On mean decomposition for summarizing conditional distributions

Celia García-Pareja[1†], Matteo Bottai[1]

[1]Unit of Biostatistics, IMM, Karolinska Institutet, Stockholm, Sweden.

[†]corresponding author: Celia García-Pareja, Unit of Biostatistics, IMM, Karolinska Institutet, Box 210, 171 77 Stockholm, Sweden. E-mail: celia.garcia.pareja@ki.se

Abstract

We propose a summary measure defined as the expected value of a random variable over disjoint subsets of its support that are specified by a given grid of proportions, and consider its use in a regression modeling framework. The obtained regression coefficients provide information about the effect of a set of given covariates on the variable's expectation in each specified subset. We derive asymptotic properties for a general estimation approach that are based on those of the chosen quantile function estimator for the underlying probability distribution. A bound on the variance of this general estimator is also provided, which relates its precision to the given grid of proportions and that of the quantile function estimator, as shown in a simulation example. We illustrate the use of our method and its advantages in two real data applications, where we show its potential for solving resource-allocation and intervention-evaluation problems.

keywords: asymptotics; compound expectation; conditional quantile function; regression model; summary measure

1 Introduction

Statistical summary measures aim to comprise relevant information from the distribution of a variable of interest. Despite their attractive simplicity, central tendency measures are not always appropriate when we intend to describe the entire distribution. The sample mean, for example, provides practical but often insufficient information. Conversely, a set of sample quantiles can provide a more detailed picture but lack information on how the distribution behaves between elements of the set.

Based on the same mathematical relation between these two extremes exploited in Wang and Zhou (2010), we propose to summarize the variable's distribution by specifying a grid of proportions that divide its mean into a sum of components. From these components, one can easily derive the variable's expected value on different segments of its support. Our approach results in a set–valued summary measure, which we refer to as compound expectation, that describes the variable's entire distribution in terms of expected values, and whose elements relate to specific fractions of the variable's support.

Following this same formulation, the compound expectation can be easily extended to a regression framework in which we summarize the variable's conditional distribution given covariates. Similarly to the univariate case, we obtain regression coefficients that measure average differences on the outcome variable over distinct fractions along its entire distribution. Other regression models that characterize entire conditional distributions have been previously proposed, e.g., by Koenker and Bassett Jr (1978), Newey and Powell (1987) or Breckling and Chambers (1988), and their advantages discussed at length (Koenker (2005), Chambers and Tzavidis (2006), Ehm et al. (2016)). However, conditional averages are still preferred in numerous applications, where measuring average behavior is essential for drawing meaningful conclusions. In this regard, our proposed regression model allows detecting different associations between outcome variable and covariates along the distribution, while preserving the appealing advantages of the mean.

All the components, which can take on positive or negative values, sum to the total expectation, and one can identify their contributions to it. Such contributions are measured as the proportion each component represents of the total mean. This feature is of convenience in numerous settings, where interests focus on how specific fractions of a variable influence its average. Examples include how premature adult-age deaths affect life expectancy (Seaman et al., 2016), how extreme warmth raises average temperature (Meehl et al., 2000), or how the most skilled increase average intelligence quotient (Wai and Putallaz, 2011).

The compound expectation can also be considered as a measure of concentration, where all components being equal indicates no concentration, and all but one components being zero indicates maximal concentration (Egghe and Rousseau, 1990). In such context, one can analyze how the variable accumulates around its different fractions, which is of use for solving resource-allocation problems. Approaches based on concentration curves, which originally appeared as a metric to study the distribution of wealth (Lorenz, 1905), have been recently used to address these problems, as exemplified by their application to transport logistics or epidemiology (Delbosc and Currie (2011), Mauguen and Begg (2016), Christopoulos et al. (2017)). Graphical tools, however, narrow modeling possibilities and require restrictions or ad hoc ordering criteria to be compared (see Shorrocks (1983), Davies and Hoy (1995) and Aaberge (2009), among others). Both these limitations are effectively overcome by the regression approach we propose in this paper.

The rest of this paper is structured as follows. In Section 2, we define the compound expectation and related quantities. In Section 3, we extend its use to a regression framework, which serves as a modeling tool in the presence of covariates. Our proposed estimation approach and its asymptotic properties are described in Section 4. In Section 5 we present a simulated data example that illustrates how different grids of proportions might reflect in final results, and in Section 6 we exemplify the application of our method to the analysis of time spent in intensive care units and weight gain. In Section 7 we conclude with some final remarks. Technical proofs and derivations can be found in the Appendix.

2 Compound expectation

Let Y be a random variable of interest with cumulative distribution function $F(y)$ and $\mathrm{E}(Y) < \infty$. The expectation of Y, μ, can be expressed in terms of its quantile function $Q(p)$ as

$$\mu = \mathrm{E}(Y) = \int_{-\infty}^{\infty} y \mathrm{d}F(y) = \int_0^1 Q(p)\mathrm{d}p.$$

Given a set of specified proportions $\{\lambda_0, \ldots, \lambda_K\}$, where $\lambda_0 = 0$, $\lambda_K = 1$ and $\lambda_{k-1} < \lambda_k$, for every $k = 1, \ldots, K$, we can think of μ as a sum of K components $\mu_1, \ldots, \mu_K$, that is,

$$\mu = \sum_{k=1}^{K} \mu_k, \quad \mu_k = \int_{\lambda_{k-1}}^{\lambda_k} Q(p)\mathrm{d}p. \tag{1}$$

Each component μ_k amounts to a specific proportion of the total expectation, which can be viewed as the contribution of the kth fraction of Y to μ. These contributions quantify the influence each fraction has on the overall mean, and allow to identify where efforts should be placed in order to, for example, increase it or decrease it. We define the kth contribution c_k as follows

$$c_k = \frac{\mu_k^+ + \mu_k^-}{\sum_{k=1}^{K}(\mu_k^+ + \mu_k^-)},$$

where $\mu_k^+ = \int_{\lambda_{k-1}}^{\lambda_k} Q^+(p)\mathrm{d}p$ and $\mu_k^- = \int_{\lambda_{k-1}}^{\lambda_k} Q^-(p)\mathrm{d}p$, with $Q^+(p) = \max(Q(p), 0)$ and $Q^-(p) = -\min(Q(p), 0)$ the positive and negative parts of $Q(p)$ respectively. Components render a picture of the distribution of Y in terms of its expected value, providing a measure of how much the specified fractions contribute to μ.

A measure of interest that can be easily derived from the decomposition presented in (1), is the expected value of Y over each considered fraction. We define the compound expectation of Y as the set $\{\overline{\mu}_k\}_{k=1}^K$, where for every $k = 1, \ldots, K$,

$$\overline{\mu}_k = \frac{\mu_k}{\lambda_k - \lambda_{k-1}}, \quad \mu = \sum_{k=1}^{K}(\lambda_k - \lambda_{k-1})\overline{\mu}_k,$$

and $\{\overline{\mu}_k\}_{k=1}^K$ measures the mean of Y for every kth fraction.

