Risk, surprise, randomization, and adversarial forecasters

Roberto Corrao*
MIT

Drew Fudenberg†
MIT

David Levine‡
EUI and WUSTL

July 21, 2023

Abstract

An *adversarial forecaster* representation sums an expected utility function and a measure of surprise that depends on an adversary's forecast. These representations are concave and satisfy a smoothness condition, and any concave preference relation that satisfies the smoothness condition has an adversarial forecaster representation. Because of concavity, the agent typically prefers to randomize. We characterize the support size of optimally chosen lotteries, and how it depends on preference for surprise.

JEL codes: D81, D91

Keywords: Surprise; random choice; adversarial forecaster

*Department of Economics, MIT, rcorrao@mit.edu

†Department of Economics, MIT, drew.fudenberg@gmail.com

‡Department of Economics, EUI and WUSTL, david@dklevine.com

We are grateful to Attila Ambrus, Oğuzhan Çelebi, Simone Cerreia-Vioglio, Philip Dybvig, Glenn Ellison, Joel Flynn, Mira Frick, Duarte Gonçalves, Ryota Ijima, Giacomo Lanzani, Fabio Maccheroni, Massimo Marinacci, Moritz Meyer-ter-Vehn, Stephen Morris, Pietro Ortoleva, Philipp Sadowski, Lorenzo Stanca, Tomasz Strzalecki, Jonathan Weinstein, and Alexander Wolitzky for helpful comments and conversations, Sivakorn Sanguanmoo for excellent research assistance, and NSF grant SES 1951056 and the Gordon B. Pye Dissertation Fellowship for financial support.

1 Introduction

Consider an agent who must choose one of their local sports team's matches to watch. They would rather watch their team win for sure than lose for sure, so if they have expected utility preferences, their most preferred match would be one where their team has probability 1 of winning. But that would be a rather boring match, and the agent would prefer to watch a match where their team is favored but not guaranteed to win, so the match has an element of suspense or surprise. Similar considerations arise in political economy in the theory of expressive voting, in which people get utility from watching a political contest, and their utility is enhanced by participation. Just as with sports matches, some may prefer a more exciting contest, so even without strategic considerations turnout is likely to be higher when the polls show a close race (see for example Levine, Modica, and Sun [2021]).

The idea that stochastic choices observed in the data may come from a deliberate desire to randomize was first advanced by Machina [1985] and is empirically supported by e.g. Agranov and Ortoleva [2017]. As expected utility does not allow a preference for randomization, we develop a minimalist departure from expected utility for which this is possible. We require that expected utility is approximately correct for comparing lotteries that are close, and that that small changes in the lottery do not change these approximations much. To ensure that the preferences are for surprise, we also require that mixtures be (weakly) preferred to the extremes, i.e. that utility is concave in probabilities.

Our definition of *continuous local expected utility* captures these three ideas. However, because the definition is not easy to work with, we introduce the *adversarial forecaster* model. Here an outcome is surprising if it is difficult to forecast in advance, where a forecast is a probability distribution over outcomes that is chosen by an adversary who attempts to minimize the forecast error. We show that this model is equivalent to continuous local expected utility.

This alternative way of describing continuous local expected utility lets us think more easily about what preferences are like - that is, what would people consider surprising under particular circumstances? It is also a powerful mathematical tool that enables us to construct classes of preferences with various properties, such as a preference for continuous densities or preferences that satisfy stochastic dominance properties. We develop and apply two large and useful classes of continuous local expected utility preferences: generalized methods of moments and transport preferences.

One tractable case is where the forecast error has a finite-dimensional parameterization. Here we show that if the forecast error is a function of k parameters and there are m moment restrictions, there is an optimal lottery with support of no more than $(k+1)(m+1)$ points. For example, in the sports case, suppose that preferences are not merely over which team wins or loses, but also over the score, where the latter can take on a continuum of values. If the forecaster is limited to predicting the mean score and there are no moment constraints, then one most preferred choice is a binary lottery between the two most extreme scores.

We then consider another tractable class of adversarial forecaster preferences, those which arise when the agent trades off the interests of different potential selves. We show that these preferences can also arise as the solution to optimal transport problems, so we call them "transport preferences." We show that optimal lotteries for these preferences can be computed by assigning to each outcome the weight of the types whose bliss points coincide with that outcome, so when the selves' preferences are more diverse, more outcomes are included in the support of the optimal lottery.[1]

We conclude our analysis by studying the monotonicity properties of adversarial forecaster preferences with respect to stochastic orders. We first show that these preferences preserve a stochastic order if and only if, for every lottery, there is a best response of the adversary that induces a utility over outcomes that respects the stochastic order. We apply this result to stochastic orders capturing risk aversion (i.e., the mean-preserving spread order) and higher-order risk aversion. In particular, we show how a preference for surprise may lead an agent with a risk-averse expected utility component to have preferences that are overall risk loving.

Related Work Our paper is related to three distinct classes of risk preference models. It is closest to other models of agents with "as-if" adversaries, e.g. Maccheroni [2002], Cerreia-Vioglio [2009], Chatterjee and Krishna [2011], Cerreia-Vioglio, Dillenberger, and Ortoleva [2015], and Fudenberg, Iijima, and Strzalecki [2015], as well as to Ely, Frankel, and Kamenica [2015], where the adversary is left implicit. When the possible outcomes are an interval of real numbers, Cerreia-Vioglio, Dillenberger, Or-

[1] In the one-dimensional case, monotone transport preferences correspond to a case of the *ordinally independent* preferences introduced by Green and Jullien [1988].

toleva, and Riella [2019] introduce a weakening of expected utility that allows optimal choices to be strictly mixed; adversarial forecaster preferences satisfy their axioms if the local utilities are strictly increasing. The adversarial forecaster model is also related to models of agents with dual selves that are not directly opposed, as in Gul and Pesendorfer [2001] and Fudenberg and Levine [2006].

The ordinally independent preferences studied in Green and Jullien [1988] have an adversarial forecaster representation provided that a supermodularity condition holds, which allows us to apply our results on optimality and monotonicity to them. The preferences induced by temporal risk in Machina [1984] are similar to adversarial forecaster preferences, but have a convex representation and so do not generate a preference for randomization.

Our analysis of monotonicity is related to the work on stochastic orders and preferences over lotteries in e.g. Cerreia-Vioglio [2009], Cerreia-Vioglio, Maccheroni, and Marinacci [2017], and Sarver [2018]. Unlike the previous results, we do not assume differentiability or finite-dimensional outcomes, and characterize monotonicity with respect to stochastic orders given a representation rather than constructing one.[2]

2 The General Model

2.1 Set Up and Definitions

We analyze preferences over lotteries that are represented by a continuous but not necessarily linear utility function V, where the lotteries $F \in \mathcal{F}$ are Borel measures over a compact metric space X of outcomes, endowed with the weak topology on measures. We say that a continuous function $w : X \to \mathbb{R}$ is a *local expected utility* at F if it is a supporting hyperplane, that is $\int w(x)d\tilde{F}(x) \geqslant V(\tilde{F})$ for every $\tilde{F} \in \mathcal{F}$ and $\int w(x)dF(x) = V(F)$. Notice that if V has a local expected utility at each F then V must be concave, so it prefers random lotteries to deterministic ones and in that sense has a preference for "surprise." This becomes a strict preference of surprise when V is strictly concave, as in some special cases that we analyze in Sections 3 and 4.

Expected utility preferences have the same local expected utility at each lottery. We weaken this to require that V has a local expected utility at every lottery F and that the local expected utility varies continuously with the lottery.

[2]See Section 5 for a more detailed discussion of these and other related results.

Definition 1. V has a *continuous local expected utility* if there is a continuous function $w : X \times \mathcal{F} \to \mathbb{R}$ such that $w(x, F)$ is a local expected utility of V at every $F \in \mathcal{F}$.

As we show in Online Appendix V, V has continuous local expected utility if and only if it is concave and Gâteaux differentiable with continuous Gâteaux derivative.[3] Moreover, the continuous local utility of V at F is a valid Gâteaux derivative for V. This observation allows us to explicitly compute the continuous local utility whenever it exists. Let δ_x denote the Dirac measure on x.

Proposition 1. *If V has continuous local expected utility $w(x, F)$, then then it is concave, and for all $F \in \mathcal{F}$ and $x \in X$ we have*

$$w(x, F) = V(F) + \left.\frac{dV((1-\lambda)F + \lambda\delta_x)}{d\lambda}\right|_{\lambda=0}$$

We now introduce a general model of preference for surprise based on a representation in the form of a zero-sum game against an adversarial forecaster. Our first result shows that these preferences are equivalent to utility functions with continuous local expected utility. Suppose that forecasts are chosen from a compact metric space Y that we call the *forecast space*. We start with a continuous function $\sigma(x, y)$ that measures the "forecast error" if x occurs and the forecast was y. This function represents the loss function of the forecaster so that we normalize it to be non-negative. We consider a forecast space rich enough so that, for any lottery F, there exists a unique forecast $\hat{y}(F)$ that minimizes the expectation of $\sigma(x, y)$ with respect to F. Moreover, since it is easy to forecast the outcome of a lottery that assigns probability 1 to a single outcome, we require that for any x, the unique minimizing forecast $\hat{y}(x)$ given the degenerate distribution δ_x that assigns probability 1 to x has forecast error 0, i.e. $\sigma(x, \hat{y}(x)) = 0$.[4] We call any function $\sigma(x, y)$ that satisfies the properties above a *forecast error*.

The adversarial forecaster tries to produce good forecasts by minimizing the expected forecast error. That is, the forecaster knows F and chooses y to min-

[3] Note that continuous local utility does not imply that there is a unique local expected utility at every point; generally there will be a continuum of local expected utilities at boundary points. Boundary points are especially important in the infinite-dimensional case since with the topology of weak convergence all points are on the boundary.

[4] Here, with abuse of notation we write $\hat{y}(x)$ in place of $\hat{y}(\delta_x)$.

imize $\int \sigma(x,y)dF(x)$. We refer to this minimum expected forecast error $\Sigma(F) = \min_{y\in Y} \int \sigma(x,y)dF(x)$ as the *suspense.*

Definition 2. A function $V : \mathcal{F} \to \mathbb{R}$ is an *adversarial forecaster utility* if

$$V(F) = \int v(x)dF(x) + \min_{y\in Y} \int \sigma(x,y)dF(x) \tag{1}$$

for some forecast space Y and forecast error function σ such that $\operatorname{argmin}_{y\in Y} \int \sigma(x,y)dF(x)$ is a singleton for all F.

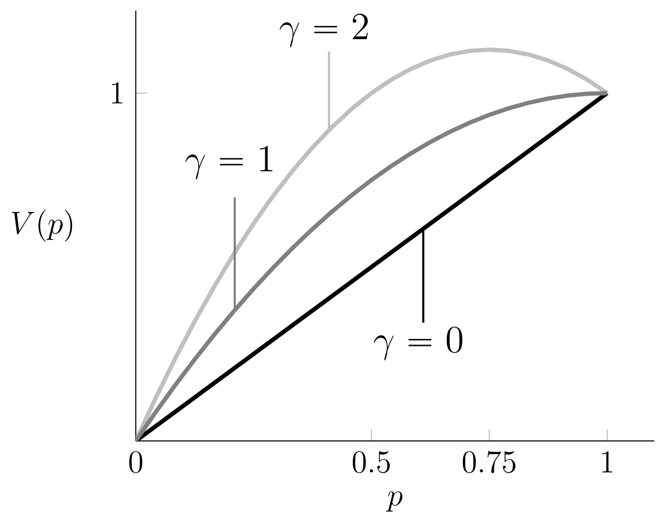
This representation can be interpreted as follows: The agent has a baseline preference over outcomes described by the expected utility function v, and a preference for surprise captured by the forecast error σ. Given a forecast of the adversary, the agent's total utility is the sum of their expected baseline utility and the expected forecast error. Equation 1 implies that V is continuous and concave, and that $V(\delta_x) = v(x)$. Note that while adversarial forecaster preferences can depart from expected utility, they do satisfy the independence axiom for comparisons of lotteries that induce the same suspense Σ.

Example 1. In a sports match, the outcome is $x = 1$ if the preferred team wins and $x = 0$ if it looses. Let p be the probability of winning, $v(x) = x$, and $\sigma(x,y) = (x-y)^2$, so the forecast error is measured by mean-squared error, where the forecast space is $Y = [0,1]$. The decision maker gets utility $v(x) = x$ plus γ times the squared error of the forecast, and the adversary's optimal choice is to forecast p, with resulting suspense $p(1-p)$, so the agent's preference over lotteries is represented by $V(p) = p + \gamma p(1-p)$. If $\gamma > 1$ and the agent can choose any value of p, the best lottery is $p = (1+\gamma)/(2\gamma)$, so the preferred team might lose, while if $0 \leqslant \gamma \leqslant 1$ the best lottery is $p = 1$. △

2.2 Equivalence of the Two Representations

Theorem 1. *A utility function has continuous local expected utility if and only if it is an adversarial forecaster utility for some forecast space and forecast error function.*

The formal proofs of this and all other results are in the appendix except where otherwise noted. Theorem 1 can be proved by noting that if V is an adversarial



$$V(p) = p + \gamma p(1-p)$$

forecaster representation, then $w(\cdot, F) = v + \sigma(\cdot, F)$ is a local expected utility of V, and that the continuity of σ implies that w is continuous. Conversely, given a representation V with continuous local expected utility w, we can set $v(x) = V(\delta_x)$, $Y = \{w(\cdot, F)\}_{F\in\mathcal{F}}$, and $\sigma(x, y) = w(x, y) - v(x)$. Because w is continuous, Y is compact, σ is continuous, and as required it has its minimum value of 0 at degenerate lotteries. Finally, we use the fact that w_V is the unique "hyperplane" (i.e., linear function) tangent to V at each F to show that σ also satisfies the uniqueness property.

2.3 Optimal lotteries

Continuous local expected utility implies the following fixed-point characterization of optimal lotteries that we use in the analysis below.

Proposition 2. *If V is an adversarial forecaster representation, then for any convex and compact set of feasible lotteries $\overline{\mathcal{F}} \subseteq \mathcal{F}$,*

$$F^* \in \operatorname*{argmax}_{F\in\overline{\mathcal{F}}} V(F) \iff F^* \in \operatorname*{argmax}_{F\in\overline{\mathcal{F}}} \int v(x) + \sigma(x, \hat{y}(F^*))dF(x). \tag{2}$$

Maximizing local expected utility is a sufficient condition for maximizing V, whether or not the local utility is continuous. The proof of necessity relies on the fact that if V has continuous local expected utility, the directional derivative of V at any lottery F in direction $\hat{F}$ is well defined and given by $\int v(x) + \sigma(x, \hat{y}(F))d\hat{F}(x)$: F^* is optimal only if the directional derivative of V at F^* in any direction is non-positive.

The necessity result fails when the local utility is not continuous. For example, if $X = [-1, 1]$ and $V(F) = \min_{y \in [-1,1]} \int_{-1}^{1} (2y-1)x dF(x)$, then $F^* = \delta_0$ is uniquely optimal over $\mathcal{F}$ for V. However, $w(x, y) = (2y-1)x$ is a local expected utility for V at F^* for every $y \in [-1, 1]$, yet F^* is strictly suboptimal for all of these local utility functions except for the one corresponding to $y = 0$.

The fixed-point condition characterizing the optimal lotteries in Proposition 2 has a clear equilibrium interpretation: The adversary chooses a forecast y given the equilibrium choice of the agent, and the agent maximizes the resulting local expected utility. The adversary's forecast is a best response if it induces the agent to choose the forecasted lottery. In particular, when $\overline{\mathcal{F}} = \mathcal{F}$, F^* is optimal if and only if $\text{supp}(F^*) \subseteq \text{argmax}_{x \in X} v(x) + \sigma(x, \hat{y}(F^*))$. In Example 1, it is easy to see that the two degenerate lotteries δ_0 and δ_1 do not satisfy this fixed-point condition when $\gamma > 1$. Instead, each optimal lottery p must assign strictly positive probability to both outcomes and, by Proposition 2, the local expected utility at p is the same for both outcomes. Some simple algebra shows that the only lottery satisfying this indifference condition is $p = (1+\gamma)/(2\gamma)$ as we show next in an extension of Example 1 to the case of a continuum of outcomes.

Example 2. Consider the setting of Example 1 but now suppose that the outcome space is continuous $X = [0, 1]$, that is the decision maker cares about the score of the game and not just who wins, so that the utility function is $V(F) = \int x dF(x) + \min_{y \in [0,1]} \int \gamma(x-y)^2 dF(x)$. This is an adversarial forecaster representation and, using Proposition 1 the local utility is $w(x, F) = x + \gamma(x - m_F)^2$, where m_F denotes the mean of F. Next, observe that for every $F \in \mathcal{F}$, the local utility $w(x, F)$ is convex in x, hence it is maximized over $\{0, 1\}$. A point mass over 0 cannot be optimal because $w(0, \delta_0) = 0 < 1 + \gamma = w(1, \delta_0)$ would contradict Proposition 2. Similarly, when $\gamma > 1$, a point mass over 1 cannot be optimal because $w(1, \delta_1) = 1 < \gamma = w(0, \delta_1)$ would yield another contradiction. Therefore, Proposition 2 implies that the optimal lottery must be binary $F_p = (1-p)\delta_0 + p\delta_1$ with $p \in (0, 1)$ defined by the indifference condition $w(0, F_p) = w(1, F_p)$, that is $p = (1+\gamma)/2\gamma$ △

Our characterization of optimal lotteries is useful for solving a rich class of problems with a unidimensional outcome space. Let $X = [0, 1]$ and, for every $F \in \mathcal{F}$, let q_F

denote the corresponding quantile function, that is, $q_F(t) = \inf\{x \in [0,1] : t \leqslant F(x)\}$. We let $\mathcal{Q}$ denote the space of lower-semicontinuous, increasing functions over X corresponding to quantiles of distributions in $\mathcal{F}$, and let $\lesssim_{st}$ denote the (incomplete) preference representing the *first-order stochastic dominance* order.[5] We consider sets of feasible distributions given by *FOSD intervals* between two deterministic outcomes $\underline{x} < \overline{x}$

$$[\underline{x}, \overline{x}]_{st} = \{G \in \mathcal{F} : \delta_{\underline{x}} \lesssim_{st} G \lesssim_{st} \delta_{\overline{x}}\}$$

We can now characterize optimal solutions of optimization problems under FOSD constraints.

Corollary 1. *If $X = [0,1]$, $\mathcal{F} = [\underline{x}, \overline{x}]_{st}$, and V has continuous local expected utility w, lottery F^* maximizes $V(F)$ over $\mathcal{F}$ if and only if its quantile function q_{F^*} satisfies*

$$q_{F^*}(t) \in \operatorname*{argmax}_{x \in [\underline{x}, \overline{x}]} v(x) + \sigma(x, \hat{y}(F^*)) \qquad \forall t \in [0,1]$$

We illustrate this result with a simple example.

Example 3. Here we extend the sport-match preferences of Example 2 to a continuum of states and risk-averse baseline preferences. We set $X = Y = [0,1]$, $v(x) = (1 - \exp(-\lambda x))/\lambda$ with $\lambda > 0$, and $\sigma(x,y) = \gamma(x-y)^2$ with $\gamma > \lambda/2$ for simplicity. The local utility at any lottery F is $w(x,F) = v(x) + \gamma(x - \overline{q}_F)^2$, and, because $\min_{x \in X} v''(x) = -\lambda$ and $\lambda < 2\gamma$, each local utility is always strictly convex in x. From Corollary 1, this implies that a lottery in $[0,1]$ is optimal if and only if its quantile function q^* satisfies

$$q^*(t) \in \operatorname*{argmax}_{x \in [0,1]} \left\{ v(x) + \gamma(x - \overline{q}^*)^2 \right\} \tag{3}$$

where $\overline{q}^* = \int_0^1 q^*(t)dt$.[6] Our restrictions on λ and γ imply that the local expected utility is strictly convex in x, so for every candidate optimal lottery F^*, the maximizers of (3) can only be 0 or 1 for every t. In particular, the unique solution to this

[5] Recall that $F \lesssim_{st} \tilde{F}$ if and only if $\int v(x)dF(x) \leqslant \int v(x)d\tilde{F}(x)$ for all increasing continuous functions v. Alternatively, $F \lesssim_{st} \tilde{F}$ if and only if $q_F(t) \leqslant q_{\tilde{F}}(t)$ for all $t \in [0,1]$.

[6] δ_0 cannot be optimal because $w(0,\delta_0) = 0 < r(\lambda) + \gamma = w(1,\delta_0)$. δ_1 is optimal if and only $w(1,\delta_1) = r(\lambda) \geqslant \gamma = w(0,\delta_1)$, which is equivalent to $1/2 + r(\lambda)/2\gamma \geqslant 1$.

maximization is

$$q^*(t) = \begin{cases} 0 & \text{if} \quad t \leqslant 1 - \overline{q}^* \\ 1 & \text{if} \quad t > 1 - \overline{q}^*. \end{cases} \tag{4}$$

where $\overline{q}^* = \min\{1/2 + r(\lambda)/2\gamma, 1\}$ is the mean of the optimal lottery and where $r(\lambda) = (1 - \exp(-\lambda))/\lambda$. When $\overline{q}^* < 1$, the variance of the optimal lottery is $1/4 - (r(\lambda)/\gamma)^2$. This s decreasing in λ and increasing in γ, which is intuitive: agents with lower lower-base line risk aversion and more taste for surprise are willing to sacrifice more expected value for higher variance. △

2.4 Stochastic Choice

The adversarial forecaster representation is concave, and often leads to randomization; a deterministic lottery is never optimal when the representation is strictly concave.[7] Many stochastic choice representations in the literature satisfy the regularity property that enlarging the choice set cannot increase the probability of pre-existing alternatives, but this is not true for adversarial forecaster preferences.[8] The next example shows how a preference for surprise reduces the agent's local risk aversion and can lead regularity to fail.

