

# Distributional Compatibility for Change of Measures

Jie Shen<sup>1</sup>, Yi Shen<sup>1</sup>, Bin Wang<sup>2</sup>, and Ruodu Wang<sup>1</sup>

<sup>1</sup>Department of Statistics and Actuarial Science, University of Waterloo, Canada

<sup>2</sup>RCSDS, NCMIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China

12th April 2019

## Abstract

In this paper, we characterize compatibility of distributions and probability measures on a measurable space. For a set of indices  $\mathcal{J}$ , we say that the tuples of probability measures  $(Q_i)_{i \in \mathcal{J}}$  and distributions  $(F_i)_{i \in \mathcal{J}}$  are compatible if there exists a random variable having distribution  $F_i$  under  $Q_i$  for each  $i \in \mathcal{J}$ . We first establish an equivalent condition using conditional expectations for general (possibly uncountable)  $\mathcal{J}$ . For a finite  $n$ , it turns out that compatibility of  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  depends on the heterogeneity among  $Q_1, \dots, Q_n$  compared with that among  $F_1, \dots, F_n$ . We show that, under an assumption that the measurable space is rich enough,  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  are compatible if and only if  $(Q_1, \dots, Q_n)$  dominates  $(F_1, \dots, F_n)$  in a notion of heterogeneity order, defined via multivariate convex order between the Radon-Nikodym derivatives of  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  with respect to some reference measures. We then proceed to generalize our results to stochastic processes, and conclude the paper with an application to portfolio selection problems under multiple constraints.

**Keywords:** change of measure, compatibility, heterogeneity order, optimization.

## 1 Introduction

### 1.1 The main problem

Change of probability measures is found ubiquitous in problems where multiple probability measures appear, with extensive theoretical treatment and applications in the fields of probability theory, statistics, decision theory, simulation, and finance.

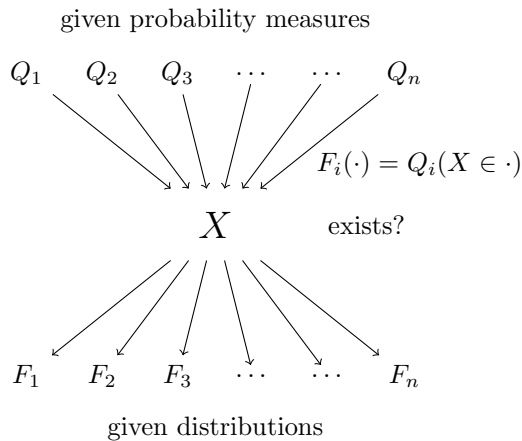
A key feature of a change of measure is that the distribution of a random variable is transformed to another one, and this serves many theoretical as well as practical purposes, such as in

the modification of a Brownian motion drift (e.g. Revuz and Yor [21]) or in importance sampling (e.g. Siegmund [24]; Glasserman and Li [14]). In view of this, a question seems natural to us: *how much* would the distribution change? We formulate this question below.

(A) Given two probability measures  $P$  and  $Q$  defined on the same measurable space  $(\Omega, \mathcal{A})$ , suppose that a random variable  $X : \Omega \rightarrow \mathbb{R}$  has a given distribution function  $F$  under  $P$ . What are the possible distributions of  $X$  under  $Q$ ?

Question (A) arises naturally if one has statistical (distributional) information about a random variable  $X$  under  $P$ , but yet she is concerned about the behaviour of  $X$  under another measure  $Q$ . A general version of question (A), the vocal focus of this paper, is the following.

(B) Given several probability measures  $Q_1, \dots, Q_n$  defined on  $(\Omega, \mathcal{A})$ , and distribution measures  $F_1, \dots, F_n$  on  $\mathbb{R}$ , does there exist a random variable  $X : \Omega \rightarrow \mathbb{R}$  such that  $X$  has distribution  $F_i$  under  $Q_i$  for  $i = 1, \dots, n$ ?



Question (B) is henceforth referred to as the *compatibility* problem for the  $n$ -tuples of measures  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$ . We give an analytical answer to question (B), and hence (A). More generally, we also address the compatibility of two infinite collections of measures.

Before describing our findings, let us look at a few intuitive cases of (B). Suppose that  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  are compatible, that is, (B) has an affirmative answer. In case that  $Q_1, \dots, Q_n$  are identical, it is clear that the respective distributions of a random variable under each  $Q_i$ ,  $i = 1, \dots, n$  are the same; thus  $F_1 = \dots = F_n$ . In case that  $Q_1, \dots, Q_n$  are mutually singular, the respective distributions of a random variable under  $Q_i$ ,  $i = 1, \dots, n$  can be arbitrary. In case that  $F_1, \dots, F_n$  are mutually singular measures on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ,  $Q_1, \dots, Q_n$  have to be also mutually singular. From the above observations, it then seems natural to us that

whether  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  are compatible depends on the *heterogeneity* (in some sense) among  $Q_1, \dots, Q_n$  compared to that of  $F_1, \dots, F_n$ . More precisely,  $Q_1, \dots, Q_n$  need to be *more heterogeneous* than  $F_1, \dots, F_n$  to allow for compatibility.

To describe the above heterogeneity mathematically, we seek help from a notion of heterogeneity order. It turns out that compatibility of  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  is closely related to multivariate convex order between the Radon-Nikodym derivatives  $(\frac{dF_1}{dF}, \dots, \frac{dF_n}{dF})$  and  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$ , where  $F$  and  $Q$  are two “reference probability measures” on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and  $(\Omega, \mathcal{A})$ , respectively. In particular, we show that question (B) has an affirmative answer only if for some measures  $F$  dominating  $(F_1, \dots, F_n)$  and  $Q$  dominating  $(Q_1, \dots, Q_n)$ ,

$$\int_{\mathbb{R}} f \left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) dF \leq \int_{\Omega} f \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) dQ$$

for all convex functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Furthermore, if the measurable space  $(\Omega, \mathcal{A})$  is rich enough, the above necessary condition is sufficient for a positive answer to (B). We then proceed to generalize our results to stochastic processes, and conclude the paper with an optimization problem related to compatibility of distributions under change of measures.

## 1.2 Relation to finance and economics

The main objective of this paper, question (B), has several deep connections to fundamental problems in finance and economics. We summarize some notable relevant points below<sup>1</sup>.

**Risk assessment under multiple scenarios.** In the evaluation of capital requirement for market risks, one often needs to assess risk models under different probability measures, e.g. stressed and non-stressed scenarios. The evaluation of a risk would then be a combination of distributions obtained under various scenarios. For a theoretical treatment of this approach and its relation to the Fundamental Review of the Trading Book, we refer to Wang and Ziegel [27]. A natural question in this context is whether one can find a risk model (represented by a random variable or a stochastic process) that has specified distributions under corresponding scenarios. For instance, one may be interested in simulating from a risk model which has a specific dynamic under a non-stressed scenario and another dynamic (e.g. with different parameters) under a stressed scenario. The existence of such a risk model is precisely question (B); see Section 4.2 for results on Brownian motions. For some other questions in the same spirit, we refer to Embrechts et al. [11, 12] where the authors address a few questions on the existence of certain models satisfying given constraints, which are raised by practitioners from the financial industry. In Section 5, we present a portfolio selection problem with constraints under multiple scenarios.

---

<sup>1</sup>We thank Marcel Nutz for suggesting the second and the third connections, and Fabio Maccheroni for helpful discussions leading to the fourth connection.

**Simultaneous mass transport.** By definition, question (B) is equivalent to the existence of a Monge mass transport from  $Q_i$  to  $F_i$  for all  $i = 1, \dots, n$  simultaneously. Optimal mass transport is an active topic with various applications in mathematical finance, in particular in the calculation of model-independent bounds; we refer to Hobson [16], Beiglböck et al. [5, 7] and Beiglböck and Juille [6] for recent advances. In the study of optimal transport, one typically looks at an optimal transport for one pair of measures. The existence of such a transport is trivial for one pair of measures, which corresponds to question (B) for  $n = 1$ . In this paper, we deal with the case  $n > 1$ , and thus *simultaneous mass transport*. Existence is no longer a trivial issue, and it has to be studied before one could discuss optimality. Admittedly, we are not aware of immediate applications of simultaneous mass transport in finance. Nevertheless, this paper serves as a starting point for future studies in this direction.

**Task assignment problem.** The third connection is a classic task assignment problem in economics. Suppose that  $\Omega$  represents a finite set of workers. Each worker  $\omega \in \Omega$  has some resources of several skills, represented by real numbers  $p_1(\omega), \dots, p_n(\omega)$ . One needs to assign workers to stations (represented by real numbers) where each station demands each skill at a specified amount, and each amount integrates to the corresponding total available resources  $\int_{\Omega} p_i(\omega) d\omega$ ,  $i = 1, \dots, n$ . The problem is whether there exists a way to assign the workers so that each station has exactly the amount of skills it demands. The continuous version of this problem is precisely question (B), where  $p_i$  represents the density of the probability measure  $Q_i$ ,  $i = 1, \dots, n$ .

**Consequentialism in decision theory.** The fourth connection is found in decision theory. An Anscombe-Aumann act (Anscombe and Aumann [1]) is a vector of distributions, resulting from a lottery (a random variable) under a set of beliefs (a collection of probability measures  $Q_1, \dots, Q_n$ ). Question (B) is equivalent to the existence of a given Anscombe-Aumann act for a pre-specified set of beliefs. Decision theorists often study preferences over the set of all Anscombe-Aumann acts without specifying the measures  $Q_1, \dots, Q_n$ , and this is referred to as an axiom of *consequentialism* (see e.g. Battigalli et al. [2]). Such an approach assumes that any choice of an act always exists, which is guaranteed by assuming the mutual singularity of  $(Q_1, \dots, Q_n)$  (Proposition 3.7 (iv) and Theorem 3.17). However, mutual singularity is not the case for many parametric models of beliefs. As such, the results in our paper are helpful to a better understanding of the decision-theoretical framework of consequentialism.

### 1.3 Notation

Throughout, we work with a fixed measurable space  $(\Omega, \mathcal{A})$ , which allows for atomless probability measures. A probability measure  $Q$  on  $(\Omega, \mathcal{A})$  is said to be atomless if for all  $A \in \mathcal{A}$

with  $Q(A) > 0$ , there exists  $B \in \mathcal{A}$ ,  $B \subset A$  such that  $0 < Q(B) < Q(A)$ . Equivalently, there exists a random variable in  $(\Omega, \mathcal{A})$  that is continuously distributed under  $Q$ . Let  $\mathcal{F}$  be the set of probability measures on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , where  $\mathcal{B}(\mathbb{R})$  stands for the Borel  $\sigma$ -algebra of  $\mathbb{R}$ , and  $\mathcal{M}_1 = \mathcal{M}_1(\Omega)$  be the set of probability measures on  $(\Omega, \mathcal{A})$ . Let  $\mathcal{L}^0(\Omega; \mathcal{Y})$  be the set of all measurable functions from  $\Omega$  to  $\mathcal{Y}$ , when the corresponding  $\sigma$ -fields are clear. For any measures  $Q, Q_1, \dots, Q_n$ , we say that  $Q$  dominates  $(Q_1, \dots, Q_n)$ , denoted by  $(Q_1, \dots, Q_n) \ll Q$ , if  $Q$  dominates  $Q_i$  for each  $i = 1, \dots, n$ .

## 2 Compatibility and an equivalent condition

We first define the main concept of this paper, compatibility problem for the two groups of measures  $(Q_i)_{i \in \mathcal{J}} \subseteq \mathcal{M}_1$  and  $(F_i)_{i \in \mathcal{J}} \subseteq \mathcal{F}$ , where  $\mathcal{J}$  is a possibly infinite set of indices.

**Definition 2.1.**  $(Q_i)_{i \in \mathcal{J}} \subseteq \mathcal{M}_1$  and  $(F_i)_{i \in \mathcal{J}} \subseteq \mathcal{F}$  are *compatible* if there exists a random variable  $X$  in  $(\Omega, \mathcal{A})$  such that  $F_i$  is the distribution of  $X$  under  $Q_i$  for each  $i \in \mathcal{J}$ .

We note that  $F$  is the distribution of  $X$  under  $Q$  if and only if  $F = Q \circ X^{-1}$ . Below we establish our first result, which leads to an equivalent condition for compatibility of  $(Q_i)_{i \in \mathcal{J}} \subseteq \mathcal{M}_1$  and  $(F_i)_{i \in \mathcal{J}} \subseteq \mathcal{F}$ .

**Theorem 2.2.** For  $(Q_i)_{i \in \mathcal{J}} \subseteq \mathcal{M}_1$ ,  $(F_i)_{i \in \mathcal{J}} \subseteq \mathcal{F}$  and  $X \in \mathcal{L}^0(\Omega; \mathbb{R})$ , assuming that there exists a probability measure in  $\mathcal{M}_1$  dominating  $(Q_i)_{i \in \mathcal{J}}$ , equivalent are:

(i)  $X$  has distribution  $F_i$  under  $Q_i$  for  $i \in \mathcal{J}$ .

(ii) For all  $Q \in \mathcal{M}_1$  dominating  $(Q_i)_{i \in \mathcal{J}}$ , the probability measure  $F = Q \circ X^{-1}$  dominates  $(F_i)_{i \in \mathcal{J}}$ , and for all  $i \in \mathcal{J}$ ,

$$\frac{dF_i}{dF}(X) = \mathbb{E}^Q \left[ \frac{dQ_i}{dQ} \middle| X \right]. \quad (2.1)$$

(iii) For some  $Q \in \mathcal{M}_1$  dominating  $(Q_i)_{i \in \mathcal{J}}$ , the probability measure  $F = Q \circ X^{-1}$  dominates  $(F_i)_{i \in \mathcal{J}}$ , and (2.1) holds.

*Proof.* (i) $\Rightarrow$ (ii): By definition,  $X$  is such that  $Q_i(X \in A) = F_i(A)$  for  $A \in \mathcal{B}(\mathbb{R})$  and  $i \in \mathcal{J}$ . Let  $Q \in \mathcal{M}_1$  such that  $Q_i \ll Q$ ,  $i \in \mathcal{J}$ . For any  $A \in \mathcal{B}(\mathbb{R})$ , if  $F(A) = 0$ , then  $Q(X \in A) = 0$ . Since  $Q_i \ll Q$ ,  $Q_i(X \in A) = F_i(A) = 0$ , we have  $F_i \ll F$  for  $i \in \mathcal{J}$ . We can verify that for any  $A \in \mathcal{B}(\mathbb{R})$  and  $i \in \mathcal{J}$ ,

$$\begin{aligned} \mathbb{E}^Q \left[ \mathbb{1}_{\{X \in A\}} \frac{dQ_i}{dQ} \right] &= Q_i(X \in A) \\ &= F_i(A) = \int_A \frac{dF_i}{dF} dF = \mathbb{E}^Q \left[ \mathbb{1}_{\{X \in A\}} \frac{dF_i}{dF}(X) \right]. \end{aligned}$$

Therefore,

$$\frac{dF_i}{dF}(X) = \mathbb{E}^Q \left[ \frac{dQ_i}{dQ} \middle| X \right], \quad i \in \mathcal{J}.$$

(ii) $\Rightarrow$ (iii): Trivial.