The components, contributions and compound expectation of an illustrative example are shown in Figure 1. We consider Y that follows a standard normal distribution and we set $\{\lambda_0, \lambda_1, \lambda_2\} = \{0, 0.3, 1\}$. If we denote the standard normal quantile function as Φ^{-1},

$$\mu = \mu_1 + \mu_2 = \int_0^{0.3} \Phi^{-1}(p)\mathrm{d}p + \int_{0.3}^{1} \Phi^{-1}(p)\mathrm{d}p = -0.35 + 0.35 = 0$$

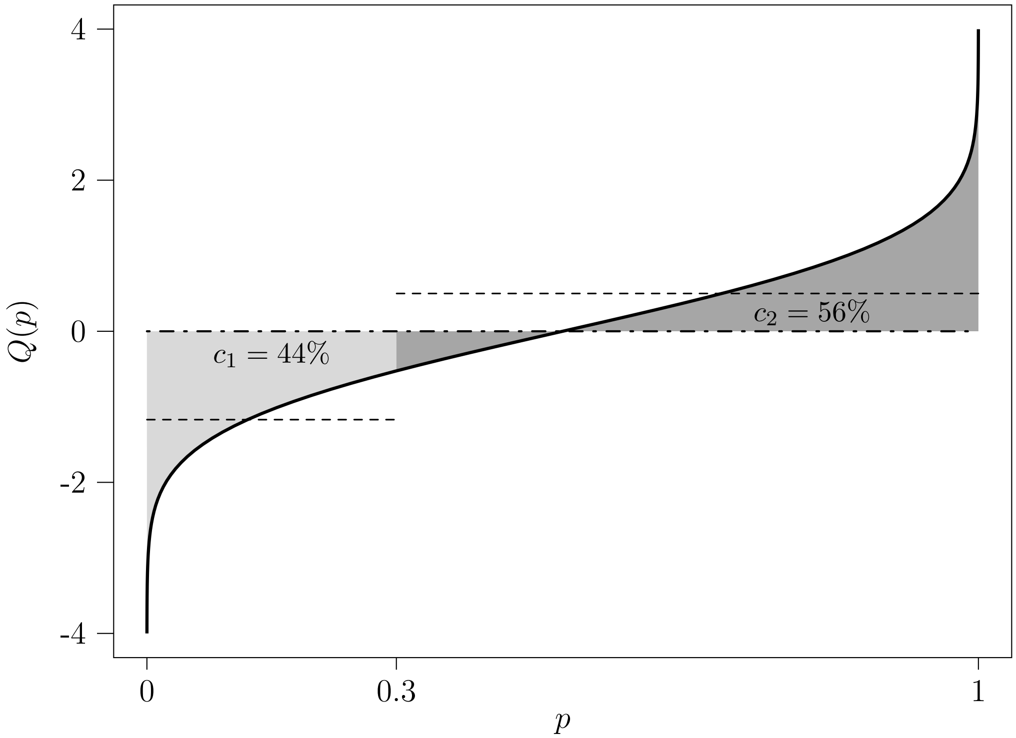

Figure 1: The light and dark gray shaded areas refer to lower and upper components, respectively, the dashed lines depict the compound expectation $\{\overline{\mu}_1, \overline{\mu}_2\}$, and the dot-dashed line depicts the total expectation. The contributions are shown in text within each component's area.

with $\{c_1, c_2\} = \{0.44, 0.56\}$ and $\{\overline{\mu}_1, \overline{\mu}_2\} = \{-1.17, 0.5\}$. In this case, 30% of Y contributes 44% to its total expectation μ and has a mean value of -1.17, while the remaining 70% contributes only 56% and has a mean value of 0.5.

The number K of components as well as the width between elements of the grid $\{\lambda_0, \ldots, \lambda_K\}$ determine the information one will retrieve from the underlying distibution. Note that the minimum grid $\{\lambda_0, \lambda_1\}$, i.e. $K = 1$, corresponds to one component representing the total expectation μ, whereas in the opposite end when $K \to \infty$, components become single quantiles.

In the following simple example, we highlight the advantages of the compound expectation over reporting single quantiles. Suppose we have two different groups of students, namely, Class A and Class B, that are graded according to a score scale that ranges from 0 to 40, and where an average of 20 is considered the minimum satisfactory grade. While Class A has a total average score of 20.3, Class B reaches only 19.6, which is below the minimum desired.

Aiming to improve the average score and performance in Class B, we take a further look at the scores' distributions in both groups. Table 1 shows the compound expectation given a $K = 10$ equally-sized components grid, $\{\lambda_0, \lambda_1, \ldots, \lambda_{10}\} = \{0, 0.1, \ldots, 1\}$, their corresponding contributions, and deciles of scores from Class A and Class B. While the deciles from both groups

Table 1: Compound expectation by deciles, overall means (last column), contributions (in %) and deciles of final scores for Class A and Class B.

Class A	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	μ
C. Expec.	10.9	12.9	15.3	17.2	18.9	20.8	22.5	24.5	27.4	32.3	20.3
Contribut.	5.38	6.36	7.55	8.48	9.32	10.3	11.1	12.1	13.5	16.0	
Deciles	11.6	13.7	16.6	18.0	19.7	21.7	23.4	25.8	29.3	37.4	

Class B	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	μ
C. Expec.	6.10	12.8	15.1	16.8	18.7	20.5	22.5	24.6	27.3	32.1	19.6
Contribut.	3.10	6.51	7.68	8.55	9.52	10.4	11.5	12.5	13.9	16.3	
Deciles	11.4	13.6	16.2	17.8	19.5	21.9	23.3	25.9	29.5	38.5	

are extremely similar and do not show any apparent differences between the two distributions, the compound expectation clearly indicates that the worst 10% of students in Class B have a lower average score than in Class A, identifying which students from Class B are more in need of an intervention (compared to Class A in this case). The bottom component summarizes all information contained in the lower tail of the distribution without the need of examining additional quantiles. This information is also conveyed by the contributions that are similar in both groups except for the bottom component, whose contribution in Class A is 2 percentage points higher than in Class B.

3 Compound expectation in a regression framework

Suppose that x is an m-dimensional vector of covariates and we want to assess its effect on each component μ_k. In this case, the components can be computed by means of the conditional quantile function, $Q(p|x)$, and for each value of x we have $\mu(x) = \sum_{k=1}^{K} \mu_k(x)$, where $\mu(x)$ is the conditional expectation of Y given x, and $\mu_k(x) = \int_{\lambda_{k-1}}^{\lambda_k} Q(p|x)\mathrm{d}p$ its kth conditional component. We focus hereafter on a specific class of all supposable $Q(p|x)$.

Let us assume that $Q(p|x)$ can be modeled in terms of x and p as follows

$$Q(p|x) = t(x)^T b(p), \tag{2}$$

where $t : \mathbb{R}^m \to \mathbb{R}^q$ is a suitable transformation of x and $b : (0,1) \to \mathbb{R}^q$ is coordinate-wise integrable, that is, its ith entry $b^i \in L^1(0,1)$ for $i = 1, \ldots, q$. Models of this form have been extensively discussed in the literature, e.g., in Koenker and Bassett Jr (1978), Efron (1991), Kim (2007), Cai and Xu (2008), Frumento and Bottai (2016) or Yuan et al. (2016), among others.

Combining (1) and (2) we have

$$\mu_k(x) = \int_{\lambda_{k-1}}^{\lambda_k} t(x)^T b(p) \ \mathrm{d}p = \int_{\lambda_{k-1}}^{\lambda_k} \left(\sum_{i=1}^{q} t^i(x) b^i(p) \right) \mathrm{d}p$$

$$
\begin{aligned}
&= \sum_{i=1}^{q} \left(\int_{\lambda_{k-1}}^{\lambda_k} t^i(x) b^i(p) \, \mathrm{d}p \right) = \sum_{i=1}^{q} \left(t^i(x) \int_{\lambda_{k-1}}^{\lambda_k} b^i(p) \, \mathrm{d}p \right) \\
&= \sum_{i=1}^{q} t^i(x) B_k^i = t(x)^T B_k, \qquad (3)
\end{aligned}
$$

where $B_k = (B_k^1, \dots, B_k^q)^T$ with $B_k^i = \int_{\lambda_{k-1}}^{\lambda_k} b^i(p) \, \mathrm{d}p$, for $i = 1, \dots, q$.