Example 4. Suppose that $X = Y \subseteq \mathbb{R}$ is an interval, that the agent's baseline utility v is concave and twice continuously differentiable, and that the agent's preference for surprise is given by $\sigma(x, y) = (x - y)^2$. As in Example 2, the continuous local utility of V is $w(x, F) = v(x) + \left(x - \int_0^1 \tilde{x} dF(\tilde{x})\right)^2$. Observe that the agent's ranking of two lotteries with the same expected value $\overline{x}$ is the same as that of an expected utility agent with utility function $w(x) = v(x) + (x - \overline{x})^2$, which is less risk averse than v. Moreover, the stochastic choice rule induced by these preferences need not satisfy Regularity. For example, if $v(x) = x$, the uniquely optimal choice for the agent from $\Delta(\{-1, 0\})$ is δ_0, so there is no suspense. In contrast, when $\Delta(\{-1, 0, 1\})$, the optimal lottery is $1/4\delta_{-1} + 3/4\delta_1$: the agent tolerates the risk of the bad outcome -1 when it can be accompanied by a larger chance of outcome 1.[9] For general v that are not too concave, i.e. when $v'' \geqslant -2$, the local utility is convex in x for all forecasts F.

[7]See Proposition 6 for a class of strictly concave adversarial forecaster representations.

[8]A stochastic choice function P satisfies *regularity* if $P(x|\overline{X}) \leqslant P(x|\overline{X}')$ for all $x \in \overline{X}' \subseteq \overline{X}$.

[9]Note that any lottery with $3/4\delta_1 > \delta_{-1}$ is preferred to a point mass at 0.

Proposition 8 below shows this implies the agent weakly prefers any mean-preserving spread $\tilde{F}$ of F to F itself. $\triangle$

Some classes of adversarial forecaster preferences do satisfy regularity. This is true for example of the weak APU of Fudenberg, Iijima, and Strzalecki [2015] when the cost function c has bounded derivatives. The weak APU representation is defined only for finite sets X; it is given by $V(F) = \sum_{x \in X} F(x)((u(x) - c(F(x)))$ where the cost function $c : [0,1] \to \mathbb{R} \cup \{\infty\}$ is strictly convex and continuously differentiable on $(0,1)$. To have continuous local expected utility we also need to assume that the derivative c' is bounded and then the local expected utility at F^* is $w(x, F^*) = u(x) - c'(F^*(x))$.[10]

2.5 Two-stage lotteries and surprise

We now apply our model to a simple example of optimal information acquisition. We consider an agent choosing among two-stage lotteries that represent distributions over both states and intermediate information. We show that the "preference for surprise" in Ely, Frankel, and Kamenica [2015] has an adversarial forecaster representation.[11] Let $\Omega = \{0,1\}$ be a binary state space. The outcomes $x = (p, \omega)$ are elements of $X = \Delta(\Omega) \times \Omega$. The agent chooses an element of the set $\overline{\mathcal{F}}$ of lotteries that satisfy the martingale constraint $\int p dF(p) = p_F$, where p_F is the marginal of F over Ω.

The lottery resolves over two time periods: In Period 1, the agent learns their interim belief $p \in \Delta(\Omega)$, and in period 2, $\omega \in \Omega$ realizes. We assume that the agent has preference for suspense in both periods. Let $E(F) = \int_0^1 \frac{1}{2}||p - p_F||^2 dF(p) = \int_0^1 p^2 dF(p) - p_F^2$, and following Ely, Frankel, and Kamenica [2015], assume that the preference for first-period suspense is $V_1(F) = g(E(F))$ for some function $g : \mathbb{R} \to \mathbb{R}$ that is twice continuously differentiable, strictly increasing, and concave, with $g(0) = 0$. The resulting utility function V_1 has continuous local utility, so it is an adversarial forecaster representation by Theorem 1. The suspense in period 2 given interim belief p is $\sum_{\omega \in \Omega} \frac{1}{2}||\delta_\omega - p||^2 p(\omega)$, and the expected period-2 suspense is

$$V_2(F) = \int g\left(\sum_{\omega \in \Omega} \frac{1}{2}||\delta_\omega - p||^2 p(\omega)\right) dF(p) = \int_0^1 g(p - p^2) dF(p).$$

[10]The stronger version of APU requires $\lim_{q \to 0} c(q) = -\infty$ which is not consistent with continuous local expected utility.

[11]Here we assume there are only two states, but it is true for any finite state space.

Finally, the agent gets direct utility equal to $\tilde{v} \in \mathbb{R}$ when the realized state is $\omega = 1$ and direct utility 0 when $\omega = 0$; the case $\tilde{v} = 0$ yields the preferences in Ely, Frankel, and Kamenica [2015].[12]

The overall utility of the agent is $V_\beta(F) = p_F\tilde{v} + (1-\beta)V_1(F) + \beta V_2(F)$, where $\beta \in [0,1]$ captures the relative importance of suspense across periods. The discussion above shows V_β has continuous local expected utility, so by Theorem 1 it admits an adversarial forecaster representation. The local utilities of V_β are:

$$w_\beta(p,\omega,F) = \omega\tilde{v} + (1-\beta)g'(D_2(F))(p^2 - p_F^2) + \beta g(p-p^2), \tag{5}$$

where $D_2(F) = \int \tilde{p}^2 dF(\tilde{p}) - p_F^2$. For every F, let F_Δ denote its marginal over interim beliefs $p \in \Delta(\Omega)$.

Proposition 3. *For every $\beta \in [0,1]$, there exists an optimal distribution F^* whose marginal over interim beliefs is supported on no more than three points. Moreover, there exist $\underline{\beta}, \overline{\beta} \in (0,1)$ with $\underline{\beta} \leqslant \overline{\beta}$ such that*

1. *When $\beta \geqslant \overline{\beta}$, $F_\Delta^* = \delta_{p_F^*}$ (so the intermediate stage reveals no information) and p_F^* is optimal if and only if iti solves $\max_{p\in[0,1]}\{p\tilde{v} + \beta g(p-p^2)\}$.*
2. *When $\beta \leqslant \underline{\beta}$, $F_\Delta^* = (1-p_F^*)\delta_0 + x_p^*\delta_1$ (the state is fully revealed) and p_F^* is optimal if and only if it solves $\max_{p\in[0,1]}\{p\tilde{v} + (1-\beta)g'(p-p^2)(p-p^2)\}$.*

Intuitively, when β=1 the agent only cares about second-period surprise so no information is revealed in the first period, and when $\beta = 0$ the agent only cares about first-period surprise so the state is revealed then; these conclusions extend to β near 0 and 1 by a continuity argument. The formulas for p_F^* follow from calculating the local utilities and applying the fixed-point characterization, as shown in Appendix A.

More generally, the agent might want to induce more than 2 posteriors. Section 3 derives a more general result on the support size of optimal distributions, and Online Appendix IV.A gives the complete solution for the case of linear g.

[12] In Ely, Frankel, and Kamenica [2015], x_F is fixed, so all the feasible two-stage lotteries induce the same prior belief over Ω, and flow utility at each period depends on the expected surprise for the next period given the current belief.

3 The Limits of Optimal Randomization

This section analyzes the extent of optimal randomization in a class of adversarial forecaster models called *generalized method of moments*, where the forecaster's loss function is parameterized by a set of moments.

3.1 Generalized Method of Moments Preferences

Suppose X is a closed bounded subset of an Euclidean space, and let S be any finite set. Given any continuous function $h : X \times S \to \mathbb{R}$, define $h(F, s) = \int h(x, s)dF(x)$ for all $s \in S$ and $F \in \mathcal{F}$. For a given h, we define the forecast space $Y = \prod_{s \in S} h(\mathcal{F}, s) \subseteq \mathbb{R}^S$, a compact set, and call it the set of *generalized moments*: these correspond to functions of the outcomes that are indexed by s. We now suppose that he adversary's goal is to match the collection of moments of F given by $h(x, s)$.

Definition 3. The loss function σ is based on the *generalized method of moments* (GMM)[13] if there is finite probability space (S, μ) and a continuous function $h : X \times S \to \mathbb{R}$ such that

$$\sigma(x, y) = \sum_{s \in S} \left(h(x, s) - y(s)\right)^2 \mu(s). \tag{6}$$

If $X \subseteq \mathbb{R}$ and $S = \{s_1, ..., s_m\}$ is a finite set of non-negative integers, we can take $h(x, s_j) = x^{s_j}$ for every $s_j \in S$, the standard method of moments.[14] The simplest case is $X \subseteq \mathbb{R}$ and $S = \{1\}$, as in Examples 1 and 4.

Proposition 4. *Any loss function σ based on the generalized methods of moments is a forecast error, and the suspense is quadratic*

$$\Sigma(F) = \int H(x, x)dF(x) - \int \int H(x, \tilde{x})dF(x)dF(\tilde{x})$$

where $H(x, \tilde{x}) = \sum_{s \in S} h(x, s)h(\tilde{x}, s)\mu(s)$.

This shows that GMM forecast errors generate quadratic utilities V (Machina [1982]) that are strictly concave. Chew, Epstein, and Segal [1991] show that strictly

[13] In econometrics, the generalized method of moments means minimizing a quadratic loss function on the data under the constraint that a number of generalized moment restrictions are satisfied.

[14] See for example Chapter 18 in Greene [2003].

concave quadratic utilities satisfy mixture symmetry, a weakening of both independence and betweenness that is more consistent with some experimental findings such as Hong and Waller [1986]. Proposition 3 in Dillenberger [2010] shows that preferences represented by quadratic utilities satisfy negative certainty independence (NCI) only if they are expected utility preferences. Therefore, when V is induced by a GMM forecast error and is strictly concave, the corresponding preference does not satisfy NCI. This is intuitive because NCI supposes the agent has a preference for deterministic outcomes.

3.2 Moment Restrictions and Bounds on Optimal Supports

We turn now to the study of optimization problems with support restrictions and moment constraints, e.g. that the expected outcome must be constant across lotteries, as is the case with fair insurance. We are mostly interested in the extent of optimal randomization, that is, in the size of the supports of optimal distributions.

To define the support restrictions formally, fix a closed subset $\overline{X} \subseteq X$ and a finite collection of k continuous functions $\Gamma = \{g_1, ..., g_k\} \subseteq C(X)$ together with the feasibility set

$$\mathcal{F}_\Gamma\left(\overline{X}\right) = \left\{F \in \Delta\left(\overline{X}\right) : \forall g_i \in \Gamma, \int g_i(x)\, dF(x) \leqslant 0\right\},$$

which we assume is non-empty. For example, if x is money, then $\int x dF(x) = 0$ is the constraint that the agent must choose a fair lottery. When the constraint set Γ is empty, the agent can pick any lottery with support $\overline{X}$.

When an expected-utility agent maximizes over $\mathcal{F}_\Gamma$, there are optimal lotteries that are extreme points of the set $\mathcal{F}_\Gamma$, and all the extreme points of this set are supported on up to $k+1$ points of $\overline{X}$. We now generalize this idea to the class of GMM preferences and show that the upper bound on the support of an optimal lottery depends on the number of moments defining the adversary's loss function as well as the number of moment restrictions.

Proposition 5. *When the agent has GMM utility with m moments and Γ contains k moment restrictions, there is an optimal lottery that puts positive probability on at most $m+k+1$ points.*

The proof is relatively simple, so we present it here. First, recall that for every

F the optimal forecast is $\hat{y}(F) = (h(F,s))_{s\in S}$ and define $\overline{Y} = \hat{y}(\mathcal{F}_\Gamma)$. Then the optimization problem becomes

$$\begin{aligned}\max_{F\in\mathcal{F}_\Gamma} V(F) &= \max_{F\in\mathcal{F}_\Gamma} \int \left\{ v(x) + \sum_{s\in S} \left(h(x,s) - h(F,s)\right)^2 \mu(s) \right\} dF(x) \\ &= \max_{\overline{y}\in\overline{Y}} \max_{F\in\overline{\mathcal{F}}:\hat{y}(F)=\overline{y}} \int \left\{ v(x) + \sum_{s\in S} \left(h(x,s) - \overline{y}(s)\right)^2 \mu(s) \right\} dF(x).\end{aligned}$$

Next, fix an optimal solution θ^* of the outer maximization problem. F^* solves the original problem and is consistent with $\overline{y}^*$ if and only if it solves

$$\max_{F\in\mathcal{F}_\Gamma(\overline{X}):\hat{y}(F)=\overline{y}^*} \int \left\{ v(x) + \sum_{s\in S} \left(h(x,s) - \overline{y}^*(s)\right)^2 \mu(s) \right\} dF(x) \tag{7}$$

which is linear in F: The agent behaves as if they were maximizing expected utility over all lotteries that have the optimal values of the relevant moments. Because the objective in (7) is linear in F, there is a solution in the set of extreme points of the set $\{F \in \mathcal{F}_\Gamma : h(F,\cdot) = \overline{y}^*\}$. This set is obtained by adding the m linear restrictions given by $\overline{y}^*$ to the set of probabilities over $\overline{X}$ that satisfy the k exogenous moment restrictions, and Winkler [1988] shows that the extreme points of this set are supported on at most $k+m+1$ points of $\overline{X}$. The next section introduces a broader class of adversarial forecaster representation that generalize GMM and for which a similar upper bound holds (see Theorem 2).

3.3 Parametric Adversarial Forecaster and Optimal Randomization

For GMM preferences, the forecast space is the set of generalized moments, that is, $Y = \prod_{s\in S} h(\mathcal{F}, s)$. Because S is finite, Y is a subset of a Euclidean space, so $\hat{y}(F) = (h(F,s))_{s\in S}$ can be interpreted as a finite-dimensional parameter that represents the best forecast for F. Parametric adversarial forecaster representations generalize these properties.

Definition 4. A forecast error σ is *parametric* if $Y \subseteq \mathbb{R}^m$ for some finite integer m, and σ and continuously differentiable in y. A utility function V is parametric if it has an adversarial forecaster representation with a parametric forecast error.

This definition is tailored for utility functions with an explicit adversarial forecaster representation (v, σ). However, the proof of Theorem 1 constructs a forecast error σ starting from a continuous local expected utility w of V. It is then straightforward to provide conditions on w that imply V is parametric.[15]

Example 5. We relax the GMM representation by allowing the forecaster to have distinct preferences regarding positive and negative surprises. For simplicity, we let $X = [0,1]$ and consider only the first moment.[16] Fix a strictly convex and twice continuously differentiable function $\rho : [-1,1] \to \mathbb{R}_+$ such that $\rho(0) = 0$, $\rho'(z) < 0$ if $z < 0$, and $\rho'(z) > 0$ if $z > 0$, and consider the preferences induced by

$$V(F) = \int_0^1 v(x)dF(x) + \min_{y \in Y} \int_0^1 \rho(x-y)dF(x),$$

where the space of parameters coincides with the space of outcomes, i.e. $Y = X$. These preferences arise from the parametric adversarial forecaster representation with forecast error $\sigma(x,y) = \rho(x-y)$ and a one-dimensional parameter space. Here $\hat{y}(\hat{F})$ is the unique minimizer in (5), and the suspense function is given by $\Sigma(F) = \int \rho(x - \hat{y}(F))dF(x)$ which can be interpreted as an index of the dispersion of F, without requiring symmetry. As we show in Section 5.2, this can lead to more "prudent" preferences. △

Example 6. Proposition 7 in Fudenberg, Iijima, and Strzalecki [2015] shows that V has an APU representation if and only if it has an AVU representation, that is,

$$V(F) = \sum_{x \in X} u(x)f(x) + \min_{y \in \mathbb{R}^X} \sum_{x \in X} \left[y(x) + \sum_{\tilde{x} \in X} \phi(y(\tilde{x})) \right] f(x) \tag{8}$$

where $\phi(z) := c^*(-z)$ where c^* is the convex conjugate of the original cost function c. Next, assume the bounded derivative condition on c and, for simplicity, that $\min_{r \in \mathbb{R}} (r + \phi(r)) = 0$.[17] We can then restrict the minimization in (8) to a compact

[15] It is sufficient that w can be written as $w(x,F) = \sigma(x, P(F))$ for some continuous functions $P : \mathcal{F} \to Y$ and $\sigma : X \times Y \to \mathbb{R}$ such that Y is a compact finite-dimensional set and σ is continuously differentiable in y.

[16] This generalizes to any arbitrary number of moments as in Section 3.1.

[17] This last assumption is only needed so that the baseline utility v from the adversarial forecaster representation coincides with u; it is satisfied for example by $\phi(r) = r^2/2 - r$.

set $Y \subseteq \mathbb{R}^X$ and define $\sigma(x, y) = y(x) + \sum_{\tilde{x} \in X} \phi(y(\tilde{x}))$ to obtain an adversarial forecaster representation.

The AVU representation in Equation 8 is an example of a parametric adversarial forecaster utility where the parameter space Y has dimension $m = |X|$. Now we generalize this model by considering uncertain taste shocks $y \in \mathbb{R}^X$ that are the same across certain classes of outcomes in X, thereby reducing the dimensionality of the parameter space. Fix a partition $\mathcal{P} = \{E_1, ..., E_m\}$ of X and a compact interval $I \subseteq \mathbb{R}$ that contains 0.[18] Define Y as the subset of I^X of vectors that are measurable with respect to the fixed partition and let $V(F)$ be defined as in (8) with $\mathbb{R}^X$ replaced by Y. Then for every partition, the utility function V has an adversarial forecaster representation with σ defined as in the original AVU representation. △

We now show that when the adversarial preferences are parametric and the feasible set is defined by a number of moment conditions, there is an optimal lottery whose support is a finite set of outcomes, and that the upper bound on this finite number of outcomes only depends on the dimension of the parameter space and on the number of moment restrictions defining the feasible set of lotteries. This result links the extent of optimal randomization, which is observable, to the parametric structure of the adversary's loss function. [19]

Consider a utility V with parametric forecast error σ and an arbitrary compact and convex set $\overline{\mathcal{F}} \subseteq \mathcal{F}$ of feasible lotteries. Define $\overline{Y} \equiv \hat{y}(\overline{\mathcal{F}})$ and observe that

$$\max_{F \in \overline{\mathcal{F}}} V(F) = \max_{F \in \overline{\mathcal{F}}} \int v(x) + \sigma(x, \hat{y}(F)) dF(x) \tag{9}$$
$$= \max_{\overline{y} \in \overline{Y}} \max_{F \in \overline{\mathcal{F}} : \hat{y}(F) = \overline{y}} \int v(x) + \sigma(x, \overline{y}) dF(x),$$

where the first equality follows from the definition of $\hat{y}(F)$, and the second equality follows by splitting the choice of the lottery in two parts: the agent chooses the desired value for the parameter $\overline{y} \in \overline{Y}$ and then chooses among the feasible distributions that are consistent with $\overline{y}$. As in the GMM case, we can fix an optimal solution $\overline{y}^*$ of the outer minimization problem and maximize $\int v(x) + \sigma(x, \overline{y}^*) dF(x)$ over the lotteries $\overline{\mathcal{F}}$ that satisfy $\hat{y}(F) = \overline{y}^*$. When $\hat{y}$ is linear and the dimension of Y is m, as in the

[18] Here we mean that the interval is large enough such that the solution of $\min_{r \in I} pr - \phi(r)$ is always in the interior of I.

[19] Example 11 in Online Appendix IV.B applies Theorem 2 to asymmetric parametric adversarial preferences that are not GMM.

GMM case, we can replicate the same steps of Proposition 5 to show that there exists an optimal lottery F^* that is supported on no more than $k + m + 1$ points of $\overline{X}$. However, these steps crucially rely on the linearity of $\hat{y}$. Next, we extend this result to nonlinear parametric utilities.

Theorem 2. *Fix a closed set* $\overline{X} \subseteq X$*,* $\{g_1, ..., g_k\} \subseteq C(X)$*, and let* $\overline{\mathcal{F}} = \mathcal{F}_\Gamma\left(\overline{X}\right)$*. Then there is a solution to (9) that assigns positive probability to no more than* $(k+1)(m+1)$ *points of* $\overline{X}$*.*

Our proof technique here is very different than that for the GMM bound in Proposition 5. The first step is Theorem 7 in the appendix, which uses the parametric transversality theorem to show that when $\overline{X}$ is finite, the bound stated in Theorem 2 holds generically for every optimal lottery. We then use an approximation argument on both the baseline utility v and the set of feasible outcomes to show that, for arbitrary $\overline{X}$, there always exists a solution with the same bound on the support.[20]

Theorem 2 has immediate implications for the nonlinear parametric examples. For the asymmetric GMM case of Example 5, there is an optimal lottery supported on no more than $2(k+1)$ points given the k moment restrictions in Γ. In our generalization of AVU in Example 6, the number of parameters coincides with the number of cells of the partition describing the uncertainty shock. This can be smaller than the cardinality of $|\overline{X}|$, so Theorem 2 yields a meaningful bound on the support of optimal lotteries. Moreover, as pointed out above, when $\overline{X}$ is finite our proof shows that *all* solutions must satisfy our upper bound. Our result then gives a testable prediction on the support of stochastic choices induced by AVU preferences with coarser shocks.[21]

3.4 Infinitely many moments and unbounded randomization

So far we have analyzed the minimal support of optimal lotteries under the assumption that the parameter space Y is finite dimensional. When Y is infinite dimensional, every optimal distribution can have "thick" (i.e. non-finite) support. We will show this for a class of GMM preferences with infinitely many relevant moments.

We extend GMM utilities by considering a compact probability space (S, μ) endowed with its Borel sigma algebra and a continuous function $h : X \times S \to \mathbb{R}$. As

[20] Doval and Skreta [2018] bounds the cardinality of the support of optimal distributions in some finite-dimensional constrained linear problems in information design using Carathéodory's theorem, which does not apply if the best response map $\hat{y}$ can be nonlinear.

