THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF FINANCE

IDENTIFYING OPTIMAL TENNIS BETTING PORTFOLIOS BASED ON VARYING RISK TOLERANCES OF BETTORS

NIMAY GODBOLE
SPRING 2020

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Finance
with honors in Finance

Reviewed and approved* by the following:

Dr. Robert Novack
Associate Professor of Business Logistics
Thesis Supervisor

Dr. Brian Davis
Clinical Assistant Professor of Finance
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

This thesis assesses the risk versus return relationship in tennis betting among the different tournaments that exist within the Association of Tennis Professionals (ATP) Tour, the main men's tennis governing body. To do so, betting data is gathered from a website well-known for this data collection, *http://www.tennis-data.co.uk/*. The purpose of this thesis is to identify optimal betting portfolios for bettors with varying levels of risk aversion. Expected returns and standard deviations of Grand Slam, Masters 1000, ATP 500, and ATP 250 tournaments were analyzed to create an optimal risky portfolio using Markowtiz's Portfolio Theory. Once that was created, optimal betting portfolios were created by betting a portion of the money in the tennis betting market and investing the rest in Vanguard's Prime Money Market Fund, VMMXX. The key finding is significant; at a risk aversion level of $A = 4$, the portfolio yields a return of 11.02 percent and a standard deviation of 15.02 percent, which is a higher return and lower standard deviation than the historical average of the S&P 500. However, further research needs to account for a larger data set and stronger statistical analysis tools to identify even more profitable betting strategies.

TABLE OF CONTENTS

List of Figures

List of Tables

Acknowledgements

I would like to thank Dr. Novack for his advice and support with regards to this thesis and for his dedication to Alpha Kappa Psi and his efforts to better the organization year after year. I would also like to thank Dr. Davis for his help during this process, for pushing me to think critically, and for his unwavering commitment to further the finance honors program.

To Dr. Simin and Dr. Gattis, FIN 406H and FIN 415 were my favorite courses I took throughout my time at Penn State and inspired me to write this thesis.

Finally, I would like to thank Melissa Paloskey for always going the extra mile to help me navigate these four years of college successfully.

Introduction

The sports betting market is expected to reach a valuation of $8 billion by 2025 (Associated Press). The business model is simple – bettors bet on outcomes they think are likely based on odds created by bookmakers. The bettors who bet on the correct outcome make money, and so do the bookmakers based on the difference between the payout to the bettors who made the right bet and the money they gained from bettors who made the wrong bet. Given that the bookmakers have such power in the market to set odds, bettors employ many strategies to gain positive returns in the betting market. The focus of this thesis is on the tennis betting market, and specifically on analyzing the relationship between expected returns and volatility among various tournaments (e.g., Grand Slams, Masters 1000, ATP 500, and ATP 250) and using a utility maximizing model to develop optimal betting portfolio for bettors given their different levels of risk aversion.

To assess the risk versus return relationship for the tennis betting market, betting odds data compiled on the website, *http://www.tennis-data.co.uk/*, from 2010 to 2019 will be analyzed using different betting strategies. This website consists of betting odds from a variety of bookmakers in various men's tennis tournaments – Grand Slams, Masters Series, ATP 500, and ATP 250 tournaments. Once this relationship is identified, a model will take into account these two risk and return factors, in addition to a variable that represents the bettors risk aversion, which will be able to recommend the optimal percentage to bet in each of the tournament categories given the bettors risk tolerance.

This paper will have the following structure: first, the literature review will give an overview of studies relating to market efficiency and the tennis betting market, betting strategies, and portfolio theory; next, the methodology of data collection, cleansing, and analysis mentioned earlier will be elaborated upon; then, a detailed discussion of results will follow; and finally, a conclusion that will finalize the thesis and offer thoughts on future areas of research.

Literature Review

The following literature review consists of an overview of research conducted in areas of market efficiency and its relation to tennis sports betting, what tennis betting strategies and models are used to predict winners and generate excess returns, and the concept of portfolio optimization in the financial markets. The sections below delve deeper into each of these areas of study, which will help establish the framework for the research presented in this thesis.

Market Efficiency and Tennis Sports Betting

Market efficiency is a notion that stems greatly from the work of Eugene Fama and his *efficient market hypothesis*. His thesis dictates that three forms of efficiencies exist in the stock market: weak form, semi-strong form, and strong form efficient, meaning that a price will fully reflect available information (Fama, 1970). Although heavily criticized, the hypothesis has been studied greatly in the stock market since its initial publication. It is a strong basis for why investors are implored to pursue passive investment strategies for long-term gain than active management.

The hypothesis, which was originally used to describe how the stock market operates, has also been tested on different sports betting markets to assess whether profitable betting strategies exist. In sports betting, market efficiency is defined by the accuracy of odds set by the bookmakers compared to the results of any given game or match. Within tennis sports betting, Candila and Scognamillo (2018) identified, through simple and brief analysis of data from more than 30,000 matches, that the tennis sports betting market violates weak-form market efficiency. They come to this conclusion through a simple betting strategy of betting on “favorites”, or players with the higher expected value (lose less money), and comparing that to betting on "longshots", or players with lower expected value (lose more money). The strategy yielded larger returns (or smaller loses) for favorites that were statistically different from zero, implying a market efficiency. Further, this implied an existence of the favorite-longshot bias, a common phenomenon in betting.

In betting, the favorite-longshot bias is a prevalent irrationality where betters overvalue "longshots" and undervalue the "favorites". This bias exists especially in the tennis sports betting market. According to Lahvicka (2013), the favorite-longshot bias is strong in high-profile tournaments, such as grand slams, matches between lower-ranked players, and later stages of a tournament. These behaviors are explained by many reasons, including the general risk-loving nature of bettors, the overestimation of longshots, and bookmakers protecting themselves from underestimation of longshots by lowering the odds to avoid significant loses (as they do not have to pay out as much money in the event of a longshot winning). This study utilizes fixed-odds betting to assess this bias, which is the same type of data used in this research. Even the study conducted by Abinzano, et al. (2016), used tennis betting exchange data to analyze the favorite-longshot bias and found this to be true.