(iii) $\Rightarrow$ (i): Suppose that (2.1) holds and  $F$  dominates  $(F_i)_{i \in \mathcal{J}}$ . One can easily verify that, for all  $A \in \mathcal{B}(\mathbb{R})$  and  $i \in \mathcal{J}$ ,

$$\begin{aligned} \mathbb{E}^{Q_i}[\mathbf{1}_{\{X \in A\}}] &= \mathbb{E}^Q \left[ \mathbf{1}_{\{X \in A\}} \frac{dQ_i}{dQ} \right] = \mathbb{E}^Q \left[ \mathbb{E}^Q \left[ \mathbf{1}_{\{X \in A\}} \frac{dQ_i}{dQ} \middle| X \right] \right] \\ &= \mathbb{E}^Q \left[ \mathbf{1}_{\{X \in A\}} \mathbb{E}^Q \left[ \frac{dQ_i}{dQ} \middle| X \right] \right] \\ &= \mathbb{E}^Q \left[ \mathbf{1}_{\{X \in A\}} \frac{dF_i}{dF}(X) \right] = F_i(A). \end{aligned}$$

Therefore,  $X$  has distribution  $F_i$  under  $Q_i$ ,  $i \in \mathcal{J}$ , thus  $(Q_i)_{i \in \mathcal{J}}$  and  $(F_i)_{i \in \mathcal{J}}$  are compatible.  $\square$

**Remark 2.3.** In the case where the index set  $\mathcal{J} = \{1, \dots, n\}$  is finite, a probability measure  $Q \in \mathcal{M}_1$  dominating  $(Q_1, \dots, Q_n)$  always exists, as we can take, for example,  $Q = \frac{1}{n}(Q_1 + \dots + Q_n)$ . As such, the existence assumption in Theorem 2.2 can be removed when  $\mathcal{J}$  is finite.

From Theorem 2.2, the necessary and sufficient condition of compatibility is the existence of  $X \in \mathcal{L}^0(\Omega; \mathbb{R})$  satisfying (2.1) for some  $Q \in \mathcal{M}_1$  dominating  $(Q_i)_{i \in \mathcal{J}}$ . This condition is not easy to verify in general. In the next sections we explore necessary and sufficient conditions, much easier to verify, based on distributional properties of the random vectors  $(\frac{dF_1}{dF}, \dots, \frac{dF_n}{dF})$  and  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$ , where  $F$  and  $Q$  are some measures dominating  $(F_1, \dots, F_n)$  and  $(Q_1, \dots, Q_n)$  respectively.

**Remark 2.4.** In the special case of  $n = 2$  and  $Q_1 \ll Q_2$ , one can take  $Q = Q_2$  in Theorem 2.2, and the two-dimensional equality in (2.1) reduces to a one-dimensional equality

$$\frac{dF_1}{dF_2}(X) = \mathbb{E}^{Q_2} \left[ \frac{dQ_1}{dQ_2} \middle| X \right].$$

### 3 Characterizing compatibility via heterogeneity order

In this section, we explore analytical conditions for compatibility of  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  based on their Radon-Nikodym derivatives with respect to some reference probability measures, which are much easier to verify than Theorem 2.2.

#### 3.1 Preliminaries on convex order

For an arbitrary probability space  $(\Gamma, \mathcal{S}, P)$ , denote by  $\mathcal{L}^1(\Gamma; \mathbb{R}^n)$  the set of all integrable  $n$ -dimensional random vectors defined on  $(\Gamma, \mathcal{S}, P)$ . Multivariate convex order is a natural notion of heterogeneity order, as defined below.

**Definition 3.1** (Convex order). Let  $(\Omega_1, \mathcal{A}_1, P_1)$  and  $(\Omega_2, \mathcal{A}_2, P_2)$  be two probability spaces. For  $\mathbf{X} \in \mathcal{L}^1(\Omega_1; \mathbb{R}^n)$  and  $\mathbf{Y} \in \mathcal{L}^1(\Omega_2; \mathbb{R}^n)$ , we write  $\mathbf{X}|_{P_1} \preceq_{\text{cx}} \mathbf{Y}|_{P_2}$ , if  $\mathbb{E}^{P_1}[f(\mathbf{X})] \leq \mathbb{E}^{P_2}[f(\mathbf{Y})]$  for all convex functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ .

For more on multi-dimensional convex order, we refer to Müller and Stoyan [19, Chapter 3] and Shaked and Shanthikumar [23, Chapter 7].

For  $\mathbf{X} \in \mathcal{L}^1(\Omega_1; \mathbb{R}^n)$  and  $\mathbf{Y} \in \mathcal{L}^1(\Omega_2; \mathbb{R}^n)$ , we use  $\mathbf{X}|_{P_1} \stackrel{d}{=} \mathbf{Y}|_{P_2}$  to represent that  $\mathbf{X}$  and  $\mathbf{Y}$  have the same distribution under  $P_1$  and  $P_2$  respectively. Clearly, if  $\mathbf{X}|_{P_1} \stackrel{d}{=} \mathbf{Y}|_{P_2}$ , then  $\mathbf{X}|_{P_1} \preceq_{\text{cx}} \mathbf{Y}|_{P_2}$  and  $\mathbf{Y}|_{P_2} \preceq_{\text{cx}} \mathbf{X}|_{P_1}$ . A key feature of convex order is its connection to conditional expectations. Below in Lemma 3.2 we quote Theorem 7.A.1 of Shaked and Shanthikumar [23] for this well-known result (an extension of Strassen's theorem, Strassen [25]); one also finds a slightly simpler formulation as Theorem 3.4.2 of Müller and Stoyan [19]. See also Hirsch et al. [15] for a construction similar to Lemma 3.2 for stochastic processes (termed *peacocks*).

**Lemma 3.2.** *For  $\mathbf{X} \in \mathcal{L}^1(\Omega_1; \mathbb{R}^n)$  and  $\mathbf{Y} \in \mathcal{L}^1(\Omega_2; \mathbb{R}^n)$ ,  $\mathbf{X}|_{P_1} \preceq_{\text{cx}} \mathbf{Y}|_{P_2}$  if and only if there exist a probability space  $(\Omega_3, \mathcal{A}_3, P_3)$  and  $\mathbf{X}', \mathbf{Y}' \in \mathcal{L}^1(\Omega_3; \mathbb{R}^n)$  such that  $\mathbf{X}'|_{P_3} \stackrel{d}{=} \mathbf{X}|_{P_1}$ ,  $\mathbf{Y}'|_{P_3} \stackrel{d}{=} \mathbf{Y}|_{P_2}$ , and  $\mathbb{E}^{P_3}[\mathbf{Y}'|\mathbf{X}'] = \mathbf{X}'$ .*

## 3.2 Heterogeneity order

As mentioned in the introduction, compatibility intuitively concerns the heterogeneity among  $(Q_1, \dots, Q_n)$  compared to  $(F_1, \dots, F_n)$ . The following lemma, based on Theorem 2.2, yields a possible way of characterizing the comparison between the two tuples of measures. More precisely, a necessary condition for compatibility is built on a convex order relation between the random vectors  $(\frac{dF_1}{dF}, \dots, \frac{dF_n}{dF})$  and  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$  for some reference probability measures  $F \in \mathcal{F}$  and  $Q \in \mathcal{M}_1$ .

**Lemma 3.3.** *If  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n$  and  $(F_1, \dots, F_n) \in \mathcal{F}^n$  are compatible, then for any  $Q \in \mathcal{M}_1$  dominating  $(Q_1, \dots, Q_n)$ , there exists  $F \in \mathcal{F}$  dominating  $(F_1, \dots, F_n)$ , such that*

$$\left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) \Big|_F \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q. \quad (3.1)$$

Moreover,  $F$  in (3.1) can be taken as  $Q \circ X^{-1}$ , where  $X$  is a random variable with distribution  $F_i$  under  $Q_i$ ,  $i = 1, \dots, n$ .

*Proof.* This lemma is directly obtained from Theorem 2.2 and Lemma 3.2. More precisely, by Theorem 2.2, there exists  $X \in \mathcal{L}^0(\Omega; \mathbb{R})$  such that

$$\left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) (X) = \mathbb{E}^Q \left[ \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big| X \right]$$

where  $F = Q \circ X^{-1}$ . Therefore,

$$\left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) \Big|_F \stackrel{d}{=} \mathbb{E}^Q \left[ \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big| X \right] \Big|_Q \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q,$$

where the last inequality follows from Lemma 3.2 by taking

$$\mathbf{X}' = \mathbb{E}^Q \left[ \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big| X \right], \mathbf{Y}' = \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right), P_3 = Q.$$

□

We summarize the necessary condition in Lemma 3.3 for compatibility by introducing the following heterogeneity order, which is shown to be a partial order in Lemma 3.5 below. In the following,  $\mathcal{M}_1(\Omega_1)$  and  $\mathcal{M}_1(\Omega_2)$  represent the sets of probability measures on two arbitrary measurable spaces  $\Omega_1$  and  $\Omega_2$ , respectively.

**Definition 3.4.**  $(P_1, \dots, P_n) \in \mathcal{M}_1^n(\Omega_1)$  is dominated by  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n(\Omega_2)$  in heterogeneity, denoted by  $(P_1, \dots, P_n) \preceq_{\text{h}} (Q_1, \dots, Q_n)$ , if

$$\left( \frac{dP_1}{dP}, \dots, \frac{dP_n}{dP} \right) \Big|_P \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q \quad (3.2)$$

for some  $P \in \mathcal{M}_1(\Omega_1)$  dominating  $(P_1, \dots, P_n)$  and  $Q \in \mathcal{M}_1(\Omega_2)$  dominating  $(Q_1, \dots, Q_n)$ .

Using the language of heterogeneity order, Lemma 3.3 says that in order for compatibility of  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n$  and  $(F_1, \dots, F_n) \in \mathcal{F}^n$ , a necessary condition is  $(F_1, \dots, F_n) \preceq_{\text{h}} (Q_1, \dots, Q_n)$ . Before discussing the sufficiency of this condition, we first establish some properties of heterogeneity order.

The following lemma implies that the choice of the reference measures  $P$  and  $Q$  in (3.2) is irrelevant; in fact, they can be conveniently chosen as the averages of the corresponding measures.

**Lemma 3.5.** For  $(P_1, \dots, P_n) \in \mathcal{M}_1^n(\Omega_1)$  and  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n(\Omega_2)$ , let  $\mathcal{M}_1^*(\Omega_1) = \{P \in \mathcal{M}_1(\Omega_1) : (P_1, \dots, P_n) \ll P\}$  and  $\mathcal{M}_1^*(\Omega_2) = \{Q \in \mathcal{M}_1(\Omega_2) : (Q_1, \dots, Q_n) \ll Q\}$ . The following are equivalent:

- (i)  $(P_1, \dots, P_n) \preceq_{\text{h}} (Q_1, \dots, Q_n)$ ; that is, (3.2) holds for some  $P \in \mathcal{M}_1^*(\Omega_1)$  and  $Q \in \mathcal{M}_1^*(\Omega_2)$ .
- (ii) For  $P = \frac{1}{n} \sum_{i=1}^n P_i$  and  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ , (3.2) holds.
- (iii) For any  $Q \in \mathcal{M}_1^*(\Omega_2)$ , there exists  $P \in \mathcal{M}_1^*(\Omega_1)$  such that (3.2) holds.

*Proof.* We proceed in the order (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) $\Rightarrow$ (iii).

(iii) $\Rightarrow$ (ii): For  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ , there exists  $P^* \in \mathcal{M}_1^*(\Omega_1)$  such that

$$\left( \frac{dP_1}{dP^*}, \dots, \frac{dP_n}{dP^*} \right) \Big|_{P^*} \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q.$$



Take the convex function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(x_1, \dots, x_n) = (x_1 + \dots + x_n)^2$ . It follows from the definition of convex order that

$$\begin{aligned} \mathbb{E}^{P^*} \left[ \left( \frac{dP_1}{dP^*} + \dots + \frac{dP_n}{dP^*} \right)^2 \right] &\leq \mathbb{E}^Q \left[ \left( \frac{dQ_1}{dQ} + \dots + \frac{dQ_n}{dQ} \right)^2 \right] \\ &= \mathbb{E}^Q[n^2] = n^2. \end{aligned}$$

On the other hand,

$$\mathbb{E}^{P^*} \left[ \frac{dP_1}{dP^*} + \dots + \frac{dP_n}{dP^*} \right] = \mathbb{E}^{P_1}[1] + \dots + \mathbb{E}^{P_n}[1] = n.$$

Hence,  $\frac{dP_1}{dP^*} + \dots + \frac{dP_n}{dP^*}$  has zero variance under  $P^*$ , which implies that it is  $P^*$ -almost surely equal to  $n$ . In other words,  $P^* = \frac{1}{n} \sum_{i=1}^n P_i$  on all sets with positive  $P^*$ -measure. Noting that moreover  $P^*$  dominates  $(P_1, \dots, P_n)$ , we must have  $P^* = \frac{1}{n} \sum_{i=1}^n P_i$ . Therefore, (3.2) holds for  $P = \frac{1}{n} \sum_{i=1}^n P_i$  and  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ .

(ii) $\Rightarrow$ (i): trivial.

(i) $\Rightarrow$ (iii): Assume (3.2) holds for some  $Q \in \mathcal{M}_1^*(\Omega_2)$  and  $P \in \mathcal{M}_1^*(\Omega_1)$ . Let  $\mathbf{Y} = (\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$ ,  $\mathbf{Z} = (\frac{dP_1}{dP}, \dots, \frac{dP_n}{dP})$ . Let  $Q'$  be another probability measure in  $\mathcal{M}_1^*(\Omega_2)$ . First, note that without loss of generality, we can assume that  $Q'$  is dominated by  $Q$ . Indeed, any general probability measure  $Q'$  can be decomposed as  $Q' = cQ'_a + (1-c)Q'_s$ , where  $c \in [0, 1]$ ,  $Q'_a$  and  $Q'_s$  are probability measures being absolutely continuous and singular with respect to  $Q$ , respectively. This implies that the distribution of  $(\frac{dQ_1}{dQ'}, \dots, \frac{dQ_n}{dQ'})$  is a mixture of the distribution of  $c^{-1}(\frac{dQ_1}{dQ'_a}, \dots, \frac{dQ_n}{dQ'_a})$  (with probability  $c$ ) and  $(0, \dots, 0)$  (with probability  $1-c$ ). It is easy to check that such a distribution has a larger convex order than  $(\frac{dQ_1}{dQ'_a}, \dots, \frac{dQ_n}{dQ'_a})$ . Thus, if we show (3.2) for  $Q'_a$ , the result also holds for  $Q'$ . In the sequel we assume  $Q'$  is dominated by  $Q$ , hence the random variable  $X = \frac{dQ'}{dQ}$  is well-defined. Let a set  $A = \{\mathbf{Y} \neq 0\}$ . Note that since  $Q'$  dominates  $(Q_1, \dots, Q_n)$ ,  $X > 0$   $Q$ -almost surely on  $A$ .  $(\frac{dQ_1}{dQ'}, \dots, \frac{dQ_n}{dQ'})$  can be then taken as  $X^{-1}\mathbf{Y}$ , where we define  $X^{-1}\mathbf{Y} = 0$  when both  $X$  and  $\mathbf{Y}$  are 0.

