

Who Gambles In The Stock Market?

ALOK KUMAR*

ABSTRACT

This study shows that people's propensity to gamble and their investment decisions are correlated. At an aggregate level, individual investors prefer stocks with lottery features, and like lottery demand, the demand for lottery-type stocks increases during economic downturns. In the cross-section, socioeconomic factors that induce greater expenditure in lotteries are associated with greater investment in lottery-type stocks. Further, these investment levels are higher in regions with favorable lottery environments. Because lottery-type stocks under-perform, gambling-related under-performance is greater among low-income investors who excessively overweight lottery-type stocks. Collectively, these results indicate that state lotteries and lottery-type stocks attract very similar socioeconomic clientele.

THE DESIRE TO GAMBLE IS DEEP-ROOTED in the human psyche. This fascination with games of chance can be traced back at least a few centuries. A complex set of biological, psychological, religious, and socioeconomic factors jointly determines an individual's propensity to gamble (e.g., France (1902), Brenner (1983), Walker (1992)). In this study, I investigate the extent to which people's overall attitudes toward gambling influence their stock investment decisions.

*Alok Kumar is at the McCombs School of Business, University of Texas at Austin, 1 University Station, B6600, Austin, TX 78712; *Phone: 512-232-6824; email: akumar@mail.utexas.edu*. I would like to thank two anonymous referees, an anonymous associate editor, Lucy Ackert, Warren Bailey, Brad Barber, Nick Barberis, Robert Battalio, Garrick Blalock, Markus Brunnermeier, Vidhi Chhaochharia, Sudheer Chava, Lauren Cohen, Shane Corwin, Josh Coval, Henrik Cronqvist, Steve Figlewski, Margaret Forster, Amit Goyal, Bing Han, Cam Harvey (the editor), David Hirshleifer, Scott Irwin, Narasimhan Jegadeesh, Danling Jiang, George Korniotis, Lisa Kramer, Charles Lee, Chris Malloy, Bill McDonald, Victor McGee, Stefan Nagel, Terrance Odean, Jerry Parwada, Allen Potoshman, Stefan Ruenzi, Kevin Scanlon, Paul Schultz, Mark Seasholes, Devin Shanthikumar, Bob Shiller, Sophie Shive, Kent Womack, Jeff Wurgler, Wei Xiong, Lei Yu, Eduardo Zambrano, Ning Zhu, and seminar participants at the Spring 2005 NBER Behavioral Finance Group Meeting, University of Notre Dame, 2005 EFA Meeting, 2006 AFA Meeting, Ohio State University, University of Texas at Austin, University of California at Los Angeles, Tuck School at Dartmouth, Columbia, and University of North Carolina at Chapel Hill for helpful discussions and valuable comments. In addition, I would like to thank Nick Crain, Jeremy Page, and Margaret Zhu for excellent research assistance, Itamar Simonson for making the investor data available to me, Brad Barber and Terrance Odean for answering numerous questions about the investor database, and Garrick Blalock for providing the state lottery expenditure data. I am grateful to Thomson Financial for access to their Institutional Brokers Estimate System (I/B/E/S) provided as part of a broad academic program to encourage earnings expectations research. Of course, I am responsible for all remaining errors and omissions.

Previous studies have emphasized the potential role of gambling in investment decisions (e.g., Friedman and Savage (1948), Markowitz (1952), Shiller (1989, 2000), Shefrin and Statman (2000), Statman (2002), Barberis and Huang (2008)). For instance, Markowitz (1952) conjectures that some investors might prefer to “take large chances of a small loss for a small chance of a large gain.” Barberis and Huang (2008) posit that investors might overweight low probability events and exhibit a preference for stocks with positive skewness.

In spite of its intuitive appeal, it has been difficult to gather direct evidence of gambling-motivated investment decisions for at least two reasons. First, people’s gambling preferences and their portfolio decisions are not directly observed. Second, a precise and well-established definition of stocks that might be perceived as instruments for gambling does not exist.

In this paper, I use individual investors’ socioeconomic characteristics to infer their gambling preferences and attempt to detect traces of gambling in their stock investment decisions. I conjecture that people’s gambling propensity, as reflected by their socioeconomic characteristics, would predict gambling behavior in other settings, including the stock market. This conjecture is motivated by recent research in behavioral economics, which demonstrates that people’s risk-taking propensity in one setting predicts risky behavior in other settings (e.g., Barsky, Juster, Kimball, and Shapiro (1997)).

I consider the most common form of gambling (state lotteries), where the identities of gamblers can be identified with greater ease and precision, and identify the salient socioeconomic characteristics of people who exhibit a strong propensity to play state lotteries. The extant evidence from lottery studies indicates that the heaviest lottery players are poor, young, and relatively less educated, single men, who live in urban areas and belong to specific minority (African-American and Hispanic) and religious (Catholic) groups. Therefore, a direct implication of my main conjecture is that investors with these specific characteristics would also invest disproportionately more in stocks with lottery features.

To formally define lottery-type stocks, I examine the salient features of state lotteries and also seek guidance from recent theoretical studies that attempt to characterize lottery-type stocks. Lottery tickets have very low prices relative to the highest potential payoff (i.e., the size of the jackpot), they have low negative expected returns, their payoffs are very risky (i.e., the prize distribution has extremely high variance), and, most importantly, they have an extremely small probability of a huge reward (i.e., they have positively skewed payoffs). In sum, for a very low cost, lottery tickets offer a tiny probability of a huge reward and a large probability of a small loss, where the probabilities of winning and losing are fixed and known in advance.

While any specific stock is unlikely to possess the extreme characteristics of state lotteries, particularly the huge reward to cost ratio, some stocks might share these features

qualitatively. To identify those stocks that could be perceived as lotteries, I consider three characteristics: (i) stock-specific or idiosyncratic volatility, (ii) stock-specific or idiosyncratic skewness, and (iii) stock price.

As with lotteries, if investors are searching for “cheap bets,” they are likely to find low-priced stocks attractive. Within the set of low-priced stocks, they would find stocks with high stock-specific skewness more attractive. And among the set of stocks that have low prices and high idiosyncratic skewness, stocks with greater idiosyncratic volatility are more likely to be perceived as lotteries because the level of idiosyncratic volatility could influence the estimates of idiosyncratic skewness. When volatility is high, investors might believe that the extreme return events observed in the past are more likely to be realized again. In contrast, if a low price-high skewness stock has low idiosyncratic volatility, the extreme return events observed in the past might be perceived as outliers, and the re-occurrence of that event is likely to be assigned a considerably lower probability.

With this motivation, I assume that individual investors would perceive low-priced stocks with high idiosyncratic volatility and high idiosyncratic skewness as lotteries. Subsequently, I use this empirical definition of lottery-type stocks to gather evidence of gambling-induced stock investment decisions among individual investors.

The empirical investigation is organized around four distinct themes. First, I compare the aggregate stock preferences of individual and institutional investors and examine whether individual investors exhibit a stronger preference for stocks with lottery features. Next, I investigate whether individual investors’ preferences for lottery-type stocks are stronger among socioeconomic groups that are known to exhibit strong preferences for state lotteries. I also directly examine whether investment levels in lottery-type stocks are higher in regions with more favorable lottery environments.¹ Third, I examine whether, similar to the demand for lotteries, the aggregate individual investor demand for lottery-type stocks increases during bad economic times. Last, I examine whether investment in lottery-type stocks has an adverse influence on portfolio performance. In particular, I investigate whether, like state lotteries, investment in lottery-type stocks is regressive, where low-income investors lose proportionately more from their gambling-motivated investments.

The main ingredient for my empirical analysis is a six-year panel of portfolio holdings and trades of a group of individual investors at a large U.S. discount brokerage house. Using this data set, I show that individual investors exhibit a strong preference for stocks with lottery features, whereas institutions exhibit a relative aversion for those stocks. Individual investors’ preferences for lottery-type stocks are distinct from their known preferences for small-cap stocks, value stocks, dividend paying stocks, and “attention grabbing” stocks (e.g., Barber and Odean (2000, 2001, 2008), Graham and Kumar (2006)). Over time, similar to lot-

tery demand, individual investors' aggregate demand for lottery-type stocks increases when economic conditions worsen. These aggregate-level results indicate that, similar to state lotteries, lottery-type stocks are more attractive to a relatively less sophisticated individual investor clientele.

Examining the cross-sectional differences within the individual investor category, I find that socioeconomic factors that induce higher expenditures in state lotteries are also associated with greater investments in lottery-type stocks. Poor, young, less educated, single men, who live in urban areas, undertake non-professional jobs, and belong to specific minority groups (African-American and Hispanic) invest more in lottery-type stocks. In addition, investors who live in regions with a higher concentration of Catholics (Protestants) have a stronger (weaker) preference for lottery-type stocks.

The results from cross-sectional analysis also indicate that local economic conditions and regional lottery environments influence the demand for lottery-type stocks. Investors who earn less than their neighbors (i.e., have lower "relative" income) and live in counties with higher unemployment rates invest relatively more in lottery-type stocks. In addition, the proportional investment in lottery-type stocks is higher in states that were early lottery adopters and have higher per capita lottery expenditures. Collectively, the cross-sectional results indicate that both state lotteries and lottery-type stocks act as complements and attract very similar socioeconomic clienteles.

Examining the portfolio performance of lottery investors, I find that investors who invest disproportionately more in lottery-type stocks experience greater under-performance. The average, annual, risk-adjusted under-performance that can be attributed to investments in lottery-type stocks is 1.10% and the level of under-performance is over 2.50% for investors who allocate at least one-third of their portfolios to lottery-type stocks. A typical investor would have improved the performance by 2.84% if she had simply replaced the lottery component of her portfolio with the non-lottery component. As a proportion of income, the degree of portfolio under-performance has a striking resemblance to the evidence from lottery studies. In both instances, the proportional level of under-performance is greater among low-income investors.

Taken together, the empirical results provide evidence of strong similarities between the behavior of state lottery players and individual investors who invest disproportionately more in stocks with lottery features. The findings are consistent with my main conjecture and indicate that a set of common personal attributes determines people's gambling preferences. Alternative explanations for these results based on local bias, investor overconfidence, media coverage, or microstructure effects have little empirical support.

I. Testable Hypotheses Motivated by Lottery Studies

To begin, I examine the empirical evidence from previous studies on state lotteries and develop the main testable hypotheses.

A. *Profile of Lottery Players*

The extant evidence from the state lottery literature indicates that both lottery participation rates and lottery expenditures are strongly influenced by people's socioeconomic characteristics (e.g., Kallick, Smits, Dielman, and Hybels (1979)). For instance, relatively poor individuals tend to spend a greater proportion of their income on lottery purchases (e.g., Clotfelter and Cook (1989), Clotfelter (2000), Rubinstein and Scafidi (2002)). In addition to income and wealth, age, education, gender, and marital status influence lottery purchases. Younger and less educated individuals find lotteries more attractive (e.g., Brenner and Brenner (1990)), and relative to women, men are more likely to participate and spend disproportionately more in lotteries. Further, single or divorced individuals are more active lottery players than people who are married (e.g., Clotfelter, Cook, Edell, and Moore (1999)).

Lottery studies also document that race, ethnicity, and religious affiliation influence people's attitudes toward lottery-playing and gambling. Specifically, both lottery participation rates and purchase levels are higher among African-American and Hispanic minority groups (e.g., Herring and Bledsoe (1994), Price and Novak (1999)). Among religious groups, Catholics and Jews are more active participants in lotteries compared to Protestants and Mormons (e.g., Tec (1964), Grichting (1986)).²

Geographically, lottery studies find that urban residents are more likely to buy lottery tickets and spend more on their lottery purchases than individuals in rural areas (e.g., Kallick, Smits, Dielman, and Hybels (1979)). Lottery participation rates and expenditures also vary significantly across the U.S., where the degree of popularity of lotteries reflects the overall social acceptability of gambling in the state (e.g., Clotfelter and Cook (1989)).

Examining the effects of broad macro-economic indicators (e.g., the unemployment rate), lottery studies demonstrate that people find the tiny probability of a large gain more attractive when economic opportunities are not very bright. As a result, during economic downturns, people are attracted more toward various forms of gambling, including state lotteries (Mikesell (1994)). For instance, during the Great Depression of the 1930s, the popularity of lottery-playing and gambling had increased dramatically in the U.S. (Brenner and Brenner (1990)). Sweden had experienced a similar phenomenon, where during the Great

Depression, gambling became extremely popular and gambling activities such as soccer pools were made legal (Tec (1964)).

B. Main Testable Hypotheses

Overall, the empirical evidence from lottery studies indicates that people's demographic characteristics and economic factors jointly determine the propensity to play lotteries. If lottery purchases and investments in lottery-type stocks are both influenced by a set of common personality attributes that determines people's gambling preferences and if people's gambling demands are not saturated, the behavior of state lottery players and lottery investors would exhibit similarities along multiple dimensions.

First, the socioeconomic characteristics of people who find state lotteries attractive would be similar to those of investors who exhibit a greater propensity to invest in stocks with lottery features. In particular, relatively poor, relatively less educated, young, single men who undertake non-professional jobs, live in urban areas, and belong to specific minority (African-American and Hispanic) and religious (Catholic) groups would invest disproportionately more in lottery-type stocks.

Second, local socioeconomic factors would influence investors' holdings of lottery-type stocks. In particular, if investors perceive stocks with lottery features as gambling devices, investors located in regions with more favorable lottery environments (states that adopted lotteries earlier and have higher per capita lottery expenditures) would tilt their portfolios more toward lottery-type stocks. In contrast, the demand for non-lottery-type stocks in those regions would be relatively weaker.

This conjecture is partially motivated by the observation that the demands for various gambling devices are positively correlated. For instance, lottery studies indicate that many types of gambling devices were legal in states that were early lottery adopters, while states without lotteries also had lower acceptability of other forms of gambling (Clotfelter and Cook (1989, Page 148)). In addition, survey evidence indicates that geographical regions with greater levels of lottery demand also exhibit stronger demands for other forms of gambling (Kallick, Smits, Dielman, and Hybels (1979)).

Additionally, because status-seeking individuals exhibit a stronger propensity to gamble to improve their upward social mobility (e.g., Friedman and Savage (1948), Brunk (1981), Brenner (1983), Becker, Murphy, and Werning (2000)), the level of investments in lottery-type stocks would be greater among investors who have a lower social status relative to their respective neighbors. Specifically, investors with lower income relative to their neighbors

would invest more in lottery-type stocks because relative income is a good proxy for relative social status and feeling of overall well-being (e.g., Luttmer (2005)).

Third, if economic conditions influence people’s gambling preferences, like state lotteries, the aggregate demand for lottery-type stocks would be higher in regions with relatively poor economic conditions (e.g., higher unemployment). Over time, as economic conditions change, the aggregate demands for state lotteries and lottery-type stocks would be correlated. In particular, like state lotteries, investors would exhibit a stronger preference for lottery-type stocks during bad economic times.

Overall, there are four distinct testable implications of my main conjecture:

H1: *Aggregate preference hypothesis:* Relative to institutions, individual investors exhibit stronger aggregate preference for lottery-type stocks.

H2: *Similar clienteles hypothesis:* The socioeconomic characteristics of lottery players and lottery investors are similar. Thus, state lotteries and lottery-type stocks act as complements.

H3: *Location and social mobility hypothesis:* Investors who live in regions with higher unemployment rates and favorable lottery environments, and have lower social status relative to their neighbors, allocate larger portfolio weights to lottery-type stocks.

H4: *Time-series hypothesis:* Similar to the demand for state lotteries, the aggregate demand for lottery-type stocks is higher during economic downturns.

In addition to testing these gambling-motivated hypotheses, I examine whether the propensity to gamble with lottery-type stocks adversely influences portfolio performance.

