Econometrica Supplementary Material

SUPPLEMENT TO "FROM AGGREGATE BETTING DATA TO INDIVIDUAL RISK PREFERENCES"

(*Econometrica*, Vol. 87, No. 1, January 2019, 1–36)

PIERRE-ANDRÉ CHIAPPORI
Department of Economics, Columbia University

BERNARD SALANIÉ
Department of Economics, Columbia University

FRANÇOIS SALANIÉ
Toulouse School of Economics, University of Toulouse Capitole and INRA

AMIT GANDHI
Department of Economics, University of Pennsylvania

APPENDIX A: EXHAUSTIVENESS AND IDENTIFICATION: THE CASE OF EXPECTED UTILITY

THE ANALYSIS IN SECTION 2 can be specialized to the case when V is a family of expected utility functionals. Normalizing again the utility of losing a bet to be zero:

$$V(R, p, \theta) = pu(R, \theta), \tag{A.1}$$

the indifference condition in Definition 1 becomes

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

and a new testable implication obtains:

PROPOSITION A.1: *If V is of the expected utility form* (A.1), *then $G(R, p, R', \theta)$ is linear in p.*

Reciprocally, if (i)–(iv) hold in Proposition 3, then linearity in p implies

$$G\big(R, p, R', \theta\big) = pg\big(R, R', \theta\big).$$

From (ii), we obtain

$$\frac{\partial}{\partial R'}\left(\frac{g_R}{g}\right) = 0,$$

or equivalently,

$$g\big(R, R', \theta\big) = A(R, \theta)u\big(R', \theta\big)$$

Pierre-André Chiappori: pc2167@columbia.edu
Bernard Salanié: bsalanie@columbia.edu
François Salanié: francois.salanie@inra.fr
Amit Gandhi: akgandhi@sas.upenn.edu

 https://doi.org/10.3982/ECTA11165

for some functions A and u. From (iii), $A(R,\theta)u(R,\theta)=1$; therefore,

$$G(R,p,R',\theta)=p\frac{u(R,\theta)}{u(R',\theta)},$$

and thus $p'u(R',\theta)=pu(R,\theta)$, so that u is the required von Neumann–Morgenstern utility function. Once more, this function is uniquely identified (up to a multiplicative constant) from the knowledge of G.

In fact, say that we normalize $u(R_m,\theta)\equiv 1$ for some R_m. Then it is easy to see that the whole family of vNM functions can be recovered by the simple (but not very practical) formula

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

APPENDIX B: ENDOGENOUS PARTICIPATION

We here provide some identification results for the case where individuals belonging to the same population of *potential* bettors decide whether or not to participate in a given race. This decision may involve horses' odds and probabilities of winning, but also psychological factors such as the thrill of gambling.[1] Recall that the space of possible preferences V is indexed by a parameter $\theta\in\Theta\equiv[0,1]$; we now let the uniform distribution characterize the population of *potential*, not *actual*, bettors. Suppose that a race is organized, and that potential bettors know the horses' probabilities $\mathbf{p}$ and odds $\mathbf{R}$. We represent participation decisions in a very flexible way: they result in a distribution of actual bettors characterized by some cumulative density function $F(\cdot|\mathbf{R},\mathbf{p})$ defined on Θ. Assume that F is continuous and strictly increasing with θ.

While F results from the demand side only, the odds $\mathbf{R}$ are equilibrium prices that are endogenously determined. Proceeding as in the case with fixed participation, the market shares $s_1,\dots,s_n$ can still be recovered from the odds using the properties of the parimutuel system:

$$\frac{1}{1-t}=\sum_j\frac{1}{R_j+1}\quad\text{and}\quad s_i(\mathbf{R})=\frac{1-t}{R_i+1}. \tag{B.1}$$

Hence, market shares only depend on the family of odds $\mathbf{R}$, and we can also define the cumulative market share

$$S_i(\mathbf{R})=\sum_{j\le i}s_j(\mathbf{R}).$$

Under single-crossing, bettors can be partitioned into intervals of types, exactly as in Lemma 2. Types in $[\theta_{i-1},\theta_i]$ bet on horse i, and the indifference condition still holds:

$$V(R_i,p_i,\theta_i)=V(R_{i+1},p_{i+1},\theta_i),\quad i=1,\dots,n-1.$$

The single-crossing assumption allows us to rewrite this condition as

$$\theta_i=H(R_i,p_i,R_{i+1},p_{i+1}),\quad i=1,\dots,n-1,$$

[1] See Conlisk (1993).

where $H(R, p, R', p')$ is increasing with its first two arguments and decreasing with its last two arguments.