[21] Online Appendix III.C provides an extension to the case of infinite X.

before, the forecast space is the compact set $Y = \{h(F, \cdot) \in C(S) : F \in \mathcal{F}\}$, [22] and the forecast error is

$$\sigma(x, y) = \int (h(x, s) - y(s))\, d\mu(s).$$

We can now extend Proposition 4 to the infinite-moment case.

Proposition 6. *Any loss function σ based on the generalized methods of (infinite) moments is a forecast error, and the suspense is quadratic*

$$\Sigma(F) = \int H(x, x) dF(x) - \int \int H(x, \tilde{x}) dF(x) dF(\tilde{x})$$

where $H(x, \tilde{x}) = \int h(x, s) h(\tilde{x}, s) d\mu(s)$. If μ has full support and $F \mapsto h(F, \cdot)$ is one-to-one, then Σ and V are strictly concave

Given an infinite GMM utility, we call H the *kernel* of the GMM representation. Next, we provide sufficient conditions for an infinite GMM utility to induce a unique optimal lottery that has full support over the outcome space. For simplicity, we consider to the one-dimensional case and do not impose exogenous moment restrictions on the feasible lotteries.

Proposition 7. *Assume that $X = [0, 1]$, $\Gamma = \varnothing$, the kernel of the GMM representation $H(x, \tilde{x}) = \int h(x, s) h(\tilde{x}, s) d\mu(s) = G(x - \tilde{x})$ is positive definite, and $H(0, \tilde{x})$ is non-negative, strictly decreasing (when positive), and strictly convex in $\tilde{x}$. Then there is a unique solution to (9), and it has full support over X.*

For the hypotheses of the theorem to be satisfied, the GMM adversary must have a sufficiently large set of forecasts, as in Example 10 in Online Appendix IV.B.[23] The proof uses Proposition 6 to obtain strict concavity of the function V, which implies that the unique optimal distribution F for V over $\mathcal{F}$ is characterized by first-order conditions which, together with the assumptions on H, imply that there cannot be an open set in X to which F assigns probability zero.

We close this section with a corollary of Theorem 2 and Proposition 7; its proof is in Online Appendix II.B.

[22]Compactness of Y follows by Arzelà–Ascoli theorem. Indeed, Y is closed because $\mathcal{F}$ is compact, it is uniformly bounded because $\mathcal{F} \times S$ is compact, and it is equicontinuous because h is (jointly) continuous.

[23]Example 7 in the next section shows how thick support arises with another adversarial forecaster preference that is not GMM.

Corollary 2. *Maintain the assumptions of Proposition 7, and let F denote the unique fully supported solution. There exists a sequence of GMM representations V^n with $|S^n| \in \mathbb{N}$, and a sequence of lotteries F^n such that each F^n is optimal for V^n, is supported on up to $|S^n|+1$ points, and $F^n \to F$ weakly, with* $\operatorname{supp} F^n \to \operatorname{supp} F = X$ *in the Hausdorff topology.*

Intuitively, as the number of moments that the adversary matches increases, the agent randomizes over more and more outcomes, up to the point that each outcome is in the support of the optimal lottery.[24]

4 Transport Utilities

This section considers a tractable class of adversarial forecaster preferences that allow the possibility of randomizations with thick support. These preferences arise when the agent trades off the interests of multiple selves with potentially heterogeneous intrinsic preferences for surprise. We show that the resulting adversarial forecaster representation has the form of the Kantorovich transport problem (hence the name) and analyze it using results from the optimal transport literature.

Formally, we take a forecast to be a continuous function y from a compact and convex finite-dimensional set X to $\mathbb{R}$, and we regard the forecast as the logarithm of a density, which we call the *score*. We define the surprise of the outcome x as $\max_\xi y(\xi) - y(x)$, that is, how much less likely the outcome is than the most likely outcome. In other words, an outcome with a lower score is more surprising and hence better. However, if the decision-maker is indifferent between all outcomes, then the forecaster can reduce surprise to 0 by forecasting that all outcomes are equally likely. Instead, we consider a decision maker with multiple selves that have heterogeneous preferences over outcomes.

We index the selves by $\theta \in X$, and represent the preferences of these different selves by a continuously differentiable *score adjustment function* $\phi(\theta, x)$, where a higher value $\phi(\theta, x) > \phi(\theta, x')$ indicates that type θ prefers the outcome x to the outcome x'. We then suppose that type θ evaluates outcomes using the preference adjusted score $y(x) - \phi(\theta, x)$, where lower adjusted scores are preferred. We continue to measure surprise in relative terms, so the surprise for a self θ at outcome x is

[24] Note that weak convergence does not imply Hausdorff convergence of the supports.

$\max_{\xi \in X}(y(\xi) - \phi(\theta, \xi)) - (y(x) - \phi(\theta, x))$, defined now in terms of preference adjusted scores.

Notice that for any particular self θ the forecaster can send the forecast $y(x) = \phi(\theta, x)$ so that this self has a uniform utility-adjusted forecast and is not surprised by anything and, as indicated, the model is not interesting if there is only one self or the selves are homogeneous. Instead, we assume that the multiple selves θ are uniformly distributed over X, denoting the uniform measure by U, and that the adversarial forecaster minimizes the average of the individual surprise over all selves. We also assume that the decision-maker maximizes the sum of a baseline continuous expected utility $v(x)$ and the expectation of the average surprise, that is,

$$V(F) = \int v(x) dF(x) + \inf_{y \in C(x)} \int \hat{\sigma}(x, y) dF(x) \tag{10}$$

where

$$\hat{\sigma}(x, y) = \int \left(\max_{\xi \in X} (y(\xi) - \phi(\theta, \xi)) - (y(x) - \phi(\theta, x)) \right) dU(\theta)$$

is the average of the individual score-adjusted surprises of the multiple selves.[25]

We say that the decision maker has *transport preferences* if the utility function is defined as in equation 10 for some $v(x)$ and $\phi(\theta, x)$. We call this transport preferences because, as we will show, the term $\inf_{y \in C(x)} \int \hat{\sigma}(x, y) dF(x)$ is isomorphic to the dual of the Kantorovich transport problem.

Transport preferences do not immediately have an adversarial forecaster representation because the function $\hat{\sigma}$ is not defined over a compact space Y, but we will show that we can restrict Y to be a compact subset of continuous functions to obtain a proper surprise function. However, we can already interpret $\hat{\sigma}$ as an aggregate measure of surprise across selves. As indicated, when the selves have different preferences over outcomes, the forecaster cannot choose a forecast that makes the utility adjusted forecasts the same for all the selves, leaving room for surprise and suspense.

Before showing that $\hat{\sigma}$ induces an adversarial forecaster utility, we report some of its structural properties that are useful for defining our surprise function.

Lemma 1. *The function $\hat{\sigma}$ is non-negative, continuous, and such that, for all $x \in X$, there exists $y \in C(X)$ such that $\hat{\sigma}(x, y) = 0$ and $\int \exp(y(\xi)) dU(\xi) = 1$.*

[25] It would be straightforward to generalize transport utility to the case where the selves are not uniformly distributed and may even have mass points. However, such preferences may not have an adversarial forecaster representation because the uniqueness property might fail.

To define the set Y of feasible forecasts, we observe that as ϕ is continuously differentiable on $X \times X$ it is K-Lipchitz for some Lipchitz constant K. As a prelude to defining Y we define Y^* to be the subset of $C(X)$ that are K-Lipchitz, and say that $y \in C(X)$ is *strongly ϕ-concave* if $y(x) = -\max_{\theta \in X} (y^*(\theta) - \phi(\theta, x))$ for some $y^* \in Y^*$. We now define the parameter space Y to be the strongly ϕ-concave functions y in $C(X)$ that satisfy the normalization $\int \exp(y(x))dU(x) = 1$. We show that Y is in compact as required.

Lemma 2. *The set Y is compact and, for every $F \in \mathcal{F}$, the problem*

$$\Sigma(F) = \min_{y \in C(X)} \int \hat{\sigma}(x, y) dF(x) \tag{11}$$

has a unique solution in Y.

This is proved in Appendix B. Denoting by σ the restriction of $\hat{\sigma}$ to $X \times Y$, we conclude that:

Theorem 3. *The function σ is a surprise function, $\Sigma(F) = \min_{y \in Y} \int \sigma(x, y) dF(x)$ is the corresponding suspense function, and for any continuous $v(x)$ the utility function V in* (10) *has an adversarial forecaster representation.*

This result follows easily from the previous two lemmas: As established, Y is the forecast space and the restriction σ over $X \times Y$ of $\hat{\sigma}$ is the forecast error. By Lemmas 1 and 2 Y is compact and σ satisfies all the properties of a forecast error. Therefore, for any continuous $v(x)$ the utility function V in (10) has an adversarial forecaster representation.

4.1 The Primal Representation and Optimal Lotteries

As indicated, transport preferences are linked to the Kantorovich optimal transportation problem through duality theory. Specifically

Theorem 4. *Suspense is the solution to choosing a probability measure $T \in \Delta(\Theta \times X)$ to solve the problem*

$$\Sigma(F) = \max_{T} \left(\int \int \phi(\theta, x) dU(\theta) dF(x) - \int \phi(\theta, x) dT(\theta, x) \right) \tag{12}$$

subject to $\int T(x, \theta) d\theta = F(x)$ and $\int T(x, \theta) dx = U(\theta)$.

Proof. The duality connection is established from the basic duality result (for example, Theorems 1.7 and 1.39 in Santambrogio, 2015) in which the primal transportation problem is to $\min_T \int \phi(\theta,x)dT(\theta,x)$ subject to $\int_X T(x,\theta)d\theta = F(x)$ and $\int_X T(x,\theta)dx = U(\theta)$ while the dual is to $\max_y \left(\int y(x)dF(x) + \int \min_{x\in X} (\phi(\theta,x) - y(x))\, dU(\theta)\right)$. When X is compact and ϕ continuous the duality theorem says that both the primal and dual problem have a solution and the two values are equal. To see how this connects to our transport preferences, rewrite the dual as

$$-\min_y \left(-\int_X y(x)dF(x) + \int_X \max_{x\in X} (-\phi(\theta,x) + y(x))\, dU(\theta)\right)$$
$$= -\Sigma(F) - \int\int \phi(\theta,x)dU(\theta)dF(x).$$

From the duality theorem it then follows that

$$\Sigma(F) = \int\int \phi(\theta,x)dU(\theta)dF(x) - \min_T \int \phi(\theta,x)dT(\theta,x).$$

■

Theorem 4 can be used to solve the problem of choosing a lottery $F \in \mathcal{F}$ when V is a transport utility. Define the correspondence

$$\Psi_\phi(\theta) = \underset{x\in X}{\operatorname{argmax}} \left\{ v(x) - \phi(\theta,x) + \int \phi(\tilde{\theta},x)dU(\tilde{\theta}) \right\} \tag{13}$$

and let $\psi \in \Psi_\phi$ denote an arbitrary measurable selection. For every measurable selection ψ, we let $U^\psi \in \mathcal{F}$ be the lottery defined as $U^\psi(\tilde{X}) = U(\psi^{-1}(\tilde{X}))$ for all measurable sets $\tilde{X}$.

Theorem 5. *If V is a transport utility with respect to v and ϕ, the set of optimal lotteries over $\mathcal{F}$ is the closure of $\{U^\psi \in \mathcal{F} : \psi \in \Psi_\phi\}$. Moreover, if $\Psi_\phi = \psi$ is single-valued, then the unique optimal lottery is U^ψ and its support is $\psi(\Theta)$.*

Equation 12 immediately implies that, for every $\psi \in \Psi_\phi$, the distribution U^ψ is optimal. The converse follows by a further application of the Kantorovich duality as shown in Appendix B. The correspondence Ψ_ϕ is single-valued when the objective function in equation 13 is strictly quasi-concave in x for every θ, as in Example 7 below, where we use Theorem 5 to solve for the optimal lottery and find sufficient conditions for it to be uniformly distributed over the entire space of outcomes. Moreover, when the favorite outcomes of the multiple selves span the entire space of outcomes, that is when $\psi(\Theta) = X$, as in the second specification of Example 7, the unique optimal lottery corresponds to a complete randomization over all the outcomes.[26]

Theorems 4 and 5 are particularly useful when the outcome space is one-dimensional and the score function of the selves satisfies a standard strict single-crossing condition. Let $q_F(t) = \inf\{x \in X : t \leqslant F(x)\}$ denote the quantile function of F defined in Section 2.3

Lemma 3. *If $X \subseteq \mathbb{R}$ and the partial derivative $\phi_x(\theta, x)$ is decreasing in θ, then*

$$\Sigma(F) = \int\int \phi(\theta, x) dU(\theta) dF(x) - \int_0^1 \phi(q_U(t), q_F(t)) dt. \tag{14}$$

.

The proof is in Appendix B. It uses the fact that $\phi_x(\theta, x)$ decreasing in θ to write the objective function in (12) as the integral with respect to T of a supermodular function, and then applies Theorem 4.3 in Galichon [2018] to rewrite the value of the primal transportation problem in terms of the optimal transportation map from selves $\theta\ in \Theta$ to outcomes $x \in X$: $\theta \mapsto q_F(U(\theta))$. Finally, the change of variable $t = U(\theta)$ applied to the second integral yields equation 14.

Observe that for every $F \in \mathcal{F}$, the quantile function q_F is nondecreasing and left-continuous. Moreover, for every nondecreasing and left-continuous q, the function $F_q(x) = \sup\{t \in [0,1] : x \geqslant q(t)\}$ is a CDF, that is, it nondecreasing, right-continuous, and such it is equal to 0 and 1 at the boundaries of X. Moreover, $F_q(x)$ is the unique CDF such that $F_q(x) \in q^{-1}(x)$ for all $x \in X$. This lets us find optimal lotteries by maximizng over the corresponding quantile functions.

Corollary 3. *Suppose that $X \subseteq \mathbb{R}$ is an interval and that $\phi_{\theta x} < 0$ and $\phi_{xx} < 0$. A lottery $F \in \mathcal{F}$ maximizes $V(F)$ if and only if*

$$q_F(t) = \operatorname*{argmax}_{x \in X} \left\{ v(x) - \phi(q_U(t), x) + \int \phi(\theta, x) dU(\theta) \right\} \tag{15}$$

[26] When there is $\psi \in \Psi_\varphi$ such that $\psi(\Theta)$ is finite, there is an optimal lottery supported on finitely many points, as in Theorem 2. Thus the number of different utility functions of the selves plays a role analogous to the number of parameters in parametric adversarial forecaster preferences.

for all $t \in [0, 1]$.

Proof. From equation 14 the problem of maximizing $V(F)$ becomes

$$\max_{F\in\mathcal{F}} V(F) = \max_{F\in\mathcal{F}} \left\{ \int v(x)dF(x) + \int\int \phi(\theta, x)dU(\theta)dF(x) - \int_0^1 \phi(q_U(t), q_F(t))dt \right\}$$
$$= \max_{F\in\mathcal{F}} \int_0^1 \left\{ v(q_F(t)) - \phi(q_U(t), q_F(t)) + \int \phi(\theta, q_F(t))dU(\theta) \right\} dt$$

where the last maximization is over all the nondecreasing and left-continuous functions $q(t)$. Finally, the result follows from the fact that the function $q(t)$ defined as the unique maximizer of problem (15) is nondecreasing (by Topkis Theorem) and continuous (by Berge Maximum Theorem). ■

Example 7. Consider a sports team example where $X = [-1, 1]$ represents the possible scores of a game, fix $\gamma \in [0, 1]$, and consider the baseline utility $v(x) = -(1 - \gamma)x^2$. We compare two cases of adversarial forecaster preferences. We start with a GMM utility with $Y = [-1, 1]$ and $\sigma(x, y) = \gamma(x - y)^2$ as in Example 3. In this case, the local expected utility is $w(x, F) = (2\gamma - 1)x^2 - 2\gamma x\overline{q}_F + \gamma\overline{q}_F^2$, where $\overline{q}_F = \int_0^1 q_F(t)dt$ is the expectation of F. When $\gamma < 1/2$, every local utility is strictly concave in x, so that the unique optimal lottery is a point mass on a single outcome which, by the fixed-point condition of Proposition 2 must be 0. When $\gamma > 1/2$ then every local utility is strictly convex, so Proposition 8 in the next section implies that the optimal lottery is supported on $\{-1, 1\}$. Moreover, the fixed-point condition of Proposition 2 implies that the expectation $\overline{q}_{F*}$ of the optimal lottery satisfies the indifference condition $w(-1, F^*) = w(1, F^*)$, so $\overline{q}_{F*} = 0$, and the optimal lottery gives probability 1/2 to -1 and 1, regardless of the specific value of $\gamma > 1/2$.

Next, we consider the transport utility induced by the multiple-selves utility function $\phi(\theta, x) = -\gamma\theta x$. In this case, we have $\varphi(\theta, x) = \gamma\theta x - (1 - \gamma)x^2$ which is strictly concave in x. Corollary 3 says that the quantile function $q_{F^*}(t)$ of the optimal lottery must solve

$$q_{F^*}(t) \in \operatorname*{argmax}_{x\in[-1,1]} \varphi(q_U(t), x) = \operatorname*{argmax}_{x\in[-1,1]} \left\{ \gamma(2t - 1)x - (1 - \gamma)x^2 \right\} \tag{16}$$

for all $t \in [0,1]$. The unique solution of (16) is

$$q_{F^*}(t) = \max\left\{-1, \min\left\{1, \frac{\gamma}{1-\gamma}(t-1/2)\right\}\right\},$$

which induces an optimal distribution that clearly depends on γ. When $\gamma \in [0, 2/3]$, the uniform distribution over $[-\gamma/2(1-\gamma), \gamma/2(1-\gamma)]$ is optimal, and for $\gamma = 0$ and $\gamma = 2/3$ we respectively have a mass point at 0 and a fully supported uniform lottery. When $\gamma \in (2/3, 1]$, the optimal lottery combines mass points at -1 and 1 with mass $1/2 - (1-\gamma)/\gamma$ each and a continuous uniform measure over $[-1, 1]$ with complementary total mass. In particular, when $\gamma = 1$ the continuous part vanishes. Observe that while the mean of the optimal lottery remains 0 as γ varies, its dispersion depends on γ. In fact, as γ increases, the optimal lottery increases with respect to the mean-preserving spread (MPS) order, varying from a Dirac probability over 0 (for $\gamma = 0$) to the maximal distribution with mean 0 over $[-1, 1]$ with respect to the MPS order (for $\gamma = 1$).[27] △

In the one-dimensional case, we can give a more explicit representation for the local utility of V that we use in Section 5.

Corollary 4. *Suppose that $X \subseteq \mathbb{R}$ is an interval and that $\phi_{\theta x} < 0$. The continuous local expected utility of $V(F)$ is given by*

$$w(x, F) = v(x) + \int \phi(\theta, x) dU(\theta) - \int_{x_F}^{x} \phi_x(T^{-1}(z), z) dz + k(F) \tag{17}$$

where $x_F = \min \operatorname{supp} F$, $T^{-1}(x)$ is the generalized inverse of the primal solution $T(\theta) = q_F(U(\theta))$, and $c(F)$ is a normalizing constant independent of x.

Proof. First, recall from the proof of Theorem 1 that for every $V(F)$ with a continuous expected utility, the local utility is given by $w(x, F) = v(x) + \sigma(x, \hat{y}(F))$. Given our assumptions, Theorems 3 and 4 implies that the suspense function of $V(F)$ is given by

$$\Sigma(F) = \min_{y \in y} \int \sigma(x, F) dF(x) = \max_{T \in \Delta(U,F)} \int \hat{\phi}(\theta, x) dT(\theta, x) \tag{18}$$

[27]See the next section for definitions of the MPS order and general integral stochastic orders.

where $\hat{\phi}(\theta, x) = \int \phi(\tilde{\theta}, x) dU(\tilde{\theta}) - \phi(\theta, x)$. Theorem 2.2 in Henry-Labordère and Touzi [2016] gives that the solution of the minimization problem in (18) is such that $\frac{\partial}{\partial x}\hat{y}(F)(x) = \hat{\phi}_x(T^{-1}(x), x)$ for all $x \in \operatorname{supp} F$ where $T(\theta) = q_F(U(\theta))$ is the solution of the maximization problem in (18). Therefore, there exists a constant $k(F)$ such that

$$\begin{aligned}\sigma(x, \hat{y}(F)) &= \int_{T(x_F)}^{x} \hat{\phi}_x(T^{-1}(z), z) dz + k(F) \\ &= \int \phi(\theta, x) dU(\theta) - \int_{T(x_F)}^{x} \phi_x(T^{-1}(z), z) dz + k(F)\end{aligned}$$

yielding the desired result. ■

4.2 Rank Dependence and Ordinal independence

We now connect transport utility to the ordinal independent preferences of Green and Jullien [1988], which is defined only for lotteries over the real line $X \subseteq \mathbb{R}$. *Ordinal independence* requires that if two distributions have the same tail, this tail can be modified without altering the preference between the distributions. Green and Julien show that the standard expected utility axioms with ordinal independence in place of the independence axiom, together with monotonicity, imply preferences have the representation $V(F) = \int_0^1 \varphi(t, q_F(x)) dt$ for some continuous real-valued utility function $\varphi(t, x)$ that is nondecreasing in x.[28] Define $\varphi(t, x) = v(x) - \phi(q_U(t), x) + \int \phi(\theta, x) dU(\theta)$, and observe that Equation 4 implies that if a transport preference is such that $\phi_x(\theta, x)$ is decreasing in θ and $\varphi_x(t, x) \geqslant 0$, then it belongs to the class of ordinal independent preferences. Conversely, if $\succsim$ is ordinal independent with $\varphi_x(\theta, x)$ increasing in θ, then it is a differentiable transport preference. This implies that if $\varphi(t, x)$ is differentiable and $\varphi_x(t, x)$ is decreasing in t, Green-Julien preferences have continuous local expected utility, so they admit an adversarial forecasting representation and have a preference for surprise; monotonicity with respect to x is not needed. The second case in Example 7 above has a differentiable transport preference that does not satisfy monotonicity, and induces lotteries with full support over $[0, 1]$.

