

November 2015

A Novel Method for Assessing Co-monotonicity: an Interplay between Mathematics and Statistics with Applications

Danang T. Qoyyimi

The University of Western Ontario

Supervisor

Dr. Ricardas Zitikis


The University of Western Ontario

Graduate Program in Statistics and Actuarial Sciences

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Danang T. Qoyyimi 2015

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

 Part of the [Multivariate Analysis Commons](#), [Probability Commons](#), and the [Statistical Methodology Commons](#)

Recommended Citation

Qoyyimi, Danang T., "A Novel Method for Assessing Co-monotonicity: an Interplay between Mathematics and Statistics with Applications" (2015). *Electronic Thesis and Dissertation Repository*. 3322.
<https://ir.lib.uwo.ca/etd/3322>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlsadmin@uwo.ca.

A NOVEL METHOD FOR ASSESSING CO-MONOTONICITY:
AN INTERPLAY BETWEEN MATHEMATICS AND STATISTICS
WITH APPLICATIONS

(Thesis format: Monograph)

by

Danang Teguh Qoyyimi

Graduate Program in Statistical and Actuarial Science

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies

The University of Western Ontario

London, Ontario, Canada

© Danang Teguh Qoyyimi 2015

Abstract

Numerous problems in econometrics, insurance, reliability engineering, and statistics rely on the assumption that certain functions are monotonic, which may or may not be true in real life scenarios. To satisfy this requirement, from the theoretical point of view, researchers frequently model the underlying phenomena using parametric and semi-parametric families of functions, thus effectively specifying the required shapes of the functions. To tackle these problems in a non-parametric way, when the shape cannot be specified explicitly but only estimated approximately, we suggest indices for measuring the lack of monotonicity in functions. We investigate properties of these indices and offer convenient computational techniques for practical use. To illustrate the new technique, we analyze a data-set of student marks on mathematics, reading and spelling. In particular, we apply our technique to determine if the marks are co-monotonic, but if not, then how much they deviate from the co-monotonic pattern. This illustrative example is for convenience only, as our technique is applicable very widely. Indeed, measuring the lack of co-monotonicity between variables plays an important role in a great variety of research areas, as noted at the beginning of this abstract.

Keywords: co-monotonicity, monotone rearrangement, convex rearrangement, association, dependence measure, education, performance evaluation.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor, Dr. Ričardas Zitikis, for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the stages of research. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides, I would like to thank my thesis committee: Professors Girma Bitsuamlak, Edward Furman, Kristina Sendova, and Douglas Woolford for their insightful comments and encouragements.

My sincere thanks also goes to the Directorate General of Human Resources for Science, Technology and Higher Education of Indonesia for providing funding for my study. Without their support it would not have been possible to complete my PhD study.

I thank my student fellows at DSAS of the University of Western Ontario for the stimulating discussions as well as for all the fun that we have had during the four years of my study. Also I thank my colleagues at the Universitas Gadjah Mada for their support.

Last but not the least, I would like to thank my family: my beloved wife and son, my parents, sisters and brothers for supporting me spiritually throughout my work on this thesis and my life in general.

Contents

Abstract	ii
Acknowledgements	iii
List of tables	vi
List of figures	vii
1 Introduction	1
1.1 Co-monotonicity	1
1.2 Convex bound of aggregate risks	5
1.3 Dependence structure via copula	7
1.4 Problem statement	10
1.5 Scope and summary of the thesis	11
2 Measuring the lack of monotonicity in functions	13
2.1 Introduction	13
2.2 An index of non-decreasingness	17
2.3 Probabilistic interpretation	21
2.4 Co-monotonically additive index	22
2.5 Main theorem	28

2.6	Concluding notes	35
3	Computing the indices	36
3.1	General considerations	36
3.2	Illustrations with insights into the indices	40
3.3	Indices of functions on finite intervals	44
4	Measuring association via lack of co-monotonicity	48
4.1	Introduction	48
4.2	Data	51
4.3	Curve fitting	52
4.3.1	Constructing \hat{h}	54
4.3.2	Bandwidth selection	56
4.4	Measuring the lack of co-monotonicity	57
4.5	Data analysis and findings	58
4.6	Comparing the LOC index with Liebscher's ζ	61
4.7	Concluding notes	67
5	Concluding remarks and future work	68
5.1	Concluding remarks	68
5.2	Future work	69
	Bibliography	70
	Curriculum vitae	83

List of tables

3.1	Indices of non-decreasingness of $h_1(t) = \sin(tM)$ and $h_2(t) = \cos(tM)$.	40
3.2	Increasing (+) and decreasing (−) regions of h_1 and h_2 .	42
3.3	Indices of non-decreasingness of $\sin(t)$ and $\cos(t)$ on $[0, M]$.	45
4.1	Summary statistics.	52
4.2	Pearson correlation coefficients.	52
4.3	Conditional-mean based LOC matrix (entries multiplied by 1,000, see Note 4.5.1).	60
4.4	Conditional-median based LOC matrix (entries multiplied by 1,000, see Note 4.5.1).	60

List of figures

2.1	Function $h_\alpha(t)$ for various α	25
2.2	The indices \mathcal{I}_h (solid-red) and \mathcal{L}_h (dashed-blue) as functions of α . . .	26
2.3	Function $g_{0.25}(t)$	27
2.4	The indices \mathcal{I}_{g_α} (solid-red) and \mathcal{L}_{g_α} (dashed-blue) as functions of α . .	28
3.1	Functions $h_1(t) = \sin(tM)$ (solid) and $h_2(t) = \cos(tM)$ (dotted). . . .	41
4.1	Frequency histograms.	52
4.2	Scatterplots and least-squares regression lines.	53
4.3	Conditional-mean (blue) and conditional-median (red) based curves. .	59
4.4	Scatterplots and fitted functions.	66

Chapter 1

Introduction

Studies of dependence structures have grown significantly in recent years. They were triggered by the fact that the assumption of independence between random variables in many situations is not reasonable. Indeed, in practice we more often find random variables that are dependent. The dependence between them may be generated by some common external mechanisms. For example, yields of agriculture are subjected to weather conditions during the growing season. Another example is life expectancy of several individuals who are exposed to common risks. If one random variable takes large values, then other random variables will tend to behave analogously. This is what in the literature is termed co-monotonicity.

1.1 Co-monotonicity

The term of co-monotonicity comes from an abbreviation of “common monotonicity”, which was introduced by Schmeidler (1986) to denote the agreement of monotonic patterns in functions (see, e.g., Denneberg, 1994; Puccetti and Wang, 2015). Two

functions are said to be co-monotonic if and only if they are increasing or decreasing simultaneously. Namely, functions f and g are co-monotonic if and only if $(f(t_1) - f(t_2))(g(t_1) - g(t_2)) \geq 0$ for every t_1 and t_2 in the common domain of f and g .

We can also find co-monotonicity in various other contexts such as vectors, sets, random variables, and probability measures. Its definition in these contexts can be adapted accordingly. For instance, two n -dimensional vectors $\mathbf{x} = \{x_i\}$ and $\mathbf{y} = \{y_i\}$ are co-monotonic if and only if $x_i \leq y_i$ for all $i = 1, \dots, n$, or $x_i \geq y_i$ for all $i = 1, \dots, n$. A set of vectors, therefore, is said to be co-monotonic if all pairs of its members are co-monotonic. In probability theory, an n -dimensional random vector $\mathbf{X} = \{X_i\}, i = 1, \dots, n$, with joint distribution function $F_{\mathbf{X}}$ is co-monotonic if and only if the support of $F_{\mathbf{X}}$ is a co-monotonic set. The existence of co-monotonic support in the actuarial literature is often interpreted as the existence of an external mechanism that moves all random coordinates X_i of \mathbf{X} in the same direction. Hence, we have another interpretation of definition of co-monotonic random vectors (cf., e.g., Denuit et al., 2005), as follows.

Definition 1.1.1 *An n -dimensional random vector \mathbf{X} is co-monotonic if and only if there exist non-decreasing functions $t_i, i = 1, \dots, n$, and a random variable Z such that $\mathbf{X} = (t_1(Z), t_2(Z), \dots, t_n(Z))$.*

One of the many interesting parts in the study of dependence structures is how to form a random vector from univariate random variables. Given univariate random variables X_1, X_2, \dots, X_n with distribution functions $F_{X_1}, F_{X_2}, \dots, F_{X_n}$, many joint distribution functions can be constructed. In the case when random variables are normally distributed, there is a correlation matrix that is used to form a multivariate normal distribution.

In general, dependence structures give insights on how to form a joint distribution function from marginals F_{X_1}, \dots, F_{X_n} . Namely, given distribution functions F_{X_i} of X_i for $i = 1, \dots, n$, we set a collection of all random vectors \mathbf{Y} whose marginals F_{Y_i} are equal to the respective marginals F_{X_i} . This set is called the Fréchet-Hoeffding class and denoted by $R_n(F_{X_1}, F_{X_2}, \dots, F_{X_n})$. For every $\mathbf{X} \in R_n(F_{X_1}, F_{X_2}, \dots, F_{X_n})$, the notation $F_{\mathbf{X}}$ refers to the joint distribution function of \mathbf{X} . The following bounds

$$\max \left\{ \sum_i F_{X_i}(x_i) - n + 1, 0 \right\} \leq F_{\mathbf{X}}(\mathbf{x}) \leq \min_i \{F_{X_i}(x_i)\} \quad (1.1)$$

always hold. In inequality (1.1), the leftmost (rightmost) part is known as the Fréchet-Hoeffding lower (upper) bound of $F_{\mathbf{X}}$ and is denoted by F_L (F_U , respectively). It has been shown that

$$F_U(\mathbf{x}) = \min_i \{F_{X_i}(x_i)\} \quad (1.2)$$

is a joint distribution function. This is the joint distribution function of a comonotonic random vector in $R_n(F_{X_1}, F_{X_2}, \dots, F_{X_n})$ (cf., e.g., Joe, 2001). Unfortunately, this is not always the case with the lower bound (1.1). In general,

$$F_L(\mathbf{x}) = \max \left\{ \sum_i F_{X_i}(x_i) - n + 1, 0 \right\} \quad (1.3)$$

is not always distribution function. The bivariate case, however, is an exception, where in this case the lower bound F_L is a joint distribution function of what is called a counter-monotonic random vector. Analogous to the definition of bivariate comonotonic random vectors, two dimensional random vectors are counter-monotonic if and only if their supports are counter-monotonic sets: a set of two dimensional vectors are counter-monotonic if and only if $(x_1 - y_1)(x_2 - y_2) \leq 0$ for all $\mathbf{x} = (x_1, x_2)$

and $\mathbf{y} = (y_1, y_2)$ in this set. Consequently, any pair of random variables (X, Y) is counter-monotonic if and only if the pair $(X, -Y)$ is co-monotonic. Similar to the case of co-monotonic random vectors, counter-monotonic random vectors minimize the expectation of supermodular functions over the class of all random vectors with identical marginals (Puccetti and Wang, 2015).

Generalizations of counter-monotonicity to higher dimensional random vectors, unfortunately, are not unique. This, by the way, is one of the reason for our concentration on the bivariate case in this thesis. It is not like the co-monotonicity that can be generalized simultaneously. Counter-monotonicity in a sense is like ordering the components of vectors in opposite directions. If the vector has two components, then we can easily do that, but by adding more components the order will be problematic. Furthermore, when some pairs have opposite directional orderings, other pairs can ruin this ordering. Intuitively for $n = 3$, if two pairs of random variables are counter-monotonic, then the remaining third may not be counter-monotonic; it could even be co-monotonic. This is getting more complex, of course, when n is higher than three. We therefore relax the notion of counter-monotonicity for higher dimensional vectors by only considering all pairs of its components. We call this notion pairwise counter-monotonicity: a random vector is pairwise counter-monotonic if and only if every pair of its two components are counter-monotonic. This, however, becomes a very special case due to the fact that the joint distribution of pairwise counter-monotonic random vector is Fréchet-Hoeffding lower bound (1.1). General discussions on counter-monotonicity can be found in Puccetti and Scarsini (2010).

In the sequel, we denote any co-monotonic random vector by $\mathbf{X}^c = (X_1^c, \dots, X_n^c)$. The c -superscript is used to indicate co-monotonicity. If \mathbf{X}^c is co-monotonic, then

there is a uniform random variable $U[0, 1]$ such that

$$\mathbf{X}^c = (F_{X_1}^{-1}(U), \dots, F_{X_n}^{-1}(U)), \quad (1.4)$$

where $F_{X_i}^{-1}$ denotes the left-inverse of the distribution function F_{X_i} . The uniform random variable U here is the external mechanism noted in Definition 1.1.1. Analogously, for $n = 2$, the counter-monotonic random vector can be expressed in a pair $(F_{X_1}^{-1}(U), F_{X_2}^{-1}(1 - U))$.

1.2 Convex bound of aggregate risks

In the area of insurance and risk management, researchers are often interested in the distribution of sums of random variables. These sums may represent aggregate claims of insurance policies, portfolios, total risks, present values of payments, and so on. We refer to Dhaene et al. (2002a,b), and Deelstra et al. (2011) for examples in insurance and finance. These variables are not necessarily independent, especially in the insurance context where the dependence is observed. The research is then directed to study dependence structures of random variables for aggregate claims.

Let X_i be the i -th individual insured risk where, together with other individual risks, it builds a risk portfolio $\mathbf{X} = (X_1, X_2, \dots, X_n)$. Researchers want to know the distribution of the sum $S = X_1 + X_2 + \dots + X_n$.

Note 1.2.1 The case when the number “N” of risks is random is very important but it can be reduced to the deterministic “n” using a well-known conditioning argument; see, e.g., Klugman et al. (2012).

The distribution of S is not easy to investigate when the random variables are not

independent. A good idea might be to replace the sum of dependent random variables by other random variables with a simpler dependent structure, called co-monotonic random vectors. Indeed, with the help of co-monotonicity we can always have the bound

$$S \leq_{cx} X_1^c + X_2^c + \cdots + X_n^c, \quad (1.5)$$

where it is in the sense of convex order (cx):

A random variable X precedes a random variable Y in convex order, denoted by $X \leq_{cx} Y$, if and only if $\mathbb{E}[X] = \mathbb{E}[Y]$ and $\mathbb{E}[(X - d)_+] \leq \mathbb{E}[(Y - d)_+]$ for every $d \in \mathbb{R}$.

The subscript ‘+’ in the second condition means $(x)_+ = x$ if $x \geq 0$ and $(x)_+ = 0$ otherwise.

In short, the sum of the components of random vector $\mathbf{X} \in R_n(F_{X_1}, F_{X_2}, \dots, F_{X_n})$ is maximal in the convex order if and only if \mathbf{X} is co-monotonic. Bound (1.5) is very attractive in the area of finance and actuarial science. It gives an insight about the distribution of aggregate risks in general, especially when the marginal distributions are known but the joint distribution is unknown or complicated.

The convex order used in bound (1.5) also suggests that co-monotonic random vectors are less favorable than those that are not co-monotonic from the perspective of risk-averse decision makers. Practical examples of this problem include life annuities, insurance portfolios, present value functions, and so on (e.g., Deelstra et al., 2011; Dhaene et al., 2002b; Kaas et al., 2000).

The convex bound for the sum of random variables is one of many examples of the use of co-monotonicity concept in actuarial science and finance. There are many other applications such as risk sharing, risk measures, optimal allocation strategies,

capital allocations, life insurance and pensions, and many others. Concise reviews of applications of co-monotonicity in actuarial science and finance can be found, e.g., in Dhaene et al. (2002a), and Deelstra et al. (2011).

1.3 Dependence structure via copula

The dependence structures in random vectors can be explored using copulas. In general, copulas are n -dimensional continuous distribution functions on $[0, 1]^n$ with uniform on $[0, 1]$ marginals. By Sklar's Theorem, copulas link univariate distribution functions with multivariate distribution functions. Numerous application of copulas have emerged in the literature and we refer to books by Nelsen (2006) and Joe (2001, 2014) for more details. Copulas have also been very successfully applied in engineering (e.g., Reddy and Ganguli, 2012; Warsido and Bitsuamlak, 2015), finance (e.g., Li, 2000; Genest et al., 2009; Sun et al., 2008), and many others.

Throughout this section we limit our attention only to two-dimensional copulas. This is in accordance with our chosen direction in this thesis to concentrate on pairwise dependencies. The definition of two-dimensional copulas can be expressed axiomatically as follows.

Definition 1.3.1 *A bivariate copula C is a non-decreasing and right-continuous function that maps the unit square $[0, 1]^2$ to the unit interval $[0, 1]$ and satisfies the following conditions:*

- (i) $\lim_{u_i \downarrow 0} C(u_1, u_2) = 0$ for $i = 1, 2$;
- (ii) $\lim_{u_1 \uparrow 1} C(u_1, u_2) = u_2$ and $\lim_{u_2 \uparrow 1} C(u_1, u_2) = u_1$;

(iii) C is supermodular, that is, the inequality

$$C(v_1, v_2) - C(u_1, v_2) - C(v_1, u_2) + C(u_1, u_2) \geq 0$$

holds for all $u_1 \leq v_1, u_2 \leq v_2$.