Following the definition given in Section 2, we refer to $\{\overline{\mu}_k(x)\}_{k=1}^K$ as the conditional compound expectation (CCE) of Y, where

$$
\overline{\mu}_k(x) = t(x)^T \overline{B}_k, \quad \overline{B}_k = \frac{B_k}{\lambda_k - \lambda_{k-1}}, \text{ for } k = 1, \dots, K.
$$

Elements of $\{\overline{\mu}_k(x)\}_{k=1}^K$ are each presented as a regression model, where $\overline{B}_k$ measures average differences across values of $t(x)$ for a specific kth fraction of population, making the class in (2) of utmost convenience in our setting.

In accord with the remark in Section 2, the CCE generalizes two regression models of interest. For $K = 1$,

$$
\overline{\mu}(x) = \int_0^1 t(x)^T b(p) \, \mathrm{d}p = \sum_{i=1}^{q} \left(t^i(x) \int_0^1 b^i(p) \, \mathrm{d}p \right) = \sum_{i=1}^{q} t^i(x) B^i,
$$

corresponds to linear regression, while

$$
\lim_{\epsilon \to 0} \overline{\mu}_k^{(\epsilon)}(x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_{\lambda_k - \epsilon}^{\lambda_k} t(x)^T b(p) \, \mathrm{d}p = \sum_{i=1}^{q} \left(t^i(x) \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_{\lambda_k - \epsilon}^{\lambda_k} b^i(p) \, \mathrm{d}p \right) = \sum_{i=1}^{q} t^i(x) b^i(\lambda_k),
$$

corresponds to the λ_k-th regression quantile.

4 Estimation and large sample properties

Let $Y_1, \dots, Y_n$ be a set of independent replicates of a random variable Y and $x_1, \dots, x_n$ the corresponding m-dimensional vectors of fixed given covariates. We define $S = [L^1(0,1)]^q$, the q-dimensional $L^1(0,1)$ product space with norm $\|b\|_S = \sum_{i=1}^q \|b^i\|_{L^1}$, for all $b \in S$.

Based on the sample $\{(Y_j, x_j)\}_{j=1}^n$, we consider an estimator of the class of conditional quantile functions presented in (2), $\widehat{Q}(p|x) = t(x)^T \hat{b}(p)$, with $\hat{b} \in S$. Our proposed estimator for $\mu_k(x)$ is then derived from (3) and $\widehat{Q}(p|x)$, that is,

$$
\hat{\mu}_k(x) = \int_{\lambda_{k-1}}^{\lambda_k} \widehat{Q}(p|x) \mathrm{d}p = t(x)^T \hat{B}_k, \qquad (4)
$$

where $\hat{B}_k = \int_{\lambda_{k-1}}^{\lambda_k} \hat{b}(p) \mathrm{d}p$ is an estimator of B_k, and all integrals are computed coordinate-wise.

In this section we show that properties of $\hat{B}_k$ are naturally inherited from those of $\hat{b}$, and thus the choice of $\hat{b}$ will completely determine the features of our estimator. Proofs for Propositions 1–4 can be found in Appendix A. Our first proposition characterizes the bias of $\hat{B}_k$ with respect to that of $\hat{b}$.

Proposition 1. *If $\hat{b}$ is an unbiased estimator of b almost everywhere, then $\hat{B}_k$ is an unbiased estimator of B_k for $k = 1, \ldots, K$.*

Remark 1. *For $\hat{B}_k$ to be unbiased, $\hat{b}$ needs not be unbiased almost everywhere but only the milder condition $\int_{\lambda_{k-1}}^{\lambda_k} \left[E\{\hat{b}(p) - b(p)\}\right] \mathrm{d}p = 0$ for $k = 1, \ldots, K$ is necessary.*

Consistency and asymptotic distributional properties of $\hat{B}_k$ can also be derived from those of $\hat{b}$, as we present in the following two propositions.

Proposition 2. *If $\hat{b} \in S$ is a consistent estimator of $b \in S$, then $\hat{B}_k$ is a consistent estimator of B_k for $k = 1, \ldots, K$.*

Proposition 3. *Let the sequence $(\tilde{b}_n)_{n\in\mathbb{N}} \subset S$, where the $\sqrt{n}\tilde{b}_n = \sqrt{n}(\hat{b}_n - b)$ converge weakly to a q-dimensional Gaussian process $\{G(p)\}_{p\in(0,1)}$ with mean $m(p) = 0$ for each $p \in (0, 1)$, and matrix-valued covariance function $R(p, s)$, such that $G \in S$. If $\hat{B}_{k,n} = \int_{\lambda_{k-1}}^{\lambda_k} \hat{b}_n(p)\mathrm{d}p$, then $\sqrt{n}(\hat{B}_{k,n} - B_k)$ is asymptotically multivariate normal with mean 0 and covariance matrix $\Sigma_k = \int_{\lambda_{k-1}}^{\lambda_k} \int_{\lambda_{k-1}}^{\lambda_k} R(p, s)\mathrm{d}p\mathrm{d}s$, with $k = 1, \ldots, K$.*

Finally, the next proposition presents an upper bound for the variance of $\hat{B}_k^i$ in terms of that of $\hat{b}^i$.

Proposition 4. *Let $\hat{b}^i \in L^1(0, 1)$ be a second order process, that is, $E\{(\hat{b}^i)^2\} < \infty$ for each $p \in (0, 1)$, then for $k = 1, \ldots, K$*

$$Var(\hat{B}_k^i) \leq (\lambda_k - \lambda_{k-1}) \int_{\lambda_{k-1}}^{\lambda_k} Var\{\hat{b}^i(p)\} \ \mathrm{d}p \ \text{for } i = 1, \ldots, q.$$

5 Grid of proportions

Both the width between elements of the grid of proportions $\{\lambda_0, \ldots, \lambda_K\}$ and its number K of components define the information we aim to summarize from the underlying distribution of the data. For example, a grid with larger K, i.e., one with more components, may provide a more detailed summary, where more fractions are distinctly described. In practice, however, increasing the number of components might worsen estimation precision and could eventually provide less detailed (or uncertain) information. The reason behind this lies on the result presented in Proposition 4, which upper bounds the variance of the CCE estimates $\hat{\mu}_k(x)$ by the variance of the underlying $\widehat{Q}(p|x)$. Segments on the variable's support where $\widehat{Q}(p|x)$ is more (less) precise will therefore yield better (worse) inferences to corresponding components. Note that the bound in Proposition 4 is derived for each covariate separately (there referred to by the index i), and for every given grid, the estimates' precision might differ across covariates.

For illustrative purposes, we analyzed 300 samples of a variable Y generated from the heteroscedastic model

$$Y = 2X + (3 + Z)\varepsilon,$$

where $X \sim \chi^2(2)$, $Z \sim$ Bernouilli(0.5) and $\varepsilon \sim \mathcal{N}(2, 2)$. CCEs were estimated given two different grids of proportions

$$\Lambda^1 = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}, \text{ and } \Lambda^2 = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$$

and using quantile regression (Koenker and Bassett Jr (1978)) for the underlying conditional quantile function. For every kth component $\mu_k(x)$, we considered a grid of order percentiles $\{p_0, p_1, \dots, p_J\}$, where $p_j = \lambda_{k-1} + jh$ for $j = 0, \dots, J$, and $h = 0.01$ the distance between consecutive elements of the grid. Note that $p_0 = \lambda_{k-1}$ and $p_J = \lambda_{k-1} + Jh = \lambda_k$. A trapezoid rule approximation yielded

$$\hat{B}^i_k = \int_{\lambda_{k-1}}^{\lambda_k} \hat{b}^i(p)\mathrm{d}p \approx \sum_{j=0}^{J-1} \frac{\hat{b}^i(p_{j+1}) - \hat{b}^i(p_j)}{2} h,$$

and subsequently

$$\hat{\mu}_k(x) = x^T \hat{B}_k = x^T \frac{h}{2} \left(\hat{b}(p_0) + 2\sum_{j=1}^{J-1} \hat{b}(p_j) + \hat{b}(p_J) \right).$$

Confidence intervals for $\hat{\mu}_k(x)$ relied on the asymptotic normality results presented in Koenker and Bassett Jr (1978) and on a discretized version of Proposition 3. Exact expressions for $\mathrm{Var}(\hat{B}_k)$ and $\mathrm{Var}\{\hat{\mu}_k(x)\}$ for grid-based estimators can be found in Appendix B.

