# Universität Stuttgart

# Fachbereich Mathematik

## Elicitation and Identification of Properties

Ingo Steinwart, Chloé Pasin, Robert Williamson, Siyu Zhang†

# Elicitation and Identification of Properties

**Ingo Steinwart**                                    INGO.STEINWART@MATHEMATIK.UNI-STUTTGART.DE
*University of Stuttgart*

**Chloé Pasin**                                                                CPASIN@ENS-CACHAN.FR
*ENS Cachan*

**Robert Williamson**                                            BOB.WILLIAMSON@ANU.EDU.AU
*ANU and NICTA*

**Siyu Zhang**[*][†]
*ENS Cachan*

## Abstract

Properties of distributions are real-valued functionals such as the mean, quantile or conditional value at risk. A property is elicitable if there exists a scoring function such that minimization of the associated risks recovers the property. We extend existing results to characterize the elicitability of properties in a general setting. We further relate elicitability to identifiability (a notion introduced by Osband) and provide a general formula describing all scoring functions for an elicitable property. Finally, we draw some connections to the theory of coherent risk measures.

**Keywords:** Elicitation, property, scoring function, identification function, risk measure, expectile.

## 1. Introduction

Probability distributions are widely used to model data. While some machine learning methods provide complete probabilistic forecasts, it is often necessary for computational or other reasons to work with partial information about the distributions (Gneiting and Katzfuss, 2014). The choice of what partial information affects the performance of methods, their usefulness and validity for real world prediction problems.

A property is a functional that assigns a real number to a probability distribution. For example, the mean, the variance, and $\tau$-quantiles are properties. Properties are often intimately related to scoring functions. It is well known (Klein and Grottke, 2008), for example, that the mean $\mu$ of a distribution $P$ on $\mathbb{R}$ can be written as $\mu = \arg\min_{t\in\mathbb{R}} \mathbb{E}_{\mathsf{Y}\sim P}(\mathsf{Y} - t)^2$. Analogously, the $\tau$-quantile $q_\tau(P)$ of $P$ satisfies $q_\tau = \arg\min_{t\in\mathbb{R}} \mathbb{E}_{\mathsf{Y}\sim P} S_\tau(\mathsf{Y}, t)$ for the so-called $\tau$-pinball function $S_\tau(t,y) = (\mathbf{1}_{t\geq y} - \tau)(t - y)$ (Gneiting, 2011). However, there are other functions $S$ that lead to the mean or the $\tau$-quantile; in the case of the mean it has been long known that such $S$ must take the form of a Bregman divergence (McCarthy, 1956; Savage, 1971; Schervish, 1989). On the other hand, there are properties such as the "conditional value at risk" $\mathrm{CVaR}_\alpha(P) = \frac{1}{1-\alpha} \int_\alpha^1 q_\beta(P)d\beta$ for which there is no such scoring function. This motivates the question: which properties are elicitable? That is, for which properties is there a suitable scoring function?

---

. † Siyu Zhang was a Master of Mathematics student from Ecole normale supérieure de Cachan visiting ANU/NICTA in 2013. He contributed to many of the results and writing in this paper but suddenly and tragically passed away before it was completed. We dedicate the paper to his memory.

Previous work has given partial answers to this question, see Gneiting (2011) and Lambert (2012) for recent summaries of much of the earlier literature. Most previous work has focused on somewhat restricted cases. For example, (Lambert et al., 2008) proved a similar result for distributions on finite sets, and in addition, they also considered vector-valued properties. Abernethy and Frongillo (2012) presented a more general treatment for *linear* properties and showed the characterization of scoring functions in terms of Bregman divergence holds in that more general setting. Finally, Lambert (2012) presented a theorem similar to one of our main results for continuous densities on compact metric spaces. Unfortunately, however, there is a flaw in his proof, see Appendix A for details. We fix this flaw and simultaneously extend the characterization of elicitable properties to classes of arbitrary bounded densities.

Expressing a property as an argmin as above is analogous (modulo a flip of sign) to the conventional M-estimators of robust statistics (Huber, 1981). By differentiating the defining nonlinear function of an M-estimator, one obtains a "Z-estimator", which is also known as an "identification function" (Osband, 1985). Our characterization of elicitable properties also includes a result regarding the associated identification function. In particular it turns out that we can construct all scoring functions from a given identification function but finding an identification function from a scoring function is not always possible.

Properties and their associated scoring functions are related to risk measures (Artzner et al., 1999; Bellini et al., 2014; Kusuoka, 2001; Rockafellar and Uryasev, 2013; Rockafellar, 2007). Connections between these measures and certain machine learning algorithms have recently appeared (Tsyarmasto and Uryasev, 2012; Gotoh et al., 2013; Gotoh and Uryasev, 2013). Another contribution of our paper is to resolve an open question that elucidates the relationship between the requirement of "coherence" of a risk measure and the elicitability of the associated property; the result shows that the expectile (Newey and Powell, 1987), which is a type of generalized quantile (Jones, 1994), is the only elicitable coherent risk measure.

We like to remark that elicitable properties are exactly those properties that, in their conditional form, can be estimated by (regularized) empirical risk minimization (ERM) algorithms. For example, SVMs with the $\tau$-pinball loss estimate the conditional $\tau$-quantiles, see Steinwart and Christmann (2011). By characterizing elicitable properties we thus describe the boundaries of this broad class of algorithms. In this respect note that even if the primal learning algorithm is not of ERM type, but its hyper-parameter selection uses cross-validation based on empirical risks, or it is eventually tested with the help of an empirical risk, the question of elicitability arises naturally.

The paper is organized as follows. In §2 we introduce properties and scoring functions, define elicitability and show the relationship between convexity of level sets of the property and the elicitability of the property. In §3 we introduce identification functions and show how scoring functions can be constructed from identification functions. In §4 we characterize which properties are elicitable. In addition, we characterize all suitable scoring functions for an elicitable property. The latter generalizes known results for particular properties such as the above mentioned ones by Abernethy and Frongillo (2012) and Gneiting (2011). We further illustrate the general theory by constructing the scoring functions for generalized quantiles, see Bellini et al. (2014). In §5 we connect elicitability and coherent risk measures by solving an open question recently raised by Ziegel (2014).

## 2. Properties and Scoring Functions

Let us begin by introducing some notations and assumptions used throughout this paper. To this end, let $A \subset \mathbb{R}$ be an interval. For technical reasons we will always equip $A$ with the Lebesgue completion $\hat{\mathcal{B}}(A)$ of the Borel $\sigma$-algebra $\mathcal{B}(A)$. We further write $\lambda$ for the Lebesgue measure on $A$. Recall that a measure $\nu$ on $A$ is strictly positive, if $\nu(O) > 0$ for all open $O \subset A$. For later use note that this is the case, if and only if $\nu((a, b]) > 0$ for all $(a, b] \subset A$.

In the following, let $(Y, \mathcal{A})$ be a measurable space, where we will think of $Y$ as a set of possible observations. We further fix a set $\mathcal{P}$ of probability measures on $(Y, \mathcal{A})$ and call $\mathcal{P}$ *topological*, if there is a topology on $\mathcal{P}$ that is induced by some vector space topology on the linear span $\mathrm{span}\,\mathcal{P}$ of $\mathcal{P}$. We further call a map $T : \mathcal{P} \to \mathbb{R}$ a property, and denote its image by $\mathrm{im}\,T$. Simple examples of properties are the mean and the variance on suitable sets $\mathcal{P}$.

Ultimately, we are interested in estimating the property $T(P)$ from observations drawn from $P$. To this end, assume that we have a so-called $\mathcal{P}$-*scoring function* $S : A \times Y \to \mathbb{R}$, that is, $\mathbb{E}_{\mathsf{Y} \sim P}\, S(t, \mathsf{Y})$ exists for all $t \in A$ and $P \in \mathcal{P}$. Note that for *loss functions*, i.e. non-negative and measurable scoring functions, the existence of the expectation is always guaranteed although it may not be finite in general. We will view $S(t, y)$ as a penalty for estimating $y \in Y$ by $t \in A$, so that smaller values $S(t, y)$ are preferred. Following this idea, we call $S$ strictly $\mathcal{P}$-consistent for the property $T : \mathcal{P} \to \mathbb{R}$, if $\mathrm{im}\,T \subset A$ and, for all $P \in \mathcal{P}$, we have

$$T(P) = \underset{t \in \mathbb{R}}{\arg\min}\, \mathbb{E}_{\mathsf{Y} \sim P} S(t, \mathsf{Y})\,. \tag{1}$$

Clearly, (1) is a minimal requirement for any ERM approach to work consistently. In general, however, we will not be able to minimize the right hand side of (1) exactly without knowing $P$, and thus we need to specify the effects of such inaccuracies. One such specification introduced by Lambert (2012) is that of order sensitivity. Recall that $S$ is said to be $\mathcal{P}$-*order sensitive for $T$*, if $\mathrm{im}\,T \subset A$ and, for all $P \in \mathcal{P}$ and all $t_1, t_2 \in A$ with either $t_2 < t_1 \leq T(P)$ or $T(P) \leq t_1 < t_2$, we have

$$\mathbb{E}_{\mathsf{Y} \sim P} S(t_1, \mathsf{Y}) < \mathbb{E}_{\mathsf{Y} \sim P} S(t_2, \mathsf{Y})\,. \tag{2}$$

Clearly, order sensitive scoring functions are consistent. Another concept introduced by Steinwart (2007) to deal with inaccuracies is *self-calibration*. Recall, that if we have a property $T$ and a $\mathcal{P}$-consistent scoring function $S$ for $T$, then $S$ is called $\mathcal{P}$-self-calibrated, if, for all $\varepsilon > 0$ and $P \in \mathcal{P}$, there exists a $\delta > 0$, such that for $t^* := T(P)$ and all $t \in A$ we have

$$\left| \mathbb{E}_{\mathsf{Y} \sim P} S(t, \mathsf{Y}) - \mathbb{E}_{\mathsf{Y} \sim P} S(t^*, \mathsf{Y}) \right| < \delta \qquad \Longrightarrow \qquad |t - t^*| < \varepsilon\,.$$

For self-calibrated $S$, every $\delta$-approximate minimizer of $\mathbb{E}_{\mathsf{Y} \sim P} S(\cdot, \mathsf{Y})$, approximates the desired property $T(P)$ with precision not worse than $\varepsilon$. The relationship between $\varepsilon$ and $\delta$ is exactly described by the so-called self-calibration function, which can often be explicitly calculated; see Steinwart (2007). In some sense order sensitivity is a global and qualitative notion while self-calibration is a local and quantitative notion. The next lemma relates order sensitivity and self-calibration.

**Lemma 1** *Let $T : \mathcal{P} \to \mathbb{R}$ be a property and $S : A \times Y \to \mathbb{R}$ be a $\mathcal{P}$-order sensitive scoring function for $T$. Then $S$ is $\mathcal{P}$-self-calibrated.*

If we start with a scoring function, we can use (1) to *define* a property $T$ whenever the optimization problem has a unique solution. In other scenarios, however, we need to start with a property and thus need to look for consistent scoring functions. This leads to the following definition.

3

**Definition 2** *A property $T : \mathcal{P} \to \mathbb{R}$ is elicitable, if there is a $\mathcal{P}$-consistent scoring function for $T$.*

One of our main goals is to characterize elicitable properties. Let us begin with the following necessary condition taken from Osband (1985), see also Lambert et al. (2008); Gneiting (2011).

**Theorem 3** *Let $T : \mathcal{P} \to \mathbb{R}$ be an elicitable property. Then, for all $P_0, P_1 \in \mathcal{P}$ with $T(P_0) = T(P_1)$ and all $\alpha \in [0, 1]$ with $(1 - \alpha)P_0 + \alpha P_1 \in \mathcal{P}$ we have*

$$T(P_0) = T(P_1) = T\big((1 - \alpha)P_0 + \alpha P_1\big).$$

*In particular, if $\mathcal{P}$ is convex, then, for all $t \in \mathbb{R}$, the level set $\{T = t\}$ is convex.*

Note that a direct consequence of Theorem 3 is that, for sufficiently large and convex $\mathcal{P}$, the variance is not elicitable.

Our next goal is to characterize the convexity of the level sets. To this end, let us recall that a function $f : X \to \mathbb{R}$ defined on some convex subset $X \subset E$ of a vector space $E$ is called *quasi-convex*, if, for all $t \in \mathbb{R}$, the sublevel sets $\{f \leq t\} := \{x \in X \colon f(x) \leq t\}$ are convex. It is well-known (Greenberg and Pierskalla, 1971) that $f$ is quasi-convex, if and only if

$$f\big((1 - \alpha)x + \alpha y\big) \leq \max\big\{f(x), f(y)\big\}$$

for all $x, y \in X$ and $\alpha \in [0, 1]$. Analogously, $f$ is called *quasi-concave*, if $-f$ is quasi-convex. Finally, $f$ is called *quasi-monotonic*, if $f$ is both quasi-convex and quasi-concave. With these notions being introduced, we can now formulate the following result.

**Theorem 4** *Let $\mathcal{P}$ be a convex and topological, and $T : \mathcal{P} \to \mathbb{R}$ be a continuous property. Then $\operatorname{im} T$ is an interval, and the following statements are equivalent:*

*i) For all $t \in \mathbb{R}$, the level set $\{T = t\}$ is convex.*

*ii) For all $t \in \mathbb{R}$, the sets $\{T < t\}$ and $\{T > t\}$ are convex.*

*iii) $T$ is quasi-monotonic.*

*iv) For all $P_0, P_1 \in \mathcal{P}$, the function $\alpha \mapsto T((1 - \alpha)P_0 + \alpha P_1)$ defined on $[0, 1]$ is monotonic.*

## 3. Identification Functions and their Relation to Scoring Functions

Consistent scoring functions are closely related to M-estimators. Similarly, the following, second tool to recover a property is closely related to so-called Z-estimators Huber (1981); confer (Dawid, 2007, p.79). To introduce it, let $T : \mathcal{P} \to \mathbb{R}$ be a property and $A \subset \mathbb{R}$ be an interval. Furthermore, let $N \in \hat{\mathcal{B}}(A)$ with $\lambda(N) = 0$ and $V : A \times Y \to \mathbb{R}$ be a function such that $V(t, \cdot) \in L_1(P)$ for all $t \in A \setminus N$ and $P \in \mathcal{P}$. Then $V$ is called a $\mathcal{P}$-*identification function for $T$*, if $\operatorname{im} T \subset A$ and

$$\mathbb{E}_{Y \sim P} V(t, Y) = 0 \qquad \Longleftrightarrow \qquad t = T(P) \tag{3}$$

for all $t \in \operatorname{\mathring{im}} T \setminus N$ and $P \in \mathcal{P}$, where $\operatorname{\mathring{im}} T$ denotes the interior of $\operatorname{im} T$. If, in addition, we have

$$\mathbb{E}_{Y \sim P} V(t, Y) > 0 \qquad \Longleftrightarrow \qquad t > T(P) \tag{4}$$

4

for all $t \in \overset{\circ}{\mathrm{im}}\, T \setminus N$ and $P \in \mathcal{P}$, then $V$ is called *oriented*. Finally, if $N = \emptyset$, then $V$ is called *strong*. Note that besides some technicalities, these notions have been taken from Gneiting (2011) who in turn credits Osband (1985). In a more implicit form, it is also used by e.g. Lambert (2012).

In the following we are not only interested in properties for which there exists a consistent scoring function, but also in properties, for which there is an identification function. This motivates the following definition.

**Definition 5** *Let $\mathcal{P}$ be a set of probability measures on $(Y, \mathcal{A})$. A property $T : \mathcal{P} \to \mathbb{R}$ is said to be* identifiable*, if there exists a $\mathcal{P}$-identification function for $T$*

For later use note that two properties $T_1, T_2$ on $\mathcal{P}$ having the same strong $\mathcal{P}$-identification function are necessarily equal, that is $T_1 = T_2$. Furthermore, multiplying an (oriented) identification function $V$ by a strictly positive weight $w : A \to (0, \infty)$ gives another (oriented) identification function $wV$. Moreover, the following lemma shows that either $V$ or $-V$ is actually oriented.

**Lemma 6** *Let $\mathcal{P}$ be a convex and topological, $T : \mathcal{P} \to \mathbb{R}$ be a continuous, quasi-monotonic property, and $V$ be a $\mathcal{P}$-identification function $T$. Then either $V$ or $-V$ is an oriented.*

Intuitively, there is a close connection between scoring and identification functions. Indeed, assume that we can naïvely take the derivative of the $S$-risks, that is

$$\frac{\partial \mathbb{E}_{\mathsf{Y} \sim P} S(t, \mathsf{Y})}{\partial t} = \mathbb{E}_{\mathsf{Y} \sim P} S'(t, \mathsf{Y}), \tag{5}$$

where $S'$ denotes the derivative of $S$ with respect to the first argument. For $t^* := T(P)$, the consistency (1) of $S$ then implies $\mathbb{E}_{\mathsf{Y} \sim P} S'(t^*, \mathsf{Y}) = 0$. Unfortunately, the required converse implication is in general not easy to show, see the discussion following Theorem 7, and, of course, (5) only holds under additional assumptions. Interestingly, however, if we start with an oriented identification function $V$ then its anti-derivative is an order sensitive scoring function, and thus consistent.

To present a corresponding formal statement we call, analogously to loss functions, a scoring function $S : A \times Y \to \mathbb{R}$ *locally Lipschitz continuous*, if for all intervals $[a, b] \subset A$ there exists a constant $c_{a,b} \geq 0$ such that, for all $t_1, t_2 \in [a, b]$ and all $y \in Y$, we have

$$\left| S(t_1, y) - S(t_2, y) \right| \leq c_{a,b} \left| t_1 - t_2 \right|.$$

Similarly, we say that a function $V : A \times Y \to \mathbb{R}$ is *locally bounded*, if, for all $[a, b] \subset A$, the restriction $V_{|[a,b] \times Y}$ of $V$ onto $[a, b] \times Y$ is bounded. Furthermore, we need to extend derivatives that are only almost everywhere defined. To make this precise, let $S : A \times Y \to \mathbb{R}$ be a function and $D \subset A \times Y$ be the set on which $S$ is differentiable in its first variable. Then the canonical extension $\hat{S}' : A \times Y \to \mathbb{R}$ of the derivative $S'$ of $S$ is defined by

$$\hat{S}'(t, y) := \begin{cases} S'(t, y) & \text{if } (t, y) \in D \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Finally, for a measure $\nu$ on $A$, an $f \in L_1(\nu)$, and $a, b \in \mathbb{R}$ we need the following notation

$$\int_a^b f \, d\nu := \mathrm{sign}(b - a) \int_{(a \wedge b, a \vee b]} f \, d\nu.$$

With these preparations we can now construct order sensitive scoring functions from identification functions.

**Theorem 7**  *Let $(Y, \mathcal{A}, \mu)$ be a $\sigma$-finite measure space, $\mathcal{P}$ be a set of $\mu$-absolutely continuous distributions on $(Y, \mathcal{A})$, and $T : \mathcal{P} \to \mathbb{R}$ be a property such that $\operatorname{im} T$ is an interval. Moreover, let $V : A \times Y \to \mathbb{R}$ be a measurable, locally bounded, and oriented $\mathcal{P}$-identification function for $T$, and $\nu$ be a measure on $A$ with $\nu \ll \lambda$ whose $\lambda$-density $w$ is locally bounded. For some fixed $t_0 \in A$ and $\kappa : Y \to \mathbb{R}$ with $\kappa \in L_1(P)$ for all $P \in \mathcal{P}$, we define $S : A \times Y \to \mathbb{R}$ by*

$$S(t, y) := \int_{t_0}^{t} V(r, y) \, d\nu(r) + \kappa(y), \qquad (t, y) \in A \times Y. \qquad (7)$$

*Then the following statements hold:*

*i)  The map $S : A \times Y \to \mathbb{R}$ is measurable and locally Lipschitz continuous. Moreover, for all $y \in Y$, the Lebesgue almost everywhere defined derivative $S'(\cdot, y) : A \to \mathbb{R}$ satisfies*

$$S'(t, y) = w(t)V(t, y). \qquad (8)$$

*In particular, its extension $\hat{S}'$ defined by (6) is a measurable and oriented $\mathcal{P}$-identification function for $T$ if and only if $w(t) > 0$ for $\lambda$-almost all $t \in \operatorname{im} T$, that is, if and only if, $\mu(\{y \in Y : \hat{S}'(t, y) \neq 0\}) > 0$ for $\lambda$-almost all $t \in \operatorname{im} T$.*

*ii)  The map $S$ is $\mathcal{P}$-order sensitive, if and only if $\nu$ is strictly positive.*

Let us assume for a moment that we are in the situation of Theorem 7. In addition, assume that $V$ is actually bounded and that $\nu$ is finite. Then, using the function $\kappa : Y \to \mathbb{R}$ defined by

$$\kappa(y) := \int_A |V(r, y)| \, d\nu(r) < \infty \qquad (9)$$

in (7) gives $S(t, y) \geq 0$ for all $t \in A$ and $y \in Y$. In other words, $S$ is an order preserving (and thus consistent) loss function.

The construction (7) is also possible for certain measures $\nu$ that are *not* Lebesgue absolutely continuous. For example, the order sensitivity can be ensured for all measures $\nu$ with $0 < \nu((t_1, t_2]) < \infty$ for all $t_1, t_2 \in \mathring{A}$ with $t_1 < t_2$, if $\nu$ additionally satisfies $\nu(N) = 0$ for the set $N$ excluded in (3). While it turns out that for such measures $S$ is still measurable, the other properties listed in *i)* may be lost. In particular, $S(\cdot, y)$ is, in general, no longer almost everywhere differentiable.

Interestingly, $\hat{S}'$ is not always an identification function for $S$ of the form (7), since there exist strictly positive measures $\nu \ll \lambda$ whose densities are not $\lambda$-almost everywhere strictly positive. For example, take an enumeration $(q_n)$ of $[0, 1] \cap \mathbb{Q}$ and consider the density $w := \mathbf{1}_A$, where $A := \bigcup_{n \geq 1} [q_n - 5^{-n}, q_n + 5^{-n}]$. Since $\lambda(A) < 1$, we then see that $w$ is not $\lambda$-almost everywhere strictly positive, but the denseness of $(q_n)$ in $[0, 1]$ shows that $\nu := wd\lambda$ is strictly positive.

## 4. Existence of Scoring and Identification Functions

In this section we show that modulo some technical assumptions, continuous, quasi-monotonic properties that are defined on the set of bounded densities are elicitable. Moreover, we characterize the set of corresponding order-sensitive scoring functions. Finally, we illustrate the developed theory by characterizing the order sensitive scoring functions for generalized quantiles.

In the following, let $(Y, \mathcal{A}, \mu)$ be a finite measure space, that is $\mu(Y) < \infty$. Recall that $(Y, \mathcal{A}, \mu)$ is separable, if there exists a countable family $(A_i) \subset \mathcal{A}$ such that, for all $A \in \mathcal{A}$ and $\varepsilon > 0$, there exists an $A_i$ such that

$$\mu(A \triangle A_i) \leq \varepsilon. \tag{10}$$

Clearly, this is satisfied if and only if the family $(\mathbf{1}_{A_i})$ is dense in the set of all measurable indicator functions with respect to the norm $\| \cdot \|_{L_p(\mu)}$, where $p \in [1, \infty)$. From the latter it is not hard to conclude that $(Y, \mathcal{A}, \mu)$ is separable, if and only if, for all $1 \leq p \leq \infty$, the space $L_p(\mu)$ is separable. Let us denote the set of bounded probability densities with respect to $\mu$ by

$$\Delta^{\geq 0} := \{ h \in L_\infty(\mu) : h \geq 0, \, \mathbb{E}_\mu h = 1 \},$$

and analogously, we write $\Delta^{>0} := \{ h \in L_\infty(\mu) : h \geq \varepsilon \text{ for some } \varepsilon > 0 \text{ and } \mathbb{E}_\mu h = 1 \}$. In the following, we always consider either $\Delta := \Delta^{\geq 0}$ or $\Delta := \Delta^{>0}$. We write

$$\mathcal{P}(\Delta) := \left\{ P : \exists h \in \Delta \text{ such that } P = h d\mu \right\}$$

for the corresponding set of probability measures on $Y$. For $p \in [1, \infty]$, we further write $\Delta_p$, when $\Delta$ is viewed as a metric space with metric $\| \cdot \|_{L_p(\mu)}$, and analogously, $\mathcal{P}(\Delta_p)$ denotes the set $\mathcal{P}(\Delta)$ equipped with the metric induced by $\| \cdot \|_{L_p(\mu)}$, that is

$$\| P_1 - P_2 \|_{L_p(\mu)} := \| h_1 - h_2 \|_{L_p(\mu)}$$

for $P_1 = h_1 d\mu$ and $P_2 = h_2 d\mu \in \Delta$. Note that the metric on $\mathcal{P}(\Delta_1)$ is the total variation norm. With these preparations we can now formulate the following technical assumption on a property.

