Randomization, Surprise, and Adversarial Forecasters

Roberto Corrao* Drew Fudenberg† David Levine‡

June 2, 2025

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. We provide several tractable classes of adversarial forecaster preferences. Concavity typically leads the agent 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, RHUL 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, Chris Shannon, Lorenzo Stanca, Tomasz Strzalecki, Jonathan Weinstein, and Alexander Wolitzky for helpful comments and conversations, Sivakorn Sanguanmoo for excellent research assistance. We thank the Gordon B. Pye Dissertation Fellowship, NSF grant SES 1951056 and SES-2417162. MIUR (PRIN 20103S5RN3). and the Leverhulme Trust for financial support.

1 Introduction

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 propose the notion of *continuous local expected utility*, which is a small and relatively tractable departure from expected utility that allows preference for randomization. Continuous local expected utility requires that expected utility is approximately correct for comparing lotteries that are close, and that small changes in the lottery do not change these approximations much. By formulating this condition in terms of supporting hyperplanes we guarantee that utility is concave in probabilities, so our representation captures a preference for surprise.

Although *continuous local expected utility* has the properties we desire, it is not easy to work with. This leads us to introduce the more tractable *adversarial forecaster* model, where the agent enjoys being "surprised," and the surprisingness of an outcome is measured by the minimized error of a fictitious adversary who tries to forecast the outcome in advance. We show that this model is equivalent to continuous local expected utility, and that a lottery is optimal for an adversarial forecaster utility if and only if it maximizes the local utility evaluated at that lottery. This alternative way of describing continuous local expected utility lets us bring our intuitions to bear: it is easier to evaluate what people would consider surprising under particular circumstances than the abstract question of how local utility might be expected to vary with the lottery they choose. 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 (GMM) preferences and transport preferences.

In GMM preferences, the forecast error has a finite-dimensional parameterization. In this case, 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. The bounded support size of GMM representations is not a necessary consequence of continous local utility. In particular, optimal lotteries can have "thick" (i.e. uncountable) support in adversarial forecaster preferences that arise when the agent trades off the interests of different potential selves. Moreover, 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. 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 respect 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. The adversarial forecaster representation imposes more differentiability properties than those models because of its assumption that for each lottery the forecaster has a unique choice that maximizes surprise. These differentiability properties and the concavity of the representation let us characterize optimal lotteries via first-order conditions. When the possible outcomes are an interval of real numbers, Cerreia-Vioglio, Dillenberger, Ortoleva, 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 preferences studied in Quiggin (1982), Green and Jullien (1988), and Galichon and Henry (2012) all have adversarial forecaster representations provided that a supermodularity condition holds. 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.

[1] In the one-dimensional case, monotone preferences of this sort correspond to *ordinally independent* preferences (Green and Jullien (1988)) with concave utility.

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

We study utility functions that are concave and approximately linear, a modest departure from the linearity of expected utility theory. This section defines the relevant notion of continuous local expected utility and introduces an alternative formulation, adversarial forecaster utility, which decomposes utility into an expected utility component and a preference for being surprised. Our main result is that these two formulations are equivalent, so optimal lotteries can be characterized in terms of the relative preferences for surprise. This gives us a powerful tool for constructing examples and classes of examples and analyzing their properties.

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 probability 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* (EU) of V 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 $\int w(x)dF(x) = V(F)$.

Note that if V has a local EU w at F and $\int w(x)dF(x) \geqslant \int w(x)d\tilde{F}(x)$, then $V(F) \geqslant V(\tilde{F})$, so that for the class of lotteries that have lower local EU than F, V ranks F versus alternative lotteries *exactly* according to their local EU. This is an important difference with the notion of local utility in (Machina, 1982), where the local EU only *approximately* ranks lotteries the same way as the original utility V. Below we compare our notion to Machina's more in detail.

Note that if V has a local expected utility at each F then V must be weakly concave, so the induced preferences over lotteries are convex: If the decision maker

[2] See Section 6 for a more detailed discussion of these and other related results.

is indifferent between lotteries F and $\tilde{F}$, they weakly prefer any mixture of the two, i.e they have a preference for randomization.[3] This is a trait for which there is overwhelming evidence (Prelec, 1990), but it is not always consistent with popular models of decision under risk such as expected utility, rank-dependent utility, and prospect theory.

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. As we will show, this yields a tractable representation that can be interpreted as reflecting a taste for surprise.

Definition 1. V has *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 IV, V has continuous local expected utility if and only if it is concave and Gâteaux differentiable with continuous Gâteaux derivative. This is a weaker form of differentiability than in Machina, 1982, but it rules out preferences with a strictly convex utility representation, such as the convex quadratic utility of Chew, Epstein, and Segal, 1991, and even those convex preferences (i.e., quasiconcave utilities) that do not always admit a Gâteaux differentiable representation, such as cautious expected utility in Cerreia-Vioglio, Dillenberger, and Ortoleva, 2015.[4]

The continuous local utility of V at F is a valid Gâteaux derivative for V, which lets us 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 it is concave, and the con-*

[3]This becomes a strict preference for surprise when V is strictly concave, as in some special cases we analyze in Sections 4 and 5. See for example Cerreia-Vioglio, 2009 for a general analysis of convex preferences under risk.

[4]For example, when their maxmin representation is with respect to a finite set of utilities, the representation in Cerreia-Vioglio, Dillenberger, and Ortoleva, 2015 is not Gâteaux differentiable.

tinuous local expected utility is[5]

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

We now introduce a representation where the agent prefers lotteries whose outcomes are difficult to predict, in the sense that even the best prediction has a large expected error. To formalize this, we use the device of a fictitious adversarial forecaster who picks a forecast over outcomes to minimize the expected forecast error given the agent's chosen lottery.

Definition 2. A *forecast* is an element y of a compact metric space Y. A continuous function $\sigma : X \times Y \to \mathbb{R}$ is a *forecast error* if it is non-negative, $\hat{y}(F) := \operatorname{argmin}_{y \in Y} \int \sigma(x,y)dF(x)$ is a singleton for all $F \in \mathcal{F}$, and $\hat{y}(x) := \hat{y}(\delta_x)$ satisfies $\sigma(x, \hat{y}(x)) = 0$ for all $x \in X$.

This definition allows for quite general forecast spaces. Perhaps the simplest case is where $X \subseteq \mathbb{R}$ and Y is the convex hull of X as in Example 1, so that a forecast corresponds to an expected value of x. We also consider the cases when the forecast is on both the mean and variance of x. We also allow fairly general forecast errors; in Example 1 we use the familiar squared error. We normalize the error to be non-negative, and assume that for any lottery F there is a unique forecast $\hat{y}(F)$ that minimizes the expectation of $\sigma(x,y)$ with respect to F.[6] We interpret σ as the loss function of the adversarial forecaster, and as with the typical loss functions in statistics (e.g., Huber (2011)) we require there is a unique optimal forecast for each lottery. Moreover, since it is easy to forecast the outcome of a lottery that assigns probability 1 to a single outcome, we require that the unique minimizing forecast $\hat{y}(x)$ given a degenerate distribution that assigns probability 1 to x has forecast error 0, i.e. $\sigma(x, \hat{y}(\delta_x)) = 0$. We call $\sigma(x, \hat{y}(F))$ the *surprise* of the decision maker at outcome x.

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

[5]The result says there is a unique way to specify a continuous local expected utility function. This does not imply that there is a unique local expected utility at each point; generally, there will be a continuum of local expected utilities at boundary points. With the topology of weak convergence, all points are on the boundary of $\Delta(X)$ when X is infinite.

[6]One sufficient condition is that $Y \subseteq \mathbb{R}^k$ is convex and $\sigma(x,y)$ is strictly convex in y for every x.

$\int \sigma(x,y)dF(x)$. We refer to the minimum value $\Sigma(F) = \min_{y\in Y} \int \sigma(x,y)dF(x)$ as the *suspense* of lottery F; it is also the expected surprise of the agent at lottery F.[7] Let $C(X)$ denote the set of continuous real-valued functions over X.

Definition 3. A function $V : \mathcal{F} \to \mathbb{R}$ is an *adversarial forecaster utility* if

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

for some $v \in C(X)$, forecast space Y, and forecast error function σ.

This representation can be interpreted as follows: The agent has a preference over outcomes described by the baseline 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 lottery's suspense. Equation 1 implies that V is continuous and concave, and that $V(\delta_x) = v(x)$. Note that adversarial forecaster preferences satisfy the independence axiom for comparisons of lotteries that induce the same suspense, but do not do so in general. Note also that these preferences do not need to respect first-order stochastic dominance: As in the next example, the agent might prefer a risky (and hence exciting) option to a sure thing that stochastically dominates it.

Example 1. In a sports match, the outcome is $x = 1$ if the preferred team wins and $x = -1$ if it loses. Here lotteries can be represented by the probability $p \in [0,1]$ that the preferred team wins. Assume that $v(x) = x$ and $\sigma(x,y) = (x-y)^2$, so the forecast error is measured by mean-squared error, where the forecast space is $Y = [-1,1]$. 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 $y = 2p-1$, the expected value of the lottery chosen. With this, the resulting suspense is $4p(1-p)$, the variance of the chosen lottery, and the agent's overall utility over lotteries is represented by $V(p) = 2p - 1 + \gamma 4p(1-p)$. Simple algebra gives that the optimal lottery is $p(\gamma) := \min\{1/2 + 1/\gamma 4, 1\}$: higher preference for surprise (i.e., higher γ) implies lower optimal winning probability for the preferred team. As illustrated in Figure 1, if $\gamma > 1/2$ and the agent can choose any $p \in [0,1]$, the best lottery $p(\gamma)$ is such that the preferred team might lose, while if $\gamma \leqslant 1/2$ the best lottery assigns probability

[7] As in Ely, Frankel, and Kamenica (2015), surprise is a function of realized outcomes and suspense is a measure of uncertainty for the outcome computed before its realization.

one to the preferred team winning the match. △

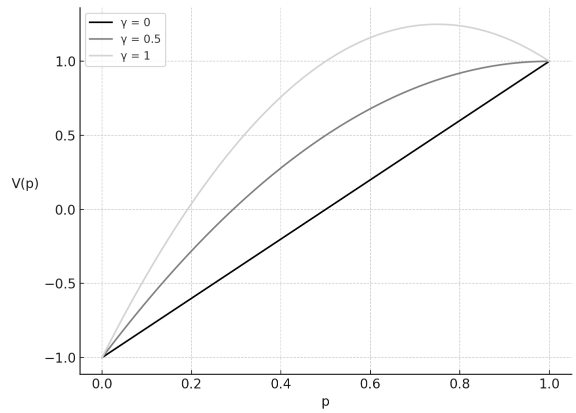

Figure 1: $V(p) = 2p - 1 + 4\gamma p(1-p)$

2.2 Equivalence of the Two Representations

We now show that the two representations developed above are in fact equivalent.

