ONLINE APPENDIX

This appendix presents background material and results from additional tests which support the main results of the paper.

A. RACETRACK BETTING IN FINLAND

The betting company Fintoto has a legal monopoly for horse race betting in Finland. On-track and off-track betting opportunities are available. Off-track betting takes place in betting shops and on Fintoto's online betting platform with approximately half of the total revenue originating from online betting (2012 figures). When a bettor wishes to wager at Fintoto's online platform, a betting account must be registered. Upon registration, the bettor's name, address and social security number must be provided. Betting is possible once money has been transferred from a bank account to the betting account. Placing bets is allowed until a race is 'closed', i.e. the horses start running. Money transfers between accounts are possible at any time. In general, these transfers are free of charge, but a €0.3 fee is applied on transfers less than €15.

Races are organized almost every day. Across the country there are 20 tracks. Vermo, located in Helsinki, is the main racetrack ('the central racetrack'). 19 regional racetracks are located mostly in major cities. In terms of turnover, the most popular days are Saturday and Wednesday. Vermo always hosts on Wednesdays, but the racetrack on Saturday changes each week. In general, regional racetracks organize races on weekday evenings and during the daytime on weekends. In addition, there are also so-called 'lunch-races' around noon on weekdays. Usually a race meeting consists of ten races. While Finnish races are the most popular among bettors, betting on horse races in other countries is available as well. However, their contribution to turnover is marginal compared to local racetracks.

There are various bet types available for each race. Race-level types of bet include Win Bet (the winner of a race), Quinella (the two horses that finish first irrespective of their order) and Place (the chosen horse finishes first, second or third). In addition, Trifecta (first, second and third in the correct order) is available but only on some races. Cross-race or multiple race bet types include Double (the winners of two races) and T4 (winners of four races). In addition, T65 (the winners of six races) is available on Wednesdays and T75 (the winners of seven races) on Saturdays. Cross-race bets provide betting opportunities with very long-odds. Picking six winners in a row, for instance, is difficult because the number of possible outcomes is very large, and therefore, the variance of odds for these outcomes is very high. Thus, success is more likely in race-level bets rather than in cross-race bets.

The organizer's take-out rates vary across different betting types. For race-level bets, the take-out rate is 15 % for Win Bet, 15 % for Place, 21 % for Quinella, and 25 % for Trifecta. Therefore, with respect to the theoretical expected value, Win Bet and Place are the most profitable while Trifecta is the least profitable type of bet from the bettor's perspective. In cross-race betting types the take-out rate is 25 % for Double, and 35 % for other cross-race betting types.

B. RISK MEASURES

The proposed risk measures are calculated as follows. Assume a two-outcome model, where a bettor chooses a single outcome and if this outcome occurs, he wins the bet and loses otherwise. A prospect can be defined as $Y = [w, p; -b, (1-p)]$, where *p* is the probability of an outcome *w*, and *b* is the amount of the bet. The outcome is denoted by the net gain $w = b(O-1)$ where O stands for the odds. As a result, the expected value of the prospect can be expressed as

$$E[Y] = pw + (1-p)(-b) = b(pO-1). \qquad \text{(OA.1)}$$

Although the 'true' winning probability *p* is unobservable, the information contained in the odds is used here to estimate *p* and calculate the 'exact' expected value for each bet.

In Eq. (OA.1), the size of the stake *b* and the chosen betting type (the take-out rate varies between types of bet) define the expected value of a gamble: higher stakes and a higher take-out rate decrease the expected value making the gamble more 'costly' to the bettor and *vice versa*. In other words, since the take-out rates range only between 15 % (Win Bet and Place) and 25 % (Trifecta), the amount bet accounts for most of the variation in expected value across bets observed.

Taking variance as a measure for a bet's riskiness, the variance of a bet can be written as

$$VAR(Y) = E(Y^2) + [E(Y)]^2 = (bO)^2 p(1-p). \qquad \text{(OA.2)}$$

It is straightforward to see that variance is increasing in the betting amount, b, and the chosen odds, O.

