Rational Expectations at the Racetrack : Testing Expected Utility Using Prediction Market Prices

Amit Gandhi *
University of Chicago

November 28, 2006

Abstract

Empirical studies have cast doubt on one of the bedrocks of applied economic modeling - the expected utility hypothesis. Economists have documented pricing anomalies, like the long-shot bias in prediction markets (low probability events are priced too high), that are inconsistent with representative agent models. In this paper, we show that the inconsistency is due to the representative agent assumption, and not to the expected utility hypothesis. When agents differ in their information sets and risk preferences, we show that trader heterogeneity can easily explain the observed pattern of price variation across betting and prediction markets. In particular, the long shot bias is found to be due to a group of traders, whom we dub the "risk-averting grandmas", who make up about 40 percent of the trading group and bet on the top favorite in a race in exchange for a premium. We show also that the expected utility hypothesis outperforms the main "behavioral" alternatives, rank dependent expected utility, and cumulative prospect theory.

*I am extremely grateful to Pierre Andre Chiappori, Jeremy Fox, Luke Froeb, Amil Petrin, Philip Reny, Bernard Salanie, and Francois Salanie for their advice and encouragement while pursing this project. Workshop participants from the Chicago IO, Econometrics, and Theory seminars, and the USC Theory seminar are also to thank for their many constructive comments.

1 Introduction

A fundamental assumption about human behavior used in modern economic modelling is the expected utility hypothesis (EUH). In its most basic form, the EUH is a hypothesis about the nature of individual preferences for risky prospects, i.e., lotteries over monetary outcomes. The EUH maintains that probability enters linearly into an economic agent's preference for risk, leading the agent to act so as to maximize his expected utility when faced with a choice among lotteries. Thus the key device that expected utility theory uses to explain an individual's attitude towards risk is the utility function over wealth.

However, the empirical validity of the EUH has been subject to vigorous debate since its inception (for a historical review, see e.g., Starmer (2000)). In particular, the assumption that probability enters linearly into the calculus of comparing lotteries has been called into question by a series of well documented experimental effects and examples.[1] Outside of the laboratory, the well known "favorite-longshot bias" in racetrack and betting markets (Griffith, 1949; Thaler and Ziemba, 1988), which finds that betting on a horse more favored to win by the market is more profitable on average than betting on a horse less favored, was one the first phenomena to motivate the idea that people treat probability nonlinearly - i.e., favorites appear undervalued by the market and longshots appear overvalued, suggesting that people underweight large probabilities and overweight small probabilities. These examples and numerous others like them[2] have led to the development of a voluminous literature on "non-expected utility" theories (well over a dozen such theories exist, see e.g., Fishburn (1988)), which attempt to relax expected utility theory in ways that better fit the experimental evidence.

[1] For example, experimental effects such as the "Allais paradox" (Segal, 1987), the common ratio effect, and the preference reversal effect (Karni and Safra, 1987) have all been put forth as evidence that people nonlinearly distort probabilities during decision making.

[2] A seminal work that experimentally uncovered many violations of expected utility theory in a systematic way is Kahneman and Tversky (1979), which is today the second most cited paper in economics (Kim et al., 2006).

In this paper, we show how an emerging class of markets known as "prediction markets" (Wolfers and Zitzewitz, 2004) can be used to examine the expected utility hypothesis against real world market data. Examples of prediction markets include the odds market at horse racetracks, as well as the more recent online exchanges such as Tradesports, Betfair, and the Iowa Electronic Market. Simply put, these are markets that price uncertain events. Thus for example, "Hilary Clinton wins the '08 election" is an uncertain event, and by allowing people to buy and sell an asset that pays 1 dollar in the event that she wins and zero dollars otherwise, the market price of the asset puts a price on the event.

Economists (e.g., Plott et al. (2003); Manski (2006); Wolfers and Zitzewitz (2006)) have recently become interested in the relationship between the price of an event in a prediction market and the underlying probability that the event will occur (e.g., if the Hilary Clinton asset has a price of .30 dollars, what is the relationship between .30 and her true chance of success). The favorite longshot bias at horse racetracks reflects a particular price/probability relationship - high priced horses (the favorites) are undervalued relative to their chance of success, and low priced horses (longshots) are overvalued. Different price/probability relationships have been discovered at racetracks in different countries

We develop a general equilibrium model of betting and prediction markets and show that in equilibrium, the price/probability relationship is determined by the distribution of risk attitudes in the betting population, i.e., by the pattern of equilibrium trade between heterogeneous risk types. We then show how our general equilibrium model allows us to use the *actual* price/probability relationship that is revealed by standard prediction market data to both nonparametrically identify and structurally estimate the distribution of risk attitudes among bettors. By comparing the structural estimates derived from different theories of decision making under risk, we can explicitly test how well expected utility explains the market data relative to the behavioral alternatives. Our analysis of the odds data from all US racetracks over a three year period (2001-2003) finds that simple EU functional forms, such as CRRA and CARA preferences, do a very good job of explaining the price/probability relationship observed in the data. The favorite longshot bias in particular is generated by a natural exchange between risk lovers and risk averters : risk lovers overbet longshots in order to finance the incentive for risk averters to bet on favorites. Moreover, we do not find evidence for nonlinear probability behavior by bettors, contrary to the conclusions from the experimental literature.

A key contribution of our empirical analysis is to show that by controlling for the heterogeneity of risk preferences, the expected utility hypothesis is capable of explaining away apparent pricing anomalies, such as the favorite longshot bias, in a sensible way. In order to explain the favorite longshot bias, the literature to date has relied upon a representative bettor assumption. A representative bettor assumes that all bettors in the market have the same preferences for risk. In equilibrium, the odds on each horse are such that the representative bettor is indifferent between betting on each horse in the race (since otherwise some horses would receive no bets, which cannot happen in equilibrium).

Clearly, the favorite longshot bias is inconsistent with a risk neutral representative bettor, since such a bettor would strictly prefer betting on the horse with the highest expected return (which under the favorite longshot bias is the horse most favored to win). Thus in order to neoclassicaly explain the favorite longshot bias with a representative bettor, this bettor must be a risk loving expected utility maximizer, who is willing to accept the lower expected return from betting on longshots because of the larger potential upside these bets offer (Weitzman, 1965).[3]

The few papers to date that test expected utility against *market* data (as opposed to experimental data, where it is already well established that expected utility has failings) have made fundamental use of the representative agent model. In the seminal paper to economet-

[3]In a similar result, (Quandt, 1986) shows that the favorite longshot bias is a necessary consequence of risk loving, mean-variance expected utility maximizing behavior among bettors (not necessarily a representative bettor).

rically compare expected and non-expected utility (EU and non-EU) theories against racetrack odds, Jullien and Salanie (2000) find that a non-EU representative agent empirically outperforms a risk loving EU representative agent. Using a different empirical methodology, but maintaining the representative bettor model, Snowberg and Wolfers (2005) arrive at a similar conclusion as to the empirical superiority of non-EU preferences in explaining the pattern of prices at the racetrack. Thus these market studies point in the direction of rejecting expected utility theory, a fact that has been duly noted by the behavioral literature Camerer (2000).

However, as we show, by allowing for risk averse bettors to enter the population and trade with the risk loving bettors, the apparent superiority of non-EU preferences disappears. Thus in the context of a richer *general* equilibrium model of market prices, expected utility generates a much more sensible explanation of the market data.

A key step in both our theoretical and empirical analysis is observing that betting/prediction markets are structurally identical to product differentiated markets of the kind that have been extensively studied in the industrial organization literature. Essentially, the horses in a race can be viewed as being "vertically" differentiated from one another by their probability of winning, and bettors differ from one another by the their "willingness to pay for quality", i.e., their risk aversion. This connection between prediction markets and product differentiated markets motivates our general equilibrium approach and empirical strategy, which we now preview.

A Preview of the Model and Empirical Strategy

As already mentioned, betting and prediction markets come in a variety of popular forms, ranging from the odds market at horse racetracks, to the more recent online exchanges such as Tradesports and the Iowa Electronic Market. The common thread tying together these markets is that they are single period, ex-ante markets for the trade of a complete set of Arrow-Debreu securities. For simplicity, we shall use the language of "horse races" to describe prediction markets more generally. Thus, consider a race with n horses running. Betting on horse i to win is equivalent to buying an Arrow-Debreu security that pays off 1 dollar in the event horse i wins, and 0 dollars otherwise. The "price of horse i" is the price of this Arrow-Debreu security. There are n such securities in the betting market, and n such prices. Actual prices at the racetrack are typically presented in the more familiar form of betting odds : the odds on a horse is related to the inverse of its Arrow-Debreu price. Thus "cheap" horse (i.e., "longshots") have long odds, and "expensive" (i.e., "favorites") horse have short odds.

The key to our equilibrium approach is recognizing that betting and prediction markets are essentially product differentiated markets of the kind that have been extensively studied in the industrial organization literature (e.g., Berry et al. (1995)). In a given race, horses can be viewed as differing both by the probability that they will win p, and by their price R. That is, horses in a market can be viewed as being differentiated "vertically" along the quality dimension p, and given its price R, a horse can be represented simply as a price/probability pair (R, p). In this way, betting markets come as close as any market to offering consumers a menu $G = \{(R_1, p_1), \ldots, (R_n, p_n)\}$ of simple lotteries akin to those used in choice experiments, providing a natural laboratory to test theories of individual choice under risk. More critically, due to the one period nature of a betting market (and the geographic distance between tracks), it is possible to view prices the prices $(R_1^k, \ldots, R_n^k)$ across different markets $k = 1, \ldots, K$ (i.e. across different races) as being determined independent of one another.[4] This stands in contrast to traditional financial securities, such as stocks, whose returns are clearly dynamically linked across markets.

We model a betting/prediction market as a standard "textbook" Arrow-Debreu security

[4]More precisely, since different markets price different events, each market is only open for a short period preceding the race, and arbitrage is unfeasible since one can only buy tickets, then one can legitimately consider each race in isolation.

market. Thus given the exogenous qualities of the horses, i.e., the probability distribution $(p_1, \ldots, p_n)$, we model the market prices $(R_1, \ldots, R_n)$ as being determined by a competitive, rational expectations equilibrium. However we introduce an identifying assumption into the Arrow-Debreu framework that has proven extremely useful in the industrial organization literature on product differentiation (Bresnahan, 1987; Berry et al., 1995), namely the assumption of discrete choice behavior by consumers. That is, we postulate a population of bettors T with a distribution $\mathbf{P}_V$ over their risk preferences $V(R, p)$ for simple gambles (R, p). A bettor $t \in T$ chooses to bet his "endowment" (the amount of money alloted for the race) on the preferred price/probability combination (R, p) (i.e., a horse) offered by the market. Such discrete choice behavior is consistent with how people seem to place bets in these markets (Thaler and Ziemba, 1988).

Our first main result shows that by introducing the discrete choice assumption into the usual Arrow-Debreu framework, we can uniquely solve for equilibrium prices under very general assumptions on the distribution of preferences and the distribution of information among agents. In particular, we show that for any distribution of preferences $\mathbf{P}_V$ satisfying weak regularity conditions (the distribution is atomless, all consumers preferences are continuous and increasing), regardless of the particular distribution of information (so long as there is enough information in the market), there exists unique equilibrium prices. This result has two important consequences for our empirical strategy. First, unique equilibrium prices exist without requiring us to make any parametric assumptions about the functional form of bettor preferences, i.e., assumptions such as requiring each bettor $t \in T$ to have preferences $V_t(R, p)$ of the form $V(R, p, \theta_t)$ for some finite dimensional parameter $\theta_t \in \Theta$. Thus the equilibrium theory is consistent with a wide range of underlying preference theories, such as those suggested by EU, RDEU, etc. Second, the equilibrium solution of the model gives rise to a reduced form relationship $\mathbf{R}(p_1, \ldots, p_n)$ between the qualities of the horses $(p_1, \ldots, p_n)$ and the market clearing prices $(R_1, \ldots, R_n)$ in a race.