By Lemma 3.2, there exists a probability space  $(\Omega', \mathcal{A}', \eta)$  and random vectors  $\mathbf{Y}', \mathbf{Z}'$ , such that  $\mathbf{Y}'|_\eta \stackrel{d}{=} \mathbf{Y}|_Q$ ,  $\mathbf{Z}'|_\eta \stackrel{d}{=} \mathbf{Z}|_P$ , and  $\mathbb{E}^\eta[\mathbf{Y}'|\mathbf{Z}'] = \mathbf{Z}'$ . Furthermore, we can obviously choose  $(\Omega', \mathcal{A}', \eta)$  to contain a random variable  $X'$  such that  $(X', \mathbf{Y}')|_\eta \stackrel{d}{=} (X, \mathbf{Y})|_Q$ . On  $(\Omega', \mathcal{A}')$ , define a new probability measure  $\eta'$  by  $\frac{d\eta'}{d\eta} = X'$ , then  $(X', \mathbf{Y}')|_{\eta'} \stackrel{d}{=} (X, \mathbf{Y})|_Q$ . For any bounded measurable function  $f$ ,

$$\mathbb{E}^\eta[f(\mathbf{Z}')\mathbf{Z}'] = \mathbb{E}^\eta[f(\mathbf{Z}')\mathbf{Y}'] = \mathbb{E}^\eta \left[ f(\mathbf{Z}') \left( \frac{\mathbf{Y}'}{X'} \right) X' \right] = \mathbb{E}^{\eta'} \left[ f(\mathbf{Z}') \left( \frac{\mathbf{Y}'}{X'} \right) \right],$$

where, again,  $X' = 0$  implies  $\mathbf{Y}' = 0$ , and in this case  $\frac{\mathbf{Y}'}{X'}$  is set to be 0. Hence

$$\begin{aligned}\mathbb{E}^\eta[f(\mathbf{Z}')\mathbf{Z}'] &= \mathbb{E}^{\eta'} \left[ f(\mathbf{Z}') \mathbb{E}^{\eta'} \left[ \frac{\mathbf{Y}'}{X'} \middle| \mathbf{Z}' \right] \right] \\ &= \mathbb{E}^\eta \left[ f(\mathbf{Z}') \mathbb{E}^{\eta'} \left[ \frac{\mathbf{Y}'}{X'} \middle| \mathbf{Z}' \right] X' \right] \\ &= \mathbb{E}^\eta \left[ f(\mathbf{Z}') \mathbb{E}^{\eta'} \left[ \frac{\mathbf{Y}'}{X'} \middle| \mathbf{Z}' \right] \mathbb{E}^\eta[X' | \mathbf{Z}'] \right].\end{aligned}$$

Therefore, we must have

$$\mathbb{E}^{\eta'} \left[ \frac{\mathbf{Y}'}{X'} \middle| \mathbf{Z}' \right] = \frac{\mathbf{Z}'}{\mathbb{E}^\eta[X' | \mathbf{Z}']}$$

$\eta$ -almost surely. Define measure  $P'$  by  $\frac{dP'}{dP}(z) = \mathbb{E}^\eta[X' | \mathbf{Z}' = \mathbf{Z}(z)] =: V(z)$ . Note that since

$$\begin{aligned}\int \frac{dP'}{dP}(z) dP(z) &= \int \mathbb{E}^\eta[X' | \mathbf{Z}' = \mathbf{Z}(z)] dP(z) \\ &= \mathbb{E}^\eta[\mathbb{E}^\eta[X' | \mathbf{Z}']] = \mathbb{E}^\eta[X'] = \mathbb{E}^Q[X] = 1,\end{aligned}$$

$P'$  is a probability measure. Then we have  $(\frac{dP_1}{dP'}, \dots, \frac{dP_n}{dP'}) = \frac{\mathbf{Z}'}{V}$ . Define probability measure  $\eta''$  by  $\frac{d\eta''}{d\eta} = \mathbb{E}^\eta[X' | \mathbf{Z}']$ . Since the relation between  $\mathbf{Z}'$ ,  $\eta$  and  $\eta''$  is in parallel with that between  $\mathbf{Z}$ ,  $P$  and  $P'$ , we have

$$\frac{\mathbf{Z}'}{V} \Big|_{P'} \stackrel{d}{=} \frac{\mathbf{Z}'}{\mathbb{E}^\eta[X' | \mathbf{Z}']} \Big|_{\eta''}.$$

However, for any test function  $g$ ,

$$\begin{aligned}\mathbb{E}^{\eta''}[g(\mathbf{Z}')] &= \int g(\mathbf{Z}') \frac{d\eta''}{d\eta} d\eta \\ &= \int g(\mathbf{Z}') \mathbb{E}^\eta[X' | \mathbf{Z}'] d\eta = \mathbb{E}^\eta[g(\mathbf{Z}')X'] = \mathbb{E}^{\eta'}(g(\mathbf{Z}')), \end{aligned}$$

hence  $\mathbf{Z}'|_{\eta'} \stackrel{d}{=} \mathbf{Z}'|_{\eta''}$ . Thus,  $\frac{\mathbf{Z}'}{\mathbb{E}^\eta[X' | \mathbf{Z}']}$ , as a function of  $\mathbf{Z}'$ , also has the same distribution under  $\eta'$  and  $\eta''$ . Consequently, we have

$$\left( \frac{dP_1}{dP'}, \dots, \frac{dP_n}{dP'} \right) \Big|_{P'} = \frac{\mathbf{Z}'}{V} \Big|_{P'} \stackrel{d}{=} \frac{\mathbf{Z}'}{\mathbb{E}^\eta[X' | \mathbf{Z}']} \Big|_{\eta'}.$$

Also, recalling that  $(X', \mathbf{Y}')|_{\eta'} \stackrel{d}{=} (X, \mathbf{Y})|_{Q'}$ ,

$$\left( \frac{dQ_1}{dQ'}, \dots, \frac{dQ_n}{dQ'} \right) \Big|_{Q'} = \frac{\mathbf{Y}'}{X} \Big|_{Q'} \stackrel{d}{=} \frac{\mathbf{Y}'}{X'} \Big|_{\eta'}.$$

The proof is finished by noting that

$$\begin{aligned}&\mathbb{E}^{\eta'} \left[ \frac{\mathbf{Y}'}{X'} \middle| \frac{\mathbf{Z}'}{\mathbb{E}^\eta[X' | \mathbf{Z}']} \right] \\ &= \mathbb{E}^{\eta'} \left[ \mathbb{E}^{\eta'} \left[ \frac{\mathbf{Y}'}{X'} \middle| \mathbf{Z}' \right] \middle| \frac{\mathbf{Z}'}{\mathbb{E}^\eta[X' | \mathbf{Z}']} \right] \\ &= \frac{\mathbf{Z}'}{\mathbb{E}^\eta[X' | \mathbf{Z}']},\end{aligned}$$

and applying Lemma 3.2 with random vectors  $\frac{\mathbf{Z}'}{\mathbb{E}^\eta[X' | \mathbf{Z}]}$ ,  $\frac{\mathbf{Y}'}{X'}$  and measure  $\eta'$ .  $\square$

**Remark 3.6.** While the definition of heterogeneity order is given simply by using the convex order between the Radon-Nikodym derivatives of the measures, Lemma 3.5 shows that the choice of the reference measures, hence the exact form of the Radon-Nikodym derivatives, does not affect the order. This explains our motivation to introduce the notion of heterogeneity order as a partial order between two groups of measures rather than between two groups of random variables.

Some simple and intuitive properties of heterogeneity order are summarized in the following proposition. These properties justify the term “heterogeneity” in the order  $\preceq_h$ .

**Proposition 3.7.** *For  $(P_1, \dots, P_n) \in \mathcal{M}_1^n(\Omega_1)$  and  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n(\Omega_2)$ , the following holds.*

- (i) *If  $P_1, \dots, P_n$  are identical, then  $(P_1, \dots, P_n) \preceq_h (Q_1, \dots, Q_n)$ .*
- (ii) *If  $Q_1, \dots, Q_n$  are identical, and  $(P_1, \dots, P_n) \preceq_h (Q_1, \dots, Q_n)$ , then  $P_1, \dots, P_n$  are also identical.*
- (iii) *If  $Q_1, \dots, Q_n$  are equivalent, and  $(P_1, \dots, P_n) \preceq_h (Q_1, \dots, Q_n)$ , then  $P_1, \dots, P_n$  are also equivalent.*
- (iv) *If  $Q_1, \dots, Q_n$  are mutually singular, then  $(P_1, \dots, P_n) \preceq_h (Q_1, \dots, Q_n)$ .*
- (v) *If  $P_1, \dots, P_n$  are mutually singular, and  $(P_1, \dots, P_n) \preceq_h (Q_1, \dots, Q_n)$ , then  $Q_1, \dots, Q_n$  are also mutually singular.*

*Proof.* (i) It is straightforward to verify that

$$\left( \frac{dP_1}{dP_1}, \dots, \frac{dP_n}{dP_1} \right) \Big|_{P_1} \stackrel{d}{=} (1, \dots, 1) \Big|_{P_1} \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q$$

for any  $Q \in \mathcal{M}_1(\Omega_2)$  that dominates  $(Q_1, \dots, Q_n)$ . As a result, we have  $(P_1, \dots, P_n) \preceq_h (Q_1, \dots, Q_n)$ .

(ii) By  $(P_1, \dots, P_n) \preceq_h (Q_1, \dots, Q_n)$  and Lemma 3.5, we have

$$\left( \frac{dP_1}{dP}, \dots, \frac{dP_n}{dP} \right) \Big|_P \preceq_{\text{cx}} (1, \dots, 1) \Big|_{Q_1} \tag{3.3}$$

holds for some  $P \in \mathcal{M}_1(\Omega_1)$  dominating  $(P_1, \dots, P_n)$ . By Lemma 3.2, (3.3) further implies  $dP_i/dP = 1$   $P$ -almost surely for  $i = 1, \dots, n$ ; thus  $P_1, \dots, P_n$  are identical.

(iii) Let  $P = \frac{1}{n} \sum_{i=1}^n P_i$  and  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ .  $(P_1, \dots, P_n) \preceq_h (Q_1, \dots, Q_n)$  implies that, for each  $i = 1, \dots, n$ ,

$$\frac{dP_i}{dP} \Big|_P \preceq_{\text{cx}} \frac{dQ_i}{dQ} \Big|_Q.$$

Note that  $Q(dQ_i/dQ = 0) = 0$  as  $Q_1, \dots, Q_n$  are equivalent. By Lemma 3.2, we know  $P(dP_i/dP = 0) = 0$ , which implies  $P \ll P_i$ . Thus,  $P_1, \dots, P_n$  are equivalent.

(iv) As  $Q_1, \dots, Q_n$  are mutually singular, there exists a partition  $\{\Omega_1, \dots, \Omega_n\} \subseteq \mathcal{A}$  of  $\Omega$  such that  $Q_i(\Omega_i) = 1$ ,  $i = 1, \dots, n$ . Let  $P = \frac{1}{n} \sum_{i=1}^n P_i$  and  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ . Note that

$$\left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) = n \times (\mathbf{1}_{\Omega_1}, \dots, \mathbf{1}_{\Omega_n})$$

takes values in the vertices of the simplex

$$S = \left\{ (s_1, \dots, s_n) \in \mathbb{R}_+^n : \sum_{i=1}^n s_i = n \right\},$$

and  $(\frac{dP_1}{dP}, \dots, \frac{dP_n}{dP})$  takes values in  $S$ . Furthermore,

$$\mathbb{E}^P \left[ \left( \frac{dP_1}{dP}, \dots, \frac{dP_n}{dP} \right) \right] = (1, \dots, 1) = \mathbb{E}^Q \left[ \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \right].$$

By the Choquet-Meyer Theorem (Choquet and Meyer [9]; see Section 10 of Phelps [20]), stating that among random vectors distributed in a simplex, the maximal elements with respect to convex order are supported over the vertices of the simplex, we have

$$\left( \frac{dP_1}{dP}, \dots, \frac{dP_n}{dP} \right) \Big|_P \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q.$$

(v) Using the notation in (iv),  $(\frac{dP_1}{dP}, \dots, \frac{dP_n}{dP})$  takes values in the vertices of the simplex  $S$ , and  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$  takes values in  $S$ . Therefore, by the Choquet-Meyer Theorem again, in order for  $(P_1, \dots, P_n) \preceq_h (Q_1, \dots, Q_n)$  to hold,  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$  has to be distributed over the vertices of the simplex  $S$ , and therefore,  $Q_1, \dots, Q_n$  are mutually singular.  $\square$

### 3.3 Almost compatibility

In Section 3.2, we see that a necessary condition for compatibility of  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n$  and  $(F_1, \dots, F_n) \in \mathcal{F}^n$  is  $(F_1, \dots, F_n) \preceq_h (Q_1, \dots, Q_n)$ . A natural question is whether (and with what additional assumptions) the above condition is also sufficient. This boils down (via Theorem 2.2) to the question of, given

$$\left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) \Big|_F \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q,$$

where  $F = \frac{1}{n} \sum_{i=1}^n F_i$  and  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ , constructing a random variable  $X$  with distribution  $F$  under  $Q$  such that

$$\left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) (X) = \mathbb{E}^Q \left[ \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big| X \right]. \quad (3.4)$$

Such problem is similar to Lemma 3.2, and more generally, the martingale construction in Strassen [25] or Hirsch et al. [15], albeit we need to construct  $X$  in the pre-specified space  $(\Omega, \mathcal{A}, Q)$ . Therefore, the existence of  $X$  satisfying (3.4) naturally depends on the probability

space  $(\Omega, \mathcal{A}, Q)$ . As a simple example, if  $F$  is a continuous distribution and one of  $Q_1, \dots, Q_n$  is not atomless, then there does not exist a random variable  $X$  with distribution  $F$  under each of  $Q_1, \dots, Q_n$ , although  $(F, \dots, F) \preceq_h (Q_1, \dots, Q_n)$  by Proposition 3.7 (i).

It seems then natural to assume that each of  $Q_1, \dots, Q_n$  is atomless. Below we give a counter example showing that this condition is still insufficient.

**Example 3.8.** Let  $\Omega = [0, 1]$ ,  $\mathcal{A} = \mathcal{B}([0, 1])$ ,  $Q_2 = \lambda$  be the Lebesgue measure,  $\frac{dQ_1}{dQ_2}(t) = 2t$ ,  $t \in [0, 1]$ ,  $F_2 = \lambda$  on  $[0, 1]$  and  $\frac{dF_1}{dF_2}(x) = |4x - 2|$ ,  $x \in [0, 1]$ . For this setting we have  $(F_1, F_2) \preceq_h (Q_1, Q_2)$  but  $(F_1, F_2)$  and  $(Q_1, Q_2)$  are not compatible. The details of these statements are given in Appendix A.1.

Example 3.8 suggests that the atomless condition, combined with the heterogeneity order  $(F_1, \dots, F_n) \preceq_h (Q_1, \dots, Q_n)$ , is not sufficient for compatibility of  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$ . Nevertheless, in this section we show that, assuming  $Q_1, \dots, Q_n$  are atomless,  $(F_1, \dots, F_n) \preceq_h (Q_1, \dots, Q_n)$  is sufficient for *almost compatibility*, a weaker notion than compatibility, which we introduce below. Denote by  $D_{\text{KL}}(\cdot \parallel \cdot)$  the Kullback-Leibler divergence between probability measures. Recall that  $D_{\text{KL}}(P \parallel Q)$  is defined as  $\int \log(dP/dQ)dP$  for  $P \ll Q$ .

**Definition 3.9.**  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n$  and  $(F_1, \dots, F_n) \in \mathcal{F}^n$  are *almost compatible*, if for any  $\varepsilon > 0$ , there exists a random variable  $X_\varepsilon$  in  $(\Omega, \mathcal{A})$  such that for each  $i = 1, \dots, n$ , the distribution of  $X_\varepsilon$  under  $Q_i$ , denoted by  $F_{i,\varepsilon}$ , is absolutely continuous with respect to  $F_i$ , and satisfies  $D_{\text{KL}}(F_{i,\varepsilon} \parallel F_i) < \varepsilon$ .

The following theorem characterizes almost compatibility via heterogeneity order in Definition 3.4, assuming each of  $Q_1, \dots, Q_n$  is atomless.

**Theorem 3.10.** *Suppose that  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n$ ,  $(F_1, \dots, F_n) \in \mathcal{F}^n$  and each of  $Q_1, \dots, Q_n$  is atomless.  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  are almost compatible if and only if  $(F_1, \dots, F_n) \preceq_h (Q_1, \dots, Q_n)$ .*

The proof of Theorem 3.10 is a bit lengthy, and is postponed to Appendix A.2 of the paper.

**Remark 3.11.** The Kullback-Leibler divergence in Definition 3.9 is not the only possible choice to provide an equivalent condition in Theorem 3.10. Indeed, the condition for necessity can be weakened to the convergence in probability of  $dF_{i,\varepsilon}/dF_i$  to 1 as  $\varepsilon \rightarrow 0$ , by using Fatou's lemma and the fact that a sequence converging in probability has a subsequence converging almost surely; the proof for sufficiency implies results as strong as the uniform convergence of  $dF_{i,\varepsilon}/dF_i$  to 1. Consequently, the Kullback-Leibler divergence used in the definition of the almost compatibility can be replaced by a series of other conditions, including:

- (i)  $dF_{i,\varepsilon}/dF_i \xrightarrow{P} 1$ ;
- (ii)  $dF_{i,\varepsilon}/dF_i \xrightarrow{\text{a.s.}} 1$ ;
- (iii)  $F_{i,\varepsilon}$  converges to  $F_i$  in total variation, and  $F_{i,\varepsilon} \ll F_i$ ;
- (iv) The Rényi divergence of order  $\infty$  between  $F_{i,\varepsilon}$  and  $F_i$  converges to 0 as  $\varepsilon \rightarrow 0$ ,

among others, without altering the result of Theorem 3.10.

