Three Essays on Risk Willingness

by

Matthew Leonel Jaramillo, M.B.A., M.S.

A Dissertation

In

Personal Financial Planning

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
The Requirements for
The Degree of

DOCTOR OF PHILOSOPHY

Approved

Donald Lacombe, Ph.D.
Chair of Committee

Mitzi Lauderdale, J.D., Ph.D., CFP®

Michael Guillemette, Ph.D., CFP®

Laura Ricaldi, Ph.D.

Michael Farmer, Ph.D.

Mark Sheridan, Ph.D.
Dean of the Graduate School

December 2022

Table of Contents

Abstract

In today's environment, consumers are increasingly responsible for their retirement accumulation and decumulation decisions. There are products in the market that can assist consumers with these decisions, and the effectiveness of these products is dependent upon individual preferences. Risk willingness is one of the fundamental aspects of retirement planning, and as such, is an important factor that can match consumers with appropriate products in the market to suit their preferences. Annuities are insurance products designed to help mitigate certain retirement risks, such as longevity and investment risks. If a risk averse consumer chooses to transfer such risks from themselves to an insuring company, they may find more satisfaction in their retirement income accumulation and decumulation decisions.

It is important to properly identify, measure, and compare an individual consumer's or household's financial risk willingness across important demographic measures. Often, a problem when comparing across demographics is that comparative groups are imbalanced by representation, response, and/or measurement. Methods that do not control for this imbalance when comparing one group to another often will have biased estimates, spurious results, and inconclusive hypothesis testing. Propensity score matching allows for comparison across groups that are balanced based upon relevant and theoretically motivated variables. When used appropriately, a researcher can confidently compare across important demographic groups in order to better estimate any potential differences that may have an impact on retirement accumulation and decumulation decisions.

List of Tables

List of Figures

Introduction

Annuities are insurance products designed to mitigate risks associated with outliving one's financial assets and investment risk. As with any insurance product, a consumer's preference will dictate their choice to either self-insure, transfer, or avoid those risks. It is often assumed that the more risk averse an individual reports to be, the more satisfaction they may find in either avoiding or transferring elements of risks associated with their current circumstances. In terms of retirement income satisfaction, those who are most risk averse may find the most satisfaction in financial insurance products that help insure against potential adverse effects.

The demographics of a household may be one of the most impactful determining factors of the willingness to take financial risks. The landscape of the United States within the last ten years, with respect to marital rights between individuals, has changed drastically. In 2015, the United States Supreme Court struck down all prior legislation aimed at preventing same-sex couples from receiving marital rights, including rights related to retirement planning. Prior to the Court's decision, couples in same-sex relationships had to engage in complex financial planning procedures in order to ensure their wishes were met. Today, same-sex couples are an increasing make-up of the population in the United States.

It is assumed men and women have differing levels of financial risk willingness, yet prior research on the subject often failed to take into account the imbalance of the distribution of the respondents to questions related to financial risk willingness. The studies that do not balance the distribution of responses do not truly report the difference in risk willingness between the genders.

CHAPTER 2

Retirement Income Satisfaction, Annuity Ownership, and Risk Willingness

Introduction

There is a sustained trend in today's retirement planning environment wherein individuals and households have been increasingly responsible for their asset accumulation and decumulation decisions. Until the 1970's, defined benefit retirement plans (DB), accounts which are professionally managed and provide a defined benefit throughout the retirement horizon of an individual, were a common employee retainment tool that allowed professionals to manage the accumulation and decumulation decisions on behalf of the employee. After the passing of the Employee Retirement Security Act of 1974, employers and defined benefit fiduciaries were subject to new regulation that altered familiar practices and set minimal standards in the name of consumer protection. Afterwards, employers began to explore alternatives to the traditional defined benefit plans. The answer was found in defined contribution plans (DC).

Today, the majority of retirement plans offered by employers and utilized by employees and individuals are defined contribution plans and individual retirement accounts (IRAs). Defined contribution plans allow employees/individuals to defer a portion of their earnings to their separate retirement accounts, oftentimes tax deferred, and may offer a tax incentive to the employer for matching the contribution up to a certain percentage. IRAs are separate retirement accounts for individuals which may offer annual tax incentives and tax deferred growth to the account holder who makes contributions, subject to various qualifications.

In addition to employer sponsored DC plans and IRAs, the insurance product of an annuity offers an individual investment account that grows tax-deferred and can be accessed without penalty beginning at age $59^1/_2$. What makes an annuity different from the aforementioned accounts is the available funding and payout options, and these available options shift certain risks from the investor to the insuring company. Annuity funding can be deferred, invested in over-time then accessed at or near retirement; or an annuity can be immediate, as a lump sum of cash is offered in exchange for certain payout choice. In the simplest forms, the annuity payout can be an exhaustive stream of payments for defined time, a consistent amount until the funds are exhausted, a lump sum withdrawal, or an individual can annuitize the account balance and receive payments for the remainder of their life. In any case, it is apparent that much of the accumulation and decumulation management is handled by the insuring company which offers the annuity.

It is estimated that there are 3.5 times the amount of retirement assets in defined contribution plans and IRAs when compared to the total amount of retirement assets held in defined benefit plans (Munnell, 2014). The popularity of self-managed plans and individual accounts has transferred multiple risks, including investment and longevity risks, to the account owners who are responsible for the management of their retirement accumulation and decumulation decisions over their lifecycle. Finding the optimal combination of retirement funding and financial products and services to accompany an investor’s level of risk aversion is fundamental to successful retirement planning. This study focuses on how annuity ownership, moderated by financial risk tolerance, impacts retirement income satisfaction. The hypothesis of this study is those who are more risk averse and own an annuity will find more satisfaction with their retirement income when compared to those who are less risk averse and own an annuity. Conversely, those who are less risk averse will find less satisfaction with their retirement income if they have investments in an annuity. The decision making under uncertainty over the life-cycle theory is the theoretical framework for this study.

Literature Review

The life-cycle theory of consumption suggests that individuals prefer to maximize their utility through consistent, intertemporal time allocation decisions which will smooth the consumption path over their lifecycle, subject to desired preferences (e.g. risk aversion). In order to maintain smooth consumption throughout the lifecycle, families will tend to borrow when young, save during middle age, and then dissave during retirement (Finke and Huston, 2003). During the accumulation phase of the lifecycle, investors are presented with various risks regarding asset location decisions, asset allocation decisions, and estimates of expected longevity for all related parties. While the outcomes of these decisions are uncertain, a more risk averse individual is expected to maximize their utility by finding a savings and consumption path with less potential overall volatility over their lifecycle. In an efficient marketplace and one in which participants prefer less risk to more risk, those who are willing to take greater risk are rewarded with greater expected return, and investors who receive the greatest reduction in utility from a loss in financial wealth are considered risk averse (Finke and Huston, 2003).

Risk aversion is commonly linked to some relative measure, oftentimes wealth. Risk aversion is often described as the marginal change of the slope of the utility function relative to the current level of wealth (Finke and Guillemette, 2016). If a person's utility of the expected value of an unknow outcome is greater than their expected utility from the unknow outcome itself, they are said to be risk averse (Pratt and Schlaifer, 1964). This is exhibited by the degree of concavity of their utility function relative to the origin. The stronger the degree of concavity relative to the origin, the higher the level of risk aversion. If a person's utility of the expected value of an unknown outcome is less than their expected utility from the unknown outcome itself, they are said to be risk-seeking (Dyer and Sarin, 1992). This behavior is measured by the degree of convexity of the function relative to the origin.

In the context of investments, volatility is often measured by the standard deviation of the investment portfolio payouts. A safer portfolio will result in less variation in returns, which implies a potential narrower payout interval in the future that is relative to the amount of real dollars invested throughout the investment timeline (Finke and Guillemette, 2016). A riskier portfolio will result in more variation in returns, which implies a potential larger future payout interval, relative to the amount of real dollars invested throughout the investment timeline (Finke and Guillemette, 2016). Grable and Lytton (1998) state that increased levels of education and being male are the better predictors of willingness to take financial risks. Alternatively, Lee and Hanna (1995) suggests that the relationship between income and willingness to take financial risk may be a function of wealth when estimating the impact of financial loss. The consequences in terms of utility loss are greatest among those with the least wealth; however, accumulated wealth may not be the most accurate indicator of ability to withstand financial loss (Finke and Huston, 2003).

The choice of how individuals choose to accumulate and decumulate assets in order to maximize their marginal utility of consumption is of personal preference. As individuals increasingly bear the responsibility of managing their own accumulation and decumulation decisions, there are products available that can potentially simplify the process. Annuities are insurance products designed to help mitigate certain risks for the investor, namely longevity and investment risks. In the accumulation phase, annuities can be either deferred or immediate. In either case, the amount invested is used to purchase accumulation units, another term for shares, in the individual account. These accumulation units are used to invest in specific types of investments, depending on the risk structure of the underlying annuity.

Longevity risk is defined as the risk of outliving one's financial assets during retirement. Through an annuity, this risk is hedged by the insurance company with the existence of mortality credits. Not all who invest financial assets into an annuity will receive the full pay out amount, due to early death. The amount remaining is redistributed to those who have annuitized and live longer than expected. Annuities add an element of certainty to asset decumulation decisions by transferring longevity risk and certain elements of investment risk, discussed above, from consumer to the market (Bernartzi et al., 2011).

Milevsky (2006) states that expected longevity is a function of gender, wealth, and subjective health status. If expected longevity continues to increase, consumer demand theory suggests an associated increase in the market demand for products that can help mitigate risks associated with longer retirement horizons. Bernartzi et al. (2011) estimates that 10% of men retiring at age 65 can expect to live an additional 27 years, and the same ratio for women can expect to live an additional 30 years after retiring at 65. Yaari (1965) states that in the absence of a bequest motive, an individual can benefit more from annuitizing all of their retirement wealth compared to annuitizing a lower percentage or none.

If an individual is defined as risk averse, decision making under uncertainty suggests a movement away from self-insurance towards the purchase of services and available insurance products from the marketplace to help mitigate the associated risks. Mitchell et al. (1999) finds that as risk aversion increases, the quantity demanded for annuities increases, even as price increases. The purchase of an annuity or the annuitization of financial assets within a retirement account decreases the probability of exhausting financial assets in retirement (Milevsky et al., 2006). The ability to transfer longevity and investment risks from the individual to the marketplace helps reduce the psychological and behavior risks of mis-managing money in retirement (Milevsky, 2005).

Andersson and Svensson (2007) studies cognitive constraints and their ability to impact the assessment of risk reduction. The study suggests that when made salient of the risks of mortality, revealed preferences become apparent. Additionally, the study suggests that the higher the cognitive ability, the more people are able to understand the associated probabilities of an outcome, and the more willing they are to act in line proportionately to the magnitude of the risk at hand.

Financial literacy affects an individual’s ability to make sound financial decisions which can impact their utility in retirement (Lusardi, 2008; Lusardi & Mitchell, 2014). Measures of excessive debt load and low financial numeracy scores have shown to be related negatively to retirement plan contribution rates (Lusardi et al., 2010). Asset accumulation decisions can be improved upon going forward, but during the decumulation phase, these same decisions cannot be improved upon retroactively. Liabson et al. (2008) find that learning is not monotonic, and this implies that an individual's ability to improve upon their decision-making ability is limited to some point in time. Additionally, overconfidence in one's ability to make financial decisions can have a detrimental impact on retirement satisfaction (Agnew and Szykman, 2011). Estimating a smooth asset decumulation pattern subject to an uncertain life expectancy takes a higher degree of financial literacy and knowledge. A familiarity with existing financial products to match such challenges is also necessary.

An annuity is a retirement planning tool that can help increase the marginal utility of consumption for risk averse individuals due to the natural consumption smoothing aspect of the product. Individuals with higher levels of risk aversion and who choose to utilize annuities are hypothesized to report higher levels of retirement satisfaction. In essence, investing in an annuity represents a transfer of wealth to an annuity provider in exchange for certainty of income. One component of overall retirement satisfaction includes how individuals choose to receive and manage their income during retirement. Do people prefer being relatively rich at retirement, with a large amount of their own money readily available to be spent as flexibly as they wish? Or would they rather have the comfort of knowing they have a steady income in perpetuity (Pannis, 2003)? The hypothesis of this study suggests that annuity ownership's impact on retirement income satisfaction is moderated by an individual's level of risk aversion/risk willingness. As the level of risk aversion increases, it is assumed retirement income satisfaction will be positively impacted by annuity ownership. The opposite is true for those who are less risk averse. As the level of risk willingness increases, it is assumed that retirement income satisfaction will be impacted negatively by annuity ownership.

Data

The Survey of Consumer Finances (SCF) is a cross-sectional, triennial survey that is supported by the Federal Reserve Board in partnership with the Department of Treasury. It is a nationally representative survey, which over-samples wealthy households. This analysis uses the 2016 and 2019 SCF waves. Each of the two waves contain data for approximately 5,300 households, includes a total of 5 implicates for each household response, and uses a variety of methods for imputing any missing data within each implicate when necessary. The unit of observation is the primary economic unit. The primary economic unit is the economically dominant individual within the household who reports aggregate household information. Any individual demographic information is representative of the primary economic unit respondent. The final sample contains 9,675 (48,375) observations.

Dependent Variable

The dependent variable in the analysis is an ordered scale measuring retirement income satisfaction. The survey asks, "“Using any number from one to five, where one equals totally inadequate and five equals very satisfactory, how would you rate the retirement income you receive (or expect to receive) from all sources? Including 401(k) accounts and all other types of pensions?” (Survey of Consumer Finances, 2016 & 2019). Respondents are asked to rate their current or expected level of retirement income satisfaction on a scale of one to five (1-5), with one (1) being totally inadequate and five

(5) being totally adequate. Approximately 9% percent of the sample reports being very dissatisfied (1) with their retirement income, and 30% of the sample reports being moderately satisfied (3). Approximately 26% of the sample reports their satisfaction with their retirement income is totally adequate (5).

Independent Variables

The independent variables in the analysis include: a self-reported financial risk willingness measure, an annuity ownership indicator, an interaction variable between annuity ownership and the financial risk willingness measure, age, age-squared, income, income-squared, education level, employment status, log of net worth, gender, marital status, an indicator for the presence of children, bequest motivation, race/ethnicity, and wave variables to identify each wave.

The variable for annuity ownership is a dichotomous variable, yes (1) or no (0). Approximately 7.2% of the sample reported owning an annuity. Annuities may allow for a perpetuity of consistent income for the life of the purchaser which provides a certainty to projected future consumption, subject to mortality. The lifecycle model states that individuals prefer a smooth consumption path throughout their life cycle, and as discussed previously, annuities transfer investment and longevity risks to the market, allow for a certain stream of income, and have been shown to reduce the probability of outliving financial assets. The expectation of annuity ownership's impact on retirement income satisfaction is ambiguous, all else equal.