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 all results are in the appendix except where otherwise noted. It is easy to see that if V is an adversarial forecaster representation, then $w(\cdot, F) = v + \sigma(\cdot, \hat{y}(F))$ is a local expected utility of V, and 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 $\sigma(x, \hat{y}(F)) = w(x, F) - v(x)$. Because w is continuous, Y is compact, σ is continuous, and σ attains its minimum value of 0 at degenerate lotteries. Finally, by Proposition 1, $w(\cdot, F)$ is the unique affine function tangent to V at each F, so σ satisfies the uniqueness property.[8]

Theorem 1 shows that, among the preferences that exhibit a taste for randomization, those with a continuous local utility are exactly those that capture a taste for

[8]When X is finite, so that $\mathcal{F}$ is finite-dimensional, the concavity and continuity of V are equivalent to a generalized version of adversarial forecaster utility where the minimum in equation 1 is replaced by an infimum and the uniqueness property is not necessarily satisfied. Corrao, Fudenberg, and Levine (2024) shows that this infimum cannot in general be strengthened to a minimum even when V also satisfies *best outcome independence* (cf. Maccheroni (2002)).

surprise in the sense of our adversarial forecaster model. This is important because it allows us to distinguish them from other quasi-concave utilities over lotteries arise from caution, such as Cerreia-Vioglio, Dillenberger, and Ortoleva, 2015. In other words, while both caution and preference for surprise imply a preference for randomization, the implications in terms of utility representation and optimal choices are different.

3 Implications for Choice

This section illustrates the implications of adversarial forecaster utility with three applications. The first gives a characterization of optimal choices, the second relates the model to models of stochastic choice, and the third examines the timing of information revelation.

3.1 Optimal lotteries

Our analysis below makes extensive use of the following result, which extends the usual first-order condition for maximization to our infinite-dimensional setting. It can be thought of as a "fixed-point" characterization of optimal lotteries, because it shows that an optimal lottery maximizes the local expected utility $v(x) + \sigma(x, \hat{y}(F^*))$ which depends on the chosen lottery F^*.

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

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

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 V has continuous local expected utility.[9] The proposition says that if F^* is optimum, it is also optimal with respect to the expected utility function $w(x, F^*) = v(x) + \sigma(x, \hat{y}(F^*)$. To see how this works, consider Example 1. Here it is

[9]For example, if $X = Y = [-1, 1]$ and $V(F) = \min_{y \in [-1,1]} \int_{-1}^{1} (2y - 1) x dF(x)$, then $F^* = \delta_0$ is 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 $y = 0$.

easy to see that the two degenerate lotteries δ_{-1} and δ_1 do not satisfy this condition when $\gamma > 1/2$. 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 $p(\gamma) = 1/2 + 1/\gamma 4$.

Proposition 2 can be directly used to link the shape of the forecast error function (or equivalently the local utility) to whether the decision maker randomizes, and if so, how.

Corollary 1. *Suppose that X is a compact and convex subset of an Euclidean space and V has continuous local expected utility w.*

1. *If $w(x, F)$ is strictly quasiconcave in x for every $F \in \mathcal{F}$, then any optimal lottery F^* over $\mathcal{F}$ is degenerate.*

2. *If $w(x, F)$ is strictly quasiconvex in x for every $F \in \mathcal{F}$, then any optimal lottery F^* over $\mathcal{F}$ is supported on the extreme points $ext(X)$ of X.*

This result follows from the fact that F^* is optimal if and only if $\text{supp}(F^*) \subseteq \text{argmax}_{x \in X} w(x, F^*)$. Therefore, when each $w(x, F)$ is strictly quasiconcave, each candidate optimal lottery must be supported on the single maximizer of the local utility at that lottery. Similarly, when $w(x, F)$ is strictly quasiconvex, each candidate optimal lottery must be supported on the extreme points of X. We know that risk averse individuals have a preference for degenerate lotteries, and risk loving individuals for extremal points. This generalizes to quasi-concavity provided the local expected utility functions have that property. In the strictly quasi-convex case, the solution can be degenerate, as in the next example.

Example 2. Here we extend the sport-match preferences of Example 1 by allowing risk-averse (CARA) baseline preferences. We set $X = Y = [-1, 1]$, $v(x) = (1 - \exp(-\lambda x))/\lambda$ with $\lambda > 0$, and $\sigma(x, y) = \gamma(x - y)^2$ with $\gamma > \lambda/2$. The local utility at any lottery F is $w(x, F) = v(x) + \gamma(x - m_F)^2$, and, because $\min_{x \in X} v''(x) = -\lambda$ and $\lambda < 2\gamma$, each local utility is strictly convex in x. From Corollary 1, a lottery F^* is optimal only if it is supported on -1 or 1. In Figure 2, we plot the local utilities $w(x, F)$ for $\lambda = \gamma = 1$ at the degenerate distributions over -1, 1, and at the lottery

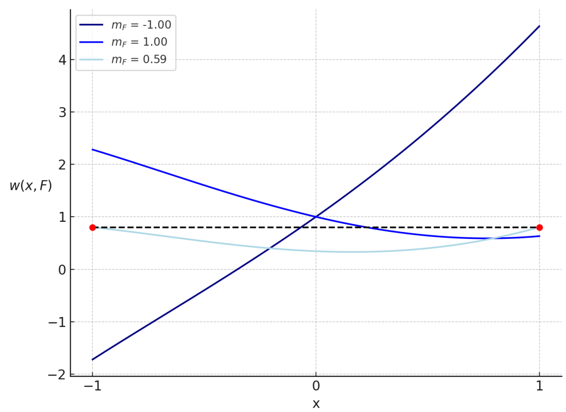

Figure 2: $w(x,F) = (1-\exp(-x)) + (x-m_F)^2$

over $\{-1,1\}$ with average $m_F = 0.59$.[10] The unique solution to this maximization is

$$q^*(t) = \begin{cases} -1 & \text{if} \quad t \leqslant \frac{1-\overline{q}^*}{2} \\ 1 & \text{if} \quad t > \frac{1-\overline{q}^*}{2}. \end{cases} \tag{3}$$

where $\overline{q}^* = \min\{r(\lambda)/\gamma, 1\}$ is the mean of the optimal lottery and where $r(\lambda) = (\exp(\lambda) - \exp(-\lambda))/4$. This corresponds to the binary lottery on $\{-1,1\}$ that assigns probability $\frac{1+\overline{q}^*}{2}$ to 1. The variance of this lottery is $1-(\overline{q}^*)^2$; when $\overline{q}^* < 1$, the variance is $1-(r(\lambda)/\gamma)^2$, which is decreasing in λ and increasing in γ: agents with lower baseline risk aversion and more taste for surprise are willing to sacrifice more expected value for higher variance. In Figure 2, $r(\lambda)/\gamma = 0.59$ and the corresponding local utility is indifferent between -1 and 1. $\triangle$

3.2 Stochastic Choice

To relate the adversarial forecaster representation to stochastic choice, suppose that the agent has a finite set $\tilde{\mathcal{F}} \subset \mathcal{F}$ of feasible lotteries, and can implement any randomization over these lotteries, so they can choose any lottery over outcomes given by $F_\lambda = \sum_{\tilde{F}\in\tilde{\mathcal{F}}} \tilde{F}\lambda(\tilde{F})$ for $\lambda \in \Delta(\tilde{\mathcal{F}})$. We assume that the agent reduces compound

[10] δ_{-1} cannot be optimal because $w(-1,\delta_{-1}) = (1-\exp(\lambda))/\lambda < (1-\exp(-\lambda))/\lambda + 4\gamma = w(1,\delta_{-1})$. δ_1 is optimal if and only $w(1,\delta_1) = (1-\exp(-\lambda))/\lambda \geqslant (1-\exp(\lambda))/\lambda 4\gamma = w(0,\delta_1)$, which is equivalent to $r(\lambda)/\lambda \geqslant \gamma$, where $r(\lambda) = (\exp(\lambda) - \exp(-\lambda))/4$. In Figure 2, we have $\lambda = \gamma = 1$, so $r(\lambda)/\lambda = 0.59 < 1 = \gamma$.

lotteries and so only cares about the final lottery F_λ. The concavity of the adversarial forecaster representation then implies that each compound lottery F_λ is weakly preferred to at least one lottery in the support of λ. This preference is strict when V is strictly concave.[11] Machina (1985) pointed out via examples that intrinsic preferences for randomization, as modeled by convex preferences, generate stochastic optimal choices without the shocks to utility or information. In a setting with real-valued outcomes, Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella, 2019 characterizes deliberately stochastic choices generated by preferences that are convex and monotone with respect to first-order stochastic dominance (FOSD). In particular, it shows that a necessary condition for deliberately stochastic choice is a violation of the regularity property that enlarging the choice set cannot increase the probability of pre-existing alternatives.[12]

The next example shows how a preference for surprise can also lead to violations of regularity with real-valued outcomes.

Example 3. Suppose that $X = Y = [-1, 1]$, that the agent's baseline utility is $v(x) = x$, , and that the agent's preference for surprise is $\sigma(x, y) = (x - y)^2$. As in Example 2, the continuous local utility of V is $w(x, F) = v(x) + (x - m_F)^2$, where $m_F = \int_{-1}^{1} \tilde{x} dF(\tilde{x})$. 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, when the set of feasible lotteries is $\Delta\big(\{-1, 0\}\big)$, the unique optimal choice is δ_0, so there is no suspense. In contrast, when the feasible lotteries are $\Delta\big(\{-1, 0, 1\}\big)$, 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.[13]

$\triangle$

Some classes of adversarial forecaster preferences do satisfy regularity. This is true for example of the weak APU of Fudenberg, Iijima, and Strzalecki (2015). The

[11]See Proposition 4 for a class of strictly concave adversarial forecaster representations.

[12]Recall that convex preferences correspond to value functions that are quasi-concave, and that a stochastic choice function $P : \mathcal{X} \to \Delta(X)$, where $\mathcal{X} \subseteq 2^X$ is the collection of feasible menus, satisfies *regularity* if $P(x|\overline{X}) \leqslant P(x|\overline{X}')$ for all $x \in \overline{X}' \subseteq \overline{X}$.

[13]Note that any binary lottery with $p\delta_1 + (1-p)\delta_{-1}$ is strictly preferred to a point mass at 0 provided that $p > (3 - \sqrt{5})/4$. Thus the example satisfies the sufficient condition for the failure of regularity in Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella, 2019 Theorem 2.

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 $F(x)$ denotes the probability mass function of F and 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 $w(x, F^*) = u(x) - c'(F^*(x))$.[14]

4 The Bounds of Optimal Randomization

This section analyzes the extent of optimal randomization in a class of adversarial forecaster models called *generalized method of moments*, which is based on the idea that the forecaster makes a prediction by targeting certain moments of the lottery. In a sense, these are the simplest examples of non-linear preferences with continuous local expected utility, because they are always quadratic. We first study the case where the adversarial forecaster only cares about a finite number of moments and show that the extent of optimal randomization is bounded by the number of moments. We then show that as the number of moments grows to infinity, the extent of randomization can increase to the point where optimal lotteries randomize over the entire space of outcomes.

4.1 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 the adversary's goal is to match the collection of moments of F given by $h(x, s)$.

Definition 4. The loss function σ is based on the *generalized method of moments* (GMM)[15] if there is a finite probability space (S, μ) and a continuous function h :

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

[15] 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.