[28] This generalizes the rank-dependent representations of Quiggin [1982] and Yaari [1987], where $\varphi(\theta, x) = w(\theta)v(x)$. See Green and Jullien, 1988 for a discussion regarding the additional behavior predictions that are allowed by the more general ordinal independent representation.

5 Monotonicity and behavior

This section characterizes monotonicity with respect to stochastic orders (e.g. first-order stochastic dominance, second-order stochastic dominance, and the mean-preserving spread order) in terms of the properties of the adversary's best response in the adversarial expected utility representation, and uses the characterization to analyze (higher-order) risk aversion and correlation aversion. These applications use the sufficient condition for monotonicity that we give in our characterization. The necessary condition shows the properties that the adversarial representation must have when the preferences of the agent are assumed to be monotone to begin with.

5.1 Stochastic orders and monotonicity

We start with the definition of the stochastic order induced by a set of continuous real-valued functions.

Definition 5. Fix a set $\mathcal{W} \subseteq C(X)$.

(i) The stochastic order $\gtrsim_{\mathcal{W}}$ is defined as:

$$F \gtrsim_{\mathcal{W}} \tilde{F} \iff \int w(x)dF(x) \geqslant \int w(x)d\tilde{F}(x) \quad \forall w \in \mathcal{W}. \tag{19}$$

(ii) A utility V *preserves* $\gtrsim_{\mathcal{W}}$ if for all $F, \tilde{F} \in \mathcal{F}$, $F \gtrsim_{\mathcal{W}} \tilde{F}$ implies $V(F) \geqslant V(\tilde{F})$.

Stochastic orders have been extensively used in decision theory to capture some monotonicity properties of behavior. For example, when $x \in \mathbb{R}$ represents monetary outcomes, the class of increasing functions generates the first-order stochastic dominance relation, and a preference that preserves this order is monotone increasing with respect to the realized wealth. Similarly, the class of convex functions generates the MPS order, and a preference that preserves this order is monotone increasing with respect to mean-preserving spreads. Conversely, a preference that preserves the stochastic order generated by concave functions would exhibit risk aversion.

Proposition 8. *Let V be an adversarial forecaster representation with baseline utility function v and surprise function σ), and fix a set $\mathcal{W} \subseteq C(X)$. Then V preserves $\gtrsim_{\mathcal{W}}$ if and only if $v + \sigma(\cdot, F) \in \langle \mathcal{W} \rangle$ for all $F \in \mathcal{F}$.*

Proposition 8 underlies the application to risk aversion in the next section. In this application, preferences are monotone with respect to the MPS order via Corollary 8, and so the optima are the feasible distributions that are maximal in the MPS order.

Similarly, we can apply Proposition 8 to the transport preferences introduced in Section 4. Given $X = [0,1]$, let $\mathcal{F}^* \subseteq \mathcal{F}$ denote the set of full-support and absolutely continuous probability measures on X.

Corollary 5. *Suppose that $X \subseteq \mathbb{R}$ is an interval, let V be a transport preference such that $\phi_{\theta x} < 0$, and fix a set $\mathcal{W} \subseteq C(X)$. Then V preserves $\succsim_{\mathcal{W}}$ if and only if, for all $F \in \mathcal{F}$, $w_0(x,F) = v(x) + \int \phi(\theta,x)dU(\theta) - \int_{x_F}^{x} \phi_x((q_F \circ U)^{-1}(z), z)dz$ is an element of $\langle \mathcal{W} \rangle$.*

Under the assumptions on ϕ, the local utility of V is equal to $w_0(x,F)$ up to a constant $k(F)$ that is independent of x. Given that by definition the set $\langle \mathcal{W} \rangle$ is closed with respect to constant translations, Proposition 8 then yields Corollary 5. The corollary also implies that supermodular ordinally independent preferences are monotone with respect to the MPS order if ϕ is convex in x.

5.2 Risk aversion and adversarial forecasters

Now we use the monotonicity result to show how a preference for surprise can alter the agent's higher-order risk preference. Theorem 2 shows there are optimal lotteries in $\mathcal{F}$ that are supported on at most two points. Moreover, because the local expected utility of the agent is $w(x,F) = v(x) + \rho\left(x - \hat{y}(F)\right)$, with second derivative $w''(x,F) = v''(x) + \rho''(x - \hat{y}(F))$, Corollary 8 implies that V preserves the MPS order when v is not too concave. This implies that the optimal distributions have the form $p^*\delta_1 + (1-p^*)\delta_0$ for some $p^* \in [0,1]$. And then the fixed-point characterization of optimality in Proposition 2 can be used to explicitly compute p^*, as we show in Online Appendix IV.B.

Consider the asymmetric loss function $\rho(z) = \lambda(\exp(z) - z)$, $\lambda \geqslant 0$ of Example 5. The relevant statistic is $\hat{y}(F) = \log\left(\int_0^1 \exp(x)dF(x)\right)$, that is, the (normalized) cumulant generating function evaluated at 1. With this loss function the agent prefers a positive surprise $x > \hat{y}(F)$ to a negative surprise $x < \hat{y}(F)$ of the same absolute value. The second derivative of the local expected utility at an arbitrary lottery F is $w''(x,F) = v''(x) + \lambda \exp(x - \hat{y}(F))$, so the agent is more risk averse over outcomes that are concentrated around $\hat{y}(F)$. The n-th order derivative of each local utility is $w^{(n)}(x, F) = v^{(n)}(x) + \lambda \exp(x - \hat{y}(F))$, so for λ high enough, $w^{(n)} > 0$. From Proposition 8, this implies that higher enjoyment for surprise induces preferences that are monotone with respect to the stochastic orders induced by smooth functions whose derivatives are positive. For example, as formalized in Menezes, Geiss, and Tressler [1980], aversion to downside risk, that is *prudence*, is equivalent to preserving the order $\gtrsim_{\mathcal{W}_3^+}$ induced by the smooth functions with positive third derivative $\mathcal{W}_3^+$, which is the case whenever λ is high.[29] As an example, suppose $v(x) = 1 - \exp(-ax)/a$ for $a > 0$. If there is no preference for surprise, the agent has standard CARA EU preferences. As λ increases, the sign of the even derivatives of the local expected utilities switches from negative to positive, while the signs of the odd derivatives remain positive, so the agent shifts from risk averse to risk loving, and their prudence increases.[30]

Next, we compare the risk attitudes of an adversarial forecaster utility $V(F)$ and the expected utility $v(x) = V(\delta_x)$ in the case $X = [0, 1]$. To do this we recall the notion of relative risk attitudes for non expected utility preferences introduced in Chew, Karni, and Safra [1987]. Given an expected utility function $v_0 \in C(X)$, F is a *simple compensated spread* of $\tilde{F}$ with respect to v_0 if

$$\int_0^1 \phi(v_0(x))dF(x) \geqslant \int_0^1 \phi(v_0(x))d\tilde{F}(x)$$

for all convex and continuous functions $\phi : v_0(X) \to \mathbb{R}$. Observe that this implies that $\int_0^1 v_0(x)dF(x) = \int_0^1 v(x)d\tilde{F}(x)$, that is, the expected utility preference given by v_0 is indifferent between F and $\tilde{F}$.

Definition 6. A continuous utility V is *relatively more risk loving* than a continuous expected utility v_0 if $V(F) \geqslant V(\tilde{F})$ whenever F is a simple compensated spread of $\tilde{F}$ with respect to v_0.

Proposition 9. *Fix an adversarial forecaster utility V with representation v and σ and a continuous expected utility v_0. Then V is relatively more risk loving than v_0 if and only if for all $F \in \mathcal{F}$ there exists a continuous and convex function $\phi_F : v_0(X) \to$*

[29] A sufficient condition for all the local expected utilities to have strictly positive n-th derivative is that $\lambda > \tilde{v}^{(n)} exp(1)$, where $\tilde{v}^{(n)} = \max_{x \in X} |v^{(n)}(x)|$.

[30] In Online Appendix IV.C, we use this CARA example to analyze the effect of preference for surprise on risk-aversion of order $n > 3$.

$\mathbb{R}$ *such that*

$$v(x) + \sigma(x, \hat{y}(F)) = \phi_F(v_0(x))$$

Corollary 6. *Fix an adversarial forecaster utility V with representation v and σ, V is relatively more risk loving than its baseline utility v if and only if for all $F \in \mathcal{F}$ there exists a continuous and convex function $\phi_F : v(X) \to \mathbb{R}$ such that*

$$\sigma(x, \hat{y}(F)) = \phi_F(v(x)) \tag{20}$$

A sufficient condition for the condition in this corollary is that the baseline utility v is strictly increasing and concave and that the surprise function is increasing and convex in x.[31] The next example applies the Corollary without requiring that σ is monotone.

Example 8. Consider a utility-based version of the one-moment GMM preferences:

$$V(F) = \int v(x)dF(x) + \lambda \min_{y \in Y} \int (v(x) - y)^2 dF(x)$$

where $Y \equiv v(X)$ and $\lambda \geqslant 0$. In this case, the (unique) relevant moment coincides with the baseline utility v, that is, the adversarial forecaster tries to predict the realized utility of the agent. In this case, we have

$$\sigma(x, \hat{y}(F)) = \lambda \left(v(x) - \int v(\tilde{x})dF(\tilde{x}) \right)^2$$

which satisfies (20). Therefore for every $\lambda > 0$ the adversarial forecaster utility V is relatively more risk-loving than the baseline expected utility v, regardless of the risk attitudes of the latter. △

5.3 Higher-order Risk aversion and surprise

Eeckhoudt and Schlesinger [2006] formalize the idea that an agent is averse to higher-order risks through the comparison of pairs of lotteries that only differ for their n-th order risk. If at any wealth level the agent prefers the lottery with less n-th order

[31] To see this in general, observe that when v is strictly increasing and concave then v^{-1} is strictly increasing and convex. If in addition σ is increasing and convex in x, then we can rewrite $\sigma(x, y) = \sigma(v^{-1}(v(x)), y)$ implying that the condition in 20 is satisfied by the continuous and convex function $\sigma_F(t) = \sigma(v^{-1}(t)\hat{y}(F))$.

risk, they say the preferences exhibit *risk apportionment* of order n. In our setting with general continuous preferences, a sufficient condition for risk apportionment of order n is monotonicity with respect to the n-th order stochastic dominance relation $\gtrsim_{\mathcal{W}_{SD_n}}$ where

$$\mathcal{W}_{SD_n} = \left\{u \in C^n(X) : \forall m \leqslant n, sgn(u^{(m)}) = (-1)^{m-1}\right\}.$$

Agents with risk apportionment of order n for all n are called *mixed risk averse*. Most participants in the experiment of Deck and Schlesinger [2014], make choices that are consistent with mixed risk aversion (at their current wealth levels), but almost 20% make risk-loving choices. These participants are mixed risk loving, which means they are consistent with risk apportionment of order for odd n but not even n.

As an example, suppose $v(x) = 1 - \exp(-ax)/a$ for $a > 0$. If there is no preference for surprise, that is $\lambda = 0$, the agent is mixed risk averse, as most of the risk-averse subjects in Deck and Schlesinger [2014]. However, as λ increases the sign of the even derivatives of the local expected utilities switches from negative to positive, while the sign of the odd derivatives remains positive, so the agent shifts from mixed risk averse to mixed risk loving. Moreover, if $a > 1$, then higher-order derivatives will be more affected by an increased taste for surprise, while the opposite is true if $a < 1$.

5.4 Repeated choices and correlation aversion

When the space of outcomes is multidimensional, our model also covers the case where the adversary can observe the realization of one dimension before choosing their action. Consider $X = X_0 \times X_1$ where X_0 is finite and X_1 is an arbitrary compact subset of Euclidean space. Assume that the adversary takes two actions $(y_0, y_1) \in Y = Y_0 \times Y_1$, where the adversary takes the first action y_0 with no additional information about F, and then takes the second action after observing the realization of x_0. Assume that both Y_0 and Y_1 are compact subsets of Euclidean space. Here the set of strategies of the adversary is $Y = Y_0 \times Y_1^{X_0}$, which is compact. Moreover, assume that. for every F, the adversary has a unique optimal strategy, as prescribed by the definition of our adversarial forecaster model.

These preferences capture the idea of aversion to correlation between x_0 and x_1, which is well documented in experiments (see for example Andersen et al. [2018]). Intuitively, the agent would tend to avoid lotteries with a high correlation between x_0

and x_1, since this means the adversary is well informed about the residual distribution of x_1 when choosing y_1. The next example formalizes this using Proposition 8.

Example 9. Let $X_0 = \{0,1\}$, $X_1 = [0,1]$, $v(x_0, x_1) = v_0(x_0) + v_1(x_1)$, and assume that the adversary tries to minimize mean squared error, so $\sigma_0(x_0, F_0) = \left(x_0 - \int \tilde{x_0} dF_0(\tilde{x_0})\right)^2$ and $\sigma_1(x_1, F_1|x_0) = \left(x_1 - \int \tilde{x_1} dF_1(\tilde{x_1}|x_0)\right)^2$, where F_0 and $F_1(\cdot|x_0)$ respectively denote the marginal and the conditional distributions of F. Then $\sigma(x_0, x_1, F) = \sigma_0(x_0, F_0) + \sigma_1(x_1, F_1|x_0)$, so the local expected utility is $w(x_0, x_1, F) = v(x_0) + v(x_1) + \sigma(x_0, x_1, F)$. We model the agent's preference for correlation between x_0 and x_1 through the monotonicity properties of their preference with respect to the supermodular and submodular order. Intuitively, preferences that preserve the supermodular order favor lotteries with high positive correlation between x_0 and x_1 because their local expected utilities are supermodular, and vice versa for the submodular order. Following Shaked and Shanthikumar [2007] (Section 9.A.4), F dominates G in the submodular (resp. supermodular) order if $F \gtrsim G$ whenever $\int w(x) dF(x) \geqslant \int w(x) dG(x)$ for all functions $w \in C(X)$ that are differentiable in x_1 and such that $\frac{\partial}{\partial x_1} w(1, x_1) - \frac{\partial}{\partial x_1} w(0, x_1) \leqslant 0$ (resp. $\geqslant$). Therefore, the submodular and supermodular order are examples of stochastic order introduced in Definition 5, where the relevant sets of functions are those ones that satisfy the partial derivative condition above. For every F, the corresponding partial derivatives for the local utility at F are

$$\frac{\partial}{\partial x_1} w(1, x_1, F) - \frac{\partial}{\partial x_1} w(0, x_1, F) = -2 \left(\int \tilde{x_1} dF_1(\tilde{x_1}|1) - \int \tilde{x_1} dF_1(\tilde{x_1}|0) \right).$$

Thus by Proposition 8, the agent's preference preserves the submodular order for all F such that $\int \tilde{x_1} dF_1(\tilde{x_1}|1) > \int \tilde{x_1} dF_1(\tilde{x_1}|0)$, and at each such lottery they would be better off by decreasing the amount of positive correlation between x_0 and x_1. By similar reasoning, the agent would prefer to decrease the amount of negative correlation between x_0 and x_1 at each lottery F such that $\int \tilde{x_1} dF_1(\tilde{x_1}|1) < \int \tilde{x_1} dF_1(\tilde{x_1}|0)$.[32] Combining these facts, we see that the agent has the highest utility with distributions such that $\int \tilde{x_1} dF_1(\tilde{x_1}|1) = \int \tilde{x_1} dF_1(\tilde{x_1}|0)$, so that the best conditional forecast is independent of x_0. △

We leave a more detailed analysis of correlation aversion under the adversarial

[32] This last claim follows from the fact that the preference of the agent preserves the supermodular order over such lotteries.

expected utility model for future research.[33]

6 Conclusion

Adversarial forecaster preferences arise naturally in many settings. It allows the interpretation of random choice as a preference for surprise, and also allows sharp characterizations of the optimal "amount" (i.e., support size) of randomization and of various monotonicity properties.

In ongoing work we consider a more general "adversarial expected utility representation" that inherits many of the optimality and monotonicity properties of the adversarial forecaster representation, but does not require continuous local utility. This allows us to consider cases where the adversary has only finitely many actions or where the loss function has kinks, as in Example the absolute-deviation loss example mentioned right after Proposition 2.

In addition to the lottery-choice setting of this paper, the adversarial expected utility representation can also be applied to settings where the agent first chooses a distribution of qualities or outcomes and then chooses an allocation rule or an information-revelation policy.

Appendix A: Sections 2 and 3

Here we prove the main results in Sections 2 and 3. The omitted proofs from these and all the other sections are in Online Appendix II.A. The proofs of the ancillary results that are first stated in this section are in Online Appendix II.B.

Lemma 4. *If V has a continuous local expected utility $w(x, F)$, then*

$$\int w(x,F)d\tilde{F}(x) - \int w(x,F)dF(x) = \lim_{\lambda\downarrow 0}\frac{V((1-\lambda)F+\lambda\tilde{F})-V(F)}{\lambda}.$$

for all $F, \tilde{F} \in \mathcal{F}$.

[33] Stanca [2021] analyzes correlation aversion under uncertainty as opposed to risk.

Proof. Fix F and $\tilde{F}$, and for $0 < \lambda \leqslant 1$ and $\overline{F} = (1-\lambda)F + \lambda\tilde{F}$ define

$$\Delta(\lambda) = \frac{V(\overline{F}) - V(F)}{\lambda}.$$

Since $w(x, F)$ is a local expected utility function at F, $\int w(x,F)d\overline{F}(x) - V(F) \geqslant V(\overline{F}) - V(F)$ so

$$\Delta(\lambda) = \frac{V(\overline{F}) - V(F)}{\lambda} \leqslant \frac{\int w(x,F)d\overline{F}(x) - V(F)}{\lambda} = \int w(x,F)d\tilde{F}(x) - \int w(x,F)dF(x).$$

On the other hand since $w(x, \overline{F})$ is a local utility function at $\overline{F}$, $\int w(x,\overline{F})dF(x) - V(\overline{F}) \geqslant V(F) - V(\overline{F})$ so

$$\Delta(\lambda) = \frac{V(\overline{F}) - V(F)}{\lambda} \geqslant \frac{V(\overline{F}) - \int w(x,\overline{F})dF(x)}{\lambda}$$

$$= \frac{\int w(x,\overline{F})\left(d\overline{F}(x) - dF(x)\right)}{\lambda} = \int w(x,\overline{F})d\tilde{F}(x) - \int w(x,\overline{F})dF(x)$$

$$\to \int w(x,F)d\tilde{F}(x) - \int w(x,F)dF(x)$$

since $w(x, \overline{F})$ is continuous in $\overline{F}$. Putting these together we have

$$\int w(x,F)d\tilde{F}(x) - \int w(x,F)dF(x) \leqslant \lim_{\lambda\downarrow 0} \Delta(\lambda) \leqslant \int w(x,F)d\tilde{F}(x) - \int w(x,F)dF(x)$$

which yields the statement. ■

Lemma 5. *Let V have continuous local expected utility w. For all $F, \tilde{F}, \overline{F} \in \mathcal{F}$ such that there exists $\mu > 0$ with $F + \mu(\tilde{F} - \overline{F}) \in \mathcal{F}$,*

$$DV(\tilde{F} - \overline{F}) := \int w(x,F)d\tilde{F}(x) - \int w(x,F)d\overline{F}(x) = \lim_{\lambda\downarrow 0} \frac{V(F + \lambda(\tilde{F} - \overline{F})) - V(F)}{\lambda}.$$

Proof. Choose $\mu > 0$ as in the statement and observe that

$$\begin{aligned}\lim_{\lambda\downarrow 0}\frac{V(F+\lambda(\tilde{F}-\overline{F}))-V(F)}{\lambda} &= \frac{1}{\mu}\lim_{\lambda\downarrow 0}\frac{V((1-\lambda/\mu)F+(\lambda/\mu)(F+\mu(\tilde{F}-\overline{F}))-V(F)}{\lambda/\mu}\\ &= \frac{1}{\mu}\left(\int w(x,F)dF(x)-\int w(x,F)d(F+\mu(\tilde{F}-\overline{F}))(x)\right)\\ &= \int w(x,F)d\tilde{F}(x)-\int w(x,F)d\overline{F}(x)\end{aligned}$$

where the second equality follows by Lemma 4. ■

We can now prove Proposition 1 and Theorem 1.

Proof of Proposition 1. Assume that V has continuous local expected utility $w(x, F)$. As argued in the main text, V is concave. Lemma 5 implies that $D(x, F) = w(x, F) + \int w(x, F)dF(x) = w(x, F) + V(F)$, where the second equality follows from the properties of $w(x, F)$. This implies that $D(x, F)$ is well defined and continuous. ■

Proof of Theorem 1. (If). Let v and σ correspond to the adversarial forecaster representation of V. The map $w : \mathcal{F} \to C(X)$ given by $w(x, F) = v(x) + \sigma(x, \hat{y}(F))$ is a continuous local utility of $V(F) = \min_{\tilde{F}\in\mathcal{F}} \int w(x, \tilde{F})dF(x)$, so that V has continuous local expected utility.

(Only if). Let $w(x, F)$ denote the continuous local expected utility of V, and define $Y = \{w(\cdot, F)\}_{F\in\mathcal{F}} \subseteq C(X)$. Since $X, \mathcal{F}$ are compact and w is continuous, it follows that Y is closed, bounded, and equicontinuous, so it is compact. For all $y = w(\cdot, F)$ and $x \in X$, define $u(x, y) = w(x, F)$ and observe that it is continuous. For all $F \in \mathcal{F}$ and for all $\tilde{y} \in Y$,

$$V(F) = \int w(x, F)dF(x) \leqslant \int u(x, \tilde{y})dF(x),$$

where both the equality and the inequality follow because $w(\cdot, F)$ is a local expected utility of V at F and the definition of Y. This implies that $V(F) = \min_{y\in Y} \int u(x, y)dF(x)$.

