THE JOURNAL OF FINANCE • VOL. LXIV, NO. 4 • AUGUST 2009

Who Gambles in the Stock Market?

ALOK KUMAR*

ABSTRACT

This study shows that the propensity to gamble and investment decisions are correlated. At the 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, lottery investment levels are higher in regions with favorable lottery environments. Because lottery-type stocks underperform, gambling-related underperformance is greater among low-income investors who excessively overweight lottery-type stocks. These results indicate that state lotteries and lottery-type stocks attract very similar socioeconomic clienteles.

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.

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

*Alok Kumar is at the McCombs School of Business, University of Texas at Austin. 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; Sudheer Chava; Vidhi Chhaochharia; 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 Poteshman; 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 University, 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 its 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.

(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 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. Specifically, I conjecture that people's gambling propensity, as reflected by their socioeconomic characteristics, predicts gambling behavior in other settings, including the stock market. This conjecture is motivated by recent research in behavioral economics that demonstrates that people's risk-taking propensity in one setting predicts risky behavior in other settings (e.g., Barsky et al. (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 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 are likely to 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 perceive low-priced stocks with high idiosyncratic volatility and high idiosyncratic skewness as lotteries. Therefore, 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. Finally, 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 data set for my empirical analysis is a 6-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 lottery 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 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 nonprofessional jobs, and belong to specific minority groups (African-American and

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

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 state lotteries and lottery-type stocks act as complements and attract very similar socioeconomic clienteles.

Turning to the portfolio performance of lottery investors, I find that investors who invest disproportionately more in lottery-type stocks experience greater underperformance. The average, annual, risk-adjusted underperformance that can be attributed to investments in lottery-type stocks is 1.10% and the level of underperformance 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 performance by 2.84% if she had simply replaced the lottery component of her portfolio with the nonlottery component. As a proportion of income, the degree of portfolio underperformance has a striking resemblance to the evidence from lottery studies. In both instances, the proportional level of underperformance 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.

The balance of the paper is organized as follows. In the next section, I use the salient findings from the literature on state lotteries to develop the key testable hypotheses. In Section II, I describe the data sources. In Section III, I formally define lottery-type stocks and using the definition of lottery-type stocks, in Sections IV to VII, I present the main empirical results. I conclude in Section VIII with a brief summary.

I. Testable Hypotheses Motivated by Lottery Studies

In this section, I examine the empirical evidence from previous studies on state lotteries and develop this paper's main testable hypotheses.

A. *Profile of Lottery Players*

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 et al. (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)). Beyond income and wealth, age, education, gender, and marital status influence lottery purchases. In particular, younger and less educated individuals find lotteries more attractive (e.g., Brenner and Brenne (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 et al. (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 et al. (1979)). Lottery participation rates and expenditures also vary significantly across the United States, 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 macroeconomic 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 United States. (Brenner and Brenner (1990)). Sweden 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 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 gambling preferences, and if people's gambling demands are not saturated, then the

[2] Other forms of gambling such as casino gambling do not have stable and well-defined demographic characteristics. See Section A of the Internet Appendix for a brief discussion. An Internet Appendix for this article is online in the "Supplements and Datasets" section at http://www.afajof.org/supplements.asp.

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 should be similar to those of investors who exhibit a greater propensity to invest in stocks with lottery features. In particular, relatively poor, less educated, young, single men who undertake nonprofessional jobs, live in urban areas, and belong to specific minority (African-American and Hispanic) and religious (Catholic) groups are expected to 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) can be expected to tilt their portfolios more toward lottery-type stocks. In contrast, the demand for nonlottery-type stocks in those regions should be relatively weaker. This conjecture is partially motivated by the observation that the demand levels 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)). In addition, survey evidence indicates that geographical regions with greater levels of lottery demand also exhibit stronger levels of demand for other forms of gambling (Kallick et al. (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 should be greater among investors who have a lower social status relative to their respective neighbors. Specifically, investors with lower income relative to their neighbors are expected to invest more in lottery-type stocks because relative income is a good proxy for relative social status and a feeling of overall well-being (e.g., Luttmer (2005)).

Third, if economic conditions influence an individual's gambling preference, then like state lotteries, the aggregate demand for lottery-type stocks should be higher in regions with relatively poor economic conditions (e.g., higher unemployment). Over time, as economic conditions change, the aggregate levels of demand for state lotteries and lottery-type stocks should be correlated. In particular, like state lotteries, investors are likely to 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 who 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 outlined above, 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 period. There are a total of 77,995 investors in the database, of which 62,387 trade 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 sample investors' racial and ethnic characteristics, education level, and immigrant status, 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.

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

[4] I thank Garrick Blalock for providing the lottery sales data. See Blalock et al. (2007) for additional details about the data.

[5] The 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/.

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 monthly price, return, volume turnover, and market capitalization data from the Center for Research on Security Prices (CRSP), and quarterly book value of common equity 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 presents definitions and sources of the variables used in the empirical analysis.