$X \times S \to \mathbb{R}$ such that

$$\sigma(x,y) = \sum_{s \in S} (h(x,s) - y(s))^2 \mu(s). \tag{4}$$

Proposition 4 below shows that any loss function σ based on the generalized methods of moments is a forecast error, and moreover the associated suspense is quadratic. 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 $s_j \in S$, yielding the standard method of moments.[16] The simplest case is $X \subseteq \mathbb{R}$ and $S = \{1\}$, as in Examples 1 and 3.

4.2 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 focus on the extent of optimal randomization, that is, 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

$$\mathcal{F}_\Gamma = \left\{ F \in \Delta\left(\overline{X}\right) : \forall g_i \in \Gamma, \int g_i(x)\, dF(x) \leqslant 0 \right\}, \tag{5}$$

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.[17] 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 at most $k+1$ 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 3. *When the agent has GMM utility with m moments and Γ contains*

[16] See for example Chapter 18 in Greene (2003).

[17] Equality constraints can be incorporated in (5) by considering both $g_i(x)$ and $-g_i(x)$.

k moment restrictions, there is an optimal lottery that puts positive probability on at most $m+k+1$ points. Moreover, if X is finite, then all optimal lotteries put positive probability on at most $m+k+1$ points.

The proof of the first statement is relatively simple, so we present it here. First, given the forecast error in equation 4, 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

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

Now fix an optimal solution $\overline{y}^*$ of the outer maximization problem.[18] F^* solves the original problem and is consistent with $\overline{y}^*$ if and only if it solves

$$\max_{F\in\mathcal{F}_\Gamma:\hat{y}(F)=\overline{y}^*} \int \left\{ v(x) + \sum_{s\in S} \left(h(x,s) - \overline{y}^*(s)\right)^2 \mu(s) \right\} dF(x) \tag{6}$$

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 (6) 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 $\overline{X}$, as claimed.[19]

Section 4.4 introduces a class of adversarial forecaster representation that generalizes GMM which a generalization of Proposition 3 holds (see Theorem 3 in Appendix A).

[18] An optimizer $\overline{y}^* \in \overline{Y}$ exists because the function

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

is upper semicontinuous by Berge Maximum theorem.

[19] Winkler's result is reported in Theorem 7 in Online Appendix I.B.

4.3 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 before, the forecast space is the compact set $Y = \{h(F, \cdot) \in C(S) : F \in \mathcal{F}\}$,[20] and the forecast error is

$$\sigma(x, y) = \int (h(x, s) - y(s))\, d\mu(s). \tag{7}$$

We now show that σ based on the generalized methods of moments is a forecast error, and moreover, the associated suspense is quadratic.

Proposition 4. *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

When $X \subseteq \mathbb{R}$, the generalized moments $h(x, s) = \exp(sx), -s_0 \leqslant s \leqslant s_0$ with $s_0 > 0$, correspond to the moment generating function and so induce a one-to one mapping. Proposition 4 implies that the GMM preference $V(F)$ induced by this class is strictly concave, thereby exhibiting a strict preference for randomization. For example, Theorem 2 in Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella (2019) implies that the stochastic choice induced by these preferences is in general non-degenerate. Moreover, because strict concavity of V is inconsistent with EU,

Chew, Epstein, and Segal (1991) show that 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

[20] The Arzelà–Ascoli theorem implies that Y is compact: 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 continuous.

negative certainty independence (NCI) only if they are expected utility preferences.[21] Therefore, when V is induced by a GMM forecast error and its continuous local utility is not constant, the corresponding preference does not satisfy NCI. This is intuitive, because NCI corresponds to a preference for certain outcomes, which is the opposite of a preference for surprise.

When a GMM utility has infinitely many moments, we call H its *kernel*. 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 the one-dimensional case and do not impose exogenous moment restrictions on the feasible lotteries.

Proposition 5. *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 optimal lottery, 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 9 in Online Appendix III.[22] The proof uses Proposition 4 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 Proposition 5; its proof is in Online Appendix I

Corollary 2. *Maintain the assumptions of Proposition 5, 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 at most $|S^n| + 1$ 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 every outcome is in the support of the optimal lottery.

[21] NCI says that if the agent prefers a lottery to a certain outcome, this ranking is not reversed by mixing each option with a third lottery.

[22] In Example 6 below thick support arises a different sort of adversarial forecaster preference.

4.4 Parametric Adversarial Forecaster and Randomization

For GMM preferences, the forecast space is the set of generalized moments, $\prod_{s\in S} h(\mathcal{F}, s)$. 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 and let us relax the symmetric loss function of the GMM case.

Definition 5. A forecast error σ is *parametric* if $Y \subseteq \mathbb{R}^m$ for some finite integer m, and σ is continuously differentiable in y. A function $V : \mathcal{F} \to \mathbb{R}$ is a *parametric adversarial forecaster utility* 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.[23]

Example 4. This example relaxes the GMM representation by allowing the forecaster to have different preferences regarding positive and negative surprises. Let $X = [0,1]$, set $Y = X$, and 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 $z > 0$. The utility function

$$V(F) = \int_0^1 v(x)dF(x) + \min_{y\in Y} \int_0^1 \rho(x-y)dF(x), \tag{8}$$

arise from the parametric adversarial forecaster representation with forecast error $\sigma(x,y) = \rho(x-y)$. Here $\hat{y}(\hat{F})$ is the unique minimizer in (8), and the suspense function is $\Sigma(F) = \int \rho(x - \hat{y}(F))dF(x)$ which can be interpreted as an index of the dispersion of F, without requiring symmetry. As we show in Section 6.2, this can lead to more "prudent" preferences than those induced $\rho(z) = z^2$. △

Example 5. Proposition 7 in Fudenberg, Iijima, and Strzalecki (2015) shows that V

[23] It is sufficient that $w(x,F) = \tilde{w}(x, P(F))$ for some continuous functions $P : \mathcal{F} \to Y$ and $\tilde{w} : X \times Y \to \mathbb{R}$ with Y is a compact finite-dimensional set and $\tilde{w}$ continuously differentiable in y.

has an APU representation if and only if it has an AVU representation, that is,

$$V(F) = \sum_{x \in X} u(x)F(x) + \min_{y \in \mathbb{R}^X} \sum_{x \in X} \left[y(x) + \sum_{\tilde{x} \in X} \phi(y(\tilde{x})) \right] F(x) \tag{9}$$

where $\phi(z) := c^*(-z)$ and c^* is the convex conjugate of the original cost function c. If c' is bounded and$\min_{r \in \mathbb{R}} (r + \phi(r)) = 0$,[24] we can restrict the minimization in (9) to a compact subset of $\mathbb{R}^X$ 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 9 is an example of a parametric adversarial forecaster utility where the parameter space Y has dimension $m = |X|$. In the spirit of the nested logit model, we generalize the APU representation by considering uncertain taste shocks $y \in \mathbb{R}^X$ that are the same across certain classes of outcomes in X, reducing the dimensionality of the parameter space.[25] Fix a partition $\mathcal{P} = \{E_1, ..., E_m\}$ of X and a compact interval $I \subseteq \mathbb{R}$ that contains 0 and is large enough that the solution of $\min_{r \in I} pr - \phi(r)$ is in the interior of I. Let Y be the subset of I^X of vectors that are measurable with respect to the fixed partition, and let $V(F)$ be as in (9) with $\mathbb{R}^X$ replaced by Y. Then for every partition, the utility function V has an adversarial forecaster representation. △

Theorem 3 in Appendix A extends the support bounds of Proposition 3 to parametric adversarial preferences that are not GMM, as in Example 4. For the asymmetric GMM case of Example 4, we show 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 5, 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 3 yields a meaningful bound on the support of optimal lotteries. Because all solutions must satisfy the upper bound when X is finite, our result gives a testable prediction on the support of stochastic choices induced by AVU preferences with perfectly correlated shocks.[26]

[24] 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 $\phi(r) = r^2/2 - r$.

[25] Nested logit divides items into groups, with correlated value shocks within each group. Here we consider the case where items within the same group have perfectly correlated shocks.

[26] Online Appendix II.B provides an extension to the case of infinite X.

4.5 Application: Preference for surprise and the timing of information

We now study the implications of adversarial forecaster preferences for the timing of 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) (EFK) has an adversarial forecaster representation.[27] EFK assumed that the agent does not directly care about the outcome itself; our approach makes it easy to model agents that directly care about the outcome.

To relate the adversarial forecaster representation to stochastic choice, suppose that the agent has a finite set $\tilde{\mathcal{F}} \subset \mathcal{F}$ of feasible lotteries, and can implement any randomization over these lotteries, so they can choose any lottery over outcomes given by $F_\lambda = \sum_{\tilde{F} \in \tilde{\mathcal{F}}} \tilde{F} \lambda(\tilde{F})$ for $\lambda \in \Delta(\tilde{\mathcal{F}})$. We assume that the agent reduces compound lotteries and so only cares about the final lottery F_λ. The concavity of the adversarial forecaster representation then implies that each compound lottery F_λ is weakly preferred to at least one lottery in the support of λ. This preference is strict when V is strictly concave.[28] the outcomes as well as surprise.

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. Following EFK, we assume that the agent has preference for suspense in both periods, and assume that the preference for first-period suspense is $V_1(F) = g(E(F))$ for $E(F) = \int_0^1 \frac{1}{2}||p - p_F||^2 dF(p) = \int_0^1 p^2 dF(p) - p_F^2$ and 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

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

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

[28]See Proposition 4 for a class of strictly concave adversarial forecaster representations.

Finally, we generalize EFK so that 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).[29]

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:

$$w_\beta(p,\omega,F) = \omega\tilde{v} + (1-\beta)g'(D_2(F))(p^2 - p_F^2) + \beta g(p - p^2), \tag{10}$$

where $D_2(F) = \int \tilde{p}^2 dF(\tilde{p}) - p_F^2$. As we show in Proposition 11 in Online Appendix II.D, when β is near 1 (so the agent mostly cares about second-period surprise) the optimum is to reveal no information in the first period so the set of interim beliefs is a singleton, and when β is near 0, so the agent mostly carers about first-period surprise, the state is fully revealed then. Finally, for intermediate values of β the optimum can have 3 different interim beliefs (see Online Appendix III).

5 Transport Utilities: A multi-self model of preferences for surprise

This section considers a tractable class of adversarial forecaster utilities that can generate 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 dual Kantorovich transport problem (hence the name) and analyze it using results from the optimal transport literature. This lets us give a simple sufficient condition for optimal lotteries to have thick support, and provide a detailed analysis of the one-dimensional case.

[29] 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.

5.1 Definition of Transport Utility

As in the infinite-dimensional GMM case of Section 4.3, we let the forecast space be infinite-dimensional, a key to obtaining the strict optimality of lotteries with thick (i.e., infinite) supports. Formally, we let the outcome space X be a compact and convex finite-dimensional set, and take a forecast to be a continuous real function y over X, which we regard as a *score function*. For example, the score $y(x)$ of outcome x can represent the forecaster's estimate of the likelihood of x in the form of the logarithm of a density of x.