As we indicated before, Λ^1 might seem preferable when one aims to retrieve a more detailed summary of the data. However, the 95% confidence interval for the top component of Y given X (with Z set to 0) is extremely large (Figure 2 (A)), providing little information about the corresponding estimate. Intuitively, such imprecision relates to the skewness of the distribution of X, which translates to less precise estimates of higher order quantiles. If we consider results given Λ^2 instead (Figure 2 (C)), estimation of the top component of Y given X is more precise, providing more detailed (although from a larger fraction) information.

We now observe that, given Λ^1, estimation of the top component of Y given Z (with X set to 0) is considerably more precise (Figure 2 (B)), specially when comparing it with other components' estimates. For the sake of completeness, we also show the compound expectation estimates of Y given Z, given the coarser grid Λ^2 (Figure 2 (D)).

We have summarized the effect that different grids might have on observed results, how it might vary across covariates, and underlined its close relation with both the chosen conditional quantile estimator and the underlying distribution of the data. Arguably, a finer grid of proportions (Λ_1) might not be optimal for obtaining a detailed summary of Y given X from our specific example, while it might be preferable when it comes to summarize Y given Z. On the contrary, a coarser grid (Λ_2) would yield more precise estimates but over larger fractions of population. Systematic approaches for obtaining optimal grids of proportions, in the sense of those that aim to find the best trade-off between precision and number of components, are highly dependable on the definition of how to exploit effectively the information contained in the data and are therefore not considered here. However, some comments on this point are included in our final remarks section (Section 7).

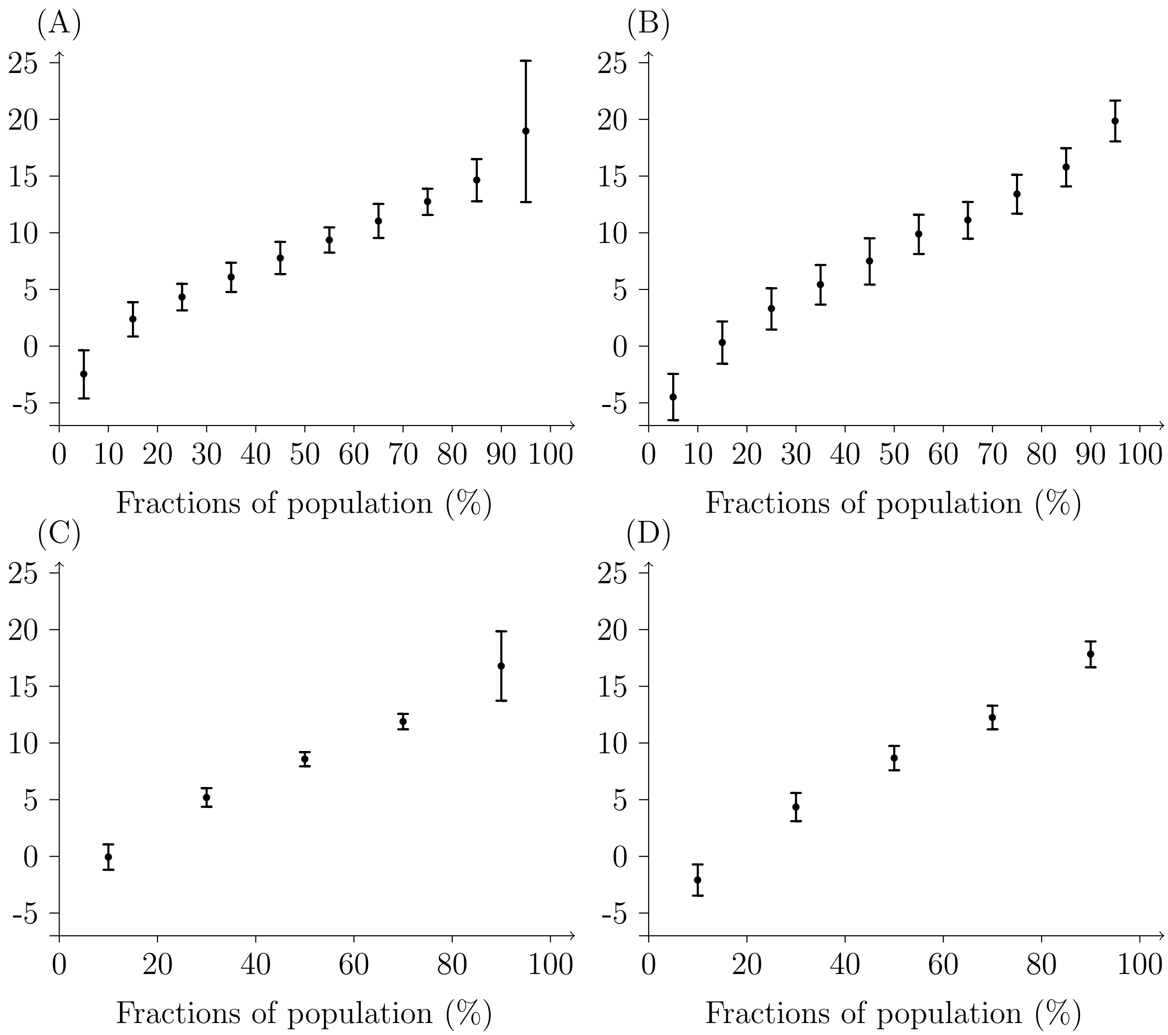

Figure 2: Estimated CCE (dots) with 95% confidence intervals (line segments) of Y given $X = 1$ (one unit increase) or $Z = 1$ (binary variable category), given the grids $\Lambda_1 = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ (panels A and B) and $\Lambda_2 = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ (panels C and D).

6 Real-data applications

6.1 Average time spent in intensive care units

The distribution of time spent in intensive care units (ICUs) is usually extremely skewed. A great majority of patients have relatively short stays whereas few present health complications and need to stay considerably longer. The latter are of interest because they experience worse medical outcomes and account for a major part of resources' allocation. In current studies, patients are often classified according to their length-of-stay and analyzed in separate groups, (see for example, Stricker et al. (2003), Trottier et al. (2007), Kramer and Zimmerman (2010)), providing an ideal setting in which to apply our method. CCE estimates can thoroughly describe time spent in ICUs without the need of defining ad-hoc cut-points that hinder meaningful groups' comparisons, and will allow to identify which and how much patients influence average length-of-stay.

Given the grid $\Lambda = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ and following the estimation strategy presented in Section 5, we estimated the CCE of length-of-stay in ICUs of 31,828 middle-aged patients (45 to 65 years old) in Sweden, in relation with their gender, age and Charlson index (Charlson et al. (1987)). The Charlson index is an integer score scale that assesses severity of existing comorbidities and is used as a prognostic indicator for disease outcome predictions. Higher Charlson scores indicate more severe comorbid conditions. Our study population was grouped in 2 categories classified with respect to their Charlson scores, namely, those with Charlson index lower or equal to 1 (mild comorbidities) and those with Charlson index larger than 1 (severe comorbidities). Further details on the data are available in Rimes-Stigare et al. (2015).