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 6 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 total skewness into idiosyncratic and systematic components. Specifically, 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

[6] Additional details on the TAQ small-trades data set, including the detailed procedure for identifying small trades, are available in Barber et al. (2009).

[7] The risk factors are obtained from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/. The Daniel et al. (1997) characteristics-based performance benchmarks are available at http://www.smith.umd.edu/faculty/rwermers/ftpsite/Dgtw/coverpage.htm.

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 6 months of daily stock returns. The factors employed in the multifactor models are *RMRF* (excess market return), *SMB* (size factor), *HML* (value factor), and *UMD* (momentum factor). The data sources are as follows: (i) ARDA: Association of Religion Data Archives, (ii) Brokerage: Large U.S. discount brokerage, (iii) BLS: Bureau of Labor Statistics, (iv) 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.

Variable Name	Description	Source
Panel A: Stock Characteristics Reported in Table II		
Percentage of market	Weight of a stock category in the aggregate market portfolio constructed using all common stocks (share codes 10 and 11) in CRSP.	CRSP
Total volatility	Standard deviation of daily stock returns.	CRSP
Idiosyncratic volatility	Standard deviation of the residual from a four-factor model.	CRSP
Total skewness	Scaled measure of the third moment of daily stock returns.	CRSP
Idiosyncratic skewness	Scaled measure of the third moment of the residual obtained by fitting a two-factor (*RMRF* and $RMRF^2$) model.	CRSP
Systematic skewness	Coefficient of the squared market factor in the skewness regression.	CRSP
Stock price	End-of-month stock price.	CRSP
Firm size	End-of-month market capitalization (price × shares outstanding).	CRSP
Book-to-market ratio	Ratio of the book-value and the market capitalization of the firm.	CRSP
Past 12-month return	Total monthly stock return during the past 12 months.	CRSP
RMRF, *SMB*, and *HML* betas	The loadings on *RMRF*, *SMB*, and *HML* factors in a three-factor model, respectively.	KFDL
Amihud illiquidity	Absolute daily returns per unit of trading volume (Amihud (2002)).	CRSP
Monthly volume turnover	Shares traded divided by the number of shares outstanding.	CRSP
Firm age	Number of years since the stock first appears in CRSP.	CRSP
Percentage dividend paying	Proportion of firm in a stock category that paid a dividend at least once during the previous 1 year.	CRSP
% Without analyst coverage	Proportion of firm in a stock category without analyst coverage.	IBES
Mean number of analysts	Mean number of analysts per stock.	IBES
% Institutional ownership	Percentage of total shares outstanding owned by 13F institutions.	13f
Panel B: Additional Variables Used in Stock-Level Regressions (Table III)		
Dividend paying dummy	Set to one if the stock paid dividends during the past 1 year.	CRSP
S&P500 dummy	Set to one if the stock belongs to the S&P500 index.	CRSP
Nasdaq dummy	Set to one if the stock belongs to the Nasdaq index.	CRSP

(*continued*)

Table I—*Continued*

Variable Name	Description	Source
Panel C: Variables Used in Investor-Level Regressions (Table IV)		
Demographic characteristics		
Wealth	Total net worth of the investor.	Brokerage
Income	Annual household income.	Brokerage
Age	Age of the head of the household.	Brokerage
Education	Proportion of residents in investor's zip code with a Bachelor's or higher educational degree.	Census
Professional dummy	Set to one if the investor belongs to one of the professional (managerial or technical) job categories.	Brokerage
Retired dummy	Set to one if the head of the household is retired.	Brokerage
Male dummy	Set to one if the head of the household is male.	Brokerage
Married dummy	Set to one if the head of the household is married.	Brokerage
Investment experience	Number of days since the brokerage account opening date.	Brokerage
Taxable account dummy	Set to one if the investor holds only taxable accounts.	Brokerage
Tax deferred account dummy	Set to one if the investor holds only tax-deferred (IRA or Keogh) retirement accounts.	Brokerage
Location-based measures		
Catholic (Protestant) dummy	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
African American-White ratio	Ratio of African-Americans and Whites in the investor's zip code.	Census
Hispanic-White ratio	Ratio of Hispanics and Whites in the investor's zip code.	Census
Proportion foreign born	Proportion of foreign born residents in the investor's zip code.	Census
Income relative to neighbors	Difference between the investor's annual income and the mean income of sample investors located within 25 miles of her zip code.	Brokerage
Urban dummy	Set to one if the investor resides within 100 miles of one of the largest 20 U.S. metropolitan areas.	Brokerage
County unemployment rate	Unemployment rate in the investor's county of residence.	BLS
State lottery expenditure	Mean annual per capita expenditure in state lotteries in the investor's state of residence.	LOTAG
State lottery age	Number of years since the lottery adoption date in the investor's state of residence.	LOTAG
Portfolio characteristics		
Initial portfolio size	Size of investor's stock portfolio when she enters the sample.	Brokerage
Monthly portfolio turnover	Average of buy and sell turnover rates.	Brokerage
Portfolio diversification	Portfolio variance divided by the average variance of all stocks in the portfolio.	Brokerage
Portfolio dividend yield	Sample period average dividend yield of the investor's portfolio.	Brokerage