The connection between copulas and bivariate distributions is given by the following theorem, which is due to Sklar (1959).

Theorem 1.3.1 *Let $F_{\mathbf{X}} \in R_2(F_{X_1}, F_{X_2})$ have continuous marginals F_{X_1} and F_{X_2} . Then there exists a unique copula C such that, for all $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$,*

$$F_{\mathbf{X}}(\mathbf{x}) = C(F_{X_1}(x_1), F_{X_2}(x_2)). \quad (1.6)$$

Conversely, if C is a copula and F_{X_1} and F_{X_2} are distribution functions, then the function $F_{\mathbf{X}}$ defined by the above equation is a bivariate distribution function with margins F_{X_1} and F_{X_2} .

Theorem 1.3.1 asserts that copulas bridge marginal distribution functions with joint distribution functions. Hence, given marginals and copulas, we obtain joint distribution functions. Conversely, given joint distribution functions we derive copulas. This relation simplifies the investigation of dependence structures of bivariate random variables. For example, in Section 1.1 we have Fréchet-Hoeffding upper bound with the joint distribution function given by $F_{\mathbf{X}}(x_1, x_2) = \min\{F_{X_1}(x_1), F_{X_2}(x_2)\}$. Thus, a copula for this set of bivariate random variables is

$$C_U(u, v) = \min\{u, v\}.$$

Analogously, the copula for the Fréchet-Hoeffding lower bound and independent bivariate random vectors are

$$C_L(u, v) = \max\{u + v - 1, 0\}$$

and

$$C_I(u, v) = uv,$$

respectively. Adapted from the Fréchet-Hoeffding bound, we have the lower and upper bounds for copulas C as follows

$$\max\{u + v - 1, 0\} \leq C(u, v) \leq \min\{u, v\}. \quad (1.7)$$

We conclude this section by giving formulas of classical dependence measures in the form of copulas. As we know, for every bivariate random vector we have Pearson's, Spearman's, Kendall's correlation coefficients as its dependence measures. Theorem 1.3.1 leads us to the fact that given a joint distribution function $F_{\mathbf{X}}$ with marginals F_{X_1} and F_{X_2} , we always have the bivariate copula

$$C(u, v) = F_{\mathbf{X}}(F_{X_1}^{-1}(u), F_{X_2}^{-1}(v)), \quad (1.8)$$

where $F_{X_1}^{-1}$ and $F_{X_2}^{-1}$ are the quantile functions of X_1 and X_2 respectively. Consequently, for example, the Kendall's correlation coefficient r_K can be expressed as

$$r_K(\mathbf{X}) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1. \quad (1.9)$$

Analogously, the Spearman's r_S and Pearson's r_P can be derived from the copula C :

$$r_S(\mathbf{X}) = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3 \quad (1.10)$$

and

$$r_P(\mathbf{X}) = \frac{1}{\sqrt{\text{Var}[X_1]\text{Var}[X_2]}} \int_0^1 \int_0^1 (C(u, v) - uv) dF_{X_1}^{-1}(u) dF_{X_2}^{-1}(v). \quad (1.11)$$

1.4 Problem statement

In general, dependence analysis progresses in at least two directions. One group of researchers work on quantifying dependencies, developing dependence measures, and analysing measurements. Examples of such research include classical dependence measures such as Pearson's, Spearman's, and Kendall's correlation coefficients, multivariate dependence measures, and so on (e.g., Diers et al., 2012; Scarsini, 1984; Schweizer and Wolff, 1981).

Another group of researchers work on implications of dependence structures on complex systems. The aforementioned convex order of the sum of random variables is an example of research in this category (e.g., Dhaene et al., 2002a; Lehmann, 1966; Schweizer and Wolff, 1981). Other examples include civil engineering (e.g., Warsido and Bitsuamlak, 2015), finance (e.g., Genest et al., 2009; Li, 2000; Sun et al., 2008), and many others.

It has been shown that the concept of co-monotonicity plays significant roles in the areas of statistics, finance, and actuarial science. The co-monotonic form can be used as a substitute of unknown, or at least hard to derive, dependence structures (see the convex order problem in Section 1.1). If for every random vector \mathbf{X} in

$R_n(F_{X_1}, F_{X_2}, \dots, F_{X_n})$ the sum of its components can be replaced by the sum of components of co-monotonic random vector, then is there any restriction such that this replacement ruins the accuracy? There are some cases, of course, when random vectors depart significantly from their co-monotonic forms. In such cases replacing random vectors with their co-monotonic forms can be problematic. However, there are practical applications such as annuities considered by Kaas et al. (2000) and Dhaene et al. (2002a,b) that do not have the aforementioned issue because, naturally, the series of present values are co-monotonic, or close to co-monotonic. Motivated by such problems, it is now natural to raise the topic of quantification of the distance of random vectors from their co-monotonic forms, which brings us to the main topic of this thesis.

1.5 Scope and summary of the thesis

As we have already noted earlier, we deliberately limit our research only to the analysis of bivariate random variables. In this case, relations between random variables can be explained by curves. We elucidate dependence structures using monotonicity (or lack of it) of such curves, which we investigate with the help of monotone rearrangements. The three and higher dimensional cases can, with much more complex mathematics, be investigated as well, which is our future work. Next, a summary of the remaining chapters follows.

In Chapter 2, we discuss indices for measuring the lack of monotonicity in functions. These indices are built on the fact that non-monotonic patterns manifest themselves when there are discrepancies of the functions from their monotone rearrangements. These discrepancies are utilized to measure distances of functions from

their monotonic patterns. Properties of such measures are discussed in this chapter. In Chapter 3, computational algorithms, examples, and illustrations are provided.

In Chapter 4, we use the monotonicity indices introduced in Chapter 2 to compare associations between study subjects in education. Perfect associations can be represented by co-monotonic random vectors while the monotonicity indices can be used to quantify discrepancies between random vectors and their co-monotonic forms. In this case, the indices are used to measure the lack of co-monotonicity. For example, given our illustrative example of student marks on various subjects, we succeed in extracting information on how the subjects are related. Among the necessary technical details, we discuss and utilize curve fitting as well as convenient ways for calculating the indices.

Chapter 5 provides concluding remarks and a number of ideas for future research.

Chapter 2

Measuring the lack of monotonicity in functions

2.1 Introduction

In a number of problems such as assessing co-monotonicity, developing statistical tests, dealing with demand and production functions in economics, modeling mortality and longevity of populations, researchers often face the need to know whether certain functions are monotonic (e.g., non-decreasing) or not, and if not, then they wish to assess their degree of non-monotonicity. Due to this reason, in this chapter we suggest and explore several indices for measuring the lack of non-decreasingness in functions.

While determining monotonicity can be a standard, though perhaps quite difficult, exercise of checking the sign of the first derivative over the region of interest, assessing

¹This chapter in a condensed form has been published in Qoyyimi, D. T. and Zitikis, R. (2014). Measuring the lack of monotonicity in functions, *Mathematical Scientist* 39(2): 107–117.

the lack of monotonicity in non-monotonic functions has gotten much less attention in the literature (e.g., Davydov and Zitikis, 2005; Yitzhaki and Schechtman, 2012). To illustrate problems where monotonicity, or lack of it, matters, we next present four specific illustrative applications.

Application 2.1.1 Monotone likelihood ratio (MLR) families play important roles in areas of statistics such as constructing uniformly powerful hypothesis tests, confidence bounds and regions. In short, a family of absolutely continuous cumulative distribution functions (cdf's) $\{F_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ is MLR if for every $\theta_1 < \theta_2$, the two cdf's F_{θ_1} and F_{θ_2} are distinct and the ratio $f_{\theta_1}(\mathbf{x})/f_{\theta_2}(\mathbf{x})$ of the corresponding densities is an increasing function of a statistic $T(\mathbf{x}) \in \mathbb{R}$, where $\mathbf{x} = (x_1, \dots, x_n)$ is a generic n -dimensional observation. For more details on the MLR families and their uses in statistics, we refer to, e.g., Chapter 4 of Bickel and Doksum (2001).

Application 2.1.2 The presence of a deductible $d \geq 0$ often changes the profile of insurance losses (e.g. Brazauskas et al., 2009, 2015). Because of this and other reasons, given two losses X and Y , which may not be observable, decision makers wish to determine whether the observable losses $X_d = [X \mid X > d]$ and $Y_d = [Y \mid Y > d]$ are stochastically (ST) ordered, say $X_d \leq_{\text{ST}} Y_d$ for every $d \geq 0$. Denuit et al. (2005) show on p. 124 that this ordering is equivalent to determining whether the ratio $S_Y(x)/S_X(x)$ is a non-decreasing function in x , where S_X and S_Y are the survival functions of X and Y , respectively. We conclude this example by noting that this ordering is known in the literature (e.g., Denuit et al., 2005; Marshall et al., 2011) as the hazard rate (HR) ordering, and is succinctly denoted by $X \leq_{\text{HR}} Y$.

Application 2.1.3 More generally than in the previous example, one may wish to determine whether for every deductible $d \geq 0$ and every policy limit $L > d$, the

observable insurance losses $X_{d,L} = [X \mid d \leq X \leq L]$ and $Y_{d,L} = [Y \mid d \leq Y \leq L]$ are stochastically ordered, say, $X_{d,L} \leq_{ST} Y_{d,L}$. We find on pp. 127–128 in Denuit et al. (2005) that this problem is equivalent to determining whether the ratio $f_Y(x)/f_X(x)$ is a non-decreasing function in x over the union of the supports of X and Y , where f_X and f_Y are the density functions of X and Y , respectively. This ordering is known in the literature (e.g., Denuit et al., 2005; Marshall et al., 2011) as the likelihood ratio (LR) ordering and is succinctly denoted by $X \leq_{LR} Y$. For further details on various stochastic orderings and their manifold applications, we refer to Levy (2006), Li and Li (2013), and Shaked and Shantikumar (2007).

Application 2.1.4 Let \mathcal{X}_+ denote the set of all non-negative random variables X representing insurance losses. The premium calculation principle (pcp) is a functional $\pi : \mathcal{X}_+ \rightarrow [0, \infty]$. Furman and Zitikis (2008a, 2009) have specialized this general premium to the weighted pcp π_w defined by the equation

$$\pi_w[X] = \frac{\mathbb{E}[Xw(X)]}{\mathbb{E}[w(X)]}, \tag{2.1}$$

where $w : [0, \infty) \rightarrow [0, \infty)$ is a weight function specified by the decision maker, or implied by certain axioms. The functional

$$\pi_w : \mathcal{X}_+ \rightarrow [0, \infty]$$

satisfies the non-negative loading property whenever the weight function w is non-decreasing (cf. Lehmann, 1966). This is one of the very basic properties that insurance premiums need to satisfy. For further information on this topic, we refer to Sendov et al. (2011). For a concise overview of pcp's, we refer to, e.g., Young (2004). For

detailed results and their proofs, we refer to, e.g., Denuit et al. (2005).

We next briefly present a few more topics and related references where monotonicity, or lack of it, of certain functions plays an important role:

- Growth curves (e.g., Bebbington et al., 2009; Chernozhukov et al., 2009; Panik, 2014).
- Mortality curves (e.g., Bebbington et al., 2011; Gavrilov and Gavrilova, 1991).
- Positive regression dependence and risk sharing (e.g., Barlow and Proschan, 1974; Bebbington et al., 2007; Dana and Scarsini, 2007; Lehmann, 1966).
- Portfolio construction, capital allocations, and co-monotonicity (e.g., Dhaene et al., 2002a,b, 2006; Furman and Zitikis, 2008b).
- Decision theory and stochastic ordering (e.g., Denuit et al., 2005; Egozcue et al., 2013; Levy, 2006; Shaked and Shantikumar, 2007).
- Engineering reliability and risks (e.g., Barlow and Proschan, 1974; Bebbington et al., 2008; Lai and Xie, 2006; Li and Li, 2013; Singpurwalla, 2006).

One unifying feature of these diverse works is that they impose monotonicity requirements on certain functions, which are generally unknown, and thus researchers seek for statistical models and data for determining their shapes. To illustrate the point, we recall, for example, the work of Bebbington et al. (2011) who specifically set out to determine whether mortality continues to increase or starts to decelerate after a certain species related late-life age. This is known in the literature as the late-life mortality deceleration phenomenon. Hence, we can rephrase the phenomenon as a question: is the mortality function always increasing? Naturally, we do not elaborate

on this topic any further in this chapter, referring the interested reader to Bebbington et al. (2011, 2014), and references therein.

To verify the monotonicity of functions such as those noted in the above examples, researchers quite often assume that the functions belong to some parametric or semiparametric families. One may not, however, be comfortable with this element of subjectivity and thus prefers to rely solely on data to make a judgement. Under these circumstances, verifying monotonicity becomes a non-parametric problem, whose solution asks for an index that, for example, takes on the value 0 when the function under consideration is non-decreasing and on positive values otherwise. In the following sections we shall introduce and discuss two such indices; both of them are useful, but due to different reasons.

2.2 An index of non-decreasingness

Perhaps the most obvious definition of an index of non-decreasingness is based on the notion of non-decreasing rearrangement that was introduced by Hardy et al. (1934). For instance, non-decreasing rearrangement of function $h : [0, 1] \rightarrow \mathbb{R}$, is defined by

$$I_h(t) = \inf\{x \in \mathbb{R} : G_h(x) \geq t\} \quad \text{for all } t \in [0, 1],$$

where

$$G_h(x) = \lambda\{s \in [0, 1] : h(s) \leq x\} \quad \text{for all } x \in \mathbb{R},$$

with λ denoting the Lebesgue measure. Hence, any distance between the original function h and its non-decreasing rearrangement I_h can serve an index of non-decreasingness. Of course, there are many distances in function spaces, and thus

many indices, but we shall concentrate here on the L_1 -distance due to its attractive geometric interpretation and other properties.

Note 2.2.1 In dependence studies, the non-decreasing rearrangement of a function h can be interpreted as a co-monotonic form of h viewed as a relationship between two random variables. In this case, any distance of the function h from its non-decreasing rearrangement can be interpreted as the distance of a two-dimensional random vector from its co-monotonic form. Studies of dependence structures in the forms of rearrangements can be found, e.g., in Puccetti and Wang (2015) and Rüschendorf (1983).

Throughout this chapter, we assume that h is integrable on its domain of definition. The following proposition will be useful for next discussion.

Proposition 2.2.1 *Function $h : [0, 1] \rightarrow \mathbb{R}$ is non-decreasing if and only if the equation $I_h(t) = h(t)$ holds for λ -almost all $t \in [0, 1]$. If h is left-continuous, then it is non-decreasing if and only if $I_h(t) = h(t)$ for all $t \in [0, 1]$.*

Proof Assume first that $I_h(t) = h(t)$ for λ -almost all $t \in [0, 1]$. Since the function I_h is non-decreasing, then the function h must be non-decreasing as well.

Conversely, suppose that the function h is non-decreasing. Then from the definition of $G_h(x)$, we have the equation

$$G_h(x) = \sup\{t \in [0, 1] : h(t) \leq x\} \tag{2.2}$$

and thus, in turn, from the definition of $I_h(t)$, we have the equation $I_h(t) = \lim_{s \uparrow t} h(s)$. Consequently, I_h is left-continuous and the equation $I_h(t) = h(t)$ holds at every continuity point $t \in [0, 1]$ of the function h . Since the set of all discontinuity points

of every non-decreasing function can only be at most of λ -measure zero, the converse of Proposition 2.2.1 follows. This finishes the entire proof of Proposition 2.2.1. ■

The function I_h is known in the literature as the generalized inverse of the function G_h , and is thus frequently denoted by G_h^{-1} . Throughout this discussion, however, we prefer using the notation I_h to emphasize the fact that this is a weakly increasing, that is, non-decreasing function. In probability and statistics, researchers would call I_h the quantile function of the ‘random variable’ h . In the literature on function theory and functional analysis (e.g., Chong and Rice, 1971; Day, 1970; Ghossoub, 2015; Korenovskii, 2007, and references therein) the function I_h is usually called the non-decreasing equimeasurable rearrangement of h .

We are now in the position to give a rigorous definition of the earlier noted L_1 -based index of non-decreasingness, which is

$$\mathcal{I}_h = \int_0^1 |I_h(t) - h(t)| dt.$$

The index \mathcal{I}_h takes on the value 0 if and only if the function h is non-decreasing. The proof of this fact is based on the well-known property (cf., e.g., Proposition 2.2.1) that h is non-decreasing if and only if the equation $I_h(t) = h(t)$ holds for λ -almost all $t \in [0, 1]$.

It is instructive to mention here that the notion of monotone rearrangement has been very successfully used in a number of areas:

- Efficient insurance contracts (e.g., Carlier and Dana, 2005; Dana and Scarsini, 2007).
- Rank-dependent utility theory (e.g. Carlier and Dana, 2003, 2008, 2011; Quiggin, 1982, 1993).