Almost compatibility has a practical implication for optimization problems. Suppose that  $Q_1, \dots, Q_n$  are atomless. For optimization problems of the form

$$\sup\{\phi(P \circ Y^{-1}) : Y \in \mathcal{L}^0(\Omega; \mathbb{R}) \text{ has distribution } F_i \text{ under } Q_i, i = 1, \dots, n\},$$

where  $\phi : \mathcal{F} \rightarrow [-\infty, \infty]$  is a functional, it suffices to consider

$$\sup\{\phi(F) : F \in \mathcal{F}, (F_1, \dots, F_n, F) \preceq_h (Q_1, \dots, Q_n, P)\},$$

as long as  $\phi$  is continuous with respect to any of the convergence types listed in Remark 3.11.

### 3.4 Equivalence of heterogeneous order and compatibility

In view of the discussions in Section 3.3,  $(F_1, \dots, F_n) \preceq_h (Q_1, \dots, Q_n)$  is not sufficient for compatibility of  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$ , but sufficient for almost compatibility if each of  $Q_1, \dots, Q_n$  is atomless. In this section, we seek for a slightly stronger condition on the  $n$ -tuple  $(Q_1, \dots, Q_n)$ , under which compatibility and almost compatibility coincide.

**Definition 3.12.**  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n$  is *conditionally atomless* if there exist  $Q \in \mathcal{M}_1$  dominating  $(Q_1, \dots, Q_n)$  and  $X \in \mathcal{L}^0(\Omega; \mathbb{R})$  such that under  $Q$ ,  $X$  is continuously distributed and independent of  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$ .

Clearly, if  $(Q_1, \dots, Q_n)$  is conditionally atomless, then each of  $Q_1, \dots, Q_n$  is atomless, since a continuous random variable under  $Q$  is also continuous under each  $Q_1, \dots, Q_n$ .

**Remark 3.13.** If  $Q_1, \dots, Q_n$  are mutually singular and each of them is atomless, then  $(Q_1, \dots, Q_n)$  is conditionally atomless. This can be seen directly by constructing a uniform random variable  $U_i$  on  $[0, 1]$  under  $Q_i$  for  $i = 1, \dots, n$ , and writing  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ . As  $Q_1, \dots, Q_n$  are mutually singular, there exists a partition  $\{\Omega_1, \dots, \Omega_n\} \subset \mathcal{A}$  of  $\Omega$  such that  $Q_i(\Omega_i) = 1, i = 1, \dots, n$ . Then the random variable  $U = \sum_{i=1}^n U_i \mathbf{1}_{\Omega_i}$  is uniformly distributed and independent of  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$  under  $Q$ .

Before approaching the main results of this section, we recall some basic facts about conditional distributions. For random vectors  $\mathbf{T}$  and  $\mathbf{S}$  defined on a probability space  $(\Omega, \mathcal{A}, P)$  and taking values in  $\mathbb{R}^m$  and  $\mathbb{R}^n$ , respectively, the conditional distribution of  $\mathbf{T}$  given  $\mathbf{S}$  (under  $P$ ), denoted by  $\mathbf{T}|\mathbf{S}$ , is a mapping from  $\mathcal{B}(\mathbb{R}^m) \times \Omega$  to  $\mathbb{R}$ , such that for each  $\omega \in \Omega$ ,  $\mathbf{T}|\mathbf{S}(\cdot, \omega)$  is a probability measure on  $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ , and for each  $A \in \mathcal{B}(\mathbb{R}^m)$ ,  $\mathbf{T}|\mathbf{S}(A, \cdot) = P(\mathbf{T} \in A | \sigma(\mathbf{S}))$   $P$ -almost surely. We write  $\mathbf{T}|\mathbf{S}(\omega)$  for the probability measure  $\mathbf{T}|\mathbf{S}(\cdot, \omega)$ , and  $\mathbf{T}|\mathbf{S}(\omega)_P$  when it is necessary to specify the probability measure  $P$ . Moreover, there exists a version of  $\mathbf{T}|\mathbf{S}$  for which the conditional distribution only depends on the value of  $\mathbf{S}$ , i.e.,  $\mathbf{T}|\mathbf{S}(\omega_1) = \mathbf{T}|\mathbf{S}(\omega_2)$  whenever  $\mathbf{S}(\omega_1) = \mathbf{S}(\omega_2)$ . We will always use this version. For an event  $E \in \mathcal{A}$ , the conditional probability of  $E$  given  $\mathbf{S} = s$ , denoted by  $P(E | \mathbf{S} = s)$ , should be understood as  $P[E | \sigma(\mathbf{S})](\omega)$  for  $\omega$  satisfying  $\mathbf{S}(\omega) = s$ .

With the help of conditional distributions, we first note that the independence in Definition 3.12 is not essential and can be replaced by continuity of the conditional distribution. Moreover, similarly to heterogeneity order, the reference probability measure  $Q$  can always be taken as  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ .

**Proposition 3.14.** *For  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n$ , the following are equivalent:*

- (i)  $(Q_1, \dots, Q_n)$  is conditionally atomless.
- (ii) For  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ , there exists a continuous random variable in  $(\Omega, \mathcal{A})$  independent of  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$  under  $Q$ .
- (iii) There exists  $X \in \mathcal{L}^0(\Omega; \mathbb{R})$  such that for some  $Q \in \mathcal{M}_1$  which dominates  $(Q_1, \dots, Q_n)$  (equivalently, for  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ ), a version of the conditional distribution  $X | \mathbf{Y}$  is everywhere continuous under  $Q$  where  $\mathbf{Y} = (\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$ .

*Proof.* Note that (iii) has two versions: one states the existence of  $Q$  and the other specifies  $Q$ . It is trivial to see that (ii) implies (i) and both versions of (iii). It remains to show (iii)  $\Rightarrow$  (i)  $\Rightarrow$  (ii).

We first show (i)  $\Rightarrow$  (ii). Assume  $(Q_1, \dots, Q_n)$  is conditionally atomless. As a result, there exist  $Q' \in \mathcal{M}_1$  and a random variable  $X$ , such that  $X$  and  $\mathbf{Y} := (\frac{dQ_1}{dQ'}, \dots, \frac{dQ_n}{dQ'})$  are independent under  $Q'$ . For  $i = 1, \dots, n$ ,  $A \in \mathcal{B}(\mathbb{R})$  and  $B \in \mathcal{B}(\mathbb{R}^n)$ ,

$$\begin{aligned} Q_i(X \in A, \mathbf{Y} \in B) &= \mathbb{E}^{Q'} \left[ \frac{dQ_i}{dQ'} \mathbb{1}_{\{X \in A\}} \mathbb{1}_{\{\mathbf{Y} \in B\}} \right] \\ &= \mathbb{E}^{Q'} [\mathbb{1}_{\{X \in A\}}] \mathbb{E}^{Q'} \left[ \frac{dQ_i}{dQ'} \mathbb{1}_{\{\mathbf{Y} \in B\}} \right] = Q'(X \in A) Q_i(\mathbf{Y} \in B). \end{aligned}$$

The independence between  $X$  and  $\mathbf{Y}$  also implies that

$$Q_i(X \in A) = \mathbb{E}^{Q'} \left[ \frac{dQ_i}{dQ'} \mathbb{1}_{\{X \in A\}} \right] = \mathbb{E}^{Q'} \left[ \frac{dQ_i}{dQ'} \right] \mathbb{E}^{Q'} [\mathbb{1}_{\{X \in A\}}] = Q'(X \in A).$$

Thus,  $X$  has the same distribution under  $Q_i$ ,  $i = 1, \dots, n$ . Let  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ , and note that  $X$  also has the same distribution under  $Q$ . Moreover,

$$Q_i(X \in A, \mathbf{Y} \in B) = Q'(X \in A)Q_i(\mathbf{Y} \in B) = Q_i(X \in A)Q_i(\mathbf{Y} \in B),$$

which means that  $X$  and  $\mathbf{Y}$  are independent under  $Q_i$  for  $i = 1, \dots, n$ . For any  $A \in \mathcal{B}(\mathbb{R})$  and  $B \in \mathcal{B}(\mathbb{R}^n)$ ,

$$\begin{aligned} Q(X \in A, \mathbf{Y} \in B) &= \frac{1}{n} \sum_{i=1}^n Q_i(X \in A, \mathbf{Y} \in B) \\ &= \frac{1}{n} \sum_{i=1}^n Q_i(X \in A)Q_i(\mathbf{Y} \in B) \\ &= Q(X \in A) \frac{1}{n} \sum_{i=1}^n Q_i(\mathbf{Y} \in B) = Q(X \in A)Q(\mathbf{Y} \in B), \end{aligned}$$

and hence  $X$  and  $\mathbf{Y}$  are independent under  $Q$ . As a result,  $X$  is also independent of

$$\frac{\mathbf{Y}}{\|\mathbf{Y}\|_1} = \frac{1}{n} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right)$$

under  $Q$ , where  $\|\cdot\|_1$  is the Manhattan norm on  $\mathbb{R}^n$ . Therefore, we conclude that  $X$  and  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$  are independent under  $Q$ .

Next we prove (iii) $\Rightarrow$ (i). Take  $X$  as in (iii) and let  $F_\omega$  be the distribution function of  $X|\mathbf{Y}(\omega)$ , and define  $X' : \Omega \rightarrow \mathbb{R}$  by  $X'(\omega) = F_\omega(X(\omega))$ . It is fundamental, though a bit lengthy, to check that  $X'$  is a random variable; moreover,  $X'|\mathbf{Y}$  almost surely follows a uniform distribution on  $[0, 1]$ . As a result,  $X'$  is a continuous random variable independent of  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$  under  $Q$ . Consequently, both versions of (iii) imply (i).  $\square$

**Remark 3.15.** As a byproduct of the above proof, we note that if a random variable  $X$  is independent of  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$  under a probability measure  $Q$ , then  $X$  is also independent of  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$  under each of  $Q_1, \dots, Q_n$ . Moreover,  $X$  has the same distribution under  $Q_1, \dots, Q_n$  and  $Q$ .

**Remark 3.16.** Right before the publication of this paper, a new preprint Delbaen [10] introduces the concept of a conditionally atomless  $\sigma$ -field which turns out to be closely related to our notion of conditionally atomless measures in Definition 3.12. For the connection and the differences between the two formulations, see the discussions in Delbaen [10]<sup>2</sup>.

Now we turn back to our main target, compatibility of  $(F_1, \dots, F_n)$  and  $(Q_1, \dots, Q_n)$ . As discussed in Section 3.3, to show compatibility one needs to construct a random variable  $X$  in  $(\Omega, \mathcal{A})$  such that

$$\left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) (X) = \mathbb{E}^Q \left[ \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \middle| X \right].$$

---

<sup>2</sup>We thank Freddy Delbaen for pointing out the preprint and for very useful discussions.



It turns out that the assumption that  $(Q_1, \dots, Q_n)$  is conditionally atomless allows for such a construction.

**Theorem 3.17.** *Suppose that  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n$  is conditionally atomless and  $(F_1, \dots, F_n) \in \mathcal{F}^n$ .  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  are compatible if and only if  $(F_1, \dots, F_n) \preceq_h (Q_1, \dots, Q_n)$ .*

The key step to prove Theorem 3.17 is the following lemma, which might be of independent interest.

**Lemma 3.18.** *Let  $\mathbf{X} = (X_1, \dots, X_m)$  and  $\mathbf{Y} = (Y_1, \dots, Y_n)$  be random vectors defined on probability spaces  $(\Omega_1, \mathcal{A}_1, P_1)$  and  $(\Omega_2, \mathcal{A}_2, P_2)$ , respectively, and  $f$  be a measurable function from  $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$  to  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ . If the convex order relation  $f(\mathbf{X})|_{P_1} \preceq_{\text{cx}} \mathbf{Y}|_{P_2}$  holds, and there exists a continuous random variable  $U$  defined on  $(\Omega_2, \mathcal{A}_2, P_2)$  independent of  $\mathbf{Y}$ , then there exists a random vector  $\mathbf{W} = (W_1, \dots, W_m)$  defined on  $(\Omega_2, \mathcal{A}_2, P_2)$ , such that  $\mathbf{W}|_{P_2} \stackrel{d}{=} \mathbf{X}|_{P_1}$ , and*

$$f(\mathbf{W}) = \mathbb{E}^{P_2}[\mathbf{Y}|\mathbf{W}].$$

*Proof.* Since  $f(\mathbf{X})|_{P_1} \preceq_{\text{cx}} \mathbf{Y}|_{P_2}$ , by Lemma 3.2, there exists a probability space  $(\Omega', \mathcal{A}', P')$  and random vectors  $\mathbf{Z}, \mathbf{Y}'$  defined on it and taking values in  $\mathbb{R}^n$ , such that  $\mathbf{Z}|_{P'} \stackrel{d}{=} f(\mathbf{X})|_{P_1}$ ,  $\mathbf{Y}'|_{P'} \stackrel{d}{=} \mathbf{Y}|_{P_2}$ , and  $\mathbf{Z} = \mathbb{E}^{P'}[\mathbf{Y}'|\mathbf{Z}]$ .

Construct random vectors  $\mathbf{X}'' = (X_1'', \dots, X_m'')$  and  $\mathbf{Y}'' = (Y_1'', \dots, Y_n'')$  on a (possibly different) probability space  $(\Omega'', \mathcal{A}'', P'')$ , such that  $\mathbf{X}''|_{P''} \stackrel{d}{=} \mathbf{X}|_{P_1}$  and the conditional distributions satisfy  $\mathbf{Y}''|\mathbf{X}''(\omega'')_{P''} = \mathbf{Y}'|\mathbf{Z}(\omega')_{P'}$  for all  $\omega', \omega''$  satisfying  $\mathbf{Z}(\omega') = f(\mathbf{X}''(\omega''))$ . It is easy to see that  $\mathbf{Y}''|_{P''} \stackrel{d}{=} \mathbf{Y}|_{P_2}$ , and

$$\mathbb{E}^{P''}[\mathbf{Y}''|\mathbf{X}''](\omega'') = \mathbb{E}^{P'}[\mathbf{Y}'|\mathbf{Z}](\omega') = \mathbf{Z}(\omega') = f(\mathbf{X}''(\omega'')), \quad \text{for } P''\text{-a.s. } \omega'' \in \Omega''.$$

What is left is therefore to construct a random vector  $\mathbf{W}$  on  $(\Omega_2, \mathcal{A}_2, P_2)$  such that  $(\mathbf{W}, \mathbf{Y})|_{P_2} \stackrel{d}{=} (\mathbf{X}'', \mathbf{Y}'')|_{P''}$ . The idea is similar to the proof of Theorem 3.10. Indeed, for  $\ell = 0, 1, \dots$  and  $h = (h_1, \dots, h_m) \in \mathbb{Z}^m$ , consider the distribution of  $\mathbf{Y}''$  restricted on the event  $\{X_i'' \in [h_i 2^{-\ell}, (h_i + 1) 2^{-\ell}], 1 \leq i \leq m\}$ . It has a density function, denoted by  $\psi_{\ell, h}(\mathbf{y})$ ,  $\mathbf{y} \in \mathbb{R}^n$ , with respect to the unconditional distribution of  $\mathbf{Y}''$ . Without loss of generality, assume  $U$  follows a uniform distribution on  $[0, 1]$ . Then for each  $\mathbf{y}$  and  $\ell = 0, 1, \dots$ , we divide  $[0, 1]$  into disjoint intervals  $\{I_{\ell, h}(\mathbf{y})\}_{h \in \mathbb{Z}^m}$ , such that  $|I_{\ell, h}(\mathbf{y})| = \psi_{\ell, h}(\mathbf{y})$ . Moreover, we can make  $\{I_{\ell', h}(\mathbf{y})\}_{h \in \mathbb{Z}^m}$  a refinement of  $\{I_{\ell, h}(\mathbf{y})\}_{h \in \mathbb{Z}^m}$  for any  $\ell' > \ell$ . Then define random vector  $\mathbf{W}_\ell = (W_{\ell, 1}, \dots, W_{\ell, m})$  by

$$W_{\ell, i} = h_i 2^{-\ell} \text{ for } U \in I_{\ell, h}(\mathbf{Y}), \quad i = 1, \dots, m.$$

Let  $\mathbf{W} = \lim_{\ell \rightarrow \infty} \mathbf{W}_\ell$ . The point-wise limit exists due to the completeness of  $\mathbb{R}^m$ .