The measure for risk willingness is reported on an ordered scale from 1 to 10, (1) being the lowest level of risk willingness and (10) being the highest level of risk willingness. The question states, "On a scale from zero to ten, where zero is not willing to take risks and ten is very willing to take risks, what number would you (and your {husband/wife/partner}) be on the scale?" (Survey of Consumer Finances, 2016 & 2019). Categories 0 and 1 have been combined. Higher levels of risk willingness have been associated with an increase in participation in investing activities, such as stock market participation, and in turn, participation in the stock market has been associated with higher levels of financial asset values over the same period of time. Retirement income satisfaction is expected to be associated positively with higher levels of risk willingness.

The main variable of interest in this study includes an interaction variable which includes: (0 or 1) annuity ownership and the different levels (1 – 10) of the risk willingness measure. Positive annuity ownership values are interacted with each level of the risk willingness measure, creating 10 interaction variables. Annuity ownership at the lowest level of risk willingness (1) is the reference group for these variables. Annuitizing financial assets insures against longevity risk, the risk of outliving one's financial assets. Annuities do so by transferring this risk from the individual or household to the insuring company. As individuals increasingly become less willing to take and manage their own financial risks, there suggests an increasing attractiveness for annuity ownership, and the opposite is true for those who are increasingly willing to take financial risks. The transfer of risk provides the highest utility of marginal consumption for an individual or household who is less willing to take financial risks by creating a less volatile consumption path over the lifecycle and in turn, securing a certain stream of retirement income. It is expected that those who are less willing to take financial risks and who own an annuity will report higher levels of retirement income satisfaction.

Age is measured continuously from age 18+ and is representative of the respondent's age at the time of the survey year. The average age in this sample is 52 years with a standard deviation of approximately 16 years. Retirement income is the sum of all forgone consumption saved for retirement purposes, and as age increases, there is more opportunity to increase retirement savings through continued contributions and/or through positive returns on investments. An age-squared variable is also included in this analysis to control for potential non-linearity with respect to age and the dependent variable. An increase in the variables age and age-squared is expected to be associated positively and negatively, respectively, with all levels of retirement income satisfaction.

The respondent's education level is separated into seven categories that each represent the highest level of educational attainment: did not graduate, high school diploma, some college, associate degree, bachelor's degree, and, master's degree, and professional/doctorate degree. A dichotomous variable is created for each category, yes (1) or no (0), and the "Did Not Graduate" category is the reference group for education. An increase in education has been empirically shown to have the potential to increase current and future wage earnings, and an increase in earnings can lead to higher levels of deferred or forgone consumption which can be utilized in retirement. Higher levels of education are expected to be associated positively with higher levels of retirement income satisfaction.

The log of net worth is measured continuously in this analysis and is calculated as the difference between all reported financial assets and all reported financial liabilities. Financial assets include all of the following when applicable: certificate of deposits, transaction accounts, money market funds, savings bonds, directly held stocks and bonds, cash value of life insurances, other miscellaneous assets, all residences, business interests, and vehicles. Financial liabilities include all of the following when applicable: all forms of secured debt, lines of credit, credit card balances, installment loans, and other miscellaneous debt. The lifecycle model states that individuals prefer a smooth consumption path throughout their life cycle. Independent of some inheritance or windfall of income, low levels of net worth produce low levels of overall retirement income, limit financial opportunities, and increase the probability of borrowing for necessity in retirement. Higher levels of net worth allow for a smoother consumption path by producing higher levels of overall retirement income, increase financial opportunities, and reduce the probability of borrowing out of necessity in retirement. Higher levels of net worth are expected to be associated positively with higher levels of retirement income satisfaction.

The variable for annual income is measured continuously in 2019 real dollars for both waves. Respondents are asked to report the total dollar amount of annual income for the entire household. The median annual income for the sample is $44,000, and the average annual income for the sample is $160,045. As mentioned previously, the SCF oversamples higher income households. An income-squared variable is included in the analysis to control for any potential non-linear association between income and the dependent variable, all else equal. Assuming annuities are normal goods, an increase in income will lead to an increase in the quantity demand for annuities. Increased levels of income are expected to be associated positively with higher levels of retirement income satisfaction.

Employment status is a categorical variable that includes employed, retired, and other. The retired category includes fully and partially retired individuals. The other category represents individuals who do not consider themselves retired or employed, such as homemakers. A dichotomous variable is created for each category of employment status, and the other status is the reference category for the group. Approximately 70% of the sample report to be employed; 19% of the sample report to be retired; and 11% of the sample report to be in the other category. When compared to the reference group "other", respondents who are employed or retired are expected to be associated positively with higher levels of retirement income satisfaction, as they are still in the labor force or they have reached the decumulation of financial assets stage of their life cycle.

Gender is a dichotomous variable, (1) for male and (0) for female. Gender is determined by the response of the primary economic unit (PEU). Approximately 20% of the PEU responses report to be female, and approximately 80% of the PEU responses report to be male. Female is the reference group for the analysis. The variable gender's expected positive or negative association with retirement income satisfaction is ambiguous.

A dichotomous variable is created to represent the presence of children, (1) yes and (0) no children. No children (0) is the reference category in this analysis. The presence of children can cause limited household resources to be allocated amongst more individuals. The presence of children is expected to be associated negatively with higher levels of retirement income satisfaction and positively with lower levels of retirement income satisfaction.

Marital status is defined as the current, legal marital status. There are 6 categories: Married, Separated, Divorced, Widowed, and Never Married. A dichotomous variable is created to represent current marital status: (1) married or (0) not married. If the PEU response is not married, the response is coded as not married. The not married category is the reference group. The variable marital status' expected positive or negative association with retirement income satisfaction is ambiguous.

Race is a categorical variable that includes White, Black, Hispanic, and Other. A dichotomous variable is created for each category, and White is the reference group in this analysis. Approximately 73% of this sample report to be White.

Bequest motivation is controlled for in this study. A dichotomous variable is created to represent the respondent's intention to leave a bequest: (0) for no bequest motivation and (1) for a bequest motivation. Bequests motivations for others represent forgone consumption or some form of windfall in the event of death, often life insurance proceeds. The decision to not exhaust all available funds in retirement for the purpose of another has a significant impact on decumulation decisions in retirement.

Wave variables are also included to control for each survey year. The 2016 wave is the reference wave. Summary statistics are reported in Table 1.

Model and Results

This study uses a pooled cross-sectional analysis of the 2016 and 2019 waves from the Survey of Consumer Finances. An ordered probit regression is estimated via maximum likelihood, and marginal effects are used to measure and interpret the association between retirement income satisfaction (1 – 5) and the interaction of annuity ownership and a (1 – 10) scaled risk willingness measure, with 1 being the least willing to take financial risks and 10 being the highest level of risk willingness.

$$y_i^* = x_i\beta + e$$

Where y_i^* is the latent variable representing retirement income satisfaction

Retirement Income Satisfaction (RIS) $_i = 1$, *if* $RIS_i^* \leq \mu_1$ *(Totally Inadequate),*
Retirement Income Satisfaction (RIS) $_I = 2$, *if* RIS_i^* $\mu_1 < RIS_i^* \leq \mu_2$,
Retirement Income Satisfaction (RIS) $_i = 3$, *if* RIS_i^* $\mu_2 < RIS_i^* \leq \mu_3$,
Retirement Income Satisfaction (RIS) $_i = 4$, *if* RIS_i^* $\mu_3 < RIS_i^* \leq \mu_4$,
Retirement Income Satisfaction (RIS) $_i = 5$ *if* $RIS_i^* > \mu_4$ *(Very Satisfactory),*

The results of the analysis are reported in Table 2. The results do show a statistically significant relation between the interaction variable of interest and retirement income satisfaction. When compared to the base category, owning an annuity at the lowest reported level of risk willingness, those who choose to annuitize financial assets and reported having risk willingness levels of 2 and 4 - 10 are positively associated with the top two levels of retirement income satisfaction (4 – 5) and negatively associated with the lower three levels of retirement income satisfaction (1 – 3). When compared to the reference group, those who reported owning an annuity at the low/mid-level (3/10) did not test a statistically significant association with retirement income satisfaction.

These results suggest that owning an annuity at most levels of risk willingness is negatively associated with the lower levels of retirement income satisfaction and positively associated with the higher levels of retirement income satisfaction. Although the decision to own an annuity reduces flexibility with the accumulation and decumulation decisions of assets, this study suggests that annuity ownership interacted with different levels of risk willingness positively affect retirement income satisfaction. Annuities can reduce portfolio volatility, especially during the decumulation phase of the retirement life cycle. Reduced volatility can limit both downside and upside return potential, which lower risk tolerance individuals who are less comfortable with risk may find more satisfaction. For individuals and households with reported higher levels of risk willingness, annuity ownership can secure a certain stream of income during retirement as a part of their overall portfolio allocation.

Table 3 and Table 4 provide Akaike and Bayesian information criteria measures for models with and without the interaction variables, respectively. When comparing two competing models, these measures provide an estimate for model performance, a lower score is preferred. The results of these tests indicate that the inclusion of the interaction variables improves the model's performance.

Table 5 analyzes annuity ownership by differing levels of annual income and wealth, separately. A tabulation of annuity ownership by annual income shows a non-linear relation between the two. A lower proportion of annuities are owned by low income respondents, and the proportion of annuities owned increases as the intervals of income increase until annual income reaches $500,000. After this point, the proportion of annuities owned by income interval sharply declines. This may be explained as annuities are deemed to be normal goods until a certain income level, and after the point of inflection, annuities may be viewed as inferior goods to higher income individuals. With respect to the relation between wealth intervals and the proportion of annuity ownership of respondents in this sample, the relation seems to be less variable. The proportion of annuity ownership increases as each wealth interval increases, though not perfectly linear. Of the 3,516 respondents who own an annuity, 2,145 are owned by those whose wealth is greater than $1million dollars.

The decision making under uncertainty over the lifecycle theory suggests that individuals prefer a smoother consumption path, and as risk aversion increases, risk averse consumers will look to increase their marginal utility of consumption by minimizing their risk exposure. There are many options available in the market that allow consumers to transfer the many types of risks associated with asset accumulation and decumulation decisions. Annuity ownership allows risk averse individuals to transfer market volatility and longevity risk to insurance companies which may be better suited and prepared when compared to the willing consumer to handle these risks through risk pooling. Continued research on how insurance products affect different satisfaction measures can provide practitioners with the positive direction in matching the needs of the consumers with the proper products available in the marketplace that may be able to satisfy the risk averse or risk tolerant individual or household.

Conclusion

In today's retirement planning environment, individuals and households are responsible for how they manage their retirement accumulation and decumulation decisions throughout their lifecycle. The life-cycle hypothesis under uncertainty suggests that individuals prefer to maximize their utility through consistent, intertemporal time allocation decisions in order to smooth the consumption path over the lifecycle, subject to desired preferences (e.g. risk tolerance). In order to maintain smooth consumption throughout the lifecycle, families will tend to borrow when young, save during middle age, and then dissave during retirement (Finke and Huston, 2003). The choice on how individuals choose to accumulate and decumulate assets in order to maximize their marginal utility of consumption is function of personal preference and product and service availability. In general, the process of annuitizing financial assets helps to reduce uncertainty by establishing a certain income stream, mitigating longevity risk, and reducing other volatilities related to asset accumulation and decumulation decisions.

This study uses a pooled cross-sectional analysis of the 2016 and 2019 waves from the Survey of Consumer Finances. An ordered probit regression is estimated via maximum likelihood, and marginal effects are calculated to measure and interpret the association between retirement income satisfaction, measured increasingly on a 1 – 5 scale, and the interaction of annuity ownership and a 1 – 10 increasing scale risk willingness measure. The results indicate a statistically significant, positive association between respondents who own an annuity at most levels of risk willingness and higher levels of retirement income satisfaction. The results suggest that individuals will receive more retirement income satisfaction by transferring a portion of their wealth or income into an annuity.

While the results of this study and many other studies suggest annuity ownership is associated positively with higher levels of retirement income satisfaction and/or overall financial satisfaction, the reality is very few individuals actually utilize annuities within their investment portfolios. Annuities are insurance products designed to mitigate certain risks for the purchaser, specifically longevity risk. If research suggests that risk aversion often increases as age increases, then there is a puzzle as to why so few individuals or households look to annuities to help mitigate the risk of outliving financial assets. One explanation is concerned with the motivation to leave a bequest to one's heirs and the uncertainty of the overall payout of a traditional pure-life annuity. Yaari (1965) states the single unavoidable cost of purchasing an annuity is the forgone cost to bequeath that wealth. Lockwood (2012) suggests that while bequest motives may explain why people do not annuitize all of their wealth, prevailing literature cannot explain why most people do not annuitize any of their wealth. An issue with annuitization of wealth is even though annuities are risk products designed for the risk averse, the inherent uncertainty of one's own death layers another level of uncertainty concerning the purchase and overall payout of an annuity. In essence, a risk averse individual is asked to make a definitive choice between two uncertain events: timing of death and an uncertain payout.

Bonds may be viewed as an alternative to annuities, as both provide for a systematic stream of income. Bonds are often a more popular asset class included in investment portfolios when compared to annuities, yet bonds do not provide a true hedge against longevity risk. And when viewed as pure investments as opposed to insurance against longevity risk, annuities fall second tier in investor preference to bonds. When compared to the popularity of equities, annuities are a third-class passenger, and this adds to the puzzle. If investors are assumed to be inherently some level of risk averse and equities have a long history of increased volatility when compared to annuities, why do individuals greatly prefer equities to annuities? The equity premium puzzle compounds the annuity puzzle.

Another explanation as to why so few people choose annuities is the inflexibility on decumulation decisions after wealth has been transferred to an annuity. According to Pannis (2003), do people prefer being relatively rich at retirement, with a large amount of their own money readily available to be spent as flexibly as they wish? Or would they rather have the comfort of knowing they have a steady income in perpetuity? As is evident, the uncertainty of the future after one surrenders wealth to an insurance company may supersede the certainties guaranteed by an annuity. The results of this study are consistent with prior research that people are positively associated with higher levels of retirement income satisfaction when an annuity is part of their investment portfolio, hence the puzzle.

Additional results of this study are consistent with prior research and theory and suggest, all else equal, that consumers who own an annuity, have a higher education level, have higher levels of annual income, have a higher risk tolerance, who are older, and have higher levels of wealth are associated positively with higher levels of retirement income satisfaction and associated negatively with lower levels of retirement income satisfaction. The decision to own an annuity is in essence a decision on how to manage the utilization of one's wealth in retirement. Because higher levels of wealth are shown to potentially contribute to higher levels of income in retirement, one must be careful to assume that annuity ownership in itself leads to higher levels of retirement income satisfaction. This study implies that when managing the conversion of wealth to income in retirement, one is likely to be more satisfied with the process if it is done in a manner that is consistent with their risk tolerance level. Over the life-cycle, continued assessments of risk tolerance should be up to date as preferences and life changes may alter one's risk tolerance as age increases.

The limitations of any research that looks into annuities is the low popularity of these products and the limited practical data that just does not exist when people fail to utilize a product. Only approximately 7% of the sample used in this study reported to utilize annuities. Additionally, identifying the risk willingness for the household when only one member reports aggregate economic data is another challenge to this study. There is little confusion when there is only one person in the household who can report a true risk willingness measure, but when there are potential multiple decision-makers within a household, identifying a true risk willingness measure becomes more convoluted. Also, risk aversion is often measured relative to net worth, and in this study, there is no indicator for the amount of net worth invested in an annuity for those respondents who report to have investments in an annuity. This information would further support how net worth, risk aversion, and annuity ownership impact retirement income satisfaction.