Following Golec and Tamarkin (1998) and Cain and Peel (2004), the skewness of a bet is

$$SKEW(Y) = E(Y^3) - 3E(Y)VAR(Y) - [E(Y)]^3 = \frac{[1+b(pO-1)]^3(1-p)(1-2p)}{p^2}. \qquad \text{(OA.3)}$$

Since most gamblers bet on several horses within the same betting type, the expected value is a multinomial prospect which can be written as

$$E[Y] = \sum_{i=1}^{m} p_i w_i + \sum_{i=1}^{m} [1-p_i](-b_i), \qquad \text{(OA.4)}$$

where m is the number of horses a bettor can wager on. The variance of the multinomial prospect is

$$VAR(Y) = \sum_{i=1}^{m} (b_i O_i)^2 (p_i - p_i^2) - \sum_{i \neq j}^{m} b_i b_j (O_i)(O_j) p_i p_j = VAR(Y_i) - COV(Y_i, Y_j). \qquad \text{(OA.5)}$$

The skewness of the multinomial prospect is

$$SKEW(Y) = \sum_{i=1}^{m} [p_i (b_i O_i - 1)^3 - b_i^3] - 3E(Y)VAR(Y) - [E(Y)]^3, \qquad \text{(OA.6)}$$

where $E(Y)$ and $VAR(Y)$ are defined in Eq. (OA.4) and Eq. (OA.5). Further, the variances of prospects for all bet types (Win Bet, Place[1], Quinella, and Trifecta) are computed separately for each race. The total variance of bets for each race is a sum of the variances of all bet types as explained in an example in Section 3.

C. DATA

The data, which were provided by Fintoto, provide information on all bets placed online on a single day, Wednesday, August 1st 2012. Figure OA.I illustrates the structure of this day. It was possible to bet on harness

[1] In Place bet, the exact odds for an outcome are unobservable, but the range between the minimum and the maximum is observable. This follows from the nature of the data set because realized odds depend on the finishing order of horses. To calculate the probability of an outcome, we use the average of the range as the probability estimate. The range is also the information bettors observe before they make their own gambling choices.

horse races at three racetracks on Fintoto's online betting platform on that day. These racing events took place in different time frames. Seinäjoki racetrack hosted the so-called 'lunch-races' from noon until 1 p.m., Enghien race track in France hosted races from 2.50 p.m. until 6.35 p.m. In terms of bet volume, the most popular races were at Vermo from 6.30 p.m. until 9.30 p.m. As to the number of bettors, almost 75 % of all online bettors wagered on the Vermo races, and 80 % of them wagered *only* on those races. Vermo was also the final race meeting available to Fintoto bettors that day. A race meeting at Vermo includes 10 harness horse races, in which it is possible to bet race-level bets and cross-race bets. Win Bet, Place and Quinella betting types were available for all races and Trifecta in races 5 and 9 only. The available cross-race betting types were Double (in races 6 and 7), T4 (from race 7 to 10), and T65 (from race 2 to 7).