If we defined the surprise of the outcome x as $\max_\xi y(\xi) - y(x)$, that is, the score difference between the outcome with the highest ex-ante score and that realized outcome, the adversary could reduce the surprise to 0 by choosing a constant y. Instead, we consider a decision maker with multiple selves that have heterogeneous preferences over outcomes. We index the selves by $\theta \in \Theta \equiv X$, 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 self θ at outcome x is $\max_{\xi \in X} (y(\xi) - \phi(\theta, \xi)) - (y(x) - \phi(\theta, x))$.

Notice that for any particular self θ the forecaster can send the forecast $y(x) = \phi(\theta, x)$ so that θ has a uniform utility-adjusted forecast and is not surprised by anything. Instead, we assume that the selves θ are uniformly distributed over X, 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,

$$V(F) = \int v(x) dF(x) + \inf_{y \in C(x)} \int \hat{\sigma}(x, y) dF(x) \tag{11}$$

where

$$\hat{\sigma}(x, y) = \int \left(\max_{\xi \in X} (y(\xi) - \phi(\theta, \xi)) - (y(x) - \phi(\theta, x)) \right) dU(\theta) \tag{12}$$

is the expected value of the score-adjusted surprises of the multiple selves with respect to the uniform measure U.[30] We say that $V(F)$ is a *transport utility* if it satisfies 11

[30] The results extend to any measure that can be represented by a density. When the distribution

for some $v(x)$ and $\phi(\theta, x)$ 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.

5.2 Adversarial Forecaster Representation

Transport utilities 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 valid surprise function. To do this, we define Y^* to be the K-Lipchitz real-valued functions on X, 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 $y^* \in Y^*$.[31] We then define the forecast space Y to be the strongly ϕ-concave functions y in $C(X)$ that satisfy the normalization $\int \exp(y(x)) dU(x) = 1$, and let σ denote the restriction of $\hat{\sigma}$ to $X \times Y$.[32]

Theorem 2. *The function σ is a forecast error, $\Sigma(F) = \min_{y \in Y} \int \sigma(x, y) dF(x)$ is the corresponding suspense function, and the utility function V in equation 11 is an adversarial forecaster utility.*

This result follows from Lemma 9 and Lemma 10 in Appendix B, which show that Y and σ satisfy all the properties of a forecast space and a forecast error function.

5.3 The Primal Representation and Optimal Lotteries

As indicated, transport preferences are linked to the Kantorovich optimal transportation problem through duality theory, which lets us give a simple sufficient condition for optimal lotteries to have thick support. First, we establish the basic duality result:

Lemma 1. *Suspense is the solution to choosing a probability measure $T \in \Delta(\Theta \times X)$ to solve the problem*

$$\Sigma(F) = \max_{T \in \Delta(U,F)} \left(\int \int \phi(\theta, x) dU(\theta) dF(x) - \int \phi(\theta, x) dT(\theta, x) \right) \tag{13}$$

where $\Delta(U, F)$ is the set of joint distributions T such that $\int T(\theta, x) d\theta = F(x)$ and $\int T(\theta, x) dx = U(\theta)$.

---

of θ has mass points, the uniqueness property required by adversarial forecaster preferences can fail.

[31] Sinceϕ is continuously differentiable on $X \times X$, it is Lipschitz continuous.

[32] This normalization is needed to bound the space of forecasts. It is consistent with the interpretation of the score as the logarithm of a density function.

We use this duality result in Theorem 9 in Appendix B to derive the set of optimal lotteries under general transport utilities. Lemma 1 also helps us find conditions on ϕ that make transport utilities strictly concave and so exhibit a strict preference for randomization. We say that ϕ satisfies the *twist condition* if the map $\theta \mapsto \nabla_x \phi(\theta, x)$ is injective for all $x \in X$. For example, the twist condition is satisfied in the one-dimensional case $X \subseteq \mathbb{R}$ when ϕ is twice differentiable with $\phi_{x\theta} < 0$, a condition that we exploit in Corollary 4 below.

Corollary 3. *If ϕ satisfies the twist condition, the transport utility $V(F)$ in equation 11 is strictly concave.*

Because finite-dimensional parametric adversarial forecaster utilities are not strictly concave when X is infinite, this corollary shows there is no intersection between them and the transport utilities that satisfy the twist condition.[33] Strict concavity also differentiates transport utilities with the twist condition from other models in the literature with a maxmin representation such as Maccheroni, 2002 and Cerreia-Vioglio, Dillenberger, and Ortoleva, 2015, which are not strictly concave in general.

5.4 The One-Dimensional Case

Let $X \subseteq \mathbb{R}$ and, for every lottery F, let $q_F(t) = F^{-1}(t)$ denote its the quantile function, where $F^{-1}(t) := \inf \{x \in X : t \leqslant F(x)\}$ denotes the generalized inverse of F. Each $q_F(t)$ is nondecreasing and left-continuous. Conversely, for every nondecreasing and left-continuous q, the function $F_q(x) = \sup \{t \in [0,1] : x \geqslant q(t)\}$ is a CDF: it is nondecreasing, right-continuous, and equal 1 at the largest point of X. Moreover, $F_q(x)$ is the unique CDF such that $F_q(x) \in q^{-1}(x)$ for all $x \in X$. As shown in Appendix B, this lets us find optimal lotteries by maximizing over the corresponding quantile functions.[34]

Corollary 4. *Suppose that $X \subseteq \mathbb{R}$ is an interval and that $\phi_{\theta x} < 0$. A lottery $F \in \mathcal{F}$ maximizes $V(F)$ if and only if*

$$q_F(t) \in \operatorname*{argmax}_{x \in X} \left\{ v(x) - \phi(q_U(t), x) + \int \phi(q_U(z), x) dz \right\} \tag{14}$$

[33] Finite-dimensional parametric adversarial forecaster utilities are not strictly concave when X is infinite because a finite set of parameters is not sufficient to pin down a unique element of $\mathcal{F}$, and $V(\lambda F + (1-\lambda)\tilde{F}) = \lambda V(F) + (1-\lambda)V(\tilde{F})$ for all lotteries $F, \tilde{F}$ such that $\hat{y}(F) = \hat{y}(\tilde{F})$.

[34] Online Appendix II.C characterizes optimal lotteries for general transport preferences.

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

Example 6. 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. Consider first a GMM utility with $Y = [-1,1]$ and $\sigma(x,y) = \gamma(x-y)^2$ as in Example 2. In this case, the adversarial forecaster utility function is

$$V(F) = \mathbb{E}_F[v(x)] + \gamma \text{Var}_F(x) \tag{15}$$

where $\text{Var}_F(x)$ is the variance of F. 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 $\overline{q}_F = \int_0^1 q_F(t)dt$. 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 Proposition 2 must be 0. When $\gamma > 1/2$ then every local utility is strictly convex, so Proposition 6 in the next section implies that the optimal lottery is supported on $\{-1,1\}$. Moreover, 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.

Now consider the transport utility induced by the multiple-selves utility function $\phi(\theta,x) = -\gamma\theta x$. Lemma 11 implies that the corresponding adversarial forecaster utility function is $V(F) = \mathbb{E}_F[v(x)] + \gamma \max_{T \in \Delta(F,U)} \text{Cov}_T(\theta,x)$, where $\Delta(F,U)$ is the set of joint distributions over $X \times \Theta$ with marginals F and U and $\text{Cov}_T(\theta,x)$ is the covariance between x and θ under T.

Corollary 4 says that the quantile function $q_{F*}(t)$ of the optimal lottery solves

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

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 depends on γ and has thick (i.e. uncountable support for all $\gamma \in (0,1)$.[35] △

[35] When γ is less than 2/3 the optimum has no mass points, as $\gamma \to 1$ the probability assigned to mass points goes to 1.

6 Monotonicity and behavior

Preferences that preserve stochastic orders capture the idea that individuals prefer lotteries that are better according to the stochastic order. 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, a preference that preserves the stochastic order generated by concave functions will exhibit risk aversion. Here we give necessary and sufficient conditions for the adversarial forecaster model to preserve a stochastic order, and give applications to absolute and relative risk aversion of various orders. Online Appendix V shows how our results help characterize aversion to correlation in risks across time periods.

6.1 Stochastic orders and monotonicity

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

Definition 6. Fix a set $\mathcal{W} \subseteq C(X)$.

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

$$F \gtrsim_{\mathcal{W}} \tilde{F} \iff \int w(x)dF(x) \geqslant \int w(x)d\tilde{F}(x) \quad \forall w \in \mathcal{W}. \tag{17}$$

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

Let $\langle \mathcal{W} \rangle$ denote the smallest closed convex cone containing $\mathcal{W}$. Because the adversarial forecaster utility has a max-min representation with local utility at F given by $v + \sigma(\cdot, \hat{y}(F))$ and this coincides with the Gâteaux derivative of V at F, we can apply Theorem 1 in Cerreia-Vioglio, Maccheroni, and Marinacci (2017) to obtain the following characterization.[36]

Proposition 6. *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 $v + \sigma(\cdot, \hat{y}(F)) \in \langle \mathcal{W} \rangle$ for all $F \in \mathcal{F}$.*

[36] Theorem 1 and Lemma 1 in Cerreia-Vioglio, Maccheroni, and Marinacci, 2017 are stated under the assumption that $X \subseteq \mathbb{R}$. However, an inspection of their proof shows that the same results hold for any compact metric space X. Therefore, we omit the proof of Proposition 6.

This result implies that for adversarial forecaster preferences to be consistent with FOSD and SOSD order when X is a subset of the reals, it is sufficient that, for all $F \in \mathcal{F}$, the function $v + \sigma(\cdot, \hat{y}(F))$ is respectively increasing and increasing and concave. This result can also be used to check whether the adversarial forecaster representation favors mean-preserving spreads; in this case it is enough to check V preserves the convex order. For example, in Example 3, when $v'' \geqslant -2$, the local utility is convex in x for all forecasts F. Thus Proposition 6 implies that the agent weakly prefers any mean-preserving spread $\tilde{F}$ of F to F.

Now we apply Proposition 6 to the transport utilities introduced in Section 5. Given $X \subseteq \mathbb{R}$, 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 $\gtrsim_{\mathcal{W}}$ if and only if, for all $F \in \mathcal{F}^*$, $w_0(x, F) \in \langle \mathcal{W} \rangle$, where*

$$w_0(x, F) = v(x) + \int \phi(\theta, x) dU(\theta) - \int_{x_F}^{x} \phi_x(T^{-1}(z), z) dz,$$

$x_F = \min \operatorname{supp} F$, and $T^{-1}(z)$ is the generalized inverse of the primal solution $T(\theta) = q_F(U(\theta))$.

Under the maintained 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 6 then yields Corollary 5.

The corollary gives easy-to-check conditions on v and ϕ such that the transport utility is consistent with a stochastic order. For example, when v is convex and ϕ is convex in x and submodular in (θ, x), we have that $w_0''(x, F) \geqslant 0$ for all F, implying that V preserves the convex order preferring mean-preserving spreads of an arbitrary lottery to the lottery itself. We discuss this more in detail in the next section.

