Adversarial forecasters, surprises and randomization

Roberto Corrao* — MIT
Drew Fudenberg† — MIT
David Levine‡ — EUI and WUSTL

November 28, 2022

Abstract

In an *adversarial forecaster* model, utility over lotteries is the sum of an expected utility function and a "surprise term" measuring how surprising the outcome is given the forecast made by an adversarial forecaster who attempts to find the forecast that minimizes the forecast error. We show that an adversarial forecaster model gives rise to preferences that are concave and satisfy a form of differentiability condition, and that any preference relation that has a concave representation that satisfies the differentiability condition arises from an adversarial forecaster model. Because of concavity, the agent typically prefers to randomize. We characterize the support size of optimally chosen lotteries, and how it depends on preferences for surprise. We then show that the preferences induced by a sequential game against an adversary with an arbitrary set of feasible actions have an adversarial forecaster representation with weaker continuity properties, and that they admit an adversarial forecaster representation if and only if the adversary has a unique best response to each lottery.

*Department of Economics, MIT, rcorrao@mit.edu
†Department of Economics, MIT, drew.fudenberg@gmail.com
‡Department of Economics, EUI and WUSTL, david@dklevine.com

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

1 Introduction

Consider a specific type of lottery: an agent must choose one of their local sports team's matches to watch. They care only about whether their team wins or loses, and prefer to watch their team win for sure than lose for sure. Some theories of preferences over lotteries assume stochastic dominance or monotonicity, which implies that the agent's most preferred match is one where their team is guaranteed to win. But that would be a rather boring match, and the agent might prefer to watch a match where their team is favored but not guaranteed to win. For this reason, we might wish to reject the axiom of monotonicity. Similar considerations arise in political economy in the theory of expressive voting, in which people get utility from watching a political contest, and their utility is enhanced by participation. Just as in the theory of sports matches, some may prefer a more exciting contest, so even without strategic considerations turnout is likely to be higher when the polls show a close race (see for example Levine, Modica, and Sun [2021]).

Suppose, then, that the agent has a preference for being surprised, and that their overall utility is the sum of a utility from which team wins, and a measure of how surprising the outcome is. An outcome is surprising if it is difficult to forecast in advance, where a forecast is a probability distribution over outcomes that is chosen by an adversarial forecaster who attempts to minimize the forecast error. We refer to this minimized forecast error as the surprise, and we assume that forecasting the true lottery always minimizes the forecast error. We refer to this as the *adversarial forecaster* representation.

We say that a preference has *continuous local expected utility* if there is a linear functional, i.e. an expected utility, that continuously varies with the distribution considered and that "supports" the preference at each lottery. We show that an adversarial forecaster model gives rise to preferences that have continuous local expected utility. In fact, this is the only restriction the adversarial forecaster representation imposes: Any preference relation that has continuous local utility can be generated by an adversarial forecaster model. Moreover, preferences with local expected utility are concave in probabilities, so a preference for surprise rationalizes stochastic choice.

The adversarial forecaster model lets us impose additional restrictions on preferences in a natural way through the forecast error. One important class of examples is when the forecast error corresponds to the widely used method of moments technique.

We show that this results in a quadratic - hence easy to analyze - utility function.

A natural application of the idea of surprise to is to story-telling. We first consider one such application: a receiver who likes an exciting story. Our model of an exciting story is a more general version of Ely, Frankel, and Kamenica [2015]. In that model, the agent cared about a particular kind of surprise, and not at all about the outcome. In our model, the agent also cares about the state, so a sender that maximizes the receiver's total utility designs the initial distribution over states as well as how information is revealed. As in Ely, Frankel, and Kamenica [2015] we find that the optimal information policy for a given distribution over states does not depend on preferences over states. However, the optimal distribution over states does depend on the receiver's state preferences, and thus so does the chosen information policy.

To better understand preference for surprise and the extent of deliberate randomization of the agent, we study optimal lotteries in the special case where the forecast error has a finite-dimensional parameterization, e.g. a function of a finite number of moments. We apply this to settings where the agent chooses a lottery subject to moment restrictions, such as that lottery's expected value equals the endowment. We show that when the parameter space is finite dimensional, there is always an optimal lottery that has finite support. Specifically, if the forecast error is a function of k parameters and there are m moment restrictions, there is an optimal lottery with support of no more than $(k + 1)(m + 1)$ points. For example, in the sports case, suppose that preferences are not merely over which team wins or loses, but also over the score, where the latter can take on a continuum of values. If the forecaster is limited to predicting the mean score and there are no moment constraints, then one most preferred choice is a binary lottery between two scores.

We then consider another tractable class of adversarial forecaster preferences, those which arise when the agents trades of the interests of different potential selves. We show that these preferences can also arise as the solution to optimal transport problems, so we call them "transport preferences." We show how the extent of randomization induced by transport preferences depends on the number of different utility functions the selves might have, and how transport preferences relate to the *ordinally independent* preferences introduced by Green and Jullien [1988].

By definition, adversarial forecaster preferences have continuous local utility. We show that dropping the continuity requirement allows exactly the class of *adversarial expected utility* preferences, where the adversary acts to minimize the agent's utility, but has actions other than forecasts and the agent and adversary's utility functions have a more general form. Moreover, we show that adversarial expected utility preferences admit an adversarial forecaster representation if and only if the adversary has a unique best response to each lottery.

We study the monotonicity properties of this more general model with respect to stochastic orders, and apply them to the question of how preference for surprise is reflected in attitudes towards risk. First we show that these preferences preserve a stochastic order if and only if, for every lottery, there is a best response of the adversary that induces a utility over outcomes that reflects the stochastic order. We then 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 preferences for surprise may lead an agent with a risk-averse expected utility component to have preferences that are overall risk loving. We then show how the adversarial expected utility model can be used to capture correlation aversion. Intuitively, the agent optimally chooses distributions that minimize the correlation between outcomes to maximize the residual uncertainty of an adversary who observes one of them.

Related Work Our paper is related to three distinct types of decision theoretic models of risk preference. 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.

It 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]. Our work complement the analyses of optimization problems with non expected utility preferences in Cerreia-Vioglio, Dillenberger, and Ortoleva [2020] and Loseto and Lucia [2021] by characterizing the optimal lotteries and bounding the size of their supports in more general environments. Our work is also related to the ordinally independent preferences studied in Green and Jullien [1988], which extend the rank-dependent utility model of Quiggin [1982]. Ordinally independent preferences have an adversarial expected utility representation provided that a supermodularity condition holds, which allows us to apply our results on optimality and monotonicity to them.

Our work is related to the study of induced preferences due to temporal risk, as in Machina [1984], where the agent chooses a lottery over outcomes and then chooses an action without observing the lottery's realization, which convex preferences over the first-stage choice. In our model, the agent chooses a lottery and then an adversary chooses an action without observing the realization. Unlike Machina's induced preferences, ours have a concave representation and a preference for randomization.

Finally, our analysis of monotonicity is related to the work on stochastic orders and preferences over lotteries in e.g. Cerreia-Vioglio [2009] and Cerreia-Vioglio, Maccheroni, and Marinacci [2017]. 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.[1]

2 Adversarial Forecasters

This section introduces the *adversarial forecaster model*, in which the agent has preferences over lotteries of outcomes that depend on both the expected utility of the lottery's outcome and a measure of surprise.

2.1 The Model

The agent plays a sequential move game against an adversarial forecaster. The agent moves first, and chooses a lottery $F \in \mathcal{F}$, the set of Borel measures on a compact metric space X of outcomes, or a compact subset of them. We endow $\mathcal{F}$ with the topology of weak convergence, which makes it metrizable and compact. Then the adversary observes F and chooses a *forecast* $\hat{F} \in \mathcal{F}$, that is, a probabilistic statement about how likely different outcomes are. We study the agent's preference over lotteries of outcomes that is induced by backward induction in this sequential game.

Let δ_x denote the Dirac measure on x.

Definition 1. (i) We say that $\sigma : X \times \mathcal{F} \to \mathbb{R}$ is a *forecast error* if $\sigma(x, \delta_x) = 0$ for all $x \in X$, σ is continuous, and $\int \sigma(x, F) dF(x) \leqslant \int \sigma(x, \hat{F}) dF(x)$ for all $F, \hat{F} \in \mathcal{F}$.

(ii) The *surprise* of lottery F given the forecast error σ is $\Sigma(F) = \min_{\hat{F} \in \mathcal{F}} \int \sigma(x, \hat{F}) dF(x) = \int \sigma(x, F) dF(x)$.[2]

[1] See Section 7 for a more detailed discussion of these and other related results.

[2] Note that our definition of surprise is more general than that of Ely, Frankel, and Kamenica

Definition 1 requires that there is no surprise when the realized outcome was predicted by the forecaster to have probability 1, and that the forecast F minimizes the expected forecast error when the true lottery is F, but it does not require that this is the only minimizer. Observe that the forecast error is always non-negative, since $\sigma(x, F) \geqslant \sigma(x, \delta_x) = 0$ for all $F \in \mathcal{F}$ and $x \in X$. One example is $X = \{0, 1\}$ and $\sigma(x, F) = \left(x - \int x dF(x)\right)^2$, so the forecast error is measured by mean-squared error. We illustrate this functional form in Example 1 below. Note also that because Σ is the minimum over a collections of linear functionals it is concave, and that $\Sigma(\delta_x) = 0$ for any x. Theorem 1 will sharpen this to provide a necessary and sufficient condition.

Let $C(X)$ denote the space of continuous real functions over X, endowed with the topology induced by the supnorm.

Definition 2. Preference $\succsim$ has an *adversarial forecaster representation* if they can be represented by a function V satisfying

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

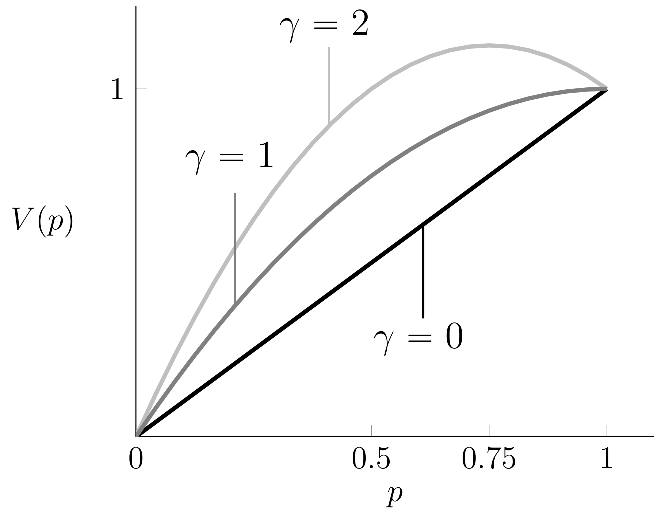
where σ is a forecast error and $v \in C(X)$.

This representation can be interpreted as follows: The agent has a baseline preference over outcomes described by the expected utility function utility v, and a preference for surprise captured by the forecast error σ. Given a forecast $\hat{F}$ of the adversary, the agent's total utility is the sum of their expected baseline utility and the expected forecast error, so the forecast error σ is also the *loss function* of the adversary.

Equation 1 shows that V is continuous and concave, and that $V(\delta_x) = v(x)$. Note that while a preference with an adversarial forecaster representation generally departs from expected utility, it does satisfy the independence axiom for comparisons of lotteries that induce the same surprise.

Example 1. In a sports match, the outcome is $x = 1$ if the preferred team wins and $x = 0$ if it looses. Let p be the probability of winning, $\hat{F}$ be the forecast, and let $\gamma(x - \int \tilde{x} d\hat{F}(\tilde{x}))^2$ measure the realized forecast error given the forecast $\hat{F}$. The decision maker gets utility $v(x) = x$ plus γ times the squared error of the forecast, and the adversary's optimal choice is to forecast p, variance $p(1-p)$, so the agent's preference over lotteries is represented by $V(p) = p + \gamma p(1-p)$. If $\gamma > 1$ and the

[2015].



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

agent can choose any value of p, the best lottery is $p = (1+\gamma)/(2\gamma)$, so that the preferred team might lose, while if $0 \leqslant \gamma \leqslant 1$ the best lottery is $p = 1$. △

2.2 Local Expected Utility

Suppose that preferences can be represented by a continuous utility function V. We say that $w \in C(X)$ is a *local expected utility* of V at F if it is a supporting hyperplane: that is, for every $\tilde{F} \in \mathcal{F}$, we have $\int w(x)d\tilde{F}(x) \geqslant V(\tilde{F})$ with $\int w(x)dF(x) = V(F)$. The function V has a *local expected utility* if there is at least one local expected utility at each F. Any function that has a local expected utility is concave, and a local expected utility at F is a supergradient of V at F.[3] Moreover, when V has a local expected utility w at F, if $\int w(x)dF(x) \geqslant \int w(x)d\tilde{F}(x)$ (resp. $>$), then $V(F) \geqslant V(\tilde{F})$ (resp. $>$), which explains the name we adopt for this supporting hyperplane.[4]

We say that V has *continuous local expected utility* if there is a continuous function $w : X \times \mathcal{F} \to \mathbb{R}$ such that $w(\cdot, F)$ is a local expected utility of V at F. This is our main "differentiability" condition for arbitrary representations V. In fact, as we show in Online Appendix V, V has a continuous local expected utility if and only if it is

[3]See e.g. Aliprantis and Border [2006] p. 264. Local utility, unlike concavity, requires there are supporting hyperplanes at boundary points. Machina [1982] uses a different definition that is neither weaker nor stronger than ours; see Online Appendix V.

[4]This follows from the concavity of V. See Online Appendix V for a formal proof.

concave and Gâteaux differentiable with continuous Gâteaux derivative. Note that continuous local utility does not imply that there is a unique local expected utility at every point; generally there will be a continuum of local expected utilities at boundary points.[5]

Theorem 1. *Let $\succsim$ be a preference over $\mathcal{F}$. The following are equivalent:*

(i) Preference $\succsim$ admits an adversarial forecaster representation.

(ii) Preference $\succsim$ has a representation V with a continuous local expected utility.

The proof of this and all other results is in the appendix except where otherwise noted. In the appendix we obtain this result as a consequence of the more general Theorem 8, but it can be proved more directly by noting that if V has an adversarial forecaster representation, then $V(F) = \int v(x)dF(x) + \int \sigma(x,F)dF(x)$ for every F, which implies that $w(\cdot,F) = v + \sigma(\cdot,F)$ is a local expected utility of V. In turn, the joint continuity of σ implies that w is continuous, yielding that V has a continuous local expected utility. Conversely, given a representation V with continuous local expected utility w, we can set $v(x) = V(\delta_x)$ and $\sigma(x,F) = w(x,F) - v(x)$. Because w is continuous, $\int w(x)d\tilde{F}(x) \geqslant V(\tilde{F})$, and $\int w(x)dF(x) = V(F)$, it follows that $\sigma(x,F)$ is a valid forecast error: It is continuous, minimized at $\hat{F} = F$, and is 0 on deterministic lotteries. Thus V admits a representation as in equation 1.

The next proposition provides a fixed-point characterization of optimal lotteries that we use in the analysis below.

Proposition 1. *If V has an adversarial forecaster representation, then for any convex and compact set $\overline{\mathcal{F}} \subseteq \mathcal{F}$,*

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

The discussion above shows maximizing local expected utility is a sufficient condition for a maximum, whether or not the local utility is continuous. The proof of necessity relies on the fact that if V has a continuous local expected utility, the directional derivative of V at any lottery F in direction $\hat{F}$ is well defined and

[5] Boundary points are especially important in the infinite-dimensional case since with the topology of weak convergence all points are on the boundary.

given by $\int v(x) + \sigma(x, F)d\hat{F}(x)$: F^* is optimal only if the directional derivative of V at F^* in any direction is non-positive. The necessity result fails when the local utility is not continuous. For example, suppose that $X = [-1, 1]$ and $V(F) = \min_{y \in [-1,1]} \int_{-1}^{1} (2y - 1)xdF(x)$, which is an example of the adversarial expected utility representation analyzed in Section 6. Then $F^* = \delta_0$ is uniquely optimal over $\mathcal{F}$ for V. However, $w(x, y) = (2y - 1)x$ is a local expected utility for V at F^* for every $y \in [-1, 1]$, yet F^* is strictly suboptimal for all of these local utility functions except for the one corresponding to $y = 0$.

The fixed-point condition characterizing the optimal lotteries in Proposition 1 has a clear equilibrium interpretation: The adversary chooses a forecast $\hat{F}$ given the equilibrium choice of the agent, and the agent maximizes the resulting local expected utility. The adversary's forecast is a best response if it induces the agent to choose the forecasted lottery. In particular, when $\overline{\mathcal{F}} = \mathcal{F}$, F^* is optimal if and only if $\text{supp}(F^*) \subseteq \text{argmax}_{x \in X}\, v(x) + \sigma(x, F^*)$. This is reminiscent of the personal equilibrium and preferred personal equilibrium of Kőszegi and Rabin [2006], with the important difference that when there are multiple fixed points they are all optima, so that the refinement to the preferred equilibria is vacuous.

The adversarial forecaster representation is concave, so typically the optimal lottery will be strictly mixed.[6] When X is an interval of real numbersCerreia-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. [7] Also, Theorem 1 implies that the Additive Perturbed Utility (APU) preferences of Fudenberg, Iijima, and Strzalecki [2015] (which are only defined for finite X) have an adversarial forecaster representation.[8]

On the other hand, adversarial forecaster representations are not necessarily APU. In particular, choices generated by APU preferences satisfy the *regularity* property that enlarging the choice set cannot increase the probability of pre-existing alternatives. [9] The next example shows how regularity can fail with adversarial forecaster

[6]See Proposition 2 for a class of *strictly concave* adversarial forecaster representations.