For any given  $\mathbf{y}$ , any  $\ell = 0, 1, \dots$  and  $h \in \mathbb{Z}^m$ ,

$$\begin{aligned} & P_2(\mathbf{W}_i \in [h_i 2^{-\ell}, (h_i + 1) 2^{-\ell}), 1 \leq i \leq m | \mathbf{Y} = \mathbf{y}) \\ &= P_2(\mathbf{W}_{\ell, i} = h_i 2^{-\ell}, 1 \leq i \leq m | \mathbf{Y} = \mathbf{y}) \\ &= \psi_{\ell, h}(\mathbf{y}) = P''(\mathbf{X}_i'' \in [h_i 2^{-\ell}, (h_i + 1) 2^{-\ell}), 1 \leq i \leq m | \mathbf{Y}'' = \mathbf{y}). \end{aligned}$$

Since  $\{[h_i 2^{-\ell}, (h_i + 1) 2^{-\ell})\}_{h \in \mathbb{Z}^m, \ell=0,1,\dots}$  forms a basis for  $\mathcal{B}(\mathbb{R}^m)$ , we have  $\mathbf{W} | \mathbf{Y}(\omega)$  under  $P_2$  equals  $\mathbf{X}'' | \mathbf{Y}''(\omega'')$  for any  $\omega \in \Omega$  and  $\omega'' \in \Omega''$  satisfying  $\mathbf{Y}(\omega) = \mathbf{Y}''(\omega'')$ . Moreover, recall that  $\mathbf{Y} |_{P_2} \stackrel{d}{=} \mathbf{Y}'' |_{P''}$ . As a result, we conclude that  $(\mathbf{W}, \mathbf{Y}) |_{P_2} \stackrel{d}{=} (\mathbf{X}'', \mathbf{Y}'') |_{P''}$ .  $\square$

*Proof of Theorem 3.17.* Necessity is guaranteed by Lemma 3.3. We only show sufficiency. Suppose that  $(F_1, \dots, F_n) \preceq_h (Q_1, \dots, Q_n)$ . We shall show that  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  are compatible. By Lemma 3.5,

$$\left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) \Big|_F \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q$$

for  $F = \frac{1}{n} \sum_{i=1}^n F_i$  and  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ . Since  $(Q_1, \dots, Q_n)$  is conditionally atomless,  $Q_1, \dots, Q_n$  are all atomless, so is  $Q$ . Hence there exists a random variable  $X'$  defined on  $(\Omega, \mathcal{A})$ , such that  $F = Q \circ X'^{-1}$ . As a result,

$$\left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) (X') \Big|_Q \stackrel{d}{=} \left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) \Big|_F \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q.$$

Applying Lemma 3.18 with  $f(x) = \left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right)(x)$ , there exists a random variable  $X$  defined on  $(\Omega, \mathcal{A})$ , such that

$$\left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) (X) = \mathbb{E}^Q \left[ \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big| X \right],$$

which, by Theorem 2.2, implies compatibility.  $\square$

**Remark 3.19.** As shown in Theorem 3.17, compatibility is closely related to heterogeneity order  $\preceq_h$ , and hence it defines a partial order. The direction of the order comes from the fact that a measurable mapping needs not to be a bijection. As multiple points are mapped to a same image, the ‘‘heterogeneity’’ between measures decreases. However, if we require the mapping to be a bijection, then compatibility becomes an equivalence relation. Indeed, in this case Theorem 3.17 would be applicable to both directions, which means that (3.2) holds for both directions, with  $P = \frac{1}{n} \sum_{i=1}^n P_i$  and  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ . As a result, we must have

$$\left( \frac{dP_1}{dP}, \dots, \frac{dP_n}{dP} \right) \Big|_P \stackrel{d}{=} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q.$$

Moreover, the proof of Theorem 3.17 actually shows that, assuming both tuples of measures are conditionally atomless, the above condition is not only necessary but also sufficient to guarantee the existence of a bijection linking  $(P_1, \dots, P_n)$  to  $(Q_1, \dots, Q_n)$ .

**Remark 3.20.** As a simple consequence of Theorem 3.17, in the case where  $n = 2$  and  $Q_1 \ll Q_2$ , if  $(Q_1, Q_2)$  and  $(F_1, F_2)$  are compatible, then  $F_1 \ll F_2$  and

$$\frac{dF_1}{dF_2} \Big|_{F_2} \preceq_{\text{cx}} \frac{dQ_1}{dQ_2} \Big|_{Q_2}.$$

The converse is also true if, in addition,  $(Q_1, Q_2)$  is conditionally atomless. Therefore, the heterogeneity order condition becomes one-dimensional, and is easy to check. Chapter 3 of Shaked and Shanthikumar [23] contains several classic methods to check  $X|_P \preceq_{\text{cx}} Y|_Q$  for arbitrary random variables  $X$  and  $Y$  and probability measures  $P$  and  $Q$ .

Below we discuss a few special cases of compatible  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n$  and  $(F_1, \dots, F_n) \in \mathcal{F}^n$  based on the heterogeneity order condition, in particular in the context of Proposition 3.7 and Theorem 3.17. We shall see how our main results are consistent with natural intuitions.

1. Assume that  $Q_1, \dots, Q_n$  are identical. The natural intuition is that the respective distributions  $F_1, \dots, F_n$  of a random variable under  $Q_1, \dots, Q_n$  have to be identical as well. Indeed, by Lemma 3.3, compatibility implies  $(F_1, \dots, F_n) \preceq_{\text{h}} (Q_1, \dots, Q_n)$ . By Proposition 3.7 (ii),  $F_1, \dots, F_n$  are identical.

2. Assume that  $Q_1, \dots, Q_n$  are mutually singular, and each of them is atomless. The natural intuition here is that the respective distributions  $F_1, \dots, F_n$  of any random variable under  $Q_1, \dots, Q_n$  are arbitrary. Proposition 3.7 (iv) suggests that  $(F_1, \dots, F_n) \preceq_{\text{h}} (Q_1, \dots, Q_n)$  holds for any  $(F_1, \dots, F_n) \in \mathcal{F}^n$ . Moreover,  $(Q_1, \dots, Q_n)$  is conditionally atomless, as seen in Remark 3.13. Therefore, by Theorem 3.17, a mutually singular tuple of atomless probability measures on  $(\Omega, \mathcal{A})$  is compatible with an arbitrary tuple of distributions on  $\mathbb{R}$ .

3. Assume that  $F_1, \dots, F_n$  are mutually singular. The natural intuition here is that the probability measures  $Q_1, \dots, Q_n$  have to be also mutually singular to allow for compatibility. Similarly to the previous case, this is justified by Theorem 3.17 and Proposition 3.7 (v).

4. Assume that  $F_1, \dots, F_n$  are identical, and  $(Q_1, \dots, Q_n)$  is conditionally atomless. Proposition 3.7 (i) gives  $(F_1, \dots, F_n) \preceq_{\text{h}} (Q_1, \dots, Q_n)$ . It follows from Theorem 3.17 that  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  are compatible. We conclude that, as long as  $(Q_1, \dots, Q_n)$  is conditionally atomless, for any distribution  $F \in \mathcal{F}$ , there exists a random variable  $X$  which has distribution  $F$  under each of  $Q_i, i = 1, \dots, n$ . Indeed, as  $(Q_1, \dots, Q_n)$  is conditionally atomless, there exists  $Q$  dominating  $(Q_1, \dots, Q_n)$  and an  $F$ -distributed random variable  $X$  under  $Q$  independent of  $(\frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ})$ . Remark 3.15 then implies that  $X$  also has distribution  $F$  under each  $Q_1, \dots, Q_n$ .

5. Assume that  $Q_1, \dots, Q_n$  are equivalent. Intuitively, the respective distributions  $F_1, \dots, F_n$  of any random variable under  $Q_1, \dots, Q_n$  have to be equivalent. This fact is implied by Proposition 3.7 (iii).

**Remark 3.21.** A notion similar to heterogeneity order is useful in comparison of statistical experiments, an area of study originated by Blackwell (Blackwell [3, 4]); the interested reader is referred to Le Cam [17] and Torgersen [26] for summaries.

## 4 Distributional compatibility for stochastic processes

### 4.1 General results

In this section we extend our results to stochastic processes with sample paths which are continuous from right with left limits (càdlàg). For a (finite or infinite) closed interval  $I \subseteq \mathbb{R}$ , let  $D(I)$  be the Skorokhod space on  $I$ , i.e., the space of all càdlàg functions defined on  $I$ . Let  $\mathcal{D}_I$  be the Borel  $\sigma$ -field of the Skorokhod topology  $J_1$ . Denote by  $\mathcal{G}_I = \mathcal{M}_1(D(I))$  the set of probability measures on  $(D(I), \mathcal{D}_I)$ . Our first step is to generalize the definition of compatibility to this setting, which follows in a natural way.

**Definition 4.1.** For a closed interval  $I \subseteq \mathbb{R}$ , we say  $(Q_i)_{i \in \mathcal{J}} \subseteq \mathcal{M}_1$  and  $(G_i)_{i \in \mathcal{J}} \subseteq \mathcal{G}_I$  are *compatible* if there exists a càdlàg stochastic process defined on  $(\Omega, \mathcal{A})$ , denoted by  $X = \{X(t)\}_{t \in I}$ , such that for each  $i \in \mathcal{J}$ , the distribution of  $X$  under  $Q_i$  is  $G_i$ .

The following is a parallel result to Theorem 2.2, which shares the same proof.

**Proposition 4.2.** *Let  $I \subseteq \mathbb{R}$  be a closed interval,  $(Q_i)_{i \in \mathcal{J}} \subseteq \mathcal{M}_1$  and  $(G_i)_{i \in \mathcal{J}} \subseteq \mathcal{G}_I$ . A stochastic process  $X$  has distribution  $G_i$  under  $Q_i$  for  $i \in \mathcal{J}$  if and only if for all  $Q \in \mathcal{M}_1$  dominating  $(Q_i)_{i \in \mathcal{J}}$ ,  $G = Q \circ X^{-1}$  dominates  $(G_i)_{i \in \mathcal{J}}$ , and for all  $i \in \mathcal{J}$ ,*

$$\frac{dG_i}{dG}(X) = \mathbb{E}^Q \left[ \frac{dQ_i}{dQ} \middle| \sigma(X) \right].$$

Then we have, parallel to Theorem 3.17:

**Theorem 4.3.** *Suppose that  $(Q_1, \dots, Q_n) \in \mathcal{M}_1^n$  is conditionally atomless,  $I \subseteq \mathbb{R}$  is a closed interval, and  $(G_1, \dots, G_n) \in \mathcal{G}_I^n$ .  $(Q_1, \dots, Q_n)$  and  $(G_1, \dots, G_n)$  are compatible if and only if  $(G_1, \dots, G_n) \preceq_h (Q_1, \dots, Q_n)$ .*

*Proof.* In the proof of Lemma 3.3, no structure of the real line has been used. As a result, Lemma 3.3 can be directly generalized to the case of stochastic processes, with  $(G_1, \dots, G_n) \in \mathcal{G}_I^n$  replacing  $(F_1, \dots, F_n) \in \mathcal{F}^n$ . For the other direction, the proof is similar to that of Theorem 3.17. The only difference is that  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  is replaced by  $(D(I), \mathcal{D}_I)$ . A careful check of the proofs of Theorem 3.17 and of Lemma 3.18 shows, however, that they only rely on the completely metrizable structure of  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  to guarantee the existence and uniqueness of the limit of the constructed sequence of random variables. Since  $(D(I), \mathcal{D}_I)$  is also completely metrizable, the

proofs naturally extend to the case of stochastic processes. More precisely, order the rational numbers in  $I$  as  $\mathbb{Q} \cap I = \{t_1, t_2, \dots\}$ . Then we replace the refining partition of the real line  $\{[h2^{-\ell}, (h+1)2^{-\ell}), h \in \mathbb{Z}\}_{\ell=0,1,\dots}$  with the refining partition of  $D(I)$ :  $\{X(t_i) \in [h_{\ell,i}2^{-\ell+i}, (h_{\ell,i}+1)2^{-\ell+i}), i = 1, \dots, \ell, h_{\ell,i} \in \mathbb{Z}\}_{\ell=1,2,\dots}$ . The rest follows in the same way as in the proofs of Theorem 3.17 and of Lemma 3.18.  $\square$

**Remark 4.4.** The proof of Theorem 4.3 sheds light upon a more general result, where  $(G_1, \dots, G_n)$  are probability measures defined on a Polish space  $Y$  equipped with the Borel  $\sigma$ -field. In particular, let  $\{y_i\}_{i=1,2,\dots}$  be a dense subset of  $Y$ , then the sequence of partitions

$$\{\{y \in Y : d(y, y_i) \in [h_{\ell,i}2^{-\ell+i}, (h_{\ell,i}+1)2^{-\ell+i}), i = 1, \dots, \ell, h_{\ell,i} = 0, 1, \dots\},$$

$\ell = 0, 1, \dots$ , can be used to replace  $\{[h2^{-\ell}, (h+1)2^{-\ell}), h \in \mathbb{Z}\}_{\ell=0,1,\dots}$  in the proof of Lemma 3.18. The rest follows exactly in the same way as in that proof. Consequently, a general version of Theorem 3.17 can be stated using the setting of a Polish space instead of  $\mathbb{R}$ . However, due to the lack of a natural order and metric as in  $\mathbb{R}$ , a rigorous proof directly for the general case of a Polish space would be notationally heavy and also less intuitive for the readers who are not familiar with Polish spaces. As such, we present Theorem 3.17 under the setting of  $\mathbb{R}$ , which is also the focus of this paper, and use this remark for a discussion for the general setting, after seeing the proof of Theorem 4.3.

## 4.2 Relation to the Girsanov Theorem

In this section we investigate how much the drift of a Brownian motion may vary under a change of measure as in the classic Girsanov Theorem. We keep in mind that, the distribution of a Brownian motion (with respect to its natural filtration) with a deterministic drift process only depends on this drift. On the other hand, Brownian motions with stochastic drift processes are not identified by the distribution of the drift processes. Due to this reason, we consider only Brownian motions with deterministic drift processes here.

Throughout this section, let  $P \in \mathcal{M}_1$  and  $B = \{B_t\}_{t \in [0, T]}$  be a  $P$ -standard Brownian motion. Furthermore, for a  $[0, T]$ -square integrable deterministic process  $\theta = \{\theta_t\}_{t \in [0, T]}$ , define

$$\frac{dQ_\theta}{dP} = e^{\int_0^T \theta_t dB_t - \frac{1}{2} \int_0^T \theta_t^2 dt},$$

and let  $G_\theta$  be the distribution measure of a Brownian motion with drift process  $\theta$ . The Girsanov Theorem says that  $B$  is a Brownian motion with drift process  $\theta$  and volatility 1 under  $Q_\theta$  (certainly, this statement is also true for adapted drift processes). Thus,  $(P, Q_\theta)$  and  $(G_0, G_\theta)$  are compatible. It is clear that distribution measures of Brownian motions with different non-random volatility terms are mutually singular, and hence they are not compatible with  $(P, Q_\theta)$ .