Having identified that the tennis betting market is inefficient due to biases, understanding different profit-generating strategies and models to predict tennis match results is important for the remainder of the thesis.

Tennis Betting Strategies and Match Forecasting Models

McHale and Morton (2011) create a model, called the Bradley-Terry Model, to forecast men's tennis match results using historical match results and surface of the match. They compare this model to other standard forecasting models to discover that recent form of the player and match score were the most crucial in producing the most accurate forecasts. Then, they tested this model for betting purposes and found that using it will generate more profits than using other models. However, the issue is that the win percentage for bets is well below fifty percent, implying high volatility, and that the few bets won were at long odds.

In addition to assessing efficiency of the tennis betting market, Candila and Scognamillo (2018) proposed a new procedure of normalization that created "unbiased" odds, which would allow investors to bet more confidently on players, allowing them to generate better returns than those obtained by using other models. The assumption here is that because the favorite-longshot bias exists because bookmakers are not setting real odds, but odds that will help them maximize their profits (or minimize their losses), a new set of implied probabilities need to be calculated in order to assess the correct odds. When the two models are compared, the Bradley-Terry Model and the new normalization procedure called the CaSco Normalization, the CaSco Normalization model generates better returns (or smaller losses) because of the fact that the model considers the betting quotes in its model, while the BT model does not consider them.

Finally, Lyócsa and Výrost (2017) assess market efficiency of the tennis betting market by analyzing forty different betting strategies to check if there are differences in them and whether it is possible to use these to generate profits. The study finds that is it possible to generate positive returns through these rules, but these returns diminish when data-snooping bias is taken into consideration. Data-snooping bias is essentially finding patterns in data that can be shown as significant out of context. However, the study highlights the high level of volatility that exists even when positive returns are possible.

Thus, all these studies show that: a) there are many different strategies and models that can be used to generate returns and forecast match results; and, b) even though returns are positive in many instances, there is significant volatility present in those returns that bettors must take into account when developing the right betting strategy for themselves.

Next, it is important to understand why and how volatility has been accounted for by investors in the stock market and why it is important to consider it in the tennis betting market as well.

Portfolio Optimization

It is clear from studying previous literature that volatility has not been accounted for while analyzing the tennis betting market. However, it is important to take a look at the foundations of the tradeoff between average returns and volatility from a financial perspective in order to apply it to a betting context.

In Fabozzi, Frank J., et. al. (2008), Harry Markowitz, the father of the modern portfolio theory, states that investors will find their optimal risky portfolio by maximizing the Sharpe ratio, which is the slope of a risk versus return graph. The next step is to determine the utility maximizing level of investment in the risky asset by taking into consideration the investor's risk aversion. This will allow the investor to weigh the risky asset in their portfolio appropriately given their investment in a risk-free asset, e.g. a treasury bill. The result being that the more risk averse the investor, the lower the volatility in the portfolio for lower the returns, and the less risk averse the investor, the higher the volatility in the portfolio for potentially greater returns (Markowitz 1952).

No studies have directly examined portfolio optimization in a sports betting context from the modern portfolio theory perspective. Assessing this will allow bettors to create an optimal betting portfolio with the appropriate expected return and volatility given their level of risk tolerance.

Conclusion

In summary, the literature suggests that market inefficiency exists, specifically within the tennis betting market, as identified by evidence of the favorite-longshot bias. However, profitability for bettors is still difficult to attain given the high volatility in returns, as demonstrated by the many different betting strategies and predictive models with applications in the tennis betting context. However, this volatility is rarely considered for bettors and returns in betting are standardized on a risk-adjusted basis in any sports betting context. This is an important notion since Markowitz's portfolio optimization seeks to help investors create optimal risky portfolios based on their risk tolerance.

Thus, even though betting attracts people with higher risk-tolerance, it is important to judge them based on varying levels of risk aversion. By assessing a variety of tournaments, the risk versus reward relationship of betting returns in those tournaments, and considering the level of risk aversion of the better, bettors will be able to optimize their betting portfolios and attain returns at the risk they are comfortable with. This is especially important when sports bettors want to diversify their betting portfolios with different levels of risky bets.

The next section will dive deeper into the data to be used, the data cleansing approach, and the overall methodology of analysis to address the existing research gap.

Description of Data and Methodology

Data Source and Description

As stated earlier, the purpose of this thesis is to analyze the risk versus return relationship of tennis betting among various tennis tournaments in order to create optimal betting portfolios for bettors with varying risk tolerances. As such, data was retrieved from *http://www.tennis.co.uk*, a website containing betting odds data for a majority of Association of Tennis Professionals (ATP) Tournaments from years 2010 to 2019. The data contains the necessary information to test popular betting strategies so that expected returns and volatility can be calculated. It includes the tournament numbers, location of the tournament, date of the match, series name (i.e., Grand Slam, Masters 1000, ATP 500, or ATP 250), indoor versus outdoor court, surface type, round of tournament the match was played in, winner and loser of the match, ATP rankings, games and sets won in the match for both players, notes relating to completion of the match, odds set by varying websites before the match began, maximum and minimum odds set by the websites, and an average of the odds set by the websites. For the purpose of this thesis, the most important data points are the before match odds data of the match winners and losers for all the matches in the tournament from various betting websites, especially the average odds data, which is used instead of the maximum or minimum to increase the likelihood of the odds being the most optimal.

This study does not disseminate between the different surfaces the tournaments are played on or other factors such as different ranking metrics, head-to-head records, serve percentage, etc. The reason being is that it is assumed that the odds data encompasses all the information so that they are reflecting the true probability of a player winning the match given all these factors and more, hence the bookmakers are setting odds for the match to maximize their profits or minimize their losses.

Going deeper into the data, 22,219 matches have been captured across sixty-one tournaments from 2010-2019. However, data for certain tournaments was not available. The following table, Table 1, lists all the tournament exceptions, and Table 2 provides a breakdown of all the tournaments by their tournament category.