- Continuous-time portfolio selection (e.g., He and Zhou, 2011; Jin and Zhou, 2008).
- Statistical applications such as performance improvement of estimators (e.g., Chernozhukov et al., 2009, 2010) and optimization problems (e.g., Rüschendorf, 1983).
- Stochastic processes and probability theory (e.g., Egorov, 1990; Thilly, 1999; Zhukova, 1994, 1998).

These are just a few illustrative topics and references, but they lead us into the vast literature on monotone rearrangements and their manifold uses.

We conclude this section with a few additional properties of the index \mathcal{I}_h which will lead us naturally to the next section. First, as one would intuitively expect, any index of non-decreasingness should not change if the function $h : [0, 1] \rightarrow \mathbb{R}$ is lifted up or down by any constant $d \in \mathbb{R}$. This is indeed the case, as the equation

$$\mathcal{I}_{h+d} = \mathcal{I}_h \tag{2.3}$$

follows easily upon checking that, for every constant $d \in \mathbb{R}$, the equation

$$I_{h+d}(t) = I_h(t) + d$$

holds for every $t \in [0, 1]$. Finally, the multiplication of the function h by any non-negative constant $c \geq 0$ (so as not to change the direction of monotonicity) should only change the index by as much as it changes the slope of the function. Indeed, we have the equation

$$\mathcal{I}_{ch} = c\mathcal{I}_h \tag{2.4}$$

that follows easily upon checking that, for every constant $c \geq 0$, the equation

$$I_{ch}(t) = cI_h(t)$$

holds for every $t \in [0, 1]$.

2.3 Probabilistic interpretation

The following probabilistic interpretation of the basic quantities involved in our research will play a pivotal role, especially when devising simple proofs of a number of results. We note at the outset that the interpretation is well known and appears frequently in the literature (e.g., Carlier and Dana, 2005; Denneberg, 1994).

The interval $[0, 1]$ can be viewed as a sample space, usually denoted by Ω in probability and statistics. Furthermore, the Lebesgue measure λ can be viewed as a probability measure, usually denoted by \mathbb{P} , which is defined on the σ -algebra of all Borel subsets of $\Omega = [0, 1]$. Hence, the function $h : [0, 1] \rightarrow \mathbb{R}$ can be viewed as a random variable, usually denoted by $X : \Omega \rightarrow \mathbb{R}$ in probability and statistics. Under these notational agreements, the function G_h can be viewed as the cdf F_X of X , and, in turn, the function I_h can be viewed as the quantile function F_X^{-1} of X .

To illustrate how this probabilistic point of view works, we recall the well-known equation

$$\int_0^1 I_h(t) dt = \int_0^1 h(t) dt, \quad (2.5)$$

which we shall later use in proofs. The validity of equation (2.5) can easily be estab-

lished as follows. We start with the equation

$$\int_0^1 I_h(t)dt = \int_0^1 F_X^{-1}(t)dt. \tag{2.6}$$

Then we recall that the mean $\mathbb{E}[X]$ of X can be written as $\int_0^1 F_X^{-1}(t)dt$. Hence,

$$\int_0^1 I_h(t)dt = \mathbb{E}[X].$$

Furthermore, appealing to the probabilistic interpretation one more time, we have

$$\mathbb{E}[X] = \int_0^1 h(t)dt,$$

which establishes equation (2.5). Of course, from the purely mathematical point of view, equation (2.5) follows from the fact that h and I_h are equimeasurable functions and thus their integrals coincide. In summary, we have demonstrated that equation (2.5) holds.

2.4 Co-monotonically additive index

It is instructive to view equation (2.3) as the additivity property

$$\mathcal{I}_{h+g_0} = \mathcal{I}_h + \mathcal{I}_{g_0}, \tag{2.7}$$

where g_0 is the constant function defined by $g_0(t) = d$ for all $t \in [0, 1]$, with $d \in \mathbb{R}$ being a constant. Indeed, $\mathcal{I}_{g_0} = 0$, and thus we conclude that equations (2.3) and (2.7) are equivalent.

Note that the functions h and g_0 are co-monotonic irrespective of the value of d .

This fact follows immediately from the definition of co-monotonicity as discussed in Section 1.1. This is a well-known notion, extensively utilized in many areas, perhaps most notably in economics and insurance. For further details and references on the topic, we refer to Denneberg (1994), Dhaene et al. (2002a,b, 2006), and references therein.

Coming now back to equation (2.7), a natural question is whether the equation still holds if the constant function g_0 is replaced by any other function g that is co-monotonic with h . For this, we first recall the fact (cf. Corollary 4.6 in Denneberg, 1994) that, for every pair of co-monotonic functions h and g ,

$$I_{h+g}(t) = I_h(t) + I_g(t) \quad \text{for every } t \in [0, 1]. \quad (2.8)$$

Unfortunately, the index \mathcal{I}_h is based on the non-linear functional

$$\Delta \mapsto \int_0^1 |\Delta(t)| dt, \quad (2.9)$$

and we can thus at most have the subadditivity property:

$$\mathcal{I}_{h+g} \leq \mathcal{I}_h + \mathcal{I}_g. \quad (2.10)$$

The lack of additivity would, of course, still be the case even if we replaced the L_1 -type functional by any other L_p -type functional. Hence, we need a linear functional.

Note that by simply dropping the absolute values from functional (2.9) would not lead us to the desired outcome because the new ‘index’ would be identically equal to 0 as seen from equation (2.5). Remarkably, there is an easy way to linearize functional (2.9). This is achieved by dropping the absolute values and, very importantly,

weighting dt with the function t . These two steps lead us to the functional

$$\Delta \mapsto \int_0^1 \Delta(t)tdt$$

and thus, in turn, to the quantity

$$\mathcal{L}_h = \int_0^1 (I_h(t) - h(t))tdt, \tag{2.11}$$

but before declaring it an index of non-decreasingness, we need to verify that \mathcal{L}_h is always non-negative and takes on the value 0 if and only if the function h is non-decreasing. These are non-trivial tasks, whose solutions make up our next Theorem 2.5.1. Before formulating the theorem, we next present two illustrative examples where \mathcal{I}_h and \mathcal{L}_h are calculated and compared.

Example 2.4.1 For a fixed parameter $\alpha \in [0, 1]$, let h_α be the function on $[0, 1]$ defined by

$$h_\alpha(t) = \begin{cases} t & \text{for } t \in [0, 0.5], \\ \alpha t + (1 - \alpha)(1 - t) & \text{for } t \in (0.5, 1]. \end{cases}$$

Note that h_α is non-decreasing when $\alpha \in [0.5, 1]$, and thus $\mathcal{I}_{h_\alpha} = 0$ and $\mathcal{L}_{h_\alpha} = 0$. In this case, we are left to work only with $\alpha \in [0, 0.5)$. Figure 2.1 illustrates the function h_α when $\alpha = 0, 0.25, 0.5, 0.75$, and $\alpha = 1$.

For every α , we derive the function G_{h_α} as follows:

$$G_{h_\alpha}(x) = \begin{cases} x & \text{when } x \in [0, \alpha], \\ \frac{2 - 2\alpha}{1 - 2\alpha}x + \frac{\alpha}{2\alpha - 1} & \text{when } x \in (\alpha, 0.5], \\ 1 & \text{when } x > 0.5. \end{cases}$$

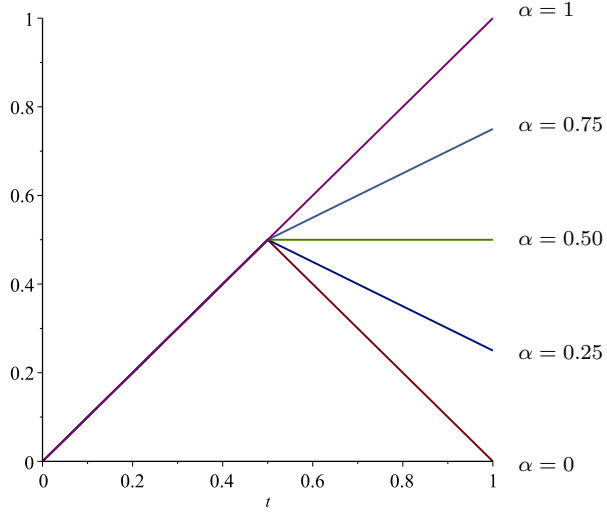


Figure 2.1: Function $h_\alpha(t)$ for various α

The non-decreasing rearrangement of h_α can then be expressed as

$$I_{h_\alpha}(t) = \begin{cases} t & \text{when } t \in [0, \alpha), \\ \frac{(1 - 2\alpha)t + \alpha}{2 - 2\alpha} & \text{when } t \in [\alpha, 1]. \end{cases}$$

Utilizing the easily checked fact that the functions h_α and I_{h_α} cross at the only point $t_c = (\alpha - 2)/(2\alpha - 3)$, we calculate the index \mathcal{I}_{h_α} as follows:

$$\begin{aligned} \mathcal{I}_{h_\alpha} &= \int_\alpha^{1/2} |I_{h_\alpha}(t) - h_\alpha(t)| dt + \int_{1/2}^{t_c} |I_{h_\alpha}(t) - h_\alpha(t)| dt + \int_{t_c}^1 |I_{h_\alpha}(t) - h_\alpha(t)| dt \\ &= \int_\alpha^{1/2} \left(t - \frac{(1 - 2\alpha)t + \alpha}{2 - 2\alpha} \right) dt + \int_{1/2}^{t_c} \left(\alpha t + (1 - \alpha)(1 - t) - \frac{(1 - 2\alpha)t + \alpha}{2 - 2\alpha} \right) dt \\ &\quad + \int_{t_c}^1 \left(\frac{(1 - 2\alpha)t + \alpha}{2 - 2\alpha} - (\alpha t + (1 - \alpha)(1 - t)) \right) dt \\ &= \frac{(1 - 2\alpha)(1 - \alpha)}{2(3 - 2\alpha)}. \end{aligned}$$

Similar arguments produce a formula for the index \mathcal{L}_{h_α} :

$$\begin{aligned} \mathcal{L}_{h_\alpha} &= \int_0^1 (I_{h_\alpha}(t) - h_\alpha(t))t dt \\ &= \int_\alpha^{1/2} \left(\frac{(1-2\alpha)t + \alpha}{2-2\alpha} - t \right) t dt + \int_{1/2}^1 \left(\frac{(1-2\alpha)t + \alpha}{2-2\alpha} - (\alpha t + (1-\alpha)(1-t)) \right) t dt \\ &= \frac{(1-2\alpha)(1-\alpha)}{24}. \end{aligned}$$

These indices as functions of α are depicted in Figure 2.2.

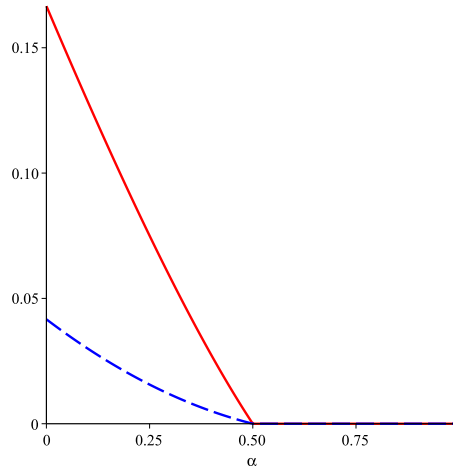


Figure 2.2: The indices \mathcal{I}_h (solid-red) and \mathcal{L}_h (dashed-blue) as functions of α .

Example 2.4.2 For a fixed parameter $\alpha \in [0, 1]$, let g_α be the function on $[0, 1]$ defined by

$$g_\alpha(t) = \begin{cases} \frac{t}{\alpha} & \text{for } t \in [0, \alpha], \\ \frac{t-1}{\alpha-1} & \text{for } t \in (\alpha, 1] \end{cases}$$

when $\alpha \in (0, 1)$,

$$g_\alpha(t) = 1 - t \quad \text{when } \alpha = 0,$$

and

$$g_\alpha(t) = t \quad \text{when} \quad \alpha = 1.$$

Figure 2.3 illustrates the function g_α when $\alpha = 0.25$. For every α , the non-decreasing

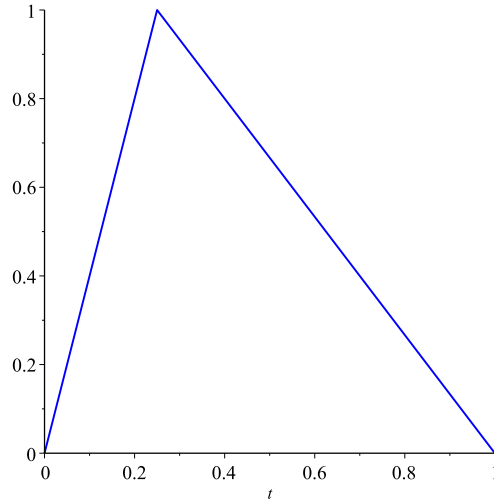


Figure 2.3: Function $g_{0.25}(t)$.

rearrangement of g_α is the identity function on $[0, 1]$, that is, $I_{g_\alpha}(t) = t$ for $t \in [0, 1]$.

The non-decreasing indices can then be calculated as follows:

$$\begin{aligned} \mathcal{I}_{g_\alpha} &= \int_0^\alpha \left(\frac{t}{\alpha} - t \right) dt + \int_\alpha^{1/(2-\alpha)} \left(\frac{t-1}{\alpha-1} - t \right) dt + \int_{1/(2-\alpha)}^1 \left(t - \frac{t-1}{\alpha-1} \right) dt \\ &= \frac{1-\alpha}{2-\alpha} \end{aligned}$$

and

$$\begin{aligned} \mathcal{L}_{g_\alpha} &= \int_0^\alpha t \left(t - \frac{t}{\alpha} \right) dt + \int_\alpha^1 t \left(t - \frac{t-1}{\alpha-1} \right) dt \\ &= \frac{1}{6}(1-\alpha). \end{aligned}$$

These indices as functions of α are depicted in Figure 2.4.

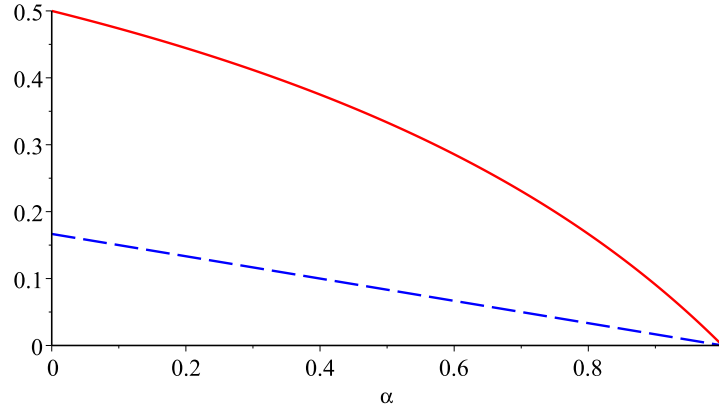


Figure 2.4: The indices $\mathcal{I}_{g\alpha}$ (solid-red) and $\mathcal{L}_{g\alpha}$ (dashed-blue) as functions of α .

2.5 Main theorem

We start this section with a few properties of \mathcal{L}_h that will be seen from the proof of the main theorem below. First, when h and g are co-monotonic, then

$$\mathcal{L}_{h+g} = \mathcal{L}_h + \mathcal{L}_g, \quad (2.12)$$

which follows from equation (2.8) and the linearity of the functional

$$\Delta \mapsto \int_0^1 \Delta(t) t dt.$$

In particular, we have $\mathcal{L}_{h+d} = \mathcal{L}_h$ for every function h and every constant $d \in \mathbb{R}$, because $\mathcal{L}_d = 0$. Next, for every non-negative constant $c \geq 0$, we have the equation

$$\mathcal{L}_{ch} = c\mathcal{L}_h, \quad (2.13)$$

which follows immediately from $I_{ch}(t) = cI_h(t)$ and the definition of \mathcal{L}_h . Furthermore, from the definitions of \mathcal{I}_h and \mathcal{L}_h we immediately obtain the bound

$$\mathcal{L}_h \leq \mathcal{I}_h, \tag{2.14}$$

which, incidentally, explains the ordering of the two curves in Figure 2.2.

Theorem 2.5.1 *For every function $h : [0, 1] \rightarrow \mathbb{R}$, the index \mathcal{L}_h is non-negative and takes on the value 0 if and only if the function h is non-decreasing.*

Proof The proof is somewhat complex, and we have thus subdivided it into three parts: First, we establish an alternative representation (equation (2.15) below) for \mathcal{L}_h on which the rest of the proof relies, and which, incidentally, clarifies how we came up with the weight t in definition (2.11). Then, in the second part, which is the longest and most complex part of the proof, we establish a certain ordering result (bound (2.16) below) that implies the non-negativity of \mathcal{L}_h . Finally, in the third part we prove that $\mathcal{L}_h = 0$ if and only if the function h is non-decreasing.