In Table 2 we show CCE estimates from where we could retrieve relevant information about average time spent in ICUs in relation with all considered factors. For instance, we observed that among the first 60% of the patients to leave the ICUs, average stay was of at most 0.85 days for any age, gender and comorbid condition. However, the top component ranged between 6.76 (for 45 years old men with mild comorbidities) and 11.3 (for 65 years old women with severe comorbidities), yielding estimated total average stays ranging from 0.89 to 1.5 days. Interpreting these results in terms of contributions (shown in Table 3), and for instance, in the reference group, we concluded that only 10% of the patients (those who stayed the longest) accounted for 73.9% of average time spent in ICUs.

Average differences across comorbidity groups can also be retrieved and are summarized in Figure 3. We observed that on average, those with severe comorbidities experienced longer stays, a trend that increased consistently across components. The largest difference, estimated at 1.86 days of average longer stay for those with severe comorbidities, was again observed amongst those on the top component, and represented a 35.2% of the total average difference between the two groups, which was only 0.49 days.

Although estimates for the top component were not as precise as for other fractions of population, all results clearly indicate a high concentration amongst the 10% with longest average stay in ICUs. Hence, the compound expectation and related quantities define a target group where to focus interventions that would yield a dramatic reduction on overall average stays and consequently average costs.

Table 2: Estimated average and CCE given the grid Λ of days spent in ICUs, in relation to comorbid condition, age and gender. The reference group refers to 45 years old females with mild comorbidities. Standard errors are displayed in parenthesis.

$\Lambda(\%)$	Reference	Severe to Mild comorb	Age	Male to Female
$0-10$	0.12(0.01)	0.06(0.01*)	0.00(0.01*)	0.00(0.01*)
$10-20$	0.14(0.05)	0.13(0.01)	0.00(0.01*)	-0.01(0.01)
$20-30$	0.18(0.05)	0.15(0.01)	0.00(0.01*)	-0.01(0.01)
$30-40$	0.29(0.04)	0.14(0.01)	0.01(0.01*)	0.00(0.01)
$40-50$	0.30(0.06)	0.19(0.01)	0.01(0.01*)	0.00(0.01)
$50-60$	0.07(0.12)	0.39(0.03)	0.02(0.01*)	0.02(0.02)
$60-70$	0.25(0.15)	0.43(0.04)	0.03(0.01*)	0.02(0.03)
$70-80$	0.21(0.25)	0.63(0.07)	0.04(0.01*)	-0.06(0.06)
$80-90$	0.80(0.56)	0.97(0.14)	0.07(0.01)	-0.20(0.13)
$90-100$	7.38(1.78)	1.86(0.41)	0.10(0.03)	-0.62(0.37)
Average	0.97(0.17)	0.49(0.04)	0.03(0.01*)	-0.08(0.04)

* denotes values that were smaller than the one displayed.

Table 3: Contributions (in %) of components of days spent in ICUs given the grid Λ, in relation to comorbid condition, age and gender. The reference group refers to 45 years old females with mild comorbidities. Bootstrapped standard errors are displayed in parenthesis.

$\Lambda(\%)$	Reference	Severe to Mild comorb	Age	Male to Female
$0-10$	1.17(0.29)	1.16(0.16)	0.07(0.07)	0.10(0.72)
$10-20$	1.56(0.55)	2.76(0.36)	1.20(0.34)	1.30(1.64)
$20-30$	2.00(0.54)	3.06(0.38)	2.28(0.40)	0.59(1.18)
$30-40$	3.23(0.61)	2.92(0.33)	2.74(0.41)	0.03(0.61)
$40-50$	3.34(0.66)	3.94(0.45)	3.76(0.61)	0.54(0.98)
$50-60$	0.76(0.64)	8.22(0.92)	6.99(0.99)	2.08(2.54)
$60-70$	2.82(1.31)	9.12(0.97)	9.23(1.06)	2.35(5.76)
$70-80$	2.36(1.71)	13.2(1.19)	15.2(1.39)	6.32(4.80)
$80-90$	8.89(4.49)	20.4(1.60)	25.7(1.87)	23.0(8.04)
$90-100$	73.9(5.64)	35.2(4.85)	32.8(4.94)	63.7(14.3)

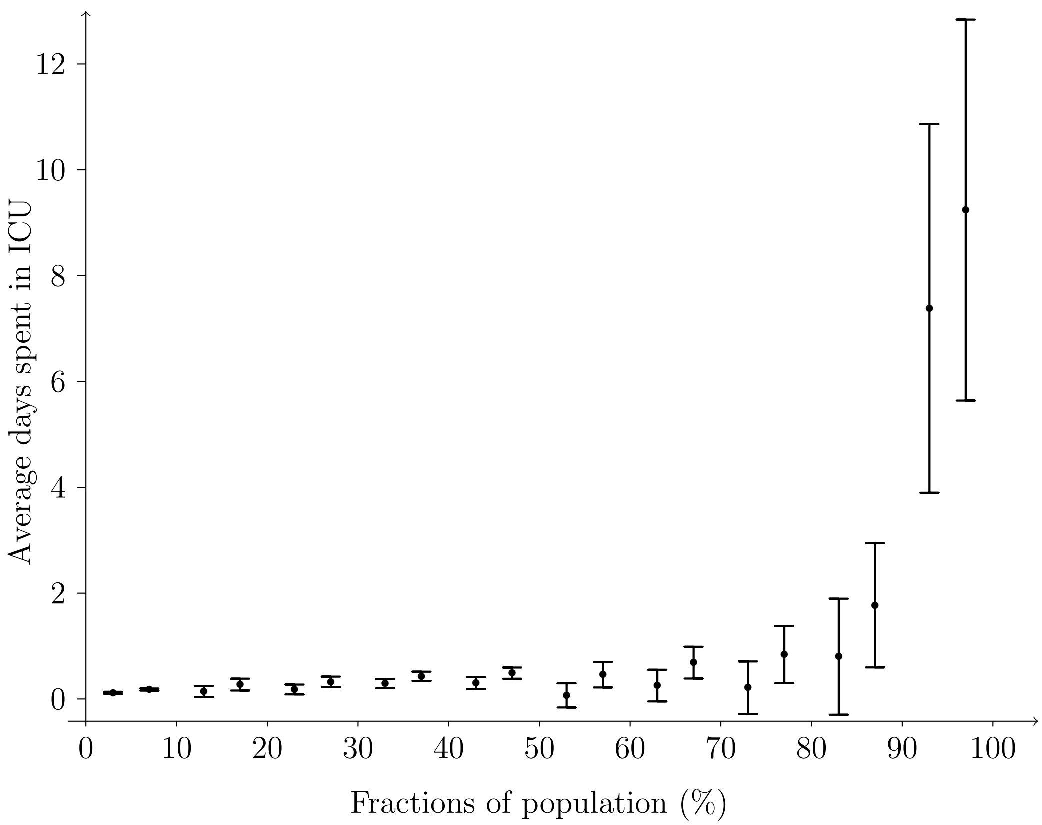

Figure 3: Estimated CCE and 95% confidence intervals of time spent in ICUs for 45 years old females across comorbid condition groups given the grid Λ. For each component we showed (from left to right) the mild and severe comorbidities groups.

6.2 Average weight gain amongst adolescents

The obesity epidemic constitutes a major health burden in numerous countries, particularly in the United States. Health-care prevention strategies often focus on adolescents, so as to revert obesity at a young age and control the spread of the disease. In studies that aim to estimate the effect of a specific intervention, a commonly used measure is average weight gain over a certain period of time (see, for example, Kornet van der Aa et al. (2017) for a review). Average weight gain estimates, however, are often misleading because they might easily take on negative values that will balance out with positive ones. In these situations, the compound expectation proves a useful measure because it allows to distinguish between fractions of the population, clearly separating those who lost weight, or in other words, those with a negative weight gain, and quantifying their contribution to the total average.