The only change introduced by endogenous participation is that the marginal type is no longer equal to the market share. Instead, we have

$$S_i(\mathbf{R}) = F\big(H(R_i, p_i, R_{i+1}, p_{i+1})|\mathbf{R}, \mathbf{p}\big), \quad i = 1, \dots, n-1. \tag{B.2}$$

Given this definition of equilibrium odds, our next task is to ensure that one can uniquely recover probabilities from odds. This requires an assumption on how participation depends on both odds and probabilities:

ASSUMPTION B.1: *Consider two races* $(\mathbf{R}, \mathbf{p})$ *and* $(\mathbf{R}, \mathbf{p}')$ *with the same odds, and a horse* $i = 1, \dots, (n-1)$ *such that*

$$p_{i+1} > p'_{i+1} \quad \textit{and} \quad \forall\, j \le i, \quad p_j \le p'_j.$$

If θ *is the highest type that bets on a horse* $j \le i$ *in race* $(\mathbf{R}, \mathbf{p})$, *then*

$$F(\theta|\mathbf{R}, \mathbf{p}) \le F\big(\theta|\mathbf{R}, \mathbf{p}'\big).$$

This assumption requires that there are proportionally more low types that decide to bet in a race in which favorites have higher probabilities of winning. This is in the spirit of our single-crossing condition on *actual* bettors, which states that lower types are more likely to be attracted by favorites. Assumption B.1 is powerful enough to imply the following result:

LEMMA B.1: *Under Assumptions* 1, 2, 3, *and* B.1, *to each family of odds* $\mathbf{R}$, *one can associate a unique race* $(\mathbf{p}, t)$.

PROOF: Note that the take t is uniquely defined by the odds $\mathbf{R}$, thanks to (B.1). Now let us proceed by contradiction. Consider two races $(\mathbf{R}, \mathbf{p})$ and $(\mathbf{R}, \mathbf{p}')$ with the same odds, but different probabilities. Without loss of generality, assume $p_1 \le p'_1$. Since $\mathbf{p} \ne \mathbf{p}'$, there must exist a horse $i < n$ such that the first condition in Assumption A holds. The type θ defined in the second condition is $H(R_i, p_i, R_{i+1}, p_{i+1})$, so we get

$$F\big(H(R_i, p_i, R_{i+1}, p_{i+1})\big)|\mathbf{R}, \mathbf{p}) \le F\big(H(R_i, p_i, R_{i+1}, p_{i+1})|\mathbf{R}, \mathbf{p}'\big).$$

Also, because H is increasing with p_i and decreasing with p_{i+1}, we have

$$H(R_i, p_i, R_{i+1}, p_{i+1}) < H\big(R_i, p'_i, R_{i+1}, p'_{i+1}\big),$$

and therefore

$$F\big(H(R_i, p_i, R_{i+1}, p_{i+1})\big)|\mathbf{R}, \mathbf{p}) < F\big(H\big(R_i, p'_i, R_{i+1}, p'_{i+1}\big)|\mathbf{R}, \mathbf{p}'\big),$$

or equivalently, $S_i(\mathbf{R}) < S_i(\mathbf{R})$ from (B.2), a contradiction. *Q.E.D.*

This result ensures that theoretically at least, we can estimate $\mathbf{p}(\mathbf{R})$ from a large enough database. This allows us to abuse notation slightly, by setting $F(\cdot|\mathbf{R}) \equiv F(\cdot|\mathbf{R}, \mathbf{p}(\mathbf{R}))$. Our equilibrium conditions then reduce to

$$S_i(\mathbf{R}) = F\big(H\big(R_i, p_i(\mathbf{R}), R_{i+1}, p_{i+1}(\mathbf{R})\big)|\mathbf{R}\big), \quad i = 1, \dots, n-1.$$

Now we want to recover information about both preferences (as captured by H) and about participation (in F) from a database of odds $\mathbf{R}$, for which the functions $\mathbf{p}(\mathbf{R})$ and $\mathbf{S}(\mathbf{R})$ are known. Unfortunately, one can interpret the above equilibrium conditions as defining $F^{-1}(\cdot|\mathbf{R})$, given any function H. If, for each race, the H term is increasing with the index i of the horse, then F verifies all relevant properties of a c.d.f., and one cannot identify H, nor F. Therefore, we have to restrict the class of F we allow for. Let us assume, for instance, that the following holds:

ASSUMPTION B.2: *There is a known function* $A(\theta|\alpha)$, *monotonic in a scalar* α, *such that, for all* θ *and* $\mathbf{R}$,

$$F(\theta|\mathbf{R}) = A\big(\theta|\alpha(\mathbf{R})\big).$$

Then the equilibrium conditions become

$$S_i(\mathbf{R}) = A\big(H\big(R_i, p_i(\mathbf{R}), R_{i+1}, p_{i+1}(\mathbf{R})\big)|\alpha(\mathbf{R})\big), \quad i = 1, \dots, n-1. \tag{B.3}$$

Now let x stand for a 4-tuple (R, p, R', p'). Choose a particular x_0, and fix the value $H(x_0)$. For all races in which a horse displays the characteristics x_0, we know the values of H and S for this horse, and since A is monotonic with respect to α, we can deduce the value of α for every such race. From this, using the same equilibrium condition, we deduce the values of H for all other horses in these races; and we can iterate the process. Under the (weak) assumption that this process ends up encompassing all horses, we thus have identified H and A, up to the constant $A(x_0)$. Therefore, the following holds:

PROPOSITION B.2: *Under Assumptions* 1, 2, 3, *and* B.2, *the functions* H *and* α *are identified up to a constant*.

Hence, our setting can accommodate varying participation. The simplest illustration of Proposition B.2 has bettor θ participate in a race with odds $\mathbf{R}$ if and only if θ is above a threshold $\alpha(\mathbf{R})$. In this case, $A(\theta|\alpha) = \theta/(1-\alpha)\mathbb{1}(\theta \geq \alpha)$ and the equilibrium conditions reduce to

$$\frac{1 - S_{i+1}}{1 - S_i} = \frac{1 - H_{i+1}}{1 - H_i}, \quad i = 1, \dots, n-2 \tag{B.4}$$

and

$$\alpha(\mathbf{R}) = \frac{H\big(R_1, p_1(\mathbf{R}), R_2, p_2(\mathbf{R})\big) - S_1(\mathbf{R})}{1 - S_1(\mathbf{R})}.$$

The first series of equations identifies the function H, up to a positive constant: if H is a solution, then, for any strictly positive constant γ, the function $1 - \gamma(1 - H)$ is also a solution. Note that the preferences V are still uniquely identified, since the marginal rate of substitution $V_{\mathbf{R}}/V_{\mathbf{p}} = H_{\mathbf{R}}/H_{\mathbf{p}}$ does not depend on γ. The last equation identifies the threshold $\alpha(\mathbf{R})$, once more up to the same γ. It is easy to see that relative participations across races, as measured by the ratio

$$\frac{1 - \alpha(\mathbf{R})}{1 - \alpha\big(\mathbf{R}'\big)},$$

are fully identified.

Under Assumption B.2, it would be natural to use the data to test intuitions about the determinants of the threshold α. We will not pursue this issue further.

APPENDIX C: A PROOF OF THE IDENTIFICATION RESULT AT THE END OF SECTION 2.3

First, notice that under first-order stochastic dominance, we can normalize V so that, for a given value R_M of odds, $V(R_M, p, \theta, \eta) = p$ for all p, θ, and η (this is just an extension of the normalization we use in our estimations; see Section 4). For the particular representation in (10), this yields $a(R_M, p) = b(R_M, p) = 0$. Fix a probability value p_M. At $(R_{i+1}, p_{i+1}) = (R_M, p_M)$, equation (11) yields a unique function a such that

$$a(R, p) = 2\big[\log(p/p_M) - b(R, p)\Phi(R, p, R_M, p_M)\big]. \tag{C.1}$$

In particular,

$$\begin{aligned} &a(R', p') - a(R, p) \\ &\quad = 2\left(\log\frac{p}{p'} - b(R, p)\Phi(R, p, R_M, p_M) + b(R', p')\Phi(R', p', R_M, p_M)\right), \end{aligned}$$

and substituting in (11) gives

$$\Phi(R, p, R', p') = \frac{b(R', p')\Phi(R', p', R_M, p_M) - b(R, p)\Phi(R, p, R_M, p_M)}{b(R', p') - b(R, p)}.$$