[7]Point (i) of their definition is satisfied because V is concave. The same argument we use in the proof of Theorem 6 below shows that point (ii) is satisfied when each $w(\cdot, F)$ is strictly increasing.

[8]This follows by combining the maxmin representation of APU preferences given in Proposition 7 of that paper and the properties of continuous local expected utilities.

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

preferences.

Example 2. Suppose that $X \subseteq \mathbb{R}$, that the agent's baseline utility v is concave and twice continuously differentiable, and that the agent has a preference for surprise given by $\sigma(x, F) = \left(x - \int \tilde{x} dF(\tilde{x})\right)^2$, so that any forecast $\hat{F}$ with the same mean as F minimizes the expected forecast error, and is a best response for the adversary. The proof of Theorem 1 implies that this generates local utility $w(x, F) = v(x) + \left(x - \int_0^1 \tilde{x} dF(\tilde{x})\right)^2$. Observe that the agent's ranking of two lotteries with the same expected value $\overline{x}$is same as those of an expected utility agent with utility function $w(x) = v(x) + (x - \overline{x})^2$, which is less risk averse than v. Moreover, the stochastic choice rule induced by these preferences need not satisfy Regularity. For example, if $v(x) = x$, the uniquely optimal choice for the agent from $\Delta\big(\{-1, 0\}\big)$ is δ_0, so there is no surprise. In contrast, when $\Delta\big(\{-1, 0, 1\}\big)$, 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.[10] For general v that are not too concave, i.e. when $v'' \geqslant -2$, the local utility is convex in x for all forecasts F. Theorem 6 below shows this implies the agent weakly prefers any mean-preserving spread $\tilde{F}$ of F to F itself. We say more about the effect of surprise on risk aversion in Section 7. △

More generally, Theorem 2 in Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella [2019] implies that whenever each $v + \sigma(\cdot, F)$ is strictly increasing and there exist two lotteries $F, \tilde{F}$ and $\lambda \in (0, 1)$ such that $V(\lambda F + (1 - \lambda \tilde{F}) > \max \left\{V(F), V(\tilde{F})\right\}$, then the induced stochastic choice does not satisfy Regularity. The second assumption holds for example when V is strictly concave and also in the preferences we analyze next, as well as in the preferences studied in Section 5.

2.3 Generalized Method of Moments

We now introduce a tractable class of adversarial forecaster representations that we apply in to information design and more general optimization problems under exogenous support and moment restrictions. Suppose X is a closed bounded subset of an Euclidean space, and let S be a compact metric space of *parameters* with the Borel sigma algebra. Given any integrable 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 call the set

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

$\{h(\cdot,s)\}_{s\in S} \subseteq C(X)$ the *generalized moments.* We assume here that the forecaster's objective is to choose a forecast $\hat{F}$ that minimizes a weighted sum of these generalized moments.

Definition 3. The loss function σ is based on the *generalized method of moments* (GMM)[11] if there is a Borel probability space (S,μ) and a continuous function $h: X \times S \to \mathbb{R}$ such that

$$\sigma(x,\hat{F}) = \int \left(h(x,s) - h(\hat{F},s)\right)^2 d\mu(s).$$

Proposition 2. *Any σ based on the generalized methods of moments is a forecast error, and the surprise 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.

This shows that GMM forecast errors can generate quadratic utilities V (Machina [1982]) that are strictly concave, and so have a strict preference for randomization.

Chew, Epstein, and Segal [1991] show that strictly concave quadratic utilities do not satisfy betweenness but satisfy mixture symmetry, a weakening of both independence and betweenness that is more consistent with some experimental findings such as Hong and Waller [1986]. Proposition 3 in Dillenberger [2010] shows that preferences represented by quadratic utilities satisfy negative certainty independence (NCI) only if they are expected utility preferences. Therefore, when V is induced by a GMM forecast error and is strictly concave, as in Proposition 2, the corresponding preference does not satisfy NCI. This is intuitive, since NCI supposes the agent has a preference for deterministic outcomes which would not generate any surprise.

Here are two classes of GMM forecast errors.

Finite Moments If $S = \{s_1, ..., s_m\}$ is a finite set of non-negative integers, we can take $h(x,s) = \prod_{i=1}^{m} x_i^{s_i}$, the standard method of moments.[12] The simplest case is the

[11] We abuse terminology here; in econometrics, the generalized method of moments minimizes a quadratic loss function on the data under the constraint that a number of generalized moment restrictions are satisfied.

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

one with only the first moment, $S = \{1\}$, as in Examples 1 and 2.

Moment Generating Function If for some $\tau > 0$ the parameter space is $S = [-\tau, \tau]^m$ we may take $h(x, s) = e^{s \cdot x}$. Here $h(F, s)$ is the moment generating function of F, where the map $F \mapsto h(F, \cdot)$ is one-to-one, so that the forecaster aims to match the entire distribution chosen by the agent. Proposition 2 shows that when μ has full support, the representation induced by this class of forecast error is strictly concave.

3 Writing a Suspenseful Novel

Ely, Frankel, and Kamenica [2015] consider how the writer of a novel who knows the ending, can best reveal information about that outcome over time.

The designer's objective is to maximize the utility of the watcher, who likes to be surprised. Here we show that the preferences they consider have an adversarial forecaster representation and can have a GMM form.[13] We also extend their analysis to let the watcher have preferences over realized outcomes, and let the broadcaster design both the distribution over states and the information revealed over time, for example both the ending of the story and how the story unfolds.

Let $\Omega = \{0, 1\}$ be a binary state space, $p \in \Delta(\Omega) = [0, 1]$ denote the probability that $s = 1$, and let $x = (\omega, p)$ be the outcome.

There are three time periods and two agents, a reader (R) and a writer (W). In Period 0, W chooses a distribution over S from a closed interval $\overline{\Delta} \subseteq [0, 1]$ (i.e., the ending of the story under some constraints) and commits to an information structure about s for Period 1 (i.e., how the story unfolds). In Period 1, R observes the signal realization, forms a posterior belief $p \in [0, 1]$, and their first-period surprise is realized. In Period 2, R observes the state realization s and their second-period surprise is realized.

Instead of working directly with the signals, we represent them with distributions over posteriors: W chooses a joint distribution $F \in \mathcal{F}$ over states and conditional beliefs of R. The feasible joint distributions are those such that, conditional on the

[13]To simplify notation we only show this for a binary state space, but it is true for any finite state space.

realization of the belief p, the induced conditional belief over Ω is equal to p itself:

$$\overline{\mathcal{F}} = \left\{ F \in \mathcal{F} : marg_S F \in \overline{\Delta}, \forall p \in \Delta(S), F(\cdot|p) = p \right\}.$$

For every, $F \in \overline{\mathcal{F}}$, we let $p_F \in \overline{\Delta}$ denote the induced probability that $\omega = 1$ and let $F_\Delta \in \Delta([0,1])$ denote the induced distribution over beliefs.[14]

In both periods, the agent has preferences for surprise. W acts to maximize the expected total utility of R which in turn depends on the surprise of the lottery, that is, its surprise. For $F \in \mathcal{F}$, the simplest version of surprise in period 1 is given by

$$V_1^L(F) = \int \frac{1}{2}||p - p_F||^2 dF_\Delta(p) = \int_0^1 p^2 dF_\Delta(p) - p_F^2.$$

As we show in Online Appendix I, this utility is equivalent to an adversarial forecaster representation with forecast error $\sigma_0(p, \hat{F}_\Delta) = \frac{1}{2}(p - \int \tilde{p} d\hat{F}_\Delta(\tilde{p}))^2$ where the forecaster only cares about the first moment of F_Δ. We refer to this as the linear case. Following Ely, Frankel, and Kamenica we assume that $V_1(F) = g(V_1^L(F))$ for some function $g : \mathbb{R} \to \mathbb{R}$ that is twice continuously differentiable, strictly increasing, and concave, with $g(0) = 0$. This makes a difference because preferences are aggregated over two periods after applying the g function. The resulting utility function V_1, while no longer quadratic, still has continuous local utility, so it is an adversarial forecaster representation by Theorem 1. The surprise in period 2 given $F \in \overline{\mathcal{F}}$ is

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

where δ_s represents the degenerate belief over s. Finally, R gets direct utility equal to $\tilde{v} \in \mathbb{R}$ when the realized state is $s = 1$ and direct utility 0 when $s = 0$; the case $\tilde{v} = 0$ yields the preferences in Ely, Frankel, and Kamenica.[15]

B wants to maximize the total utility of R, that is, to solve

$$\max_{F \in \overline{\mathcal{F}}} p_F \tilde{v} + (1-\beta) g\left(\int_0^1 p^2 dF_\Delta(p) - p_F^2 \right) + \beta \int_0^1 g(p - p^2) dF_\Delta(p). \tag{3}$$

[14] In Ely, Frankel, and Kamenica, p is fixed, so $\overline{\Delta} = \{p_0\}$.

[15] Ely, Frankel, and Kamenica have two different preference specifications for R, capturing preferences for suspense and surprise respectively. In their specification for suspense, flow utility at t depends on the expected surprise in period $t + 1$ given the belief at period t.

where $\beta \in [0,1]$ captures the relative importance of surprises across periods. Let $V_\beta(F)$ denote the total utility of W defined in equation 3. The discussion above shows V_β has a continuous local expected utility, so by Theorem 1 it admits an adversarial forecaster representation. The local utilities of V_β are:

$$w_\beta(\omega, p, F) = \omega\tilde{v} + (1-\beta)g'(D_2(F))(p^2 - p_F^2) + \beta g(p - p^2), \tag{4}$$

where $D_2(F) = \int \tilde{p}^2 dF_\Delta(\tilde{p}) - p_F^2$, and the baseline utility of R is $v_\beta(\omega, p) = V_\beta(\delta_{(\omega,p)}) = s\tilde{v} + \beta g(p-p^2)$ yielding a forecast error $\sigma_\beta(\omega, p, F) = (1-\beta)g'(D_2(F))(p^2 - p_F^2)$.

Because the total payoff of the reader depends only on the marginals of F, we can think to the writer as choosing only p_F and F_Δ given the consistency constraint. Next, we describe how the optimal marginals (p_F^*, F_Δ^*) depend on β.

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

1. *When $\beta \geqslant \overline{\beta}$, no disclosure is uniquely optimal (i.e., $F_\Delta^* = \delta_{p_F^*}$) and p_F^* is optimal if and only if it solves $\max_{p \in \overline{\Delta}} \{p\tilde{v} + \beta g(p-p^2)\}$.*

2. *When $\beta \leqslant \underline{\beta}$, full disclosure is uniquely optimal (i.e., $F_\Delta^* = (1-p_F^*)\delta_0 + p_F^*\delta_1$) and p_F^* is optimal if and only if it solves $\max_{p \in \overline{\Delta}} \{p\tilde{v} + (1-\beta)g'(p-p^2)(p-p^2)\}$.*

The proof of this result is in Online Appendix I. It is derived by computing the local expected utility of V_β at the candidate solution (p_F^*, F_Δ^*) and verifying that F_Δ^* is indeed optimal for that local expected utility by Proposition 1. Because the state is binary, each local utility is a linear combination $g'(D_2(F))p^2$ and $g(p-p^2)$, where the first term is strictly convex and the second is strictly concave For example, if $g(d) = \sqrt{d}$, then $g'(D_2(F))$ is very high for F such that F_Δ is concentrated around p_F, since in this case $D_2(F)$ is close to 0. Thus revealing no information cannot maximize V_β, since the local expected utility $w_\beta(\omega, p, F)$ is strictly convex in p. More generally, because W has nonlinear preferences over F_Δ, W might want to induce more than 2 posteriors, unlike in Bayesian persuasion with a binary state. Section 4 derives a more general result on the support size of optimal distributions.

In the linear case $g(d) = d$ with $\overline{\Delta} = [0,1]$ we can completely characterize the solution. For every $F \in \overline{\mathcal{F}}$, the total payoff of the watcher simplifies to $V_\beta(F) = p_F(\tilde{v} + \beta) - p_F^2(1-\beta) + \int (1-2\beta)p^2 dF_\Delta(p)$. The utility over realized posteriors $(1-2\beta)p^2$ is strictly concave when $\beta > 1/2$, so non- disclosure is uniquely optimal.

When $\beta < 1/2$, this term is strictly convex, so full disclosure is uniquely optimal, and when $\beta = 1/2$, R is indifferent over all the information structures. For every value of β and $\tilde{v}$, $p_F^* = \max\left\{0, \min\left\{1, \frac{\tilde{v}+\max\{\beta,1-\beta\}}{2\max\{\beta,1-\beta\}}\right\}\right\}$: the writer assigns a probability p_F^* to $\omega = 1$ that depends on the baseline value v as well as on the surprise parameter β. The nature of the optimal information structure between the two periods is always extreme (full or no-disclosure) and depends only on β. Observe that the disclosure policy in not affected by the baseline value v, hence with EFK preferences, that is with $\tilde{v} \equiv 0$, the optimal disclosure policy would be the same. However, the optimal probability of $\omega = 1$ would be $p^* = 1/2$, which is independent of the weight β. We give a more detailed analysis of the linear case in Online Appendix I .

Measures of uncertainty and information For an adversarial forecaster representation, the surprise function simplifies to $\Sigma(F) = V(F) - \int v(x)dF(x)$. The properties of V imply that Σ is concave and that $\Sigma(\delta_x) = 0$ for all $x \in X$. Frankel and Kamenica [2019] show that these are the two properties characterizing a valid measure of uncertainty, that is, a function representing the cost of uncertainty in a given decision problem. In our model, the coupled decision problem they consider is the forecasting problem faced by the adversary whose prior over outcomes coincides with the lottery F chosen by the agent. Because the Bregman divergence of Σ coincides with the forecast error σ, their Theorem 3 implies that the forecast error is what they call a valid measure of information: the forecast error generated by x given lottery F coincides with ex-post value for the adversary of observing the realized outcome as opposed to receiving no additional information.

4 Moment restrictions and optimal randomization

We turn now to the study of optimization problems with support restrictions and moment constraints, e.g. that the expected outcome must be constant across lotteries, as is the case with fair insurance. We are mostly interested in the extent of optimal randomization under preferences with adversarial forecaster representation, that is, in the size of the supports of optimal distributions. We show that with GMM preferences, this depends on the number of moments forecaster considers- the more accurately the forecaster tries to match the distribution, the more the agent has to work to make the outcome hard to predict. Moreover, this is true for a broader class of adversarial forecaster representations that nests the GMM case while allowing for non-quadratic loss functions and thus for a non-linear best response function of the adversarial forecaster.

Recall that a GMM representation is defined by a probability space (S, μ) and a continuous function $h(F, s) = \int h(x, s) dF(x)$. We may define the space of moments as the image of the map $P(F) \equiv h(F, \cdot)$, that is, $Y \equiv P(\mathcal{F}) \subseteq \mathbb{R}^S$. When S is finite, Y is a subset of a Euclidean space. If we then define a *parametric forecast error* on Y by $\hat{\sigma}(x, y) = \int (h(x, s) - y(s))^2 \, d\mu(s)$, we see that the GMM forecast error $\sigma(x, F) = \hat{\sigma}(x, P(F))$ depends on F only through $P(F)$. This lets us work with the function $\hat{\sigma}(x, y)$ instead of $\sigma(x, F)$, which is easier to study since it is strictly concave and differentiable in y. Parametric adversarial forecaster representations generalize these properties to other settings where surprise depends on the lottery only through a space Y of parameters.

Definition 4. A forecast error σ is *parametric* if there exist a set $Y \subseteq \mathbb{R}^m$, a continuous map $P : \mathcal{F} \to \mathbb{R}^m$, and a continuous function $\hat{\sigma} : X \times Y \to \mathbb{R}_+$ that is strictly concave and differentiable in y, such that $Y = P(\mathcal{F})$ and $\sigma(x, F) = \hat{\sigma}(x, P(F))$ for all $(x, F) \in X \times \mathcal{F}$.

When $\succsim$ has an adversarial forecaster representation with a parametric forecast error σ, we say that it has a *parametric representation.* In this case

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

and we let $\hat{y}(F)$ denote the (unique) parameter attaining the minimum. With any parametric forecast error, $P(F) = \hat{y}(F)$, which simplifies optimization problems as we now show.

Fix a compact and convex set $\overline{\mathcal{F}} \subseteq \mathcal{F}$ of feasible lotteries and observe that

$$\max_{F \in \overline{\mathcal{F}}} V(F) = \max_{F \in \overline{\mathcal{F}}} \int v(x) + \hat{\sigma}(x, P(F)) dF(x) \tag{5}$$

$$= \max_{\theta \in Y} \max_{F \in \overline{\mathcal{F}} : P(F) = \theta} \int v(x) + \hat{\sigma}(x, \theta) dF(x). \tag{6}$$

where the first equality follows from the fact that the expected forecast error given F is minimized at F, and the second equality follows by splitting the choice of the lottery in two parts: first the agent chooses the desired value for the parameter $\theta \in Y$

and then chooses among the feasible distributions that are consistent with θ. This program is linear in F and strictly concave in the finite dimensional parameter y, which makes it more tractable than the original problem.