A next question is whether there exists a  $P$ -standard Brownian motion which has a deterministic drift process  $\mu = \{\mu_t\}_{t \in [0, T]}$  under  $Q_\theta$ . We are interested in the values of  $\mu$  such that  $(G_0, G_\mu)$  and  $(P, Q_\theta)$  above are compatible. Here we do not assume that  $(P, Q_\theta)$  is conditionally atomless, which means that there might not be any random source other than  $B$ .

**Theorem 4.5.** *Suppose that the deterministic processes  $\theta = \{\theta_t\}_{t \in [0, T]}$  and  $\mu = \{\mu_t\}_{t \in [0, T]}$  are  $[0, T]$ -square integrable, and  $\mu_t \neq 0$  almost everywhere on  $[0, T]$ .  $(P, Q_\theta)$  and  $(G_0, G_\mu)$  are compatible if and only if*

$$\int_0^T \mu_t^2 dt \leq \int_0^T \theta_t^2 dt.$$

*Proof.* (i) Necessity. By the Girsanov Theorem, we know that  $(G_0, G_\mu)$  and  $(P, Q_\mu)$  are compatible. Using Proposition 4.2 for  $n = 2$ , we have

$$\frac{dG_\mu}{dG_0}(B) = \mathbb{E} \left[ \frac{dQ_\mu}{dP} \middle| \sigma(B) \right] = e^{\int_0^T \mu_t dB_t - \frac{1}{2} \int_0^T \mu_t^2 dt}.$$

Suppose that  $(P, Q_\theta)$  and  $(G_0, G_\mu)$  are compatible. Note that

$$e^{\int_0^T \mu_t dB_t - \frac{1}{2} \int_0^T \mu_t^2 dt} \Big|_P \stackrel{d}{=} \frac{dG_\mu}{dG_0}(B) \Big|_P \stackrel{d}{=} \frac{dG_\mu}{dG_0} \Big|_{G_0}.$$

By Theorem 4.3, we have

$$e^{\int_0^T \mu_t dB_t - \frac{1}{2} \int_0^T \mu_t^2 dt} \Big|_P \preceq_{\text{cx}} \frac{dQ_\theta}{dP} \Big|_P \stackrel{d}{=} e^{\int_0^T \theta_t dB_t - \frac{1}{2} \int_0^T \theta_t^2 dt} \Big|_P.$$

Applying the convex function  $x \mapsto x^2$ , we have

$$e^{\int_0^T \mu_t^2 dt} = \mathbb{E}[(e^{\int_0^T \mu_t dB_t - \frac{1}{2} \int_0^T \mu_t^2 dt})^2] \leq \mathbb{E}[(e^{\int_0^T \theta_t dB_t - \frac{1}{2} \int_0^T \theta_t^2 dt})^2] = e^{\int_0^T \theta_t^2 dt}$$

and hence  $\int_0^T \mu_t^2 dt \leq \int_0^T \theta_t^2 dt$ .

(ii) Sufficiency. Suppose  $\int_0^T \mu_t^2 dt \leq \int_0^T \theta_t^2 dt$ . Define a deterministic process  $\alpha = \{\alpha_t\}_{t \in [0, T]}$  by

$$\alpha_t = \inf \left\{ r \geq 0 : \int_0^r \theta_s^2 ds = \int_0^t \mu_s^2 ds \right\}.$$

It is easy to see that  $\alpha_t$  is strictly increasing in  $t$ ,  $\alpha_T \leq T$ , and furthermore,

$$\theta_{\alpha_t}^2 d\alpha_t = \mu_t^2 dt. \tag{4.1}$$

Let a stochastic process  $\hat{B} = \{\hat{B}_t\}_{t \in [0, T]}$  be given by  $d\hat{B}_t = dB_t - \theta_t dt$ . By the Girsanov Theorem,  $\hat{B}$  is a  $Q_\theta$ -standard Brownian motion. Define

$$W_t = \int_0^t \beta_{\alpha_s} dB_{\alpha_s}, \quad t \in [0, T],$$

where  $\beta = \{\beta_s\}_{s \in [0, \alpha_T]}$  is given by  $\beta_{\alpha_t} = \frac{\theta_{\alpha_t}}{\mu_t}$ ,  $t \in [0, T]$ .  $W = \{W_t\}_{t \in [0, T]}$  is clearly a Gaussian process,  $\mathbb{E}^P[W_t] = 0$ , and

$$\mathbb{E}^P[W_t W_s] = \mathbb{E}^P[W_s^2] = \int_0^s \frac{\theta_{\alpha_u}^2}{\mu_u^2} d\alpha_u = s, \quad 0 \leq s < t \leq T.$$

Therefore,  $W$  is a  $P$ -standard Brownian motion. Furthermore, for  $t \in [0, T]$ ,

$$\begin{aligned} W_t &= \int_0^t \beta_{\alpha_s} dB_{\alpha_s} = \int_0^t \beta_{\alpha_s} (d\hat{B}_{\alpha_s} + \theta_{\alpha_s} d\alpha_s) \\ &= \int_0^t \beta_{\alpha_s} d\hat{B}_{\alpha_s} + \int_0^t \beta_{\alpha_s} \theta_{\alpha_s} d\alpha_s \\ &= \int_0^t \beta_{\alpha_s} d\hat{B}_{\alpha_s} + \int_0^t \mu_s ds, \end{aligned}$$

where the last equality is due to (4.1). As  $\int_0^t \beta_{\alpha_s} d\hat{B}_{\alpha_s}$  defines a  $Q_\theta$ -standard Brownian motion, we conclude that  $W$  has distribution  $G_\mu$  under  $Q_\theta$ , and hence  $(P, Q_\theta)$  and  $(G_0, G_\mu)$  are compatible.  $\square$

We list Theorem 4.5 for the case of a constant drift term below, and look more closely at the construction of the desired stochastic process.

**Corollary 4.6.** *Let  $\theta_t = a$  and  $\mu_t = b$ ,  $t \in [0, T]$ , where  $a, b$  are two constants, and  $b \neq 0$ .  $(P, Q_\theta)$  and  $(G_0, G_\mu)$  are compatible if and only if  $b^2 \leq a^2$ .*

If  $b^2 \leq a^2$ , the process which has distribution  $G_0$  under  $P$  and distribution  $G_\mu$  under  $Q_\theta$  can be written in a simple explicit form. Let

$$W_t = \frac{a}{b} B_{(\frac{b}{a})^2 t}, \quad t \in [0, T].$$

It is clear that  $W = \{W_t\}_{t \in [0, T]}$  is a  $P$ -Brownian motion. Furthermore,

$$W_t = \frac{a}{b} B_{(\frac{b}{a})^2 t} = \frac{a}{b} \left( \hat{B}_{(\frac{b}{a})^2 t} + a \frac{b^2}{a^2} t \right) = \frac{a}{b} \hat{B}_{(\frac{b}{a})^2 t} + bt, \quad t \in [0, T].$$

In this example, it is clear that  $0 < b^2 \leq a^2$  is essential; otherwise  $W$  will not be well-defined.

## 5 Application to a portfolio optimization problem

Let  $P, Q, R$  be three probability measures defined on  $(\Omega, \mathcal{A})$ , and  $F, G$  be two probability distributions on  $\mathbb{R}$ . In this section, we investigate the optimization problem of the type

$$\begin{aligned} &\min_{H \in \mathcal{F}} H([a, \infty)) \\ &\text{s.t. } (P, Q, R) \text{ is compatible with } (F, G, H). \end{aligned} \tag{5.1}$$

Problem (5.1) is motivated by portfolio selection under multiple constraints. In a classic complete-market portfolio selection problem, an investor optimizes an objective function under the physical measure  $P$  (e.g. expected utility) subject to a budget constraint which is evaluated under a risk-neutral measure  $Q$ . For this problem, one obtains an optimal position with loss random variable  $\xi^*$ , and we denote by  $F$  (resp.  $G$ ) the distribution of  $\xi^*$  under  $P$  (resp.  $Q$ ).

The optimal position  $\xi^*$  may not be unique but its distributions under  $P$  and  $Q$  are typically unique (see e.g. Chapter 3 of Föllmer and Schied [13]). We assume, in addition, that there is a regulatory requirement set by a regulator using a measure  $R$  which may not be the same as  $P$  due to extensive usage of stress-testing in calculating regulatory capital (see e.g. Cambou and Filipovic [8]). A typical regulatory requirement is using the Value-at-Risk under stressed scenarios (see e.g. McNeil et al. [18]), that is, a loss random variable  $\xi$  has to satisfy  $R(\xi \geq a) \leq p_0$  where  $a \in \mathbb{R}$  is the capital level of the investor and  $p_0$  is a pre-specified probability level. The investor needs to determine whether a position with her desired distributions under  $P$  and  $Q$  can satisfy this constraint. That is, to determine the existence of a random variable  $\xi$ , such that

$$\xi|_P \sim F, \quad \xi|_Q \sim G, \quad \text{and } R(\xi \geq a) \leq p_0.$$

Using the framework of this paper, this is to determine the existence of a probability distribution  $H$ , such that  $(P, Q, R)$  is compatible with  $(F, G, H)$ , and  $H([a, \infty)) \leq p_0$ . It is obvious that the optimization problem (5.1) directly addresses the above issue.

To study (5.1), we assume that  $(P, Q, R)$  are conditionally atomless and  $Q, R \ll P$ . By Theorem 3.17, the compatibility in (5.1) is equivalent to the heterogeneity order  $(F, G, H) \preceq_h (P, Q, R)$ . An application of Lemma 3.5 (iii) shows this is equivalent to

$$\left( \frac{dF}{dF'}, \frac{dG}{dF'}, \frac{dH}{dF'} \right) \Big|_{F'} \preceq_{\text{cx}} \left( 1, \frac{dQ}{dP}, \frac{dR}{dP} \right) \Big|_P$$

for some  $F'$  such that  $F, G, H \ll F'$ . As  $\frac{dF}{dF'} \Big|_{F'} \preceq_{\text{cx}} 1$ ,  $F'$  must be the same as  $F$ . Hence, an equivalent condition is

$$\left( \frac{dG}{dF}, \frac{dH}{dF} \right) \Big|_F \preceq_{\text{cx}} \left( \frac{dQ}{dP}, \frac{dR}{dP} \right) \Big|_P.$$

For simplicity, we assume  $\frac{dG}{dF}$  is  $F$ -a.e. injective. As such, we can write  $H([a, \infty)) = H(\{t \in \mathbb{R} : \frac{dG}{dF}(t) \in D_a\})$  for some measurable set  $D_a \subseteq \mathbb{R}$ . By Lemma 3.2, the compatibility holds if and only if there exists some probability space  $(\Omega', \mathcal{A}', P')$  and random variables  $X', Y', Z', W'$  defined on that space, such that

$$(X', W')|_{P'} \stackrel{d}{=} \left( \frac{dG}{dF}, \frac{dH}{dF} \right) \Big|_F, \quad (Y', Z')|_{P'} \stackrel{d}{=} \left( \frac{dQ}{dP}, \frac{dR}{dP} \right) \Big|_P,$$

and

$$(X', W') = \mathbb{E}^{P'}[(Y', Z')|X'].$$

The relation  $\mathbb{E}^{P'}[(Y', Z')|X', W'] = \mathbb{E}^{P'}[(Y', Z')|X']$  is used above, which is guaranteed by  $(X', W')|_{P'} \stackrel{d}{=} (\frac{dG}{dF}, \frac{dH}{dF})|_F$  and that  $\frac{dG}{dF}$  is injective. We take  $(\Omega', \mathcal{A}', P')$  as fixed from now on, since only distributions matter in our optimization problem. We have

$$\begin{aligned} H([a, \infty)) &= \mathbb{E}^F \left[ \frac{dH}{dF} \mathbf{1}_{\{\frac{dG}{dF} \in D_a\}} \right] = \mathbb{E}^{P'} [W' \mathbf{1}_{\{X' \in D_a\}}] \\ &= \mathbb{E}^{P'} [Z' \mathbf{1}_{\{X' \in D_a\}}] = \mathbb{E}^{P'} [\mathbb{E}^{P'} [Z' \mathbf{1}_{\{X' \in D_a\}} | Y']]. \end{aligned}$$



Hence, we relax the reliance of  $W'$  in the optimization problem, and (5.1) can be rewritten as

$$\min_{(X', Y', Z')} \mathbb{E}^{P'} [Z' \mathbf{1}_{\{X' \in D_a\}}], \quad (5.2)$$

where the minimum is taken subject to the constraints

$$X'|_{P'} \stackrel{d}{=} \left. \frac{dG}{dF} \right|_F, \quad (Y', Z')|_{P'} \stackrel{d}{=} \left. \left( \frac{dQ}{dP}, \frac{dR}{dP} \right) \right|_P, \quad \text{and} \quad X' = \mathbb{E}^{P'} [Y'|X'].$$

Under  $P'$ , given the joint distribution of  $X'$  and  $Y'$ , the conditional distributions  $X'|Y' = y$  and  $Z'|Y' = y$  are both fixed for  $P'$ -almost every  $y$ . Hence, by the Hardy-Littlewood inequality (in the form of Remark 3.25 of Rüschemdorf [22]), the sub-problem, for fixed  $(X', Y')$ ,

$$\min_{Z' \in K_y} \mathbb{E}^{P'} [Z' \mathbf{1}_{\{X' \in D_a\}} | Y' = y],$$

where  $K_y$  is the set of all random variables  $Z'$  satisfying  $(Z'|Y' = y)|_{P'} \stackrel{d}{=} \left( \frac{dR}{dP} | \frac{dQ}{dP} = y \right)|_P$ , has a simple solution such that  $Z'$  given  $Y' = y$  and  $\mathbf{1}_{\{X' \in D_a\}}$  given  $Y' = y$  are counter-monotonic. Consequently, we have

$$\min_{Z' \in K_y} \mathbb{E}^{P'} [Z' \mathbf{1}_{\{X' \in D_a\}} | Y' = y] = \int_0^{p_{X', Y'}(y)} f(x|y) dx,$$

where  $f(\cdot|y)$  is the left-quantile of the distribution function of  $\frac{dR}{dP}$  given  $\frac{dQ}{dP} = y$  under  $P$ , and

$$p_{X', Y'}(y) = P'(X' \in D_a | Y' = y).$$

Define a function  $\Phi$  on  $(\mathcal{L}^0(\Omega'; \mathbb{R}))^2$  by

$$\Phi(X', Y') = \mathbb{E}^{P'} \left[ \int_0^{p_{X', Y'}(Y')} f(x|Y') dx \right] = \int_0^\infty \int_0^{p_{X', Y'}(y)} f(x|y) dx dF_Y(y),$$

where  $F_Y$  is the distribution of  $\frac{dQ}{dP}|_P$ . Clearly,  $\Phi$  is determined by the joint distribution of  $(X', Y')$  under  $P'$ . By this argument, we relax the reliance of  $Z'$  in the optimization problem (5.2). To summarize, the results in this paper allow us to transform the original optimization problem (5.1) into

$$\min_{(X, Y) \in K} \Phi(X, Y) \quad (5.3)$$

where  $K$  is the set of all random variables  $(X, Y) \in (\mathcal{L}^0(\Omega'; \mathbb{R}))^2$  satisfying

$$X|_{P'} \stackrel{d}{=} \left. \frac{dG}{dF} \right|_F, \quad Y|_{P'} \stackrel{d}{=} \left. \frac{dQ}{dP} \right|_P, \quad \text{and} \quad \mathbb{E}^{P'} [Y|X] = X. \quad (5.4)$$

**Remark 5.1.** Problem (5.3) can be seen as a generalized martingale mass transportation problem (e.g. Beiglböck et al. [5]). In a classic two-period martingale mass transportation problem, the objective is to minimize  $\mathbb{E}^\mathbb{P}[\phi(X, Y)]$  for some cost function  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$  over  $(X, Y)$  where the distributions of  $X$  and  $Y$  under some measure  $\mathbb{P}$  are known, and  $\mathbb{E}^\mathbb{P}[Y|X] = X$ . Note that

our constraints (5.4) are the same as in the classic problem. The only difference between (5.3) and the classic problem is that our objective  $\Phi$  does not have the form of an expected value of  $\phi(X, Y)$ . Rather,  $\Phi$  is determined by the joint distribution of  $(X, Y)$ . Hence,  $\Phi$  can be seen as a generalized cost functional in a mass transportation problem.

