You Are What You Bet: Eliciting Risk Attitudes from Horse Races

Pierre-André Chiappori, Amit Gandhi, Bernard Salanié and Francois Salanié

March 14, 2008

What Do We Know About Risk Preferences?

Not that much:

- intuition drawn from theory + casual observation (Arrow 65):
  - $\text{ARA}(x) = -u''(x)/u'(x)$ should be decreasing, since richer people buy more risk;
  - $\text{RRA}(x) = -xu''(x)/u'(x)$ should be close to constant, as the proportion of wealth invested in risky assets is fairly constant across wealth levels (?).
- but this completely neglects other sources of individual variations.
- financial and insurance evidence: all over the map these days, RRA from 0.5 to 50.

Experimental evidence

Points to violations of expected utility, since Allais 1953, at least "close to the edges of the triangle" (where some probabilities are small).

Also suggests that (generalized) risk aversions are very heterogeneous:

Barsky et al (QJE 1997) use survey questions, linked to actual behavior;

they report D1=2 and D9=25 for RRA, poorly explained by demographics.

Guiso-Paiella (2003) report similar findings ("massive unexplained heterogeneity").

More Evidence

On the same survey, Chiappori-Paiella (2006) uses the time dimension and finds RRA index has mean=4.2 and median=1.7.

Yet much of economics does not take this heterogeneity very seriously.

Can we document this heterogeneity on "actual" data?

Ideally. . .

We would observe a large, representative and stable population of people,
making a large number of repeated and yet uncorrelated choices in very simple risky situations.

Using Horse Bets: The Pros

A "win bet" at odds R on horse i buys an Arrow-Debreu asset for state "i wins" with net return R.

Very simple model of *vertically differentiated* varieties:

- at a given price (odds), a horse that is more likely to win is unambiguously better;
- equilibrium prices (odds) reflect the distribution of preferences towards risk and beliefs;
- ...which can be recovered if it is not too "rich".

More than 100,000 races are run in the US every year.

Using Horse Bets: The Cons

Bettors are unlikely to be a representative sample of the US population:
"they must love risk since they gamble": not so obvious;
a decision to bet may come from a "utility of gambling",
whereas the choice of what horse to bet on would be guided by risk-averse preferences.
Second problem: stable population? Races are run in very different places at at very different times.

- we can control for important observables (demographics of racetrack area, day of week)—just started;
- but not for characteristics of individual bettors;
- so we need to control for voluntary participation $\longrightarrow$ left for further work.

The Identification Question

Assume a population of bettors, stable in time (given some observed characteristics);
and look at win bets. A given bettor θ with beliefs p_θ values a \$1 bet that

- wins (net) \$$R$ with probability p_θ
- loses \$1 with probability $(1 - p_\theta)$

as $W(p_\theta, R, \theta)$.
e.g., with expected utility theory (EUT), u rebased at current wealth:

$$W(p_\theta, R, \theta) = p_\theta u(R, \theta) + (1 - p_\theta) u(-1, \theta).$$

or, for Cumulative Prospect Theory (CPT)

$$W(p_\theta, R, \theta) = G(p_\theta, \theta) u_+(R, \theta) + H(1 - p_\theta, \theta) u_-(-1, \theta).$$

Can we recover uniquely the distribution of $\theta \in \Theta$ in the population?

The Parimutuel System

All money bet is given to the winners (apart from "track take")
Therefore returns depend directly on bets; so we also have *market shares*:
in race m for each horse i

$$s_i^m(R_i^m + 1) = 1 - t^m$$

where s_i is market share of i and t^m is track take, so:

$$s_i^m = \frac{\frac{1}{R_i^m+1}}{\sum_{j=1}^{n^m} \frac{1}{R_j^m+1}}.$$

which we denote $S_i(R^m)$.

Beliefs and Information

In the parimutuel system, odds reflect market shares.
But do they reflect "true" probabilities?
Let true probabilities be $t = (p_1, \dots, p_n)$, and each bettor has an information partition on t;
Gandhi (2006): if

- the distribution of bettors is atomless
- every possible winner is desirable if its return is large enough
- **for every $t \neq t'$, there exists a bettor who can distinguish t and t'**

then there is a unique REE with returns $R_1, \dots, R_n$ that fully reveal t.