We now apply this parametric model to optimization problems with a *moment restriction*. We fix some closed (possibly finite) subset $\overline{X} \subseteq X$ and a finite collection of k continuous functions $\Gamma = \{g_1, ..., g_k\} \subseteq C(X)$ together with the feasibility set

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

which we assume is non-empty. If x is money, then $\int x dF(x) = 0$ is the budget constraint that the agent may choose any fair lottery.

The next result shows that when an adversarial forecaster representation is parametric, there is always a solution of this optimization problem whose support is a finite set of outcomes. Moreover, the upper bound on this finite number of outcomes only depends on the dimension of Y and on the number of moment restrictions defining the feasible set of lotteries.

Theorem 2. *Fix a closed set $\overline{X} \subseteq X$, $\{g_1, ..., g_k\} \subseteq C(X)$, and let $\overline{\mathcal{F}} = \mathcal{F}_\Gamma\left(\overline{X}\right)$. Then there is a solution to (5) that assigns positive probability to no more than $(k+1)(m+1)$ points of $\overline{X}$.*

When $\gtrsim$ has a GMM representation with finitely many moments m and $\Gamma = \varnothing$, the theorem implies the optimal lottery puts positive probability on at most $m+1$ points. In this case the proof is relatively simple: Because $P(F) = (h(F, s))_{s \in S}$ and $\overline{\mathcal{F}} = \Delta(\overline{X})$, equation 6 becomes

$$\max_{\theta \in Y} \max_{F \in \Delta(\overline{X}) : h(F, \cdot) = y} \int v(x) + \hat{\sigma}(x, y) dF(x)$$

Fix a $\theta^* \in Y$ that solves the outer maximization problem. Then F^* solves the original problem if and only if it solves

$$\max_{F \in \Delta(\overline{X}) : h(F, \cdot) = \theta^*} \int \left(v(x) + \hat{\sigma}(x, \theta^*)\right) dF(x) \tag{7}$$

which is linear in F: The agent behaves as if they were maximizing expected utility over all lotteries that have the optimal values of the relevant moments. Because the objective in (7) is linear in F, there is a solution in the set of extreme points of the set $\{F \in \Delta(\overline{X}) : h(F, \cdot) = \theta^*\}$. This set is obtained by adding the m moment restrictions given by θ^* to the set of probabilities over $\overline{X}$, and Winkler [1988] shows that the extreme points of this set are supported on at most $m + 1$ points of $\overline{X}$.

This proof strategy relies on the linearity of the map P, which holds for the GMM representation but not for general parametric ones. The first step of the proof of the general result (Theorem 10 in the Appendix) makes use of the transversality theorem to show that, whenever $\overline{X}$ is finite, the bound on the support stated in Theorem 2 holds generically for every optimal lottery.[16] We conclude the proof with an approximation argument on both the baseline utility v and the set of feasible outcomes to show that, for arbitrary $\overline{X}$, there always exists a solution with the same bound on the support.

When Y is infinite dimensional, the choice set can have a thicker support.

Theorem 3. *Assume that $X = [0, 1]$, $\Gamma = \varnothing$, the kernel of a generalized methods of moments forecasting forecast error $H(x, \tilde{x}) = \int h(x, s)h(\tilde{x}, s)d\mu(s) = G(x - \tilde{x})$ is positive definite, and $H(0, \tilde{x})$ is non-negative, strictly decreasing (when positive) and strictly convex in $\tilde{x}$. Then there exists a unique maximum F and it has full support over X.*

To prove this, we first invoke Proposition 2 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. Then the complementary slackness condition, together with the assumptions on H, imply that there cannot be an open set in X to which F assigns probability zero.

Theorem 2 shows that when the adversary is confined to a small set of forecasts the support of the optimum is thin. Theorem 3 gives a sufficient condition for the support of the optimum to be thick. For this condition to be satisfied the adversary must have a "large" set of forecasts. Example 5 in Online Appendix IV.A shows that the set of stochastic processes with continuous sample paths on a unit interval is large enough to generate thick support for the optimum.

We close this section with a corollary that follows from Theorems 2 and 3.

[16] This bound applies to stochastic choices from finite sets, and can be empirically tested. Online Appendix III.B provides an extension to the case of infinite X.

Corollary 1. *Maintain the assumptions of Theorem 3, and let F denote the unique fully supported solution. There exists a sequence of method of moments representations V^n with $|S^n| = m^n \in \mathbb{N}$, and a sequence of lotteries F^n such that each F^n is optimal for V^n, is supported on up to $m^n + 1$ points, and $F^n \to F$ weakly, with* $\operatorname{supp} F^n \to \operatorname{supp} F = X$ *in the Hausdorff topology.*

Intuitively, as the number of moments that the adversary matches increases, the agent randomizes over more and more outcomes, up to the point that each outcome is in the support of the optimal lottery.[17]

5 Mutiple Selves, Transport Preferences, and Adversarial Forecasters

5.1 Transport Preferences

This section considers another tractable class of adversarial forecaster preferences that arise when the agent trade offs the (potentially diverging) interests of multiple selves. This representation can be equivalently expressed as the solution of an intermediate problem known as the transport problem, rationalizing the name we give to this class.

As a simple example, suppose $X = [0, 1]$ is a sports score with 0 meaning the agent's team loses by a landslide and 1 it wins by a landslide. The agent is of two minds about the game: one part (2/3 to be exact) would like to see the other team get a good thumping ($x = 1$), while the part (1/3) feels the team has been pretty uppity lately and wouldn't mind seeing it get a good thumping ($x = 0$). The best lottery for the agent would be one that balances the preferences of the selves: a 1/3rd chance $x = 0$ and a 2/3rd chance $x = 1$. We next formally describe a game between an agent with multiple selves against an adversarial forecaster and link this model back to the simple idea just illustrated.

Let the space of outcomes X be a compact convex subset of a Euclidean space with nonempty interior, let $\theta \in \Theta = [0, 1]$ index the different selves, and suppose that each self has a baseline utility function $\phi(\theta, x)$ for outcomes. We will assume that the agent's choice of lottery is made to maximize the sum of the average expected utility

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

the selves and their average individual surprise, with the average computed using the uniform distribution U.[18]

Let $Y \subseteq C(X)$ denote the set of continuous functions on X that are normalized so that $\int \exp(-y(x))\, dx = 1$.[19] We interpret each $y \in Y$ as a probabilistic forecast of the outcome in the form of the negative log-likelihood: for any strictly positive continuous density f on X, y defined by $y(x) = -\log(f(x))$ is in Y and conversely any $y \in Y$ corresponds to a unique strictly positive continuous density.

The timing of the game against the adversarial forecaster is the same as in Section 1. First, the agents chooses a lottery F without knowing which self will be experiencing the realized payoff. Then the adversary observes F and chooses a forecast $y \in Y$, without knowing the realized outcome x or the agent's θ. Finally, the realized self gets terminal payoff $\phi(\theta, x) + \sigma_\phi(\theta, x, y)$ where

$$\sigma_\phi(\theta, x, y) \equiv \max_{\xi \in X} \{\phi(\theta, \xi) + y(\xi)\} - [\phi(\theta, x) + y(x)],$$

is the *individual forecast error* for θ given the adversarial forecast y and outcome x. To understand the definition of $\sigma_\phi(\theta, x, y)$, consider the distorted baseline utility $\phi(\theta, x) + y(x)$ for self θ, which combines the quality of the outcome $\phi(\theta, x)$ and the measure of how unlikely the outcome was according to the adversary's forecast $y(x)$. The individual forecast error is then the difference between the highest possible distorted baseline utility and the realized one. Note that this is necessarily non-negative.

With transport preferences, the adversarial forecaster minimizes the expectation of the individual forecast error given the chosen lottery F and the distribution of selves, and the agent chooses lotteries to maximize the expected sum of their baseline utility ϕ and surprise $\Sigma_\phi(F) \equiv \min_{y \in Y} \int\int_0^1 \sigma_\phi(\theta, x, y) d\theta dF(x)$.

Definition 5. A preference $\succsim$ over $\mathcal{F}$ is a *transport preference* if can be represented by

$$V(F) = \int\int_0^1 \phi(x, \theta) d\theta dF(x) + \Sigma_\phi(F) \tag{8}$$

for some bounded measurable function $\phi(\theta, x)$ and associated individual forecast error

[18]Any non-uniform distribution over selves can be replicated by having more selves with the same utility function. In fact, all the results of this section hold when selves have any distribution $Q \in \Delta(\Theta)$ that has a density.

[19]Here we are considering the integral over X with respect to the Lebesgue measure.

σ_ϕ.

Note that the simple example at the start of this section corresponds to transport preferences with $\phi(\theta, x) = (\mathbf{1}[\theta \geqslant 1/3] - 1/2)\, x$.

5.2 Continuously Differentiable Transport Preferences

The next result gives a sufficient condition for transport preferences to have an adversarial forecaster representation.

Theorem 4. *If $\succsim$ is a transport preference with a continuously differentiable ϕ, then $\succsim$ has an an adversarial forecaster representation with*

$$v(x) = \int_0^1 \phi(x, \theta) d\theta \quad \text{and} \quad \sigma(x, F) = \int_0^1 \sigma_\phi(\theta, x, \hat{y}(F)) d\theta$$

where $\hat{y}(F) \equiv \operatorname{argmin}_{y \in Y} \int \int_0^1 \sigma_\phi(\theta, x, y) d\theta dF(x)$ for every $F \in \mathcal{F}$.

In the appendix, we derive this result from a more general characterization of the the adversarial forecaster representation that we present in the next section. However, it is also possible to directly prove the result more directly from standard optimal transport duality results as follows: First, equation 8 can be rewritten as

$$V(F) = \min_{y \in Y} \left\{ \int_0^1 \max_{\xi \in X} \{\phi(\theta, \xi) + y(\xi)\} d\theta - \int y(x) dF(x) \right\}. \tag{9}$$

This expression is equivalent to the dual problem of the Kantorovich optimal transportation problem and under the assumptions of Theorem 4, there exists a unique solution $\hat{y}(F) \in Y$ that attains this minimum. Finally, we have that

$$V(F) = \int_0^1 \phi(x, \theta) d\theta + \int_0^1 \sigma_\phi(\theta, x, \hat{y}(F)) d\theta$$

and can verify that $\sigma(x, F) = \sigma_\phi(\theta, x, \hat{y}(F))$ is a valid forecast error.

Even without the additional assumptions of Theorem 4, strong duality holds in the optimal transportation problem associated to ϕ, that is, we can rewrite V as

$$V(F) = \max_{T \in \Delta(U, F)} \int \phi(\theta, x) dT(\theta, x) \tag{10}$$

where $\Delta(U, F) \subseteq \Delta(\Theta \times X)$ is the set of joint distributions over $\Theta \times X$ with marginals respectively given by U and F. We use this alternative representation to analyze the 1-dimensional case in the next section, and to derive the properties of optimal lotteries in Online Appendix III.D.

5.3 One-dimensional case and ordinal independence

In this section we assume that $X = [0, 1]$ so that we may view identify each F as a cdf, and define the generalized inverse or quantile function $F^{[-1]}(t) \equiv \inf\{x \in X : F(x) \geqslant t\}$. When the individual utilities of the multiple selves can be ordered with respect to their marginal utility, we obtain a (non separable) rank-dependent representation for transport preferences.

Proposition 4. *Let $\succsim$ be a transport preference such that ϕ is continuously differentiable with $\phi_x(\theta, x)$ increasing in θ. The adversarial forecaster representation of $\succsim$ can be equivalently written as*

$$V(F) = \int_0^1 \phi(\theta, F^{[-1]}(\theta))d\theta. \tag{11}$$

To see why this proposition holds, note that when $\phi_x(\theta, x)$ is increasing in θ an optimal solution to the transport problem problem in equation 10 is to give the highest x to the highest θ, that is, to assign $\theta = F^{[-1]}(x)$.

This alternative representation implies that $\succsim$ belongs to the class of *ordinal independent* preferences studied in Green and Jullien [1988] when ϕ is nondecreasing in x. Ordinally independent preferences relax independence to ordinal independence, which says that if two distributions have the same tail, this tail can be modified without altering the preference between these distributions. Green and Jullien [1988] show that the standard axioms of expected utility along with monotonicity and ordinal independence in place of independence imply the existence of a continuous real valued utility function $\phi(\theta, x)$ nondecreasing in x and such that preferences have the representation in equation 11. This generalizes the rank-dependent representations of Quiggin [1982] and Yaari [1987], where $\phi(\theta, x) = w(\theta)v(x)$, which rules out the Friedman-Savage paradox where risk preferences depend on the status-quo wealth level.

If $\phi(\theta, x)$ is differentiable and $\phi_x(\theta, x)$ is weakly increasing in θ, these preferences have a continuous local expected utility, so they admit an adversarial forecasting representation and have a preference for surprise; monotonicity with respect to x is not needed. The next example has a differentiable transport preference that does not satisfy monotonicity and induce lotteries with full support over $[0, 1]$.

Example 3. Suppose $\phi(\theta, x) = \theta x - x^2/2$. The unique optimal distribution over $\mathcal{F}$ given these preferences is the uniform distribution over $[0, 1]$. To see this, observe that we can recast the problem by maximizing over quantile functions rather than CDFs:

$$\max_{F^{[-1]}} \int_0^1 (tF^{[-1]}(t) - \left(F^{[-1]}(t)\right)^2/2)dt,$$

which is uniquely solved by $F^{[-1]}(t) = t$ that is, the quantile function of the uniform distribution. Thus as in Theorem 3, this representation induces optimal distributions with thick support. △

In Online Appendix III.D, we generalize this example to the entire class of transport preferences without restricting to one-dimensional outcomes.

6 Adversarial expected utility

We now generalize the adversarial forecaster representation to adversaries with other objectives than minimizing the forecast error. This provides a link between the adversarial forecaster representation and the induced preferences of an expected utility agent in a zero-sum sequential game. Moreover, it lets us deal with preferences that are not consistent with continuous local utility representations, such as in Example 4 below where the loss function of the adversary is the absolute value of the error. Finally, it clarifies the relation of adversarial preferences to other risk preferences that admit a maxmin representation, such as those in Maccheroni [2002], Cerreia-Vioglio [2009], and Cerreia-Vioglio, Dillenberger, and Ortoleva [2015].

Let X be a compact metric space of outcomes, Y a compact metric space of choices for the adversary, and $G \in \mathcal{G}$ the space of probability measures on the Borel sets of $X \times Y$, endowed with the topology of weak convergence.

We suppose that the agent has expected utility preferences and the adversary has the opposite preferences: it prefers what is least liked by the agent.

Definition 6. Preference $\succsim$ over $\mathcal{F}$ has an *adversarial expected utility* representation if there is a compact metric space Y and a continuous utility function $u : X \times Y \to \mathbb{R}$ such that $\succsim$ is represented by

$$V(F) = \min_{y \in Y} \int u(x, y) dF(x). \tag{12}$$

Adversarial expected utility is similar to Maccheroni [2002]'s maxmin model under risk, although that model assumes there is always a deterministic outcome that is preferred to all non-deterministic ones, and that Maccheroni [2002] does not assume that Y is compact.[20] The envelope representation of Chatterjee and Krishna [2011] is a particular case of adversarial expected utility representation where $X = [0, 1]$ and V satisfies stronger continuity properties.

Next we link the adversarial expected utility representation to the adversarial forecaster representation. First, we relax joint continuity in the definition of forecast errors.

Definition 7. We say that $\tilde{\sigma} : X \times \mathcal{F} \to \mathbb{R}_+$ is a *weak forecast error* if the family of functions $\{\tilde{\sigma}(\cdot, F)\}_{F \in \mathcal{F}}$ is equicontinuous over X, $\tilde{\sigma}(x, \delta_x) = 0$ for all $x \in X$, and if $\int \tilde{\sigma}(x, F) dF(x) \leqslant \int \tilde{\sigma}(x, \hat{F}) dF(x)$ for all $F, \hat{F} \in \mathcal{F}$.

When a preference $\succsim$ can be represented as in equation 1 by using a weak forecast error $\tilde{\sigma}$, we say that it has a *weak adversarial forecaster representation*. The adversarial expected utility representation and the weak adversarial forecaster representation are equivalent.

Proposition 5. *Let $\succsim$ be a preference over $\mathcal{F}$. The following conditions are equivalent*

(i) The preference $\succsim$ has a weak adversarial forecaster representation.

(ii) The preference $\succsim$ has an adversarial expected utility representation.

The intuition for the equivalence between (i) and (ii) is similar to that for Theorem 1, with joint continuity of the local expected utility function replaced by equicontinuity. In particular, given an adversarial expected utility representation of $\succsim$ with associated utility function u over $X \times Y$, we can define the local expected utility of

[20] The paper incorrectly claims the resulting representation involves a minimum that is always attained. Machina [1984] and Frankel and Kamenica [2019] make the same mistake, see Corrao, Fudenberg, and Levine [2022].

V as $w(x,F) = u(x,\hat{y}(F))$ for some (not necessarily continuous) selection $\hat{y}(F)$ from $\hat{Y}(F) \equiv \text{argmin}_{y\in Y} \int u(x,y)dF(x)$. Similarly, the corresponding weak forecast error can be defined by $\tilde{\sigma}(x,F) = w(x,F) - v(x)$ where $v(x) = V(\delta_x)$. In the next example, the agent's preferences have an adversarial expected utility representation but cannot be represented by an adversarial forecaster representation.

