Strategic Forecasting among Experts: Evidence from the Horse-Racing Tipsters

Bruno Deschamps and Olivier Gergaud *

September 2, 2005

1 Introduction

It is established since the seminal paper of Scharfstein and Stein (1990) that reputation concern induces forecasters to release biased forecasts. Since then, a growing theoretical literature has shown that, in order to maximize their reputation, forecasters may choose to release forecasts that are not consistent with their own beliefs. This can generate herding or anti-herding behavior. One of the implication of this literature is that forecasts' originality with respect to the consensus can depend on the reputation of forecasters.

For this study, we have collected a large amount of data from two main French daily horse-racing daily newspapers: Paris-Turf (PT henceforth) and Tiercé Magazine (TM henceforth). Both newspapers report every day, before each race, the tips of a set of professional and non-professional horse-racing tipsters. There are as a whole 101 tipsters : 71 professionals and 30 non-professionals (10 jockeys, 10 drivers and 10 trainers). A tip consists of an ordered list of 8 horses that are expected to be the most competitive. After each race, every tipster scores some points if his/her tip succeeded in predicting

*B. Deschamps: ECARES, Université Libre de Bruxelles, 50 avenue Roosevelt, CP 114, B-1050 Bruxelles, Belgium. Tel.: + +32 (0) 2.650. 33.75, E-mail: bdescham@ulb.ac.be. O. Gergaud: OMI-Université de Reims Champagne-Ardenne and TEAM-Université de Paris-I. Address: 57 bis, rue Pierre Taittinger, 51096 Reims Cedex, Tel.: + 33 (0) 3.26.91.38.56, Fax: + 33 (0) 3.26.91.38.69, E-mail : olivier.gergaud@univ-reims.fr.

the outcome of the race. The more precise the tip, the more points he/she scores. Both newspapers rank their tipsters according to the number of points scored during the current year. In details, there are 2 distinct constests for 65 professional tipsters (TM: 35 tipsters and PT: 30 tipsters) and 3 other PT contests for 30 non-professional tipsters (Jockeys: 10, Drivers: 10 and Trainers: 10). At last, we also have the tips for 6 professional and higly-reputed tipsters who are not involved in any contest. All these contests started January 1st 2004 and ended December 31st, 2004.

The goal of this paper is to analyze whether reputation concern induces tipsters to make more or less original and risky tips in such contests. Said differently, does tipsters' behavior depends on their reputation or not? If their goal is to maximize their reputation, they will do everything to finish well placed in the contest at the end of the year. The reputational herding literature predicts that this reputational concern can induce tipsters to be more or less original/conservative with respect to the public information depending on their position in the contest. This is what we test in the paper. Contrary to Chen and Jiang (2005), Bernhardt, Campello and Kutsoati (2004) and Zitzewitz (2001), we do not test whether experts herd or anti-herd the public information. We simply analyze whether experts change the level of originality of their forecasts when their reputation evolves.

In order to achieve this objective, we have to observe the public information. We proxy this public information by ranking -per race- each horse on his likelihood of winning the race. This likelihood is proxied by a set of 12 variables available before each race such as the form, the jockey etc. By doing so, we get another ordered list of 15-20 horses from most likely to win to most likely to lose the race. Then, the originality level of a tip is measured by comparing it to this public information. We call the distance between the tip and the public information the originality of the tip. For instance, we attribute a very low level of originality to a tip which is very close to the public information and vice-versa.

In order to investigate whether strategic forecasting behavior takes place in these contests, we estimate whether changes of reputation induce the tipsters to change the originality of their forecasts. To do so, we estimate the reputation of a tipster by his/her position in the contest. Top-ranked tipsters are assumed to have a good reputation level while bottom-ranked tipsters are assumed to have a low reputation level.