II. Data Sources

To test the gambling hypotheses, I primarily use data from a major U.S. discount brokerage house. This data set contains all trades and end-of-month portfolio positions of a set of individual investors during the 1991 to 1996 time period. There are a total of 77,995 investors in the database, of which 62,387 have traded common stocks. An average investor holds a four-stock portfolio (median is three) with an average size of \$35,629 (median is

\$13,869). For a subset of households, demographic measures, including age, income, location (zip code), total net worth, occupation, marital status, family size, gender, etc. are available. The demographic measures were compiled by Infobase Inc. in June 1997.³

I enrich the individual investor database using data from several additional sources. First, to identify the racial and ethnic characteristics, education level, and immigrant status of investors in the sample, I obtain the racial and ethnic compositions of each zip code using data from the 1990 U.S. Census. I assign each investor the appropriate zip code-level racial and ethnic characteristics. I also assume that investors who live in more educated zip codes are likely to be more educated and investors who live in zip codes with a greater proportion of foreign born people are more likely to be immigrants.

Second, to characterize the lottery environment of the state, I obtain the annual per capita lottery sales data for the 37 U.S. states in which lotteries were legal during the sample period.⁴ For a handful of states, I am also able to obtain zip code-level lottery sales data directly from the state lottery agencies. Third, I obtain the religious profile of all U.S. counties in 1990 using data from the Association of Religion Data Archives.⁵ For each county, I compute the proportion of Catholics and the proportion of Protestants. Using each investor's zip code, I assign the appropriate county-level religious characteristic to the investor.

In addition to detailed data on individual investors, I obtain quarterly institutional holdings from Thomson Financial. These data contain the end-of-quarter stock holdings of all institutions that file form 13F with the Securities and Exchange Commission. I obtain trading data from the Trade and Quote (TAQ) and the Institute for the Study of Security Markets (ISSM) databases, where small-sized trades (trade size below \$5,000) are used to proxy for retail trades.⁶

I also use data from a few other standard sources. I obtain analysts' quarterly earnings estimates from Thomson Financial's Institutional Brokers Estimate System (I/B/E/S) summary files and monthly macroeconomic data from Datastream. For each stock in the sample, I obtain the monthly prices, returns, volume turnover, and market capitalization data from the Center for Research on Security Prices (CRSP), and the quarterly book value of common equity data from COMPUSTAT. The monthly time-series of the three Fama-French factors and the momentum factor are from Kenneth French's data library while the characteristic-based performance benchmarks are from Russell Wermers' web site.⁷

***** **Table I here** *****

III. Lottery-Type Stocks

A. An Empirical Definition

Motivated by the salient features of state lotteries, I consider three stock characteristics to identify stocks that might be perceived as lotteries: (i) stock-specific or idiosyncratic volatility, (ii) idiosyncratic skewness, and (iii) stock price. At the end of month t , I compute both the idiosyncratic volatility and the idiosyncratic skewness measures using the previous six months (i.e., months $t - 6$ to $t - 1$) of daily returns data. The idiosyncratic volatility measure is the variance of the residual obtained by fitting a four-factor model to the daily stock returns time-series. To measure idiosyncratic skewness, I adopt the Harvey and Siddique (2000) method and decompose the total skewness into idiosyncratic and systematic components. Specifically, the idiosyncratic skewness is a scaled measure of the third moment of the residual obtained by fitting a two-factor model to the daily stock returns time-series, where the two factors are the excess market returns and the squared excess market returns. The stock price refers to the price at the end of month $t - 1$.

I consider all CRSP stocks and assume that stocks in the lowest k^{th} stock price percentile, the highest k^{th} idiosyncratic volatility percentile, and the highest k^{th} idiosyncratic skewness percentile are likely to be perceived as lottery-type stocks. All three sorts are carried out independently. I choose $k = 50$ to have a considerable number of lottery-type stocks in the sample, but the main results are very similar when I choose $k = 33$.

I use stock price as one of the defining characteristics of lottery-type stocks because, like lotteries, if investors are searching for cheap bets, they would naturally gravitate toward low-priced stocks. Thus, stock price is likely to be an important characteristic of stocks that might be perceived as lotteries. Within the set of low-priced stocks, investors are likely to be attracted more toward stocks that occasionally generate extreme positive returns that cannot be justified by the movements in the market. In other words, investors would find stocks with high stock-specific or idiosyncratic skewness attractive. Therefore, I use idiosyncratic skewness as the second defining characteristic of lottery-type stocks.⁸

Last, within the set of stocks that have low prices and high idiosyncratic skewness, stocks with higher stock-specific volatility are more likely to be perceived as lotteries. When idiosyncratic volatility is high, investors might believe that the extreme return events observed in the past are more likely to be repeated. In particular, if investors adopt an asymmetric weighting scheme and assign a larger weight to upside volatility and ignore or assign lower weight to downside volatility, high idiosyncratic volatility could amplify the perception of

skewness. In contrast, if a low price-high idiosyncratic skewness stock has low idiosyncratic volatility, the extreme return events observed in the past might be perceived as outliers, and the re-occurrence of an extreme return event might be assigned a low probability. Consequently, higher idiosyncratic volatility could amplify the estimates of the level of idiosyncratic skewness and the likelihood of realizing extreme positive return in the future.⁹

Strictly speaking, the three stock characteristics identify stocks that *appear* to be like lotteries, rather than stocks that are truly lotteries. Ideally, one would classify stocks with higher probability of large positive returns (i.e., positive skewness) in the *future* as lottery-type stocks. While it is conceivable that sophisticated institutional investors would be able to predict future skewness, it is unlikely that less sophisticated individual investors would be successful in identifying those predictors. They are more likely to “naïvely” extrapolate past moments into the future and would pick stocks that appear like lotteries. Because my study focuses on the investment choices of individual investors, I characterize lottery-type stocks using measures that are more likely to be used by individual investors to naïvely identify stocks with lottery features.

***** Table II here *****

B. Main Characteristics

Table II presents the sample period averages of several important characteristics of lottery-type stocks. For comparison, I also report the characteristics of non-lottery-type stocks and other remaining stocks in the CRSP universe. The non-lottery-type stock category consists of stocks that are in the highest k^{th} stock price percentile, the lowest k^{th} idiosyncratic volatility percentile, and the lowest k^{th} idiosyncratic skewness percentile. The remaining stocks are classified into the “Other Stocks” category.

Table I, Panel A provides a brief definition of the stock characteristics reported in Table II. The summary statistics indicate that lottery-type stocks have very low average market capitalization (\$31 million), low institutional ownership (7.35%), a relatively high book-to-market ratio (0.681), and lower liquidity. These stocks are also younger (mean age is about six years), have low analyst coverage (about 71% of stocks have no analyst coverage), and are mostly non-dividend-paying stocks (only 3.37% pay dividends). Given the definition of lottery-type stocks, not surprisingly, they have significantly higher volatility, higher skewness, and lower prices. Similarly, by definition, non-lottery-type stocks have diametrically opposite features, and “other stocks” have characteristics in between these two extremes.

I also find that lottery-type stocks are concentrated heavily in the energy, mining, financial services, bio-technology, and technology sectors. The industries with the lowest concentration of lottery stocks include utilities, consumer goods, and restaurants. As a group, lottery-type stocks represent 1.25% of the total stock market capitalization, but in terms of their total number, they represent about 13% of the market.

C. How Might Investors Identify Lottery-Type Stocks?

The volatility and skewness measures are difficult to compute using the standard formulas and individual investors are unlikely to compute those measures to identify lottery-type stocks. Given the clear differences between the stock characteristics of lottery-type stocks and other stocks, it is conceivable that relatively less sophisticated individual investors would use one or more of the salient stock characteristics of lottery-type stocks to identify them. Some investors might even be attracted toward certain industries, which might have strong lottery characteristics.

To formally examine whether a collection of common stock characteristics could serve as a substitute for the three lottery features, I estimate cross-sectional regressions in which one of the lottery features is the dependent variable. The set of stock characteristics reported in Table II are the independent variables. In untabulated results, I find that, although the univariate regression estimates are strong, only a handful of stock characteristics are strongly significant in a multivariate specification. When idiosyncratic skewness is the dependent variable, only 2.85% of the cross-sectional variation in skewness can be explained by these stock characteristics. Even when I include idiosyncratic volatility and stock price in the set of independent variables, the explanatory power increases to only 3.65%.¹⁰

I also examine whether the set of stock characteristics listed in Table II can serve as a substitute for idiosyncratic volatility. When idiosyncratic volatility is the dependent variable, the explanatory power is higher (= 17.33%), but a large part of the cross-sectional variation in idiosyncratic volatility is still unexplained. When I use the lottery-stock dummy as the dependent variable and estimate a logit regression, the combined explanatory power of all stock characteristics is higher (= 21.24%), but the increase is driven primarily by the presence of stock price in the dependent variable.

These regression results indicate that even a comprehensive set of stock characteristics is unlikely to serve as an effective substitute for the three lottery features. Although certain salient stock characteristics could steer investors toward lottery-type stocks, realizations of extreme returns are necessary to generate a perception of “lottery”. Investors who are

looking for cheap ways of buying a tiny probability of a very high return would most likely extrapolate the past extreme return events into the future, especially if the associated stocks have low prices and high volatility. Even if investors do not compute skewness and volatility according to the standard formulas, they would be able to discriminate between high and low volatility stocks or high and low skewness stocks. If both volatility and skewness levels are high, investors might be able to identify those stocks with even greater ease.¹¹

IV. Aggregate Preferences for Lottery-Type Stocks

In the first set of tests, I gather support for the first hypothesis (H1). Specifically, I characterize the aggregate stock preferences of individual investors and compare them with the aggregate institutional preferences.

***** **Figure 1 here** *****

A. *Lottery-Type Stocks in Aggregate Investor Portfolios*

To begin, I examine how the aggregate individual and institutional preferences for lottery-type stocks vary over time. Figure 1 shows the monthly weights allocated to lottery-type stocks in the aggregate individual and institutional portfolios, respectively. For comparison, I also show the total weight of lottery-type stocks in the aggregate market portfolio. To construct the aggregate individual investor portfolio, I combine the portfolios of all individual investors in the brokerage sample. I construct the aggregate institutional portfolio in a similar manner using the 13F institutional portfolio holdings data. I define the aggregate market portfolio by combining all stocks within the CRSP universe.

The plot indicates that, relative to the market portfolio, individual investors significantly over-weight lottery-type stocks. The average weights allocated to lottery-type stocks in the aggregate retail and market portfolios are 3.74% and 1.25%, respectively. In contrast, the average weight allocated to lottery-type stocks in the aggregate institutional portfolio is only 0.76%. The aggregate time-series results indicate that individual investors exhibit a strong preference for stocks with lottery features, while institutions exhibit a weak aversion for those stocks.

B. *Aggregate Stock Preference Measure*

To characterize investors' aggregate preferences for lottery-type stocks more accurately, I estimate stock-level pooled and cross-sectional regressions and compare the aggregate stock preferences of individual and institutional investors. In these regressions, I employ a set of stock characteristics as independent variables. The set includes measures that investors might use to identify lottery-type stocks. The stock preference in the aggregate investor portfolio is the dependent variable.

The aggregate investor preference for stock i in month t is the unexpected (or excess) portfolio weight allocated to that stock. Specifically, this measure is defined as:

$$EW_{ipt} = \frac{w_{ipt} - w_{imt}}{w_{imt}} \times 100. \quad (1)$$

Here, w_{ipt} is the actual weight assigned to stock i in the aggregate investor portfolio p in month t , and w_{imt} is the weight of stock i in the aggregate market portfolio in month t . The institutional preference for a stock is identified in an identical manner using the aggregate institutional portfolio.

If the sample investors were to randomly select stocks such that the probability of selecting a stock is proportional to its market capitalization, the weight of each stock in the aggregate investor portfolio would be equal to the weight of the stock in the aggregate market portfolio. Thus, for a given stock, a positive (negative) deviation from the expected weight in the market portfolio captures the aggregate individual investor preference (aversion) for the stock. While other benchmarks exist for measuring the expected weight of a stock in a given portfolio, I use the market-capitalization-based benchmark because it is simple and based on few assumptions.¹²

C. *Stock-Level Fama-MacBeth and Panel Regression Estimates*

In the first regression specification, the independent variables are the three measures that reflect the lottery characteristics of stocks: Idiosyncratic volatility, idiosyncratic skewness, and stock price. I estimate the regression specification at the end of each month using the Fama and MacBeth (1973) cross-sectional regression method and use the Pontiff (1996) method to correct the Fama-MacBeth standard errors for potential higher order serial correlation.¹³ To ensure that extreme values are not affecting my results, I winsorize all variables at their 0.5 and 99.5 percentile levels. I standardize both the dependent and the independent variables (mean is set to zero and the standard deviation is one) so that the coefficient estimates can be directly compared within and across regression specifications.

***** Table III here *****

The Fama-MacBeth regression estimates are presented in Table III. Column (1) reports the estimates for idiosyncratic volatility and skewness measures, and for robustness, column (2) reports the estimates for total volatility and skewness measures. The results indicate that individual investors assign a relatively larger weight to stocks with higher idiosyncratic volatility, higher idiosyncratic skewness, and lower prices. Thus, individual investors prefer to hold stocks that might be perceived as lotteries. I find that the estimates in column (2) with total volatility and skewness measures are very similar to the estimates in column (1) where I consider idiosyncratic volatility and skewness measures.

To examine which lottery characteristics have stronger influence on investors' aggregate preferences, I compare the coefficient estimates of the three characteristics that are used to define lottery-type stocks. The results indicate that idiosyncratic volatility and idiosyncratic skewness variables influence investors' aggregate preferences with similar magnitudes. Their coefficient estimates are comparable (0.056 and 0.047, respectively). I find that the stock price measure has the strongest influence on aggregate stock preferences. Specifically, the magnitude of the coefficient estimate for stock price ($= -0.191$) is more than three times stronger than the estimates of idiosyncratic volatility and idiosyncratic skewness measures.

To ensure that the stock-level regression results are not simply restating individual investors' known preferences for small-cap stocks, value stocks, dividend paying stocks, or "attention grabbing" stocks (e.g., Barber and Odean (2000, 2001, 2008), Graham and Kumar (2006)), I estimate regression specifications with several control variables. This set includes market beta, firm size, the book-to-market ratio, past twelve-month stock return, systematic skewness (or coskewness), monthly volume turnover, a dividend paying dummy, firm age, an S&P500 dummy, and a NASDAQ dummy. Similar to the three main independent variables, I measure the control variables at the end of month $t - 1$. Table I, Panels A and B provide brief definitions of all variables used in the full regression specification.

The full specification results reported in column (3) indicate that the coefficient estimates of all three lottery indicators remain significant in the presence of control variables. The coefficient estimates of control variables also have the expected signs. For instance, the coefficient estimate for *Firm Size* is strongly negative, which indicates that individual investors exhibit a preference for relatively smaller stocks. The positive coefficient estimates for the *S&P500 Dummy* and *Volume Turnover* indicate that investors exhibit a preference for relatively more visible and liquid firms. The positive coefficient for the turnover measure is also consistent with individual investors' preferences for attention-grabbing stocks. Stocks with high monthly turnover are more likely to be in the news and, thus, they are

more likely to catch the attention of individual investors. Interestingly, individual investors exhibit an aversion for stocks that have high coskewness and increase the skewness of the overall portfolio.