Example 4. Consider the same setting of Example 1, but suppose the adversary's objective is to minimize the absolute deviation, so

$$V(F) = \int_0^1 v(x)dF(x) + \min_{c\in[0,1]} \int |x-c|dF(x).$$

In this case, the relevant statistic for the adversary is the median of the chosen distribution. The median need not be unique for some F. For example, consider the class of distributions $F^\epsilon = (1/2-\epsilon)\delta_0 + (1/2+\epsilon)\delta_1$ for $\epsilon \in (-1/2, 1/2)$, and observe that every number c in $[0,1]$ is a valid median for F^ϵ at $\epsilon = 0$. If we let $\hat{c}(F)$ be an arbitrary selection from the correspondence mapping distributions to medians, then $\tilde{\sigma}(x,F) = |x - \hat{c}(F)|$ is a weak forecast error. However, the family of distributions F^ϵ introduced above shows it is not possible to construct a continuous forecast error σ, since every selection $\hat{c}(F)$ from the sets of medians will be discontinuous at $\epsilon = 0$. △

We next relate transport preferences to adversarial expected utility. As we show in the proof of Theorem 5, we can restrict the minimization problem defining transport preferences in equation 9 to a compact subset $\tilde{Y} \subseteq Y$. Therefore, transport preferences admits an adversarial expected utility representation with utility $u(x,y) = y(x) + \int \max_{\xi\in X} \{\phi(\theta,\xi) + y(\xi)\} dQ(\theta)$, and Proposition 5 implies that they also have a weak adversarial forecaster representation.

Theorem 4 shows that transport preferences have an adversarial forecaster representation when ϕ is continuously differentiable, which guarantees there is a unique minimizer for every F. The next result shows that uniqueness of the minimizer is characterizes the adversarial forecaster representations within the more general class of adversarial expected utility.

Definition 8. An adversarial representation satisfies *uniqueness* if $\hat{Y}(F)$ is a singleton for all $F \in \mathcal{F}$.

Theorem 5. *A preference $\succsim$ over $\mathcal{F}$ has an adversarial expected utility representation that satisfies uniqueness if and only if it has an adversarial forecaster representation.*

To show that $\succsim$ has an adversarial forecaster representation if it has an adversarial expected utility representation that satisfies uniqueness, we define $v(x) = V(\delta_x)$ and $\sigma(x, F) = u(x, \hat{y}(F)) - v(x)$, where the uniqueness of $\hat{y}(F)$ implies it is continuous. To prove that $\succsim$ with an adversarial forecaster representation has also an adversarial expected utility representation that satisfies uniqueness, we start from the adversarial forecaster representation and consider a modified minimization problem for the adversary that lets them pick an expected utility (i.e., a hyperplane) that supports V at F. The joint continuity of σ implies that there exists a unique supporting expected utility for every F, hence the adversary has a unique best response in this ancillary problem, yielding the result.

7 Monotonicity and behavior

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

7.1 Stochastic orders and monotonicity

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

Definition 9. Fix a set $\mathcal{W} \subseteq C(X)$.

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

$$F \succsim_{\mathcal{W}} \tilde{F} \iff \int w(x) dF(x) \geq \int w(x) d\tilde{F}(x) \quad \forall w \in \mathcal{W}. \tag{13}$$

(ii) A preference $\succsim$ *preserves* $\succsim_{\mathcal{W}}$ if for all $F, \tilde{F} \in \mathcal{F}$, $F \succsim_{\mathcal{W}} \tilde{F}$ implies $F \succsim \tilde{F}$.

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

Notice that if $\succsim$ preserves $\succsim_{\mathcal{W}}$ then it also does so for any larger set $\tilde{\mathcal{W}} \supseteq \mathcal{W}$. For every set $\mathcal{W} \subseteq C(X)$, let $\langle \mathcal{W} \rangle$ denote smallest closed convex cone containing $\mathcal{W}$ and all the constant functions. From Theorem 2 in Castagnoli and Maccheroni [1999], for every $v \in C(X)$, the expected utility preference $\succsim_v$ preserves $\succsim_{\mathcal{W}}$ if and only if $v \in \langle \mathcal{W} \rangle$. Intuitively, if v is not parallel to some function in $\mathcal{W}$ then it crosses every function in $\mathcal{W}$, and so has a less preferred point that is preferred by $\succsim_{\mathcal{W}}$.

Given an adversarial expected utility representation (Y, u), let $\mathcal{H}(\hat{Y}(F))$ denote the space of Borel probability measures over $\hat{Y}(F)$, that is, the set of mixed best responses of the adversary given F.[21] Moreover, define $u(\cdot, H) = \int u(\cdot, y) dH(y) \in C(X)$ for every probability measure $H \in \mathcal{H}$. In an adversarial expected utility representation, we can associate the utility function u with the set $\mathcal{W}_{u,Y} = \{u(\cdot, y) : y \in \hat{Y}(F), F \in \mathcal{F}\}$ and a stochastic order $\succsim_{u,Y}$ on $\mathcal{F}$. It is clear that the expected utility preference $\succsim$ represented by u preserves $\succsim_{u,Y}$, and more generally, preserves any stochastic order $\succsim_{\tilde{\mathcal{W}}}$ generated by a set $\tilde{\mathcal{W}} \supseteq \mathcal{W}_{u,Y}$. Theorem 6 provides a converse to this.

Theorem 6 (Monotonicity Theorem)**.** *Let $\succsim$ have an adversarial expected utility representation (Y, u) and fix a set $\mathcal{W} \subseteq C(X)$. The following conditions are equivalent:*

(i) The preference $\succsim$ preserves $\succsim_{\mathcal{W}}$.

(ii) For all $F \in \mathcal{F}$, there exists $H \in \mathcal{H}(\hat{Y}(F))$ such that $u(\cdot, H) \in \langle \mathcal{W} \rangle$.

An expected utility representation preserves a given stochastic order if and only if there always exists a (mixed) best response of the adversary such that the utility induced by that best response belongs to the convex cone generated by the stochastic order. The proof that (ii) implies (i) only formalizes the discussion before the theorem; the fact that (ii) implies (i) is more involved. To show this, we first observe that the

[21] Recall that $\hat{Y}(F)$ is the set of pure best responses of the adversary given F.

preference $\succsim$ preserves $\succsim_{\mathcal{W}}$ if and only if for all $F, G, \hat{G} \in \mathcal{F}$ such that $G \succsim_{\mathcal{W}} \hat{G}$, there exists $H \in \mathcal{H}(\hat{Y}(F))$ such that $\int u(x,H)dG(x) \geqslant \int u(x,H)d\hat{G}(x)$. By the Sion minmax theorem, this assertion is equivalent to the statement that there exists $H \in \mathcal{H}(\hat{Y}(F))$ such that $\succsim_{u(\cdot,H)}$ preserves $\succsim_{\mathcal{W}}$. Finally, because $u(\cdot, \hat{y}(F))$ is continuous, Theorem 2 in Castagnoli and Maccheroni [1999]) shows that $u(\cdot, \hat{y}(F)) \in \langle \mathcal{W} \rangle$.

Theorem 6 differs from other monotonicity results in the literature for preferences with concave representations because it characterizes monotonicity for a given representation, instead of constructing a representation with the desired monotonicity properties.[22]

Corollary 2. *Let $\succsim$ have an adversarial forecaster representation (v, σ) and fix a set $\mathcal{W} \subseteq C(X)$. Then $\succsim$ preserves $\succsim_{\mathcal{W}}$ if and only if $v + \sigma(\cdot, F) \in \langle \mathcal{W} \rangle$ for all $F \in \mathcal{F}$.*[23]

Corollary 2 underlies Section 3's characterizations of the optimal distributions in the application to writing a novel (Proposition 3) and the application to risk aversion in the next section. In these applications, preferences are monotone with respect to the MPS order via Corollary 2, and so the optima are the feasible distributions that are maximal in the MPS order.

Similarly, we can apply Theorem 6 to the transport preferences introduced in Section 5. Given $X = [0,1]$, let $\mathcal{F}^* \subseteq \mathcal{F}$ denote the set of full-support and absolutely continuous probability measures on X.

Corollary 3. *Let $X = [0,1]$ and let $\succsim$ be a transport preference such that ϕ is continuously differentiable with $\phi_x(\theta, x)$ increasing in θ, and fix a set $\mathcal{W} \subseteq C(X)$. The preference $\succsim$ preserves $\succsim_{\mathcal{W}}$ if and only if $w_0(x, F) = \int_0^x \phi_x(F(z), z)dz$ is an element of $\langle \mathcal{W} \rangle$ for all $F \in \mathcal{F}^*$.*

Under the assumptions on ϕ, standard results in optimal transport theory show that, at all full-support and absolutely continuous F, the local expected utility of V is $w(x, F) = w_0(x, F) + c_F$, where $c_F \in \mathbb{R}$ is a lottery-dependent constant. Corollary

[22] For example, Proposition 22 in Cerreia-Vioglio [2009] (for preferences with a quasiconcave representation), Theorem 4.2 in Chatterjee and Krishna [2011] (for preferences with a concave and Lipschitz continuous representation), and Theorem S.1 in Sarver [2018] (for preferences with a concave representation) assume that the underline preference preserves a stochastic order.

[23] When X is a compact interval in the real line, this last statement directly follows from Proposition 1 in Cerreia-Vioglio, Maccheroni, and Marinacci [2017] because, as we show in Proposition 9 in Online Appendix V, if V has an adversarial forecaster representation then it is Gâteaux differentiable with derivative $v + \sigma(\cdot, F)$. However, the proof of Theorem 6 is quite different as it does not rely on Gâteaux differentiability.

2 then yields the stated result, where the restriction to the dense set $\mathcal{F}^*$ is sufficient because the local expected utility of V is continuous. Thus, Corollary 3 implies that ordinally independent preferences are monotone with respect to the MPS order if ϕ is convex in x.

Online Appendix IV.C applies our monotonicity results to correlation aversion by examining the case where the adversary can observe the realization of one dimension of the outcome before choosing their action.

Intuitively, this leads the agent to avoid lotteries with a high correlation between dimensions because higher correlation makes it easier for the adversary to make accurate forecasts.

7.2 Application: risk aversion and adversarial forecasters

Now we use the monotonicity result to show how a preference for surprise can alter the agent's higher-order risk preferences. We consider an asymmetric version of the method of moments representation, where the forecaster is asymmetrically concerned about the direction of deviations of the realized moment from the forecast. For simplicity, we let $X = [0, 1]$ and consider only the first moment.[24]

Fix a strictly convex and twice continuously differentiable function $\rho : [-1, 1] \to \mathbb{R}_+$ such that $\rho(0) = 0$, $\rho'(z) < 0$ if $z < 0$, and $\rho'(z) > 0$ if $z > 0$, and consider the preferences over lotteries induced by

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

where the space of parameters coincide with the space of outcomes $Y = X$. These preferences arise from the parametric adversarial forecaster representation with forecast error $\sigma(x, \hat{F}) = \rho(x - \hat{y}(\hat{F}))$ where $\hat{y}(\hat{F})$ is the unique minimzer of $\int \rho(x-y)d\hat{F}(x)$, so by Theorem 2 there are optimal lotteries in $\mathcal{F}$ supported on no more than two points. In particular, $\int \rho(x - \hat{y}(F))dF(x)$ can be interpreted as an index of the dispersion of F, without requiring symmetry. Moreover, the local expected utility of the agent is $w(x, F) = v(x) + \rho\left(x - \hat{y}(F)\right)$, with second derivative $w''(x, F) = v''(x) + \rho''(x - \hat{y}(F))$ that also depends on the lottery F. Therefore, when v is not too concave, so that $w'' \geqslant 0$, Corollary 2 implies that V preserves the MPS order.

[24] It is easy to generalize this to finite or infinite numbers of moments as we did for the quadratic GMM in Section 2.3.

This implies that the optimal distributions have the form $p^*\delta_1 + (1-p^*)\delta_0$ for some $p^* \in [0,1]$. And then the fixed-point characterization of optimality in Proposition 1 can be used to explicitly compute p^*, as we show in Online Appendix IV.A.

Consider the asymmetric loss function $\rho(z) = \lambda(\exp(z) - z)$, $\lambda \geqslant 0$. The relevant statistic is $\hat{y}(F) = \log\left(\int_0^1 \exp(x) dF(x)\right)$, that is, the (normalized) cumulant generating function evaluated at 1. With this loss function the agent prefers a positive surprise $x > \hat{y}(F)$ to a negative surprise $x < \hat{y}(F)$ of the same absolute value. The second derivative of the local expected utility at an arbitrary lottery F is $w''(x,F) = v''(x) + \lambda \exp(x - \hat{y}(F))$, so the agent is more risk averse over outcomes that are concentrated around $\hat{y}(F)$. The n-th order derivative of each local utility is $w^{(n)}(x,F) = v^{(n)}(x) + \lambda \exp(x - \hat{y}(F))$, so for λ high enough, $w^{(n)} > 0$. From Theorem 6, this implies that higher enjoyment for surprise induces preferences over lotteries 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.[25] Here asymmetric preference for surprise is crucial: if the third derivatives of all the local expected utilities of V coincide with those of v, preferences for surprise do not affect higher-order risk aversion. 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, while increasing their degree of prudence.[26]

8 Conclusion

The adversarial forecaster representation arises naturally in many settings. It allows the interpretation of random choice as a preference for surprise, and also allows sharp characterizations of the optimal "amount" (i.e., support size) of randomization and of various monotonicity properties. The more general adversarial expected utility

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

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

representation inherits many of the optimality and monotonicity properties of the adversarial forecaster representation with the advantage of not requiring differentiability. This for example, allows us to consider the case where the adversary has only finitely many actions or where the loss functions has kinks, as in Example 4. In the adversarial forecaster model, the agent likes to be surprised so the adversary wants to minimize forecast error. One can also consider an adversary that tries to maximize forecast error to model agents who suffer anxiety from lotteries that are to predict, as in Caplin and Leahy [2001] and Battigalli, Corrao, and Dufwenberg [2019]. These preferences are the flipped version of our model, and generate a preference for deterministic outcomes. We leave a detailed analysis of this extension for future research.

Online Appendix III extends some of the results on the adversarial forecaster model to adversarial expected utility. Proposition 6 characterizes optimal lotteries by a fixed-point property that extends Proposition 1. Theorem 12 shows that, whenever the adversary has only k many actions, there is an optimal lottery that is a convex combination of no more than k extreme points of the set of feasible lotteries. Theorem 13 leverages this result to improve on the bound on randomization provided in Theorem 2 to $k+m$ when the adversary has k many actions. We believe that these results can be used in applications similar to the ones presented in the main text.

Finally, the adversarial forecaster and adversarial expected utility representations can 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. In ongoing work, we show that standard design problems of optimal allocation and Bayesian persuasion naturally induce adversarial expected utility preferences, so that our results can be applied there.

Appendix I: Sections 2 and 6

This section proves the results in Section 6 and then shows that Theorem 1 of Section 2 follows. It then proves Proposition 2 on the GMM representation. The omitted proofs from this section are in Online Appendix II.A.

Preliminaries

We will make use of the *Bregman divergence*, which is closely related to local expected utility. Fix a continuous V that has a local expected utility. For each $F \in \mathcal{F}$, let $\mathcal{W}_V(F) \subseteq C(X)$ denote the (nonempty) set of local expected utilities of V at F.

Definition 10. Let V be continuous and have a local expected utility. We say that $B : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_+$ is a Bregman divergence for V is

$$B(\tilde{F}, F) = V(F) - V(\tilde{F}) - \int w_F(x) d(F - \tilde{F})(x) \qquad \forall F \in \mathcal{F}$$

for some $w_F \in \mathcal{W}_V(F)$.

Definition 11. We say that $\sigma : X \times \mathcal{F} \to \mathbb{R}_+$ is a *pseudo forecast error* if $\sigma(\cdot, F)$ is continuous for all $F \in \mathcal{F}$, $\sigma(x, \delta_x) = 0$ for all $x \in X$, and if $\int \sigma(x, F) dF(x) \leqslant \int \sigma(x, \hat{F}) dF(x)$ for all $F, \hat{F} \in \mathcal{F}$.

Theorem 7. *Let V be a continuous functional. The following are equivalent:*

(i) V has a local expected utility.

(ii) There exist $v \in C(X)$ and a pseudo forecast error σ such that

$$V(F) = \int v(x) dF(x) + \min_{\hat{F} \in \mathcal{F}} \int \sigma(x, \hat{F}) dF(x) \qquad \forall F \in \mathcal{F}. \tag{14}$$

(iii) There is a separable metric space Y and a continuous function $u : X \times Y \to \mathbb{R}$ such that

$$V(F) = \min_{y \in Y} \int u(x, y) dF(x) \qquad \forall F \in \mathcal{F}.$$

If any of these conditions holds, then

1. *v is uniquely defined by $v(x) = V(\delta_x)$;*

2. *σ satisfies 14 if and only if $\sigma(x, F) = B(\delta_x, F)$ for some Bregman divergence.*

This result implies that even if multiple forecast errors are consistent with (14), the induced surprise function Σ is uniquely defined by $\Sigma(F) = V(F) - \int v(x) dF(x)$.