We find that contestants are getting more and more original as their reputation (position in the contest) goes up. This means that bottom-ranked tipsters are less original than top-ranked tispsters. This effect is clear in the case of the two most prestigious contests (professional tipsters, minimum of 30 contestants). In the other three less prestigious contests (non-professional tipsters, 10 contestants each), there is no evidence for an effect of reputation on originality. We also find that tipsters tend to be more original both at the beginning and at the end of the contest. The econometric analysis also suggests that a successful tipster in t is all the more original in $t+1$ that the number of successful tipsters in t is small and reciprocally. Finally, we learn that unsuccessful tipsters in t are all the less original in $t+1$ that the number of successful tipsters in t is large. We interpret these results as evidence of strategic forecasting behavior.

This paper can also be considered as a natural experiment for examining the risk-taking behaviors in rank-order tournaments.[1] These tournaments are usually used when absolute performance can be contaminated by common shocks (Prendergast 1999). Our dataset is suitable to study the effect of ranks on risk-taking because the degree of originality of a tip can be interpreted as the risk taken by the tipster. Our results suggest that the amount of risk in these tournaments depends indeed on the rank of the agents. In this study, contary to some predictions of the literature, we do not find that bottom-ranked tipsters take more risks. On the contrary, they react to a lower rank by taking significantly less risks.

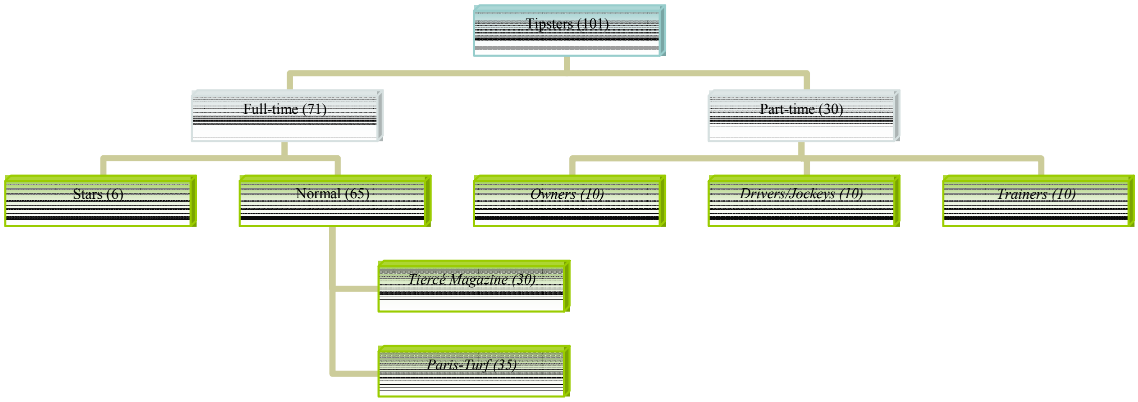
[1] For other evidence, see Lee (2004) and Knoeber and Thurman (1994).

2 The Data

2.1 Sources

The data have been collected from two main French horse-racing daily newspapers: Paris-Turf and Tiercé Magazine. Both newspapers publish tips the day before each race. More precisely, they report the tips of not-less than 101 experts. Each of them tips an ordered list of 8 horses that they expect to be the most competitive during the race.

Tipsters are of two types: professional (i.e. full-time) and non-professional (i.e. part-time). The latter category is made of three types: trainers, drivers and jockeys. 6 tipsters (1 part-time and 5 full-time) enjoy a high level of reputation in the field of Pari-Mutuel Betting (Omar Sharif for instance). A particularity is that their performances are not accounted for, given that these Superstars do not participate in any contest, contrary to "normal" tipsters who get points for each of their tips given their relevance, i.e. ability to predict the race outcome. They score some points if the top 3 (tiercé), top 4 (quarté) or top 5 (quinté) finishers are among the 8 horses they tipped. The number of points they score also depends on whether the race was easy to predict or not, and they get a special bonus when they succeed in predicting a tiercé, a quarté or a quinté in the exact order. At the end of the year, the tipster having scored the most points is declared the contest's winner. The dataset contains as a whole 95 different tipsters involved in 5 distinct annual contests/championships (January 1st, 2004 till December 31st, 2004) plus 6 Star tipsters:



Remark: contests/championships in italics.