Table 1. List of Tournaments with Missing Betting Data

Tournament Name	Data Capture
Adelaide Open	*No Data*
Stuttgart Open	*No Data*
Mallorca Open	*No Data*
Rio Open	*2014-2019*
Tokyo Open	*2010; 2012-2019*
Pune Open	*2018-2019*
Montpellier Open	*2010; 2012-2019*
Sofia Open	*2016-2019*
Cordoba Open	*2019*
New York Open	*2018-2019*
Santiago Open	*2010-2011*
Marrakech Open	*2016-2019*
Budapest Open	*2017-2019*
Estoril Open	*2010-2012; 2015-2019*
Geneva Open	*2015-2019*
Lyon Open	*2017-2019*
Eastbourne Open	*2010-2014; 2017-2019*
Los Cabos Open	*2016-2019*
Kitzbuhel Open	*2011-2019*
Winston-Salem Open	*2011-2019*
St. Petersburg Open	*2010-2013; 2015-2019*
Chengdu Open	*2016-2018*
Zhuhai Open	*2019*
Antwerp Open	*2016-2019*

Table 2. Breakdown of Tennis Tournaments by Tournament Type

Tournaments				
Grand Slams	Masters 1000	ATP 500	ATP 250	
Australian Open	*Indian Wells Masters*	*Rotterdam Open*	*Doha Open*	*Marrakech Open*
French Open	*Miami Open*	*Rio Open*	*Pune Open*	*Budapest Open*
Wimbledon	*Monte-Carlo Masters*	*Dubai Tennis Championships*	*Adelaide Open*	*Munich Open*
U.S. Open	*Madrid Open*	*Mexican Open*	*Auckland Open*	*Estoril Open*
	Italian Open	*Barcelona Open*	*Montpellier Open*	*Geneva Open*
	Canadian Open	*Queen's Club Championships*	*Sofia Open*	*Lyon Open*
	Cincinnati Masters	*Halle Open*	*Cordoba Open*	*Stuttgart Open*
	Shanghai Masters	*German Open*	*New York Open*	*'s-Hertogenbosch Open*
	Paris Masters	*Washington Open*	*Buenos Aires Open*	*Eastbourne Open*
		Beijing Open	*Santiago Open*	*Mallorca Open*
		Tokyo Open	*Marseille Open*	*Newport Open*
		Vienna Open	*Delray Beach Open*	*Bastad Open*
		Swiss Indoors	*Houston Open*	*Umag Open*
			Metz Open	*Atlanta Open*
			Chengdu Open	*Gstaad Open*
			Zhuhai Open	*Los Cabos Open*
			Moscow Open	*Kitzbuhel Open*
			Stockholm Open	*Winston-Salem Open*
			Antwerp Open	*St. Peterburg Open*

These tournament classes are picked because of their varying popularity – grand slams being the most popular and ATP 250 being the least popular. This will allow the bettor to have different options for which types of tournament classes they need to bet in more often based on their risk preferences.

Data Cleansing and Manipulation

After gathering all the data, certain data points were eliminated as certain pre-match odds were unavailable, players retired either before or during the match, or players were disqualified during the match. Player disqualification or player retirement data was discarded because these are extremely unforeseen circumstances that would not have been forecasted by the bookmaker, hence the odds do not reflect the situation. As a result, 915 data points were removed, for a total of 21,304 data points.

Additionally, although ATP 500 data was analyzed to identify the tournament style's expected return and standard deviation, ATP 500 data is excluded when identifying an optimal risky portfolio because there are 1,232 less data points (matches) – ATP 500 had 3,608 matches, while the next lowest was Grand Slams with 4,840. As a result, it can be assumed that the returns and standard deviation calculations for ATP 500 were adversely impacted by the significantly lower number of data points available.

The final step of the data manipulation was the conversion of how the odds are written. The average odds data is converted from decimal form to American form (i.e., 1.11 decimal odds become -$909, meaning that the bettor need to bet $909 to win $100, which is an 11.00 percent return). This will allow for calculation of expected returns and standard deviation when a betting strategy is applied to the matches.

Research Methodology

Having converted the decimal odds to American odds, a betting strategy needed to be employed to assess the returns of betting on matches. Using Lyócsa and Výrost's analysis of various strategies (2017), a simple betting strategy is created. Bets are only made if the player is the favorite and has a lower rank than his opponent, or $Odds^A < Odds^B \wedge Rank^A < Rank^B$.

Next, since there are more matches in ATP 250 per year than in the other tournaments, returns and standard deviations are annualized from 2010-2019 for each tournament style. Then, an overall expected return and standard deviation is calculated.

Finally, an optimal risky portfolio is calculated using Markowitz's Portfolio Optimization Theory with a risk-free rate given by Vanguard's Prime Money Market Fund, VMMXX (Vanguard). The Sharpe ratio is maximized while restricting non-negative weights of asset classes, hence no short sales are allowed as it does not make sense in a betting context. Once the optimal risky portfolio is identified, the following equations are used to calculate the optimal investment strategy for bettors with varying degrees of risk tolerances.

$$U = E(r) - .005A\sigma^2$$

$$\frac{\partial U}{\partial y} = E(r_p) - r_f - .01yA\sigma_p^2 = 0$$

$$\Rightarrow y^* = \frac{E(r_p) - r_f}{0.01A\sigma_p^2}$$

U is the utility gained, *E(r)* is the expected return, *A* is the risk aversion variable that is defined further in the next section, σ^2 is the measure of variance or square of volatility, and y^* is the weight to be invested in the optimal risky portfolio (Langford 2016).

Summary of Findings

The first step of the process was to identify a strong betting strategy that a normal, unsophisticated bettor can use. Below, in Table 3, is the data for three different betting strategies and their expected returns and standard deviations.