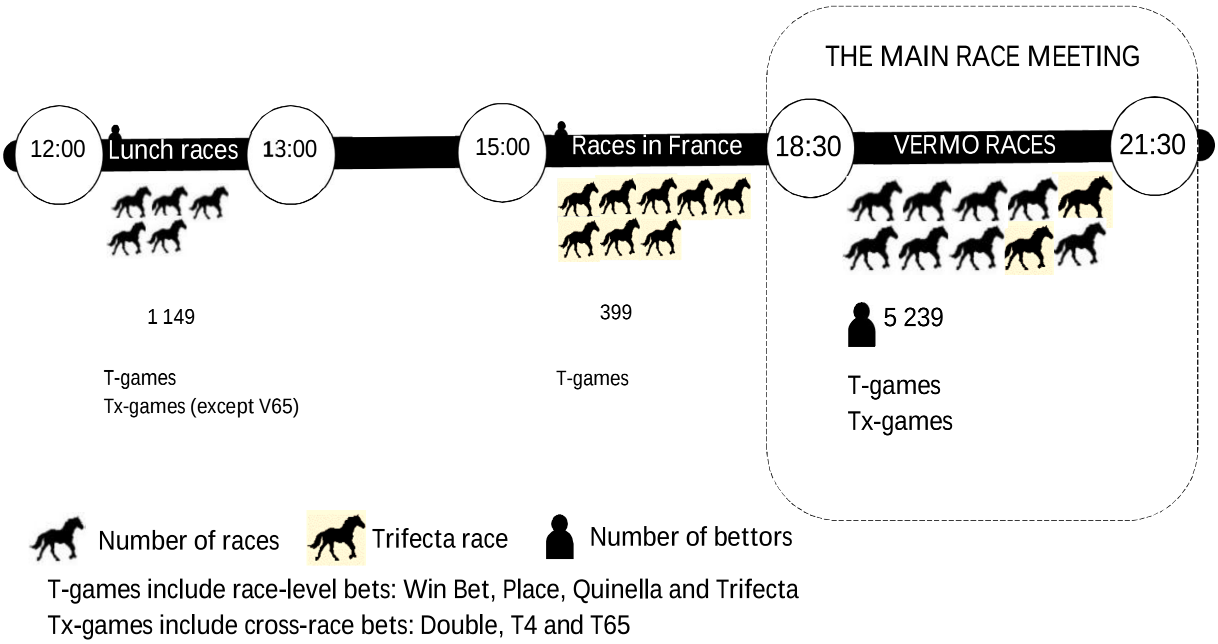

Figure OA.I. Horse Race Betting Opportunities via Fintoto on Wednesday 1st August 2012

In this study, we focus on bettors' choices in Vermo races because it is the main horse race meeting and it also ends the betting day[2]. The data include all bettors who bet on the Vermo races. Only those bettors whose records exhibit obvious errors in their information or bets (e.g. no observation for odds or the amount of bet)

[2] For example, we do not have a data set of odds for the French racetrack which is as accurate as that for the Finnish racetracks, and thus, the analysis conducted on the Vermo races is not possible for all races on August 1st 2012.

were dropped from the sample. While the data set includes accurate information on race-level bets, it does not contain any direct information on cross-race bets. The bet-level data set consists of 167,816 observations including bets placed and money transfers. However, in this study, we collapsed all bet-level information to the race level. In total, we have 5,217 bettors with 9,983 observations (races).

D. DESCRIPTIVE STATISTICS

Table OA.I reports descriptive statistics for betting types. Quinella is the most popular type of bet, with 3,706 bettors, whereas Place is the least popular with 908 bettors. The mean of the bet amount varies moderately between €4.8 and €10.9 across the betting types. The mean of variances increases with the level of difficulty associated with the type of bet in the sense that Place is the easiest and Trifecta the most difficult betting type.

Table OA.I. Descriptive statistics by bet type

Descriptive statistics by betting type				
Win bet	Number of bettors: 1265			
	Mean	Std. Dev.	Min.	Max.
Mean amount of bets per bettor per race (€)	6.9	27.7	1.0	850
Mean variance of bets per bettor per race	846.1	10873.3	0.5	371365
Place	Number of bettors: 908			
	Mean	Std. Dev.	Min.	Max.
Mean amount of bets per bettor (€)	4.8	9.4	1.0	167.5
Mean variance of bets per bettor per race	103.0	527.3	0.1	10152.6
Quinella	Number of bettors: 3706			
	Mean	Std. Dev.	Min.	Max.
Mean amount of bets per bettor (€)	8.6	14.8	1.0	300.0
Mean variance of bets per bettor per race	1352.3	9998.7	1.8	456778.0
Trifecta	Number of bettors: 2688			
	Mean	Std. Dev.	Min.	Max.
Mean amount of bets per bettor (€)	14.5	20.4	0.5	322.5
Mean variance of bets per bettor per race	4025.4	14314.6	5.4	479058.0
Total	Number of bettors: 5217			
	Mean	Std. Dev.	Min.	Max.
Mean amount of bets per bettor (€)	10.9	20.7	0.5	850.0
Mean variance of bets per bettor per race	2104.5	11569.4	0.1	456778.0

E. ROBUSTNESS TEST OF BETTING MODEL

As with all empirical studies, there are limitations in this study as well. To address these potential problems, several robustness tests were conducted. In all the results in this section, the variance of the set of bets is the risk measure used as the dependent variable in regression models.