**Definition 8** *A property $T : \mathcal{P}(\Delta) \to \mathbb{R}$ is* strictly locally non-constant*, if for all $t \in \overset{\circ}{\mathrm{im}}\, T, \varepsilon > 0$, and $P \in \{T = t\}$, there exist a $P_- \in \{T < t\}$ and a $P_+ \in \{T > t\}$ such that $\| P - P_\pm \|_{L_\infty(\mu)} \leq \varepsilon$.*

The definition above basically ensures that for each distribution we can suitably change the density to change the property. A very similar assumption is used by Lambert (2012).

Our next goal is to show that under the assumptions above, there exists an oriented identification function.

**Theorem 9** *Let $(Y, \mathcal{A}, \mu)$ be a separable and finite measure space and $T : \mathcal{P}(\Delta_1) \to \mathbb{R}$ be a continuous, quasi-monotonic property. Assume that $\mathrm{im}\, T$ is equipped with $\hat{\mathcal{B}}(\mathrm{im}\, T)$. Then, if $T$ is strictly locally non-constant, the following statements are true:*

*i) There exists a measurable and oriented $\mathcal{P}(\Delta)$-identification function $V^* : \mathrm{im}\, T \times Y \to \mathbb{R}$ for $T$ such that for Lebesgue-almost all $t \in \mathrm{im}\, T$ we have*

$$\| V^*(t, \cdot) \|_{L_\infty(\mu)} = 1 .$$

*ii) If $V : \mathrm{im}\, T \times Y \to \mathbb{R}$ is a measurable oriented $\mathcal{P}(\Delta)$-identification for $T$, then there exists a measurable $w : \mathrm{im}\, T \to (0, \infty)$ such that, for $\lambda \otimes \mu$-almost all $(t, y) \in \overset{\circ}{\mathrm{im}}\, T \times Y$, we have*

$$V(t, y) = w(t) V^*(t, y) . \tag{11}$$

iii) *If* $S : \operatorname{im} T \times Y \to \mathbb{R}$ *is a measurable, locally Lipschitz continuous, and* $\mathcal{P}(\Delta)$-*consistent scoring function for* $T$, *then, for* $\lambda \otimes \mu$-*almost all* $(t, y) \in \operatorname{im} T \times Y$, *the derivative* $S'(t, y)$ *exists. Furthermore, there exists a measurable and locally bounded* $w : \operatorname{im} T \to \mathbb{R}$, *such that, for* $\lambda \otimes \mu$-*almost all* $(t, y) \in \operatorname{im}^{\circ} T \times Y$, *we have*

$$S'(t, y) = w(t) V^*(t, y). \tag{12}$$

*Finally,* $S$ *is* $\mathcal{P}(\Delta)$-*order sensitive for* $T$, *if and only if* $w \geq 0$ *and the measure* $\nu := w d\lambda$ *is strictly positive.*

Unfortunately, the proof of the existence of $V^*$ is anything than constructive, since it relies on Hahn-Banach's theorem. Nonetheless, in specific situations $V^*$ can be found by elementary considerations. Before we can present some examples, let us first combine Theorem 9 with the results from the previous sections:

**Corollary 10** *Let* $(Y, \mathcal{A}, \mu)$ *be a separable, finite measure space and* $T : \mathcal{P}(\Delta_1) \to \mathbb{R}$ *be a continuous, strictly locally non-constant property. Then the following statements are equivalent:*

i) *For all* $t \in \operatorname{im} T$, *the level set* $\{T = t\}$ *is convex.*

ii) $T$ *is quasi-monotonic.*

iii) $T$ *is identifiable and has a bounded identification function.*

iv) $T$ *is elicitable.*

v) *There exists a non-negative, measurable, locally Lipschitz continuous scoring function that is* $\mathcal{P}(\Delta)$-*order sensitive for* $T$.

*Moreover, if* $T$ *is elicitable, then every measurable, locally Lipschitz continuous scoring function* $S : \operatorname{im} T \times \mathbb{R} \to \mathbb{R}$ *that is* $\mathcal{P}(\Delta)$-*order sensitive for* $T$ *is of the form*

$$S(t, y) = \int_{t_0}^{t} V^*(r, y) \, w(r) \, dr + \kappa(y), \qquad (t, y) \in \operatorname{im} T \times (Y \setminus N) \tag{13}$$

*where* $V^*$ *is the identification function from Theorem 9,* $t_0 \in \operatorname{im} T$, $w \geq 0$ *is measurable and locally bounded such that* $\nu := w d\lambda$ *is strictly positive,* $\kappa : Y \to \mathbb{R}$ *is a function with* $\kappa \in L_1(P)$ *for all* $P \in \mathcal{P}$, *and* $N \subset Y$ *is measurable with* $\mu(Y) = 0$.

Note that the variability of $t_0$ in (13) is actually superfluous. Indeed, if we pick a $w$ satisfying the assumptions mentioned above and we have, e.g. $t_0 < t_1$, then, for all $t \in \operatorname{im} T$, we find

$$\int_{t_0}^{t} V^*(r, y) \, w(r) \, dr = \int_{t_1}^{t} V^*(r, y) \, w(r) \, dr + \int_{t_0}^{t_1} V^*(r, y) \, w(r) \, dr,$$

and since the second integral on the right hand side does not depend on $t$ anymore, it can simply be viewed as part of the offset function $\kappa$.

To give some concrete examples of the theory developed so far, let us fix an interval $Y := [a, b]$ and equip it with the Lebesgue measure, i.e. $\mu := \lambda$. For $\tau \in (0, 1)$, recall that the $\tau$-quantile for a distribution $P \in \mathcal{P}(\Delta^{>0})$ is the unique solution $T(P) := t^* \in [a, b]$ of the set of equations

$$P\big((-\infty, t)\big) = \tau \qquad \text{and} \qquad P\big((t, \infty)\big) = 1 - \tau.$$

Clearly this $t^*$ solves the equation $(1 - \tau)\mathbb{E}_P\mathbf{1}_{(-\infty,t)} = \tau\mathbb{E}_P\mathbf{1}_{(t,\infty)}$, and consequently

$$V(t, y) := (1 - \tau)\mathbf{1}_{(-\infty,t)}(y) - \tau\mathbf{1}_{(t,\infty)}(y)\,, \qquad\qquad t, y \in [a, b],$$

is, modulo an obvious normalization constant, the only candidate for $V^*$. Moreover, the function $t \mapsto (1 - \tau)\mathbb{E}_P\mathbf{1}_{(-\infty,t)}$ is strictly increasing in $t$, while $\tau\mathbb{E}_P\mathbf{1}_{(t,\infty)}$ is strictly decreasing in $t$, and from this it is easy to conclude that $V$ is indeed a (strong) identification function for the $\tau$-quantile. Let us now find all measurable, locally Lipschitz continuous and $\mathcal{P}(\Delta^{>0})$-order sensitive scoring functions. To this end, we first observe that we can replace $V^*$ by $V$ in (13), since the weight $w$ in (12) is bounded away from zero and infinity by the specific form of $V$. Now, we set $t_0 := a$ and fix a measurable, locally bounded $w \geq 0$ such that $wd\lambda$ is strictly positive. Let us further denote the anti-derivative of $w$ by $g$, that is $g(r) := \int_a^r w(s)ds$ for $r \in [a, b]$. By the assumptions made on $w$, we then see that $g$ is non-negative, strictly increasing, and locally Lipschitz with $g(a) = 0$. Conversely, it is not hard to see that every $g$ satisfying the latter set of assumptions is an anti-derivative of the form above. Now, for $a \leq t \leq y$, we have

$$\int_a^t V(r, y)w(r)dr + \tau g(y) = -\tau\int_a^t w(r)dr + \tau g(y) = \tau\big(g(y) - g(t)\big)\,, \qquad (14)$$

while for $a \leq y \leq t$ we obtain

$$\int_a^t V(r, y)w(r)dr + \tau g(y) = \int_a^y V(r, y)w(r)dr + \int_y^t V(r, y)w(r)dr + \tau g(y)$$
$$= (1 - \tau)\big(g(t) - g(y)\big)\,. \qquad (15)$$

Combining both expressions and adding an offset function $\kappa$ gives the general form

$$S(t, y) = \big|\mathbf{1}_{(-\infty,t]}(y) - \tau\big| \cdot \big|g(t) - g(y)\big| + \kappa(y) \qquad (16)$$

of all measurable, locally Lipschitz continuous and $\mathcal{P}(\Delta^{>0})$-order sensitive scoring functions for the $\tau$-quantile. Here $g$ is an arbitrary non-negative, strictly increasing, and locally Lipschitz function on $[a, b]$. Clearly, $S$ is Lipschitz, if and only if $g$ is Lipschitz, and for such $S$, the form (16) coincides with the representation found by Lambert (2012), while for differentiable $g$ it coincides with that of Grant and Gneiting (2013). Moreover, by considering $g(r) := r$ and $\kappa = 0$, we obtain the well-known $\tau$-pinball loss. Finally, note that by (14) and (15), an $S$ of the form (16) is convex in $t$, if and only if $g$ is both concave and convex. This leads to the following corollary.

**Corollary 11** *For each interval $Y = [a, b]$, the $\tau$-pinball loss is, modulo a constant factor and an offset function, the only locally Lipschitz continuous and convex scoring function that is $\mathcal{P}(\Delta^{>0})$-order sensitive for the $\tau$-quantile.*

Our next goal is to generalize these considerations to so-called generalized quantiles considered in e.g. Bellini et al. (2014) because of their importance as a risk measure for financial applications. To this end, let $\Phi_-, \Phi_+ : [0, \infty) \to [0, \infty)$ be strictly convex and strictly increasing functions satisfying $\Phi_i(0) = 0$ and $\Phi_i(1) = 1$ for $i = \pm$. Then, for $\tau \in (0, 1)$, the generalized $\tau$-quantile of a $P \in \mathcal{P}(\Delta^{\geq 0})$ is the unique solution of

$$t^* = \underset{t \in \mathbb{R}}{\arg\min}\,(1 - \tau)\mathbb{E}_{\mathsf{Y} \sim P}\Phi_-\big((\mathsf{Y} - t)^-\big) + \tau\mathbb{E}_{\mathsf{Y} \sim P}\Phi_+\big((\mathsf{Y} - t)^+\big)\,. \qquad (17)$$

Note that unlike Bellini et al. (2014) we assumed that the functions $\Phi_-, \Phi_+$ are not only convex but strictly convex to ensure that (17) has a *unique* solution for all $P \in \mathcal{P}(\Delta^{\geq 0})$. Of course, this excludes the quantiles, which, however, we have already treated above. Probably the best-known example of generalized quantiles are expectiles, see (Newey and Powell, 1987), that correspond to the choice $\Phi_-(r) = \Phi_+(r) = r^2$ for $r \geq 0$. Clearly, generalized quantiles are elicitable, since (17) directly translates into an optimization problem of the form (1) for the score function

$$S(t,y) := (1 - \tau)\Phi_-\big((y - t)^-\big) + \tau\Phi_+\big((y - t)^+\big), \qquad t, y \in [a, b]. \qquad (18)$$

Note that for expectiles, this $S$ becomes the asymmetric least squares loss, which has recently attracted some interest in the machine learning community, see e.g. Huang et al. (2014). In the following, our goal is to characterize all order sensitive scoring functions for generalized quantiles. To keep the corresponding calculations brief, we restrict our considerations to the case $\Phi_- = \Phi_+$. Let $\Phi : \mathbb{R} \to [0, \infty)$ be the symmetric extension of $\Phi_+$, that is $\Phi(r) := \Phi_+(|r|)$ for $r \in \mathbb{R}$. Now assume that $\Phi$ is continuously differentiable, and that its derivative $\psi := \Phi'$ is absolutely continuous. Then Corollary 3 of Bellini et al. (2014) implies that the canonical extension of $S'$, which for $y \neq t$ is given by

$$S'(t,y) = (1 - \tau)\psi\big((t - y)^+\big) - \tau\psi\big((y - t)^+\big), \qquad (19)$$

is a corresponding (oriented) identification function. By some simple considerations we further find

$$\min\{1 - \tau, \tau\} \cdot \|\psi_{|[0,(b-a)/2]}\|_\infty \leq \|S'(t, \cdot)\|_\infty \leq \|\psi_{|[0,b-a]}\|_\infty$$

for all $t \in [a, b]$, and therefore the weight $w$ in (12) is bounded away from zero and infinity. In (13) we can thus replace $V^*$ by $\hat{S}'$. Now, we set $t_0 := a$ and fix a measurable, locally bounded $w \geq 0$ such that $wd\lambda$ is strictly positive. Let us further denote the anti-derivative of $w$ by $g$, that is $g(r) := \int_a^r w(s)ds$ for $r \in [a, b]$. Furthermore, we define

$$G(t,y) := \int_a^t \psi'(y - r)g(r)dr = \int_a^t \psi'(r - y)g(r)dr, \qquad t, y \in [a, b]. \qquad (20)$$

where the last identity follows from the symmetry of $\Phi$, which implies $\psi'(-r) = \psi'(r)$ for all $r \in \mathbb{R}$. Now, for $a \leq t \leq y$, we have

$$\int_a^t S'(r,y)w(r)dr + \tau G(y,y) = -\tau \int_a^t \psi(y - r)g'(r)dr = \tau\big(G(y,y) - G(t,y) - g(t)\psi(y - t)\big)$$

by integration by parts, see e.g. (Bogachev, 2007a, Corollary 5.4.3) for the case of absolutely continuous functions. Similarly, for $a \leq y \leq t$ we obtain

$$\int_a^t S'(r,y)w(r)dr + \tau G(y,y) = (1 - \tau)\big(G(y,y) - G(t,y) - g(t)\psi(y - t)\big).$$

Combining both expressions and adding an offset function $\kappa$ gives the general form

$$S(t,y) = \big|\mathbf{1}_{(-\infty,t]}(y) - \tau\big| \cdot \big(G(y,y) - G(t,y) - g(t)\psi(y - t)\big) + \kappa(y) \qquad (21)$$

of all measurable, locally Lipschitz continuous and $\mathcal{P}(\Delta^{\geq 0})$-order sensitive scoring functions for the generalized $\tau$-quantile. Here $g$ is an arbitrary non-negative, strictly increasing, and locally Lipschitz function on $[a, b]$, and $G$ is given by (20).

In some cases, the function $G$ can be explicitly calculated. For example, for expectiles, we have $\Phi(r) = r^2$, and thus $\psi(r) = 2r$ and $\psi'(r) = 2$. Consequently, $G(\cdot, y)$ equals, independently of $y$, the anti-derivative of $2g$, and (21) coincides with the characterization by Gneiting (2011). More generally, for $\Phi(r) = r^n$ with $n \in \mathbb{N}$ and $n \geq 2$, $G$ can be computed using induction. For example, for $n = 3$, we have $G(t, y) = 6|g^{-''}(y) - g^{-''}(t) - g^{-'}(t)(y - t)|$, where $g^{-'}$ and $g^{-''}$ denote the first and second anti-derivative of $g$, respectively.

Finally, note that the above calculations are an example of how to solve the following general question: Given a scoring function $S_0$ and a resulting property $T$, which other scoring functions $S$ can be used to find $T$? Note that such surrogate scoring functions $S$ may be desirable, for example, to find an efficient learning algorithm or to better control statistical behaviour, or robustness of an estimation procedure. With the developed theory, the answer to the question above is, ignoring the described technicalities, straightforward: First compute the derivative $S_0'$, then normalize it such that it becomes $V^*$, and then compute all $S$ by (13).

## 5. Expectiles

In this section we negatively answer an open question recently posed by Ziegel (2014): Is there any coherent, law-invariant, and elicitable property other than expectiles? Bellini and Bignozzi (2013) (confer Bellini et al. (2014)) have recently presented a similar result, but under stronger hypotheses.

To begin, we recall the notion of coherent risk measures, see Rockafellar and Uryasev (2013). To this end, we fix probability space $(\Omega, \mathcal{A}, \nu)$ and write $\mathcal{P} := \{P_Y : Y \in L_0(\nu)\}$ for the set of all distributions of random variables $Y : \Omega \to \mathbb{R}$. Given a property $T : \mathcal{P} \to \mathbb{R}$, we further write, in a slight abuse of notations, $T(Y) := T(P_Y)$ for all $Y \in L_0(\nu)$, where $L_0(\nu)$ denotes the space of all $\nu$-equivalence classes of measurable $Y : \Omega \to \mathbb{R}$. Thus, we can view $T$ as a map $T : L_0(\nu) \to \mathbb{R}$. In the literature, such maps that factor through $\mathcal{P}$ are called *law-invariant*. Let us consider the following features of $T$, that are assumed to be satisfied for all $Y, Y' \in L_0(\nu)$, $\lambda > 0$, and $c \in \mathbb{R}$:

**T0 (definite).** $T(0) = 0$.
**T1 (translation equivariant).** $T(Y + c) = T(Y) + c$
**T2 (positively homogeneous).** $T(\lambda Y) = \lambda T(Y)$
**T3 (subadditive).** $T(Y + Y') \leq T(Y) + T(Y')$
**T4 (monotonic).** $T(Y) \leq T(Y')$ whenever $Y \leq Y'$
**T5 (convex).** $T((1 - t)Y + tY') \leq (1 - t)T(Y) + tT(Y')$

If $-T$ satisfies **T0** to **T4**, then $T$ is called a *coherent risk measure*. The following theorem partially describes the identification function of identifiable properties satisfying some of these assumptions.

**Theorem 12** *Let $(\Omega, \mathcal{A}, \nu)$ be an atom-free measure space, $\mathcal{P} := \{P_Y : Y \in L_0(\nu)\}$, and $T : \mathcal{P} \to \mathbb{R}$ be an identifiable property. If $T$ satisfies **T0** and **T1**, then the following statements are true:*

*i) There exists a $\psi : \mathbb{R} \to \mathbb{R}$ with $\psi(0) = 0$ and $\psi(-1) = 1$ such that $V : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by*

$$V(t, y) := \psi(y - t), \qquad y, t \in \mathbb{R} \tag{22}$$

*is an oriented $\mathcal{P}$-identification function for $T$. Moreover, $\psi(s) < 0$ if and only if $s > 0$.*

*ii) If $T$ also satisfies **T2**, then we have $\psi(1)\psi(st) = \psi(s)\psi(t)$ and $\psi(s) = \psi(1)\psi(-s)$ for all $s, t > 0$. In addition, there exists an $s_0 > 0$ with $\psi(s_0) \neq \psi(1)$.*

*iii)* If $T$ also satisfies **T3** and **T4**, then $\psi$ considered in (22) is decreasing on $(0, \infty)$.

*iv)* If $T$ also satisfies **T2** and $\psi$ is decreasing on $(0, \infty)$, then there exists an $\alpha > 0$ such that $\psi$ is given by

$$\psi(s) = \begin{cases} \psi(1)s^\alpha & \text{if } s \geq 0 \\ (-s)^\alpha & \text{if } s \leq 0. \end{cases} \tag{23}$$

*v)* If $T$ also satisfies **T5**, and $\psi$ considered in (22) is continuous and surjective, then $\psi$ is concave.

With the help of the theorem above, we can now present the following result that answers the question raised by Ziegel (2014):

**Corollary 13** *Let $(\Omega, \mathcal{A}, \nu)$ be an atom-free measure space, $\mathcal{P} := \{P_Y : Y \in L_0(\nu)\}$, and $T : \mathcal{P} \to \mathbb{R}$ be an identifiable property satisfying **T0** to **T4**. Then $T$ is an $\tau$-expectile for some $\tau \geq 1/2$.*

Note that Bellini et al. (2014) only gave a partial answer to Zeigel's question. Namely, they showed that the only coherent generalized $\tau$-quantiles are expectiles.

## 6. Some Proofs

**Proof** [Lemma 1] For a fixed $\varepsilon > 0$ we define $t_\varepsilon^- := t^* - \varepsilon$ and $t_\varepsilon^+ := t^* + \varepsilon$. Furthermore, we define

$$\delta := \min\{\mathbb{E}_{Y \sim P}S(t_\varepsilon^-, Y), \mathbb{E}_{Y \sim P}S(t_\varepsilon^*, Y)\} - \mathbb{E}_{Y \sim P}S(t^+, Y).$$

Setting $t_1 := t^*$ and $t_2 := t_\varepsilon^\pm$, we then see by (2) that $\delta > 0$. Let us now fix a $t \in A$ with $t \geq t^*$ and $|\mathbb{E}_{Y \sim P}S(t, Y) - \mathbb{E}_{Y \sim P}S(t^*, Y)| < \delta$. By the definition of $\delta$, we then obtain $\mathbb{E}_{Y \sim P}S(t, Y) < \mathbb{E}_{Y \sim P}S(t_\varepsilon^+, Y)$, and using (2) we conclude that $t \leq t_\varepsilon^+$. In other words we have shown $|t - t^*| \leq \varepsilon$. The case $t \leq t^*$ can be treated analogously. ∎

**Proof** [Corollary 10] *iii)* $\Rightarrow$ *v)*. Follows from Theorem 7 and (9) by using a strictly positive, bounded and Lebesgue integrable $w : \operatorname{im} T \to \mathbb{R}$.

*v)* $\Rightarrow$ *iv)*. Trivial.

*iv)* $\Rightarrow$ *i)*. Theorem 3.

*i)* $\Rightarrow$ *ii)*. Theorem 4.

*ii)* $\Rightarrow$ *iii)*. Theorem 9.

Finally, to show 13, we assume that $T$ is elicitable and fix a measurable, locally Lipschitz continuous scoring function $S : \operatorname{im} T \times \mathbb{R} \to \mathbb{R}$ that is $\mathcal{P}(\Delta)$-order sensitive for $T$. By part *ii)* of Theorem 9 we then find a measurable and locally bounded $w : \operatorname{im} T \to [0, \infty)$ such that (12) holds and $\nu := wd\lambda$ is strictly positive. Consequently, there exists a measurable $N \subset Y$ with $\mu(Y) = 0$ such that for all $y \in Y \setminus N$ there exists a measurable $N_y \subset \operatorname{im}^\circ T$ with $\lambda(N_y) = 0$ such that (12) holds for all $t \in \operatorname{im}^\circ T \setminus N_y$. The fundamental theorem of calculus for absolutely continuous functions, see e.g. (Bogachev, 2007a, Theorems 5.3.6 and 5.4.2), yields

$$S(t, y) - S(t_0, y) = \int_{t_0}^t S'(r, y)\, dr = \int_{t_0}^t w(r)V^*(r, y)\, dr$$

for all $t \in \operatorname{im} T$ and $y \in Y \setminus N$. By setting $b(y) := S(t_0, y)$, we then see that $S$ is of the form (13). ∎

**Proof** [Corollary 13] We apply Theorem 12: since $T$ satisfies **T0** to **T4**, it has an identification function $V$ of the form (22), where $\psi$ is given by (23) for some $\alpha > 0$ and $\psi(1) < 0$. However, for $\alpha > 0$, the function $\psi$ is continuous and and surjective. **T2** and **T3** imply **T5**, and therefore, $\psi$ is concave. However, the only $\psi$ of the form (23) that is concave, is that for $\alpha = 1$ and $\psi(1) \leq -1$. In the case $\psi(1) = -1$, we immediately see that $\psi$ is the identification function of the $1/2$-expectile. Moreover, if $\psi(1) < -1$, then multiplying $\psi$ by $\frac{1}{1-\psi(1)}$, we see that $\psi$ equals the identification function for the $\tau$-expectile with $\tau = \frac{\psi(1)}{\psi(1)-1}$. Finally, using $\psi(1) < 1$ we find $\tau > 1/2$. ∎

## Acknowledgments

## References

Jacob D Abernethy and Rafael M Frongillo. A characterization of scoring rules for linear properties. *COLT 2012. Journal of Machine Learning Research — Workshop and Conference Proceedings*, 23:27.1–27.13, 2012.

Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

H. Bauer. *Measure and Integration Theory*. De Gruyter, Berlin, 2001.

F. A. Behringer. Convexity is equivalent to midpoint convexity combined with strict quasiconvexity. *Optimization*, 24:219–228, 1992.

F. Bellini, B. Klar, A. Müller, and R. E. Gianin. Generalized quantiles as risk measures. *Insurance Math. Econom.*, 54:41–48, 2014.

Fabio Bellini and Valeria Bignozzi. Elicitable risk measures. Preprint, RiskLab, Department of Mathematics, ETH Zurich, December 2013.

