Investor Clienteles and Habitat-Based Return Comovements: Direct Evidence*

Alok Kumar

McCombs School of Business

University of Texas at Austin

Jeremy Page

McCombs School of Business

University of Texas at Austin

Oliver Spalt

Department of Finance

Tilburg University

August 25, 2009

*Please address all correspondence to Alok Kumar, Department of Finance, McCombs School of Business, University of Texas at Austin, 1 University Station, B6600, Austin, TX 78712; *Phone:* 512-232-6824; *email:* akumar@mail.utexas.edu. Jeremy Page can be reached at 512-431-6157 or jeremy.page@phd.mccombs.utexas.edu. Oliver Spalt can be reached at +31 13-466-3545 or o.g.spalt@uvt.nl. We would like to thank John Griffin, George Korniotis, and seminar participants at UT-Austin for helpful discussions and valuable comments. We are responsible for all remaining errors and omissions.

Investor Clienteles and Habitat-Based Return Comovements: Direct Evidence

ABSTRACT

Using a large database of portfolio holdings and trades, we identify several investor habitats and show that investors' correlated trading activities generate return comovements among the stocks within those habitats. We focus on geography- and price-based return comovements but also consider other related forms of comovement. First, we show that both retail and institutional investors specialize and exhibit a greater propensity to buy and sell stocks that have similar characteristics. Specifically, due to local stock preference and a propensity to gamble, investors specialize in local stocks, low-priced stocks, and stocks with lottery features. Next, we show that specialization-induced correlated trading generates return comovements within geography- and price-based stock categories. The comovement patterns are stronger among stocks with lottery features, high retail concentration, and high arbitrage costs. Across time, comovements get amplified during periods of greater aggregate uncertainty and stronger consumer sentiment. Overall, our results provide *direct* empirical support for the habitat-based return comovements model proposed in Barberis, Shleifer, and Wurgler (2005).

1. Introduction

One of the fundamental goals of asset pricing theory is to understand the sources of common variation in security prices. The traditional asset pricing theory posits that return comovements are induced by correlated cash flows or changes in the discount rates. However, recent theoretical studies conjecture that trading activities of investors can induce common factors in security returns that are unrelated to fundamentals. In particular, Barberis, Shleifer, and Wurgler (2005) propose a habitat-based model of return comovements in which investors with certain attributes are attracted toward stocks with certain characteristics and investor clienteles emerge. Consequently, investors concentrate their trading within specific stock categories (i.e., habitats) and correlated trading activities of investor clienteles within their respective habitats can induce comovements in stock returns, especially when arbitrage costs are high.

In this study, we use the portfolio holdings and trading activities of both retail and institutional investors to provide *direct* evidence of habitat-based return comovements. We conjecture that investors' high propensity to hold local stocks and stocks with lottery features (low prices, high volatility, and high skewness) generate investor clienteles. Further, correlated trading within these investor habitats would induce strong return comovements among stocks within those habitats. We also posit that trading-induced return comovements would be amplified during periods of greater aggregate uncertainty due to an increase in the degree of correlated trading. Our focus is on geography- and price-based comovements which have been identified in the recent literature but we also consider other related forms of comovement.

Recent empirical studies provide evidence of different types of investor habitats. For example, Kumar and Lee (2006) show that low priced stocks have a higher concentration of retail investors. Kumar (2009b) shows that investors with certain socioeconomic characteristics are attracted toward stocks with lottery features (low prices, high volatility, and high skewness). Dorn and Huberman (2009) use German brokerage data to show that differences in risk aversion lead to the emergence of volatility-based investor clienteles. Investors with high levels of risk aversion prefer low volatility stocks, while less risk averse investors specialize in high volatility stocks. When an investor sells a stock and purchases another stock following the sale, she exhibits a greater propensity to buy a stock that has similar volatility level as the stock sold recently. Thus, investors trade in and out of stocks within their volatility-based habitats, which may generate correlations in the returns of stocks with similar volatility levels.