Further research into how to improve the accuracy of the measurement of the changes in risk tolerance over time can assist planners and consumers maximize their utility with respect to their retirement accumulation and decumulation decisions. Additionally, further understanding into the differences in cultural practices and customs can save time, effort, money, and discomfort when planners, especially in the United States, have such a diverse book of clients.

Table 1.1 Summary Statistics

Dependent Variable	**Mean**	**Standard Deviation**
Retirement Income Satisfaction (1-5)		
1	0.0915	0.2807
2	0.1056	0.3073
3	0.3057	0.4607
4	0.1834	0.3869
5	0.2625	0.4399
Independent Variables		
Annuity Ownership	0.0726	0.2483
Risk Willingness Level (1-10)		
1 - (Not Willing)	0.0459	0.5697
2	0.0748	0.2621
3	0.1138	0.3169
4	0.0985	0.2969
5	0.2145	0.4098
6	0.1272	0.3323
7	0.1437	0.3502
8	0.0981	0.2967
9	0.0315	0.1730
10 – (Substantial)	0.0586	0.2337
Interaction – Annuitization/Risk Willingness Level 1	0.0022	0.0464
Interaction – Level 2	0.0043	0.0655
Interaction – Level 3	0.0075	0.0860
Interaction – Level 4	0.0058	0.0761
Interaction – Level 5	0.0112	0.1055
Interaction – Level 6	0.0104	0.1016
Interaction – Level 7	0.0125	0.1111
Interaction – Level 8	0.0067	0.0815
Interaction – Level 9	0.0021	0.0454
Interaction – Level 10	0.0033	0.0575
Annual Income	16.00	121.89
Annual Income^2	15,113.44	873,333
Age	52.46	16.05
Age^2	3009.99	1709.93
No. Obs: 9,675 (48,375)	**Data: 2016 and 2019 Waves SCF**	

Table 1.1 Continued

Independent Variables		**Mean**	**Standard Deviation**
Retired		0.1914	0.3933
Employed		0.7025	0.4571
Other		0.1061	0.3080
Did Not Graduate		0.0730	0.0812
H.S Diploma		0.2269	0.4188
Some College		0.1783	0.3828
Associate		0.0476	0.2131
Bachelor’s		0.1238	0.2576
Master’s		0.0714	0.2576
Professional		0.0453	0.2078
Married		0.5933	0.4912
Not Married		0.4067	0.4576
Children		0.5581	0.4966
No Children		0.4419	0.4789
Male		0.7987	0.4010
Female		0.2013	0.4010
Net Worth (log)		12.8937	2.9629
Bequest (yes)		0.6814	0.4659
Bequest (no)		0.3186	0.2786
White		0.7331	0.4891
Black		0.1213	0.3265
Hispanic		0.0899	0.2861
Other		0.0555	0.2289
Wave 2016		0.5169	0.4756
Wave 2019		0.4831	0.4667
No. Obs: 9,675 (48,375)		**Data: 2016 and 2019 Waves SCF**	

Table 1.2. **Marginal Effects of Ordered Probit Regression Estimating Retirement Income Satisfaction**
Pseudo R^2: 0.1276

	Retirement Income Satisfaction	**1**	**2**	**3**	**4**	**5**
Independent Variables						
Annuity Ownership		0.0274 (0.0172)	0.0118 (0.0079)	0.0128 (0.0080)	- 0.0082 (0.0051)	- 0.0438 (0.0274)
Risk Tolerance Level						
1 (Not Willing -Base)						
2		- 0.0107*** (0 .0063)	- 0.0036*** (0.0021)	- 0.0018* (0.0010)	0.0040*** (0.0024)	0.0123*** (0.0071)
3		- 0.0411*** (0.0057)	- 0.0153*** (0.0020)	- 0.0110* (0.0013)	0.0149*** (0.0021)	0.0525*** (0.0068)
4		- 0.0516*** (0.0058)	- 0.0199*** (0.0021)	- 0.0157*** (0.0015)	0.0184*** (0.0021)	0.0689*** (0.0071)
5		- 0.0494*** (0.0054)	- 0.0189*** (0.0018)	- 0.0147*** (0.0011)	0.0177*** (0.0020)	0.0653*** 0.0063
6		- 0.0490*** 0.0057	- 0.0187*** (0.0020)	- 0.0145*** (0.0014)	0.0175*** (0.0021)	0.0647*** (0.0069)
7		- 0.0658*** (0.0055)	- 0.0266*** (0.0020)	- 0.0238*** (0.0016)	0.0227*** (0.0020)	0.0936*** (0.0069)
8		- 0.0699*** (0.0057)	- 0.0286*** (0.0022)	- 0.0265*** (0.0019)	0.0238*** (0.0021)	0.1013*** (0.0074)
9		- 0.0771*** (0.0066)	- 0.0324*** (0.0028)	- 0.0318*** (0.0034)	0.0256*** (0.0022)	0.1157*** (0.0104)
10		- 0.0813*** (0.0060)	- 0.0346*** (0.0024)	- 0.0352*** (0.0027)	0.0265*** (0.0021)	0.1247*** (0.0088)
Interaction - Annuity/Risk						
1 (Base)						
2		- 0.1175*** (0.0215)	- 0.0506*** (0.0092)	- 0.0549*** (0.0100)	0.0354*** (0.0065)	0.1877*** (0.0343)
3		- 0.0196 (0.0195)	- 0.0084 (0.0084)	- 0.0092 (0.0091)	0.0059 (0.0058)	0.0314 (0.0312)
4		- 0.0636*** (0.0205)	- 0.0274*** (0.0088)	- 0.0297*** (0.0095)	0.0192*** (0.0061)	0.1016*** (0.0327)
5		- 0.0651*** (0.0189)	- 0.0281*** (0.0081)	- 0.0304*** (0.0068)	0.0196*** (0.0057)	0.1040*** (0.0301)
6		- 0.0675*** (0.0191)	- 0.0291*** (0.0082)	- 0.0315*** (0.0089)	0.0203*** (0.0057)	0.1078*** (0.0305)
7		- 0.0696*** (0.0189)	- 0.0300*** (0.0081)	- 0.0325*** (0.0088)	0.0210*** (0.0057)	0.1111*** (0.0302)
8		- 0.0567*** (0.0203)	- 0.0244*** (0.0087)	- 0.0265*** (0.0095)	0.0171*** (0.0061)	0.0906*** (0.0325)
9		- 0.0545*** (0.0265)	- 0.0235*** (0.0114)	- 0.0255*** (0.0123)	0.0164*** (0.0080)	0.0871*** (0.0423)
10		- 0.0153** (0.0237)	- 0.0221** (0.0102)	- 0.0240** (0.0010)	0.0155** (0.0071)	0.0820** (0.0378)
No. Obs: 9,675 (48,375)	**Data: 2016 and 2019 Waves SCF**		**Standard Errors in Parenthesis**	**Level of Significance * 10% ** 5% *** 1%**		

Table 1.2 Continued

	Retirement Income Satisfaction	**1**	**2**	**3**	**4**	**5**
Annual Income		- 0.0002*** (0.0000)	- 0.0001*** (0.0000)	- 0.0001*** (0.0000)	0.0000*** (0.0000)	0.0003*** (0.0000)
Annual Income^2		- 0.000*** (0.0000)	- 0.0000*** (0.0000)	- 0.0000*** (0.0000)	0.0000*** (0.0000)	0.0000*** (0.0000)
Age		0.0067*** (0.0003)	0.0029*** (0.0001)	0.0031*** (0.0001)	- 0.0020*** (0.0001)	- 0.0108*** (0.0005)
Age^2		- 0.0000*** (0.0000)	- 0.0000*** (0.0000)	- 0.0000*** (0.0000)	0.0000*** (0.0000)	0.0000*** (0.0000)
Employment						
Other (Base)						
Employed		- 0.0249*** (0.0030)	- 0.0107*** (0.0013)	- 0.0116 (0.0014)	0.0075*** (0.0009)	0.0398*** (0.0048)
Retired		- 0.0743*** (0.0037)	- 0.0320*** (0.0016)	- 0.0347*** (0.0018)	0.0224*** (0.0011)	0.1187*** (0.0059)
Education						
Did Not Graduate (Base)						
H.S. Diploma		- 0.0182*** (0.0027)	- 0.0075*** (0.0011)	- 0.0067*** (0.0010)	0.0065*** (0.0010)	0.0259*** (0.0038)
Some College		- 0.0291*** (0.0029)	- 0.0125*** (0.0012)	- 0.0120*** (0.0012)	0.0101*** (0.0010)	0.0435*** (0.0043)
Associate		- 0.0428*** (0.0044)	- 0.0191*** (0.0021)	- 0.0204*** (0.0026)	0.0143*** (0.0014)	0.0681*** (0.0078)
Bachelor's		- 0.0429*** (0.0037)	- 0.0225*** (0.0018)	- 0.0251*** (0.0021)	0.0106*** (0.0012)	0.0809*** (0.0064)
Master's		- 0.0510*** (0.0041)	- 0.0235*** (0.0020)	- 0.0265*** (0.0026)	0.0164*** (0.0012)	0.0846*** (0.0075)
Professional		- 0.0705*** (0.0042)	- 0.0347*** (0.0023)	- 0.0450*** (0.0038)	0.0201*** (0.0010)	0.1301*** (0.0093)
Net Worth (log)		- 0.0319*** (0.0005)	- 0.0137*** (0.0002)	- 0.0149*** (0.0002)	0.0096*** (0.0002)	0.0510*** (0.0007)
Married (Not Married)		- 0.0012 (0.0021)	- 0.0005 (0.0009)	- 0.0005 (0.0009)	0.0003 (0.0006)	0.0020 (0.0033)
Children (No Children)		0.0038* (0.0022)	0.0016* (0.0009)	0.0017* (0.0007)	- 0.0011* (0.0007)	- 0.0060* (0.0032)
Male (Female)		- 0.0043* (0.0025)	- 0.0018* (0.0011)	- 0.0020* (0.0011)	0.0013* (0.0007)	0.0068* (0.0040)
No. Obs: 9,675 (48,375)	**Data: 2016 and 2019 Waves SCF**		**Standard Errors in Parenthesis**	**Level of Significance * 10%. ** 5% *** 1%**		

Table 1.2 Continued

	Retirement Income Satisfaction	1	2	3	4	5
Race/Ethnicity						
White (Base)						
Black		- 0.0325*** (0.0024)	- 0.0151*** (0.0012)	- 0.0193*** (0.0017)	0.0083*** (0.0005)	0.0587*** (0.0049)
Hispanic		0.0014 (0.0031)	0.0005 (0.0013)	0.0006 (0.0013)	- 0.0004 (0.0009)	- 0.0022 (0.0048)
Other		0.0269*** (0.0041)	0.0010*** (0.0015)	0.0092*** (0.0011)	- 0.0086*** (0.0014)	-0.0379*** (0.0053)
Bequest (No Bequest)		- 0.0384*** (0.0019)	- 0.0165*** (0.0008)	- 0.0182*** (0.0009)	0.0115*** (0.0005)	0.0616*** (0.0030)
2019 Wave (2016 Wave)		0.0158*** (0.0024)	0.0068*** (0.0010)	0.0074*** (0.0011)	- 0.0047*** (0.0007)	-0.0252*** (0.0039)
No. Obs: 9,675 (48,375)	**Data: 2016 and 2019 Waves SCF**		**Standard Errors in Parenthesis**	**Level of Significance * 10%. ** 5% *** 1%**		

Table 1.3. AIC/BIC With Interaction Variables

Observations	DF	AIC	BIC
48,375	44	128,728.8	129,115.4

Table 1.4. AIC/BIC Without Interaction Variables

Observations	DF	AIC	BIC
48,375	35	131,286.9	132,468.3

Table 1.5. Annuity Ownership by Income and Wealth

Dollar Amount	Income/Annuity	Percent	Wealth/Annuity	Percent
	No. Obs.		No. Obs.	
≤ $10K	33	~1%	55	1.5%
$10K -- $50K	512	14%	27	0.8%
$50K+ -- $100K	729	21%	61	1.7%
$100K+ -- $500K	1,297	37%	581	17%
$500K+ -- $1M	253	7%	647	18%
> $1M	692	20%	2,145	61%
Total	3,516	100%	3,516	100%

References

Agnew, J., & Szykman, L. (2011). Annuities, financial literacy and information overload. *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*, 158-178.

Benartzi, S., Previtero, A., & Thaler, R. H. (2011). Annuitization puzzles. *Journal of Economic Perspectives*, *25*(4), 143-64.

Dyer, J. S., & Sarin, R. K. (1982). Relative risk aversion. *Management Science*, *28,* 8.

Finke, M. S., & Guillemette, M. A. (2016). Measuring risk tolerance: A review of literature. *Journal of Personal Finance*, *15*(1), 63.

Finke, M. S., & Huston, S. J. (2003). The brighter side of financial risk: Financial risk tolerance and wealth. *Journal of family and economic issues*, *24*(3), 233-256.

Grable, J. E. (2000). Financial risk tolerance and additional factors that affect risk taking in everyday money matters. *Journal of business and psychology*, *14*(4), 625-630.

Grable, J., & Lytton, R. H. (1999). Financial risk tolerance revisited: the development of a risk assessment instrument☆. *Financial services review*, *8*(3), 163-181.

Hanna, S. D., Kim, K. T., & Lindamood, S. (2018). Behind the numbers: Understanding the survey of consumer finances. *Journal of Financial Counseling and Planning*, *29*(2), 410-418.

Heo et al. (2016). An Estimation of the Mediation Effect of Risk Tolerance among Marital Status, Gender, and Investing Behavior. *International Journal of Human Ecology*, 17, 1-14.

Lee, H. K., & Hanna, S. (1995). Empirical patterns of risk-tolerance. *Proceedings of the Academy of Financial Services*.

Lee, H. K., & Hanna S. (1995). Investment portfolios and human wealth. *Financial Counseling and Planning*, *6,* 147–152.

Lockwood LM. Bequest Motives and the Annuity Puzzle. Review of Economic Dynamics. 2012, Apr;15(2):226-243.

Lusardi, A. (2008). *Financial literacy: an essential tool for informed consumer choice?* (No. w14084). National Bureau of Economic Research.

Lusardi, A., Mitchell, O. S., & Curto, V. (2010). Financial literacy among the young. *Journal of consumer affairs*, *44*(2), 358-380.

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of economic literature*, *52*(1), 5-44.

Milevsky, M. A. (2006). *The calculus of retirement income: Financial models for pension annuities and life insurance*. Cambridge University Press.

Milevsky, M. A., Moore, K. S., & Young, V. R. (2006). Asset allocation and annuity purchase strategies to minimize the probability of financial ruin. *Mathematical Finance*, *16*(4), 647-671.

