

Gambling and Comovement

Alok Kumar, Jeremy K. Page, and Oliver G. Spalt*

Abstract

This study shows that correlated trading by gambling-motivated investors generates excess return comovement among stocks with lottery features. Lottery-like stocks comove strongly with one another, and this return comovement is strongest among lottery stocks located in regions where investors exhibit stronger gambling propensity. Looking directly at investor trades, we find that investors with a greater propensity to gamble trade lottery-like stocks more actively and that those trades are more strongly correlated. Finally, we demonstrate that time variation in general gambling enthusiasm and income shocks from fluctuating economic conditions induce a systematic component in investors' demand for lottery-like stocks.

I. Introduction

Identifying sources of comovement in stock returns is one of the fundamental goals of asset pricing research. The traditional view is that returns of certain stocks move together because they have correlated cash flows or are subject to systematic shifts in discount rates, or both. However, a growing number of studies provide evidence of comovement patterns in stock returns that cannot be easily attributed to comovement in cash flows or a systematic shift in discount rates (e.g., Lee, Shleifer, and Thaler (1991), Froot and Dabora (1999), Kumar and Lee (2006), and Boyer (2011)).

In this paper, we posit that active correlated trading among gambling-motivated investors induces a common factor among the returns of stocks that those investors find attractive as gambling objects.¹ Although previous research has found evidence of gambling-motivated investment decisions among both

*Kumar (corresponding author), akumar@miami.edu, University of Miami, School of Business Administration, Coral Gables, FL 33124; Page, jeremy.page@byu.edu, Brigham Young University, Marriott School of Management, Provo, UT 84602; and Spalt, o.g.spalt@uvt.nl, Tilburg University, School of Economics and Management, Tilburg 5000 LE, Netherlands. We thank Stephen Brown (the editor), James Choi, Joost Driessen, Rik Frehen, Zoran Ivkovich (the referee), Kelvin Law, Toby Moskowitz, and seminar participants at the 2011 American Finance Association Meetings (Denver, CO) for helpful comments and discussions. We also thank Brad Barber and Terrance Odean for sharing the discount brokerage and Institute for the Study of Security Markets (ISSM)/Trade and Quote (TAQ) data sets. We are responsible for all remaining errors and omissions.

¹We define gambling in the context of financial markets as the desire to seek lottery-type payoffs (i.e., extreme returns at a low cost) using financial assets.

retail and institutional investors, this study is the first to examine the potential impact of these decisions on return comovement. Our main conjecture is motivated by the habitat-based framework of return comovement (Barberis, Shleifer, and Wurgler (2005)) and the observation that gambling-motivated trading activities are concentrated among a subset of stocks with lottery-like features.

Barberis et al. (2005) argue that when groups of investors concentrate their trading within a specific habitat or category of stocks, fluctuation in those investors' sentiment can lead to nonfundamental comovement among those stocks. In addition, prior studies show that investors with a taste for gambling concentrate their trading in lottery-like stocks with high skewness and volatility and low nominal prices.² Those investors with gambling preferences trade actively,³ and their trading activities are often correlated, perhaps due to their stronger behavioral biases and because their demographic attributes are similar.⁴ Given the observed behavior of gambling-motivated investors, we conjecture that the lottery-like stocks favored by these investors would exhibit excess return comovement.

Consistent with this prediction, we find that lottery-like stocks comove strongly with one another, and provide evidence that this return comovement is generated by the correlated trading of gambling-motivated investors. First, we show that large numbers of stocks are significantly affected by gambling sentiment. We identify lottery-like stocks by constructing an index that ranks securities by how much they share features of attractive monetary gambles (i.e., low price, high volatility, and high skewness). This lottery index (LIDX) is inspired by Kumar (2009b) and has been used in other studies to identify lottery-type stocks (e.g., Doran et al. (2011), Kumar et al. (2011)). We then measure return comovement among lottery-like stocks by regressing stock returns on a portfolio of high-LIDX stocks while controlling for commonly used return factors.

We find that 14.27% of all stocks in the Center for Research in Security Prices (CRSP) universe, comprising 12.01% of the total stock market value, comove significantly with a portfolio of high-LIDX stocks. In comparison, 39.13%, 23.14%, 15.52%, and 13.18% of stocks have significant betas relative to the market, size, value, and momentum factors, comprising 78.01%, 29.76%, 39.11%, and 28.59% of the aggregate stock market value, respectively. These comparisons indicate that gambling-induced sentiment affects return comovement in a smaller but economically meaningful segment of the market.

To identify whether the return comovement we observe is actually driven by the trading of gambling-motivated investors, we test whether the degree of return

²See, for example, Kumar (2009b), Kumar, Page, and Spalt (2011), Doran, Jiang, and Peterson (2011), Coelho, John, and Taffler (2010), Dorn and Sengmueller (2009), Hoffmann and Shefrin (2011), and Grinblatt and Keloharju (2009). Gao and Lin (2011) provide evidence of substitution effects between lottery gambling and stock trading using Taiwanese data.

³For example, Dorn and Sengmueller (2009) report that 98% of German discount brokerage investors report that they "enjoy investing," 45% "enjoy taking risky positions," and 47% find that "in gambling, the fascination increases with the size of the bet." These investors hold more concentrated portfolios, the realized skewness in their portfolios is higher and, most importantly, they trade very actively.

⁴Kumar (2009a) shows that less sophisticated investors with stronger behavioral biases exhibit more correlated trading, while Kumar (2009b) shows that investors with stronger propensity to gamble are less sophisticated.

comovement among lottery-like stocks is stronger for lottery stocks that are disproportionately held by gambling-motivated investors. Capturing variation in investors' propensity to gamble is a challenge for any study on the subject because there are no direct measures of stock-level gambling activities. We use an exogenous geographical proxy for gambling propensity developed by Kumar et al. (2011), who show that the ratio of Catholics to Protestants in the local population (CPRATIO) is an effective proxy for the gambling propensity of local investors. This choice is motivated by prior evidence suggesting that religion-based differences in gambling norms influence people's actual gambling behavior.⁵

Specifically, we conjecture that CPRATIO around a firm's headquarters would reflect the gambling propensity of the firm's investors. All else equal, a firm located in a high-CPRATIO region would have an investor clientele that is more likely to gamble. This geography-based identification strategy is similar to the approach in recent studies such as Becker, Ivković, and Weisbenner (2011) and Kumar et al. (2011) and implicitly relies on the existence of local bias among shareholders (e.g., Coval and Moskowitz (1999), Huberman (2001), and Ivković and Weisbenner (2005)). Although our baseline specifications rely on the existence of local bias, we also use portfolio holdings data for institutions and retail investors to confirm our main findings in a way that does not depend on local bias.

We find that the return comovement among lottery-like stocks is strongest for lottery stocks that are located in regions where investors have stronger preferences for gambling, as proxied by CPRATIO. A natural concern is that CPRATIO picks up geographical variation in other investor, firm, or market attributes besides gambling propensity that may drive the observed return comovement, such as investor wealth or other correlates. However, our findings are robust to explicit controls for local income, education, and other geographic characteristics such as local economic conditions, as well as a variety of firm-level controls and year and industry fixed effects. Furthermore, our results are not driven by microstructure-based biases, are robust to various definitions of "local," and are not exclusive to financial centers or any particular region in the United States. In a particularly stringent test, we include metropolitan statistical area (MSA) dummies in our main regression specification. We find that even when we condition on time-invariant characteristics of the MSA where the stock is headquartered, return comovement among two otherwise identical lottery stocks is higher when the local population is more prone to gambling. This evidence is consistent with our gambling conjecture, and it is not obvious what fundamental factor would be able to explain this pattern.

We find similar results when we replace CPRATIO with a holdings-based measure of the average lottery preference of a stock's shareholders. This measure, although arguably less exogenous than CPRATIO, has the advantage of not relying on an assumption of local bias. We also show directly using a sample of retail investor holdings that return comovement is stronger when the characteristics

⁵Section A.1 of the Internet Appendix (available at www.jfqa.org) summarizes the views and attitudes of different religious denominations toward gambling and describes prior evidence of the link between religious background and gambling behavior. Also, Kumar et al. (2011) establish that the CPRATIO measure reflects the effects of religion-induced social norms and does not merely proxy for geographical characteristics (e.g., urban, financial center, level of trust, political orientation, etc.) that influence investment decisions and are also correlated with local religious composition.

of the average investor holding the stock are similar to the characteristics of lottery investors identified in Kumar (2009b). Collectively, our findings indicate that correlated trading of lottery-like stocks by investors with gambling preferences induces significant comovement among those stocks.

Overall, we find robust evidence that lottery-like stocks that are disproportionately held by gambling-motivated investors comove more strongly with other lottery stocks. This suggests that the excess return comovement we observe among lottery-like stocks is driven by the trading behavior of investors with gambling preferences, consistent with the habitat-based framework of Barberis et al. (2005). To support this interpretation, we examine investor trading behavior more directly and show that it is consistent with our gambling-based comovement hypothesis.

Specifically, we find that both retail and institutional investors in high-CPRATIO areas allocate larger portfolio weights to both local and nonlocal high-LIDX stocks, and trade them more actively. Furthermore, we show that retail investor trades are more correlated and comprise a larger fraction of trading volume when CPRATIO and LIDX are high. Institutional investors exhibit a similar but weaker pattern.⁶

In the last part of the paper, we investigate what causes gambling-motivated investors to move in and out of lottery-like stocks in a correlated way. We consider two possibilities. First, investors may be influenced by time variation in people's enthusiasm for gambling, so that as gambling sentiment rises or falls over time, gambling-motivated investors' demand for lottery-like stocks fluctuates in a correlated way. Consistent with this hypothesis, we find that the relation between CPRATIO and lottery-stock comovement is strongest when gambling enthusiasm as proxied by state lottery sales is high.

Second, gambling-motivated investors may experience correlated income shocks as macroeconomic conditions fluctuate over time, causing their demand for lottery-like stocks to fluctuate in a correlated way. Consistent with this conjecture, we find that the relation between CPRATIO and lottery-stock comovement is stronger when the local economy performs well, which allows gambling-motivated investors to increase their demand for lottery-like stocks.

These empirical findings contribute to an emerging finance literature that analyzes the relation between gambling and investment decisions. Prior research has shown that investors with a stronger propensity to gamble exhibit a strong preference for lottery-like stocks (Kumar (2009b), Kumar et al. (2011)), trade more frequently (Dorn and Sengmueller (2009), Hoffmann and Shefrin (2011), and Grinblatt and Keloharju (2009)), and substitute between lottery gambling and stock trading (Gao and Lin (2011)). Furthermore, Doran et al. (2011) show that gambling preferences of individual investors during the New Year influence the January prices and returns of assets with lottery features, and Coelho et al. (2010) show that gambling-motivated retail investors trade stocks of bankrupt firms for a shot at extreme payoffs.