This original dataset allows to address several interesting issues related to the experts' strategic behavior. In particular, we are interested in analysing their risk-taking strategy, the extent to which tipsters depart from the consensus, in what circumstances they tend to herd or anti-herd.

In order to analyse the tipsters' strategic forecasting behavior, we need a "consensus forecast" or "public information". To proxy this public information, we rank -per race- each registered horse (between 15 and 20 horses) on the basis of their likelihood of winning the race from a set of 12 dummy variables: whether or not the horse is suited to the track, whether or not he is on form, whether or not his jockey/driver performs well, etc. We assume that these information are common knowledge among all the experts even if they are published in $t-1$ along with the tips for race t. In details, we compute a sum of these 12 dummies (consensus forecast) and rank horses according to this statistics. This proxy is then used to estimate how original a forecast is by calculating how distant from the consensus forecast it is.

2.2 Descriptive Statistics

There are as a whole 25,563 different tips as follows:

Table 1: Risk-Taking statistics

Contest	Tipsters	Races	Min.	Max.	Mean	Std. Dev.
Part-time tipsters						
Drivers	10	142	0.71	9.71	3.92	1.47
Jockeys	10	39	1.29	8.57	3.78	1.25
Trainers	10	126	0.86	9.00	4.21	1.37
Full-time tipsters						
Paris-Turf	35	330	0.00	9.38	3.81	1.19
Tiercé-Mag.	30	299	0.50	9.13	3.83	1.21
Stars	6	329	0.63	8.63	3.84	1.24
Total	101	-	0.00	9.71	3.84	1.23

* A 8-horse tip corresponding exactly to the consensus forecast Top 8 horses is given a 0 here.

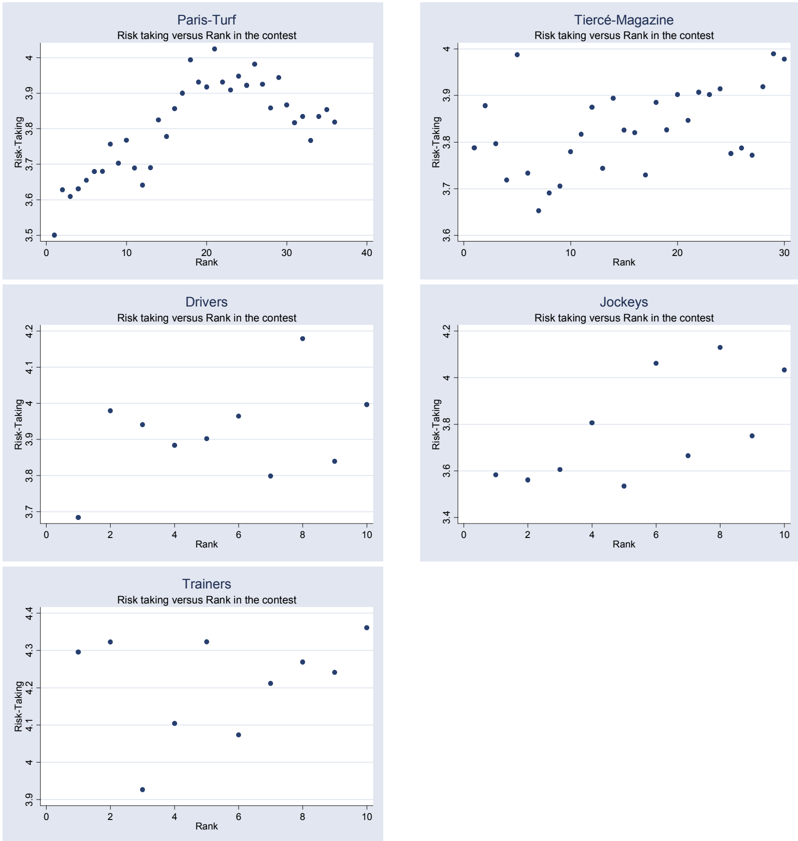
In the matter of Risk-taking, part-time tipsters are slightly more original than full-time tipsters even if their tips vary more on average (see Table 1). The fact that their reputation and revenues don't depend on their forecasting activity may explain why they behave more freely and why they forecast more originally. One may also interpret this difference from the number of tipsters involved in the contest: the smaller the contest is, the more original the forecasts are. A big contest seems to provide more incentives to tipsters who take less risk to win.