Meanwhile, a lower bound for the optimal value of (5.3) can be obtained by considering an optimization problem with a weaker constraint:

$$\min_{Y \in K'} \Phi(X, Y), \quad (5.5)$$

where  $K'$  is the set of all random variables  $Y$  satisfying  $Y|_{P'} \stackrel{d}{=} \frac{dQ}{dP}|_P$ , and

$$\mathbb{E}^{P'}[Y \mathbf{1}_{\{X \in D_a\}}] = \mathbb{E}^{P'}[X \mathbf{1}_{\{X \in D_a\}}].$$

Note that since only the joint distribution of  $X$  and  $Y$  matters, here we take  $X$  as given and reduce the problem to an optimization solely over  $Y$ . Denote by  $Y^*$  an optimal solution of (5.5), and let  $p^* = p_{X, Y^*}$ . Then for any two points  $y_1, y_2$  and  $\lambda \in (0, 1)$  satisfying  $p^*(y_1), p^*(y_2), p^*(\lambda y_1 + (1 - \lambda)y_2) \in (0, 1)$ , a variational argument leads to the first order condition

$$f(p^*(\lambda y_1 + (1 - \lambda)y_2)|\lambda y_1 + (1 - \lambda)y_2) = \lambda f(p^*(y_1)|y_1) + (1 - \lambda)f(p^*(y_2)|y_2), \quad (5.6)$$

which implies that  $f(p^*(y)|y)$  must be linear in  $y$  when  $p^*(y)$  is between 0 and 1. Combining this with the constraints

$$\mathbb{E}^{P'}[p^*(Y^*)] = P'(X \in D_a) = F([a, \infty))$$

and

$$\mathbb{E}^{P'}[Y^* p^*(Y^*)] = \mathbb{E}^{P'}[Y^* \mathbf{1}_{\{X \in D_a\}}] = \mathbb{E}^{P'}[X \mathbf{1}_{\{X \in D_a\}}]$$

generically gives a unique solution, which is a local minimum by checking the second order condition. Note that similar to (5.2), (5.5) can be rewritten as  $\min \mathbb{E}^{P'}[Z \mathbf{1}_{\{X \in D_a\}}]$ , where the minimum is taken over all the  $(Y, Z)$  such that  $(Y, Z)|_{P'} \stackrel{d}{=} (\frac{dQ}{dP}, \frac{dR}{dP})|_P$  and  $\mathbb{E}^{P'}[Y \mathbf{1}_{\{X \in D_a\}}] = \mathbb{E}^{P'}[X \mathbf{1}_{\{X \in D_a\}}]$ . As the objective  $\mathbb{E}^{P'}[Z \mathbf{1}_{\{X \in D_a\}}]$  is linear and the feasible region is convex (with respect to mixture), the local minimum must also be the global minimum for the optimization problem (5.5), providing a lower bound for the optimal value in (5.3).

In some special cases, the above lower bound can be analytically calculated, and one can construct random variables satisfying the original constraints that attain this bound. As a result, (5.3) and (5.5) have the same optimal value, and the original problem (5.1) is completely solved. We give one simple example. Let  $F$  and  $G$  be supported on  $[0, 1]$ ,  $a \geq \frac{1}{2}$ , and  $\frac{dG}{dF}$  follows a

symmetric triangular distribution under  $F$  and is decreasing:  $\frac{dG}{dF}(t) = 2 - \sqrt{2t}$  for  $t \in [0, 1/2)$  and  $\frac{dG}{dF}(t) = \sqrt{2-2t}$  for  $t \in [1/2, 1]$ .  $\frac{dQ}{dP}$  follows a uniform distribution on  $[0, 2]$ , and the conditional distribution of  $\frac{dR}{dP}$  given  $\frac{dQ}{dP}$  is uniform with linear bounds:  $\frac{dR}{dP} \Big| \frac{dQ}{dP} = y \sim \text{Unif}([cy - b, cy + b])$  for some constants  $b$  and  $c$ . In this case we construct  $(X, Y)$  such that  $X|_{P'} \stackrel{d}{=} \frac{dG}{dF} \Big|_F$ , and

$$Y|X = x \sim \begin{cases} \text{Unif}([0, 2x]) & x \in [0, 1) \\ \text{Unif}([2x - 2, 2]) & x \in [1, 2]. \end{cases}$$

$p_{X,Y}(y)$  can be derived and then it can be verified that the first order condition (5.6) is met. Consequently, the dependence given by  $(X, Y)$  is indeed optimal for problem (5.3), and the corresponding optimal value can be calculated. We omit the detail as the rest is purely computational.

## Acknowledgements

The authors are grateful to the Editor, the Associate Editor, two referees, Michel Baes, Fabio Bellini, Paul Embrechts, Fabio Maccheroni, Tiantian Mao, Alfred Müller, Marcel Nutz, Jan Obloj, Sidney Resnick, Ludger Rüschendorf, Alexander Schied and Xiaolu Tan for various helpful suggestions and discussions on an earlier version of the paper. J. Shen acknowledges financial support from the China Scholarship Council. Y. Shen and R. Wang acknowledge financial support by the Natural Sciences and Engineering Research Council (NSERC 2014-04840, RGPIN-2018-03823, RGPAS-2018-522590) of Canada. R. Wang is also grateful to FIM at ETH Zurich for supporting his visit in 2017, during which part of this paper was written.

## A Appendix

### A.1 Details in Example 3.8

Note that  $\frac{dQ_1}{dQ_2}$  is uniform on  $[0, 2]$  under  $Q_2 = \lambda$ , and  $\frac{dF_1}{dF_2}$  is also uniform on  $[0, 2]$  under  $F_2 = \lambda$ . Thus,

$$\left( \frac{dF_1}{d\lambda}, \frac{dF_2}{d\lambda} \right) \Big|_{\lambda} \stackrel{d}{=} \left( \frac{dQ_1}{d\lambda}, \frac{dQ_2}{d\lambda} \right) \Big|_{\lambda}.$$

Therefore,  $(F_1, F_2) \preceq_h (Q_1, Q_2)$ .

Next, we will see that  $(Q_1, Q_2)$  and  $(F_1, F_2)$  are not compatible. Suppose for the purpose of contradiction that  $(Q_1, Q_2)$  and  $(F_1, F_2)$  are compatible. By Theorem 2.2, there exists a random variable  $X$  in  $(\Omega, \mathcal{A})$  with a uniform distribution on  $[0, 1]$  under  $Q_2 = \lambda$  such that

$$\frac{dF_1}{d\lambda}(X) = \mathbb{E}^\lambda \left[ \frac{dQ_1}{d\lambda} \Big| X \right].$$

In addition,

$$\left. \frac{dF_1}{d\lambda}(X) \right|_{\lambda} \stackrel{d}{=} \left. \frac{dQ_1}{d\lambda} \right|_{\lambda},$$

and therefore,

$$\frac{dF_1}{d\lambda}(X) = \frac{dQ_1}{d\lambda}, \quad \lambda\text{-almost surely.}$$

From the definition of  $F_1$  and  $Q_1$ , we have, for  $\lambda$ -almost surely  $t \in [0, 1]$ ,  $|4X(t) - 2| = 2t$ . It follows that  $X(t) = (t + 1)/2$  or  $X(t) = (1 - t)/2$  for all  $t \in [0, 1]$ . Write

$$A = \left\{ t \in [0, 1] : X(t) = \frac{t+1}{2} \right\}, \quad B = \left\{ t \in [0, 1] : X(t) = \frac{1-t}{2} \right\}$$

and

$$C = \left\{ \frac{1-t}{2} : t \in A \right\}.$$

As  $X$  is  $\mathcal{B}([0, 1])$ -measurable and has distribution  $F_2$  under  $\lambda$ , we have  $A, B \in \mathcal{B}([0, 1])$  and  $\lambda(A) = \lambda(B) = 1/2$ . Note that  $\lambda(C) = 1/4$ ; however  $\lambda(C \cap X(A \cup B)) = 0$ , contradicting the fact that  $X$  has a uniform distribution on  $[0, 1]$  under  $\lambda$ .

## A.2 Proof of Theorem 3.10

*Proof. Necessity.* Assume that  $(Q_1, \dots, Q_n)$  and  $(F_1, \dots, F_n)$  are almost compatible. This means that for any  $\varepsilon > 0$ , there exists  $(F_{1,\varepsilon}, \dots, F_{n,\varepsilon})$  such that  $D_{\text{KL}}(F_{i,\varepsilon} \| F_i) < \varepsilon$  for  $i = 1, \dots, n$ , and  $(Q_1, \dots, Q_n)$  is compatible with  $(F_{1,\varepsilon}, \dots, F_{n,\varepsilon})$ . Define probability measures

$$F_\varepsilon = \frac{1}{n}(F_{1,\varepsilon} + \dots + F_{n,\varepsilon}),$$

$$F = \frac{1}{n}(F_1 + \dots + F_n)$$

and

$$Q = \frac{1}{n}(Q_1 + \dots + Q_n).$$

Note that the distribution of  $X_\varepsilon$  under  $Q$  is  $F_\varepsilon$ , where  $X_\varepsilon$  is the random variable defining the compatibility between  $(Q_1, \dots, Q_n)$  and  $(F_{1,\varepsilon}, \dots, F_{n,\varepsilon})$ . Moreover, for  $i = 1, \dots, n$ , we have  $F_{i,\varepsilon} \ll F_\varepsilon$ ,  $Q_i \ll Q$ ,  $dF_{i,\varepsilon}/dF_\varepsilon \leq n$  and  $dQ_i/dQ \leq n$ . For  $\varepsilon > 0$ , by Lemma 3.3,

$$\left( \frac{dF_{1,\varepsilon}}{dF_\varepsilon}, \dots, \frac{dF_{n,\varepsilon}}{dF_\varepsilon} \right) \Big|_{F_\varepsilon} \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q.$$

As a result, for any convex function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$\mathbb{E}^{F_\varepsilon} \left[ f \left( \frac{dF_{1,\varepsilon}}{dF_\varepsilon}, \dots, \frac{dF_{n,\varepsilon}}{dF_\varepsilon} \right) \right] \leq \mathbb{E}^Q \left[ f \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \right].$$

For  $i = 1, \dots, n$ ,

$$\frac{dF_{i,\varepsilon}}{dF_\varepsilon} = \frac{dF_i}{dF} \frac{dF_{i,\varepsilon}/dF_i}{dF_\varepsilon/dF}. \quad (\text{A.1})$$

Since  $D_{\text{KL}}(F_{i,\varepsilon} \| F_i)$  converges to 0, by Pinsker's inequality,  $F_{i,\varepsilon}$  converges to  $F_i$  in total variation, which is equivalent to  $dF_{i,\varepsilon}/dF_i$  converging in  $L^1|_{F_i}$  to 1. Hence for any sequence  $\varepsilon_m \downarrow 0$ , there exists a subsequence, which we still denote as  $\varepsilon_m \downarrow 0$  by a slight abuse of notation, such that  $dF_{i,\varepsilon_m}/dF_i$  converge to 1  $F_i$ -almost surely. It is easy to check that we have  $dF_{\varepsilon_m}/dF$  converge to 1 as well. (A.1) then implies that

$$\frac{dF_{i,\varepsilon_m}}{dF_{\varepsilon_m}} \rightarrow \frac{dF_i}{dF} \quad F_i\text{-almost surely.} \quad (\text{A.2})$$

On any set  $B \in \mathcal{B}(\mathbb{R})$  such that  $F_i(B) = 0$  but  $F(B) > 0$ , suppose  $dF_{i,\varepsilon}/dF_\varepsilon$  does not converge to  $dF_i/dF = 0$  in probability under  $F|_B$ , the measure  $F$  restricted on  $B$ . Then there exists  $\delta > 0$  and a subsequence of  $\varepsilon_m$  (again denoted as  $\varepsilon_m$ ), such that  $P^{F|_B}(dF_{i,\varepsilon_m}/dF_{\varepsilon_m} > \delta) \geq c$  for some constant  $c > 0$ . Since  $F_{\varepsilon_m}$  converges to  $F$  in total variation, for  $m$  large enough,  $P^{F_{\varepsilon_m}|_B}(dF_{i,\varepsilon_m}/dF_{\varepsilon_m} > \delta) \geq c/2$ . Hence  $F_{i,\varepsilon_m}(B) \geq \delta P^{F_{\varepsilon_m}|_B}(dF_{i,\varepsilon_m}/dF_{\varepsilon_m} > \delta) \geq \frac{c\delta}{2}$ , which contradicts the fact that  $F_{i,\varepsilon_m}$  converges to  $F_i$  in total variation. We conclude that  $dF_{i,\varepsilon}/dF_\varepsilon$  converge to  $dF_i/dF = 0$  in probability under  $F$  on set  $\{dF_i/dF = 0\}$ . Combining this result with (A.2) and taking a further subsequence allows us to replace the  $F_i$ -almost sure convergence in (A.2) by  $F$ -almost sure convergence.

For any convex function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$\mathbb{E}^{F_{\varepsilon_m}} \left[ f \left( \frac{dF_{1,\varepsilon_m}}{dF_{\varepsilon_m}}, \dots, \frac{dF_{n,\varepsilon_m}}{dF_{\varepsilon_m}} \right) \right] = \int f \left( \frac{dF_{1,\varepsilon_m}}{dF_{\varepsilon_m}}, \dots, \frac{dF_{n,\varepsilon_m}}{dF_{\varepsilon_m}} \right) dF_{\varepsilon_m}.$$

Since  $\frac{dF_{i,\varepsilon_m}}{dF_{\varepsilon_m}} \in [0, n]$ , and  $f$  is convex hence continuous,  $|f(\frac{dF_{1,\varepsilon_m}}{dF_{\varepsilon_m}}, \dots, \frac{dF_{n,\varepsilon_m}}{dF_{\varepsilon_m}})|$  is bounded. Let  $b$  be an upper bound of it. Because  $F_{\varepsilon_m}$  converges in total variation to  $F$ , we have

$$\left| \int f \left( \frac{dF_{1,\varepsilon_m}}{dF_{\varepsilon_m}}, \dots, \frac{dF_{n,\varepsilon_m}}{dF_{\varepsilon_m}} \right) dF_{\varepsilon_m} - \int f \left( \frac{dF_{1,\varepsilon_m}}{dF_{\varepsilon_m}}, \dots, \frac{dF_{n,\varepsilon_m}}{dF_{\varepsilon_m}} \right) dF \right| \leq 2b\delta(F_{\varepsilon_m}, F) \rightarrow 0 \quad (\text{A.3})$$

uniformly, where  $\delta(\cdot, \cdot)$  is the total variation distance. Moreover, by dominated convergence, we have

$$\int f \left( \frac{dF_{1,\varepsilon_m}}{dF_{\varepsilon_m}}, \dots, \frac{dF_{n,\varepsilon_m}}{dF_{\varepsilon_m}} \right) dF \rightarrow \int f \left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) dF. \quad (\text{A.4})$$

(A.3) and (A.4) together show that

$$\begin{aligned} \mathbb{E}^F \left[ f \left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) \right] &= \lim_{m \rightarrow \infty} \mathbb{E}^{F_{\varepsilon_m}} \left[ f \left( \frac{dF_{1,\varepsilon_m}}{dF_{\varepsilon_m}}, \dots, \frac{dF_{n,\varepsilon_m}}{dF_{\varepsilon_m}} \right) \right] \\ &\leq \mathbb{E}^Q \left[ f \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \right]. \end{aligned}$$

Sufficiency. Assume that  $(F_1, \dots, F_n) \preceq_{\text{h}} (Q_1, \dots, Q_n)$ . By Lemma 3.5, this means that

$$\left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) \Big|_F \preceq_{\text{cx}} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) \Big|_Q$$

holds for  $F = \frac{1}{n} \sum_{i=1}^n F_i$  and  $Q = \frac{1}{n} \sum_{i=1}^n Q_i$ .