(*continued*)

Table I—*Continued*

Variable Name	Description	Source
Panel C: Variables Used in Investor-Level Regressions (Table IV)		
Portfolio local bias	Proportion of the portfolio that is invested in stocks within 100 miles of the investor's zip code.	Brokerage
Industry concentration	Largest weight allocated to one of the 48 Fama-French industries.	Brokerage
Portfolio factor exposures	*RMRF*, *SMB*, *HML*, and *UMD* betas of the investor portfolio.	Brokerage
Panel D: Variables Used in State-Level Regressions (Table V)		
Annual per capita state lottery expenditure	Annual per capita expenditure on state lotteries in the state.	LOTAG
State unemployment	Monthly unemployment rate in the state.	BLS
Catholic (Protestant) dummy	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)		
UNEMP	U.S. monthly unemployment rate.	DS
UEI	Unexpected inflation (current inflation minus the average of the past 12 realizations).	DS
MP	Monthly growth in industrial production.	DS
RP	Monthly default risk premium (difference between Moody's Baa-rated and Aaa-rated corporate bond yields).	DS
TS	Term spread (difference between the yields of constant maturity 10-year Treasury bond and 3-month Treasury bill).	DS
EFC	Mean monthly change in analysts' earnings forecasts of lottery-type stocks.	IBES
LOTRET	Mean monthly return of a portfolio of lottery-type stocks.	CRSP
MKTRET	Monthly market return.	CRSP
Panel F: Additional Variables Used in Performance Regressions (Table VIII)		
Lottery-type stock participation dummy	Set to one if an investor buys at least one lottery-type stock during the sample period.	Brokerage
Strong lottery-type stock preference dummy	Set to one if the lottery-type stock preference measure is in the highest decile.	Brokerage

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 should 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 are likely to 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]

Finally, 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 are able to predict future skewness, it is unlikely that less sophisticated individual investors would be successful in identifying those predictors. Rather, they are more likely to "naïvely" extrapolate past moments into the future and 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.

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 nonlottery-type stocks and the other remaining stocks in the CRSP universe. The nonlottery-type stock category consists of stocks that are in the highest k^{th}

[8] Other 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.

[9] Of 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 C of the Internet Appendix.

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 nonlottery-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 nonlottery-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	Nonlottery-Type	Others
Number of stocks	1,553	1,533	8,945
Percentage of the market	1.25%	50.87%	47.88%
Total volatility	78.57	3.29	22.14
Idiosyncratic volatility	75.56	2.96	20.36
Total skewness	0.330	0.175	0.237
Systematic skewness	−0.202	−0.061	−0.110
Idiosyncratic skewness	0.731	−0.041	0.332
Stock price	\$3.83	\$31.68	\$17.51
Market beta	1.090	0.906	0.897
Firm size (in million \$)	31.41	1650.87	539.66
SMB beta	0.804	0.378	0.617
Book-to-market ratio	0.681	0.314	0.348
HML beta	0.272	0.186	0.151
Past 12-month return	16.52%	20.22%	18.14%
Amihud illiquidity	70.16	0.465	15.13
Monthly volume turnover	84.72%	64.16%	57.90%
Firm age (in years)	5.78	12.10	11.87
Percentage dividend paying	3.37%	44.59%	57.03%
Percentage without analyst coverage	71.30%	21.19%	36.87%
Mean number of analysts	3.93	12.40	6.49
Percentage institutional ownership	7.35%	49.34%	30.09%

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.

The summary statistics in Table II 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 6 years), have low analyst coverage (about 71% of stocks have no analyst coverage), and are mostly nondividend-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, nonlottery-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

[10] The 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)).

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 preferences of institutional investors.

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 figure indicates that, relative to the market portfolio, individual investors significantly overweight 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.