Part 1: Here we express \mathcal{L}_h by an alternative formula that plays a pivotal role in our subsequent considerations. For this, we first recall that, by definition, the indicator $\mathbf{1}\{S\}$ of statement S takes on the value 1 if the statement S is true and on the value 0 otherwise. With this notation, and also using Fubini's theorem, we have the equations:

$$\begin{aligned} \mathcal{L}_h &= \int_0^1 (I_h(t) - h(t)) \int_0^t ds dt \\ &= \int_0^1 \int_0^1 (I_h(t) - h(t)) \mathbf{1}\{s \leq t\} ds dt \end{aligned}$$

$$\begin{aligned}
\mathcal{L}_h &= \int_0^1 \left(\int_s^1 I_h(t) dt - \int_s^1 h(t) dt \right) ds \\
&= \int_0^1 \left(\int_0^s h(t) dt - \int_0^s I_h(t) dt \right) ds \\
&= \int_0^1 (H_h(s) - C_H(s)) ds,
\end{aligned} \tag{2.15}$$

where $H_h : [0, 1] \rightarrow \mathbb{R}$ is defined by

$$H_h(s) = \int_0^s h(t) dt,$$

and $C_{H_h} : [0, 1] \rightarrow \mathbb{R}$ is the convex rearrangement of H_h defined by

$$C_H(s) = \int_0^s I_h(t) dt.$$

The penultimate equation of (2.15) is derived from the fact that

$$\int_s^1 h(t) dt = \int_0^1 h(t) dt - \int_0^s h(t) dt$$

and

$$\int_s^1 I_h(t) dt = \int_0^1 I_h(t) dt - \int_0^s I_h(t) dt.$$

Moreover, along with equality (2.5), we have

$$\int_s^1 (I_h(t) - h(t)) dt = \int_0^s (h(t) - I_h(t)) dt.$$

The right-hand side of equation (2.15) is the desired alternative expression of \mathcal{L}_h .

Part 2: In view of expression (2.15), the non-negativity of \mathcal{L}_h follows from the bound

$$H_h(t) \geq C_{H_h}(t) \quad \text{for all } t \in [0, 1]. \tag{2.16}$$

To prove bound (2.16), we first note that every real number $y \in \mathbb{R}$ can be decomposed as the sum $w_1(y) + w_2(y)$, where $w_1(y) = \min\{y, 0\}$ and $w_2(y) = \max\{y, 0\}$. Hence,

$$\begin{aligned} I_h(s) &= F_X^{-1}(s) \\ &= w_1(F_X^{-1}(s)) + w_2(F_X^{-1}(s)). \end{aligned} \tag{2.17}$$

Now we recall (e.g., Denuit et al., 2005, Property 1.5.16(i), p. 19) that for every non-decreasing and continuous function w , we have the equation $w(F_X^{-1}(s)) = F_{w(X)}^{-1}(s)$. Since w_1 and w_2 are non-decreasing and continuous, equation (2.17) implies

$$\begin{aligned} I_h(s) &= F_{w_1(X)}^{-1}(s) + F_{w_2(X)}^{-1}(s) \\ &= I_{h_-}(s) + I_{h_+}(s), \end{aligned}$$

where $h_-(s) = w_1(h(s))$ and $h_+(s) = w_2(h(s))$. Hence,

$$\begin{aligned} \int_0^t I_h(s) ds &= \int_0^t I_{h_-}(s) ds + \int_0^t I_{h_+}(s) ds \\ &\leq \int_0^t h_-(s) ds + \int_0^t h_+(s) ds \\ &= \int_0^t h(s) ds, \end{aligned}$$

provided that

$$\int_0^t I_{h_-}(s) ds \leq \int_0^t h_-(s) ds \tag{2.18}$$

and

$$\int_0^t I_{h_+}(s)ds \leq \int_0^t h_+(s)ds. \tag{2.19}$$

We shall prove bounds (2.18) and (2.19) next.

Proof of bound (2.18). Let $X_- = \min\{X, 0\}$. We have the equation

$$\int_0^t I_{h_-}(s)ds = \int_0^t F_{X_-}^{-1}(s)ds$$

and thus the bound

$$\int_0^t I_{h_-}(s)ds \leq \int_0^t F_{YX_-}^{-1}(s)ds, \tag{2.20}$$

where Y is the random variable defined by $Y(\omega) = \mathbf{1}\{\omega \leq t\}$. To establish bound (2.20), we have used the inequality $X_- \leq YX_-$, which holds because X_- is non-positive.

Next we observe that the cdf $F_{YX_-}(x)$ takes on the value 1 at the point $x = 0$ and has a jump of a size at least as high as $1 - t$ at the point $x = 0$. Hence, the quantile function $F_{YX_-}^{-1}(s)$ is equal to 0 for at least all $s \in (t, 1)$, and so we have the equations:

$$\begin{aligned} \int_0^t F_{YX_-}^{-1}(s)ds &= \int_0^1 F_{YX_-}^{-1}(s)ds \\ &= \mathbb{E}[YX_-] \\ &= \int_0^t h_-(s)ds. \end{aligned} \tag{2.21}$$

Bound (2.20) and equations (2.21) complete the proof of bound (2.18).

Proof of bound (2.19). Let $X_+ = \max\{X, 0\}$. In our following considerations we shall need to estimate X_+ from below by ZX_+ , where Z is the random variable

defined by $Z(\omega) = \mathbf{1}\{\omega > t\}$. For this reason, we now observe that bound (2.19) is equivalent to the following one:

$$\int_t^1 I_{h_+}(s)ds \geq \int_t^1 h_+(s)ds. \tag{2.22}$$

The equivalence of the two bounds follows from the equation

$$\int_0^1 I_{h_+}(s)ds = \int_0^1 h_+(s)ds,$$

which is a consequence of equation (2.5). To establish bound (2.22), we start with the equation

$$\int_t^1 I_{h_+}(s)ds = \int_t^1 F_{X_+}^{-1}(s)ds$$

and arrive at the bound

$$\int_t^1 I_{h_+}(s)ds \geq \int_t^1 F_{ZX_+}^{-1}(s)ds. \tag{2.23}$$

The cdf $F_{ZX_+}(x)$ is equal to 0 for all $x < 0$ and has a jump of a size at least as high as t at the point $x = 0$. Hence, the quantile function $F_{ZX_+}^{-1}(s)$ is equal to 0 for at least all $s \in (0, t)$, and so we have the equations:

$$\begin{aligned} \int_t^1 F_{ZX_+}^{-1}(s)ds &= \int_0^1 F_{ZX_+}^{-1}(s)ds \\ &= \mathbb{E}[ZX_+] \\ &= \int_t^1 h_+(s)ds. \end{aligned} \tag{2.24}$$

Bound (2.23) and equations (2.24) complete the proof of bound (2.22) and thus, in turn, establish bound (2.19) as well.

Having thus proved bounds (2.18) and (2.19), we have established bound (2.16). As we have noted earlier, this implies that \mathcal{L}_h is non-negative.

Part 3: In this final part of the proof of Theorem 2.5.1, we establish the fact that \mathcal{L}_h takes on the value 0 if and only if the function h is non-decreasing. This we do in two parts.

First, we assume that h is non-decreasing. Then the function H_h is convex. Furthermore, the convex rearrangement C_{H_h} of the function H_h leaves the function H_h unchanged because H_h is convex. In summary, when h is non-decreasing, then the integral $\int_0^1 (H_h(t) - C_{H_h}(t)) dt$ and thus the index \mathcal{L}_h are equal to 0.

Moving now in the opposite direction, if the integral $\int_0^1 (H_h(t) - C_{H_h}(t)) dt$ is equal to 0, then due to the already proved bound $H_h \geq C_{H_h}$, we have $H_h(t) = C_{H_h}(t)$ for λ -almost all $t \in [0, 1]$. Consequently, the function H_h must be convex, and thus the function h must be non-decreasing. This concludes the proof of Step 3, and thus of the entire Theorem 2.5.1. ■

As we have seen in the above proof, the definition of the index \mathcal{L}_h fundamentally relies on the notion of convex rearrangement, which also prominently features in several other research areas, such as:

- Stochastic processes (e.g., Davydov, 1998; Davydov and Zitikis, 2003, 2005; Davydov and Thilly, 2007; Thilly, 1999; Zhukova, 1994).
- Convex analysis (e.g., Davydov and Vershik, 1998) with applications in areas such as the optimal transport problem (e.g., Lachièze-Rey and Davydov, 2011).

- Econometrics (e.g., Gastwirth, 1971; Giorgi, 2005; Lorenz, 1905).
- Insurance (e.g., Brazauskas et al., 2008; Greselin et al., 2009; Necir et al., 2010).

2.6 Concluding notes

Inspired by applications in a number of research areas, we have explored two indices, \mathcal{I} and \mathcal{L} , designed for measuring the lack of monotonicity in functions. The indices take on the value 0 for every non-decreasing function, and on positive values for other functions: the larger the values, the less non-decreasing the function is deemed to be. The index \mathcal{I} is simpler, but it is only subadditive for co-monotonic functions, whereas the index \mathcal{L} is more complex, but it is additive for co-monotonic functions. In bivariate case, these indices can be interpreted as the distances of random variables and their co-monotonic forms that motivate us to use it as the lack of co-monotonicity indices. We discuss this notion in more details along with its illustrative example in Chapter 4.

Chapter 3

Computing the indices

Except for very simple functions such as h_α and g_α of Examples 2.4.1 and 2.4.2, calculating the indices \mathcal{I} and \mathcal{L} is usually a tedious and time consuming task. To facilitate a practical implementation irrespective of the function h , we next develop a technique that gives numerical values of the two indices at any prescribed precision and in virtually no time.

3.1 General considerations

We start with a general observation: given two integrable functions $h, g : [0, 1] \rightarrow \mathbb{R}$, we have the bound

$$\int_0^1 |I_h(t) - I_g(t)| dt \leq \int_0^1 |h(t) - g(t)| dt, \quad (3.1)$$

¹This chapter in a condensed form has been published in Qoyyimi, D. T. and Zitikis, R. (2014). Measuring the lack of monotonicity in functions, *Mathematical Scientist* 39(2): 107–117.

which is well known (e.g., Lorentz, 1953) and has been utilized by many researchers (e.g., Chernozhukov et al., 2009; Egorov, 1990; Thilly, 1999; Zhukova, 1994).

Proof of bound (3.1) Using the probabilistic interpretation, we write the equation

$$\int_0^1 |I_h(t) - I_g(t)| dt = \int_0^1 |F_X^{-1}(t) - F_Y^{-1}(t)| dt. \quad (3.2)$$

The integral on the right-hand side of equation (3.2) is known as the Dobrushin distance between the two cdf's F_X and F_Y . The integral is equal (Dobrushin, 1970) to $\inf \mathbb{E}[|\xi - \eta|]$, where the infimum is taken over all random variables ξ and η that have finite first moments and whose cdf's are equal to F_X and F_Y , respectively. The infimum is not larger than $\mathbb{E}[|h(U) - g(U)|]$, where U is a uniform random variable on $\Omega = [0, 1]$, because the cdf's of the random variables $h(U)$ and $g(U)$ are equal to F_X and F_Y , respectively. Indeed, in the case of $h(U)$ for example, the cdf $F_{h(U)}$ of $h(U)$ is equal to $\mathbb{P}\{\omega \in \Omega : h(U(\omega)) \leq x\}$, which is equal to $\lambda\{t \in [0, 1] : h(t) \leq x\}$ because $U(\omega) = \omega$ by the definition of the uniform random variable on $\Omega = [0, 1]$. Note that $\lambda\{t \in [0, 1] : h(t) \leq x\}$ is equal to $G_h(x)$, which is in turn equal to $F_X(x)$ according to our probabilistic interpretation. Hence, $F_{h(U)} = F_X$ and, likewise, $F_{g(U)} = F_Y$. ■

Due to bound (3.1), we obviously have

$$|\mathcal{I}_h - \mathcal{I}_g| \leq 2 \int_0^1 |h(t) - g(t)| dt. \quad (3.3)$$

Likewise, we obtain the bound

$$|\mathcal{L}_h - \mathcal{L}_g| \leq 2 \int_0^1 |h(t) - g(t)| dt, \quad (3.4)$$

which holds for every pair of integrable functions $g, h : [0, 1] \rightarrow \mathbb{R}$. Just like bound (3.3), bound (3.4) helps us to develop a discretization technique for calculating the index \mathcal{L}_h numerically. More details on the technique follow next.

Namely, we shall replace g by a specially constructed estimator \widehat{h} of h such that the L_1 -distance $\int_0^1 |h(t) - \widehat{h}(t)| dt$ can be made as small as desired by choosing a sufficiently small ‘tuning’ parameter n . To this end, we proceed as follows. First, we partition the interval $[0, 1)$ into n subintervals $[(i-1)/n, i/n)$ and then choose any point t_i in each subinterval. Denote $\tau_i = h(t_i)$ and let

$$\widehat{h}(t) = \begin{cases} \tau_i & \text{when } t \in [(i-1)/n, i/n), \\ \tau_n & \text{when } t = 1. \end{cases} \quad (3.5)$$

With $\tau_{1:n} \leq \dots \leq \tau_{n:n}$ denoting the ordered values τ_1, \dots, τ_n , the function

$$G_{\widehat{h}}(x) = \lambda\{t \in [0, 1] : \widehat{h}(t) \leq x\}$$

can be written as

$$G_{\widehat{h}}(x) = \begin{cases} 0 & \text{for } x < \tau_{1:n}, \\ i/n & \text{for } x \in [\tau_{i:n}, \tau_{i+1:n}), \quad 1 \leq i \leq n-1, \\ 1 & \text{for } x \geq \tau_{n:n}. \end{cases}$$

Hence, the non-decreasing rearrangement

$$I_{\widehat{h}}(t) = \inf\{x \in \mathbb{R} : G_{\widehat{h}}(x) \geq t\}$$

can be expressed in a computationally convenient way as

$$I_{\widehat{h}}(t) = \tau_{i:n} \quad \text{for every } t \in ((i-1)/n, i/n],$$

which holds for every $i = 1, \dots, n$. This implies

$$\begin{aligned} \mathcal{I}_{\widehat{h}} &= \int_0^1 |I_{\widehat{h}}(t) - \widehat{h}(t)| dt \\ &= \frac{1}{n} \sum_{i=1}^n |\tau_{i:n} - \tau_i|. \end{aligned} \quad (3.6)$$

Likewise, to calculate $\mathcal{L}_{\widehat{h}}$, we use formula (2.11) with \widehat{h} instead of h , and then employ the above expressions for \widehat{h} and $I_{\widehat{h}}$. We obtain

$$\begin{aligned} \mathcal{L}_{\widehat{h}} &= \int_0^1 (I_{\widehat{h}}(t) - \widehat{h}(t)) t dt \\ &= \frac{1}{n^2} \sum_{i=1}^n i (\tau_{i:n} - \tau_i). \end{aligned} \quad (3.7)$$

From bounds (3.3) and (3.4), we conclude that $|\mathcal{I}_{\widehat{h}} - \mathcal{I}_h|$ and $|\mathcal{L}_{\widehat{h}} - \mathcal{L}_h|$ do not exceed $2 \int_0^1 |\widehat{h}(t) - h(t)| dt$, which converges to 0 when $n \rightarrow \infty$ irrespectively of the chosen t_i 's because the function h is integrable on $[0, 1]$. Hence, instead of calculating the usually unwieldy \mathcal{I}_h and \mathcal{L}_h , we can employ formulas (3.6) and (3.7) and easily calculate $\mathcal{I}_{\widehat{h}}$ and $\mathcal{L}_{\widehat{h}}$ instead. Choosing a sufficiently large n , we can reach any desired level of accuracy. An illustration of this procedure follows next.

3.2 Illustrations with insights into the indices

Here we calculate and interpret the indices in the case of the functions

$$h_1(t) = \sin(tM)$$

and

$$h_2(t) = \cos(tM)$$

defined on the interval $[0, 1]$, for several values of M . The functions are of course simple, but we have nevertheless visualized them in Figure 3.1 in order to facilitate our following discussion. We have used estimators (3.6) and (3.7) to calculate the indices, with the obtained values reported in Table 3.1. We see from the table that when $M = \pi/2$ and π , then irrespectively of which of the two indices we use, the function h_1 is more non-decreasing (i.e., the index value is smaller) than h_2 . The two functions are equally non-decreasing when $M = 3\pi/2$. When $M = 2\pi$, then the function h_1 is less non-decreasing (i.e., the index value is larger) than h_2 , and this is so for both indices. We shall now make sense of the numerical values by analyzing the four panels of Figure 3.1.

M	\mathcal{I}_{h_1}	\mathcal{I}_{h_2}	M	\mathcal{L}_{h_1}	\mathcal{L}_{h_2}
$\pi/2$	0.0000	0.5274	$\pi/2$	0.0000	0.1739
π	0.3183	1.2732	π	0.0870	0.4053
$3\pi/2$	1.1027	1.1027	$3\pi/2$	0.3409	0.3409
2π	1.2732	0.8270	2π	0.3618	0.2026

Table 3.1: Indices of non-decreasingness of $h_1(t) = \sin(tM)$ and $h_2(t) = \cos(tM)$.