Table 3. Analysis of Different Betting Strategies

	Odds(A) < Odds(B)		Rank(A) < Rank(B)		Odds(A) < Odds(B) ^ Rank(A) < Rank(B)	
	Exp. Ret	*SD*	*Exp. Ret*	*SD*	*Exp. Ret*	*SD*
ATP 250	1.80%	68.46%	-8.87%	66.11%	25.27%	46.75%
Grand Slam	2.72%	55.97%	-6.01%	59.74%	16.96%	36.25%
Masters 1000	0.73%	65.72%	-8.50%	68.32%	22.22%	44.20%

Having used Lyócsa and Výrost's analysis of betting strategies, it was important to ensure that the betting strategy employed to calculate returns must be easily followed by a normal bettor who does not have sophisticated means of analyzing matches (i.e., complex statistical analyses of historical data, using match forecasting models, and etc.). After initial assessment of these strategies, the third betting strategy, $Odds^A < Odds^B \wedge Rank^A < Rank^B$, is most prudent a bettor should follow to get optimal results while betting because it maximizes the expected returns while having significantly lower standard deviations.

Looking further in the three strategies, below is Table 4, which further assess the efficacy of the betting strategies. The different strategies were analyzed by looking at how many data points the strategies could be applied to and the likelihood of positive returns each strategy would result in. The results of this analysis further solidify that the third betting strategy, $Odds^A$

< $Odds^B$ ^ $Rank^A$ < $Rank^B$, is best out of the ones analyzed in this thesis for it is simple to use and provides strong returns with the lowest standard deviations.

Table 4. Analysis of Different Betting Strategies: Part 2

Rank(A) < Rank(B)				
Tournament	**Matches**	**# of Bets**	**% of Bets**	**Success Rate**
ATP 250	7391	7391	100.0%	62.85%
ATP 500	3608	3608	100.0%	66.35%
Grand Slams	4840	4840	100.0%	73.88%
Masters 1000	5465	5465	100.0%	66.44%
Total	21304	21304	100.0%	66.87%

Odds(A) < Odds(B)				
Tournament	**Matches**	**# of Bets**	**% of Bets**	**Success Rate**
ATP 250	7391	7391	100.0%	70.78%
ATP 500	3608	3608	100.0%	73.48%
Grand Slams	4840	4840	100.0%	79.92%
Masters 1000	5465	5465	100.0%	72.48%
Total	21304	21304	100.0%	73.75%

Odds(A) < Odds(B) ^ Rank(A) < Rank(B)				
Tournament	**Matches**	**# of Bets**	**% of Bets**	**Success Rate**
ATP 250	7391	4645	62.8%	87.64%
ATP 500	3608	2394	66.4%	91.02%
Grand Slams	4840	3576	73.9%	91.75%
Masters 1000	5465	3631	66.4%	91.13%
Total	21304	14246	66.9%	90.13%

Out of all the matches analyzed, the third betting strategy results in bets on 66.9 percent of them. This is because not every match will have a favored player who also has a lower rank than the opponent. The key finding here is the success rate. The success rate is the number of bets that result in a positive return. Out of 14,246 bets made, 90.13 percent of them are positive returns, while the success rate of the strategy on the top left, $Rank^A$ < $Rank^B$, is only 66.87 percent and the success rate of the strategy on the top right, $Odds^A$ < $Odds^B$, is only 73.75 percent. This speaks to the efficacy of the third betting strategy, which is that it is accurately able to identify which favorites are most likely to win. This also means that because the success rate is high, the standard deviation is lower because the expected return is not as reliant on extremely high returns that occur from unlikely events.

Using the third betting strategy to further the analysis, returns and volatility were annualized to identify correlation between the three asset classes, which is the first step in

Markowitz's portfolio optimization. Below, in Table 5, are the annualized expected returns and standard deviations.

Table 5. Annualized Returns and Standard Deviations

Date	ATP 250	Grand Slam	Masters 1000
2010	24.77%	20.50%	22.77%
2011	24.38%	14.73%	19.54%
2012	22.19%	15.91%	21.38%
2013	23.69%	14.93%	19.45%
2014	28.82%	16.64%	20.89%
2015	25.22%	15.90%	24.35%
2016	23.77%	17.39%	22.34%
2017	25.45%	14.76%	23.40%
2018	27.11%	18.81%	25.31%
2019	27.32%	20.05%	22.80%
EXP. RET.	**25.27%**	**16.96%**	**22.22%**

Date	ATP 250	Grand Slam	Masters 1000
2010	44.58%	32.84%	42.17%
2011	43.49%	36.89%	46.03%
2012	46.24%	35.76%	40.83%
2013	44.26%	36.29%	45.47%
2014	43.89%	38.44%	45.70%
2015	45.07%	36.22%	40.24%
2016	47.64%	34.79%	39.91%
2017	47.03%	36.93%	46.77%
2018	51.94%	38.61%	46.58%
2019	53.34%	35.77%	48.34%
ST. DEV.	**46.75%**	**36.25%**	**44.20%**

As seen in Table 5, returns and standard deviations for ATP 250 tournaments are higher than Grand Slams, which fits with the theory of small and large cap stocks – large cap stocks (in this case represented by Grand Slams) have significant amount of trading, hence have lower returns, while small cap stocks (represented by ATP 250) are more deemed to have greater inefficiency, resulting in higher returns and standard deviations (Stanhope 2015).

Next, a correlation matrix is created based off of the returns of the different tournaments. This is so that a covariance between the different asset classes is calculated, as seen in Table 6. This is an important step because it helps assess the relationship between the asset classes, which is crucial to finding the optimal risky portfolio's standard deviation.

Table 6. Correlation and Covariance Matrices

CorrelMat	*ATP250*	*Grand Slam*	*Masters 1000*
ATP250	1	0.369705659	0.32427486
Grand Slam	0.369705659	1	0.468296844
Masters 1000	0.32427486	0.468296844	1

CovarMat	*ATP250*	*Grand Slam*	*Masters 1000*
ATP250	0.21854554	0.076397868	0.054958439
Grand Slam	0.076397868	0.195392699	0.075045651
Masters 1000	0.054958439	0.075045651	0.131431743

It is clear from this analysis that there is very little correlation between the returns per year for the different tournaments, which means that the covariance between the asset classes is low as well.