Although I correct the Fama-MacBeth standard errors for potential higher order auto-correlations, to further ensure that the standard error estimates are not downward biased, I estimate a panel regression specification and compute month- and firm-clustered standard errors (Petersen (2008)). The estimates are reported in Table III, column (4). I find that the panel regression estimates are qualitatively similar to the Fama-MacBeth regression estimates. In absolute terms, the coefficient estimates of volatility and skewness increase, while the coefficient estimate of stock price decreases. Nevertheless, the price coefficient estimate is almost two times the estimates of volatility and skewness, and it is still the strongest determinant of individual investors' aggregate stock preferences.

Since both dependent and independent variables have been standardized, the stock-level regression estimates are easy to interpret in economic terms. EW has a mean of 1.01% and a standard deviation of 3.45% and in column (4), the *Idiosyncratic Volatility* variable has a coefficient estimate of 0.059. This estimate implies that, all else equal, a one standard deviation increase in the idiosyncratic volatility level of a stock would induce a $0.059 \times 3.45 = 0.204\%$ increase in the EW measure for that stock. In percentage terms, relative to the mean of EW , this corresponds to about a 20% increase in EW , which is economically significant.

Among the other two lottery characteristics, the coefficient estimate for *Idiosyncratic Skewness* is slightly lower ($= 0.052$) than the volatility estimate, while *Price* has a higher estimate ($= -0.108$). The mean stock price during the sample period is \$15.51 and its standard deviation is \$15.31. Thus, all else equal, two stocks with prices of \$5 and \$20 would have a $0.108 \times 3.45 = 0.373\%$ difference in their EW measures. In percentage terms, relative to the mean of EW , this corresponds to about a 37% difference in EW .

These rough calculations indicate that the statistically significant coefficient estimates for the three lottery characteristics in stock-level regressions are also economically significant. Overall, the stock level regression estimates indicate that individual investors exhibit a strong aggregate preference for stocks with lottery features, even after I account for the known determinants of their stock preferences.

D. Aggregate Institutional Stock Preferences

Due to the aggregate summing-up constraints, the aggregate individual and institutional preferences for stocks with lottery features should roughly be opposite.¹⁴ To investigate

whether these fundamental constraints hold, I estimate stock-level cross-sectional regressions to examine the aggregate institutional preferences. These regression estimates are also presented in Table III. Columns (6) and (7) report the Fama-MacBeth and panel regression estimates, respectively.

Consistent with the summing-up constraints, I find that the individual and institutional investor groups exhibit roughly opposite preferences. Most importantly, unlike individual investors, institutions exhibit a relative aversion for stocks with lottery features, and they overweight stocks with higher coskewness. The other coefficient estimates in the institutional regression are broadly consistent with the previous evidence on aggregate institutional preferences (e.g., Bennett, Sias, and Starks (2003), Frieder and Subrahmanyam (2005)).

E. Robustness Checks for Stock-Level Regression Estimates

I conduct additional tests to ensure that the stock-level regression estimates are robust. In the first test, I ensure that my results are not strongly influenced by microstructure issues or institutional constraints. The concern might be that the results are induced mechanically by the constraints faced by individual and institutional investors. For instance, individual investors might be constrained to hold lower-priced stocks due to the small size of their portfolios. Similarly, institutional constraints such as the prudent man rules might prevent them from holding lower priced stocks (Badrinath, Gay, and Kale (1989), Del Guercio (1996)).

When I re-estimate the stock-level regressions after excluding stocks that are priced below \$5, the sub-sample coefficient estimates for both individual and institutional portfolios are very similar to the reported full sample results (see columns (5) and (8)). Thus, the stock-level regression results do not appear to be mechanically induced by potential microstructure effects or investors' constraints.

In the next set of robustness tests, I introduce several interaction terms in the regression specification to capture investors' preferences for lottery-type stocks more accurately. The interaction terms reflect the definition of lottery-type stocks more accurately. For these tests, I first define high volatility, high skewness, and low price dummy variables. The high volatility dummy is set to one for stocks that are in the highest three volatility deciles. The other two dummy variables are defined in an analogous manner. Using the three dummy variables, I define four interaction terms and include them in the regression specification. For robustness, I consider specifications for both total and idiosyncratic volatility and skewness measures. The regression estimates are presented in Table III, Panel B.

The results from the extended regression specifications indicate that the individual in-

vestors assign larger weights to stocks with higher volatility and skewness levels and lower prices. The dummy variables as well as the interaction terms have positive and statistically significant estimates for the aggregate individual investor portfolio. Moreover, the idiosyncratic and total measures yield very similar results (see columns (1) and (2)). In contrast, when I re-estimate the extended regression specification for the aggregate institutional portfolio, the dummy variables and the interaction terms have negative and statistically weaker coefficient estimates.¹⁵

Taken together, the stock-level regression estimates indicate that individual investors overweight stocks that are more likely to be perceived as lotteries, while institutions underweight those stocks. Thus, like state lotteries, stocks with lottery characteristics attract a relatively less sophisticated individual investor clientele. This evidence provides strong empirical support for the first hypothesis (H1).

V. Socioeconomic Profile of Lottery Investors

In this section, to gather support for the second and third hypotheses (H2 and H3), I examine how the preference for lottery-type stocks vary cross-sectionally within the individual investor category.

A. Measuring Individual Preference for Lottery-Type Stocks

I use five distinct but related measures to capture an investor’s preference for lottery-type stocks. I compute the lottery-stock preference measures for each investor at the end of each month and use the sample period averages to quantify an investor’s overall preference for lottery-type stocks. The preference measures in month t employ the set of lottery-type stocks identified using stock price, idiosyncratic volatility, and idiosyncratic skewness measures obtained at the end of month $t - 1$.

The first measure of lottery-stock preference (LP) of investor i in month t is the *raw portfolio weight* allocated to lottery-type stocks:

$$LP_{it}^{(1)} = \frac{\sum_{j \in \mathcal{L}_{t-1}} n_{ijt} P_{jt}}{\sum_{j=1}^{N_{it}} n_{ijt} P_{jt}} \times 100, \quad (2)$$

where \mathcal{L}_{t-1} is the set of lottery-type stocks defined at the end of month $t - 1$, N_{it} is the number of stocks in the portfolio of investor i at the end of month t , n_{ijt} is the number of shares of stock j in the portfolio of investor i at the end of month t , and P_{jt} is the price of stock j in month t .

The second lottery preference measure is the *portfolio size adjusted weight* in lottery-type stocks. I define this alternative measure because, even merely due to chance, an investor with a larger portfolio could allocate a larger weight to lottery-type stocks.¹⁶ To ensure that a large weight in lottery-type stocks is not mechanically generated by a large portfolio size, I compare the weight investor i allocates to lottery-type stocks ($LP_{it}^{(1)}$) with an expected weight of lottery-type stocks in her portfolio that is determined by the size of her portfolio. For ease of interpretation, I normalize both the actual and the expected portfolio weights such that they lie between zero and one. The second lottery preference measure is defined as the percentage difference between the actual and the expected normalized weight measures:

$$LP_{it}^{(2)} = \frac{NW_{it} - ENW_{it}}{ENW_{it}} \times 100. \quad (3)$$

In this definition, the actual and the expected normalized weights in lottery-type stocks for investor i in month t are given by:

$$NW_{it} = \frac{LP_{it}^{(1)} - \min(LP_{it}^{(1)})}{\max(LP_{it}^{(1)}) - \min(LP_{it}^{(1)})}, \text{ and} \quad (4)$$

$$ENW_{it} = \frac{PSize_{it} - \min(PSize_{it})}{\max(PSize_{it}) - \min(PSize_{it})}, \quad (5)$$

respectively. Here, $PSize_{it}$ is the total size of the stock portfolio of investor i in month t , $\min(PSize_{it})$ is the minimum portfolio size of sample investors in month t , and $\max(PSize_{it})$ is the maximum portfolio size of sample investors in month t . The $\min(LP_{it}^{(1)})$ and $\max(LP_{it}^{(1)})$ measures are defined in an analogous manner using the lottery weights of sample investors in month t .

The third lottery preference measure is the *market portfolio adjusted weight* in lottery-type stocks. I compare the raw lottery preference measure ($LP_{it}^{(1)}$) to the expected weight of lottery-type stocks determined on the basis of total market capitalization of lottery-type stocks, and obtain the excess percentage weight allocated to lottery-type stocks. Specifically, the third lottery preference measure is defined as:

$$LP_{it}^{(3)} = \frac{LP_{it}^{(1)} - LP_t^{mkt}}{LP_t^{mkt}} \times 100, \quad (6)$$

where LP_t^{mkt} is the weight allocated to lottery-type stocks in the aggregate market portfolio in month t .

In the fourth lottery-type stock preference measure, I compare an investor's preference for lottery-type stocks with her preference for non-lottery type stocks and obtain a *relative*

lottery preference measure. Specifically, the $LP_{it}^{(4)}$ measure is defined as the difference between the excess percentage weight in lottery-type stocks and the excess percentage weight in non-lottery-type stocks:

$$LP_{it}^{(4)} = \frac{LP_{it}^{(2)} - NLP_{it}^{(2)}}{NLP_{it}^{(2)}} \times 100. \quad (7)$$

Since the market capitalization of the non-lottery-type stock category is significantly higher (about 40 times) than the capitalization of lottery-type stocks, it is necessary to examine the excess weight differential. The raw weight differential does not have a very meaningful interpretation.

Last, I define a lottery-type stock preference measure using investors' trades. Because the portfolio of lottery-type stocks change monthly, under the position-based measures of lottery-type stocks, a component of the total weight in lottery-type stocks reflects an investor's "passive" preference for lottery-type stocks. This is the weight allocated to those lottery-type stocks that did not have lottery features at the time of purchase.

To identify whether investors actively seek lottery-type stocks, each month, for each investor i , I compute the buy volume for lottery-type stocks (VBL_{it}) and the total buy volume for all stocks in the portfolio (VB_{it}). The *trade-based* lottery preference measure is defined as the ratio between those two trading volume measures:

$$LP_{it}^{(5)} = \frac{VBL_{it}}{VB_{it}} \times 100. \quad (8)$$

This measure reflects the active preference of investor i for lottery-type stocks in month t .

Given the similarities in their definitions, it is not surprising that the five lottery preference measures are positively correlated. The average correlation between the position-based lottery preference measures ($LP^{(1)}$ - $LP^{(4)}$) is 0.646, while the trade-based measure ($LP^{(5)}$) has a weaker correlation with the position-based measures (average correlation = 0.521).

B. Choice of Independent Variables in Investor-Level Regressions

To characterize the heterogeneity in individual investors' preferences for lottery-type stocks, I estimate investor-level cross-sectional regressions, where the dependent variable is one of the five lottery preference measures defined in equations (2)-(8).¹⁷ A set of variables that capture investors' socioeconomic characteristics, local economic conditions, and portfolio characteristics are employed as independent variables. The focus of this analysis is on the coefficient estimates of socioeconomic variables, which could provide empirical support for the second and third hypotheses (H2 and H3).

Table I, Panel C defines all the variables used in investor-level regressions. For ease of interpretation, I group the independent variables into three broad categories. The first set contains the key demographic variables that are known to explain people’s preferences for state lotteries. The second set of independent variables contains location-based demographic measures. The last set contains a number of portfolio characteristics that serve as control variables. To ensure that investors’ demographic characteristics are not just a nonlinear function of income, I also include squared income as an additional control variable.

***** Table IV here *****

C. *Investor-Level Cross-Sectional Regression Estimates*

The investor-level cross-sectional regression estimates are presented in Table IV. In specifications (1) to (5), I use one of the five lottery preference measures ($LP^{(1)}$ to $LP^{(5)}$) as the dependent variable, respectively.¹⁸ For brevity, the coefficient estimates of all control variables are suppressed.

I find that younger, less wealthy, less educated, non-professional, and single men invest disproportionately more in lottery-type stocks. The propensity to gamble with lottery-type stocks is lower among retired investors and among those who only hold tax-deferred accounts.¹⁹ Thus, the demographic attributes that induce greater lottery participation and expenditures are also associated with greater investments in lottery-type stocks.

Examining the coefficient estimates of religion, race, and ethnicity variables, I find that the *Catholic Dummy* has the strongest influence on investor’s propensity to invest in lottery-type stocks. Specifically, investors who live in counties with a relatively greater concentration of Catholics (Protestants) invest more (less) in lottery-type stocks. The investment in lottery-type stocks is also higher in zip codes with a greater concentration of minorities (African-Americans or Hispanics) and foreign born individuals.²⁰ This evidence indicates that, like state lotteries, investment in lottery-type stocks is correlated with the religious, racial, and ethnic characteristics of individual investors.

Examining the effects of other geographical factors, I find that investors who earn less than their “neighbors” (other investors who are located within a 25-mile radius) and live in urban regions exhibit stronger preference for lottery-type stocks. This evidence indicates that to some extent gambling-motivated investments are likely to be influenced by a desire to maintain or increase upward social mobility. The local economic condition, as captured by the county unemployment rate, is also associated with investors’ decisions to hold lottery-

type stocks. In particular, consistent with the evidence from lottery studies, the propensity to gamble is greater in regions with higher unemployment rates.

Another intriguing piece of evidence that emerges from the coefficient estimates of geographical factors is that investors who live in states with favorable lottery environments invest more in lottery-type stocks. The average investment in lottery-type stocks is higher in states that adopted lotteries early and have higher per capita consumption of lotteries. Thus, greater acceptability of gambling in a state is associated with greater investment in lottery-type stocks. This direct link between lottery expenditures and investments in lottery-type stocks indicates that they act as complements.

The coefficient estimates with trade-based lottery-type stock preference measure as the dependent variable are reported as specification (5) in Table IV. I find that these coefficient estimates are qualitatively similar to the estimates obtained using the position-based lottery preference measures reported in columns (1) to (4). With the trade-based measure, investor's education level, urban location, and state lottery expenditure measures are the strongest correlates of investors' propensity to invest in lottery-type stocks.

The coefficient estimates of unreported control variables are also as expected. For instance, investors who hold better diversified portfolios and exhibit a preference for high dividend yield stocks invest less in lottery-type stocks. In contrast, investors who hold portfolios with greater industry concentration exhibit stronger preference for lottery-type stocks.

Since all variables in investor-level regressions are standardized, the coefficient estimates are easy to interpret in economic terms. For instance, *Age* has a coefficient estimate of -0.044 in the first specification that uses the $LP^{(1)}$ measure with a mean of 10.36% and a standard deviation of 16.62%. The coefficient estimate of *Age* implies that, all else equal, a one standard deviation increase in the age of an investor would be associated with a $0.044 \times 16.62 = 0.73\%$ reduction in the weight allocated to lottery-type stocks. Thus, a 65-year old investor would allocate 2.19% lower weight to lottery-type stocks than he would have at the age of 30 (a three standard deviation change in age).

Another interpretation of the *Age* coefficient estimate is that the differential in the lottery weights would be 2.19% if two investors are similar on all dimensions but their age differential is 35. If the older investor is also a Protestant, she would further reduce the weight allocated to lottery-type stocks by 0.76%, and the total weight differential would be 2.95%. In percentage terms, relative to the mean of $LP^{(1)}$, there is an economically significant 28.48% reduction in lottery weight.

D. Robustness Checks for Investor-Level Regression Estimates

To examine the robustness of the investor-level regression estimates, I conduct five sets of additional tests. First, I examine whether the cross-sectional regression estimates are sensitive to microstructure issues (e.g., large bid-ask spread) that might make the identification of lottery-type stocks noisy. I redefine the lottery-type stock preference measure, where I only consider stocks with a price above \$5. These estimates are also presented in Table IV (column (6)). I find that these results are qualitatively similar to the baseline estimates reported in column (2). Thus, investors' gambling preferences rather than microstructure effects are the primary drivers of investor-level cross-sectional regression results.

