

# Identifying Risk Preferences : Vertical Differentiation at the Racetrack

Amit Gandhi

October 3, 2006

# Goals for Today

To (hopefully) present :

- ▶ A stripped down (i.e., "purely vertical") version of Gandhi's (2006) model of prediction market (aka betting market) equilibrium.
- ▶ Testable restrictions implied by the (above) equilibrium that are sufficient for non-parametrically identifying the key underlying model primitive - The distribution of risk preferences. (Joint with PA Chiappori, B Salanie, F Salanie).
- ▶ Some preliminary tests and estimates.

# What Do We Know About Risk Preferences?

Not that much:

- ▶ intuition drawn from theory + casual observation (Arrow 65):
  - ▶  $ARA(x) = -u''(x)/u'(x)$  should be decreasing, since richer people buy more risk;
  - ▶  $RRA(x) = -xu''(x)/u'(x)$  should be close to constant, as the proportion of wealth invested in risky assets is fairly constant across wealth levels (?).
- ▶ but this completely neglects composition effects, inter alia.
- ▶ financial and insurance evidence: all over the map these days, RRA from 2 to 30.

# 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”).

Yet many market model in the economics of uncertainty do not (or cannot) take this heterogeneity very seriously.

Can we document this heterogeneity, and the departures from EUT, on “actual” data?

# Prediction Markets

In the prototypical laboratory experiment or survey response, participants make discrete choices from a menu of gambles.

Prediction markets offer a “real world” experiment.

- ▶ What are they? Very simple. Single period markets for the trade of Arrow-Debreu securities. If there are  $S$  possible future states of nature, then security  $i$  pays off \$1 in the event state  $i$  occurs, and \$0 otherwise. Examples include the Iowa Electronic Markets, Tradesorts, and more traditionally, *the odds market at racetracks*.
- ▶ Recent normative interest in prediction markets centers around interpretation of the price  $r_i$  as the “market belief” about  $p_i$  - the “true” probability of occurrence of state  $i$ . How to design prediction markets to achieve perfect information aggregation, i.e.  $r_i = p_i$ . (Ledyard, Plott, Wolfers, Manski).
- ▶ Our interests are positive. Each horse  $(r_i, p_i)$  defines a simple gamble. Consumers can be reasonably described as making discrete choices from the menu of gambles  $G = \{(r_i, p_i)\}_{i=1}^n$  offered by the market. Can we identify features of bettor risk preferences from racetrack data?

# Odds Versus Prices

It is useful to reparameterize Arrow-Debreu prices  $r_i$  in terms of odds  $R_i$  :

If we invest  $M$  dollars in horse  $i$ , our net return on investment conditional on winning is

$$R_i = \frac{\frac{M(1-\tau)}{r_i} - M}{M} = \frac{1-\tau}{r_i} - 1.$$

$R_i$  is the *odds* on horse  $i$ .  $R_i \in [-1, \infty)$ .

*Key point to remember* : We can always think of  $R_i$  as  $R(r_i)$ .

# The Data

.Our data is a large number of races  $m = 1, \dots, M$

A race  $m$  consists of

- ▶ a number of horses  $n^m$
- ▶ a vector of odds  $R_i^m$  for  $i = 1, \dots, n^m$
- ▶ the index  $f^m$  of the horse that won race  $m$ .

Thus we cannot (directly) observe

- ▶ Individual level choices
- ▶ The probabilities of winning  $(p_1, \dots, p_n) \in \Delta^{n-1}$ .

How do we fill this void?

# The Parimutuel Mechanism

If  $(s_1, \dots, s_n) \in \Delta^{n-1}$  are the distribution of bets (i.e. aggregate demands) over horses, then the odds are determined by the market clearing (“parimutuel”) rule

$$R_i = \frac{1 - \tau}{s_i} - 1.$$

Thus the Arrow-Debreu price  $r_i = s_i$ . Market shares = prices! E.g., If 20 percent of all dollars bet are allotted to horse  $i$ , then  $r_i = .20$ . In particular then  $(r_1, \dots, r_n) \in \Delta^{n-1}$ .