V. I. Bogachev. *Measure Theory, Vol. I*. Springer-Verlag, Berlin, 2007a.

V. I. Bogachev. *Measure Theory, Vol. II*. Springer-Verlag, Berlin, 2007b.

C. Castaing and M. Valadier. *Convex Analysis and Measurable Multifunctions*. Springer, Berlin, 1977.

A.Philip Dawid. The geometry of proper scoring rules. *Annals of the Institute of Statistical Mathematics*, 59(1):77–93, March 2007.

N. Dinculeanu. *Vector Integration and Stochastic Integration in Banach Spaces*. John Wiley & Sons, New York, 2000.

Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

Tilmann Gneiting and Matthias Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and its Applications*, 1:125–151, 2014.

Jun-ya Gotoh and Stan Uryasev. Support vector machines based on convex risk functionals and general norms. Technical Report 2013-6, Risk Management and Financial Engineering Lab, University of Florida, November 2013.

Jun-ya Gotoh, Akiko Takeda, and Rei Yamamoto. Interaction between financial risk measures and machine learning methods. *Computational Management Science*, pages 1–38, June 2013. doi: 10.1007/s10287-013-0175-5.

K. Grant and T. Gneiting. *Consistent scoring functions for quantiles*, volume Volume 9, pages 163–173. Institute of Mathematical Statistics, Beachwood, Ohio, 2013.

H. J. Greenberg and W. P. Pierskalla. A review of quasi-convex functions. *Oper. Res.*, 19:1553–1570, 1971.

Xiaolin Huang, Lei Shi, and Johan A.K. Suykens. Asymmetric least squares support vector machine classifiers. *Computational Statistics and Data Analysis*, 70:395–405, 2014.

P.J. Huber. *Robust Statistics*. Wiley, New York, 1981.

M.C. Jones. Expectiles and M-quantiles are quantiles. *Statistics and Probability Letters*, 20:149–153, 1994.

Ingo Klein and Michael Grottke. On J.M. Keynes' "The Principal Averages and the Laws of Error which Lead to Them" - Refinement and Generalisation. Technical Report 07/2008, Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung, Friedrich-Alexander-Universität, 2008.

S. Kusuoka. On law-invariant coherent risk measures. *Adv. Math. Econ*, 3:83–95, 2001.

Nicolas Lambert, David Pennock, and Yoav Shoham. Eliciting properties of probability distributions. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 129–138, 2008.

Nicolas S. Lambert. Elicitation and evaluation of statistical forecasts. Preprint, Stanford Graduate School of Business, April 2012.

John McCarthy. Measures of the value of information. *Proceedings of the National Academy of Sciences*, 42:654–655, 1956.

R. E. Megginson. *An Introduction to Banach Space Theory*. Springer, New York, 1998.

Whitney K. Newey and James L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847, July 1987.

Kent Harold Osband. *Providing Incentives for Better Cost Forecasting*. PhD thesis, University of California, Berkeley, 1985.

A. Pietsch. *Eigenvalues and s-Numbers*. Geest & Portig K.-G., Leipzig, 1987.

R. Tyrell Rockafellar. Coherent approaches to risk in optimization under uncertainty. *Tutorials in operations research*, pages 38–61, 2007.

R. Tyrell Rockafellar and Stan Uryasev. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, 18(1):33–53, 2013.

Leonard J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

M.J. Schervish. A general method for comparing probability assessors. *The Annals of Statistics*, 17 (4):1856–1879, 1989.

I. Steinwart. How to compare different loss functions. *Constr. Approx.*, 26:225–287, 2007.

I. Steinwart and A. Christmann. *Support Vector Machines*. Springer, New York, 2008.

I. Steinwart and A. Christmann. Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*, 17:211–225, 2011.

Peter Tsyarmasto and Stan Uryasev. Advanced risk measures in estimation and classification. In *Proceedings of the International Workshop "Stochastic Programming for Implementation and Advanced Applications"*, pages 114–118, Vilnius, July 2012. URL http://www.moksloperiodika.lt/STOPROG_2012/abstract/114-118-Tsyar.pdf.

Johanna F. Ziegel. Coherence and elicitability. *Mathematical Finance*, 2014. To appear.

## Appendix A. Comparison with (Lambert, 2012)

Let us briefly summarize the differences between (Lambert, 2012) and the present paper:

- As mentioned earlier, (Lambert, 2012) considers distributions with continuous densities on compact metric spaces. Our results of Section 4 require significantly weaker assumptions on the considered distributions. In fact, we do not need any sort of topological assumption.

- Lambert only considers Lipschitz continuous scoring functions in the representation and requires a scoring function to begin with (instead of an identification function). However, this is a lesser difference as we think that his paper could be easily changed along these lines.

- There is a bug in his proof on page 46, 4 lines from the bottom of the page. Here, Lambert has a family of normalized functionals $\Phi_\theta$ such that $\Phi_\theta(f) \to \Phi_{\theta_0}(f)$ for all $f \in \ker \Phi_{\theta_0}(f)$. He then claims that this implies $|\Phi_\theta(v)| \to 1$ for some $v$ with $\Phi_{\theta_0}(v) = 1$.

  In finite dimensional spaces this is true: by compactness there exists a subsequence $\Phi_{\theta_n}$ that converges to some $\Phi$ in norm and thus $\|\Phi\| = 1$. Our Lemma 33 shows $\Phi = \alpha \Phi_{\theta_0}$ for some $\alpha \in \mathbb{R}$ and by comparing norms we obtain $|\alpha| = 1$. Moreover, the last arguments actually hold for all convergent subsequences, and thus the assertion follows.

In infinite dimensional spaces, this argument no longer works: Depending on the involved spaces, one only gets weak- or weak*-convergent subsequences (or nets) and their limit $\Phi$ does not need to satisfy $\|\Phi\| = 1$, but only $\|\Phi\| \leq 1$, where $< 1$ is not just a rare pathological case but more the rule than the exception.

Unfortunately, this bug has far reaching consequences. Indeed, Lambert needs the convergence above to find a measurable version of $V^*$. While considering measurability is often viewed as a technical detail left to mathematicians, it is, in this case, at the core of the entire characterization (13), and Lambert is actually very aware of this, too. Indeed, this measurability is needed in (30), analogously to Lambert's proof, to apply Fubini's theorem. Without this change of integration, it cannot be proven that $S$ is order sensitive.

It is not clear at all to us how to repair this bug within Lambert's proof. In this paper we thus take a completely different route, which is laid out in detail in Section G.

## Appendix B. Proofs Related to Scoring Functions

**Proof** [Theorem 3] Let $S : A \times Y \to \mathbb{R}$ be a $\mathcal{P}$-consistent scoring function for $T$. Moreover, let $P_0, P_1 \in \mathcal{P}$ and $\alpha \in [0, 1]$ satisfy both $t^* := T(P_0) = T(P_1)$ and $P_\alpha := (1 - \alpha)P_0 + \alpha P_1 \in \mathcal{P}$. For $t \in A$, we then have

$$
\begin{aligned}
\mathbb{E}_{\mathsf{Y} \sim P_\alpha} S(t^*, \mathsf{Y}) &= (1 - \alpha)\mathbb{E}_{\mathsf{Y} \sim P_0} S(t^*, \mathsf{Y}) + \alpha \mathbb{E}_{\mathsf{Y} \sim P_1} S(t^*, \mathsf{Y}) \\
&\leq (1 - \alpha)\mathbb{E}_{\mathsf{Y} \sim P_0} S(t, \mathsf{Y}) + \alpha \mathbb{E}_{\mathsf{Y} \sim P_1} S(t, \mathsf{Y}) \\
&= \mathbb{E}_{\mathsf{Y} \sim P_\alpha} S(t, \mathsf{Y}) \,.
\end{aligned}
$$

Consequently, $t^*$ minimizes $\mathbb{E}_{\mathsf{Y} \sim P_\alpha} S(\,\cdot\,, \mathsf{Y})$, and by (1) we thus find $t^* = T(P_\alpha)$. ∎

**Lemma 14** *Let $E$ be a topological vector space, $X \subset E$ be a convex subset and $f : X \to \mathbb{R}$ be a continuous function. Then the following statements are equivalent:*

*i) For all $t \in \mathbb{R}$, the level sets $\{f = t\}$ are convex.*

*ii) For all $t \in \mathbb{R}$, the sets $\{f < t\}$ and $\{f > t\}$ are convex.*

*iii) The function $f$ is quasi-monotonic.*

*iv) For all $x_0, x_1 \in X$, the function $h : [0, 1] \to \mathbb{R}$ defined by $\alpha \mapsto f((1 - \alpha)x_0 + \alpha x_1)$ is monotonic.*

**Proof** [Lemma 14] *i)* $\Rightarrow$ *ii)*. By symmetry, it suffices to consider the case $\{f < t\}$. Let us assume that $\{f < t\}$ is not convex. Then there exist $x_0, x_1 \in \{f < t\}$ and an $\alpha \in (0, 1)$ such that for $x_\alpha := (1 - \alpha)x_0 + \alpha x_1$ we have $x_\alpha \notin \{f < t\}$, that is $f(x_\alpha) \geq t$. Now, we first observe that, for $t_0 := f(x_0) < t$ and $t_1 := f(x_1) < t$, we have $t_0 \neq t_1$, since $t_0 = t_1$ would imply $f(x_\alpha) \in \{f = t_0\} \subset \{f < t\}$ by the assumed convexity of the level set $\{f = t_0\}$. Let us assume without loss of generality that $t_0 < t_1$. Then we have $t_1 \in (f(x_0), f(x_\alpha))$, and thus the intermediate value theorem applied to the continuous map $\beta \mapsto f((1 - \beta)x_0 + \beta x_\alpha)$ on $(0, 1)$ yields a $\beta^* \in (0, 1)$ such that for $x^* := (1 - \beta^*)x_0 + \beta^* x_\alpha$ we have $f(x^*) = t_1$. Let us define $\gamma := \frac{(1 - \beta^*)\alpha}{1 - \beta^* \alpha}$. Then we have $\gamma \in (0, 1)$ and $x_\alpha = (1 - \gamma)x^* + \gamma x_1$. By the assumed convexity of $\{f = t_1\}$, we thus conclude that $f(x_\alpha) \in \{f = t_1\} \subset \{f < t\}$, i.e. we have found a contradiction.

*ii)* $\Rightarrow$ *iii)*. This follows from $\{f \geq t\} = \bigcap_{t' < t}\{f > t'\}$ and $\{f \leq t\} = \bigcap_{t' > t}\{f < t'\}$.

*iii)* $\Rightarrow$ *i)*. This follows from $\{f = t\} = \{f \leq t\} \cap \{f \geq t\}$.

*iii)* $\Rightarrow$ *iv)*. Let us pick $x_0, x_1 \in X$, and without loss of generality, we may assume that $f(x_0) \leq f(x_1)$. For $\alpha \in [0, 1]$ we further write $x_\alpha := (1-\alpha)x_1 + \alpha x_0$, so that the quasi-monotonicity implies $f(x_0) \leq f(x_\alpha) \leq f(x_1)$. Now assume that $h$ is not monotonic. Then $h$ is not increasing and hence there exist $\alpha_0 \leq \alpha_1$ with $f(x_{\alpha_1}) < f(x_{\alpha_0})$. We write $t := \frac{1}{2}f(x_{\alpha_0}) + \frac{1}{2}f(x_{\alpha_1})$, which ensures $f(x_0) \leq f(x_{\alpha_1}) < t < f(x_{\alpha_0})$. By the continuity of $\alpha \to f(x_\alpha)$ there then exist $\alpha^* \in [0, \alpha_0)$ and $\alpha^{**} \in (\alpha_0, \alpha_1)$ such that $x_{\alpha^*}, x_{\alpha^{**}} \in \{f = t\}$. However, the already shown implication *iii)* $\Rightarrow$ *i)* ensures the convexity of $\{f = t\}$, which in turn implies $f(x_{\alpha_0}) = t$ by $\alpha_0 \in (\alpha^*, \alpha^{**})$. This is a contradiction, and hence $h$ is monotonic.

*iv)* $\Rightarrow$ *iii)*. Let us assume without loss of generality that $\alpha \mapsto f((1 - \alpha)x_0 + \alpha x_1)$ is increasing on $[0, 1]$. Then we obtain

$$\min\{f(x_0), f(x_1)\} = f(x_0) \leq f((1 - \alpha)x_0 + \alpha x_1) \leq f(x_1) = \max\{f(x_0), f(x_1)\},$$

and hence $f$ is both quasi-convex and quasi-concave. ∎

**Proof** [Theorem 4] The equivalence follows directly from Lemma 14. Moreover $\mathcal{P}$ is convex and thus connected. The continuity of $T$ then shows that $\operatorname{im} T$ is connected, too, and hence $\operatorname{im} T$ is an interval. ∎

## Appendix C. Proofs Related to Identification Functions

**Proof** [Lemma 6] Let us fix a $t \in \operatorname{im}^\circ T \setminus N$. If $t = T(P)$ for all $P \in \mathcal{P}$, there is nothing to prove, and hence we may assume without loss of generality that there exists a $P \in \mathcal{P}$ with $t \neq T(P)$. By (3) we conclude that $\mathbb{E}_{\mathsf{Y} \sim P} V(t, \mathsf{Y}) \neq 0$. Let us focus on the case $t > T(P)$ and $\mathbb{E}_{\mathsf{Y} \sim P} V(t, \mathsf{Y}) > 0$ since the remaining three cases can be treated analogously. Let us first show that, for all $Q \in \mathcal{P}$, we have

$$t > T(Q) \qquad \Longrightarrow \qquad \mathbb{E}_{\mathsf{Y} \sim Q} V(t, \mathsf{Y}) > 0. \tag{24}$$

To this end, we assume the converse, that is, there exists a $Q \in \mathcal{P}$ with $t > T(Q)$ and $\mathbb{E}_{\mathsf{Y} \sim Q} V(t, \mathsf{Y}) \leq 0$. For $\alpha \in [0, 1]$ we consider $P_\alpha := \alpha P + (1 - \alpha)Q$ and

$$h(\alpha) := \mathbb{E}_{\mathsf{Y} \sim P_\alpha} V(t, \mathsf{Y}) = \alpha \mathbb{E}_{\mathsf{Y} \sim P} V(t, \mathsf{Y}) + (1 - \alpha)\mathbb{E}_{\mathsf{Y} \sim Q} V(t, \mathsf{Y}).$$

Then $P, Q \in \{T < t\}$ together with Theorem 4 implies $P_\alpha \in \{T < t\}$ for all $\alpha \in [0, 1]$, while our assumptions ensure $h(0) \leq 0$ and $h(1) > 0$. Since $h$ is continuous, the intermediate value theorem gives an $\alpha^* \in [0, 1)$ with $h(\alpha^*) = 0$, that is $\mathbb{E}_{\mathsf{Y} \sim P_{\alpha^*}} V(t, \mathsf{Y}) = 0$. By (3) we conclude that $P_{\alpha^*} \in \{T = t\}$, which contradicts the earlier found $P_{\alpha^*} \in \{T < t\}$, i.e. we have shown (24).

Let us now show that, for all $Q \in \mathcal{P}$, we have

$$t < T(Q) \qquad \Longrightarrow \qquad \mathbb{E}_{\mathsf{Y} \sim Q} V(t, \mathsf{Y}) < 0. \tag{25}$$

Let us assume the converse, i.e. that there is a $Q \in \mathcal{P}$ with $t < T(Q)$ and $\mathbb{E}_{\mathsf{Y} \sim Q} V(t, \mathsf{Y}) \geq 0$. By (3), we can exclude the case $\mathbb{E}_{\mathsf{Y} \sim Q} V(t, \mathsf{Y}) = 0$, and hence we have $\mathbb{E}_{\mathsf{Y} \sim Q} V(t, \mathsf{Y}) > 0$. Let us

define $P_\alpha$ and $h(\alpha)$ as above. Then $h(0) > 0$ and $h(1) > 0$ imply $h(\alpha) > 0$ for all $\alpha \in [0,1]$ by Theorem 4. Let us now consider $g(\alpha) := T(P_\alpha)$ for $\alpha \in [0,1]$. The continuity of $T$ guarantees that $g : [0,1] \to \mathbb{R}$ is continuous, while the assumed $T(P) < t < T(Q)$ gives $g(1) < t < g(0)$. The intermediate value theorem then shows that there exists an $\alpha^* \in [0,1]$ such that $g(\alpha^*) = t$, that is $T(P_{\alpha^*}) = t$. By (3) we thus find $\mathbb{E}_{\mathsf{Y} \sim P_{\alpha^*}} V(t, \mathsf{Y}) = 0$, that is $h(\alpha^*) = 0$. Since the latter contradicts the earlier found $h(\alpha^*) > 0$, we have shown (25).

By combining (3) with (24) and (25), we then see that $V$ is an oriented identification function. ∎

**Lemma 15** *Let $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$ be $\sigma$-finite measure spaces and $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. For $\omega_1 \in \Omega_1$ we define $A_{\omega_1} := \{\omega_2 : (\omega_1, \omega_2) \in A\}$. Then $A_{\omega_1}$ is measurable. Moreover, we have $\mu_1 \otimes \mu_2(A) = 0$ if and only if $\mu_2(A_{\omega_1}) = 0$ for $\mu_1$-almost all $\omega_1 \in \Omega_1$.*

**Proof** The measurability of the set $A_{\omega_1}$ follows e.g. from (Bogachev, 2007a, Proposition 3.3.2). By Tonelli's theorem and the measurability of $A$ we further conclude that

$$\mu_1 \otimes \mu_2(A) = \int_{\Omega_1} \int_{\Omega_2} \mathbf{1}_A(\omega_1, \omega_2) \, d\mu_2(\omega_2) d\mu_1(\omega_1) = \int_{\Omega_1} \mu_2(A_{\omega_1}) \, d\mu_1(\omega_1) \,.$$

Now the equivalence easily follows. ∎

**Lemma 16** *Let $(Y, \mathcal{A}, \mu)$ be a $\sigma$-finite measure space and $A$ be an interval that is equipped with $\hat{\mathcal{B}}(A)$. Let $S : A \times Y \to \mathbb{R}$ be a measurable and locally Lipschitz continuous function and*

$$D := \left\{ (t, y) \in \mathring{A} \times Y : \exists S'(t, y) \right\}.$$

*Then, the following statements are true:*

*i) The set $D$ is $\hat{\mathcal{B}}(A) \otimes \mathcal{A}$-measurable and of full measure, i.e. $\lambda \otimes \mu((A \times Y) \setminus D) = 0$. Moreover, for all $y \in Y$, the set $D_y := \{t \in \mathring{A} : (t, y) \in D\}$ is measurable and satisfies $\lambda(\mathring{A} \setminus D_y) = 0$.*

*ii) The canonical extension $\hat{S}' : A \times Y \to \mathbb{R}$ defined by (6) is measurable and locally bounded.*

*iii) Let $P$ be a $\mu$-absolutely continuous probability measure such that $S(t, \cdot) \in L_1(P)$ for all $t \in \mathbb{R}$. Then, there exists a measurable $N \subset \mathring{\mathrm{im}}\, T$ with $\lambda(N) = 0$, which is independent of $P$, such that the function $R_P : \mathrm{im}\, T \to \mathbb{R}$ defined by*

$$R_P(t) := \mathbb{E}_{\mathsf{Y} \sim P} S(t, \mathsf{Y}), \qquad t \in \mathrm{im}\, T,$$

*is differentiable at all $t \in \mathring{\mathrm{im}}\, T \setminus N$ and its derivative is given by*

$$R_P'(t) = \mathbb{E}_{\mathsf{Y} \sim P} \hat{S}'(t, \mathsf{Y}). \tag{26}$$

*Furthermore, we have $\mu(Y \setminus D_t) = 0$ for all $t \in \mathring{\mathrm{im}}\, T \setminus N$, where $D_t := \{y : (t, y) \in D\}$.*

**Proof** *i).* Let us fix an interval $[a,b] \subset \mathring{A}$ and a $y \in Y$. Then $S(\cdot, y)_{|[a,b]}$ is Lipschitz continuous, and therefore absolutely continuous. By (Bogachev, 2007a, Proposition 5.3.4), we conclude that $S(\cdot, y)_{|[a,b]}$ is of bounded variation and hence $S'(t, y)$ exists for Lebesgue almost all $t \in \mathring{A}$ by (Bogachev, 2007a, Theorem 5.2.6). Moreover, using the local Lipschitz continuity of $S$ and the completeness of $\mathbb{R}$ it is elementary to show that

$$D = \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{\varepsilon,\delta \in [-\frac{1}{k}, \frac{1}{k}] \cap \mathbb{Q} \setminus \{0\}} A_{n,\varepsilon,\delta}$$

where

$$A_{n,\varepsilon,\delta} := \left\{ (t,y) \in \mathring{A} \times Y : t+\varepsilon, t+\delta \in \mathring{A} \text{ and } \left| \frac{S(t+\varepsilon, y) - S(t, y)}{\varepsilon} - \frac{S(t+\delta, y) - S(t, y)}{\delta} \right| \leq \frac{1}{n} \right\}.$$

By the measurability of $S$, all sets $A_{n,\varepsilon,\delta}$ are measurable, and hence so is $D$. Let us write $Z := (\mathring{A} \times Y) \setminus D$ as well as $Z_y := \{t \in \mathring{A} : (t, y) \in Z\} = \{t \in \mathring{A} : \neg \exists S'(t, y)\} = \mathring{A} \setminus D_y$. By Lemma 15 and the measurability of $D$, all $Z_y$ are measurable and our previous considerations showed $\lambda(Z_y) = 0$ for all $y \in Y$, so that by Lemma 15 we find $\lambda \otimes \mu(Z) = 0$.

*ii).* Our first observation is that the measurability of $\hat{S}'$ is a direct consequence of the measurability of the set $D$ considered above. Let us now pick an interval $[a,b] \subset A$ and a pair $(t, y) \in \mathring{A} \times Y$ with $t \in [a,b]$. Note that if $t = a$, then $a \in \mathring{A}$, and hence there exists an $\varepsilon > 0$ such that $[a-\varepsilon, b] \subset \mathring{A} \subset A$ and, of course, $t \in (a-\varepsilon, b)$. Moreover, if $\hat{S}'$ turns out to be bounded on $[a-\varepsilon, b] \times Y$, then it is also bounded on $[a,b] \times Y$ and therefore we may assume without loss of generality that $t > a$. By the same argument we may also assume $t < b$, that is $t \in (a, b)$. Now, if $(t, y) \notin D$, then $\hat{S}'(t, y) = 0$ and hence there is nothing to prove. Moreover, if $(t, y) \in D$, then $S'(t, y)$ exists and for an arbitrary non-vanishing sequence $t_n \to 0$ we have

$$S_n(t, y) := \frac{S(t+t_n, y) - S(t, y)}{t_n} \to S'(t, y) \tag{27}$$

for $n \to \infty$. Without loss of generality we may assume that $t + t_n \in (a, b)$ for all $n \geq 1$. Then, using the local Lipschitz constant $c_{a,b} \geq 0$ we find

$$\left| S(t+t_n, y) - S(t, y) \right| \leq c_{a,b} |t_n|,$$

and hence we obtain first $|S_n(t, y)| \leq c_{a,b}$ and then $|S'(t, y)| \leq c_{a,b}$.

*iii).* By our previous considerations and Lemma 15 we first note that there exists a measurable $N \subset \mathring{\text{im}} T$ with $\lambda(N) = 0$ and $\mu(Y \setminus D_t) = 0$ for all $t \in \mathring{\text{im}} T \setminus N$. Let us pick a $t \in \mathring{\text{im}} T \setminus N$. Then $P \ll \mu$ implies $P(Y \setminus D_t) = 0$. Let us further fix a $y \in D_t$. Then we have previously seen that $|S_n(t, y)| \leq c_{a,b}$, and obviously, we have $c_{a,b} \in L_1(P)$. Therefore, (27) and Lebesgue's theorem of dominated convergence shows

$$\lim_{n \to \infty} \frac{R_P(t+t_n) - R_P(t)}{t_n} = \lim_{n \to \infty} \mathbb{E}_{\mathsf{Y} \sim P} S_n(t, \mathsf{Y}) = \mathbb{E}_{\mathsf{Y} \sim P} \hat{S}'(t, \mathsf{Y}), \tag{28}$$

that is, we have shown (26). ∎

**Proof** [Theorem 7] Before we begin with the actual proof, let us first note that the integral in (7) is defined and finite for all $t \in A$ and $y \in Y$, since for fixed $y \in Y$, the function $r \mapsto w(r)V(r,y)$ is bounded on $(t_0 \wedge t, t_0 \vee t]$.

