \max(\min(y, c), -c)$ depending on a tuning constant $c > 0$ which can be easily generalized to take into account a region delimited by a lower and an upper delimiting curves:

$$\psi(\hat{B}_{1i}^{HT}(t)) = \max\left(\min(\hat{B}_{1i}^{HT}(t), U(t)), L(t)\right) \quad \text{for all } t \in [0, T].$$

Remark that L needs not to be $-U$. We propose to use the following truncation function,

$$\psi_\alpha(\hat{B}_{1i}^{HT}(t)) = \max\left(\min\left(\hat{B}_{1i}^{HT}(t), \mu_{\hat{B}}(t) + \alpha(U(t) - \mu_{\hat{B}}(t))\right), \mu_{\hat{B}}(t) + \alpha(L(t) - \mu_{\hat{B}}(t))\right),$$

where α is an unknown positive dilatation parameter that controls the size of the central region (Sun and Genton (2011)) and $\mu_{\hat{B}}$ is the mean of the estimated conditional bias over the sample. In practice, the delimiting curves $\mu_{\hat{B}}(t) + \alpha(U(t) - \mu_{\hat{B}}(t))$ and $\mu_{\hat{B}}(t) + \alpha(L(t) - \mu_{\hat{B}}(t))$ are smoothed, using a mobile averaging technique, in order to avoid a too irregular truncation.

Figure 4 displays the mechanism of global truncation based on modified band depth. The upper (U) and the lower (L) curves delimiting the α central area are plotted in blue. A conditional bias curve is plotted in red. Parts of this curve lying outside of the central area, plotted in red dotted line, will be truncated and replaced by the corresponding parts of the bound curves. We can remark that the central zone constructed in this way reflects the daily seasonality of the data. In Figure 5, we plot central areas constructed according to the suggested methods: pointwise, spherical PCA and based on modified band depth. We can remark on this plot that the central area based on modified band depth is not symmetric and is much larger than the other two areas.

Using the global truncation function ψ_α and relation (9), we can construct a new robust estimator as follows

$$\hat{t}_Y^{(R4)} = \hat{t}_Y(t) + \sum_{i \in s} \psi_\alpha(B_{1i}^{HT}(t)) - \sum_{i \in s} B_{1i}^{HT}(t), \quad t \in [0, T]. \quad (26)$$

The dilatation factor α allows to control the trade-off between bias and variance: for α small, the curves are strongly truncated meaning large bias and small variance whereas for α large, the curves are less truncated meaning less bias and larger variance. To determine the value of the truncation parameter α , we can use the functional minimax approach:

$$\alpha_{opt} = \arg \min_{\alpha} \max_{i \in s} \frac{1}{D} \sum_{d=1}^D \left| \hat{B}_{1i}^{HT}(t_d) + \Delta_\alpha(t_d) \right|$$

or the q th power criterion introduced in Section 3.2:

$$\alpha_{opt}^{alt} = \arg \min_{\alpha \geq 0} \frac{1}{D} \sum_{d=1}^D \sum_{i \in s} \left| \hat{B}_{1i}^{HT}(t_d) + \Delta_\alpha(t_d) \right|^q, \quad (27)$$

