What are the Odds?
A Measure of the Small Sample Problems

Kurt Rotthoff*
Seton Hall University
Stillman School of Business

Danielle Zanzalari
Seton Hall University

John Jasina
Claflin University
School of Business

Fall 2010

Abstract

Decisions on whether to retain recent hires are often limited by small sample size. We empirically assess whether uncertainty in employee retention decisions could be significantly reduced by increasing sample size. Using a unique data set from professional tennis matches to measure small sample outcomes, we find little difference in giving three chances, relative to five chances, in determining innate ability.

JEL: L23, J21, D21, L83

---

* Kurt Rotthoff at: Kurt.Rotthoff@shu.edu or Rotthoff@gmail.com, Seton Hall University, JH 621, 400 South Orange Ave, South Orange, NJ 07079. John Jasina: jjasina@claflin.edu, Claflin University, 400 Magnolia St., Orangeburg, SC 29115 and Danielle Zanzalari: danielle.zanzalari@gmail.com. We would like to thank Anca Cotet, Patrick McLaughlin, and Hillary Morgan for helpful comments. Any mistakes are our own.

I. Introduction

There is a problem of uncertainty with small sample sizes. However, small sample problems occur frequently and cannot be avoided in many instances. One small sample problem in business is the decision to retain talent. In the corporate setting, managers have to decide how many opportunities to give employees to prove they are an asset to the company. This problem is particularly challenging because often only small samples of work inform the decision, especially with new employees.

Small sample problems are nothing new. This issue has been addressed in Guilkey and Salemi (1982), Mankiw and Shapiro (1986), and Nelson and Kim (1993). Small sample problems are found in employee discrimination rulings (Piette and White 1999) and NCAA rules violations (Ruibal 2009). We expand the empirical applications of small sample problems by using a unique dataset from tennis to address this problem. Comparing outcomes in professional tennis matches, we test outcome differences in three and five-set matches. We find that when the sample is small, minor differences in sample size do little to increase the probability of the better person winning.

The next section establishes the employment problem, followed by a more specific discussion of tennis. Section IV discusses the data and methodology, followed by the results. The last section concludes.

II. The Employment Problem

One key decision for managers is what employees to keep, promote, and fire. The arduous part of that decision is it often needs to be made where there are relatively few measureable, and often noisy, performance statistics of the employee's skills.

The ability to decipher the top talent is highly valuable; this study aims to increase the ability for managers to make these difficult decisions. If an employee messes up two of the first three projects, should you give them a few more chances to see if there is a problem with the employee or is it possibly randomness? This study looks at a unique dataset of small samples to see how increasing the number of observations increases the probability of finding true ability.

In this study we use data from professional tennis to see if the number of sets played in a match changes the probability of the higher ranked player winning. It is commonly understood that increasing the sample size, n, increases the precision of the estimates. This study helps analyze this precision by analyzing the different outcomes in three-set tennis matches, relative to a five-set matches.

III. Tennis

One of biggest differences between men's and women's tennis is the number of sets in a match. Women's tennis always play three set matches. Men's tennis, for smaller tournaments, play three set matches whereas for Grand Slam tournaments play five set matches. Playing more sets in a given match is hypothesized to increase the probability of a 'true outcome'. We are defining a true outcome as an outcome where the higher ranked player, used as a proxy for innate ability, wins. This occurs because the difference in structure, having to win three sets versus only two, decreases the probability of an upset.

Del Corral (2009) finds that attendance is affected by competitive balance and that competitive balance for men's tennis is lower than women's by 10%. Likewise, Gilsdorf (2008) finds that the higher seed wins most often in men's tennis. Gilsdorf comes to this conclusion by using Rosen's (1986) sequential elimination-style tournament model to test head to head results between specific players in tennis. Neither of these studies looks at the varying number of sets between men's and women's tennis. Magnus and Klaassen (1999b) measure the probability of each player winning in the final set in a tennis match at the Wimbledon. This is most relevant to our study because they recognize differences in match structure; implying that it is more difficult for unseeded men to beat a seed than it is for unseeded woman.