Lemma 1. *Suppose $F^n \to F$ and that $w^n \to w$. Then $\int w^n(x) dF^n(x) \to \int w(x) dF(x)$. Moreover, if V is continuous with continuous local expected utility and if each w^n is a local expected utility for F^n, then w is a local expected utility for F.*

Lemma 2. *Let V have a continuous local expected utility w. For all $F, \tilde{F}, \overline{F} \in \mathcal{F}$ such that there exists $\mu > 0$ with $F + \mu(\tilde{F} - \overline{F}) \in \mathcal{F}$, we have*

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

Section 6

Proof of Proposition 5. (i) implies (ii). Define $\mathcal{W}_{v,\sigma} = cl\left(\{v + \sigma(\cdot, F)\}_{F \in \mathcal{F}}\right)$, where cl denotes the closure operation, and $M = \max_{F \in \mathcal{F}} |V(F)|$. For every $F \in \mathcal{F}$, we have $\max_{x \in X} |v(x) + \sigma(x, F)| \leqslant M$, so $\max_{x \in X} |w(x)| \leqslant M$ for all $w \in \mathcal{W}_{v,\sigma}$. Next, because X is compact, v is uniformly continuous and the family $\{\sigma(\cdot, F)\}_{F \in \mathcal{F}}$ is uniformly equicontinuous over X, so there is a continuous function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\omega(0) = 0$ and$|v(x) + \sigma(x, F) - v(x') - \sigma(x', F)| \leqslant \omega(d(x, x'))$ for every $x, x' \in X$ and $F \in \mathcal{F}$. Thus the family of functions $\mathcal{W}_{v,\sigma}$ is equicontinuous over X and, by the Arzaela-Ascoli theorem, a compact metric space. This implies that $V(F) = \min_{w \in \mathcal{W}_{v,\sigma}} \int w(x) dF(x)$, and, by setting $Y = \mathcal{W}_{v,\sigma}$ and $u(x, y) = y(x)$, that $\succsim$ admits a representation as in equation 12.

(ii) implies (i). By assumption $\succsim$ has a representation V as in equation 12. Define $\mathcal{W} = \{u(\cdot, y)\}_{y \in Y} \subseteq C(X)$. Continuity of u implies that $\mathcal{W}$ is uniformly bounded and equicontinuous, hence compact. Next, for every F fix an arbitrary $\hat{y}(F) \in \operatorname{argmin}_{y \in Y} \int u(x, y) dF(x)$ and define $w(x, F) = u(x, \hat{y}(F))$ for all x and F. Because $\{w(\cdot, F)\}_{F \in \mathcal{F}} \subseteq \mathcal{W}$, the family of function $\{w(\cdot, F)\}_{F \in \mathcal{F}}$ is equicontinuous in x. Moreover, by construction $w(\cdot, F)$ is a local expected utility of V for every F, that is $w(\cdot, F) \in \mathcal{W}_V(F)$. Theorem 7 thus implies there is a $v \in C(X)$ and a pseudo surprise function σ such that V can be written as in equation 14. In particular, by point 2 of Theorem 7, σ satisfies 14 if and only if $\sigma(x, F) = B(\delta_x, F)$ for some Bregman divergence of V. Let B_w be the Bregman divergence induced by $\{w(\cdot, F)\}_{F \in \mathcal{F}}$ and let $\sigma(x, F) = B_w(\delta_x, F) = w(x, F) - v(x)$, hence the family of functions $\{\sigma(\cdot, F)\}_{F \in \mathcal{F}}$ is equicontinuous. This implies that $\succsim$ has a weak adversarial forecaster representation.

We now prove the main representation result for the adversarial forecaster model; it implies Theorem 5, which asserts the equivalence of conditions (i) and (iii).

Theorem 8. *Consider a preference $\succsim$ over $\mathcal{F}$. The following are equivalent:*

(i) $\succsim$ has an adversarial forecaster representation

(ii) $\succsim$ *can be represented by a function* V *with continuous local expected utility*

(iii) $\succsim$ *has an adversarial expected utility representation that satisfies uniqueness.*

Proof of Theorem 8.

(i) implies (ii). Let v and σ correspond to the adversarial forecaster representation of $\succsim$. The map $w_V : \mathcal{F} \to C(X)$ given by $w_V(x, F) = v(x) + \sigma(x, F)$ is a continuous local utility of $V(F) = \min_{\tilde{F}\in\mathcal{F}} \int w_V(x, \tilde{F})dF(x)$, so that V represents $\succsim$ and has a continuous local expected utility.

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

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

where both the equality and the inequality follow from the fact that $W_V(\cdot, F)$ is a local expected utility of V at F and the definition of Y. This implies that $V(F) = \min_{y\in Y} \int u(x, y)dF(x) =$. It remains to show that $\int u(x, y)dF(x)$ has a unique minimum over y. Suppose that for some F there is a $\tilde{F} \neq F$ such that $V(F) = \int w_V(x, \tilde{F})dF(x)$. For every $\lambda \in (0, 1)$, define $F_\lambda = \lambda\tilde{F} + (1 - \lambda)F$. Then because V is concave and the w_V are local expected utility functions, for all $\lambda \in [0, 1]$

$$\begin{aligned}\lambda V(\tilde{F}) + (1-\lambda)V(F) \leqslant V(F_\lambda) &\leqslant \lambda \int w_V(x, \tilde{F})d\tilde{F}(x) + (1-\lambda)\int w_V(x, \tilde{F})dF(x) \\ &= \lambda V(\tilde{F}) + (1-\lambda)V(F),\end{aligned}$$

so that

$$V(F_\lambda) = \int w_V(x, \tilde{F})dF_\lambda(x) \tag{15}$$

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

Moreover,

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

so that, by rearranging the terms,

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

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

$$V(\tilde{F}) = \int w_V(x, F_\mu) d\tilde{F}(x). \tag{16}$$

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

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

where the first equality follows by (16), the second equality by Lemma 2, the inequality by the properties of w_V, the third equality by (15), and the last equality by the properties of w_V again. This implies that $w_V(\tilde{x}, F_\mu) \leqslant w_V(\tilde{x}, \tilde{F})$. Similarly,

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

where the first equality follows by the properties of w_V, the second equality follows by Lemma 2, the inequality by the properties of w_V, and the third and the last equality by (16). This implies that $w_V(\tilde{x}, \tilde{F}) \leqslant w_V(\tilde{x}, F_\mu)$ and we conclude that $w_V(\tilde{x}, F_\mu) = w_V(\tilde{x}, \tilde{F})$. Since this is true for all $\mu > 0$ and w_V is continuous it holds also in the limit: $w_V(\tilde{x}, F) = w_V(\tilde{x}, \tilde{F})$. Given that $\tilde{x}$ was arbitrary, the minimizer is unique, which proves that V has an adversarial expected utility representation that satisfies uniqueness.

(iii) implies (i). We next show that if $\succsim$ has an adversarial expected utility representation that satisfies uniqueness, then it has an adversarial forecaster representation. Let Y and u denote the adversarial expected utility representation of $\succsim$. For all $F \in \mathcal{F}$, let $\hat{y}(F) \in Y$ denote the unique minimizer of $\int u(x, \tilde{y})dF(x)$. Define $v(x) = \min_{y \in Y} u(x, y)$, $\sigma(x, F) = u(x, y(F)) - v(x)$, and $V(F) = \int v(x)dF(x) + \int \sigma(x, F)dF(x)$. Observe that, by construction, we have $V(F) = \min_{y \in Y} \int u(x, y)dF(x)$, hence V represents $\succsim$. Finally, fix $F, \tilde{F} \in \mathcal{F}$ and observe that

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

showing that σ is a forecast error. ■

Section 2

Proof of Theorem 1. This is the equivalence between (i) and (ii) in Theorem 8. ■

Proof of Proposition 1. (If). This direction follows immediately from the discussion before the proposition. See Propositions 6 in Online Appendix III.A and 8 in Online Appendix V for alternative proofs that can also be applied to the more general adversarial expected utility model.

(Only if). Fix an optimal lottery F^* for V over $\overline{\mathcal{F}}$ and assume that there exists $\hat{F}$ that is strictly better than F^* for an expected utility agent with utility $v + \sigma(\cdot, F^*)$. Due to convexity of $\overline{\mathcal{F}}$, F^* is also optimal for V when restricted on the segment between F^* and $\hat{F}$, implying that the directional derivative of V at F^* in direction $\hat{F}$ is negative, contradicting $\hat{F}$ strictly preferred to F^* for the expected utility $v+\sigma(\cdot, F)$. ■

Proof of Proposition 2. This follows from the following three lemmas. The first two are standard and we relegate their proof to Online Appendix II.

Lemma 3. *$\sigma(x, F)$ defined by a methods of moments forecast is a forecast error.*

Lemma 4. *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 $(\delta_z - F)$ *given by*

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

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

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

Lemma 5. *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 4 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 for some $\hat{s}$ we have $\int h(x,\hat{s})dM(x) \neq 0$. Since h is continuous there is an open set $\tilde{S} \subseteq S$ such that $\hat{s} \in \tilde{S}$ and $\int h(x,s)dM(x) \neq 0$ for all $s \in \tilde{S}$. Since μ assigns positive probability to open sets of S this implies that

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

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

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

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

Appendix II: Section 4

The proofs of the ancillary results stated in this section of appendix in Online Appendix II.B, together with the proof of Corollary 1 stated in Section 4 of the main text.

Appendix II.A: General characterization

In this section, we fix an arbitrary adversarial expected utility representation (Y, u). The parametric adversarial forecaster representation $(Y, v, \hat{\sigma})$ considered in Section 4 is the case where $u(x, y) = v(x) + \hat{\sigma}(x, y)$.

First, we consider arbitrary convex and compact subsets $\overline{\mathcal{F}} \subseteq \mathcal{F}$ of feasible lotteries. Let $\mathcal{H}$ denote the set of probability measures over Y, and $ext(\overline{\mathcal{F}})$ the set of extreme points of $\overline{\mathcal{F}}$. By Choquet's theorem, for all $F \in \overline{\mathcal{F}}$, there exists $\lambda \in \Delta\left(ext\left(\overline{\mathcal{F}}\right)\right)$ such that $F = \int \tilde{F} d\lambda(\tilde{F})$. For every $\overline{\mathcal{F}}$, define $V^*\left(\overline{\mathcal{F}}\right) = \max_{F \in \overline{\mathcal{F}}} V(F)$. By Sion's minmax theorem,

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

Next we characterize the optimal lotteries given an arbitrary feasibility set $\overline{\mathcal{F}}$. Let $\Lambda_F \subseteq \Delta\left(ext\left(\overline{\mathcal{F}}\right)\right)$ be the set of probability measures over extreme points that satisfy $F = \int \tilde{F} d\lambda\left(\tilde{F}\right)$ for F.

Theorem 9. *Fix $\hat{H} \in \arg\min_{H \in \mathcal{H}} \max_{F \in ext\left(\overline{\mathcal{F}}\right)} \int\int u(x, y)\, dF(x)\, dH(y)$. 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 $\tilde{F} \in \bigcup_{\lambda \in \Lambda_{\hat{F}}} \operatorname{supp} \lambda$, $V(\hat{F}) = \int\int u(x, y)\, d\tilde{F}(x)\, d\hat{H}(y)$.*

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

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

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

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

By combining the first two chains of inequalities, we have

$$\int\int u\left(x,y\right) d\hat{F}\left(x\right) d\hat{H}\left(y\right) \geqslant \int\int u\left(x,y\right) d\tilde{F}\left(x\right) d\hat{H}\left(y\right) \qquad \forall \tilde{F} \in ext\left(\overline{\mathcal{F}}\right). \tag{17}$$

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

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

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

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

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

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

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

which contradicts equation (17).

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

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

Note that when $\overline{\mathcal{F}} = \Delta\left(\bar{X}\right)$ for some closed subset $\bar{X}$, the extreme points $ext\left(\overline{\mathcal{F}}\right) = \bar{X}$ are simply point masses over the set of feasible outcomes. In this case, Theorem 9 implies that F is optimal if and only if $V(F) \geqslant \int u(x,y)d\hat{H}(y)$ for all $x \in \overline{X}$, with equality for $x \in \operatorname{supp} F$.

Appendix II.B: Section 4

For the rest of this section, we fix a closed subset $\overline{X} \subseteq X$ and a finite collection of functions $\Gamma = \{g_1, ..., g_k\} \subset C(\overline{X})$. As in the main text, we consider $\mathcal{F}_\Gamma(\overline{X}) \subseteq \mathcal{F}$. By Theorem 2.1 in Winkler [1988], $\tilde{F} \in ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$ if and only if $\tilde{F} \in \mathcal{F}_\Gamma(\overline{X})$ and $\tilde{F} = \sum_{i=1}^{p} \alpha_i \delta_{x_i}$ for some $p \leqslant k+1$, $\alpha \in \Delta\left(\{1, ..., p\}\right)$, and $\{x_1, ..., x_p\} \subseteq \bar{X}$ such that the vectors $\{\left(g_1\left(x_i\right), ..., g_k\left(x_i\right), 1\right)\}_{i=1}^{p}$ are linearly independent. For every finite subset of extreme points $\mathcal{E} \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$, define

$$
\overline{X}_{\mathcal{E}} = \bigcup_{\tilde{F} \in \mathcal{E}} \operatorname{supp} \tilde{F} \subseteq \overline{X},
$$

which is finite from Winkler's theorem. We identify $co(\mathcal{E})$ with the subset of $\mathcal{F}_\Gamma(\overline{X})$ composed of all convex combinations of extreme points in $\mathcal{E}$. Recall that $\hat{Y}(F) \equiv \operatorname{argmin}_{y \in Y} \int u(x,y)dF(x)$, and that (Y, u) satisfies the uniqueness property if $\hat{Y}(F)$ is a singleton for all $F \in \mathcal{F}$ (see Definition 8).

Theorem 10. *Fix a finite set $\mathcal{E} \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$, and suppose that Y has the structure of an m-dimensional manifold with boundary, that u is continuously differentiable in y, and that Y and u satisfy the uniqueness property. We have:*

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

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

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

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

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

where, for every $y \in Y$, $u(\tilde{F}_\ell, y) = \int u(x,y) d\tilde{F}_\ell(x)$ and $w(\tilde{F}_\ell) = \int w(x) d\tilde{F}_\ell(x)$. For every $(y,v) \in Y \times \mathbb{R}$, the derivative of $U_{\overline{\mathcal{E}}}$ with respect to $w \in \mathbb{R}^{\overline{X}_{\mathcal{E}}}$ is a $(m+2) \times r$ matrix whose ℓ-th row coincides with the probability vector $\tilde{F}_\ell$, and because the $\left\{\tilde{F}_1, ..., \tilde{F}_{m+2}\right\}$ are extreme points of $\mathcal{F}_\Gamma(\overline{X})$, this matrix has full rank, so the total derivative of $U_{\overline{\mathcal{E}}}$ has full rank as well. Hence by the parametric transverality theorem,[27] for an open dense full measure subset of $\mathbb{R}^{\overline{X}_{\mathcal{E}}}$, denoted $\mathcal{W}(\overline{\mathcal{E}})$, the manifold $(y,v) \mapsto u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell)$ intersects zero transversally. Since $\dim(Y \times \mathbb{R}) < m+2$,there is no (y,v) that solve $u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell) = 0$ for all $\ell \leqslant m+2$. And since $\mathcal{E}$ has finitely many subsets $\overline{\mathcal{E}}$ of $m+2$ extreme points, the intersection $\mathcal{W} = \bigcap_{\overline{\mathcal{E}}} \mathcal{W}(\overline{\mathcal{E}})$ is open dense and of full measure since it is the finite intersection of full measure sets. Thus, for $w \in \mathcal{W}$ and for all $y \in Y$ and $v \in \mathbb{R}$, $u(\tilde{F}_\ell, y) - v + w(\tilde{F}_\ell) = 0$ for at most $m+1$ extreme points in $\mathcal{E}$.

Next, fix $w \in \mathcal{W}$, $F^* \in \operatorname{argmax}_{F \in co(\mathcal{E})} V_w$, and $\lambda \in \Lambda_{F^*}$. By Theorem 9, for all $\tilde{F} \in \operatorname{supp} \lambda \subseteq \mathcal{E}$, we have $u(\tilde{F}, H_w) - V_w(F^*) + w(\tilde{F}) = 0$. By the previous part of the proof and Theorem 9, we then have $|\operatorname{supp} \lambda| \leqslant m+1$. Therefore, F_w is the linear combination of up to $m+1$ extreme points in $\mathcal{E}$. Each $\tilde{F} \in \mathcal{E}$ is supported on up to $k+1$ points of $\overline{X}_{\mathcal{E}}$, so F_w is supported on up to $(m+1)(k+1)$ points of $\overline{X}_{\mathcal{E}}$.