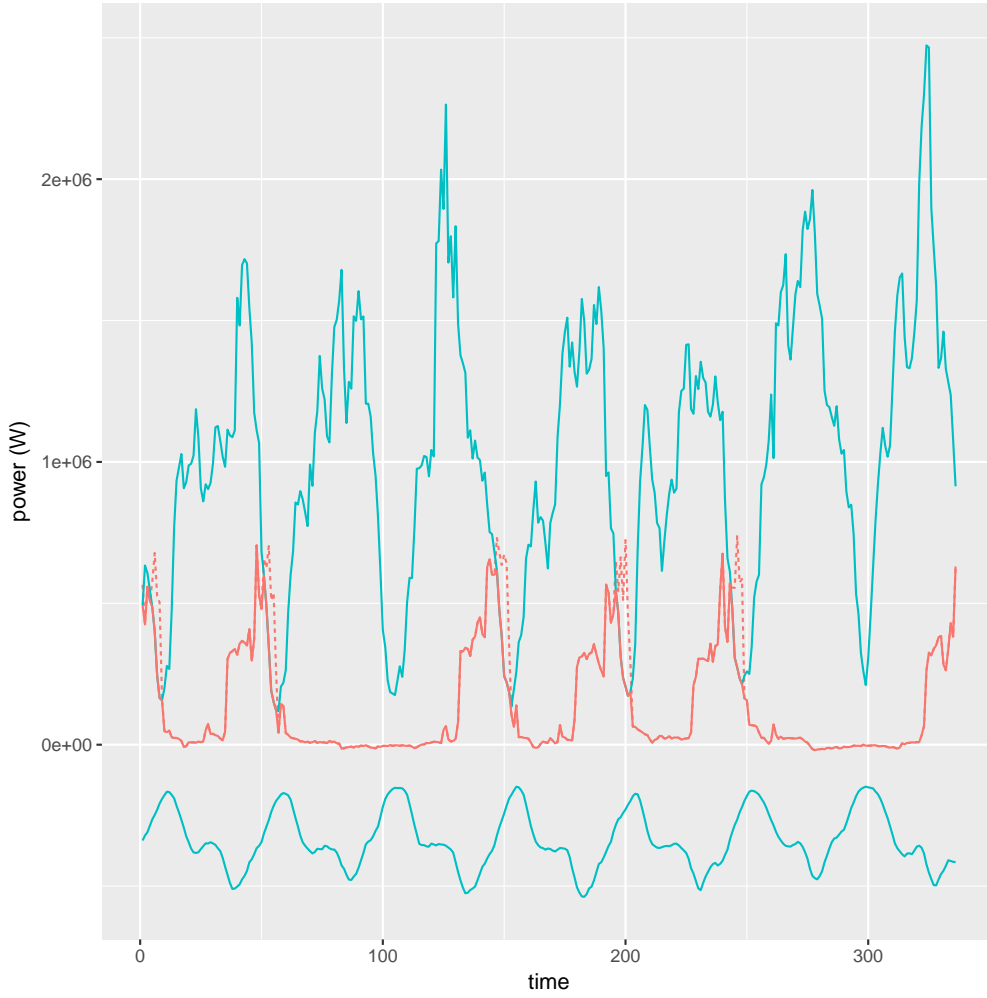


Figure 4: The upper (U) and the lower (L) curves delimiting the α central area are plotted in blue. A conditional bias curve is plotted in red and the truncated part of this curve is plotted in red dotted line.

where

$$\Delta_{\alpha}(t) = \sum_{i \in s} \left(\psi_{\alpha}(\hat{B}_{1i}^{HT}(t)) - \hat{B}_{1i}^{HT}(t) \right).$$

The optimum values α_{opt} and α_{opt}^{alt} are obtained numerically by a Newton-Raphson algorithm.

6 Mean square error estimation

In this section, we derive approximate point-wise estimators of the mean square errors of the robust estimators. For a given time instant t , the mean square error (MSE) can be

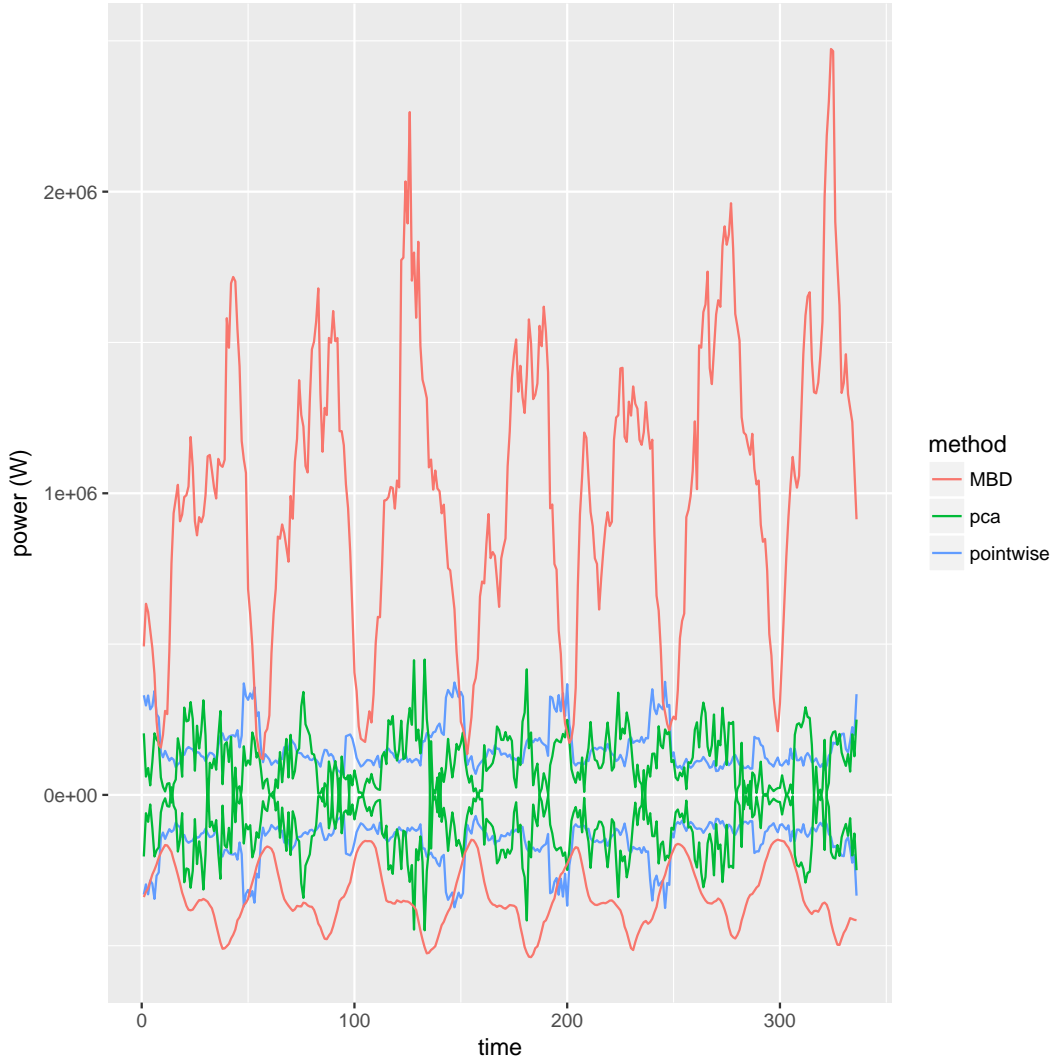


Figure 5: Central areas constructed according to three methods: point-wisely (blue line), robust PCA (red line) and functional modified band depth (green line).