Although these studies document that gambling behavior influences investment decisions, our paper is the first to directly investigate the impact of gambling

⁶The weak evidence of correlated trading could be due to the coarseness of the institutional data (quarterly as opposed to monthly).

attitudes on return comovement. Our results also extend the recent literature on nonfundamentals-based return comovement and show that gambling-induced sentiment is an important source of comovement in stock returns. In economic terms, we show that stocks affected by gambling-induced sentiment comprise more than 12% of the overall stock market value. More broadly, our findings contribute to the debate on the importance of investors' behavioral biases for aggregate market outcomes.

Our study is related to the work of Kumar et al. (2011), who introduce CPRATIO as a proxy for gambling propensity. Using this gambling proxy, Kumar et al. (2011) document that in regions with a higher CPRATIO, investors exhibit a stronger propensity to hold lottery-type stocks, broad-based employee stock option plans are more popular, the first-day return following an initial public offering is higher, and the magnitude of the negative lottery-stock premium is larger. Subsequent studies have used the CPRATIO measure to identify the role of gambling or skewness preferences in various corporate finance and asset pricing settings. For example, Shu, Sulaeman, and Yeung (2012) use the same religion-based proxy to study mutual fund risk-taking behavior. Similarly, Schneider and Spalt (2016) use CPRATIO to show that the skewness preferences of chief executive officers (CEOs) generate inefficient internal capital allocations in conglomerates.

In contrast to these related studies, we document that lottery-like stocks exhibit significant excess return comovement, and we use CPRATIO to establish that this comovement is induced by trading activities of investors with strong gambling preferences. None of the earlier studies, including Kumar et al. (2011), analyzes return comovement or provides direct evidence of gambling-motivated correlated trading and its impact on return comovement.

The remainder of the paper is organized as follows: We summarize the data and our methods in Section II. Section III presents the main empirical results, Section IV provides additional evidence using direct investor-level data, and Section V identifies the potential determinants of gambling sentiment. We conclude in Section VI with a brief discussion.

II. Data and Methods

We use data from multiple sources to test our gambling-based conjectures. In this section, we identify those sources and briefly summarize the key measures. The Appendix provides further details about all variables, and Table 1 reports the summary statistics for those variables.

A. Lottery Characteristics of Individual Stocks

We measure the attractiveness of a stock as a gambling object using LIDX. To construct this index, we assign all stocks from the CRSP into vigintiles (20 bins) each year by price, idiosyncratic volatility, and idiosyncratic skewness. Bin 20 contains stocks from the lowest price group and the highest volatility, and skewness groups. For each stock, the price, volatility, and skewness vigintile bin scores are added to produce a score between 3 and 60. This score is then scaled between 0 and 1 using $LIDX = (\text{Score} - 3) / (60 - 3)$. A higher value of LIDX for a

TABLE 1
Summary Statistics

Table 1 reports summary statistics for variables used in the empirical analysis. All variables are defined in the Appendix. The sample period is from 1980 to 2005 for most variables.

Panel A. Summary Statistics

Variables	Mean	SD	Percentiles			Percentiles		N
			10th	25th	Median	75th	90th	
LOTTERY_STOCK_BETA	0.31	0.17	0.91	-0.62	-0.23	0.73	1.46	65,981
LOW_PRICE_BETA	0.30	0.16	0.88	-0.61	-0.23	0.71	1.43	65,981
HIGH_VOLATILITY_BETA	0.30	0.16	0.89	-0.60	-0.23	0.71	1.43	65,981
HIGH_SKEWNESS_BETA	0.46	0.26	1.42	-0.99	-0.35	1.11	2.21	65,981
LIDX	0.50	0.53	0.22	0.18	0.33	0.68	0.77	65,981
AVG_LOTTERY_PREFERENCE (RETAIL)	0.38	0.37	0.13	0.23	0.29	0.47	0.55	52,332
AVG_LOTTERY_PREFERENCE (INST)	0.30	0.29	0.07	0.23	0.25	0.34	0.41	61,545
RETURN _(-1,0)	0.19	0.05	0.95	-0.50	-0.24	0.38	0.86	65,981
TURNOVER	0.10	0.06	0.14	0.01	0.03	0.12	0.22	65,981
ln(FIRM_AGE)	4.93	4.91	0.84	3.78	4.25	5.54	6.03	65,981
ln(MKT_CAP)	18.52	18.40	2.20	15.72	16.91	20.07	21.50	65,981
MB_RATIO	2.69	1.61	3.58	0.64	0.98	2.88	5.37	65,981
CPRATIO	2.19	1.92	1.66	0.30	0.72	3.20	4.76	65,981
INDUSTRY_CLUSTER_DUMMY	0.71	1.00	0.46	0.00	0.00	1.00	1.00	65,981
TOTAL_POPULATION	4.81	2.88	4.91	0.55	1.25	6.36	14.33	65,981
AVG_EDUCATION	0.29	0.28	0.07	0.21	0.24	0.35	0.40	65,981
MALE_FEMALE_RATIO	0.95	0.94	0.04	0.91	0.92	0.98	1.00	65,981
MARRIED	0.51	0.51	0.05	0.46	0.48	0.53	0.56	65,981
MINORITY	0.27	0.26	0.11	0.12	0.18	0.34	0.42	65,981
AGE	33.62	33.60	2.44	30.61	31.75	35.36	36.75	65,981
URBAN	0.95	0.97	0.05	0.89	0.94	0.99	0.99	65,981
STATE_MACRO_INDEX	0.44	0.35	0.74	-0.39	-0.04	0.86	1.37	65,981
STATE_LOTTERY_SALES	256.54	226.18	174.44	88.28	136.99	314.78	443.35	34,679

Panel B. Correlation Matrix of Main Variables

Variables	LOTTERY_BETA	LOW_PRICE_BETA	HIGH_VOLATILITY_BETA	HIGH_SKEWNESS_BETA	LIDX	CPRATIO
LOTTERY_BETA	1.00					
LOW_PRICE_BETA	0.97	1.00				
HIGH_VOLATILITY_BETA	0.97	0.95	1.00			
HIGH_SKEWNESS_BETA	0.90	0.87	0.86	1.00		
LIDX	0.38	0.38	0.38	0.34	1.00	
CPRATIO	0.05	0.05	0.05	0.05	0.06	1.00

stock indicates that the stock is more attractive to investors who enjoy speculative trading and gambling.

B. Return Comovement Measures

The main dependent variables in our empirical analysis are stock-level measures of return comovement. In particular, we compute the degree to which a stock comoves with an index of lottery-type stocks. We compute the annual return comovement measure for each stock i by estimating the following time-series regression:

$$(1) \quad R_{it} - R_{ft} = \beta_0 + \beta_1 \text{CHAR_IDX}_{it} + \beta_2 \text{RMRF}_t + \beta_3 \text{SMB}_t + \beta_4 \text{HML}_t + \beta_5 \text{UMD}_t + \varepsilon_{it}.$$

Here, CHAR_IDX is a return index relative to which the comovement is measured. For example, to measure the return comovement relative to an index of low-priced stocks (i.e., the low-price habitat), we define a low-price index (LOW_PRICE), which is the portfolio return of stocks priced below the 30th New York Stock Exchange (NYSE) percentile of price at the end of the prior year, less the risk-free

rate. In this calculation, we exclude stock i from the index to avoid spurious correlations. Our main variable of interest (β_1) from this regression is the comovement measure of stock i relative to the chosen return index.

Because our central conjecture is that investors' gambling activities induce excess comovement, we focus on β_1 estimates when the CHAR_IDX in equation (1) is an equal-weighted return index of lottery stocks.⁷ We refer to this β_1 estimate interchangeably as *lottery-stock beta* or *lottery comovement measure* in the rest of the paper. Here, lottery stocks are defined as stocks with LIDX values above the 70th percentile. For robustness, we also consider separate comovement measures based on each of the three lottery characteristics that comprise the definition of lottery stocks: low price, high volatility, and high skewness.⁸

We compute betas relative to the CHAR_IDX, the 3 Fama–French (1993) factors (RMRF, SMB, and HML), and the Carhart (1997) UMD factor. We estimate these betas using annual, stock-by-stock time-series regressions using daily data to produce a panel of annual stock-level betas.

C. Geographic Proxy for Gambling Propensity

We proxy for variation in investors' gambling propensity using geographic variation in religious composition. To measure religious composition, we collect data on religious adherence using the Churches and Church Membership files from the Association of Religion Data Archives (ARDA). The data set contains county-level statistics for 133 Judeo–Christian church bodies, including information on the number of churches and the number of adherents of each church. We aggregate membership data from the individual churches to obtain the proportion of Catholics (CATH) and the proportion of Protestants (PROT) in each MSA.

Our primary measure of shareholder gambling propensity is the ratio of Catholics to Protestants (CPRATIO) in the MSA where the firm is headquartered. Using CPRATIO of the firm's location as a measure of shareholder gambling preferences relies on two assumptions: i) that the gambling attitudes of the prevailing local religious group give rise to social norms that influence the behavior of investors in the area, and ii) that there exists some degree of local bias, so that the preferences and trading behavior of local investors impact stock returns. Later, we show complementary results using measures of shareholder preferences based on investor holdings data, which do not rely on the local bias assumption.

Because the CPRATIO measure is based on firm location, we control for other geographic characteristics in our tests. We obtain additional county-level demographic characteristics from the U.S. Census Bureau, which we then aggregate to the MSA level. This set includes the total population of the county, the county-level education (the proportion of county population above age 25 with

⁷We have also replicated our main results using a value-weighted version of the LIDX index and obtain very similar results. A potential issue with value-weighted lottery indices is that, by construction, they assign larger weight on large companies. Because large companies are less likely to be perceived as attractive bets, an equal-weighted index is likely to be a better proxy for gambling attractiveness.

⁸We find qualitatively similar results when we use long–short portfolios to obtain the return comovement estimates. For example, we obtain similar results when we use a lottery-minus-nonlottery factor (LMN), which is the difference between the returns of portfolios of lottery stocks and nonlottery stocks, respectively.

a bachelor's degree or higher), the male–female ratio in the county, the proportion of households in the county with a married couple, the minority population (the non-White proportion of county population), the per capita income of county residents, the median age of county residents, and the proportion of the county residents who live in urban areas.