Our second main result shows that this reduced form relationship $\mathbf{R}(p_1, \ldots, p_n)$ between prices and probabilities is invertible. This invertibility reflects the rational expectations nature of the equilibrium : agents can invert equilibrium prices to learn the probabilities in a race. The existence of this inverse reduced form relationship $\mathbf{p}(R_1, \ldots, R_n)$ provides the basic key to estimating the structural model. As the econometrician, what we can actually observe in the data are the prices $(R_1^k, \ldots, R_n^k)$ and the index of the winning horse i_w^k across a sample of races $k = 1, \ldots, K$. Thus we cannot directly observe the qualities p_i of the horses $i = 1, \ldots, n$ in the race, and hence cannot directly identify the reduced form $\mathbf{R}(p_1, \ldots, p_n)$ from the data. However, since we can observe a draw from the probability distribution $(p_1, \ldots, p_n)$ in the form of the horse that wins the race, the data do identify the inverse reduced form $\mathbf{p}(R_1, \ldots, R_n)$.

Putting our two main results together gives us our estimation strategy. Suppose we make a parametric assumption about the form of risk preferences. Thus for every bettor $t \in T$, $V_t(R, p) = V(R, p, \theta_t)$ for some $\theta_t \in \Theta \subset \mathbb{R}^m$. Then the only unknown primitive of the structural model is the distribution F over bettor types θ. Through our equilibrium theory, any such F implies an inverse reduced form relationship $\mathbf{p}(R_1, \ldots, R_n; F)$. We can thus estimate the unknown F by maximum likelihood since the winning horse in a race is an outcome of the multinomial trial $\mathbf{p}(R_1, \ldots, R_n; F)$, and the trials across races are independent of one another.

This estimation strategy hinges critically on solving the reduced form of the model. While our equilibrium theory supports the unique existence of the inverse reduced form relationship between prices and probabilities, the actual estimation of the distribution F depends upon our ability to solve for this relationship. While we could pursue numerical methods to achieve this solution, the sheer size of the number of races at our disposal (all North American races over a 3 year period, constituting some 200,000 races), and the fact that the average number of horses is close to 10 (and hence we are solving for on average 10 unknowns $(p_1, \ldots, p_n)$ in each race), makes it extremely expensive just to compute the likelihood function for a single F.

In our final set of results before turning to the empirical analysis, we show that when we restrict the heterogeneity of risk preferences to be "one dimensional", the inverse reduced form admits a simplification that makes the estimation problem tractable. One dimensional heterogeneity means that individuals differ along a single dimension $\theta \in \mathbb{R}$, which is completely natural in our setting since the horses in a race differ along a single vertical dimension, namely the probability of winning p. In the same spirit as the industrial organization literature on vertical differentiation (Shaked and Sutton, 1982; Bresnahan, 1987), we assume that the type θ orders individuals in terms of their price sensitivity, which in our setting translates into "willingness to take risk". Such a parametric structure on preferences causes the inverse reduced form $\mathbf{p}(R_1, \ldots, R_n)$ to decompose in a convenient way that allows us to nonparametrically identify and estimate the distribution F. The standard EU functional forms such as CRRA and CARA are cases of one dimensional preferences.

In our empirial analysis, we compare CRRA/CARA preferences to the main behavioral alternative. Among the most studied of the non-expected utility theories are rank dependent expected utility (RDEU) (Quiggin, 1982) and cumulative prospect theory (CPT) (Tversky and Kahneman, 1992). The key device that these theories use to describe an individual's attitude toward risk is the individual's probability weighting function, which transforms the probabilities that define a lottery into decision weights.[5] Thus rank dependent and cumulative prospect theory can be thought of as the "duals" to the EU model (Yaari, 1987) - they attempt to describe risk attitudes through preferences that are nonlinear in probability rather than nonlinear in wealth. Nonlinear probability weighting accounts for the main experimental anomalies in the literature (Starmer, 2000), which has led to repeated calls from

[5] Cumulative prospect theory generalizes expected utility one step beyond probability weighting by allowing for the asymmetric treatment of losses and gains. We do not explore this aspect of CPT in the current paper and is a topic of ongoing research.

the experimental community to abandon expected utility theory in applied economic modelling (see e.g., Rabin and Thaler (2001)). Yet simple expected utility representations, such as time separable, constant relative risk averse (CRRA) preferences, continue to dominate the literature on asset pricing, macroeconomics, contract theory, etc (Chiappori, 2006).

Using a data set consisting of all North American races over a three year period (2001-2003), we estimate two different models of one dimensional preference heterogeneity. Under expected utility theory, heterogeneity of risk attitudes is generated by allowing for individual differences in the curvature of utility. Thus in the first model, types θ have different curvatures in their utility for wealth $u_\theta(w)$ (captured through a power function, i.e., a CRRA functional form $u_\theta(w) = w^\theta$) and act according to EU theory (i.e., they maximize $V_\theta(p, R) = pu_\theta(R)$). Rank dependent and cumulative prospect theory on the other hand suggest that individual differences arise through differences in the curvature of the probability weighting function (Gonzalez and Wu, 1999). Thus in the second model, agents have different curvature in their probability weighting function $G_\theta(p)$ (also captured through a power function, i.e., $G(p) = p^\theta$), and act according to the probability weighting theory (i.e., they maximize $V_\theta(p, R) = G_\theta(p)u(R)$).

In the EU model, we find there to be economically significant heterogeneity of risk preferences : there are a large group (40 percent of the population) of risk averting "grandmas" who generally back the top favorite in a race, and then there is everyone else, who are risk loving, and generally back the remaining longshots. As we show, this form of trade between risk averters and risk lovers is not an accident, but rather reflects a key restriction of the expected utility hypothesis : If bettors are expected utility maximizers, then equilibrium prices exhibit the favorite longshot bias if and only if all the risk averters in the population back the top favorite in a race. Since our data are in fact characterized by the favorite longshot bias, we see this restriction coming out in our estimates. Thus if expected utility theory were in fact misspecified, then by turning to the non-EU probability weighting model, we should see support for even more risk aversion in the population. However in the second model, we find no such evidence - all bettors have perfectly linear probability weighting, causing the estimated model to collapse to a homogeneous risk loving EU population, which is empirically outperformed by the first model. Thus the "curvature of utility" theory of risk preferences proposed by EU is better supported by the data than the "curvature of probability weighting" theory proposed by RDEU/CPT, in sharp contrast to the experimental evidence.

Just how well do the predictions of the CRRA model fare? We compare the inverse reduced form of our estimated model $\mathbf{p}(R_1, \ldots, R_n; \hat{F})$ to a flexibly specified multinomial model $\mathbf{p}(R_1, \ldots, R_n; \hat{\beta})$, where β is a vector of parameters and $\hat{\beta}$ are its estimated values. The idea of the flexible model is to estimate the "true" reduced form contained in the data that does not make any structural assumptions. Another test of expected utility theory, and CRRA/CARA preferences in particular, is to test how much of the explanatory power of the flexible reduced form our structural model's reduced form is able to recover. We find a very strong result - virtually all of the R^2 from the flexible reduced form is recovered by our structural model with CRRA preferences. Said another way, if one wanted to write down an arbitrary statistical model for predicting the probabilities of winning from the market prices, one can hardly do better than use our structural model (i..e Arrow-Debreu theory) with CRRA bettors to derive this relationship.

Related literature Our approach to estimating risk preferences and testing the EUH against market data has important antecedents in terms of both style and substance in the economics literature. In terms of substance, we follow in the work of Jullien and Salanie (2000), who first recognized betting markets as a natural test bed for theories of decision making under risk. They also employ maximum likelihood methods for estimating different models of risk preferences. However their model of racetrack prices lacked any heterogeneity of information or preferences, leading them to estimate the preferences of a "representative bettor". Thus although they find statistically significant departures from expected utility theory, it is unclear whose preferences these departures represent, and more importantly, what is their economic significance. In terms of style, our approach closely resembles the pioneering work of Bresnahan (1987). Like us, Bresnahan (1987) models a vertically differentiated goods market (in his case, automobiles), and explicitly solves for the reduced form of a structural equilibrium model (in his case, oligopoly supply and demand). However his interests lie in testing which supply side assumptions (competition or collusion) gave the best empirically performing reduced form equations. Since there is no supply side in our exchange economy, our interests rather lie in testing assumptions about the demand side.

2 A General Equilibrium Model of Betting Markets

2.1 The Pricing Puzzle

Two basic facts about betting markets that have thus far defied a unified explanation by any economic model are the simultaneous efficiency and bias of prices (Sauer, 1998). The betting odds at racetracks have been shown to be quite informationally efficient in the sense that no information beyond the final market prices on each horse in a race is needed to predict the probability of each horse winning the race. Nevertheless, a horse's Arrow-Debreu price is a biased estimate of its probability of winning : in North American tracks, the prices on "favorites" (i.e., expensive horses) systematically underestimate the probability of winning, and the prices on "longshots" (i.e., cheap horses) systematically overstate, a pattern known as the *favorite-longshot bias.* Different nonlinear relationships between prices and probabilities, such as the reverse favorite-longshot bias, have been discovered in other countries.

Our general equilibrium approach is able to capture these basic empirical realities quite handily. The basic story behind the equilibrium is the following. Before the market opens at a given race, nature determines a state $(p_1, \ldots, p_n)$, the state being a probability distribution over the n horses running the race, i.e., a roulette wheel. Bettors come to the market with potentially different information concerning the underlying state, and they trade, using both their private information *and* market prices to update their beliefs (that is, bettors have rational expectations). The equilibrium prices/odds $(R_1, \ldots, R_n)$ that prevail at market close allow bettors to perfectly infer the underlying state (that is, the equilibrium is fully revealing). After the close of the market, a spin of nature's roulette wheel $(p_1, \ldots, p_n)$ determines the winning horse i_w in the race.

Under very general regularity conditions on the distribution of information and the distribution of preferences in the betting population, we show that there exists a unique fully revealing rational expectations price equilibrium in the betting market. The rational expectations equilibrium (REE) is a function $\mathbf{R}(p_1, \ldots, p_n)$ that maps any possible state of nature (i.e., a roulette wheel) $(p_1, \ldots, p_n)$ to a vector of market clearing prices $(R_1, \ldots, R_n)$. The fully revealing property of the REE means that the function $\mathbf{R}$ is invertible, with inverse $\mathbf{p}(R_1, \ldots, R_n)$ allowing bettors to perfectly infer the race's roulette wheel $(p_1, \ldots, p_n)$ from the market prices.[6] The unique existence of such a fully revealing REE thus explains the observed informational efficiency of betting market prices : in equilibrium, market prices are sufficient for perfectly inferring the true probability distribution over the horses in a race. Of course, since the REE will not generally be the identity map, the model readily allows for the observed "bias", or difference between prices and probabilities in a race.

2.2 Market Clearing Prices

As we did in the introduction, we shall continue to use the language of horse races to describe betting and prediction markets more generally. Consider a race with n horses running, with

[6] A special case of the model is when all bettors have private information that perfectly informs them of the state.

the *outcome* of the race being defined by the winning horse. An ex-ante market (i.e., before the race is run) is open for the trade of n Arrow-Debreu securities. A unit of security i buys 1 dollar in the event that horse i wins the race, and 0 dollars otherwise. Let r_i denote the Arrow-Debreu price of security i, and let M_i denote the total number of dollars in the market spent on purchasing security i. Since purchasing security i is equivalent to betting on horse i, we can equivalently refer to M_i as the total number of dollars bet on horse i. Define the *market share of horse* i, denoted s_i, to be the aggregate budget share of security i, i.e.,

$$s_i = \frac{M_i}{M_1 + \cdots + M_n}.$$

Finally, let τ denote the participatory tax per dollar bet, commonly called the track take. We now establish the following simple result that plays a central role throughout the paper.

Proposition 2.1 *The security market clears if and only if* $r_i = s_i$ *for each* $i = 1, \ldots, n$.

Proof Market clearing means the supply of dollars equals the demand of dollars in each of the possible n outcomes of the race. This happens if and only if

$$\begin{aligned} & \frac{(1-\tau)M_i}{r_i} = (1-\tau)(M_1 + \cdots + M_n) \quad (\forall i) \\ \iff \quad & r_i = s_i \quad (\forall i). \quad \blacksquare \end{aligned} \tag{1}$$

Thus a necessary condition for the security market to clear is that $\sum_{i=1}^{n} r_i = 1$, i.e., the Arrow-Debreu prices add up to 1.