# Distribution of Preferences

Assume a population of bettors  $T$ , stable in time (given some observed characteristics)—*participation is left for future work*.

Take one of them: (s)he values a \$1 bet that

- ▶ wins (net)  $\$R$  with probability  $p$
- ▶ loses \$1 with probability  $(1 - p)$

as  $V(p, R)$ .

e.g., with expected utility theory (EUT),  $u$  rebased at current wealth:

$$V(p, R) = pu(R) + (1 - p)u(-1).$$

or, for rank dependent expected utility (RDEU)

$$V(p, R) = g(p)u(R) + (1 - g(p))u(-1).$$

Let  $P_V$  be the distribution of preferences over  $T$ .

# Equilibrium

- ▶ Assume each  $t \in T$  has common and unbiased beliefs  $(p_1, \dots, p_n) \in \Delta^{n-1}$ .
- ▶ Each  $t \in T$  bets 1 dollar, and makes a utility maximizing discrete choice from the menu of gambles  $G = \{(R(r_i), p_i)\}_{i=1}^n$ .

Then the market share of horse  $i$  is

$$s_i = \Pr(V(p_i, R(r_i)) \geq V(p_j, R(r_j)) \quad \forall j = 1, \dots, n),$$

where the probability is over  $V$  in the population of bettors. But parimutuel mechanism sets  $r_i = s_i$ .

Thus equilibrium prices  $(r_1, \dots, r_n) \in \Delta^{n-1}$  solve

$$r_i = \Pr(V(p_i, R(r_i)) \geq V(p_j, R(r_j)) \quad \forall j = 1, \dots, n).$$

# A Simple Case : 1D-SC Preferences

Suppose all possible references  $V(p, R)$  can be written as instances of  $W(p, R, \theta)$  where

- ▶  $\theta \in \Theta = (\theta_{min}, \theta_{max}) \subset \mathbb{R}$  with continuous, strictly increasing cdf  $F(\theta)$ .
- ▶ The “master function”  $W$  is continuous in all arguments, strictly increasing in  $p$  and  $R$ , and for each  $\theta$ , strictly minimized at  $p = 0$ .
- ▶ **Condition (SC):** The marginal rate of substitution  $W'_R/W'_p$  increases in  $\theta$ .

(SC) means that larger  $\theta$ 's prefer longer odds;

# Equilibrium in the 1D-SC “vertical” model

Fix beliefs  $p_1 > \dots > p_n$ . Consider prices  $r_1 > \dots > r_n$ . Assuming 1D-SC, then the set of  $\theta$ 's who bet on horse  $i$  is some interval  $[\theta_{i-1}, \theta_i]$  where

$$\theta_0 = 0 < \theta_1 < \dots < \theta_{n-1} < \theta_n = 1,$$

and for  $i = 1, \dots, (n - 1)$ ,

$$W(p_i, R(r_i), \theta_i) = W(p_{i+1}, R(r_{i+1}), \theta_i) \quad (I_i).$$

## Equilibrium pricing function $r_i(p_1, \dots, p_n)$

However, by the parimutuel mechanism, we know that

$$r_i(\theta) = F(\theta_i) - F(\theta_{i-1}) \quad \forall i = 1, \dots, n,$$

which we can invert to obtain

$$\theta_i(r) = F^{-1}\left(\sum_{j=1}^i r_j\right) \quad \forall i = 1, \dots, n-1.$$

Then, recursively use the  $(n-1)$  marginal conditions  $l_1, \dots, l_{n-1}$  to uniquely solve for  $(r_1, \dots, r_n) \in \Delta^{n-1}$ . That is, solving the  $i^{\text{th}}$  marginal condition for  $r_{i+1}$  yields

$$r_{i+1} = r(r_i, p_i, p_{i+1}, \theta_i(r_1, \dots, r_i))$$

For any choice  $r_1^*$ , build  $r_2(r_1^*), r_3(r_2(r_1^*)), \dots$ . By monotonicity of  $r$  there exists unique  $r_1^*$  such that

$$r_1^* + \sum_{i=2}^n r_i(r_1^*) = 1.$$

## Inverse Equilibrium Pricing Function $p_i(R_1, \dots, R_n)$

Suppose we observe odds  $R_1 < \dots < R_n$  (i.e. prices  $r_1 > \dots > r_n$ ). Then once again, by the parimutuel mechanism, we know the locations of the  $(n - 1)$  marginal bettors  $\theta_i(R)$ . Solving the  $i^{\text{th}}$  marginal condition for  $p_{i+1}$  yields