*i).* For $[a,b] \subset A$, $t_1, t_2 \in [a,b]$ with $t_1 < t_2$, and $y \in Y$ we obtain

$$\left| S(t_1, y) - S(t_2, y) \right| \leq \|V_{|[a,b] \times Y}\|_\infty \int_{t_1}^{t_2} w(r) dr \leq \|V_{|[a,b] \times Y}\|_\infty \cdot \|w_{|[a,b]}\|_\infty \cdot |t_1 - t_2|,$$

and therefore $S$ is indeed locally Lipschitz continuous. In particular, $S(\cdot, y) : A \to \mathbb{R}$ is continuous for all $y \in Y$. Let us now show that $S(t, \cdot) : Y \to \mathbb{R}$ is measurable for all $t \in A$. Without loss of generality we consider the case $t_0 \leq t$, only. For an arbitrary but fixed $P \in \mathcal{P}$, the function $(r, y) \mapsto \mathbf{1}_{(t_0, t]}(r) w(r) V(r, y)$ is bounded and since it is only non-zero on a set of finite measure $\lambda \otimes P$, we find $((r, y) \mapsto \mathbf{1}_{(t_0, t]}(r) w(r) V(r, y)) \in L_1(\lambda \otimes P)$. Fubini's theorem, see e.g. (Bauer, 2001, Corollary 23.7) then gives $S(t, \cdot) \in L_1(P)$. In particular, $\mathbb{E}_{Y \sim P} S(t, Y)$ exists and the function $S(t, \cdot) : Y \to \mathbb{R}$ is measurable. Now, the measurability of $S$ follows from the continuity of $S(\cdot, y) : A \to \mathbb{R}$ and the measurability of $S(t, \cdot) : Y \to \mathbb{R}$ with the help of (Castaing and Valadier, 1977, Lemma III.14 on p. 70) and the fact that intervals are Polish spaces, cf. (Bauer, 2001, p. 157).

To show the assertions around (8), we first observe by Lemma 16 that, for given $y \in Y$, the derivative $S'(t, y)$ exists for Lebesgue almost all $t \in \mathring{A}$ and the extension $\hat{S}'$ is locally bounded. The formula (8) follows from (Bogachev, 2007a, Theorems 5.3.6 and 5.4.2).

To characterize when $\hat{S}'$ is an oriented identification function, we observe by Lemma 15 that there exists an $N \subset \mathring{A}$ with $\lambda(N) = 0$, such that, for all $t \in \mathring{A} \setminus N$, the derivative $S'(t, y)$ satisfies (8) for $\mu$-almost all $y \in Y$. Let us pick a $t \in \mathring{A} \setminus N$ and a $P \in \mathcal{P}$. Since $P \ll \mu$ and $\hat{S}'$ is locally bounded, we then find $\hat{S}'(t, \cdot) \in L_1(P)$ and

$$\mathbb{E}_{Y \sim P} \hat{S}'(t, Y) = \mathbb{E}_{Y \sim P} w(t) V(t, Y) = w(t) \mathbb{E}_{Y \sim P} V(t, Y). \tag{29}$$

Now the first characterization immediately follows, since $\mathbb{E}_{Y \sim P} V(t, Y) \neq 0$ for $\lambda$-almost all $t$ by (3). To show the second characterization, we first observe that $\mu(\{y \in Y : \hat{S}'(t, y) \neq 0\}) = 0$ implies $\mathbb{E}_{Y \sim P} \hat{S}'(t, Y) = 0$ by $P \ll \mu$. Now, if $\hat{S}'$ is an oriented identification function, then we have already seen that $\mathbb{E}_{Y \sim P} \hat{S}'(t, Y) \neq 0$ for $\lambda$-almost all $t \in A$ by (4), and hence we obtain $\mu(\{y \in Y : \hat{S}'(t, y) \neq 0\}) > 0$ for $\lambda$-almost all $t \in A$. Conversely, if we start with the latter, we can solve (8) for $w(t)$ to find $w > 0$ Lebesgue almost surely.

*ii).* We pick a $P = h d\mu \in \mathcal{P}$ and write $t^* := T(P)$. For $t_1, t_2 \in \operatorname{im} T$ with $t_2 < t_1 \leq t^*$ we obtain by Tonelli's theorem together with $\|\mathbf{1}_{(t_0, t_1]} V\|_\infty < \infty$ and $h \in L_1(\mu)$ that $\mathbf{1}_{(t_0, t_1]} V \in L_1(\nu \otimes P)$ and $\mathbf{1}_{(t_0, t_2]} V \in L_1(\nu \otimes P)$, and thus also $\mathbf{1}_{(t_2, t_1]} V \in L_1(\nu \otimes P)$. Fubini's theorem hence implies

$$\mathbb{E}_{Y \sim P} S(t_1, Y) - \mathbb{E}_{Y \sim P} S(t_2, Y) = \int_Y \int_{t_2}^{t_1} V(r, y) \, d\nu(r) \, dP(y)$$
$$= \int_{t_2}^{t_1} \mathbb{E}_{Y \sim P} V(r, Y) \, d\nu(r). \tag{30}$$

Now, if $\nu((t_1, t_2]) > 0$, then $\mathbb{E}_{Y \sim P} V(r, Y) < 0$ for all $r \in (t_2, t_1] \setminus N$, where $N$ is the set excluded in (4), ensures that the last integral is strictly negative, and hence $\mathbb{E}_{Y \sim P} S(t_1, Y) < \mathbb{E}_{Y \sim P} S(t_2, Y)$ follows. Conversely, if $\nu((t_1, t_2]) = 0$, then (30) implies $\mathbb{E}_{Y \sim P} S(t_1, Y) = \mathbb{E}_{Y \sim P} S(t_2, Y)$, and hence $S$ is not order sensitive. The second case, $t_2 > t_1 \geq t^*$, can be treated analogously. ∎

## Appendix D. Proofs Related to Existence of Identification Function

Our first goal is to show that we can apply all results from Appendices F and G. Note that the material in these two appendices is entirely independent of the rest of the paper and thus it is no problem using forward references to results of these appendices as soon as we have checked that all assumptions made in these appendices are met. This is done in the following lemma.

**Lemma 17** *Let $(Y, \mathcal{A}, \mu)$, $\Delta$, and $T : \mathcal{P}(\Delta) \to \mathbb{R}$ be as in Theorem 9. We fix a $p \in [1, \infty)$, write $B := \Delta$, and consider the map $\Gamma : B \to \mathbb{R}$ defined by $\Gamma(h) := T(hd\mu)$. Furthermore, we consider $h_\star := \mu(Y)^{-1}\mathbf{1}_Y \in B$, the set $A := -h_\star + B$, the norm $\| \cdot \|_F := \| \cdot \|_\infty$, the space $E := L_p(\mu)$, and the functional $\varphi' := (\mathbb{E}_\mu)_{|L_p(\mu)}$. Then, the assumptions **G1** to **G5** of Appendix F are satisfied and we have $H = L_\infty(\mu)$. Moreover, if $p \in (1, \infty)$, then the assumptions **G6** to **G9** of Appendix G are also satisfied.*

**Proof** Clearly, the set $B$ is convex and non-empty. In addition, the expectation $\mathbb{E}_\mu : L_1(\mu) \to \mathbb{R}$ is continuous, and, since $\mu$ is finite, its restriction $\varphi'$ onto $L_p(\mu)$ is continuous with respect to $\| \cdot \|_p$. Furthermore, we clearly have $B \subset \{\varphi' = 1\}$, and thus **G1** is satisfied. Moreover, $H = L_\infty(\mu)$ is obvious.

To check **G2**, we first observe that $\| \cdot \|_E \leq \| \cdot \|_F$ on $F := \operatorname{span} A \subset L_\infty(\mu)$, and thus we can apply Lemma 21 to obtain $F \subset \ker \varphi'$. For $f \in F$ with $\|f\|_\infty < \mu(Y)^{-1}$, we first find $h_\star + f \geq \varepsilon > 0$ with $\varepsilon := \mu(Y)^{-1} - \|f\|_\infty$, and thus $h_\star + f \in \operatorname{cone} \Delta$. Hence, there exist a $c \in \mathbb{R}$ and a $g \in \Delta$ such that $h_\star + f = cg$, and this yields $\mathbb{E}_\mu(h_\star + f) = \mathbb{E}_\mu cg = c$. On the other hand, $F \subset \ker \varphi'$ implies $\mathbb{E}_\mu(h_\star + f) = \mathbb{E}_\mu h_\star = 1$, and thus we conclude that $c = 1$. This yields $h_\star + f = g \in \Delta = B$, and consequently, we have $f \in -h_\star + B = A$. In other words, we have shown $\mu(Y)^{-1}B_F \subset A$, and thus $0 \in \mathring{A}^F$.

To show **G3**, we first consider the case $\Delta = \Delta^{\geq 0}$. Here we pick a $g \in L_\infty(\mu)$ and define $g^+ := \max\{g, 0\}$ and $g^- := \max\{-g, 0\}$. This gives $g^+, g^- \in \operatorname{cone} B$ and

$$\|g^+\|_{L_p(\mu)}^p + \|g^-\|_{L_p(\mu)}^p = \|g\|_{L_p(\mu)}^p.$$

Using $a + b \leq 2^{1-1/p}(a^p + b^p)^{1/p}$ we then obtain the cone assumption **G3** for $K = 2^{1-1/p}$. In the case $\Delta = \Delta^{>0}$ we first observe that there is nothing to prove for $g = 0$. Let us thus fix a $g \in L_\infty(\mu)$ with $g \neq 0$. We define $\varepsilon := \mu(Y)^{-1/p}\|g\|_{L_p(\mu)} > 0$ and consider $g_\varepsilon^+ := g^+ + \varepsilon\mathbf{1}_Y$ and $g_\varepsilon^- := g^- + \varepsilon\mathbf{1}_Y$. Clearly, this gives both $g = g_\varepsilon^+ - g_\varepsilon^-$ and $g_\varepsilon^\pm \in \Delta^{>0}$. Moreover, we have

$$\|g_\varepsilon^+\|_{L_p(\mu)} + \|g_\varepsilon^-\|_{L_p(\mu)} \leq \|g^+\|_{L_p(\mu)} + \|g^-\|_{L_p(\mu)} + 2\varepsilon\mu(Y)^{1/p}$$
$$\leq K\big(\|g^+\|_{L_p(\mu)}^p + \|g^-\|_{L_p(\mu)}^p\big)^{1/p} + 2\|g\|_{L_p(\mu)}$$
$$= (K + 2) \cdot \|g\|_{L_p(\mu)}.$$

To check **G4** it suffices to observe that the assumed continuity of $T : \mathcal{P}(\Delta_1) \to \mathbb{R}$ immediately implies the continuity of $T : \mathcal{P}(\Delta_p) \to \mathbb{R}$. Furthermore, in Theorem 9 we assume that $T$ is strictly locally non-constant, which directly translates into **G5**. Moreover, $E = L_p(\mu)$ is a Banach space, and hence **G7** is satisfied. In addition, $\mu$ is assumed to be separable, and hence the space $L_{p'}(\mu)$ is separable. Using $L_p'(\mu)$ is isometrically isomorphic to $L_{p'}(\mu)$, we conclude that $L_p'(\mu) = E'$ is separable, i.e. **G8** is satisfied. Furthermore, $H = L_\infty(\mu)$ is dense in $L_p(\mu)$, so that **G9** is satisfied, too.

It remains to prove **G6**. Let us begin by showing that $B = \Delta$ is a $\mathcal{B}(E)$-measurable subset of $E$. To this end, we consider the sets

$$B_{t,m} := \{h \in L_p(\mu) : t \le h \le m\}.$$

Then, for each $t \ge 0$ and $m \in \mathbb{N}$ with $t \le m$, the set $B_{t,m}$ is closed in $E = L_p(\mu)$. Indeed, if $(h_n) \subset B_{t,m}$ is a sequence converging to some $h \in L_p(\mu)$, that is $\|h_n - h\|_{L_p(\mu)} \to 0$, then there exists a subsequence $(h_{n_k})$ that converges $\mu$-almost surely to $h$. Since $t \le h_n \le m$ for all $n \ge 1$, we then obtain $t \le h \le m$. Therefore $B_{t,m}$ is also $\mathcal{B}(E)$-measurable, and so is the set

$$B_0 := \bigcup_{m=1}^{\infty} B_{0,m}.$$

In addition, $K := \{h \in L_p(\mu) : \mathbb{E}_\mu = 1\}$ is closed in $L_p(\mu)$ since we have already seen that $\varphi'$ is continuous, and therefore this set is also $\mathcal{B}(E)$-measurable. Now the measurablity of $\Delta^{\ge 0}$ follows from $\Delta^{\ge 0} = B_0 \cap K$. The measurability of $\Delta^{> 0}$ follows analogously by the identity $\Delta^{> 0} = \tilde{B}_0 \cap K$, where

$$\tilde{B}_0 := \bigcup_{m,n=1}^{\infty} B_{1/n,m}.$$

Let us finally show that **G6** is satisfied, that is, $\Gamma^{-1}(I)$ is a Borel measurable subset of $E$. To this end, we first observe that $I = \Gamma(\overset{\circ}{B})$ is open and since $\Gamma : B \to \mathbb{R}$ is continuous with respect to $\|\cdot\|_E$, the set $\Gamma^{-1}(I)$ is open in the metric space $(B, \|\cdot\|_E)$. Since the topology of the latter space is the trace topology of $\|\cdot\|_E$ on $B$ we conclude that there is an $\|\cdot\|_E$-open subset $O$ of $E$ such that $\Gamma^{-1}(I) = B \cap O$. Now the assertion follows from the previously established measurability of $B$. ∎

**Lemma 18** *Let $1 \le p, q < \infty$ and $(\Omega, \mu)$ be a $\sigma$-finite measure space. Furthermore, let $\varphi \in L'_\infty(\mu)$ such that the restrictions*

$$\varphi : (L_\infty(\mu) \cap L_p(\mu), \|\cdot\|_p) \to \mathbb{R}$$
$$\varphi : (L_\infty(\mu) \cap L_q(\mu), \|\cdot\|_p) \to \mathbb{R}$$

*are continuous. Let $\varphi_p \in L'_p(\mu)$ and $\varphi_q \in L'_q(\mu)$ be the corresponding unique extensions of $\varphi$ and $f_p \in L_{p'}(\mu)$ and $f_q \in L_{q'}(\mu)$ be the representing functions for $\varphi_p$ and $\varphi_q$. Then for $\mu$-almost all $\omega \in \Omega$ we have $f_p(\omega) = f_q(\omega)$.*

**Proof** [Lemma 18] For $h \in L_\infty(\mu)$ our assumptions yield

$$\int_\Omega f_p h \, d\mu = \varphi_p(h) = \varphi(h) = \varphi_q(h) = \int_\Omega f_q h \, d\mu.$$

Now the assertion follows from considering $h := \mathbf{1}_{\{f_p > f_q\}}$ and $h := \mathbf{1}_{\{f_p < f_q\}}$. ∎

**Lemma 19** *Let $(Y, \mathcal{A}, \mu)$ be a finite and separable measure space and $\Delta$ be either $\Delta^{\ge 0}$ or $\Delta^{> 0}$. Moreover, let $T : \mathcal{P}(\Delta) \to \mathbb{R}$ be a strictly locally non-constant and quasi-monotonic property and $V$ be an oriented $\mathcal{P}(\Delta)$-identification function for $T$. Then, for all $p \in [1, \infty)$, the following statements are equivalent:*

   *i)* $T : \mathcal{P}(\Delta_p) \to \mathbb{R}$ *is continuous.*

   *ii)* *For Lebesgue-almost all $t \in \operatorname{im} T$ we have $V(t, \cdot) \in L_{p'}(\mu)$.*

**Proof** [Lemma 19] Let us fix a Lebesgue zero set $N \subset \operatorname{im}^{\circ} T$ such that (3) and (4) hold for all $t \in \operatorname{im}^{\circ} T \setminus N$. For $t \in \operatorname{im}^{\circ} T \setminus N$, we then have

$$\{T = t\} = \{P \in \mathcal{P}(\Delta) : \mathbb{E}_{\mathsf{Y} \sim P} V(t, \mathsf{Y}) = 0\} \tag{31}$$

$$\{T \geq t\} = \{P \in \mathcal{P}(\Delta) : \mathbb{E}_{\mathsf{Y} \sim P} V(t, \mathsf{Y}) \leq 0\} \tag{32}$$

$$\{T \leq t\} = \{P \in \mathcal{P}(\Delta) : \mathbb{E}_{\mathsf{Y} \sim P} V(t, \mathsf{Y}) \geq 0\}. \tag{33}$$

Now let $B := \Delta$, $H := \operatorname{span} \Delta = L_\infty(\mu)$, and $\Gamma : B \to \mathbb{R}$ defined by $\Gamma(h) := T(h d\mu)$. Moreover, for $t \in \operatorname{im}^{\circ} T \setminus N$ we define the linear functional $\tilde{z}'_t : H \to \mathbb{R}$ by

$$\tilde{z}'_t(h) := \int_Y V(t, y)\, h(y) d\mu(y),$$

where we note that $V(t, \cdot) \in L_1(h d\mu)$ for all $h \in \Delta$ and $\operatorname{span} \Delta = L_\infty(\mu)$ ensure that $\tilde{z}'_t$ is actually well-defined.

   *i)* $\Rightarrow$ *ii).* Since we assume that $T : \mathcal{P}(\Delta_p) \to \mathbb{R}$ is continuous, we conclude that from Lemma 17 that **G1** to **G5** are satisfied, and hence we can apply Theorem 22. For $t \in \operatorname{im}^{\circ} T \setminus N$, let $z'_t \in (H, \|\cdot\|_{L_p(\mu)})'$ be the separating functional obtained by the latter theorem. We then obtain

$$\ker z'_t \cap B = \{T = t\} = \ker \tilde{z}'_t \cap B$$

by (31). By Lemma 32 we conclude that $\tilde{z}'_t \neq 0$ and $\ker z'_t = \operatorname{span}\{T = t\} = \ker \tilde{z}'_t$, and since the former is closed with respect to $\|\cdot\|_{L_p(\mu)}$, so is the latter. However, this implies that $\tilde{z}'_t$ is continuous with respect to $\|\cdot\|_{L_p(\mu)}$. Since $H = L_\infty(\mu)$ is dense in $L_p(\mu)$ and $V(t, \cdot)$ is the representing function of $\tilde{z}'_t$, we finally conclude that $V(t, \cdot) \in L_{p'}(\mu)$ by Lemma 18.

   *ii)* $\Rightarrow$ *i).* Without loss of generality we may assume that $V(t, \cdot) \in L_{p'}(\mu)$ for all $t \in \operatorname{im}^{\circ} T \setminus N$. Then $\tilde{z}'_t$ is continuous with respect to $\|\cdot\|_{L_p(\mu)}$ for $t \in \operatorname{im}^{\circ} T \setminus N$, and therefore the sets $\{\tilde{z}'_t \leq 0\}$ and $\{\tilde{z}'_t \geq 0\}$ are closed in $(H, \|\cdot\|_{L_p(\mu)})$. This shows that $B \cap \{\tilde{z}'_t \leq 0\}$ and $B \cap \{\tilde{z}'_t \geq 0\}$ are closed in $B$ with respect to $\|\cdot\|_{L_p(\mu)}$, and using (32) and (33) we conclude that $\{T \geq t\}$ and $\{T \leq t\}$ are closed in $\mathcal{P}(\Delta_p)$ for all $t \in \operatorname{im}^{\circ} T \setminus N$. Moreover, for $t \in N$, we find a sequence $t_n \in \operatorname{im}^{\circ} T \setminus N$ with $t_n \searrow t$ since $N$ is a Lebesgue zero set and $\operatorname{im}^{\circ} T$ is open. This gives

$$\{T \leq t\} = \bigcap_{n \geq 1} \{T \leq t_n\} \tag{34}$$

and hence $\{T \leq t\}$ is closed. Analogously, we find that $\{T \geq t\}$ is closed for all $t \in N$. Let us finally consider the possible endpoints of the interval $\operatorname{im} T$. For example, if $t = \min \operatorname{im} T$ exists, then we find by an argument identical to (34) that $\{T \leq t\}$ is closed, and $\{T \geq t\} = \mathcal{P}(\Delta_p)$ is also closed in $\mathcal{P}(\Delta_p)$. Summing up, the sets $\{T \leq t\}$ and $\{T \geq t\}$ are closed in $\mathcal{P}(\Delta_p)$ for all $t \in \operatorname{im} T$, and hence $T$ is both lower- and upper-semicontinuous with respect to $\|\cdot\|_{L_p(\mu)}$, i.e. $T : \mathcal{P}(\Delta_p) \to \mathbb{R}$ is continuous. ∎

**Proof** [Theorem 9] *i).* Let us fix a $p \in [1, \infty)$. By Lemma 17 we know that **G1** to **G5** are satisfied, and **G6** to **G9** are additionally satisfied if $p > 1$. Consequently, we can apply Theorem 22 and, if $p > 1$, also Theorem 34. For all $t \in I = \overset{\circ}{\mathrm{im}}\, T$, let $z'_t \in (H, \|\cdot\|_{L_p(\mu)})'$ be the functional provided by Theorem 22. Since $H = L_\infty(\mu)$ is dense in $L_p(\mu)$, each $z'_t$ can be uniquely extended to a functional $\hat{z}'_t \in L_p(\mu)'$ and, in addition, this extension satisfies $\|\hat{z}'_t\|_{L_p(\mu)'} = 1$. Now let $\iota_p : L_{p'}(\mu) \to L_p(\mu)'$ be the isometric isomorphism defined by

$$\iota_p g(f) := \int_Y g f d\mu, \qquad g \in L_{p'}(\mu),\ f \in L_p(\mu).$$

Then $V_p(t, \cdot) := -\iota_p^{-1} \hat{z}'_t$, $t \in \mathbb{R}$, defines an oriented $\mathcal{P}(\Delta)$-identification function for $T$, since for $P = h\mu \in \mathcal{P}(\Delta)$ we have

$$\mathbb{E}_{\mathsf{Y} \sim P} V_p(t, \mathsf{Y}) = \int_Y V_p(t, y) h(y) d\mu(y) = \iota_p V_p(t, \cdot)(h) = -\hat{z}'_t(h).$$

Note that the definition of $V_p$ actually depends on the chosen $p$. Of course, eventually, we are only interested in $V_1$, but we will see below that for establishing the measurability of $V_1$, it actually makes sense to consider $V_p$ for $p > 1$, too. For later use we further note that, given an oriented $\mathcal{P}(\Delta)$-identification $\tilde{V}$ for $T$ with

$$\|\tilde{V}(t, \cdot)\|_{L_{p'}(\mu)} = 1$$

for Lebesgue almost all $t \in \overset{\circ}{\mathrm{im}}\, T$, we have $\iota_p \tilde{V}(t, \cdot) = -\hat{z}'_t$ for such $t$ by the uniqueness of $z'_t$ in Theorem 22. For Lebesgue almost all $t \in \overset{\circ}{\mathrm{im}}\, T$ we thus have

$$V_p(t, \cdot) = \tilde{V}(t, \cdot) \qquad \mu\text{-almost everywhere.} \tag{35}$$

Our next goal is to show that there exists a measurable modification of $V_1$. Here, we will proceed in two steps. In the first step we show that there is such a modification for $V_p$ if $p > 1$. Based on this, the measurable modification of $V_1$ is then found in the second step.

Now, let $p > 1$ be fixed and $Z : (I, \hat{\mathcal{B}}(I)) \to (E', \mathcal{B}(E'))$ be the map obtained by Theorem 34 for $E := L_p(\mu)$ and $I := \overset{\circ}{\mathrm{im}}\, T$. Then we have $Z(t) = -\hat{z}'_t = \iota_p V_p(t, \cdot)$ for all $t \in I$, and thus the map $t \mapsto V_p(t, \cdot)$ is $(\hat{\mathcal{B}}(I), \mathcal{B}(L_{p'}(\mu)))$-measurable. Let us now fix a finite measure $\nu$ on $\hat{\mathcal{B}}(I)$, that has a strictly positive Lebesgue density. Then $Z$ is Bochner $\nu$-integrable by Theorem 34, and hence so is the map $t \mapsto V_p(t, \cdot)$. By (Pietsch, 1987, Proposition 6.2.12) we then obtain a $(\hat{\mathcal{B}}(I) \otimes \mathcal{A}, \mathcal{B}(\mathbb{R}))$-measurable map $\tilde{V}_p : \overset{\circ}{\mathrm{im}}\, T \times Y \to \mathbb{R}$ such that, for all Lebesgue-almost all $t \in \overset{\circ}{\mathrm{im}}\, T$, we have

$$\mu\big(\{y \in Y : V_p(t, y) \neq \tilde{V}_p(t, y)\}\big) = 0. \tag{36}$$

Since for $t \in \overset{\circ}{\mathrm{im}}\, T$ satisfying (36), we have $\iota_p V_p(t, \cdot) = \iota_p \tilde{V}_p(t, \cdot)$, this map $\tilde{V}_p$ is a measurable and oriented $\mathcal{P}(\Delta)$-identification function for $T$.