Now we prove point 2. Because $\mathcal{W}$ is dense in $\mathbb{R}^{\overline{X}_{\mathcal{E}}}$, there exists a sequence $w^n \in \mathcal{W}$ such that $w^n(x) \to 0$ for all $x \in \overline{X}_{\mathcal{E}}$, and a sequence of corresponding optimal lotteries F^n with support of no more than $(m+1)(k+1)$ points of $\overline{X}_{\mathcal{E}}$. Choose a convergent subsequence of $F^n \to F$, and observe that lotteries with no more than $(m+1)(k+1)$

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

points of support cannot converge weakly to a lottery with larger support. Finally, because V_w is continuous with respect to w, the Berge Maximum Theorem implies that F solves $\max_{F \in co(\mathcal{E})} V_0(F)$, concluding the proof. ■

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

Proof of Theorem 2. Fix a parametric adversarial forecaster representation $(Y, v, \hat{\sigma})$, and define $u = v + \sigma$. By Definition 4, the adversarial expected utility representation (Y, u) is such that Y has the structure of an m-dimensional manifold with boundary, u is continuously differentiable in y, and Y and u satisfy the uniqueness property. By Theorem 10 and Lemma 6, there exists a solution F^* that is supported on no more than $(k+1)(m+1)$ points of $\overline{X}$. ■

Proof of Theorem 3. Stationarity implies that $H(x,x)$ is constant, so the directional derivatives from Lemma 4 simplify to

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

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

$$\int H(z,x)dF(x) \geqslant \int H(x,\tilde{x})dF(x)dF(\tilde{x}) \text{ for all } z \in X. \tag{18}$$

This implies the complementary slackness condition: if there exists $z \in A$ such that z satisfies (18) with strict inequality, then $F(A) = 0$.[28]

Next we show that for any $0 < a \leqslant 1$ and interval $A = [0, a)$ there is $z \in A$ such that $\int H(z,x)dF(x)) = \int H(x,\tilde{x})dF(x)dF(\tilde{x})$. By continuity this implies

[28] If there is such z with $F(A) > 0$, then by continuity of H, there is an open set $\tilde{A} \subseteq A$ containing z with $F(\tilde{A}) > 0$, and every $x \in \tilde{A}$ satisfies (18) with strict inequality. Then $\int H(x,\tilde{x})dF(x)dF(\tilde{x}) = \int_{\tilde{A}} \int_X H(x,\tilde{x})dF(\tilde{x})dF(x) + \int_{\tilde{A}^c} \int_X H(x,\tilde{x})dF(\tilde{x})dF(x) > F(\tilde{A}) \int H(x,\tilde{x})dF(x)dF(\tilde{x}) + (1 - F(\tilde{A})) \int H(x,\tilde{x})dF(x)dF(\tilde{x}) = \int H(x,\tilde{x})dF(x)dF(\tilde{x})$, a contradiction.

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

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

Appendix IV: Section 7

Recall that $\hat{Y}(F) = \operatorname{argmin}_{y \in Y} \int u(x,y)dF(x)$, and let $\mathcal{H}(\hat{Y}(F)) \subseteq \mathcal{H}$ denote the probability measures over $\hat{Y}(F)$.

Proof of Theorem 6. We only prove the equivalence between (i) and (ii) since the other implications are explained in the main text. (i) implies (ii). As a preliminary step we show that, for every $F \in \mathcal{F}$ and for every $G, \hat{G} \in \mathcal{F}$ such that $G \gtrsim_{\mathcal{W}} \hat{G}$, there exists $y \in \hat{Y}(F)$ such that $\int u(x,y)d\hat{G}(x) \leqslant \int u(x,y)dG(x)$. Observe that $\lambda G + (1-\lambda)F \gtrsim_{\mathcal{W}} \lambda\hat{G} + (1-\lambda)F$, for all $\lambda \in (0,1]$. By hypothesis, this implies that $V(\lambda G + (1-\lambda)F) \geqslant V(\lambda\hat{G} + (1-\lambda)F)$, for all $\lambda \in (0,1]$. Next, consider a sequence $\lambda_n \to 0$. For every $n \in \mathbb{N}$, fix two any $\hat{y}_n \in \hat{Y}(\lambda_n\hat{G} + (1-\lambda_n)F)$, and $y_n \in \hat{Y}(\lambda_n G + (1-\lambda_n)F)$. Observe that for every $n \in \mathbb{N}$,

$$\begin{aligned}\int u(x,\hat{y}_n)d(\lambda_n\hat{G} + (1-\lambda_n)F)(x) &= V(\lambda_n\hat{G} + (1-\lambda_n)F) \\ \leqslant V(\lambda_n G + (1-\lambda_n)F) &= \int u(x,y_n)d(\lambda_n G + (1-\lambda_n)F)(x) \\ &\leqslant \int u(x,\hat{y}_n)d(\lambda_n G + (1-\lambda_n)F)(x)\end{aligned}$$

where the last inequality follows since $y_n \in \hat{Y}(\lambda_n G + (1-\lambda_n)F)$. This implies that

$$\lambda_n \int u(x,\hat{y}_n)d\hat{G}(x) + (1-\lambda_n)\int u(x,\hat{y}_n)dF(x) \leqslant \lambda_n \int u(x,\hat{y}_n)dG(x) + (1-\lambda_n)\int u(x,\hat{y}_n)dF(x),$$

which in turn gives $\int u(x,\hat{y}_n)d\hat{G}(x) \leqslant \int u(x,\hat{y}_n)dG(x)$. Take a subsequence $\hat{y}_n$ converging to y. By Lemma 1 $y \in \hat{Y}(F)$ and $\int u(x,y)d\hat{G}(x) \leqslant \int u(x,y)dG(x)$ as desired.

Next, fix $F \in \mathcal{F}$ and define the subset of the signed measures on X in the weak topology $\mathcal{M} = \left\{G - \hat{G} : G, \hat{G} \in \mathcal{F}, G \gtrsim_{\mathcal{W}} \hat{G}\right\}$; for every $M \in \mathcal{M}$, there exists $y \in \hat{Y}(F)$ such that $\int u(x,y)dM(x) \geqslant 0$. Let $\mathcal{U}(F)$ denote the convex hull of $\left\{u(\cdot,y) : y \in \hat{Y}(F)\right\}$. Since $\hat{Y}(F)$ is compact so is $\mathcal{U}(F)$, so $\max_{w\in\mathcal{U}(F)} \int w(x)dM(x)$ exists, and is nonnegative for all $M \in \mathcal{M}$. Thus $\inf_{M\in\mathcal{M}} \max_{w\in\mathcal{U}(F)} \int w(x)dM(x) \geqslant 0$. Now we show that $\mathcal{M}$ is convex and compact. Fix $M, M' \in \mathcal{M}$ and $\lambda \in [0,1]$, and probability measures $G, G', \hat{G}, \hat{G}'$ such that $G \gtrsim_{\mathcal{W}} \hat{G}$, $G' \gtrsim_{\mathcal{W}} \hat{G}'$, such that $M = G - \hat{G}$ and $M' = G' - \hat{G}'$. From the definition of $\gtrsim_{\mathcal{W}}$, $\lambda G + (1-\lambda)G' \gtrsim_{\mathcal{W}} \lambda\hat{G} + (1-\lambda)\hat{G}'$, so $\lambda M + (1-\lambda)M' \in \mathcal{M}$. Moreover, the subset in $\mathcal{F}\times\mathcal{F}$ of points $G, \hat{G}$ such that $G \gtrsim_{\mathcal{W}} \hat{G}$ is closed so it is compact. As subtraction is continuous, $\mathcal{M}$ is the continuous image of a compact set, so it is also compact. Given that $\mathcal{U}(F)$ and $\mathcal{M}$ are compact and convex, and the objective function is bilinear and continuous in each argument separately, the Sion minmax Theorem implies that $\max_{w\in\mathcal{U}(F)} \min_{M\in\mathcal{M}} \int w(x)dM(x) \geqslant 0$.

Letting $v \in \mathcal{U}(F)$ be a solution, we see that $G \gtrsim_{\mathcal{W}} \hat{G}$ implies $\int v(x)d\left(G - \hat{G}\right)(x) \geqslant 0$, that is $\gtrsim_v$ preserves $\gtrsim_{\mathcal{W}}$. Hence, because v continuous, Theorem 2 in Castagnoli and Maccheroni [1999] implies that $v \in \langle\mathcal{W}\rangle$.

(ii) implies (i). Consider $F, G \in \mathcal{F}$ such that $F \gtrsim_{\mathcal{W}} G$, and a probability distribution H over $\hat{Y}(F)$ such that $\int u(x,y)dH(y) \in \langle\mathcal{W}\rangle$. Then $y \in \hat{Y}(F)$ implies $V(G) \leqslant \int u(x,y)dG(x)$ and $\int u(x,y)dF(x) = V(F)$. By Fubini's theorem this implies $V(G) \leqslant \int\int u(x,y)dH(y)dG(x)$ and $\int\int u(x,y)dH(y)dF(x) = V(F)$. Since $\int u(x,y)dH(y) \in \langle\mathcal{W}\rangle$ and $G \preceq_{\mathcal{W}} F$, it follows that

$$V(G) \leqslant \int\int u(x,y)dH(y)dG(x) \leqslant \int\int u(x,y)dH(y)dF(x) = V(F). \quad \blacksquare$$

References

Aliprantis, C. D. and K. C. Border (2006). *Infinite dimensional analysis*. Springer.

Battigalli, P., R. Corrao, and M. Dufwenberg (2019). "Incorporating belief-dependent motivation in games." *Journal of Economic Behavior & Organization*, 167, 185–218.

Caplin, A. and J. Leahy (2001). "Psychological expected utility theory and anticipatory feelings." *The Quarterly Journal of Economics*, 116, 55–79.

Castagnoli, E. and F. Maccheroni (1999). *Generalized stochastic dominance and unanimous preferences*. Pitagora.

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.

— (2020). "An explicit representation for disappointment aversion and other betweenness preferences." *Theoretical Economics*, 15, 1509–1546.

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. Levine (2022). "On Concave Functions over Lotteries."

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.

Frankel, A. and E. Kamenica (2019). "Quantifying information and uncertainty." *American Economic Review*, 109, 3650–80.

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.

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

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

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

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

Hong, C. S. and W. S. Waller (1986). "Empirical tests of weighted utility theory." *Journal of mathematical psychology*, 30, 55–72.

Kőszegi, B. and M. Rabin (2006). "A model of reference-dependent preferences." *The Quarterly Journal of Economics*, 121, 1133–1165.

Levine, D. K., S. Modica, and J. Sun (2021). "Twin Peaks: Expressive Voting and Soccer Hooliganism."

Loseto, M. and A. Lucia (2021). "Deliberate Randomization under Risk." *Available at SSRN 3811073*.

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

Machina, M. J. (1982). "" Expected Utility" Analysis without the Independence Axiom." *Econometrica*, 277–323.

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

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

Quiggin, J. (1982). "A theory of anticipated utility." *Journal of economic behavior & organization*, 3, 323–343.

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

Sarver, T. (2018). "Dynamic Mixture-Averse Preferences." *Econometrica*, 86, 1347–1382.

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

Yaari, M. E. (1987). "The dual theory of choice under risk." *Econometrica: Journal of the Econometric Society*, 95–115.

Online Appendix I: Proofs omitted from the main appendix

This appendix gives proofs of the secondary results stated in the main text.

Online Appendix I.A: Section 3

We first prove Proposition 3, and then spell out the details for the linear case $g(d) = d$ that were sketched in the main text.

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.

Proof of Proposition 3. First, observe that Proposition 1 implies that that $F^* \in \operatorname{argmax}_{F \in \overline{\mathcal{F}}} V_\beta(F)$ if and only if $F^* \in \operatorname{argmax}_{F \in \mathcal{F}(x_0)} \int w_\beta(x, F^*)dF(x)$.

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

$$\Delta(p_F^*, q^*) = \left\{ F_\Delta \in \Delta(S) : \int p^2 dF(p) = p_F^*, \int p^2 dF(p) = q^* \right\}.$$

Consider the maximization problem:

$$\max_{F_\Delta \in \Delta(p_F^*, q^*)} \int g(p - p^2)dF_\Delta(p). \tag{19}$$

If F_Δ is feasible for Problem 3, 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 19, so any solution F_Δ of Problem 19 is also a solution of Problem 3. Finally, observe that $\Delta(p_F^*, q^*)$ is a moment set with $k = 2$ moment conditions. The objective function of Problem 19 is linear in F_Δ, so it follows from Theorem 2.1. in Winkler [1988] that there is solution of Problem 19, and hence of Problem 3, that is supported on no more than three points of $\Delta(S)$, concluding the proof of the first statement.

Next, assume that (p_F^*, F_Δ^*) is optimal, that is there exists an optimal $F^* \in \overline{\mathcal{F}}$ whose marginals are given by (p_F^*, F_Δ^*). By the initial claim and equation 4, (p_F^*, F_Δ^*)

solve

$$\max_{p\in\overline{\Delta},F_\Delta\in\Delta(S):\int\tilde{p}dF(\tilde{p})=p}\left\{p\tilde{v}+(1-\beta)g'\left(D_2(F^*)\right)\int(\tilde{p}^2-p^2)dF_\Delta(\tilde{p})+\beta\int g(\tilde{p}-\tilde{p}^2)dF_\Delta(p)\right\}$$

$$=\max_{p\in\overline{\Delta}}\left\{p\tilde{v}-(1-\beta)g'\left(D_2(F^*)\right)p^2+\max_{F_\Delta:\int\tilde{p}dF(\tilde{p})=p}\left[\left[\int(1-\beta)g'\left(D_2(F^*)\right)\tilde{p}^2+\beta g(\tilde{p}-\tilde{p}^2)dF_\Delta(\tilde{p})\right]\right]\right\} \tag{20}$$

$$=\max_{p\in\overline{\Delta}}\left\{p\tilde{v}-(1-\beta)g'\left(D_2(F^*)\right)p^2+cav(\xi_{\beta,F^*})(p)\right\}$$

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

1. When $\beta\geqslant\overline{\beta}$, ξ_{β,F^*} is strictly concave so that $cav(\xi_{\beta,F^*})=\xi_{\beta,F^*}$. By Corollary 2 in Kamenica and Gentzkow [2011], the inner maximization problem in equation 20 is uniquely solved by $F_\Delta=\delta_p$, that is, no disclosure is uniquely optimal. This implies that $F^*_\Delta=\delta_{p^*_F}$. Next, we have that $p\tilde{v}-(1-\beta)g'\left(D_2(F^*)\right)p^2+\xi_{\beta,F^*}(p)=p\tilde{v}+\beta g(p-p^2)$. Given that the optimal (p^*_F,F^*_Δ) are arbitrary, the statement follows.

2. When $\beta\leqslant\underline{\beta}$, ξ_{β,F^*} is strictly convex. By Corollary 2 in Kamenica and Gentzkow [2011], the inner maximization problem in equation 20 is uniquely solved by $F_\Delta=(1-p)\delta_0+p\delta_1$, that is, full disclosure is uniquely optimal, and $cav(\xi_{\beta,F^*})(\tilde{p})=(1-\beta)g'\left(D_2(F^*)\right)\tilde{p}$. This implies that $F^*_\Delta=(1-p^*_F)\delta_0+p^*_F\delta_1$. Next, we have that $p\tilde{v}-(1-\beta)g'\left(D_2(F^*)\right)p^2+cav(\xi_{\beta,F^*})(p)=p\tilde{v}+(1-\beta)g'(D_2(F^*))(p-p^2)$. Given that the optimal (p^*_F,F^*_Δ) are arbitrary, the statement follows. ■

The linear case Consider the setting of Section 3 with an arbitrary finite state space Ω and $X=\Omega\times\Delta(\Omega)$. As before, the broadcaster chooses a joint distribution $F\in\mathcal{F}$ over states and conditional beliefs of the watcher, where the feasible joint distributions are those such that the marginal over states is the feasible set $\overline{\Delta}\subseteq\Delta(\Omega)$ and the conditional distribution over states given the belief p is equal to p itself.

The preferences of the watcher over joint distributions of states and beliefs have an adversarial forecaster representation, where preferences over states are given by utility function $v\in\mathbb{R}^\Omega$, and the forecast error given the realization $x=(\omega,p)$ and the forecast $\hat{F}$ of the adversary is

$$\sigma_\beta((\omega, p), \hat{F}) = (1-\beta)\sigma_0(p, \hat{F}_\Delta) + \beta\sigma_1(\omega, \hat{F}(\cdot|p)).$$

Here $\hat{F}_\Delta$ and $\hat{F}(\cdot|p)$ are respectively the marginal distribution over $\Delta(\Omega)$ and the conditional distribution over Ω given p, while σ_0 and σ_1 are forecast errors for the outcome spaces $X_0 = \Delta(\Omega)$ and $X_1 = \Omega$ respectively, and $\beta \in [0,1]$ a parameter capturing the relative importance of interim and ex post surprise.

Clearly, σ_β satisfies all the properties of Definition 1. Indeed,

$$\sigma_\beta((\omega, p), \delta_{(\omega,p)}) = (1-\beta)\sigma_0(p, \delta_p) + \beta\sigma_1(\omega, \delta_\omega) = 0$$

because σ_0 and σ_1 are forecast errors, and for every $\hat{F} \in \mathcal{F}$,

$$\begin{aligned}\int \sigma_\beta(x, F)dF(x) &= (1-\beta)\int \sigma_0(p, F_\Delta)dF_\Delta(p) + \beta \int\int \sigma_1(\omega, F(\cdot|p))dF(\omega|p)dF_\Delta(p) \\ &\leqslant \int \sigma_\beta(x, \hat{F})dF(x)\end{aligned}$$

With this, the preferences of the watcher over joint lotteries F are given by $V_\beta(F) = \int v(s)dF(\omega,p) + \min_{\hat{F}\in\mathcal{F}} \int \sigma_\beta((\omega,p),\hat{F})dF(\omega,p)$. The broadcaster solves $\max_{F\in\overline{\mathcal{F}}} V(F)$.