Finally, the optimal risky portfolio can be calculated, as shown in Table 7.

Table 7. Optimal Risky Portfolio Analysis

Asset Class	Exp. Ret.	SD	Weights
ATP 250	25.27%	46.75%	41.22%
Masters 1000	22.22%	44.20%	30.24%
Grand Slam	16.96%	36.25%	28.54%
	21.98%	33.26%	100.00%

	SD	Exp. Ret.
Rf	0.00%	2.00%

Sharpe	60.06%

Rf is the risk-free rate that bettors invest their cash holdings. The expected return of this is the expected return of the underlying asset that the bettors use – as stated earlier, this is VMMXX's 1-year yield for the past year. A money market fund is used as the risk-free rate because it is highly liquid and bettors can hold their cash in the account that is not being used to bet.

The weights of the optimal risky portfolio were calculated by maximizing the Sharpe Ratio while ensuring that the weights remain non-negative.

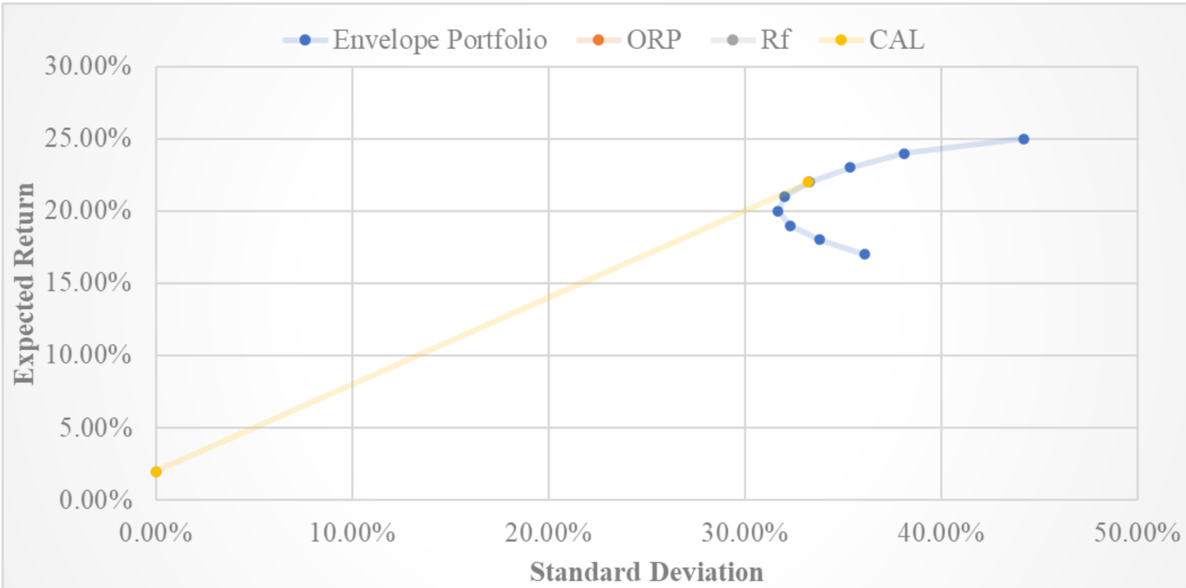
Based on the analysis, 41.22 percent of the bets need to occur in ATP 250 tournaments, 30.24 percent need to occur in Masters 1000 tournaments, and 28.54 percent need to occur in Grand Slams. This yields an expected return of the optimal risky portfolio of 21.98 percent and a standard deviation of 33.26 percent, assuming that bettors will bet year-round. Interestingly, the standard deviation of the portfolio is less than the standard deviation of any asset class. This is because there is very low correlation between the returns of the different tournaments, as the correlation matrix above suggests. Hence, there is little covariance between the different asset classes, resulting in the forecasted standard deviation to be lower than any of the asset classes. This furthers the notion that diversification of bets will not only help maximize returns, but also minimize volatility in them.

To further examine this relationship, various portfolios are created to understand the risk versus return relationship. Below is the data and graph showcasing the results.

Table 8. Envelope Portfolios

Envelope Port	Annual σ	Annual Mean	W(ATP250)	W(M1000)	W(GS)
#1	36.14%	17.00%	0.00%	0.72%	99.28%
#2	33.82%	18.00%	6.24%	9.88%	83.89%
#3	32.32%	19.00%	15.03%	14.99%	69.98%
#4	31.71%	20.00%	23.82%	20.11%	56.06%
#5	32.05%	21.00%	32.62%	25.23%	42.15%
#6	33.30%	22.00%	41.41%	30.35%	28.24%
#7	35.37%	23.00%	50.20%	35.47%	14.33%
#8	38.13%	24.00%	59.00%	40.59%	0.41%
#9	44.19%	25.00%	91.10%	8.90%	0.00%
ORP	33.26%	21.98%	41.22%	30.24%	28.54%

Figure 1: Envelope Portfolios and CAL



The envelope portfolio shows that the minimum variance occurs when the expected return is at twenty percent, with bettors weighing their portfolio bets with 23.82 percent for ATP 250, 20.11 percent for Masters 1000, and 56.06 percent for Grand Slams. Hence, a rational bettor should not bet solely on Grand Slams or attempt to get expected returns of less than twenty percent because they will not be minimizing their volatility for a given expected return. For example, if a bettor can create envelope portfolio #5, which gives them twenty-one percent expected return for 32.05 percent standard deviation, they wouldn't create envelope portfolio #3, which gives them nineteen percent expected return for 32.32 percent standard deviation.

The CAL, or the capital allocation line, represents the investment opportunities for the bettor – this means that the bettor can create any portfolio from the risk-free rate to the optimal risky portfolio, with the slope of the Sharpe Ratio, by varying the weights of the investments. To go beyond the CAL and attain higher returns, the bettor would have to borrow money. For the scope of this thesis, those returns and standard deviations were not analyzed.

Once the optimal risky portfolio was calculated, bettors risk aversion was taken into consideration to identify the mix of their investment in the risk-free asset and the portfolio. Using the utility formula mentioned in the research methodology section, the weights to be invested in the optimal risky portfolio are calculated. Table 9 shows the results.