Persisting Differences in Beliefs

The above theorem gives some foundation for assuming common, correct beliefs.
Persisting differences in beliefs can be captured

1. through agent-dependent nonlinear proba weighting functions $G(p, \theta)$
2. through transformations $p_\theta = h(p, \theta)$
3. through "random choice".

1 and 2 can be identified nonparametrically but must be constant across races for a given agent
It is the reverse for 3.
So far we allow for 1 only.

The Data

Our data is a large number of races $m = 1, \ldots, M$
Data on a race m consists of

- a number of horses n^m
- a vector of odds R_i^m for $i = 1, \ldots, n^m$
- the index f^m of the horse that won race m;
- some covariates X^m (**omitted in what follows**).

Empirical Strategy

Suppose (for simplicity) all races have exactly n horses
and we observe an infinity of races, so that
for every possible vector of odds $R = (R_1, \dots, R_{n-1})$

- we can estimate $p_i(R)$ for $i = 1, \dots, n-1$ by the proportion of such races won by horse i:

$$p_i(R) \simeq \frac{\sum_{R^m=R}(f^m = i)}{\sum_{R^m=R} 1}.$$

- we know that by definition,

$$S_i(R) = \Pr\left(\{\theta | W(p_i(R), R_i, \theta) \geq W(p_j(R), R_j, \theta) \ \ \forall j\}\right). \qquad (E)$$

The Nexus of the Problem

For any $R = (R_1, \dots, R_n)$, denote $\Theta(i, R)$ the subset of Θ such that

$$\forall j = 1, \dots, n, \;\; W(p_i(R), R_i, \theta) \geq W(p_j(R), R_j, \theta).$$

Then we know the probability of all such sets when the returns (R_i) vary freely (given enough data!)
Is it enough? I.e. is this a *probability-determining family* for Θ so that it identifies the distribution of θ?

Conjecture

Yes, it is "usually" enough to insure that the following condition holds:

The Many-Races Assumption on a subset A: take a subset A of Θ, and assume that there exists a race R and a horse i such that $\Theta_i(R) \subset A$.

Then assume two candidate probabilities on Θ with pdfs f and g. Say they differ on a subset A, with $f(\theta) > g(\theta)$ on A.

By MRA, take $\Theta_i(R) \subset A$; then f puts greater probability than g on $\Theta_i(R)$, which contradicts

$$\int_{\Theta_i(R)} f(\theta) d\theta = \int_{\Theta_i(R)} g(\theta) d\theta = S_i(R).$$

(Also extends to measures with isolated atoms.)

The Curse of Dimensionality

Note that MRA requires **separability:**
for any $\theta \neq \theta'$, there exists a race R and horse i such that $\theta \in \Theta_i(R)$ but not θ'.

And separability cannot hold if Θ is more than n-dimensional (or $(n-1)$ if the track take does not vary across races.)

A Simpler Case: One-dimensional Heterogeneity

Assume that Θ is a subset of $\mathbb{R}$, and that $n \geq 4$. We impose a single-crossing condition:

Condition (SC): each $W(.,.,\theta)$ is increasing in p and R, and the marginal rate of substitution W'_R/W'_p increases in θ.

(SC) means that larger θ's prefer longer odds;

(SC) implies (MRA), but it is **much too strong**:

e.g. if Joe is more risk-averse than Jim on favorites, he also is on outsiders.

But it makes things simpler at this early stage. . .

What We Can Prove under these Assumptions

Theorem: let F_0 be the true cdf of θ on an interval Θ of $\mathbb{R}$; denote $\underline{\theta} = F_0^{-1}(1/n)$. Then

- the data uniquely identify $\underline{\theta}$, and $F_0(\theta)$ above $\underline{\theta}$;
- the assumption that all preferences belong to $W(.,.,\Theta)$ above $\underline{\theta}$ is testable.

From now on, look at the equivalent problem: F_0 known (we take it to be uniform on $[0,1]$), we look for the master function W for $\theta > 1/n$.