In our study we use data from every professional tennis match, set, and tournament from 2007-2009 to measure the probability of true outcomes in men's and women's tennis. To accurately measure these aspects, rankings are important in our analysis. We only look at matches in which at least one player is ranked. Tennis seedings are generated each tournament by a computer and based on past performances, regardless of the surface on which previous tournaments were played. Rankings are, therefore, recent and relatively accurate at the start of each tournament. We also recognize that each player may play best on a particular court surface (Du Bois and Heyndels 2007 and del Corral 2009), so we also control for different court surfaces.

To find the true outcomes we measure the impact of the five set match structure. We use a difference-in-difference (DID) to control for the self selection problem: the best players will choose to enter smaller tournaments, which are three set matches, at a different rate than their entrance into Grand Slam tournaments, which are five sets (for men). The DID allows us to compare the relative difference in men's matches with those of women's matches, controlling for any fundamental differences in the men's and women's play.[1]

IV. Methodology and Data

In order to estimate these differences we use data from all professional tennis matches from 2007-2009, gathered from ESPN.com.[2] Using match level data we look at all matches that include at least one ranked player, using tournament rankings, and drop any matches where there is a walk-over.[3] This leaves us with a total of 9,425 tennis matches. Of those matches: 7,552 matches are non-Grand Slam tournament matches (3,967 men's and 3,585 women's matches) and 1,873 Grand Slam matches (1003 men's and 870 women's matches).

We use the player ranking as an estimate of true, or innate, ability. This gives us a measure of who is the better player. To get a true outcome we use a probit to measure if the Higher Ranked player Wins, *HRW*, one, zero otherwise; controlling for the player's tournament ranking and different court surfaces (Following Lallemand, Plasman and Rycx 2005).[4]

---

[1] Given the model setup, we also do not have to worry about the underdog effect found in Harbaugh and Klumpp (2005).
[2] 2007 is the earliest year publically available match by match data is available.
[3] A walk-over is when one player withdraws from the match for an injury or other reason.
[4] This will give an accurate relative ranking. It is possible to also use the national ranking at the time of the tournament. However national rankings are considered when tournament rankings are devised. One worry is that if some of the talent does not show up at a smaller (non-Grand Slam) tournament, the relative rankings will be different. This issue will be controlled for in the setup of the difference-in-difference estimation.

$$HRW_{ij} = \beta_0 + \beta_1 Clay_{ij} + \beta_2 Grass_{ij} + \beta_3 Supreme_{ij} + \beta_4 Carpet_{ij} + \beta_5 CloseRanks_{ij}$$

$$+ \beta_6 BothRanked_{ij} + \beta_7 FinalMatch_{ij} + \beta_8 Y + \beta_9 M + \varepsilon \qquad (1)$$

The use of tournament ranking or seeding is similar to Magnus and Klaassen (1999a and 1999b) who look at Wimbledon outcomes.[5] We control for two players that are within five rankings of each other, *CloseRanks*, if both players are ranked, *BothRanked*, and if it's the final match of a tournament, *FinalMatch*.[6] This is done for each match, *i*, in each tournament, *j*. For each regression we use also control for monthly, *M*, and yearly, *Y*, fixed effects, consistent with the model specification used in Sunde (2003).

Du Bois and Heyndels (2007) find that men's and women's matches are equally (un-)predictable. However del Corral (2009) finds that the changing in the number of seeded players affected men's and women's grand slam tournaments differently. Using DID estimation allows us to control for any fundamental differences in Grand Slam and non-Grand Slam tournaments as well as fundamental differences in men and women's play.

The traditional DID model is set up by:

$$[(\text{treated group})_w - (\text{control})_w] - [(\text{treated group})_m - (\text{control})_m] \qquad (2)$$

where the treated group is the non-Grand Slam tournaments and the control group is Grand Slam tournaments. The first group is the women's matches, *w*, and the second group is male matches, *m*. This DID approach allow us to find the true affect of increasing the number of observations in a small sample, even when the major events have more competition (Rohm, Chatterjee and Habibullah 2004).

---

[5] They do not control for playing surface or tournament characteristics since they only use Wimbledon data.

[6] This has also been tested as any player within 10 ranks, with similar results.