During our 1980–2005 sample period, the county-level religion and demographic data are available only for years 1980, 1990, and 2000. Following the approach in the recent literature (e.g., Alesina and La Ferrara (2000), Hilary and Hui (2009)), we linearly interpolate the data to obtain the values for intermediate years. We aggregate the county-level data to the MSA level for each year by averaging across all counties in the MSA with a firm headquarters in our data set. The average measure is a weighted average, where the weight is based on the number of firms in the respective county.

D. Other Data Sources

We obtain daily stock returns and financial information for all stocks in the CRSP/Compustat merged database with share codes 10 and 11 for the years 1980–2005. The Fama–French factor returns and risk-free rates come from Professor Kenneth French's data library.⁹

We gather data from several additional sources to construct other variables used in our analysis. Specifically, we use data from a major U.S. discount brokerage house, which contain all trades and end-of-month portfolio positions of a sample of individual investors during the 1991–1996 time period.¹⁰ We also use trade data from the TAQ/ISSM database, as well as institutional holdings data from 13(f) filings provided by Thomson Reuters.

To measure local macroeconomic conditions, we follow Korniotis and Kumar (2013) and use an index consisting of the equal-weighted average of three state-level economic indicators that are likely to move with the state-level business cycle. This set includes the growth rate of state labor income, the relative state unemployment rate, and a state-level version of the housing collateral ratio used in Lustig and van Nieuwerburgh (2010).

The state-level labor income data are obtained from the U.S. Bureau of Economic Analysis (BEA), and the growth rate is defined as the log difference between the state income in a given quarter and the state income in the same quarter in the last year. Relative state unemployment is defined as the ratio of the current state unemployment rate to the moving average of the state unemployment rates in the previous 16 quarters, using state unemployment data from the U.S. Bureau of Labor Statistics (BLS). The housing collateral ratio (hy) is defined as the log ratio of housing equity to labor income, constructed using the Lustig and van Nieuwerburgh (2005) method.

In some of our tests we also use state-level annual per capita lottery sales data for the 37 U.S. states in which lotteries were legal during the sample period.¹¹

⁹The data library is available at <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>

¹⁰See Barber and Odean (2000) for additional details about the brokerage data.

¹¹We thank Garrick Blalock for providing the lottery sales data. See Blalock, Just, and Simon (2007) for additional details about the data.

E. Summary Statistics

Panel A of Table 1 presents summary statistics for the main variables used in our analysis. Our main dependent variables, the return comovement measures, are computed annually, so the unit of observation in the sample is at the stock-year level. The sample includes 66,752 stock-year observation for 8,935 unique stocks. The sample period is from 1980 to 2005 for most variables.

Panel B of Table 1 reports correlations among the return comovement measures and the two main explanatory variables, CPRATIO and LIDX. The betas for lottery stock, low price, high volatility, and high skewness are highly correlated with one another. This finding partially reflects the mild correlations among the measures of share price, idiosyncratic skewness, and idiosyncratic volatility.¹²

III. Gambling-Induced Sentiment and Return Comovement

In this section we present our main empirical results. Our key conjecture is that correlated trading by gambling-motivated investors generates excess return comovement among lottery-like stocks. This conjecture leads to two main testable hypotheses. First, lottery-like stocks would exhibit excess return comovement, controlling for other recognized return factors. Second, if this return comovement is driven by the trading of gambling-motivated investors, then it should be strongest for lottery stocks that are disproportionately held by investors with gambling preferences.

A. Return Comovement among Lottery-Like Stocks

We begin by documenting gambling-induced return comovement in a large cross section of stocks. To quantify the economic importance of this gambling-related return comovement, we identify the set of stocks that have a significant beta estimate relative to the lottery-stock index in equation (1). Panel A of Table 2 reports the beta estimates obtained using equation (1), averaged first cross-sectionally in each year and then across years. We report the beta estimates for the entire sample, as well as for subsets of stocks defined by quintiles of LIDX. We compare the beta estimates relative to the lottery-stock index and its constituents with factor loadings on the standard risk factors.

The beta estimates reported in Table 2 indicate that lottery-stock betas are statistically significant and are comparable in magnitude to the coefficient estimates for the other 4 factors. Furthermore, lottery-stock betas increase significantly with LIDX, whereas loadings on other factors exhibit an opposite pattern. This evidence indicates that high-LIDX stocks comove strongly with other lottery-like stocks. In Panel B we report analogous comovement measures estimated separately for each of the three lottery characteristics. Similar to the lottery-stock

¹²We also examine the correlations among the CHAR_IDX indices and the standard risk factors (i.e., RMRF, SMB, HML, and UMD). We find that the CHAR_IDX indices are strongly and positively correlated with the market factor as well as the SMB factor. The correlations are between 0.65 and 0.77. In contrast, CHAR_IDX indices are negatively correlated with the HML and UMD factors. These correlations range between -0.20 and -0.32 .

TABLE 2
Comovement Estimates Using Lottery-Stock Indices

Table 2 reports average coefficients from annual stock-level time-series regressions of daily firm returns on an equal-weighted index of lottery-like stocks and the RMRF, SMB, HML, and UMD factors. The index of lottery-like stocks consists of stocks above the 70th percentile of LIDX, a measure of stocks' lottery features that takes high values for stocks with low price, high volatility, and high skewness. LIDX is defined in the Appendix. In Panel B, we replace the high-LIDX index with low-price, high-volatility, and high-skewness indices, defined based on the 30th and 70th percentiles, respectively. We suppress the other factor betas for brevity in Panel B. We report average coefficients for all stocks and for subsets of stocks defined by quintiles of LIDX. We average the coefficients cross-sectionally in each year and then across years. *t*-statistics are reported in parentheses below the coefficient estimates, and the sample period is from 1980 to 2005.

Indices	All Stocks	LIDX Quintile				
		Low	2	3	4	High
<i>Panel A. Lottery-Stock Index Beta Estimates</i>						
Lottery-stock index	0.336 (37.42)	-0.076 (-8.84)	0.007 (0.71)	0.187 (20.77)	0.489 (24.45)	0.884 (29.21)
RMRF	0.663 (74.51)	0.957 (70.87)	0.954 (71.40)	0.838 (85.49)	0.554 (24.67)	0.155 (4.85)
SMB	0.479 (48.33)	0.414 (26.33)	0.650 (41.92)	0.692 (42.87)	0.509 (24.94)	0.159 (5.36)
HML	0.057 (3.21)	0.187 (5.51)	0.078 (2.87)	0.033 (1.12)	0.057 (3.18)	0.024 (2.27)
UMD	-0.041 (-4.91)	-0.038 (-1.87)	-0.058 (-3.87)	-0.071 (-4.21)	-0.042 (-3.30)	-0.026 (-2.43)
<i>Panel B. Beta Estimates Relative to Lottery-Stock Index Components</i>						
LOW_PRICE_INDEX	0.323 (29.94)	-0.071 (-8.36)	0.006 (0.67)	0.175 (17.17)	0.464 (20.05)	0.864 (24.78)
HIGH_VOLATILITY_INDEX	0.324 (33.10)	-0.082 (-9.37)	0.007 (0.78)	0.185 (17.92)	0.482 (26.19)	0.848 (27.59)
HIGH_SKEWNESS_INDEX	0.498 (31.66)	-0.081 (-6.40)	0.036 (2.63)	0.299 (20.03)	0.709 (21.29)	1.244 (26.46)

betas, these comovement measures are both economically and statistically significant, and significantly stronger for high-LIDX stocks.

We find that high-LIDX stocks not only comove significantly with one another, but are also less sensitive to the other standard factors. This is evident in Panel A of Table 2, as the significance of the RMRF, SMB, HML, and UMD betas declines substantially as we move from the lowest- to the highest-LIDX quintiles. Furthermore, in untabulated results we test the joint significance of the RMRF, SMB, HML, and UMD betas. Among the highest-LIDX quintile, for 86.65% of stock-years we are unable to reject the null hypothesis that the betas associated with the RMRF, SMB, HML, and UMD factors are jointly equal to 0. In contrast, we fail to reject the null for only 22.00% of stock-years in the lowest-LIDX quintile. Taken together, the results in Table 2 suggest that lottery-like, high-LIDX stocks comove together and are relatively less sensitive to the standard factors.

To examine the economic significance of our findings, we compare the significance of the comovement generated by the gambling activities of investors with the comovement levels generated by the standard asset pricing factors. We find that during our sample period, 14.27% of stocks have a significant (at the 5% level) exposure to gambling-induced sentiment (i.e., a significant coefficient β_1 in equation (1)). In comparison, 39.14%, 23.14%, 15.52%, and 13.18% of stocks have significant betas relative to the market, size, value, and momentum factors, respectively. Furthermore, stocks with significant lottery-stock betas comprise 12.01% of the aggregate stock market value, whereas stocks with significant betas

relative to the market, size, value, and momentum factors comprise 78.01%, 29.76%, 39.11%, and 28.59%, respectively.

These beta estimates indicate that the breadth of the impact of gambling-based sentiment on return comovement is comparable to the importance of the standard asset pricing factors. The gambling-based sentiment affects a smaller segment of the market, but it is economically meaningful.¹³

B. Identifying the Role of Gambling Sentiment

To identify whether the return comovement we observe is indeed driven by the trading of gambling-motivated investors, we test whether the degree of comovement among lottery-like stocks is stronger for lottery stocks that are disproportionately held by investors with gambling preferences. We use CPRATIO in the MSA where the stock is headquartered to proxy for the degree to which gambling-motivated investors own and trade the stock.

Specifically, we estimate ordinary least squares (OLS) regressions in which the lottery comovement measures (i.e., the β_1 estimate from equation (1)) are the dependent variable. The main independent variable is the $\text{CPRATIO} \times \text{LIDX}$ interaction term. The estimates of this interaction term test our second main hypothesis, that comovement among lottery stocks (those with high LIDX) is strongest when the gambling propensity of local investors is high (i.e., CPRATIO is high). In all regressions, we control for the direct effects of CPRATIO and LIDX.

A natural concern in interpreting the $\text{CPRATIO} \times \text{LIDX}$ interaction term is that CPRATIO might capture geographical variation in other investor, firm, or market attributes besides gambling propensity that could be driving the observed return comovement, such as investor wealth or other correlates. We therefore consider several firm attributes and characteristics of the firm's geographic location as control variables. Firm-level variables include the return and turnover of the stock over the prior year, the age of the firm, its market capitalization, and the market-to-book ratio. The geographic variables are a comprehensive set of MSA-level demographic variables including total population, average educational attainment, average age, and the fraction of households that are male, married, or minorities.