Prices at the racetrack are not customarily quoted in terms of the Arrow-Debreu prices r_i, but are rather quoted in terms of the odds R_i on horse i. The odds R_i are defined as net profit per dollar bet on horse i in the event i wins the race. Thus if the odds on a horse are quoted as 2, and you bet 5 dollars on the horse, then if the horse wins you receive 15 dollars, your net profit being $(5)(2) = 10$. While the Arrow-Debreu prices are not explicitly quoted, they are nevertheless implicitly being quoted through the odds. That is, the odds $(R_1, \ldots, R_n)$ at a race implicitly define Arrow-Debreu prices $(r_1, \ldots, r_n)$, where

$$R_i = \frac{(1-\tau)}{r_i} - 1 \quad (\forall i).$$

Thus the odds market at a racetrack implicitly defines a textbook one-period and complete Arrow-Debreu securities market.

Using the market clearing condition (1), we have that market clearing odds are

$$R_i = \frac{(1-\tau)}{s_i} - 1. \tag{2}$$

The market clearing condition (2) is in fact how betting odds are institutionally determined at the racetrack, and parimutuel betting systems more generally. In view of Proposition 2.2, we can understand the so called "parimutuel mechanism" expressed by (2) as a method of setting the odds in a race so to ensure that the implicit Arrow-Debreu security market clears.

While the prices determined by (1) are market clearing, Arrow-Debreu equilibrium requires that they also be consistent with utility maximizing behavior on the part of the bettors at the track. That is, equilibrium occurs at prices $(r_1, \ldots, r_n)$ if aggregate demand at these prices results in each horse i's market share equaling r_i. That is, in equilibrium, people bet on horses in proportions equal to the prices. In order to model aggregate demand and explore the equilibrium problem, we now turn to the issue of preferences.

2.3 Preferences

Suppose a bettor has beliefs $(p_1, \ldots, p_n) \in \Delta^{n-1}$ over the possible outcomes of the race (where Δ^{n-1} is the $(n-1)$ dimensional simplex, i.e., the set of probability distributions over the horses), and is deciding which horse to back with the M dollars the bettor has alloted for the race. If the market odds are $(R_1, \ldots, R_n)$, then from the point of view of the bettor, each horse i in the race can be thought of as a simple gamble (R_i, p_i), which yields a gain of R_i per dollar bet with probability p_i, and yields a loss of -1 per dollar bet with probability $(1 - p_i)$. Thus from the point of view of the bettor, the market offers a choice among a menu of n gambles $G = \{(R_1, p_1), \ldots, (R_n, p_n)\}$.[7]

Assumption 2.2 (The Space of Preferences) *We postulate the existence of a* stable *(across races) continuum of consumers T. Each consumer $t \in T$ has a complete, continuous, transitive, and strictly monotonic preference relation $\succsim_t$ over simple gambles $(R, p) \in \mathbb{R}_+ \times [0, 1]$. The strictly worst gambles for any $t \in T$ are any gambles of the form $(R, 0)$.*

Thus each consumer t's preference relation can be represented by a continuous utility function $V_t : \mathbb{R}_+ \times [0, 1] \to \mathbb{R}$ that is strictly increasing in a gamble's net rate of return from winning R (the first argument of V_t) and probability of winning p (the second argument of V_t). In addition each consumer t's utility function is strictly minimized whenever $p = 0$, i.e., $V_t(0, R) = V_t(0, R')$ and $V_t(0, R) < V(p, R')$ for any returns R, R' and any probability $p > 0$. That is, the strictly worst gamble for any consumer is one that has no probability of winning, regardless of the return from winning (since this return is never realized).

Let $\mathbf{V} \subset \mathbb{R}^{\mathbb{R}_+ \times [0,1]}$ be the set of all such utility functions. We endow $\mathbf{V}$ with the relative product topology, otherwise known as the topology of pointwise convergence. Let the measurable sets in $\mathbf{V}$ be the Borel subsets (the σ-algebra of subsets generated by the open sets) in this topology. Our population T gives rise to a probability measure $\mathbf{P}_V$ over the space $\mathbf{V}$. The probability measure $\mathbf{P}_V$ describes the distribution of consumer preferences for gambles (R, p).

[7] More generally, we can allow the menu G to include the "no trade" option of not betting, which is equivalent to a gamble that offers a net rate of return zero with probability one. The equilibrium analysis we present still goes forward largely unchanged if the no trade option was included. To ease the current exposition, we leave the no trade option out of the choice set.

Our two final assumptions are mild regularity assumptions on the distribution of preferences $\mathbf{P}_V$. The first assumption requires that the probability measure $\mathbf{P}_V$ be sufficiently continuous, or atomless, so as to not permit a positive mass of consumers to be indifferent between two distinct gambles when at least one of the gambles has a non-zero probability of winning.

Assumption 2.3 (Continuity) *For any two distinct gambles (R_i, p_i) and (R_j, p_j) with p_i or p_j greater than 0 (or both), the number of consumers indifferent between gamble i and j has a probability measure of zero. More precisely, if $p_i > 0$ or $p_j > 0$ then*

$$\mathbf{P}_V(\{V \in \mathbf{V} : V(R_i, p_i) = V(R_j, p_j)\}) = 0.$$

If all consumers $t \in T$ have *common* beliefs $(p_1, \ldots, p_n) \in \Delta^{n-1}$, then for any odds $(R_1, \ldots, R_n)$, the market offers bettors a *common* menu of gambles G.

$$G = \{(R_1, p_1), \ldots, (R_n, p_n)\} \subset \mathbb{R}_+ \times [0, 1].$$

The subset of the population T that prefers the i^{th} gamble from such a common set G is denoted

$$S_i = \{V \in \mathbf{V} : V(R_i, p_i) \geq V(R_j, p_j) \text{ for all } j \neq i\}.$$

The share of the population T that prefers the i^{th} gamble from the common set G is thus

$$q_i(G) = q_i(R_1, \ldots, R_n; p_1, \ldots, p_n) = \mathbf{P}_V(S_i).[8] \tag{3}$$

We refer to q_i as the *market share* of the i^{th} gamble from G (i.e., if the market offered a choice of gambles from the common set G, then q_i is the share of the population T that

[8] S_i is measurable based on the topology on $\mathbf{V}$.

chooses the i^{th} gamble).

Our last assumption, desirability, requires that for any gamble $g = (p_g, R_g)$ with nonzero probability of winning $p_g > 0$, and for any finite common set of gambles G with $g \in G$, it is always possible to induce some positive mass of the population to prefer g from G by making g's return from winning R_g sufficiently large. If the EUH were true, then desirability would be automatically satisfied if, for example, any nonzero fraction of the population were risk loving. We now state these assumptions more precisely.

Assumption 2.4 (Desirability) *For any menu of common gambles, if $p_i > 0$, then it is possible to make R_i large enough such that the market share of horse i becomes positive.*

This completes our assumptions on the preferences of bettors. As can be seen, we have made only weak regularity assumptions on the space of preferences (monotonicity, continuity, etc) and the distribution of preferences (atomlessness, etc). Thus nearly any theory of risk preferences can be fit inside our space, which is an important point since main goal of the equilibrium theory to follow is to help us recover the "correct" preference theory from the data.

2.4 Information

Following the suggestion of Figlewski (1979), we consider the information problem facing bettors at the market as one involving a mixture of risk and uncertainty. In a given race, there exists some "true" objective probability distribution $(p_1, \ldots, p_n)$ that determines the outcome of the race, and thus bettors face a well defined risk as to which horse will win. That is, the winner of the race is truly determined by the spin of a roulette wheel governed by the probability distribution $(p_1, \ldots, p_n)$. However betters are *uncertain* about the distribution $(p_1, \ldots, p_n)$.

In order to capture this uncertainty, we introduce the state space $S = \hat{\Delta}^{n-1}$, where

$\hat{\Delta}^{n-1}$ denotes the set of probability distributions in the interior of Δ^{n-1} such that no two probabilities are equal, i.e., $p_i \neq p_j$ if $i \neq j$. The interpretation is that in each race, nature chooses a state in S, which is the true roulette wheel that determines the winning horse. The state is such that a horse always has a nonzero chance of winning, and no two horses are statistically identical.

After nature chooses a state and before the market opens, bettors receive private information concerning the state. We model this private information in the usual way by giving each bettor $t \in T$ an information partition I^t over S. Thus when nature chooses the state $s \in S$, bettor t cannot distinguish between the occurrence of s and the occurrence of any other element inside $I^t(s) \subset S$. Since our population is a continuum, we make the following reasonable assumption on the distribution of information.

Assumption 2.5 (Law of Large Numbers) *For any two distinct states s' and s, a non-measure zero mass of bettors can distinguish between the occurrence of s and s', i.e., $s' \neq s$ implies $\{t \in T : I^t(s) \neq I^t(s')\}$ has nonzero mass (w.r.t. $\mathbf{P}_V$).*

Thus in a hypothetical world of pooled information, where every bettor's private information I^t is public knowledge, each bettor exactly knows the state s, and hence knows nature's roulette wheel $(p_1, \ldots, p_n) \in \hat{\Delta}^{n-1}$.

2.5 Fully Revealing Rational Expectations Equilibrium

We now address the problem of price equilibrium in the odds market at horse racetracks (which recall is our metaphor for betting and prediction markets more generally). We have already shown in Section 2.2 that the odds market implicitly defines a "textbook" Arrow-Debreu securities market. Thus we seek equilibrium prices $(r_1, \ldots, r_n)$ for the n securities in the market. Since we have a continuum of consumers and thus price taking behavior on the part of bettors is appropriate, we seek a competitive price equilibrium. Moreover, since bettors have imperfect information, they may use prices in addition to their private information to help infer the state (since different states of nature may command different market prices even if the states cannot be distinguished by an individual bettor). The empirical regularity that the prices at racetrack are *sufficient* to econometrically uncover the underlying probability distribution over the horses suggest that the information revealing function of prices is perfect.

These observation taken together lead us to seek a fully revealing rational expectations equilibrium (REE) of the betting market model. We briefly review the fully revealing REE concept as it applies to our model, but for a more general treatment, see (Mas-Colell et al., 1995, ch. 17)). First, we hypothetically assume a world of pooled information in which all bettors know each other's private information. By the Law of Large Numbers assumption on the distribution of information, every bettor would perfectly know the state s in such a world. When every bettor perfectly knows the state, we show the unique existence of a usual competitive price equilibrium $(r_1^s, \ldots, r_n^s)$. We then consider the map $s \mapsto (r_1^s, \ldots, r_n^s)$ from states of nature to the pooled information price equilibrium. We show that this map is one to one, and thus constitutes a fully revealing REE. The idea is that while our world of pooled information was hypothetical, the fully revealing REE makes it a reality through the inverting of prices.

2.5.1 Pooled Information

Thus assume a world of pooled information, in which all bettors know each other's private information, and thus know the state of nature $(p_1, \ldots, p_n) \in \hat{\Delta}^{n-1}$ from which the winning horse is drawn. Together with the market odds $(R_1, \ldots, R_n)$, this gives rise to a menu of gambles $G = \{(R_1, p_1), \ldots, (R_n, p_n)\}$, which is common to all the bettors in T. We seek a vector of Arrow-Debreu prices $(r_1, \ldots, r_n) \in \Delta^{n-1}$ that is consistent with market clearing and utility maximization on the part of bettors. We can "plug-in" for the market clearing condition (1), and simplify the problem to seeking a vector of market shares $(s_1, \ldots, s_n) \in \Delta^{n-1}$ that is consistent with utility maximization. Thus a la (2), the set of gambles available in the market has the form $G = \{(R(s_i), p_i)\}_{i \in \{1,\ldots,n\}}$.

Given such a set G, the market share of the i^{th} horse as we have already examined, is given by

$$q_i(R(s_1), \ldots, R(s_n); p_1, \ldots, p_n).$$

Thus the market is in equilibrium when, for some market shares $(s_1^*, \ldots, s_n^*)$,

$$s_i^* = q_i(R(s_1^*), \ldots, R(s_n^*); p_1, \ldots, p_n) \quad \text{for} \quad i = 1 \ldots, n. \tag{4}$$

If such a vector of market shares exist, then by the market clearing condition (1), $(r_1^*, \ldots, r_n^*) = (s_1^*, \ldots, s_n^*)$ constitute an vector of equilibrium Arrow-Debreu prices.