From Table 2, we learn that full-time tipsters score more frequently than part-time tipsters. Surprisingly, the 6 Superstars are not the most efficient tipsters with an average frequency of success of only 10%. The best tipsters are thus those involved in the two main contests (PT and TM). We see here that tips are more frequently successful when the tipster is a full-time one, when it is involved in a contest and when this contest is a large and prestigious one.

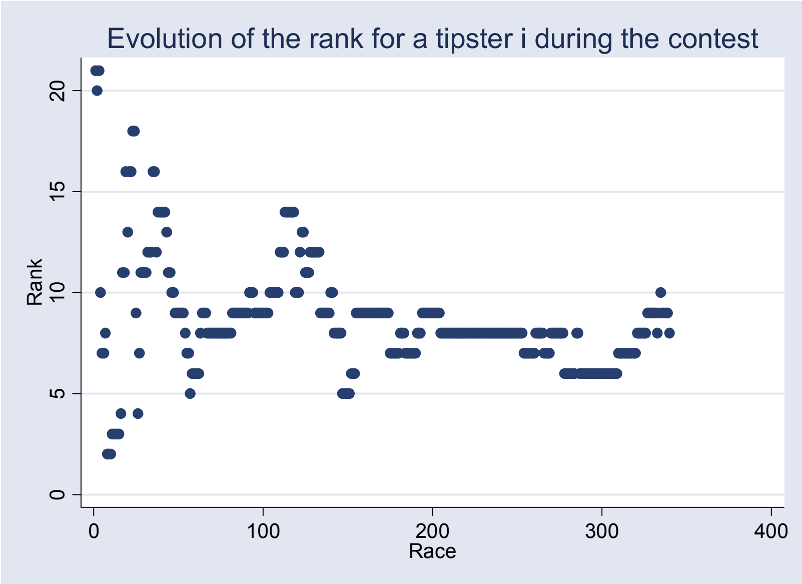
Table 2: Frequency of Success

Contest	Contestants	%
Part-time tipsters		
Drivers	10	21
Jockeys	10	14
Trainers	10	6
Full-time tipsters		
Paris-Turf	35	33
Tiercé-Mag.	30	32
Stars	6	10
Total	101	29

With the following Graphs, we see how Originality vary with the average ranking computed -contest by contest- over the whole contest. Top-ranked tipsters tend to be less original than middle-ranked tipsters and bottom-ranked tipsters. Let's remark also an interesting quadratic shape for PT.



This positive relationship between Originality and Ranking is confirmed by a simple OLS regression between the two variables (see Table 3 - Model 0). Let's finally remark with Graph 2 that the standard deviation of the rank tends to decrease during the contest. The rank doesn't vary much during the second half of the contest.



3 Results

3.1 Fixed-Effects Regressions

In this section, *Originality* (O) is regressed against several measures of *Success* using a series of fixed-effect analyses[2]. Letting $i = 1, ..., I$ index the tipsters, $t = 1, ..., T$ index the races (and thus time), the basic model is:

$$O_{it} = \alpha + \beta Rk_{it-1} + \epsilon_{it1} \tag{1}$$

With Rk_{it-1}, the relative rank of individual i in period $t-1$ and contest c[3]. By relative rank we mean the absolute rank divided by the number of tipsters involved in the contest (either 35, 30 or 10). Strangely, β is non-significant in Model (2), a model in which unobservable individual heterogeneity is controlled for (see Table 3). It is likely the indication that the rank captures this individual heterogeneity and perhaps the talent of tipsters.