expressed as

$$MSE_p(\hat{t}_Y^R(t)) = V_p(\hat{t}_Y^R(t)) + E_p(\hat{t}_Y^R(t) - \hat{t}_Y(t))^2 - V_p(\hat{t}_Y^R(t) - \hat{t}_Y(t)).$$

Similarly to Gwet and Rivest (1992) and Beaumont et al. (2013), we suggest the following point-wise mean square error estimator:

$$\widehat{MSE}_p(t) = v_p(\hat{t}_Y^R(t)) + \max\left[0, (\hat{t}_Y^R(t) - \hat{t}_Y(t))^2 - v_p(\hat{t}_Y^R(t) - \hat{t}_Y(t))\right]. \quad (28)$$

where $v_p(\hat{t}_Y^R(t))$ and $v_p(\hat{t}_Y^R(t) - \hat{t}_Y(t))$ are design-consistent estimators of $V_p(\hat{t}_Y^R(t))$ and $V_p(\hat{t}_Y^R(t) - \hat{t}_Y(t))$.

Using relation (9), we can write the robust estimator $\hat{t}_Y^R(t)$ and $\hat{t}_Y^R(t) - \hat{t}_Y(t)$ as follows

$$\begin{aligned}\hat{t}_Y^R(t) &= \sum_{i \in s} d_i (Y_i(t) + Z_{ic(t)}(t)) \quad \text{and} \\ \hat{t}_Y^R(t) - \hat{t}_Y(t) &= \sum_{i \in s} d_i Z_{ic(t)}(t),\end{aligned}$$

where $Z_{ic(t)}(t) = \pi_i(\psi_{c(t)}(B_{1i}^{HT}(t)) - B_{1i}^{HT}(t))$. For simple sampling designs for which the first and second order inclusion probabilities are known, we can use the Horvitz-Thompson variance estimator,

$$v_p(\hat{t}_Y^R(t)) = \sum_{i \in s} \sum_{j \in s} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \frac{\hat{Z}_{i,c(t)}(t) Y_j(t) + \hat{Z}_{j,c(t)}(t) Y_i(t)}{\pi_i \pi_j}, \quad (29)$$

where $\hat{Z}_{i,c(t)}(t) = \pi_i(\psi_{c(t)}(\hat{B}_{1i}^{HT}(t)) - \hat{B}_{1i}^{HT}(t))$. A variance estimator is obtained for $\hat{t}_Y^R(t) - \hat{t}_Y(t)$ by a similar procedure. For the robust estimator $\hat{t}_Y^{(R4)}$ given in (26) (section 5.1) computed by using functional truncation methods based on depth, a variance estimator may be computed by using (29) with $\hat{Z}_{i\alpha}(t) = \pi_i[\psi_\alpha(\hat{B}_{1i}^{HT}(t)) - \hat{B}_{1i}^{HT}(t)]$.

Using linearization techniques, we can write for the robust estimator $\hat{t}_Y^{(R2)}$ given in (23):

$$N^{-1}(\hat{t}_Y^{(R2)}(t) - t_Y) \simeq N^{-1} \left(\sum_{i \in s} d_i u_i - \sum_{i \in U} u_i \right) + N^{-1} \sum_{k=1}^K \left[(\hat{F}_k^R - F_k) \mathbf{v}_k + (\hat{\mathbf{v}}_k - \mathbf{v}_k) F_k \right]$$

where $u_i = N\mathbf{\Gamma}^{-1} \left(\frac{Y_i - m}{\|Y_i - m\|} \right)$, $i \in U$ is the linearized variable of m (see Chaouch and Goga (2012)) with $\mathbf{\Gamma}$ given in (14). We also have

$$N^{-1} \sum_{k=1}^K (\hat{\mathbf{v}}_k - \mathbf{v}_k) F_k \simeq N^{-1} \left(\sum_{i \in s} d_i \sum_{k=1}^K F_k \tilde{\mathbf{v}}_{i,k} - \sum_{i \in U} \sum_{k=1}^K F_k \tilde{\mathbf{v}}_{i,k} \right),$$

where $\tilde{\mathbf{v}}_{i,k} = \sum_{\ell \neq i} \langle Y_k - m, \mathbf{v}_i \rangle \langle Y_k - m, \mathbf{v}_\ell \rangle \mathbf{v}_\ell / (\lambda_i - \lambda_\ell) \|Y_k - m\|^2$ is the linearized variable of \mathbf{v}_k obtained with similar arguments as in Cardot et al. (2010). We also have

$$\begin{aligned} & N^{-1} \sum_{k=1}^K (\hat{F}_k^R - F_k) \mathbf{v}_k \\ & \simeq N^{-1} \sum_{i \in s} d_i \sum_{k=1}^K (\langle Y_i - m, \mathbf{v}_k \rangle + \pi_i(\psi_{c_k}(B_{1i,k}^F) - B_{1i,k}^F)) \mathbf{v}_k - N^{-1} \sum_{i \in U} \sum_{k=1}^K \langle Y_i - m, \mathbf{v}_k \rangle \mathbf{v}_k. \end{aligned}$$

The variance estimator can then be computed for

$$\hat{Z}_{ic}(t) = \hat{u}_i(t) + \sum_{k=1}^K \left(\hat{F}_k \hat{\mathbf{v}}_{i,k} + \langle Y_i - \hat{m}, \hat{\mathbf{v}}_k \rangle \hat{\mathbf{v}}_k + \pi_i(\psi_{c_k}(\hat{B}_{1i,k}^F) - \hat{B}_{1i,k}^F) \hat{\mathbf{v}}_k \right).$$

For the third robust estimator $\hat{t}_Y^{(R3)}$ given in (25) based on projection on known basis function ϕ_1, \dots, ϕ_Q , the variance estimator is obtained for $Z_{ic}(t) = \sum_{q=1}^Q (Y_i + \pi_i(\psi_{c_q}(\hat{B}_{1i,q}) - \hat{B}_{1i,q})) \phi_q(t)$.