Mitchell, O. S., Poterba, J. M., Warshawsky, M. J., & Brown, J. R. (1999). New evidence on the money's worth of individual annuities. *American economic review*, *89*(5), 1299 - 1318.

Munnell, A. H., & Chen, A. (2013). 401 (k)/IRA Holdings in 2013: An Update from the SCF. *Age*, *55*(14), 1-14.

Panis, C. W. (2003). Annuities and Retirement. *RAND Working Paper*, 03-17.

Pratt, J. W., Raiffa, H., & Schlaifer, R. (1964). The foundations of decision under uncertainty: An elementary exposition. *Journal of the American statistical association*, *59*(306), 353-375.

Sund, B., Svensson, M., & Andersson, H. (2017). Demographic determinants of incident experience and risk perception: Do high-risk groups accurately perceive themselves as high-risk?. *Journal of risk research*, *20*(1), 99-117.

Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies*, *32*(2), 137-150.

Chapter 3

A Comparison of Risk Willingness Between Same-Sex and Different-Sex Couples: A Quasi-Experimental Approach

Introduction

The composition of household decision makers in the United States has been and is increasingly diverse. The study of that composition through the lens of same-sex household decision makers has not been as diverse. While the literature on household decision making is thoroughly abundant, the breakdown and study of the household composition with respect to the sexual orientation of those decision makers is lacking. It may not be entirely appropriate to assume that traditional financial planning assumptions in household decision making, such as risk tolerance measures, are homogeneous across the different types of household compositions. It may easily be understood why many research areas relating to same-sex couples have been stifled or overlooked. Throughout history, this subject matter has been deemed taboo, criticized, or simply passed over. As recent as 1996, the United States passed the Defense of Marriage Act resulting from over a decade of efforts and support from members of congress. The purpose of this legislation was aimed to maintain the definition of a civil union or marriage to be restricted to "traditional" male-female unions. Not only were marriages and civil unions restricted at the federal level, but benefits at the federal level were also restricted, including Social Security Survivor's benefits and the filing of joint tax returns. In 2015 the Supreme Court of the United States struck down all prior legislation restricting the union of same-sex couples, legalized it across the United States, and required the recognition of same-sex marriages across state lines. As a result, the financial planning process for same-sex couples has shifted, as all couples now are given basic marital rights that impact the financial planning process.

The impact of the change of this law in the United States has far reaching implications, particularly in the field of financial planning where a significant number of decisions are the product of the legality, effort, and desire of both decision makers in the household. One of the beginning steps in the financial planning process is to determine the risk tolerance(s) of the household decision makers in an effort to align their investment decisions with their level of comfort. The purpose of this paper is to estimate any potential difference in risk aversion between same-sex couples and different sex couples. Identifying differences in risk aversion amongst participants is fundamental to the financial planning process. The theory of decision making under uncertainty is the underlying theoretical foundations used for this study. The formal hypothesis of this study suggests there is no difference in the average risk tolerance measure between same-sex and different sex households.

Literature Review

Studies that focus on the comparison of same-sex relationships to different-sex relationships are critical because same-sex couples are demographically distinct from different-sex couples. Individuals who report to be in same-sex couples are younger, more educated, more likely to be employed, less likely to have children, and slightly more likely to be female than individuals in different-sex couples (Gates 2013b; Umberson et al., 2016).

With respect to labor force participation of women, it is estimated that 76% of women in same-sex couples are in the workforce as opposed to 62% of women in different-sex couples (Badgett et al. 2021). Antecol and Steinberger (2013) suggest the differential gap is explained by a difference in fertility levels and by the fact that women in same-sex couples are more likely to be primary earners compared to women in different-sex couples. Badgett et al. (2021) finds that same-sex couples comprised of women are penalized from a double gender pay gap. Their study states that even though labor force participation is higher, female same-sex couples had the lowest personal income on average when compared to all other couple types.

Carpenter and Gates (2008) studies partnership rates of heterosexual and gay men and heterosexual women and lesbians in California, and the results from their study find that heterosexual men and women are overwhelmingly more likely to report being or have been married or in a long-term partnership when compared to their counterparts. Using data from American Community Survey, Badgett et al. (2021) finds that heterosexual men and women are more likely to report having children, overwhelmingly biological, and that discrimination against same-sex couples by foster care and/or adoption agencies may also contribute to the discrepancy in the number of children between different-sex and same-sex couples. Few studies have focused on analyzing differences in risk aversion between same-sex and different-sex couples.

Carter and Bao (2005) states that the study of risk aversion has implications in many different fields including: brand choice, information search, preference for gambles, decision framing, financial portfolio management, and insurance purchases, among others. The study continues to state that because people use different methods to categorize people on different aspects of risk aversion, it is difficult to compare results across studies. Some of the most common methods of measuring risk aversion include choice dilemma studies, self-reporting measures, perceived risk studies, and gamble studies.

Kogan and Wallach (1964) use the Choice Dilemma Questionnaire (CDQ) which relies on measuring attitude towards risk when faced with making a choice in dilemma. One deficiency of this method is that this method does not explicitly such important issues as construct validation nor issues of reliability such as internal consistency (Carter and Bao, 2005). Gambles studies often provide unrealistic scenarios and often fail to elicit typical decision behaviors (MacCrimmon and Wehrung, 1984). Self-reporting risk measures often fail to provide adequate initial standards by which risk aversion can be measured and only ask the respondent to assess their overall risk aversion, when the studies focus may be financial or otherwise. It has been empirically proven that people have different risk aversion attitudes towards different scenarios. Perceived risk studies focus on the perceived uncertainty of outcomes and the perceived negative consequences associated with an outcome (Bauer, 1960). While the perceived uncertainty of outcomes is directly relative to measuring risk aversion, the perceived negative consequences associated with an outcome may be subject to anchoring bias; therefore, initial conditions for each individual in the study must be accounted for. One can see the problem of heterogeneous variability for each individual.

Risk aversion may be conceptualized as a preference for maintaining a certain level of consumption over uncertain consumption even if the expected value of the uncertain consumption exceeds that of the level of certain wealth (Finke and Huston, 2003). While the risk averse individual may choose certainty over uncertainty even at a loss, a study by Hanna and Chen (1997) revealed that in all scenarios tested, a portfolio consisting of stocks outperformed a less volatile portfolio over the same 20-year period.

Assessing financial risk aversion is one of the first steps in the financial planning process. Risk aversion, with respect to investing and financial products, represents an investor's willingness to take or avoid financial risks. In more technical terms, risk aversion is often described as the marginal change of the slope of the utility function relative to the current level of wealth (Finke and Guillemette, 2016). If a person's utility of the expected value of an unknown outcome is greater than their expected utility from the unknow outcome itself, they are said to be risk averse (Pratt and Schlaifer, 1964). Conversely, if a person's utility of the expected value of an unknown outcome is less than their expected utility from the unknown outcome itself, they are said to be risk-seeking (Dyer and Sarin, 1992). If there is no difference between the expected value of an unknow outcome and expected utility from the unknown outcome itself, they are considered to be risk neutral.

In the context of investments, volatility is often measured by the standard deviation of the investment portfolio payouts. A safer portfolio will result in less variation in returns, which implies a potential narrower payout interval in the future that is relative to the amount of real dollars invested throughout the investment timeline (Finke and Guillemette, 2016). A riskier portfolio will result in more variation in returns, which implies a potential larger future payout interval, relative to the amount of real dollars invested throughout the investment timeline (Finke and Guillemette, 2016). Previous studies have examined how gender differences impact risk aversion including:

longevity expectations, workforce participation, wealth accumulation, access to higher paying jobs, access to retirement plans, and education (Heo et al., 2016). When taken together, these factors put women at higher risk than men of having financial problems (Fonseca et al., 2012; Heo et al., 2016).

Grable and Lytton (1998) find that higher levels of education and being male are the strongest predictors of willingness to take financial risks. Alternatively, Lee and Hanna (1995) argue that the relationship between income and willingness to take financial risk may be attributed to the importance of wealth when estimating the negative consequences of financial loss. The consequences in terms of utility loss are greatest among those with the least wealth; however, accumulated wealth may not be the most accurate indicator of ability to withstand financial loss (Finke and Huston, 2003). Income and risk aversion have also show to be indicators of ability to withstand financial loss. While men are generally more willing to invest a larger portion of their investment portfolio in equities, women often feel more comfortable investing in assets that are subject to less volatility (Heo et al., 2016). According to Grable (2000), this difference in willingness to invest in equities accounts for at least 10% of the gap in lifetime wealth accumulation between men and women.

As mentioned previously, most empirical studies that include marriage or any type of living arrangements have taken only the traditional definition, male-female, into consideration. As laws in the United States and other countries have changed to provide equal rights and benefits to all with respect to their sexual orientation, more and more households are reporting different definitions of marriage or living arrangement compared to what was considered “traditional”. To assume, or simply overlook, that the gender composition of household decision makers is homogeneous or not important is incorrect. A test of differences in risk aversion between same-sex and different-sex household decision makers is extremely relevant.

A common dilemma when isolating the comparison of interest between two groups is ensuring the sampling procedure or assignment is random, and the distribution of the group characteristics, covariates, are balanced. If not balanced, the study will produce estimates that are not representative of the population due to selection bias. Selection bias may occur when observed or unobserved covariates are not accounted for in a statistical model or controlled for in the design which result in spurious estimates of causal effects (Rosenbaum, 2010). In observational studies for causal effects, treatments are assigned to experimental units without the benefit of randomization (Rosenbaum and Rubin, 1984). As a result, treatment groups may differ systematically with respect to relevant characteristics and, therefore, may not be directly comparable, resulting in biased estimates (Rosenbaum and Rubin, 1984). Observational studies must consistently deal with data that is often collected without random assignment, and many research designs fail to adequately control for selection bias at the analysis level through the balancing of covariates between the groups of interest. Fortunately, there exists a scalar function of the covariates, namely the propensity score, that summarizes the information required to balance the distribution of the covariates (Rosenbaum and Rubin, 1984).

With respect to random assignment, many times survey participants cannot be assigned to a group as it is too difficult, unethical, or simply impossible. For example, Higgins et al. (2011) studies how racial profiling impacted the propensity to be searched after a traffic stop between two groups, Blacks and Whites and Hispanics and Whites. In this scenario, it is evident that random assignment to an ethnicity or race is impossible. When random assignment is not feasible or controlled for, the result is poor research design that has limited capability to make inferential or causal statement and conclusions.

Ideally, an experimental design approach that allows for random assignment can better examine the current vs. the counterfactual, control vs. treatment. This type of design would produce estimates that have much stronger inferential and/or causal interpretations. According to the counterfactual framework for modeling causal effects, the true treatment effect for the group of interest is the difference between the treated outcome and the counterfactual (Holland, 1986; Rubin, 1974). As it is impossible to observe both the current and the counterfactual simultaneously, a reasonable alternative is to estimate an Average Treatment Effect (ATE) for the population (Rubin, 1974; Winship and Morgan, 1999). In order to correctly apply propensity score matching to estimate an Average Treatment Effect and Average Treatment Effect of the Treated (ATT), several assumptions must be met: the conditional independence assumption (CIA), the stable unit treatment value assumption (SUTVA), and the common support assumption (CSA).

The conditional independence assumption states that assignment to the treatment group is independent of the treatment effect conditional on a set of observed covariates, propensity score (Rubin, 1980). Formally, treatment assignment and the observed covariates are conditionally independent given the propensity score, that is

$$\mathbf{x \perp z \mid e(x)}$$

where: x = observed covariates

z = assignment condition (treatment or control)

$e\,(x)$ = propensity score (Rosenbaum and Rubin, 1983).

The stable unit treatment value assumption (SUTVA) states that the outcome does not depend on the assignment procedure, randomized or self-selected, and the treatment is the same for all participants in the treatment group (Holmes, 2014; Rosenbaum and Rubin, 1983). Cox (1958) states that the observation of one unit should be unaffected by the particular assignment of treatment to the other units. This states that when a treatment subject is matched with a control subject, both have the same likelihood of being assigned to either group, and all participants within the treatment group receive the same treatment. This assumption ensures that the estimates for the average treatment effect and the average treatment effect of the treated are consistent across the treated group.

The common support assumption requires that there is enough overlap or common support of the distributions between the treatment group and control group with respect to their propensity scores. Smith and Todd (2005) explain this by identifying the range of propensity scores that have a positive density within both distributions, control and treatment. This analysis can be estimated in several ways but are not limited to: using comparable histograms to visually inspect overlap, comparing minima and maxima values of propensity scores, and using inferential statistics to test significant difference between the distributions of the two groups. This analysis uses inferential statistics to test common support.

Data

The data used in this study derives from the Survey of Consumer Finances' (SCF) 2016 and 2019 waves, which are pooled together. These waves were selected as they both contained a better measurable variation of the dependent variable, risk willingness, that was not available in the previous waves. Each wave of the SCF is a cross-sectional, triennial survey that is supported by the Federal Reserve Board in Partnership with the Department of Treasury. The National Opinion Research Center (NORC) at the University of Chicago has been responsible for collecting the data since 1992. Each wave of the SCF randomly selects individuals from different economic strata to participate in the voluntary study that collectively produces nationally representative data sets. Each wave of the SCF oversamples wealthy households in the United States.

The unit of measurement is the primary economic unit (PEU). The PEU is the self-reported, economically dominant individual or couple within the household who reports aggregate household information on demographics, income sources, housing characteristics, and a number of attitudinal and expectation questions (Hanna et. al, 2017). Any individual demographic information is representative of the PEU respondent. To control for missing data and to conceal the identity of each respondent, the SCF introduces a total of 5 implicates, inclusive, into each wave for every PEU respondent (Hanna et. al, 2017). This study uses all implicates from each wave and applies the appropriate weights to ensure proper variance estimates. The total number of initial observations in this analysis is 10,646 (53,230) PEU responses. The final analysis includes 102 (510) total response.

Dependent Variable

The dependent variable, or outcome variable, in this analysis is a self-reported measure of financial risk willingness scaled from 0 – 10. The question asks, "On a scale from zero to ten, where zero is not at all willing to take risks and ten is very willing to take risks, what number would you (and your {husband/wife/partner}) be on the scale?" (Survey of Consumer Finances, 2016 & 2019). This response that measured as such on a scale from 1 – 10 continuously. Categories 0 and 1 are combined. An average of self-reported risk willingness from the two samples, control and treatment, is used for the Mann-Whitney U Test of comparison in the second stage of the analysis.

In the first stage of the analysis, a control and treatment group must be identified, different-sex and same-sex respectively. The identification of the decision maker(s) based upon whether they are a same sex or different-sex household is the main variable of interest in this study. Each respondent's gender is identified, and the gender of their spouse/partner is also identified to create a dichotomous variable that represents the household status as a same sex or different-sex household that is determined by the PEU. The variable SSH (Same Sex Household) is coded as 0 if they are a different-sex household, control, and SSH is coded as 1 if they are a same-sex household, treatment. There are 310 same sex households identified in the 2016 wave, and 270 same sex households identified in the 2019 wave. There is a total of 510 same sex households in the pooled analysis sample. In total, same sex households comprise 1.8% of the initial analysis sample.