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

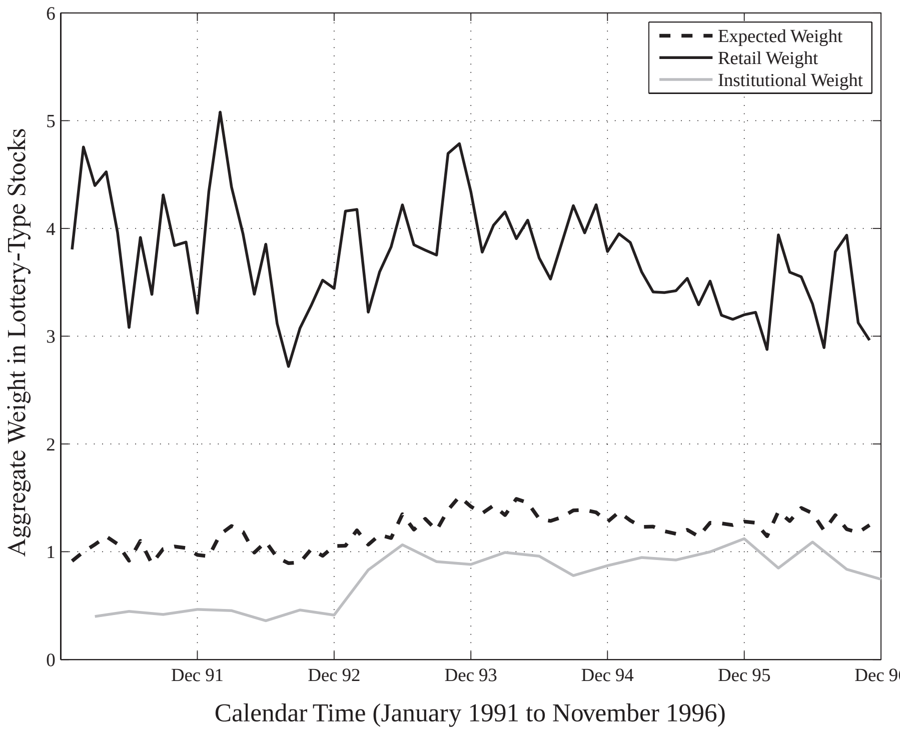

Figure 1. Aggregate weight in lottery-type stocks over time. This figure shows the time series of the actual weights allocated to lottery-type stocks in the aggregate individual and institutional investor portfolios. The expected lottery weight time series, which reflects the weight allocated to lottery-type stocks in the aggregate market portfolio, is also shown. 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.

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

$$EW_{ipt} = \frac{w_{ipt} - w_{imt}}{w_{imt}} \times 100. \tag{1}$$

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 I 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 (the 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.

The Fama–MacBeth regression estimates are presented in Table III. Column (1) reports the estimates for idiosyncratic volatility and skewness measures; 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 the 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 similarly influence investors' aggregate preferences as their coefficient estimates are comparable

[12] For 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.

[13] For 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 ($DW \approx 2$).

Table III

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

This table reports the Fama and MacBeth (1973) 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 are 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 parentheses 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 (the mean is set to zero and the standard deviation is one).

Panel A: Baseline Estimates

	Individuals					Institutions		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.003	0.003	0.007			−0.041		
	(10.67)	(9.09)	(3.23)			(−6.31)		
Idiosyncratic or total volatility	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)
Idiosyncratic or total skewness	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)
Stock price	−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)
Market beta			0.111	0.155	0.100	−0.006	−0.008	0.002
			(6.79)	(8.69)	(8.03)	(−2.47)	(−2.18)	(0.37)
Log(firm size)			−0.189	−0.200	−0.183	0.185	0.220	0.156
			(−10.41)	(−10.79)	(−8.81)	(4.95)	(10.57)	(3.16)
Book-to-market ratio			−0.071	−0.086	−0.064	0.052	0.058	0.065
			(−7.63)	(−10.92)	(−6.32)	(4.12)	(5.58)	(3.28)
Past 12-month stock return			−0.015	−0.013	−0.021	0.024	0.030	0.012
			(−2.51)	(−1.95)	(−2.53)	(3.40)	(6.96)	(1.87)
Systematic skewness			−0.012	−0.017	−0.011	0.020	0.014	0.001
			(−3.36)	(−3.15)	(−2.34)	(2.03)	(2.26)	(0.06)
Monthly volume turnover			0.125	0.133	0.150	−0.032	−0.038	−0.037
			(8.72)	(6.77)	(9.51)	(−6.51)	(−5.07)	(−3.70)
Dividend paying dummy			−0.069	−0.100	−0.074	0.014	0.016	0.012
			(−7.62)	(−11.53)	(−7.37)	(4.37)	(2.31)	(3.14)
Firm age			−0.038	−0.069	−0.047	0.014	0.016	0.011
			(−6.63)	(−6.65)	(−7.17)	(3.20)	(2.75)	(2.02)
S&P500 dummy			−0.004	−0.005	−0.008	0.012	0.012	0.014
			(−2.51)	(−1.90)	(−1.96)	(3.25)	(4.83)	(4.22)

(continued)

Table III—*Continued*

Panel A: Baseline Estimates

Variable	Individuals (1)	(2)	(3)	(4)	(5)	Institutions (6)	(7)	(8)
Nasdaq dummy			0.033	0.024	0.030	−0.016	−0.023	−0.004
			(2.96)	(4.47)	(3.50)	(−3.44)	(−5.81)	(−1.12)
(Mean) Number of observations	5,979	5,979	5,310	377,010	256,813	4,238	101,761	78,028
(Mean) Adjusted R^2	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)
High volatility dummy	0.051	0.053	−0.013	−0.019
	(5.22)	(6.90)	(−2.32)	(−2.17)
High skewness dummy	0.046	0.042	−0.032	−0.042
	(3.25)	(3.39)	(−3.03)	(−2.16)
Low price dummy	0.092	0.107	−0.034	−0.031
	(6.51)	(6.09)	(−3.17)	(−3.00)
High volatility × high skewness	0.074	0.079	−0.010	−0.011
	(6.32)	(5.37)	(−1.87)	(−1.96)
High volatility × low price	0.028	0.025	−0.005	−0.004
	(8.71)	(7.01)	(−1.16)	(−1.13)
High skewness × low price	0.017	0.016	−0.007	−0.007
	(3.27)	(3.70)	(−1.03)	(−0.92)
High skewness × high volatility × low price	0.046	0.048	−0.037	−0.033
	(4.26)	(4.62)	(−2.08)	(−1.88)

(Other coefficient estimates have been suppressed.)