In other related economic settings, Graham and Kumar (2006) provide evidence of tax-induced or age-induced dividend clienteles. Battalio and Mendenhall (2005) demonstrate the existence of less sophisticated investor clientele that holds systematically biased earnings expectations. Blackburn, Goetzmann, and Ukhov (2007) show that distinct investor clienteles based on risk preferences exist and affect the pricing of stocks within their respective habitats. Similarly, in corporate settings, Hartzell and Starks (2003) and Frieder and Subrahmanyam (2006) document clientele effects in executive compensation.

Recent studies also provide evidence consistent with the habitat-based view of return comovements. For example, Kumar and Lee (2006) study the comovement patterns induced by the trading activities of retail investors. Using retail trades at a large U.S. discount brokerage house, they show that retail demand is significantly correlated across stocks, stock returns are sensitive to fluctuations in overall retail demand, and this sensitivity is stronger among stocks in which retail investors trade more, such as low priced stocks.[1] In contrast, Pirinsky and Wang (2004) and Sun (2008) examine the effects of institutional clienteles on return comovements. Further, Pirinsky and Wang (2006) use headquarter location changes to show that the returns of stocks within a certain geographical area comove more strongly, while Green and Hwang (2009) use stock splits to show that the comovement patterns are stronger among similar-priced stocks. Greenwood (2008) shows that stocks that are over-weighted in the Nikkei 225 index comove strongly with the stocks in the index and weakly with stocks outside of the index.

Our paper extends these recent empirical studies on investor habitats and habitat-induced return comovements along several dimensions. First, we show that the evidence of volatility-based

[1] Also, see Barber, Odean, and Zhu (2009b) for evidence of correlated trading among retail investors.

specialization presented in Dorn and Huberman (2009) generalizes to other stock attributes and to institutional investors. We also examine the potential asset pricing implications of volatility specialization and provide evidence of return comovements within volatility-based categories. Second, we extend the Kumar and Lee (2006) evidence to a longer time period and identify the mechanisms (i.e., local bias and gambling preferences) that induce comovements in stock returns. In addition, rather than focus on either retail or institutional clienteles, we compare the effects of retail and institutional investors on return comovements. Third, we provide direct empirical support for the conjectures in Pirinsky and Wang (2006) and Green and Hwang (2009). They posit that geography- and price-based comovements are induced by investors' preference for local stocks and stocks with certain price levels, respectively, but neither study directly links the stock preferences and trading activities of investors to comovement patterns in stock returns.

In the first part of the paper, we use the portfolio holdings and trades of both retail and institutional investors and show that investors specialize in local stocks and stocks with lottery features. Next, we show that the trading in and out of price- and geography-based stock categories generates return comovements among stocks within those categories. A part of the comovement patterns in local stocks are induced by the gambling behavior of investors. We find that stocks with lottery features move more strongly with other stocks in the geographical vicinity.

Collectively, our results provide *direct* evidence of geography- and price-based comovements induced by the local bias and gambling preferences of investors. Our evidence also indicates that the gambling behavior of investors is an important source of comovement patterns in stock returns. In broader terms, our results highlight the importance of a habitat-based framework of return comovements proposed in Barberis, Shleifer, and Wurgler (2005).

The rest of the paper is organized as follows. In the next section, we summarize our key hypotheses. We provide evidence of investor habitats and correlated trading in Section 3. In Section 4, we provide evidence of trading-induced comovements and in Section 5 we identify the mechanisms that induce those comovement patterns in stock returns. We conclude in Section 6 with a brief summary.

2. Related Research and Testable Hypotheses

Our main objective is to test the theoretical predictions of the habitat-based model of return comovements proposed in Barberis, Shleifer, and Wurgler (2005, hereafter BSW). The key prediction of the model is that investors concentrate their trading within specific stock categories (i.e., habitats) and correlated trading activities of investors within their respective habitats can induce comovements in stock returns, especially when arbitrage costs are high. In this section, we outline a set of testable hypotheses based on the broad theoretical predictions of the BSW model. We also develop several direct extensions of this model and examine whether gambling activities of investors and stock-specific or market-wide uncertainty influence trading-induced return comovements within investor habitats. The evidence from these empirical tests allows us to shed light on the mechanisms that lead to the emergence of investor habitats.