[2] The Hausman test rejects systematically the rendom-effects specifications.
[3] Index c is dropped here and in what follows for convenience.

We use thereafter the other -available and continuous- measure of *Success* instead of Rk_{it-1}: the number of points the tipster gets after each successful race. In Model (2), *Success* is either cumulative (S_{it-1}) or not (s_{it-1}). S_{it-1} is proxied by the total number of points (P_{it}) a tispter accumulates, race by race, given the relevance of its successive tips up to period $t-1$: $S_{it-1} = \sum_{t=0}^{t-1} P_{it}$. s_{it-1} is a dummy variable which takes the value 1 when the previous tip is successful, otherwise 0. Model (2) has the following structure:

$$O_{it} = \alpha + \beta S_{it-1} + \gamma s_{it-1} + \delta T + \epsilon_{it2} \tag{2}$$

PT and TM have different assessment systems. PT has also adopted a particular system for its three part-time tipsters contests (Jockeys, Drivers and Trainers). As a whole, we have three distinct assessment systems (PT full-time, PT part-time and TM). We take this aspect into account by interacting the continuous variable S_{it-1} with the following dummies: PT, TM, Part-time (Jockeys + Drivers + Trainers) and Stars. Finally, to introduce the 6 Star-tipsters in the analysis, we have simulated for each of them a series of points randomly using a standard normal distribution.

Model (2) exhibits a positive and significant relationship between *Originality* and cumulative success for PT and TM and a negative one for Part-time tipsters (significant at the 5% level). Forecasters are getting more and more original as they go up in the ranking. Top-ranked tispsters appear to be more original than middle-ranked and bottom-ranked tipsters. Hopefully, the 6 Stars get a non-significant coefficient. We also get a significant and negative effect for Time, meaning that tipsters are more original on average at the beginning of the contest and that this originality tends to decrease over the contest.

s_{it-1} is surprisingly non-significant. This is probably due to the fact that the effect of the last outcome (success or failure) is not captured correctly when

Success is expressed in absolute terms, i.e. when one doesn't control for other tipsters' successes or failures. Our intuition here is that the reaction to the last outcome is more likely relative than absolute. Indeed, one may wonder how does tipster i react to other tipsters' outcomes. We test this intuition by introducing two variables, one for the relative success (Rs_{it-1}) and another one for the relative failure (Rf_{it-1}) of tipster i at period $t-1$:

$$\begin{aligned} Rs_{it-1} &= \sum_{j \neq i} s_{jt-1}/N-1 \qquad & if s_{it-1} = 1 \\ Rf_{it-1} &= \sum_{j \neq i} s_{jt-1}/N-1 \qquad & if s_{it-1} = 0 \end{aligned}$$

Rs_{it-1}, and Rf_{it-1} are proportions defined on an interval [0,1]. When Rs_{it-1} = 0, tipster i is the only successful tipster among the N tipsters[4] involved in the contest at period $t-1$. $Rs_{it-1} = 1$ means that everyboby is successful in $t-1$. $Rf_{it-1} = 0$ means that noboby has won in period $t-1$. $Rf_{it-1} = 1$ means that tipster i is the only loser among the N tipsters involved in the contest at period $t-1$. Hence, a value of Rs_{it-1} (Rf_{it-1}) close to 1 (0) may be interpreted as a "banal" success (failure) given that a lot of (a few) contestants have a successful outcome. These variables replace s_{it-1} in Model (3):

$$O_{it} = \alpha + \beta S_{it-1} + \gamma_S Rs_{it-1} + \gamma_F Rf_{it-1} + \delta T + \epsilon_{it3} \qquad (3)$$

Model (3) shows that tipsters react significantly to relative success as expected (negative sign). A strong relative success (Rs_{it-1} close to 0) is proved to boost originality of tipster i at period t. Interestingly, we also get that tipsters are less original in t when they lose in $t-1$ and that the outcome of the race was relatively easy to predict (Rf_{it-1} close to 1) in comparison with a "banal" failure.