In the second robustness test, I examine whether the preference for lottery-type stocks reflects an informational advantage rather than a preference for gambling. Ivković and Weisbenner (2005) find that individual investors exhibit a preference for stocks in their vicinity, perhaps because they have better information about those stocks. Motivated by their evidence, I examine whether investment in local lottery-type stocks reflects an informational advantage.

Specifically, I compute the portfolio weight allocated to lottery-type stocks using only investors' local stocks (stocks that are within 100 miles of the investor's location) and re-estimate the investor-level cross-sectional regression. The results indicate that investors who prefer lottery-type stocks do not differentiate between local and non-local lottery-type stocks (see column (7)). The coefficient estimates with local lottery weights are very similar to those obtained using total lottery weights (see columns (1)-(5)). The similarities in these results indicate that gambling preferences rather than local bias-induced informational advantage influence the cross-sectional relation between lottery preferences and socioeconomic characteristics.²¹

In the third set of robustness tests, I entertain the possibility that a large portfolio weight in lottery-type stocks is a reflection of investor overconfidence rather than an indicator of strong lottery preference. Each of the three lottery characteristics used to define lottery-type stocks could potentially induce greater overconfidence. In particular, stocks with high idiosyncratic volatility are harder to value, provide noisier feedback, and could amplify investors' behavioral biases such as overconfidence. Volatility and skewness are positively correlated and, thus, skewness could have a similar effect on investor overconfidence. Further, higher levels of valuation uncertainty (e.g., higher levels of intangible assets) associated with low priced stocks could induce greater overconfidence (e.g., Daniel, Hirshleifer, and Subrahmanyam (1998, 2001), Hirshleifer (2001), Kumar (2008)).

To distinguish between overconfidence- and gambling-based explanations, first, I examine

whether investors who allocate a larger weight to lottery-type stocks also trade actively. Since active trading is one of the defining features of overconfidence, a positive lottery weight-turnover relation would be consistent with the conjecture that investors over-weight lottery-type stocks due to their higher levels of overconfidence. I find that investors who invest in lottery-type stocks at least once during the sample period (lottery participants) trade less frequently. The average monthly portfolio turnover of non-participants and participants are 7.05% and 6.23%, respectively. Within the group of investors who hold lottery-type stocks, portfolio turnover declines monotonically with lottery weight. For the five lottery-weight ($LP^{(1)}$) sorted quintiles, the average turnover rates are 11.34%, 7.58%, 5.37%, 4.23%, and 2.91%, respectively. If portfolio turnover is a reasonable proxy for overconfidence, this evidence indicates that larger investment in lottery-type stocks is unlikely to be induced by overconfidence.

In the second overconfidence test, I exclude investors whose portfolio turnover is in the highest quintile (active traders) and re-estimate the investor-level regression for a sub-sample of investors who trade moderately and are unlikely to exhibit the overconfidence bias. If overconfidence induces a strong relation between socioeconomic characteristics and lottery preferences, this relation would be considerably weaker for the sub-sample of investors who are unlikely to exhibit overconfidence. The sub-sample results are presented in column (8) of Table IV. I find that the sub-sample estimates are very similar to the full-sample estimates reported in column (2), which indicates that the relation between investors' socioeconomic characteristics and lottery preferences is unlikely to reflect overconfidence.

In the third overconfidence test, I examine whether overconfidence has an incremental ability to explain investors' decision to hold lottery-type stocks. For this test, I define an *Overconfidence Dummy*, which is set to one for investors who belong to the highest portfolio turnover quintile and the lowest risk-adjusted performance quintile. The measure is defined under the assumption that overconfident investors would trade most actively and those trades would hurt their portfolio performance the most. When I include the *Overconfidence Dummy* in investor-level regression specifications, I find that it has a significantly positive estimate in all instances. For instance, in specification (2), the *Overconfidence Dummy* has a significantly positive estimate (estimate = 0.079, t -statistic = 7.85) and the other coefficient estimates reported in Table IV remain very similar.²² This evidence indicates that overconfidence has an incremental ability to explain investors' preference for stocks with lottery features.²³

In addition to these new results, the main investor-level regression results presented in Table IV do not have a meaningful economic interpretation under the overconfidence-based explanation. For example, high levels of overconfidence in high unemployment regions or greater degree of overconfidence among Catholics is not predicted by any overconfidence

theory, but these results are strongly consistent with the evidence from lottery studies and have a natural interpretation under the gambling hypothesis.

In the fourth robustness test, I examine whether investors over-weight lottery stocks not because of their gambling preferences but merely because those stocks are in the news more often. Specifically, I re-estimate the investor-level regression, where the dependent variable is the first lottery preference measure, but the set of lottery-type stocks excludes stocks that have turnover in the highest quintile and are more likely to be in the news. In untabulated results, I find that the investor-level regression estimates for this sub-sample are qualitatively very similar to the full-sample estimates. For instance, the coefficient estimate for *Wealth* is -0.043 (t -statistic = -3.27), *Education* has an estimate of -0.084 (t -statistic = -4.49), and the *Catholic Dummy* has a strong positive estimate (coefficient = 0.056 , t -statistic = 4.56). This evidence indicates that news is not the primary channel through which investors identify lottery-type stocks. Investors with socioeconomic characteristics of lottery players over-weight even those lottery-type stocks that are less likely to be in the news.

In the last set of robustness tests, I investigate whether one of the lottery characteristics or some combination of those characteristics are more important for explaining investors' gambling preferences. The results are discussed in Section A.3 of the online appendix. The evidence indicates that stock price is the most important lottery characteristic, followed by idiosyncratic skewness, and the least important characteristic is idiosyncratic volatility.

E. Regional Gambling Preferences and Investments in Lottery-Type Stocks

Although the evidence from investor-level regressions indicate that local lottery environment influences the propensity to gamble with lottery-type stocks, the relation is identified with some noise because the variables used in the regression model are measured at different levels of aggregation. For greater accuracy, I re-examine the influence of local lottery environment on lottery investments using variables that are defined at the same (either zip code or state) level of aggregation.

Focusing on the relation between the aggregate measures of lottery expenditure and investment in lottery-type stocks, I find that the correlation between per capita state-level lottery expenditure and mean state-level portfolio weight in lottery-type stocks is significantly positive (correlation = 0.303 , p -value = 0.035). The correlation between lottery age and the mean state-level portfolio weight in lottery-type stocks is even stronger (correlation = 0.417 , p -value = 0.014). Surprisingly, the correlations between the mean state-level portfolio weight in non-lottery-type stocks and lottery environment measures (per capita lottery sales and

lottery age) are significantly negative (correlations are -0.172 and -0.284 , and the p -values are 0.054 and 0.033 , respectively).²⁴

Because state-level lottery sales data might be a crude proxy for regional lottery environment, I obtain zip code-level lottery sales data for several states. In the empirical exercise, I focus on the zip code-level lottery sales data for California that has the largest (about 27%) proportion of sample investors. Unfortunately, the zip code-level data are available only for the more recent time period (2005 and 2006). In spite of the non-overlapping time periods for lottery sales and brokerage data sets, I find that the zip code-level per capita lottery sales and the zip code-level investment in lottery-type stocks are positively correlated (correlation = 0.106 , p -val = 0.035). Moreover, when I sort zip codes using the per capita lottery sales measure, the $LP^{(2)}$ lottery preference measure for the lowest and the highest lottery sales deciles are 45.19% and 84.94%, respectively.

These correlation estimates indicate that the mean investment levels in lottery-type stocks are higher in regions with favorable lottery environments. In light of the extant evidence from lottery studies, this evidence indicates that individual investors are likely to perceive stocks with lottery features as valid gambling devices.²⁵

***** Table V here *****

To further quantify the relation between regional lottery environments and investments in lottery-type stocks, I estimate state-level panel regressions, where the dependent variable is the mean state-level preference for lottery-type stocks. The independent variables capture the lottery environment and the socioeconomic characteristics of the state. Table I, Panel D defines these state-level independent variables. I estimate one regression specification for each of the five lottery preference measures. The panel regression estimates are reported in Table V, where following Petersen (2008), I use state- and month-clustered standard errors to compute the t -statistics.

Consistent with the investor-level regression estimates and the correlation estimates, I find that regional lottery environment influences investors' preference for lottery-type stocks (see column (1)). The proportional investment in lottery-type stocks is higher in states with favorable lottery environments and higher unemployment rates. Adding the religion variables to the regression specification (see column (2)) does not change those estimates considerably. More importantly, I find that the effect of religious affiliation on lottery investment is evident even in the aggregate state-level regressions. The mean investment in lottery-type stocks is higher (lower) in states with a stronger concentration of Catholics (Protestants). Even when I consider other lottery preference measures (specifications (3)-(6)), the coefficient estimates are remarkably similar to the baseline estimates reported in column (2).

The correlation estimates and state-level regression estimates indicate that the demand for state lotteries and the mean state-level investment in lottery-type stocks are associated with a common set of socioeconomic characteristics. The state-level results also indicate that state lotteries do not saturate the aggregate gambling demands of state investors. Overall, the results from state-level regressions, in conjunction with the evidence from investor-level regressions, provide strong support for the second and third hypotheses (H2 and H3).

VI. Time Variation in Lottery Preferences

In this section, I test the fourth hypothesis (H4), which posits that lottery demand and aggregate demand for lottery-type stocks would be correlated over time because they are induced by common economic factors. In particular, like aggregate lottery demand, individual investors' aggregate demand for lottery-type stocks would increase during economic downturns.

A. Time Series Regression Model

I examine the time variation in the aggregate demand for lottery-type stocks by estimating the following time-series regression model:

$$\begin{aligned}
EBSI_t = b_0 &+ b_1 UNEMP_{t-1} + b_2 UEI_{t-1} + b_3 MP_{t-1} + b_4 RP_{t-1} + b_5 TS_{t-1} \\
&+ b_6 EFC_{t-1} + b_7 EFC_t \\
&+ b_8 MKTRET_{t-1} + b_9 MKTRET_t + b_{10} LOTRET_{t-1} + b_{11} LOTRET_t \\
&+ b_{12} EBSI_{t-1} + \varepsilon_t.
\end{aligned} \tag{9}$$

The dependent variable in the model is the excess buy-sell imbalance ($EBSI$) for lottery-type stocks in a given month. This measure captures the change in investors' bullishness toward lottery-type stocks relative to the change in their bullishness toward other remaining stocks. It is defined as $EBSI_t = LBSI_t - OBSI_t$, where $LBSI_t$ is the month- t buy-sell imbalance of a portfolio of lottery stocks, and $OBSI_t$ is the month- t buy-sell imbalance of a portfolio that contains other remaining stocks.²⁶ The portfolios of lottery-type stocks and other stocks are defined at the end of month $t - 1$.²⁷

The independent variables in the regression specification include the following five macroeconomic variables that vary significantly over the business cycle (Chen, Roll, and Ross (1986), Ferson and Schadt (1996)): Monthly U.S. unemployment rate ($UNEMP$), unexpected inflation (UEI), monthly growth in industrial production (MP), monthly default risk premium

(*RP*), and the term spread (*TS*). To examine whether investors' trading behavior is influenced by changes in the expected future cash-flows of lottery-type stocks, I use revisions in analysts' forecasts of future earnings (*EFC*) as a proxy for changes in investors' expectations about future cash-flows.²⁸

Additionally, investors are known to be sensitive to past returns. They might trade in response to recent market returns or returns from lottery-type stocks (e.g., Odean (1999), Barber and Odean (2008)). To capture the effects of returns on investors' trading activities, I include the market (*MKTRET*) and the lottery portfolio returns (*LOTRET*) as additional independent variables. Last, I use the one-month lagged *EBSI* variable as an explanatory variable to control for potential serial correlation in that measure. Table I, Panel E provides additional details on all variables used in the time-series regressions.

***** Table VI here *****

B. Time Series Regression Estimates using the Brokerage Data

First, I use the brokerage sample to estimate the time-series regression model. Although the January 1991 to November 1996 sample period is short, the macroeconomic variables exhibit considerable time variation during this period. Thus, if investors' propensity to invest in lottery-type stocks is influenced by changes in macroeconomic conditions, the trading intensity would vary over time and the relation between macroeconomic indicators and trading intensity could be identified.

The time series regression estimates are presented in Table VI. The results indicate that higher unemployment rates are associated with greater relative demand shifts for lottery-type stocks (coefficient estimate = 0.189, *t*-statistic = 2.54). Furthermore, *EBSI* is higher when the default risk premium is higher to compensate for the relatively poor state of the economy (coefficient estimate = 0.124, *t*-statistic = 2.75). The remaining three macroeconomic variables have statistically insignificant coefficient estimates.

The time-series regression estimates also indicate that *EFC*, which proxies for investors' changing expectations about future cash-flows, has insignificant coefficient estimates. This evidence indicates that changes in investors' trading activities in lottery-type stocks are unlikely to be driven by changing expectations about stock fundamentals. The coefficient estimates of control variables are also as expected. For instance, the lagged *EBSI* variable has a positive coefficient estimate, which indicates that there is persistence in investors' differential demand shifts.²⁹

In economic terms, the time series regression estimates are significant. During the sample period, the U.S. unemployment rate has a mean of 6.40% and a standard deviation of 0.78%, while *EBSI* has a mean of -0.75% and a standard deviation of 8.06%. Because all variables have been standardized, the unemployment rate increases of one percentage point (say, from 5% to 6%) corresponds to a 1.28 standard deviation increase in unemployment. Thus, one percentage point change in unemployment rate corresponds to a $1.28 \times 0.189 \times 8.06 = 1.95\%$ increase in *EBSI*. In absolute terms, this increase is more than 2.5 times the mean of *EBSI* and represents an economically significant shift.

C. Time Series Regression Estimates using “Small-Trades” Data

To examine the robustness of the time series regression estimates, I construct a proxy for retail trading using the ISSM and TAQ databases and re-estimate the time-series regression. One of the main advantages of the ISSM/TAQ data is that they are available from 1983 to 2000, which is considerably longer than the six-year brokerage sample. The macroeconomic variables exhibit greater variation during the 18-year period and, therefore, their potential influence on the aggregate demand for lottery-type stocks could be identified more accurately.

The small trades data capture retail trading quite well because the *BSI* time-series computed using the small-trades data is positively correlated with the *BSI* time-series obtained using the brokerage data. The correlations between the two *BSI* time-series for lottery-type stocks and other stocks are 0.504 and 0.533, respectively. Even the *EBSI* time-series obtained using the two samples have a strong, positive correlation of 0.526. These correlation estimates indicate that the small trades data from ISSM/TAQ capture retail trading reasonably well.

The time-series regression estimates indicate that the coefficient estimates obtained using the small-trades data are qualitatively very similar to those obtained using the brokerage sample (see Table VI, column (6)). For example, with the small-trades data, the coefficient estimates of lagged *UNEMP*, lagged *RP*, and *LOTRET* variables are 0.135 (t -stat = 3.14), 0.112 (t -stat = 3.16), and 0.816 (t -stat = 11.02), respectively. In comparison, using the brokerage data, the corresponding coefficient estimates are 0.189 (t -stat = 2.54), 0.124 (t -stat = 2.75), and 0.427 (t -stat = 2.07), respectively. These comparisons indicate that the time-series relation between lottery demand and economic indicators identified using the relatively short brokerage sample is quite robust.

Collectively, the time-series regression estimates using brokerage and ISSM/TAQ data indicate that, like lottery demand, investors’ propensity to buy lottery-type stocks is higher

during economic downturns. This evidence is consistent with the fourth hypothesis (H4).

VII. Lottery Preferences and Portfolio Performance

In the last set of tests, I investigate whether investment in lottery-type stocks has a positive or an adverse influence on portfolio performance. I also examine whether investment in lottery-type stocks is regressive where portfolio under-performance related to investment in lottery-type stocks decreases with income.