Next, consider the binary state case $\Omega = \{0,1\}$, $\Delta(\Omega) = [0,1]$, with $\overline{\Delta} = [0,1]$ and the forecast errors: $\sigma_0(p, \hat{F}_\Delta) = \frac{1}{2}(p - \int \tilde{p}d\hat{F}_\Delta(\tilde{p}))^2$ and $\sigma_1(\omega, \hat{p}) = (s - \hat{p})^2$. Also assume that the watcher gets utility $\tilde{v} \in \mathbb{R}$ when the state is equal to $\omega = 1$. For every feasible lottery $F \in \overline{\mathcal{F}}$ let $p_F \in [0,1]$ denote induced probability that $\omega = 1$ and let F_Δ the marginal over $\Delta(\Omega)$. The definition of $\mathcal{F}$ implies that $p_F = \int p dF_\Delta(p)$. The total payoff of the watcher simplifies to

$$\begin{aligned}V_\beta(F) &= \tilde{v}p_F + (1-\beta)\int (p-p_F)^2 dF_\Delta(p) + \beta\int p(1-p)dF_\Delta(p) \\ &= p_F(\tilde{v}+\beta) - p_F^2(1-\beta) + \int (1-2\beta)p^2 dF_\Delta(p),\end{aligned}$$

which is W's payoff in Section 3 when g is linear $g(d) = d$. Therefore, the maximization problem of the broadcaster simplifies to

$$\max_{F\in\overline{\mathcal{F}}} V_\beta(F) = \max_{p\in[0,1]} \left\{ p(\tilde{v}+\beta) - p^2(1-\beta) + \max_{F_\Delta \in \Delta[0,1]: \int \tilde{p}dF_\Delta(\tilde{p}) = p} \int (1-2\beta)\tilde{p}^2 dF_\Delta(\tilde{p}) \right\}$$

When $\beta < 1/2$, the integrand in the inner maximization is strictly convex, so full disclosure is uniquely optimal. When $\beta > 1/2$, the integrand in the inner maximization is strictly concave, so that no disclosure is uniquely optimal. When $\beta = 1/2$, then the corresponding term disappears, and the watcher is indifferent over all the information structures. And simple computations show that $p_F^* = \max\left\{0, \min\left\{1, \frac{\tilde{v}+\max\{\beta,1-\beta\}}{2\max\{\beta,1-\beta\}}\right\}\right\}$ solves the outer maximization problem. $\triangle$

Online Appendix I.B: Section 5

Proof of Theorem 4. Let $\succsim$ is a transport preference with a continuously differentiable ϕ, and recall that $Y \subseteq C(X)$ is the set of continuous functions such that $\int \exp(-y(x))\, dx = 1$. We first prove that if ϕ is continuous, then $\succsim$ has an an adversarial expected utility representation. First, observe that

$$V(F) = \int \int_0^1 \phi(x, \theta) d\theta dF(x) + \Sigma_\phi(F) = \min_{y \in Y} \left\{ \int_0^1 y^\phi(\theta) d\theta - \int y(x) dF(x) \right\}, \quad (21)$$

where $y^\phi(\theta) = \max_{\xi \in X} \{\phi(\theta, \xi) + y(\xi)\}$. Observe that the restriction $\int \exp(-\tilde{y}(\xi)) d\xi = 1$ defining the set elements of the set Y is irrelevant in the previous minimization problem because, for every $y \in Y$ and $c \in \mathbb{R}$ the function $y + c$ attains the same value as y. This and Proposition 1.11 in Santambrogio [2015] together with its proof, imply that

$$V(F) = \min_{y \in \tilde{Y}} \left\{ \int_0^1 y^\phi(\theta) d\theta - \int y(x) dF(x) \right\} \quad (22)$$

for some compact set $\tilde{Y} \subseteq Y$. This implies that $\succsim$ has an adversarial expected utility representation $(\tilde{Y}, u)$ where $u(x, y) = \int_0^1 y^\phi(\theta) d\theta - y(x)$.

Next, because both X and Θ are compact and convex with nonempty interior and U has full support, Proposition 7.18 in Santambrogio [2015] implies that the solution $y \in C(X)$ to the intermediate minimization problem in equation 22 is unique up to an additive constant. In turn, there exists a unique $y \in \tilde{Y}$ that satisfies the normalization $\int \exp(-\tilde{y}(\xi)) d\xi = 1$ and that solves the problem in equation 22 restricted to $\tilde{Y}$. Theorem then implies that $\succsim$ has an adversarial expected utility representation with uniqueness, hence that it also has an adversarial forecaster representation. Finally, given that the continuous local expected utility of V is $w(x, F) = \int_0^1 y^\phi(\theta) d\theta - \int y_F(x) dF(x)$ for all F, where $y_F \in \tilde{Y}$ is the unique solution of the minimization problem in equation 22, and $V(\delta_x) = \int_0^1 \phi(\theta, x) d\theta$ for all

$x \in X$, the formulas for v and σ given the statement follow. ■

Proof of Proposition 4. From the proof of Theorem 5 we know that $\succsim$ admits a representation V as in equation 21. By Proposition 1.11 in Santambrogio we have $V(F) = \max_{T \in \Delta(U,F)} \int \phi(\theta, x) dT(\theta, x)$ for all $F \in \mathcal{F}$. Next, because $X = [0,1]$, U is atomless, and ϕ is continuously differentiable with ϕ_x increasing in θ, Theorem 2.9 in Santambrogio [2015] implies that $V(F) = \int_0^1 \phi(\theta, F^{[-1]}(\theta)) d\theta$. ■

Online Appendix II: Ancillary results

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

Online Appendix II.A: Ancillary results for Appendix I

Proof of Theorem 7. It is immediate that under (ii), condition (iii) for V is obtained by setting $Y = \{v + \sigma(\cdot, F)\}_{F \in \mathcal{F}}$ and $u(x, y) = y(x)$. It is also immediate that (iii) implies (i) since, for all $F \in \mathcal{F}$ and $y \in \hat{Y}(F)$, we have that $u(x, y)$ is a local expected utility of V at F. We next prove that (i) implies (ii). Because V has a local expected utility, $\mathcal{W}_V(F) \neq \varnothing$ for all $F \in \mathcal{F}$. Fix $w_F \in \mathcal{W}_V(F)$ for all $F \in \mathcal{F}$ and let B denote the corresponding Bregman divergence as defined in Definition 10. Observe that for every F we have

$$\begin{aligned}\int B(\delta_x, F) dF(x) &= V(F) - \int V(\delta_x) dF(x) - \int w_F(x) dF(x) + \int w_F(x) dF(x) \\ &= V(F) - \int V(\delta_x) dF(x),\end{aligned}$$

so $V(F) = \int V(\delta_x) dF(x) + \int B(\delta_x, F) dF(x)$. Now define $v(x) = V(\delta_x)$ and $\sigma(x, F) = B(\delta_x, F)$ for all x and F. Given that V is continuous, it follows that v is continuous. Next, we show that σ is a pseudo forecast error. First, observe that, for every F,

$$\sigma(x, F) = V(F) - v(x) - \int w_F(\tilde{x}) dF(\tilde{x}) + w_F(x)$$

is continuous in x since v and w_F are continuous. Second, $\sigma(x, \delta_x) = B(\delta_x, \delta_x) = 0$ for every x. Finally, fix $F, \tilde{F} \in \mathcal{F}$ and observe that

$$\begin{aligned}\int \sigma(x, \tilde{F}) dF(x) &= V(\tilde{F}) - \int v(x) dF(x) - \int w_{\tilde{F}}(x) d\tilde{F}(x) + \int w_{\tilde{F}}(x) dF(x) \\ &\geqslant V(F) - \int v(x) dF(x) = \int \sigma(x, F) dF(x),\end{aligned}$$

where the inequality follows since $w_{\tilde{F}} \in \mathcal{W}_V(\tilde{F})$. This shows that σ is a pseudo forecast error. Thus $V(F) = \int v(x) dF(x) + \min_{\hat{F} \in \mathcal{F}} \int \sigma(x, \hat{F}) dF(x)$, as desired. (ii) implies (i).

Next, we prove point 1. Assume that there exist $\hat{v} \neq v$ that satisfy equation 14 for V, possibly with respect to a different pseudo forecast errors σ and $\hat{\sigma}$. Then we have $v(x) = V(\delta_x) = \hat{v}(x) + \min_{\hat{F} \in \mathcal{F}} \hat{\sigma}(x, \hat{F}) = \hat{v}(x) + \hat{\sigma}(x, \delta_x) = \hat{v}(x)$, a contradiction.

We finally prove point 2. First let $\sigma(x, F) = B(\delta_x, F)$ for some Bregman divergence of V. It follows from the proof of (i) implies (ii) that σ satisfies 14 for V. Conversely, assume that a pseudo forecast error σ satisfies 14 for V. Fix F, and for every F and x, define $w_F(x) = v(x) + \sigma(x, F)$. Given that σ is a pseudo forecast error, we have

$$V(F) = \int v(x) dF(x) + \int \sigma(x, F) dF(x) = \int w_F(x) dF(x).$$

Next, fix $\tilde{F} \in \mathcal{F}$ and observe that

$$V(\tilde{F}) \leqslant \int v(x) d\tilde{F}(x) + \int \sigma(x, F) d\tilde{F}(x) = \int w_F(x) d\tilde{F}(x).$$

This proves that $w_F \in \mathcal{W}_V(F)$. Because F was arbitrary, it follows that w_F is a local expected utility for V. Consider the corresponding Bregman divergence B and observe that, for every $\tilde{F} \in \mathcal{F}$,

$$\begin{aligned}B(\tilde{F}, F) &= V(F) - V(\tilde{F}) - \int (v(x) - \sigma(x, F))\, d(F - \tilde{F})(x) \qquad \forall F \in \mathcal{F} \\ &= \int \left(\sigma(x, F) - \sigma(x, \tilde{F})\right) d\tilde{F}(x)\end{aligned}$$

where the second equality follows from equation 14. With this, we have $B(\delta_x, F) = \sigma(x, F)$ for every x. Given that F was arbitrarily chosen, the implication follows. ■

Proof of Lemma 1. Write

$$\int w^n(x)dF^n(x) - \int w(x)dF(x) = \int (w^n(x) - w(x))\, dF^n(x) + \int w(x)d(F^n(x) - F(x)).$$

For the second term $\int w(x)d(F^n(x) - F(x)) \to 0$ by the definition of weak convergence. Analyzing the first term

$$\int (w^n(x) - w(x))\, dF^n(x) \leqslant \sup |w^n(x) - w(x)| \int dF^n(x) = \sup |w^n(x) - w(x)| \to 0.$$[29]

Finally, we wish to show that if $F^n \to F$ and w^n are local expected utility functions for F^n with $w^n \to w$ then w is a local expected utility function for F. Suppose we are given $\int w^n(x)d\tilde{F}(x) \geqslant V(\tilde{F})$ and $\int w^n(x)dF^n(x) = V(F^n)$. We have $\int w(x)d\tilde{F}(x) \geqslant V(\tilde{F})$ by the definition of weak convergence. It remains to show that $\int w(x)dF(x) = V(F)$. As $V(F)$ is continuous so it suffices to show that $\int w^n(x)dF^n(x) = \int w(x)dF(x)$. This follows directly from the first result. ■

Lemma 7. *If V has a continuous local expected utility $w(x, F)$, then*

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

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

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

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

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

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

[29] This highlights an important difference between positive and signed measures. In the case of a signed measure it is not true that $\int (w^n(x) - w(x))\, dF^n(x) \leqslant \sup |w^n(x) - w(x)| \int dF^n(x)$ and in fact the lemma is false for signed measures on infinite dimensional spaces.

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

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

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

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

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

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

which yields the statement. ■

Proof of Lemma 2. 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 7. ■

Proof of Lemma 3. 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 continous in x we have $F^n \to F$ implies that $h_{F^n}(s)$ converges pointwise to $h_n(s)$. Hence $\left(h(x,s) - \int h(\tilde{x},s)dF^n(\tilde{x})\right)^2$ converges pointwise to $\left(h(x,s) - \int h(\tilde{x},s)dF(\tilde{x})\right)^2$. Given that h is square-integrable over (S,μ), the dominated convergence theorem implies that

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

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

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

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

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

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

Proof of Lemma 4. 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

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

We next extend V to the space of signed measures by

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

and observe that the cross term is

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

so that

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

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

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

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

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

Online Appendix II.B: Ancillary results for Appendix II

Before proving Lemma 6, we state and prove an intermediate result.

Lemma 8. *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], we have that the subset $\Delta_0(\mathcal{F}_e)$ of finitely supported probability measures over $\mathcal{F}_e$ is dense in $\Delta(\mathcal{F}_e)$. In turn this implies the statement. $\blacksquare$

Proof of Lemma 6. Let $\hat{F}$ solve $\max_{F\in\mathcal{F}_\Gamma(\overline{X})} V(F)$. By Lemma 8, there exists a sequence $\hat{F}^n \to \hat{F}$ such that, for every $n \in \mathbb{N}$, we have $\hat{F}^n \in co(\mathcal{E}^n)$ for some finite set $\mathcal{E}^n \subseteq ext(\mathcal{F}_\Gamma(\overline{X}))$. By Theorem 10, 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. $\blacksquare$

Proof of Corollary 1. By Theorem 15.10 in Aliprantis and Border [2006], there exists a sequence of finitely supported $\mu^n \in \Delta(S)$ such that $\mu^n \to \mu$. The GMM adversarial forecaster representation V^n induced by $(h, \mu^n$ satisfies the assumptions of Theorem 2 by defining $Y^n = \prod_{s\in\operatorname{supp}\mu^n} h(X,s) \subseteq \mathbb{R}^{m^n}$, where $m_n = |\operatorname{supp}\mu^n|$, so for every $n \in \mathbb{N}$, there exists a solution F^n of the problem $\max_{F\in\Delta(\overline{X})} V^n(F)$ that is supported on up to $m_n + 1$ points of $\overline{X}$. Because the constraint set $\Delta(\overline{X})$ is compact and V is continuous, the Berge maximum theorem implies that all the accumulation points of the sequence F^n are solutions of the problem $\max_{F\in\Delta(\overline{X})} V(F)$, where V

is the GMM adversarial forecaster representation induced by h and μ. Theorem 3 established that this problem has a unique full-support solution F, so F is the unique accumulation point of F^n. Because $\overline{X}$ is compact, the sequence $\operatorname{supp} F^n$ converges to some set $\hat{X} \subseteq \overline{X}$ in the Haussdorf sense. By Box 1.13 in Santambrogio [2015], $F^n \to F$ implies that $\operatorname{supp} F \subseteq \hat{X}$, and, given that $\operatorname{supp} F = X$, it follows that $\operatorname{supp} F^n \to \overline{X}$. ■

Online Appendix III: Optimization

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

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

Proposition 6. *Let V have 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 (ii) immediately yields Proposition 1 as a corollary since, due Theorem 5, if V has an adversarial forecaster representation, then it also has an adversarial expected utility representation where $\hat{Y}(F)$ is a singleton for every F. 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 as in Cerreia-Vioglio [2009] but restricting to finite set of utilities (which corresponds to a finite Y in our notation).

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

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

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

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

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

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

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

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

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

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

Online Appendix III.B: finite Y

This section states and proves additional results on the optimization problem of Section 4. Fix an arbitrary compact and convex set $\overline{\mathcal{F}} \subseteq \mathcal{F}$ of feasible lotteries. We start with a simple lemma that establishes the existence of a saddle pair (F^*, y^*).

Lemma 9. *There exists $F^* \in \overline{\mathcal{F}}$ and $y^* \in Y$ such that*

$$\int u(x, y^*)dF^*(x) = V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F) \tag{23}$$

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

For every (F^*, y^*) as in Lemma 9, define the set

$$\overline{\mathcal{F}}(F^*, y^*) = \left\{ F \in \overline{\mathcal{F}} : \forall y \in Y \backslash \{y^*\}, \int u(x, y)dF(x) \geqslant \int u(x, y)dF^*(x) \right\} \tag{24}$$

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

Lemma 10. *Fix (F^*, y^*) as in Lemma 9 and a nonempty, closed, and convex set $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$. The set* $\operatorname{argmax}_{F \in K} \int u(x, y^*)dF(x)$ *is nonempty, convex, and closed.*

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

We next state and prove a general, yet simple, result about the existence of maximizers of Problem 23 that are extreme points of convex, closed sets $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$.

Lemma 11. *Fix (F^*, y^*) as in Lemma 9 and a nonempty, closed, and convex set $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$ such that $F^* \in K$. We have*

$$\operatorname*{argmax}_{F \in K} \int u(x, y^*)dF(x) \subseteq \operatorname*{argmax}_{F \in \overline{\mathcal{F}}} V(F). \tag{25}$$

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

Proof. Fix $F^* \in \mathcal{F}$ and $y^* \in Y$ as in Lemma 9 and a nonempty, closed, and convex set $K \subseteq \overline{\mathcal{F}}(F^*, y^*)$. Let $\hat{F} \in \text{argmax}_{F \in K} \int u(x, y^*) dF(x)$. We need to show that $V(\hat{F}) = V(F^*)$. Observe that

$$\int u(x,y) d\hat{F}(x) \geqslant \int u(x,y) dF^*(x) \qquad \forall y \in Y \backslash \{y^*\} \tag{26}$$

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

$$\int u(x,y^*) d\hat{F}(x) \geqslant \int u(x,y^*) dF^*(x) \tag{27}$$

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

$$\int u(x,y) d\hat{F}(x) \geqslant \int u(x,y) dF^*(x) \geqslant V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F) \tag{28}$$

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

Lemma 11 is not very insightful per se since the set $\overline{\mathcal{F}}(F^*, y^*)$ depends on the particular choice of (F^*, y^*). However, whenever we can find a set K as in the statement of Lemma 11 whose extreme points satisfy interesting properties, the theorem lets us conclude that there is an optimizer of the original problem with those properties. We next apply this strategy to optimization problems with additional structure on $\overline{\mathcal{F}}$ and on Y by relying on known characterizations of extreme points of sets of probability measures. For completeness, we report here the original results mentioned.