As an illustrative example, we estimated the CCE of weight gain amongst adolescents given the grid $\Lambda = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$, in relation with their age and obesity condition (obese versus non-obese). We analyzed a subset of $7,410$ adolescents from the National Longitudinal Survey of Youth (NLSY97) Cohort (Bureau of Labor Statistics, 2014) that followed young people in the United States (aged 13 to 17 on their first interview) through yearly or biennial interviews in which multiple aspects of their lives were reported. We considered their weight (in kilograms) at two different time points (as reported in 1997 and 2002) and computed their difference so that we could estimate average weight gain 5 years after the first interview, and supposed they had been under a certain weight-control intervention that we aim to evaluate.

Table 4 shows large disparities on estimated average weight gain across different fractions of population. For example in the reference group (non-obese 15 years old), CCE estimates ranged from -1.17 (bottom component) to 33.3 Kg (top component) with a total average of 13.3 Kg of weight gain. Contributions of these components to the total average were 0.86% and 24.6%, respectively (Table 5).

Regarding differences between the obese and non-obese groups we observe that, on average, the difference in weight gain was only of 1.27 Kg, that is, hinting almost no effect of the intervention when comparing both groups. However, once divided into fractions of population, the bottom and top components represent a contribution of more than 20% each to the total average difference, with estimates of -6.08 and 7.97 Kg, respectively. In other words, while on average it might seem that the difference between the obese and non-obese groups is almost negligible, a closer inspection revealed larger differences, and that those are highly concentrated around the 10% of adolescents that lost and gained the most. In this case, the intervention would be producing almost opposite effects in the fractions corresponding to the bottom and top components. Those two groups, that represent the lower and upper tail of the distribution of weight gain difference, are comprehensively summarized by their respective components defining target groups in which to modify the intervention.

7 Final remarks

A major strength of our approach lies on the possibility of choosing any quantile function estimator for the underlying distribution of the data, ensuring that all the compound expectation's

Table 4: Estimated average and CCE of weight gain given the grid Λ, in relation to obesity condition and age at baseline (1997). The reference group refers to 15 years old (mean age at baseline) non-obese adolescents. Standard errors are displayed in parenthesis.

$\Lambda(\%)$	Reference	Obese to Non-obese	Age
$0-10$	$-1.17(0.56)$	$-6.08(0.48)$	$-1.54(0.14)$
$10-20$	$3.96(0.33)$	$-1.6(0.29)$	$-1.78(0.11)$
$20-30$	$6.27(0.32)$	$-0.25(0.31)$	$-2.13(0.09)$
$30-40$	$8.49(0.40)$	$0.68(0.31)$	$-2.39(0.11)$
$40-50$	$10.7(0.42)$	$1.37(0.30)$	$-2.77(0.13)$
$50-60$	$13.1(0.44)$	$1.89(0.33)$	$-3.19(0.11)$
$60-70$	$15.9(0.44)$	$2.11(0.32)$	$-3.45(0.12)$
$70-80$	$19.2(0.48)$	$2.68(0.36)$	$-3.69(0.15)$
$80-90$	$23.3(0.69)$	$3.93(0.50)$	$-3.84(0.19)$
$90-100$	$33.3(0.91)$	$7.97(0.67)$	$-4.18(0.24)$
Average	$13.3(0.17)$	$1.27(0.13)$	$-2.89(0.05)$

Table 5: Contributions (in %) of components of weight gain given the grid Λ, in relation to to obesity condition and age at baseline (1997). The reference group refers to 15 years old (mean age at baseline) non-obese adolescents. Bootstrapped standard errors are displayed in parenthesis.

$\Lambda(\%)$	Reference	Obese to Non-obese	Age
$0-10$	$0.86(0.76)$	$21.3(2.52)$	$5.33(0.63)$
$10-20$	$2.93(0.22)$	$5.69(1.23)$	$6.14(0.33)$
$20-30$	$4.63(0.20)$	$0.86(0.78)$	$7.35(0.23)$
$30-40$	$6.27(0.27)$	$2.39(1.07)$	$8.24(0.26)$
$40-50$	$7.86(0.26)$	$4.78(0.97)$	$9.56(0.31)$
$50-60$	$9.70(0.30)$	$6.61(0.93)$	$11.0(0.25)$
$60-70$	$11.8(0.25)$	$7.38(0.78)$	$11.9(0.21)$
$70-80$	$14.2(0.28)$	$9.36(0.85)$	$12.8(0.28)$
$80-90$	$17.2(0.42)$	$13.8(1.33)$	$13.3(0.44)$
$90-100$	$24.6(0.64)$	$27.9(1.82)$	$14.4(0.62)$

estimation properties are to be inherited, and thus providing theoretically sound inference techniques. The large scope and variety of available methods, e.g., Portnoy (2003) and Frumento and Bottai (2017) for censored data, or Koenker (2004), Geraci and Bottai (2007) and Wang and Fygenson (2009) for longitudinal data, among others, provide our method with flexibility to adapt to almost any scenario.

Because quantiles are invariant to monotone transformations, models that estimate quantiles of these transformations from the original data might also be used. In such case, however, the estimated regression coefficients are often not directly interpretable and results are to be provided for each specific value of the covariates.

The grid of proportions also provides flexibility to our measure, allowing to summarize the underlying distribution in as many fractions deemed appropriate. While in the present work we only considered a fixed given number and width of fractions of population in which to estimate the compound expectation, systematic approaches that choose optimally amongst all possible grids of proportions could also be explored. Systematic approaches could include, for example, finding the number and width of components that minimized the mean squared error between the underlying quantile function and the compound expectation estimates or that maintained the precision of the compound expectation estimates constant across components.

Finally, our measure has shown to be of great use in practical applications where average measures are of interest, and a more detailed look at the underlying distribution of the data reveals relevant information that might have been overlooked otherwise.

Appendix A

Proof of Proposition 1. Because

$$\mathrm{E}(\hat{B}_k - B_k) = \mathrm{E}\left\{\int_{\lambda_{k-1}}^{\lambda_k} \hat{b}(p)\mathrm{d}p - \int_{\lambda_{k-1}}^{\lambda_k} b(p)\mathrm{d}p\right\} = \int_{\lambda_{k-1}}^{\lambda_k} \left[\mathrm{E}\{\hat{b}(p)\} - b(p)\right] \mathrm{d}p,$$

if $\mathrm{E}\{\hat{b}(p) - b(p)\} = 0$ almost everywhere then $\mathrm{E}(\hat{B}_k - B_k) = 0$, for $k = 1, \ldots, K$.