The influence of other racetracks. Although participation rates in other racetracks were low compared to the Vermo races during the race meeting, it is nevertheless worthwhile checking that betting in another race track does not have a significant influence on our results. To detect the potential impact, a subsample that includes bettors who only bet on Vermo was created. The same regressions as in Table 5 were carried out using this subsample. The difference between the estimates from the original data set and the subsample was tested by comparing unstandardized coefficient estimates from the two samples.

The impact of favorite-longshot bias (FLB). It is well-established in the literature that subjective probability estimates calculated from odds are biased by FLB (e.g Snowberg and Wolfers 2010). Therefore, FLB may cause some bias in the probability estimates used when variance and skewness are calculated. We conducted a robustness check for this possible bias. The probability estimates are 'corrected' using the results given by Snowberg and Wolfers (2010) who estimated Prelec's (1998) parametric probability function parameter for FLB from a large data set of the US pari-mutuel horse race betting markets[3]. Because their findings suggest that low probabilities are overweighted and high probabilities are underweighted, we then correct our probability estimates to match these results. More specifically, Prelec's one-parameter weighting function is expressed by the form $w(p) = \exp\left(-(-\ln p)^{\alpha}\right)$ where α is a probability weighting parameter with the value 0.928 estimated by Snowberg and Wolfers (2010). Finally, we calculated 'a corrected variance' which is regressed on the covariates that were applied in the betting model.

[3]Although the probability weighting parameter estimate from the US horse race betting markets may not be exactly the same in Finland, the empirical literature confirms that FLB is present in the Finnish racetracks as well (Suhonen 2011; Kanto et al. 1992).

The effect of correlations between choices across betting types. As noted above the summation of variances for betting types may cause some choices to be correlated with each other, i.e., we cannot observe whether bettors pick the same horse for different betting types in the same race. In consequence, we compute a 'simple variance estimate' which is the difference between the worst outcome (all bets are lost) and the best outcome (the largest gain is attained) for each race. This simple variance estimate does not take into account any information regarding probability estimates, and therefore, it is more robust to the cross-correlations between choices over betting types. We run the regression model using this 'simple variance estimate' as a dependent variable.

The effect of unobservable cross-race bets. While cross-race bets are not directly observable in the data, they may still affect estimates in the betting model. However, since we are able to track changes in the betting account balance before bets are placed in each race, this value may be compared to the initial account balance. In other words, the bettor's position (gain or loss) is contrasted with the account value at the beginning of the race meeting. This indicates how far the bettor is from his 'reference point' when he places bets in each race. Consequently, this includes all changes that cause deviations including betting amounts on and possible gains from the cross-race bets. Using this indirect information, we calculated new independent variables, which are similar to the betting model, for the robustness test.

Table OA.II shows estimated coefficients for the robustness test regression models. The left panel of Table OA.II reports the results for the possible effect of excluding races at other tracks. The signs of the coefficient estimates do not differ from the betting model. The only difference is that the dummy variable for the loss domain is no longer statistically significant in the subsample. However, the difference between the betting model estimate and the subsample estimate is not statistically significant ($p = 0.26$). Moreover, the coefficient estimate on the size of cumulative loss remains negative and highly significant. Hence, it is unlikely that betting in other racetracks has much impact on the estimates of the betting model. The right panel of Table OA.II reports estimates for the robustness test concerning FLB. Again, the signs and statistical significance of the estimates are in line with the betting model, which suggests that FLB has no meaningful impact on the results[4]. In addition, estimates of the robustness test for possible correlations between chosen betting types are presented in the middle panel of Table OA.II. This 'simple variance estimator' yields results that are qualitatively similar to the betting model because there are no differences in signs or the statistical significance of estimates between the betting model and the robustness check model.

Table OA.II. Robustness Tests for Other Racetracks, FLB and Correlations between Chosen Bet Types.