In addition, we include a dummy variable that equals 1 if the firm is located in an industry cluster and an index of state-level macroeconomic activity. We include year dummies and industry dummies based on the 48 Fama–French (1997) industries in all of our regressions and cluster standard errors at the firm level.¹⁴

Panel A of Table 3 presents our baseline results. Across all four comovement measures, the CPRATIO-LIDX interaction is highly significant and positive, indicating that excess comovement is strongest for lottery-like stocks that are located in regions where gambling propensity is high. Consistent with the findings in

¹³When we measure significance at the 1% or 10% level, all of these proportions change, but the qualitative picture remains unchanged.

¹⁴We cluster by firm in our baseline regressions because our main variable of interest, the CPRATIO-LIDX interaction, varies at the firm-year level. Our main results are robust to clustering at the MSA level as well.

Table 2, the baseline effect on LIDX shows that comovement with the high-LIDX portfolio is generally stronger among lottery stocks. Interestingly, the baseline effect on CPRATIO is negative, which shows that we are not simply capturing

TABLE 3
Return Comovement Regression Estimates: Baseline Results

Table 3 reports coefficient estimates from regressions of return comovement measures on measures of lottery characteristics (LIDX) and the lottery preferences of local investors (CPRATIO). The comovement measures are based on equation (1), estimated using equal-weighted LIDX, price, volatility, and skewness indices, respectively. LIDX is an index (range 0 to 1) of stocks' lottery characteristics, which takes high values for stocks with high idiosyncratic skewness and volatility and low nominal price. CPRATIO is the ratio of Catholics to Protestants in the MSA where the firm is headquartered. Other control variables include the stock's past 12-month return, share turnover, the natural log of firm age in months, the natural log of firm market capitalization, market/book ratio, a dummy for whether the firm is located in an industry cluster, an index of state macroeconomic conditions, and demographic characteristics of the MSA where the firm is located. These demographic characteristics include total population, education, male-female ratio, proportion of married households, proportion of minority residents, median age, and proportion of residents living in urban areas. All variables are defined in the Appendix. Panel A shows results using LIDX to measure the stock's attractiveness to investors with lottery preferences. Panel B shows results when MSA dummies are included in the specification used in Panel A. *t*-statistics, clustered by firm in Panel A and by MSA in Panel B, are reported in parentheses below the coefficient estimates. All specifications include industry and year dummies. The sample period is from 1980 to 2005.

Independent Variables	Dependent Variable: COMOVEMENT_BETA			
	Lottery Stock	Low Price	High Volatility	High Skewness
<i>Panel A. Baseline Regressions</i>				
CPRATIO × LIDX	0.070 (5.98)	0.075 (6.49)	0.068 (5.99)	0.095 (5.41)
CPRATIO	-0.035 (-5.64)	-0.038 (-6.06)	-0.033 (-5.53)	-0.050 (-5.30)
LIDX	0.827 (23.54)	0.753 (21.93)	0.843 (24.68)	1.038 (19.04)
RETURN _(-1,0)	0.007 (1.49)	0.012 (2.70)	-0.003 (-0.64)	0.000 (-0.04)
TURNOVER	0.213 (5.88)	0.139 (3.89)	0.288 (8.14)	0.377 (6.61)
FIRM_AGE	0.002 (0.37)	0.006 (0.99)	-0.003 (-0.48)	-0.013 (-1.43)
ln(MCAP)	-0.080 (-26.84)	-0.084 (-29.00)	-0.071 (-24.82)	-0.125 (-27.50)
MB_RATIO	0.013 (9.72)	0.011 (8.64)	0.012 (9.27)	0.021 (10.11)
INDUSTRY_CLUSTER	0.000 (0.01)	-0.002 (-0.25)	-0.002 (-0.25)	0.007 (0.48)
TOTAL_POPULATION	0.001 (1.02)	0.001 (0.96)	0.000 (0.37)	0.003 (1.38)
AVG_EDUCATION	0.150 (1.80)	0.148 (1.82)	0.189 (2.33)	0.225 (1.77)
MALE_FEMALE_RATIO	-0.415 (-2.29)	-0.417 (-2.36)	-0.393 (-2.22)	-0.800 (-2.85)
MARRIED	-0.075 (-0.59)	-0.092 (-0.74)	-0.016 (-0.13)	-0.129 (-0.67)
MINORITY	-0.013 (-0.25)	0.009 (0.17)	0.013 (0.26)	-0.046 (-0.59)
MEDIAN_AGE	-0.002 (-0.53)	-0.002 (-0.54)	-0.001 (-0.19)	-0.001 (-0.11)
URBAN	0.187 (2.04)	0.181 (2.07)	0.191 (2.13)	0.300 (2.16)
STATE_MACRO_INDEX	0.010 (1.44)	0.009 (1.43)	0.011 (1.74)	0.021 (2.08)
Industry dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.169	0.172	0.167	0.142
<i>N</i>	65,981	65,981	65,981	65,981

(continued on next page)

TABLE 3 (continued)
Return Comovement Regression Estimates: Baseline Results

Independent Variables	Dependent Variable: COMOVEMENT_BETA			
	Lottery Stock	Low Price	High Volatility	High Skewness
<i>Panel B. Comovement Regression Estimates with Location Fixed Effects</i>				
CPRATIO × LIDX	0.069 (7.02)	0.072 (7.50)	0.068 (7.17)	0.094 (6.38)
LIDX	0.818 (25.17)	0.748 (23.54)	0.832 (26.28)	1.027 (20.26)
RETURN _(-1,0)	0.008 (1.60)	0.013 (2.81)	-0.003 (-0.55)	0.000 (0.04)
TURNOVER	0.212 (5.86)	0.139 (3.89)	0.285 (8.09)	0.378 (6.66)
FIRM_AGE	0.002 (0.27)	0.005 (0.91)	-0.003 (-0.54)	-0.015 (-1.64)
ln(MCAP)	-0.080 (-26.87)	-0.084 (-28.96)	-0.071 (-24.82)	-0.125 (-27.41)
MB_RATIO	0.013 (9.73)	0.011 (8.63)	0.012 (9.27)	0.021 (10.06)
INDUSTRY_CLUSTER	-0.001 (-0.09)	-0.003 (-0.21)	-0.003 (-0.22)	0.015 (0.74)
STATE_MACRO_INDEX	0.018 (2.34)	0.019 (2.57)	0.016 (2.07)	0.036 (3.00)
Industry dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.170	0.173	0.168	0.143
N	65,981	65,981	65,981	65,981

higher comovement in high-CPRATIO regions across all stocks. This evidence highlights the strength of our identification strategy that relies on the interaction between LIDX and CPRATIO. By focusing on this interaction term we are able to exploit both the variation in gambling propensity across MSAs and the variation in lotteryiness across stocks *within* an MSA. Collectively, the results in Panel A are consistent with our gambling-based comovement hypothesis.

The regression estimates are also significant economically. As a specific example, consider the Denver–Aurora MSA, which has an average CPRATIO of about 0.72 and is therefore at the 25th CPRATIO percentile. The Chicago–Naperville–Joliet MSA is at the 75th percentile with a CPRATIO of 3.2. Furthermore, a stock in the 75th percentile of LIDX would have an LIDX value of 0.68 compared with a stock in the 25th percentile of LIDX with an LIDX value of 0.33. Our estimates in Table 3 imply that a lottery-like stock in Chicago comoves much more with other lottery-like stocks than a low-LIDX stock in Denver. In particular, the lottery-stock beta of a high-LIDX stock in Chicago is 0.33 higher than that of a low-LIDX stock in Denver, which is economically significant relative to the mean estimate of 0.31 for lottery-stock betas. Similarly, the effect of moving from the 25th to the 75th percentile of LIDX would be 0.37 in Chicago compared to 0.31 in Denver. Thus, the effect of a stock's lottery characteristics on excess comovement is significantly stronger in MSAs with higher CPRATIOS, where investors exhibit a greater appetite for lottery-like payoffs.

Another way to see that our results are economically significant is to note that the adjusted R^2 of the regressions increases by 16%–22% (on average 19%

across all dimensions) when we include LIDX, CPRATIO, and the interaction term in the specification. Given the large number of controls we use and the fact that we include year and industry dummies, this increase in explanatory power is economically meaningful.

In spite of the large set of control variables we use in these regressions, it is possible that our results could be spuriously driven by some other unobserved geographic characteristic. To address this potential concern, in Panel B of Table 3 we run the same set of regressions as in Panel A but introduce MSA fixed effects in the regression specifications. These extended specifications, which are similar in spirit to Rajan and Zingales (1998), test whether the relation between LIDX and excess return comovement is stronger in high-CPRATIO areas after controlling for differences in average comovement across MSAs. Note that in these regressions, all of our MSA-level variables are effectively captured by the MSA fixed effects.¹⁵ Other features of the regression specifications are unchanged. We continue to include firm controls and industry and year dummies in addition to the MSA fixed effects. With the addition of MSA fixed effects, we also cluster the standard errors at the MSA level.

Panel B of Table 3 shows that our results are qualitatively unchanged and similar in magnitude to our baseline estimates reported in Panel A. This evidence shows that our main results cannot be easily explained by an unobserved time-invariant factor at the MSA level and further supports our conjecture that the CPRATIO–LIDX interaction captures differential gambling attitudes across stocks that differ in their attractiveness as a gamble. Overall, the results in Table 3 provide strong evidence consistent with our conjecture that gambling-motivated correlated trading induces comovement patterns in stock returns.

C. Robustness Checks

We show in Internet Appendix Table A1 that our main results reported in Table 3 are robust to a large battery of checks. We reestimate different versions of our baseline regression in Panel A of Table 3 and show only the estimates of CPRATIO–LIDX interaction for conciseness. The details of these tests are presented in Section A.2 of the Internet Appendix, but we provide a brief summary in this section.

We repeat our analysis using an alternate set of betas, where we estimate lottery-stock betas as well as the betas for low price, high volatility, and high skewness, controlling for market and industry factors, rather than RMRF, SMB, HML, and UMD. We next include a large set of additional control variables to capture comovements induced by fundamental factors. We also consider different versions of the CPRATIO variable. To examine whether microstructure biases affect our results, we exclude all stocks with a price below \$5; additionally, we control for liquidity using Amihud's (2002) illiquidity measure.