Proof of Proposition 2. Let $(\hat{b}_n)_{n\in\mathbb{N}} \subset S$ be a sequence of consistent estimators of $b \in S$, that is,

$$\lim_{n\to\infty} \Pr\left(\|\hat{b}_n - b\|_S > \varepsilon\right) = 0.$$

Consider the sequence $\hat{B}_{k,n} = \int_{\lambda_{k-1}}^{\lambda_k} \hat{b}_n(p)\mathrm{d}p$ of estimators of B_k for $k = 1, \ldots, K$. Then,

$$\begin{aligned}\left\|\hat{B}_{k,n} - B_k\right\|_\infty &= \max_{i\in\{1,\ldots,q\}} \left|\int_{\lambda_{k-1}}^{\lambda_k} \left(\hat{b}_n^i(p) - b^i(p)\right) \mathrm{d}p\right| \\ &\leq \sum_{i=1}^{q} \left|\int_{\lambda_{k-1}}^{\lambda_k} \left(\hat{b}_n^i(p) - b^i(p)\right) \mathrm{d}p\right| \leq \sum_{i=1}^{q} \int_{\lambda_{k-1}}^{\lambda_k} \left|\hat{b}_n^i(p) - b^i(p)\right| \mathrm{d}p \\ &\leq \sum_{i=1}^{q} \int_0^1 \left|\hat{b}_n^i(p) - b^i(p)\right| \mathrm{d}p = \left\|\hat{b}_n - b\right\|_S,\end{aligned}$$

and because $\left\|\hat{B}_{k,n} - B_k\right\|_\infty \geq 0$, we have

$$\lim_{n\to\infty} \Pr\left(\|\hat{B}_{k,n} - B_k\|_\infty > \varepsilon\right) = 0.$$

Proof of Proposition 3. Consider the linear continuous operator

$$\begin{aligned} T_k : S &\longrightarrow \mathbb{R}^q \\ b &\longmapsto \int_{\lambda_{k-1}}^{\lambda_k} b(p)\ \mathrm{d}p, \end{aligned}$$

where the integral is computed coordinate-wise. That T_k is continuous follows easily from

$$\|T_k(b)\|_\infty = \left\|\int_{\lambda_{k-1}}^{\lambda_k} b(p)\ \mathrm{d}p\right\|_\infty = \max_{i\in\{1,\ldots,q\}} \left|\int_{\lambda_{k-1}}^{\lambda_k} b^i(p)\ \mathrm{d}p\right| \leq \sum_{i=1}^q \|b^i\|_{L^1} = \|b\|_S.$$

Because $\sqrt{n}\tilde{b}_n \to G$ in S, the continuous mapping theorem (Kallenberg, 1997, Th. 3.27, p. 54) yields

$$\sqrt{n}\int_{\lambda_{k-1}}^{\lambda_k} \tilde{b}_n(p)\ \mathrm{d}p \to \int_{\lambda_{k-1}}^{\lambda_k} G(p)\ \mathrm{d}p,\ \text{ that is, } \sqrt{n}(\hat{B}_{k,n} - B_k) \to \int_{\lambda_{k-1}}^{\lambda_k} G(p)\ \mathrm{d}p,$$

where $\to$ denotes weak convergence.

Let us now compute the distribution of $\int_{\lambda_{k-1}}^{\lambda_k} G(p)\ \mathrm{d}p$. We have

$$\int_{\lambda_{k-1}}^{\lambda_k} G(p)\ \mathrm{d}p = \lim_{L\to+\infty} \frac{\lambda_k - \lambda_{k-1}}{L} \sum_{l=1}^L G_l,$$

where $G_l = G(\lambda_{k-1} + lh)$ with $h = L^{-1}(\lambda_k - \lambda_{k-1})$.

Because $G = (G^1, \ldots, G^q)$ is a Gaussian process, we know that the vector

$$\sum_{l=1}^L G_l = \left(\sum_{l=1}^L G_l^1, \ldots, \sum_{l=1}^L G_l^q\right)$$

is multivariate normally distributed, with mean 0 and variance covariance matrix

$$\mathrm{Var}\left(\sum_{l=1}^L G_l\right) = \mathrm{E}\left(\sum_{l=1}^L \sum_{r=1}^L G_l^T G_r\right) = \sum_{l=1}^L \sum_{r=1}^L \mathrm{E}(G_l G_r) = \sum_{l=1}^L \sum_{r=1}^L R(l, r).$$

The fact that $\sum_{l=1}^L G_l$ has mean 0 follows immediately from $m(p) = 0$ for all $p \in (0, 1)$.

We consider now the normalised multivariate normal $Z_L \sim \mathcal{MN}(0, I_q)$ where I_q is the q-dimensional identity matrix, $Z_L = \Sigma_{K,L}^{-1/2} L^{-1}(\lambda_k - \lambda_{k-1}) \sum_{l=1}^L G_l$ and $\Sigma_{K,L} = L^{-2}(\lambda_k -$

$\lambda_{k-1})^2 \sum_{l=1}^{L} \sum_{r=1}^{L} R(l,r)$. It is clear that $Z_L \to Z$ with $Z \sim \mathcal{MN}(0, I_q)$. Applying now Slutsky's theorem we have

$$\frac{\lambda_k - \lambda_{k-1}}{L} \sum_{l=1}^{L} G_l \sim \mathcal{MN}(0, \Sigma_K),$$

where

$$\Sigma_k = \lim_{L \to +\infty} \left(\frac{\lambda_k - \lambda_{k-1}}{L} \right)^2 \sum_{l=1}^{L} \sum_{r=1}^{L} R(l,r) = \int_{\lambda_{k-1}}^{\lambda_k} \int_{\lambda_{k-1}}^{\lambda_k} R(p,s) \mathrm{d}p \mathrm{d}s,$$

concluding our proof.

Proof of Proposition 4.

$$\begin{aligned}
\mathrm{Var}(\hat{B}_k^i) &= \mathrm{E}\left(\left| \hat{B}_k^i - \mathrm{E}(\hat{B}_k^i) \right|^2 \right) = \mathrm{E}\left(\left| \int_{\lambda_{k-1}}^{\lambda_k} \left[\hat{b}^i(p) - \mathrm{E}\{\hat{b}^i(p)\} \right] \mathrm{d}p \right|^2 \right) \\
&\leq \mathrm{E}\left(\left[\int_{\lambda_{k-1}}^{\lambda_k} \left| \hat{b}^i(p) - \mathrm{E}\{\hat{b}^i(p)\} \right| \mathrm{d}p \right]^2 \right) \\
&\overset{(1)}{\leq} \mathrm{E}\left[\int_{\lambda_{k-1}}^{\lambda_k} \left| \hat{b}^i(p) - \mathrm{E}\{\hat{b}^i(p)\} \right|^2 \mathrm{d}p \int_{\lambda_{k-1}}^{\lambda_k} 1^2 \mathrm{d}p \right] \\
&= (\lambda_k - \lambda_{k-1}) \int_{\lambda_{k-1}}^{\lambda_k} \mathrm{E}\left[\left| \hat{b}^i(p) - \mathrm{E}\{\hat{b}^i(p)\} \right|^2 \right] \mathrm{d}p \\
&= (\lambda_k - \lambda_{k-1}) \int_{\lambda_{k-1}}^{\lambda_k} \mathrm{Var}\{\hat{b}^i(p)\} \mathrm{d}p,
\end{aligned}$$

where in (1) we use Hölders' inequality.

Appendix B

For any grid-based quantile function estimator like, for example, the one obtained from regression quantiles, the $q \times q$ variance-covariance matrices for the compound expectation regression coefficients $\hat{B}_k = (\hat{B}_k^1, \ldots, \hat{B}_k^q)$ for each $k = 1, \ldots, K$ computed as in Section 6 are given by

$$\begin{aligned}
\mathrm{Var}\left(\hat{B}_k\right) = \mathrm{Var}&\left\{ \frac{h}{2} \left(\hat{b}(p_0) + 2 \sum_{j=1}^{J-1} \hat{b}(p_j) + \hat{b}(p_J) \right) \right\} = \frac{h^2}{4} \left\{ \mathrm{Var}\left(\hat{b}(p_0) \right) \right. \\
&+ 4\mathrm{Var}\left(\sum_{j=1}^{J-1} \hat{b}(p_j) \right) + \mathrm{Var}\left(\hat{b}(p_J) \right) + 2\mathrm{Cov}\left(\hat{b}(p_0), \sum_{j=1}^{J-1} \hat{b}(p_j) \right) \\
&+ 2\mathrm{Cov}^T\left(\hat{b}(p_0), \sum_{j=1}^{J-1} \hat{b}(p_j) \right) + \mathrm{Cov}\left(\hat{b}(p_0), \hat{b}(p_J) \right) + \mathrm{Cov}^T\left(\hat{b}(p_0), \hat{b}(p_J) \right)
\end{aligned}$$