(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 on stock price (= −0.191) is more than three times stronger than the estimates of the 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, book-to-market, the past 12-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 $t-1$.

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 on *Firm Size* is strongly negative, which indicates that individual investors exhibit a preference for relatively smaller stocks. The positive coefficients on *S&P500 dummy* and *Volume Turnover* indicate that investors exhibit a preference for relatively more visible and liquid firms. The positive coefficient on the turnover measure is also consistent with individual investors' preferences for attention-grabbing stocks as stocks with high monthly turnover are more likely to be in the news and thus 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 autocorrelations, 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 (2009)). 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. The variable *EW* has a mean of 1.01% and a standard deviation of 3.45%, and in column (4) *Idiosyncratic Volatility* 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 be roughly opposite.[14] To investigate whether these fundamental constraints hold, I estimate stock-level cross-sectional regressions to examine 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 previous evidence on aggregate institutional preferences (e.g., Bennett et al. (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 prudent man rules might prevent them from holding lower-priced stocks (Badrinath et al. (1989), Del Guercio (1996)).

When I re-estimate the stock-level regressions after excluding stocks that are priced below $5, the subsample 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 precisely. 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.

[14] Very 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.

The results from the extended regression specifications indicate that the individual investors 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 varies 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 the 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,

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

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

[15] The 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 subsample coefficient estimates are very similar to the full-sample estimates.

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 ($LP^{(1)}_{it}$) 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:

$$LP^{(2)}_{it} = \frac{NW_{it} - ENW_{it}}{ENW_{it}} \times 100. \tag{3}$$

In equation (3), the actual and the expected normalized weights in lottery-type stocks for investor i in month t are given by

$$NW_{it} = \frac{LP^{(1)}_{it} - min\left(LP^{(1)}_{it}\right)}{max\left(LP^{(1)}_{it}\right) - min\left(LP^{(1)}_{it}\right)} \tag{4}$$

and

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

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 the sample investors in month t, and $max(PSize_{it})$ is the maximum portfolio size of the sample investors in month t. The $min(LP^{(1)}_{it})$ and $max(LP^{(1)}_{it})$ 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 ($LP^{(1)}_{it}$) 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

$$LP^{(3)}_{it} = \frac{LP^{(1)}_{it} - LP^{mkt}_{t}}{LP^{mkt}_{t}} \times 100, \tag{6}$$

[16] An 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.

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 nonlottery-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 nonlottery-type stocks:

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

Since the market capitalization of the nonlottery-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.

Finally, I define a lottery-type stock preference measure using investors' trades. Because the portfolio of lottery-type stocks changes 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 these two trading volume measures:

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

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) to (8).[17] A set of variables that capture investors' socioeconomic

[17] Examining 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 B of the Internet Appendix for additional details.

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).

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.

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.[18] For brevity, the coefficient estimates of all control variables are suppressed.

I find that younger, less wealthy, less educated, nonprofessional 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 the religion, race, and ethnicity variables, I find that *Catholic dummy* has the strongest influence on an 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. 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

[18] As 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.

[19] I 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.

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

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 the 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) to (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 parentheses 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)
Intercept	−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)
Wealth	−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)
Age	−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)
Zip code education	−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)
Professional dummy	−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)
Retired dummy	−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)
Male dummy	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)
Married dummy	−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)
Investment experience	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)
Taxable account only	0.034	0.027	0.027	0.018	0.045	0.030	0.029	0.023
dummy	(3.58)	(2.24)	(2.20)	(1.79)	(4.76)	(2.48)	(2.88)	(1.89)
Tax deferred acc. only	−0.022	−0.053	−0.015	−0.025	−0.029	−0.067	−0.048	−0.033
dummy	(−2.39)	(−4.67)	(−1.31)	(−2.54)	(−3.24)	(−5.24)	(−4.05)	(−1.95)
Catholic county	0.053	0.049	0.043	0.035	0.032	0.058	0.078	0.044
dummy	(5.13)	(3.69)	(3.68)	(3.85)	(2.77)	(3.98)	(7.39)	(3.79)
Protestant county	−0.046	−0.035	−0.027	−0.024	−0.030	−0.051	−0.040	−0.051
dummy	(−3.19)	(−3.38)	(−3.69)	(−2.12)	(−3.38)	(−4.19)	(−4.54)	(−4.24)
Zip code Afr.	0.028	0.025	0.028	0.023	0.019	0.021	0.023	0.030
Am.–White ratio	(3.51)	(3.67)	(2.64)	(3.28)	(2.14)	(2.11)	(3.19)	(2.96)
Zip code Hispanic–	0.034	0.031	0.040	0.028	0.020	0.046	0.036	0.034
White ratio	(2.97)	(3.07)	(3.56)	(3.19)	(2.20)	(3.00)	(3.28)	(3.38)
Zip code prop foreign	0.020	0.017	0.011	0.021	0.011	0.007	0.016	0.020
born	(2.28)	(2.01)	(1.64)	(2.18)	(1.45)	(0.74)	(1.89)	(2.18)