Theorem 2.6 *In a world of pooled information, for any state of nature $(p_1, \ldots, p_n) \in \hat{\Delta}^{n-1}$, there exists unique equilibrium odds $(R(s_1^*), \ldots, R(s_n^*))$, with market shares $(s_1^*, \ldots, s_n^*) \in \hat{\Delta}^{n-1}$.*

2.5.2 Rational Expectations

Thus for any state of nature $(p_1, \ldots, p_n) \in \hat{\Delta}^{n-1}$, there exists a unique n-tuple of pooled information equilibrium odds of the form

$$(R_1^*, \ldots, R_n^*) = (R(s_1^*), \ldots, R(s_n^*)) \quad \text{for} \quad (s_1^*, \ldots, s_n^*) \in \hat{\Delta}^{n-1}. \tag{5}$$

Let us denote this mapping by $\mathbf{R}(p_1, \ldots, p_n)$. We now wish to show that this mapping is one to one, and hence constitutes a unique fully revealing REE. The problem boils down to the following : given any n-tuple of odds $(R_1^*, \ldots, R_n^*)$ of the form (5), can we uniquely recover the commonly known state $(p_1, \ldots, p_n)$ that supports the odds in equilibrium? That is, can we uniquely solve the system of equations in $(p_1, \ldots, p_n) \in \hat{\Delta}^{n-1}$,

$$s_i^* = q_i(R_1^*, \ldots, R_n^*; p_1, \ldots, p_n) \quad \text{for} \quad i = 1, \ldots, n. \tag{6}$$

Theorem 2.7 *For any n-tuple of odds $(R_1^*, \ldots, R_n^*)$ of the form* (5)*, there exists a unique probability distribution $(p_1^*, \ldots, p_n^*) \in \hat{\Delta}^{n-1}$, consisting of distinct and nonzero probabilities, that solves* (6)*.*

Thus the function $\mathbf{R}(p_1, \ldots, p_n)$ constitutes the unique fully revealing REE of the model. We express the inverse of the REE as $\mathbf{p}(R_1, \ldots, R_n)$, which maps any vector of observable and distinct odds $(R_1, \ldots, R_n)$ to the underlying state $(p_1, \ldots, p_n) \in \hat{\Delta}^{n-1}$. It is this inverse function that bettors in our model use to infer nature's roulette wheel from market prices. It is straightforward to show that this inverse pricing function satisfies the symmetry condition that for any $i = 1, \ldots, n$,

$$\mathbf{p}_i(R_i, R_{-i}) = \mathbf{p}_i(R_i, Q_{-i}), \tag{7}$$

where Q_{-i} is a permutation of the elements in R_{-i}.

3 The Empirical Strategy

The market equilibrium of our model naturally generates price variation across races. Races exogenously differ in the number of horses running n, and the underlying state of nature $(p_1, \ldots, p_n) \in \hat{\Delta}^{n-1}$. From this exogenous variation, the REE pricing function $\mathbf{R}$ generates observable odds $(R_1, \ldots, R_n)$ that vary across races. In each race, a spin of the roulette wheel $(p_1, \ldots, p_n)$ determines the index of the winning horse, which we denote by i_w.

Thus the REE $\mathbf{R}(p_1, \ldots, p_n)$ expresses the reduced form of our structural model - it is the mapping from the exogenously varying primitives (states of nature) to the endogenously determined equilibrium prices $(R_1, \ldots, R_n)$. Moreover this reduced form is invertible, with inverse $\mathbf{p}(R_1, \ldots, R_n)$. The essential link between our equilibrium theory and the data lies in the fact that the inverse reduced form can be recovered using the variation in standard racetrack data. Our data consist of a sample of races $k = 1, \ldots, K$. For each race k, we observe the number of horses running n^k, the vector of equilibrium odds $(R_1^k, \ldots, R_{n^k}^k)$, and the index of the winning horse i_w^k. This data thus nonparametrically pins down the inverse reduced form : $\mathbf{p}_i(R_1, \ldots R_n)$ is consistently estimated by the fraction of times horse i wins in the subset of races having prices $(R_1, \ldots, R_n)$.

We now address the question of what structural parameters of our model we can identify through knowledge of the reduced form. The unknown primitive recall is the distribution of preferences $\mathbf{P}_V$. Suppose we parameterize this distribution by assuming $V_t(R, p) = V(R, p, \theta_t)$ for $\theta_t \in \Theta \subset \mathbb{R}^n$. Such a parameterization may be suggested by a particular preference theory, such as expected utility theory, or rank dependent utility theory, etc. We then have an unknown continuous distribution F over the types $\theta \in \Theta$. We thus pose two natural questions.

1. Does knowledge of the reduced form $\mathbf{R}$ (or equivalently, its inverse $\mathbf{p}$) nonparametrically identify the distribution F? That is, do different F's generate different $\mathbf{p}$'s in our equilibrium model?

2. If the distribution over types F is identified, then how can we estimate it from our data?

We now address both of these questions for the case where the heterogeneity of risk preferences is presumed to be "one dimensional". While seemingly restrictive, a model of one dimensional heterogeneity is the common way that individual differences in risk preferences are conceived : people differ in a "willingness to take risk" attitude, and this difference in attitude results in differences in choice behavior. For example, Dohmen et al. (2005) find that survey questions asking people to rank themselves in terms of "willingness to take risk" on an 11-point scale turn out to be good predictors for actual behavior in subsequent choice experiments. Thus we can view a model of one dimensional heterogeneity as a good approximation. In expected utility modelling, the CARA and CRRA functional are both cases of one dimensional preferences. Moreover, since our horses are only differentiated in one dimension, namely quality (i.e. probability of winning), then it is natural to assume that bettors differ in the single dimension of "willingness to pay for quality", which in our setting translates into "willingness to take risk".

4 One Dimensional Heterogeneity

In this section, I present our model of betting market equilibrium for the special case that heterogeneity across bettors in preferences for risk can be reduced to a single dimensional type θ that orders individuals in terms of "willingness to take risk". I show that in this one dimensional model, the inverse reduce form $\mathbf{p}(R_1, \ldots, R_n)$ has a special recursive structure. Using this recursive structure, we can both nonparametrically identify the distribution F over types and estimate F by way of a numerically tractable maximum likelihood estimator.

4.1 One dimensional preferences

We now introduce more assumptions on the space of preferences and the distribution of preferences in addition to what was laid out in Section 2.3. This additional structure we shall refer to as the "one dimensional model".

Assumption 4.1 (The Space of Preferences) *We have a parameterized utility function over gambles*

$$V : \mathbb{R}_+ \times [0,1] \times \mathbb{R} \to \mathbb{R},$$

where $V(R, p, \theta)$ is the utility that a person of type θ receives from consuming the gamble

(R,p). V is continuous in (R,p,θ), and for each $\theta \in \mathbb{R}$, $V(R,p,\theta)$ is strictly increasing in p and R, and strictly minimized at $p=0$. The domain of θ is $\mathbb{R}$, although any interval valued domain is equally applicable.

Assumption 4.2 (The Distribution of Preferences) *The type θ is distributed according to a continuous CDF F that is strictly increasing over the support of θ, (and thus F admits a strictly increasing inverse CDF F^{-1}).*

Assumption 4.3 (Single Crossing) *The function $V(R,p,\theta)$ satisfies the single crossing condition that for any two gambles (R_1,p_1) and (R_2,p_2) with $p_1 > p_2$, and for some θ such that*

$$V(R_2,p_2,\theta) \geq V(R_1,p_1,\theta),$$

then $\theta' > \theta$ implies that

$$V(R_2,p_2,\theta') > V(R_1,p_1,\theta').$$

Now consider a menu of n gambles $G = \{(R_1,p_1),\ldots,(R_n,p_n)\}$, and let the subset of the population who chooses (R_i,p_i) from G be denoted

$$I_i = \left\{\theta \in \mathbb{R} : V(R_i,p_i,\theta) = \max_{j\in\{1,\ldots,n\}} V(R_j,p_j,\theta)\right\}.$$

Thus the share of the population who choose (R_i,p_i) from G is $q_i(G) =$

$$q_i(R_1,\ldots,R_n;p_1,\ldots,p_n) = \mathbf{P}_\theta(I_i). \tag{8}$$

Clearly, each I_i is a closed set.

Now suppose that $S = \{i_1,\ldots,i_m\} \subset \{1,\ldots,n\}$ indexes the gambles in G that receive non-zero market share, i.e., $i \in S$ iff $q_i(G) > 0$. It is fairly straightforward to show the following result, which is largely driven by the single crossing assumption.

Theorem 4.4 *There exists* $-\infty < \theta_1 < \cdots < \theta_{m-1} < \infty$ *such that*

$$I_{i_1} = (-\infty, \theta_1], I_{i_2} = [\theta_1, \theta_2], \ldots, I_{i_m} = [\theta_m, \infty)$$

and

$$V_{\theta_1}(R_{i_1}, p_{i_1}) = V_{\theta_1}(R_{i_2}, p_{i_2}), \ldots, V_{\theta_{m-1}}(R_{i_{m-1}}, p_{i_{m-1}}) = V_{\theta_{m-1}}(R_{i_m}, p_{i_m}) \tag{9}$$

where

$$R_{i_1} < \cdots < R_{i_m} \quad \textit{and} \quad p_{i_1} > \cdots > p_{i_m}.$$

As a simple corollary to the theorem, the market shares $q_i(G)$ for $i \in S$ can be expressed as

$$\begin{aligned} q_{i_1}(G) &= F(\theta_1) \\ q_{i_2}(G) &= F(\theta_2) - F(\theta_1) \\ &\vdots \\ q_{i_m}(G) &= 1 - F(\theta_{m-1}). \end{aligned}$$

Thus in summary, given a menu of gambles G, and the indices $S = \{i_1, \ldots, i_m\}$ of gambles receiving non-zero market share, the market shares $q_{i_j}(G)$ for $j = 1, \ldots, m$ are determined by the locations $\theta_1, \ldots, \theta_{m-1}$ of the *marginal bettors*, where the marginal bettor θ_j satisfies the *marginal condition* (9) of being indifferent between the gambles (R_{i_j}, p_{i_j}) and $(R_{i_{j+1}}, p_{i_{j+1}})$.

4.2 Equilibrium

Recall that by our market clearing conditions (1) and (2), the market share s_i alloted to each horse i determines its return R_i through

$$R_i = R(s_i) = \frac{1 - \tau}{s_i} \quad \text{where} \quad \sum_{i=1}^{n} s_i = 1. \tag{10}$$

In market equilibrium, the share alloted to each horse is such that

$$s_i = q_i(R(s_1), \ldots, R(s_n); p_1, \ldots, p_n) \quad \text{for} \quad i = 1, \ldots, n. \tag{11}$$

I now show if we observe the returns $(R_1, \ldots, R_n)$ in a betting market, which by (10) implies that we also observe the market shares $(s_1, \ldots, s_n)$, then we can recover the locations $\theta_1, \ldots, \theta_{n-1}$ of the marginal bettors. We can thus solve for the underlying state $(p_1, \ldots, p_n)$ that satisfies (11) by finding the probability distribution over horses that satisfies the marginal conditions (9). Let us see precisely how this works.

Step 1 : Recovering the Marginal Bettors

Suppose we observe the returns $(R_1, \ldots, R_n)$ in a market, which are ordered such that

$$R_1 < \cdots < R_n.$$

As a necessary condition for equilibrium, it must be that

$$p_1 > \cdots > p_n.$$

Since all the returns are finite, each horses receives a positive market share. Thus there exist marginal bettors $\theta_1, \ldots, \theta_{n-1}$ such that

$$I_1 = (-\infty, \theta_1], I_2 = [\theta_1, \theta_2], \ldots, I_n = [\theta_{n-1}, \infty).$$

and thus

$$
\begin{aligned}
s_1 &= F(\theta_1) \\
s_2 &= F(\theta_2) - F(\theta_1) \\
&\vdots \\
s_n &= 1 - F(\theta_{n-1}).
\end{aligned} \tag{12}
$$

Since the shares $s_1, \ldots, s_n$ are observable via (10), we can invert the system (12) to locate the marginal bettors $\theta_1, \ldots, \theta_{n-1}$, i.e.,

$$
\begin{aligned}
\theta_1 &= F^{-1}(s_1) \\
\theta_2 &= F^{-1}(s_1 + s_2) \\
&\vdots \\
\theta_{n-1} &= F^{-1}(s_1 + \cdots + s_{n-1}).
\end{aligned} \tag{13}
$$

Step 2 : Recovering the Probabilities

Having recovered the marginal bettors $\theta_1, \ldots, \theta_{n-1}$, we can find the probabilities

$$p_1 = \mathbf{p}_1(R_1, \ldots, R_n), \ldots, p_n = \mathbf{p}(R_1, \ldots, R_n)$$

that satisfy the equilibrium condition (11) by finding the probability distribution $(p_1, \ldots, p_n)$ that satisfies the marginal conditions (9). The solution to this problem admits a very convenient recursive structure, which we now reveal.