¹⁵There is some time-series variation in our demographic variables. However, this variation is small compared to the cross-sectional variance. What is important to note is that even though CPRATIO itself does not vary much in the time series, and therefore might not be separately identified when we include the MSA dummies, the interaction term between CPRATIO and LIDX is still identified.

We also consider additional geographic robustness checks. First, we include a dummy for the 10 largest MSAs to show that our results are not driven merely by large cities and financial centers. Second, we present results where we exclude New York, which is by far the largest MSA and also a high-CPRATIO location. Last, we exclude each of the four U.S. Census Bureau divisions (South, West, North-East, and Mid-West) from the sample.

In all of these alterations of the baseline specification, we find very similar results. Overall, the evidence from these robustness checks indicates that our key finding that excess return comovement is strongest for high-LIDX stocks located in high-CPRATIO areas is robust. Thus, there is considerable support for our gambling-based comovement hypothesis, and our findings are less likely to be explained by other alternative conjectures.

D. Alternate Measures of Investors' Gambling Preferences

Although CPRATIO is a relatively exogenous and broadly available proxy for the degree to which a stock is held and traded by gambling-motivated investors, it has the disadvantage of being somewhat indirect and relying on an assumption of local bias. For robustness, we therefore replace CPRATIO with an alternate measure of the lottery preferences of a stock's investors. Specifically, we measure investors' lottery preferences by observing their portfolio holdings during the prior year.

Although our holdings-based average lottery preference measure is arguably less exogenous than the location-based CPRATIO, it has the advantage of capturing the gambling preference of a stock's investors more directly, without requiring the assumption of local bias. In Internet Appendix Table A1 we report abbreviated results using two versions of this holdings-based lottery preference measure, one based on retail investor holdings and another based on institutional investor holdings. In both cases, we find that lottery-stock comovement is significantly stronger for high-LIDX stocks that have a higher concentration of investors with stronger gambling preferences.

In a separate, related test we show that the degree to which a stock comoves with the high-LIDX portfolio is significantly related to investor clientele characteristics that are associated with gambling propensity. Although the previous analysis has demonstrated this using CPRATIO as a proxy for gambling propensity, in Table 4 we show similar results for a broader set of investor characteristics that have been previously identified as predictors of gambling propensity (see Kumar (2009b)). We use the portfolio holdings of brokerage investors to obtain the average characteristics of the retail investors who hold each stock. We then regress the return comovement measures on these measures of each stock's retail investor clientele.

We find that excess return comovement is stronger for stocks that are held more often by investors whose demographic characteristics imply a greater propensity to gamble. For example, excess return comovement is higher for stocks that are held more often by investors from higher-CPRATIO areas; younger investors; lower-income, nonprofessional, unmarried, and male investors; and investors with lower education levels and more concentrated portfolios. Again, this

TABLE 4
Gambling Clienteles and Return Comovement

Table 4 reports cross-sectional regression estimates, where the dependent variable is a return comovement measure. The betas are defined in the Appendix. The main independent variables are the following characteristics of the retail investors who trade the stock: CPRATIO, age, annual income, education, professional occupation, gender (PROPORTION_MALE), marital status (PROPORTION_MARRIED), proportion African American, proportion Hispanic, proportion Foreign Born, proportion Urban (located within 100 miles of the top 25 U.S. metropolitan regions), average state-level lottery sales, and portfolio concentration (Herfindahl index of portfolio weights). The clientele characteristic is the value-weighted average characteristic of retail investors who trade the stock during the brokerage sample period. To measure these variables, we use data on the portfolio holdings, trading, and demographics of individual investors at a large U.S. discount brokerage house over the 1991–1996 time period. Average investor characteristics from 1996 are used for subsequent years through the end of the sample period (2005). *t*-statistics are reported in parentheses below the coefficient estimates.

Independent Variables	Dependent Variable: COMOVEMENT_MEASURE			
	Lottery Stock	Low Price	High Volatility	High Skewness
AVG_CPRATIO	0.006 (2.81)	0.007 (2.97)	0.005 (3.34)	0.005 (2.99)
AGE	-0.001 (-2.40)	-0.001 (-2.35)	-0.001 (-2.54)	-0.001 (-1.96)
INCOME	-0.000 (-0.60)	-0.000 (-0.38)	-0.000 (-0.46)	-0.000 (-0.22)
PROPORTION_PROFESSIONAL	-0.007 (-1.10)	-0.009 (-1.44)	-0.003 (-1.12)	-0.002 (-0.40)
PROPORTION_MALE	0.032 (2.90)	0.029 (2.87)	0.034 (2.99)	0.042 (2.75)
PROPORTION_MARRIED	-0.010 (-1.37)	-0.012 (-1.56)	-0.008 (-1.63)	-0.013 (-1.23)
PORTFOLIO_CONCENTRATION	0.053 (4.72)	0.044 (3.66)	0.036 (3.49)	0.050 (3.10)
AVG_EDUCATION	-0.002 (-4.34)	-0.001 (-4.12)	-0.001 (-4.06)	-0.002 (-4.11)
PROPORTION_AFRICAN_AMERICAN	-0.001 (-1.22)	-0.001 (-1.18)	-0.001 (-1.49)	-0.001 (-1.32)
PROPORTION_FOREIGN_BORN	-0.000 (-0.90)	-0.000 (-0.98)	-0.000 (-0.70)	-0.001 (-0.96)
PROPORTION_URBAN	0.002 (0.24)	-0.002 (-0.07)	0.003 (0.20)	0.003 (0.18)
AVG_STATE_LOTTO_SALES	0.010 (1.62)	0.011 (2.12)	0.012 (1.80)	0.014 (1.66)
Industry dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.076	0.067	0.076	0.057
<i>N</i>	32,101	32,101	32,101	32,101

evidence provides further direct support for our conjecture that return comovement patterns are at least partly driven by the behavior of gambling-motivated investors.

IV. Direct Evidence Using Investor-Level Data

In order for gambling-motivated investors to cause return comovement among lottery stocks, it is necessary that those investors concentrate their trading within the habitat of lottery-like stocks and that their trades of lottery stocks are correlated. In this section, we analyze the trading activities of both retail and institutional investors to test directly whether these conditions are satisfied. We look at retail and institutional investors separately, because previous research suggests that they differ in their sophistication, in their ability to trade based on information, and in the degree to which behavioral biases would influence their

trading behavior. We also provide additional evidence based on investor-level data to demonstrate that CPRATIO is an effective proxy for gambling attitudes.

A. Gambling and Turnover

We start our analysis by providing some additional validation for the choice of CPRATIO as a proxy for people's gambling attitudes. In our main analysis, we use CPRATIO as a proxy for the level of trading in a particular stock by gambling-motivated investors, based on our conjecture that CPRATIO captures geographical variation in investors' gambling propensity. Table 5 provides evidence consistent with this conjecture. The table reports regressions of portfolio holdings and trading measures on CPRATIO and other stock characteristics. Specifically, we test whether CPRATIO is related to investors' propensity to invest in stocks with lottery-like characteristics (which we measure by the weighted average LIDX of stocks in the investors' portfolios) and by the percentage of the portfolio allocated to stocks in the top two quintiles of LIDX. Furthermore, we test whether CPRATIO is related to portfolio turnover, both overall and particularly among high-LIDX stocks.

Panel A of Table 5 reports results for retail investor portfolios, and Panel B shows evidence from institutional investor portfolios. We find that both retail and

TABLE 5
Gambling Preferences and Trading Behavior

Table 5 reports regressions of portfolio decisions, trading, and trading correlation measures on the CPRATIO of the investor's location and other investor characteristics. In Panels A and B, the dependent variable is either the value-weighted average LIDX of stocks in the portfolio, the proportion of the portfolio allocated to stocks in the top 2 quintiles of LIDX, portfolio turnover, or turnover of high-LIDX stocks. The regressions control for portfolio and demographic characteristics including portfolio size, turnover, total population, education, male-female ratio, the proportion married, the proportion of minorities, median age, and urban proportion, as well as portfolio concentration and industry concentration in Panel B. Demographic variables correspond to the area where the investor is located, and all variables are defined in the Appendix. The coefficients on control variables are suppressed for brevity. Panel A reports results for retail investor trading, and Panel B reports results for institutional investor trading. In Panel C, the dependent variable is RTC, which is the beta from a regression of the BSI of small trades on the average BSI of stocks in certain categories based on lottery characteristics. We measure retail trading correlation with respect to lottery stocks, low-price stocks, high-volatility stocks, high-skewness stocks, and local stocks. Panel D reports the results for ITC and institutional trading proportion (ITP) measures. The regressions in Panels C and D include all of the controls from the regressions in Table 3, Panel A, and the coefficients from the control variables are suppressed for brevity. *t*-statistics, clustered by firm, are reported in parentheses below the coefficient estimates. The sample period is 1991–1996 in Panel A, 1980–2005 in Panels B and D, and 1983–2000 in Panel C.

Independent Variables	Dependent Variable			
	Average LIDX	High-LIDX Weight	Turnover	High-LIDX Turnover
<i>Panel A. Retail Investor Behavior</i>				
CPRATIO	0.396 (6.46)	0.616 (7.04)	0.086 (2.60)	0.011 (2.51)
Portfolio and demographic controls	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.018	0.009	0.001	0.234
<i>N</i>	1,974,638	1,974,638	1,974,638	1,974,638
<i>Panel B. Institutional Investor Behavior</i>				
CPRATIO	0.502 (3.40)	0.200 (2.63)	0.181 (4.33)	0.068 (2.23)
Portfolio and demographic controls	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.296	0.126	0.138	0.115
<i>N</i>	54,914	54,914	54,914	40,846

(continued on next page)

TABLE 5 (continued)
 Gambling Preferences and Trading Behavior

Independent Variables	Dependent Variable: RTC				RTP
	Lottery Stock	Low Price	High Volatility	High Skewness	
<i>Panel C. Retail Trading Correlation Regression Estimates</i>					
CPRATIO × LIDX	0.013 (2.12)	0.008 (1.20)	0.013 (2.02)	0.011 (1.57)	1.025 (5.39)
All controls	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.033	0.027	0.031	0.035	0.267
N	36,519	36,513	36,517	36,465	36,900
<i>High-Macroeconomic-Index Subsample</i>					
CPRATIO × LIDX	0.016 (2.27)	0.012 (1.71)	0.016 (2.25)	0.012 (1.71)	3.255 (4.14)
<i>Low-Macroeconomic-Index Subsample</i>					
CPRATIO × LIDX	-0.003 (-0.32)	0.001 (0.13)	0.000 (-0.02)	0.001 (0.10)	-0.876 (0.74)
<i>Panel D. Institutional Trading Correlation Regression Estimates</i>					
CPRATIO × LIDX	0.023 (1.80)	0.027 (1.43)	0.032 (1.41)	0.001 (0.06)	0.010 (2.07)
All controls	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.001	0.001	0.001	0.004	0.409
N	59,508	59,507	59,507	59,506	65,253

institutional investors in high-CPRATIO areas invest more heavily in high-LIDX stocks.¹⁶ In addition, investors in those regions trade more actively, particularly among high-LIDX stocks. This evidence suggests that CPRATIO is likely to capture variation in investors' tendency to trade actively within gambling habitats, which could generate excess return comovement among stocks with lottery characteristics.