(*continued*)

Table IV—*Continued*

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income relative to neighbors	−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)
Urban dummy	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)
County unemployment rate	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)
State lottery expenditure	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)
State lottery age	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)
Number of investors	21,194	21,194	21,194	21,194	18,650	21,194	21,194	16,955
Adjusted R^2	(0.043)	(0.058)	(0.035)	(0.052)	(0.031)	(0.061)	(0.055)	(0.040)

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. Local economic conditions, as captured by a county's unemployment rate are 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 that 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 using the 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, 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, which implies that, all else equal, a one-standard deviation increase in the age of an investor is associated with a $0.044 \times 16.62 = 0.73\%$ reduction in the weight allocated to lottery-type stocks.[21] Thus, a 65-year-old investor would allocate 2.19% lower weight to lottery-type stocks than he would have allocated 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 is 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 such that 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 the 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 nonlocal 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) to (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.[22]

[21] The $LP^{(1)}$ measure has a mean of 10.36% and a standard deviation of 16.62%.

[22] I 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

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 (2009)).

To distinguish between overconfidence- and gambling-based explanations, I first 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 nonparticipants and participants is 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 subsample 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 subsample of investors who are unlikely to exhibit overconfidence. The subsample results are presented in column (8) of Table IV. I find that the subsample 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.

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.

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 *Overconfidence dummy* in investor-level regression specifications, I find that it has a significantly positive estimate in all instances. For instance, in specification (2), *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.[23] This evidence indicates that overconfidence has an incremental ability to explain investors' preference for stocks with lottery features.[24]

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 a 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 subsample are qualitatively very similar to the full-sample estimates. For instance, the coefficient on *Wealth* is −0.043 (t-statistic = −3.27), *Education* has an estimate of −0.084 (t-statistic = −4.49), and *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 C of the Internet Appendix. The evidence indicates that stock price is the most important lottery characteristic, followed by idiosyncratic skewness. The least important lottery characteristic appears to be idiosyncratic volatility.

[23] This result is not mechanically induced. See Section D of the Internet Appendix for further details.

[24] I 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 subsample of lottery stocks that have moderate levels of intangible assets and are less likely to be associated with 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.

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

Although the evidence from investor-level regressions indicates that the 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 the 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 nonlottery-type stocks and the 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).[25]

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, which has the largest (about 27%) proportion of sample investors. Unfortunately, zip code–level data are available only for more recent years (2005 and 2006). In spite of the nonoverlapping time periods for the 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-value = 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 is 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.[26]

[25] Given 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 nonlottery-type stock categories.

[26] While 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.

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 parentheses 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)
Annual per capita state lottery	0.057	0.050	0.107	0.037	0.109	0.121
expenditure	(4.99)	(3.66)	(3.36)	(2.04)	(3.69)	(5.49)
State lottery age	0.136	0.107	0.228	0.109	0.192	0.174
	(2.66)	(2.440	(8.98)	(7.63)	(6.45)	(6.92)
Monthly state unemployment	0.068	0.062	0.026	0.047	0.126	0.129
rate	(6.85)	(2.56)	(2.17)	(2.18)	(5.98)	(2.02)
Catholic state dummy		0.231	0.192	0.096	0.232	0.259
		(6.44)	(7.91)	(3.65)	(9.85)	(2.92)
Protestant state dummy		−0.094	−0.114	−0.116	−0.176	−0.098
		(−2.65)	(−3.65)	(−3.45)	(−6.04)	(−4.10)
(Estimates of control variables have been suppressed.)						
Number of observations	2,236	2,236	2,236	2,236	2,236	2,236
Adjusted R^2	0.024	0.047	0.097	0.067	0.092	0.057

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. 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 (2009), 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 demand 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 is 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 should 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

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

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 the other remaining stocks.[27] The portfolios

[27] In 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.

of lottery-type stocks and other stocks are defined at the end of month $t-1$.[28]

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.[29]

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 1-month lagged *EBSI* variable as an explanatory variable to control for potential serial correlation in that measure.

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 should vary over time and the relation between macroeconomic indicators and trading intensity may 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.

[28] The 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.

[29] If 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.