Independent Variables

The log of income is measured continuously in this analysis. Income is defined as the total inflow of monies into the household as reported by the PEU, and income does not represent losses due to investments or other balance sheet inventories. Therefore, any value less than zero was omitted from this analysis, and it is interpreted that the PEU did not answer the question correctly. The average income in this analysis is $160,045, the SCF oversamples wealthy households.

Gender is a dichotomous variable. Gender in this analysis is coded as 1 if the PEU is male and 0 if the PEU is female. The gender variable is used to better help match groups in the first stage of the matching procedure. Approximately 80% of the PEUs who respond in the sample are male. In the SCF, the question for marital status asks, "What is your current legal marital status? Are you married, separated, divorced, widowed, or have you never been married?" The 5 categories are divided and combined into 2 categories, married and not married. If the legal status does not equate married, the respondent is considered not married. A dichotomous variable is created for marital status, 1 for married and 0 otherwise. Approximately 60% of the sample reports to be married.

The SCF asks, "What is the highest level of school completed or the highest degree you have received?" There are 16 different categories available for response. Categories 1 – 8 are combined and create the "did not graduate" category, as they represent a level that has not achieved a high school diploma. Responses 9 – 16 represent the categories which correspond to the highest degree achieved, and these categories included: high school diploma, some college but no degree, two categories for associate degree, bachelor's degree, master's degree, professional school degree, and doctorate degree. The two categories for an associate degree are combined, and the professional school and doctorate degree categories are combined also. The variable for education is categorical with 7 total categories. A dichotomous variable is created for each category, and the variable is coded as 1 if the PEU responds to that category and 0 otherwise.

This analysis measures age continuously. Each PEU's age is determined by subtracting their date of birth from the year of the wave in which they participated. The average age of the responding PEU in this analysis is 54 years of age. Race in this analysis is measured in 4 different categories, and a dichotomous variable is created for each category that is coded as 1 if the PEU responds to that category and 0 otherwise. The four variables in this analysis are: White, Black, Hispanic, and Other.

The presence of children in this analysis is dichotomous variable, and the variable coded as 1 if the respondent(s) have children and 0 otherwise. Considerations for a bequest is also controlled for in this analysis. The PEU was asked if they plan to leave an inheritance or estate for their heirs, and if they answered yes, the variable is coded as 1 and 0 otherwise. In many studies, the correlation between children and leaving a bequest is high, but children are not the only recipients of bequests. Many organizations often benefit from bequests also. A correlation matrix analysis comparing bequests motivations and children based on this sample reveals a correlation coefficient of 0.0143. This indicates an acceptable level of correlation between the covariates for analysis purposes.

The log of net worth is measured continuously in this analysis. Net worth represents the sum of total assets minus total liabilities. Included in the calculation of total assets are financial and non-financial assets. Included in the calculation of total liabilities are all lines of credit and loans, mortgage(s), etc. A positive net worth is a result of total assets being greater than total liabilities.

A wave variable is created to identify each wave category used in the analysis. There is a total of 2 waves, and the waves are coded chronologically.

Model

This analysis uses a 1-for-1 nearest neighbor without replacement propensity score matching technique, t-tests, and the Mann-Whitney U test (MWU test) to estimate the potential difference in risk tolerance between households that identify as same-sex decision makers and those that do not. Propensity score matching is used to help control for selection bias that may arise from the differences in distributions of the independent variables used when comparing different groups. This allows correction and minimization of selection bias at the design level as opposed to covariate adjustment at the analysis level.

In addition to propensity score matching and the MWU-test of significance (Table 2), this analysis reports an average treatment effect (ATE) and an average treatment effect of the treated (ATT) (Table 3), which are better treatment effects than what may be obtained from using other methods that do not control adequately for selection bias (Heckman et al., 1997; Higgins et al., 2011).

Propensity score matching is a two-stage process. In the first stage, Stata's *psmatch2* command utilizes a probit regression model which is estimated via maximum likelihood, a logistic regression model is also appropriate, to determine the propensity of all respondents for experiencing a treatment of interest (Higgins et al., 2011). The treatment of interest for this study is same-sex household status, and the control in this study is different-sex household status. In each wave of the SCF, same-sex households made up between approximately 0.5% and 1.5% of the sample. Propensity score matching is used to help balance the matching covariates between the treatment and control groups that result from non-random sampling. It is possible that the balance of the covariates are not improved, or even worsened, after the matching procedure; therefore, it is necessary to test the balance of the treatment and control groups before the matching process and after the matching process by utilizing a measure of standardized bias. The standardized bias is calculated by taking the difference in the proportions of a cofounder variable of interest in each the control and treatment groups and dividing by the pooled standard deviation of the two groups and multiplied by 100 (Mayne et al, 2015).

While a t-test can be used to measure the difference in two group means, it does not measure any potential difference in the distributions of the covariates. Measuring potential imbalance of the distributions of covariates prior to matching identifies the presence of selection bias. Distributions of covariates are likely balanced if there is no relation between the treatment conditions and the covariates or no relation between propensity scores and the covariates (Rosenbaum & Rubin, 1984). In most cases, the distributions will not be equal when the sample selection is non-random.

The propensity score is determined as such (Rosenbaum & Rubin, 1983):

$$p(T) = Pr\{T=1|S\} = E\{T|S\}, (1)$$

where p(T) is the propensity to being identified between the treatment group or the control group in this study, T indicates that a household identifies as either a same sex household or different-sex household, and S is a vector that contains the covariates upon which the two groups will be matched (Higgins et al., 2011). The propensity score is determined through the use of a probit regression that assesses the likelihood of identifying with either the treatment group or the control group. The vector of covariates used to match the two groups include a variety of socioeconomic factors which are based on the responses from the Primary Economic Unit (PEU): income, age, education, net worth, marital status, ethnicity, gender, presence of children, bequest motivation, and a wave variable to control for each wave.

In the second stage of the process, the treatment and control groups are matched to the counterparts, same sex vs. different-sex status, based upon the propensity scores derived from the first stage to establish two comparative samples. This study uses a 1-for-1 nearest neighbor, without replacement matching technique. After matching, the sample size is reduced to 510 responses, 255 for the control group and 255 for the treatment group. To estimate the balance of the distributions of the covariates after the matching procedure, a standardized bias estimate is conducted, reported in Table 3, which measures the mean difference relative to the variability of the values in the covariate distribution (Rosenbaum & Rubin, 1985). For continuous covariates, the STANDARDIZED BIAS estimate is measured by dividing the difference in means between the two groups by the pooled standard deviation of the groups multiplied by 100 (Clark, 2015). For binary categorical variables, the STANDARDIZED BIAS estimate is the difference between the proportions of the characteristics in each of two groups divided by the pooled standard deviation multiplying by 100 (Austin, 2009).

Once it has been determined that the two groups have been adequately balanced and selection bias has been as minimized as possible, this study uses a Mann-Whitney U test as the hypothesis test to compare the average difference in risk tolerances between the matched same-sex households and different-sex households. The Mann-Whitney U test is used in place of a standard t-test as it is the t-test's non-parametric counterpart and accounts for independent variances for the control and treatment groups.

Additionally, an average treatment effect (ATE) and an average treatment effect of the treated (ATT) are also reported, Table 3. These two statistics provide counterfactual estimates for the treatment and control groups. The average treatment effect (ATE) is defined as the difference of averages between two groups that are estimated when compared to their estimated counterfactual counterparts: what would be the difference in average outcome if all participants were under the treatment assumption and what would be the difference in average outcome if all participants were under the control assumption. These averages are estimated as such: the treatment group is weighted by the inverse of the propensity score derived from the first stage of the analysis to produce a counterfactual estimate for all participants in the treatment group; the control group is weighted by one minus the propensity score derived from the first stage of the analysis to produce a counterfactual estimate for all participants in the control group.

The average treatment effect of the treated (ATT) only considers a counterfactual estimate of the treatment group, not the control group. The ATT is defined as the difference in averages between two groups that are estimated when the treatment group is compared to its estimated counterfactual counterpart: what would be the average outcome if only the treatment group remained under the treatment assumption and what would be the average outcome if the treatment group were under the control assumption. In essence, a measure of difference of the treatment group's average risk tolerance and its estimated counterfactual's average risk tolerance. The counterfactual estimates are calculated as such: the treatment group is weighted by the inverse of the propensity score derived from the first stage of the analysis to produce a counterfactual estimate for all participants in the treatment group.

Results

Prior to the propensity score matching process, the mean standard bias estimate is 15.4%, and after the matching process, the mean standard bias estimate is reduced to 4.6%. Prior to the propensity score matching process, the median standard bias is estimated to be 10.1%, and after the matching process, the median standardized bias estimate is reduced to 4.6%. Harder et al. (2010) recommends a standardized bias estimate less than 20% to be considered a suitable balance between comparison groups, and Caliendo and Koepening (2008) recommends a standardized bias estimate of less than 5% to be considered a suitable balance. While other studies have recommended additional varying acceptable measures of standardized bias, this analysis supports Harder et al. (2010)'s recommendation. A list of covariates with before and after measurements of standardized bias are reported in Table 1. Figure 1 illustrates a comparison of standardized bias for all covariates before and after the matching process.

Prior to the matching process, 11 of the 18 matching covariates had a statistically significant difference in means between the control and treatment groups. Additionally, STANDARDIZED BIAS estimates in 3 of the 18 matching covariates were over the 20% acceptable threshold. After the matching process, only 1 of the 18 matching covariates had a statistically significant difference in the means between the control and treatment groups. More importantly, 18 of the 18 matching covariates were below the 20% STANDARDIZED BIAS acceptable threshold. Figure 1 illustrates a comparison of the

STANDARDIZED BIAS percentages across covariates before and after the matching process.

After the matching procedure, the Mann-Whitney U test is used to estimate any statistically significant difference in the average risk tolerance between the control group and the treatment group after the matching process. The null hypothesis for this analysis is the average risk tolerance for both the treated and control groups is equal. The average reported risk tolerance for the sample prior to matching is 5.39/10; the average reported risk tolerance for the control group after matching is 5.54/10; the average risk tolerance for the treatment group after matching is 5.87/10. The results of the MWU test are listed in Table 2 and are as follows where Risk Tolerance is the scaled dependent variable representing risk tolerance reported on a scale from 1 - 10, and sss is the identifying variable between the control (0) and the treatment (1):

Ho: Risk Tolerance (sss = 0) = Risk Tolerance (sss = 1)

$z = -1.867$

$\text{Prob} > |z| = 0.0618$

Based upon this sample, there is not a statistically significant difference in the average reported risk tolerances between the control and treatment groups after matching at the 5% level of significance; therefore, the null hypothesis cannot be rejected at the 5% alpha level. Alternatively, the null hypothesis can be rejected at the 10% level of significance, and it can be assumed that there is a statistically significance difference in the average reported risk tolerance between the control and treatment groups after matching at the 10% alpha level. The average risk tolerance score for the control group, DS-couples, is 5.54/10, and the average risk tolerance score for the treatment group, SS-couples, is 5.87/10. The difference in reported risk tolerance is 0.33 higher for the treatment group with a standard error of approximately 0.15.

As mentioned previously, the average treatment effect (ATE) is defined as the difference between what would be the average outcome if all participants were under the treatment assumption and what would be the average outcome if all participants were under the control assumption. In this analysis, the ATE is 0.258 with a standard error of 0.063. Based on this sample, it can be inferred that same-sex households have a higher average reported risk tolerance level when compared to their counterparts, significant at the 1% alpha level.

The average treatment effect on the treated (ATT), conversely, does not consider the difference in the potential counterfactuals of all participants. Instead, the ATT only considers the difference in the potential counterfactuals of the participants who report to be in the treatment group. Based on this sample, the ATT is 0.487 with a standard error of 0.134. This estimate is significant at the 1% alpha level. The results of the ATE and ATT are listed in Table 3.

Conclusion

This analysis uses data from the 2016 and 2019 waves of the Survey of Consumer Finances to test the potential difference in risk tolerance between couples who identify as same-sex household financial decision makers and couples who identify as different-sex household financial decision makers. The ending analysis sample size is 102 (510) primary economic unit responses.

In an attempt to minimize selection bias resulting from a non-random sample, this analysis utilizes a two-stage propensity score matching and testing process. First, a propensity score is used to balance the covariates used to match the two groups of interest. The propensity score matching method used is a 1-for-1, nearest neighbor matching without replacement technique. Second, a Mann-Whitney U test is used to test the difference in means between the control and treatment groups. To ensure the matching process is successful in its purpose, the measure of standardized bias is reported twice, one prior to matching and one after matching, in order to estimate the reduction of selection bias at the design level.

The results indicate there is a statistically significant difference in average reported risk tolerance between same-sex couples and different-sex couples at the 10% alpha level, with same-sex couples having the higher average reported risk tolerance. The average risk tolerance score for the control group, DS-couples, is 5.54/10, and the average risk tolerance score for the treatment group, SS-couples, is 5.87/10. The Average Treatment Effect, a measure of the counterfactual of all respondents, indicates same-sex couples have a 0.258 higher average reported risk tolerance compared to different-sex couples. The Average Treatment Effect of the Treated, a measure of the counterfactual only for those in the treatment group, indicates same-sex couples report to have a 0.487 higher average risk tolerance in general.

Same-sex couples are demographically distinct from different-sex couples; therefore, it is important not to assume that risk aversion will be the same between the two groups. Measuring risk aversion is an initial step in the financial planning process, and everything subsequent in the process is a function of the client's risk aversion.

One limitation of this research is that it is difficult to accurately measure a household's risk aversion when potentially only one person is providing a response for all members of the household. Even if two or more measures of risk aversion are reported, can it be assumed that a simple average of the two represents the actual risk aversion of the household members? An inclusion of the bargaining power, relative to financial risk willingness, among all members of the household would help provide more accurate average estimates for both the control and treatment groups. Another limitation of this study is the relatively small sample sizes of the control and treatment groups. Those in the survey who were identified as same-sex households only comprise 1.8% of the initial sample. Pooling more waves together in this analysis may have bolstered the number of same-sex respondents, yet this was impossible as previous waves of the SCF utilized different measures of risk willingness than were used in this study. While both groups were matched on gender as one of the cofounding variables, identifying the sex of the same-sex groups and measuring and comparing risk willingness accordingly may give more insight into the differences in reported risk aversion when compared to different-sex couples. Unfortunately, reducing the sample size of the control and treatment groups further may have provided less reliable estimates.

A survey from the Nations Endowment for Financial Education found that 30% of LGTBQ+ individuals have faced bias, discrimination, and/or exclusion in the financial services sector. An additional survey, LGBTQ+ Money Study, found that these individuals are less likely to use important financial tools, do not feel ready to make important financial decisions, more likely to carry financial stress, less likely to be prepared for a financial emergency, more likely to have significant student loan debt and credit card debt, and more likely to face discrimination when compared to their non-LGBTQ+ counterparts. The results from this study suggest that financial planners will benefit by being aware and understanding the nature of diversity amongst their clients. While prior studies have ignored or overlooked such measures of diversity, these results emphasize the importance of propensity score matching in quasi-experimental procedures as a proper way to estimate factors that may contribute to differences in measures of risk aversion and other relatable fields of study. The use of propensity score matching to help balance the distribution of comparison variables has significant implications for the field of personal financial planning, as comparison groups of interest are often of different size and characteristics. The importance of studying diversity is to counter the underrepresentation of certain socio-demographic groups. When control measures are not in place to account for such underrepresentation, biased estimates may lead to spurious conclusion, and in fact, may do more damage than good. It is important in all scientific fields of study to dedicate research efforts towards all groups that make up society.