Intuition

Given (SC), if we order odds as $R_1 \leq \ldots \leq R_n$ then the set of θ's who bet on horse i is some interval

$$\Theta_i(R) = [\theta_{i-1}(R), \theta_i(R)]$$

where $\theta_0(R) = 0, \theta_n(R) = 1$ and for $i = 1, \ldots, n-1$,

$$W(p_i(R), R_i, \theta_i(R)) = W(p_{i+1}(R), R_{i+1}, \theta_i(R)) \qquad (I_i).$$

With F_0 uniform on $[0,1]$, we can estimate the $\theta_i(R)$'s using

$$S_i(R) = \theta_i(R) - \theta_{i-1}(R)$$

Note that since horse 1 is by definition the favorite, his market share is larger than $1/n$, so $\theta_1 > 1/n$ always.
Beyond that: intervals are probability-determining sets on $\mathbb{R}$...so we are done and there is nothing to test?
Not quite: our assumptions on derivatives have consequences.

Some Notation

First define $\Gamma(v, R, \theta)$ by

$$\Gamma(W(p, R, \theta), R, \theta) \equiv p :$$

Then the indifference condition

$$W(p_i(R), R_i, \theta_i(R)) = W(p_{i+1}(R), R_{i+1}, \theta_i(R)) \qquad (I_i).$$

becomes

$$p_{i+1}(R) = \Gamma(W(p_i(R), R_i, \theta_i(R)), R_{i+1}, \theta_i(R)) \qquad (J_i).$$

So $p_{i+1}(R)$ depends on its n arguments (and i, and n) only through the 4 numbers

$$p_i(R), R_i, R_{i+1}, \theta_i(R). \qquad (IC)$$

Testable by "regressing" $p_{i+1}(R)$ on

$$p_i(R), \theta_i(R), R_i, R_{i+1} \text{ and } R_{i+2}, \ldots, R_n, R_1, \ldots, R_{i-3}, i, n,$$

and testing that "the coefficients in the second group are all zero".

Another Equality Condition

Look again:

$$P_{i+1}(p_i, R_i, \theta_i, R_{i+1}) \equiv p_{i+1}(R) = \Gamma(W(p_i(R), R_i, \theta_i(R)), R_{i+1}, \theta_i(R))$$

The "marginal rate of substitution" between $p_i(R)$ and R_i, i.e.

$$\frac{\frac{\partial P_{i+1}}{\partial p_i(R)}}{\frac{\partial P_{i+1}}{\partial R_i}}$$

does not depend on R_{i+1}; call this condition (MRS).

Sufficiency?

If (IC) and (MRS) hold then we can write

$$p_{i+1}(R) = G(H(p_i(R), R_i, \theta_i(R)), R_{i+1}, \theta_i(R))$$

for some functions G and H.
We would like to identify H to W and G to Γ (up to an increasing transform);
But we also need to check that

$$H'_p > 0, \; H'_R > 0, \; H'_R/H'_p \text{ increases in } \theta,$$

and

$$G \text{ increases in } H \;\text{ and decreases in } \theta_i(R).$$

Sufficiency

These additional conditions turn out to boil down to:

$$P_{i+1} \text{ increases in } p_i \text{ and in } R_i; \qquad (V_1)$$

$$P_{i+1} \text{ decreases in } R_{i+1}; \qquad (V_2)$$

and the MRS of P_{i+1} in (p_i, R_i), i.e.

$$\frac{\frac{\partial P_{i+1}(R)}{\partial R_i}}{\frac{\partial P_{i+1}}{\partial p_i}}$$

increases in $\theta_i(R)$ (call this (V_3)).
Adding these conditions $(V_1), (V_2), (V_3)$ to (IC) and (MRS) yields a set of **necessary and sufficient conditions for identification** (up to an increasing transformation $w(p, R, \theta) = F(W(p, R, \theta), \theta)$)
If the model is well-specified!.