6.2 Risk aversion and prudence under adversarial forecasters

To see that preference for surprise can alter the agent's risk preference, consider a parametric adversarial asymmetric forecaster utility with $X = Y = [0, 1]$ as in

Example 4 with loss function $\rho(z)$, and a baseline utility function $v(x)$:

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

Theorem 3 in Appendix A 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 $w''(x,F) = v''(x) + \rho''(x - \hat{y}(F))$, Proposition 6 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 Proposition 2 can be used to explicitly compute p^*.

Suppose in particular that $\rho(z) = \lambda \exp(z) - \lambda z$ for some $\lambda > 0$. The local expected utility is $w(x,F) = v(x) + \lambda \exp(x - \hat{y}(F)) - \lambda(x - \hat{y}(F))$, where $\hat{y}(F) = \log\left(\int_0^1 \exp(x)dF(x)\right)$, 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 $\hat{y}(F)$.

Similarly, preference for surprise can alter the agent's higher-order risk preference. 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 6, 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.[37] 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. Online Appendix VI briefly reviews higher-order risk aversion and explains how it is affected by a preference for surprise.

[37] 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 $\tilde{v}^{(n)} = \max_{x \in X} |v^{(n)}(x)|$.

6.3 Relative risk aversion and adversarial forecasters

In this section, we assume that X is a compact interval of real numbers and compare the risk attitudes of an adversarial forecaster utility $V(F)$ and the baseline expected utility $v(x) = V(\delta_x)$ that comes from ignoring the suspense term of V. We first recall the notion of relative risk attitudes introduced in Chew, Karni, and Safra (1987) for non-EU preferences.

Definition 7. For all $v \in C(X)$ and $F, \tilde{F} \in \mathcal{F}$, we say that F is a *simple compensated spread* of $\tilde{F}$ with respect to v if $\int v(x)dF(x) = \int v(x)d\tilde{F}(x)$ and there exists $x_0 \in X$ such that $F(x) \geqslant \tilde{F}(x)$ for all $x < x_0$ and $F(x) \leqslant \tilde{F}(x)$ for all $x \geqslant x_0$. A continuous utility V is *more risk loving* than a continuous expected utility v if $V(F) \geqslant V(\tilde{F})$ whenever F is a *simple compensated spread* of $\tilde{F}$ with respect to v.

In other words, F is a *simple compensated spread* of $\tilde{F}$ with respect to v if an expected-utility agent with utility v is indifferent between these two lotteries and F increase in the upper tail and decreases in the lower tail with respect to $\tilde{F}$. We can use Definition 7 to compare the risk attitudes of an adversarial forecaster utility with those of an agent with the same baseline utility v but with no preference for surprise. This isolates the role of the forecast error σ in the agent's attraction for risk.

Corollary 6. *Fix an adversarial forecaster utility V with representation v and σ. Then V is more risk loving than 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*

$$\sigma(x, \hat{y}(F)) = \phi_F(v(x)). \tag{18}$$

This result follows from combining Proposition 3 in Cerreia-Vioglio, Maccheroni, and Marinacci, 2017 and our Proposition 6. The former implies that V is more risk loving than v if and only if V preserves $\gtrsim_{\mathcal{W}_v}$ where $\mathcal{W}_v$ is the set of functions $\phi(v(x))$ where $\phi(t)$ is continuous and convex. By Proposition 6 this is equivalent to $v + \sigma(\cdot, \hat{y}(F)) \in \mathcal{W}_v$ for all $F \in \mathcal{F}$, which is equivalent to 18. A sufficient condition for equation 18 is that the baseline utility v is strictly increasing and concave and that the surprise function is increasing and convex in x.[38] The next example applies Corollary 6 to GMM preferences.

[38] To see this, observe that v^{-1} is strictly increasing and convex when v is strictly increasing and concave. If in addition σ is increasing and convex in x, we can rewrite $\sigma(x, y) = \sigma(v^{-1}(v(x)), y)$, so 18 is satisfied by the continuous and convex function $\phi_F(t) = \sigma(v^{-1}(t), \hat{y}(F))$.

Example 7. Consider the GMM preferences $V(F) = \int v(x)dF(x) + \lambda \min_{y \in Y} \int (v(x) - y)^2 dF(x)$, with $Y \equiv v(X)$ and $\lambda \geqslant 0$. Here the adversarial forecaster tries to predict the realized utility of the agent, so $\sigma(x, \hat{y}(F)) = \lambda \left(v(x) - \int v(\tilde{x})dF(\tilde{x})\right)^2$, which satisfies (18). Thus for every $\lambda > 0$ the adversarial forecaster utility V is relatively more risk-loving than the baseline expected utility v. △

7 Conclusion

Adversarial forecaster preferences arise naturally in many settings. They allow the interpretation of random choice as a preference for surprise, and also allow sharp characterizations of the optimal "amount" (i.e., support size) of randomization and of various monotonicity properties. Ongoing work considers 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 lets us consider cases where the adversary has only finitely many actions or where the loss function has kinks, as it does with an absolute-deviation forecast error. This more general 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, 3, and 4

Here we prove the main results in Sections 2 and 4. The proofs of the ancillary results that are first stated in this section are in Online Appendix I.A.

Lemma 2. *If V has continuous local expected utility $w(x, F)$, then for all $F, \tilde{F} \in \mathcal{F}$:*

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

Proof. (1) This is immediate, as by definition $\int w(x, F)dF(x) = V(F) \leqslant \int w(x, \tilde{F})dF(x)$ for all $F, \tilde{F} \in \mathcal{F}$.

(2) 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, $\int w(x,F)d\overline{F}(x) \geqslant V(\overline{F})$ so

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

Similalrly, since $w(x,\overline{F})$ is a local utility function at $\overline{F}$, $\int w(x,\overline{F})dF(x) \geqslant V(F)$, so

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

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

as $\lambda \to 0$, since $w(x,\overline{F})$ is continuous in $\overline{F}$. Putting these together yields

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

which yields the statement. ■

Lemma 3. *Let V have continuous local expected utility $w(x,F)$. For all $F, \tilde{F}, \overline{F} \in \mathcal{F}$ such that there exists $\mu > 0$ with $F + \mu(\tilde{F} - \overline{F}) \in \mathcal{F}$,*

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

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 2. ■

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 3 implies that $D(F, (\delta_x - F)) = w(x, F) - \int w(x, F)dF(x) = w(x, F) - V(F)$, where the second equality follows from the properties of $w(x, F)$. This implies that $D(F, (\delta_x - F))$ is well-defined and continuous and that $w(x, F) = V(F) + D(F, (\delta_x - F))$ as desired. ∎

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

(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, Y is closed, bounded, and equicontinuous, so it is compact. For all $y = w(\cdot, F)$ and $x \in X$, define $v(x, y) = w(x, F)$ and observe that it is continuous. For all $F \in \mathcal{F}$ and for all $\tilde{y} \in Y$, $V(F) = \int w(x, F)dF(x) \leqslant \int v(x, \tilde{y})dF(x)$, 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 $V(F) = \min_{y \in Y} \int v(x, y)dF(x)$.

It remains to show that $\int v(x, y)dF(x)$ has a unique minimum over $y \in Y$. Suppose that for some F there is a $\tilde{y} \neq \hat{y}(F)$ such that $V(F) = \int v(x, \tilde{y})dF(x)$. By the definition of Y, there exists $\tilde{F} \in \mathcal{F}$ such that $v(\cdot, \tilde{y}) = w(\cdot, \tilde{F})$. For every $\lambda \in [0, 1]$, define $F_\lambda = \lambda \tilde{F} + (1 - \lambda)F$. Then, for all $\lambda \in [0, 1]$

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

where the first inequality follows from concavity of V, the second inequality because $w(x, \tilde{F})$ is a local utility of V, the first equality by the definition of F_λ, and the last equality because $V(F) = \int w(x, \tilde{F})dF(x)$. Thus

$$V(\tilde{F}) + (1 - \lambda)V(F) = V(F_\lambda) = \int w(x, \tilde{F})dF_\lambda(x). \tag{19}$$

Next, fix $\mu \in (0, 1)$. By rearranging terms in (19),

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

Conversely, because V is concave $V(\tilde{F}) \leqslant \int w(x, F_\mu) d\tilde{F}(x)$. Together with the line above this implies

$$V(\tilde{F}) = \int w(x, F_\mu) d\tilde{F}(x). \tag{20}$$

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

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

where the first equality follows by (20), the second equality by Lemma 3, the inequality by the properties of w, the third equality by (19), and the last equality by the properties of w. 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 3, the inequality by the properties of w, and the third and the last equality by (20). 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 in the limit: $w(\tilde{x}, F) = w(\tilde{x}, \tilde{F}) = v(\tilde{x}, \tilde{y})$. Given that $\tilde{x}$ was arbitrary, the minimizer is unique, which proves that V is an adversarial expected utility representation that satisfies uniqueness.

Proof of Proposition 2. (If). For all $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} \int v(x) + \sigma(x, \hat{y}(F^*)) dF(x)$ and $F \in \overline{\mathcal{F}}$, $V(F^*) = \int v(x) + \sigma(x, \hat{y}(F^*)) dF^*(x) \geqslant \int v(x) + \sigma(x, \hat{y}(F^*)) dF(x) \geqslant V(F)$, where the first equality and last inequality follow from the definition of continuous local utility and the fact that $v(x) + \sigma(x, \hat{y}(F))$ is a continuous local utility of V, and the first inequality follows by assumption. This implies that $F^* \in \text{argmax}_{F\in\overline{\mathcal{F}}} V(F)$.[39] (Only if). Fix an optimal lottery F^* for V over $\overline{\mathcal{F}}$ and assume that there exists $\hat{F} \in \overline{\mathcal{F}}$ that is strictly better than F^* for an expected utility agent with utility $v+\sigma(\cdot,\hat{y}(F^*))$. 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 the directional derivative of V at F^* in direction $\hat{F}$ is negative, which, by Lemma 3 and Theorem 1, contradicts $\hat{F}$ strictly preferred to F^* for expected utility function $v + \sigma(\cdot, \hat{y}(F))$. ■

Proof of Corollary 1. By Proposition 2, F^* maximizes $V(F)$ over $\mathcal{F}$ if and only if $F^* \in \text{argmax}_{F\in\mathcal{F}} \int v(x)+\sigma(x,\hat{y}(F^*))dF(x)$, that is, if and only if $x \in \text{argmax}_{\tilde{x}\in X} v(\tilde{x})+\sigma(\tilde{x},\hat{y}(F^*))$ for all $x \in \text{supp}(F^*)$. Assume that $w(x,F)$ is strictly quasiconcave in x for all $F \in \mathcal{F}$ and assume by contradiction that $x, x' \in \text{supp}(F^*)$ with $x \neq x'$. The set $\text{argmax}_{\tilde{x}\in X} v(\tilde{x}) + \sigma(\tilde{x},\hat{y}(F^*))$ must be a singleton and therefore x and x' cannot be both optimal, contradicting the optimality of F^*. Next assume that $w(x,F)$ is strictly quasiconvex in x for all $F \in \mathcal{F}$ and assume by contradiction that $x \in \text{supp}(F^*)$ such that $x \notin \text{ext}(X)$. Because $w(x,F)$ is strictly quasiconvex, there exists $x' \in textext(X)$ such that $w(x,F) < w(x',F)$, implying that $x \notin \text{argmax}_{\tilde{x}\in X} v(\tilde{x})+\sigma(\tilde{x},\hat{y}(F^*))$, which contradicts the optimality of F^*. ■

Proof of Proposition 4. The result follows from the following three lemmas. The first two are standard and are proved in Online Appendix II.A. Recall that here we allow the set S to be any compact metric space.