B. Gambling and Correlated Retail Trading

The second necessary condition for our proposed gambling-comovement relation is the existence of a systematic component in investors' trades (i.e., their trades must be correlated). In this section we provide evidence that the correlations among retail trades are stronger for high-LIDX stocks in high-CPRATIO areas.

To assess the degree of correlated trading, we construct a measure of retail trading correlation (RTC) using small trades from the ISSM and the TAQ databases. Following recent studies, we use trade size to identify retail trades and

¹⁶Kumar (2009b) shows that gambling-motivated investors do not distinguish between local and nonlocal lottery-type stocks. They overweight both local and nonlocal lottery-type stocks. Similar to that evidence, in unreported results, we find that gambling-motivated investors in high-CPRATIO regions trade both local and nonlocal lottery-type stocks. This evidence suggests that gambling-based trading activities of investors can generate comovement patterns even among stocks that are not local to them.

obtain a time series of retail buy–sell imbalance for each stock.¹⁷ We then measure the degree to which retail demand for a particular stock is correlated with the retail demand for lottery-type stocks, low-priced stocks, local stocks, and so forth. The RTC measure is defined as the coefficient from a regression of the retail buy–sell imbalance (BSI) of the individual stock on the average BSI of, say, low-priced stocks.

Specifically, we measure the level of correlation in the trading activities of retail investors by estimating time-series regressions of the following form:

$$(2) \quad \text{BSI}_{it} = \beta_0 + \beta_1 \text{BSI}_{pt} + \beta_2 \text{RMRF}_t + \varepsilon_{it},$$

where BSI_{it} is the period- t BSI of stock i , BSI_{pt} is the period- t BSI of portfolio p , and RMRF_t is the period- t market return excess over the risk-free rate. The period- t BSI for stock i is defined as

$$(3) \quad \text{BSI}_{it} = \frac{\text{VB}_{it} - \text{VS}_{it}}{\text{VB}_{it} + \text{VS}_{it}},$$

where VB_{it} and VS_{it} are the period- t dollar buy and sell trading volumes of stock i , respectively. BSI_{pt} is the equal-weighted average of the period- t BSI of all stocks that belong to portfolio p . We estimate the time-series regression annually using monthly data.

The results reported in Panel C of Table 5 indicate that retail trades are more correlated when a stock has stronger lottery characteristics and is located in high-CPRATIO regions where investors have a stronger gambling propensity. The results are qualitatively similar but weaker when we use the institutional trading correlation (ITC) measure as the dependent variable. This measure is defined similar to the RTC measure, where we use quarterly changes in portfolio holdings of institutions to measure institutional trades. The estimates reported in the last columns of Panels C and D indicate that the levels of retail trading as well as institutional trading are high for stocks that have lottery features and are located in high-CPRATIO regions.

Interestingly, when we split the sample into years with relatively good and bad economic conditions, we find that retail participation is higher. Plus, we find that the correlation in retail demand is higher among lottery-like stocks that are located in higher-CPRATIO areas. This evidence suggests that fluctuations in economic conditions may help generate a systematic component in investors' demand for lottery-like stocks, which is necessary for investors' trading to induce excess comovement.

Overall, the trading level and trading correlation regression estimates demonstrate that gambling-induced sentiment exists. This evidence provides additional

¹⁷We use the retail trading data for the 1983–2000 time period obtained from ISSM/TAQ, where we use small-sized trades (trade size \leq \$5,000) to proxy for retail trades. Like Barber, Odean, and Zhu (2009), we use the ISSM/TAQ data only until 2000 because the assumption that small trades proxy for retail trading is less likely to be valid after 2000. In particular, the introduction of decimalized trading in January 2001 and extensive order-splitting by institutions due to reduced trading costs make small trade size a less reliable proxy for retail trading after 2000. See Hvidkjaer (2008) or Barber et al. (2009) for additional details about the ISSM/TAQ data sets, including the procedure for identifying small trades.

direct support to our conjecture that gambling-motivated trading of retail and institutional investors induces comovement in stock returns.

V. Sources of Gambling Sentiment

So far, our results indicate that lottery-like stocks comove strongly with one another, and this return comovement is strongest among lottery stocks located in high-CPRATIO regions, where investors have a greater propensity to gamble. Furthermore, CPRATIO captures variation in the degree of correlation in retail demand for lottery stocks, and such correlation in demand is necessary for gambling-motivated investors to generate excess return comovement among lottery stocks. We consider two possible mechanisms that may cause investors' demand for lottery stocks to fluctuate in a systematic way.

First, investors may be influenced by time variation in people's enthusiasm for gambling. As the population's attitudes or enthusiasm toward gambling rise and fall over time, investors' demand for lottery-like stocks may fluctuate with it, generating broad correlation in their trades for lottery stocks. Second, gambling-motivated investors may experience correlated income shocks as macroeconomic conditions fluctuate over time, causing their demand for lottery-like stocks to fluctuate in a correlated way. We provide evidence for both of these mechanisms. Specifically, we demonstrate that the strength of the relation between CPRATIO and lottery stock comovement varies with both general gambling sentiment and macroeconomic conditions.

A. Evidence Using Lottery Ticket Sales as a Gambling Proxy

We measure general gambling enthusiasm using state-level per capita lottery ticket sales.¹⁸ An advantage of using state-level lottery sales data is that it gives us more power by allowing for both time-series and cross-sectional variation in the population's enthusiasm for gambling. We first split our sample for each year into above- and below-median groups based on per capita lottery ticket sales in the state where the firm is located. We then reestimate our baseline regressions using these subsamples.

Panel A of Table 6 presents the results; for brevity, we do not report the estimates of other control variables, although they are included in the regressions. The patterns across the two subsamples are striking. The $CPRATIO \times LIDX$ estimates are substantially higher in the subsample with high lottery ticket sales and are insignificant in the subgroup with low lottery sales. We find consistent results when we estimate a regression with a triple interaction between CPRATIO, LIDX, and a HIGH_LOTTO_SALES that is 1 if the state-level lottery ticket sales in the year are above median.

¹⁸There is considerable geographical dispersion in the types of state lotteries that are available, the states and institutions that offer them, and the accuracy of the lottery sales data. However, for a subset of states and years, we were able to obtain reliable state-level lottery sales data. We have data for 37 states during the 1990–2002 period. These data have also been used in other studies (e.g., Kumar (2009b)), and we have no reason to believe that the sample is biased.

These results indicate that the relation between CPRATIO and lottery stock comovement is stronger when lottery ticket sales are high. This evidence is consistent with our conjecture that general gambling sentiment influences investors' demand for lottery stocks, which consequently generates a systematic component in the trading activities of gambling-motivated investors.

TABLE 6
Lottery Sales, Local Economy, and Gambling-Induced Comovement

Table 6 reports coefficient estimates from regressions of return comovement measures on measures of lottery characteristics (LIDX) and the lottery preferences of local investors (CPRATIO), for subsamples based on local lottery sales and local (state) macroeconomic conditions. LIDX is an index (range 0 to 1) of stocks' lottery characteristics, which takes high values for stocks with high idiosyncratic skewness and volatility and low nominal price. CPRATIO is the ratio of Catholics to Protestants in the MSA where the firm is headquartered. Other control variables (suppressed for brevity) include the stock's past 12-month return, share turnover, the natural log of firm age in months, the natural log of firm market capitalization, market/book ratio, a dummy for whether the firm is located in an industry cluster, an index of state macroeconomic conditions, and demographic characteristics of the MSA where the firm is located. These demographic characteristics include total population, education, male–female ratio, proportion of married households, proportion of minority residents, median age, and the proportion of residents living in urban areas. All variables are defined in the Appendix. Panel A shows results when we partition the sample according to state-level lottery ticket sales. The top panel shows results for the subsample of firms located in states with above-median per capita lottery sales. The middle panel shows results for the subsample of firms located in states with below-average lottery sales. The bottom panel includes a triple interaction with CPRATIO, LIDX, and state lottery sales. *t*-statistics, clustered by firm, are reported in parentheses below the coefficient estimates. Panel B presents results when we partition the sample by local economic conditions, measured using an index defined as the equal-weighted average of the housing collateral ratio, change in unemployment, and change in local income growth (all measured at the state level). The top panel shows results for the subsample of firms located in states with above-median macroeconomic conditions. The middle panel shows results for the subsample of firms located in states with below-median macro conditions. The bottom panel includes a triple interaction with CPRATIO, LIDX, and state macroeconomic conditions. All specifications include industry and year dummies. The sample period is 1990–2002 for Panel A, and 1980–2005 for Panel B.