Let us now find the modification for $V_1$. To this end, we fix some arbitrary $p > 1$ and continue considering the map $Z : (I, \hat{\mathcal{B}}(I)) \to (E', \mathcal{B}(E'))$ for $E := L_p(\mu)$. For $H = \mathrm{span}\,\Delta$ we then have

$$Z(t)_{|H} = -z'_t, \qquad t \in I,$$

where $z'_t \in H' \cap (H, \|\cdot\|_{L_p(\mu)})'$ is the separating functional obtained by Theorem 22. For $\Gamma$ defined in Lemma 17, we consequently have

$$\{\Gamma = t\} = \{Z(t)_{|H} = 0\} \cap B. \tag{37}$$

Now let $\tilde{z}_t' \in H' \cap (H, \|\cdot\|_{L_1(\mu)})'$ be the separating functional obtained by Theorem 22 in the case "$p = 1$". Then we have

$$\{\Gamma = t\} = \{\tilde{z}_t' = 0\} \cap B\,, \tag{38}$$

so that Lemma 33 together with (37) and (38) gives an $\alpha(t) \neq 0$ with

$$Z(t)_{|H} = -\alpha(t)\tilde{z}_t'\,, \qquad\qquad t \in I. \tag{39}$$

Moreover, both functionals have opposite orientation, and hence we actually have $\alpha(t) > 0$. Since, for fixed $t \in I$, the functional $\tilde{z}_t'$ is continuous with respect to $\|\cdot\|_{L_1(\mu)}$ on $H$, the same is thus true for $Z(t)_{|H}$. Now, using the separability of $\mu$, there exists a countable subset $L \subset L_\infty(\mu)$ that is dense in $L_1(\mu)$. With the help of this subset we obtain

$$\big\| Z(t)_{|H} : (H, \|\cdot\|_{L_1(\mu)}) \to \mathbb{R} \big\| = \sup_{h \in B_{L_1(\mu)} \cap L_\infty(\mu)} \big| Z(t)_{|H}(h) \big| = \sup_{h \in B_{L_1(\mu)} \cap L_\infty(\mu) \cap L} \big| Z(t)_{|H}(h) \big|$$

by the continuity of $Z(t)_{|H}$ on $H$ with respect to $\|\cdot\|_{L_1(\mu)}$. Moreover, for fixed $h \in B_{L_1(\mu)} \cap L_\infty(\mu) \cap L$ the map $t \mapsto Z(t)_{|H}(h) = Z(t)(h)$ is $(\hat{\mathcal{B}}(I)), \mathcal{B}(\mathbb{R}))$-measurable by the above measurability of $Z$ and $h \in L_p(\mu)$. Since countable suprema over measurable functions are measurable, we conclude that the map

$$t \mapsto \big\| z(t)_{|H} : (H, \|\cdot\|_{L_1(\mu)}) \to \mathbb{R} \big\|$$

is $(\hat{\mathcal{B}}(I)), \mathcal{B}(\mathbb{R}))$-measurable. Using (39) we further have

$$\big\| Z(t)_{|H} : (H, \|\cdot\|_{L_1(\mu)}) \to \mathbb{R} \big\| = |\alpha(t)| \cdot \|\tilde{z}_t' : (H, \|\cdot\|_{L_1(\mu)}) \to \mathbb{R}\| = \alpha(t)\,,$$

and hence $t \mapsto \alpha(t)$ is $(\hat{\mathcal{B}}(I)), \mathcal{B}(R))$-measurable. Now recall that $\tilde{V}_p(t, \cdot)$ is, for Lebesgue almost $t \in \overset{\circ}{\mathrm{im}}\, T$, a representation of $Z(t)$, that is

$$\tilde{V}_p(t, \cdot) = \iota_p^{-1} Z(t)\,.$$

Let us pick a $t \in \overset{\circ}{\mathrm{im}}\, T$. By construction $V_1(t, \cdot)$ is then a representation of the extension $\hat{z}_t'$ of $\tilde{z}_t'$ to $L_1(\mu)$, and hence $\alpha(t)V_1(t, \cdot) = -\iota_1^{-1}(\alpha(t)\hat{z}_t')$. Furthermore, (39) shows that on the dense subspace $H = L_\infty(\mu)$, the functionals $Z(t)$ and $\alpha(t)\hat{z}_t'$ coincide. By Lemma 18 we conclude that

$$\tilde{V}_p(t, \cdot) = \alpha(t)V_1(t, \cdot)$$

$\mu$-almost surely. Consequently, $\tilde{V}_1(t, y) := \frac{\tilde{V}_p(t,y)}{\alpha(t)}$, where $(t, y) \in \mathrm{im}\, T \times Y$, defines a measurable and oriented $\mathcal{P}(\Delta)$-identification function for $T$ with $\|\tilde{V}_1(t, \cdot)\|_{L_\infty(\mu)} = 1$ for Lebesgue-almost all $t \in \overset{\circ}{\mathrm{im}}\, T$.

*ii).* Let $V^*$ be a measurable and oriented $\mathcal{P}(\Delta)$-identification function for $T$ obtained in *i)*. We have already seen in (35) that $V^*$ is $\mu \otimes \lambda$-almost surely unique. Moreover, let $V$ be another measurable and oriented $\mathcal{P}(\Delta)$-identification function for $T$. Since $T : \mathcal{P}(\Delta_1) \to \mathbb{R}$ is continuous, we see by Lemma 19 that $V(t, \cdot) \in L_\infty(\mu)$ for Lebesgue almost all $t \in \mathrm{im}\, T$. Moreover, the definition of an identification function immediately gives $V(t, \cdot) \neq 0$ for Lebesgue almost all $t \in \mathrm{im}\, T$. For $t \in \mathrm{im}\, T$ we write

$$w(t) := \|V(t, \cdot)\|_{L_\infty(\mu)}\,, \tag{40}$$

25

if $V(t, \cdot) \in L_\infty(\mu)$ and $V(t, \cdot) \neq 0$, and $w(t) := 1$ otherwise. This gives $w(t) > 0$ for all $t \in \operatorname{im} T$, and thus

$$\tilde{V}(t, y) := \frac{V(t, y)}{w(t)}, \qquad (t, y) \in \operatorname{im} T \times Y,$$

defines another oriented $\mathcal{P}(\Delta)$-identification function for $T$. Since we further have $\|\tilde{V}(t, \cdot)\|_{L_\infty(\mu)} = 1$ for Lebesgue almost all $t \in \operatorname{im}^\circ T$, Equation (35) gives (11).

Finally, to show that $w$ is measurable, we fix a countable and dense subset $D$ of $B_{L_1(\mu)}$ and fix a Lebesgue zero set $N \subset \operatorname{im} T$ with $V(t, \cdot) \in L_\infty(\mu)$ for all $t \in \operatorname{im} T \setminus N$. Moreover, we consider the maps $\bar{V} := \mathbf{1}_{\operatorname{im} T \setminus N} V$ and $\bar{w} : \operatorname{im} T \to \mathbb{R}$ defined by

$$\bar{w}(t) := \|\bar{V}(t, \cdot)\|_{L_\infty(\mu)} = \sup_{h \in D} |\langle \iota_1 \bar{V}(t, \cdot), h \rangle|, \qquad t \in \operatorname{im} T.$$

Then, for each $h \in D$, the map $t \mapsto \langle \iota \bar{V}(t, \cdot), h \rangle$ is $(\hat{\mathcal{B}}(\operatorname{im} T), \mathcal{B}(\mathbb{R}))$-measurable by the assumed measurability of $V$, and hence so is $\bar{w}$. However, our construction ensures $w = \bar{w}$ Lebesgue almost surely, and thus $w$ is $(\hat{\mathcal{B}}(\operatorname{im} T), \mathcal{B}(\mathbb{R}))$-measurable, too.

*iii).* The existence of $S'$ outside a measurable set $Z \subset \operatorname{im}^\circ T \times Y$ with $\lambda \otimes \mu(Z) = 0$ follows from Lemma 16.

To show (12), let $N \subset \operatorname{im}^\circ T$ be the $\lambda$-zero set considered around (26). Our first goal is to show that, for all $t \in \operatorname{im}^\circ T \setminus N$ and all $P \in \mathcal{P}(\Delta)$, we have

$$t = T(P) \qquad \Longrightarrow \qquad \mathbb{E}_{\mathsf{Y} \sim P} \hat{S}'(t, \mathsf{Y}) = 0. \tag{41}$$

To this end, we fix a $P \in \mathcal{P}(\Delta)$ with $T(P) \in \operatorname{im}^\circ T \setminus N$ and consider the function $R_P : \operatorname{im} T \to \mathbb{R}$ defined in Lemma 16. Then the map $R_P : \operatorname{im} T \to \mathbb{R}$ has a global minimum at $t^* := T(P)$, since $S$ is $\mathcal{P}(\Delta)$-consistent for $T$. Moreover, $R_P : \operatorname{im} T \to \mathbb{R}$ is differentiable at $t^*$ by Lemma 16, and hence we obtain $R'_P(t^*) = 0$. Equation (26) then yields (41).

Now let $V^*$ be the oriented identification function obtained in part *i).* Without loss of generality we may assume that the $\lambda$-zero set $N$ obtained above is such that $\|V^*(t, \cdot)\|_{L_\infty(\mu)} = 1$ and both (3) and (4) hold for $V^*$ and all $t \in \operatorname{im}^\circ T \setminus N$. For a fixed $t \in \operatorname{im}^\circ T \setminus N$, we can thus consider $\hat{z}'_t := (\iota_1 \hat{S}'(t, \cdot))_{|H}$ and $z_t := (\iota_1 V^*(t, \cdot))_{|H}$. Here we note that $\hat{S}'$ is locally bounded by Lemma 16, and thus $\hat{S}'(t, \cdot) \in L_\infty(\mu)$. Furthermore, let again $\Gamma$ be the map considered in Lemma 17. By (41) we then know that $\{\Gamma = t\} \subset \ker \hat{z}'_t \cap B$, while (3) ensures $\{\Gamma = t\} = \ker z'_t \cap B$. Consequently, Lemma 33 gives a $w(t) \in \mathbb{R}$ with $\hat{z}'_t = w(t) z'_t$, and from this we immediately obtain (12). For later purposes, let us write $w(t) := 0$ for $t \in N$.

To show that $w$ is locally bounded, we fix an interval $[a, b] \subset \operatorname{im} T$. By Lemma 16 we then know that there exists a constant $c_{a,b}$ such that $|S'(t, y)| \leq c_{a,b}$ for all $t \in [a, b] \setminus N$ and $y \in D_t$, where $D_t := \{y : \exists S'(t, y)\}$. As above, we may further assume that $\|V^*(t, \cdot)\|_{L_\infty(\mu)} = 1$ for all $t \in [a, b] \setminus N$. By (12) we conclude that $|w(t)| = \|S'(t, \cdot)\|_{L_\infty(\mu)} \leq c_{a,b}$ for all $t \in [a, b] \setminus N$.

To show the measurability of $w$, we fix a $1 < p < \infty$ and a countable dense subset $L \subset L_p(\mu)$. By (Pietsch, 1987, Proposition 6.2.12) and the measurability of $\hat{S}'$ we then know that the map $[a, b] \to L_{p'}(\mu)$ defined by $t \mapsto \hat{S}'(t, \cdot)$ is $(\hat{\mathcal{B}}(I), \mathcal{B}(L_{p'}(\mu)))$-measurable, and hence the map $t \mapsto |\langle \hat{S}'(t, \cdot), f \rangle|$ is $(\hat{\mathcal{B}}(I), \mathcal{B}(\mathbb{R}))$-measurable for a fixed $f \in L$. For $t \in \operatorname{im}^\circ T \setminus N$ we further have

$$\|\hat{S}'(t, \cdot)\|_{L_{p'}(\mu)} = \sup_{f \in B_{L_p(\mu)} \cap L} |\langle \hat{S}'(t, \cdot), f \rangle| < \infty,$$

and hence $t \mapsto \|\hat{S}'(t, \cdot)\|_{L_{p'}(\mu)}$ is also measurable. Analogously, we obtain the measurability of $t \mapsto \|V^*(t, \cdot)\|_{L_{p'}(\mu)}$, and since $\|V^*(t, \cdot)\|_{L_{p'}(\mu)} \neq 0$ for all $t \in [a, b] \setminus N$, we finally obtain the measurability of $w$ by (12).

Our next goal is to show that $w \geq 0$, if $S$ is order sensitive. To this end, we fix a $P \in \mathcal{P}(\Delta)$ with $T(P) \in \overset{\circ}{\mathrm{im}}\, T$. Since $S$ is $\mathcal{P}(\Delta)$-order sensitive for $T$, the map $t \mapsto R_P(t)$ considered in Lemma 16 is decreasing on $(-\infty, T(P)] \cap \mathrm{im}\, T$, and consequently, we have $R'_P(t) \leq 0$ for all $t \in (-\infty, T(P)] \cap (\mathrm{im}\, T \setminus N)$. On the other hand, using (26) and (12) we conclude that

$$R'_P(t) = w(t)\mathbb{E}_{y \sim P} V^*(t, y)$$

for all $t \in \overset{\circ}{\mathrm{im}}\, T \setminus N$, where $N \subset \overset{\circ}{\mathrm{im}}\, T$ is the Lebesgue zero set considered around (26). Now, the orientation of $V^*$ gives $\mathbb{E}_{y \sim P} V^*(t, y) < 0$ for all $t \in (-\infty, T(P)) \cap (\overset{\circ}{\mathrm{im}}\, T \setminus N)$, and hence we find $w(t) \geq 0$ for such $t$. For the remaining $t$ recall that we set $w(t) = 0$ above.

Let us now assume that $w \geq 0$. Then it remains to show that the measure $\nu := wd\lambda$ satisfies $\nu((t_1, t_2]) > 0$ for all $t_1, t_2 \in \mathrm{im}\, T$ with $t_1 < t_2$, if and only if $S$ is order sensitive. To this end, observe that by (12) there exists a measurable $Z \subset Y$ with $\mu(Z) = 0$ such that for all $y \in Y \setminus Z$ there exists a measurable $N_y \subset \overset{\circ}{\mathrm{im}}\, T$ with $\lambda(N_y) = 0$ such that (12) holds for all $t \in \overset{\circ}{\mathrm{im}}\, T \setminus N_y$. The fundamental theorem of calculus for absolutely continuous functions, see e.g. (Bogachev, 2007a, Theorems 5.3.6 and 5.4.2), yields

$$S(t, y) - S(t_0, y) = \int_{t_0}^{t} S'(r, y)\, dr = \int_{t_0}^{t} w(r)V^*(r, y)\, dr$$

for all $t \in \mathrm{im}\, T$ and $y \in Y \setminus Z$. By setting $b(y) := S(t_0, y)$, we then see that $S$ is of the form (7). Now the assertion follows by part *ii)* of Theorem 7. ∎

## Appendix E. Proofs Related to Expectiles

**Lemma 20** *Let $\varphi : (0, \infty) \to (0, \infty)$ be a group homomorphism, that is $\varphi(st) = \varphi(s)\varphi(t)$ for all $s, t \in (0, \infty)$. If $\varphi$ is increasing and there exists an $s_0 \in (0, \infty)$ with $\varphi(s_0) \neq 1$, then there exists an $\alpha > 0$ such that $\varphi(s) = s^\alpha$ for all $s \in (0, \infty)$. In particular, $\varphi$ is continuous and surjective.*

**Proof** [Lemma 20] Note that from $\varphi(s^n) = (\varphi(s))^n$, which follows by simple induction, we obtain $(\varphi(t))^{1/n} = \varphi(t^{1/n})$ by setting $t = s^n$. Combining both yields $\varphi(s^q) = (\varphi(s))^q$ for all $s > 0$ and $q \in \mathbb{Q}$ with $q > 0$. Moreover, since we have $\varphi(s)\varphi(s^{-1}) = 1$, this identity also holds for $q \in \mathbb{Q}$ with $q < 0$, and for $q = 0$ it is obviously satisfied. Let us define

$$D := \{s_0^q : q \in \mathbb{Q}\}$$

and $\alpha := \frac{\ln \varphi(s_0)}{\ln s_0}$. Note that we have $\alpha > 0$ since $\varphi$ is assumed to be increasing and $\varphi(s_0) \neq 0$. Then the definition of $\alpha$ yields $s_0^\alpha = \varphi(s_0)$, and thus $\varphi(s_0^q) = (\varphi(s_0))^q = s_0^{\alpha q}$, that is $\varphi(t) = t^\alpha$ for all $t \in D$. Since $D$ is dense in $(0, \infty)$ it hence remains to show that $\varphi$ is continuous. Note that the latter follows from the continuity at 1, since $s_n \to s$ implies $s_n s^{-1} \to 1$ and thus $\varphi(s_n)\varphi(s^{-1}) = \varphi(s_n s^{-1}) \to \varphi(1) = 1$.

To show the continuity at 1, we first observe that $\varphi(s_0)\varphi(s_0^{-1}) = 1$ implies $\varphi(s_0^{-1}) \neq 1$, and hence we may assume without loss of generality that $s_0 > 1$. Let us now assume that $\varphi$ is not

27

continuous at 1. Then, there exists a sequence $(t_n) \subset (0, \infty)$ such that $t_n \to 1$ and $\varphi(t_n) \not\to 1$. This implies that $\varphi(t_n^{-1}) = (\varphi(t_n))^{-1} \not\to 1$, and hence we may assume without loss of generality that $t_n > 1$ for all $n \geq 1$. In addition, we may clearly assume that $t_n \searrow 1$. Now, $\varphi(t_n) \not\to 1$ yields an $\varepsilon > 0$ such that $\varphi(t_n) \geq 1 + \varepsilon$ for all $n \geq 1$. Let us pick a $t > 1$. Then there exists an $n \geq 1$ such that $t_n < t$, and thus $\varphi(t_n) \leq \varphi(t)$, that is $\varphi(t) \geq 1 + \varepsilon$ for all $t > 1$. On the other hand, we have $\varphi(s_0^{1/n}) = s_0^{\alpha/n} \to 1$, i.e. we have found a contradiction. ∎

**Proof** [Theorem 12] *i).* Let us fix an oriented $\mathcal{P}$-identification function $\tilde{V} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ for $T$. We begin by some preliminary considerations on $\tilde{V}$. To this end, we first note that for $\mathsf{Y} := 0$ and $y \in \mathbb{R}$ we find $T(y) = T(\mathsf{Y} + y) = T(\mathsf{Y}) + y = y$. For $\mathsf{Y}' := y$ and $t \in \mathbb{R}$ we further have $\mathbb{E}_\nu \tilde{V}(t, \mathsf{Y}') = \tilde{V}(t, y)$. By $T(\mathsf{Y}') = y$ and the definition of oriented identification functions we conclude that $\tilde{V}(t, y) = 0$ if and only if $t = y$, as well as, $\tilde{V}(t, y) > 0$ if and only if $t > y$. With these preparations, we now fix some $y_1 < t < y_2$ and define

$$p := \frac{\tilde{V}(t, y_1)}{\tilde{V}(t, y_1) - \tilde{V}(t, y_2)}.$$

Our preliminary considerations show both $\tilde{V}(t, y_1) > 0$ and $-\tilde{V}(t, y_2) > 0$, and thus we find $p \in (0, 1)$. Since $(\Omega, \mathcal{A}, \nu)$ is atom-free, there then exists an $A \in \mathcal{A}$ with $\nu(A) = 1 - p$. Let us consider the random variable

$$\mathsf{Y} := y_1 \mathbf{1}_A + y_2 \mathbf{1}_{\Omega \setminus A}. \tag{42}$$

An easy calculation shows that

$$\mathbb{E}_\nu \tilde{V}(t, \mathsf{Y}) = \tilde{V}(t, y_1)(1 - p) + \tilde{V}(t, y_2)p = -\frac{\tilde{V}(t, y_1)\tilde{V}(t, y_2)}{\tilde{V}(t, y_1) - \tilde{V}(t, y_2)} + \frac{\tilde{V}(t, y_2)\tilde{V}(t, y_1)}{\tilde{V}(t, y_1) - \tilde{V}(t, y_2)} = 0,$$

and thus $T(\mathsf{Y}) = t$. For $s \in \mathbb{R}$ this yields $T(\mathsf{Y} + s) = T(\mathsf{Y}) + s = t + s$, and hence we find

$$0 = \mathbb{E}_\nu \tilde{V}(t + s, \mathsf{Y} + s) = -\frac{\tilde{V}(t + s, y_1 + s)\tilde{V}(t, y_2)}{\tilde{V}(t, y_1) - \tilde{V}(t, y_2)} + \frac{\tilde{V}(t + s, y_2 + s)\tilde{V}(t, y_1)}{\tilde{V}(t, y_1) - \tilde{V}(t, y_2)}.$$

From the latter we easily conclude that

$$\frac{\tilde{V}(t, y_1)}{\tilde{V}(t, y_2)} = \frac{\tilde{V}(t + s, y_1 + s)}{\tilde{V}(t + s, y_2 + s)} \tag{43}$$

for all $y_1 < t < y_2$ and $s \in \mathbb{R}$. Now, for $y < t$, we have $y - t < 0 < 1$, and hence (43) implies

$$\frac{\tilde{V}(0, y - t)}{\tilde{V}(0, 1)} = \frac{\tilde{V}(t, y - t + t)}{\tilde{V}(t, 1 + t)} = \frac{\tilde{V}(t, y)}{\tilde{V}(t, t + 1)}. \tag{44}$$

Analogously, for $y > t$, we have $-1 < 0 < y - t$, and hence (43) implies

$$\frac{\tilde{V}(0, -1)}{\tilde{V}(0, y - t)} = \frac{\tilde{V}(t, -1 + t)}{\tilde{V}(t, y - t + t)} = \frac{\tilde{V}(t, t - 1)}{\tilde{V}(t, y)}. \tag{45}$$

Finally, $-1 < 0 < 1$ together with (43) implies

$$\frac{\tilde{V}(0,-1)}{\tilde{V}(0,1)} = \frac{\tilde{V}(t,-1+t)}{\tilde{V}(t,1+t)} = \frac{\tilde{V}(t,t-1)}{\tilde{V}(t,t+1)} \tag{46}$$

for all $t \in \mathbb{R}$. Let us write $w(t) := \tilde{V}(t, t-1)$ and $\psi(r) := \frac{\tilde{V}(0,r)}{\tilde{V}(0,-1)}$ for $r, t \in \mathbb{R}$. Clearly, this gives both $\psi(0) = 0$ and $\psi(-1) = 1$, and combining (44) with (46) we further find

$$\tilde{V}(t,y) = w(t)\psi(y-t) \tag{47}$$

for $y < t$, while (45) gives (47) for $y > t$. Moreover, for $y = t$ our preliminary considerations yield $\tilde{V}(t,y) = 0 = w(t)\psi(y - t)$, and thus (47) holds for all $y, t \in \mathbb{R}$. Finally, we have $w(t) > 0$ for all $t \in \mathbb{R}$, so that $V = \tilde{V}/w$ is an oriented $\mathcal{P}$-identification function for $T$.