$$+2\text{Cov}\left(\sum_{j=1}^{J-1}\hat{b}(p_j),\hat{b}(p_J)\right)+2\text{Cov}^T\left(\sum_{j=1}^{J-1}\hat{b}(p_j),\hat{b}(p_J)\right)\Bigg\}$$

$$=\frac{h^2}{4}\Bigg[\text{Var}\left(\hat{b}(p_0)\right)+\text{Var}\left(\hat{b}(p_J)\right)+\text{Cov}\left(\hat{b}(p_0),\hat{b}(p_J)\right)$$

$$+\text{Cov}^T\left(\hat{b}(p_0),\hat{b}(p_J)\right)+2\sum_{j=1}^{J-1}\Big\{\text{Cov}\left(\hat{b}(p_0),\hat{b}(p_j)\right)+\text{Cov}^T\left(\hat{b}(p_0),\hat{b}(p_j)\right)$$

$$+\text{Cov}\left(\hat{b}(p_j),\hat{b}(p_J)\right)+\text{Cov}^T\left(\hat{b}(p_j),\hat{b}(p_J)\right)\Big\}+4\sum_{j=1}^{J-1}\text{Var}\left(\hat{b}(p_j)\right)$$

$$+4\sum_{j=2}^{J-2}\sum_{l=j+1}^{J-1}\Big\{\text{Cov}\left(\hat{b}(p_l),\hat{b}(p_j)\right)+\text{Cov}^T\left(\hat{b}(p_l),\hat{b}(p_j)\right)\Big\}\Bigg],$$

where $\text{Var}\{\hat{b}(p_j)\}$ denotes the variance-covariance matrix for the p_jth quantile regression coefficients and $\text{Cov}\{\hat{b}(p_j),\hat{b}(p_l)\}$ the covariance matrix between the p_jth and p_lth quantile regression coefficients.

The variance of each kth conditional component follows easily.

$$\text{Var}\{\hat{\mu}_k(x)\}=\text{Var}\left(x^T\hat{B}_k\right)=\text{Var}\left(\sum_{i=1}^{q}x^i\hat{B}_k^i\right)=$$

$$=\sum_{i=1}^{q}(x^i)^2\text{Var}\left(\hat{B}_k^i\right)+2\sum_{i=1}^{q-1}\sum_{j=i+1}^{q}x^ix^j\text{Cov}\left(\hat{B}_k^i,\hat{B}_k^j\right),$$

where $\text{Var}\left(\hat{B}_k^i\right)$ and $\text{Cov}\left(\hat{B}_k^i,\hat{B}_k^j\right)$ are entries from the matrix $\text{Var}\left(\hat{B}_k\right)$.

Acknowledgements

The authors would like to thank Pol del Aguila Pla and Erin Gabriel for fruitful discussions that helped to improve the quality of this paper. This project was partially funded by the KID doctoral grant from Karolinska Institutet.

References

Aaberge, R. (2009). Ranking intersecting Lorenz curves. *Social Choice and Welfare*, 33(2):235–259.

Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika*, 75(4):761–771.

Bureau of Labor Statistics, U. D. (2014). National Longitudinal Survey of Youth 1979 Cohort, 1979-2002 (rounds 1-6). *Center for Human Resource Research, The Ohio State University.*

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 103(484):1595–1608.

Chambers, R. and Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, 93(2):255–268.

Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*, 40(5):373–383.

Christopoulos, K. A., Hartogensis, W., Glidden, D. V., Pilcher, C. D., Gandhi, M., and Geng, E. H. (2017). The Lorenz curve: a novel method for understanding viral load distribution at the population level. *AIDS*, 31(2):309–310.

Davies, J. and Hoy, M. (1995). Making inequality comparisons when Lorenz curves intersect. *The American Economic Review*, 85(4):980–986.

Delbosc, A. and Currie, G. (2011). Using Lorenz curves to assess public transport equity. *Journal of Transport Geography*, 19(6):1252–1259. Special section on Alternative Travel futures.

Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, 1(1):93–125.

Egghe, L. and Rousseau, R. (1990). Elements of concentration theory. In *Proceedings of the Second International Conference on Bibliometrics, Scientometrics and Infometrics*, London, Ontario.

Ehm, W., Gneiting, T., Jordan, A., and Krger, F. (2016). Of quantiles and expectiles: consistent scoring functions, choquet representations and forecast rankings. *J. R. Stat. Soc. B*, 78(3):505–562.

Frumento, P. and Bottai, M. (2016). Parametric modeling of quantile regression coefficient functions. *Biometrics*, 72:74–84.

Frumento, P. and Bottai, M. (2017). Parametric modeling of quantile regression coefficient functions with censored and truncated data. *Biometrics*, 0. Preprint.

Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostat*, 8(1):140–154.

Kallenberg, O. (1997). *Foundations of Modern Probability.* Springer Science & Business Media.

Kim, M.-O. (2007). Quantile regression with varying coefficients. *The Annals of Statistics*, 35(1):92–108.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 46(1):33–50.

Kornet van der Aa, D. A., Altenburg, T. M., Randeraadvan der Zee, C. H., and Chinapaw, M. J. M. (2017). The effectiveness and promising strategies of obesity prevention and treatment programmes among adolescents from disadvantaged backgrounds: a systematic review. *Obesity Reviews*, 18(5):581–593.

Kramer, A. A. and Zimmerman, J. E. (2010). A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay. *BMC Medical Informatics and Decision Making*, 10(1):1–16.

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 9(70):209–219.

Mauguen, A. and Begg, C. B. (2016). Using the Lorenz curve to characterize risk predictiveness and etiologic heterogeneity. *Epidemiology*, 27(4):531–537.

Meehl, G. A., Zwiers, F., Evans, J., Knutson, T., Mearns, L., and Whetton, P. (2000). Trends in extreme weather and climate events: Issues related to modeling extremes in projections of future climate change. *Bulletin of the American Meteorological Society*, 81(3):427–436.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, 55(4):819–847.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, 98(464):1001–1012.

Rimes-Stigare, C., Frumento, P., Bottai, M., Mårtensson, J., Martling, C.-R., Walther, S. M., Karlström, G., and Bell, M. (2015). Evolution of chronic renal impairment and long-term mortality after de novo acute kidney injury in the critically ill; a swedish multi-centre cohort study. *Critical Care*, 19(1):221.

Seaman, R., Leyland, A. H., and Popham, F. (2016). Increasing inequality in age of death at shared levels of life expectancy: A comparative study of Scotland and England and Wales. *SSM - Population Health*, 2:724–731.

Shorrocks, A. F. (1983). Ranking income distributions. *Economica*, 50(197):3–17.

Stricker, K., Rothen, H. U., and Takala, J. (2003). Resource use in the ICU: short- vs. long-term patients*. *Acta Anaesthesiologica Scandinavica*, 47(5):508–515.

Trottier, V., McKenney, M. G., Beninati, M., Manning, R., and Schulman, C. I. (2007). Survival after prolonged length of stay in a trauma intensive care unit. *Journal of Trauma and Acute Care Surgery*, 62(1):147–150.

Wai, J. and Putallaz, M. (2011). The Flynn effect puzzle: A 30-year examination from the right tail of the ability distribution provides some missing pieces. *Intelligence*, 39(6):443–455.

Wang, H. J. and Fygenson, M. (2009). Inference for censored quantile regression models in longitudinal studies. *The Annals of Statistics*, 37(2):756–781.

Wang, H. J. and Zhou, X.-H. (2010). Estimation of the retransformed conditional mean in health care cost studies. *Biometrika*, 97(1):147–158.

Yuan, Y., Chen, N., and Zhou, S. (2016). Modeling regression quantile process using monotone B-splines. *Technometrics*, 0. Preprint.