Lemma 4. *Let Y be a compact set of a Euclidean space. The function $\sigma(x,y)$ defined in equation 7 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 5. *Let $H(x,\tilde{x}) = \int h(x,s)h(\tilde{x},s)d\mu(s)$. Then*

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

[39] See Proposition10 in Online Appendix II.A for an alternative proof that can also be applied to the more general adversarial expected utility model.

The directional derivatives $DV(F, \delta_z - F)$ for directions $(\delta_z - F)$ at F are

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

When $F \mapsto h(F, \cdot)$ is one-to-one we have an additional property:

Lemma 6. *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 5 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

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

Hence for $V(F)$ to be strictly convex it suffices that $\int h(x,s)dM(x) \neq 0$ for any signed measure M with $\int dM(x) = 0$. From the Jordan decomposition, $M = \lambda(F - \tilde{F})$ 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 1 to 1, this implies $F = \tilde{F}$ and $M = 0$. ■

Now we extend the support bounds of Proposition 3 to parametric adversarial preferences. 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 $\overline{Y} = \hat{y}(\overline{\mathcal{F}})$. The same steps as for GMM show that $\max_{F\in\overline{\mathcal{F}}} V(F) = \max_{\overline{y}\in\overline{Y}} \max_{F\in\overline{\mathcal{F}}:\hat{y}(F)=\overline{y}} \int v(x) + \sigma(x,\overline{y})dF(x)$.[40] 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 $\hat{y}(F) = \overline{y}^*$.

Theorem 3. *Let V be a parametric adversarial forecaster utility. 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 an optimal lottery for V over $\overline{\mathcal{F}}$ that assigns positive probability to at most $(k+1)(m+1)$ points of $\overline{X}$.*

[40]As in the GMM case, a maximizer exists because the function $R(\overline{y}) = \max_{F\in\overline{\mathcal{F}}:\hat{y}(F)=\overline{y}} \int v(x) + \sigma(x,\overline{y})dF(x)$ is upper semicontinuous and $\overline{Y}$ is compact because the function $\hat{y}$ is continuous.

The proof of Theorem 3 uses the following two results. Let $\mathcal{H}$ denote the set of probability measures over Y. Let $ext(\overline{\mathcal{F}})$ denote the extreme points of any convex and compact $\overline{\mathcal{F}} \subseteq \mathcal{F}$. Let $\Lambda_F \subseteq \Delta\left(ext\left(\overline{\mathcal{F}}\right)\right)$ be the probability measures over extreme points that satisfy $F = \int \tilde{F} d\lambda(\tilde{F})$.[41]

Lemma 7. *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 $\tilde{F} \in \bigcup_{\lambda \in \Lambda_{\hat{F}}} \operatorname{supp} \lambda$, $V(\hat{F}) = \int\int u(x,y)\, d\tilde{F}(x)\, d\hat{H}(y)$.*

Now fix a closed subset $\overline{X} \subseteq X$ and a finite collection of functions $\Gamma = \{g_1, ..., g_k\} \subset C(\overline{X})$, and consider $\mathcal{F}_\Gamma(\overline{X}) \subseteq \mathcal{F}$. By Theorem 7 in Online Appendix I.B (cf. 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 (Theorem 7 in Online Appendix). Recall that $\hat{Y}(F) \equiv \operatorname{argmin}_{y \in Y} \int u(x,y) dF(x)$.

Theorem 4. *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 $\hat{Y}(F)$ is a singleton for all $F \in \mathcal{F}$. Then:*

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.

[41] This set is non-empty from Choquet's theorem.

Proof of 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,[42] 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 Lemma 7, 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 Lemma 7, we then have $|\operatorname{supp} \lambda| \leqslant m+1$. Therefore, F_w is the linear combination of at most $m+1$ extreme points in $\mathcal{E}$. Each $\tilde{F} \in \mathcal{E}$ is supported on at most $k+1$ points of $\overline{X}_{\mathcal{E}}$, so F_w is supported on at most $(m+1)(k+1)$ points of $\overline{X}_{\mathcal{E}}$.

Proof of 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, F solves $\max_{F \in co(\mathcal{E})} V_0(F)$, concluding the proof. ■

Lemma 8. *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 $(m+1)(k+1)$*

[42] See e.g. Guillemin and Pollack (2010).

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 3. Fix a parametric adversarial forecaster representation $(Y, v, \hat{\sigma})$, and define $u = v + \sigma$. By Definition 5, 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 4 and Lemma 8, there exists a solution F^* that is supported on no more than $(k+1)(m+1)$ points of $\overline{X}$. ∎

Proof of Proposition 5. Because $H(x, \tilde{x}) = G(x - \tilde{x})$, it follows that $H(x, x) = G(0)$ is constant, so the directional derivatives from Lemma 5 simplify to

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

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

$$\int H(z,x)dF(x) \geqslant \int H(x,\tilde{x})dF(x)dF(\tilde{x}) \text{ for all } z \in X. \tag{21}$$

This implies the complementary slackness condition: if there exists $z \in A$ such that z satisfies (21) with strict inequality, then $F(A) = 0$.[43]

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 $\int H(0,x)dF(x)) = \int H(x,\tilde{x})dF(x)dF(\tilde{x})$ 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 $\left\{x' \in X : \int H(x',x)dF(x) > \int H(x,\tilde{x})dF(x)dF(\tilde{x})\right\}$, 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, i.e. $\int H(x,\tilde{x})dF(x)dF(\tilde{x}) > 0$, implies that $H(a,x) > 0$ for an non-empty interval $x \in [a,b]$. Since $H(0,\tilde{x})$ is decreasing and $H(a,a) = \max_{\tilde{x}} H(a,\tilde{x})$,

[43] 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 (21) 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.

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 hold on a positive measure subset of A^c, 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 $(a+b)/2$. ■

Appendix B: Sections 5 and 6

We start with the lemmas that prove Theorem 2 , and then prove Lemma 10, Corollary 3, Theorem 9, Corollaries 4 and 5, and Propositions **??** and **??**. In the proofs of this appendix, we make extensive use of standard results on optimal transport. All of these results are restated in Online Appendix I.B.

Lemma 9. *The function $\hat{\sigma}$ defined in equation 12 is non-negative, continuous, and such that, for all $x \in X$, there exists $y \in C(X)$ with $\hat{\sigma}(x,y) = 0$ and $\int \exp(y(\xi))dU(\xi) = 1$.*

Proof. 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

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

■

Lemma 10. *The set Y is compact and, for every $F \in \mathcal{F}$, we have*

$$\Sigma(F) := \min_{y\in C(X)} \int \hat{\sigma}(x,y)dF(x) = \inf_{y\in C(X)} \int \hat{\sigma}(x,y)dF(x). \tag{22}$$

Moreover, the minimization problem in (22) *has a unique solution in* Y*.*

Proof. The strongly ϕ-concave functions are K-Lipschitz so Y is equicontinuous. Together with the constraint $\int \exp(y(\xi))dU(\xi) = 1$ this implies that Y is totally bounded, so 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 $y^n(x) = -\max_{\theta \in X}(y^{*n}(\theta) - \phi(\theta, x))$. 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-Lipschitz, there is a subsequence $y^{*n} \to y^*$ that is also K-Lipschitz. Convergence and continuity imply that $y(x) = -\max_{\theta \in X}(y^*(\theta) - \phi(\theta, x))$, so y is strongly ϕ-concave and Y is closed.

We next show that equation 22 has 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

$$\sup_{y^* \in C(X), y \in C(X)} \left(\int y^*(\theta)dU(\theta) + \int y(x)dF(x) \right) : y^*(\theta) - y(x) \leqslant \phi(\theta, x) \quad \forall(\theta, x). \tag{23}$$

For every feasible pair (y^*, y), we have $-y^*(\theta) \geqslant y(x) - \phi(\theta, x)$, so $-y^*(\theta) \geqslant \max_{\xi} y(\xi) - \phi(\theta, \xi)$. This means that if the alternative problem has a solution y the original problem has the same solution. The alternative problem is the dual of the Kantorovitch transport problem and we draw upon results from that literature.

Proposition 7 in Online Appendix I.B (cf. Proposition 1.11 in Santambrogio (2015)) shows that because X is compact and ϕ is continuous, Problem 23 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 8 in Online Appendix I.B (cf. 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 and Y is normalized, equation 22 has exactly one solution in Y. ■

Proof of Theorem 2. This follows immediately from Lemma 9 and Lemma 10. ■

Proof of Lemma 1. Fix a continuous function $\varphi(\theta, x)$ and $F \in \mathcal{F}$.Theorem 5 in Online Appendix I.B (cf. Theorem 1.39 in Santambrogio (2015)), usually called the Duality Theorem of Optimal Transport, says that the value of the problem $\min_T \int \varphi(\theta, x) dT(\theta, x)$ subject to $\int_X T(\theta, x) d\theta = F(x)$ and $\int_X T(\theta, x) dx = U(\theta)$ is equal to the value of the dual

$$\max_{y \in C(X)} \left(\int y(x) dF(x) + \int \min_{x \in X} \left(\varphi(\theta, x) - y(x) \right) dU(\theta) \right),$$

and that both problems have solutions. To connect this to transport preferences, define $\varphi(\theta, x) = \phi(\theta, x) - \int \phi(\tilde{\theta}, \tilde{x}) dU(\tilde{\theta}) dF(\tilde{x})$ and observe that the suspense function $\Sigma(F)$ can be rewritten as

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

The duality theorem yields $\Sigma(F) = \int\int \phi(\theta, x) dU(\theta) dF(x) - \min_T \int \phi(\theta, x) dT(\theta, x)$. ■

Proof of Corollary 3. Lemma 1 implies that $V(F) = \int v(x) dF(x) + \Sigma(F)$ where Σ is defined by equation 13. Because the marginal over Θ is the uniform distribution and ϕ satisfies the twist condition, Proposition 9 in Online Appendix I.B (cf. Proposition 7.19 in Santambrogio (2015)) implies that $\Sigma(F)$ is strictly concave, so $V(F)$ is strictly concave. ■

Before proving Corollary 4, we state and prove an ancillary lemma.