*ii)*. For $y_1 < t < y_2$, we again consider the random variable $\mathsf{Y}$ given by (42). Then the assumed homogeneity of $T$ gives $T(s\mathsf{Y}) = sT(\mathsf{Y}) = st$ for all $s > 0$, and thus we obtain

$$0 = \mathbb{E}_\nu \tilde{V}(st, s\mathsf{Y}) = -\frac{\tilde{V}(st, sy_1)\tilde{V}(t, y_2)}{\tilde{V}(t, y_1) - \tilde{V}(t, y_2)} + \frac{\tilde{V}(st, sy_2)\tilde{V}(t, y_1)}{\tilde{V}(t, y_1) - \tilde{V}(t, y_2)}.$$

The latter together with (47) implies

$$\frac{\tilde{V}(t, y_1)}{\tilde{V}(t, y_2)} = \frac{\tilde{V}(st, sy_1)}{\tilde{V}(st, sy_2)}$$

for all $y_1 < t < y_2$ and $s \in \mathbb{R}$. For $r_1 := y_1 - t$ and $r_2 := y_2 - t$, Equation (47) thus gives us

$$\frac{\psi(r_1)}{\psi(r_2)} = \frac{\psi(sr_1)}{\psi(sr_2)} \tag{48}$$

for all $r_1 < 0 < r_2$ and $s > 0$. In particular, for $r_1 := -1$ and $r_2 := 1$, we get $\psi(s) = \psi(1)\psi(-s)$ for all $s > 0$. Similarly, for $r_1 := -1$ and $r_2 := s > 0$ we find

$$\frac{\psi(-1)}{\psi(s)} = \frac{\psi(-s)}{\psi(s^2)} = \frac{\psi(s)}{\psi(1)\psi(s^2)},$$

and thus we obtain $\psi(s)\psi(s) = \psi(1)\psi(s^2)$ for all $s > 0$. Furthermore, considering $r_1 := -t$ and $r_2 := s$ for $s, t > 0$, we find

$$\frac{\psi(t)}{\psi(s)} = \frac{\psi(1)\psi(-t)}{\psi(s)} = \frac{\psi(1)\psi(-st)}{\psi(s^2)} = \frac{\psi(1)\psi(st)}{\psi(s)\psi(s)}$$

and hence the functional equations are proven. Let us finally assume that $\psi(s) = \psi(1)$ for all $s > 0$. Note that this yields $\psi(s) = \psi(-1)$ for all $s < 0$. Our goal is to show that $V$ given by (22) is not an identification function for $T$, which means that we can exclude this case altogether. To this end, we assume the converse, that is, $V$ given by (22) is an identification function for $T$. Let us again consider the variable $\mathsf{Y}$ given by (42), where this time we set $y_1 := -1$, $t := 0$, and

$y_2 := 1$. Moreover, we replace the generic identification function $\tilde{V}$ by $V$. Then we already know that $T(\mathsf{Y}) = 0$. However, we also have

$$\mathbb{E}_\nu(1/2, \mathsf{Y}) = V(1/2, -1)(1 - p) + V(1/2, 1)p = \psi(-1 - 1/2)(1 - p) + \psi(1 - 1/2)p$$
$$= \psi(-1)(1 - p) + \psi(1)p$$
$$= \mathbb{E}_\nu(0, \mathsf{Y}) = 0,$$

and thus we conclude that $T(\mathsf{Y}) = 1/2$, since $V$ was assumed to be an identification function.

*iii).* Let us assume that $\psi$ is not decreasing on $(0, \infty)$. Then there exists $0 < y_1 < y_2$ with $\psi(y_1) < \psi(y_2)$. For a fixed $y_3 < 0$ we now define

$$p := \frac{-\psi(y_3)}{\psi(y_1) - \psi(y_3)} \qquad \text{and} \qquad q := \frac{\psi(y_2)}{\psi(y_2) - \psi(y_3)}.$$

Note that $\psi(y_1) < 0$ and $-\psi(y_3) < 0$ imply $p \in (0, 1)$, and analogously we ensure $q \in (0, 1)$. Moreover, we have $1 - q > p$, since $\psi(y_1) - \psi(y_3) < \psi(y_2) - \psi(y_3) < 0$ implies $\frac{1}{\psi(y_2) - \psi(y_3)} < \frac{1}{\psi(y_1) - \psi(y_3)}$, and thus

$$1 - q = \frac{-\psi(y_3)}{\psi(y_2) - \psi(y_3)} > \frac{-\psi(y_3)}{\psi(y_1) - \psi(y_3)} = p.$$

Since $(\Omega, \mathcal{A}, \nu)$ is atom-free, there then exist disjoint $A, B \in \mathcal{A}$ with $\nu(A) = p$ and $\nu(B) = q$. Let us consider the random variables

$$\mathsf{Y}_1 := y_1 \mathbf{1}_A + y_3 \mathbf{1}_{\Omega \setminus A}$$
$$\mathsf{Y}_2 := y_3 \mathbf{1}_B + y_2 \mathbf{1}_{\Omega \setminus B}.$$

Since $A \subset \Omega \setminus B$, we then have $\mathsf{Y}_1 = y_1 < y_2 = \mathsf{Y}_2$ on $A$. Similarly, $B \subset \Omega \setminus A$ implies $\mathsf{Y}_1 = y_3 = \mathsf{Y}_2$ on $B$, and on the remaining set $\Omega \setminus (A \cup B)$, we have $\mathsf{Y}_1 = y_3 < y_2 = \mathsf{Y}_2$. Consequently, we have $\mathsf{Y}_1 \leq \mathsf{Y}_2$ and thus

$$T(\mathsf{Y}_1 - \mathsf{Y}_2) \leq T(0) = 0. \tag{49}$$

On the other hand, we have

$$\mathbb{E}_\nu V(0, \mathsf{Y}_1) = \psi(y_1)p + \psi(y_3)(1 - p) = -\frac{\psi(y_1)\psi(y_3)}{\psi(y_1) - \psi(y_3)} + \frac{\psi(y_3)\psi(y_1)}{\psi(y_1) - \psi(y_3)} = 0$$

and

$$\mathbb{E}_\nu V(0, \mathsf{Y}_2) = \psi(y_3)q + \psi(y_2)(1 - q) = \frac{\psi(y_3)\psi(y_2)}{\psi(y_2) - \psi(y_3)} - \frac{\psi(y_2)\psi(y_3)}{\psi(y_2) - \psi(y_3)} = 0.$$

Consequently, we find $T(\mathsf{Y}_1) = T(\mathsf{Y}_2) = 0$, and thus we obtain $0 = T(\mathsf{Y}_1) \leq T(\mathsf{Y}_1 - \mathsf{Y}_2) + T(\mathsf{Y}_2) = T(\mathsf{Y}_1 - \mathsf{Y}_2)$. Together with (49) we conclude that $T(\mathsf{Y}_1 - \mathsf{Y}_2) = 0$. Now consider the random variable $\mathsf{Y}_3 := (y_1 - y_2)\mathbf{1}_A$. Then $\mathsf{Y}_3 \leq 0$ implies $T(\mathsf{Y}_3) \leq T(0) = 0$. On the other hand, the construction yields $\mathsf{Y}_3 = y_1 - y_2 = \mathsf{Y}_1 - \mathsf{Y}_2$ on $A$, $\mathsf{Y}_3 = 0 = \mathsf{Y}_1 - \mathsf{Y}_2$ on $B$, and

$\mathsf{Y}_3 = 0 > y_3 - y_2 = \mathsf{Y}_1 - \mathsf{Y}_2$ on $\Omega \setminus (A \cup B)$. Consequently, we have $\mathsf{Y}_3 \geq \mathsf{Y}_1 - \mathsf{Y}_2$, and thus $T(\mathsf{Y}_3) \geq T(\mathsf{Y}_1 - \mathsf{Y}_2) = 0$. Together, these considerations show $T(\mathsf{Y}_3) = 0$, which in turn leads to

$$0 = \mathbb{E}_\nu V(0, \mathsf{Y}_3) = \psi(y_1 - y_2)p + \psi(0)(1-p) = \psi(y_1 - y_2)p \,.$$

Now, $p \neq 0$ gives $\psi(y_1 - y_2) = 0$, which contradicts $\psi(y_1 - y_2) > 0$.

*iv).* Let us define $\varphi : (0, \infty) \to (0, \infty)$ by $\varphi(s) := \frac{\psi(s)}{\psi(1)}$. By part *ii)* we then know that $\varphi$ is a group homomorphism and that there exists an $s_0 > 0$ with $\varphi(s_0) \neq 1$. Moreover, since $\psi$ is assumed to be decreasing on $(0, \infty)$, the map $\varphi$ is increasing, and hence Lemma 20 tells us that there exists an $\alpha > 0$ such that $\varphi(s) = s^\alpha$ for all $s > 0$. Now, (23) follows from part *ii)*.

*v).* Let us assume that $\psi$ is not concave. Since $\psi$ is assumed to be continuous, we conclude by (Behringer, 1992, Theorems 8 and 10) that $\psi$ is not mid-point concave, i.e. there exist $y_1, y_2 \in \mathbb{R}$ such that

$$\psi\Big(\frac{y_1 + y_2}{2}\Big) < \frac{\psi(y_1)}{2} + \frac{\psi(y_2)}{2} \,.$$

Now, $\psi$ is assumed to be surjective, and thus we find a $y_3 \in \mathbb{R}$ such that

$$-\frac{\psi(y_1)}{2} - \frac{\psi(y_2)}{2} < \psi(y_3) < -\psi\Big(\frac{y_1 + y_2}{2}\Big) \,.$$

For some fixed, disjoint $A, B \in \mathcal{A}$ with $\nu(A) = \nu(B) = 1/4$, we consider the random variables

$$\mathsf{Y}_1 := y_1 \mathbf{1}_A + y_2 \mathbf{1}_B + y_3 \mathbf{1}_{\Omega \setminus (A \cup B)}$$
$$\mathsf{Y}_2 := y_2 \mathbf{1}_A + y_1 \mathbf{1}_B + y_3 \mathbf{1}_{\Omega \setminus (A \cup B)} \,.$$

For $V$ given by (22) this construction yields

$$\mathbb{E}_\nu V(0, \mathsf{Y}_1) = \frac{\psi(y_1)}{4} + \frac{\psi(y_2)}{4} + \frac{\psi(y_3)}{2} > 0 \,,$$

and analogously, $\mathbb{E}_\nu V(0, \mathsf{Y}_2) > 0$. Since $V$ is an oriented identification function, we conclude that $T(\mathsf{Y}_1) < 0$ and $T(\mathsf{Y}_2) < 0$. Furthermore, we have

$$\mathbb{E}_\nu V\Big(0, \frac{\mathsf{Y}_1 + \mathsf{Y}_2}{2}\Big) = \frac{1}{4}\psi\Big(\frac{y_1 + y_2}{2}\Big) + \frac{1}{4}\psi\Big(\frac{y_1 + y_2}{2}\Big) + \frac{\psi(y_3)}{2} < 0$$

and thus $T(\frac{\mathsf{Y}_1 + \mathsf{Y}_2}{2}) > 0$. Together, these consideration give

$$\frac{T(\mathsf{Y}_1)}{2} + \frac{T(\mathsf{Y}_2)}{2} < 0 < T\Big(\frac{\mathsf{Y}_1 + \mathsf{Y}_2}{2}\Big) \,,$$

which contradicts the assumed convexity of $T$. ∎

## Appendix F. An Abstract Separation Theorem

The goal of this Appendix is to present a rather generic separation result for Banach spaces. Note that the results of this appendix are entirely independent of all results presented so far with the exception of Lemma 14, which itself is independent of the rest of the paper.

Let us begin by fixing some notations. To this end, let $(E, \| \cdot \|_E)$ be a normed space. We write $E'$ for its dual and $B_E$ for its closed unit ball. Moreover, for an $A \subset E$ we write $\mathring{A}^E$ for the interior of $A$ with respect to the norm $\| \cdot \|_E$. Furthermore, $\mathrm{span}\, A$ denotes the space spanned by $A$ and $\mathrm{cone}\, A := \{\alpha x : \alpha \geq 0, x \in A\}$ denotes the cone generated by $A$. In addition, we need to make the following assumptions:

**G1 (Simplex face).** $(E, \| \cdot \|_E)$ is a normed space and $B \subset B_E$ is a non-empty, convex set for which there exists a $\varphi' \in E'$ such that $B \subset \{\varphi' = 1\}$. We write $H := \mathrm{span}\, B$.

**G2 (Non-empty relative interior).** For a fixed $x_\star \in B$, we define $A := -x_\star + B$ and

$$F := \mathrm{span}\, A$$

We assume that there exists a norm $\| \cdot \|_F$ on $F$ such that $\| \cdot \|_E \leq \| \cdot \|_F$ and $0 \in \mathring{A}^F$.

**G3 (Cone decomposition).** There exists a constant $K > 0$ such that for all $z \in H$ there exist $z^-, z^+ \in \mathrm{cone}\, B$ such that $z = z^+ - z^-$ and

$$\|z^-\|_E + \|z^+\|_E \leq K\|z\|_E \,.$$

**G4 (Continuous, quasi-monotonic functional).** We have a quasi-monotonic functional $\Gamma : B \to \mathbb{R}$ that is $\| \cdot \|_E$-continuous.

**G5 (Strictly locally non-constant).** We write $I := \Gamma(\mathring{B})$ for the interior of the image of $B$ under $\Gamma$. We assume that, for all $r \in I$, $\varepsilon > 0$, and $x \in \{\Gamma = r\}$, there exist $x^- \in \{\Gamma < r\}$ and $x^+ \in \{\Gamma > r\}$ such that $\|x - x^-\|_F \leq \varepsilon$ and $\|x - x^+\|_F \leq \varepsilon$.

Before we can formulate our separation result, we need to define a norm on $H$. This is done in the following lemma.

**Lemma 21** *Let **G1** be satisfied, and suppose that all assumptions except $0 \in \mathring{A}^F$ of **G2** are satisfied, too. Then, the space $F$ satisfies $F \subset \ker \varphi'$. In particular, we have $x_\star \notin F$ and*

$$H = F \oplus \mathbb{R}x_\star \,.$$

*If we equip $H$ with the norm $\| \cdot \|_H$, defined by*

$$\|y + \alpha x_\star\|_H := \|y\|_F + \|\alpha x_\star\|_E$$

*for all $y + \alpha x_\star \in F \oplus \mathbb{R}x_\star$, then, we have $\| \cdot \|_E \leq \| \cdot \|_H$ on $H$, $\| \cdot \|_F = \| \cdot \|_H$ on $F$.*

**Proof** Let us fix a $y \in F$. Since $F = \mathrm{span}(-x_\star + B)$, there then exists $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $x_1, \ldots, x_n \in B$ such that $y = \sum_{i=1}^n \alpha_i(-x_\star + x_i)$. By the linearity of $\varphi'$, this yields

$$\langle \varphi', y \rangle = \sum_{i=1}^n \alpha_i \big( \langle \varphi', x_i \rangle - \langle \varphi', x_\star \rangle \big) = 0 \,,$$

where in the last step we used $\langle \varphi', x_i \rangle = 1 = \langle \varphi', x_\star \rangle$. The second assertion follows from the first and $\langle \varphi', x_\star \rangle = 1$. Now, we immediately obtain $F \cap \mathbb{R}x_\star = \{0\}$, and thus $F \oplus \mathbb{R}x_\star$ is indeed a direct sum. Moreover, the equality $F \oplus \mathbb{R}x_\star = \mathrm{span}\, B$ follows from

$$\sum_{i=1}^n \alpha_i(-x_\star + x_i) + \alpha_0 x_\star = \sum_{i=1}^n \alpha_i x_i + \Big( \alpha_0 - \sum_{i=1}^n \alpha_i \Big) x_\star \,,$$

which holds for all $n \in \mathbb{N}$, $\alpha_0, \ldots, \alpha_n \in \mathbb{R}$, and $x_1, \ldots, x_n \in B$. Now, $\| \cdot \|_H$ can be constructed in the described way. Here we note, that the definition of $\| \cdot \|_H$ resembles a standard way of defining norms on direct sums, and thus $\| \cdot \|_H$ is indeed a norm. Furthermore, $\| \cdot \|_E \leq \| \cdot \|_H$ immediately follows from the construction of $\| \cdot \|_H$ and the assumed $\| \cdot \|_E \leq \| \cdot \|_F$. In addition, $\| \cdot \|_F = \| \cdot \|_H$ on $F$ is obvious. ∎

With the help of these assumptions we can now formulate the generic separation result for Banach spaces that will be used in the proof of Theorem 9. Note that in its formulation as well as in the following results we write $\|z'\|_{E'} = 1$ for the norm of a functional $z' \in (H, \| \cdot \|_E)'$.

**Theorem 22** *Assume that **G1** to **G5** are satisfied. Then, for all $r \in I$, there exists exactly one $z'_r \in (H, \| \cdot \|_E)'$ such that $\|z'_r\|_{E'} = 1$, and*

$$\{\Gamma < r\} = \{z'_r < 0\} \cap B$$
$$\{\Gamma = r\} = \{z'_r = 0\} \cap B$$
$$\{\Gamma > r\} = \{z'_r > 0\} \cap B \,.$$

In the remainder of this section, we prove Theorem 22. To this end, we assume, if not stated otherwise, throughout this section that the conditions **G1** to **G5** are satisfied. Moreover, on $B$ we consider both the metric $d_E$ induced by $\| \cdot \|_E$ and the metric $d_F$ induced by $\| \cdot \|_F$ via translation, that is

$$d_F(x_1, x_2) := \|(-x_\star + x_1) - (-x_\star + x_2)\|_F = \|x_1 - x_2\|_F = \|x_1 - x_2\|_H \,, \qquad x_1, x_2 \in B,$$

where the last identity follows from Lemma 21.

Before we can actually prove Theorem 22, we need a couple of intermediate results. We begin with some simple consequences of the assumptions **G1** to **G5**. Our first result in this direction shows that the space $H$ can be generated from $F$ and an *arbitrary* element of $B$.

**Lemma 23** *For all $x_0 \in B$ we have $F \oplus \mathbb{R}x_0 = H$.*

**Proof** [Lemma 23] By $\varphi'(x_0) = 1$ and the inclusion $F \subset \ker \varphi'$ established in Lemma 21, we see that $x_0 \notin F$, and hence $F \cap \mathbb{R}x_0 = \{0\}$.

The inclusion $F \oplus \mathbb{R}x_0 \subset H$ follows from the equality $H = \operatorname{span} B$ established in Lemma 21 and

$$\sum_{i=1}^{n} \alpha_i(-x_\star + x_i) + \alpha_0 x_0 = \sum_{i=0}^{n} \alpha_i x_i - \sum_{i=1}^{n} \alpha_i x_\star \,,$$

which holds for all $n \in \mathbb{N}$, $\alpha_0, \ldots, \alpha_n \in \mathbb{R}$, and $x_1, \ldots, x_n \in B$.

To prove the converse inclusion, we first note that $-x_\star = (-x_\star + x_0) - x_0 \in F \oplus \mathbb{R}x_0$ implies $\mathbb{R}x_\star \subset F \oplus \mathbb{R}x_0$. Since we also have $F \subset F \oplus \mathbb{R}x_0$, we conclude by Lemma 21 that $H = F \oplus \mathbb{R}x_\star \subset F \oplus \mathbb{R}x_0$. ∎

The following, trivial result compares the metrics $d_E$ and $d_F$. The only reason why we state this lemma explicitly is that we need its results several times, so that it becomes convenient to have a reference.

**Lemma 24** *The identity map* $\mathrm{id} : (B, d_F) \to (B, d_E)$ *is Lipschitz continuous. In particular, open, respectively closed, sets with respect to $d_E$ are also open, respectively closed, with respect to $d_F$.*

**Proof** [Lemma 24] The assumed inequality $\| \cdot \|_E \leq \| \cdot \|_F$ immediately implies $d_E(x_1, x_2) \leq d_F(x_1, x_2)$ for all $x_1, x_2 \in B$, and thus the identity map $\mathrm{id} : (B, d_F) \to (B, d_E)$ is indeed Lipschitz continuous. The other assertions are a direct consequence of this continuity. ∎

The next lemma shows that the cone decomposition **G3** makes it easier to decide whether a linear functional is continuous.

**Lemma 25** *A linear map $z' : H \to \mathbb{R}$ is continuous with respect to $\| \cdot \|_E$, if and only if for all sequences $(z_n) \subset \mathrm{cone}\, B$ with $\|z_n\|_E \to 0$ we have $\langle z', z_n \rangle \to 0$.*

**Proof** [Lemma 25] "$\Rightarrow$": Since $\mathrm{cone}\, B \subset H$ by the definition of $H$, this implication is trivial.

"$\Leftarrow$": By the linearity of $z'$ it suffices to show that $z'$ is $\| \cdot \|_E$-continuous in 0. To show the latter, we fix a sequence $(z_n) \subset H$ with $\|z_n\|_E \to 0$. Since $H = \mathrm{span}\, B$, there then exist sequences $(z_n^-), (z_n^+) \subset \mathrm{cone}\, B$ with $z_n = z_n^+ - z_n^-$ and $\|z_n^-\|_E + \|z_n^+\|_E \leq K\|z_n\|_E$. Consequently, we obtain $\|z_n^-\|_E \to 0$ and $\|z_n^+\|_E \to 0$, and thus our assumption together with the linearity of $z'$ yields $\langle z', z_n \rangle = \langle z', z_n^+ \rangle - \langle z', z_n^- \rangle \to 0$ ∎

The following result collects properties of the sets $\{\Gamma < r\}$ and $\{\Gamma > r\}$ we wish to separate.

**Lemma 26** *The image $\Gamma(B)$ is an interval, and, for all $r \in \mathring{\Gamma}(B)$, the sets $\{\Gamma < r\}$ and $\{\Gamma > r\}$ are convex and open in $B$ with respect to both $d_E$ and $d_F$.*

**Proof** [Lemma 26] Clearly, the sets $\{\Gamma < r\}$ and $\{\Gamma > r\}$ are open with respect to $d_E$, since $\Gamma$ is assumed to be continuous with respect to $d_E$. By Lemma 24, the sets are then also open with respect to $d_F$. Since $B$ is convex, it is connected, and thus $\Gamma(B)$ is connected by the continuity of $\Gamma$. Moreover, the only connected sets in $\mathbb{R}$ are intervals, and hence $\Gamma(B)$ is an interval. Finally, the convexity of the sets $\{\Gamma < r\}$ and $\{\Gamma > r\}$ directly follows from Lemma 14. ∎

Our next goal is to investigate relative interiors of subsets of $A$. We begin with a result that shows the richness of $\mathring{A}^F$.

**Lemma 27** *For all $r \in I$, there exists an $x \in \{\Gamma = r\}$ such that $-x_\star + x \in \mathring{A}^F$.*

**Proof** [Lemma 27] If $x_\star \in \{\Gamma = r\}$ there is nothing to prove, and hence we may assume without loss of generality that $x_\star \in \{\Gamma > r\}$. Let us write $r^\star := \Gamma(x_\star)$. Now, since $r \in I$ and $I$ is an open interval by Lemma 26, there exists an $s \in I$ with $s < r$. Let us fix an $x_0 \in \{\Gamma = s\}$. Then, for $\lambda \in [0, 1]$ we consider $x_\lambda := \lambda x_\star + (1 - \lambda)x_0$. Then we have $\Gamma(x_0) = s < r < r^\star = \Gamma(x_\star)$, and thus the intermediate theorem shows that there exists a $\lambda \in (0, 1)$ with $\Gamma(x_\lambda) = r$. Our goal is to show that this $x_\lambda$ satisfies $-x_\star + x_\lambda \in \mathring{A}^F$. To this end, we recall that $0 \in \mathring{A}^F$ gives an $\varepsilon > 0$ such that for all $y \in F$ satisfying $\|y\|_F \leq \varepsilon$ we actually have $y \in A$. Let us write $\delta := \lambda \varepsilon$. Then it

suffices to show that, for all $y \in F$ satisfying $\| -x_\star + x_\lambda - y \|_F \le \delta$, we have $y \in A$. Consequently, let us fix such a $y \in F$. For

$$\tilde{x} := x_\star + \frac{y - (1-\lambda)(-x_\star + x_0)}{\lambda}$$

we then have $y = \lambda(-x_\star + \tilde{x}) + (1-\lambda)(-x_\star + x_0)$. By the convexity of $A$ and $-x_\star + x_0 \in A$, it thus suffices to show $-x_\star + \tilde{x} \in A$. However, the latter follows from

$$\begin{aligned}
\| -x_\star + \tilde{x} \|_F &= \lambda^{-1} \| y - (1-\lambda)(-x_\star + x_0) \|_F \\
&= \lambda^{-1} \| y - x_\lambda + x_\star \|_F \\
&\le \lambda^{-1} \delta \,,
\end{aligned}$$

and thus the assertion is proven. ∎

Our last elementary result shows that having non-empty relative interior in $A$ implies a non-empty relative interior in $F$. This result will later be applied to translates of the open, non-empty sets $\{\Gamma < r\}$ and $\{\Gamma > r\}$.