Wooldridge (2000, pg. 415) allows us to measure the statistical significance for the DID estimation:

$$HRW_{ij} = \beta_0 + \beta_1 Men_{ij} + \beta_2 GrandSlam_{ij} + \delta_1 (Men * GrandSlam)_{ij} + \beta_3 Clay_{ij}$$
$$+ \beta_4 Grass_{ij} + \beta_5 Supreme_{ij} + \beta_6 Carpet_{ij} + \beta_7 CloseRanks_{ij} + \beta_8 BothRanked_{ij}$$
$$+ \beta_9 FinalMatch_{ij} + \beta_{10} Y + \beta_{11} M + \varepsilon \quad (3)$$

If δ_1 is statistically different from zero, then the high ranked player has a statistically different probability of winning if they play three or five set match.

V. Results

Looking at the results of equation 1, we find that different court surfaces have an insignificant impact, however if the two players are within five ranks of each other we get the expected decrease in the probability of a HRW. If both players are ranked, but not within five ranks, the probability of the HRW does not change, except for in women's Grand Slam matches.

Table I – Output from Equation 1

The regressions of equation 1, looking at the probability of the higher ranked player winning the match. This is a probit regression with the marginal effects reported.

Dependent Variable: HRW	Male Grand Slam	Female Grand Slam	Male non-Grand Slam	Female non-Grand Slam
Clay	-0.002 (0.03)		0.013 (0.51)	-0.029 (1.09)
Grass	-0.024 (0.57)	-0.05 (1.17)	-0.01 (0.13)	0.022 (0.27)
Supreme			-0.069 (0.99)	-0.156 (2.07)*
Carpet			-0.043 (1.23)	-0.107 (1.90)
CloseRanks	-0.195 (3.03)**	-0.165 (2.53)*	-0.08 (3.26)**	-0.146 (5.48)**
BothRanked	-0.054 (1.47)	-0.087 (2.15)*	0.058 (1.81)	-0.018 (0.54)
FinalMatch	0.004 (0.04)	-0.275 (1.84)	-0.071 (1.96)	-0.051 (1.33)
Month FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Obs Probability	0.7996	0.78966	0.69776	0.70349
Pseudo R2	0.0173	0.0227	0.0076	0.0140
Observations	1003	870	3967	3585
Pr>Chi2	0.0437	0.0092	0.0086	0.0000

Absolute value of z statistics in parentheses
* significant at 5%; ** significant at 1%

The results from these regressions are used in a DID model, as established in equation 2, in table II.

Table II - One player ranked, from table I

	Non-Grand slam (Treated Group)	Grand Slam (Control)	Difference	Difference-in-Difference
Women	0.703	0.790	0.087	
Men	0.698	0.800	0.102	0.015

This shows that when one player is ranked, the probability of the HRW is 70.3% for women in a non-Grand Slam tournament and 79% in a Grand Slam tournament (a difference of 8.7%). For men in a non-Grand Slam tournament, the HRW 69.8% of the time. These tournaments all play three set matches. However when men play a Grand Slam tournament, which is a five set match, the probability of the HRW is 80% (a difference of 10.2%). Controlling for the differences in men and women, while simultaneously controlling for the differences in the Grand Slam and non-Grand Slam play, it is shown that the increased probability from moving to a five set match, from a three set match, is 1.5%. So increasing the number of observations only has a small marginal effect on the probability of a true outcome.

Next we estimate the affect when both players are ranked (table III).

Table III - Players are closely ranked, within five rankings, from table I

	Non-Grand slam	Grand slam	Diff	Dif-in-Dif
Women	0.557	0.538	-0.019	
Men	0.610	0.607	-0.013	0.006

When both players are closely ranked increasing the number of sets increases the true outcome by 0.6%. This follows our expectations that more sets increase the probability of a true outcome and is consistent with Boulier and Stekler (1999), however the effect continues to be small.

To test the statistical significance of the above DID estimation (reported in Tables II and III) the results of equation 3 are reported in table IV.