Constructing the Indifference Curves

Given the estimated $P_{i+1}(p_i, \theta_i, R_i, R_{i+1})$ function, we fix $\theta_i(R) = \theta$;
for any point in the $(p_i(R), R_i) = (p, R)$ plane we know that the indifference curve of any representation of $W(p, R, \theta)$ has slope

$$\frac{\frac{\partial P_{i+1}}{\partial p_i}}{\frac{\partial P_{i+1}}{\partial R_i}}(p_i, \theta_i, R_i, R_{i+1})$$

(for any value of R_{i+1}).
This gives a **test for misspecification:**
Once the indifference curve for θ that goes through (p, R) is constructed, choose some odds R' and compute $p' = P_{i+1}(p, \theta, R, R')$;
then (p', R') should lie on that same indifference curve.

Assume $W(p, R, \theta) = F(pu(R, \theta), \theta)$; then we get

$$p_{i+1}(R) = p_i(R) \frac{u(R_i, \theta_i(R))}{u(R_{i+1}, \theta_i(R))}$$

Thus EUT yields two additional conditions; define $\psi_{i+1} = \log\left(P_{i+1}/p_i(R)\right)$:

$$\psi_{i+1} \text{ only depends on } \theta_i(R), R_i \text{ and } R_{i+1} \qquad (EU_1)$$

and

$$\frac{\partial^2 \psi_{i+1}}{\partial R_i \partial R_{i+1}} = 0 \qquad (EU_2).$$

Testing Expected Utility

$(EU_1), (EU_2)$ complete the set of necessary and sufficient conditions under expected utility
and then we can estimate the vNM utility function "nonparametrically":

- fix $u(-1, \theta) = 0$ and $u(R_0, \theta) = 1$ for some R_0 and all θ
- then $u(R, \theta) = E(p_i(R)/p_{i+1}(R) | R_{i+1} = R, R_i = R_0, \theta_i(R) = \theta)$.

Testing Homogeneous Risk Preferences

An easy one: just add

$$\frac{\partial P_{i+1}}{\partial \theta_i} = 0.$$

(Visually: just plot the indifference curves through some (p, R) for various θ's).

Estimating Probabilities

First specify a flexible functional form for $p_i(R) = P(R_i, (R_{-i}))$:

$$p_i = \frac{e^{q_i}}{\sum_{j=1}^{n} e^{q_j}}$$

with, e.g.

$$q_i(R) = \sum_{k=1}^{K} a_k(R_i, \alpha) T_k(R_{-i})$$

and

- the T_k's are symmetric functions—we take $\sum_i 1/(1+R_i)^k$;
- the a_k's are estimated at quantiles of R_i and cubically splined.

Then maximize over α the log-likelihood

$$\sum_{m=1}^{M} \log p_{f^m}(R^m, \alpha).$$

The Favorite-longshot Bias

If market shares were equal to probabilities
(as they would with risk-neutral bettors)
we would have $N_i \equiv 0$, where N_i is the "normalized gain on horse i in its race":

$$N_i = p_i(R_i + 1)\sum_{j=1}^{n} \frac{1}{R_j + 1}.$$

The favorite-longshot bias is the empirical fact that N_i is larger for favorites than for longshots.

As Expected

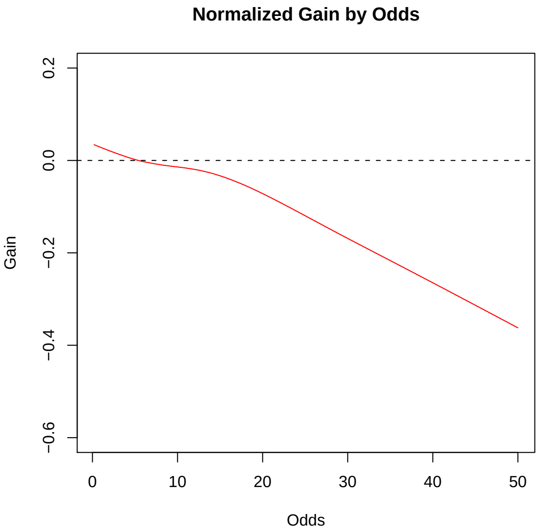


Figure: Normalized Expected Gains

We analyze in fact the "differential gain" $D_i = N_i - N_{i+1}$, which would be zero with
risk-neutral bettors.
It has positive mean in fact (the favorite-longshot bias);
and a sizable dispersion (standard error 0.083).