Panel (a) is clear: the increasing function h_1 has its index zero, and the decreasing function h_2 has a positive index.

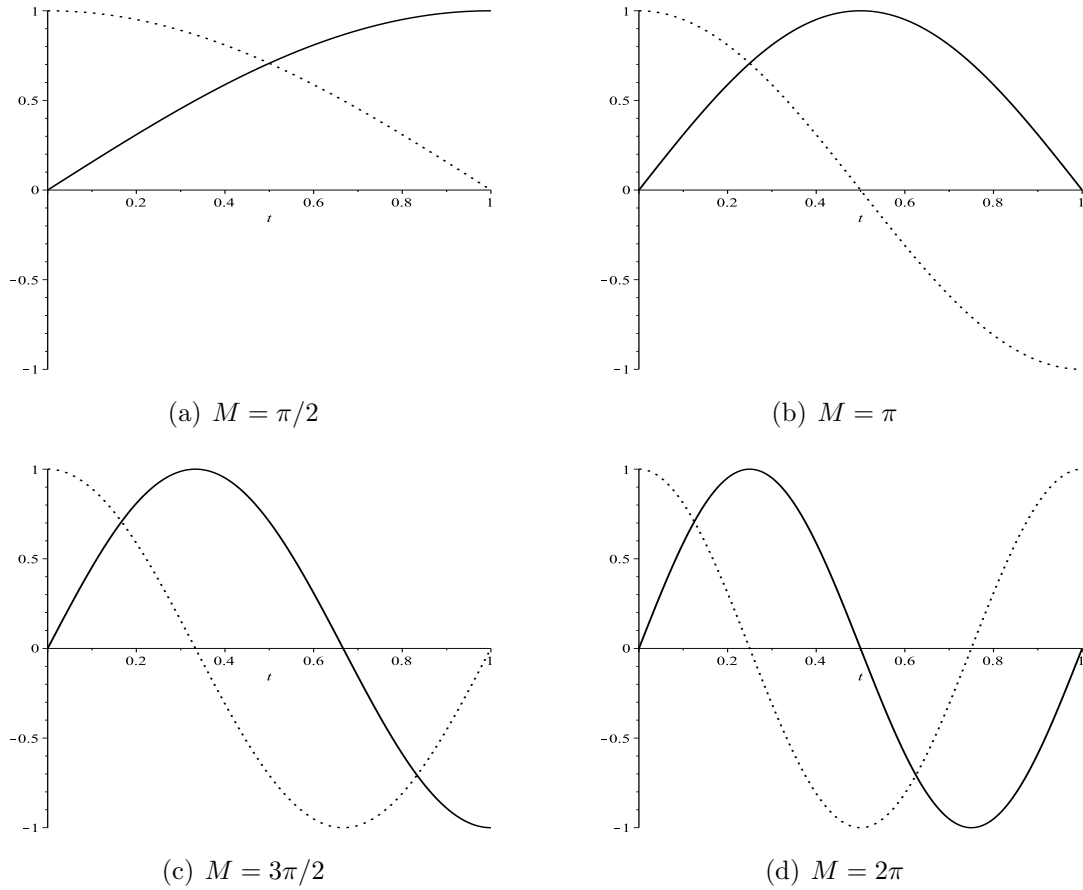


Figure 3.1: Functions $h_1(t) = \sin(tM)$ (solid) and $h_2(t) = \cos(tM)$ (dotted).

In panel (b), the function h_1 is increasing in the first half of the interval $[0, 1]$ and the function h_2 is always decreasing. Not surprisingly, therefore, any of the two indices of the function h_1 is smaller than the corresponding index of h_2 .

In panel (c), the two functions have the same \mathcal{I} -indices, as well as the same \mathcal{L} -indices, and the reason for this is based on the general property that if $g(t) = -h(1-t)$ for all $t \in [0, 1]$, then $I_g(t) = -I_h(1-t)$ for all $t \in [0, 1]$. Hence, the equations $\mathcal{I}_g = \mathcal{I}_h$ and $\mathcal{L}_g = \mathcal{L}_h$ hold. In words, if we flip h upside-down and also from left to right, then the value of any of the two indices will not change. This is why the two functions in panel (c) have the same \mathcal{I} -indices as well as the same \mathcal{L} -indices.

The results corresponding to panel (d) are more challenging to explain. To proceed, we adopt the following route: We subdivide the interval $(0, 1]$ into four equal subintervals as follows:

$$[0, 1) = \bigcup_{k=1}^{2M/\pi} \left[\frac{k-1}{2M/\pi}, \frac{k}{2M/\pi} \right); \tag{3.8}$$

recall that $M = 2\pi$ in this case. By reshuffling these four subintervals, we can reconstruct the function h_2 out of the corresponding pieces of the function h_1 , and we can of course do so the other way around. This one-to-one mapping between the two functions may wrongly suggest that the indices of the two functions should be the same, but they are obviously not, as we see from Table 3.1. With some tinkering we realize, however, that this is so because the original order of the aforementioned pieces of the function h_1 is such that this function is more ‘wiggly’ (i.e., follows the pattern ‘increase-decrease-increase’) than the function h_2 (i.e., follows the pattern ‘decrease-increase’). Naturally now, since more wiggly functions tend to be less monotonic, the function h_1 has a larger index than the function h_2 . Table 3.2 summarizes this point of view for all of the four panels of Figure 3.1.

	Panel (a)	Panel (b)	Panel (c)	Panel (d)
$h_1(t)$	+	+-	+ - - [= +-]	+ - - + [= + - +]
\mathcal{I}_{h_1}	0	0.3183	1.1027	1.2732
$h_2(t)$	-	-+	- - + [= - +]	- - + + [= - +]
\mathcal{I}_{h_2}	0.5274	1.2732	1.1027	0.8270

Table 3.2: Increasing (+) and decreasing (-) regions of h_1 and h_2 .

Another illustration relates to Example 2.4.2. Let X and Y be two random variables that are uniformly distributed on the unit interval $[0, 1]$. We pair X and Y through the equation $Y = g_\alpha(X)$, where g_α is the function defined in Example 2.4.2,

and whose definition for the reader's convenience is reminded next:

$$g_\alpha(t) = \begin{cases} \frac{t}{\alpha} & \text{for } t \in [0, \alpha], \\ \frac{t-1}{\alpha-1} & \text{for } t \in (\alpha, 1] \end{cases}$$

when $\alpha \in (0, 1)$,

$$g_\alpha(t) = 1 - t \quad \text{when } \alpha = 0,$$

and

$$g_\alpha(t) = t \quad \text{when } \alpha = 1.$$

We have already calculated the indices \mathcal{I}_{g_α} and \mathcal{L}_{g_α} for all $\alpha \in [0, 1]$, and now want to see how the indices are related to dependence measures. To facilitate the discussion, we recall the formulas

$$\mathcal{I}_{g_\alpha} = \frac{1 - \alpha}{2 - \alpha} \tag{3.9}$$

and

$$\mathcal{L}_{g_\alpha} = \frac{1}{6}(1 - \alpha). \tag{3.10}$$

These indices as functions of α are depicted in Figure 2.4. When $\alpha = 0$, we have the pair $(X, 1 - X)$, which is counter-monotonic as discussed in Section 1.1. When $\alpha = 1$, we have the pair (X, X) which is co-monotonic. As α grows from 0 to 1, the random vector (X, Y) moves from being counter-monotonic $(X, 1 - X)$ to co-monotonic (X, X) . Since in the bivariate case the joint distribution of any counter-monotonic random vector is the Fréchet-Hoeffding lower bound, then this is the farthest form from being co-monotonic. In this case, the index should be maximal among all $\alpha \in [0, 1]$. Indices (3.9) and (3.10), as well as illustrations in Figure 2.4, support this

argument. Also, as α grows, the random vector (X, Y) approaches the co-monotonic form, and so the distance from co-monotonicity gradually diminishes.

3.3 Indices of functions on finite intervals

Suppose now that we want to measure the lack of non-decreasingness of a function defined on $[a, A] \subset \mathbb{R}$. Since shifting to the left or to the right does not change the shape of the function, and thus its degree of non-decreasingness, we thus redefine the function onto the interval $[0, M]$ by simply replacing its argument t by $t - a$, where $M = A - a$. Therefore, without loss of generality, from now on we work with any integrable function f defined on the interval $[0, M]$, for some $M > 0$. We note at the outset that we cannot reduce our task to the interval $[0, 1]$ by simply replacing its argument t by tM because such an operation would inevitably distort the degree of non-decreasingness.

Hence, given a function $f : [0, M] \rightarrow \mathbb{R}$, we proceed by first defining its non-decreasing rearrangement by the formula

$$I_{f,M}(t) = \inf\{x \in \mathbb{R} : G_{f,M}(x) \geq t\} \quad \text{for all } t \in [0, M],$$

where

$$G_{f,M}(x) = \lambda\{t \in [0, M] : f(t) \leq x\} \quad \text{for all } x \in \mathbb{R}.$$

Our first index of non-decreasingness of the function $f : [0, M] \rightarrow \mathbb{R}$ is then defined by

$$\mathcal{I}_{f,M} = \int_0^M |I_{f,M}(t) - f(t)| dt. \quad (3.11)$$

Furthermore, with

$$H_{f,M}(t) = \int_0^t f(s)ds$$

and

$$C_{H_{f,M}}(t) = \int_0^t I_{f,M}(s)ds$$

for all $t \in [0, M]$, we define the second index of non-decreasingness of f by the formula

$$\mathcal{L}_{f,M} = \int_0^M (H_{f,M}(t) - C_{H_{f,M}}(t)) dt. \quad (3.12)$$

We shall next illustrate the two indices using the functions $\sin(t)$ and $\cos(t)$ defined on the four domains $[0, \pi/2]$, $[0, \pi]$, $[0, 3\pi/2]$, and $[0, 2\pi]$. The values of the two indices are given in Table 3.3. Since this example mimics that of Section 3.2,

M	$\mathcal{I}_{\sin,M}$	$\mathcal{I}_{\cos,M}$	M	$\mathcal{L}_{\sin,M}$	$\mathcal{L}_{\cos,M}$
$\pi/2$	0.0000	0.8284	$\pi/2$	0.0000	0.4292
π	1.0000	4.0000	π	0.8584	4.0000
$3\pi/2$	5.1962	5.1962	$3\pi/2$	7.5708	7.5708
2π	8.0000	5.1962	2π	14.2832	8.0000

Table 3.3: Indices of non-decreasingness of $\sin(t)$ and $\cos(t)$ on $[0, M]$.

various interpretations there apply here as well. In short, we see from the table that irrespectively of which of the two non-decreasing indices we use, the index of non-decreasingness of $\sin(t)$ is smaller than that of $\cos(t)$ on the domains $[0, \pi/2]$ and $[0, \pi]$. The two functions have the same non-decreasingness indices on $[0, 3\pi/2]$. Finally, on the domain $[0, 2\pi]$, the index of non-decreasingness of the function $\sin(t)$ is greater than that of $\cos(t)$, irrespectively of which of the two indices we use, which implies that $\sin(t)$ is less non-decreasing than $\cos(t)$ on $[0, 2\pi]$.

We have used a discretization technique to calculate the values reported in Ta-

ble 3.3. The technique is a modification of that of Section 3.1. To explain the modification, in Theorem 3.3.1 below we establish a connection between the pair of the earlier introduced indices on the interval $[0, 1]$ and the pair of the current ones on the interval $[0, M]$.

Theorem 3.3.1 *Let $f : [0, M] \rightarrow \mathbb{R}$ for some $M > 0$, and let $h : [0, 1] \rightarrow \mathbb{R}$ be the function defined by $h(t) = f(tM)$ for all $t \in [0, 1]$. Then*

$$\mathcal{I}_{f,M} = M\mathcal{I}_h \quad \text{and} \quad \mathcal{L}_{f,M} = M^2\mathcal{L}_h. \quad (3.13)$$

Proof Since

$$G_h(x) = \frac{1}{M}G_{f,M}(x),$$

we have $I_h(t) = I_{f,M}(tM)$ for all $t \in [0, 1]$. Hence,

$$\begin{aligned} \mathcal{I}_h &= \int_0^1 |I_{f,M}(tM) - f(tM)| \, dt \\ &= \frac{1}{M}\mathcal{I}_{f,M}, \end{aligned}$$

which establishes the first equation of (3.13). To prove the second equation, we first check that

$$H_h(t) = \frac{1}{M}H_{f,M}(tM)$$

and $C_{H_h}(t) = C_{H_{f,M}}(tM)/M$. Consequently,

$$\begin{aligned} \mathcal{L}_h &= \frac{1}{M} \int_0^1 (H_{f,M}(tM) - C_{H_{f,M}}(tM)) \, dt \\ &= \frac{1}{M^2}\mathcal{L}_{f,M}. \end{aligned}$$

This establishes the second equation of (3.13), and concludes the proof. ■

We are now in the position to introduce estimators $\widehat{\mathcal{I}}_{f,M}$ and $\widehat{\mathcal{L}}_{f,M}$ of the indices $\mathcal{I}_{f,M}$ and $\mathcal{L}_{f,M}$, respectively. Namely, with $h(t) = f(tM)$ and using formulas (3.6) and (3.7), we have

$$\widehat{\mathcal{I}}_{f,M} = \frac{M}{n} \sum_{i=1}^n |\tau_{i:n} - \tau_i| \quad (3.14)$$

and

$$\widehat{\mathcal{L}}_{f,M} = \left(\frac{M}{n}\right)^2 \sum_{i=1}^n i (\tau_{i:n} - \tau_i), \quad (3.15)$$

where $\tau_{1:n} \leq \dots \leq \tau_{n:n}$ denote the ordered values

$$\tau_i = f(t_i M), \quad i = 1, \dots, n.$$

We used formulas (3.14) and (3.15) to obtain the numerical values of the two indices reported in Table 3.3, where we set $n = 100,000$ in order to have a mesh sufficiently fine to achieve the desired accuracy level of four decimal digits.

Chapter 4

Measuring association via lack of co-monotonicity

4.1 Introduction

Measuring association, or lack of it, between variables has fascinated researchers for centuries. A considerable impetus to this research was given by Sir Francis Galton who in 1885 published his empirical and theoretical developments on regression and correlation. Inspired by that work, a decade later Karl Pearson proposed a coefficient to measure correlation between variables, which is nowadays known as the Pearson correlation coefficient. For detailed historical notes on this coefficient and many of its interpretations, we refer to Rodger and Nicewander (1988), and Stanton (2001).

The mathematical simplicity and thus interpretability of the Pearson correlation coefficient have encouraged researchers to use it in a variety of areas where measuring

¹A small portion of this chapter in a condensed form has been published in Qoyyimi, D. T. and Zitikis, R. (2015). Measuring association via lack of co-monotonicity: the LOC index and a problem of educational assessment, *Dependence Modeling* 3(1): 83–97.

association between variables is of interest. But just like any other synthetic measure, the Pearson correlation coefficient also has a number of limitations, such as high sensitivity to outliers (e.g., Abdullah, 1990; Shevlyakov and Smirnov, 2011), its reliance on linearity (e.g., Rodger and Nicewander, 1988), and so on.

In many practical situations, however, we encounter problems that are poorly described by linear relationships, and thus measuring association (or lack of it) using the Pearson correlation coefficient may not be prudent. Hence, a number of alternative ways have emerged in the literature, including the rank correlation coefficients of Kendall and Spearman (e.g., Kendall and Stuart, 1961; Nelsen, 2006), the correlation coefficients of Gini (e.g., Gupta, 1999; Nelsen, 2006) and Blomqvist (1950). Concisely, these coefficients provide different counting and aggregation rules of concordant and discordant pairs of bivariate data:

Two pairs (x_i, y_i) and (x_j, y_j) are concordant if either $x_i < x_j$ and $y_i < y_j$,
or $x_i > x_j$ and $y_i > y_j$.

For detailed and illuminating discussions of these coefficients, we refer to Section 5.1 of Nelsen (2006). For recent discussions, methodological and applied developments on copulas, we refer to Embrechts (2009), Jaworski et al. (2010, 2013), and references therein.

The concordance notion leads immediately to the notion of co-monotonicity that has deep roots in mathematics (cf., e.g., Denneberg, 1994; Schmeidler, 1986, and reference therein). This notion has turned out to be particularly useful in economics, finance, and insurance. For details and references on the topic, we refer to, e.g., Dhaene et al. (2002b,a, 2006), Vyncke (2004), and references therein.

A number of indices for measuring dependence, concordance, and co-monotonicity have been proposed in the literature (e.g., Dhaene et al., 2012, 2014; Koch and

de Schepper, 2011; Liebscher, 2014, and references therein). All of them are concerned with different aspects of dependence but nevertheless – as intended by the authors – fall into a large class of concordance coefficients that possess certain ‘desirable’ characteristics or properties (e.g., Mari and Kotz, 2001; Nelsen, 2006; Schweizer and Wolff, 1981; Scarsini, 1984, and references therein). In particular, among those characteristics is a symmetry (or interchangeability, permutation, etc.) condition, which in the context of this thesis is not desirable and would even be misleading, due to the very reason that explanatory and response variables are not symmetric (interchangeable). Hence, for measuring the lack of, or departure from, co-monotonicity between pairs of variables, none of the aforementioned coefficients can truly serve our purpose.