This can be rewritten as

$$b(R, p) = b(R', p')\frac{\Phi(R, p, R', p') - \Phi(R', p', R_M, p_M)}{\Phi(R, p, R', p') - \Phi(R, p, R_M, p_M)},$$

which identifies b up to a multiplicative constant. Take an arbitrary (R_1, p_1), and denote $b_1 = b(R_1, p_1)$. Then $b(R, p) = b_1\Omega(R, p)$, where

$$\Omega(R, p) \equiv \frac{\Phi(R, p, R_1, p_1) - \Phi(R_1, p_1, R_M, p_M)}{\Phi(R, p, R_1, p_1) - \Phi(R, p, R_M, p_M)}$$

is identified. Plugging this into (C.1) gives us

$$a(R, p) = 2\big[\log(p/p_M) - b_1\Omega(R, p)\Phi(R, p, R_M, p_M)\big],$$

and our normalization $a(R_M, p) \equiv 0$ implies

$$\log p = \log p_M + b_1\Omega(R_M, p)\Phi(R_M, p, R_M, p_M)$$

for all p. Substituting into the definition of V in (10) gives us

$$\log V = \log p + 2\theta\big[\log(p/p_M) - b_1\Omega(R, p)\Phi(R, p, R_M, p_M)\big] + \eta b_1\Omega(R, p)$$

and we can replace the terms in $\log p$ to get

$$\begin{aligned} \log V = {} & \log p_M + b_1\big[(2\theta + 1)\Omega(R_M, p)\Phi(R_M, p, R_M, p_M) \\ & - 2\theta\Omega(R, p)\Phi(R, p, R_M, p_M) + \eta\Omega(R, p)\big]. \end{aligned}$$

In this formula for V, the bracketed term is identified, and it fully determines the value of the marginal rate of substitution V_p/V_R.

Therefore, the mapping from (θ, η) to the marginal rate of substitution V_p/V_R is identified from the data.

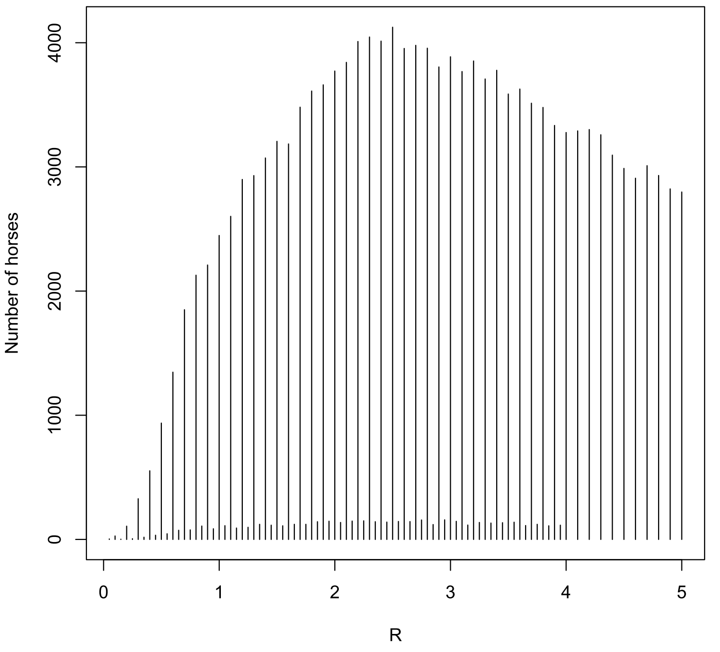

FIGURE S1.—Distribution of odds, $R \leq 5$.

APPENDIX D: ROUNDING

Odds are rounded on the track. We illustrate the rounding mechanism in Figure S1: for odds below 4, odds are sometimes quoted in twentieths, but tenths are much more likely (e.g., $R = 2.1$ and $R = 2.2$ are much more likely than $R = 2.15$). For longer odds, the spacing of odds becomes coarser, but the same pattern still obtains (e.g., $R = 26.25$ is much less likely than $R = 26.0$ or $R = 26.5$).

APPENDIX E: VIOLATIONS OF SINGLE-CROSSING

Figure S2 gives some information on how these violations depend on θ_i and on the risk-neutral probability (which is a decreasing function of odds). The color of the dots on Figure S2 represent the magnitude of the violation, and their size represents their statistical significance. The violations appear to be concentrated in two areas. The first one is the diagonal line, which corresponds to bettors indifferent between the favorite and the second favorite in a race. The second problem region is the right corner, where betting is on longshots.

APPENDIX F: DETAILED ESTIMATION RESULTS

In the tables in this section, the column labeled "Degrees" refers to the degrees of polynomials whose product is a single term in the sums in (13); for example, for the homogeneous case, $(1, 0, 2)$ means the product of a polynomial in R of degree 1, and a polynomial in p of degree 2. For the heterogeneous case, $(4, 1, 2)$ means a product of a polynomial in $R|\theta$ of degree 4, a polynomial in θ of degree 1, and a polynomial in p of degree 2.