Table IV – Statistical Significance for DID

	HRW
Male	-0.002 (0.19)
GrandSlam	0.088 (4.45)**
Men*GrandSlam δ_1	0.012 (0.48)
Other Controls	Yes
Fixed Effects	Yes
Observations	9425
Absolute value of z statistics in parentheses	
* significant at 5%; ** significant at 1%	

The variable of interest is insignificant, δ_1, showing there is no statistically significant difference in three and five set matches tested above. Although we find a small increase in the probability of a true outcome, with additional sets, this increase is not statistically different from zero.

VI. Conclusion

Increasing the number of observations increases the probability of a true outcome in professional tennis. However, we are surprised to find that increasing the number of sets has a relatively small and statistically insignificant affect on the high ranked player winning. Increasing matches from three sets to five sets adds substantial time to a match, while having no statistical affect on the accuracy of the outcome.

This can be applied to the corporate setting, although three chances or projects may seem insufficient for deciding the fate of an employee within a company, increasing to five chances may be more of a hassle than a help in formulating an accurate decision.

Works Cited

Boulier, Bryan L., Stekler, H.O., 1999. Are Sports Seedings Good Predicators?: An Evaluation *International Journal of Forecasting* 15:83-91

del Corral, Julio, 2009. Competitive Balance and Match Uncertainty in Grand-Slam Tennis *Journal of Sports Economics* 10(6): 563-581

Du Bois, Cind and Bruno Heyndels, 2007. It's a Different Game You Go to Watch: Competitive Balance in Men's and Women's Tennis, *European Sport Management Quarterly* 7(2):167-185

Gilsdorf, Keith F., Sukhatme, Vasant A., 2008 Testing Rosen's Sequential Elimination Tournament Model, *Journal of Sports Economics* 9(3): 287-303

Guilkey, David K. and Michael K. Salemi, 1982. Small Sample Properties of Three Tests for Granger-Causal Ordering in a Bivariate Stochastic System *The Review of Economics and Statistics* Vol. 64, No. 4, pp. 668-680

Harbaugh, Rick and Tilman Klumpp, 2005. Early Round Upsets and Championship Blowouts *Economic Inquiry* Vol. 43, Is 2, 316-329

Lallemand, T., Plasman, R., & Rycx, F. 2005. Women and Competition in Elimination Tournaments: Evidence from Professional Tennis Data, *IZA Discussion Paper 1843*, retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.9461&rep=rep1&type=pdf.

Magnus, J.R. and Klaassen, F.J.G.M. 1999a. On the advantage of serving first in a tennis set: four years at Wimbledon, *The Statistician*, 48, 247-256.

Magnus, J.R. and Klaassen, F.J.G.M. 1999b. The final set in a tennis match: four years at Wimbledon, *Journal of Applied Statistics*, 26 (4), 461-468.

Mankiw, N. Gregory and Matthew D. Shapiro, 1986 Do We Reject Too Often? Small Sample Bias in Tests of Rational Expectations *Economics Letters* 20: 139-145

Nelson, Charles R. and Myung J. Kim, 1993. Predictable Stock Returns: The Role of Small Sample Bias *The Journal of Finance* Vol. 48, No. 2, pp. 641-661

Piette, Michael J. and Paul F. White, 1999. Approaches For Dealing With Small Sample Sizes In Employment Discrimination Litigation *Journal of Forensic Economics* Vol. 12, No. 1, Winter Available at SSRN: http://ssrn.com/abstract=164995

Rohm, Andrew J., Sangit Chatterjee, and Mohamed Habibullah, 2004. Strategic Measure of Competitiveness for Ranked Data *Managerial and Decision Economics* 25, 103-108

Rosen, Sherwin, 1986. Prizes and Incentives in Elimination Tournaments *The American Economic Review* Vol. 76, No. 4, pp. 701-715

Ruibal, Sal, 2009. Study: NCAA Penalties Harsher for Non-BCS, Historically Black Schools *USA Today*http://www.usatoday.com/sports/college/2009-09-22-punishment-study_N.htm

Sunde, W. 2003. Potential, Prizes and Performance: Testing Tournament Theory with Professional Tennis Data, *IZA Discussion Paper 947*, retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.9351&rep=rep1&type=pdf.

Wooldridge, Jeffery W. 2000 Introductory Economics: A modern Approach *South Western College Publishing*