Table VI

Macroeconomic 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) to (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. Newey and West (1987) adjusted t-statistics for the coefficient estimates are reported in parentheses below the estimates.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.001	0.972	1.399	0.972	0.023	−0.010
	(0.20)	(1.29)	(1.55)	(1.10)	(0.25)	(−0.16)
Lagged *UNEMP*	0.202				0.189	0.135
	(2.74)				(2.54)	(3.14)
Lagged *UEI*		−0.091			−0.072	−0.053
		(−1.36)			(−0.77)	(−1.05)
Lagged *MP*		−0.013			0.016	−0.022
		(−0.69)			(0.20)	(−0.45)
Lagged *RP*		0.507			0.124	0.112
		(4.59)			(2.75)	(3.16)
Lagged *TS*		−0.114			−0.012	−0.066
		(−0.44)			(−0.50)	(−0.91)
Lagged *EFC*			0.034		0.071	0.014
			(0.78)		(1.59)	(1.44)
EFC			0.003		−0.014	0.005
			(0.40)		(−0.17)	(0.32)
Lagged *LOTRET*				0.102	−0.037	−0.029
				(1.02)	(−0.12)	(−0.32)
LOTRET				0.498	0.427	0.816
				(3.54)	(2.07)	(11.02)
Lagged *MKTRET*				−0.047	−0.017	−0.054
				(−1.33)	(−0.51)	(−0.83)
MKTRET				−0.127	−0.111	−0.219
				(−1.56)	(−1.45)	(−3.30)
Lagged *EBSI*					0.412	0.215
					(3.99)	(3.00)
Number of Months	71	71	71	71	70	215
Adjusted R^2	0.085	0.204	0.005	0.181	0.396	0.493

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 the control variables are also as expected. For instance, lagged *EBSI* has a positive coefficient estimate, which indicates that there is persistence in investors' differential demand shifts.[30]

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, a 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 6-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 may 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* are 0.135 (t-statistic = 3.14), 0.112 (t-statistic = 3.16), and 0.816 (t-statistic = 11.02), respectively. In comparison, using the brokerage data, the corresponding coefficient estimates are 0.189 (t-statistic = 2.54), 0.124 (t-statistic = 2.75), and 0.427 (t-statistic = 2.07), respectively. These comparisons indicate that the time-series relation between lottery demand and

[30] For 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.

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 underperformance 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, then like lottery players, lottery investors should be willing to invest in lottery-type stocks, even when they are expected to underperform. Furthermore, the magnitude of the underperformance 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.

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 portfolios of "nonlottery stocks" and "other stocks." All three stock portfolios are 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 nonlottery and other stock categories. Specifically, relative to the nonlottery stock portfolio, the annualized raw, characteristic-adjusted, and risk-adjusted performance differentials are -7.96%, -4.98%, and -7.10%, respectively. 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,

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, nonlottery 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), standard deviation of monthly portfolio returns (*SD*), 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 et al. (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 parentheses below the estimates.

Portfolio	MeanRet	*SD*	CharAdjRet	Alpha	RMRF	SMB	HML	UMD	Adj. R^2
Lottery (L)	0.472	7.934	−0.375	−0.552	1.204	1.130	−0.049	−0.442	0.880
			(−2.95)	(−3.22)	(18.90)	(15.15)	(−0.78)	(−8.05)	
Nonlottery	1.135	4.025	0.040	0.041	0.920	−0.123	0.102	−0.008	0.963
(NL)			(0.47)	(0.84)	(28.39)	(−8.16)	(5.77)	(−0.82)	
Others (O)	1.033	4.644	−0.026	−0.033	0.959	0.099	−0.103	−0.010	0.981
			(−1.12)	(−0.83)	(18.39)	(7.90)	(−6.82)	(−1.19)	
L–NL	−0.663	5.882	−0.415	−0.592	0.284	1.253	−0.151	−0.433	0.728
	(−2.95)		(−3.14)	(−3.12)	(6.15)	(12.13)	(−2.17)	(−6.65)	
L–O	−0.562	4.846	−0.349	−0.519	0.244	1.031	0.051	−0.431	0.685
	(−2.57)		(−2.93)	(−3.08)	(5.96)	(10.61)	(0.83)	(−5.96)	

I find that lottery-type stocks earn at least 4% lower average annual returns.[31]

B. Investment in Lottery-Type Stocks and Portfolio Performance

The lower average returns of lottery-type stocks suggest that greater investment in lottery-type stocks is likely to be associated with greater average underperformance. The exact magnitude of portfolio underperformance, however, depends 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 underperformance that is associated with investment in lottery-type stocks, I estimate the degree of underperformance 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 nonlottery portfolio component of an investor's portfolio by

[31] For 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 IA.I of the Internet 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 to 1996 sample period. See Section E of the Internet Appendix for an additional discussion.

the market portfolio.[32] 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 underperformance (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 underperform 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 the investor's age, investment experience, annual household income, plus 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.

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) to (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) to (8), respectively.[33]

The coefficient estimates from specification (1) indicate that the average annual risk-adjusted underperformance is 3.00% (0.250 × 12) for an investor who trades lottery-type stocks at least once during the sample period.[34] The level of underperformance is significant (0.189 × 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 underperformance is greater for investors who allocate a larger portfolio weight to lottery-type stocks. The incremental annual risk-adjusted underperformance is 3.16% (0.263 × 12) for an investor who increases the investment in lottery-type

[32] I thank an anonymous referee for suggesting this test.

[33] As 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.

[34] The low adjusted R^2s in the cross-sectional regressions are consistent with the evidence in Barber and Odean (2001, p. 280).