Bootstrap

Approximation by bootstrap of the variance estimator v_p used in (28) is possible. We consider the without replacement bootstrap introduced by Gross (1980) for simple random sampling without replacement and that can be extended easily to stratified simple random sampling. The method consists in creating a pseudo-population U^* by duplicating each unit $i \in s$, $d_i = 1/\pi_i$ times. Several methods have been proposed to deal with the situation when d_i is not integer. We consider here the population bootstrap as suggested by Booth et al. (1994) which consists in completing U^* by a simple random sampling of size $N - [N/n]$. From this pseudo-population, we select B replication samples s^* of size n according to the initial sampling design. The bootstrap variance estimator of the robust estimator $\hat{t}_Y^R(t)$ is the empirical variance of $\hat{t}_Y^R(t)$ computed over the replication samples:

$$v_p^{boot}(\hat{t}_Y^R(t)) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{t}_Y^{R,b}(t) - \frac{1}{B} \sum_{b=1}^B \hat{t}_Y^{R,b}(t) \right)^2. \quad (30)$$

The value of the cut-off tuning parameter c is computed in each replication using the minimax approach. However, as the robust estimator based on the minimax approach is built using minima and maxima (which are "non linearizable" functions), we may have poor results for estimates based on population bootstrap.

We also consider the generalized bootstrap studied by Bertail and Combris (1997). For this bootstrap method, the sample of individuals is kept unchanged but the sampling weights are replicated. More precisely, we generate random weights w_i^{*b} with $b = 1, \dots, B$ and B large, such as $E(w_i^{*b}) = N^{-1}$, $\text{Var}(w_i^{*b}) = (1 - \pi_i)N^{-2}$ and $\text{Cov}(w_i^{*b}, w_j^{*b}) = (1 - \pi_i\pi_j/\pi_{ij})N^{-2}$, $i \neq j$. In practice, w_i^{*b} may be simulated from a multivariate normal law with moments given above. The parameters of interest are written as functions of means and means of type $\mu_Y = \sum_{i \in U} Y_i/N$, that estimated at each replication b by $\hat{\mu}_Y^b = \sum_{i \in s} w_i^{*b} d_i Y_i$. Formula (30) is next used to obtain a variance estimator of the robust estimators suggested in this paper.

7 An illustration with real dataset

The methods and estimators studied in this paper are illustrated on data from the Irish Commission for Energy Regulation (CER) Smart Metering Project that was conducted in 2009-2010 (CER, 2011)². This dataset contains thousands of electricity load curves of residential clients observed every half-hour during one year. We have selected from this dataset $N = 3994$ load curves without missing data and the electricity consumption recorded over one week, from the 18th to the 24th of January 2010. So, we have $D = 336$ points in time. The interest parameter is the total consumption electricity during this week.

²The data are available on request at the address:
<http://www.ucd.ie/issda/data/commissionforenergyregulation/>

We consider two sampling designs: simple random sampling (SRS) without replacement and stratified random sampling with SRS within strata (STR). For the stratified sampling, strata are built by considering the total electricity consumption over the second semester of 2009. We have built 5 strata, containing respectively 1270, 898, 770, 659 and 397 statistical units. The first strata corresponds to meters with small levels of consumption whereas the last one is associated to the meters with the largest levels of consumption. In this scenario, there are no "strata jumpers". We consider two other STR samplings with 10% strata jumpers (STR-SJ10) and respectively, with 20% "strata jumpers" (STR-SJ20). These "strata jumpers" are simulated by selecting randomly, with equal probabilities, some units in the population and then affecting them to a wrong stratum, which is also chosen randomly with equal probabilities. For each scenario, we have considered three sample sizes: $n = 40, 100$ and respectively, $n = 400$ and the sample sizes within strata are computed according to the optimal allocation taking the consumption of the previous week as auxiliary information.

7.1 Performance of the suggested robust estimators of the total consumption curve

We evaluate the performances and compare the different estimators presented in previous sections for various situations: different sampling designs, "strata jumpers" rates, sample sizes. The estimators considered here are:

- the Horvitz-Thompson (HT) estimator;
- the point-wise robust estimator $\hat{t}_Y^{(R1)}$ given by (9) with the tuning constant c chosen by the minimax pointwise criterion (minimax pointwise) and the q th ($q = 4, 10$) power criteria (qth pointwise);
- the robust estimator $\hat{t}_Y^{(R2)}$ given by (23) and based on spherical PCA with the minimax criteria (robust PCA) and $K = 5$ principal components;
- the robust estimator $\hat{t}_Y^{(R3)}$ given by (25) and based on wavelet expansions with the minimax criteria (robust wave)³;
- the robust estimator $\hat{t}_Y^{(R4)}$ given by (26) with the global truncation function based on the modified band depth, minimax criteria (MBD).