We have organized the rest of the chapter as follows. In Section 4.2 we describe a classical data-set of Thorndike and Thorndike-Christ (2010), which is of our primary interest, and then visualize the data using scatterplots with superimposed classical least-squares regression lines. In Section 4.3 we fit curves to bivariate data using several powerful methods available in the literature, which is a precursor to our use of an index for measuring *lack of co-monotonicity* (LOC). The definition of the LOC index is provided in Section 4.4. In Section 4.5 we utilize the LOC index to analyze the data-set of Thorndike and Thorndike-Christ (2010). In Section 4.6 we discuss the difference between the LOC index and that of Liebscher (2014), and the conclusion is given in Section 4.7.

4.2 Data

To facilitate full transparency of our reasoning and adopted methodology, we use a publicly available data-set of Thorndike and Thorndike-Christ (2010). The set consists of marks of 52 sixth grade students on three study subjects: Mathematics, Reading, and Spelling. The students belonged to two classes, taught by two teachers, who administered tests on the three subjects. For each student and for each study subject, the teachers reported the number of correct answers and used them to assess each student's achievement on each of the three subjects.

For our analysis, we first normalize the marks to the unit interval $[0, 1]$ by dividing the number of correct answers by the total number of items (i.e., questions or problems) on the tests: 65 items for Mathematics, 45 for Reading, and 80 for Spelling. Hence, throughout the chapter we deal with functions

$$h : [0, 1] \rightarrow [0, 1]$$

that model association between pairs of study subjects, which we denote by X and Y , connected via the hypothetical equation $y = h(x)$ with h estimated from data (topic of Section 4.3). Summary statistics and histograms of the normalized marks are reported in Table 4.1 and Figure 4.1. In Figure 4.2 we have depicted the corresponding six scatterplots, which provide valuable insights into relationships between paired variables (e.g., Best et al., 2006; Cleveland et al., 1982; Meyer and Shinar, 1992, and references therein). Even though we argue that the relationships between student marks on all pairs of study subjects are non-linear, it is nevertheless instructive to start considerations with the classical least-squares regression lines, which we have depicted in Figure 4.2, and to also report values of the Pearson correlation coefficient

Summary statistics	Mathematics	Reading	Spelling
Minimum	0.2923	0.4667	0.4750
1st quartile	0.5077	0.6833	0.6375
2nd quartile (median)	0.5846	0.7778	0.7188
3rd quartile	0.6769	0.8667	0.8000
Mean	0.5873	0.7654	0.7192
Maximum	0.9231	0.9778	0.9500
Standard deviation	0.1373	0.1233	0.1129

Table 4.1: Summary statistics.

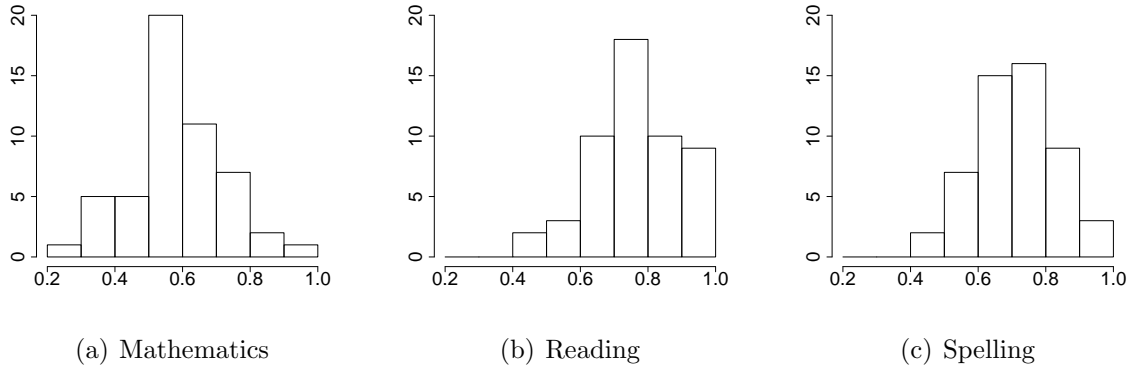


Figure 4.1: Frequency histograms.

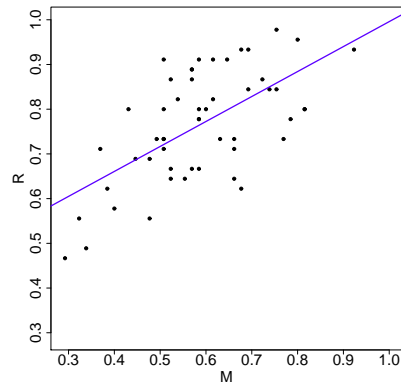
values, which we have done in Table 4.2.

	Mathematics	Reading	Spelling
Mathematics	1.000000	0.622224	0.146615
Reading	0.622224	1.000000	0.642215
Spelling	0.146615	0.642215	1.000000

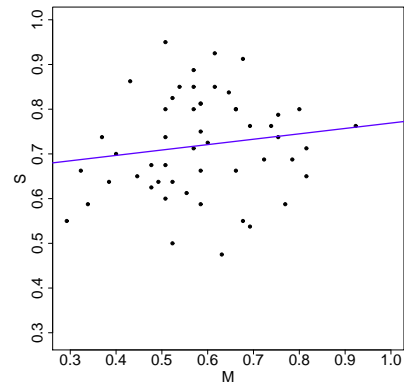
Table 4.2: Pearson correlation coefficients.

4.3 Curve fitting

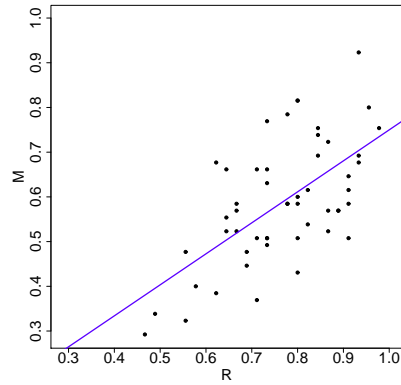
Here we discuss curve fitting to scatterplots – and we have six of them (see Figure 4.2) – which is a precursor to calculating the LOC index, which is a topic of Section 4.4



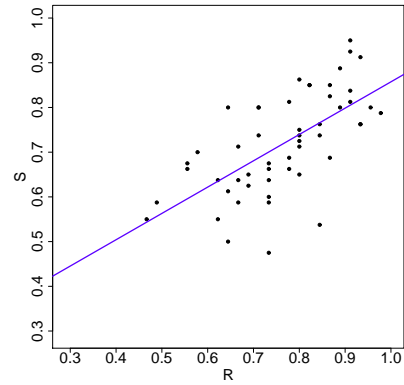
(a) Mathematics-Reading



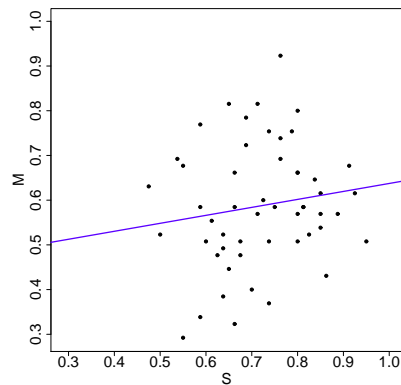
(b) Mathematics-Spelling



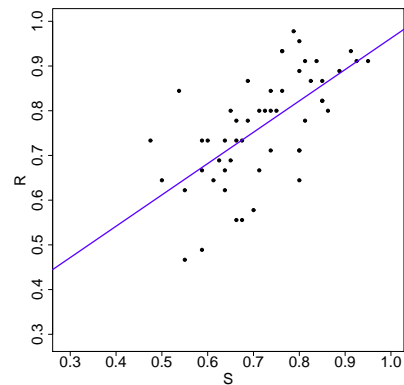
(c) Reading-Mathematics



(d) Reading-Spelling



(e) Spelling-Mathematics



(f) Spelling-Reading

Figure 4.2: Scatterplots and least-squares regression lines.

below.

A number of approaches have been developed for fitting curves to bivariate data. The parametric approach is one of them, which includes popular models such as linear, generalized linear, nonlinear, parametric growth curve, and many other ones (see, e.g., Pan and Fang, 2002; Seber and Wild, 1989). The disadvantage of this approach, especially in the context of the present problem, is that the shape and form of the functions to be fitted are difficult to guess, and thus involves an element of subjectivity that we want to avoid. Hence, we opt for the non-parametric approach, which is sometimes referred to as scatterplot smoothing (Ruppert et al., 1995).

In general, there are two broad non-parametric approaches for fitting curves to bivariate data: one is based on the conditional mean and another on a conditional quantile, such as the conditional median. Both methods have their own advantages and disadvantages, and we shall illustrate both of them. We note at the outset that in the case of the conditional quantile, we shall restrict our attention to the conditional median that serves a natural alternative to the mean when data are skewed. Some further details and references on the two methods will be provided in Section 4.3.1 below, with their actual use for analyzing the data of Thorndike and Thorndike-Christ (2010) exhibited in Section 4.5.

4.3.1 Constructing \hat{h}

The *conditional-mean* approach is based on the assumption that a good model for h is given by the conditional mean, and thus

$$h(x) = \mathbb{E}[Y|X = x]. \quad (4.1)$$

Given a scatterplot consisting of n pairs (x_i, y_i) , the local linear estimate – which is our choice among many other ones available in the literature – for estimating $h(x)$ is given by

$$\widehat{h}(x) = \widehat{\beta}_0,$$

where $\widehat{\beta}_0$ is a solution to the minimization problem

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n L(y_i - (\beta_0 + \beta_1(x_i - x))) K\left(\frac{x_i - x}{b}\right); \quad (4.2)$$

throughout this chapter we work with the standard normal kernel K . Details and references on the bandwidth b selection will be provided in Section 4.3.2 below. As to the loss function L , in the conditional-mean case we use the quadratic loss function $L(x) = x^2$, which is a natural choice because the expected quadratic loss is minimized at the mean. In the case of the conditional-median approach, an analogous argument leads us toward the absolute loss function $L(x) = |x|$.

We note in passing that this estimate naturally arises from the fact (recall here the local constant regression method of Nadaraya-Watson model) that $h(x)$ defined by equation (4.1) solves the minimization problem $\mathbb{E}[(Y - \beta_0)^2 | X = x]$ with respect to β_0 . The additional quantity $\beta_1(x_i - x)$ in objective function (4.2) is included to diminish the asymptotic bias of the estimate, if compared to the bias arising from the Nadaraya-Watson method (Fan, 1992). For further properties of the local linear estimate, we refer to Simonoff (1996), Wand and Jones (1995), and references therein.

It is also natural to use the *conditional-quantile* approach (Koenker, 2005), which is based on the assumption that a good model for $h(x)$ is given by the conditional quantile, and thus

$$h(x) = Q_{Y|X=x}(\tau)$$

for some $\tau \in (0, 1)$. An estimate $\hat{h}(x)$ of $h(x)$ stems from the minimization problem of (4.2) using the loss function $L(x)$ that is equal to τx for all $x \geq 0$ and $(1-\tau)(-x)$ for all $x < 0$. Upon recalling that throughout this problem we set $\tau = 0.5$, in the conditional-median case we therefore work with the absolute loss function $L(x) = 0.5|x|$. The factor 0.5 is of course irrelevant in our considerations as it does not influence the result of minimization problem (4.2).

4.3.2 Bandwidth selection

The construction of bandwidth b is based on how good the resulting estimator $\hat{h}(x)$ of $h(x)$ is, and for this task it is customary to use the mean integrated squared error (MISE)

$$\text{MISE}(\hat{h}) = \int \mathbb{E} \left[(\hat{h}(x) - h(x))^2 \mid x_1, x_2, \dots, x_n \right] w(x) dx \quad (4.3)$$

with some weight function w that ensures convergence of the integral (Ruppert and Wand, 1994). Specifically, the bandwidth is chosen so that it asymptotically minimizes the MISE. There are of course other good ways to choose the bandwidth but we shall not delve deeply into this subject here and just note some of the facts that we shall utilize in our data-driven computations.

Namely, we follow Fan and Gijbels (2000), Ruppert and Wand (1994), and Ruppert et al. (1995) when using the *conditional-mean* approach. We start out with the asymptotic optimal bandwidth given by formula (3.21) in Fan and Gijbels (1996, p. 68). To facilitate its practical implementation, we use the direct plug-in method proposed by Ruppert et al. (1995, pp. 1262–1263). In the latter reference, the resulting bandwidth is denoted by \hat{h}_{DPI} , which in this chapter is denoted by \hat{b} to avoid a possible notational confusion with the estimate \hat{h} of h .

When using the *conditional-median* approach, we follow Yu and Jones (1997), who show that the optimal bandwidth in this case is equal to the estimate \widehat{b} from the conditional-mean approach multiplied by

$$\left\{ \frac{\tau(1-\tau)}{\phi(\Phi^{-1}(\tau))^2} \right\}^{1/5},$$

where $\tau = 1/2$ due to our median based approach. The ϕ in the above quantity is the standard normal density, and Φ^{-1} is the standard normal quantile function. Hence, in summary, the optimal bandwidth under the conditional-median approach is

$$\widehat{b}(\pi/2)^{1/5}. \tag{4.4}$$

4.4 Measuring the lack of co-monotonicity

In view of the above discussion, we can now assume that for any given scatterplot we have constructed a well-fitting function

$$\widehat{h} : [0, 1] \rightarrow [0, 1].$$

If the function happens to be increasing, then we say that the random variables X and Y have co-monotonic movements, but if not, then we want to assess how much the function deviates from the increasing pattern. This we accomplish using an index that takes value 0 when \widehat{h} is increasing and some positive value otherwise: the more the function deviates from the increasing pattern, the larger the value. The index is the one we introduced and discussed in Chapter 2. The second index (\mathcal{L}) is used in this application.

4.5 Data analysis and findings

We work with six scatterplots, and to each of them we fit two curves: one using the conditional-mean approach and the other one using the conditional-median approach. In both cases, we use the same mathematical notation \hat{h} but when plotting in Figure 4.3, we use different colors to distinguish the two cases. The technicalities of curve fitting follow next, for which we use the R software (R Core Team, 2013).

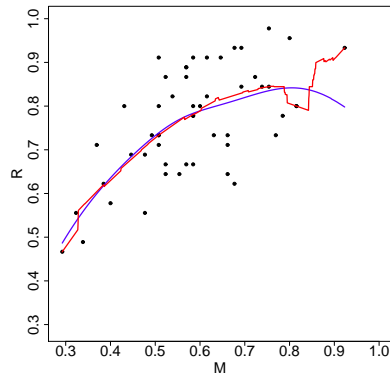
In the case of the *conditional-mean* approach, we use the local linear kernel regression method as discussed in Section 4.3.1. To aid us with computations, we use the R package `Kernsmooth` (Wand and Ripley, 2014) with the function `dpill` assigned for selecting the optimal bandwidth and the function `locpoly` (with `degree=1`) for curve fitting. We set the grid size to 1,000.

In the case of the *conditional-median* approach, we use the R package `quantreg` (Koenker, 2015) with the function `lpqr` used to obtain \hat{h} with $\tau = 1/2$ and $m = 1,000$. We see from the six panels of Figure 4.3 that all the estimates \hat{h} are more jiggly than those arising from the conditional-mean approach. Definitely, we can improve them with more work and a more sophisticated tuning of the parameters, but this would beat our purpose of showing that we can easily calculate the LOC index irrespective of how much irregular the function is.

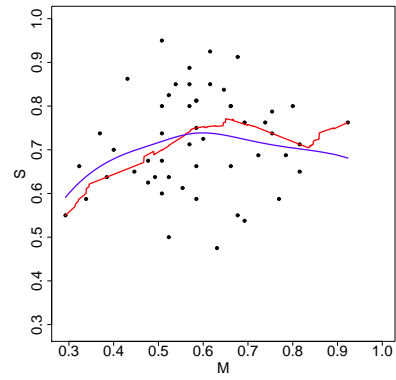
Based on our visual assessment, no function in Figure 4.3 appears to be increasing over its entire domain of definition. Nevertheless, we may argue that some of them are more increasing than others. To substantiate this claim, we employ the LOC index discussed in Section 4.4. The following terminology is useful.