It remains to show that $\int u(x, y)dF(x)$ has a unique minimum over y. Suppose that for some F there is a $\tilde{F} \neq F$ such that $V(F) = \int w(x, \tilde{F})dF(x)$. For every $\lambda \in (0, 1)$, define $F_\lambda = \lambda\tilde{F} + (1-\lambda)F$. Then because V is concave and the w are local expected utility functions, for all $\lambda \in [0,1]$

$$\lambda V(\tilde{F}) + (1-\lambda)V(F) \leqslant V(F_\lambda) \leqslant \lambda \int w(x,\tilde{F})d\tilde{F}(x) + (1-\lambda)\int w(x,\tilde{F})dF(x)$$
$$= \lambda V(\tilde{F}) + (1-\lambda)V(F),$$

so that

$$V(F_\lambda) = \int w(x,\tilde{F})dF_\lambda(x) \tag{21}$$

Next, fix $\mu \in (0,1)$. By the properties of w, $V(\tilde{F}) \leqslant \int w(x,F_\mu)d\tilde{F}(x)$, so

$$\lambda V(\tilde{F}) + (1-\mu)V(F) = V(F_\mu) = \int w(x,F_\mu)dF_\mu(x)$$
$$= \mu \int w(x,F_\mu)d\tilde{F}(x) + (1-\mu)\int w(x,F_\mu)dF(x)$$

so that, by rearranging the terms,

$$V(\tilde{F}) = \int w(x,F_\mu)d\tilde{F}(x) + \frac{(1-\mu)}{\mu}\left(\int w(x,F_\mu)dF(x) - V(F)\right) \geqslant \int w(x,F_\mu)d\tilde{F}(x)$$

where the last inequality follows because $\mu \in (0,1)$ and $\int w(x,F_\mu)dF(x) \geqslant V(F)$. With this,

$$V(\tilde{F}) = \int w(x,F_\mu)d\tilde{F}(x). \tag{22}$$

Fix $\tilde{x} \in X$. Since $\mu > 0$, there exists $\lambda \in (0,\mu)$ such that $F_\mu + \lambda(\delta_{\tilde{x}} - \tilde{F}) \in \mathcal{F}$. Therefore,

$$w(\tilde{x},F_\mu) - V(\tilde{F}) = w(\tilde{x},F_\mu) - \int w(x,F_\mu)d\tilde{F}(x) = \lim_{\lambda\downarrow 0}\frac{V(F_\mu + \lambda(\delta_{\tilde{x}} - \tilde{F})) - V(F_\mu)}{\lambda}$$
$$\leqslant \lim_{\lambda\downarrow 0}\frac{\int w(x,\tilde{F})d\left(F_\mu + \lambda(\delta_{\tilde{x}} - \tilde{F})\right)(x) - V(F_\mu)}{\lambda}$$
$$= \int w(x,\tilde{F})d\left(\delta_{\tilde{x}} - \tilde{F}\right)(x) = w(\tilde{x},\tilde{F}) - V(\tilde{F}),$$

where the first equality follows by (22), the second equality by Lemma 5, the inequality by the properties of w, the third equality by (21), and the last equality by the properties of w again. This implies that $w(\tilde{x}, F_\mu) \leqslant w(\tilde{x}, \tilde{F})$. Similarly,

$$\begin{aligned} w(\tilde{x}, \tilde{F}) - V(\tilde{F}) = w(\tilde{x}, \tilde{F}) - \int w(x, \tilde{F}) d\tilde{F}(x) &= \lim_{\lambda \downarrow 0} \frac{V(\tilde{F} + \lambda(\delta_{\tilde{x}} - \tilde{F})) - V(\tilde{F})}{\lambda} \\ &\leqslant \lim_{\lambda \downarrow 0} \frac{\int w(x, F_\mu) d\left(\tilde{F} + \lambda(\delta_{\tilde{x}} - \tilde{F})\right)(x) - V(\tilde{F})}{\lambda} \\ &= \int w(x, F_\mu) d\left(\delta_{\tilde{x}} - \tilde{F}\right)(x) = w(\tilde{x}, F_\mu) - V(\tilde{F}), \end{aligned}$$

where the first equality follows by the properties of w, the second equality follows by Lemma 5, the inequality by the properties of w, and the third and the last equality by (22). This implies that $w(\tilde{x}, \tilde{F}) \leqslant w(\tilde{x}, F_\mu)$, so $w(\tilde{x}, F_\mu) = w(\tilde{x}, \tilde{F})$. Since this is true for all $\mu > 0$ and w is continuous it holds also in the limit: $w(\tilde{x}, F) = w(\tilde{x}, \tilde{F})$. Given that $\tilde{x}$ was arbitrary, the minimizer is unique, which proves that V is an adversarial expected utility representation that satisfies uniqueness. Now we show that if $\gtrsim$ has an adversarial expected utility representation that satisfies uniqueness, then it has an adversarial forecaster representation. Let Y and u denote the adversarial expected utility representation of $\gtrsim$. For all $F \in \mathcal{F}$, let $\hat{y}(F) \in Y$ denote the unique minimizer of $\int u(x, \tilde{y}) dF(x)$. Define $v(x) = \min_{y \in Y} u(x, y)$, $\sigma(x, F) = u(x, y(F)) - v(x)$, and $V(F) = \int v(x) dF(x) + \int \sigma(x, F) dF(x)$. Observe that, by construction $V(F) = \min_{y \in Y} \int u(x, y) dF(x)$, hence V represents $\gtrsim$. Finally, fix $F, \tilde{F} \in \mathcal{F}$ and observe that

$$\begin{aligned} \int \sigma(x, F) dF(x) &= \int u(x, y(F)) dF(x) - \int v(x) dF(x) \\ &\leqslant \int u(x, y(\tilde{F})) dF(x) - \int v(x) dF(x) = \int \sigma(x, \tilde{F}) dF(x) \end{aligned}$$

showing that σ is a forecast error. ■

Proof of Proposition 2. (If). This direction follows immediately from the discussion before the proposition.[34] (Only if). Fix an optimal lottery F^* for V over $\overline{\mathcal{F}}$ and assume that there exists $\hat{F}$ that is strictly better than F^* for an expected utility agent with utility $v + \sigma(\cdot, F^*)$. Due to convexity of $\overline{\mathcal{F}}$, F^* is also optimal when maximizing V over the lotteries in the segment between F^* and $\hat{F}$. This implies that

[34]See Propositions 10 in Online Appendix III.A and 11 in Online Appendix V for alternative proofs that can also be applied to the more general adversarial expected utility model.

the directional derivative of V at F^* in direction $\hat{F}$ is negative, which contradicts $\hat{F}$ strictly preferred to F^* for expected utility function $v + \sigma(\cdot, F)$. ■

Before proving Proposition 3 we introduce some additional notation. For every $F \in \mathcal{F}$, define $\xi_{\beta,F} : [0,1] \to \mathbb{R}$ as $\xi_{\beta,F}(\tilde{p}) = (1-\beta)g'(D_2(F))\,\tilde{p}^2 + \beta g(\tilde{p} - \tilde{p}^2)$ and let $cav(\xi_{\beta,F})$ denote its concavification.

Proof of Proposition 3. First, observe that Proposition 2 implies that that $F^* \in \text{argmax}_{F \in \overline{\mathcal{F}}} V_\beta(F)$ if and only if $F^* \in \text{argmax}_{F \in \overline{\mathcal{F}}} \int w_\beta(x, F^*) dF(x)$.

We now prove the first part of the statement. Let $\beta \in [0,1]$, fix an arbitrary optimal distribution F^* with marginals (p_F^*, F_Δ^*), and denote $q^* = \int p^2 dF_\Delta^*(p)$. Define

$$\Delta(p_F^*, q^*) = \left\{ F_\Delta \in \Delta[0,1]) : \int p^2 dF_\Delta(p) = p_F^*, \int p^2 dF_\Delta(p) = q^* \right\}.$$

Consider the maximization problem:

$$\max_{F_\Delta \in \Delta(p_F^*, q^*)} \int g(p - p^2) dF_\Delta(p). \tag{23}$$

If F_Δ is feasible, it yields a weakly higher utility than F_Δ^* because F_Δ has the same second moment as F_Δ^* and the latter is feasible for Problem 23, so any solution F_Δ of Problem 23 is also a solution of the original problem. Finally, observe that $\Delta(p_F^*, q^*)$ is a moment set with $k = 2$ moment conditions. The objective function of Problem 23 is linear in F_Δ, so it follows from Theorem 2.1. in Winkler [1988] that there is solution of Problem 23, and hence of the original problem, that is supported on no more than three points of $\Delta([0,1])$, concluding the proof of the first statement.

Next, assume that there exists an optimal $F^* \in \overline{\mathcal{F}}$ whose marginals are given by (p_F^*, F_Δ^*). By the initial claim and equation 5, (p_F^*, F_Δ^*) solve

$$\begin{aligned}
&\max_{p \in \overline{\Delta}, F_\Delta \in \Delta([0,1]) : \int \tilde{p} dF(\tilde{p}) = p} \left\{ p\tilde{v} + (1-\beta)g'(D_2(F^*)) \int (\tilde{p}^2 - p^2) dF_\Delta(\tilde{p}) + \beta \int g(\tilde{p} - \tilde{p}^2) dF_\Delta(p) \right\} \\
&= \max_{p \in \overline{\Delta}} \left\{ p\tilde{v} - (1-\beta)g'(D_2(F^*))\,p^2 + \max_{F_\Delta : \int \tilde{p} dF(\tilde{p}) = p} \left[\int (1-\beta)g'(D_2(F^*))\,\tilde{p}^2 + \beta g(\tilde{p} - \tilde{p}^2) dF_\Delta(\tilde{p}) \right] \right\} \qquad (24) \\
&= \max_{p \in \overline{\Delta}} \left\{ p\tilde{v} - (1-\beta)g'(D_2(F^*))\,p^2 + cav(\xi_{\beta,F^*})(p) \right\}
\end{aligned}$$

Given the assumptions on g and given that $\overline{\Delta}$ is compact, there exist $\underline{\beta}, \overline{\beta} \in (0,1)$ with $\underline{\beta} \leqslant \overline{\beta}$ such that ξ_{β,F^*} is strictly concave over $\overline{\Delta}$ for all $\beta \geqslant \overline{\beta}$ and ξ_{β,F^*} is strictly convex over $\overline{\Delta}$ for all $\beta \leqslant \underline{\beta}$. We now prove points 1 and 2.

1. When $\beta \geqslant \overline{\beta}$, ξ_{β,F^*} is strictly concave so that $cav(\xi_{\beta,F^*}) = \xi_{\beta,F^*}$. By Corollary 2 in Kamenica and Gentzkow [2011], the inner maximization problem in equation 24 is uniquely solved by $F_\Delta = \delta_p$, that is, no disclosure is uniquely optimal. This implies that $F^*_\Delta = \delta_{p^*_F}$. Because $p\tilde{v} - (1-\beta)g'\left(D_2(F^*)\right)p^2 + \xi_{\beta,F^*}(p) = p\tilde{v} + \beta g(p - p^2)$ and the optimal (p^*_F, F^*_Δ) are arbitrary, the statement follows.

2. When $\beta \leqslant \underline{\beta}$, ξ_{β,F^*} is strictly convex. By Corollary 2 in Kamenica and Gentzkow [2011], the inner maximization problem in equation 24 is uniquely solved by $F_\Delta = (1-p)\delta_0 + p\delta_1$, that is, full disclosure is uniquely optimal, and $cav(\xi_{\beta,F^*})(\tilde{p}) = (1-\beta)g'\left(D_2(F^*)\right)\tilde{p}$. This implies that $F^*_\Delta = (1-p^*_F)\delta_0 + p^*_F\delta_1$. Next, $p\tilde{v} - (1-\beta)g'\left(D_2(F^*)\right)p^2 + cav(\xi_{\beta,F^*})(p) = p\tilde{v} + (1-\beta)g'(D_2(F^*))(p - p^2)$. Given that the optimal (p^*_F, F^*_Δ) are arbitrary, the statement follows. ∎

Proof of Proposition 6. This follows from the following three lemmas. The first two are standard and are proved in Online Appendix II.A.

Lemma 6. *$\sigma(x, F)$ defined by a method of moments forecast is a forecast error.*

Given $F, \tilde{F} \in \mathcal{F}$, the direction $\tilde{F} - \overline{F}$ is *relevant* at F if for some $\lambda > 0$ the signed measure $F + \lambda(\tilde{F} - \overline{F}) \geqslant 0$ is an ordinary measure.

Lemma 7. *Let $H(x, \tilde{x}) = \int h(x,s)h(\tilde{x},s)d\mu(s)$. Then*

$$V(F) = \int H(x,x)dF(x) - \int\int H(x,\tilde{x})dF(x)dF(\tilde{x})$$

with directional derivatives for relevant directions $(\delta_z - F)$ at F given by

$$DV(F)(\delta_z - F) =$$

$$H(z,z) - \int H(x,x)dF(x) - 2\left[\left[\int H(z,x)dF(x)\right) - \int H(x,\tilde{x})dF(x)dF(\tilde{x})\right].$$

We now allow the set S to be any compact metric space. When $F \mapsto h(F,\cdot)$ is one-to-one we have an additional property:

Lemma 8. *If $F \mapsto h(F,\cdot)$ is one-to-one and μ assigns positive probability to open sets of S then $V(F)$ is strictly concave.*

Proof. From Lemma 7 it suffices to prove that the positive semi-definite quadratic form $\int\int H(x,\tilde{x})dM(x)dM(\tilde{x})$ is positive definite on the linear subspace of signed measures where $\int dM(x) = 0$. Recall that $H(x,\tilde{x}) = \int h(x,s)h(\tilde{x},s)d\mu(s)$, and suppose that $\int h(x,\hat{s})dM(x) \neq 0$ for some $\hat{s}$. Since h is continuous there is an open set $\tilde{S} \subseteq S$ such that $\hat{s} \in \tilde{S}$ and $\int h(x,s)dM(x) \neq 0$ for all $s \in \tilde{S}$. Since μ assigns positive probability to open sets of S this implies that

$$\int\int H(x,\tilde{x})dM(x)dM(\tilde{x}) = \int\left[\left(\int h(x,s)dM(x)\right)\int h(\tilde{x},s)dM(\tilde{x})\right]\mu(s)ds > 0.$$

Hence it suffices for $V(F)$ to be strictly convex that $\int h(x,s)dM(x) \neq 0$ for any signed measure M with $\int dM(x) = 0$. Using the Jordan decomposition we may write $M = \lambda(F - \tilde{F})$ where $F, \tilde{F}$ are probability measures and $\lambda > 0$ if $M \neq 0$. Hence $\int h(x,s)dM(x) = 0$ for $M \neq 0$ if and only if for all s

$$h(F,s) = \int h(x,s)dF(x) = \int h(x,s)d\tilde{F}(x) = h_{\tilde{F}}(s).$$

Since $h \to h(F,\cdot)$ is one-to-one this implies $F = \tilde{F}$ and $M = 0$. ■

To prove Theorem 2 we use a sequence of intermediate results. To begin, we fix an arbitrary parametric adversarial forecaster representation V, and define $u(x,y) = v(x) + \hat{\sigma}(x,y)$. Let $\mathcal{H}$ denote the set of probability measures over Y.

For any convex and compact subset $\overline{\mathcal{F}} \subseteq \mathcal{F}$ of lotteries, let $ext(\overline{\mathcal{F}})$ denote the set of extreme points of $\overline{\mathcal{F}}$. By Choquet's theorem, for all $F \in \overline{\mathcal{F}}$, there exists $\lambda \in \Delta\left(ext\left(\overline{\mathcal{F}}\right)\right)$ such that $F = \int \tilde{F} d\lambda(\tilde{F})$. Let $\Lambda_F \subseteq \Delta\left(ext\left(\overline{\mathcal{F}}\right)\right)$ be the set of probability measures over extreme points that satisfy $F = \int \tilde{F} d\lambda\left(\tilde{F}\right)$ for F.

Theorem 6. *Fix* $\hat{H} \in \arg\min_{H\in\mathcal{H}} \max_{F\in ext(\overline{\mathcal{F}})} \int\int u(x,y)\, dF(x)\, dH(y)$. *Then* $\hat{F} \in \arg\max_{F\in\overline{\mathcal{F}}} V(F)$ *if and only if for all* $\tilde{F} \in ext\left(\overline{\mathcal{F}}\right)$, $V(\hat{F}) \geqslant \int\int u(x,y)\, d\tilde{F}(x)\, d\hat{H}(y)$, *and, for all* $\tilde{F} \in \bigcup_{\lambda\in\Lambda_{\hat{F}}} \operatorname{supp}\lambda$, $V(\hat{F}) = \int\int u(x,y)\, d\tilde{F}(x)\, d\hat{H}(y)$.

Note that when $\overline{\mathcal{F}} = \Delta\left(\bar{X}\right)$ for some closed subset $\bar{X}$, the extreme points $ext\left(\overline{\mathcal{F}}\right) = \bar{X}$ are simply point masses over the set of feasible outcomes. In this case, Theorem 6 implies that F is optimal if and only if $V(F) \geqslant \int u(x,y)d\hat{H}(y)$ for all $x \in \overline{X}$, with equality for $x \in \operatorname{supp} F$.

Now we fix a closed subset $\overline{X} \subseteq X$ and a finite collection of functions $\Gamma = \{g_1, ..., g_k\} \subset C(\overline{X})$. As in the main text, we consider $\mathcal{F}_\Gamma(\overline{X}) \subseteq \mathcal{F}$. By Theorem 2.1

in Winkler [1988], $\tilde{F} \in ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$ if and only if $\tilde{F} \in \mathcal{F}_\Gamma(\overline{X})$ and $\tilde{F} = \sum_{i=1}^p \alpha_i \delta_{x_i}$ for some $p \leqslant k+1$, $\alpha \in \Delta(\{1,...,p\})$, and $\{x_1,...,x_p\} \subseteq \bar{X}$ such that the vectors $\{(g_1(x_i),...,g_k(x_i),1)\}_{i=1}^p$ are linearly independent. For every finite subset of extreme points $\mathcal{E} \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$, define $\overline{X}_\mathcal{E} = \bigcup_{\tilde{F}\in\mathcal{E}} \operatorname{supp} \tilde{F} \subseteq \overline{X}$, which is finite from Winkler's theorem. We identify $co(\mathcal{E})$ with the subset of $\mathcal{F}_\Gamma(\overline{X})$ composed of all convex combinations of extreme points in $\mathcal{E}$. Recall that $\hat{Y}(F) \equiv \operatorname{argmin}_{y\in Y} \int u(x,y)dF(x)$, and that (Y,u) satisfies the uniqueness property if $\hat{Y}(F)$ is a singleton for all $F \in \mathcal{F}$.

Theorem 7. *Fix a finite set $\mathcal{E} \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$, and suppose that Y has the structure of an m-dimensional manifold with boundary, that u is continuously differentiable in y, and that Y and u satisfy the uniqueness property. We have:*

1. *For an open dense full measure set of $w \in \mathcal{W} \subseteq \mathbb{R}^{\overline{X}_\mathcal{E}}$, every lottery F that solves $\max_{\tilde{F}\in co(\mathcal{E})} \min_{y\in Y} \int (u(x,y) + w(x))d\tilde{F}(x)$ has finite support on no more than $(k+1)(m+1)$ points of $\overline{X}_\mathcal{E}$.*

2. *There exists a lottery F that solves $\max_{\tilde{F}\in co(\mathcal{E})} \min_{y\in Y} \int u(x,y)d\tilde{F}(x)$ and has finite support on no more than $(k+1)(m+1)$ points of $\overline{X}_\mathcal{E}$.*

Proof. Let $|\mathcal{E}| = n$ and $|\overline{X}_\mathcal{E}| = r \leqslant n(k+1)$. Because $|\operatorname{supp} \tilde{F}| \leqslant k+1$ for every $\tilde{F} \in ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$, both statements are trivial if $(m+1) \geqslant n$. For $(m+1) < n$, for every $w \in \mathbb{R}^{\overline{X}_\mathcal{E}}$, define $u_w(x,y) = u(x,y) + w(x)$ and $V_w(F) = \min_{y\in Y} \int u_w(x,y)dF(x)$, and fix $H_w \in \arg\min_{H\in\mathcal{H}} \max_{F\in\mathcal{E}} \int\int u_w(x,y)\, dF(x)\, dH(y)$. For every $w \in \mathbb{R}^{\overline{X}_\mathcal{E}}$, the uniqueness property implies that $H_w = \hat{y}(F_w) \in Y$ for some $F_w \in \operatorname{argmax}_{F\in co(\mathcal{E})} V_w(F)$, and the expectation of each w with respect to each $F \in co(\mathcal{E})$ is well defined since $\operatorname{supp} F \subseteq \overline{X}_\mathcal{E}$ by construction.

We first prove point 1. Fix an arbitrary subset of $m+2$ extreme points $\overline{\mathcal{E}} = \left\{\tilde{F}_1,...,\tilde{F}_{m+2}\right\} \subseteq \mathcal{E}$ and consider the map $U_{\overline{\mathcal{E}}} : Y \times \mathbb{R} \times \mathbb{R}^{\overline{X}_\mathcal{E}} \to \mathbb{R}^{m+2}$ defined by

$$U_{\overline{\mathcal{E}}}(y,v,w)_\ell = u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell) \qquad \forall \ell \in \{1,...,m+2\}$$

where, for every $y \in Y$, $u(\tilde{F}_\ell, y) = \int u(x,y)d\tilde{F}_\ell(x)$ and $w(\tilde{F}_\ell) = \int w(x)d\tilde{F}_\ell(x)$. For every $(y,v) \in Y \times \mathbb{R}$, the derivative of $U_{\overline{\mathcal{E}}}$ with respect to $w \in \mathbb{R}^{\overline{X}_\mathcal{E}}$ is a $(m+2) \times r$ matrix whose ℓ-th row coincides with the probability vector $\tilde{F}_\ell$, and because the $\left\{\tilde{F}_1,...,\tilde{F}_{m+2}\right\}$ are extreme points of $\mathcal{F}_\Gamma(\overline{X})$, this matrix has full rank, so the total derivative of $U_{\overline{\mathcal{E}}}$ has full rank as well. Hence by the parametric transverality theorem,[35] for an open dense full measure subset of $\mathbb{R}^{\overline{X}_{\mathcal{E}}}$, denoted $\mathcal{W}(\overline{\mathcal{E}})$, the manifold $(y, v) \mapsto u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell)$ intersects zero transversally. Since $\dim(Y \times \mathbb{R}) < m+2$, there is no (y, v) that solve $u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell) = 0$ for all $\ell \leqslant m + 2$. And since $\mathcal{E}$ has finitely many subsets $\overline{\mathcal{E}}$ of $m + 2$ extreme points, the intersection $\mathcal{W} = \bigcap_{\overline{\mathcal{E}}} \mathcal{W}(\overline{\mathcal{E}})$ is open, dense, and of full measure, since it is the finite intersection of full-measure sets. Thus, for $w \in \mathcal{W}$ and for all $y \in Y$ and $v \in \mathbb{R}$, $u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell) = 0$ for at most $m + 1$ extreme points in $\mathcal{E}$.