By our assumptions on V, for any choice of $p_1 \in [0, 1]$, there are unique values of $p_2^*(p_1), \ldots, p_n^*(p_1)$ that satisfy the marginal conditions (9). The functions $p_i^*(p_1)$ for $i = 2, \ldots, n$ are continuous and increasing. Our problem is to find p_1 such that

$$p_1 + p_2^*(p_1) + \cdots + p_n^*(p_1) = 1. \tag{14}$$

Notice that $p_1 = 0$ implies that $p_2^*(p_1) = 0, \ldots, p_n^*(p_1) = 0$ since for every θ, $V_\theta(R,p)$ is strictly minimized at $p = 0$. Likewise, it is clear that $p_1 = 1$ implies

$$p_1 + p_2^*(p_1) + \cdots + p_n^*(p_1) > 1.$$

Thus for a unique $p_1 \in (0,1)$ we have that (14) is satisfied.

Example : Expected Utility Theory

If bettor preferences follow expected utility theory, then not only do the $p_i^*(p_1)$ functions admit a closed form, but we can also analytically solve for p_1 in (14).

Suppose preferences in the population satisfy expected utility theory. Then

$$V_\theta(R,p) = pU_\theta(R) + (1-p)U_\theta(-1).$$

Now consider the marginal condition for the i^{th} marginal bettor θ_i, namely

$$\begin{aligned} & V_{\theta_i}(R_i, p_i) = V_{\theta_i}(R_{i+1}, p_{i+1}) \\ \Rightarrow \quad & p_i U_{\theta_i}(R_i) + (1-p_i)U_{\theta_i}(-1) = p_{i+1}U_{\theta_i}(R_{i+1}) + (1-p_{i+1})U_{\theta_i}(-1) \\ \Rightarrow \quad & p_{i+1} = \left(\frac{U_{\theta_i}(R_i) - U_{\theta_i}(-1)}{U_{\theta_i}(R_{i+1}) - U_{\theta_i}(-1)} \right) p_i. \end{aligned}$$

For $i = 1, \ldots, (n-1)$, define

$$c_i = \prod_{j=1}^{i} \frac{U_{\theta_j}(R_j) - U_{\theta_j}(-1)}{U_{\theta_j}(R_{j+1}) - U_{\theta_j}(-1)}.$$

Then we have that

$$
\begin{aligned}
p_2^*(p_1) &= c_1 p_1 \\
p_3^*(p_1) &= c_2 p_1 \\
&\vdots \\
p_n^*(p_1) &= c_{n-1} p_1.
\end{aligned}
$$

Thus (14) becomes

$$p_1 + c_1 p_1 + \cdots + c_{n-1} p_1 = 1,$$

which yields

$$p_1 = \frac{1}{1 + c_1 + \cdots + c_{n-1}}.$$

Thus for a given family of expected utility theoretic preferences V_θ, and a given CDF F over θ, and a given race $R_1 > \cdots > R_n$, we can analytically solve for the underlying state of nature $(p_1, \ldots, p_n) \in \hat{\Delta}^{n-1}$ in market equilibrium, i.e. analytically solve for the inverse REE of the model.

4.3 Nonparametric Identification

We now show that the one dimensional model nonparametrically identifies the underlying distribution of preferences F up to a tight equivalence class. Let us consider a fixed number of horses n. Define the similarity relation $\sim$ over the possible CDF's of the model as follows. For any two CDF's F and G over θ, we say $F \sim G$ iff $F^{-1}(x) = G^{-1}(x)$ for all $x \in (1/n, 1)$. Thus two CDF's are similar if all of their quantiles greater than $1/n$ are the same. Let the inverse reduced form of model under F and G be denoted $\mathbf{p}^F$ and $\mathbf{p}^G$ respectively. We now show that any two $\sim$ distinct CDF's generate different reduced form relationships between the prices and probabilities in a race.

Proposition 4.5 *If F and G are both continuous and strictly increasing CDF's over θ, then their respective reduced forms are different if and only if the CDF's are not $\sim$ similar.*

Proof We prove sufficiency. Necessity follows similarly. Suppose that F and G are not $\sim$ similar and thus $F^{-1}(x) \neq G^{-1}(x)$ for some $x > 1/n$. Then we can find a race with odds $(R(s_1), \ldots, R(s_n))$ where $s_1 = x_1$. Thus the first marginal bettor under F, θ_1^F, differs from the first marginal bettor under G, θ_1^G. Suppose $(p_1^F, \ldots, p_n^F)$ is the unique state that supports these odds in equilibrium under the distribution F, and thus p_2^F solves for p_2 in

$$V_{\theta_1^F}(R_1, p_1^F) = V_{\theta_1^F}(R_2, p_2).$$

Then by single crossing, it cannot also be the case that p_2^F also solves for p_2 in

$$V_{\theta_1^G}(R_1, p_1^F) = V_{\theta_1^G}(R_2, p_2).$$

Thus the inverse reduced form of the model under F, $\mathbf{p}^F(R_1, \ldots, R_n)$ cannot equal the inverse reduced under G, $\mathbf{p}^G(R_1, \ldots, R_n)$ (in particular they must differ at our chosen $(R(s_1), \ldots, R(s_n))$).

4.4 Estimation

Using (14), the solution of the underlying state of nature from the observed prices in a race becomes a one dimensional problem - searching for p_1. This can be accomplished numerically in a tractable way (using bisection for instance). For the case when the one dimensional preferences satisfy expected utility theory, then we have shown that the state can be solved for analytically. Whether numerically or analytically, the structure of the one dimensional model allows us to easily compute the inverse reduced form relationship of the model $\mathbf{p}(R_1, \ldots, R_n; F)$ given the CDF F. Using our sample of races $k = 1, \ldots, K$ (the notation for our data was given in Section 3), we can thus express the log-likelihood of any strictly increasing and continuous CDF F in the following way,

$$LL(F) = \sum_{k=1}^{K} \log(\mathbf{p}_{i_w^k}(R_1^k \ldots, R_{n^k}^k; F)).$$

A maximum likelihood estimator $\hat{F}$ of the true distribution of preferences F_0 maximizes the function $LL(F)$.

5 The Favorite Longshot Bias

Before we turn to the empirical analysis, we briefly explore the empirical content of the expected utility hypothesis as it applies to our data. What about the EUH is being tested by our empirical strategy? We now show that for the one dimensional model, the EUH makes a strong prediction about the favorite longshot bias, which is a key pattern present in the data.

The favorite-longshot bias (FLB) is an empirical relationship between the odds on a horse and the profitability from betting on a horse (see Jullien and Salanie (2002) for a history of its discovery). Equivalently, the FLB implies a relationship between the true probability of the horse winning and its Arrow-Debreu price : prices under-predict probabilities for favorites and over-predict probabilities for longshots. The FLB is important for our purposes because, as we show below, it forces the expected utility hypothesis to make a strong prediction as to the composition of bettors at the track. It is this prediction that helps us understand what the data are testing when we compare expected utility theory to non-expected utility theory.

Consider a race with state $(p_1, \ldots, p_n)$ and odds $(R_1, \ldots, R_n)$. Recall that we can re-express prices in terms of Arrow-Debreu prices $(r_1, \ldots, r_n)$ as shown in section 2.2. Ignoring track take (since it does not substantively change the analysis), the expected return from betting on horse i is :

$$
\begin{aligned}
ER_i &= p_i R_i + (1 - p_i)(-1) \\
&= \tfrac{p_i}{r_i} - 1.
\end{aligned}
\tag{15}
$$

Order horses from most expensive ("favorites") to cheapest ("longshots"), i.e., $R_1 < \cdots < R_n$. The *favorite-longshot bias* (FLB for short) is the empirical finding that ER_i is decreasing in i. In terms of (15), the FLB Implies that r_i underestimates p_i for market favorites, and overestimates p_i for longshots.

We now state the key lemma for understanding the FLB.

Lemma 5.1 *In a race, suppose $R_1 > \cdots > R_n$ and $p_1 < \cdots < p_n$, and consider the menu of gambles $G = \{(R_1, p_1), \ldots, (R_n, p_n)\}$. If ER_i for $i = 1, \ldots, n$ is constant, then $j > i$ implies the gamble (R_j, p_j) is related to (R_i, p_i) by a mean-preserving spread.*

Proof If $i < j$, the distribution function $F_i(x)$ of the gamble (R_i, p_i) lies strictly below the distribution function $F_j(x)$ of the gamble (R_j, p_j) for $x < R_i$, and lies strictly above for $x > R_i$. Thus since the distribution functions are single crossing and have the same mean, (R_j, p_j) is related to (R_i, p_i) by a mean-preserving spread. ∎

Remark As a simple corollary to the lemma, if $j > i$ and $ER_i \geq ER_j$, then a risk averter strictly prefers the gamble (R_i, p_i) to (R_j, p_j). Likewise, if $ER_j \leq ER_i$, then a risk lover strictly prefers the gamble (R_j, p_j) to (R_i, p_i). The intuition for the corollary is simple. If we start at the "right prices" $r_i = p_i$, and hence ER_i is constant across i, then by the lemma, $i < j$ implies that a risk averter strictly prefers a bet on horse i to horse j, and a risk lover strictly prefers a bet on horse j to horse i. If we shift up the mean of the favorite i, the risk averter continues to strictly prefer i. If we shift up the mean of the longshot j, the risk lover continues to strictly prefer j.

We now pose the following question. Let us assume the one dimensional model of betting market equilibrium. Suppose we know that the FLB holds true in a race. What does that teach us about the composition of bettors at the track? The following prediction is a key consequence of the expected utility hypothesis.

Theorem 5.2 *A necessary and sufficient condition for the FLB to arise in equilibrium is that all of the risk averters in the population bet on the top favorite. In particular, the fraction of risk averters in the population cannot be greater than or equal to p_1, the probability of winning for the top favorite.*

Proof Let us prove sufficiency. Suppose all of the risk averters bet on horse 1, the top favorite. Then all of the marginal bettors θ_i for $i = 1, \ldots, (n-1)$ are risk lovers. However, using the corollary, in order for each such θ_i to be indifferent between the gamble (R_i, p_i) and the gamble (R_{i+1}, p_{i+1}), it must be the case that $ER_i > ER_{i+1}$ for all i. But this is the FLB, and we have sufficiency.

Consider now necessity. Suppose not all of the risk averters bet on the favorite. Then we know that the first marginal bettor, θ_1, is a risk lover. However then it is not possible for the FLB to hold, since the FLB implies that $ER_1 > ER_2$. But by the corollary, θ_1 would then strictly prefer to bet on horse 1, and thus he cannot be the marginal bettor. Hence we have necessity.

In particular then, since the FLB requires $F(\theta_1) < p_1$ (since otherwise $ER_1 \leq ER_j$ for some $j > 1$ by (15)), it cannot be the case that the fraction of risk averters in the population is greater than p_1. ∎

Thus the EUH makes a strong prediction as to the composition of bettors at the track if the FLB is to hold. The EUH predicts that there is a chunk of risk averting "grandmas" who back favorites, and then everyone else (risk lovers) who sort across the remaining horses. This is the essential empirical content of the EUH in our model, for it limits the amount of risk aversion possible in the population if the race data that are characterized by the FLB (which they are). Since risk aversion in non-expected utility theories need not respect second order stochastic dominance, which is the key property used in the above proofs, these theories do not face the same upper bound on the fraction of risk averters. Thus one way to understand how the EUH can be tested is to see whether the amount of risk aversion allowable by EUH is a binding constraint. That is, we can estimate an expected utility model of preference heterogeneity, and in so far as expected utility is misspecified, then estimating a non-expected utility model should yield a greater percentage of risk aversion in the population and thereby improve the fit to the data.