We draw $I = 5000$ samples according to each sampling strategy and for each estimator \hat{t}_Y of t_Y , we compute the relative bias (RB) and the relative mean square error (RMSE):

$$RB(\hat{t}_Y(t_d)) = 100 \frac{EMC[\hat{t}_Y(t_d)] - t_Y(t_d)}{t_Y(t_d)}, \quad d = 1, \dots, D$$

$$RMSE(\hat{t}_Y(t_d)) = 100 \frac{MSE_{MC}[\hat{t}_Y(t_d)]}{MSE_{MC}[\hat{t}_Y^{HT}(t_d)]}, \quad d = 1, \dots, D$$

³wavelets Daubechies Least Asymmetric, 10

where $E_{MC}[\hat{t}_Y(t_d)] = \sum_{i=1}^I \hat{t}_Y^{(i)}(t_d)/I$ and $MSE_{MC}(\hat{t}_Y(t_d)) = \sum_{i=1}^I (\hat{t}_Y^{(i)}(t_d) - t_Y(t_d))^2/I$ are the Monte-Carlo expectation and mean square error of $\hat{t}_Y(t_d)$ computed over the $I = 5000$ samples and $t_Y(t_d)$ is the real value of the total curve at instant t_d . In order to assess the global performance, we consider the mean value, over time, of these indicators

$$RB = \frac{1}{D} \sum_{d=1}^D RB(\hat{t}_Y(t_d)) \quad \text{and} \quad RMSE = \frac{1}{D} \sum_{d=1}^D RMSE(\hat{t}_Y(t_d)).$$

Estimator RB (%)	SRS (size=)			STR J0 (size=)			STR J10 (size=)			STR J20 (size=)		
	40	100	400	40	100	400	40	100	400	40	100	400
minimax pointwise	-9	-6	-3	-2	-2	-1	-4	-3	-1	-5	-4	-2
4th pointwise	-4	-2	-1	-1	-1	0	-2	-1	0	-2	-1	0
10th pointwise	-7	-4	-2	-2	-2	-1	-3	-2	-1	-4	-3	-1
Robust PCA	-7	-5	-2	-1	-1	0	-3	-2	-1	-3	-3	-1
Robust wave	-7	-5	-2	-2	-1	0	-3	-2	-1	-3	-3	-1
MBD	-8	-5	-2	0	0	0	-2	-1	-1	-3	-2	-1

Table 1: Relative bias (RB in %).

Estimator RMSE (%)	SRS (size=)			STR J0 (size=)			STR J10 (size=)			STR J20 (size=)		
	40	100	400	40	100	400	40	100	400	40	100	400
minimax pointwise	74	85	96	97	96	98	89	88	91	84	86	91
4th pointwise	85	92	98	98	97	99	93	94	97	91	93	97
10th pointwise	77	86	96	97	96	98	90	89	92	85	88	93
Robust PCA	73	83	95	97	97	98	87	86	89	82	84	90
Robust wave	72	83	94	94	93	95	85	83	88	81	82	89
MBD	75	86	97	100	99	100	89	87	89	85	85	90

Table 2: Relative MSE (RMSE in %).

The results are reported in Tables 1 and 2. We can note that the use of robust methods lead to important precision gains particularly when the sample size is small. For the SRS design and for the best robust method, the global error is reduced by 28% when the sample size is 40, by 17% when the sample size is 100. Moreover the robust methods never deteriorate significantly the global precision. The performances of these robust estimators are quite similar. Nevertheless, the functional methods based on wavelets or spherical PCA are slightly better, followed by the robust estimator built with the global truncation function based on the modified band depth.

However, robust methods tend to underestimate the population total curve because the outliers, whose influence is reduced, are often units with large values. So the robust methods lead to a negative bias of a few percents. This bias is larger for more imprecise sampling designs.

For stratified samplings without strata jumpers, the use of robust approaches do not lead to much improvement. This result is not surprising since a good stratification permits

to reduce the influence of large units during the sampling phase. We also remark that, in this situation, the Horvitz-Thompson estimator is nearly as effective as the less accurate robust approach (MBD). We also note that the relative bias in that case is very small (less than 2%) which could mean that the conditional biases are almost not truncated.

On the contrary, in presence of strata jumpers, the use of robust methods permits to improve significantly the precision, especially when the strata jumpers rate is high. The observed gains are approximately 15% in presence of 10% of strata jumpers.

On this simulation study, the minimax criterion for the choice of the tuning constant gives better result than the q th power criterion. As expected, the performances of the robust estimators built on minimax and the q th power criteria are very similar for q large. We plot in Figure 6, the relative mean square error along time for the suggested estimators and SRS sample of size $n = 100$. We can remark that RMSE varies much over time.