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

which is increasing in  $p_i$ . Then there exists unique  $p_1^*$  such that  
Thus there exists unique  $p_1^*$  such that

$$p_1^* + \sum_{i=2}^n p_i(p_1^*) = 1.$$

# Identification (finally) : What We Can Prove

**Theorem:** let  $F_0$  be the true cdf of  $\theta$  on an interval  $\Theta$  of  $\mathbb{R}$ ; then

- ▶ the data uniquely identify  $F_0$ ;
- ▶ the assumption that all preferences belong to  $W(.,., \Theta)$  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  $W$ .

## Some Notation

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

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

$\Gamma$  increases in  $v$ , decreases in  $R$ , and  $\Gamma''_{R\theta} < 0$  by (SC).

Then use change of variables:

$$\phi_1 = p_i(R); \phi_2 = \theta_i(R); \phi_3 = R_i; \phi_4 = R_{i+1};$$

complete with  $R_1, \dots, R_{i-3}$  and  $R_{i+2}, \dots, R_n$  if non-empty

and define  $\pi_{i+1}(\phi) = p_{i+1}(R)$ . The data pin down  $\pi_{i+1}$ .

(Intuitively, first estimate  $p_{i+1}(R)$  and then back out  $\pi_{i+1}(\phi)$  by “regressing”  $p_{i+1}(R)$  on

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

# An example

Say the race has 8 horses and we look at  $i = 6$  (the penultimate longshot):

the list of variables  $R = (R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8)$

turns into the new variables  $\phi = (p_6(R), \theta_6(R), R_6, R_7, R_1, R_2, R_3, R_8)$ .

# The New Indifference Condition

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

becomes

$$\pi_{i+1}(\phi) = \Gamma(W(\phi_1, \phi_3, \phi_2), \phi_4, \phi_2) \quad (J_i).$$

Immediate consequence:

$$\pi_{i+1} \text{ does not depend on } i, \quad \text{and} \quad \frac{\partial \pi_{i+1}}{\partial \phi_k} = 0 \text{ for } k > 4. \quad (IC)$$

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

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

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

# Another Equality Condition

Look again:

$$\pi_{i+1}(\phi) = \Gamma(W(\phi_1, \phi_3, \phi_2), \phi_4, \phi_2) \quad (J_i).$$

The “marginal rate of substitution” between  $\phi_1$  and  $\phi_3$ , i.e.

$$\frac{\frac{\partial \pi_{i+1}}{\partial \phi_1}}{\frac{\partial \pi_{i+1}}{\partial \phi_3}}$$

does not depend on  $\phi_4$ ; call this condition (*MRS*).

# Sufficiency?

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

$$\pi_{i+1}(\phi) = G(H(\phi_1, \phi_3, \phi_2), \phi_4, \phi_2)$$

for some functions  $G$  and  $H$ .

We would like to identify  $H$  to  $W$  and  $G$  to  $\Gamma$  (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

$G$  increases in  $H$  and decreases in  $\phi_4$ .