6 Empirical Analysis

We now turn to estimating the distribution of risk preferences in the betting population. In order to do so by the methodology we have developed, we must assume a parametric form for $V(R, p, \theta)$. The two competing hypothesis concerning the nature of risk preferences described in Section 1, namely expected utility and probability weighting (i.e. RDEU/CPT), suggest distinct parametric strategies. Under expected utility, different types θ differ in their risk preferences due to differences in the curvature of their utility. We can capture such behavior through the representation $V(R, p, \theta) = pu_\theta(R) = pR^\theta$, i.e., a power utility specification, which is a case of constant relative risk averse (CRRA) preferences. Thus higher θ types are more risk loving, and thus the specification satisfies our single crossing condition.

On the other hand, probability weighting theories emphasize the nonlinear distorting of probability that occurs during decision making as the main explanatory device for risk attitudes. We can introduce a probability weighting function $G_\theta(p)$ for $p \in [0, 1]$ that is strictly increasing and satisfies $G(0) = 0$ and $G(1) = 1$, and specify $V(R, p, \theta) = G_\theta(p)u(R)$ (which corresponds to a rank dependent preference specification as explained in Jullien and

Salanie (2000)). Thus G distorts a gamble's probability of winning into a decision weight. We specify $G_\theta(p) = p^{1/\theta}$. Thus larger θ types have more concave G's, and thus overweight the probability of winning, i.e. they have more risk loving attitudes. Thus this specification also satisfies the single crossing specification. Here we see the duality between probability weighting and expected utility quite clearly : risk love under expected utility corresponds to convex u, whereas risk love under probability weighting corresponds to concave G.

There are a number ways we can flexibly specify the distribution F over types for each of the two models (recall that the distribution is nonparametrically identified within each model). Here we choose to illustrate the results using a simple two parameter specification for the inverse distribution function $F^{-1}(x;\beta)$, where β is a two dimensional parameter vector (our basic findings are robust to the complexity of the specification of F.). Let us choose an upper and lower bound for θ, denoted as θ_{min} and θ_{max} respectively, which we can be sure will not be binding (virtually no one in the population lives outside the bounds). Then we specify

$$F^{-1}(x;\beta_1,\beta_2) = \theta_{min} + (\theta_{max} - \theta_{min}) f(x;\beta_1,\beta_2),$$

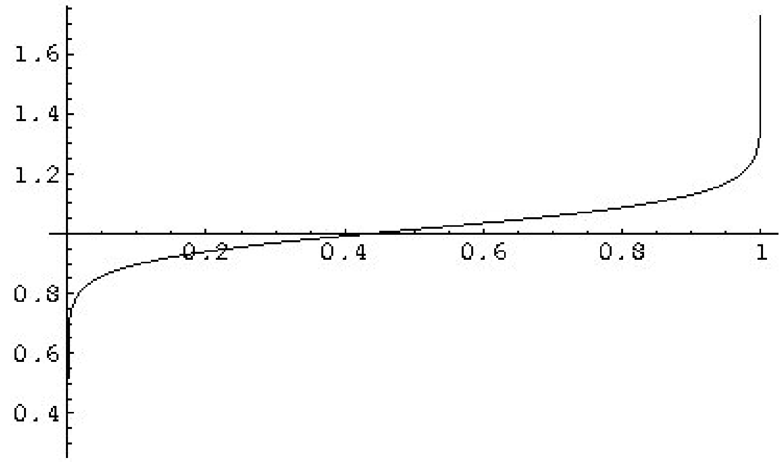
where $f(x;\beta_1,\beta_2)$ is a nondecreasing function from $[0,1]$ to $[0,1]$ with parameters β_1 and β_2. One such f is provided by Prelec (1998), which has the form

$$f(x;\beta_1,\beta_2) = \exp(-\beta_1(-ln(x))^{\beta_2}).$$

An advantage of this specification is that it allows us to nest the possibility that there is no heterogeneity in the population, which corresponds to $\beta_2 = 0$, in which case the population is a point mass at $\theta_{min} + (\theta_{max} - \theta_{min})\exp(-\beta_1)$.

Below we present the estimated distribution F_{EU}^{-1} for the expected utility model. As can be seen, close to 40 percent of the population is risk averse (with θ ranging from .8 to 1) and the remainder of the population is risk loving, with θ ranging from 1 to 1.6. Surprisingly,

Figure 1: The estimated inverse CDF F_{EU}^{-1}



there is no evidence for extreme risk loving in the population, which might be thought to be the true for racetrack bettors. Rather, there is evidence of a chunk of (mild) risk averters, who are the grandmas that bet on top favorites (the average market share for the top favorite in a race is 32 percent, with a standard deviation of .08 percent), and there is everyone else, who spread themselves over the remaining horses.

Let us now present the estimated distribution F_{PW}^{-1} for the probability weighting model. Here we find a strong result : the estimated values of β_2^{PW} is of the order $.1^{-4}$ and is not significantly different than 0. The estimated value of β_1 is such that the distribution F_{PW}^{-1} becomes a point mass over $\theta = 1$, and thus the entire population behaves linearly with respect to probability. We parameterized u in the probability weighting model to have a CARA functional form, and the estimated u is slightly risk loving with a CARA coefficient of -.01 (again, the finding is robust to the specification of u, which can also be estimated nonparametrically). Thus instead of seeing more risk averse behavior in the probability weighting model, which recall is our prediction if the expected utility hypothesis did not hold, we see the estimated probability weighting model collapsing to a representative expected utility maximizing risk loving agent. Moreover this representative expected utility maximizer is (as can be expected) empirically outperformed (both in terms of likelihood value and R^2) by the heterogeneous expected utility specification. Thus the data supports the explanation that risk attitudes amongst bettors are generated by differences in the curvature of utility, rather than the probability weighting explanation.

We present one more test of the adequacy of the expected utility hypothesis, and the CRRA functional form assumption in particular. Recall that our estimated distribution F_{EU}^{-1} implies an inverse reduced form $\mathbf{p}(R_1, \ldots, R_n; F^{EU})$ of the betting market model. We can also estimate a flexible functional form $\mathbf{p}(R_1, \ldots, R_n; \alpha)$ that predicts the probability of a horse's success given the vector of prices in a race, which depends on a vector of parameters α. The idea of the flexible reduced form is to capture the "true" relationship between probabilities and prices in the data without making any structural assumptions. The question is, how much of this true relationship can our structural model with CRRA preferences recover? There are a number of multinomial regression diagnostics that measure how well one model fares against another, but we focus here on the simplest, namely R^2 (where residuals are formed by taking the binary outcome of winning for each horse in each race minus its predicted probability). Using a specification developed in a follow up piece (Chiappori et al., 2006), we use

$$\mathbf{p}_i(R_1, \ldots, R_n) = \frac{e^{q_i}}{\sum_{j=1}^{n} e^{q_j}}$$

with, e.g.

$$q_i(R_1, \ldots, R_n) = \sum_{k=1}^{K} a_k(R_i, \alpha) T_k(R_{-i})$$

and the T_k's are symmetric sieve functions. In the same fashion as estimating the distribution

F in our structural model, we maximize over α in the log-likelihood

$$\sum_{k=1}^{K} \log \mathbf{p}_{i_w^k}(R_1^k, \ldots, R_{n^k}^k, \alpha).$$

Once again we find a very strong result. Virtually all (99 percent) of the R^2 from the flexible specification is recovered by our structural model with estimated preferences F_{EU}^{-1}. This is particularly striking since our structural estimate involved only two parameters (the parameters of the CDF), whereas the flexible specification has well over 30 parameters. Thus simple CRRA preferences alongside our Arrow-Debreu equilibrium theory appear to completely explain the relationship between prices and probabilities contained in the data. Said another way, if one wanted to forecast the probabilities of winning of the horses in a race using the market odds, one can hardly do better than use our structural model with CRRA heterogeneity to derive this forecast.

7 Conclusion

This paper is motivated by the widening gulf between the application of the expected utility hypothesis to modelling economic phenomena of empirical interest, and the existing empirical evidence concerning the validity of the expected utility hypothesis itself. The basic reason for this impasse is that experimental tests of economic hypotheses have long been viewed with skepticism by neoclassical economists. In his famous defense of the neoclassical approach to economics, Milton Friedman (Friedman, 1953) argued that the test of a hypothesis should not be based on its assumptions, but rather based on its predictions in the settings under which the hypothesis is proclaimed to work. As used by economists, the EUH describes the behavior of agents in real world markets, where opportunities for self selection, feedback from the environment, and learning are present, not the behavior of experimental subjects in the laboratory. Alongside equilibrium theory, the EUH makes predictions about the market outcomes (e.g., prices and quantities) that results from the interactions of these agents. A neoclassical examination of the EUH thus should be based on testing those predictions against market data, rather than directly testing the EU assumptions (e.g., the independence axiom) against experimental data.

In this paper, we take up the question of testing the expected utility hypothesis against real world market data. We have shown that the class of markets that have come to be known as "prediction markets", a traditional example being the odds market at racetracks, are a well suited source of data for such an analysis. In the first part of the paper, we developed a general equilibrium model with heterogeneous information, heterogeneous preferences, and rational expectations, to explain the relationship between the probabilities and prices of the uncertain events that define a prediction market. It was found that the key determinant to this relationship is the distribution of preferences. In the second part of the paper, we set out to estimate the distribution of preferences amongst bettors at American racetracks using odds data from a three year span at all American racetracks. We develop our empirical machinery (nonparametric identification and a maximum likelihood estimator) under the assumption of "one dimensional" preference heterogeneity, which when augmented by the expected utility hypothesis, explains the favorite longshot bias as the result of trade between risk averters (who bet on the top favorite in a race) and risk lovers (who spread themselves out over the remaining horses). Finally, we estimate the distribution of risk attitudes using both the expected utility approach and the probability weighting approaches to describing preferences for risk, and find strong evidence that the former rather than the latter is the more accurate description of bettor preferences. Future research aims at estimating our structural model under multidimensional preference heterogeneity, and estimating risk preferences in different market settings, using for example Tradesports contract data.

8 Appendix : Proof of Existence and Uniqueness of the Fully Revealing Rational Expectations Equilibrium

In order to prove Theorem 2.6 and Theorem 2.7, we first establish a series of intermediate results on the behavior of the market share functions given our assumptions on the preferences.

Lemma 8.1 *If $\mathbf{P}_V$ is continuous, then for any finite set of gambles G with at least one gamble having nonzero probability of winning,*

$$\mathbf{P}_V(S_i \cap S_j) = 0 \textit{ for every } i \neq j.$$

Proof If $p_i = 0$, then $S_i = \emptyset$ because each $V \in \mathbf{V}$ is strictly minimized at $p = 0$. Otherwise, $p_i > 0$, and for $i \neq j$, $S_i \cap S_j$ is a subset of

$$\{V \in \mathbf{V} : V(R_i, p_i) = V(R_j, p_j)\},$$

which by continuity has measure 0 under $\mathbf{P}_V$. ∎

Lemma 8.2 *If $\mathbf{P}_V$ is continuous, then for any finite set of distinct gambles G with at least one gamble in G having nonzero probability of winning, the sum of the market shares equals 1, i.e.,*

$$\sum_{i=1}^{n} q_i(R_1, \ldots, R_n; p_1, \ldots, p_n) = 1.$$

Proof Recall that

$$\sum_{i=1}^{n} q_i(R_1, \ldots, R_n; p_1, \ldots, p_n) = \sum_{i=1}^{n} \mathbf{P}_V(S_i).$$

Moreover, since G is finite, each $V \in \mathbf{V}$ attains a maximum over G, and thus

$$\bigcup_{i=1}^{n} S_i = \mathbf{V}.$$

However by Lemma 8.1, we have that for all $i \neq j$,

$$\mathbf{P}_V(S_i \cap S_j) = 0.$$

Thus

$$\sum_{i=1}^{n} \mathbf{P}_V(S_i) = \mathbf{P}_V\left(\bigcup_{i=1}^{n} S_i\right) = 1. \quad \blacksquare$$

Lemma 8.3 *If $\mathbf{P}_V$ is continuous, then for any n-tuple of distinct probabilities $(p_1, \ldots, p_n)$,*

$$q_i(R_1, \ldots, R_n; p_1, \ldots, p_n)$$

is a continuous function in $(R_1, \ldots, R_n) \in \mathbf{R}^n$. Furthermore, for any n-tuple of distinct returns $(R_1, \ldots, R_n)$,

$$q_i(R_1, \ldots, R_n; p_1, \ldots, p_n)$$

is a continuous function in $(p_1, \ldots, p_n) \in \Delta^{n-1}$.