Next, fix $w \in \mathcal{W}$, $F^* \in \operatorname{argmax}_{F \in co(\mathcal{E})} V_w$, and $\lambda \in \Lambda_{F^*}$. By Theorem 6, for all $\tilde{F} \in \operatorname{supp} \lambda \subseteq \mathcal{E}$, $u(\tilde{F}, H_w) - V_w(F^*) + w(\tilde{F}) = 0$. By the previous part of the proof and Theorem 6, we then have $|\operatorname{supp} \lambda| \leqslant m + 1$. Therefore, F_w is the linear combination of up to $m + 1$ extreme points in $\mathcal{E}$. Each $\tilde{F} \in \mathcal{E}$ is supported on up to $k + 1$ points of $\overline{X}_{\mathcal{E}}$, so F_w is supported on up to $(m + 1)(k + 1)$ points of $\overline{X}_{\mathcal{E}}$.

Now we prove point 2. Because $\mathcal{W}$ is dense in $\mathbb{R}^{\overline{X}_{\mathcal{E}}}$, there exists a sequence $w^n \in \mathcal{W}$ such that $w^n(x) \to 0$ for all $x \in \overline{X}_{\mathcal{E}}$, and a sequence of corresponding optimal lotteries F^n with support of no more than $(m + 1)(k + 1)$ points of $\overline{X}_{\mathcal{E}}$. Choose a convergent subsequence of $F^n \to F$, and observe that lotteries with no more than $(m+1)(k+1)$ points of support cannot converge weakly to a lottery with larger support. Finally, because V_w is continuous with respect to w, the Berge Maximum Theorem implies that F solves $\max_{F \in co(\mathcal{E})} V_0(F)$, concluding the proof. ■

Lemma 9. *Suppose that for every finite set $\mathcal{E} \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$ there exists a lottery $F_{\mathcal{E}}$ that solves $\max_{F \in co(\mathcal{E})} V(F)$ and has finite support on no more than $(m+1)(k+1)$ points of $\overline{X}$. Then there exists a lottery F^* that solves $\max_{F \in \mathcal{F}_\Gamma(\overline{X})} V(F)$ and that has finite support on no more than $(m + 1)(k + 1)$ points of $\overline{X}$.*

Proof of Theorem 2. Fix a parametric adversarial forecaster representation $(Y, v, \hat{\sigma})$, and define $u = v + \sigma$. By Definition 4, the adversarial expected utility representation (Y, u) is such that Y has the structure of an m-dimensional manifold with boundary, u is continuously differentiable in y, and Y and u satisfy the uniqueness property. By Theorem 7 and Lemma 9, there exists a solution F^* that is supported on no more than $(k + 1)(m + 1)$ points of $\overline{X}$. ■

[35] See e.g. Guillemin and Pollack [2010].

Proof of Proposition 7. Stationarity implies that $H(x,x)$ is constant, so the directional derivatives from Lemma 7 simplify to

$$DV(F)(\delta_z - F) = -2\left[\int H(z,x)dF(x)) - \int H(x,\tilde{x})dF(x)dF(\tilde{x})\right].$$

Since $V(F)$ is continuous and concave on a compact set the maximum exists, and is characterized by the condition that no directional derivative is positive, which is

$$\int H(z,x)dF(x) \geqslant \int H(x,\tilde{x})dF(x)dF(\tilde{x}) \text{ for all } z \in X. \tag{25}$$

This implies the complementary slackness condition: if there exists $z \in A$ such that z satisfies (25) with strict inequality, then $F(A) = 0$.[36]

Next we show that for any $0 < a \leqslant 1$ and interval $A = [0,a)$ there is $z \in A$ such that $\int H(z,x)dF(x)) = \int H(x,\tilde{x})dF(x)dF(\tilde{x})$. By continuity this implies $\int H(0,x)dF(x)) = \int H(x,\tilde{x})dF(x)dF(\tilde{x})$ and by symmetry $\int H(1,x)dF(x)) = \int H(x,\tilde{x})dF(x)dF(\tilde{x})$. Suppose instead that for all $z \in A$ $\int H(z,x)dF(x) > \int H(x,\tilde{x})dF(x)dF(\tilde{x})$, and take $a \in X$ to be the supremum of the set $\{x' \in X : \int H(x',x)dF(x) > \int H(x,\tilde{x})dF(x)dF(\tilde{x})\}$, so that $\int H(a,x)dF(x) = \int H(x,\tilde{x})dF(x)dF(\tilde{x})$. By complementary slackness $F(A) = 0$. Positive definiteness, that is $\int H(x,\tilde{x})dF(x)dF(\tilde{x}) > 0$, implies that $H(a,x) > 0$ for some non-trivial interval $x \in [a,b]$. Since $H(0,\tilde{x})$ is decreasing and $H(a,a) = \max_{\tilde{x}} H(a,\tilde{x})$, it follows that $H(a,x) > H(0,x)$. Hence $\int H(x,\tilde{x})dF(x)dF(\tilde{x}) = \int H(a,x)dF(x) > \int H(0,x)dF(x))$, violating the first order condition at $z = 0$.

Finally, suppose there is a non-trivial open interval $A = (a,b)$ such that $F(A) = 0$. We may assume w.l.o.g. that $\int H(a,x)dF(x) = \int H(x,\tilde{x})dF(x)dF(\tilde{x})$, $\int H(b,x)dF(x)) = \int H(x,\tilde{x})dF(x)dF(\tilde{x})$. Then for $x \notin A$ by strict convexity either $(1/2)(H(a,x) + H(b,x)) > H((a+b)/2,x)$ or both the left-hand side and the right-hand side are equal to zero. The latter cannot be true for a positive measure set of $x \notin A$, so $\int H(x,\tilde{x})dF(x)dF(\tilde{x}) = (1/2)\left(\int H(a,x)dF(x) + \int H(b,x)dF(x)\right) > \int H((a+b)/2,x)dF(x))$ violating the first order condition at $(a+b)/2$. ∎

[36] If there is $z \in A$ with $F(A) > 0$, then there is an open set $\tilde{A} \subseteq A$ containing z with $F(\tilde{A}) > 0$, and every $x \in \tilde{A}$ satisfies (25) with strict inequality. Then $\int\int H(x,\tilde{x})dF(x)dF(\tilde{x}) = \int_{\tilde{A}}\int_X H(x,\tilde{x})dF(\tilde{x})dF(x) + \int_{\tilde{A}^c}\int_X H(x,\tilde{x})dF(\tilde{x})dF(x) > F(\tilde{A})\int\int H(x,\tilde{x})dF(x)dF(\tilde{x}) + (1 - F(\tilde{A}))\int\int H(x,\tilde{x})dF(x)dF(\tilde{x}) = \int\int H(x,\tilde{x})dF(x)dF(\tilde{x})$, a contradiction.

Appendix B: Section 4

Proof of Lemma 1. The continuity of $\hat{\sigma}(x,y)$ follows from the uniform continuity of $\phi(\theta,x)$. Non-negativity of $\hat{\sigma}(x,y)$ is obvious. For given x to find y such that $\sigma(x,y)=0$ and $\int \exp(y(\xi))dU(\xi)=1$ choose

$$y(\xi) = -\max_{\theta\in X}\left(\phi(\theta,x)-\phi(\theta,\xi)\right) + \frac{1}{\log\left(\int \exp(-\max_{\theta\in X}\left(\phi(\theta,x)-\phi(\theta,\tilde{\xi})\right))dU(\tilde{\xi})\right)}.$$

By construction $\int \exp(y(\xi))dU(\xi)=1$ and $x\in\arg\max_{\xi} y(\xi)-\phi(\theta,\xi)$ so

$$\int\left(\left(\max_{\xi} y(\xi)-\phi(\theta,\xi)\right)-(y(x)-\phi(\theta,x))\right)dU(\theta)=0.$$

■

Proof of Lemma 2. The strongly ϕ-concave functions are K-Lipshitz so the family Y is equicontinuous. Together with the constraint $\int \exp(y(\xi))dU(\xi)=1$ this implies that Y is totally bounded. Hence any sequence $y^n\in Y$ has a subsequence that converges to some $y\in C(X)$. To show that Y is closed, let $y^{*n}\in Y^*$ be such that $y^n(x)=-\max_{\theta\in X}\left(y^{*n}(\theta)-\phi(\theta,x)\right)$. Since the sequence y^n is bounded and the sequence y^{*n} is equicontinuous, the sequence y^{*n} is also bounded. And because the sequence y^{*n} is K-Lipchitz, there is a subsequence $y^{*n}\to y^*$ that is also K-Lipchitz. Convergence and continuity imply that $y(x)=-\max_{\theta\in X}\left(y^*(\theta)-\phi(\theta,x)\right)$, that is, y is strongly ϕ-concave, so indeed Y is closed.

We next show that equation 11 has a exactly one solution in Y. We have

$$\inf_{y\in C(X)}\left\{\int\left(\max_{\xi}(y(\xi)-\phi(\theta,\xi))\right)-(y(x)-\phi(\theta,x))\,dU(\theta)dF(x)\right\}$$

$$=-\sup_{y\in C(X)}\left\{\int\left(-\max_{\xi}(y(\xi)-\phi(\theta,\xi))\right)dU(\theta)+\int y(x)dF(x)\right\}+\left[\int\phi(\theta,x)dU(\theta)dF(x)\right]$$

where the final term does not depend on y. Consider the alternative problem

$$\sup_{y^*\in C(X),y\in C(X)}\int y^*(\theta)dU(\theta)+\int y(x)dF(x) \tag{26}$$

subject to $y^*(\theta) - y(x) \leqslant \phi(\theta, x)$ for all (θ, x). It follows that $-y^*(\theta) \geqslant y(x) - \phi(\theta, x)$ implying $-y^*(\theta) \geqslant \max_\xi y(\xi) - \phi(\theta, \xi)$. This means that if the alternative problem has a solution y then the original problem has the same solution. The alternative problem is the extensively studied dual of the Kantorovitch transport problem and we draw upon results from that literature.

We say that $y \in C(X)$ is ϕ-concave if $y(x) = -\max_{\theta \in X} (y^*(\theta) - \phi(\theta, x))$ for some $y^* \in C(X)$. Proposition 1.11 in Santambrogio [2015] shows that because X is compact and ϕ is continuous, Problem 26 has a solution (y^*, y) where y is ϕ-concave with respect to y^* and $y^*(\theta) = -\max_{\xi \in X} (y(x) - \phi(\theta, \xi))$. This last step implies that y^* is K-Lipschitz and therefore that the solution is strongly ϕ-concave. Since the objective function is invariant to adding a constant to y and subtracting it from y^*, at least one such solution satisfies the normalization $\int \exp(y(\xi)) dU(\xi) = 1$.

Proposition 7.18 in Santambrogio [2015] shows that because ϕ is continuously differentiable, X is the closure of a bounded connected open set, and the uniform measure over θ has full support on X, all ϕ-concave solutions differ only by additive constants. Since strong ϕ-concavity implies ϕ-concavity the fact that the set Y is normalized shows that there is exactly one solution in Y. ∎

In what follows we often use the notation $\Delta(U, F)$, where for all $F \in \mathcal{F}$,

$$\Delta(U, F) = \{T \in \Delta(\Theta \times X) : marg_\Theta T = U, marg_X T = F)\}.$$

Proof of Theorem 5. By Theorem 4,

$$\max_{F \in \mathcal{F}} V(F) = \max_{T \in \Delta(\Theta \times X): marg_\Theta T = U} \int \left\{ v(x) - \phi(\theta, x) + \int \phi(\tilde{\theta}, x) dU(\tilde{\theta}) \right\} dT(\theta, x),$$

which immediately implies that $F \in \operatorname{argmax}_{\tilde{F} \in \mathcal{F}} V(F)$ if and only if there exists $T \in \Delta(\Theta \times X)$ with marginals given by U and F such that $T(G_\phi) = 1$, where $G_\phi = Gr(\Psi_\phi) \subseteq \Theta \times X$ is the graph of the correspondence Ψ_ϕ. In turn, this is equivalent to $0 \geqslant \inf_{T \in \Delta(U,F)} \{1 - T(G_\phi)\}$. Let G_ϕ^c denote the complement of G_ϕ. Theorem 1.27 in Villani [2021] gives

$$\inf_{T \in \Delta(U,F)} T(G_\phi^c) = \sup \left\{ F(A) - U(A^{G_\phi^c}) : A \subseteq X \quad \text{is closed} \right\},$$

where $A^{G^c_\phi} = \{\theta \in \Theta : \exists x \in A, (\theta, x) \in G_\phi\}$. Therefore, $F \in \operatorname{argmax}_{\tilde{F} \in \mathcal{F}} V(\tilde{F})$ is also equivalent to

$$0 \geqslant \sup \left\{ F(A) - U(A^{G^c_\phi}) : A \subseteq X \quad \text{is closed} \right\}$$

which in turn is equivalent to

$$U(\Psi^\ell_\phi(A)) \geqslant F(A) \tag{27}$$

for all closed $A \subseteq X$, where $\Psi^\ell_\phi(A) = \{\theta \in \Theta : \Psi_\phi(\theta) \cap A \neq \varnothing\}$ is the lower-inverse of the correspondence Ψ_ϕ evaluated at A. Also, observe that the class of closed sets $A \subseteq X$ is a π-class of the Borel sigma-algebra of X. Therefore, the inequality in (27) holds for all measurable sets $A \subseteq X$.

So far we have shown that

$$\operatorname*{argmax}_{F \in \mathcal{F}} V(F) = \left\{ F \in \mathcal{F} : F(A) \leqslant U(\Psi^\ell_\phi(A)) \quad \text{for all measurable} \quad A \right\}.$$

Finally, because U is atomless, Corollary 3.4 in Castaldo, Maccheroni, and Marinacci [2004] says that the right-hand side of the last equation is equal to the closure of $\{U \circ \psi^{-1} \in \mathcal{F} : \psi \in \Psi_\phi\}$, yielding the desired result. ∎

Proof of Lemma 3. By Theorem 4 we have

$$\begin{aligned}
\Sigma(F) &= \max_{T \in \Delta(U,F)} \left(\int \int \phi(\theta, x) dU(\theta) dF(x) - \int \phi(\theta, x) dT(\theta, x) \right) \\
&= \max_{T \in \Delta(U,F)} \left(\int \hat{\phi}(\theta, x) dT(\theta, x) \right)
\end{aligned}$$

where we defined $\hat{\phi}(\theta, x) = \int \phi(\tilde{\theta}, x) dU(\tilde{\theta}) - \phi(\theta, x)$. Given our assumption of ϕ_x, we have that $\hat{\phi}(\theta, x)$ is supermodular. Because of this and that U is atomless, Theorem 4.3 in Galichon [2018] can be directly applied to conclude that

$$\begin{aligned}
\max_{T \in \Delta(U,F)} \left(\int \hat{\phi}(\theta, x) dT(\theta, x) \right) &= \int \hat{\phi}(\theta, q_F(U(\theta))) dU(\theta) = \int_0^1 \hat{\phi}(q_U(t), q_F(t)) dt \\
&= \int \int \phi(\theta, x) dU(\theta) dF(x) - \int_0^1 \phi(q_U(t), q_F(t)) dt
\end{aligned}$$

where the second equality follows from the change of variable formula by setting $t = U(\theta)$, and the third equality follows from the definition of $\hat{\phi}(\theta, x)$. ■

References

Agranov, M. and P. Ortoleva (2017). "Stochastic choice and preferences for randomization." *Journal of Political Economy*, 125, 40–68.

Aliprantis, C. D. and K. C. Border (2006). *Infinite Dimensional Analysis*. Springer.

Cerreia-Vioglio, S. (2009). "Maxmin expected utility on a subjective state space: Convex preferences under risk."

Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva (2015). "Cautious expected utility and the certainty effect." *Econometrica*, 83, 693–728.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2019). "Deliberately stochastic." *American Economic Review*, 109, 2425–45.

Cerreia-Vioglio, S., F. Maccheroni, and M. Marinacci (2017). "Stochastic dominance analysis without the independence axiom." *Management Science*, 63, 1097–1109.

Chatterjee, K. and R. V. Krishna (2011). "A nonsmooth approach to nonexpected utility theory under risk." *Mathematical Social Sciences*, 62, 166–175.

Chew, S. H., L. G. Epstein, and U. Segal (1991). "Mixture symmetry and quadratic utility." *Econometrica*, 139–163.

Dillenberger, D. (2010). "Preferences for one-shot resolution of uncertainty and Allais-type behavior." *Econometrica*, 78, 1973–2004.

Doval, L. and V. Skreta (2018). "Constrained information design: Toolkit." *arXiv preprint arXiv:1811.03588*.

Ely, J., A. Frankel, and E. Kamenica (2015). "Suspense and surprise." *Journal of Political Economy*, 123, 215–260.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015). "Stochastic choice and revealed perturbed utility." *Econometrica*, 83, 2371–2409.

Fudenberg, D. and D. K. Levine (2006). "A dual-self model of impulse control." *American Economic Review*, 96, 1449–1476.

Galichon, A. (2018). *Optimal transport methods in economics*. Princeton University Press.

Green, J. R. and B. Jullien (1988). "Ordinal independence in nonlinear utility theory." *Journal of Risk and Uncertainty*, 1, 355–387.

Greene, W. H. (2003). *Econometric Analysis*. Pearson Education.

Guillemin, V. and A. Pollack (2010). *Differential topology.* Vol. 370. American Mathematical Soc.

Gul, F. and W. Pesendorfer (2001). “Temptation and self-control.” *Econometrica*, 69, 1403–1435.

Henry-Labordère, P. and N. Touzi (2016). “An explicit martingale version of the one-dimensional Brenier theorem.” *Finance and Stochastics*, 20, 635–668.

Hong, C. S. and W. S. Waller (1986). “Empirical tests of weighted utility theory.” *Journal of Mathematical Psychology*, 30, 55–72.

Levine, D. K., S. Modica, and J. Sun (2021). “Twin peaks: expressive voting and soccer hooliganism.”

Loseto, M. and A. Lucia (2021). “Deliberate randomization under risk.” *Available at SSRN 3811073.*

Maccheroni, F. (2002). “Maxmin under risk.” *Economic Theory*, 19, 823–831.

Machina, M. J. (1982). ““Expected utility” analysis without the independence axiom.” *Econometrica*, 277–323.

— (1984). “Temporal risk and the nature of induced preferences.” *Journal of Economic Theory*, 33, 199–231.

— (1985). “Stochastic choice functions generated from deterministic preferences over lotteries.” *The economic journal*, 95, 575–594.

Menezes, C., C. Geiss, and J. Tressler (1980). “Increasing downside risk.” *American Economic Review*, 70, 921–932.

Quiggin, J. (1982). “A theory of anticipated utility.” *Journal of Economic Behavior & Organization*, 3, 323–343.

Santambrogio, F. (2015). “Optimal transport for applied mathematicians.” *Birkäuser, NY*, 55, 94.

Sarver, T. (2018). “Dynamic mixture-averse preferences.” *Econometrica*, 86, 1347–1382.

Winkler, G. (1988). “Extreme points of moment sets.” *Mathematics of Operations Research*, 13, 581–587.

Yaari, M. E. (1987). “The dual theory of choice under risk.” *Econometrica*, 95–115.

Online Appendix I: Section 5

Proof of Proposition 8. In Proposition 12 in Online Appendix V, we show that the function V is Gâteaux differentiable with derivative given by the local utility $w(x, F)$ as in Proposition 1. Theorem 1 then implies that the local utility is given by $w(x, F) = v(x) + \sigma(x, \hat{y}(F))$ for every $F \in \mathcal{F}$. With this, exactly the same argument of Proposition 1 in Cerreia-Vioglio, Maccheroni, and Marinacci [2017] yields the desired result. ■

Proof of Proposition 9. In Proposition 12 in Online Appendix V, we show that the function V is Gâteaux differentiable with derivative given by the local utility $w(x, F)$ as in Proposition 1. Theorem 1 then implies that the local utility is given by $w(x, F) = v(x) + \sigma(x, \hat{y}(F))$ for every $F \in \mathcal{F}$. With this, the result follows from Proposition 3 in Cerreia-Vioglio, Maccheroni, and Marinacci [2017]. ■

Online Appendix II: Ancillary results

This appendix gives proofs of the ancillary results stated in the main appendix.