The Differential Gain

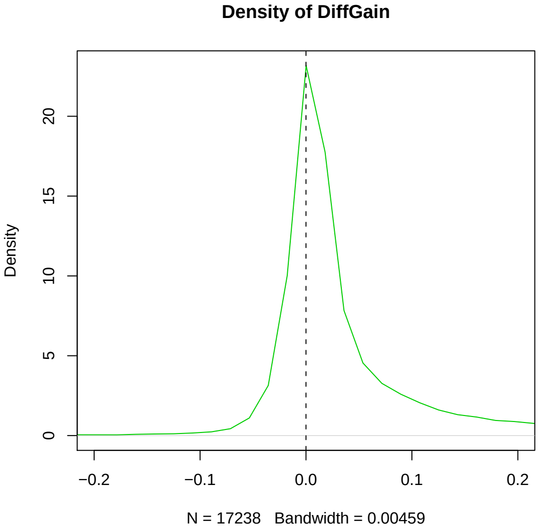


Figure: Density of Differential Gain

The Differential Gain 2

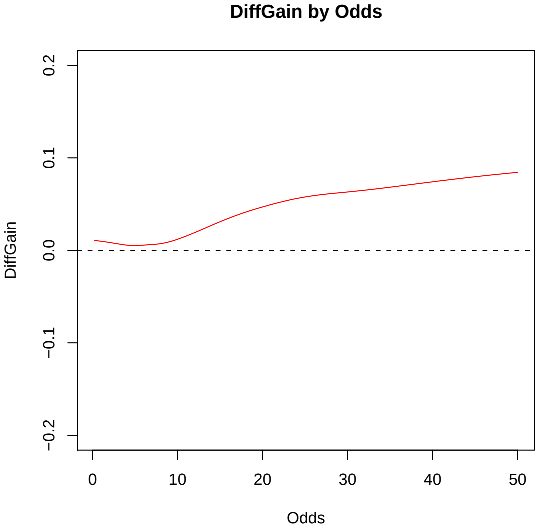


Figure: The Differential Gain Increases with Odds R_i

Results with Expected Utility Bettors

We normalize $u(-1,\theta) \equiv 0$ and $u(0,\theta) \equiv 1$;
then any (analytic) u expands as

$$u(R,\theta) = (R+1)(1 + R\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} a_{kl} R^k \theta^l).$$

We write $p_i u(R_i, \theta_i(R)) = p_{i+1} u(R_i, \theta_i(R))$
which boils down to a linear regression with

$$y_i = D_i$$

and

$$x_i(k,l) = (N_{i+1} R_{i+1}^{k+1} - N_i R_i^{k+1}) \theta_i(R)^l.$$

for $k = 0, \ldots, K$ and $l = 0, \ldots, L$.
Imposing $L = 0$ gives the representative bettor model of Jullien-Salanié.

Estimated vNM u for a Representative Bettor

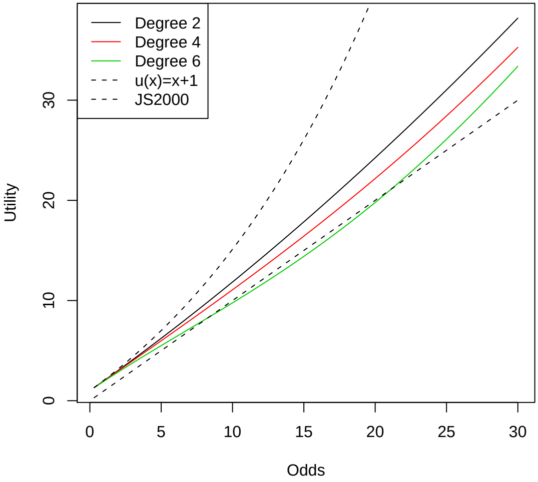


Figure: Polynomials of order 2, 4, and 6

Moving to Heterogeneous Bettors

Caveat 1: the favorite almost always gets at least 25% of the money, so we cannot say anything about bettors with θ lower than that.