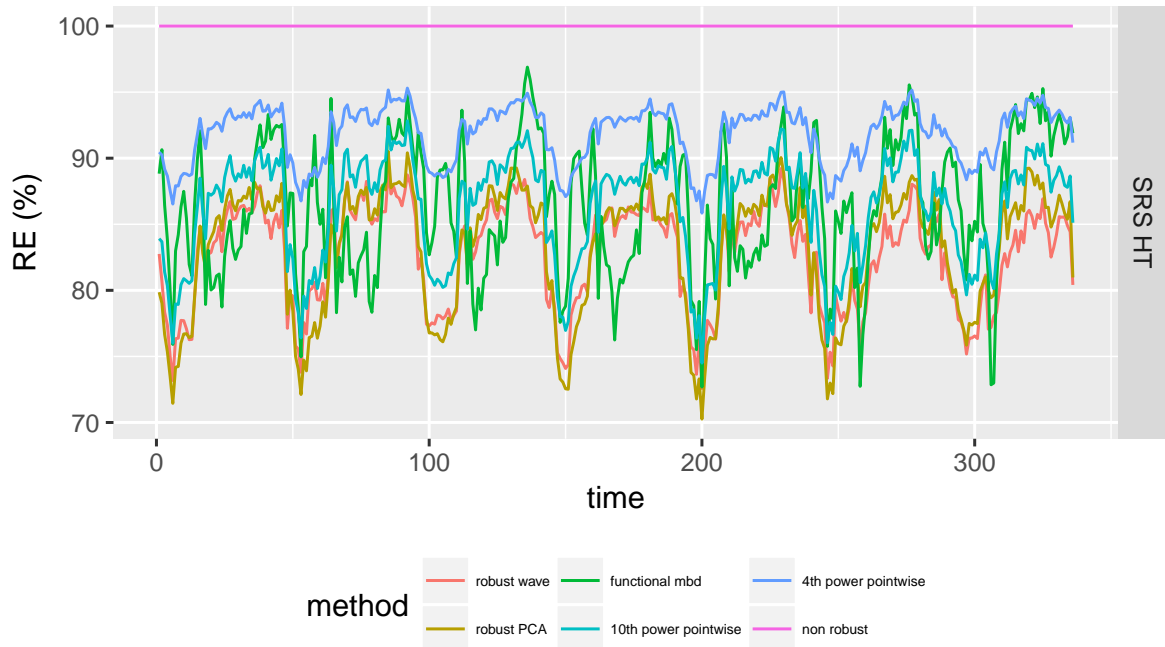


Figure 6: Evolution of the RMSE over time for different methods and SRS of size $n = 100$

We have also computed the execution time for the suggested methods. The point-wise robust method is the fastest robust method: for example, for $n = 100$, its mean execution time for one simulation is around 6×10^{-2} seconds. The projection methods are slightly slower but they never take more than 2×10^{-1} seconds and finally, the functional MBD estimator is around 3×10^{-1} second always for a sample of 100 individuals. Moreover, this computation time only moderately increases when the sample size increases.

7.2 A comparison of the mean square error estimators

We compare the linearization approaches with the population bootstrap and the generalized bootstrap. For the bootstrap methods, we consider $B = 1000$ replications. We compute, by means of $I = 5000$ simulations, the estimators of MSE for the following total estimators: \hat{t}_{HT} the usual Horvitz-Thompson estimator, $\hat{t}_{HT}^{(R1)}$ the point wise estimator robustified via the minimax criterion, $\hat{t}_{HT}^{(R1)}$ robustified via the 10th power criterion; $\hat{t}_Y^{(R2)}$ based on robust PCA and robustified by the minimax criterion, and the total estimators based on a wavelet expansion, $\hat{t}_Y^{(R3)}$, robustified via the minimax criterion. The relative bias of the estimators of MSE are given in Table 3.

	SRS			STR J10		
	Gen. Boot.	Gross' Boot.	Lin.	Gen. Boot.	Gross' Boot.	Lin.
HT	-1	4	-1	-1	0	-1
Minimax	23	4	-27	20	0	-17
10th power	25	9	-21	23	2	-15
RPCA	24	5	-32	22	2	-26
Wavelet	25	6	-23	25	3	-21

Table 3: Relative Bias of \widehat{MSE} for samples of size $n = 100$, with different MSE estimation procedures.

We can note that, as expected, all the estimators of the mean squared error provide reasonable results for the non robust Horvitz-Thompson estimator. We also note that, in our particular context, the MSE estimators based on linearization lead to a significant underestimation when the estimator is robust: this underestimation is about 20% when the tuning constant is determined by the new criterion and about 30% for the minimax criterion. This was expected because we do not take into account the variability due to the data driven selection of the value of the tuning constant. On the contrary, we observe a strong overestimation of the variance of robust estimators for generalized bootstrap whereas Gross' bootstrap seems to gives satisfactory results for all the scenarios.

As far as computation time is concerned, the MSE estimation based on linearization is quite fast, around a few tenth of second, whereas the bootstraps are significantly slower, around 20 seconds for the generalized bootstrap and 80 seconds for Gross' bootstrap.

8 Concluding remarks

Three types of robust estimators have been proposed in this work in order to adapt, from the univariate to the functional case, robust estimation techniques in finite populations:

- Point-wise robust estimators built by truncation of the conditional bias at each instant.
- Robust estimation based on dimension reduction methods.

- Global functional truncation methods based on statistical depth.

These approaches have been compared on the estimation of totals of load electricity curves. The comparisons have shown that robust methods lead to a noticeable improvement of the precision, especially when the estimation is the most imprecise (small sample sizes, sampling designs which do not include any auxiliary information or presence of very heterogeneous units in a same stratum). When the precision of the non robust estimators is already satisfying (larger sample sizes or relevant stratification), the precision gains are smaller. However, a very important fact is that robust methods never deteriorate the quality of the estimation.

We can also rank, in our simulation study, the different approaches according to their performances. The robust estimators based on wavelets expansion or on robust PCA are the most effective, followed by pointwise robust estimators and then global functional truncation based on the notion of depth.

The corresponding mean squared errors can be estimated using linearization or bootstrap. Gross' bootstrap seems to give satisfactory results but is computationally intensive whereas linearization-based techniques are much faster but may lead to noticeable underestimations.

We have also proposed a new criterion for choosing the tuning constant based on the q th power of the conditional bias. Its application on a real dataset showed that the minimax criterion is more effective than this new criterion.