Future research may explore differences in risk aversion within the same-sex household financial decision makers household, men-men v. women-women. Additional insight into contributing factors of differences in risk aversion can better assist financial planners meet the specific needs of their clients.

Table 2.1: Descriptive Statistics (Averages) and Before/After Matching Standardized Bias Measurement

	Before Propensity Score Matching				**After Propensity Score Matching**			
Covariates	**Same Sex (SS)**	**Non-SS**	**SB (%)**	**t-test**	**Same Sex (SS)**	**Non-SS**	**SB (%)**	**t-test**
Dependent								
Risk Tolerance	5.69	5.39	8.0	1.85	5.87	5.54	13.3	**2.18***
Demographics								
Income	11.84	11.76	5.0	1.02	11.84	11.82	1.0	0.15
Gender (Male)	0.50	0.82	-71.7	**-18.64***	0.50	0.51	-1.8	-0.25
Marital Status	0.53	0.62	-18.6	**-4.24***	0.53	0.49	6.4	1.00
Age	50.67	54.23	-25.3	**-5.62***	50.67	50.20	3.3	0.05
Net Worth	12.62	12.87	-8.9	-1.94	12.62	12.83	-2.1	-1.08
White	0.80	0.75	12.4	**2.67***	0.80	0.79	0.5	0.09
Black	0.08	0.11	-9.1	-1.93	0.08	0.06	6.7	1.54
Hispanic	0.07	0.09	-6.9	-1.48	0.07	0.07	-0.7	-0.12
Other (Race)	0.04	0.06	-5.8	-1.31	0.05	0.05	-0.8	-0.14
Children	0.64	0.78	-33.6	**-8.91***	0.64	0.63	-0.4	-0.07
Bequest	0.62	0.69	-16.2	**-3.73***	0.62	0.60	4.6	0.71
Education								
Did Not Grad.	0.04	0.07	-17.5	**-3.40***	0.04	0.05	-5.9	-1.07
Diploma	0.15	0.19	-10.1	**-2.61***	0.15	0.13	5.2	0.88
Some College	0.08	0.14	-14.2	**-2.97***	0.08	0.06	6.9	1.29
Assoc.	0.10	0.10	0.3	0.06	0.10	0.13	-9.9	1.49
Bachelor's	0.28	0.25	5.5	1.25	0.28	0.29	-1.3	-0.21
Master's	0.21	0.14	19.0	**4.65***	0.21	0.25	-7.7	-1.12
Professional	0.12	0.09	8.5	**2.01***	0.12	0.08	12.2	**2.00***

Data: Survey of Consumer Finance   Number of Obs.: 510   **@ 5% Alpha**

Table 2.2: Mann-Whitney Hypothesis Test Results

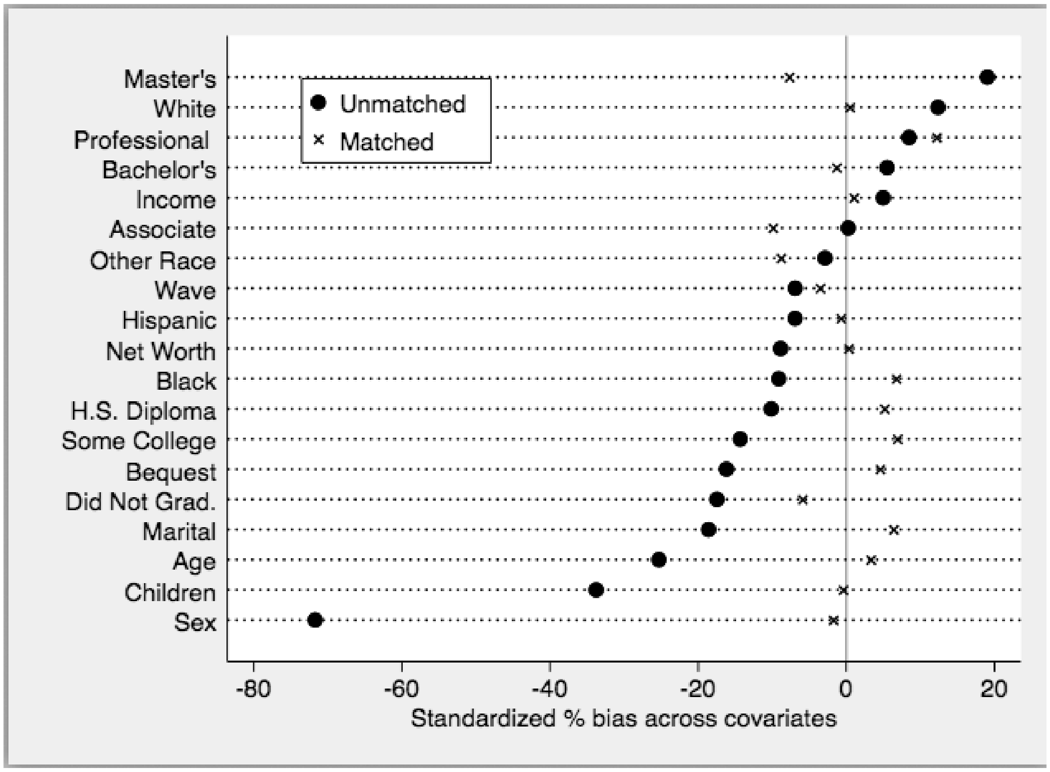
Null	**Ho:**Risk Tolerance (sss = 0) = Risk Tolerance (sss = 1)
Alternative	**Ha:**Risk Tolerance (sss = 0) /= Risk Tolerance (sss = 1)
	z = -1.867*
	Prob > | z | = 0.0618
*** indicates significance @ 10%**	

Table 2.3: Average Treatment Effect (ATE) and Average Treatment Effect of the Treated (ATT) of Reported Risk Tolerance Difference Between Same-Sex and Different-Sex Couples

Average Treatment Effect (ATE)		**Average Treatment Effect of the Treated (ATT)**	
Estimate	Standard Error	Estimate	Standard Error
0.258*	0.063	**0.487***	0.134

*** Indicates Significance at 1% Alpha**

Figure 2.1: Comparison of Standardized Bias Percentage Before and After Matching



References

Antecol, H., & Steinberger, M. D. (2013). Labor supply differences between married heterosexual women and partnered lesbians: A semi-parametric decomposition approach. *Economic Inquiry*, *51*(1), 783-805.

Austin, P. C. (2009). Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Communications in statistics simulation and computation*, *38*(6), 1228-1234.

Badgett, M. V., Carpenter, C. S., & Sansone, D. (2021). LGBTQ economics. *Journal of Economic Perspectives*, *35*(2), 141-70.

Bauer, Raymond A. (1960), "Consumer Behavior as Risk Taking," in *Dynamic Marketing for a Changing World*, ed. Robert S. Hancock, Chicago: American Marketing Association, 389-398.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, *22*(1), 31-72.

Carpenter, C., & Gates, G. J. (2008). Gay and lesbian partnership: Evidence from California. *Demography*, *45*(3), 573-590.

Carter A. Mandrik and Yeqing Bao (2005) ,"Exploring the Concept and Measurement of General Risk Aversion", in NA – Advances in Consumer Research Volume 32, eds.

Geeta Menon and Akshay R. Rao, Duluth, MN : Association for Consumer Research, Pages: 531-539.

Clark, M.H. 2015. Propensity Scoring. In Smelser, N. J., and Baltes, P. B. (Eds.). (2001). *International encyclopedia of the social & behavioral sciences* (Vol. 19). Second edition, Amsterdam: Elsevier, 140-146.

Cox, D.R. (1958). *Planning of experiments*. Oxford, UK: Wiley.

Dyer, J. S., & Sarin, R. K. (1982). Relative risk aversion. *Management Science*, *28,* 8.

Finke, M. S., & Guillemette, M. A. (2016). Measuring risk tolerance: A review of literature. *Journal of Personal Finance*, *15*(1), 63.

Finke, M. S., & Huston, S. J. (2003). The brighter side of financial risk: Financial risk tolerance and wealth. *Journal of family and economic issues*, *24*(3), 233-256.

Fonseca, R., Mullen, K. J., Zamarro, G., & Zissimopoulos, J. (2012). What explains the gender gap in financial literacy? The role of household decision making. *Journal of Consumer Affairs*, *46*(1), 90-106.

Gates, G. J. (2013). Same sex and different sex couples in the American Community Survey: 2005-2011.

Grable, J. E. (2000). Financial risk tolerance and additional factors that affect risk taking in everyday money matters. *Journal of business and psychology*, *14*(4), 625-630.

Grable, J., & Lytton, R. H. (1999). Financial risk tolerance revisited: the development of a risk assessment instrument☆. *Financial services review*, *8* (3), 163-181.

Hanna, S., & Chen, P. (1997). Subjective and objective risk tolerance: Implications for optimal portfolios. *Financial Counseling and Planning*, *8*(2), 17–26.

Hanna, S. D., Kim, K. T., & Lindamood, S. (2018). Behind the numbers: Understanding the survey of consumer finances. *Journal of Financial Counseling and Planning*, *29*(2), 410 418.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological methods*, *15*(3), 234.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training program. *The review of economic studies*, *64*(4), 605-654.

Heo et al. (2016). An Estimation of the Mediation Effect of Risk Tolerance among Marital Status, Gender, and Investing Behavior. *International Journal of Human Ecology*, 17, 114.

Higgins, G. E., Jennings, W. G., Jordan, K. L., & Gabbidon, S. L. (2011). Racial profiling in decisions to search: A preliminary analysis using propensity-score matching. *International Journal of Police Science & Management*, *13*(4), 336 - 347.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, *81*(396), 945-960.

Holmes, W.M. (2014). *Using propensity scores in quasi-experimental designs.* Thousand Oaks, CA: Sage.

Kogan, Nathan and Michael A. Wallach (1964), *Risk Taking: A Study in Cognition and Personality*, New York: Holt, Rinehart & Winston.

Lee, H. K., & Hanna S. (1995). Investment portfolios and human wealth. *Financial Counseling and Planning, 6*, 147–152.

Mayne, S. L., Lee, B. K., & Auchincloss, A. H. (2015). Evaluating propensity score methods in a quasi-experimental study of the impact of menu-labeling. *PloSone*, *10* (12), e0144962.

Pratt, J. W., Raiffa, H., & Schlaifer, R. (1964). The foundations of decision under uncertainty: An elementary exposition. *Journal of the American statistical association*, *59*(306), 353 - 375.

Rosenbaum, P. R. (2010). Missing covariate values. *Design of observational studies. New York: Springer Science+ Business Media*, 193-4.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, *79*(387), 516-524.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33-38.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

Rubin, D. B. (1980). Bias reduction using Mahalanobis-metric matching. *Biometrics*, 293-298.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of econometrics*, *125*(1-2), 305-353.

Umberson, D., Thomeer, M. B., Kroeger, R. A., Lodge, A. C., & Xu, M. (2015). Challenges and opportunities for research on same-sex relationships. *Journal of Marriage and Family*, *77*(1), 96-111.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual review of sociology*, *25*(1), 659-706.

Chapter 4

A Comparison of Risk Willingness Between Men and Women: A Quasi-Experimental Approach

Introduction

It has been well documented that there are significant differences between genders with respect to multiple decision-making criteria. For example, studies suggest that women tend to be better when compared to men on delaying gratification (Dittrich and Leipold, 2014), less likely to enter competitive situations (Charness and Gneezy, 2012), less likely to engage in frequent market trades (Fellner and Maciejovsky, 2007), and generally are more likely to be risk averse (Sebai, 2014). The conclusions drawn from these studies have been based on both experimental and observational methodologies based on data from developing and developed nations. Observational methodologies have the benefit of reduced costs, readily-available-large data sets, socially and morally acceptable data collection procedures, and often take much less time to conduct when compared to experimental studies. One limitation of the conclusions derived from observational methodologies is the limited confidence in causal interpretation. Experimental design methodologies may have their own limitations, but researchers have more confidence in the causal conclusions estimated through the use of experimental design studies.

Risk and uncertainty are of great importance in almost every financial decision (Sebai, 2014). While there are many factors that may contribute to the heterogeneity of decision makers when estimating financial risk aversion, gender is a repeating factor that is present in almost every study. When comparing confounding variables for causal interpretation, such as gender, it is important that the distributions of the variables be balanced across comparison groups; otherwise, any estimate may be biased and lead to inefficient hypothesis testing. This study differs from previous research as it uses a quasi-experimental design, namely propensity score matching, to estimate the average difference in reported risk aversion between men and women. This study also uses propensity score matching to estimate treatment effects that are designed to infer causal relationships between gender and risk aversion. The theoretical framework of this study is decision making under uncertainty.

Literature Review

There has been an extensive amount of research focused on the association between financial risk aversion and explanatory variables, and how they may impact aversion to financial risk taking. The importance and general implication of these studies is that, all else being equal, conservative strategies result in less retirement income on average when compared to more aggressive investment strategies (Berggren and Gonzales, 2010). If women are more risk averse when planning for retirement, are expected to have a longer life expectancy, and are penalized by the gender pay gap, these combined factors suggest that women may have a higher probability of experiencing financial difficulties in retirement. Risk averse individuals may benefit more from working with a financial advisor, which in turn may help reduce any anxiety associated with making financial decisions.

Risk aversion has a direct impact on portfolio allocation decisions for investors who choose to participate in the market. Portfolio theory suggests that utility-maximizing individuals assume varying degrees of risk that are influenced by a number of factors including: age, wealth, income, marital status, education, and individual preferences (Arano et al., 2009). Gender has often been associated as a measure for individual preferences, yet previous research is often hesitant to include gender as a theoretical foundation for an influencing factor to risk aversion.

Sebai (2014) measures gender's impact on financial risk aversion through the use of a series of OLS and Logistic regressions. The study uses data from a Tunisian brokerage firm and focuses on the propensity of men and women holding short or long positions in the stock market. Additionally, through the use of OLS regression analysis, the study focuses on the impact gender has on reported risk aversion. The results suggest that education is an important factor for both men and women when estimating risk aversion, but education has a different effect when assessed on gender. As education levels increase, women are less likely to hold long positions, and the opposite is true for men. When measuring reported risk aversion levels, education is negatively associated with risk aversion for men and positively associated with risk aversion for women. These results are consistent with the findings of Dwyer et al. (2002) and Atkinson et al. (2003), but the results do not account adequately for selection bias and the effect the bias may have on hypothesis testing.