Lemma 11. *If $X \subseteq \mathbb{R}$ and the partial derivative $\phi_x(\theta, x)$ is decreasing in θ, then*

$$\Sigma(F) = \int\int \phi(q_U(t), q_F(z)) dt dz - \int_0^1 \phi(q_U(t), q_F(t)) dt. \tag{24}$$

Proof. By Lemma 1,

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

where $\hat{\phi}(\theta, x) = \int \phi(\tilde{\theta}, x) dU(\tilde{\theta}) - \phi(\theta, x)$. Because ϕ_x is decreasing in θ, and U is atomless, Theorem 6 in Online Appendix I.B (cf. Theorem 4.3 in Galichon (2018)) implies 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 $\hat{\phi}(\theta, x)$. ■

Proof of Corollary 4. From equation 24 the problem of maximizing $V(F)$ becomes

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

This immediately implies that if $F \in \mathcal{F}$ is such that $q_F(t)$ is a maximizer of problem (14) for all $t \in [0, 1]$, then F is optimal for V. Conversely, assume that F is optimal for V and, by contradiction, that $q_F(t_0)$ is not a maximizer of problem (14) at t_0. Next, let $\hat{q}(t)$ be defined as the pointwise minimum of the argmax correspondence of problem (14). By Lemma 17.30 in Aliprantis and Border, 2006 this function is lower semicontinuous, and by Theorem 4' in Milgrom and Shannon, 1994 it is non-decreasing. This implies that $\hat{q}(t)$ is the quantile function of a lottery $\hat{F} \in \mathcal{F}$. This implies that $V(\hat{F}) > V(F)$, yielding a contradiction. ■

Proof of Proposition ??. Fix φ twice continuously differentiable with $\varphi_{tx} > 0$ and define $\phi(\theta, x) = -\varphi(\theta, x)$ and $v(x) = \int_0^1 \varphi(\theta, x) dU(\theta)$. The induced transport utility is:

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

where the second equality follows from Lemma 1, the third equality from the definitions of $v(x)$ and $\phi(\theta, x)$, and the third equality by Theorem 6 in Online Appendix I.B (cf. Theorem 4.3 in Galichon (2018)). This yields the first part of the statement. The second part of the statement follows from the same steps as above by defining $\varphi(\theta, x) = v(x) - \phi(x, \theta) + \int \phi(x, \theta) dU(\theta)$. ■

Proof of Proposition ??. Define $\phi(x, \theta) = -(\theta \circ x)$ and $v(x) = (\int \theta dU(\theta) \circ x)$. The transport utility induced by $\phi(\theta, x)$ and $v(x)$ so defined is:

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

where the second equality follows from Lemma 1 and the third equality from the definitions of $v(x)$ and $\phi(\theta, x)$. This yield the statement. ■

Proof of Proposition 6. In Proposition 13 in Online Appendix IV, we show that 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 $w(x, F) = v(x) + \sigma(x, \hat{y}(F))$ 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 Corollary 5. First, recall from the proof of Theorem 1 that for every $V(F)$ with continuous expected utility, the local utility is $w(x, F) = v(x) + \sigma(x, \hat{y}(F))$. Theorems 2 and 1 imply that the suspense function of $V(F)$ is

$$\Sigma(F) = \min_{y \in y} \int \sigma(x, F) dF(x) = \max_{T \in \Delta(U,F)} \int \hat{\phi}(\theta, x) dT(\theta, x), \tag{25}$$

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 $T(\theta) = q_F(U(\theta))$ of the minimization problem in (25) satisfies $\frac{\partial}{\partial x}\hat{y}(F)(x) = \hat{\phi}_x(T^{-1}(x), x)$ for all $x \in \operatorname{supp} F$. Thus there is a constant $k(F)$ such that

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

The continuous local utility of V is $w(x, F) = v(x) + \sigma(x, \hat{y}(F)) + k(F) := w_0(x, F) + k(F)$. Thus by 6, for every set $\mathcal{W} \subseteq C(X)$, V preserves $\gtrsim_{\mathcal{W}}$ if and only if $w(\cdot, F) \in \langle \mathcal{W} \rangle$ for all F, which is equivalent to $w_0(\cdot, F) \in \langle \mathcal{W} \rangle$ for all F. ■

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.

Corrao, R., D. Fudenberg, and D. K. Levine (2024). "On concave functions over lotteries." *Journal of Mathematical Economics*, 110, 102936.

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

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.

Galichon, A. and M. Henry (2012). "Dual theory of choice with multivariate risks." *Journal of Economic Theory*, 147, 1501–1516.

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.* Amer. Math. 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.

Loseto, M. and A. Lucia (2021). "Deliberate randomization under risk." *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.

Milgrom, P. and C. Shannon (1994). "Monotone comparative statics." *Econometrica: Journal of the Econometric Society*, 157–180.

Prelec, D. (1990). "A "Pseudo-endowment" effect, and its implications for some recent nonexpected utility models." *Journal of Risk and Uncertainty*, 3, 247–259.

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.

Online Appendix I: Ancillary results

This appendix gives proofs of the ancillary results stated in the main appendix and states some useful results from optimal transportation theory.

Online Appendix I.A: Ancillary results for Appendix A

Lemma 4. Let Y be a compact set of a Euclidean space. The function $\sigma(x, y)$ defined in equation 7 is a forecast error.

Proof of Lemma 4. 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,

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

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

Lemma 5. 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

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

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

Proof of Lemma 5. 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

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

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

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

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

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

We next state and prove a more general version of Lemma 7 that considers 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 $V^*\left(\overline{\mathcal{F}}\right) = \max_{F \in \overline{\mathcal{F}}} V(F)$.

By Sion's minmax theorem,

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

Lemma 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 Lemma 7. 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

$$\int\int u\left(x,y\right)d\hat{F}\left(x\right)d\hat{H}\left(y\right)\geqslant\int\int u\left(x,y\right)d\tilde{F}\left(x\right)d\hat{H}\left(y\right)\qquad\forall\tilde{F}\in ext\left(\overline{\mathcal{F}}\right).\tag{26}$$

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

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

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

which contradicts equation (26).
Conversely, fix $\tilde{F} \in ext\left(\overline{\mathcal{F}}\right)$ and observe that the implication follows by

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

Before proving Lemma 8, we state and prove an intermediate result.

Lemma 12. *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 $\Delta(\mathcal{F}_{e})$, which implies the statement. ■

Lemma 8. 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 8. Let $\hat{F}$ solve $\max_{F \in \mathcal{F}_\Gamma(\overline{X})} V(F)$. By Lemma 12, 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 $\mathcal{E}^n \subseteq \text{ext}(\mathcal{F}_\Gamma(\overline{X}))$. By Theorem 4, 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 2. Maintain the assumptions of Proposition 5, 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 at most $m^n + 1$ 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, μ^n) satisfies the assumptions of Theorem 3 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 $\max_{F \in \Delta(X)} V^n(F)$ that is supported on at most $m_n + 1$ points of X. Because the constraint set $\Delta(X)$ is compact and V is continuous, all the accumulation points of the sequence F^n are solutions of the problem $\max_{F \in \Delta(X)} V(F)$, where V is the GMM adversarial forecaster representation induced by h and μ. Proposition 5 established that this problem has a unique full-support solution F, so F is the unique accumulation point of F^n. Because X is compact, the sequence $\operatorname{supp} F^n$ converges to some set $\hat{X} \subseteq X$ in the Haussdorf sense. By Box 1.13 in Santambrogio (2015), $F^n \to F$ implies that $\operatorname{supp} F \subseteq \hat{X}$, and so $\operatorname{supp} F^n \to \overline{X}$ because $\operatorname{supp} F = X$. ∎

Online Appendix I.B: Theorems cited in the main appendix

Proposition 7 (Proposition 1.11 in Santambrogio, 2015)**.** *Suppose that X and Y are compact, $c(x, y)$ is a continuous function, and $\mu \in \Delta(X)$ and $\nu \in \Delta(Y)$. Then there exists a solution $(\varphi^*, \psi^*) \in C(X) \times C(Y)$ to the optimal transport dual problem*

$$\max_{\varphi \in C(X), \psi \in C(Y),\ \textit{such that}\ c(x,y) \geqslant \varphi(x)+\psi(y)} \int \varphi d\mu(x) + \int \psi d\nu(x) \tag{27}$$

where φ is c-concave with respect to φ^* and $\varphi^*(y) = \min_{x \in X} (c(x, y) - \varphi(x))$.

Theorem 5 (Theorem 1.39 in Santambrogio, 2015)**.** *Suppose that X and Y are Polish spaces and that $c : X \times Y \to \mathbb{R}$ is uniformly continuous and bounded. Then the problem in (27) admits a solution and its value is equal to $\Gamma_c(\mu, \nu) := \min_{\pi \in \Delta(\mu,\nu)} \int c(x, y) d\pi(x, y)$.*

Proposition 8 (Proposition 7.18 in Santambrogio, 2015)**.** *Assume that $X = Y$ is the closure of a bounded connected open set of $\mathbb{R}^n$, that $c : X \times X \to \mathbb{R}$ is continuously differentiable, and that at least one of the probability measures $\mu, \nu \in \Delta(X)$ is supported on the whole X. Then the solution $\varphi^* \in C(X)$ in Proposition 7 is unique up to additive constants.*

Proposition 9 (Proposition 7.19 in Santambrogio, 2015)**.** *Under the same assumptions of Proposition 8, if in addition, ν is absolutely continuous with respect to the Lebesgue measure and c satisfies the twist condition, then $\Gamma_c(\mu, \nu)$ is strictly convex in μ.*

Theorem 6 (Theorem 4.3 in Galichon, 2018)**.** *Assume that X and Y are compact intervals in the real line, that $c(x, y)$ is strictly submodular, and fix $\mu \in \Delta(X)$ and $\nu \in \Delta(Y)$ such that ν has no mass points. Then the primal problem $\Gamma_c(\mu, \nu)$ admits a unique solution and this solution is deterministic and equal to $T(y) = q_\mu(F_\nu(y))$, where q_μ is the quantile function associated with μ and F_ν is the CDF associated with ν.*

Theorem 7 (Theorem 2.1 in Winkler, 1988)**.** *Let $(X, \mathcal{X})$ be a measurable space and let $\overline{\mathcal{F}} \subseteq \mathcal{F}$ be a simplex of probability measures whose extreme points are Dirac measures. Fix measurable functions $g_1, ..., g_n$ over X and real number $c_1, ..., c_n$. Consider the set*

$$\mathcal{H} = \left\{ F \in \overline{\mathcal{F}} : \forall i \in \{1, ..., n\}, g_i \textit{ is } F\textit{-integrable and } \int g_i(x) dF(x) \leqslant c_i \right\}.$$

Then $\mathcal{H}$ is convex and each of its extreme points is supported on up to $n+1$ points of X.

Online Appendix II: Optimization

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

Online Appendix II.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 $\tilde{F} \in \overline{\mathcal{F}}$,

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

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

(i) implies (ii). Fix $F^* \in \text{argmax}_{F\in\overline{\mathcal{F}}} V(F)$. 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 $w^* \in \text{argmin}_{w\in co(\mathcal{W})} R(w)$. Observe that

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

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

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

where the last equality follows from the Sion minmax theorem because $\overline{\mathcal{F}}$ is compact and convex. Thus $F^* \in \text{argmax}_{F\in\overline{\mathcal{F}}} \int w^*(x)dF(x) = \text{argmax}_{F\in\overline{\mathcal{F}}} \int\int u(x,y)dH(y)dF(x)$.
■

Online Appendix II.B: Robust solutions

This section shows that the finite-support property of Theorem 3 holds generically for optimal lotteries for a parametric adversarial forecaster V over $\overline{\mathcal{F}}$ that are "robust" in the following sense. For every $F \in \mathcal{F}_\Gamma(\overline{X})$, we call a sequence as in Lemma 12 a *finitely approximating sequence of F*.