Proof We prove the first part of the theorem (continuity in $(R_1, \ldots, R_n)$). The second part (continuity in $(p_1, \ldots, p_n)$) follows similarly to the first.

Let us fix any n-tuple of distinct probabilities $(p_1, \ldots, p_n)$. Now consider any n-tuple of returns $(R_1, \ldots, R_n)$. Define a function $F : \mathbf{V} \to \{0, 1\}$ as

$$F(V) = \prod_{j \neq i} \mathbf{1}\left[V(R_i, p_i) \geq V(R_j, p_j)\right],$$

where $\mathbf{1}(\cdot)$ is the indicator function. Then clearly

$$q_i(R_1, \ldots, R_n; p_1, \ldots, p_n) = \int F(V)\mathbf{P}_V(dV).$$

Now consider any sequence of n-tuples of returns $\{(R_1^t, \ldots, R_n^t)\}_{t\in\mathbb{N}_+}$ that converges to $(R_1, \ldots, R_n)$. For each $t \in \mathbb{N}_+$, define $F^t : \mathbf{V} \to \{0, 1\}$ as

$$F^t(V) = \prod_{j\neq i} \mathbf{1}\left[V(R_i^t, p_i) \geq V(R_j^t, p_j)\right],$$

and thus,

$$q_i(R_1^t, \ldots, R_n^t; p_1, \ldots, p_n) = \int F^t(V)\mathbf{P}_V(dV).$$

We need to establish that $q_i(R_1^t, \ldots, R_n^t; p_1, \ldots, p_n) \xrightarrow{t} q_i(R_1, \ldots, R_n; p_1, \ldots, p_n)$.

For every $V \in \mathbf{V} - S_i$ we have $F(V) = 0$. Thus for every $V \in \mathbf{V} - S_i$,

$$V(R_i, p_i) < V(R_j, p_j) \text{ for some } j \neq i.$$

By continuity of every utility function $V \in \mathbf{V}$, we have that for every $V \in \mathbf{V} - S_i$,

$$F^t(V) \xrightarrow{t} F(V).$$

On the other hand, for every $V \in S_i$, $F(V) = 1$. Since at least one $p_i > 0$, then by Lemma 8.1, for almost every $V \in S_i$,[9]

$$V(R_i, p_i) > V(R_j, p_j) \text{ for every } j \neq i.\text{[10]}$$

[9] The statement "for almost every $V \in S_i$" means "for all $V \in S_i$ except possibly in a subset $S \subset S_i$ with $\mathbf{P}_V(\mathrm{S}) = 0$".
[10] $S_i = \{V \in \mathbf{V} : V(R_i, p_i) > V(R_j, p_j) \text{ for every } j \neq i\} \cup_{j\neq i} (S_i \cap S_j)$.

Once again, by continuity of every utility function $V \in \mathbf{V}$, we have that for almost every $V \in S_i$,

$$F^t(V) \xrightarrow{t} F(V).$$

Thus for almost every $V \in \mathbf{V}$, $F^t(V)$ converges pointwise to $F(V)$. By Lebesgue's dominated convergence theorem,

$$q_i(R_1^t, \ldots, R_n^t; p_1, \ldots, p_n) \xrightarrow{t} q_i(R_1, \ldots, R_n; p_1, \ldots, p_n).$$

The proof of continuity in $(p_1, \ldots, p_n)$ follows similarly. The requirement in the theorem that $(p_1, \ldots, p_n)$ range over Δ^{n-1} is overly restrictive. it is only used to ensure that $(p_1, \ldots, p_n) \in \Delta^{n-1}$ implies at least one $p_i > 0$, which allows all the steps from continuity in returns to be repeated. ∎

We now state a definition.

Monotonicity For any n-tuple of distinct probabilities $(p_1, \ldots, p_n)$, any n-tuple of returns $(R_1, \ldots, R_n)$, and any strict subset $\mathbf{I} \subset \{1, \ldots, n\}$, consider a change to the returns from winning appearing in the choice set G that weakly increases the returns of the gambles indexed by $\mathbf{I}$ and weakly decreases the returns of the remaining gambles. This change leads to a new choice set $G^* = \{(R_i^*, p_i)\}_{i \in \{1, \ldots, n\}}$ with

$$R_i^* \geq R_i \text{ for } i \in \mathbf{I}$$

and

$$R_i^* \leq R_i \text{ for } i \notin \mathbf{I}.$$

We say that the distribution of consumer preferences $\mathbf{P}_V$ satisfies *monotonicity in return* if the sum of the shares of the gambles indexed by $\mathbf{I}$ weakly increase as a result of the change in returns, and the sum of the shares of the gambles indexed by $\{1, \ldots, n\} - \mathbf{I}$ weakly decrease as a result of the change in returns. That is

$$\sum_{i \in \mathbf{I}} q_i^* \geq \sum_{i \in \mathbf{I}} q_i$$

and

$$\sum_{i \notin \mathbf{I}} q_i^* \leq \sum_{i \notin \mathbf{I}} q_i.$$

Remark The definition of *monotonicity in probability* is stated similarly, except for any n-tuple of distinct returns $(R_1, \ldots, R_n)$, and any n-tuple of probabilities $(p_1, \ldots, p_n) \in \Delta^{n-1}$[11], we consider a weak increase of the probabilities of winning of the gambles for a strict subset of the gambles, and a weak decreases of the probabilities of winning for the remaining gambles, producing a new n-tuple of probabilities $(p_1^*, \ldots, p_n^*) \in \Delta^{n-1}$. The distribution $\mathbf{P}_V$ satisfies *monotonicity in probability* if such a change results in an increase in the of sum the shares of the gambles for which the probabilities increased, and a decrease in the sum of the shares of the gambles for which the probabilities decreased.

Lemma 8.4 *If $\mathbf{P}_V$ is continuous, then it satisfies monotonicity in return and monotonicity in probability.*

Proof We prove the theorem for monotonicity in returns. A similar argument follows for monotonicity in probability. Let $(p_1, \ldots, p_n)$, $(R_1, \ldots, R_n)$, and $(R_1^*, \ldots, R_n^*)$ be the n-tuples described in the definition of monotonicity. Similarly to the proof of Theorem 8.3, define

$$F_i(V) = \prod_{j \neq i} \mathbf{1}\left[V(R_i, p_i) \geq V(R_j, p_j)\right] \quad \text{and} \quad F_i^*(V) = \prod_{j \neq i} \mathbf{1}\left[V(R_i^*, p_i) \geq V(R_j^*, p_j)\right].$$

[11]We restrict the domain of n-tuple of probabilities to the simplex so as to ensure at least one probability is nonzero.

By the linearity of the integral operation

$$\sum_{i\in\mathbf{I}} q_i(R_1,\ldots,R_n;p_1,\ldots,p_n) = \int \sum_{i\in\mathbf{I}} F_i(V)\mathbf{P}_V(dV)$$

and

$$\sum_{i\in\mathbf{I}} q_i(R_1^*,\ldots,R_n^*;p_1,\ldots,p_n) = \int \sum_{i\in\mathbf{I}} F_i^*(V)\mathbf{P}_V(dV).$$

Since at least one $p_i > 0$, then by Lemma 8.1, for almost every $V \in \mathbf{V}$, $\sum_{i\in\mathbf{I}} F_i(V)$ equals 0 or 1.[12] However by the monotonicity of each $V \in \mathbf{V}$, $\sum_{i\in\mathbf{I}} F_i(V) = 1$ implies $\sum_{i\in\mathbf{I}} F_i^*(V) \geq 1$. Thus for almost every $V \in \mathbf{V}$,

$$\sum_{i\in\mathbf{I}} F_i^*(V) \geq \sum_{i\in\mathbf{I}} F_i(V)$$

and since the integral is an increasing linear operation,

$$\sum_{i\in\mathbf{I}} q_i(R_1^*,\ldots,R_n^*;p_1,\ldots,p_n) \geq \sum_{i\in\mathbf{I}} q_i(R_1,\ldots,R_n;p_1,\ldots,p_n). \quad \blacksquare$$

Lemma 8.5 *Consider any n-tuple of distinct, nonzero probabilities $(p_1,\ldots,p_n)$, and any subset $\mathbf{I} \subset \{1,\ldots,n\}$. If $\{(R_1^t,\ldots,R_n^t)\}_{t\in\mathbb{N}}$ is a sequence of n-tuples of returns with $\{R_i^t\}_{t\in\mathbb{N}}$ for $i \in \mathbf{I}$ nondecreasing and converging to ∞, and $\{R_i^t\}_{t\in\mathbb{N}}$ for $i \notin \mathbf{I}$ converging to $\bar{R}_i$, then then there exists a positive integer M such that for all $t > M$,*

$$\sum_{i\in\mathbf{I}} q_i^t > 0.$$

That is, at least one of the gambles indexed by $\mathbf{I}$ has a market share greater than 0 for all n-tuples of returns far along enough in the sequence.

[12] The set of V for which $\sum_{i\in\mathbf{I}} F_i(V) > 0$ equals $\cup_{i\neq j; i,j\in\mathbf{I}}(S_i \cap S_j)$

Proof Consider fixing $R_i = \bar{R}_i$ for $i \notin \mathbf{I}$, and let the returns for the gambles indexed by $i \in \mathbf{I}$ follow the sequence $\{R_i^t\}_{t \in \mathbb{N}}$. The resulting sequence of market shares, which we denote as $\bar{q}_i^t$ can be shown by desirability and monotonicity in return to satisfy

$$\lim_{t \to \infty} \sum_{i \in \mathbf{I}} \bar{q}_i^t > 0.^{13}$$

Thus there exists a positive integer N such that for all $t \geq N$,

$$\sum_{i \in \mathbf{I}} \bar{q}_i^t > 0.$$

In particular then,

$$\sum_{i \in \mathbf{I}} \bar{q}_i^N > 0.$$

Since the q_i functions are continuous in the n-tuple of returns, we can find an $\epsilon > 0$ such that $|\hat{R}_i - \bar{R}_i| < \epsilon$ for all $i \notin \mathbf{I}$ implies

$$\sum_{i \in \mathbf{I}} \hat{q}_i^N > 0.$$

By assumption we can find an N' such that $t > N'$ implies $|R_i^t - \bar{R}_i| < \epsilon$ for all $i \notin \mathbf{I}$. Taking $M = \max\{N, N'\}$ thus ensures that $t > M$ implies

$$\sum_{i \in \mathbf{I}} q_i^t > 0.^{14} \quad \blacksquare$$

8.1 Proof of Theorem 2.6

We prove the result in three steps. In the first step, we introduce an upper bound $\bar{R}$ on the odds payable by a gamble in the market, and show that an equilibrium exists under

[13] Once again we know by monitinicity in return that the limit exists.

[14] More precisely, $t > M$ implies $\sum_{i \in \mathbf{I}} q_i(R_{\mathbf{I}}^t, R_{-\mathbf{I}}^t) \geq \sum_{i \in \mathbf{I}} q_i(R_{\mathbf{I}}^N, R_{-\mathbf{I}}^t) > 0$

this restriction by Brouwer's fixed point theorem. This follows from continuity of $\mathbf{P}_V$ which drives the continuity of the q_i. In the second step, we show that it is always possible to raise the upper bound $\bar{R}$ high enough such that it is not binding in equilibrium, and thus an equilibrium of the form (4) exists. This result is driven by the desirability assumption. Lastly we show that the equilibrium is unique, which is driven by monotonicity in return (a consequence of continuity).

We shall assume there is an upper bound $\bar{R}$ on the net returns payable by a gamble. Under this "restriction" to the parimutuel mechanism, the the market shares s_i determine the market returns $\bar{R}_i$ through

$$\bar{R}_i(s_i) = \min\left(\frac{1-\tau}{s_i} - 1, \bar{R}\right).$$

Thus whenever $s_i \leq (1-\tau)/(1+\bar{R})$, the restriction $\bar{R}$ on the odds is binding.