In Model (4), we introduce S^2_{it-1} and T^2, respectively the square of S_{it-1} and T to see whether the relationship between *Originality*, time and cumulative

[4] With N, the total number of successful tipsters in contest c.

success is linear or not. The relationship appears quadratic for TM only and linear for the other contests (PT and Part-time). The introduction of these square terms does not add much to the results and even causes some damages (see for instance Part-time S_{it-1}) except perhaps in the case of Time. The coefficient of T^2 is positive and significant, indicating that tipsters tend to be more original both at the beginning and at the end of the contest than in the middle of the contest. A satisfying model seems to be one in which we keep the square for Time only (Model 5). In this model, cumulative success is non-significant for part-time tipsters. With this specification, we get that γ_S is significantly different from γ_F at the 5% level: Fstat(1, 25444) = 3.90. This indicates that tipsters react more in terms of originality in case of failure than in case of success in $t-1$ ($\widehat{\gamma}_F > \widehat{\gamma}_S$). A Wooldridge test for autocorrelation in panel data on the same model fails to reject the null hypothesis of no first-order autocorrelation: F(1, 100) = 0.373.

These results suggest that there is a clear and strong tournament effect at least in the case of PT and TM main contests. Contestants are getting more and more original as their reputation in the contest goes up. Tips appear more original both at the beginning and at the end of the contest.

3.2 Originality versus Rank

In the former section we have analyzed the effect of success on originality by pooling all the observations. In this section, we present very preliminary evidence suggesting that the effect of success on originality depends on whether tipsters are top-ranked or bottom-ranked.

The reason why we believe that the effect of success on originality depends on the rank is that the incentives to take risks are no the same everywhere in the contest. First, the goal of tipsters is obviously to win the contest. Indeed, a victory is magnified in the newspapers. So, the tipsters who are not too far behind the first place could have incentives to make risky tips in order to become first. Second, the bottom-ranked tipsters can lose their job or their place in the contest if they finish at the last place. These tipsters could then take less risks in order to score points and avoid the last place.

3.2.1 Predictions

Our predictions are thus the following:

For top-ranked experts: the better ranked an expert is, the higher the reward of gaining one position is. Top-ranked experts react thus to success by being more original and deviating more from the consensus forecast.

For bottom-ranked experts, the lower ranked an expert is, the higher the cost of loosing one position is. Bottom-ranked experts react thus to failure by being more original and deviating more from the consensus forecast.

3.2.2 Testing the predictions

Let us now test these predictions. We only consider Paris-Turf for the moment but the results seem to hold with Tiercé Magazine as well. Letting $i = 1, ..., I$ be index for tipsters, $j = 1, ..., 356$ index for ranks and $t = 1, ..., T$ index for races, the basic model is:

$$O_{jt} = \alpha_j + \beta_j Rk_{jt-1} + \varepsilon_{jt} \tag{4}$$

When estimated rank by rank, Model 4 produces a set of 35 values of β that are plotted in Figure 1 against he rank. For instance, $\beta > 0$ and $j = 1$ means that, on average, the tipster who is ranked first reacts to success by making a more original tip.