Since our simulation studies have shown that the use of robust methods seems to be particularly relevant for small sample sizes, a natural extension of the work presented here is robust estimation of curves for small areas as considered in the PhD dissertation of De Moliner (2017). However, robust estimation for small areas is a challenging issue. Indeed, aggregating robust small domain estimates lead to overall estimators that may have a large bias, as noted in Rivest and Hidiroglou (2004), Favre-Martinoz et al. (2015) and Clark et al. (2017). Another difficulty is the fact that aggregated domain estimates may not be consistent with the population total estimate. To overcome this difficulty, one can use the approach suggested in Favre-Martinoz et al. (2015) based on a calibration technique.

Appendix

We suppose that the sample size n and the population size N become large. We consider a sequence of growing and nested populations U_N with size N tending to infinity and a sequence of samples s_N of size n_N drawn from U_N according to the sampling design $p_N(s_N)$. The first and second order inclusion probabilities are respectively denoted by π_{kN} and π_{klN} . For simplicity of notations and when there is no ambiguity, we drop the subscript N . To prove our asymptotic results we need to introduce the following assumptions.

A1. We assume that $\lim_{N \rightarrow \infty} \frac{n}{N} = \pi \in (0, 1)$.

A2. We assume that $\min_{k \in U} \pi_k \geq \lambda > 0$, $\min_{k \neq l \in U} \pi_{kl} \geq \lambda^* > 0$ and

$$\pi_{kl} = \pi_k \pi_l \left\{ 1 - \frac{(1 - \pi_k)(1 - \pi_l)}{D(\pi)} [1 + o(1)] \right\}$$

uniformly in k and l , where $D(\pi) = \sum_U \pi_i(1 - \pi_i)$.

A3. There are two positive constants C_2 and C_3 and $\beta > 1/2$ such that, for all N and for all $(r, t) \in [0, T] \times [0, T]$,

$$\frac{1}{N} \sum_{k \in U} (Y_k(0))^2 < C_2 \quad \text{and} \quad \frac{1}{N} \sum_{k \in U} (Y_k(t) - Y_k(r))^2 < C_3 |t - r|^{2\beta}.$$

Assumptions **A1** and **A2** are classical hypotheses in survey sampling and deal with the first and second order inclusion probabilities. They are satisfied for high entropy sampling designs with fixed size (see for example Hájek (1964)). They directly imply that $cn \leq D(\pi) \leq n$, for some strictly positive constant c . Assumption **A3** is a regularity condition on the individual trajectories. Even if point-wise consistency, for each fixed value of t , can be proven without any condition on β , this regularity condition is required to get the uniform convergence of the mean estimator (see Cardot and Josserand (2011)).

Proposition 8.1. *Suppose that **A1** and **A3** are fulfilled and the sampling design is simple random sampling without replacement or suppose that hypotheses **A1-A3** are fulfilled. Then*

$$\sup_{t \in [0, T]} \left| \sum_{i \in s} d_i A_i(t) - \sum_{i \in U} A_i(t) \right| = O_p(n^{-1/2}).$$

Proof. Recall that, for $t \in [0, T]$,

$$A_i(t) = \frac{-1}{1 - \pi_i} \sum_{j \in U, j \neq i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_j} Y_j(t).$$

For simple random sampling without replacement, $\pi_i = n/N$ and $\pi_{ij} = n(n-1)/(N(N-1))$ for $i \neq j$, and we have that (with $d_i = 1/\pi_i$),

$$\sum_{i \in s} d_i A_i(t) - \sum_{i \in U} A_i(t) = \frac{1}{N-1} (t_Y(t) - \hat{t}_Y(t)) \quad (31)$$

The result is then a direct consequence of Proposition 3.1 in Cardot and Josserand (2011).

Consider now the more general case of fixed-size high entropy sampling designs. Introducing the approximation to the second order inclusion probabilities in A_i we get after some algebra

$$\sum_{i \in s} d_i A_i(t) - \sum_{i \in U} A_i(t) \approx \frac{1}{D(\pi)} \left(\sum_U \pi_i(1 - \pi_i) Y_i(t) - \sum_s d_i \pi_i(1 - \pi_i) Y_i(t) \right), \quad t \in [0, T]. \quad (32)$$

The weighted trajectories $\pi_i(1 - \pi_i) Y_i(t)$, $t \in [0, T]$ also satisfy assumption **A3** and the result is a consequence of Proposition 3.1 in Cardot and Josserand (2011) (see also Cardot et al. (2014)). \square

References

Beaumont, J.-F., Haziza, D., and Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, 100(3):555–569.