Independent Variables	Dependent Variable: COMOVEMENT_BETA			
	Lottery Stock	Low Price	High Volatility	High Skewness
<i>Panel A. Lottery-Ticket-Sales-Based Subsamples and Interactions</i>				
<i>High State-Level Lottery Ticket Sales</i>				
CPRATIO × LIDX	0.081 (4.31)	0.086 (4.65)	0.077 (4.20)	0.095 (3.31)
CPRATIO	−0.048 (−4.45)	−0.049 (−4.65)	−0.044 (−4.19)	−0.057 (−3.51)
LIDX	0.730 (10.04)	0.689 (9.73)	0.729 (10.27)	1.086 (9.58)
<i>Low State-Level Lottery Ticket Sales</i>				
CPRATIO × LIDX	−0.005 (−0.19)	−0.003 (−0.12)	−0.008 (−0.33)	0.003 (0.09)
CPRATIO	−0.011 (−0.85)	−0.009 (−0.69)	−0.010 (−0.80)	−0.029 (−1.50)
LIDX	0.979 (14.27)	0.939 (13.96)	0.975 (14.49)	1.308 (12.38)
<i>State-Level Lottery Ticket Sales Interactions</i>				
CPRATIO × LIDX × HIGH_LOTTO_SALES	0.083 (2.82)	0.088 (3.04)	0.082 (2.86)	0.093 (2.05)
CPRATIO × LIDX	−0.001 (−0.05)	−0.001 (−0.04)	−0.004 (−0.15)	0.005 (0.13)
CPRATIO × HIGH_LOTTO_SALES	−0.015 (−1.10)	−0.015 (−1.14)	−0.019 (−1.49)	−0.010 (−0.49)
LIDX × HIGH_LOTTO_SALES	−0.239 (−2.71)	−0.240 (−2.79)	−0.244 (−2.81)	−0.248 (−1.77)
CPRATIO	−0.029 (−2.52)	−0.030 (−2.62)	−0.024 (−2.14)	−0.044 (−2.51)
LIDX	1.003 (15.07)	0.964 (14.74)	1.002 (15.41)	1.367 (13.36)
HIGH_LOTTO_SALES	0.107 (2.91)	0.111 (3.08)	0.118 (3.28)	0.106 (1.85)

(continued on next page)

TABLE 6 (continued)
 Lottery Sales, Local Economy, and Gambling-Induced Comovement

Independent Variables	Dependent Variable: COMOVEMENT_BETA			
	Lottery Stock	Low Price	High Volatility	High Skewness
<i>Panel B. State-Level-Macroeconomic-Condition-Based Subsamples and Interactions</i>				
<i>High State-Level Macroeconomic Index</i>				
CPRATIO × LIDX	0.125 (5.07)	0.136 (5.64)	0.121 (5.06)	0.169 (4.51)
CPRATIO	-0.067 (-4.97)	-0.072 (-5.39)	-0.066 (-5.01)	-0.090 (-4.28)
LIDX	0.806 (11.97)	0.717 (10.86)	0.819 (12.51)	1.046 (9.92)
<i>Low State-Level Macroeconomic Index</i>				
CPRATIO × LIDX	0.057 (4.58)	0.061 (4.94)	0.055 (4.55)	0.077 (4.14)
CPRATIO	-0.030 (-4.48)	-0.032 (-4.76)	-0.029 (-4.48)	-0.043 (-4.22)
LIDX	0.790 (19.92)	0.719 (18.63)	0.810 (21.02)	0.971 (15.86)
<i>State-Level Macroeconomic Index Interactions</i>				
CPRATIO × LIDX × HIGH_MACRO_INDEX	0.086 (3.22)	0.094 (3.59)	0.084 (3.25)	0.115 (2.85)
CPRATIO × LIDX	0.058 (4.68)	0.062 (5.06)	0.056 (4.66)	0.080 (4.32)
CPRATIO × HIGH_MACRO_INDEX	-0.050 (-4.13)	-0.055 (-4.57)	-0.049 (-4.19)	-0.066 (-3.60)
LIDX × HIGH_MACRO_INDEX	-0.059 (-0.87)	-0.080 (-1.20)	-0.061 (-0.93)	-0.050 (-0.47)
CPRATIO	-0.025 (-3.90)	-0.026 (-4.12)	-0.024 (-3.81)	-0.037 (-3.84)
LIDX	0.836 (21.78)	0.770 (20.55)	0.851 (22.80)	1.038 (17.45)
HIGH_MACRO_INDEX	0.026 (0.90)	0.035 (1.22)	0.034 (1.20)	0.017 (0.37)

B. Impact of Local Economic Conditions

We perform a similar set of tests to examine whether the strength of the relation between CPRATIO and lottery stock comovement varies with economic conditions. As with the analysis based on lottery ticket sales, we take advantage of both time-series and cross-sectional variation in economic conditions. We measure economic conditions at the state level using a state-level macroeconomic index proposed in Korniotis and Kumar (2013). This measure of economic activity is defined as an equal-weighted measure of standardized state-level income growth, relative state unemployment, and the state-level housing collateral ratio.

We use variation in economic conditions to capture correlated income shocks that would affect the ability of gambling-motivated investors to allocate capital to their preferred lottery-like stocks. Conditional on the level of the national economy, when local economic conditions in high-CPRATIO areas are relatively strong, their higher current or future income allows local investors to become more important to the return-generating process of local stocks.¹⁹ For retail

¹⁹Note that this conjecture is not contrary to evidence that gambling appetite usually increases during bad economic times, because the prior statement is conditional on income. We are making the inverse claim. Holding overall economic conditions constant, we look at the income effect.

investors in particular, higher income levels could increase retail participation. In addition, local institutions could become more important if local economic conditions are correlated with inflows into these funds.

Panel B of Table 6 shows that the relation between CPRATIO and lottery stock comovement is stronger in state-years when local economic conditions are relatively good. For all of our measures of lottery stock comovement, the coefficient on the $\text{CPRATIO} \times \text{LIDX}$ term is more than twice as large in the years when local economic conditions are relatively good, and the specification with a triple interaction confirms that these differences are statistically significant.

In sum, the evidence in Table 6 indicates that gambling-motivated investors play a stronger role in generating return comovement among lottery-like stocks when local lottery ticket sales are higher, or when local economic conditions are relatively good. These findings suggest that both time variation in local gambling attitudes and income shocks due to fluctuating economic conditions serve as coordinating mechanisms that generate systematic components in investors' demand for lottery stocks.

VI. Summary and Conclusion

This paper examines whether the gambling-motivated trading activity of retail and institutional investors (i.e., gambling-induced sentiment) is a significant source of comovement among stock returns. We find that stocks with lottery-like features comove strongly, while being relatively unaffected by commonly recognized return factors. In economic terms, these effects are large. The set of stocks that exhibit gambling-induced comovement comprises more than 12% of the overall stock market value.

Using a religion-based proxy for investors' gambling propensity, we find robust evidence that the return comovement among lottery-like stocks is strongest for lottery stocks that are held and traded more by investors with a higher propensity to gamble. An alternative measure of investors' gambling preferences based on their past portfolio allocations yields similar results. Return comovement is also higher among stocks that are held more intensely by investors with demographic characteristics that are known to be associated with gambling behavior.

We also examine trading and holdings data directly to confirm that our religion-based gambling proxy captures variation in the type of investor behavior necessary to generate excess return comovement in a habitat-based trading framework. Specifically, we find that investors in areas with a high ratio of Catholics to Protestants, whom we expect to have a higher propensity to gamble, overweight lottery-like stocks in their portfolios and trade lottery stocks more actively. In addition, we find that retail demand for lottery stocks is more strongly correlated for lottery stocks that are located in areas where investors have a higher propensity to gamble.

We examine two possible sources that may contribute to a systematic component in investors' demand for lottery-like stocks. First, investors may be influenced by broad variation over time in people's enthusiasm for gambling, so that as general gambling sentiment rises or falls over time, investors' demand for lottery-like stocks fluctuates in a correlated way. Second, gambling-motivated

investors may experience correlated income shocks as macroeconomic conditions fluctuate over time, causing their demand for lottery-like stocks to fluctuate in a correlated way. We find evidence supporting both of these potential sources of systematic gambling sentiment.

Taken together, these results indicate that gambling-motivated trading is a significant source of comovement in stock returns. The gambling-motivated activities of both retail and institutional investors induce excess comovement in stock returns. The empirical findings extend the literature on gambling and investment decisions. Most of the previous literature has documented that gambling is important for investment decisions. Our paper is the first to directly investigate the impact of gambling attitudes on stock return comovement. Our findings also extend the recent literature on nonfundamentals-based return comovement and demonstrate that gambling behavior is an important determinant of comovement in stock returns. More broadly, our study contributes to the debate on the importance of investors' behavioral biases for aggregate market outcomes.

Our findings have important implications for portfolio decisions of investors. In particular, investors who are not aware of the existence of excess comovement among lottery-type stocks may underestimate the riskiness of those stocks and may overweight those stocks in their portfolios. In future research, it would be interesting to estimate the economic costs of these potential distortions in investor portfolios.

Appendix. Brief Variable Definitions and Sources

In this Appendix we briefly define the main variables used in the empirical analysis. The data sources are: i) 13(f): 13(f) institutional portfolio holdings data from Thomson Reuters, ii) ARDA: Association of Religion Data Archives, iii) BEA: U.S. Bureau of Economic Analysis, iv) BLS: U.S. Bureau of Labor Statistics, v) Brokerage: large U.S. discount brokerage, vi) Census: U.S. Census County files, vii) Compustat, viii) CRSP: Center for Research in Security Prices, ix) Estimated: estimated by the authors, x) ISSM/TAQ: Institute for the Study of Security Markets (ISSM) and the Trade and Quote (TAQ) databases, xi) LOTAG: state lottery agencies. Table 1 reports the summary statistics for all variables.

1. Return Comovement Measures (Betas)

All return comovement measures are estimated as β_1 from the regression

$$r_{it} - r_{ft} = \beta_0 + \beta_1 \text{CHAR_IDX}_{it} + \beta_2 \text{RMRF}_t + \beta_3 \text{SMB}_t + \beta_4 \text{HML}_t + \beta_5 \text{UMD}_t + \varepsilon_{it},$$

where CHAR_IDX_{it} is the day- t return of an equal-weighted portfolio of stocks in a certain category, excluding stock i . The betas are estimated annually using daily data. The compositions of the category portfolios used for each comovement measure are described in the following section.

2. Stock Category Definitions

LOTTERY_STOCK: Stocks above the 70th NYSE percentile of the lottery characteristics index (LIDX), which is calculated from stocks' price, volatility, and skewness as described below.

LOW_PRICE: Stocks priced below the 30th NYSE percentile of price at the end of the prior year.

HIGH_VOLATILITY: Stocks above the 70th NYSE percentile of total daily return volatility over the prior year.

HIGH_SKEWNESS: Stocks above the 70th NYSE percentile of total daily return skewness over the prior year.

3. Stock Characteristics

VOLATILITY: Total volatility (standard deviation) of daily returns measured over the year (estimated). *Source:* CRSP.

SKEWNESS: Total skewness of daily returns measured over the year (estimated). *Source:* CRSP.

LIDX: Lottery characteristics index; stocks are assigned to vigintiles (semi-deciles) by price, volatility, and skewness (where 20 is the lowest price group and the highest volatility and skewness groups). The price, volatility, and skewness vigintile assignments are added for each stock to produce a score ranging from 3 to 60, which is then scaled to range from 0 to 1 using $(\text{Score} - 3)/(60 - 3)$ (estimated). *Source:* CRSP.

AVG_LOTTERY_PREFERENCE (RETAIL): Value-weighted average lottery preference of the stock's shareholders from the brokerage data set, where lottery preference is defined as the average LIDX of stocks in the household's portfolio over the prior 12 months. *Source:* Brokerage.