Table 9. Optimal Portfolios Based on Risk Aversion Levels

Risk Aversion (A)	y*	Exp. Ret.	SD
1	1.805771248	38.08%	60.06%
2	0.902885624	20.04%	30.03%
3	0.601923749	14.03%	20.02%
4	0.451442812	11.02%	15.02%
5	0.36115425	9.22%	12.01%
6	0.300961875	8.01%	10.01%
7	0.257967321	7.15%	8.58%

y is the weight of the risky portfolio*
1-y is the weight of investment in VMMXX*

As the data suggests, when the risk aversion factor, *A*, is 1, the bettor is expected to borrow. However, an investor cannot likely borrow at the same risk-free rate, hence that data point is ignored. The rest of the data reveals significant insights into the risk versus return relationship by varying the amount the bettors keeps as cash holdings in a money market fund and bet on tennis matches. The definition of different values of *A* are shown in the spectrum below.

Figure 2. Risk Aversion Scale

A = 1 ——————————————— A = 7

No Risk Aversion *High Risk Aversion*

Although this spectrum ends at $A = 7$, the value of *A* is not constrained on the far end, or $A \geq 0$. The values of *A* are also not defined by Markowitz; hence the bettor has to rate themselves based on their evaluation of how risky they are. However, several factors influence *A*; for example, the age of the bettor and their experiences with betting are two factors that can impact a bettor's risk aversion. The analysis of maximizing utility based on risk tolerance resulted in one key finding:

At risk aversion of $A = 4$, the bettor bets approximately forty-five percent of the money given the ratios of the optimal risky portfolio and the remaining fifty-five percent is invested in the VMMXX, yielding a return of 11.02 percent and a standard deviation of 15.02 percent. Comparing this to the average return of the S&P 500 index, which has yielded, on average, 9.8 percent over the past ninety years (CNBC 2017) and has a standard deviation of 15.2 percent since 1926 to 2017 (Reuters 2019), it would be a better strategy to actively bet in the tennis betting market and invest in a money market fund. Additionally, there are no fees associated with betting, as the profit for the bookmakers is built into the bet, which makes this option even more appealing considering the fees investors are charged to have their accounts managed. Further, by following the simple betting strategy, the bettor does not need to have knowledge of the sport either, which allows there to be low barriers to entry into the tennis betting market.

In conclusion, bettors in the tennis betting market should break down their bets by the tournament style they are betting in. It is apparent that different tournaments yield varying expected returns and standard deviations, which is in line with the theory of market efficiency of small and large cap stocks – it can be assumed that Grand Slams have significant amount of betting, hence have lower returns, while ATP 250 tournaments have lesser betting, and are more deemed to have greater inefficiency, resulting in higher returns and standard deviations. Further, due to the lack of correlation between the returns of the tournaments, the covariance is impacted. This results in the standard deviation of the optimal risky portfolio being less than any of the individual tournaments. Bettors are encouraged to diversify their bets by betting in these different tournament structures to maximize their returns and minimize volatility. Investing in the money market fund and the optimal risky portfolio does result in higher returns and lower standard deviation than the stock market.

Shortcomings and Conclusion

Research Shortcomings

There are shortcomings with the research that include data availability and general research analysis.

First, the data analyzed in this thesis is from 2010-2019. This thesis certainly could have explored data from much earlier, however due to time constraints, gathering that data was difficult. Additionally, as mentioned in the data description section above, there were many exceptions with the availability of the data. Particularly, this was the case for ATP 250 tournaments, which do not get as much coverage. As a result, there could have been significantly more data points for ATP 250 tournaments that could have impacted the expected return and standard deviation calculations. Additionally, there was significantly less data for ATP 500 tournaments than others. Future research should seek to do a deeper dive into the past and include odds data from 2000 onwards.

Second, the analysis of the data was done by coming up with a simple betting strategy that any bettor can follow. Due to a lack of statistical tools, other betting strategies were not considered while analyzing return data, such as creating a bootstrap distribution or using match forecasting models to predict winners. Future research should examine different betting strategies that yield strong positive returns. Additionally, the asset classes that were created can be differed – analysis can be done based on the different rounds of the tournament (i.e., 1^{st} round, 2^{nd} round, quarter-finals, and etc.) to understand whether those yield different returns and standard deviations.

Research Conclusion

Unlike past literature on tennis betting, this research put a significant emphasis on volatility when compared to the returns generated by the betting strategy. Understanding that all bettors do not have the same characteristics or interests, the goal of the thesis was to identify what an optimal risky tennis portfolio looks like and how bettors can create an optimal portfolio by combining their investment in a risk-free asset and betting in the tennis market. It is fair to assume that they are able to invest in a highly liquid money market as bettors also tend to have margin accounts so that their cash grows.

As such, the research focus was primarily on identifying whether meaningful returns can be generated. The findings suggest that it would, in fact, be more beneficial to invest approximately fifty-five percent of the money in a money market fund and forty-five percent of the money is bet in ratio with the optimal risky portfolio than to invest in the stock market. However, as mentioned in the previous section, additional analysis of betting strategies should be done that enables a greater percentage of matches to be bet on for potentially higher returns.