- Bertail, P. and Combris, P. (1997). Bootstrap généralisé d'un sondage. *Annales d'Économie et de Statistique*, 46:49–83.
- Booth, J. G., Butler, R. W., and Hall, P. (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association*, 89(428):1282–1289.
- Brown, B. (1983). Statistical use of the spatial median. *Journal of the Royal Statistical Society, B*, 45:25–30.
- Cardot, H., Chaouch, M., Goga, C., and Labruère, C. (2010). Properties of design-based functional principal components analysis. *Journal of Statistical Planning and Inference*, 140(1):75–91.
- Cardot, H., Degras, D., and Josserand, E. (2013a). Confidence bands for Horvitz-Thompson estimators using sampled noisy functional data. *Bernoulli*, 19:2067–2097.
- Cardot, H., Dessertaine, A., Goga, C., Josserand, É., and Lardin, P. (2013b). Comparison of different sample designs and construction of confidence bands to estimate the mean of functional data: An illustration on electricity consumption. *Survey Methodology*, 39:283–301.
- Cardot, H., Goga, C., and Lardin, P. (2014). Variance estimation and asymptotic confidence bands for the mean estimator of sampled functional data with high entropy unequal probability sampling designs. *Scandinavian J. of Statistics*, 41:516–534.
- Cardot, H., Goga, C., Lardin, P., et al. (2013c). Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data. *Electronic Journal of Statistics*, 7:562–596.
- Cardot, H. and Josserand, E. (2011). Horvitz-Thompson estimators for functional data: asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika*, 98:107–118.
- Chambers, R. L. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*, 81(396):1063–1069.
- Chaouch, M. and Goga, C. (2012). Using complex surveys to estimate the L_1 -median of a functional variable: application to electricity load curves. *International Statistical Review*, 80(1):40–59.
- Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.*, 91:862–872.
- Clark, R., Kobic, P., and Smith, P. (2017). A comparison of two robust estimation methods for business surveys. *International Statistical Review*, 85(2):270–289.

- Cuesta-Albertos, J. A., Fraiman, R., and Ransford, T. (2006). Random projections and goodness-of-fit tests in infinite-dimensional spaces. *Bulletin of the Brazilian Mathematical Society*, 37(4):477–501.
- Deroyon, T. and Favre-Martinoz, C. (2018). Comparison of conditional-bias and Kokic-Bell methods for Poisson and stratified sampling. *Survey Methodology*, to appear.
- De Moliner, A. (2017). *Estimation robuste de courbes de consommation électrique moyennes par sondage pour de petits domaines en présence de valeurs manquantes (in French)*. PhD thesis, Université de Bourgogne Franche-Comté.
- Degras, D. (2014). Rotation sampling for functional data. *Statistica Sinica*, 24(3):1075–1095.
- Favre-Martinoz, C. (2015). *Estimation robuste en population finie et infinie*. PhD thesis, Université de Rennes 1, France.
- Favre-Martinoz, C., Haziza, D., and Beaumont, J. (2015). A method for determining the cut-off points for winsorized estimators with application to domain estimation. *Survey Methodology*, 41:51 – 77.
- Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test*, 10(2):419–440.
- Gervini, D. (2008). Robust functional estimation using the spatial median and spherical principal components. *Biometrika*, 95:587–600.
- Gervini, D. (2012). Outlier detection and trimmed estimation for general functional data. *Statistica Sinica*, 22:1639–1660.
- Gross, S. (1980). Median estimation in sample surveys. In *ASA Proceedings of Survey Research*, pages 181–184.
- Gwet, J.-P. and Rivest, L.-P. (1992). Outlier resistant alternatives to the ratio estimator. *Journal of the American Statistical Association*, 87(420):1174–1182.
- Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, 35:1491–1523.
- Hyndman, R. J. and Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *J. Comput. Graph. Statist.*, 19:29–45.
- Jolliffe, I. T. (2002). *Principal component analysis*. Springer Series in Statistics. Springer-Verlag, New York, second edition.
- Kemperman, J. (1987). The median of a finite measure on a Banach space. In: *Dodge, Y. (Ed.), Statistical Data Analysis Based on the L_1 Norm and Related Methods, North-Holland, Amsterdam*, pages 217–230.

- Kokic, P. and Bell, P. (1994). Optimal winsorizing cutoffs for a stratified finite population estimator. *Journal of Official Statistics*, 10:419–419.
- Lardin-Puech, P., Cardot, H., and Goga, C. (2014). Analysing large datasets of functional data: a survey sampling point of view. *Journal de la Société Française de Statistique*, 155(4):70–94.
- Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T., and Cohen, K. L. (1999). Robust principal component analysis for functional data. *Test*, 8:1–73.
- López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734.
- Mallat, S. (1998). *A wavelet tour of signal processing*. Academic Press, Inc., San Diego, CA.
- Muñoz-Pichardo, J., Muñoz-García, J., Moreno-Rebollo, J., and Pino-Mejías, R. (1995). A new approach to influence analysis in linear models. *Sankhyā: The Indian Journal of Statistics, Series A*, 57:393–409.
- Pfeffermann, D. and Rao, C. (2009). *Handbook of statistics - Sample surveys: Designs, Methods and Applications*, volume 29A. Elsevier.
- Ramsay, J.-O. and Silverman, B.-W. (2005). *Functional Data Analysis*. Springer Series in Statistics, New York, second edition.
- Rivest, L. and Hidiroglou, M. (2004). Outlier treatment for disaggregated estimates. In American Statistical Association, Alexandria, V., editor, *Proceedings of the Survey Research Methods Section*, pages 4248–4256.
- Rivest, L.-P. (1994). Statistical properties of winsorized means for skewed distributions. *Biometrika*, 81(2):373–383.
- Small, C. (1990). A survey of multidimensional medians. *International Statistical Review*, 58:263–277.
- Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20:316–334.
- Vardi, Y. and Zhang, C.-H. (2000). The multivariate L_1 -median and associated data depth. *Proc. Natl. Acad. Sci. USA*, 97(4):1423–1426.
- Weiszfeld, E. (1937). Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tôhoku Mathematical Journal*, 43:355–386.
- Welsh, A. and Ronchetti, E. (1998). Bias-calibrated estimation from sample surveys containing outliers. *Journal of the Royal Statistical Society, B*, 60:413–428.