Berggren and Gonzales (2010) studies gender's effect on financial decision making and overconfidence in financial decision making. The study is focused on students at the Umea School of Business in Sweden, estimated to be the most gender equal country in the world (Global Gender Gap, 2007). The results suggest that when compared to women, men are more likely to report to be risk takers, more likely to invest in financial markets, more likely to invest a windfall into the stock market, and more likely to make a risky gamble for profit. While this study has the benefit of being representative of one of the most gender-neutral countries in the world, it is limited to only the students at the Umea School of Business. According to Berggen and Gonzales (2010), a major limitation of the study is that it may not be readily generalized to the population as a whole.

Sapienza et al. (2009) studies circulating testosterone levels in men and women and how these levels impact career choice. According to their study, higher levels of testosterone in humans has been shown to enhance the motivation for competition and dominance, reduce fear, alter the balance between sensitivity to punishment and reward, and be positively associated with behavior such as gambling and alcohol use. The results of their study suggest that when men and women are analyzed separately, the negative association between circulating testosterone levels and risk aversion among men was weak and not statistically significant. Alternatively, the negative association between circulating testosterone levels and risk aversion was stronger and statistically significant for women. A limitation to this research is the gender imbalance of participants. Of the 550 participants, 381 participants were male, and 169 participants were female.

Age is an associated demographic variable when estimating risk aversion. Bakshi and Chen (1994) suggests that as age increases, risk aversion increases also. The findings do not suggest a pure linear association. Riley and Chow (1992) suggests that risk aversion is negatively associated with age until age 65, at which risk aversion is positively associated with an increase in age. In general, it is assumed that investors reduce the proportion of their assets held in risky investments as age increases, presumably as they have less time to recover from significant losses to their investment portfolios (Arano et al., 2009). A limitation to the Arano et al. (2009) study is that of the overall sample, only 19.78% of the respondents are women.

A common issue when analyzing risk aversion differences between men and women is the disproportion in survey responses of men and women, often a disparity in favor of men of up to four times the response rates. One method applied at the analysis level used to balance the disproportion is the Blinder-Oaxaca decomposition technique. Georgellis and Wall (2005) use this technique when analyzing risk aversion by studying transition rates from salaried employees to self-employment, and the results suggests there is a lower transition rate for women when compared to men. Fossen (2009) studies risk aversion by comparing gender difference in self-employment and also uses the Blinder-Oaxaca decomposition at the analysis level for correction purposes. The results indicate the estimated degree of risk aversion is low for men and higher, still moderate, for women (Fossen, 2009).

As the research continues to suggest that women tend to be more risk averse when compared to men, several European countries have taken said research and applied the results to rebuilding their financial systems after the world-wide market crash in 2008. After declaring bankruptcy in 2008, Iceland turned to two women to rebuild its financial system “after the banking empire built by its your male business-schooled elite collapsed” (Maxield et al., 2010). A study based out of France after the 2008 market crash suggests that women are more risk averse and are better able than men to “safely” husband financial resources (Ferrary, 2009).

Opposed to the few studies aforementioned that use selection bias correction techniques at the analysis level, the focus of this study is the use of propensity score matching to estimate causal effects and to correct for selection bias at the design level in order to better analyze differences in risk aversion between men and women. Selection bias can be present as a result of several circumstances, but one of the most common is when the sampling procedure or assignment is not random. If the assignment to a covariate group is not randomly assigned, an imbalance amongst the covariates is present, and a study may produce estimates that are not representative of the population. If selection bias is not accounted for at the design level or corrected for at the analysis level, the results may produce spurious estimates of causal effects (Rosenbaum, 2010). In observational studies for causal effects, treatments are assigned to experimental units without the benefit of randomization (Rosenbaum and Rubin, 1984). As a result, treatment groups may differ systematically with respect to relevant characteristics and, therefore, may not be directly comparable, resulting in biased estimates (Rosenbaum and Rubin, 1984).

Differing from observational studies, an experimental design approach can allow for random assignment, which is a better approach to examine the current vs. the counterfactual (i.e. control vs. treatment) and can produce estimates designed for causal interpretations. The counterfactual framework for modeling causal effects suggests that the true treatment effect for the group of interest is the difference between the treated outcome and the counterfactual (Holland, 1986; Rubin, 1974). It is well known that it is unreasonable to assume someone can observe the current and the counterfactual at the same time. A solution to this observational dilemma is to estimate an Average Treatment

Effect (ATE) for the population and an Average Treatment Effect for Treated (Rubin, 1974; Winship and Morgan, 1999). In order to produce these estimates and to properly apply propensity score matching, several assumptions must be met: the conditional independence assumption (CIA), the stable unit treatment value assumption (SUTVA), and the common support assumption (CSA).

The conditional independence assumption states that assignment to the treatment group is independent of the treatment effect conditional on a set of observed covariates, propensity score (Rubin, 1980). Formally, treatment assignment and the observed covariates are conditionally independent given the propensity score, that is

$$\mathbf{x \perp z \mid e(x)}$$

where: x = observed covariates

z = assignment condition (treatment or control)

e (x) = propensity score (Rosenbaum and Rubin, 1983).

The stable unit treatment value assumption (SUTVA) states that the outcome does not depend on the assignment procedure, randomized or self-selected, and the treatment is the same for all participants in the treatment group (Holmes, 2014; Rosenbaum and Rubin, 1983). The common support assumption assures there is enough common support of the distributions between the treatment and control groups with respect to their propensity scores. Smith and Todd (2005) explain this by identifying the range of propensity scores that have a positive density within both distributions, control and treatment.

Data

The data used in this study derives from the Survey of Consumer Finances' (SCF) 2016 and 2019 waves, which are pooled together. Each wave of the SCF is a cross-sectional, triennial survey that is supported by the Federal Reserve Board in Partnership with the Department of Treasury. The National Opinion Research Center (NORC) at the University of Chicago has been responsible for collecting the data since 1992. Each wave of the SCF randomly selects individuals from different economic strata to participate in the voluntary study that collectively produces nationally representative data sets. Each wave of the SCF oversamples wealthy households in the United States.

The unit of measurement is the primary economic unit (PEU). The PEU is the self-reported, economically dominant individual or couple within the household who reports aggregate household information on demographics, income sources, housing characteristics, and a number of attitudinal and expectation questions (Hanna et. al, 2017). Any individual demographic information is representative of the PEU respondent. To control for missing data and to conceal the identity of each respondent, the SCF introduces a total of 5 implicates, inclusive, into each wave for every PEU respondent (Hanna et. al, 2017). This study uses all implicates from each wave and applies the appropriate weights to ensure proper variance estimates. Values that do no align with appropriate responses are dropped from the analysis. The total number of initial observations in this analysis is 10,646 (53,230) PEU responses. The final analysis includes 1,764 (8,820) total responses, 882 (4,410) in each the control and treatment groups.

Dependent Variable

The dependent variable, or outcome variable, in this analysis is a self-reported measure of financial risk willingness scaled from 0 – 10. The question asks, "On a scale from zero to ten, where zero is not at all willing to take risks and ten is very willing to take risks, what number would you (and your {husband/wife/partner}) be on the scale?" (Survey of Consumer Finances, 2016 & 2019). This response that measured as such on a scale from 1 – 10 continuously. Categories 0 and 1 are combined for coding and distribution purposes. Averages of risk willingness from the control sample and treatment sample are used for the Mann-Whitney U Test of Means comparison in the second stage of the analysis.

The treatment and control variables used in the second stage of the analysis are based on the available identifying categories for gender, women and men. Woman is the treatment variable, and man is the control variable. The 2016 and 2019 waves of the SCF are limited to these two responses. The gender variable is coded as 0 for man and 1 for woman. In the 2016 wave there are 21,843 men PEU respondents and 5,670 women PEU respondents. In the 2019 wave there are 20,670 men PEU respondents and 5,047 women PEU respondents. Women PEU respondents comprise of approximately 20% of the sample before matching.

Independent Variables

Income is a continuous variable. Income is defined as the total inflow of monies into the household as reported by the PEU, and income does not represent losses due to investments or other balance sheet inventories. Therefore, any value less than zero was omitted from this analysis, and it is interpreted that the PEU did not answer the question correctly. The median income is $44,000, and the mean income is $160,045.

Marital status is a dichotomous variable in this analysis, married (1) or not married (0). In the SCF, the question for marital status asks, “What is your current legal marital status? Are you married, separated, divorced, widowed, or have you never been married?” If the legal status does not equate married, the respondent is considered not married.

The variable for education is categorical and totals 7 categories which corresponds to the highest level of education obtained by the PEU: did not graduate high school, high school graduate, some college experience, associate degree, bachelor’s degree, master’s degree, and professional degree. For each category, a dichotomous variable is created, and if the education level corresponds to the category, the variable is coded as a 1 or 0 otherwise.

A dichotomous variable for White, Black, Hispanic, and Other-race is created for each category and coded as 1 if the respondent identifies as such or 0 otherwise. Age is measured continuously in this analysis. The respondent’s age is determined by subtracting their birth year from the year of their survey wave, 2016 or 2019. To control for the presence of children, regardless of how many, a dichotomous variable is created and coded as 1 for children or 0 otherwise.

Considerations for a bequest is controlled for in this analysis. The PEU was asked if they plan to leave an inheritance or estate for their heirs, and if they answered yes, the variable is coded as 1 and 0 otherwise.

Net worth is a continuous variable, and the log of net worth is used. After taking the log of net worth, the sample size is reduced from 53,230 PEU responses to 48,375 PEU responses, implicates included. A wave variable is included to control for each respondent's survey year.

Model

This study uses a 1-for-1 nearest neighbor without replacement propensity score matching technique, t-tests, and the Mann-Whitney U test to estimate the potential difference in financial risk willingness between men and women. Propensity score matching is used to balance the distributions of the independent covariates used to derive the propensity score and establish the control and treatment groups. This method helps to reduce selection bias by correcting the imbalance in the distributions of the independent covariates at the design level as opposed to making corrections at the analysis level.

In addition to propensity score matching, t-tests, and the Mann-Whitney U test of significance, this study reports the average treatment effect (ATE) and the average treatment effect of the treated (ATT), which are better treatment effects than what may be obtained from using other methods that do not adequately control for selection bias (Heckman et al., 1997; Higgins et al., 2011).

This quasi-experimental study design is a two-stage process. In the first stage, Stata's *psmatch2* command utilizes a probit regression model estimated via maximum likelihood to determine the propensity score of all respondents for experiencing a treatment of interest (Higgins et al., 2011). The treatment group of this study is women, and the control group of this study is men. Women comprise of approximately 20% of the respondents for each survey wave. Because the assignment to the treatment and control groups may not be random, propensity score matching is used to help balance the matching covariates used to establish both groups. It is possible that the balance of the covariates are not improved, or even worsened, after the matching procedure; therefore, it is necessary to test the balance of the treatment and control groups before the matching process and after the matching process by utilizing a measure of standardized bias. The standardized bias is calculated by taking the difference in the proportions of a cofounder variable of interest in each the control and treatment groups and dividing by the pooled standard deviation of the two groups and multiplied by 100 (Mayne et al, 2015).

Often times a t-test is used to measure the difference of means between two groups, but the test does not measure any potential imbalance of the distributions of covariates used. It is important to measure the imbalance of distributions before and after the matching process to identify the presence of significant selection bias at either stage. Distributions of covariates are likely balanced if there is no relation between the treatment conditions and the covariates or no relation between propensity scores and the covariates (Rosenbaum & Rubin, 1984). In most cases, the distributions will not be equal when the sample selection is non-random. The propensity score is determined as such (Rosenbaum & Rubin, 1983):

$$p(T) = \Pr\{T=1|S\} = E\{T|S\}, (1)$$

where p(T) is the propensity to being identified between the treatment group or the control group of the study, T indicates that a PEU identifies as either male or female, and S is a vector that contains the covariates upon which the two groups will be matched (Higgins et al., 2011). As mentioned previously, a probit regression is used to derive a propensity score that assesses the likelihood of identifying with either the control or treatment groups. The vector of covariates used in the matching process includes variables suggested by theory and also control for relevant socioeconomic factors such as: income, age, education, log of net worth, marital status, ethnicity, presence of children, and a wave variable identifying each PEU's survey year.

After matching, the analysis sample size is reduced from 9,675 (48,375) responses to 1,764 (8,820) total responses, 882 (4,410) in each the control and treatment groups. To estimate the balance of the distributions of the covariates after the matching procedure, a standardized bias estimate is conducted which measures the mean difference relative to the variability of the values in the covariate distribution (Rosenbaum & Rubin, 1985). For continuous covariates, the standardized bias estimate is measured by dividing the difference in means between the two groups by the pooled standard deviation of the groups multiplied by 100 (Clark, 2015). For binary categorical variables, the standardized bias estimate is the difference between the proportions of the characteristics in each of two groups divided by the pooled standard deviation multiplying by 100 (Austin, 2009). The acceptable range of estimated standardized bias is from 5% - 20%. This study used the upward bound of 20% as acceptable for continued estimation.

In the second stage, respondents are matched to their counterparts derived from the propensity score and establish the two comparative groups, control and treatment groups. A Mann-Whitney U test is used to compare the difference in average risk willingness measures between men and women. A Mann-Whitney U test is recommended, and contrary to the standard t-test, accounts for independent variances for the control and treatment groups. The formal hypothesis states there is no difference in average risk willingness measures between men and women.

In addition to the Mann-Whitney U test (Table 2), this study reports an average treatment effect (ATE) and an average treatment effect of the treated (ATT) (Table 3). These statistical measures produce counterfactual risk willingness estimates for the control and treatment groups. The average treatment effect (ATE) is a counterfactual estimate for all respondents, which can be used to predict the causality of the treatment effect when considering the entire sample. The average treatment effect of the treated (ATT) is a counterfactual estimate which can be used to predict causality of the treatment effect when considering only the treatment group. These averages are estimated as such: the treatment group is weighted by the inverse of the propensity score derived from the first stage of the analysis to produce a counterfactual estimate for all participants in the treatment group; the control group is weighted by one minus the propensity score derived from the first stage of the analysis to produce a counterfactual estimate for all participants in the control group.

Results

Prior to the propensity score matching process, the mean standardized bias estimate is 27.3%, and after the matching process, the mean standardized bias estimate is reduced to 3.4%. Prior to the matching process, the median standardized bias is estimated to be 12.4%, and after the matching process, the median standardized bias is estimated to be 1.7%. Table 1 reports the before and after matching measures of standardized bias. Figure 1 displays a visual representation of the before and after matching measures of standardized bias. As mentioned previously, this study uses the 20% standardized bias estimate threshold as recommended by Harder et al. (2010). Different studies have recommended different standardized bias threshold estimates. For example, Caliendo and Koepening (2008) have recommended the standardized bias estimate threshold be set as low as 5%. A list of the dependent variable and related covariate estimates with before and after measurements of standardized bias are reported in Table 1.

Prior to the matching process, 17 out of the 17 matching covariates had a statistically significant difference in means between the two comparative groups, and 6 out of the 17 matching covariates has exceeded the 20% standardized bias estimate threshold. After the matching process, only 7 out of the 17 matching covariates have a statistically significant difference in means between the two comparative groups, and 0 out of the 17 matching covariates have a standardized bias estimate measure above the 20% acceptable threshold.