On the one hand, if people with strong gambling preferences find the small possibility of a very large return attractive, like lottery players, lottery investors would be willing to invest in lottery-type stocks, even when they are expected to under-perform. Furthermore, the magnitude of the under-performance induced by investment in lottery-type stocks might be greater among investors with stronger lottery preferences (e.g., low-income investors).

On the other hand, although lottery-type stocks earn lower average performance, there is significant heterogeneity in their performance. It is therefore possible that investors with strong gambling preferences are able to identify lottery-type stocks with superior performance, assign larger weight to those lottery-type stocks, and generate higher overall returns from their lottery investments. In this scenario, greater allocation in lottery-type stocks would be induced by an informational advantage rather than investors' pure gambling preferences.

***** Table VII here *****

A. Performance of Lottery-Type Stocks

Prior to estimating the potential economic costs associated with investments in lottery-type stocks, I examine the performance of lottery-type stocks. For comparison, I also report the performance of “non-lottery stocks” and “other stocks” portfolios. All three stock portfolios have been previously defined in Section III. Table VII reports the characteristics and performance of the three value-weighted portfolios.

The performance estimates indicate that lottery-type stocks earn significantly lower average returns, relative to both non-lottery and other stock categories. Specifically, relative to the non-lottery stock portfolio, the annualized raw, characteristic-adjusted, and risk-adjusted performance differentials are -7.96% , -4.98% , and -7.10% , respectively. And relative to the “other stocks” portfolio, the annualized raw, characteristic-adjusted, and risk-adjusted

performance differentials are -6.74% , -4.19% , and -6.23% , respectively. Thus, irrespective of the benchmark used and irrespective of the performance measure used, I find that lottery-type stocks earn at least 4% lower average annual returns.³⁰

B. Investment in Lottery-Type Stocks and Portfolio Performance

The lower average returns of lottery-type stocks suggests that greater investment in lottery-type stocks is likely to be associated with greater average under-performance. The exact magnitude of portfolio under-performance, however, would depend upon the subset of lottery stocks chosen by the investor, the weights allocated to those lottery-type stocks, and the holding periods of those stocks.

To isolate the level of under-performance that is associated with investment in lottery-type stocks, I estimate the degree of under-performance that is generated in a well-diversified market portfolio when a part of that portfolio is replaced by the lottery-stock component of an investor’s portfolio. This method is equivalent to replacing the non-lottery portfolio component of an investor’s portfolio by the market portfolio.³¹ I conduct this exercise for every investor who holds lottery-type stocks and obtain the risk-adjusted performance of investor-specific hypothetical portfolios.

I find that the average annualized risk-adjusted under-performance (the four-factor alpha) of hypothetical portfolios is 1.10% and it increases almost monotonically with lottery weight. For instance, investors who allocate one third of their portfolios to lottery-type stocks under-perform by about 2.50% annually on a risk-adjusted basis.

To better examine the relation between investors’ propensity to invest in lottery-type stocks and portfolio performance, I estimate cross-sectional regressions, where the dependent variable is the performance of an investor’s hypothetical portfolio. The main independent variables in these performance regressions are a lottery-stock participation dummy, one of the five lottery-stock preference measures ($LP^{(1)}-LP^{(5)}$), and a strong lottery preference dummy to capture potential nonlinearity in the lottery-type stock preference and performance relation. The strong lottery preference dummy is set to one for investors whose portfolio weights in lottery-type stocks are in the highest decile.

The performance regression specification also includes the known determinants of portfolio performance as control variables. This set contains demographic variables, including investor’s age, investment experience, the annual household income, zip code education level, a male dummy, a retired dummy, and race/ethnicity identifiers. I also consider the following four portfolio characteristics as control variables: Initial portfolio size, monthly portfolio

turnover, dividend yield of the portfolio, and portfolio diversification. Table I, Panels A and F provide additional details on these measures.

***** Table VIII here *****

The performance cross-sectional regression estimates are reported in Table VIII, where I use clustered standard errors to account for cross-sectional dependence within zip codes. In specifications (1)-(4), I consider the first lottery preference. To ensure the robustness of the performance regression estimates, I consider lottery preference measures $LP^{(2)}$ - $LP^{(5)}$ in specifications (5)-(8), respectively.³²

The coefficient estimates from specification (1) indicate that the average annual, risk-adjusted under-performance is 3.00% (0.250×12) for an investor who trades lottery-type stocks at least once during the sample period.³³ The level of under-performance is significant ($0.189 \times 12 = 2.27\%$), even after I account for other known determinants of portfolio performance (see specification (2)).

The estimates from specifications (3) and (4) indicate that the degree of under-performance is greater for investors who allocate a larger portfolio weight to lottery-type stocks. The incremental annual, risk-adjusted under-performance is 3.16% (0.263×12) for an investor who increases the investment in lottery-type stocks by one standard deviation. Furthermore, if the investor is in the highest lottery weight decile, the portfolio under-performs by an additional 0.83% (0.069×12) annually and the total annual, risk-adjusted underperformance is 3.99%.

When I consider other lottery preference measures (specifications (5)-(8)), the coefficient estimates are qualitatively similar. Even when I consider a trade-based lottery preference measure (specification (8)), the degree of under-performance that can be attributed to investment in lottery-type stocks is economically significant. The incremental annual, risk-adjusted under-performance is 3.04% for an investor who increases the buying intensity in lottery-type stocks by one standard deviation and belongs to the highest portfolio weight decile. Collectively, the performance regression estimates indicate that portfolios of investors who invest more in lottery-type stocks experience greater under-performance, even when I account for other known determinants of portfolio performance.

Because low-income and less wealthy investors invest disproportionately more in lottery-type stocks (see Table IV), their lottery-type stock investments generate larger under-performance, both when measured in dollar terms and when measured as a proportion of the total annual income. As a proportion of income, the degree of portfolio under-performance due to investments in lottery-type stocks has striking resemblance to the evidence from state

lottery studies. In both instances, the proportional level of under-performance decreases with income. Therefore, like state lotteries, lottery-type stocks appear to be regressive.

C. Performance of Lottery and Non-Lottery Portfolio Components

To better identify the mechanisms that generate portfolio under-performance, I compare the performance of lottery and non-lottery components of investor portfolios. If investors who prefer lottery-type stocks are “bad” investors in general, both the lottery and the non-lottery portfolio components of their portfolios would perform poorly and the two performance measures would be positively correlated. In contrast, if the under-performance of the lottery component of the portfolio is induced by certain specific behavioral biases that are induced exclusively or get amplified by lottery-type stocks, then the correlation between the lottery and the non-lottery portfolio components would be zero. The third possibility is that investors hold a layered portfolio that contains a large well-diversified component along with a small component containing lottery-type stocks (e.g., Shefrin and Statman (2000)). In this scenario, the non-lottery portfolio component would perform relatively better than the lottery component and the two performance measures would be negatively correlated.

In the first test, I compute the average correlation between the performance of lottery and non-lottery components of investors’ portfolios. Every month, I decompose the total portfolio performance of each investor into the performance of lottery-type stocks and non-lottery stocks and compute the time-series correlation between the two performance measures.³⁴ I find that the average lottery-non-lottery performance correlation is weak and mildly positive. The mean correlation estimate is 0.004 (t -statistic = 3.83) and the median is 0.001. This evidence is weakly consistent with the conjecture that investors earn low returns from their investments in lottery-type stocks due to their overall lack of financial sophistication.

In the second test, I compare the performance levels of lottery and non-lottery portfolio components. Specifically, I estimate the additional return a lottery investor would have earned if she had simply replaced the lottery component of her portfolio with the non-lottery component. I compute the performance of this hypothetical portfolio and examine the difference between the performance levels of the actual and the hypothetical portfolios. This performance differential would reflect both the relative under-performance of lottery-type stocks and the additional biases that an investor might exhibit when she holds lottery-type stocks.³⁵ I find that a typical lottery investor would have been able to earn an average of $0.237 \times 12 = 2.84\%$ higher annual returns if she simply replaced her lottery investments with her non-lottery investments.

To examine whether the potential for performance improvement is greater among investors with stronger gambling preferences, I estimate a cross-sectional regression where the performance differential between the actual and the hypothetical portfolios is the dependent variable. The set of independent variables is identical to the performance regression estimated earlier. I find that the level of relative under-performance is greater among investors who allocate a larger portfolio weight to lottery-type stocks (see Table VIII, specification (9)). The incremental annual, risk-adjusted relative under-performance is 2.08% (0.173×12) for an investor who increases the weight in lottery-type stocks by one standard deviation.

Taken together, these performance comparisons indicate that individual investors earn lower returns from their lottery investments. This under-performance of the lottery-type stock component of investor portfolios reflects both the under-performance of lottery-type stocks and investor' behavioral biases.

VIII. Summary and Conclusion

This paper shows that the gambling preferences of individual investors are reflected in their stock investment decisions. Using monthly portfolio holdings and trading data from a large U.S. brokerage house, I find that individual investors invest disproportionately more in stocks that have the qualitative features of state lotteries. Within the individual investor category, socioeconomic factors that induce greater expenditure in state lotteries are also associated with greater investments in lottery-type stocks. And similar to lottery demand, individual investors' demand for lottery-type stocks increases when economic conditions worsen.

Investors who invest disproportionately more in lottery-type stocks experience greater under-performance and the degree of portfolio under-performance resembles the evidence from lottery studies. In both instances, the level of under-performance as a proportion of income is greater among low-income investors.

Overall, these empirical findings indicate that state lotteries and lottery-type stocks act as complements and attract very similar socioeconomic clienteles. There are striking similarities between the behavior of state lottery players and individual investors who invest disproportionately more in stocks with lottery features.

The finding that socioeconomic characteristics of individual investors influence their stock preferences is not totally surprising because the psychological, social, economic, political, and religious identities of an individual supersede her identity as an investor. Portfolio

choice models that recognize this potential link could better explain the portfolio decisions of individual investors. Further, if socioeconomic characteristics influence portfolio choice, those characteristics could also be reflected in stock prices. For instance, the return generating process of a stock with an older investor clientele might be influenced by the preferences and biases that are unique to older investors. Similarly, it is easy to imagine a Catholic stock and a Protestant stock, an African-American stock and a White stock, or a Democrat stock and a Republican stock.

In broader terms, the evidence in the paper suggests that the link between changes in socioeconomic environment and the stock market behavior might be stronger than currently believed. For example, on the one hand, as the U.S. population becomes older, the aggregate level of gambling-motivated trading in financial markets could decline, which in turn could affect the equilibrium returns, volume, and volatility of lottery-type stocks. On the other hand, as gambling attains wider acceptability in the society and the level of gambling activities increases, the level of speculative trading in financial markets could rise. These social shifts could be associated with higher levels of trading, higher volatility, and lower average returns. Future research could examine how the interactions among different social processes influence the stock market behavior.

References

- Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.
- Badrinath, S. G., Gerald D. Gay, and Jayant R. Kale, 1989, Patterns of institutional investment, prudence, and the managerial “safety-net” hypothesis, *Journal of Risk and Insurance* 56, 605–629.
- Barber, Brad M., and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773–806.
- , 2001, Boys will be boys: Gender, overconfidence, and common stock investment, *Quarterly Journal of Economics* 116, 261–292.
- , 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies* 21, 785–818.

- , and Ning Zhu, 2008, Do noise traders move markets?, *Review of Financial Studies* Forthcoming.
- Barberis, Nicholas, and Ming Huang, 2008, Stocks as lotteries: The implications of probability weighting for security prices, *American Economic Review* Forthcoming.
- Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro, 1997, Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study, *Quarterly Journal of Economics* 112, 537–579.
- Becker, Gary S., Kevin M. Murphy, and Ivan Werning, 2000, Status, lotteries and inequality, Working Paper No. 160, George J. Stigler Center for the Study of the Economy and the State, University of Chicago.
- Bennett, James A., Richard W. Sias, and Laura T. Starks, 2003, Greener pastures and the impact of dynamic institutional preferences, *Review of Financial Studies* 16, 1203–1238.
- Blalock, Garrick, David R. Just, and Daniel H. Simon, 2007, Hitting the jackpot or hitting the skids: Entertainment, poverty, and the demand for state lotteries, *American Journal of Economics and Sociology* 66, 545–570.
- Brenner, Reuven, 1983, *History – The Human Gamble* (University of Chicago Press, Chicago, IL).
- , and Gabrielle A. Brenner, 1990, *Gambling and Speculation* (Cambridge University Press, Cambridge, UK).
- Brunk, Gregory G., 1981, A test of the Friedman-Savage gambling model, *Quarterly Journal of Economics* 96, 341–348.
- Brunnermeier, Markus K., and Jonathan A. Parker, 2005, Optimal expectations, *American Economic Review* 95, 1092–1118.
- Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2001, Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices, *Journal of Financial Economics* 61, 345–381.
- Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–403.
- Clotfelter, Charles T., 2000, Do lotteries hurt the poor? Well, yes and no, Working Paper, Terry Sanford Institute of Public Policy, Duke University, April 2000.

- , and Philip J. Cook, 1989, *Selling Hope: State Lotteries in America* (Harvard University Press, Cambridge, MA).
- , Julie A. Edell, and Marian Moore, 1999, State lotteries at the turn of the century: Report to the national gambling impact study commission, Working Paper (April), Duke University.
- Daniel, Kent D., Mark Grinblatt, Sheridan Titman, and Russell Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* 52, 1035–1058.
- Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 1998, A theory of overconfidence, self-attribution, and security market under- and over-reactions, *Journal of Finance* 53, 1835–1885.
- , 2001, Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* 56, 921–965.
- Del Guercio, Diane, 1996, The distorting effect of the prudent man law on institutional equity investments, *Journal of Financial Economics* 40, 31–62.
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.
- Ferson, Wayne E., and Rudi W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425–461.
- France, Clemens J., 1902, The gambling impulse, *American Journal of Psychology* 13, 364–407.
- Frieder, Laura, and Avanidhar Subrahmanyam, 2005, Brand perceptions and the market for common stock, *Journal of Financial and Quantitative Analysis* 40, 57–85.
- Friedman, Milton, and Leonard J. Savage, 1948, The utility analysis of choices involving risk, *Journal of Political Economy* 56, 279–304.
- Graham, John R., and Alok Kumar, 2006, Do dividend clienteles exist? Evidence on dividend preferences of retail investors, *Journal of Finance* 61, 1305–1336.
- Grichting, W.L., 1986, The impact of religion on gambling in Australia, *Australian Journal of Psychology* 38, 45–58.

- Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263–1295.
- Herring, Mary, and Timothy Bledsoe, 1994, A model of lottery participation: Demographics, context, and attitudes, *Policy Studies Journal* 22, 245–257.
- Hirshleifer, David A., 2001, Investor psychology and asset prices, *Journal of Finance* 56, 1533–1597.
- Ivković, Zoran, and Scott Weisbenner, 2005, Local does as local is: Information content of the geography of individual investors' common stock investments, *Journal of Finance* 60, 267–306.
- Kallick, Maureen, Daniel Smits, Ted Dielman, and Judith Hybels, 1979, A survey of American gambling attitudes and behavior, Survey Research Center Research Report, Institute for Social Research, University of Michigan.
- Kumar, Alok, 2008, Hard-to-value stocks, behavioral biases, and informed trading, *Journal of Financial and Quantitative Analysis* Forthcoming.
- , and Charles M.C. Lee, 2006, Retail investor sentiment and return comovements, *Journal of Finance* 61, 2451–2486.
- Luttmer, Erzo F.P., 2005, Neighbors as negatives: Relative earnings and well-being, *Quarterly Journal of Economics* 120, 963–1002.
- Markowitz, Harry, 1952, The utility of wealth, *Journal of Political Economy* 60, 151–158.
- Mikesell, John L., 1994, State lottery sales and economic activity, *National Tax Journal* 47, 165–171.
- Odean, Terrance, 1999, Do investors trade too much?, *American Economic Review* 89, 1279–1298.
- Petersen, Mitchell A., 2008, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* Forthcoming.
- Polkovnichenko, Valery, 2005, Household portfolio diversification: A case for rank-dependent preferences, *Review of Financial Studies* 18, 1467–1502.
- Pontiff, Jeffrey, 1996, Costly arbitrage: Evidence from closed-end funds, *Quarterly Journal of Economics* 111, 1135–1151.