Appendix A

Tennis Betting Data Sample

Tournament	Date	Series	WRank	LRank	AvgW	AvgL	American OddsW	American OddsL	Returns on Favorites
Qatar Exxon Mobil Open	1/4/2010	ATP250	29	72	1.1902	4.5906	-525.7623554	359.06	19.02%
Qatar Exxon Mobil Open	1/4/2010	ATP250	60	19	3.718	1.2634	271.8	-379.6507213	-100.00%
Qatar Exxon Mobil Open	1/4/2010	ATP250	53	99	1.242	3.934	-413.2231405	293.4	24.20%
Qatar Exxon Mobil Open	1/4/2010	ATP250	57	N/A	1.03325	11.6875	-3007.518797	1068.75	3.33%
Qatar Exxon Mobil Open	1/4/2010	ATP250	56	84	1.3292	3.302	-303.7667072	230.2	32.92%
Qatar Exxon Mobil Open	1/4/2010	ATP250	37	54	1.372	2.987	-268.8172043	198.7	37.20%
Qatar Exxon Mobil Open	1/4/2010	ATP250	62	69	2.345	1.5738	134.5	-174.2767515	-100.00%
Qatar Exxon Mobil Open	1/4/2010	ATP250	90	31	1.3878	2.928	-257.8648788	192.8	38.78%
Qatar Exxon Mobil Open	1/5/2010	ATP250	N/A	241	2.624	1.467	162.4	-214.1327623	-100.00%
Qatar Exxon Mobil Open	1/5/2010	ATP250	122	41	2.436	1.5418	143.6	-184.569952	-100.00%
Qatar Exxon Mobil Open	1/5/2010	ATP250	49	70	1.2332	4.022	-428.8164666	302.2	23.32%
Qatar Exxon Mobil Open	1/5/2010	ATP250	101	291	1.2052	4.372	-487.3294347	337.2	20.52%
Qatar Exxon Mobil Open	1/5/2010	ATP250	2	93	1.03	11.86	-3333.333333	1086	3.00%
Qatar Exxon Mobil Open	1/5/2010	ATP250	1	86	1.013	17.5	-7692.307692	1650	1.30%
Qatar Exxon Mobil Open	1/5/2010	ATP250	6	132	1.0676	8.388	-1479.289941	738.8	6.76%
Qatar Exxon Mobil Open	1/6/2010	ATP250	122	N/A	1.4506	2.672	-221.9263205	167.2	45.06%
Qatar Exxon Mobil Open	1/6/2010	ATP250	101	60	2.032	1.764	103.2	-130.8900524	-100.00%
Qatar Exxon Mobil Open	1/6/2010	ATP250	1	53	1.021	13.4	-4761.904762	1240	2.10%
Qatar Exxon Mobil Open	1/6/2010	ATP250	29	57	1.513	2.513	-194.9317739	151.3	51.30%
Qatar Exxon Mobil Open	1/6/2010	ATP250	6	56	1.1262	5.97	-792.3930269	497	12.62%
Qatar Exxon Mobil Open	1/6/2010	ATP250	2	62	1.0202	13.5	-4950.49505	1250	2.02%
Qatar Exxon Mobil Open	1/6/2010	ATP250	37	40	1.4624	2.637	-216.2629758	163.7	46.24%
Qatar Exxon Mobil Open	1/6/2010	ATP250	90	49	1.6016	2.276	-166.2234043	127.6	60.16%
Qatar Exxon Mobil Open	1/7/2010	ATP250	29	101	1.429	2.75	-233.1002331	175	42.90%
Qatar Exxon Mobil Open	1/7/2010	ATP250	1	90	1.049	9.5	-2040.816327	850	4.90%
Qatar Exxon Mobil Open	1/7/2010	ATP250	6	37	1.2444	3.974	-409.1653028	297.4	24.44%
Qatar Exxon Mobil Open	1/8/2010	ATP250	2	29	1.032	11.66	-3125	1066	3.20%
Qatar Exxon Mobil Open	1/8/2010	ATP250	6	1	2.823	1.4156	182.3	-240.6159769	-100.00%
Qatar Exxon Mobil Open	1/9/2010	ATP250	6	2	2.17	1.689	117	-145.137881	-100.00%
Heineken Open	1/10/2010	ATP250	332	263	2.198	1.6528	119.8	-153.1862745	-100.00%
Heineken Open	1/11/2010	ATP250	34	56	1.5382	2.444	-185.8045336	144.4	53.82%
Heineken Open	1/11/2010	ATP250	28	54	1.3748	3.018	-266.8089648	201.8	37.48%
Heineken Open	1/11/2010	ATP250	30	48	1.559	2.382	-178.8908766	138.2	55.90%
Heineken Open	1/11/2010	ATP250	31	456	1.2816	3.576	-355.1136364	257.6	28.16%
Heineken Open	1/11/2010	ATP250	209	266	1.635	2.214	-157.480315	121.4	63.50%
Heineken Open	1/11/2010	ATP250	33	43	1.5538	2.392	-180.5706031	139.2	55.38%
Heineken Open	1/11/2010	ATP250	60	86	1.6442	2.206	-155.2312946	120.6	64.42%
Heineken Open	1/12/2010	ATP250	675	59	2.597	1.4836	159.7	-206.7824648	-100.00%
Heineken Open	1/12/2010	ATP250	58	57	1.7	2.12	-142.8571429	112	70.00%
Heineken Open	1/12/2010	ATP250	27	235	1.1146	6.308	-872.600349	530.8	11.46%
Heineken Open	1/12/2010	ATP250	67	32	2.503	1.5164	150.3	-193.6483346	-100.00%
Heineken Open	1/12/2010	ATP250	31	332	1.239	3.98	-418.4100418	298	23.90%
Heineken Open	1/12/2010	ATP250	27	34	1.9866	1.7744	-101.3581999	-129.1322314	98.66%
Heineken Open	1/13/2010	ATP250	33	30	1.9664	1.7976	-103.4768212	-125.3761284	96.64%
Heineken Open	1/13/2010	ATP250	16	60	1.2054	4.3126	-486.8549172	331.26	20.54%

Bibliography

Associated Press. “Sports Betting Market Expected to Reach $8 Billion by 2025.” MarketWatch, MarketWatch, 4 Nov. 2019, www.marketwatch.com/story/firms-say-sports-betting-market-to-reach-8-billion-by-2025-2019-11-04.

Candila, Vincenzo, and Antonio Scognamillo. “Estimating the Implied Probabilities in the Tennis Betting Market: A New Normalization Procedure.” International Journal of Sport Finance, vol. 13, 2018, pp. 225–242.

Fabozzi, Frank J., et al. “Portfolio Selection.” Handbook of Finance, 2008, doi:10.1002/9780470404324.hof002001.

