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

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

June 25, 2008

# Are Attitudes Towards Risk Homogeneous?

Much empirical work in macroeconomics and finance assumes that they are.
Yet there is mounting experimental evidence that risk attitudes are massively heterogeneous:
Barsky et al (*QJE* 1997) use survey questions, linked to actual behavior;
they report D1=2 and D9=25 for relative risk aversion, poorly explained by demographics.
Guiso-Paiella (2003) report similar findings ("massive unexplained heterogeneity").
On the same survey, Chiappori-Paiella (2007) uses the time dimension and finds RRA index has mean=4.2 and median=1.7.
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 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 $\longrightarrow$ (mostly) left for further work.

# The Identification Question

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

- wins (net) $\$R$ with probability $p_\theta$
- 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) with no betting as reference point:

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

*Can we recover uniquely the distribution of $\theta$ in 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 $t = (p_1, \ldots, p_n)$, and each bettor has an information partition on the set of possible $t$'s;
Gandhi (2008): 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 $R_1, \ldots, R_n$ that fully reveal $t$.
(How we get there is a mystery, as always!)

# 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 $R = (R_1, \ldots, R_{n-1})$

- we can estimate $p_i(R)$ for $i = 1, \ldots, n-1$ 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,

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

# A Simple Case: One-dimensional Heterogeneity

Assume that $\Theta$ 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 $\theta$.

(SC) means that larger $\theta$'s prefer longer odds: e.g. for expected utility,

$$\frac{pW'_p}{W'_R} = \frac{u}{u'_R} = \text{fear-of-ruin},$$

so (SC) says that *fear-of-ruin* decreases in $\theta$.

(SC) 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... (and decreasing risk-aversion implies decreasing fear-of-ruin).

# What We Can Prove under these Assumptions

From now on, look at the equivalent problem: $\theta$ is uniformly distributed on $[0, 1]$, we look for the master function $W$.

**Theorem:**

- the data uniquely identify $W(., .\theta)$ for $\theta > 1/n$;
- the assumption of (one-dimensional heterogeneity + (SC)) is testable.
- restrictions about the $W$ functions are too.

# Intuition

Given (SC), if we order odds as $R_1 \leq \ldots \leq R_n$ then the set of $\theta$'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 $\theta$ 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.

# Intervals are probability-determining on $\mathbb{R}$

... the market share of a horse maps into the set of preferences that choose it

... under our assumptions this set maps into an interval of $\mathbb{R}$

... and we know the measure of all such intervals, essentially all intervals in $(1/n, 1]$

... and we apply the theorem in the title.

# Estimating and testing

The indifference condition

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

can be rewritten as

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

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

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

# The Model is Identified and Testable

**Testable:** that only 4 of the *n* arguments matter,
+ separability implications,
+ monotonicity implications,
and more implications if we restrict admissible *V*'s (e.g. expected utility.)
**Identifiable:** up to the obvious increasing transformation.
i.e. we recover the distribution (over $\theta$) of the MRS of risk and return (or fear-of-ruin.)

# Going Further: Testing Expected Utility

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

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

# Expected Utility: Step 1—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 $\alpha$ the log-likelihood

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

We use the boundary condition:

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

so we can estimate the vNM utility function "nonparametrically iteratively" for any given $\theta$:

- start from $u^1(R, \theta) = 1/(R+1)$ for instance;
- then

$$u^{m+1}(r, \theta) = E\left(\frac{p_i(R)}{p_{i+1}(R)} u^m(R_i, \theta_i(R)) | R_{i+1} = r, \theta_i(R) = \theta\right).$$

- renormalize so the average $u^{m+1}$ is one.