- Price, Donald I., and E. Shawn Novak, 1999, The tax incidence of three Texas lottery games: Regressivity, race, and education, *National Tax Journal* 52, 741–751.
- Rubinstein, Ross, and Benjamin Scafidi, 2002, Who pays and who benefits? Examining the distributional consequences of the Georgia lottery for education, *National Tax Journal* 55, 223–238.
- Shefrin, Hersh M., and Meir Statman, 2000, Behavioral portfolio theory, *Journal of Financial and Quantitative Analysis* 35, 127–151.
- Shiller, Robert J., 1989, *Market Volatility* (MIT Press, Cambridge, MA.).
- , 2000, *Irrational Exuberance* (Princeton University Press, Princeton, NJ).
- Statman, Meir, 2002, Lottery players/Stock traders, *Financial Analysts Journal* 58, 14–21.
- Tec, Nechama, 1964, *Gambling in Sweden* (Bedminster Press, Totowa, NJ).
- Tversky, Amos, and Daniel Kahneman, 1992, Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty* 5, 297–323.
- Walker, Michael B., 1992, *The Psychology of Gambling* (Pergamon Press, Oxford, UK.).

Notes

¹I assume that a state that adopted state lotteries earlier and has a higher per capita lottery expenditure has a favorable lottery environment.

²Other forms of gambling such as casino gambling do not have stable and well-defined demographic characteristics. See Section A.1 of the online appendix for a brief discussion.

³Additional details on the individual investor database are available in Barber and Odean (2000) and Barber and Odean (2001).

⁴I thank Garrick Blalock for providing the lottery sales data. See Blalock, Just, and Simon (2007) for additional details about the data.

⁵The 1990 U.S. Census data are available at <http://www.census.gov/main/www/cen1990.html>. The 1990 county-level religion data are available at <http://www.thearda.com/>.

⁶Additional details on the TAQ small-trades data set, including the detailed procedure for identifying small trades, are available in Barber, Odean, and Zhu (2008).

⁷The risk factors are obtained from <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>. The Daniel, Grinblatt, Titman, and Wermers (1997) characteristics-based performance benchmarks are available at <http://www.smith.umd.edu/faculty/rwermers/ftpsite/Dgtw/coverpage.htm>.

⁸Other mechanisms can generate a preference for skewness. For instance, over-weighting of very low probability events (e.g., the probability of winning a lottery jackpot) can induce a preference for skewness (Tversky and Kahneman (1992), Polkovnichenko (2005), Barberis and Huang (2008)). Brunnermeier and Parker (2005) show that anticipatory utility (e.g., dream utility) can generate a preference for skewness in portfolio decisions.

⁹Of course, high volatility is not a characteristic that is unique to state lotteries. Other forms of gambling such as casinos also share this feature. For robustness, I examine the sensitivity of my main results by defining lottery-type stocks without the volatility characteristic. See Section A.3 of the online appendix.

¹⁰The inability of stock characteristics to explain the cross-sectional heterogeneity in skewness is consistent with the evidence in previous studies that attempt to predict skewness (e.g., Chen, Hong, and Stein (2001)).

¹¹Another channel through which investors might be driven toward lottery-type stocks is the news media. I examine this conjecture in Section V.D.

¹²For instance, one might conjecture that all stocks, irrespective of their size, would have an equal probability of being chosen. Thus, all stocks would have an expected weight of $1/N$, where N is the number of stocks available in the market.

¹³For each independent variable, I estimate an autoregressive model using the time-series of its coefficient estimates. The standard error of the intercept in this model is the autocorrelation corrected standard error of the coefficient estimate. The order of the autoregressive model is chosen such that its Durbin-Watson statistic is close to two. I find that three lags are usually sufficient to eliminate the serial correlation in errors ($DW \approx 2$).

¹⁴Very small institutions and very large and wealthy individual investors are not appropriately represented in the sample. Therefore, the summing-up constraints are not expected to hold perfectly.

¹⁵The stock-level regression results do not merely reflect the regional preferences of investors from California (27% of the sample) or the hedging preferences of mutual fund investors. When I re-estimate the stock-level regressions after excluding investors who reside in California or investors who hold mutual funds, I find that the sub-sample coefficient estimates are very similar to the full sample estimates.

¹⁶An investor holding a larger portfolio would hold a greater number of stocks and, thus, she is more likely to select stocks from the subset of lottery-type stocks. This choice need not reflect a preference for lottery-type stocks. I find that the correlation between the $LP(1)$ measure and portfolio size is significantly positive. However, the portfolio size-based adjustment used to define the $LP(2)$ measure eliminates this mechanically-induced correlation between portfolio size and portfolio weight allocated to lottery-type stocks.

¹⁷Examining the lottery-type stock participation rates, I find that the overall participation rate is about 35% and it does not vary significantly across the income and wealth categories. See Section A.2 of the online appendix for additional details.

¹⁸As before, I winsorize all variables at their 0.5 and 99.5 percentile levels and standardize both the dependent and the independent variables. I use clustered standard errors to account for cross-sectional dependence within zip codes. The estimates are very similar when I assume that data are clustered by counties or states.

¹⁹I also experiment with an interaction dummy variable in the regression specification that is set to one for investors who are retired and hold only tax-deferred accounts. I find that this interaction dummy has a marginally negative coefficient estimate (estimate = -0.014 , t -statistic = -1.56). The evidence indicates that retired investors with only tax-deferred accounts are extra cautious and allocate lower weights to lottery-type stocks.

²⁰The Catholic and Hispanic measures are positively correlated (correlation = 0.137) but they are not substitutes for each other.

²¹I also explicitly examine whether investors have superior information about local lottery-type stocks. If investors are informed, the local lottery-type stocks they buy should outperform the local lottery-type stocks they sell. However, I find that the average k -day returns following purchases is lower than the average k -day returns following sales. For $k = 5, 10, 21, 42, 63, 84, 105, 126,$ and 252 , the average post-trade buy-sell return differentials are -0.25% , -0.34% , -0.26% , -0.99% , -1.32% , -1.47% , -1.75% , -2.95% , and -5.64% , respectively. This evidence is inconsistent with the local bias-induced information asymmetry hypothesis.

²²This result is not mechanically induced. See Section A.4 of the online appendix for further details.

²³I also conduct two additional tests to entertain the overconfidence hypothesis. In the first test, I define an alternative overconfidence proxy (the difference between the average k -day returns following stock sales and purchases). Next, I consider a sub-sample of lottery stocks that have moderate levels of intangible assets and are less likely to be associated overconfidence. The relation between socioeconomic characteristics and lottery weight remains strong in both cases and I do not find evidence consistent with the overconfidence

hypothesis.

²⁴Given the positive correlation with lottery-type stocks, this negative correlation is not mechanically induced because lottery-type stocks represent only a small segment of the aggregate portfolio and there is a large “other stocks” category between the lottery-type and non-lottery-type stock categories.

²⁵While I find a strong correlation between per capita lottery sales and investment in lottery-type stocks within a region, I am unable to establish a causal link. To establish causality, one could use lottery advertisement expenses in a region as an instrument. The regional advertisement expense is likely to be an effective instrument because it would be correlated with regional lottery sales but there is no obvious link between the advertising measure and investment in lottery-type stocks within a region. Unfortunately, lottery advertising data are confidential and are not available from state lottery agencies. I thank an anonymous referee for suggesting this instrument and the associated test.

²⁶In my analysis, the composition of portfolios of lottery-type stocks and other stocks changes every month. However, the time-series regression estimates are very similar if those two portfolios are defined at the beginning of each year or the beginning of the sample period and held fixed during the entire year or the entire sample period, respectively.

²⁷The buy-sell imbalance (*BSI*) of portfolio p in month t is defined as: $BSI_{pt} = \frac{100}{N_{pt}} \sum_{i=1}^{N_{pt}} BSI_{it}$, where the *BSI* for stock i in month t is defined as $BSI_{it} = \frac{\sum_{j=1}^{D_t} (VB_{ijt} - VS_{ijt})}{\sum_{j=1}^{D_t} (VB_{ijt} + VS_{ijt})}$. Here, D_t is the number of days in month t , VB_{ijt} is the buy volume (measured in dollars) for stock i on day j in month t , VS_{ijt} is the sell volume (measured in dollars) for stock i on day j in month t , and N_{pt} is the number of stocks in portfolio p formed in month t . See Kumar and Lee (2006) for further details of the *BSI* measure, including a discussion about why an equal-weighted *BSI* measure is more appropriate for capturing shifts in investor sentiment.

²⁸If trading in lottery-type stocks is motivated mainly by investors’ gambling preferences, investors would not pay much attention to the fundamentals. Nevertheless, to choose a subset of stocks from the larger set of lottery-type stocks, they *might* consider the fundamentals.

²⁹For robustness, I consider additional lags of the *EBSI* variable in the regression specification. In untabulated results, I find that those lagged variables have statistically insignificant coefficient estimates. I also experiment with other regression specifications that include contemporaneous values of macroeconomic variables, lagged unemployment rates measured over a quarter, and innovations in unemployment rates. These estimates are qualitatively similar to the reported results.

³⁰For robustness, I also estimate Fama-MacBeth cross-sectional regressions to examine the performance of lottery-type stocks and find that lottery-type stocks earn lower average risk-adjusted returns. The results are reported in Table A.I of the online appendix. In these tests, I use a longer time period (1980 to 2005) to obtain more accurate estimates of the characteristics and performance of the three portfolios, but I also report the estimates for the 1991-1996 sample period. See Section A.5 of the online appendix for an additional discussion.

³¹I thank an anonymous referee for suggesting this test.

³²As before, to allow for direct comparisons among the coefficient estimates, I standardize all independent variables, and to keep the discussion focused on the incremental effects of investors’ preference for lottery-type stocks, I suppress the coefficient estimates of control variables.

³³The low adjusted R^2 in the cross-sectional regressions are consistent with the evidence in Barber and Odean (2001, Page 280).

³⁴The correlation is only defined for investors who hold both lottery and non-lottery type stocks.

³⁵When investors invest in lottery-type stocks, they might exhibit new types of biases or the biases that they exhibit with non-lottery stocks get amplified due to stocks’ lottery characteristics.

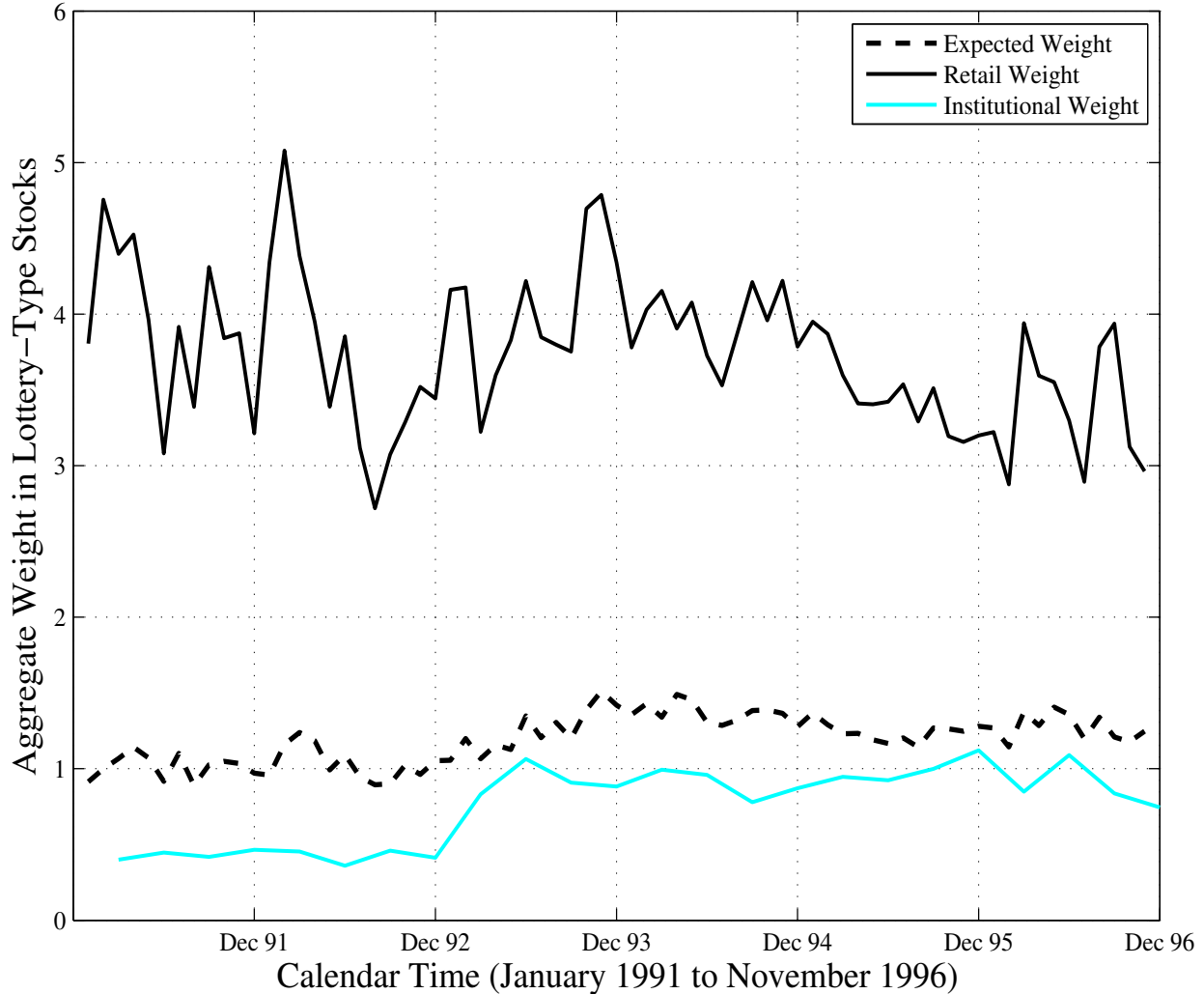


Figure 1. Aggregate weight in lottery-type stocks over time. This figure shows the time series of the actual weights allocated to lottery-type stocks in the aggregate individual and institutional investor portfolios. The expected lottery weight time-series is also shown, which reflects the weight allocated to lottery-type stocks in the aggregate market portfolio. The aggregate individual investor portfolio is formed by combining the portfolios of all individual investors in the brokerage sample. The aggregate institutional portfolio is constructed in an analogous manner using the 13F institutional portfolio holdings data. The aggregate market portfolio is obtained by combining all stocks in the CRSP universe. The stocks in the lowest k^{th} price percentile, highest k^{th} idiosyncratic volatility percentile, and highest k^{th} idiosyncratic skewness percentile are identified as lottery-type stocks. For the plot, $k = 50$. Additional details on the definition of lottery-type stocks are available in Section III.A. The individual investor data are from a large U.S. discount brokerage house for the period 1991 to 1996, while the institutional holdings data are from Thomson Financial.