Definition 4.5.1 *Given two functions $\hat{g}, \hat{h} : [0, 1] \rightarrow [0, 1]$, we say that*

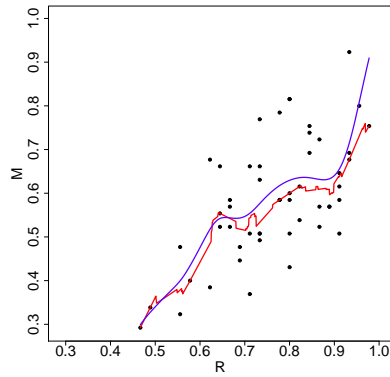
- (1) \hat{g} deviates from increasing pattern by the amount $\mathcal{L}(\hat{g})$;



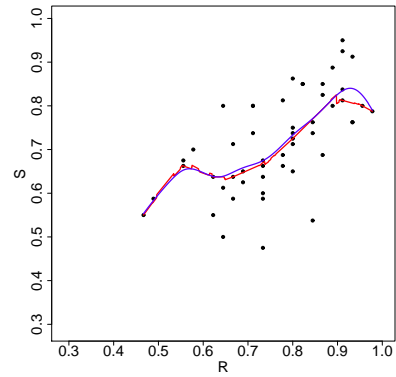
(a) Mathematics-Reading



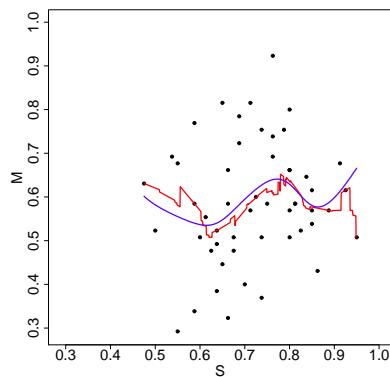
(b) Mathematics-Spelling



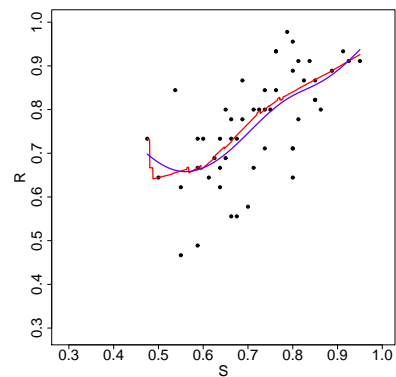
(c) Reading-Mathematics



(d) Reading-Spelling



(e) Spelling-Mathematics



(f) Spelling-Reading

Figure 4.3: Conditional-mean (blue) and conditional-median (red) based curves.

- (2) \hat{g} deviates less from increasing pattern than \hat{h} when $\mathcal{L}(\hat{g}) < \mathcal{L}(\hat{h})$; and
- (3) pairs (v_i, w_i) , $i = 1, \dots, n$, exhibit less LOC than pairs (x_i, y_i) , $i = 1, \dots, m$, when $\mathcal{L}(\hat{g}) < \mathcal{L}(\hat{h})$, where \hat{g} arises from the pairs (v_i, w_i) and \hat{h} from (x_i, y_i) .

Following the guidelines of Section 3.1, we produce the step-wise approximation D_m of the function \hat{h} . Then we calculate the index $\hat{\mathcal{L}}(\hat{D}_m)$ according to formula (3.7). Findings in the form of 'LOC matrices' are presented in Tables 4.3 and 4.4, whose entries are the values of the LOC index: the larger the value, the more the

X	Y		
	Mathematics	Reading	Spelling
Mathematics	0.000000	0.231814	1.759735
Reading	0.007202	0.000000	0.097565
Spelling	0.855971	0.145532	0.000000

Table 4.3: Conditional-mean based LOC matrix (entries multiplied by 1,000, see Note 4.5.1).

X	Y		
	Mathematics	Reading	Spelling
Mathematics	0.000000	0.286703	0.923108
Reading	0.007911	0.000000	0.163541
Spelling	2.197968	0.175055	0.000000

Table 4.4: Conditional-median based LOC matrix (entries multiplied by 1,000, see Note 4.5.1).

corresponding pairs deviate from the co-monotonic pattern.

The LOC matrix is, naturally, asymmetric, and it should be such in order to match the asymmetry that we see in the respective paired panels of Figure 4.2. For example, the entry 0.231814 in Table 4.3 is the value (multiplied by 1,000) of the LOC index for Mathematics-Reading, whereas 0.007202 is the value (multiplied by 1,000) of the LOC index for Reading-Mathematics.

Note 4.5.1 We have multiplied all the original LOC-index values by 1,000 to avoid recording too many decimal zeros in the tables; this does not matter when interpreting our results because the LOC-index is relative as we see from parts (2) and (3) of Definition 4.5.1.

Naturally, one may also wish to know how much a given study subject influences the other ones, which leads us in the direction of causality (e.g., Cheng, 1997; Pearl, 2009, and references therein), which at this stage of our research we want to avoid discussing. Nevertheless, the reader may wish to draw some conclusions from Tables 4.3 and 4.4, as well as from the scatterplots of Figure 4.3. Note that even though the corresponding entries of Tables 4.3 and 4.4 are different, the causality-type conclusions that we may infer from both of them would not contradict each other. This may not always be the case, especially if data are considerably skewed. In the case of the data that we are exploring, however, the descriptive statistics and histograms in Section 4.2 suggest fairly symmetric distributions of all the three study subjects.

4.6 Comparing the LOC index with Liebscher's ζ

Liebscher's (2014) suggestion for determining whether co-movements of random variables follow an increasing pattern is philosophically closest to our current research. Specifically, given a pair of random variables, say X and Y , whose cdf's we denote by F and G , respectively, Liebscher's (2014) coefficient of *monotonically increasing dependence* is

$$\zeta_{X,Y} = 1 - \frac{1}{c_\psi} \mathbb{E}[\psi(F(X) - G(Y))], \quad (4.5)$$

where

$$c_\psi = 2 \int_0^1 (1-u)\psi(u)du$$

is the normalizing constant, and ψ can be any non-negative and symmetric around 0 function on the interval $[-1, 1]$ such that $\psi(0) = 0$. For example, $\psi(x) = x^2/2$, which we shall use in a moment.

Various properties and extensions of this index have been discussed by Liebscher (2014), from which we see that, to a certain degree, the index can be used for tackling our problem. Yet, due to a different goal set out by Liebscher (2014), his index does not truly serve our needs because it is 1) symmetric with respect to X and Y as we have noted earlier, and 2) based on rank scatterplots, whereas our problem relies on raw-data scatterplots, which can be considerably different from rank-based scatterplots as we shall see from graphs in later in this section.

Naturally, to understand $\zeta_{X,Y}$ we only need to understand its expectation-based part, which under the quadratic function

$$\psi(x) = \frac{x^2}{2}$$

is equal to

$$I_{X,Y} = \frac{1}{2} \mathbb{E} \left[(F(X) - G(Y))^2 \right].$$

Note 4.6.1 The quantity $I_{X,Y}$ is closely related to the Spearman rank correlation coefficient, denoted here by $S_{X,Y}$, which is, by definition, equal to the Pearson correlation coefficient between $F(X)$ and $G(Y)$. Hence, we easily check the equation $S_{X,Y} = 1 - 12 I_{X,Y}$.

Next we work with a scatterplot (x_i, y_i) , $i = 1, \dots, n$, which we view as our ‘pop-

ulation'. To avoid computational complications that inevitably arise when dealing with ranks when some of the x_i 's or y_i 's are equal, throughout the rest of this section we work under the assumption

$$x_i \neq x_j \quad \text{and} \quad y_i \neq y_j \quad \text{whenever} \quad i \neq j. \quad (4.6)$$

Note 4.6.2 Assumption (4.6) is violated by the data-set of Thorndike and Thorndike-Christ (2010). However, this is not an issue because we can always add negligible noise – e.g., independent and identically distributed normal random variables with means 0 and very small standard deviations, say 10^{-5} – and make all the marks unequal without *practically* changing their numerical values.

Let F_n and G_n be the marginal cdf's defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i \leq x\}$$

and

$$G_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{y_i \leq x\}.$$

Under this 'finite population' scenario, the quantity $I_{X,Y}$ becomes

$$I_{n,x,y} = \frac{1}{2n} \sum_{i=1}^n (F_n(x_i) - G_n(y_i))^2.$$

Let $x_{1:n} < \dots < x_{n:n}$ be the ordered values of x_1, \dots, x_n , and let $y_{(1)}, \dots, y_{(n)}$ be the corresponding induced ordered values. In other words, the original pairs (x_i, y_i) have been ordered according to their first coordinates and the resulting pairs are now

$(x_{i:n}, y_{(i)})$. With the notation

$$r_i = nG_n(y_{(i)}) \quad (4.7)$$

we have

$$\begin{aligned} I_{n,x,y} &= \frac{1}{2n} \sum_{i=1}^n (F_n(x_{i:n}) - G_n(y_{(i)}))^2 \\ &= \frac{1}{2n} \sum_{i=1}^n \left(\frac{i}{n} - \frac{r_i}{n} \right)^2 \\ &= \frac{1}{2n^3} \sum_{i=1}^n (i - r_i)^2, \end{aligned} \quad (4.8)$$

where we used the equation $F_n(x_{i:n}) = i/n$.

Next we construct a function $h_n^0 : [0, 1] \rightarrow [0, 1]$ such that $\mathcal{L}(h_n^0)$ is equal to the right-hand side of equation (4.8) or, in other words, such that

$$\mathcal{L}(h_n^0) = I_{n,x,y}. \quad (4.9)$$

Namely, for every $i = 1, \dots, n$, let

$$h_n^0(t) = \frac{r_i}{n} \quad \text{for all } t \in \left(\frac{i-1}{n}, \frac{i}{n} \right]. \quad (4.10)$$

The LOC index of the function h_n^0 is

$$\begin{aligned} \mathcal{L}(h_n^0) &= \sum_{i=1}^n \left(\frac{i}{n} - \frac{r_i}{n} \right) \int_{(i-1)/n}^{i/n} t dt \\ &= \frac{1}{2n^3} \sum_{i=1}^n (i - r_i)(2i - 1) \\ &= \frac{1}{2n^3} \sum_{i=1}^n (i - r_i)^2, \end{aligned} \quad (4.11)$$

where we used the equations $\sum_{i=1}^n i = \sum_{i=1}^n r_i$ and $\sum_{i=1}^n i^2 = \sum_{i=1}^n r_i^2$. This establishes equation (4.9) and helps us to connect the index \mathcal{L} with Liebscher's ζ .

For this, we first observe that the set of equations

$$h_n^0\left(\frac{i}{n}\right) = \frac{r_i}{n}, \quad i = 1, \dots, n, \quad (4.12)$$

is equivalent to the set

$$h_n^0(F_n(x_{i:n})) = G_n(y_{(i)}), \quad i = 1, \dots, n,$$

which is in turn equivalent to the set of equations

$$h_n^0(F_n(x_i)) = G_n(y_i), \quad i = 1, \dots, n.$$

This implies that Liebscher's ζ is the LOC index \mathcal{L} of the step-wise function h_n^0 , which originates from the *rank-based* scatterplot $(F_n(x_i), G_n(y_i))$ and not from the original scatterplot $(x_{i:n}, y_{(i)})$. This also explains a considerable difference between the meanings of the two indices. To support our conclusions, we have depicted the two scenarios in Figure 4.4, where we have used Mathematics (with added noise) as the 'explanatory' variable and Reading (with added noise) as the 'response.'

Consequently, in order to decide whether the problem at hand would be better served by the index \mathcal{L} or Liebscher's ζ , we first need to decide whether the solution of the problem should rely on the original scatterplot $(x_{i:n}, y_{(i)})$, $i = 1, \dots, n$, or on the rank-based scatterplot $(F_n(x_{i:n}), G_n(y_{(i)}))$; the latter is of course equivalent to the scatterplot

$$\left(\frac{i}{n}, \frac{r_i}{n}\right), \quad i = 1, \dots, n.$$

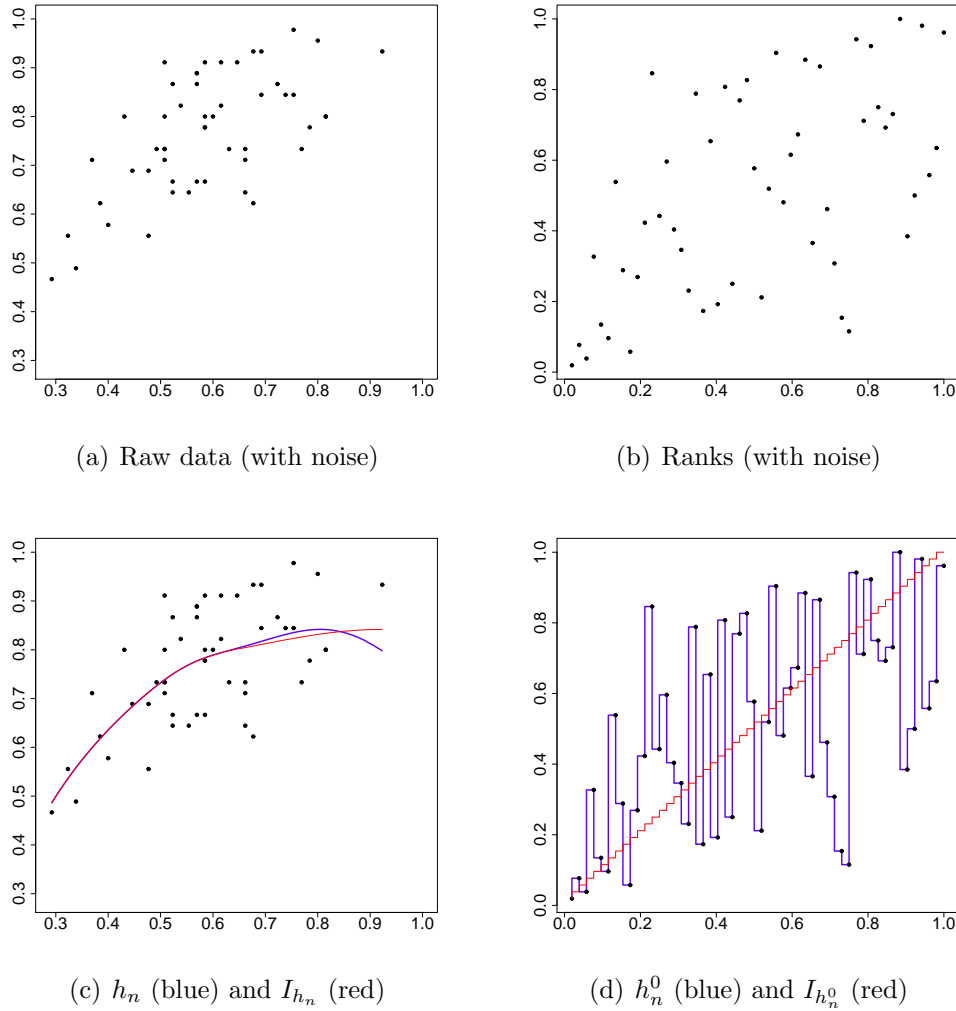


Figure 4.4: Scatterplots and fitted functions.

If the association between student *rankings* according to their marks is of primary interest, with no consideration to causality, then Liebscher’s ζ is an appropriate index. If, however, the *marks* themselves are of primary interest, as is the case in the current problem, and keeping in mind that the marks are not interchangeable random variables with respect to causality, then we should rely on the original scatterplot

$$(x_{i:n}, y_{(i)}), \quad i = 1, \dots, n$$

and use the LOC index \mathcal{L} .

4.7 Concluding notes

The herein proposed index for measuring the lack of co-monotonicity between pairs of variables is capable of measuring the extent to which the variables deviate from co-monotonic patterns. The LOC index is designed to work with all relationships, including non-linear and non-monotonic. The performance of the index has been illustrated using the Thorndike and Thorndike-Christ (2010) data-set consisting of student marks on three study subjects.

Chapter 5

Concluding remarks and future work

5.1 Concluding remarks

This thesis has focused on measuring the departure from co-monotonicity patterns produced by bivariate random elements. While studies of co-monotonicity and its consequences for complex systems have been extensively analyzed in recent years, measuring the lack of it, to the best of our knowledge, has been missing in these discussions. Co-monotonic forms, in bivariate case, can be explained by the existence of non-decreasing functions that satisfactorily fit the patterns. Measuring the lack of co-monotonicity, therefore, can be interpreted as measuring the distance of the functions from their monotone rearrangements. The lack of co-monotonicity can then be measured by the distance of functions from their non-decreasing rearrangements. We have proposed two indices for measuring the lack of monotonicity in functions: one index is based on the L_1 -norm and another one is a Gini-type index. Their properties

and computational algorithms have also been provided.

To illustrate the herein developed technique, we have used the lack of monotonicity index to study associations between study subjects in education. Specifically, given student marks in various subjects, we have explored the problem of comparing various study subjects with other subjects. We have calculated the lack of co-monotonicity index that measures the distance of bivariate random vectors from their co-monotonic forms and made appropriate conclusions.

5.2 Future work

Our technique allows us to explore the dependence structures of bivariate random vectors in terms of appropriately constructed functions. This facilitates a very broad applicability to a great variety of problems, which include:

- testing hazard rate dominance,
- testing pairwise counter-monotonicity and generalized counter-monotonicity,
- investigating pairwise co-monotonicity in financial data,
- investigating co-monotonicity of high dimensional vectors, and many others.