# Sufficiency!

These additional conditions turn out to boil down to:

$$\pi_{i+1} \text{ increases in } \phi_1 \text{ and in } \phi_3; \quad (V_1)$$

$$\pi_{i+1} \text{ decreases in } \phi_4; \quad (V_2)$$

and the MRS of  $\pi_{i+1}$  in  $(\phi_1, \phi_3)$ , i.e.

$$\frac{\frac{\partial \pi_{i+1}}{\partial \phi_3}}{\frac{\partial \pi_{i+1}}{\partial \phi_1}}$$

increases in  $\phi_2$  (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

◀ Back

Given the estimated  $\pi_{i+1}(\phi_1, \phi_2, \phi_3, \phi_4)$  function, we fix  $\phi_2 = \theta$ ; for any point in the  $(\phi_1, \phi_3) = (p, R)$  plane we know that the indifference curve of any representation of  $W(p, R, \theta)$  has slope

$$\frac{\frac{\partial \pi_{i+1}}{\partial \phi_1}}{\frac{\partial \pi_{i+1}}{\partial \phi_3}}(\phi_1, \phi_2, \phi_3, \phi_4)$$

(for any value of  $\phi_4$ ).

This gives a **test for misspecification**:

Once the indifference curve for  $\theta$  that goes through  $(p, R)$  is constructed, choose some odds  $R'$  and compute

$$p' = \pi_{i+1}(p, \theta, R, R');$$

then  $(p', R')$  should lie on that same indifference curve.

# Going Further: Cumulative Prospect Theory

CPT is equivalent to

$$\frac{\partial^2 \log \frac{\partial W}{\partial R}}{\partial p \partial R} = 0$$

for *one* representation of  $W$ .

Not straightforward to test (nonparametrically).

## Expected utility is easier

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

$$\pi_{i+1}(\phi) = \phi_1 \frac{u(\phi_3, \phi_2)}{u(\phi_4, \phi_2)}$$

Thus EUT yields two additional conditions; define

$$\psi_{i+1}(\phi) = \log(\pi_{i+1}(\phi)/\phi_1):$$

$$\psi_{i+1}(\phi) \text{ only depends on } \phi_2, \phi_3 \text{ and } \phi_4 \quad (EU_1)$$

and

$$\frac{\partial^2 \psi_{i+1}}{\partial \phi_3 \partial \phi_4} = 0 \quad (EU_2).$$

# 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 all  $\theta$
- ▶ then  $u(R, \theta) = E(p_i/p_{i+1} | R_{i+1} = R, R_i = R_0, \theta)$ .

# Testing Homogeneous Risk Preferences

An easy one: just add

$$\frac{\partial \pi_{i+1}(\phi)}{\partial \phi_2} = 0.$$

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

# Empirical Strategy: 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

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

Let  $\hat{\alpha}$  be the estimate.

# Testing

Proceed further: we already have the probabilities  $\hat{p}_i^m = P_i(R^m, \hat{\alpha})$ ;  
Now we specify the  $\pi_{i+1}$  function:

$$P_{i+1}^m - P_i^m = \sum_{k=0}^K A_k^m (R_{i+1}^m - R_i^m)^k + \varepsilon,$$

with the  $A_k$  functions of  $(\theta_i(R^m), R_i^m, \hat{p}_i^m)$  and “additional regressors” (excluded by our theory).

**Result 1:** with  $K = 2$ , the  $R^2$  of this regression is 0.9990.

**Result 2:** without the “additional regressors”, it is still 0.9988.

So we very much fail to reject (IC).

# Measuring 1-dimensional-Real Heterogeneity

If we take out the  $\theta_i(R^m)$  factors in the  $A_k$ 's, the  $R^2$  falls to 0.9782.

Hugely significant statistically (337,000 horses...)  
rather less in pragmatic terms.

# Rejecting the SC1 Model

We use the estimated  $\pi$  function to check that

- ▶ it increases in  $\phi_1$ : yes for 96% of horses;
- ▶ it increases in  $\phi_3$ : yes for 98% of horses;
- ▶ the resulting MRS decreases in  $\theta$ : **no for 17% of horses;**
- ▶ it does not depend on  $\phi_4$ : **in fact  $\phi_4$  explains 4% of the variation in MRS.**

Thus the one-dimensional *plus* single-crossing model does not fit the data well as it should (still not bad - checking robustness of regression specification).

# Is Expected Utility a Good Approximation?

In this (**tentatively rejected**) model it turns out to be:  
The additional restrictions are not rejected.