Online Appendix II.A: Ancillary results for Appendix I

Proof of Lemma 6. We must show that σ is non-negative, weakly continuous, that $\sigma(x, x) = 0$ and that $\int \sigma(x, F) dF(x) \leqslant \int \sigma(x, G) dF(x)$. Non-negativity is obvious. Since $h(x, s)$ is continuous in x we have $F^n \to F$ implies that $h_{F^n}(s)$ converges pointwise to $h_n(s)$. Hence $\left(h(x, s) - \int h(\tilde{x}, s) dF^n(\tilde{x})\right)^2$ converges pointwise to $\left(h(x, s) - \int h(\tilde{x}, s) dF(\tilde{x})\right)^2$. Given that h is square-integrable over (S, μ), the dominated convergence theorem implies that

$$\int \left(h(x, s) - \int h(\tilde{x}, s) dF^n(\tilde{x}) \right)^2 d\mu(s) \to \int \left(h(x, s) - \int h(\tilde{x}, s) dF(\tilde{x}) \right)^2 d\mu(s).$$

For the last property, $\sigma(x,x) = \int (h(x,s) - h(x,s))^2 \, d\mu(s) = 0$, and so

$$\int \sigma(x,G) dF(x) = \int \int (h(x,s) - h_G(s))^2 \, d\mu(s) dF(x) = \int \left(\int (h(x,s) - h_G(s))^2 \, dF(x) \right) d\mu(s).$$

Since mean square error is minimized by the mean for each s,

$$h(F,s) = \int h(x,s) dF(x) \in \arg\min_{H \in \mathbb{R}} \int (h(x,s) - H)^2 \, dF(x)$$

implying that $\int \sigma(x,F) dF(x) \leqslant \int \sigma(x,G) dF(x)$. ■

Proof of Lemma 7. By definition $V(F) = \int\int (h(x,s) - h(F,s))^2 \, d\mu(s) dF(x)$, and simple manipulations show this is equal to

$$\int H(x,x) dF(x) - \int \int \int [h(x,s) h(\tilde{x},s) d\mu(s)] \, dF(x) dF(\tilde{x}).$$

We extend V to the space of signed measures by

$$V(F+M) = \int H(x,x) d\left(F(x) + M(x)\right) - \int \int H(x,\tilde{x}) d\left(F(x) + M(x)\right) d\left(F(\tilde{x}) + M(\tilde{x})\right)$$

and observe that the cross term is

$$-2 \int \left(\int H(x,\tilde{x}) dF(\tilde{x}) \right) dM(x) = -2 \int \int h(x,s) h(\tilde{x},s) d\mu(s) dF(\tilde{x}) dM(x)$$

so that

$$V(F+M) = V(F) + \int \left[H(x,x) - 2 \int h(x,s) h(\tilde{x},s) d\mu(s) dF(\tilde{x}) \right] dM(x) - \int \int H(x,\tilde{x}) dM(x) dM(\tilde{x}).$$

This enables us to compute the directional derivatives. The directional derivative in the direction $M = \delta_z - F$ is given as

$$DV(F)(\delta_z - F) = \int \left[\int h^2(x,s) d\mu(s) - 2 \int h(x,s) h(\tilde{x},s) d\mu(s) dF(\tilde{x}) \right] (d\delta_z - dF(x))$$

$$= \int h^2(z,s) d\mu(s) - 2 \int h(z,s) h(\tilde{x},s) d\mu(s) dF(\tilde{x})$$

$$-\int h^2(x,s)dF(x)d\mu(s) + 2\int h(x,s)h(\tilde{x},s)d\mu(s)dF(\tilde{x})dF(x). \qquad \blacksquare$$

Online Appendix II.B: Ancillary results for Appendix B

We next restate and prove Theorem 6. Moreover, we relax the original assumptions by considering an arbitrary adversarial expected utility representation (Y, u) of V, and an arbitrary convex and compact set of feasible lotteries $\overline{\mathcal{F}} \subseteq \mathcal{F}$. Define $V^*\left(\overline{\mathcal{F}}\right) = \max_{F\in\overline{\mathcal{F}}} V\left(F\right)$. By Sion's minmax theorem,

$$V^*\left(\overline{\mathcal{F}}\right) = \max_{F\in\overline{\mathcal{F}}} \min_{y\in Y} \int u\left(x,y\right) dF\left(x\right) = \min_{H\in\mathcal{H}} \max_{F\in ext\left(\overline{\mathcal{F}}\right)} \int\int u\left(x,y\right) dF\left(x\right) dH\left(y\right).$$

Theorem 7. Fix $\hat{H} \in \arg\min_{H\in\mathcal{H}} \max_{F\in ext\left(\overline{\mathcal{F}}\right)} \int\int u\left(x,y\right) dF\left(x\right) dH\left(y\right)$. Then $\hat{F} \in \arg\max_{F\in\overline{\mathcal{F}}} V\left(F\right)$ if and only if for all $\tilde{F} \in ext\left(\overline{\mathcal{F}}\right)$, $V(\hat{F}) \geqslant \int\int u\left(x,y\right) d\tilde{F}\left(x\right) d\hat{H}\left(y\right)$, and, for all $\tilde{F} \in \bigcup_{\lambda\in\Lambda_{\hat{F}}} \operatorname{supp}\lambda$, $V(\hat{F}) = \int\int u\left(x,y\right) d\tilde{F}\left(x\right) d\hat{H}\left(y\right)$.

Proof of Theorem 6. Fix $\hat{H}$ as in the statement. Then fix $\hat{F} \in \arg\max_{F\in\overline{\mathcal{F}}} V\left(F\right)$, $\tilde{F} \in ext\left(\overline{\mathcal{F}}\right)$, and observe that

$$\begin{aligned}
&\int\int u\left(x,y\right) d\tilde{F}\left(x\right) d\hat{H}\left(y\right) \leqslant \max_{F\in ext\left(\overline{\mathcal{F}}\right)} \int\int u\left(x,y\right) dF\left(x\right) d\hat{H}\left(y\right) \\
&= \min_{H\in\mathcal{H}} \max_{F\in ext\left(\overline{\mathcal{F}}\right)} \int\int u\left(x,y\right) dF\left(x\right) dH\left(y\right) = V^*\left(\overline{\mathcal{F}}\right) = V\left(\hat{F}\right),
\end{aligned}$$

yielding the first part of the desired condition. Next, observe that

$$\begin{aligned}
V^*\left(\overline{\mathcal{F}}\right) &= \max_{F\in ext\left(\overline{\mathcal{F}}\right)} \int\int u\left(x,y\right) dF\left(x\right) d\hat{H}\left(y\right) \\
&\geqslant \int\int u\left(x,y\right) d\hat{F}\left(x\right) d\hat{H}\left(y\right) \geqslant \min_{H\in\mathcal{H}} \int\int u\left(x,y\right) d\hat{F}\left(x\right) dH(y) = V^*\left(\overline{\mathcal{F}}\right),
\end{aligned}$$

Combining the first two chains of inequalities yields

$$\int\int u\left(x,y\right) d\hat{F}\left(x\right) d\hat{H}\left(y\right) \geqslant \int\int u\left(x,y\right) d\tilde{F}\left(x\right) d\hat{H}\left(y\right) \qquad \forall\tilde{F} \in ext\left(\overline{\mathcal{F}}\right). \tag{28}$$

Next, fix $\lambda \in \Lambda_{\hat{F}}$, $F^* \in \operatorname{supp} \lambda$, and assume toward a contradiction that

$$V\left(\hat{F}\right) > \int\int u\left(x,y\right) dF^*\left(x\right) d\hat{H}\left(y\right).$$

It follows that $\int \left(\int u\left(x,y\right) d\tilde{F}\left(x\right)\right) d\hat{H}\left(y\right) d\lambda\left(\tilde{F}\right) = \int u\left(x,y\right) d\hat{F}\left(x\right) d\hat{H}\left(y\right)$ $\geqslant V\left(\hat{F}\right) > \int\int u\left(x,y\right) dF^*\left(x\right) d\hat{H}\left(y\right)$, so there exists $F^\star \in \operatorname{supp}\lambda$ and $\varepsilon > 0$ such that

$$\int\int u\left(x,y\right) dF^\star\left(x\right) d\hat{H}\left(y\right) > \int\int u\left(x,y\right) d\tilde{F}\left(x\right) d\hat{H}\left(y\right)$$

for all $\tilde{F} \in \operatorname{supp}\lambda \cap B_\varepsilon\left(F^*\right)$, where $B_\varepsilon\left(F^*\right) \subseteq \mathcal{F}$ is the ball of radius ε (in the Kantorovich-Rubinstein metric) centered at F^*.

Next, define the probability measure $\lambda^\star = \lambda(B_\varepsilon\left(F^*\right))\delta_{F^\star} + (1 - \lambda\left(B_\varepsilon\left(F^*\right)\right))\,\lambda\left(\cdot|B_\varepsilon\left(F^*\right)^c\right)$ and the lottery $F_{\lambda^\star} = \int \tilde{F} d\lambda^\star(\tilde{F})$. Then

$$\begin{aligned}
&\int\int u\left(x,y\right) dF_{\lambda^\star}\left(x\right) d\hat{H}\left(y\right) = \int\left(\int u\left(x,y\right) d\tilde{F}\left(x\right)\right) d\hat{H}\left(y\right) d\lambda^\star\left(\tilde{F}\right) \\
&= \lambda(B_\varepsilon\left(F^*\right))\int u\left(x,y\right) dF^\star\left(x\right) + \left(1 - \lambda\left(B_\varepsilon\left(F^*\right)\right)\right)\int\left(\int u\left(x,y\right) d\tilde{F}\left(x\right)\right) d\hat{H}\left(y\right) d\lambda\left(\tilde{F}|B_\varepsilon\left(F^*\right)^c\right) \\
&> \lambda(B_\varepsilon\left(F^*\right))\int\left(\int u\left(x,y\right) d\tilde{F}\left(x\right)\right) d\hat{H}\left(y\right) d\lambda\left(\tilde{F}|B_\varepsilon\left(F^*\right)\right) \\
&+ \left(1 - \lambda\left(B_\varepsilon\left(F^*\right)\right)\right)\int\left(\int u\left(x,y\right) d\tilde{F}\left(x\right)\right) d\hat{H}\left(y\right) d\lambda\left(\tilde{F}|B_\varepsilon\left(F^*\right)^c\right) \\
&= \int\left(\int u\left(x,y\right) d\tilde{F}\left(x\right)\right) d\hat{H}\left(y\right) d\lambda\left(\tilde{F}\right) = \int\int u\left(x,y\right) d\hat{F}\left(x\right) d\hat{H}\left(y\right)
\end{aligned}$$

which contradicts equation (28).

Conversely, fix $\tilde{F} \in ext\left(\overline{\mathcal{F}}\right)$ and observe that the implication follows by

$$\begin{aligned}
V\left(\hat{F}\right) &\geqslant \max_{\tilde{F} \in ext\left(\overline{\mathcal{F}}\right)} \int\int u\left(x,y\right) d\tilde{F}\left(x\right) d\hat{H}\left(y\right) \\
&= \min_{H \in \mathcal{H}} \max_{F \in ext\left(\overline{\mathcal{F}}\right)} \int\int u\left(x,y\right) dF\left(x\right) dH\left(y\right) = V^*\left(\hat{\mathcal{F}}\right) \geqslant V\left(\hat{F}\right). \qquad \blacksquare
\end{aligned}$$

Before proving Lemma 9, we state and prove an intermediate result.

Lemma 10. *For every $F \in \mathcal{F}_\Gamma(\overline{X})$, there exists a sequence $F^n \to F$ such that each F^n is the convex combination of finitely many points in $ext(\mathcal{F}_\Gamma(\overline{X}))$.*

Proof. Define $\mathcal{F}_e = ext(\mathcal{F}_\Gamma(\overline{X}))$ and endow it with the relative topology. This makes $\mathcal{F}_e$ metrizable. Next, by the Choquet's theorem, $\mathcal{F}_\Gamma(\overline{X})$ can be embedded in the set $\Delta(\mathcal{F}_e)$ of Borel probability measures over $\mathcal{F}_e$. By Theorem 15.10 in Aliprantis and Border [2006], the subset $\Delta_0(\mathcal{F}_e)$ of finitely supported probability measures over $\mathcal{F}_e$ is dense in $\Delta(\mathcal{F}_e)$. In turn, this implies the statement. ■

Lemma 6. Suppose that for every finite set $\mathcal{E} \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$ there exists a lottery $F_\mathcal{E}$ that solves $\max_{F\in co(\mathcal{E})} V(F)$ and has finite support on no more than $(m+1)(k+1)$ points of $\overline{X}$. Then there exists a lottery F^* that solves $\max_{F\in\mathcal{F}_\Gamma(\overline{X})} V(F)$ and that has finite support on no more than $(m+1)(k+1)$ points of $\overline{X}$.

Proof of Lemma 9. Let $\hat{F}$ solve $\max_{F\in\mathcal{F}_\Gamma(\overline{X})} V(F)$. By Lemma 10, there exists a sequence $\hat{F}^n \to \hat{F}$ such that, for every $n \in \mathbb{N}$, $\hat{F}^n \in co(\mathcal{E}^n)$ for some finite set $\mathcal{E}^n \subseteq ext(\mathcal{F}_\Gamma(\overline{X}))$. By Theorem 7, for every $n \in \mathbb{N}$, there exists a lottery $F^n \in co(\mathcal{E}^n)$ that is supported on no more that $(k+1)(m+1)$ points of $\overline{X}$ and such that $V(F^n) \geqslant V(\hat{F}^n)$. Given that $\mathcal{F}_\Gamma(\overline{X})$ is compact, there exists a subsequence of F^n that converges to some lottery $F^* \in \mathcal{F}_\Gamma(\overline{X})$. Since each F^n has support on at most $(k+1)(m+1)$ points, the same is true for F^*. And since V is continuous $V(F^n) \to V(F^*)$ and $V(\hat{F}_n) \to V(\hat{F})$ hence $V(F^*) \geqslant V(\hat{F})$, F^* is optimal. ■

Corollary 1. Maintain the assumptions of Proposition 7, and let F denote the unique fully supported solution. There exists a sequence of method of moments representations V^n with $|S^n| = m^n \in \mathbb{N}$, and a sequence of lotteries F^n such that each F^n is optimal for V^n, is supported on up to $m^n + 1$ points, and $F^n \to F$ weakly, with $\operatorname{supp} F^n \to \operatorname{supp} F = X$ in the Hausdorff topology.

Proof of Corollary 2. By Theorem 15.10 in Aliprantis and Border [2006], there exists a sequence of finitely supported $\mu^n \in \Delta(S)$ such that $\mu^n \to \mu$. The GMM adversarial forecaster representation V^n induced by $(h, \mu^n$ satisfies the assumptions of Theorem 2 by defining $Y^n = \prod_{s\in\operatorname{supp}\mu^n} h(X,s) \subseteq \mathbb{R}^{m^n}$, where $m_n = |\operatorname{supp}\mu^n|$, so for every $n \in \mathbb{N}$, there exists a solution F^n of the problem $\max_{F\in\Delta(\overline{X})} V^n(F)$ that is supported on up to $m_n + 1$ points of $\overline{X}$. Because the constraint set $\Delta(\overline{X})$ is compact and V is continuous, the Berge maximum theorem implies that all the accumulation points of the sequence F^n are solutions of the problem $\max_{F\in\Delta(\overline{X})} V(F)$, where V is the GMM adversarial forecaster representation induced by h and μ. Proposition 7 established that this problem has a unique full-support solution F, so F is the unique accumulation point of F^n. Because $\overline{X}$ is compact, the sequence $\operatorname{supp} F^n$ converges to some set $\hat{X} \subseteq \overline{X}$ in the Haussdorf sense. By Box 1.13 in Santambrogio [2015], $F^n \to F$ implies that $\operatorname{supp} F \subseteq \hat{X}$, and, given that $\operatorname{supp} F = X$, it follows that $\operatorname{supp} F^n \to \overline{X}$. ■

Online Appendix III: Optimization

This appendix collects additional optimization results for adversarial forecaster and adversarial expected utility representation that are of independent interest.

Online Appendix III.A: Optimal lotteries in the adversarial EU model

Here we provide two alternative characterizations of optimal lotteries under the adversarial expected utility model.

Proposition 10. *Let V be an adversarial expected utility representation (Y, u) and let $\overline{\mathcal{F}} \subseteq \mathcal{F}$ be a convex and compact set. The following are equivalent:*

(i) $F^* \in \operatorname{argmax}_{F\in\overline{\mathcal{F}}} V(F)$

(ii) There exists $H \in \mathcal{H}(\hat{Y}(F^*))$ such that $F^* \in \operatorname{argmax}_{F\in\overline{\mathcal{F}}} \int\int u(x,y)dH(y)dF(x)$.

(iii) For all $F \in \overline{\mathcal{F}}$, there exists $y \in \hat{Y}(F^*)$ such that $\int u(x,y)dF^*(x) \geqslant \int u(x,y)dF(x)$.

The equivalence between (i) and (iii) is similar to Proposition 1 in Loseto and Lucia [2021], with the important difference that they consider quasiconcave representations and restrict to a finite set of utilities (which corresponds to a finite Y in our notation).

Proof. As a preliminary step, define $\mathcal{W} = \{u(\cdot, y)\}_{y\in Y}$ and observe that it is compact since u is continuous.

The equivalence between (ii) and (iii) is a standard application of the Wald-Pearce Lemma, so we only prove the equivalence between (i) and (ii).

(ii) implies (i). Let $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} \int\int u(x,y)dH(y)dF(x)$ for some $H \in \mathcal{H}(\hat{Y}(F^*))$. For all $\tilde{F} \in \overline{\mathcal{F}}$, we have

$$V(F^*) = \int\int u(x,y)dH(y)dF^*(x) \geqslant \int\int u(x,y)dH(y)d\tilde{F}(x) \geqslant V(\tilde{F}),$$

yielding that $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} V(F)$.

(i) implies (ii). Fix $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} V(F)$. Define $R : C(X) \to \mathbb{R}$ as $R(w) = \max_{F \in \overline{\mathcal{F}}} \int w(x)dF(x)$ and let $co(\mathcal{W})$ denote the closed convex hull of $\mathcal{W}$, which is also compact. Because $\overline{\mathcal{F}}$ is compact, R is continuous. Fix $w^* \in \operatorname{argmin}_{w \in co(\mathcal{W})} R(w)$. Observe that

$$\begin{aligned} &\min_{w \in co(\mathcal{W})} \int w(x)dF^*(x) = \max_{F \in \overline{\mathcal{F}}} \min_{w \in co(\mathcal{W})} \int w(x)dF(x) = \min_{w \in co(\mathcal{W})} \max_{F \in \overline{\mathcal{F}}} \int w(x)dF(x) \\ &= \max_{F \in \overline{\mathcal{F}}} \int w^*(x)dF(x) \geqslant \int w^*(x)dF^*(x) \geqslant \min_{w \in co(\mathcal{W})} \int w(x)dF^*(x) \end{aligned}$$

This shows that $w^* \in \operatorname{argmin}_{w \in co(\mathcal{W})} \int w(x)dF^*(x)$, that is, there exists $H \in \mathcal{H}(\hat{Y}(F^*))$ such that $w^*(x) = \int u(x,y)dH(y)$. Next, observe that

$$\begin{aligned} \max_{F \in \overline{\mathcal{F}}} \min_{w \in co(\mathcal{W})} \int w(x)dF(x) &= \max_{F \in \overline{\mathcal{F}}} V(F) = V(F^*) = \min_{w \in \mathcal{W}} \int w(x)dF^*(x) \\ &\leqslant \int w^*(x)dF^*(x) \leqslant \max_{F \in \overline{\mathcal{F}}} \int w^*(x)dF(x) \\ &= \min_{w \in co(\mathcal{W})} \max_{F \in \overline{\mathcal{F}}} \int w(x)dF(x) = \max_{F \in \overline{\mathcal{F}}} \min_{w \in co(\mathcal{W})} \int w(x)dF(x), \end{aligned}$$

where the last equality follows from Sion minamx theorem given that $\overline{\mathcal{F}}$ is compact and convex. This yields $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} \int w^*(x)dF(x) = \operatorname{argmax}_{F \in \overline{\mathcal{F}}} \int\int u(x,y)dH(y)dF(x)$.
■

Online Appendix III.B: finite Y

This section states and proves additional results on the optimization problem of Section 3. Fix an arbitrary compact and convex set $\overline{\mathcal{F}} \subseteq \mathcal{F}$ of feasible lotteries. We start with a simple lemma that establishes the existence of a saddle pair (F^*, y^*).

Lemma 11. *There exists* $F^* \in \overline{\mathcal{F}}$ *and* $y^* \in Y$ *such that*

$$\int u(x,y^*)dF^*(x) = V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F) \tag{29}$$

Proof. Because $\overline{\mathcal{F}}$ is compact and V is continuous in the weak topology, there exists $F^* \in \overline{\mathcal{F}}$ such that $V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F)$. And because Y is compact and u is continuous in y, there exists $y^* \in Y$ such that $\int u(x,y^*)dF^*(x) = V(F^*)$, yielding the statement. ■

For every (F^*, y^*) as in Lemma 11, define the set

$$\overline{\mathcal{F}}(F^*, y^*) = \left\{ F \in \overline{\mathcal{F}} : \forall y \in Y \backslash \{y^*\}, \int u(x,y)dF(x) \geqslant \int u(x,y)dF^*(x) \right\} \tag{30}$$

Observe that $\overline{\mathcal{F}}(F^*, y^*)$ is nonempty since it contains F^*, and convex since it is defined by (possibly infinitely many) linear inequalities. In addition, $\overline{\mathcal{F}}(F^*, y^*)$ is the intersection of closed sets since $u(.,y)$ is a continuous function for all $y \in Y \backslash \{y^*\}$, so it too is closed.

Lemma 12. *Fix* (F^*, y^*) *as in Lemma 11 and a nonempty, closed, and convex set* $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$. *The set* $\operatorname{argmax}_{F \in K} \int u(x,y^*)dF(x)$ *is nonempty, convex, and closed.*

Proof. Given that K is nonempty, convex, and closed, hence compact, and the map $F \mapsto \int u(x,y^*)dF(x)$ is linear and continuous, the statement immediately follows. ■

We next state and prove a general, yet simple, result about the existence of maximizers of Problem 29 that are extreme points of convex, closed sets $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$.