Definition 8. 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 8. *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 uses the following lemma.

Lemma 13. *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 $\hat{w}^n \to w_{|\hat{X}}$. Fix $n \in \mathbb{N}$ large enough that $B_{1/n}(\hat{x}) \cap B_{1/n}(\hat{x}') = \varnothing$ for all $\hat{x}, \hat{x}' \in \hat{X}$.[44] 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. ■

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

Proof of Theorem 8. Without loss of generality, assume that $\overline{X} = \bigcup_{F \in \mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F$.[45] 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 $\mathcal{E}^n \uparrow \overline{\mathcal{E}}$. By construction, $\overline{X}_{\mathcal{E}^n} \uparrow \overline{X}$.[46] 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 4. By Lemma 13 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

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

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

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

Because $w_{|\overline{X}_{\mathcal{E}^n}} \in \hat{\mathcal{W}}^n$ for every $n \in \mathbb{N}$, Theorem 4 implies that F^n is supported on at most $(k+1)(m+1)$ points of $\overline{X}_{\mathcal{E}^n}$, and because $F^n \to F^*$, it follows that F is supported on at most $(k+1)(m+1)$ points of $\overline{X}$. Given that F^* and w were arbitrarily chosen, the result follows. ■

7.1 Online Appendix II.C: Optimal lotteries under general transport preferences

Lemma 1 can be used to solve the problem of choosing a lottery $F \in \mathcal{F}$ when V is a transport utility. Define the correspondence

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

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

[46] This follows from the fact that $\overline{X} = \bigcup_{F \in \mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F$ by assumption. See also footnote 45.

For every measurable selection $\psi \in \Psi_\phi$ and measurable set $\tilde{X}$, let $U^\psi(\tilde{X}) = U(\psi^{-1}(\tilde{X}))$. In what follows, we let $\Delta(U, F)$ denote the set of joint probability measures with marginals U and F.

Theorem 9. *If V is a transport utility with respect to v and ϕ, the set of optimal lotteries over $\mathcal{F}$ is the closure of $\left\{U^\psi \in \mathcal{F} : \psi \in \Psi_\phi\right\}$. Moreover, if $\Psi_\phi = \psi$ is single-valued, then the unique optimal lottery is U^ψ and its support is $\psi(\Theta)$.*

Proof. By Lemma 1,

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

This implies that $F \in \text{argmax}_{\tilde{F} \in \mathcal{F}} V(F)$ 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. 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 $A^{G_\phi^c} = \{\theta \in \Theta : \exists x \in A, (\theta, x) \in G_\phi\}$. Therefore, $F \in \text{argmax}_{\tilde{F} \in \mathcal{F}} V(\tilde{F})$ is equivalent to $0 \geqslant \sup\left\{F(A) - U(A^{G_\phi^c}) : A \subseteq X \quad \text{is closed}\right\}$, which is equivalent to $U(\Psi_\phi^\ell(A)) \geqslant F(A)$ for all closed $A \subseteq X$, where $\Psi_\phi^\ell(A) = \{\theta \in \Theta : \Psi_\phi(\theta) \cap A \neq \varnothing\}$ 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 $U(\Psi_\phi^\ell(A)) \geqslant F(A)$ holds for all measurable sets $A \subseteq X$.

We have shown that F is optimal if and only if $F(A) \leqslant U(\Psi_\phi^\ell(A))$ for all measurable A, i.e. $\text{argmax}_{F \in \mathcal{F}} V(F) = \left\{F \in \mathcal{F} : F(A) \leqslant U(\Psi_\phi^\ell(A)) \text{ for all measurable } A\right\}$. And 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. ■

Online Appendix II.D: Formal result from Section 4.5

Proposition 11. *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 it 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* $\max_{p\in[0,1]}\{p\tilde{v} + (1-\beta)g'(p-p^2)(p-p^2)\}$.

Before proving Proposition 11 we introduce some additional notation. For every $F \in \mathcal{F}$, define $\xi_{\beta,F} : [0,1] \to \mathbb{R}$ as $\xi_{\beta,F}(\tilde{p}) = (1-\beta)g'\left(D_2(F)\right)\tilde{p}^2 + \beta g(\tilde{p} - \tilde{p}^2)$ and let $cav(\xi_{\beta,F})$ denote its concavification. Also, observe that Proposition 2 implies that $F^* \in \text{argmax}_{F\in\overline{\mathcal{F}}} V_\beta(F)$ if and only if $F^* \in \text{argmax}_{F\in\overline{\mathcal{F}}} \int w_\beta(x, F^*)dF(x)$.

Proof of Proposition 11. To show there is an optimal lottery with support on at most three points, 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:

$$\max_{F_\Delta \in \Delta(p_F^*, q^*)} \int g(p-p^2)dF_\Delta(p). \tag{28}$$

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 28, so any solution F_Δ of Problem 28 is also a solution of the original problem, and $\Delta(p_F^*, q^*)$ is a moment set with 2 moment conditions. The objective function of Problem 28 is linear in F_Δ, so it follows from Theorem 2.1. in Winkler (1988) that there is solution of Problem 28, and hence of the original problem, that is supported on no more than three points.

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 10, (p^*_F, F^*_Δ) solve

$$
\begin{aligned}
&\max_{p\in\overline{\Delta}, F_\Delta\in\Delta([0,1]):\int\tilde{p}dF(\tilde{p})=p} p\tilde{v} + (1-\beta)g'\left(D_2(F^*)\right)\int(\tilde{p}^2 - p^2)dF_\Delta(\tilde{p}) \\
&+ \beta\int g(\tilde{p}-\tilde{p}^2)dF_\Delta(p) \\
&= \max_{p\in\overline{\Delta}} p\tilde{v} - (1-\beta)g'\left(D_2(F^*)\right)p^2 \\
&+ \max_{F_\Delta:\int\tilde{p}dF(\tilde{p})=p}\left[\int(1-\beta)g'\left(D_2(F^*)\right)\tilde{p}^2 + \beta g(\tilde{p}-\tilde{p}^2)dF_\Delta(\tilde{p})\right] \\
&= \max_{p\in\overline{\Delta}}\left\{p\tilde{v} - (1-\beta)g'\left(D_2(F^*)\right)p^2 + cav(\xi_{\beta,F^*})(p)\right\}
\end{aligned}
\tag{29}
$$

Given the assumptions on g and 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 29 is uniquely solved by $F_\Delta = \delta_p$, so $F^*_\Delta = \delta_{p^*_F}$. Because $p\tilde{v} - (1-\beta)g'\left(D_2(F^*)\right)p^2 + \xi_{\beta,F^*}(p) = p\tilde{v} + \beta g(p - p^2)$ 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 29 is uniquely solved by $F_\Delta = (1-p)\delta_0 + p\delta_1$. Here $cav(\xi_{\beta,F^*})(\tilde{p}) = (1-\beta)g'\left(D_2(F^*)\right)\tilde{p}$, so $F^*_\Delta = (1-p^*_F)\delta_0 + p^*_F\delta_1$. Because $p\tilde{v} - (1-\beta)g'\left(D_2(F^*)\right)p^2 + cav(\xi_{\beta,F^*})(p) = p\tilde{v} + (1-\beta)g'(D_2(F^*))(p-p^2)$, the statement follows. ■

Online Appendix III: Additional examples

This section presents two examples. In the first one, 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 optimal lottery under the asymmetric adversarial forecaster preferences of Section 6.2.

Example 8. Given the optimization problem, we need to maximize the function $V(F) = 0.5V_1(F) + 0.5V_2(F)$ over all distributions F in $\mathcal{F}$, where $\mathcal{F}$ is the set of all distributions over $[0,1]$ with no more than three points in their support. The function $V(F)$ is a weighted sum of two components: 1. $V_1(F) = ((F))^{1/2}$, the square root of the variance of F. 2. $V_2(F) = \sum_{i=1}^{3} \sqrt{x_i - x_i^2} p_i$, the expectation of $\sqrt{x - x^2}$ with respect to F. We represent F as a discrete distribution with probabilities p_1, p_2, p_3 and corresponding values x_1, x_2, x_3 (each $x_i \in [0, 1]$). The constraints are:

1. $0 \leqslant p_i \leqslant 1$ for $i = 1, 2, 3$.
2. $\sum_{i=1}^{3} p_i = 1$.
3. $0 \leqslant x_i \leqslant 1$ for $i = 1, 2, 3$.

The optimal solution F numerically found is $\{(0.124; 1/4), (0.146; 1/4)(0.146; 1/2)\}$. The maximum value of $V(F)$ is approximately 0.354. These values represent the distribution F within $\mathcal{F}$ that maximizes $V(F) = 0.5V_1(F) + 0.5V_2(F)$, under the given constraints.

Example 9 (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 $H(0, \tilde{x}) = \exp -\tilde{x}$, which is non-negative, strictly decreasing and strictly convex. △

Online Appendix IV: 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 12. *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.[47]

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 10 shows that continuous local expected utility does not imply Fréchet differentiability.

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

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[48] 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 2 with the language of Gâteaux derivatives just introduced.

Proposition 13 (Lemma 2 in Appendix 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.*

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 10. Consider a function V with a *Yaari's dual representation*, that is, $V(F) = \int xd(g(F))(x)$ 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. △

[48] Here we follow Huber (2011) and subsequent authors and modify the definition of the Gâteaux derivative to only consider directions that lie within the set of probability measures.

Online Appendix V:Repeated choices and correlation aversion

In experimental settings, participants appear to be averse to correlation between risks across time periods (see for example Andersen et al. (2018)). Here we show that allowing the adversarial forecaster to make second-period forecasts after observing the first-period realizations is a special case of the one-period adversarial forecaster model and that the induced preferences can exhibit correlation aversion.

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 action $y_0 \in X_0$ with no additional information about F, and then takes $y_1 \in X_1$ after observing the realization of x_0, where 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. The agent knows that the adversary picks $y_1 \in Y_1$ conditional on the realization of x_0, and their induced preferences over lotteries on $X_0 \times X_1$ take into account the adversary's conditional best response.

Example 11. 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 $F_1(\cdot|x_0)$ denote the marginal and the conditional distributions of F. Mapping this to the one-period model,$\sigma(x_0, x_1, F) = \sigma_0(x_0, F_0) + \sigma_1(x_1, F_1|x_0)$, 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), we say that 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. $\geqslant$). 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 6, 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 F such that $\int \tilde{x_1} dF_1(\tilde{x_1}|1) < \int \tilde{x_1} dF_1(\tilde{x_1}|0)$, because the agent's preference preserves the supermodular order over such lotteries. Combining these facts shows the agent's utility is only maximized if $\int \tilde{x_1} dF_1(\tilde{x_1}|1) = \int \tilde{x_1} dF_1(\tilde{x_1}|0)$, so the best conditional forecast is independent of x_0. △

As an example, suppose $v(x) = 1 - \exp(-ax)/a$ for $a > 0$ and that the forecaster's loss function is $\rho(z) = \exp(\lambda z) - \lambda z$ for some $\lambda > 0$. 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 $a < 1$.

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.

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.

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.