AVG_LOTTERY_PREFERENCE (INST): Value-weighted average lottery preference of the stock's institutional shareholders, where lottery preference is defined as the average LIDX of stocks in the institution's portfolio over the prior 4 quarters. *Source:* 13(f).

RETURN_(-1,0): 12-month stock return over the prior year. *Source:* CRSP.

TURNOVER: Average monthly share turnover (share volume/shares outstanding) over the prior year. *Source:* CRSP.

ln(FIRM_AGE): Natural log of the number of months since the stock appeared on CRSP. *Source:* CRSP.

ln(MCAP): Natural log of price \times shares outstanding, in millions. *Source:* CRSP.

MB_RATIO: Ratio of market value of equity to book value of equity. *Source:* Compustat.

AVG_CPRATIO: CPRATIO in county where shareholder resides, averaged across the stock's retail shareholders in the brokerage data set. *Source:* Brokerage.

INCOME: Average income of stock's retail shareholders. *Source:* Brokerage.

PROPORTION_PROFESSIONAL: Proportion of the stock's retail shareholders with a professional occupation. *Source:* Brokerage.

PROPORTION_MALE: Proportion of the stock's retail shareholders that are male. *Source:* Brokerage.

PROPORTION_MARRIED: Proportion of the stock's retail shareholders that are married. *Source:* Brokerage.

PORTFOLIO_CONCENTRATION: Average portfolio concentration of the stock's retail shareholders, measured by the Herfindahl index of the investor's portfolio weights. *Source:* Brokerage.

AVG_EDUCATION: Proportion of residents over age 25 with a bachelor's degree or higher in the county where the shareholder resides, averaged across the stock's retail shareholders in the brokerage data set. *Source:* Census, Brokerage.

PROPORTION_AFRICAN_AMERICAN: Proportion of African-Americans in the county where the shareholder resides, averaged across the stock's retail shareholders in the brokerage data set. *Source:* Census, Brokerage.

PROPORTION_FOREIGN_BORN: Proportion of residents that are foreign-born in the county where the shareholder resides, averaged across the stock's retail shareholders in the brokerage data set. *Source:* Census, Brokerage.

PROPORTION_URBAN: Proportion of the stock's retail shareholders that live within 100 miles of one of the top 25 U.S. metropolitan areas. *Source:* Census, Brokerage.

AVG_STATE_LOTTO_SALES: Annual per capita lottery sales in the state where the shareholder resides, averaged across the stock's retail shareholders in the brokerage data set. *Source:* LOTAG, Brokerage.

4. Regional Characteristics

CPRATIO: Ratio of Catholic population to Protestant population in the MSA. *Source:* ARDA.

INDUSTRY_CLUSTER_DUMMY: MSA-level measure equal to 1 if 10% or more of the market capitalization of firms located in the MSA is from a single industry, and 10% or more of that industry's market capitalization is located in that MSA. *Source:* CRSP, Compustat.

TOTAL_POPULATION: Total population of the MSA. *Source:* Census.

AVG_EDUCATION: Proportion of MSA residents over age 25 with a bachelor's degree or higher. *Source:* Census.

MALE_FEMALE_RATIO: Ratio of males to females in the MSA. *Source:* Census.

MARRIED: Proportion of households in the MSA with a married couple. *Source:* Census.

MINORITY: Proportion of MSA residents who are non-White. *Source:* Census.

AGE: Median age in the MSA. *Source:* Census.

URBAN: Proportion of MSA residents living in urban areas. *Source:* Census.

STATE_MACRO_INDEX: Equal-weighted average of three state-level measures: the growth rate of state labor income, the relative state unemployment rate, and a state-level version of the housing collateral ratio (*hy*). See Korniotis and Kumar (2011) for details on the individual macroeconomic measures. *Source:* BEA, BLS.

HIGH_MACRO_INDEX: Equals 1 when STATE_MACRO_INDEX is above median.

STATE_LOTTERY_SALES: Annual per capita lottery sales in the state. *Source:* LOTAG.

HIGH_LOTTO_SALES: Equals 1 when STATE_LOTTERY_SALES is above median.

5. Trading Measures

RTC: Retail trading correlation; all RTC measures are estimated as β_1 from the regression

$$BSI_{it} = \beta_0 + \beta_1 CHAR_IDX_t + \beta_2 RMRF_t + \varepsilon_{it},$$

where BSI_{it} is the BSI of small trades in stock i during month t , and $CHAR_IDX_t$ is the equal-weighted mean BSI of stocks in a certain category, excluding stock i . The BSI beta is estimated annually using monthly data. The compositions of the category portfolios used for each trading correlation measure are identical to those used for the comovement measures. *Source:* ISSM/TAQ.

RTP: Retail trading proportion; ratio of dollar volume of small trades to total dollar volume of trades in the stock. *Source:* ISSM/TAQ.

ITC: Institutional trading correlation; analogous to the RTC measure but using institutional trades, as proxied by quarterly changes in portfolio holdings. *Source:* 13(f).

ITP: Institutional trading proportion; ratio of dollar volume of large institutional trades to total dollar volume of trades in the stock. Quarterly changes in portfolio holdings are used to define institutional trades. *Source:* 13(f).

References

- Alesina, A., and E. La Ferrara. "Participation in Heterogeneous Communities." *Quarterly Journal of Economics*, 115 (2000), 847–904.
- Amihud, Y. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5 (2002), 31–56.
- Barber, B. M., and T. Odean. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance*, 55 (2000), 773–806.
- Barber, B. M.; T. Odean; and N. Zhu. "Do Noise Traders Move Markets?" *Review of Financial Studies*, 22 (2009), 151–186.
- Barberis, N.; A. Shleifer; and J. Wurgler. "Comovement." *Journal of Financial Economics*, 75 (2005), 283–317.
- Becker, B.; Z. Ivković; and S. Weisbenner. "Local Dividend Clienteles." *Journal of Finance*, 66 (2011), 655–683.
- Blalock, G.; D. R. Just; and D. H. Simon. "Hitting the Jackpot or Hitting the Skids: Entertainment, Poverty, and the Demand for State Lotteries." *American Journal of Economics and Sociology*, 66 (2007), 545–570.

- Boyer, B. "Style-Related Comovement: Fundamentals or Labels?" *Journal of Finance*, 66 (2011), 307–332.
- Carhart, M. M. "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52 (1997), 57–82.
- Coelho, L.; K. John; and R. J. Taffler. "Bankrupt Firms: Who's Buying?" Working Paper, University of Warwick (2010).
- Coval, J. D., and T. J. Moskowitz. "Home Bias at Home: Local Equity Preference in Domestic Portfolios." *Journal of Finance*, 54 (1999), 2045–2073.
- Doran, J. S.; D. Jiang; and D. R. Peterson. "Gambling Preference and the New Year Effect of Assets with Lottery Features." *Review of Finance*, 16 (2011), 685–731.
- Dorn, D., and P. Sengmueller. "Trading as Entertainment." *Management Science*, 55 (2009), 591–603.
- Fama, E. F., and K. R. French. "Industry Costs of Capital." *Journal of Financial Economics*, 43 (1997), 153–193.
- Froot, K. A., and E. M. Dabora. "How Are Stock Prices Affected by the Location of Trade?" *Journal of Financial Economics*, 53 (1999), 189–216.
- Gao, X., and T.-C. Lin. "Do Individual Investors Trade Stocks as Gambling? Evidence from Repeated Natural Experiments." Working Paper, University of Hong Kong (2011).
- Grinblatt, M., and M. Keloharju. "Sensation Seeking, Overconfidence, and Trading Activity." *Journal of Finance*, 64 (2009), 549–578.
- Hilary, G., and K. W. Hui. "Does Religion Matter in Corporate Decision Making in America?" *Journal of Financial Economics*, 93 (2009), 455–473.
- Hoffmann, A., and H. Shefrin. "Online Investors: What They Want, What They Do, and How Their Portfolios Perform." Working Paper, Santa Clara University (2011).
- Huberman, G. "Familiarity Breeds Investment." *Review of Financial Studies*, 14 (2001), 659–680.
- Hvidkjaer, S. "Small Trades and the Cross-Section of Stock Returns." *Review of Financial Studies*, 21 (2008), 1123–1151.
- Ivković, Z., and S. Weisbender. "Local Does as Local Is: Information Content of the Geography of Individual Investors' Common Stock Investments." *Journal of Finance*, 60 (2005), 267–306.
- Korniotis, G. M., and A. Kumar. "Do Behavioral Biases Adversely Affect the Macro-Economy?" *Review of Financial Studies*, 24 (2011), 1513–1559.
- Korniotis, G. M., and A. Kumar. "State-Level Business Cycles and Local Return Predictability." *Journal of Finance*, 68 (2013), 1037–1096.
- Kumar, A. "Hard-to-Value Stocks, Behavioral Biases, and Informed Trading." *Journal of Financial and Quantitative Analysis*, 44 (2009a), 1375–1401.
- Kumar, A. "Who Gambles in the Stock Market?" *Journal of Finance*, 64 (2009b), 1889–1933.
- Kumar, A., and C. M. C. Lee. "Retail Investor Sentiment and Return Comovements." *Journal of Finance*, 61 (2006), 2451–2486.
- Kumar, A.; J. Page; and O. Spalt. "Religious Beliefs, Gambling Attitudes, and Financial Market Outcomes." *Journal of Financial Economics*, 102 (2011), 671–708.
- Lee, C. M. C.; A. Shleifer; and R. H. Thaler. "Investor Sentiment and the Closed-End Fund Puzzle." *Journal of Finance*, 46 (1991), 75–109.
- Lustig, H., and S. van Nieuwerburgh. "Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective." *Journal of Finance*, 60 (2005), 1167–1219.
- Lustig, H., and S. van Nieuwerburgh. "How Much Does Household Collateral Constrain Regional Risk Sharing?" *Review of Economic Dynamics*, 13 (2010), 265–294.
- Rajan, R. G., and L. Zingales. "Financial Dependence and Growth." *American Economic Review*, 88 (1998), 559–586.
- Schneider, C., and O. G. Spalt. "Conglomerate Investment, Skewness, and the CEO Long-Shot Bias." *Journal of Finance*, forthcoming (2016).
- Shu, T.; J. Sulaeman; and P. E. Yeung. "Local Religious Beliefs and Mutual Fund Risk-Taking Behaviors." *Management Science*, 58 (2012), 1779–1796.