Table I
Brief Definitions and Sources of Main Variables

This table briefly defines the main variables used in the empirical analysis. All volatility and skewness estimates are obtained using six months of daily stock returns. The factors employed in the multi-factor models are *RMRF* (excess market return), *SMB* (size factor), *HML* (value factor), and *UMD* (momentum factor). The data sources are: (i) ARDA: Association of Religion Data Archives, (ii) Brokerage: Large U.S. discount brokerage, (iii) BLS: Bureau of Labor Statistics, (iiii) Census: 1990 U.S. Census, (v) CRSP: Center for Research on Security Prices, (vi) DS: Datastream, (vii) IBES: Institutional Brokers Estimate System from Thomson Financial, (viii) KFDL: Kenneth French’s data library, (ix) LOTAG: State lottery agencies, and (x) 13F: 13F institutional portfolio holdings data from Thomson Financial.

Panel A: Stock Characteristics Reported in Table II

Variable Name	Description	Source
<i>Percentage of Market</i>	Weight of a stock category in the aggregate market portfolio constructed using all common stocks (share codes 10 and 11) in CRSP.	CRSP
<i>Total Volatility</i>	Standard deviation of daily stock returns.	CRSP
<i>Idiosyncratic Volatility</i>	Standard deviation of the residual from a four-factor model.	CRSP
<i>Total Skewness</i>	Scaled measure of the third moment of daily stock returns.	CRSP
<i>Idiosyncratic Skewness</i>	Scaled measure of the third moment of the residual obtained by fitting a two-factor (<i>RMRF</i> and $RMRF^2$) model.	CRSP
<i>Systematic Skewness</i>	Coefficient of the squared market factor in the skewness regression.	CRSP
<i>Stock Price</i>	End-of-month stock price.	CRSP
<i>Firm Size</i>	End-of-month market capitalization (price \times shares outstanding).	CRSP
<i>Book-To-Market Ratio</i>	Ratio of the book-value and the market capitalization of the firm.	CRSP
<i>Past 12-Month Return</i>	Total monthly stock return during the past twelve months.	CRSP
<i>RMRF, SMB, and HML Betas</i>	The loadings on <i>RMRF</i> , <i>SMB</i> , and <i>HML</i> factors in a three-factor model, respectively.	KFDL
<i>Amihud Illiquidity</i>	Absolute daily returns per unit of trading volume (Amihud (2002)).	CRSP
<i>Monthly Volume Turnover</i>	Shares traded divided by the number of shares outstanding.	CRSP
<i>Firm Age</i>	Number of years since the stock first appears in CRSP.	CRSP
<i>Percentage Dividend Paying</i>	Proportion of firm in a stock category that paid a dividend at least once during the previous one year.	CRSP
<i>% Without Analyst Coverage</i>	Proportion of firm in a stock category without analyst coverage.	IBES
<i>Mean Number of Analysts</i>	Mean number of analysts per stock.	IBES
<i>% Institutional Ownership</i>	Percentage of total shares outstanding owned by 13F institutions.	13F

Panel B: Additional Variables used in Stock-Level Regressions (Table III)

<i>Dividend Paying Dummy</i>	Set to one if the stock paid dividend during the past one year.	CRSP
<i>S&P500 Dummy</i>	Set to one if the stock belongs to the S&P500 index.	CRSP
<i>NASDAQ Dummy</i>	Set to one if the stock belongs to the NASDAQ index.	CRSP

Panel C: Variables used in Investor-Level Regressions (Table IV)

Variable Name	Description	Source
Demographic Characteristics		
<i>Wealth</i>	Total net worth of the investor.	Brokerage
<i>Income</i>	Annual household income.	Brokerage
<i>Age</i>	Age of the head of the household.	Brokerage
<i>Education</i>	Proportion of residents in investor's zip code with a Bachelor's or higher educational degree.	Census
<i>Professional Dummy</i>	Set to one if the investor belongs to one of the professional (managerial or technical) job categories.	Brokerage
<i>Retired Dummy</i>	Set to one if the head of the household is retired.	Brokerage
<i>Male Dummy</i>	Set to one if the head of the household is male.	Brokerage
<i>Married Dummy</i>	Set to one if the head of the household is married.	Brokerage
<i>Investment Experience</i>	Number of days since the brokerage account opening date.	Brokerage
<i>Taxable Account Dummy</i>	Set to one if the investor holds only taxable accounts.	Brokerage
<i>Tax Deferred Account Dummy</i>	Set to one if the investor holds only tax-deferred (IRA or Keogh) retirement accounts.	Brokerage
Location-Based Measures		
<i>Catholic (Protestant) Dummy</i>	Set to one if the proportion of Catholics (Protestants) in the county of investor's residence is greater than the mean proportion of Catholics (Protestants) across the U.S. counties.	ARDA
<i>African American-White Ratio</i>	Ratio of African-Americans and Whites in the investor's zip code.	Census
<i>Hispanic-White Ratio</i>	Ratio of Hispanics and Whites in the investor's zip code.	Census
<i>Proportion Foreign Born</i>	Proportion of foreign born residents in the investor's zip code.	Census
<i>Income Relative to Neighbors</i>	Difference between the investor's annual income and the mean income of sample investors located within 25 miles of her zip code.	Brokerage
<i>Urban Dummy</i>	Set to one if the investor resides within 100 miles of one of the largest twenty U.S. metropolitan areas.	Brokerage
<i>County Unemployment Rate</i>	Unemployment rate in the investor's county of residence.	BLS
<i>State Lottery Expenditure</i>	Mean annual per capita expenditure in state lotteries in investor's state of residence.	LOTAG
<i>State Lottery Age</i>	Number of years since the lottery adoption date in the investor's state of residence.	LOTAG
Portfolio Characteristics		
<i>Initial Portfolio Size</i>	Size of investor's stock portfolio when she enters the sample.	Brokerage
<i>Monthly Portfolio Turnover</i>	Average of buy and sell turnover rates.	Brokerage
<i>Portfolio Diversification</i>	Portfolio variance divided by the average variance of all stocks in the portfolio.	Brokerage
<i>Portfolio Dividend Yield</i>	Sample period average dividend yield of the investor portfolio.	Brokerage
<i>Portfolio Local Bias</i>	Proportion of the portfolio that is invested in stocks within 100 miles of the investor's zip code.	Brokerage
<i>Industry Concentration</i>	Largest weight allocated to one of the 48 Fama-French industries.	Brokerage
<i>Portfolio Factor Exposures</i>	<i>RMRF</i> , <i>SMB</i> , <i>HML</i> , and <i>UMD</i> betas of the investor portfolio.	Brokerage

Panel D: Variables used in State-Level Regressions (Table V)

<i>Annual Per Capita State Lottery Expenditure</i>	Annual per capita expenditure on state lotteries in the state.	LOTAG
<i>State Unemployment</i>	Monthly unemployment rate in the state.	BLS
<i>Catholic (Protestant) Dummy</i>	Set to one if the proportion of Catholics (Protestants) in the state of investor's residence is greater than the mean proportion of Catholics (Protestants) across all U.S. states.	ARDA

Panel E: Time-Series Regression Variables (Table VI)

Variable Name	Description	Source
<i>UNEMP</i>	U.S. monthly unemployment rate.	DS
<i>UEI</i>	Unexpected inflation (current inflation minus the average of the past twelve realizations).	DS
<i>MP</i>	Monthly growth in industrial production.	DS
<i>RP</i>	Monthly default risk premium (difference between Moody's Baa-rated and Aaa-rated corporate bond yields).	DS
<i>TS</i>	Term spread (difference between the yields of constant maturity ten-year treasury bond and three-month treasury bill).	DS
<i>EFC</i>	Mean monthly change in analysts' earnings forecasts of lottery-type stocks.	IBES
<i>LOTRET</i>	Mean monthly return of a portfolio of lottery-type stocks.	CRSP
<i>MKTRET</i>	Monthly market return.	CRSP

Panel F: Additional Variables used in Performance Regressions (Table VIII)

<i>Lottery-Type Stock Participation Dummy</i>	Set to one if an investor buys at least one lottery-type stock during the sample period.	Brokerage
<i>Strong Lottery-Type Stock Preference Dummy</i>	Set to one if the lottery-type stock preference measure is in the highest decile.	Brokerage

Table II
Basic Characteristics of Lottery-Type Stocks

This table reports the mean monthly characteristics of lottery-type stocks, measured during the 1991 to 1996 sample period. For comparison, the characteristics of non-lottery-type stocks and stocks that do not belong to either of the two categories (i.e., other stocks) are also reported. The stocks in all three categories are defined at the end of each month using all stocks in the CRSP universe. The stocks in the lowest k^{th} price percentile, highest k^{th} idiosyncratic volatility percentile, and highest k^{th} idiosyncratic skewness percentile are identified as lottery-type stocks. Similarly, stocks in the highest k^{th} price percentile, lowest k^{th} idiosyncratic volatility percentile, and lowest k^{th} idiosyncratic skewness percentile are identified as non-lottery-type stocks. For the results reported in the table, $k = 50$. Additional details on the definition of lottery-type stocks are available in Section III.A and all reported measures are defined in Table I, Panel A.

Measure	Lottery-Type	Non-Lottery-Type	Others
<i>Number of Stocks</i>	1,553	1,533	8,945
<i>Percentage of the Market</i>	1.25%	50.87%	47.88%
<i>Total Volatility</i>	78.57	3.29	22.14
<i>Idiosyncratic Volatility</i>	75.56	2.96	20.36
<i>Total Skewness</i>	0.330	0.175	0.237
<i>Systematic Skewness</i>	-0.202	-0.061	-0.110
<i>Idiosyncratic Skewness</i>	0.731	-0.041	0.332
<i>Stock Price</i>	\$3.83	\$31.68	\$17.51
<i>Market Beta</i>	1.090	0.906	0.897
<i>Firm Size (in Million \$)</i>	31.41	1650.87	539.66
<i>SMB Beta</i>	0.804	0.378	0.617
<i>Book-To-Market Ratio</i>	0.681	0.314	0.348
<i>HML Beta</i>	0.272	0.186	0.151
<i>Past 12-Month Return</i>	16.52%	20.22%	18.14%
<i>Amihud Illiquidity</i>	70.16	0.465	15.13
<i>Monthly Volume Turnover</i>	84.72%	64.16%	57.90%
<i>Firm Age (in Years)</i>	5.78	12.10	11.87
<i>Percentage Dividend Paying</i>	3.37%	44.59%	57.03%
<i>Percentage Without Analyst Coverage</i>	71.30%	21.19%	36.87%
<i>Mean Number of Analysts</i>	3.93	12.40	6.49
<i>Percentage Institutional Ownership</i>	7.35%	49.34%	30.09%

Table III
Aggregate Stock Preferences of Individual and Institutional Investors:
Stock-Level Regression Estimates

This table reports the Fama-MacBeth cross-sectional regression estimates (columns (1), (2), (3), and (6)) and the panel regression estimates with time fixed effects (columns (4), (5), (7), and (8)) for the aggregate individual and institutional portfolios. Panel B reports panel regression estimates from an extended specification that includes the independent variables from Panel A along with the variables shown in Panel B. The dependent variable in these regressions is the excess weight assigned to a stock in the aggregate individual or institutional portfolio (see equation (1) in Section IV.B). All independent variables are measured at the end of month $t - 1$ and have been defined in Table I, Panels A and B. Total volatility and skewness measures are used in column (2) of Panel A and columns (2) and (4) of Panel B. In the Fama-MacBeth regression estimation, I use the Pontiff (1996) method to correct the Fama-MacBeth standard errors for potential higher-order serial correlation in the coefficient estimates. In the panel regression estimation, to account for potential serial and cross correlations, I compute firm- and month-clustered standard errors. The t -statistics, obtained using corrected standard errors, are reported in smaller font below the estimates. I winsorize all variables at their 0.5 and 99.5 percentile levels. Both the dependent variable and the independent variables have been standardized (mean is set to zero and the standard deviation is one).

Panel A: Baseline Estimates

Variable	Individuals (1-5)					Institutions (6-8)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Intercept</i>	0.003	0.003	0.007			-0.041		
	10.67	9.09	3.23			-6.31		
<i>Idio or Total Volatility</i>	0.056	0.055	0.049	0.059	0.046	-0.044	-0.048	-0.051
	7.45	7.24	5.13	8.85	5.66	-4.37	-5.32	-5.78
<i>Idio or Total Skewness</i>	0.047	0.049	0.038	0.052	0.049	-0.071	-0.070	-0.066
	5.06	5.27	5.01	9.28	6.14	-5.55	-4.04	-3.69
<i>Stock Price</i>	-0.191	-0.190	-0.137	-0.108	-0.124	0.061	0.062	0.059
	-8.99	-9.93	-9.77	-7.97	-8.73	5.64	8.26	8.51
<i>Market Beta</i>			0.111	0.155	0.100	-0.006	-0.008	0.002
			6.79	8.69	8.03	-2.47	-2.18	0.37
<i>Log(Firm Size)</i>			-0.189	-0.200	-0.183	0.185	0.220	0.156
			-10.41	-10.79	-8.81	4.95	10.57	3.16
<i>Book-To-Market Ratio</i>			-0.071	-0.086	-0.064	0.052	0.058	0.065
			-7.63	-10.92	-6.32	4.12	5.58	3.28
<i>Past 12-Month Stock Return</i>			-0.015	-0.013	-0.021	0.024	0.030	0.012
			-2.51	-1.95	-2.53	3.40	6.96	1.87
<i>Systematic Skewness</i>			-0.012	-0.017	-0.011	0.020	0.014	0.001
			-3.36	-3.15	-2.34	2.03	2.26	0.06

Variable	Individuals (1-5)				Institutions (6-8)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Monthly Volume Turnover</i>			0.125	0.133	0.150	-0.032	-0.038	-0.037
			8.72	6.77	9.51	-6.51	-5.07	-3.70
<i>Dividend Paying Dummy</i>			-0.069	-0.100	-0.074	0.014	0.016	0.012
			-7.62	-11.53	-7.37	4.37	2.31	3.14
<i>Firm Age</i>			-0.038	-0.069	-0.047	0.014	0.016	0.011
			-6.63	-6.65	-7.17	3.20	2.75	2.02
<i>S&P 500 Dummy</i>			-0.004	-0.005	-0.008	0.012	0.012	0.014
			-2.51	-1.90	-1.96	3.25	4.83	4.22
<i>NASDAQ Dummy</i>			0.033	0.024	0.030	-0.016	-0.023	-0.004
			2.96	4.47	3.50	-3.44	-5.81	-1.12
<i>(Mean) Number of Observations</i>	5,979	5,979	5,310	377,010	256,813	4,238	101,761	78,028
<i>(Mean) Adjusted R²</i>	0.049	0.050	0.116	0.103	0.114	0.109	0.132	0.141

Panel B: Robustness Test Results (Panel Regression Estimates)

Variable	Individuals (1-2)		Institutions (3-4)	
	(1)	(2)	(3)	(4)
<i>High Volatility Dummy</i>	0.051	0.053	-0.013	-0.019
	5.22	6.90	-2.32	-2.17
<i>High Skewness Dummy</i>	0.046	0.042	-0.032	-0.042
	3.25	3.39	-3.03	-2.16
<i>Low Price Dummy</i>	0.092	0.107	-0.034	-0.031
	6.51	6.09	-3.17	-3.00
<i>High Volatility × High Skewness</i>	0.074	0.079	-0.010	-0.011
	6.32	5.37	-1.87	-1.96
<i>High Volatility × Low Price</i>	0.028	0.025	-0.005	-0.004
	8.71	7.01	-1.16	-1.13
<i>High Skewness × Low Price</i>	0.017	0.016	-0.007	-0.007
	3.27	3.70	-1.03	-0.92
<i>High Skewness × High Volatility × Low Price</i>	0.046	0.048	-0.037	-0.033
	4.26	4.62	-2.08	-1.88

(Other coefficient estimates have been suppressed.)