By Lemma 3.2, there exists a probability space  $(\Omega', \mathcal{A}', Q')$  and random vectors  $\mathbf{Y}' = (Y'_1, \dots, Y'_n), \mathbf{Z}' = (Z'_1, \dots, Z'_n)$  defined on  $(\Omega', \mathcal{A}', Q')$ , such that

$$\begin{aligned} (Y'_1, \dots, Y'_n) &\stackrel{d}{=} \left( \frac{dQ_1}{dQ}, \dots, \frac{dQ_n}{dQ} \right) =: \mathbf{Y} = (Y_1, \dots, Y_n), \\ (Z'_1, \dots, Z'_n) &\stackrel{d}{=} \left( \frac{dF_1}{dF}, \dots, \frac{dF_n}{dF} \right) =: \mathbf{Z} = (Z_1, \dots, Z_n), \end{aligned}$$

and

$$\mathbb{E}^{Q'} [Y'_i | Z'_i] = Z'_i, \quad i = 1, \dots, n.$$

Given  $m = 0, 1, \dots$ , define random vector  $\mathbf{Y}_m = (Y_{m,1}, \dots, Y_{m,n})$  by

$$Y_{m,i} = \begin{cases} 0 & \text{if } Y_i = 0 \\ \exp(2^{-m} \lfloor 2^m \log(Y_i) \rfloor) & \text{otherwise} \end{cases}$$

for  $i = 1, \dots, n$ . Similarly we define  $\mathbf{Y}'_m, \mathbf{Z}_m$  and  $\mathbf{Z}'_m$  for  $\mathbf{Y}', \mathbf{Z}$  and  $\mathbf{Z}'$ , respectively. Note that

$$\begin{aligned} \mathbb{E}^{Q'} [Y'_{m,i} | Z'_{m,i}] &\in \left[ \exp(-2^{-m}) \mathbb{E}^{Q'} [Y'_i | Z'_{m,i}], \mathbb{E}^{Q'} [Y'_i | Z'_{m,i}] \right] \\ &\subseteq \left[ \exp(-2^{-m}) Z'_{m,i}, \exp(2^{-m}) Z'_{m,i} \right] \end{aligned}$$

for  $i = 1, \dots, n$ .

Each of  $Q_1, \dots, Q_n$  is atomless, and so is  $Q$ . As a result, we can divide  $\Omega$  into disjoint sets  $A_{k,j}^m$ , where  $k = (k_1, \dots, k_n) \in (\mathbb{Z} \cup \{-\infty\})^n$  and  $j = (j_1, \dots, j_n) \in (\mathbb{Z} \cup \{-\infty\})^n$ , such that  $Y_{m,i}(\omega) = \exp(k_i 2^{-m})$  for  $\omega \in A_{k,j}^m$  and  $i = 1, \dots, n$ ,

$$Q(A_{k,j}^m) = Q'(Y'_{m,i} = \exp(k_i 2^{-m}), Z'_{m,i} = \exp(j_i 2^{-m}), i = 1, \dots, n).$$

Here we follow the tradition that  $\exp(-\infty) = 0$  for ease of notation. Define random vector  $\mathbf{Z}''_m$  on  $(\Omega, \mathcal{A}, Q)$  by  $Z''_{m,i}(\omega) = \exp(j_i 2^{-m})$  for  $\omega \in A_{k,j}^m$ , then  $(\mathbf{Y}_m, \mathbf{Z}''_m)|_Q \stackrel{d}{=} (\mathbf{Y}'_m, \mathbf{Z}'_m)|_{Q'}$ .

Let  $I_d$  be the identity random variable on  $(R, \mathcal{B}(\mathbb{R}))$ . For  $\ell = 0, 1, \dots$  and  $h \in \mathbb{Z}$ , denote by  $\varphi_{\ell,h}^m(z)$  the conditional probability under  $F$  of the event  $I_d \in [h2^{-\ell}, (h+1)2^{-\ell})$  given  $Z_m = z$ :

$$\varphi_{\ell,h}^m(z) = F(I_d \in [h2^{-\ell}, (h+1)2^{-\ell}) | Z_m = z).$$

Then for any  $\ell = 0, 1, \dots$ ,  $A_{k,j}^m$  can be further divided into disjoint subsets  $A_{k,j,\ell,h}^m$ , such that  $Q(A_{k,j,\ell,h}^m) = Q(A_{k,j}^m) \varphi_{\ell,h}^m(\exp(j 2^{-m}))$ . Moreover, the partitions can be made such that  $\{A_{k,j,\ell',h}^m\}_{h \in \mathbb{Z}}$  is a refinement of  $\{A_{k,j,\ell,h}^m\}_{h \in \mathbb{Z}}$  for any  $\ell' > \ell$  and any given  $m, k, j$ . Define  $X_{m,\ell}(\omega) = h2^{-\ell}$  for  $\omega \in A_{k,j,\ell,h}^m$ , and  $X_m = \lim_{\ell \rightarrow \infty} X_{m,\ell}$ . The limit exists since it is easy to check that  $X_{m,\ell}$  is increasing with respect to  $\ell$ . Note that  $X_{m,\ell}$  is conditionally independent of  $\mathbf{Y}_m$  given  $\mathbf{Z}''_m$ , hence  $X_m$  is also conditionally independent of  $\mathbf{Y}_m$  given  $\mathbf{Z}''_m$ .

By construction, for any  $A \in \mathbb{R}^n$ ,  $\ell = 0, 1, \dots$ , and  $h \in \mathbb{Z}$ ,

$$\begin{aligned}
& Q(\mathbf{Z}_m'' \in A, X_{m,\ell'} \in [h2^{-\ell}, (h+1)2^{-\ell}]) \\
&= Q(\mathbf{Z}_m'' \in A, X_{m,\ell} = h2^{-\ell}) \\
&= \sum_{\substack{k \\ j: \exp(j2^{-m}) \in A}} Q(A_{k,j,\ell,h}^m) \\
&= \sum_{\substack{k \\ j: \exp(j2^{-m}) \in A}} Q(A_{k,j}^m) \varphi_{\ell,h}^m(\exp(j2^{-m})) \\
&= \sum_{j: \exp(j2^{-m}) \in A} Q(\mathbf{Z}_m'' = \exp(j2^{-m})) \varphi_{\ell,h}^m(\exp(j2^{-m})) \\
&= \sum_{j: \exp(j2^{-m}) \in A} F(\mathbf{Z}_m = \exp(j2^{-m})) F([h2^{-\ell}, (h+1)2^{-\ell}] | \mathbf{Z}_m = \exp(j2^{-m})) \\
&= F(\mathbf{Z}_m^{-1}(A) \cap [h2^{-\ell}, (h+1)2^{-\ell}])
\end{aligned} \tag{A.5}$$

for all  $\ell' \geq \ell$ . Thus,  $\mathbf{Z}_m$ , restricted on interval  $[h2^{-\ell}, (h+1)2^{-\ell}]$ , has the same distribution as  $\mathbf{Z}_m''$ , restricted on set  $X_{m,\ell'}^{-1}([h2^{-\ell}, (h+1)2^{-\ell}])$ . Note that  $X_{m,\ell'}^{-1}([h2^{-\ell}, (h+1)2^{-\ell}])$  is the same set for any  $\ell' \geq \ell$ , hence  $\mathbf{Z}_m$  restricted on interval  $[h2^{-\ell}, (h+1)2^{-\ell}]$  also has the same distribution as  $\mathbf{Z}_m''$  restricted on  $X_m^{-1}([h2^{-\ell}, (h+1)2^{-\ell}])$  for all  $m = 0, 1, \dots$ . Because the collection of sets  $\{[h2^{-\ell}, (h+1)2^{-\ell}]\}_{h \in \mathbb{Z}, \ell=0,1,\dots}$  forms a basis for  $\mathcal{B}(\mathbb{R})$ ,  $\mathbf{Z}_m$  restricted on any Borel set  $B$  has the same distribution as  $\mathbf{Z}_m''$  restricted on  $X_m^{-1}(B)$ . Therefore we conclude that  $\mathbf{Z}_m'' = \mathbf{Z}_m \circ X_m$   $Q$ -almost surely. Moreover, by taking  $A = \mathbb{R}^n$  in (A.5), it follows that  $Q(X_{m,\ell'} \in [h2^{-\ell}, (h+1)2^{-\ell}]) = F([h2^{-\ell}, (h+1)2^{-\ell}])$  for all  $\ell' \geq \ell$ . A similar reasoning as above then shows that  $F = Q \circ X_m^{-1}$ .

For any  $A \in \mathcal{B}$  and any  $i = 1, \dots, n$ ,

$$Q_i(X_m \in A) = \int_{X_m^{-1}(A)} Y_i dQ. \tag{A.6}$$

It is easy to see that

$$\int_{X_m^{-1}(A)} Y_{m,i} dQ \leq \int_{X_m^{-1}(A)} Y_i dQ \leq \exp(2^{-m}) \int_{X_m^{-1}(A)} Y_{m,i} dQ. \tag{A.7}$$

Moreover,

$$\begin{aligned}
& \int_{X_m^{-1}(A)} Y_{m,i} dQ \\
&= \sum_j Q\left(X_m \in A | \mathbf{Z}_m'' = e^{j2^{-m}}\right) \sum_k e^{k_i 2^{-m}} Q\left(\mathbf{Y}_m = e^{k 2^{-m}}, \mathbf{Z}_m'' = e^{j2^{-m}}\right) \\
&= \sum_j Q\left(X_m \in A | \mathbf{Z}_m'' = e^{j2^{-m}}\right) Q\left(\mathbf{Z}_m'' = e^{j2^{-m}}\right) \mathbb{E}^Q[Y_{m,i} | \mathbf{Z}_m'' = e^{j2^{-m}}] \\
&= \sum_j Q\left(X_m \in A | \mathbf{Z}_m'' = e^{j2^{-m}}\right) Q\left(\mathbf{Z}_m'' = e^{j2^{-m}}\right) \mathbb{E}^{Q'}[Y_{m,i}' | \mathbf{Z}_m' = e^{j2^{-m}}]
\end{aligned}$$

$$\begin{aligned}
&\geq \sum_j Q\left(X_m \in A \mid \mathbf{Z}_m'' = e^{j2^{-m}}\right) Q\left(\mathbf{Z}_m'' = e^{j2^{-m}}\right) \exp(j_i 2^{-m} - 2^{-m}) \\
&= \sum_j F\left(A \mid \mathbf{Z}_m = e^{j2^{-m}}\right) F\left(\mathbf{Z}_m = e^{j2^{-m}}\right) \exp(j_i 2^{-m} - 2^{-m}) \\
&\geq \exp(-2^{-m}) \sum_j \exp(j_i 2^{-m}) F\left(A \cap \{\mathbf{Z}_m = e^{j2^{-m}}\}\right) \\
&= \exp(-2^{-m}) \int_A Z_{m,i} dF \\
&\geq \exp(-2^{-m+1}) \int_A Z_i dF \\
&= \exp(-2^{-m+1}) F_i(A),
\end{aligned}$$

where the first equality holds since  $X_m$  is independent of  $\mathbf{Y}_m$  given  $\mathbf{Z}_m''$ , and the fourth equality holds because  $Q \circ X_m^{-1} = F$  and  $\mathbf{Z}_m \circ X_m = \mathbf{Z}_m''$ . Symmetrically,

$$\int_{X_m^{-1}(A)} Y_{m,i} dQ \leq \exp(2^{-m}) F_i(A). \tag{A.8}$$

Combining (A.6)-(A.8), we have

$$Q_i(X_m \in A) \in [\exp(-2^{-m+1}) F_i(A), \exp(2^{-m+1}) F_i(A)].$$

Since this holds for any  $A \in \mathcal{B}(\mathbb{B})$ , we conclude that  $Q_i \circ X_m^{-1}$  is absolutely continuous with respect to  $F_i$ , and  $dQ_i \circ X_m^{-1} / dF_i \in [\exp(-2^{-m+1}), \exp(2^{-m+1})]$ . It is easy to see that  $D_{\text{KL}}(Q_i \circ X_m^{-1} \| F_i)$  converges to 0 as  $m \rightarrow \infty$ .  $\square$

## References

- [1] Anscombe, F. J., Aumann, R. J.: A definition of subjective probability. *Annals of Mathematical Statistics* **34**(1), 199–205 (1963)
- [2] Battigalli, P., Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M.: Mixed extensions of decision problems under uncertainty. *Economic Theory* **63**(4), 827–866 (2017)
- [3] Blackwell, D.: Comparison of experiments. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* **1**, 93–102 (1951)
- [4] Blackwell, D.: Equivalent comparisons of experiments. *Annals of Mathematical Statistics* **24**(2), 265–272 (1953)
- [5] Beiglböck, M., Henry-Labordère, P., Penkner, F.: Model-independent bounds for option prices: a mass transport approach. *Finance and Stochastics* **17**(3), 477–501 (2013)
- [6] Beiglböck, M., Juillet, N.: On a problem of optimal transport under marginal martingale constraints. *Annals of Probability* **44**(1), 42–106 (2016)



- [7] Beiglböck, M., Nutz, M., Touzi, N.: Complete duality for martingale optimal transport on the line. *Annals of Probability* **45**(5), 3038–3074 (2017)
- [8] Cambou, M., Filipovic, D.: Model uncertainty and scenario aggregation. *Mathematical Finance* **27**(2), 534–567 (2017)
- [9] Choquet, G., Meyer, P. A.: Existence et unicité des représentations intégrales dans les convexes compacts quelconques. *Annales de l’institut Fourier* **13**(1), 139–154 (1963)
- [10] Delbaen, F.: Common monotonicity and time consistency for Lebesgue continuous monetary utility functions. arXiv:1904.04522 (2019)
- [11] Embrechts, P., Hofert, M., Wang, R.: Bernoulli and tail-dependence compatibility. *Annals of Applied Probability* **26**(3), 1636–1658 (2016)
- [12] Embrechts, P., McNeil, A., Straumann, D.: Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond* **1**, 176–223 (2002)
- [13] Föllmer, H., Schied, A.: *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter, Berlin, Fourth Edition (2016)
- [14] Glasserman, P., Li, J.: Importance sampling for portfolio credit risk. *Management Science* **51**(11), 1643–1656 (2005)
- [15] Hirsch, F., Profeta, C., Roynette, B., Yor, M.: *Peacocks and Associated Martingales, with Explicit Constructions*. Springer Science and Business Media (2011)
- [16] Hobson, D.: The Skorokhod embedding problem and model-independent bounds for option prices. *Paris-Princeton Lectures on Mathematical Finance 2010* 267–318. Springer, Berlin (2011)
- [17] Le Cam, L.: Comparison of experiments - a short review. *Lecture Notes-Monograph Series* **30** 127-138 (1996)
- [18] McNeil, A. J., Frey, R., Embrechts, P.: *Quantitative Risk Management: Concepts, Techniques and Tools, Revised Edition*. Princeton University Press, Princeton (2015)
- [19] Müller, A., Stoyan, D.: *Comparison Methods for Statistical Models and Risks*. Wiley, New York (2002)
- [20] Phelps, R. R.: *Lectures on Choquet’s Theorem, Second ed.* Springer-Verlag, Berlin, Heidelberg (2001)
- [21] Revuz, D., Yor, M.: *Continuous Martingales and Brownian Motion*. Springer Science and Business Media (2013).
- [22] Rüschendorf, L.: *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg (2013)
- [23] Shaked, M., Shanthikumar, J. G.: *Stochastic Orders*. Springer Science and Business Media

(2007)

- [24] Siegmund, D.: Importance sampling in the Monte Carlo study of sequential tests. *Annals of Statistics* **4**(4) 673–684 (1976)
- [25] Strassen, V.: The existence of probability measures with given marginals. *Annals of Mathematical Statistics* **36**(2), 423–439 (1965)
- [26] Torgersen, E. N.: *Comparison of Statistical Experiments*. Cambridge University Press, Cambridge (1991)
- [27] Wang, R. and Ziegel, J. F.: Scenario-based risk evaluation. SSRN: 3235450 (2018)