Fama, Eugene F. “Efficient Capital Markets: A Review of Theory and Empirical Work.” The Journal of Finance, vol. 25, no. 2, 1970, p. 383., doi:10.2307/2325486.

“Is the Stock Market More Volatile Now Than Ever Before?” Reuters, Thomson Reuters, 30 Apr. 2019, www.reuters.com/article/idUSWAOA9NUAIRCF192L.

Lahvicka, Jiri. “What Causes the Favorite-Longshot Bias? Further Evidence from Tennis.” SSRN Electronic Journal, 2013, doi:10.2139/ssrn.2287335.

Langford, Charles K. “The Risk Aversion Coefficient.” The Risk Aversion Coefficient | Desjardins Online Brokerage, 4 Aug. 2016, www.disnat.com/en/learning/expert-articles/charles-k-langford/the-risk-aversion-coefficient.

Lyócsa, Štefan, and Tomáš Výrost. “To Bet or Not to Bet: a Reality Check for Tennis Betting Market Efficiency.” Applied Economics, vol. 50, no. 20, 2017, pp. 2251–2272., doi:10.1080/00036846.2017.1394973.

Markowitz, Harry. “Portfolio Selection.” The Journal of Finance, vol. 7, no. 1, Mar. 1952, pp. 77–91.

Mchale, Ian, and Alex Morton. “A Bradley-Terry Type Model for Forecasting Tennis Match Results.” International Journal of Forecasting, vol. 27, no. 2, 2011, pp. 619–630., doi:10.1016/j.ijforecast.2010.04.004.

Michaelsantoli. “The S&P 500 Has Already Met Its Average Return for a Full Year, but Don't Expect It to Stay Here.” CNBC, CNBC, 19 June 2017, www.cnbc.com/2017/06/18/the-sp-500-has-already-met-its-average-return-for-a-full-year.html.

Stanhope, Ehren, and Chris Meredith. “Inefficiency Breeds Opportunity in Small Cap Equities.” O'Shaughnessy Asset Management, July 2015, www.osam.com/pdfs/whitepapers/_4_Commentary_InefficiencyBreedsOpportunitySmallCapEquities.pdf.

“Vanguard Mutual Fund Profile.” Vanguard, investor.vanguard.com/mutual-funds/profile/VMMXX.

ACADEMIC VITA

Nimay R. Godbole

EDUCATION

The Pennsylvania State University | Schreyer Honors College — **University Park, PA**
Smeal College of Business | Bachelor of Science in Finance — *Class of 2020*
Dean's List 7/7 Semesters

Institut Américain Universitaire — **Aix-en-Provence, FR**
The School of Business & International Relations — *May 2017 – Jul 2017*
SAT scores: *Math 770 | Writing 760 | Reading 660*

PROFESSIONAL EXPERIENCE

Strategy&, Part of the PwC Network — **New York, NY**
Health Industries Advisory | Summer Associate — *Jun 2019 – Aug 2019*

- Created a prioritization framework by matching strategic value and level of complexity of more than 25 initiatives to help develop a 7-year implementation roadmap for a connected health platform
- Synthesized critical insights from the client regarding the necessary capabilities of the platform, which seeks to improve consumer experience, by strategizing the content of workshop deliverables

DePuy Synthes, Companies of Johnson & Johnson — **West Chester, PA**
Trauma and CMF Division | Supply Planning Co-Op — *Dec 2017 – Jul 2018*

- Identified over $10M in excess inventory and reduced lead-time of essential items by 30% through analysis of daily backorders, resulting in increased efficiency of the Supply Planning department
- Generated an additional $2M in revenue by ensuring over a 99% compliance rate of a government contract and maintaining communication across multiple business functions
- Increased membership and community service events of the Intern/Co-Op Association (ICA) by 200% through spearheading a new organization structure and improving communication flow

Sibling Relationships in Emerging Adulthood: Shared Leisure and Relationship Quality — **University Park, PA**
Recreation, Parks, and Tourism Management Department | Undergraduate Research Assistant — *Oct 2016 – Dec 2017*

- Presented at the Canadian Conference on Leisure Research at University of Waterloo after reviewing more than 20 studies and developing an abstract in collaboration with lead researchers
- Analyzed qualitative data of 7 focus groups to identify correlations between variables such as increased shared digital leisure among certain dyads led to more amicable relationships

LEADERSHIP EXPERIENCE

Schreyer Consulting Group — **University Park, PA**
President (Dec 2018 – Dec 2019) — *Jan 2017 – May 2020*

- Initiated contact with new firms to increase SCG sponsored events by 50%, while overseeing the planning of treks to corporate offices and serving as a liaison between consulting firms and Schreyer administration
- Developed interviewing preparation resources and a Schreyer-sponsored case competition to enhance more than 2,000 students' exposure to consulting and networking opportunities

Alpha Kappa Psi Professional Business Fraternity — **University Park, PA**
Vice President (Dec 2018 – Dec 2019) — *Feb 2017 – May 2020*

- Maintained a 100% job placement for graduating seniors by forming relationships with Fortune 100 firms across various industries that provided tailored professional and funding opportunities for the chapter
- Raised $41K for the Penn State Dance Marathon, a 17% increase from previous year, by actively managing 4 chair positions to increase chapter involvement in philanthropy

Global Brigades, Business Chapter — **University Park, PA**
Director of Education (May 2017 – Dec 2017) — *Dec 2016 – Dec 2017*

- Educated more than 30 members about the economic and social implications of assisting communities in Central America and West Africa by creating interactive presentations and case studies
- Assisted with the empowerment of a small business community in Honduras by helping over 50 rural community members and children become financially literate through games and presentations

AWARDS AND INTERESTS

- *Schreyer Honors College Academic Excellence Scholarship, Evan Pugh Scholar Award, Penn State Provost Award, Donald T. Houpt Award, and AP Scholar Award (2015, 2016)*
- 1st Place – Johnson & Johnson Penn State Case Competition, 4th place – Princeton University Graduate Case Competition
- Interests include cricket, tennis, blackjack, international cuisine, hiking the Himalayas, Gordon Ramsay videos, and card throwing