Lemma 13. *For any* (F^*, y^*) *as in Lemma 11 and nonempty, closed, and convex set* $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$ *such that* $F^* \in K$,

$$\operatorname*{argmax}_{F \in K} \int u(x,y^*)dF(x) \subseteq \operatorname*{argmax}_{F \in \overline{\mathcal{F}}} V(F). \tag{31}$$

In particular, there exists $F_0 \in ext(K)$ *such that* $V(F_0) = V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F)$.

Proof. Fix $F^* \in \mathcal{F}$ and $y^* \in Y$ as in Lemma 11 and a nonempty, closed, and convex set $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$. Let $\hat{F} \in \text{argmax}_{F \in K} \int u(x, y^*) dF(x)$. We need to show that $V(\hat{F}) = V(F^*)$. Observe that

$$\int u(x,y) d\hat{F}(x) \geqslant \int u(x,y) dF^*(x) \qquad \forall y \in Y \backslash \{y^*\} \tag{32}$$

since $\hat{F} \in K \subseteq \overline{\mathcal{F}}(F^*, y^*)$. Moreover,

$$\int u(x,y^*) d\hat{F}(x) \geqslant \int u(x,y^*) dF^*(x) \tag{33}$$

since $\hat{F} \in \text{argmax}_{F \in K} \int u(x, y^*) dF(x)$ and $F^* \in K$. Then for all $y \in Y$, we have that

$$\int u(x,y) d\hat{F}(x) \geqslant \int u(x,y) dF^*(x) \geqslant V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F) \tag{34}$$

and in particular that $V(\hat{F}) \geqslant \max_{F \in \overline{\mathcal{F}}} V(F)$. Given that $\hat{F} \in \overline{\mathcal{F}}$, we must have $V(\hat{F}) = V(F^*)$, so $\hat{F} \in \text{argmax}_{F \in \overline{\mathcal{F}}} V(F)$. This proves the first part of the theorem. The second part immediately follows from the Bauer maximum principle since the map $F \mapsto \int u(x, y^*) dF(x)$ is linear over the convex set K. ■

Lemma 13 is not very insightful per se since the set $\overline{\mathcal{F}}(F^*, y^*)$ depends on the particular choice of (F^*, y^*). However, whenever we can find a set K as in the statement of Lemma 13 whose extreme points satisfy interesting properties, the theorem lets us conclude that there is an optimizer of the original problem with those properties. We now apply this strategy to optimization problems with additional structure on $\overline{\mathcal{F}}$ and on Y by relying on known characterizations of extreme points of sets of probability measures. For completeness, we report here the original results mentioned.

Theorem 8 (Proposition 2.1 in Winkler [1988])**.** *Fix a convex and closed set $\overline{\mathcal{F}} \subset \mathcal{F}$, an affine function $\Lambda : \overline{\mathcal{F}} \to \mathbb{R}^{n-1}$, and a convex set $C \subset \Lambda(\overline{\mathcal{F}})$. The set $\Lambda^{-1}(C)$ is convex and every extreme point of $\Lambda^{-1}(C)$ is a convex combination of at most n extreme points of $\overline{\mathcal{F}}$.*

We can combine this result with Lemma 13 to obtain the following result.

Theorem 9. *Suppose that Y has m elements. There exists a solution $F^* \in \text{argmax}_{F \in \overline{\mathcal{F}}} V(F)$ that is a convex combination of at most m extreme points of $\overline{\mathcal{F}}$.*

Proof. Fix (F^*, y^*) as in Lemma 11. Observe that $|Y\backslash\{y^*\}| = m - 1$ by assumption. For simplicity we write $Y\backslash\{y^*\} = \{y_1, ..., y_m - 1\}$. Define the map $\Lambda : \overline{\mathcal{F}} \to \mathbb{R}^{m-1}$ as

$$\Lambda(F)_i = \int u(x, y_i) dF(x) \qquad \forall i \in \{1, ..., m-1\} \tag{35}$$

Also define the convex set

$$C \equiv \Lambda(\overline{\mathcal{F}}(F^*, y^*)) \subseteq \Lambda(\overline{\mathcal{F}}) \tag{36}$$

It is easy to see that $\Lambda^{-1}(C) = \overline{\mathcal{F}}(F^*, y^*)$. By Theorem 8 it follows that every extreme point of $\overline{\mathcal{F}}(F^*, y^*)$ is a convex combination of at most n extreme points of $\overline{\mathcal{F}}$. Finally, the statement follows by a direct application of Theorem 13. ■

The next result sharpens Theorem 2 for the case where Y is finite.

Theorem 10. *Suppose that Y is finite with m elements. For every closed $\overline{X} \subseteq X$, there exists an optimal lottery F^* for the problem in equation 9 that has finite support on no more than $k + m$ points of $\overline{X}$.*

Proof. Let $\overline{\mathcal{F}} = \mathcal{F}_\Gamma(\overline{X})$ for some closed $\overline{X} \subseteq X$, and fix (F^*, y^*) as in Lemma 11. The set $\overline{\mathcal{F}}(F^*, y^*)$ is defined by $k + m - 1$ moment restrictions: k moments restrictions from Γ and $m - 1$ from the definition of $\overline{\mathcal{F}}(F^*, y^*)$. By Lemma 13 there exists $F^* \in \text{ext}(\overline{\mathcal{F}}(F^*, y^*))$ such that $V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F)$. By Winkler's Theorem the each $\tilde{F} \in \overline{\mathcal{F}}(F^*, y^*)$ is supported on up to $k + m$ points of $\overline{X}$ as desired. ■

Online Appendix III.C: Robust solutions

This section shows that the finite-support property of Theorem 2 generically holds for all solutions of the optimization problem in equation 9 that are "robust" in the following sense. For every $F \in \mathcal{F}_\Gamma(\overline{X})$, we call a sequence as in Lemma 10 a *finitely approximating sequence of F.*

Definition 7. Fix $w \in C(\overline{X})$ and a lottery F that solves

$$\max_{F \in \mathcal{F}_\Gamma(\overline{X})} \min_{y \in Y} \int u(x, y) + w(x) dF(x)$$

We say that F is a *robust solution at* w if

$$F^n \in \underset{\tilde{F} \in co(\mathcal{E}^n)}{\operatorname{argmax}} \left\{ \min_{y \in Y} \int u(x, y) + w(x) dF(x) \right\}$$

for some approximating sequence $F^n \in co(\mathcal{E}^n)$ of F, with $\mathcal{E}^n$ being any finite set of extreme points generating F^n.

In words, an optimal lottery F is robust if it can be approximated by a sequence of lotteries that are generated by finitely many extreme points and that are optimal within the set of lotteries generated by the same extreme points.

Theorem 11. *Suppose that Y is an m-dimensional manifold with boundary, that u is continuously differentiable in y, and that Y and u satisfy the uniqueness property. For an open dense set of $w \in \overline{\mathcal{W}} \subseteq C(\overline{X})$, every robust solution at w has finite support on no more than $(k+1)(m+1)$ points of $\overline{X}$.*

The proof will use the following lemma.

Lemma 14. *Fix a finite set $\hat{X} \subseteq \overline{X}$ and an open dense subset $\hat{\mathcal{W}}$ of $\mathbb{R}^{\hat{X}}$. The set*

$$\overline{\mathcal{W}} = \left\{ w \in C(\overline{X}) : w_{|\hat{X}} \in \hat{\mathcal{W}} \right\}$$

is open and dense in $C(\overline{X})$, where $w_{|\hat{X}}$ denotes the restriction of w on $\hat{X}$.

Proof. Because $\hat{\mathcal{W}}$ is open, so is $\overline{\mathcal{W}}$. Fix $w \in C(\overline{X})$. Given that $w_{|\hat{X}} \in \mathbb{R}^{\hat{X}}$, there exists a sequence $\hat{w}^n \in \hat{\mathcal{W}}$ such that $\hat{w}^n \to w_{|\hat{X}}$. Next, fix $n \in \mathbb{N}$ large enough that $B_{1/n}(\hat{x}) \cap B_{1/n}(\hat{x}') = \varnothing$ for all $\hat{x}, \hat{x}' \in \hat{X}$.[37] By Urysohn's Lemma (see Lemma 2.46 in Aliprantis and Border [2006]), for every $\hat{x} \in \hat{X}$, there exists a continuous function $v^n_{\hat{x}}$ such that $v^n_{\hat{x}}(x) = 0$ for all $x \in \overline{X} \backslash B_{1/n}(\hat{x})$ and $v^n_{\hat{x}}(\hat{x}) = 1$. Now define the continuous function

$$w^n(x) = w(x)(1 - \max_{\hat{x} \in \hat{X}} v^n_{\hat{x}}(x)) + \sum_{\hat{x} \in \hat{X}} \hat{w}^n(x) v^n_{\hat{x}}(x).$$

Because $w^n \in \overline{\mathcal{W}}$, $\hat{X}$ is finite, and $\overline{X}$ is compact, $w^n \to w$ as desired. ■

[37]Here, $B_{1/n}(\hat{x})$ is the open ball centered at $\hat{x}$ and of radius $1/n$.

Proof of Theorem 11. Without loss of generality, we assume that $\overline{X} = \bigcup_{F \in \mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F$.[38] Define $\overline{\mathcal{E}} = cl\left(ext\left(\mathcal{F}_\Gamma(\overline{X})\right)\right)$ and consider an increasing sequence of finite sets of extreme points $\mathcal{E}^n \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$ such that $\mathcal{E}^n \uparrow \overline{\mathcal{E}}$. Observe that, by construction, $\overline{X}_{\mathcal{E}^n} \uparrow \overline{X}$.[39] For every $n \in \mathbb{N}$, let $\hat{\mathcal{W}}^n$ the open dense subset of $\mathbb{R}^{\overline{X}_{\mathcal{E}^n}}$ that satisfies the property of point 2 in Theorem 7. By Lemma 14 the set

$$\overline{\mathcal{W}}^n = \left\{ w \in C(\overline{X}) : w_{|\overline{X}_{\mathcal{E}^n}} \in \hat{\mathcal{W}}^n \right\}$$

is an open dense subset of $C(\overline{X})$. By the Baire category theorem (see Theorem 3.46 in Aliprantis and Border [2006]), the set $\overline{\mathcal{W}} = \bigcap_{n \in \mathbb{N}} \overline{\mathcal{W}}^n$ is dense in $C(\overline{X})$.

Next, fix $w \in \overline{\mathcal{W}}$ and a robust optimal lottery F^* for

$$\max_{F \in \mathcal{F}_\Gamma(\overline{X})} \min_{y \in Y} \int u(x, y) + w(x) dF(x)$$

It follows that F^* is the weak limit of a sequence of solutions F^n of the problem

$$\max_{F \in co(\mathcal{E}^n)} \min_{y \in Y} \int u(x, y) + w(x) dF(x)$$

In particular, given that, for every $n \in \mathbb{N}$ $w_{|\overline{X}_{\mathcal{E}^n}} \in \hat{\mathcal{W}}^n$, Theorem 7 implies that F^n is supported on up to $(k+1)(m+1)$ points of $\overline{X}_{\mathcal{E}^n}$. Because $F^n \to F^*$, it follows that F is supported on up to $(k+1)(m+1)$ points of $\overline{X}$. Given that F^* and w were arbitrarily chosen, the result follows. ■

Online Appendix IV: Additional applications

Online Appendix IV.B: Additional examples

This section presents two examples. In the first, there are GMM preferences that have a strictly concave representation and give rise to an optimal lottery with full support. The second example illustrates most of the main results in the text by solving an

[38] Assume not, then we could just consider lotteries over the closed set $\overline{X}' = cl\left(\bigcup_{F \in \mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F\right)$.

[39] This follows from the fact that $\overline{X} = \bigcup_{F \in \mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F$ by assumption. See also footnote 38.

optimal lottery under the asymmetric adversarial forecaster preferences of Section 5.2.

Example 10 (Weiner Process Example)**.** We interpret $x \in [0, 1]$ as time. While it is natural to think of $h(\cdot, s)$ as a random function of s with distribution induced by F, there is a dual interpretation in which we think of $h(x, \cdot)$ as a random function of x (a random field) with distribution induced by μ. In this interpretation, the $H(x, \tilde{x})$ are the second (non-central) moments of that random variable between different points $x, \tilde{x}$ in the random field. If, for example, $X = [0, 1]$, then this random field is a stochastic process, and $H(x, \tilde{x})$ the second moments of the process h between times $x, \tilde{x}$. It is well known that continuous time Markov process are equivalent to stochastic differential equations and that an underlying measure space S and measure μ can be found for each such process. Specifically, consider the process generated by the stochastic differential equation $dh = -h + dW$ where W is the standard Weiner process on (S, μ) and the initial condition $h(0, s)$ has a standard normal distribution. Then the distribution of the difference between $h(x, \cdot)$and $h(\tilde{x}, \cdot)$ depends only on the time difference $\tilde{x} - x$, and in particular $H(x, \tilde{x}) = \int h(x, s)h(\tilde{x}, s)d\mu(s) = G(x - \tilde{x})$. In this case $H(0, \tilde{x}) = e^{-\tilde{x}}$, which is non-negative, strictly decreasing and strictly convex. △

Example 11 (Optimal lotteries under asymmetric forecast error)**.** Let $X = [0, 1]$ and consider the parametric adversarial forecaster preferences with asymmetric loss function $\rho(z) = \exp(\lambda z) - \lambda z$ and linear baseline utility $v(x) = \overline{v}x$ for some $0 < \overline{v} < 1$ and $\lambda > 0$. In this case, the best response of the adversary is $\hat{x}(F) = \frac{1}{\lambda} \ln \left(\int_0^1 \exp(\lambda x) dF(x) \right)$ and the continuous local utility function is $w(x, F) = \overline{v}x + \exp(\lambda(x - \hat{x}(F))) - \lambda(x - \hat{x}(F))$, which is convex for every F. Corollary 8 then implies that the preference induced by this adversarial forecaster representation preserves the MPS order. Now consider maximizing the V defined by the loss function above over the entire simplex $\mathcal{F}$. Because the preference preserves the MPS order, Theorem 2 shows that the optimal distributions are supported on 0 and 1, that is, $F = p\delta_1 + (1 - p)\delta_0$ for some $p \in [0, 1]$. By Proposition 2, the optimal probability p^* solves

$$\max_{p \in [0,1]} \overline{v}p + p\left(\exp\left(\lambda(1 - \hat{x}(p^*))\right) - \lambda(1 - \hat{x}(p^*))\right) + (1 - p)(\exp\left(-\lambda\hat{x}(p^*)\right) + \lambda\hat{x}(p^*)). \tag{37}$$

If there is an interior solution, the agent is indifferent over any $p \in [0,1]$. This is the case only if the solution is the p^*_{int} defined by

$$\overline{v} + \exp\left(\lambda(1 - \hat{x}(p^*_{int}))\right) - \lambda = \exp\left(-\lambda \hat{x}(p^*_{int})\right)$$

which is equivalent to

$$p^*_{int} = \frac{1}{(\lambda - \overline{v})} - \frac{1}{(\exp(\lambda) - 1)}.$$

Therefore, the overall solution is $p^* = \min\{1, \max\{0, p^*_{int}\}\}$. Clearly the solution is increasing with respect to the baseline utility parameter $\overline{v}$, but the effect of the asymmetry parameter λ is ambiguous. △

Online Appendix V: Adversarial forecasters, local utilities, and Gâteaux derivatives

In this section, we discuss the relationship between our notion of local utility and the one in Machina [1982]. This is closely related to the differentiability properties of a function V with a continuous local expected utility, which we also discuss.

Fix a continuous functional $V : \mathcal{F} \to \mathbb{R}$. Recall that V has a local expected utility if, for every $F \in \mathcal{F}$ there exists $w(\cdot, F) \in C(X)$ such that $V(F) = \int w(x,F)dF(x)$ and $V(\tilde{F}) \leqslant \int w(x,F)d\tilde{F}(x)$ for all $\tilde{F} \in \mathcal{F}$, We say that this local expected utility is continuous if w is continuous in (x, F).

Proposition 11. *Let $\succsim$ admit a representation V with a local expected utility w and, for every $F \in \mathcal{F}$, let $\succsim_F$ denote the expected utility preference induced by $w(\cdot, F)$. Then $F \succsim_F \tilde{F}$ (resp. $F \succ_F \tilde{F}$) implies that $F \succsim \tilde{F}$ (resp. $F \succ \tilde{F}$).*

Proof. The first implication follows from $V(F) = \int w(x,F)dF(x) \geqslant \int w(x,F)d\tilde{F}(x) \geqslant V(\tilde{F})$. To prove the second, let $V(\tilde{F}) \geqslant V(F)$ and observe that $\int w(x,F)d\tilde{F}(x) \geqslant V(\tilde{F}) \geqslant V(F) = \int w(x,F)dF(x)$, implying that $\tilde{F} \succsim_F F$ as desired. ■

Machina [1982] introduced the concept of local utilities for a preference over lotteries with $X \subseteq \mathbb{R}$. For ease of comparison, we make assume here that $X = [0,1]$ for the rest of this section. Machina [1982] says that V has a local utility if, for every $F \in \mathcal{F}$, there exists a function $m(\cdot, F) \in C(X)$ such that

$$V(\tilde{F}) - V(F) = \int m(x,F)d(\tilde{F} - F)(x) + o(||\tilde{F} - F||),$$

where $||\cdot||$ is the L_1-norm. This is equivalent to assuming V is *Fréchet differentiable* over $\mathcal{F}$, a strong notion of differentiability.[40]

Our notion of local expected utility is neither weaker nor stronger than Fréchet differentiablility. If V has continuous local expected utility, then it is concave, which is not implied by Fréchet differentiability. Conversely, Example 12 shows that continuous local expected utility does not imply Fréchet differentiability.

Now we discuss the relationship between continuous local expected utility and the weaker notion of *Gâteaux differentiability*, which has been used to extend Machina's notion of local utility to functions that are not necessarily Fréchet differentiable.

In particular, Chew, Karni, and Safra [1987] develops a theory of local utilities for rank-dependent preferences and Chew and Nishimura [1992] extends it to a broader class. Recall that V is Gâteaux differentiable[41] at F if there is a $w(\cdot, F) \in C(X)$ such that

$$\int w(x,F)d\tilde{F}(x) - \int w(x,F)dF(x) = \lim_{\lambda \downarrow 0} \frac{V((1-\lambda)F + \lambda\tilde{F}) - V(F)}{\lambda}.$$

If $w(\cdot, F)$ is the Gâteaux derivative of V at F we can define the directional derivative operator $DV(F)(\tilde{F} - \overline{F}) = \int w(x,F)d\tilde{F}(x) - \int w(x,F)d\overline{F}(x)$. We can restate Lemma 4 with the language of Gâteaux derivatives just introduced.

Proposition 12 (Lemma 4 in Online Appendix II.A)**.** *If V has continuous local expected utility $w(x,F)$, then V is Gâteaux differentiable and $w(\cdot,F)$ is the Gâteaux derivative of V at F, for all F.*

Corollary 7. *V has continuous local expected utility if and only if it is concave and Gâteaux differentiable with continuous Gâteaux derivative.*

[40] The notion of Fréchet differentiability depends on the norm used. Here, following Machina, we use the L_1-norm.

[41] Here we follow Huber [2011] and subsequent authors and adapt the standard definition of the Gâteaux derivative to only consider directions that lie within the set of probability measures.

We conclude by providing an example of a class of preferences that have continuous local expected utility but not a local utility in Machina's sense.

Example 12. Consider a function V with a *Yaari's dual representation*, that is, $V(F) = \int x d(g(F))(x)$ for some continuous, strictly increasing, and onto function $g : [0,1] \to [0,1]$. In addition, assume that g is strictly convex and continuously differentiable, for example $g(t) = t^2$. By Lemma 2 in Chew, Karni, and Safra [1987], V is not Fréchet differentiable, but since $V(F) = \int_0^1 1 - g(F(x))dx$, it is strictly concave in F. Moreover, by Corollary 1 in Chew, Karni, and Safra [1987], V is Gâteaux differentiable with Gâteaux derivative $w(x,F) = \int_0^x g'(F(z))dz$, which is continuous in (x,F). Therefore, by Corollary 7, V has continuous local expected utility and, by Theorem 1, it admits an adversarial forecaster representation. △

References

Aliprantis, C. D. and K. C. Border (2006). *Infinite Dimensional Analysis*. Springer.

Andersen, S. et al. (2018). "Multiattribute utility theory, intertemporal utility, and correlation aversion." *International Economic Review*, 59, 537–555.

Castaldo, A., F. Maccheroni, and M. Marinacci (2004). "Random correspondences as bundles of random variables." *Sankhyā: The Indian Journal of Statistics*, 409–427.

Chew, S. H., E. Karni, and Z. Safra (1987). "Risk aversion in the theory of expected utility with rank dependent probabilities." *Journal of Economic Theory*, 42, 370–381.

Chew, S. H. and N. Nishimura (1992). "Differentiability, comparative statics, and non-expected utility preferences." *Journal of Economic Theory*, 56, 294–312.

Deck, C. and H. Schlesinger (2014). "Consistency of higher order risk preferences." *Econometrica*, 82, 1913–1943.

Eeckhoudt, L. and H. Schlesinger (2006). "Putting risk in its proper place." *American Economic Review*, 96, 280–289.

Galichon, A. (2018). *Optimal transport methods in economics*. Princeton University Press.

Huber, P. J. (2011). "Robust statistics." *International encyclopedia of statistical science*. Springer, 1248–1251.

Kamenica, E. and M. Gentzkow (2011). "Bayesian persuasion." *American Economic Review*, 101, 2590–2615.

Shaked, M. and J. G. Shanthikumar (2007). *Stochastic orders*. Springer.

Stanca, L. (2021). “Recursive preferences, correlation aversion, and the temporal resolution of uncertainty.”

Villani, C. (2021). *Topics in optimal transportation*. Vol. 58. American Mathematical Soc.

Winkler, G. (1988). “Extreme points of moment sets.” *Mathematics of Operations Research*, 13, 581–587.