Theorem 11 (Proposition 2.1 in Winkler [1988])**.** *Fix a convex and closed set $\overline{\mathcal{F}} \subset \mathcal{F}$, an affine function $\Lambda : \overline{\mathcal{F}} \to \mathbb{R}^{n-1}$, and a convex set $C \subset \Lambda(\overline{\mathcal{F}})$. The set $\Lambda^{-1}(C)$ is convex and every extreme point of $\Lambda^{-1}(C)$ is a convex combination of at most n extreme points of $\overline{\mathcal{F}}$.*

We can combine this result with Lemma 11 to obtain the following result.

Theorem 12. *Suppose that Y has m elements. There exists a solution $F^* \in \text{argmax}_{F \in \overline{\mathcal{F}}} V(F)$ that is a convex combination of at most m extreme points of $\overline{\mathcal{F}}$.*

Proof. Fix (F^*, y^*) as in Lemma 9. Observe that $|Y \backslash \{y^*\}| = m - 1$ by assumption. For simplicity we write $Y \backslash \{y^*\} = \{y_1, ..., y_m - 1\}$. Define the map $\Lambda : \overline{\mathcal{F}} \to \mathbb{R}^{m-1}$ as

$$\Lambda(F)_i = \int u(x, y_i) dF(x) \qquad \forall i \in \{1, ..., m-1\} \tag{29}$$

Also define the convex set

$$C \equiv \Lambda(\overline{\mathcal{F}}(F^*, y^*)) \subseteq \Lambda(\overline{\mathcal{F}}) \tag{30}$$

It is easy to see that $\Lambda^{-1}(C) = \overline{\mathcal{F}}(F^*, y^*)$. By Theorem 11 it follows that every extreme point of $\overline{\mathcal{F}}(F^*, y^*)$ is a convex combination of at most n extreme points of $\overline{\mathcal{F}}$. Finally, the statement follows by a direct application of Theorem 11. ■

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

Theorem 13. *Suppose that Y is finite with m elements. For every closed $\overline{X} \subseteq X$, there exists an optimal lottery F^* for the problem in equation 5 that has finite support on no more than $k + m$ points of $\overline{X}$.*

Proof of Theorem 13. Let $\overline{\mathcal{F}} = \mathcal{F}_\Gamma(\overline{X})$ for some closed $\overline{X} \subseteq X$, and fix (F^*, y^*) as in Lemma 9. The set $\overline{\mathcal{F}}(F^*, y^*)$ is defined by $k + m - 1$ moment restrictions: k moments restrictions from Γ and $m - 1$ from the definition of $\overline{\mathcal{F}}(F^*, y^*)$. By Lemma 11 there exists $F^* \in \text{ext}(\overline{\mathcal{F}}(F^*, y^*))$ such that $V(F^*) = \max_{F \in \overline{\mathcal{F}}} V(F)$. By Winkler's Theorem the each $\tilde{F} \in \overline{\mathcal{F}}(F^*, y^*)$ is supported on up to $k+m$ points of $\overline{X}$ as desired. ■

Online Appendix III.C: Robust solutions

This section shows that the finite-support property of Theorem 2 generically holds for all solutions of the optimization problem in equation 5 that are "robust" in the following sense. For every $F \in \mathcal{F}_\Gamma(\overline{X})$, we call a sequence as in Lemma 8 a *finitely approximating sequence of F.*

Definition 12. 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

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

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

The proof will use the following lemma.

Lemma 12. *Fix a finite set $\hat{X} \subseteq \overline{X}$ and an open dense subset $\hat{\mathcal{W}}$ of $\mathbb{R}^{\hat{X}}$. The set*

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

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

Proof. Because $\hat{\mathcal{W}}$ is open, so is $\overline{\mathcal{W}}$. Fix $w \in C(\overline{X})$. Given that $w_{|\hat{X}} \in \mathbb{R}^{\hat{X}}$, there exists a sequence $\hat{w}^n \in \hat{\mathcal{W}}$ such that $\hat{w}^n \to w_{|\hat{X}}$. Next, fix $n \in \mathbb{N}$ large enough so that $B_{1/n}(\hat{x}) \cap B_{1/n}(\hat{x}') = \varnothing$ for all $\hat{x}, \hat{x}' \in \hat{X}$.[30] 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

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

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

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

Proof of Theorem 14. Without loss of generality, we assume that $\overline{X} = \bigcup_{F \in \mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F$.[31] Define $\overline{\mathcal{E}} = cl\left(ext\left(\mathcal{F}_\Gamma(\overline{X})\right)\right)$ and consider an increasing sequence of finite sets of extreme points $\mathcal{E}^n \subseteq ext\left(\mathcal{F}_\Gamma(\overline{X})\right)$ such that $\mathcal{E}^n \uparrow \overline{\mathcal{E}}$. Observe that, by construction, we have $\overline{X}_{\mathcal{E}^n} \uparrow \overline{X}$.[32] 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 10. By Lemma 12 the set

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

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

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

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

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

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

In particular, given that, for every $n \in \mathbb{N}$, we have $w_{|\overline{X}_{\mathcal{E}^n}} \in \hat{\mathcal{W}}^n$, Theorem 10 implies that F^n is supported on up to $(k+1)(m+1)$ points of $\overline{X}_{\mathcal{E}^n}$. Given that $F^n \to F^*$, it follows that F is supported on up to $(k+1)(m+1)$ points of $\overline{X}$. Given that F^* and w were arbitrarily chosen, the result follows. ∎

Online Appendix III.D: Optimal lotteries under transport preferences

Here we consider the problem of choosing a lottery $F \in \mathcal{F}$ when $\succsim$ is a transport preference with representation given by ϕ.[33] Define the correspondence $\Psi_\phi(\theta) =$

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

[32] This follows from the fact that $\overline{X} = \bigcup_{F \in \mathcal{F}_\Gamma(\overline{X})} \operatorname{supp} F$ by assumption. See also footnote 31.

[33] Some of the results presented in this section can be extended to the case where there are additional feasibility constraints such as the moment constraints considered in the previous section. However, we leave a formal analysis of these case for future research.

$\text{argmax}_{x\in X}\,\phi(\theta,x)$ and let $\psi \in \Psi_\phi$ denote an arbitrary measurable selection.

Proposition 7. *If $\gtrsim$ is a transport preference with continuous ϕ, then the set of optimal lotteries for $\gtrsim$ over $\mathcal{F}$ is the closure of $\{U \circ \psi^{-1} \in \mathcal{F} : \psi \in \Psi_\phi\}$.*

Proof of Proposition 7. By the Proof of Proposition 4, we have

$$\max_{F\in\mathcal{F}} V(F) = \max_{T\in\Delta(\Theta\times X):marg_\Theta T=U} \int \phi(\theta,x)dT(\theta,x),$$

which immediately implies that $F \in \text{argmax}_{\tilde{F}\in\mathcal{F}} V(F)$ if and only if there exists $T \in \Delta(U,F)$ such that $T(G_\phi) = 1$, where $G_\phi = Gr(\Psi_\phi) \subseteq \Theta \times X$ is the graph of the correspondence Ψ_ϕ. In turn, this is equivalent to $0 \geqslant \inf_{T\in\Delta(U,F)} \{1 - T(G_\phi)\}$ and, by an application of the Kantorovich duality for $\{0,1\}$-valued costs (see Theorem 1.27 in Villani [2021]), it is also equivalent to

$$U(\Psi^\ell_\phi(A)) \geqslant F(A)$$

for all closed $A \subseteq X$, where $\Psi^\ell_\phi(A) = \{\theta \in \Theta : \Psi_\phi(\theta) \cap A \neq \varnothing\}$ is the lower-inverse of the correspondence Ψ_ϕ evaluated at A. Finally, because U is atomless, Corollary 3.4 in Castaldo, Maccheroni, and Marinacci [2004] implies that this is equivalent to the fact that F is in the closure of $\{U \circ \psi^{-1} \in \mathcal{F} : \psi \in \Psi_\phi\}$. ■

The characterization of optimality follows by first rewriting the problem as

$$\max_{F\in\mathcal{F}} V(F) = \max_{T\in\Delta(\Theta\times X):marg_\Theta T=U} \int \phi(\theta,x)dT(\theta,x),$$

which immediately implies that, for every $\psi \in \Psi_\phi$, the distribution $U \circ \psi^{-1}$ is optimal. The converse follows by a further application of the Kantorovich duality.

Proposition 7 result yields the following corollaries.

Corollary 4. *Let $\gtrsim$ be a transport preference with representation ϕ such that $\Psi_\phi = \psi$ is single valued. The unique optimal lottery is $U \circ \psi^{-1}$.*

The assumption of the corollary is satisfied, for example, when $\phi(\theta,x)$ is strictly quasi-concave in x for every θ, as in Example 3 where the optimal lottery is uniformly distributed over the entire space of outcomes.

Corollary 5. *Let $\gtrsim$ be a transport preference with representation ϕ. For every $\psi \in \Psi_\phi$, there is an optimal lottery F that assigns probability 1 to $\psi(\Theta)$.*

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

Online Appendix IV: Additional applications

Online Appendix IV.A: Additional examples

This section presents two examples. In the first, there are GMM preferences that have a strictly concave representation and give rise to an optimal lottery with full support. The second example illustrates most of the main results in the text by solving an optimal lottery under the asymmetric adversarial forecaster preferences of Section 7.2.

Example 5 (Weiner Process Example)**.** We interpret $x \in [0, 1]$ as time. While it is natural to think of $h(\cdot, s)$ as a random function of s with distribution induced by F, there is a dual interpretation in which we think of $h(x, \cdot)$ as a random function of x (a random field) with distribution induced by μ. In this interpretation, the $H(x, \tilde{x})$ are the second (non-central) moments of that random variable between different points $x, \tilde{x}$ in the random field. If, for example, $X = [0, 1]$, then this random field is a stochastic process, and $H(x, \tilde{x})$ the second moments of the process h between times $x, \tilde{x}$. It is well known that continuous time Markov process are equivalent to stochastic differential equations and that an underlying measure space S and measure μ can be found for each such process. Specifically, consider the process generated by the stochastic differential equation $dh = -h + dW$ where W is the standard Weiner process on (S, μ) and the initial condition $h(0, s)$ has a standard normal distribution. Then the distribution of the difference between $h(x, \cdot)$and $h(\tilde{x}, \cdot)$ depends only on the time difference $\tilde{x} - x$, and in particular $H(x, \tilde{x}) = \int h(x, s)h(\tilde{x}, s)d\mu(s) = G(x - \tilde{x})$. In this case $H(0, \tilde{x}) = e^{-\tilde{x}}$, which is non-negative, strictly decreasing and strictly convex. $\triangle$

Example 6 (Optimal lotteries under asymmetric forecast error)**.**

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

$$\max_{p\in[0,1]} \overline{v}p + p\left(\exp\left(\lambda(1 - \hat{x}(p^*))\right) - \lambda(1 - \hat{x}(p^*))\right) + (1-p)(\exp\left(-\lambda\hat{x}(p^*)\right) + \lambda\hat{x}(p^*)). \tag{31}$$

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

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

which is equivalent to

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

Therefore, the overall solution is $p^* = \min\{1, \max\{0, p^*_{int}\}\}$. Clearly, the solution is increasing the the baseline utility parameter $\overline{v}$. However, the effect of the asymmetric parameter λ is ambiguous: (say more rpecicelsy what is meant here)

$\triangle$

Online Appendix IV.B: Risk preferences and surprise

Eeckhoudt and Schlesinger [2006] formalize the idea that an agent is averse to higher-order risks through the comparison of pairs of lotteries that only differ for their n-th order risk. If at any wealth level the agent prefers the lottery with less n-th order risk, they say the preferences exhibit *risk apportionment* of order n. In our setting with general continuous preferences, a sufficient condition for risk apportionment of order n is monotonicity with respect to the n-th order stochastic dominance relation

$\gtrsim_{\mathcal{W}_{SD_n}}$ where

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

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

As an example, suppose $v(x) = 1 - \exp(-ax)/a$ for $a > 0$. If there is no preference for surprise, that is $\lambda = 0$, the agent is mixed risk averse, as most of the risk averse subjects in Deck and Schlesinger [2014]. However, as λ increases the sign of the even derivatives of the local expected utilities switches from negative to positive, while the sign of the odd derivatives remains positive, so the agent shifts from mixed risk averse to mixed risk loving. Moreover, if the agent is very risk averse, that is, $a > 1$, then higher-order derivatives will be more affected by an increased taste for surprise, while the opposite is true if the agent is not very risk averse, that is, $a < 1$.

Online Appendix IV.C: Repeated choices and correlation aversion

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

$$V(F) = \min_{y \in Y} \int u(x, y_0, y_1(x_0)) dF(x)$$

still admit an adversarial expected utility representation. These preferences capture the idea of aversion to correlation between x_0 and x_1, which is well documented in experiments (see for example Andersen et al. [2018]). Intuitively, the agent would tend to avoid lotteries with high correlation between x_0 and x_1, since this means the adversary is well informed about the residual distribution of x_1 when choosing y_1. The next example formalizes this using Theorem 6.

Example 7. Let $X_0 = \{0,1\}$, $X_1 = [0,1]$, $v(x_0,x_1) = v_0(x_0) + v_1(x_1)$, and assume that the adversary tries to minimize mean squared error, so $\sigma_0(x_0,F_0) = \left(x_0 - \int \tilde{x_0} dF_0(\tilde{x_0})\right)^2$ and $\sigma_1(x_1,F_1|x_0) = \left(x_1 - \int \tilde{x_1} dF_1(\tilde{x_1}|x_0)\right)^2$, where F_0 and $F_1(\cdot|x_0)$ respectively denote the marginal and the conditional distributions of F. Then $\sigma(x_0,x_1,F) = \sigma_0(x_0,F_0)+\sigma_1(x_1,F_1|x_0)$, so the local expected utility is $w(x_0,x_1,F) = v(x_0)+v(x_1)+\sigma(x_0,x_1,F)$. We model the agent's preference for correlation between x_0 and x_1 through the monotonicity properties of their preference with respect to the supermodular and submodular order. Intuitively, preferences that preserve the supermodular order favor lotteries with high positive correlation between x_0 and x_1 because their local expected utilities are supermodular, and vice versa for the submodular order. Following Shaked and Shanthikumar [2007] (Section 9.A.4), F dominates G in the submodular (resp. supermodular) order if $F \gtrsim G$ whenever $\int w(x)dF(x) \geqslant \int w(x)dG(x)$ for all functions $w \in C(X)$ that are differentiable in x_1 and such that $\frac{\partial}{\partial x_1} w(1,x_1) - \frac{\partial}{\partial x_1} w(0,x_1) \leqslant 0$ (resp. $\geqslant$). Therefore, the submodular and supermodular order are examples of stochastic order introduced in Definition 9, where the relevant sets of functions are those ones that satisfy the partial derivative condition above. For every F, the corresponding partial derivatives for the local utility at F are

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

Thus by Theorem 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 a similar reasoning, the agent would prefer to decrease the amount of negative correlation between x_0 and x_1 at each lottery F such that $\int \tilde{x_1} dF_1(\tilde{x_1}|1) < \int \tilde{x_1} dF_1(\tilde{x_1}|0)$.[34] Combining these facts, we see that the agent has highest utility with distributions such that $\int \tilde{x_1} dF_1(\tilde{x_1}|1) = \int \tilde{x_1} dF_1(\tilde{x_1}|0)$, so that the best conditional forecast is independent of x_0. △

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

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

expected utility model for future research.[35]

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

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

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

Proposition 8. *Let $\gtrsim$ admit a representation V with a local expected utility w and, for every $F \in \mathcal{F}$, let $\gtrsim_F$ denote the expected utility preference induced by $w(\cdot, F)$. Then $F \gtrsim_F \tilde{F}$ (resp. $F \succ_F \tilde{F}$) implies that $F \gtrsim \tilde{F}$ (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 $\tilde{F} \gtrsim_F F$ 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.[36]

Our notion of local expected utility is neither weaker nor stronger than Fréchet differentiablility. If V has a continuous local expected utility, then it is concave, which

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

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

is not implied by Fréchet differentiability. Conversely, Example 8 below shows that continuous local expected utility does not imply Fréchet differentiability.

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

In particular, Chew, Karni, and Safra [1987] develops a theory of local utilities for rank-dependent preferences and Chew and Nishimura [1992] extends it to a broader class. Recall that V is Gâteaux differentiable [37] 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 $DV(F)(\tilde{F} - \overline{F}) = \int w(x,F)d\tilde{F}(x) - \int w(x,F)d\overline{F}(x)$ and we say that 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 in fact an ordinary measure. We can then restate Lemma 7 with the language of Gâteaux derivatives just introduced.

Proposition 9. *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 6. *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 an important class of preferences that have a continuous local expected utility but not a local utility in Machina's sense.

Example 8. 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

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

in (x, F). Therefore, by Corollary 6, V has a continuous local expected utility and, by Theorem 1, it admits an adversarial forecaster representation. △

References

Aliprantis, C. D. and K. C. Border (2006). *Infinite dimensional analysis*. Springer.

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

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

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

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

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

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

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

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

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

Stanca, L. (2021). "Recursive Preferences, Correlation Aversion, and the Temporal Resolution of Uncertainty."

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

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