This figure suggests that there exists a clear relationship between the rank of an expert and his/her reaction to the success. The correlation coefficient between β_j and ranks is indeed -0.65 and the slope is significantly negative. Top-ranked experts react to success by taking more risk while bottom-ranked experts react to success by taking less risks. These results are clearly consistent

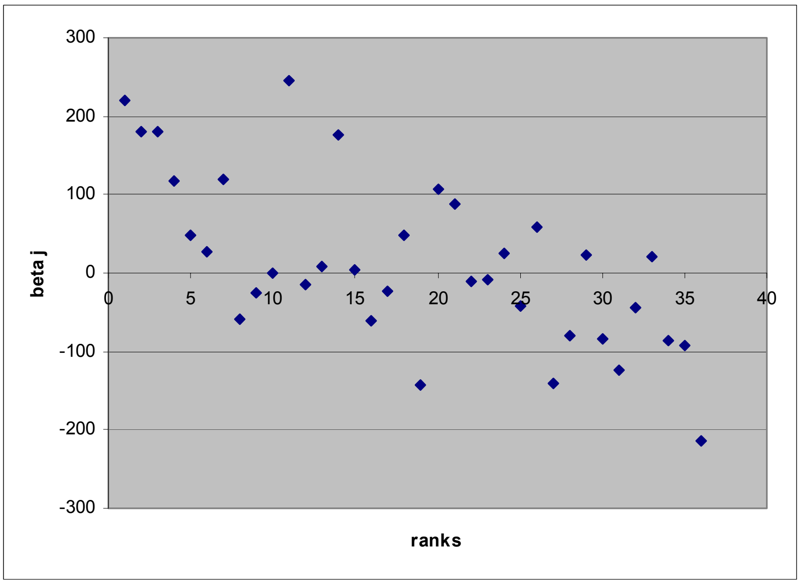


Figure 1:

with our predictions. Our interpretation of these results is that the level of originality can be explained by career concern motives. The extent to which experts herd the consensus strongly depends on their recent successes.

This is however not a proof that there exists a contest effect that depends on the rank. Indeed, there is a potential endogeneity problem in the sense that the way a tipster reacts to relative success can affect his/her rank in the contest. The causality of the relationship between ranks and β_j is therefore not obvious. Do expert choose, as we hope, a β depending on their ranks? Or is it simply the rank that is determined by the β they have chosen? In this section we argue why that the causality goes from ranks to β.

First, could it be that the relationship between β_j and ranks is only caused by individual fixed effects? If it were the case, we would expect β to vary more across individuals than across ranks. But this is not the case as $Var(\beta_j) = 0.111$ and $Var(\beta_i) = 0.082$. $\beta_i's$ are obtained from a series of Equation 4 estimated

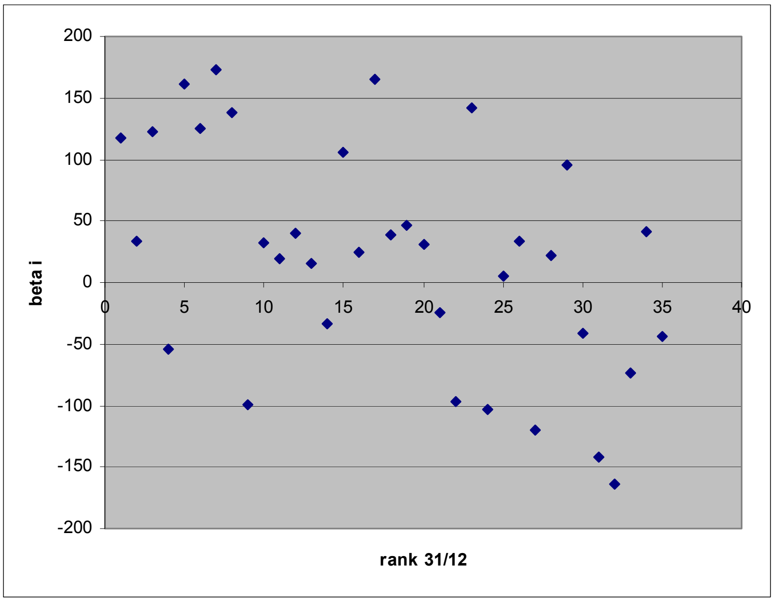


Figure 2:

on i, that is to say tipster by tipster instead of rank by rank like in the case of $\beta_j's$. This is one more indication that the rank has an effect on the behavior. But the main reason to believe in a contest effect comes from Figure 2.