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) to (8), the dependent variable is the risk-adjusted performance measure (four-factor alpha) of a hypothetical portfolio that is formed by replacing the nonlottery component of an investor portfolio by the market portfolio. Lottery-type stocks are 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 nonlottery component of her portfolio. The set of independent variables includes a participation dummy, lottery-type stock preference measure, and 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) to (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 the investor's age, investment experience, annual household income, plus 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 are 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 parentheses below the estimates.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	−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)
Lott-stock	−0.250	−0.189							
part. dummy	(−4.59)	(−3.32)							
Lott-type			−0.402	−0.263	−0.299	−0.226	−0.237	−0.193	−0.173
stock pref			(−8.51)	(−4.86)	(−3.88)	(−3.13)	(−3.29)	(−2.88)	(−2.73)
Strong lott.			−0.069	−0.048	−0.079	−0.078	−0.075	−0.070	−0.008
pref. dummy			(−2.36)	(−2.17)	(−1.75)	(−1.73)	(−1.71)	(−2.16)	(−0.36)
Control variables	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
	(Estimates of control variables have been suppressed.)								
Number of investors	40,476	27,565	34,588	26,204	25,229	25,229	25,229	25,229	25,229
Adjusted R^2	0.009	0.039	0.017	0.041	0.039	0.042	0.040	0.036	0.045

stocks by one standard deviation. Furthermore, if the investor is in the highest lottery weight decile, the portfolio underperforms 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) to (8)), the coefficient estimates are qualitatively similar. Even when I consider a trade-based lottery preference measure (specification (8)), the degree of underperformance that can be attributed to investment in lottery-type stocks is economically significant. The incremental annual risk-adjusted underperformance 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 underperformance, 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 underperformance, 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 underperformance due to investments in lottery-type stocks has a striking resemblance to the evidence from state lottery studies. In both instances, the proportional level of underperformance decreases with income. Therefore, like state lotteries, lottery-type stocks appear to be regressive.

C. Performance of Lottery and Nonlottery Portfolio Components

To better identify the mechanisms that generate portfolio underperformance, I compare the performance of lottery and nonlottery components of investor portfolios. If investors who prefer lottery-type stocks are "bad" investors in general, both the lottery and the nonlottery portfolio components of their portfolios would perform poorly and the two performance measures should be positively correlated. In contrast, if the underperformance 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 nonlottery portfolio components should 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 nonlottery portfolio component should perform relatively better than the lottery component and the two performance measures should be negatively correlated.

In the first test, I compute the average correlation between the performance of lottery and nonlottery components of investors' portfolios. Every month, I decompose the total portfolio performance of each investor into the performance of lottery-type stocks and nonlottery stocks and compute the time-series correlation between the two performance measures.[35] I find that the average lottery-nonlottery 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 nonlottery 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 nonlottery component. I compute the performance of this

[35] The correlation is only defined for investors who hold both lottery- and nonlottery-type stocks.

hypothetical portfolio and examine the difference between the performance levels of the actual and the hypothetical portfolios. Such performance differential would reflect both the relative underperformance of lottery-type stocks and the additional biases that an investor might exhibit when she holds lottery-type stocks.[36] I find that a typical lottery investor would have been able to earn $0.237 \times 12 = 2.84\%$ higher annual returns on average if she simply replaced her lottery investments with her nonlottery 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 underperformance 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 underperformance 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 underperformance of the lottery-type stock component of investor portfolios reflects both the underperformance of lottery-type stocks and investors' 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 underperformance and the degree of portfolio underperformance resembles the evidence from lottery studies. In both instances, the level of underperformance 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.

[36] When investors invest in lottery-type stocks, they might exhibit new types of biases or the biases that they exhibit with nonlottery stocks get amplified due to stocks' lottery characteristics.

The finding that socioeconomic characteristics of individual investors influence their stock preferences is not entirely 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 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 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 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.

Barber, Brad M., and Terrance Odean, 2001, Boys will be boys: Gender, overconfidence, and common stock investment, *Quarterly Journal of Economics* 116, 261–292.

Barber, Brad M., and Terrance Odean, 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.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2009, Do noise traders move markets? *Review of Financial Studies* 22, 151–186.

Barberis, Nicholas, and Ming Huang, 2008, Stocks as lotteries: The implications of probability weighting for security prices, *American Economic Review* 98, 2066–2100.

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).

Brenner, Reuven, 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.

Clotfelter, Charles T., and Philip J. Cook, 1989, *Selling Hope: State Lotteries in America* (Harvard University Press, Cambridge, MA).

Clotfelter, Charles T., Philip J. Cook, 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, 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.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 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, 2009, Hard-to-value stocks, behavioral biases, and informed trading, *Journal of Financial and Quantitative Analysis* (forthcoming).

Kumar, Alok, 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.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite heteroskedasticity and auto-correlation consistent variance–covariance matrix, *Econometrica* 55, 703–708.

Odean, Terrance, 1999, Do investors trade too much? *American Economic Review* 89, 1279–1298.

Petersen, Mitchell A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

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).

Shiller, Robert J., 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).