Caveat 2: for larger values of θ, we cannot safely say anything about utility of favorites;

e.g. for $\theta_i = 0.9$, the odds R_i and R_{i+1} are almost always larger than 10.

Estimated vNM u for Heterogeneous Bettors ($l = 0, 1$)

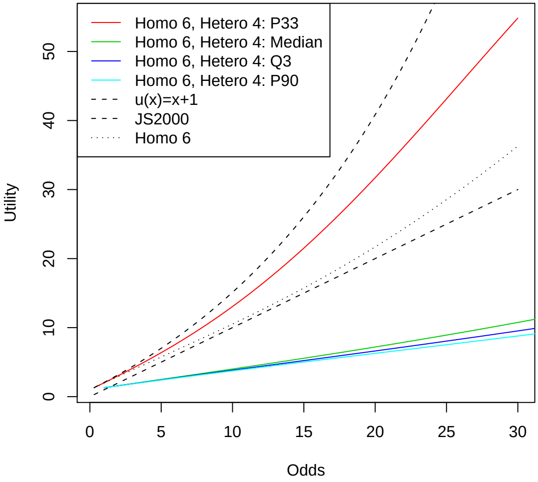


Figure: Polynomials $k \leq 6$

What of Risk Aversion?

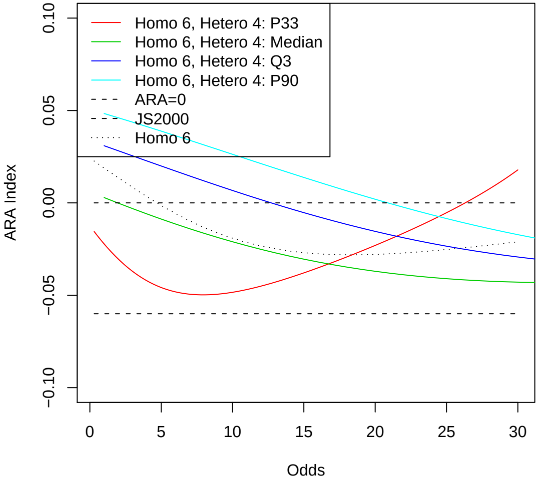


Figure: Polynomials $K = 6, L = 1$

Testing the Various Hypotheses

So far, informally (but close to $400,000$ horses in $50,000$ races...)
We just look at the residual standard error and the R^2 in the regressions.

Explaining the Differential Gain D_i

Model	Residual SE	R^2
Representative Risk Neutral	0.0831	0
Representative, $K = 6$	0.0542	0.638
Heterogeneous, $K = 6$ and $L = 1$	0.0505	0.686

Assuming expected utility,
heterogeneity of preferences only seems to matter for the low θ's, who tend to bet on favorites.
The other dimensions of betting behavior seem to be well explained by a representative bettor (somewhat more risk-averse than if homogeneity is imposed).
But we only explain less than 70% of the variance using expected utility.

Does Single Crossing Hold?

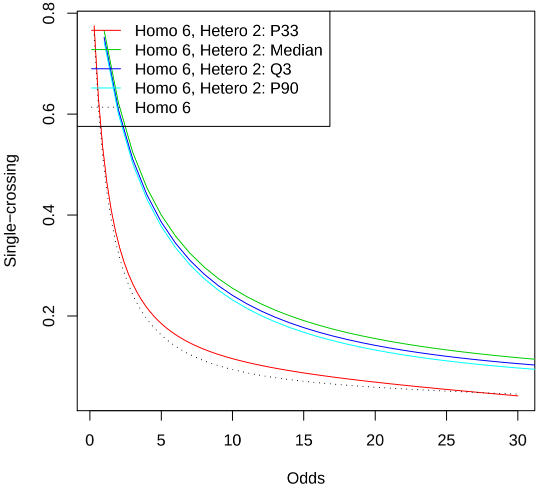
We assumed (translated in the expected utility world)

$$\frac{u'_R(R,\theta)}{u(R,\theta)}$$

increases in θ for all R.

We did not impose it for estimation, so we plot it with our estimates.

Single Crossing is not Rejected



To do list

- non-expected utility;
- modelling bettor participation in a particular race.