Once the control and treatment groups have been established by the use of propensity score matching and the standardized bias estimates have been reduced, a Mann-Whitney U test is used to estimate any statistically significant difference in the average risk tolerance between men and women, control and treatment groups respectively. The null hypothesis for this analysis is that there is no statistical difference in an average risk tolerance measure between men and women. The average reported risk tolerance for the entire analysis sample, men and women combined, is 5.39/10; the average reported risk tolerance for the control group, men, is 5.57/10; the average reported risk tolerance for the treatment group, women, is 4.65/10. The results of the Mann-Whitney U test, reported in Table 2, which tests our formal hypothesis are as follows where Risk Tolerance is the dependent variable representing the reported risk tolerance score of the PEU which is scaled between 1- 10, and Sex is the identifying variable establishing the control (0) and the treatment (1) groups:

Ho: Risk Tolerance (Sex = 0) = Risk Tolerance (Sex = 1)

z = 37.76

Prob > | z | = 0.000

Based upon this sample, there is a statistically significant difference in the average reported risk tolerance between men and women, acceptable at the 1% alpha level. Therefore, the null hypothesis can be rejected. The difference in average reported risk tolerance is -0.92 with a standard error of approximately 0.02 for the treatment group when compared to the control group. Women report to have a 0.92 lower score when compared to men.

In addition to testing the difference in average risk tolerance between men and women, this study report two measures of treatment effects, the Average Treatment Effect (ATE) and the Average Treatment Effect of the Treated (ATT). These estimates represent counterfactual estimates that can better help to infer causality when using observational data for research purposes. The ATE estimates counterfactual risk tolerance estimates for both the control and treatment groups which then are used to estimate the causal impact of the treatment of interest on the entire sample. The ATT estimates counterfactual risk tolerance estimates for only the treatment group which are then used to estimate the causal impact of the treatment of interest for only the treatment group. In this analysis, the ATE is reported to be -0.761 with a standard error of 0.14, and the ATT is reported to be -0.871with a standard error of 0.06. Both estimates are significant at the 1% level of alpha. Both estimates are reported in table 3.

Conclusion

The focus of this study is to test the difference in reported average financial risk willingness between men and women. The analysis uses pooled cross-sectional data from the 2016 and 2019 waves of the Survey of Consumer Finances. A two-stage technique using propensity score matching and a Mann-Whitney U test is used establish a quasi-experimental research design. In the first stage, a probit model is used to estimate a propensity score that is derived from a list of theoretically and empirically motivated covariates. The analysis uses a 1-for-1 nearest neighbor without replacement to define a control and treatment group, men and women respectively, used for the second stage testing. In the second stage, a Mann-Whitney U test of mean differences is used to test the following hypothesis: there is no difference in the average reported risk willingness between men and women. The results of the Mann-Whitney U test indicate that women when compared to men, on average, report a 0.92/10 lower risk willingness score, significant at the 1% alpha level. The Mann-Whitney U test is used in place of a standard t-test as the Mann-Whitney U test accounts for standard error corrections and is the non-parametric counterpart to the t-test. Estimates of standardized bias are reported before the matching process and after the matching process to estimate the proportional imbalance and necessary corrections of the distributions of the covariates used in the analysis. Prior to the matching process, 17 of 17 covariates used to derive the propensity score reported statistically significant average differences between the control and treatment groups through the use of a t-test. After the matching process, only 1 of 17 covariates used to derive the propensity score reported a statistically significant average difference between the control and treatment groups. Measures of standardized bias before and after the matching process indicate that the mean standardized bias was reduced from 27.3%, prior to matching, to 3.4%, post matching process. Treatment effects are calculated to infer causal effects, one of the entire sub-sample (ATE) and another for only the treatment group (ATT). The ATE estimate is reported to be -0.76/10 with a standard error of 0.014, indicating that when the entire sub-sample is estimated against their counterfactual risk willingness scores, the treatment of interest suggests an average risk willingness score of 0.768/10 lower. The ATT estimate is reported to be -0.87/10 with a standard error of 0.59, indicating that when the treatment group is compared to their counterfactual risk willingness scores, the treatment of interest suggests an average risk willingness score of 0.87/10 lower.

This study continues the well-established research of analyzing differences in financial risk willingness between men and women. This study adds to the current research as it uses a quasi-experimental design approach that is designed to balance the distributions of covariates while reducing standardized bias in order to ensure more accurate estimations, standard error estimates, hypothesis testing, and causal inferences. As mentioned previously, many studies that look at comparing risk aversion measures between men and women are not representative a general population, do not balance the distributions of the main variables of interest, do not correct for standardized bias among covariates at the design level or analysis level, or do not offer treatment effects for causal interpretations. Therefore, non-randomization selection of survey respondents can lead to an imbalance of the distributions of the covariates used in analysis which, in turn, can lead to spurious estimations, inaccurate hypothesis testing, and invalid interpretations and causal estimates.

One particular limitation to this research is prevalent because identifying the balance of household bargaining power between a household member and the PEU respondent, the economically dominant financial decision-maker, can be difficult when only one person reports aggregate household information. When the respondent reports to be single in household status, this problem is eliminated, but when the household is made up of more than one person, the problem is evident. One solution to this problem was to limit the sample to only single PEU respondents. This study chose to incorporate both single and married/partnered individuals as the combination is a better representation of the general population and limiting the sample accordingly would have reduced the sample, prior to matching, by almost 45%.

A positive implication of this research is the continued study of risks to successful retirement planning. If women are more risk averse when planning for retirement, are expected to have a longer life expectancy, and are penalized by the gender pay gap, these combined factors suggest that women may have a higher probability of experiencing financial difficulties in retirement. Because of the empirical methods used in this study, the results of the hypothesis testing and treatment effect measures are deemed to be more reliable. Future researchers can better rely on the results of this study as they continue to look for causes and solutions to common retirement planning issues, particularly when a portion of the population faces discrimination in almost every aspect of wealth accumulation and representation in the fields of business and financial planning. Future research on the long-term effects of how a conservative investment portfolio and a conservative decumulation pattern can impact retirement satisfaction can be beneficial for planners when describing to their clients the importance of assessing their financial risk willingness.

Table 3.1: Descriptive Statistics (Averages) and Before/After Matching Standardized Bias Measurement

	Before Propensity Score Matching				After Propensity Score Matching			
Covariates	**Men**	**Women**	**SB (%)**	**t-test**	**Men**	**Women**	**SB (%)**	**t-test**
Dependent								
Risk Tolerance	5.62	4.66	-42.8	**-35.89***	5.57	4.65	27.3	**37.32***
Demographics								
Income	2.0 e+05	39,890	-16.2	**-10.77***	47,282	39,890	-0.7	**-4.00***
Marital Status	0.74	0.09	-177.4	**-134.0***	0.09	0.09	0.0	0.03
Age	54.17	54.78	3.80	**3.34***	54.29	54.78	3.0	1.87
Net Worth	13.27	11.18	-74.7	**-62.21***	11.13	11.18	1.9	1.31
White	0.77	0.64	-28.9	**-26.64***	0.62	0.64	4.0	2.50
Black	0.08	0.22	39.6	**38.84***	0.23	0.22	-4.5	-2.51
Hispanic	0.09	0.08	5.10	**4.45***	0.09	0.10	-1.9	-1.22
Other (Race)	0.04	0.06	-11.2	**-8.85***	0.04	0.03	1.5	1.20
Children	0.79	0.76	-6.3	**-5.43***	0.77	0.76	-3.4	-2.22
Bequest	0.61	0.68	-17.4	**-3.89***	061	0.62	3.6	0.98
Education								
Did Not Grad.	0.07	0.09	7.40	**6.48***	0.09	0.09	0.2	0.16
Diploma	0.19	0.21	4.90	**4.25***	0.21	0.21	0.5	0.30
Some College	0.12	0.17	15.8	**-14.09***	0.18	0.17	-0.8	-0.51
Assoc.	0.09	0.13	12.8	**11.42***	0.14	0.13	-3.9	-2.35
Bachelor’s	0.27	0.22	-12.0	**-9.98***	0.21	0.22	1.2	0.80
Master’s	0.14	0.12	-7.01	**-5.77***	0.12	0.12	1.3	0.92
Professional	0.10	0.04	-23.3	**-17.78***	0.04	0.04	1.2	1.01

Table 3.2: Mann-Whitney Hypothesis Test Results

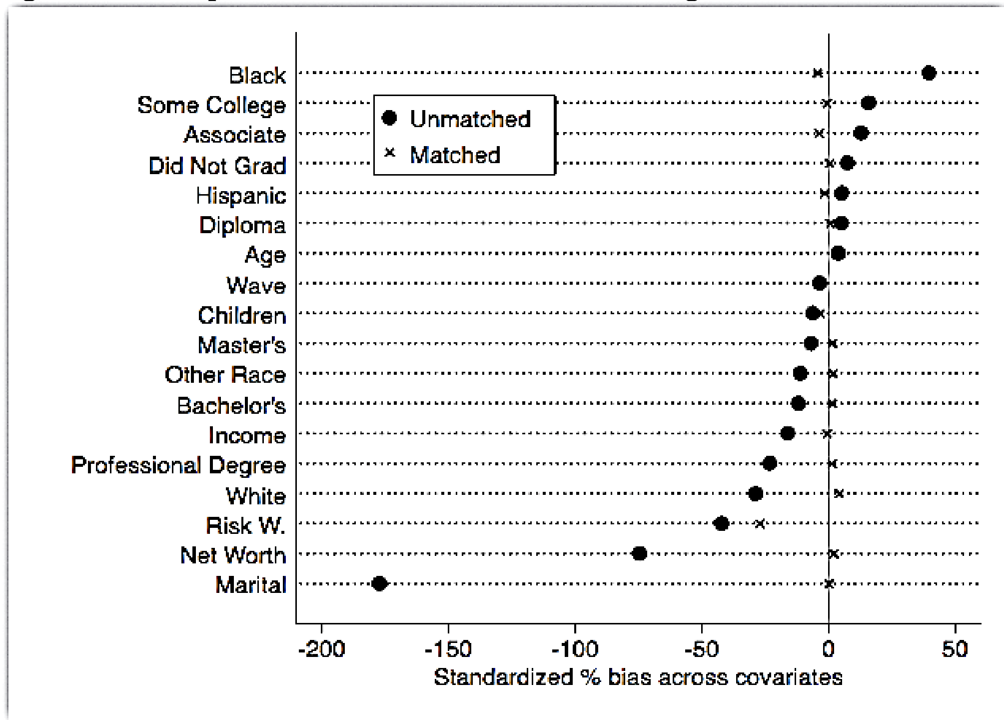
Null	**Ho:** Risk Tolerance (M= 0) = Risk Tolerance (W= 1)
Alternative	**Ha:** Risk Tolerance (M = 0) /= Risk Tolerance (W = 1)
	z = 37.76*
	Prob > | z | = 0.0000
*** indicates significance @ 1% alpha**	

Table 3.3: Average Treatment Effect (ATE) and Average Treatment Effect of the Treated (ATT) of Reported Risk Tolerance Difference Between Men and Women

Average Treatment Effect (ATE)		**Average Treatment Effect of the Treated (ATT)**	
Estimate	Standard Error	Estimate	Standard Error
-0.768*	0.014	**-0.871***	0.059

*** Indicates Significance at 1% Alpha**

Figure 3.1: Comparison of Standardized bias Percentage Before and After Matching



References

Arano, K., Parker, C., & Terry, R. (2010). Gender-based risk aversion and retirement asset allocation. *Economic Inquiry*, *48*(1), 147-155.

Atkinson, S. M., S. Boyce Baird, and M. B. Frye, 2003, Do Female Mutual Fund Managers Manage Differently? Journal of Financial Research, 26 (1), 1-18.

Bakshi, G. S., & Chen, Z. (1994). Baby boom, population aging, and capital markets. *Journal of business*, 165-202.

Berggren, J., & Romualdo, G. (2010). Gender difference in financial decision making: A quantitative study of risk aversion and overconfidence between the genders.

Caliendo, M., Fossen, F. M., & Kritikos, A. S. (2009). Risk attitudes of nascent entrepreneurs new evidence from an experimentally validated survey. *Small business economics*, *32*(2), 153-167.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1), 50-58.

Dittrich, M., and K. Leipold, 2014, Gender Differences in Time Preferences, Economics Letters, 122 (3), 413-415.

Dwyer, P. D., J. H. Gilkeson, and J. A. List, 2002, Gender Differences in Revealed Risk Taking: Evidence from Mutual Fund Investors, Economics Letters, 76 (2), 151-158.

Fellner, G., and B. Maciejovsky, 2007, Risk Attitude and Market Behavior: Evidence from Experimental Asset Markets, Journal of Economic Psychology, 28 (3), 338-350.

Georgellis, Y., & Wall, H. J. (2005). Gender differences in self employment. *International review of applied economics*, *19*(3), 321-342.

Hanna, S. D., Kim, K. T., & Lindamood, S. (2018). Behind the numbers: Understanding the survey of consumer finances. *Journal of Financial Counseling and Planning*, *29*(2), 410.

Maxfield, S., Shapiro, M., Gupta, V., & Hass, S. (2010). Gender and risk: women, risk taking and risk aversion. *Gender in Management: An International Journal*, *25*(7), 586-604.

Riley Jr, W. B., & Chow, K. V. (1992). Asset allocation and individual risk aversion. *Financial Analysts Journal*, *48*(6), 32-37.

Rosenbaum, Paul. (2010). Design of Observational Studies. 10.1007/978-1-4419-1213-8.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, *79*(387), 516-524.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33-38.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

Rubin, D. B. (1980). Bias reduction using Mahalanobis-metric matching. *Biometrics*, 293-298.

Sapienza, P., Zingales, L., & Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences*, *106*(36), 15268-15273.

Sebai, S. (2014). Further evidence on Gender differences and their impact on Risk Aversion. *Journal of Business Studies Quarterly*, *6*(1), 308.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators?. *Journal of econometrics*, *125*(1-2), 305-353.

Conclusion

While it may be assumed that those who are more willing to take risks may find less satisfaction with the use insurance products, this study reports that, regardless of one's financial risk willingness, annuities are positively associated with higher levels of retirement income satisfaction at almost any level of financial risk willingness. Even as many studies report higher levels of retirement satisfaction with the utilization of annuities, many consumers in the United States do not participate in the annuity market.

The demographic composition of the household, with respect to same-sex versus different-sex couples, has a profound impact on the financial risk willingness of households. Those who report to be in a same-sex couple, on average, are younger, more educated, more likely to be female, less likely to have children, and are more likely to take financial risks. Over the long run, all else equal, a riskier portfolio offers a wider return interval that allows for potential higher portfolio returns, but it also has the potential to lead to larger losses to the financial portfolio.

If women, on average, earn less compared to their male counterparts, are less likely to be given the opportunity to participate in executive compensation programs, are more likely to be penalized for fertility decisions, are more likely to live longer, and are less likely to take financial risks, these factors combined have the potential to create significant retirement planning issues for women. Measuring the difference in financial risk willingness between the genders can provide an insight into which factors may have the most devastating impact on retirement accumulation and decumulation decisions.