**Lemma 28** *Let $K \subset A$ be an arbitrary subset with $\mathring{K}^A \ne \emptyset$, that is $K$ has non-empty relative $\| \cdot \|_F$-interior in $A$. Then, for all $y \in \mathring{K}^A$, there exists a $\delta_y \in (0, 1/2]$ such that $(1-\delta)y \in \mathring{K}^F$ for all $\delta \in (0, \delta_y]$. In particular, we have $\mathring{K}^F \ne \emptyset$.*

**Proof** [Lemma 28] By the assumed $0 \in \mathring{A}^F$, there exists an $\varepsilon_0 \in (0, 1]$ such that $\varepsilon_0 B_F \subset A$. Moreover, the assumption $y \in \mathring{K}^A$ yields an $\varepsilon_1 \in (0, \varepsilon_0]$ such that

$$(y + \varepsilon_1 B_F) \cap A \subset K \,. \tag{50}$$

We define $\delta_y := \varepsilon_1/(\varepsilon_1 + \|y\|_F)$. Then, it suffices to show that

$$(1-\delta)y + \varepsilon_1 \delta B_F \subset K \tag{51}$$

for all $\delta \in (0, \delta_y]$. To show the latter, we fix a $y_1 \in \varepsilon_1 \delta B_F$. An easy estimate then shows that $\| -\delta y + y_1 \|_F \le \delta \|y\|_F + \|y_1\|_F \le \delta(\|y\|_F + \varepsilon_1) \le \varepsilon_1$, and hence we obtain

$$(1-\delta)y + y_1 = y - \delta y + y_1 \in (y + \varepsilon_1 B_F) \,.$$

By (50) it thus suffices to show $(1-\delta)y + y_1 \in A$. Now, if $y_1 = 0$, then the latter immediately follows from $(1-\delta)y + y_1 = (1-\delta)y + \delta \cdot 0$, the convexity of $A$, and $0 \in A$. Therefore, it remains to consider the case $y_1 \ne 0$. Then we have

$$\frac{\varepsilon_0}{\|y_1\|_F} y_1 \in \varepsilon_0 B_F \subset A \,,$$

and $\frac{\|y_1\|_F}{\varepsilon_0} \le \frac{\varepsilon_1 \delta}{\varepsilon_0} \le \delta$. Consequently, the convexity of $A$ and $0 \in A$ yield

$$(1-\delta)y + y_1 = (1-\delta)y + \frac{\|y_1\|_F}{\varepsilon_0}\left(\frac{\varepsilon_0}{\|y_1\|_F} y_1\right) + \left(\delta - \frac{\|y_1\|_F}{\varepsilon_0}\right) \cdot 0 \in A \,,$$

and hence (51) follows. ∎

Our next goal is to move towards the proof of Theorem 22. This is done in a couple of intermediate results that successively establish more properties of certain, separating functionals.

We begin with a somewhat crude separation of convex subsets in $A$ that have an non-empty relative interior.

**Lemma 29** *Let* $K_-, K_+ \subset A$ *be two convex sets with* $\mathring{K}_\pm^A \neq \emptyset$ *and* $K_- \cap \mathring{K}_+^F = \emptyset$. *Then there exist a* $y' \in F'$ *and an* $s \in \mathbb{R}$ *such that*

$$K_- \subset \{y' \leq s\} \qquad and \qquad \mathring{K}_-^F \subset \{y' < s\}\,,$$
$$K_+ \subset \{y' \geq s\} \qquad and \qquad \mathring{K}_+^F \subset \{y' > s\}\,.$$

*Moreover, if* $s \leq 0$, *then we actually have* $\mathring{K}_-^A \subset \{y' < s\}$, *and, if* $s \geq 0$, *we have* $\mathring{K}_+^A \subset \{y' > s\}$.

**Proof** [Lemma 29] By Lemma 28 and the assumed $\mathring{K}_\pm^A \neq \emptyset$ we find $\mathring{K}_\pm^F \neq \emptyset$. By a version of the Hahn-Banach separation theorem, see e.g. (Megginson, 1998, Thm. 2.2.26), there thus exist a $y' \in F'$ and an $s \in \mathbb{R}$ such that

$$K_- \subset \{y' \leq s\}$$
$$K_+ \subset \{y' \geq s\}$$
$$\mathring{K}_+^F \subset \{y' > s\}\,.$$

Let us first show $\mathring{K}_-^F \subset \{y' < s\}$. To this end, we fix a $y_1 \in \mathring{K}_-^F$ and a $y_2 \in \mathring{K}_+^F$. Since $\mathring{K}_-^F$ is open in $F$, there then exists an $\lambda \in (0, 1)$ such that

$$\lambda y_2 + (1 - \lambda)y_1 = y_1 + \lambda(y_2 - y_1) \in \mathring{K}_-^F \subset K_-\,.$$

From the latter and the already obtained inclusions we conclude that

$$s \geq \langle y', \lambda y_2 + (1 - \lambda)y_1 \rangle = \lambda \langle y', y_2 \rangle + (1 - \lambda)\langle y', y_1 \rangle > \lambda s + (1 - \lambda)\langle y', y_1 \rangle\,.$$

Now, some simple transformations together with $\lambda \in (0, 1)$ yield $\langle y', y_1 \rangle < s$, i.e. we have shown $\mathring{K}_-^F \subset \{y' < s\}$.

Let us now show that $s \leq 0$ implies $\mathring{K}_-^A \subset \{y' < s\}$. To this end, we use contraposition, that is, we assume that there exists a $y \in \mathring{K}_-^A$ with $\langle y', y \rangle \geq s$. Since $\mathring{K}_-^A \subset K_-$, the already established inclusion $K_- \subset \{y' \leq s\}$ then yields $\langle y', y \rangle = s$. Moreover, by Lemma 28 there exists a $\delta > 0$ such that $(1 - \delta)y \in \mathring{K}_-^F$. From the previously established $\mathring{K}_-^F \subset \{y' < s\}$ we thus obtain

$$s > \langle y', (1 - \delta)y \rangle = (1 - \delta)s\,.$$

Clearly, this yields $\delta s > 0$, and since $\delta > 0$, we find $s > 0$. The remaining implication can be shown analogously. ∎

The next result refines the separation of Lemma 29 under additional assumptions on the sets that are to be separated.

**Proposition 30** *Let $K_-, K_0, K_+ \subset A$ be mutually disjoint, non-empty convex sets with $\mathring{K}_\pm^A = K_\pm$ and $A = K_- \cup K_0 \cup K_+$. Furthermore, assume that, for all $y \in K_0$ and $\varepsilon > 0$, we have $K_- \cap (y + \varepsilon B_F) \neq \emptyset$ and $K_+ \cap (y + \varepsilon B_F) \neq \emptyset$. Then there exist a $y' \in F'$ and an $s \in \mathbb{R}$ such that*

$$K_- = \{y' < s\} \cap A$$
$$K_0 = \{y' = s\} \cap A$$
$$K_+ = \{y' > s\} \cap A.$$

**Proof** [Proposition 30] We begin by proving $K_0 = \{y' = s\} \cap A$ with the help of Lemma 29. To this end, we first observe that we clearly have $\mathring{K}_\pm^A = K_\pm \neq \emptyset$ and $K_- \cap \mathring{K}_+^F \subset K_- \cap K_+ = \emptyset$. Consequently, Lemma 29 provides a $y' \in F'$ and an $s \in \mathbb{R}$ that satisfy the inclusions listed in Lemma 29. Our first goal is to show $K_0 \subset \{y' = s\} \cap A$. To this end, we fix a $y \in K_0$. Since $K_- \cap (y + \varepsilon B_F) \neq \emptyset$ for all $\varepsilon > 0$, we then find a sequence $(y_n) \subset K_-$ such that $y_n \to y$. By Lemma 29 we then obtain

$$\langle y', y \rangle = \lim_{n \to \infty} \langle y', y_n \rangle \leq s,$$

i.e. $y \in \{y' \leq s\} \cap A$. Using $K_+ \cap (y + \varepsilon B_F) \neq \emptyset$ for all $\varepsilon > 0$, we can analogously show $y \in \{y' \geq s\} \cap A$, and hence we obtain $y \in \{y' = s\} \cap A$.

To show the inclusion $\{y' = s\} \cap A \subset K_0$, we assume without loss of generality that $s \geq 0$. Let us now fix a $y \in A \setminus K_0$, so that our goal becomes to show $y \notin \{y' = s\} \cap A$. Now, if $y \in K_+$, we obtain $\langle y', y \rangle > s$, since we have seen in Lemma 29 that $s \geq 0$ implies $K_+ = \mathring{K}_+^A \subset \{y' > s\}$. Therefore, it remains to consider the case $y \in K_-$. Let us fix a $y_1 \in K_+$. Then we have just seen that $\langle y', y_1 \rangle > s$. For $\lambda \in [0, 1]$ we now define $y_\lambda := \lambda y_1 + (1 - \lambda)y$. Now, if there is a $\lambda \in (0, 1)$ with $\langle y', y_\lambda \rangle = s$, we obtain

$$s = \langle y', \lambda y_1 + (1 - \lambda)y \rangle = \lambda \langle y', y_1 \rangle + (1 - \lambda)\langle y', y \rangle > \lambda s + (1 - \lambda)\langle y', y \rangle,$$

that is $\langle y', y \rangle < s$. Consequently, it remains to show the existence of such a $\lambda \in (0, 1)$. Let us assume the converse, that is $x_\lambda \in K_- \cup K_+$ for all $\lambda \in (0, 1)$. Since $y_0 = y \in K_-$ and $y_1 \in K_+$, we then have

$$x_\lambda \in K_- \cup K_+ \tag{52}$$

for *all* $\lambda \in [0, 1]$. Let us now consider the map $\psi : [0, 1] \to A$ defined by $\psi(\lambda) := y_\lambda$. Clearly, $\psi$ is continuous, and since $K_\pm = \mathring{K}_\pm^A$, the pre-images $\psi^{-1}(K_-)$ and $\psi^{-1}(K_+)$ are open, and, of course, disjoint. Moreover, by $\psi(0) = y_0 = y \in K_-$ and $\psi(1) = y_1 \in K_+$, they are also non-empty, and (52) ensures $\psi^{-1}(K_-) \cup \psi^{-1}(K_+) = [0, 1]$. Consequently, we have found a partition of $[0, 1]$ consisting of two open, non-empty sets, i.e. $[0, 1]$ is not connected. Since this is obviously false, we found a contradiction finishing the proof of $\{y' = s\} \cap A \subset K_0$.

To prove the remaining two equalities, let us again assume without loss of generality that $s \geq 0$. By Lemma 29, we then know $K_+ = \mathring{K}_+^A \subset \{y' > s\} \cap A$. Conversely, for $y \in \{y' > s\} \cap A$ we have already shown $y \notin K_0$, and by the inclusion $K_- \subset \{y' \leq s\}$ established in Lemma 29 we also know $y \notin K_-$. Since $A = K_- \cup K_0 \cup K_+$, we conclude that $y \in K_+$. Consequently, we have also shown $K_+ = \{y' > s\} \cap A$, and the remaining $K_- = \{y' < s\} \cap A$ now immediately follows. ∎

The next result shows the existence of a separating functional considered in Theorem 22. Its proof heavily relies on the preceding results.

**Theorem 31** *For all $r \in I$ there exists an $z' \in H'$ such that*

$$\{\Gamma < r\} = \{z' < 0\} \cap B$$
$$\{\Gamma = r\} = \{z' = 0\} \cap B$$
$$\{\Gamma > r\} = \{z' > 0\} \cap B.$$

*Moreover, $z'$ is actually continuous with respect to $\|\cdot\|_E$.*

**Proof** [Theorem 31] Let us consider the sets

$$K_- := -x_\star + \{\Gamma < r\}$$
$$K_0 := -x_\star + \{\Gamma = r\}$$
$$K_+ := -x_\star + \{\Gamma > r\}.$$

Our first goal is to show that these sets satisfy the assumptions of Proposition 30. To this end, we first observe that $\{\Gamma < r\} \subset B$ immediately implies $K_- \subset -x_\star + B = A$, and the same argument can be applied to $K_0$ and $K_+$. Moreover, they are mutually disjoint since the defining level sets are mutually disjoint, and since $r \in \mathring{\Gamma}(B)$ they are also non-empty. The equality $A = K_- \cup K_0 \cup K_+$ follows from $B = \{\Gamma < r\} \cup \{\Gamma = r\} \cup \{\Gamma > r\}$, and the convexity of $K_-$ and $K_+$ is a consequence of the convexity of $\{\Gamma < r\}$ and $\{\Gamma > r\}$ established in Lemma 26. Similarly, the convexity of $K_0$ follows from the assumed quasi-monotonicity of $\Gamma$ by Lemma 14. Moreover, by Lemma 26, the set $\{\Gamma < r\}$ is open in $B$ with respect to $d_F$, and since the metric spaces $(B, d_F)$ and $(A, \|\cdot\|_F)$ are isometrically isomorphic via translation with $-x_\star$, we see that $K_-$ is open in $A$ with respect to $\|\cdot\|_F$. This shows $\mathring{K}_-^A = K_-$, and $\mathring{K}_+^A = K_+$ can be shown analogously. Finally, observe that for $x \in \{\Gamma = r\}$, $\varepsilon > 0$, and $y := -x_\star + x$ we have

$$\begin{aligned} K_- \cap (y + \varepsilon B_F) &= \left(-x_\star + \{\Gamma < r\}\right) \cap \left(-x_\star + x + \varepsilon B_F\right) \\ &= \left(-x_\star + \{\Gamma < r\}\right) \cap \left(-x_\star + x + \varepsilon B_H\right) \\ &= -x_\star + \left(\{\Gamma < r\} \cap (x + \varepsilon B_H)\right) \\ &\neq \emptyset, \end{aligned}$$

where in the second step we used the fact $\|\cdot\|_F = \|\cdot\|_H$ on $A \subset F$, see Lemma 21. Obviously, $K_- \cap (y + \varepsilon B_F) \neq \emptyset$ can be shown analogously, and hence, the assumptions of Proposition 30 are indeed satisfied.

Now, let $y' \in F'$ and $s \in \mathbb{R}$ be according to Proposition 30. Moreover, let $\hat{y}' \in H'$ be the extension of $y'$ to $H$ that is defined by

$$\langle \hat{y}', y + \alpha x_\star \rangle := \langle y', y \rangle$$

for all $y + \alpha x_\star \in H = F \oplus \mathbb{R}x_\star$. Clearly, $\hat{y}'$ is indeed an extension of $y'$ to $H$ and the continuity of $\hat{y}'$ on $H$ follows from

$$|\langle \hat{y}', y + \alpha x_\star \rangle| = |\langle y', y \rangle| \leq \|y'\| \cdot \|y\|_F \leq \|y'\| \cdot \|y + \alpha x_\star\|_H.$$

With the preparations, we now define an $z' \in H'$ by

$$\langle z', z \rangle := -s\langle \varphi', z \rangle + \langle \hat{y}', z - \langle \varphi', z \rangle x_\star \rangle, \qquad z \in H.$$

Indeed, $z'$ is obviously linear. Moreover, the restriction $\varphi'_{|H}$ of $\varphi'$ to $H$ is continuous with respect to $\|\cdot\|_H$, since Lemma 21 ensured $\|\cdot\|_E \leq \|\cdot\|_H$ on $H$, and consequently we obtain $z' \in H'$.

Let us show that $z'$ is the desired functional. To this end, we first observe that the inclusion $F \subset \ker \varphi'$ established in Lemma 21 together with $x_\star \in B \subset \{\varphi' = 1\}$ yields $x_\star + F \subset \{\varphi' = 1\}$. For $x \in x_\star + F \subset H$ this gives

$$\langle z', x \rangle = -s\langle \varphi', x \rangle + \langle \hat{y}', x - \langle \varphi', x \rangle x_\star \rangle = -s + \langle \hat{y}', x - x_\star \rangle = -s + \langle y', x - x_\star \rangle.$$

Moreover, recall that we have $x \in B$ if and only if $-x_\star + x \in A$, and hence we obtain

$$\begin{aligned} \{z' = 0\} \cap B &= \{x \in B : \langle y', x - x_\star \rangle = s\} \\ &= \{x \in B : -x_\star + x \in \{y' = s\}\} \\ &= x_\star + \{y \in A : y \in \{y' = s\}\} \\ &= x_\star + (\{y' = s\} \cap A) \\ &= x_\star + K_0 \\ &= \{\Gamma = r\}. \end{aligned}$$

The remaining equalities $\{\Gamma < r\} = \{z' < 0\} \cap B$ and $\{\Gamma > r\} = \{z' > 0\} \cap B$ can be shown analogously.

Let us finally show that the functional $z'$ found so far is actually continuous with respect to $\|\cdot\|_E$. Let us assume the converse. By Lemma 25, there then exists a sequence $(z_n) \subset \operatorname{cone} B$ with $\|z_n\|_E \to 0$ and $\langle z', z_n \rangle \not\to 0$. Picking a suitable subsequence and scaling it appropriately, we may assume without loss of generality that either $\langle z', z_n \rangle < -1$ for all $n \geq 1$, or $\langle z', z_n \rangle > 1$ for all $n \geq 1$. Let us consider the first case, only, the second case can be treated analogously. We begin by picking an $x_0 \in \{\Gamma > r\} = \{z' > 0\} \cap B$. This yields $\alpha := \langle z', x_0 \rangle > 0$. Moreover, since $(z_n) \subset \operatorname{cone} B$ and $z_n \neq 0$ by the assumed $\langle z', z_n \rangle < -1$, we find sequences $(\alpha_n) \subset (0, \infty)$ and $(x_n) \subset B$ such that $z_n = \alpha_n x_n$ for all $n \geq 1$. Our first goal is to show that $\alpha_n \to 0$. To this end, we observe that $x_n \in B \subset \{\varphi' = 1\}$ implies $1 = |\langle \varphi', x_n \rangle| \leq \|\varphi'\| \cdot \|x_n\|_E$, and hence we obtain

$$|\alpha_n| \leq |\alpha_n| \cdot \|\varphi'\| \cdot \|x_n\|_E = \|\varphi'\| \cdot \|z_n\|_E \to 0.$$

For $n \geq 1$, we define $\beta_n := \frac{1}{1+\alpha\alpha_n}$. Our considerations made so far then yield both $\beta_n \to 1$ and $\beta_n \in (0, 1)$ for all $n \geq 1$. By the definition of $\alpha$ and the assumptions made on $(z_n)$, this yields

$$\langle z', \beta_n(x_0 + \alpha z_n) \rangle = \beta_n(\alpha + \alpha\langle z', z_n \rangle) < 0 \tag{53}$$

for all $n \geq 1$. On the other hand, $x_0 \in \{\Gamma > r\}$ ensures $\frac{\Gamma(x_0)-r}{2} > 0$, and by the $\|\cdot\|_E$-continuity of $\Gamma$, there thus exists a $\delta > 0$ such that, for all $x \in B$ with $\|x - x_0\|_E \leq \delta$, we have

$$\left|\Gamma(x) - \Gamma(x_0)\right| \leq \frac{\Gamma(x_0) - r}{2}.$$

For such $x$, a simple transformation then yields $\Gamma(x) \geq \frac{\Gamma(x_0)+r}{2} > r$, and thus we find

$$\{x \in B : \|x - x_0\|_E \leq \delta\} \subset \{\Gamma > r\} = \{z' > 0\} \cap B.$$

To find a contradiction to (53), it thus suffices to show that

$$\beta_n(x_0 + \alpha z_n) \in \{x \in B : \|x - x_0\|_E \leq \delta\} \tag{54}$$

for all sufficiently large $n$. To prove this, we first observe that

$$\beta_n(x_0 + \alpha z_n) = \beta_n x_0 + \frac{\alpha \alpha_n}{1 + \alpha \alpha_n} x_n = \beta_n x_0 + (1 - \beta_n) x_n \,,$$

and since $\beta_n \in (0, 1)$, the convexity of $B$ yields $\beta_n(x_0 + \alpha z_n) \in B$. Finally, we have

$$\|x_0 - \beta_n(x_0 + \alpha z_n)\|_E \leq (1 - \beta_n)\|x_0\|_E + \alpha \beta_n \|z_n\|_E \to 0$$

since $\beta_n \to 1$ and $\|z_n\|_E \to 0$. Consequently, (54) is indeed satisfied for all sufficiently large $n$, which finishes the proof. ∎

Theorem 31 has shown the existence of a functional separating the level sets of $\Gamma$. Our next and final goal is to show that this functional is unique modulo normalization. To this end, we need the following lemma, which shows that the null space of a separating functional is completely determined by the set $\{\Gamma = r\}$.

**Lemma 32** *Let $r \in I$ and $z : H \to \mathbb{R}$ be a linear functional satisfying $\{\Gamma = r\} = B \cap \ker z'$. Then we have $\ker z' = \operatorname{span}(\ker z' \cap B) = \operatorname{span}\{\Gamma = r\}$ and $z' \neq 0$.*

**Proof** [Lemma 32] Since $\ker z'$ is a subspace, the inclusion $\operatorname{span}(\ker z' \cap B) \subset \ker z'$ is obvious.

To prove the converse inclusion, we fix an $z \in \ker z'$. Moreover, using Lemma 27, we fix an $x_0 \in \{\Gamma = r\} = B \cap \ker z'$ satisfying $-x_\star + x_0 \in \mathring{A}^F$. By $z \in \ker z' \subset H$ and Lemma 23, which showed $H = F \oplus \mathbb{R}x_0$, there then exist a $y \in F$ and an $\alpha \in \mathbb{R}$ such that $z = y + \alpha x_0$. Obviously, it suffices to show both $\alpha x_0 \in \operatorname{span}(\ker z' \cap B)$ and $y \in \operatorname{span}(\ker z' \cap B)$. Now, $\alpha x_0 \in \operatorname{span}(\ker z' \cap B)$ immediately follows from $x_0 \in \operatorname{span}(\ker z' \cap B)$, and for $y = 0$ the second inclusion is trivial. Therefore, let us assume that $y \neq 0$. Since $-x_\star + x_0 \in \mathring{A}^F$, there then exists an $\varepsilon > 0$ such that for all $y' \in F$ with $\| - x_\star + x_0 - y'\|_F \leq \varepsilon$ we have $y' \in A$. Writing $\hat{y} := \frac{\varepsilon}{\|y\|_F} y$, we have $\hat{y} \in F$ by the assumed $y \in F$, and thus also $\tilde{y} := -x_\star + x_0 + \hat{y} \in F$. Moreover, our construction immediately yields $\| - x_\star + x_0 - \tilde{y}\|_F = \varepsilon$, and hence we actually have $\tilde{y} \in A = -x_\star + B$. Consequently, we have found $x_0 + \hat{y} = \tilde{y} + x_\star \in B$. On the other hand, the assumed $x_0 \in \ker z'$ implies $\alpha x_0 \in \ker z'$, and thus we find $y \in \ker z'$ by $z \in \ker z'$ and $z = y + \alpha x_0$. Using both $x_0, y \in \ker z'$, we thus obtain $x_0 + \hat{y} \in \ker z'$, which together with the already established $x_0 + \hat{y} \in B$ shows $x_0 + \hat{y} \in \operatorname{span}(\ker z' \cap B)$. Since $x_0 \in B \cap \ker z'$ by assumption we therefore finally find the desired $y \in \operatorname{span}(\ker z' \cap B)$ by the definition of $\hat{y}$.

Finally, assume that $z' = 0$. By **G5** there then exists an $x \in \{\Gamma < r\}$. Then the assumed $z' = 0$ implies $x \in \ker z'$ while $\{\Gamma < r\} \subset B$ implies $x \in B$. This yields $x \in B \cap \ker z' = \{\Gamma = r\}$, which contradicts the assumed $x \in \{\Gamma < r\}$. ∎

The next lemma shows that modulo orientation, two normalized separating functionals are equal.

**Lemma 33** *Let $r \in I$ and $z_1', z_2' \in H'$ such that $\{\Gamma = r\} = B \cap \ker z_1'$ and $\{\Gamma = r\} \subset B \cap \ker z_2'$. Then there exists an $\alpha \in \mathbb{R}$ such that $z_2' = \alpha z_1'$, and if $\{\Gamma = r\} = B \cap \ker z_2'$, we actually have $\alpha \neq 0$.*

**Proof** [Lemma 33] Our assumptions guarantee $B \cap \ker z'_1 \subset B \cap \ker z'_2 \subset \ker z'_2$, and thus Lemma 32 yields $\ker z'_1 \subset \ker z'_2$. Moreover, Lemma 32 shows $z'_1 \neq 0$, which in turn gives a $z_0 \in H$ with $z_0 \notin \ker z'_1$. For $z \in H$, an easy calculation then shows that

$$z - \frac{\langle z'_1, z \rangle}{\langle z'_1, z_0 \rangle} z_0 \in \ker z'_1 \subset \ker z'_2 \,.$$

and hence we conclude that $\langle z'_2, z \rangle = \frac{\langle z'_1, z \rangle}{\langle z'_1, z_0 \rangle} \langle z'_2, z_0 \rangle$. In other words, for $\alpha := \frac{\langle z'_2, z_0 \rangle}{\langle z'_1, z_0 \rangle}$, we have $z'_2 = \alpha z'_1$. Finally, $\{\Gamma = r\} = B \cap \ker z'_2$ implies $z'_2 \neq 0$ by Lemma 32, and hence we conclude that $\alpha \neq 0$. ∎

With these preparations, we can finally present the proof of Theorem 22. Because of all the preliminary work, this proof actually reduces to a few lines.