Method						
Fixed Effects Model						
	Subsample – only Vermo Bettors		Simple Variance Estimate		FLB Corrected Variance Estimate	
Dependent Variable	*Ln*(Var)		*Ln*(Simple Var)		*Ln*(FLB Corrected Var)	
Independent Variable	Coefficient	Robust Stand. Error	Coefficient	Robust Stand. Error	Coefficient	Robust Stand. Error
GainD	0.1951***	0.0581	0.2267***	0.0742	0.1673***	0.0537
LossD	-0.0325	0.0440	-0.1746***	0.0591	-0.1190***	0.0420
Cgain	0.0004	0.0007	0.0002	0.0008	0.0002	0.0006
Closs	-0.0058***	0.0013	-0.0022*	0.0012	-0.0031***	0.0010
Open Bets	-0.0094***	0.0015	-0.0096***	0.0016	-0.0098***	0.0014
Race Rank	-0.0806***	0.0165	-0.1098***	0.0207	-0.0742***	0.0146
Transfer	0.0014***	0.0005	0.0014***	0.0005	0.0014***	0.0003
Constant	4.5082***	0.0542	7.3644***	0.0630	4.8179***	0.0456
Race Dummies	Yes		Yes		Yes	
No. Bettors	4209		5217		5217	
Obs.	13744		18346		18346	
Within R^2	0.23		0.26		0.25	
Notes: Coefficients for race dummies are not reported. Statistical significance: * < 0.1, ** < 0.5, and *** < 0.01.						

Table OA.III reports estimates for the robustness test concerning cross-race bets. Again, the model that is corrected for these bets yields estimates whose signs and statistical significance are the same as in the betting model. However, the magnitude of the estimates for the gain domain and for the loss domain are larger in the latter model. Most likely this results from increased variation in the independent variables in the corrected model, because they contain all deviations including cross-race bets and their possible gains.

[4]As an additional robustness test, the model was estimated with a slightly more extreme value of α = 0.85 which did not alter the results qualitatively. Moreover, there is ample evidence of heterogeneity in the probability weighting function between individuals (e.g. Gonzalez and Wu (1999)), which we are not able to model here.

Overall, the robustness tests indicate that the results obtained in the betting model are fairly robust to alternative model specifications. In particular, estimates for the domain of gains and incurring losses are stable across all models. The estimate for the domain of losses is the only variable of interest which exhibits some sensitivity in terms of the absolute values of estimates across different models. Further, control variables are also stable across alternative specifications.

Table OA.III. Robustness test for cross-race bets

Method		
Fixed Effects Model		
Cross-race Bets Correction Model		
Dependent Variable	*Ln*(Var)	
Independent Variable	Coef.	Robust Stand. error
Above the reference point	0.2693***	0.0475
Below the reference point	-0.2674***	0.0340
Cumulatively above the reference point	-0.0001	0.0002
Cumulatively below the reference point	-0.0031***	0.0009
Open Bets	-0.0062***	0.0011
Race Rank	-0.0630***	0.0130
Transfer	0.0007**	0.0003
Constant	4.5214***	0.0406
Race Dummies	Yes	
No. Bettors	5217	
Obs.	18346	
Within R^2	0.23	

Notes: Coefficients for race dummies are not reported. Statistical significance: * < 0.1, ** < 0.5, and *** < 0.01.

References

Cain M, Peel DA (2004) Utility and the Skewness of Return in Gambling. *Geneva Papers on Risk & Ins. Theory* 29:145-163.

Golec J, Tamarkin M (1998) Bettors Love Skewness, Not Risk, at the Horse Track. *J. Political Econom.* 106(1):205 – 225.

Gonzalez R, Wu G (1999) On the shape of the probability weighting function. *Cognit. Psychology* 38:129-166.

Kanto A, Rosenqvist G, Suvas A (1992) On utility function estimation of racetrack bettors. *J. Econom. Psych.* 13:491-498.

Prelec D (1998) The Probability Weighting Function. *Econometrica* 66(3):497-527.

Suhonen N (2011) Market Efficiency in Finnish Harness Horse Racing. *Finnish Econom. Papers* 24(1):55-63.

Snowberg E, Wolfers J (2010) Explaining the Favorite-Long Shot Bias: Is It Risk-Love or Misperceptions? *J. Political Econom.* 118(4):723-46.