Some of these problems, such as testing hazard rate dominance, have already been explored by me but not included in this thesis due to the required additional work and time constraints. They will become a part of my future research and graduate student supervision programs at Universitas Gadjah Mada, where I already have a permanent position.

Bibliography

- Abdullah, M. B. (1990). On a robust correlation coefficient. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 39(4):455–460.
- Barlow, R. E. and Proschan, F. (1974). *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt Rinehart and Winston, New York.
- Bebbington, M., Green, R., Lai, C.-D., and Zitikis, R. (2014). Beyond the Gompertz law: exploring the late-life mortality deceleration phenomenon. *Scandinavian Actuarial Journal*, 2014(3):1–19.
- Bebbington, M., Hall, A. J., Lai, C.-D., and Zitikis, R. (2009). Dynamics and phases of kiwifruit (*Actinidia deliciosa*) growth curves. *New Zealand Journal of Crop and Horticultural Science*, 37(3):179–188.
- Bebbington, M., Lai, C.-D., and Zitikis, R. (2007). Reliability of modules with load-sharing components. *Journal of Applied Mathematics and Decision Sciences*, 2007(Article ID 43565):1–18.
- Bebbington, M., Lai, C.-D., and Zitikis, R. (2008). Reduction in mean residual life in the presence of a constant competing risk. *Applied Stochastic Models in Business and Industry*, 24(1):51–63.

- Bebbington, M., Lai, C.-D., and Zitikis, R. (2011). Modelling deceleration in senescent mortality. *Mathematical Population Studies*, 18(1):18–37.
- Best, L. A., Hunter, A. C., and Stewart, B. M. (2006). Perceiving relationships: a physiological examination of the perception of scatterplots. In Barker-Plummer, D., Cox, R., and Swoboda, N. (Eds.), *Diagrammatic Representation and Inference*, pp. 244–257. Springer-Verlag, Berlin, Heidelberg.
- Bickel, P. J. and Doksum, K. A. (2001). *Mathematical Statistics: Basic Ideas and Selected Topics* (2nd ed.). Prentice-Hall, Upper Saddle River, NJ.
- Blomqvist, N. (1950). On a measure of dependence between two random variables. *Annals of Mathematical Statistics*, 21(4):593–600.
- Brazauskas, V., Jones, B. L., Puri, M. L., and Zitikis, R. (2008). Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference*, 138(11):3590–3604.
- Brazauskas, V., Jones, B. L., and Zitikis, R. (2009). When inflation causes no increase in claim amounts. *Journal of Probability and Statistics*, 2009(Article ID 943926):1–10.
- Brazauskas, V., Jones, B. L., and Zitikis, R. (2015). Trends in disguise. *Annals of Actuarial Science*, 9(1):58–71.
- Carrier, G. and Dana, R.-A. (2003). Pareto efficient insurance contracts when the insurer’s cost function is discontinuous. *Economic Theory*, 21(4):871–893.
- Carrier, G. and Dana, R.-A. (2005). Rearrangement inequalities in non-convex insurance models. *Journal of Mathematical Economics*, 41(4-5):483–503.

- Carlier, G. and Dana, R.-A. (2008). Two-persons efficient risk-sharing and equilibria for concave law-invariant utilities. *Economic Theory*, 36(2):189–223.
- Carlier, G. and Dana, R.-A. (2011). Optimal demand for contingent claims when agents have law invariant utilities. *Mathematical Finance*, 21(2):169–201.
- Cheng, P. W. (1997). From covariation to causation: a causal power theory. *Psychological Review*, 104(2):367–405.
- Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575.
- Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.
- Chong, K. M. and Rice, N. M. (1971). *Equimeasurable Rearrangements of Functions*. Queen’s Papers in Pure and Applied Mathematics – No. 28. Queen’s University, Kingston, Ontario.
- Cleveland, W. S., Diaconis, P., and McGill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. *Science*, 216(4550):1138–1141.
- Dana, R.-A. and Scarsini, M. (2007). Optimal risk sharing with background risk. *Journal of Economic Theory*, 133(1):152–176.
- Davydov, Y. (1998). Convex rearrangements of stable processes. *Journal of Mathematical Science*, 92(4):4010–4016.

- Davydov, Y. and Thilly, E. (2007). Convex rearrangements of Levy process. *ESAIM: Probability and Statistics*, 11(1):161–172.
- Davydov, Y. and Vershik, A. M. (1998). Réarrangements convexes des marches aléatoires. *Annales de l'Institut Henri Poincaré*, 34(1):73–95.
- Davydov, Y. and Zitikis, R. (2003). Generalized Lorenz Curves and Convexifications of Stochastic Processes. *Journal of Applied Probability*, 40(4):906–925.
- Davydov, Y. and Zitikis, R. (2005). An index of monotonicity and its estimation: a step beyond econometric applications of the Gini index. *Metron*, 63(3):351–372.
- Day, P. W. (1970). *Rearrangement of Measurable Functions*. PhD thesis, California Institute of Technology.
- Deelstra, G., Dhaene, J., and Vanmaele, M. (2011). An overview of comonotonicity and its applications in finance and insurance. In Di Nunno, G. and Øksendal, B. (Eds.), *Advanced Mathematical Methods for Finance*, pp. 155–179. Springer, Berlin.
- Denneberg, D. (1994). *Non-additive Measure and Integral*. Springer, Dordrecht.
- Denuit, M., Dhaene, J., Goovaerts, M. J., and Kaas, R. (2005). *Actuarial Theory for Dependent Risk: Measures, Orders and Models*. Wiley, Chicester.
- Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., and Vyncke, D. (2002a). The concept of comonotonicity in actuarial science and finance: applications. *Insurance: Mathematics and Economics*, 31(2):133–161.

- Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., and Vyncke, D. (2002b). The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, 31(1):3–33.
- Dhaene, J., Linders, D., Schoutens, W., and Vyncke, D. (2012). The herd behavior index: a new measure for the implied degree of co-movement in stock markets. *Insurance: Mathematics and Economics*, 50(3):357–370.
- Dhaene, J., Linders, D., Schoutens, W., and Vyncke, D. (2014). A multivariate dependence measure for aggregating risks. *Journal of Computational and Applied Mathematics*, 263(1):78–87.
- Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., Tang, Q., and Vyncke, D. (2006). Risk measures and comonotonicity: a review. *Stochastic Models*, 22(4):573–606.
- Diers, D., Eling, M., and Marek, S. D. (2012). Dependence modeling in non-life insurance using the Bernstein copula. *Insurance: Mathematics and Economics*, 50(3):430–436.
- Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distribution. *Theory of Probability and Its Applications*, 15(3):458—486.
- Egorov, V. A. (1990). A functional law of iterated logarithm for ordered sums. *Theory of Probability and its Applications*, 35(2):342–347.
- Egozcue, M., García, L. F., and Zitikis, R. (2013). An optimal strategy for maximizing the expected real-estate selling price: accept or reject an offer? *Journal of Statistical Theory and Practice*, 7(3):596–609.

- Embrechts, P. (2009). Copulas: a personal view. *Journal of Risk and Insurance*, 76(3):639–650.
- Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87(420):998–1004.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall, London.
- Fan, J. and Gijbels, I. (2000). Local polynomial fitting. In Schimek, M. (Ed.), *Smoothing and Regression: Approaches, Computation, and Application*, pp. 229–276. Wiley, New York.
- Furman, E. and Zitikis, R. (2008a). Weighted premium calculation principles. *Insurance: Mathematics and Economics*, 42(1):459–465.
- Furman, E. and Zitikis, R. (2008b). Weighted risk capital allocations. *Insurance: Mathematics and Economics*, 43(2):263–269.
- Furman, E. and Zitikis, R. (2009). Weighted pricing functionals with applications to insurance. *North American Actuarial Journal*, 13(4):483–496.
- Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica*, 39(6):1037–1039.
- Gavrilov, L. A. and Gavrilova, N. S. (1991). *The Biology of Life Span: a Quantitative Approach*. Harwood Academic Publishers, New York.
- Genest, C., Gendron, M., and Bourdeau-Brien, M. (2009). The advent of copulas in finance. *European Journal of Finance*, 15(7-8):609–618.

- Ghossoub, M. (2015). Equimeasurable rearrangements with capacities. *Mathematics of Operation Research*, 40(2):429–445.
- Giorgi, G. M. (2005). Gini's scientific work: an evergreen. *Metron*, 63(3):299–315.
- Greselin, F., Puri, M. L., and Zitikis, R. (2009). L-functions, processes, and statistics in measuring economic inequality and actuarial risks. *Statistics and Its Interface*, 2(2009):227–245.
- Gupta, S. D. (1999). Gini association and pseudo Lorenz curve. *Communications in Statistics—Theory and Methods*, 28(9):2181–2199.
- Hardy, G. H., Littlewood, J. E., and Pólya, G. (1934). *Inequalities*. Cambridge University Press, Cambridge.
- He, X. D. and Zhou, X. Y. (2011). Portfolio choice via quantiles. *Mathematical Finance*, 21(2):203–231.
- Jaworski, P., Durante, F., and Härdle, W. K. (2013). *Copulae in Mathematical and Quantitative Finance: Proceedings of the Workshop Held in Carcow, 10-11 July 2012*. Springer, Berlin.
- Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T. (2010). *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*. Springer, Berlin.
- Jin, H. and Zhou, X. Y. (2008). Behavioral portfolio selection in continuous time. *Mathematical Finance*, 18(3):385–426.
- Joe, H. (2001). *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, New York.

- Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall/CRC, Boca Raton, Florida.
- Kaas, R., Dhaene, J., and Goovaerts, M. J. (2000). Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics*, 27(2):151–168.
- Kendall, M. G. and Stuart, A. (1961). *The Advanced Theory of Statistics Vol. 2: Inference and Relationship* (3th ed.). Hafner Publishing Company, New York.
- Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*, (4th ed.). Wiley, Hoboken, NJ.
- Koch, I. and de Schepper, A. (2011). Measuring comonotonicity in m-dimensional vectors. *ASTIN Bulletin*, 41(1):191–213.
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York.
- Koenker, R. (2015). quantreg: Quantile Regression. R package version 5.11.
- Korenovskii, A. (2007). *Mean Oscillations and Equimeasurable Rearrangements of Functions*. Springer, Berlin.
- Lachièze-Rey, R. and Davydov, Y. (2011). Rearrangements of Gaussian fields. *Stochastic Processes and their Applications*, 121(11):2606–2628.
- Lai, C.-D. and Xie, M. (2006). *Stochastic Ageing and Dependence for Reliability*. Springer, New York.
- Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, 37(5):1137–1153.

- Levy, H. (2006). *Stochastic Dominance: Investment Decision Making under Uncertainty* (2nd ed.). Springer, New York.
- Li, D. X. (2000). On default correlation. *Journal of Fixed Income*, 9(4):43–54.
- Li, H. and Li, X. (2013). *Stochastic Orders in Reliability and Risk: In Honor of Professor Moshe Shaked*. Lecture Notes in Statistics Vol. 208. Springer, New York.
- Liebscher, E. (2014). Copula-based dependence measures. *Dependence Modelling*, 2(1):49–64.
- Lorentz, G. (1953). An inequality for rearrangements. *American Mathematical Monthly*, 60(3):176–179.
- Lorenz, M. (1905). Methods of measuring the concentration of wealth. *Publication of the American Statistical Association*, 9(70):209–219.
- Mari, D. D. and Kotz, S. (2001). *Correlation and Dependence*. Imperial College Press, London.
- Marshall, A. W., Olkin, I., and Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications* (2nd ed.). Springer, New York.
- Meyer, J. and Shinar, D. (1992). Estimating correlation from scatterplots. *Human Factors*, 34(3):335–349.
- Necir, A., Rassoul, A., and Zitikis, R. (2010). Estimating the conditional tail expectation in the case of heavy-tailed losses. *Journal of Probability and Statistics*, 2010(Article ID 596839):1–17.

- Nelsen, R. B. (2006). *An Introduction to Copulas* (2nd ed.). Springer, New York.
- Pan, J. and Fang, K. (2002). *Growth Curve Models and Statistical Diagnostics*. Springer, New York.
- Panik, M. J. (2014). *Growth Curve Modeling: Theory and Applications*. Wiley, New York.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press, New York.
- Puccetti, G. and Scarsini, M. (2010). Multivariate comonotonicity. *Journal of Multivariate Analysis*, 101(1):291–304.
- Puccetti, G. and Wang, R. (2015). Extremal dependence concepts. *Available at SSRN: <http://ssrn.com/abstract=2436392>*, 2015(435844):1–34.
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.
- Quiggin, J. (1993). *Generalized Expected Utility Theory: The Rank Dependent Model*. Kluwer, Dordrecht.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Reddy, M. J. and Ganguli, P. (2012). Bivariate flood frequency analysis of upper Godavari river flows using Archimedean copulas. *Water Resources Management*, 26(14):3995–4018.
- Rodger, J. L. and Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *American Statistician*, 42(1):59–66.

- Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270.
- Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, 22(3):1346–1370.
- Rüschendorf, L. (1983). Solution of a statistical optimization problem by rearrangement methods. *Metrika*, 30(1):55–61.
- Scarsini, M. (1984). On measures of concordance. *Stochastica*, 8(3):201–218.
- Schmeidler, D. (1986). Integral representation without additivity. *Proceeding of the American Mathematical Society*, 97(2):255–261.
- Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *Annals of Statistics*, 9(4):879–885.
- Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. Wiley, New York.
- Sendov, H. S., Wang, Y., and Zitikis, R. (2011). Log-supermodularity of weight functions, ordering weighted losses, and the loading monotonicity of weighted premiums. *Insurance: Mathematics and Economics*, 48(2):257–264.
- Shaked, M. and Shantikumar, J. G. (2007). *Stochastic Orders*. Springer, New York.
- Shevlyakov, G. and Smirnov, P. (2011). Robust estimation of the correlation coefficient: an attempt of survey. *Austrian Journal of Statistics*, 40(1):147–156.
- Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. Springer, New York.

- Singpurwalla, N. (2006). *Reliability and Risk: A Bayesian Perspective*. Wiley, Chichester.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.
- Stanton, J. (2001). Galton, Pearson, and the peas: a brief history of linear regression for statistics instructors. *Journal of Statistics Education*, 9(3).
- Sun, W., Rachev, S., Fabozzi, F. J., and Kalem, P. S. (2008). A new approach to modeling co-movement of international equity markets: evidence of unconditional copula-based simulation of tail dependence. *Empirical Economics*, 36(1):201–229.
- Thilly, E. (1999). *Réarrangements Convexes des Trajectoires de Processus Stochastiques*. PhD thesis, Université de Lille I.
- Thorndike, R. M. and Thorndike-Christ, T. (2010). *Measurement and Evaluation in Psychology and Education* (8th ed.). Prentice Hall, Boston.
- Vyncke, D. (2004). Comonotonicity. In Teugels, J. and Sundt, B. (Eds.), *Encyclopedia of Actuarial Science*, pp. 302–305. Wiley, Chichester.
- Wand, M. and Ripley, B. (2014). KernSmooth: Functions for Kernel Smoothing. R package version 2.23-13.
- Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall/CRC, London.
- Warsido, W. P. and Bitsuamlak, G. T. (2015). Synthesis of wind tunnel and climatological data for estimating design wind effects: a copula based approach. *Structural Safety*, 57(2015):8–17.

- Yitzhaki, S. and Schechtman, E. (2012). Identifying monotonic and non-monotonic relationships. *Economics Letters*, 116(1):23–25.
- Young, V. R. (2004). Premium principles. In Teugels, J. L. and Sundt, B. (Eds.), *Encyclopedia of Actuarial Science*, pp. 1–9. Wiley, Chichester.
- Yu, K. and Jones, M. C. (1997). A comparison of local constant and local linear regression quantile estimators. *Computational Statistics & Data Analysis*, 25(2):159–166.
- Zhukova, E. E. (1994). *Monotone and Convex Rearrangements of Functions and Stochastic Processes*. PhD thesis, Saint-Petersburg University, Saint-Petersburg.
- Zhukova, E. E. (1998). Increasing permutations of random processes. *Journal of Mathematical Sciences*, 88(1):43–52.

Curriculum vitae

Name: Danang Teguh Qoyyimi

**Post-Secondary
Education and
Degrees:** Universitas Gadjah Mada
Yogyakarta, Indonesia
2006–2008 M.Sc.

University of Western Ontario
London, ON
2011–2015 Ph.D.

**Honours and
Awards:** Directorate General of Higher Education of Indonesia
Scholarship for Ph.D study
2011–2015

Young researcher award
at the International Conference on AISC
The University of North Carolina Greensboro
October 10–12, 2014

**Related Work
Experience:** Lecturer
Universitas Gadjah Mada, Indonesia
2008–present

Publications:

1. Qoyyimi, D. T. and Zitikis, R. (2015). Measuring association via lack of co-monotonicity: the LOC index and a problem of educational assessment. *Dependence Modeling*, 3(1): 83–97.
2. Qoyyimi, D. T. and Zitikis, R. (2014). Measuring the lack of monotonicity in functions. *Mathematical Scientist*, 39(2): 107–117.