The figure represents the β_i for every expert and their final rank in the contest, i.e. their rank on the 31st of December. It shows that there exists a negative relationship between the final rank in the contest and β_i. However, the correlation between β_i and rank31/12 is only -0.50, which is much lower than the correlation between β_j and the ranks (-0.65). If there was no contest effect we would expect $corr(\beta_i, rank31/12) > corr(\beta_j, ranks)$. Indeed, experts' position in the ranking changes often. Therefore, a smoothing effect of β_i across ranks should appear if there was no contest effect. The individuals fixed effect seem thus enable to explain the entirety of the difference of β_J between ranks. This is a strong indication that there exists a contest effect.

4 Conclusion

TBW

References

[1] Bernhardt, D., Campello, M. and Kutsoati, E. (2004). "Who Herds?" mimeo, University of Illinois.

[2] Chen, Q. and Jiang, W. (2005). "Analysts' Weighting of Private and Public Information." *Review of Financial Studies*, forthcoming.

[3] Knoeber, C. and Thurman, W. (1994) "Testing the Theory of Tournaments: An Empirical Analysis of Broiler Production." *Journal of Labor Economics*, 12(4), 155-179.

[4] Lee, J. (2004) "Prize and Risk-Taking Strategy in Tournaments: Evidence from Professional Poker Players." IZA Discussion Paper 1345.

[5] Prendergast, C. (1999). "The Provision of Incentives in Firms." *Journal of Economic Literature*, 37(1), 7-63.

[6] Scharfstein, D. and Stein, J. (1990). "Herd Behavior and Investment." *American Economic Review*, 80(3), 465-479.

Table 3 : OLS and Fixed-Effects Regressions for the relationship between Originality and Success (Dep. Var. : O_{it}*1000)

Variables	Notation	OLS	Fixed-Effects Regressions				
		(0)	(1)	(2)	(3)	(4)	(5)
Rank	Rk_{it-1}	202.452	-41.523				
		(7.53)**	(0.84)				
Time	T			-155.645	-156.999	-516.699	-750.048
				(2.21)*	(2.23)*	(2.47)*	(5.35)**
Time²	T²					329.673	530.393
						(2.00)*	(4.89)**
Last outcome							
Absolute	s_{it-1}			-18.025			
				(1.05)			
Relative & Success	$Rs_{it-1}*s_{it-1}$				-58.306	-69.370	-69.225
					(2.12)*	(2.51)*	(2.51)*
Relative & Failure	$Rs_{it-1}*(1-s_{it-1})$				-139.502	-151.996	-152.440
					(3.15)**	(3.43)**	(3.44)**
Cumulative Success	S_{it-1}						
TM				0.021	0.021	-0.024	0.025
				(3.13)**	(3.09)**	(1.24)	(3.70)**
TM²						0.000	
						(2.82)**	
PT				0.070	0.070	0.063	0.085
				(3.03)**	(3.02)**	(1.00)	(3.64)**
PT²						0.000	
						(0.26)	
Part-time				-0.025	-0.025	-0.027	-0.018
				(1.98)*	(1.99)*	(0.95)	(1.46)
Part-time²						0.000	
						(0.26)	
Stars (simulated)				-0.002	-0.002	-0.023	-0.002
				(0.34)	(0.35)	(0.95)	(0.37)
Stars² (simulated)						0.000	
						(0.89)	
Constant		3,737.212	3,864.071	3,827.618	3,850.873	3,961.996	3,951.535
		(234.35)**	(143.88)**	(220.56)**	(204.45)**	(139.21)**	(141.58)**
Observations		25563	25563	25553	25553	25553	25553
R-squared		0.0022	0.0022	0.0010	0.0007	0.0000	0.0003

Absolute value of t statistics in parentheses

* significant at 5%; ** significant at 1%