Table IV
Investor Characteristics and Preference for Lottery-Type Stocks:
Cross-Sectional Regression Estimates

This table reports the estimates of investor-level cross-sectional regressions, where the dependent variable is a measure of investor's preference for lottery-type stocks. The lottery-type stock preference measures are defined in Section V.A and all explanatory variables are defined in Table I, Panel C. In specifications (1)-(5), one of the lottery-type stock preference measures ($LP^{(1)}$ - $LP^{(5)}$, respectively) is used as the dependent variable. In specification (6), the dependent variable is the $LP^{(2)}$ measure, but stocks with price below \$5 are excluded from the analysis. In specification (7), I use the $LP^{(1)}$ preference measure for local lottery-type stocks only. In specification (8), I also use the $LP^{(1)}$ preference measure, but I exclude active traders (investors with portfolio turnover in the top quintile) from the sample. In all specifications, the set of control variables includes portfolio size, monthly portfolio turnover, portfolio diversification, local bias, portfolio dividend yield, portfolio industry concentration, the four factor exposures of the portfolio, and squared income. For brevity, the coefficient estimates of these control variables are suppressed. The t -statistics for the coefficient estimates are reported in smaller font below the estimates. Both the dependent and independent variables have been standardized such that each variable has a mean of zero and a standard deviation of one.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Intercept</i>	-0.019	0.008	-0.006	-0.007	-0.009	0.020	-0.010	-0.013
	-1.30	0.82	-0.68	-0.48	-0.91	1.48	-0.97	-0.90
<i>Wealth</i>	-0.052	-0.081	-0.051	-0.068	-0.059	-0.110	-0.057	-0.040
	-3.64	-5.30	-4.40	-4.43	-4.95	-4.51	-3.97	-3.33
<i>Age</i>	-0.044	-0.063	-0.051	-0.064	-0.038	-0.114	-0.048	-0.039
	-4.43	-5.77	-3.54	-5.76	-3.42	-7.15	-4.38	-3.82
<i>Zip Code Education</i>	-0.052	-0.061	-0.046	-0.060	-0.036	-0.126	-0.051	-0.073
	-5.29	-6.23	-4.68	-5.99	-3.65	-3.71	-4.11	-5.27
<i>Professional Dummy</i>	-0.041	-0.023	-0.035	-0.027	-0.011	-0.059	-0.021	-0.036
	-3.70	-2.06	-3.17	-2.48	-1.97	-2.75	-1.98	-2.24
<i>Retired Dummy</i>	-0.037	-0.041	-0.027	-0.040	-0.011	-0.048	-0.036	-0.039
	-2.99	-3.33	-2.20	-3.23	-1.85	-2.71	-3.03	-2.17
<i>Male Dummy</i>	0.033	0.024	0.023	0.036	0.034	0.046	0.027	0.029
	3.12	2.16	2.10	3.45	3.03	2.73	2.51	2.59
<i>Married Dummy</i>	-0.016	-0.010	-0.023	-0.011	-0.021	-0.030	-0.023	-0.011
	-1.55	-1.10	-2.19	-1.19	-1.99	-2.06	-2.16	-1.72
<i>Investment Experience</i>	0.067	0.064	0.081	0.045	0.004	0.019	0.063	0.079
	7.23	6.96	8.79	4.79	0.41	2.76	6.43	5.49
<i>Taxable Account Only Dummy</i>	0.034	0.027	0.027	0.018	0.045	0.030	0.029	0.023
	3.58	2.24	2.20	1.79	4.76	2.48	2.88	1.89
<i>Tax Deferred Acc. Only Dummy</i>	-0.022	-0.053	-0.015	-0.025	-0.029	-0.067	-0.048	-0.033
	-2.39	-4.67	-1.31	-2.54	-3.24	-5.24	-4.05	-1.95

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Catholic County Dummy</i>	0.053	0.049	0.043	0.035	0.032	0.058	0.078	0.044
	5.13	3.69	3.68	3.85	2.77	3.98	7.39	3.79
<i>Protestant County Dummy</i>	-0.046	-0.035	-0.027	-0.024	-0.030	-0.051	-0.040	-0.051
	-3.19	-3.38	-3.69	-2.12	-3.38	-4.19	-4.54	-4.24
<i>Zip Code Afr. Am.-White Ratio</i>	0.028	0.025	0.028	0.023	0.019	0.021	0.023	0.030
	3.51	3.67	2.64	3.28	2.14	2.11	3.19	2.96
<i>Zip Code Hispanic-White Ratio</i>	0.034	0.031	0.040	0.028	0.020	0.046	0.036	0.034
	2.97	3.07	3.56	3.19	2.20	3.00	3.28	3.38
<i>Zip Code Prop Foreign Born</i>	0.020	0.017	0.011	0.021	0.011	0.007	0.016	0.020
	2.28	2.01	1.64	2.18	1.45	0.74	1.89	2.18
<i>Income Relative To Neighbors</i>	-0.040	-0.042	-0.031	-0.044	-0.094	-0.103	-0.045	-0.037
	-3.61	-4.97	-3.07	-5.14	-8.65	-8.34	-4.92	-3.52
<i>Urban Dummy</i>	0.030	0.032	0.025	0.018	0.017	0.004	0.029	0.015
	2.32	3.15	2.81	2.75	2.31	1.28	2.66	1.85
<i>County Unemployment Rate</i>	0.035	0.026	0.030	0.027	0.017	0.027	0.031	0.034
	3.94	2.91	3.14	2.99	2.50	2.30	4.14	2.64
<i>State Lottery Expenditure</i>	0.031	0.026	0.027	0.030	0.023	0.025	0.027	0.024
	3.17	2.66	2.97	3.11	1.96	2.15	2.46	2.77
<i>State Lottery Age</i>	0.044	0.046	0.037	0.036	0.031	0.039	0.051	0.030
	4.60	4.75	3.75	3.70	2.51	2.69	3.97	3.12
<i>Number of Investors</i>	21,194	21,194	21,194	21,194	18,650	21,194	21,194	16,955
<i>Adjusted R²</i>	0.043	0.058	0.035	0.052	0.031	0.061	0.055	0.040

Table V
State-Level Preference for Lottery-Type Stocks:
Panel Regression Estimates

This table reports the estimates from state-level panel regressions with month fixed effects. The dependent variable is the average weight allocated to lottery-type stocks by brokerage investors in state i in month t . In specifications (1) and (2), the first lottery preference measure is used. Specifications (3)-(6) use lottery preference measures $LP^{(2)}$ - $LP^{(6)}$, respectively. The lottery-type stock preference measures are defined in Section V.A. The set of control variables includes mean investor age, mean household income, squared income, mean education level, proportion of male population in the state, proportion married, proportion African American, proportion Hispanic, proportion foreign born, and mean local bias of investors in the state. For brevity, the coefficient estimates of the control variables have been suppressed. Additional details on main independent variables are provided in Table I, Panel D. I use state- and month-clustered standard errors to compute the t -statistics. The t -statistics for the coefficient estimates are reported in smaller font below the estimates. Both the dependent and independent variables have been standardized such that each variable has a mean of zero and a standard deviation of one.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
<i>Annual Per Capita State Lottery Expenditure</i>	0.057	0.050	0.107	0.037	0.109	0.121
	4.99	3.66	3.36	2.04	3.69	5.49
<i>State Lottery Age</i>	0.136	0.107	0.228	0.109	0.192	0.174
	2.66	2.44	8.98	7.63	6.45	6.92
<i>Monthly State Unemployment Rate</i>	0.068	0.062	0.026	0.047	0.126	0.129
	6.85	2.56	2.17	2.18	5.98	2.02
<i>Catholic State Dummy</i>		0.231	0.192	0.096	0.232	0.259
		6.44	7.91	3.65	9.85	2.92
<i>Protestant State Dummy</i>		-0.094	-0.114	-0.116	-0.176	-0.098
		-2.65	-3.65	-3.45	-6.04	-4.10
<i>(Estimates of control variables have been suppressed.)</i>						
<i>Number of Observations</i>	2,236	2,236	2,236	2,236	2,236	2,236
<i>Adjusted R²</i>	0.024	0.047	0.097	0.067	0.092	0.057

Table VI
Macro-Economic Conditions and Demand Shifts:
Time Series Regression Estimates

This table reports the estimation results for the time-series regression model defined in equation (9). The dependent variable is the excess buy-sell imbalance ($EBSI$) in month t . Among the independent variables, $UNEMP_t$ is the U.S. unemployment rate in month t , UEI_t is the unexpected inflation in month t , MP_t is the monthly growth in industrial production, RP_t is the monthly risk premium, TS_t is the term spread, EFC_t is the mean change in analysts' earnings forecasts of lottery-type stocks in month t , $MKTRET_t$ is the monthly market return, and $LOTRET_t$ is the mean monthly return on lottery-type stocks. Table I, Panel E provides additional details on the independent variables. In specifications (1)-(5), $EBSI$ is computed using the individual investor data from a large U.S. discount brokerage house for the 1991 to 1996 period. In specification (6), I use a proxy for retail trading obtained from the ISSM and TAQ databases for the 1983 to 2000 period. Additional details on the regression specification are available in Section VI.A. Both the dependent variable and the independent variables have been standardized. The Newey-West adjusted t -statistics for the coefficient estimates are reported in smaller font below the estimates.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
<i>Intercept</i>	0.001	0.972	1.399	0.972	0.023	-0.010
	0.20	1.288	1.549	1.099	0.25	-0.16
<i>Lagged UNEMP</i>	0.202				0.189	0.135
	2.74				2.54	3.14
<i>Lagged UEI</i>		-0.091			-0.072	-0.053
		-1.36			-0.77	-1.05
<i>Lagged MP</i>		-0.013			0.016	-0.022
		-0.69			0.20	-0.45
<i>Lagged RP</i>		0.507			0.124	0.112
		4.59			2.75	3.16
<i>Lagged TS</i>		-0.114			-0.012	-0.066
		-0.44			-0.50	-0.91
<i>Lagged EFC</i>			0.034		0.071	0.014
			0.78		1.59	1.44
<i>EFC</i>			0.003		-0.014	0.005
			0.40		-0.17	0.32
<i>Lagged LOTRET</i>				0.102	-0.037	-0.029
				1.02	-0.12	-0.32
<i>LOTRET</i>				0.498	0.427	0.816
				3.54	2.07	11.02
<i>Lagged MKTRET</i>				-0.047	-0.017	-0.054
				-1.33	-0.51	-0.83
<i>MKTRET</i>				-0.127	-0.111	-0.219
				-1.56	-1.45	-3.30
<i>Lagged EBSI</i>					0.412	0.215
					3.99	3.00
<i>Number of Months</i>	71	71	71	71	70	215
<i>Adjusted R²</i>	0.085	0.204	0.005	0.181	0.396	0.493

Table VII
Performance of Lottery-Type Stocks:
Time Series Regression Estimates

This table reports the characteristics and performance of three value-weighted portfolios for the 1980 to 2005 period: lottery-type stocks, non-lottery stocks, and other stocks. The construction of these three stock portfolios is described in Section III. The following performance measures are reported: Mean monthly portfolio return (MeanRet), the standard deviation of monthly portfolio returns (StdDev), characteristic-adjusted mean monthly portfolio return (CharAdjRet), and the intercept (Alpha) as well as the factor exposures (RMRF, SMB, HML, and UMD are the exposures to the market, size, value, and momentum factors, respectively) from a four-factor model. The characteristic-adjusted returns are computed using the Daniel, Grinblatt, Titman, and Wermers (1997) method. Only stocks with CRSP share code 10 and 11 are included in the analysis. The t -statistics for the coefficient estimates are reported in smaller font below the estimates.

Portfolio	MeanRet	StdDev	CharAdjRet	Alpha	RMRF	SMB	HML	UMD	Adj. R^2
<i>Lottery (L)</i>	0.472	7.934	-0.375	-0.552	1.204	1.130	-0.049	-0.442	
			<small>-2.95</small>	<small>-3.22</small>	<small>18.90</small>	<small>15.15</small>	<small>-0.78</small>	<small>-8.05</small>	
<i>Non-Lottery (NL)</i>	1.135	4.025	0.040	0.041	0.920	-0.123	0.102	-0.008	
			<small>0.47</small>	<small>0.84</small>	<small>28.39</small>	<small>-8.16</small>	<small>5.77</small>	<small>-0.82</small>	
<i>Others (O)</i>	1.033	4.644	-0.026	-0.033	0.959	0.099	-0.103	-0.010	
			<small>-1.12</small>	<small>-0.83</small>	<small>18.39</small>	<small>7.90</small>	<small>-6.82</small>	<small>-1.19</small>	
<i>L-NL</i>	-0.663	5.882	-0.415	-0.592	0.284	1.253	-0.151	-0.433	
			<small>-2.95</small>	<small>-3.14</small>	<small>-3.12</small>	<small>6.15</small>	<small>12.13</small>	<small>-2.17</small>	
<i>L-O</i>	-0.562	4.846	-0.349	-0.519	0.244	1.031	0.051	-0.431	
			<small>-2.57</small>	<small>-2.93</small>	<small>-3.08</small>	<small>5.96</small>	<small>10.61</small>	<small>0.83</small>	

Table VIII
Preference for Lottery-Type Stocks and Portfolio Performance:
Cross-Sectional Regression Estimates

This table reports the estimates for performance cross-sectional regressions. In specifications (1)-(8), the dependent variable is the risk-adjusted performance measure (four-factor alpha) of a hypothetical portfolio that is formed by replacing the non-lottery component of an investor portfolio by the market portfolio. Lottery-type stocks have been defined in Section III.A. In specification (9), the dependent variable is the performance differential between the actual and a hypothetical portfolio that is defined by replacing the lottery component of an investor portfolio by the non-lottery component of her portfolio. The set of independent variables include a participation dummy, lottery-type stock preference measure, and a strong lottery-type stock preference dummy. In specifications (3) and (4), the main independent variable is the $LP^{(1)}$ lottery-stock preference measure. Specifications (5)-(8) use one of the lottery-type stock preference measures ($LP^{(2)}$ - $LP^{(5)}$, respectively) as the main independent variable. The set of control variables includes investor's age, investment experience, the annual household income, zip code education level, a male dummy, a retired dummy, two race/ethnicity dummies, initial portfolio size, monthly portfolio turnover, dividend yield of the portfolio, and portfolio diversification. For brevity, the coefficient estimates of the control variables have been suppressed. The lottery-type stock preference measures have been defined in Section V.A and other variables have been defined in Table I, especially Panel F. Clustered standard errors are used to account for potential cross-sectional dependence within zip codes. All independent variables have been standardized. The t -statistics for the coefficient estimates are reported in smaller font below the estimates.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Intercept</i>	-0.368	-0.273	-0.317	-0.309	-0.291	-0.292	-0.292	-0.294	-0.157
	-6.78	-3.09	-6.19	-3.89	-3.74	-2.76	-2.73	-2.76	-2.43
<i>Lott-Stock Part. Dummy</i>	-0.250	-0.189							
	-4.59	-3.32							
<i>Lott-Type Stock Pref</i>			-0.402	-0.263	-0.299	-0.226	-0.237	-0.193	-0.173
			-8.51	-4.86	-3.88	-3.13	-3.29	-2.88	-2.73
<i>Strong Lott. Pref. Dummy</i>			-0.069	-0.048	-0.079	-0.078	-0.075	-0.070	-0.008
			-2.36	-2.17	-1.75	-1.73	-1.71	-2.16	-0.36
<i>Control Variables</i>	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
<i>(Estimates of control variables have been suppressed.)</i>									
<i>Number of Investors</i>	40,476	27,565	34,588	26,204	25,229	25,229	25,229	25,229	25,229
<i>Adjusted R²</i>	0.009	0.039	0.017	0.041	0.039	0.042	0.040	0.036	0.045