**Proof** [Theorem 22] The existence of $z'_r$ has already be shown in Theorem 31. To show that $z'_r$ is unique, we assume that there is another $\tilde{z}'_r$ that enjoys the properties of $z'_r$. Then Lemma 33 gives an $\alpha \neq 0$ with $z'_r = \alpha \tilde{z}'_r$. The imposed normalization $\|z'_r\|_{E'} = 1 = \|\tilde{z}'_r\|_{E'}$ implies $|\alpha| = 1$, and the orientation of $z'_r$ and $\tilde{z}'_r$ on $\{\Gamma < r\}$ excludes the case $\alpha = -1$. Thus we have $z'_r = \tilde{z}'_r$. ∎

## Appendix G. Measurable Dependence of the Separating Hyperplanes

In this section we show that the separating functional found in Theorem 22 depends measurably on the level $r$ provided that some additional assumptions are satisfied.

In the following we always assume that **G1** to **G5** are satisfied. Moreover, $z'_r \in (H, \| \cdot \|_E)'$ denotes the unique separating functional found in Theorem 22. In addition to **G1** to **G5**, we consider the following assumptions:

**G6 (Measurability).** The pre-image $\Gamma^{-1}(I)$ is a Borel measurable subset of $E$.

**G7 (Completeness).** The space $E$ is a Banach space.

**G8 (Separability).** The dual space $E'$ is separable.

**G9 (Denseness).** The space $H = \operatorname{span} B$ is dense in $E$ with respect to $\| \cdot \|_E$.

The following theorem essentially shows that under these additional assumptions the map $r \mapsto z'_r$ is measurable. To formulate it, we write $\mathcal{B}(X)$ for the Borel $\sigma$-algebra of a given topological space $X$. Moreover, we equip the interval $I$ with the Lebesgue completion $\hat{\mathcal{B}}(I)$ of the Borel $\sigma$-algebra $\mathcal{B}(I)$.

**Theorem 34** *Assume that **G1** to **G9** are satisfied. Then for every $r \in I$ there exists exactly one $\hat{z}'_r \in E'$ such that $(\hat{z}'_r)_{|H} = z'_r$. Moreover, the map $Z : (I, \hat{\mathcal{B}}(I)) \to (E', \mathcal{B}(E'))$ defined by*

$$Z(r) := -\hat{z}'_r$$

*is measurable and satisfies $\|Z(r)\|_{E'} = 1$ for all $r \in I$. Moreover, for all finite measures $\nu$ on $\hat{\mathcal{B}}(I)$, the map $Z$ is Bochner $\nu$-integrable.*

To prove Theorem 34 we again need a couple of preliminary results. Most of these results consider, in one form or the other, the following function $\Psi : I \to [0, \infty)$ defined by

$$\Psi(r) := \inf_{z' \in S^+} \sup_{x \in \{\Gamma = r\}} |\langle z', x \rangle|, \qquad r \in I, \qquad (55)$$

where $S^+ := \{z' \in E' : \|z'_{|H}\|_{E'} = 1 \text{ and } \langle z', x_\star \rangle \geq 0\}$.

Our first result shows that the functionals found in Theorem 22 are essentially the only minimizers of the outer infimum in (55).

**Lemma 35** *Assume that **G1** to **G5** are satisfied. Then, for all $r \in I$, we have $\Psi(r) = 0$, and there exists a $z' \in S^+$ such that*

$$\Psi(r) = \sup_{x \in \{\Gamma = r\}} |\langle z', x \rangle|.$$

*Moreover, for every $z' \in S^+$ satisfying this equation, we have the following implications*

$$
\begin{aligned}
\Gamma(x_\star) < r & \qquad \Rightarrow \qquad & z'_{|H} = -z'_r \\
\Gamma(x_\star) = r & \qquad \Rightarrow \qquad & z'_{|H} = \pm z'_r \\
\Gamma(x_\star) > r & \qquad \Rightarrow \qquad & z'_{|H} = z'_r.
\end{aligned}
$$

**Proof** [Lemma 35] To show the existence of $z' \in S^+$, we assume without loss of generality that $\Gamma(x_\star) \geq r$. Then the unique separating functional $z'_r \in (H, \| \cdot \|_E)'$ found in Theorem 22 satisfies

$$\sup_{x \in \{\Gamma = r\}} |\langle z'_r, x \rangle| = 0,$$

and since $\Psi(r) \geq 0$, we conclude that

$$\Psi(r) = \sup_{x \in \{\Gamma = r\}} |\langle z'_r, x \rangle| = 0.$$

In addition, $\Gamma(x_\star) \geq r$ implies $\langle z'_r, x_\star \rangle \geq 0$. Extending $z'_r$ to a bounded linear functional $z' \in E'$ with the help of Hahn-Banach's extension theorem, see e.g. (Megginson, 1998, Theorem 1.9.6), then yields $z' \in S^+$, and as a by-product of the proof, we have also established $\Psi(r) = 0$.

To show the implications, we restrict our considerations to the case $\Gamma(x_\star) < r$, the remaining two cases can be treated analogously. Then the already established $\Psi(r) = 0$ yields $\langle z', x \rangle = 0$ for all $x \in \{\Gamma = r\}$, that is $\{\Gamma = r\} \subset B \cap \ker z'$. Since $\|z'_r\|_{E'} = 1 = \|z'_{|H}\|_{E'}$, we then conclude by Lemma 33 and Theorem 22 that $z'_r = -z'_{|H}$ or $z'_r = z'_{|H}$. Assume that the latter is true. Then $\Gamma(x_\star) < r$ implies $0 > \langle z'_r, x_\star \rangle = \langle z', x_\star \rangle \geq 0$, and hence we have found a contradiction. Consequently, we have $z'_r = -z'_{|H}$. ■

Our next goal is to show that there exists a measurable selection of the minimizers of the function $\Psi$. To this end, we first need to show that the inner supremum is measurable. To show this, let us now consider the functions $\Phi_n : I \times E' \to \mathbb{R}$, $n \in \mathbb{N} \cup \{\infty\}$ defined by

$$\Phi_n(r, z') := \sup_{x \in \{\Gamma = r\} \cap nB_E} |\langle z', x \rangle|, \qquad (r, z') \in I \times E'.$$

The following lemma shows that $\Phi_n$ is continuous in the second variable.

**Lemma 36** *Assume that **G1** to **G5** are satisfied. Then, for all $n \in \mathbb{N}$ and $r \in I$, the map $\Phi_n(r, \cdot) : E' \to \mathbb{R}$ is continuous.*

**Proof** [Lemma 36] For $z_1', z_2' \in E'$ the triangle inequality for suprema yields

$$
\begin{aligned}
\left| \Phi_n(r, z_1') - \Phi_n(r, z_2) \right| &= \left| \sup_{x \in \{\Gamma = r\} \cap nB_E} \left| \langle z_1', x \rangle \right| - \sup_{x \in \{\Gamma = r\} \cap nB_E} \left| \langle z_2', x \rangle \right| \right| \\
&\leq \sup_{x \in \{\Gamma = r\} \cap nB_E} \left| \langle z_1', x \rangle - \langle z_2', x \rangle \right| \\
&\leq \| z_1' - z_2' \|_{E'} \cdot n \,.
\end{aligned}
$$

Now the assertion easily follows. ∎

The next lemma shows that the function $\Phi_n$ is measurable in the first variable, provided that some technical assumptions are met.

**Lemma 37** *Assume that **G1** to **G7** are satisfied and $E$ is separable. Then, for all $n \in \mathbb{N}$ and $z' \in E'$, the map $\Phi_n(\cdot, z') : I \to \mathbb{R}$ is $\hat{\mathcal{B}}(I)$-measurable.*

**Proof** [Lemma 37] Let us write $B_0 := \Gamma^{-1}(I) \cap nB_E$. Note that $nB_E$ is closed and thus $\mathcal{B}(E)$-measurable. Since $\Gamma^{-1}(I)$ is $\mathcal{B}(E)$-measurable by **G6**, we conclude that $B_0$ is $\mathcal{B}(E)$-measurable. Consequently, $\mathbf{1}_{E \setminus B_0} : E \to \mathbb{R}$ is $\mathcal{B}(E)$-measurable, and the extension $\hat{\Gamma} : E \to \mathbb{R}$ defined by

$$
\hat{\Gamma}(z) := \begin{cases} \Gamma(z) & \text{if } z \in B_0 \\ 0 & \text{otherwise.} \end{cases}
$$

is also $\mathcal{B}(E)$-measurable. Consequently, the map $h : I \times E \to \mathbb{R}^2$ defined by

$$
h(r, z) := \left( \hat{\Gamma}(z) - r, \, \mathbf{1}_{E \setminus B_0}(z) \right), \qquad (r, z) \in I \times E
$$

is $\mathcal{B}(I) \otimes \mathcal{B}(E)$-measurable. Moreover, note that the definition of $h$ yields

$$
\{ z \in E : h(r, z) = 0 \} = \{ z \in B_0 : \Gamma(z) = r \} = \{ \Gamma = r \} \cap nB_E \,.
$$

For $F : I \to 2^E$ defined by

$$
F(r) := \left\{ z \in E : h(r, z) \in \{0\} \right\},
$$

we thus find $F(r) = \{ \Gamma = r \} \cap nB_E$ for all $r \in I$. Finally, $\xi : I \times E \to \mathbb{R}$ defined by $\xi(r, z) := |\langle z', z \rangle|$ is continuous and thus $\mathcal{B}(I \times E)$-measurable. Moreover, we have $\mathcal{B}(I \times E) = \mathcal{B}(I) \otimes \mathcal{B}(E)$ by (Bogachev, 2007b, Lemma 6.4.2) since $I$ and $E$ are both separable, and thus $\xi$ is $\mathcal{B}(I) \otimes \mathcal{B}(E)$-measurable. Since separable Banach spaces are Polish spaces, (Castaing and Valadier, 1977, Lemma III.39 on p. 86) then shows that the map

$$
r \mapsto \sup_{z \in F(r)} \xi(r, z)
$$

is $\hat{\mathcal{B}}(I)$-measurable. From the latter we easily obtain the assertion. ∎

With the help of the two previous results, the next result now establishes the desired measurability of $\Phi$. Unfortunately, it requires a stronger separability assumption than the preceding lemmas.

**Corollary 38** *Assume that **G1** to **G8** are satisfied. Then $\Phi_\infty : I \times E' \to \mathbb{R}$ is $\hat{\mathcal{B}}(I) \otimes \mathcal{B}(E')$-measurable.*

**Proof** [Corollary 38] Let us first recall, see e.g. (Megginson, 1998, Theorem 1.10.7), that dual spaces are always Banach spaces. Consequently, $E'$ is a Polish space. Moreover, the separability of $E'$ implies the separability of $E$, see e.g. (Megginson, 1998, Theorem 1.12.11), and hence the map $\Phi_n(\,\cdot\,, z') : I \to \mathbb{R}$ is $\hat{\mathcal{B}}(I)$-measurable for all $z' \in E'$ and $n \in \mathbb{N}$ by Lemma 37. Since $\Phi_n(r, \,\cdot\,) : E' \to \mathbb{R}$ is continuous for all $r \in I$ and $n \in \mathbb{N}$ by Lemma 36, we conclude that $\Phi_n$ is a Carathéodory map. Moreover, $E'$ is Polish, and thus $\Phi_n$ is $\hat{\mathcal{B}}(I) \otimes \mathcal{B}(E')$-measurable for all $n \in \mathbb{N}$, see e.g. (Castaing and Valadier, 1977, Lemma III.14 on p. 70). Finally, we have $\Phi_\infty(r, z') = \lim_{n\to\infty} \Phi_n(r, z')$ for all $(r, z') \in I \times E'$, and hence $\Phi_\infty$ is also $\hat{\mathcal{B}}(I) \otimes \mathcal{B}(E')$-measurable. ∎

The next result shows that we can find the minimizers of the infimum used in the definition of $\Psi : I \to [0, \infty)$ in a measurable fashion.

**Theorem 39** *Assume that **G1** to **G8** are satisfied. Then there exists a measurable map $\zeta : (I, \hat{\mathcal{B}}(I)) \to (E', \mathcal{B}(E'))$ such that, for all $r \in I$, we have $\zeta(r) \in S^+$ and*

$$\Psi(r) = \sup_{x \in \{\Gamma = r\}} |\langle \zeta(r), x \rangle| \, .$$

**Proof** [Theorem 39] Let us first show that $S^+$ is closed. To this end, we pick a sequence $(z'_n) \subset S^+$ that converges in norm to some $z' \in E'$. Then $\langle z'_n, x_\star \rangle \geq 0$ immediately implies $\langle z', x_\star \rangle \geq 0$. To show that $\|z'_{|H}\|_{E'} = 1$ we first observe that, for $x \in H$ with $\|x\|_E \leq 1$, we easily find

$$|\langle z', x \rangle| = \lim_{n\to\infty} |\langle z'_n, x \rangle| \leq 1 \, ,$$

and thus $\|z'_{|H}\|_{E'} \leq 1$. To show the converse inequality, we pick, for all $n \geq 1$, an $x_n \in H$ with $\|x_n\|_E \leq 1$ such that $1 - 1/n \leq |\langle z'_n, x_n \rangle| \leq 1$. Then we obtain

$$\big|\langle z', x_n \rangle - 1\big| \leq \big|\langle z' - z'_n, x_n \rangle\big| + \big|\langle z'_n, x_n \rangle - 1\big| \leq \|z' - z'_n\|_{E'} + 1/n \, ,$$

and since the right hand-side converges to 0, we find $\|z'_{|H}\|_{E'} \geq 1$. Consequently, we have shown $z \in S^+$, and therefore, $S^+$ is indeed closed. From the latter, we conclude that $\mathbf{1}_{E'\setminus S^+} : E' \to \mathbb{R}$ is $\mathcal{B}(E')$-measurable. Moreover, Corollary 38 showed that $\Phi_\infty : I \times E' \to \mathbb{R}$ is $\hat{\mathcal{B}}(I) \otimes \mathcal{B}(E')$-measurable, and consequently, the map $h : I \times E' \to \mathbb{R}^2$ defined by

$$h(r, z) := \big(\mathbf{1}_{E'\setminus S^+}(z'), \, \Phi_\infty(r, z')\big), \qquad (r, z') \in I \times E',$$

is also $\hat{\mathcal{B}}(I) \otimes \mathcal{B}(E')$-measurable. We define $F : I \to 2^{E'}$ by

$$F(r) := \big\{z' \in E' : h(r, z') = 0\big\}, \qquad r \in I.$$

Note that our construction ensures

$$F(r) = \{z \in S^+ : \Phi_\infty(r, z') = 0\} = \left\{z' \in S^+ : \Psi(r) = \sup_{x \in \{\Gamma = r\}} |\langle z', x \rangle|\right\}, \tag{56}$$

where in the last step we used the equality $\Psi(r) = 0$ established in Lemma 35. Moreover, the latter lemma also showed $F(r) \neq \emptyset$ for all $r \in I$, and since $E'$ is Polish, Aumann's measurable selection principle, see (Steinwart and Christmann, 2008, Lemma A.3.18) or (Castaing and Valadier, 1977, Theorem III.22 on p. 74) yields a a measurable map $\zeta : (I, \hat{\mathcal{B}}(I)) \to (E', \mathcal{B}(E'))$ with $\zeta(r) \in F(r)$ for all $r \in I$. Then (56) shows that $\zeta$ is the desired map. ∎

With these preparations, we can finally prove Theorem 34. The basic idea behind this proof is to combine Lemma 35 and Theorem 39.

**Proof** [Theorem 34] Since $z'_r$ is a bounded linear functional on $(H, \|\cdot\|_E)$, and $H$ is dense in $E$ by **G9**, the existence of the unique extension follows from e.g. (Megginson, 1998, Theorem 1.9.1). Moreover, this theorem also shows that $\|Z(r)\|_{E'} = \|\hat{z}'_r\| = \|z'_r\| = 1$.

Let us now consider the measurable selection $\zeta : I \to E'$ from Theorem 39. Furthermore, we fix an $r \in I$. If $r > \Gamma(x_\star)$, then Lemma 35 shows that $\zeta(r)_{|H} = -z'_r$, and thus $\zeta(r) = -\hat{z}'_r$. Analogously, $r < \Gamma(x_\star)$ implies $\zeta(r) = \hat{z}'_r$, and in the case $r = \Gamma(x_\star)$ we have either $\zeta(r) = -\hat{z}'_r$ or $\zeta(r) = \hat{z}'_r$. From these relations it is easy to obtain the desired measurability of $Z : (I, \hat{\mathcal{B}}(I)) \to (E', \mathcal{B}(E'))$.

Since the image $Z(I)$ is separable by the separability of $E'$, we further see by (Dinculeanu, 2000, Theorem 8, page 5) that $Z$ is an $E$-valued measurable function in the sense of Bochner integration theory. Finally, we have already seen that $\|Z(\cdot)\|_{E'}$ is bounded and hence $Z$ is indeed Bochner $\nu$-integrable for all finite measures $\nu$ on $\hat{\mathcal{B}}(I)$. ∎

Ingo Steinwart
Universität Stuttgart
Fachbereich Mathematik
Pfaffenwaldring 57
70569 Stuttgart
Germany
**E-Mail:** `INGO.STEINWART@MATHEMATIK.UNI-STUTTGART.DE`

Chloé Pasin
ENS Cachan
**E-Mail:** `CPASIN@ENS-CACHAN.FR`

Robert Williamson
ANU and NICTA
**E-Mail:** `BOB.ILLIAMSON@ANU.EDU.AU`

Siyu Zhang†
ENS Cachan

## Erschienene Preprints ab Nummer 2007/2007-001

2014-008 *Steinwart, I.:*  Measuring the Capacity of Sets of Functions in the Analysis of ERM

2014-007 *Steinwart, I.:*  Convergence Types and Rates in Generic Karhunen-Loève Expansions with Applications to Sample Path Properties

2014-006 *Steinwart, I.; Pasin, C.; Williamson, R.; Zhang, S.:*  Elicitation and Identification of Properties

2014-005 *Schmid, J.; Griesemer, M.:*  Integration of Non-Autonomous Linear Evolution Equations

2014-004 *Markhasin, L.:*  $L_2$- and $S_{p,q}^r B$-discrepancy of (order 2) digital nets

2014-003 *Markhasin, L.:*  Discrepancy and integration in function spaces with dominating mixed smoothness

2014-002 *Eberts, M.; Steinwart, I.:*  Optimal Learning Rates for Localized SVMs

2014-001 *Giesselmann, J.:*  A relative entropy approach to convergence of a low order approximation to a nonlinear elasticity model with viscosity and capillarity

2013-016 *Steinwart, I.:*  Fully Adaptive Density-Based Clustering

2013-015 *Steinwart, I.:*  Some Remarks on the Statistical Analysis of SVMs and Related Methods

2013-014 *Rohde, C.; Zeiler, C.:*  A Relaxation Riemann Solver for Compressible Two-Phase Flow with Phase Transition and Surface Tension

2013-013 *Moroianu, A.; Semmelmann, U.:*  Generalized Killling spinors on Einstein manifolds

2013-012 *Moroianu, A.; Semmelmann, U.:*  Generalized Killing Spinors on Spheres

2013-011 *Kohls, K; Rösch, A.; Siebert, K.G.:*  Convergence of Adaptive Finite Elements for Control Constrained Optimal Control Problems

2013-010 *Corli, A.; Rohde, C.; Schleper, V.:*  Parabolic Approximations of Diffusive-Dispersive Equations

2013-009 *Nava-Yazdani, E.; Polthier, K.:*  De Casteljau's Algorithm on Manifolds

2013-008 *Bächle, A.; Margolis, L.:*  Rational conjugacy of torsion units in integral group rings of non-solvable groups

2013-007 *Knarr, N.; Stroppel, M.J.:*  Heisenberg groups over composition algebras

2013-006 *Knarr, N.; Stroppel, M.J.:*  Heisenberg groups, semifields, and translation planes

2013-005 *Eck, C.; Kutter, M.; Sändig, A.-M.; Rohde, C.:*  A Two Scale Model for Liquid Phase Epitaxy with Elasticity: An Iterative Procedure

2013-004 *Griesemer, M.; Wellig, D.:*  The Strong-Coupling Polaron in Electromagnetic Fields

2013-003 *Kabil, B.; Rohde, C.:*  The Influence of Surface Tension and Configurational Forces on the Stability of Liquid-Vapor Interfaces

2013-002 *Devroye, L.; Ferrario, P.G.; Györfi, L.; Walk, H.:*  Strong universal consistent estimate of the minimum mean squared error

2013-001 *Kohls, K.; Rösch, A.; Siebert, K.G.:*  A Posteriori Error Analysis of Optimal Control Problems with Control Constraints

2012-018 *Kimmerle, W.; Konovalov, A.:*  On the Prime Graph of the Unit Group of Integral Group Rings of Finite Groups II

2012-017 *Stroppel, B.; Stroppel, M.:*  Desargues, Doily, Dualities, and Exceptional Isomorphisms

2012-016 *Moroianu, A.; Pilca, M.; Semmelmann, U.:* Homogeneous almost quaternion-Hermitian manifolds

2012-015 *Steinke, G.F.; Stroppel, M.J.:* Simple groups acting two-transitively on the set of generators of a finite elation Laguerre plane

2012-014 *Steinke, G.F.; Stroppel, M.J.:* Finite elation Laguerre planes admitting a two-transitive group on their set of generators

2012-013 *Diaz Ramos, J.C.; Dominguez Vázquez, M.; Kollross, A.:* Polar actions on complex hyperbolic spaces

2012-012 *Moroianu; A.; Semmelmann, U.:* Weakly complex homogeneous spaces

2012-011 *Moroianu; A.; Semmelmann, U.:* Invariant four-forms and symmetric pairs

2012-010 *Hamilton, M.J.D.:* The closure of the symplectic cone of elliptic surfaces

2012-009 *Hamilton, M.J.D.:* Iterated fibre sums of algebraic Lefschetz fibrations

2012-008 *Hamilton, M.J.D.:* The minimal genus problem for elliptic surfaces

2012-007 *Ferrario, P.:* Partitioning estimation of local variance based on nearest neighbors under censoring

2012-006 *Stroppel, M.:* Buttons, Holes and Loops of String: Lacing the Doily

2012-005 *Hantsch, F.:* Existence of Minimizers in Restricted Hartree-Fock Theory

2012-004 *Grundhöfer, T.; Stroppel, M.; Van Maldeghem, H.:* Unitals admitting all translations

2012-003 *Hamilton, M.J.D.:* Representing homology classes by symplectic surfaces

2012-002 *Hamilton, M.J.D.:* On certain exotic 4-manifolds of Akhmedov and Park

2012-001 *Jentsch, T.:* Parallel submanifolds of the real 2-Grassmannian

2011-028 *Spreer, J.:* Combinatorial 3-manifolds with cyclic automorphism group

2011-027 *Griesemer, M.; Hantsch, F.; Wellig, D.:* On the Magnetic Pekar Functional and the Existence of Bipolarons

2011-026 *Müller, S.:* Bootstrapping for Bandwidth Selection in Functional Data Regression

2011-025 *Felber, T.; Jones, D.; Kohler, M.; Walk, H.:* Weakly universally consistent static forecasting of stationary and ergodic time series via local averaging and least squares estimates

2011-024 *Jones, D.; Kohler, M.; Walk, H.:* Weakly universally consistent forecasting of stationary and ergodic time series

2011-023 *Györfi, L.; Walk, H.:* Strongly consistent nonparametric tests of conditional independence

2011-022 *Ferrario, P.G.; Walk, H.:* Nonparametric partitioning estimation of residual and local variance based on first and second nearest neighbors

2011-021 *Eberts, M.; Steinwart, I.:* Optimal regression rates for SVMs using Gaussian kernels

2011-020 *Frank, R.L.; Geisinger, L.:* Refined Semiclassical Asymptotics for Fractional Powers of the Laplace Operator

2011-019 *Frank, R.L.; Geisinger, L.:* Two-term spectral asymptotics for the Dirichlet Laplacian on a bounded domain

2011-018 *Hänel, A.; Schulz, C.; Wirth, J.:* Embedded eigenvalues for the elastic strip with cracks

2011-017 *Wirth, J.:* Thermo-elasticity for anisotropic media in higher dimensions

2011-016 *Höllig, K.; Hörner, J.:* Programming Multigrid Methods with B-Splines

2011-015 *Ferrario, P.:* Nonparametric Local Averaging Estimation of the Local Variance Function