Since the return vector $(\bar{R}(s_1), \ldots, \bar{R}(s_n))$ is clearly a continuous function of the market shares $(s_1, \ldots, s_n)$, and since the market share function q_i are continuous in returns by Theorem 8.3, we have $f : \Delta^{n-1} \to \Delta^{n-1}$ given by

$$f_i(s_1, \ldots, s_n) = q_i(\bar{R}(s_1), \ldots, \bar{R}(s_n); p_1, \ldots, p_n) \quad \text{for} \quad i = 1 \ldots, n,$$

is a continuous function. By the Brouwer fixed point theorem, the map f has a fixed point $(\bar{s}_1, \ldots, \bar{s}_n)$, which is thus an equilibrium of the restricted parimutuel market.

If the upper bound $\bar{R}$ is not binding for any of the $\bar{s}_i$, then clearly the fixed point satisfies the property (4) of being an equilibrium in the unrestricted parimutuel market. We now show that it is possible to raise the bar $\bar{R}$ sufficiently high so that it is not binding for the corresponding restricted equilibrium $(\bar{s}_1. \ldots, \bar{s}_n)$.

Suppose that this was not true. Then there exists a sequence of upper bounds $\{\bar{R}^t\}$ monotonically converging to ∞ with a corresponding sequence of equilibria $\{(\bar{s}_1^t, \ldots, \bar{s}_n^t)\} \subset$

Δ^{n-1} where for each t the upper bound $\bar{R}^t$ is binding for at least one $\bar{s}_i$. Since this sequence of market shares lives in a compact space, we can find a convergent subsequence $\{(\bar{s}_1^{t_k}, \ldots, \bar{s}_n^{t_k})\}$ converging to $(\bar{s}_1, \ldots, \bar{s}_n)$, with $\bar{s}_i = 0$ for at least one i (which follows from the fact that the restriction is binding for each t_k).

Let $\mathbf{I} \subset \{1, \ldots, n\}$ be the strict subset of indices i for which $\bar{s}_i = 0$. Then for each $i \in \mathbf{I}$, the sequence $\{\bar{R}(\bar{s}_i^{t_k})\}$ converges to ∞,[15] and without loss of generality we can say it converges to ∞ monotonically.[16]

However by desirability, this situation is not possible. It would be mean that there is a sequence of n-tuples of returns $\{(R_1^m, \ldots, R_n^m\}$ with $\{R_i^m\}$ monotonically converging to ∞ for $i \in \mathbf{I}$ and $\{R_i^m\}$ converging to finite R_i for $i \notin \mathbf{I}$, and

$$\lim_{m\to\infty} \sum_{i\in\mathbf{I}} q_i(R_1^m, \ldots, R_n^m; p_1, \ldots, p_n) = 0,$$

which contradicts lemma 8.5. Thus it must be the case that we can find a large enough upper bound $\bar{R}$ such that $\bar{R}$ is not binding for the equilibrium $(\bar{s}_1, \ldots, \bar{s}_n)$. These market shares thus satisfy the condition for $(s_1^*, \ldots, s_n^*)$ in (4). Moreover, these equilibrium market shares must also be located in the interior of Δ^{n-1} (because $\bar{R}$ is non-binding), i.e., $\bar{s}_i > 0$ for all i. It also follows that since the probability distribution $(p_1, \ldots, p_n)$ involved distinct probabilities, the equilibrium market shares $(\bar{s}_1, \ldots, \bar{s}_n)$ must be distinct, since otherwise one gamble in the market would dominate another, thereby causing the latter to have zero market share, which would contradict the fact the equilibrium shares lies in the interior of the simplex.

We now address uniqueness. Suppose that there exist two n-tuples of market shares

[15]Since $\min\left(\frac{1-\tau}{s_i^{t_k}} - 1, \bar{R}^{t_k}\right)$ goes to ∞.

[16]We can always take a subsequence to assure monotonic convergence

$(\bar{s}_1, \ldots, \bar{s}_n)$ and $(s_1^*, \ldots, s_n^*)$ that satisfy the equilibrium condition (4). Then for $i = 1, \ldots, n$,[17]

$$\bar{s}_i = q_i(R(\bar{s}_1), \ldots, R(\bar{s}_n)) \quad \text{and} \quad s_i^* = q_i(R(s_1^*), \ldots, R(s_n^*)).$$

Since both equilibrium tuples are located in the simplex, it must be the case that for some nonempty strict subset $\mathbf{I} \subset \{1, \ldots, n\}$, $s_i^* \leq \bar{s}_i$ for all $i \in \mathbf{I}$ with a strict inequality for at least one $i \in \mathbf{I}$, and and $s_i^* \geq \bar{s}_i$ for all $i \notin \mathbf{I}$ with a strict inequality for at least one $i \notin \mathbf{I}$. This implies that

$$\sum_{i \in \mathbf{I}} s_i^* < \sum_{i \in \mathbf{I}} \bar{s}_i,$$

However we also have that $R(s_i^*) \geq R(\bar{s}_i)$ for $i \in \mathbf{I}$ and $R(s_i) \leq R(\bar{s}_i)$ for $i \notin \mathbf{I}$, which by monotonicity in return of $\mathbf{P}_V$, implies that

$$\sum_{i \in \mathbf{I}} s_i^* \geq \sum_{i \in \mathbf{I}} \bar{s}_i.$$

This is a contradiction, and thus the equilibrium is unique. ∎

8.2 Proof of Theorem 2.7

Since the R_i^* are assumed distinct, and $\mathbf{P}_V$ satisfies continuity, then we that for any $(p_1, \ldots, p_n) \in \Delta^{n-1}$,

$$\sum_{i=1}^{n} s_i^* - q_i(R_1^*, \ldots, R_n^*; p_1, \ldots, p_n) = 0.^{18} \tag{16}$$

Furthermore, by continuity once again, $q_i(R_1^*, \ldots, R_n^*; p_1, \ldots, p_n)$ is continuous in $(p_1, \ldots, p_n)$ over Δ^{n-1}. Now consider the following continuous self map over Δ^{n-1} (where for simplicity

[17] We suppress the probabilities $(p_1, \ldots, p_n)$ in the following notation because they remain the same.
[18] Note that $\sum s_i^* = 1$ because the s_i are (observed) market shares.

we write $R^* = (R_1^*, \dots, R_n^*))$. For $(p_1, \dots, p_n) \in \Delta^{n-1}$,

$$p_i \mapsto \frac{p_i + \max(0, s_i^* - q_i(R^*; p_1, \dots, p_n))}{\sum_{j=1}^n (p_j + \max(0, s_j^* - q_j(R^*; p_1, \dots, p_n)))} \quad \text{for} \quad i = 1, \dots, n. \tag{17}$$

By the Brouwer fixed point theorem, this map must have a fixed point $(p_1^*, \dots, p_n^*) \in \Delta^{n-1}$.

Moreover, this fixed point must satisfy $q_i(R^*, p_1^*, \dots, p_n^*) = s_i^*$ for $i = 1, \dots, n$. If these are equalities are not satisfied, then by (16) we have that for at least one i we have $s_i^* > q_i(R^*, p_1^*, \dots, p_n^*)$, and for at least one j we have $s_j^* < q_j(R^*, p_1^*, \dots, p_n^*)$. Thus

$$\sum_{i=1}^n p_i^* + \max(0, s_i^* - q_i(R^*, p_1^*, \dots, p_n^*)) > 1,$$

and under the the mapping (17), p_j^* must get sent to a strictly smaller number, which violates the fact that $(p_1^*, \dots, p_n^*)$ is a fixed point. Thus $(p_1^*, \dots, p_n^*)$ solves (6)).

Since each s_i^* is nonzero and distinct, it must be the case that each p_i^* is nonzero and distinct. Otherwise some gamble, indexed by i say, would be dominated (either it has zero probability or it has the same probability as another gamble but lower return), and thus $q_i(R^*, p_1^*, \dots, p_n^*) = 0$, which is inconsistent with the fact the market share of the i^{th} gamble is $s_i^* > 0$.

The uniqueness of $(p_1^*, \dots, p_n^*)$ follows from monotonicity in a manner parallel to that used in proving the uniqueness in Theorem 2.6, except exploiting monotonicity in probability instead of monotonicity in return (both of which recall follows from continuity). ∎

References

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–90, July 1995.

Timothy F Bresnahan. Competition and collusion in the american automobile industry: The 1955 price war. *Journal of Industrial Economics*, 35(4):457–82, June 1987.

Colin Camerer. Prospect theory in the wild. In Daniel Kahneman and Amos Tversky, editors, *Choices, Values, and Frames*. Cambridge University Press, 2000.

Pierre Andre Chiappori. Relative risk aversion is constant : Evidence from panel data. 2006.

Pierre-Andre Chiappori, Amit Gandhi, Bernard Salanie, and Francois Salanie. Nonparametric identification of risk preferences. 2006.

Thomas Dohmen, Armin Falk, David Huffman, Uwe Sunde, Jrgen Schupp, and Gert G. Wagner. Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey. IZA Discussion Papers 1730, Institute for the Study of Labor (IZA), September 2005.

Stephen Figlewski. Subjective information and market efficiency in a betting market. *Journal of Political Economy*, 87(1):75–88, 1979.

P.C. Fishburn. *Nonlinear Preference and Utility Theory*. Johns Hopkins University Press, Baltimore MD, 1988.

Milton Friedman. The methodology of positive economics. In *Essays on Positive Economics*. University of Chicago Press, 1953.

Richard Gonzalez and George Wu. On the shape of the probability weighting function,. *Cognitive Psychology*, 38(1):129–166, February 1999.

R.M. Griffith. Odds adjustment by american horse-race bettors. *American Journal of Psychology*, 62(2):290–294, 1949.

Bruno Jullien and Bernard Salanie. Estimating preferences under risk: The case of racetrack bettors. *Journal of Political Economy*, 108(3):503–530, 2000.

Bruno Jullien and Bernard Salanie. Empirical evidence on the preferences of racetrack bettors. Technical report, 2002. mimeo.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decisions under risk. *Econometrica*, 47:262–291, 1979.

Edi Karni and Zvi Safra. "preference reversal' and the observability of preferences by experimental methods. *Econometrica*, 55(3):675–85, May 1987.

E. Han Kim, Adair Morse, and Luigi Zingales. What has mattered to economics since 1970. 2006.

Charles F. Manski. Interpreting the predictions of prediction markets. *Economics Letters*, 91(3):425–429, June 2006.

Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

Charles R. Plott, Jorgen Wit, and Winston C. Yang. Parimutuel betting markets as information aggregation devices: experimental results. *Economic Theory*, 22(2):311–351, 09 2003.

Drazen Prelec. The probability weighting function. *Econometrica*, 66(3):497–528, May 1998.

Richard E Quandt. Betting and equilibrium. *The Quarterly Journal of Economics*, 101(1): 201–07, 1986.

J. Quiggin. A theory of anticipated utility. *Journal of Economic Behaviour and Organization*, 3:323–343, 1982.

Matthew Rabin and Richard H. Thaler. Anomalies: Risk aversion. *Journal of Economic Perspectives*, 15(1):219–232, 2001.

Raymond D. Sauer. The economics of wagering markets. *Journal of Economic Literature*, 36(4):2021–2064, December 1998.

Uzi Segal. Some remarks on quiggin's anticipated utility. *Journal of Economic Behavior & Organization*, 8(1):145–154, March 1987.

Avner Shaked and John Sutton. Relaxing price competition through product differentiation. *Review of Economic Studies*, 49(1):3–13, January 1982.

Eric Snowberg and Justin Wolfers. Explaining the favorite longshot bias : Is it risk-love or misperceptions. Technical report, 2005. mimeo.

Chris Starmer. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, 38(2):332–382, 2000.

Richard H. Thaler and William T. Ziemba. Anomalies: Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives*, 2(2):161–174, 1988.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

Martin Weitzman. Utility analysis and group behavior: An empirical study. *Journal of Political Economy*, 73(1):18–26, 1965.

Justin Wolfers and Eric Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, Spring 2004.

Justin Wolfers and Eric Zitzewitz. Interpreting prediction market prices as probabilities. NBER Working Papers 12200, National Bureau of Economic Research, Inc, May 2006.

M.E. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.