# 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

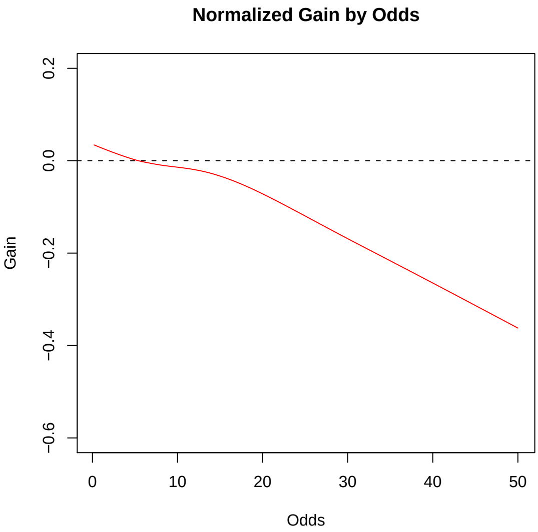
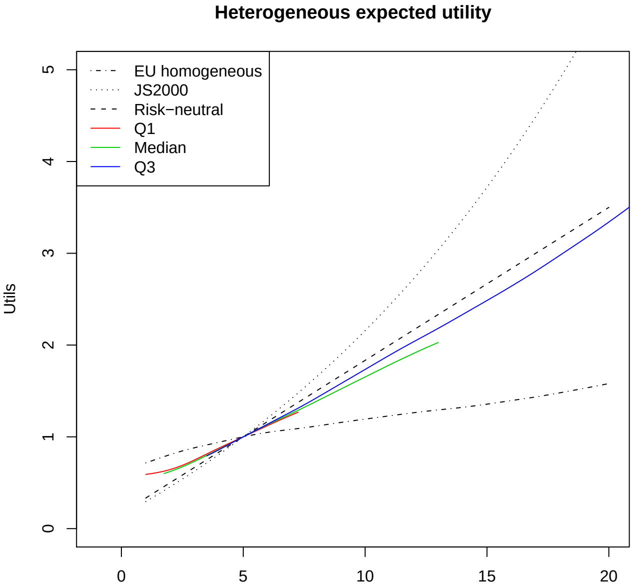
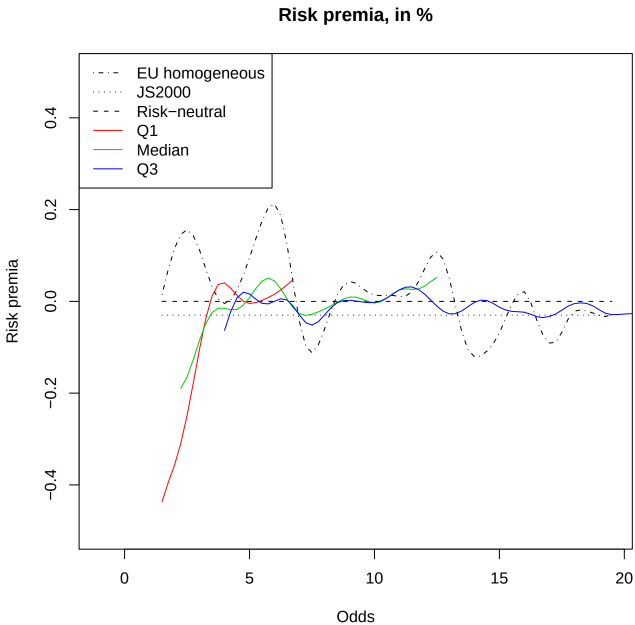



Figure: Normalized Expected Gains

# Results with Expected Utility Bettors



# What of Risk Aversion?



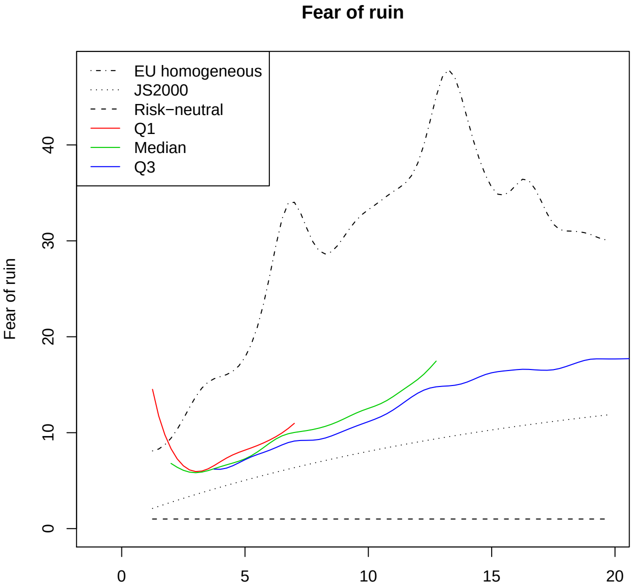
# Does Single Crossing Hold?

We assumed (translated in the expected utility world)

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

decreases in $\theta$ for all $R$.
We did not impose it for estimation, so we plot it with our estimates.

# Single Crossing is not Rejected



# More generally

The identification approach suggests an iterative estimation procedure *that does not rely on preestimating probabilities*:
At step $s$, say we have approximations of probability of winning $p^s_{i,c}$ for each horse $i$ in any race $c$;
Then we update for any horse $j > 1$ in a race $d$:

$$p^{s+1}_{j,d} = \Pr(i \text{ won race } c | R_{i,c} = R_{j,d}, R_{i-1,c} = R_{j-1,d}, p^s_{i,c} = p^s_{j,d}, \theta_i(R^c) =$$

(and completing for $j = 1$ by adding-up constraint.)
Remember the equation

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

If the iterations abvove converge, then they converge to the true probabilities, and the RHS gives us the $W$ function.

# To do list

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