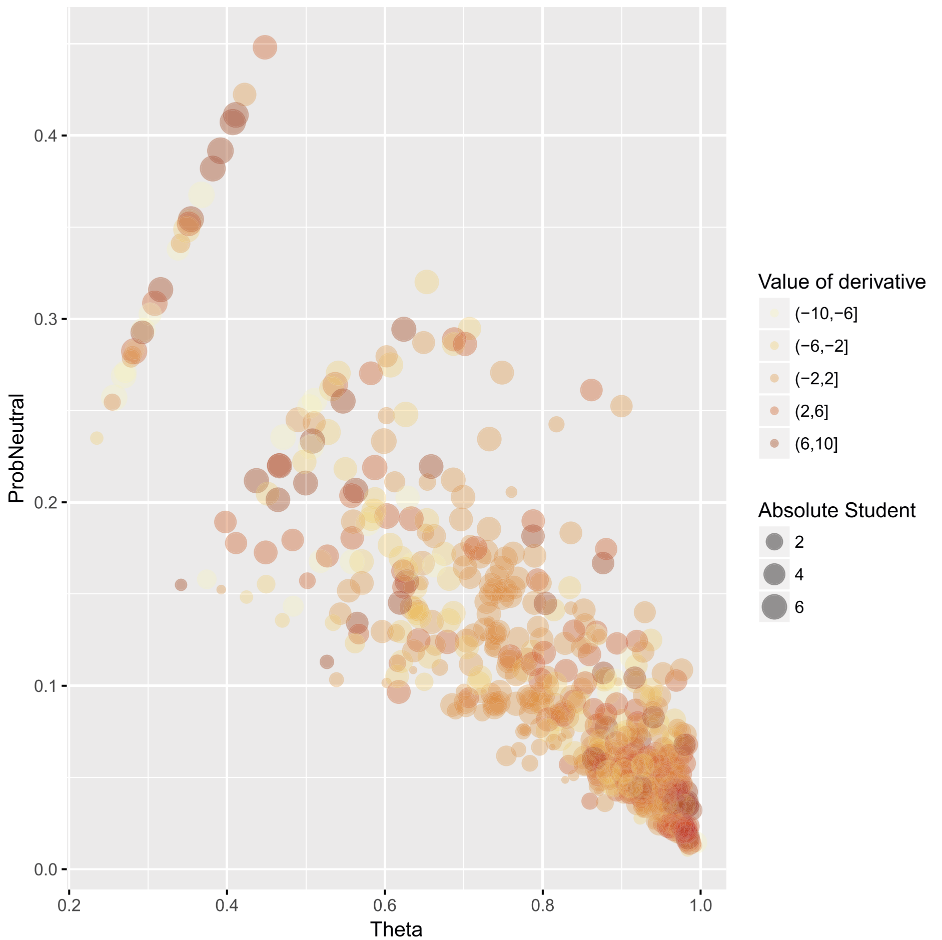

FIGURE S2.—Derivative of NF with respect to θ in the heterogeneous expected utility case.

TABLE S1

PARAMETER ESTIMATES FOR HOMOGENEOUS EU

	Degrees	Estimates	Standard Errors	Student
1	(1, 0, 0)	0.121	0.017	6.911
2	(12, 0, 0)	−0.039	0.011	3.532

TABLE S2

PARAMETER ESTIMATES FOR HETEROGENEOUS EU

	Degrees	Estimates	Standard Errors	Student
1	(1, 0, 0)	0.123	0.018	7.045
2	(2, 1, 0)	0.018	0.004	4.554
3	(3, 1, 0)	−0.015	0.004	4.092

TABLE S3

PARAMETER ESTIMATES FOR HOMOGENEOUS YAARI

	Degrees	Estimates	Standard Errors	Student
1	(0, 0, 1)	−0.031	0.005	6.735
2	(0, 0, 3)	−0.008	0.002	4.028

TABLE S4

PARAMETER ESTIMATES FOR HOMOGENEOUS RDEU

	Degrees	Estimates	Standard Errors	Student
1	(0, 0, 2)	−0.019	0.003	5.946
2	(1, 0, 0)	0.119	0.017	6.862

TABLE S5

PARAMETER ESTIMATES FOR HOMOGENEOUS (AND HETEROGENEOUS) NEU

	Degrees	Estimates	Standard Errors	Student
1	(1, 0, 2)	0.098	0.008	11.682

REFERENCES

CONLISK, J. (1993): "The Utility of Gambling," *Journal of Risk and Uncertainty*, 6, 255–275. [2]

Co-editor Lars Peter Hansen handled this manuscript.

Manuscript received 19 October, 2012; final version accepted 15 June, 2018; available online 3 October, 2018.