While the BSW model is very general and applies to different types of assets and different types of investor habitats, we focus on two types of stock return comovements which have been identified in the recent literature: geography- and price-based comovements. Pirinsky and Wang (2006) use headquarter location changes to show that the returns of stocks within a certain geographical area comove together, while Green and Hwang (2009) use stock splits to show that the returns of similar-priced stocks comove more strongly. We use portfolio holdings and trades of both retail and institutional investors and examine directly whether the evidence of geography- and price-based comovements are induced by the local bias and gambling preferences of retail and institutional investors.[2]

2.1 Retail and Institutional Clienteles

The BSW model assumes that investor habitats exist but it does not explicitly specify whether the comovement patterns would be stronger within retail or institutional habitat. In the first set of tests, we examine the relative impact of retail and institutional investors on return comovements. While the trading activities of both retail and institutional investors can potentially generate comovements among the returns of stocks within their respective habitats, we conjecture that the impact of retail investors is likely to be stronger for several reasons.

First, if behavioral biases induce correlated trading (e.g., Barber, Odean, and Zhu (2009b)) and retail investors exhibit stronger behavioral biases due to their lower sophistication level relative to institutions, the trades of retail investors would exhibit stronger correlations and are more likely to induce comovements. Second, stocks that attract retail investors have higher levels of uncertainty. Higher uncertainty levels would amplify investors' behavioral biases, generate stronger trading correlations, and consequently lead to higher return comovements. Third, stocks with greater retail concentration are likely to smaller and less visible firms with higher arbitrage costs. Even if the intensity of trading correlation among these stocks is not high, the

[2]As mentioned in the Introduction, previous studies have provided evidence consistent with the BSW model. However, there is still considerable debate about the source (risk or investor sentiment) of the documented patterns in return comovements. For a recent overview of this literature, see Kasch and Sarkar (2009).

impact of trading correlation on comovements is likely to be stronger for these stocks.

More formally, we conjecture that:

H1a: *Correlated retail trading:* Trades of retail investors would be more strongly correlated than those of institutions.

H1b: *Retail clienteles and comovements:* Price- and geography-based return comovements would be stronger among stocks that are dominated by retail investors.

If institutions attempt to exploit potential mispricing generated by retail investors, they might blunt the effects of retail trading on returns. In our empirical tests, we directly account for the potential effects of informed institutional trading on comovements.

2.2 Comovement Mechanisms: Local Preference and Gambling

To gather stronger empirical support for the BSW model, we identify the mechanisms generate price- and geography-based return comovements. Pirinsky and Wang (2006) conjecture that correlated local trading would induce comovements in returns. Specifically, the returns of stocks located within a certain geographical region would be more strongly correlated due to local stock preference of investors. However, the local stock preference of investors could be induced by either familiarity or superior information. Further, the type of local stock preference could vary across different investor types. For example, the local stock preference of institutions is more likely to be information-based while the local stock preference of relatively less sophisticated retail investors is more likely to be based on familiarity. If the local stock preference of investors is information driven, higher level of local investment need not induce stronger return correlations.

Because the relation between local stock preference and return comovements depend upon several factors, we use local stock preference measures of both retail and institutional investors to better quantify the local bias - return comovement relation. Specifically, we posit that:

H2a: *Retail local preference and comovements:* Local return comovements would be stronger among stocks with more local retail clienteles due to their familiarity-based local stock preference.

H2b: *Institutional local preference and comovements:* Local return comovements would be weaker among stocks with more local institutional clienteles due to their information-based local stock preference.

In addition to local stock preference, we conjecture that gambling behavior of investors is likely to be an important determinant of geography- and price-based return comovements. Investors who hold lottery-type stocks trade more actively (e.g., (Kumar 2009b)). Those investors are also likely to exhibit stronger trading correlations due to their lower levels of sophistication and stronger behavioral biases. Further, lottery-type stocks have higher arbitrage costs and are held primarily by retail investors. More active and more correlated trading by the investor clienteles of lottery-type stocks is likely to induce stronger correlations among their returns, especially because moderating arbitrage forces would be weak due to lower institutional interest and high arbitrage costs.

> **H3a:** *Gambling and comovements:* Return comovements would be strong among stocks with strong lottery features and are held by investors with a strong propensity to gamble.

In addition to its effect on lottery-type stocks, gambling-motivated activities is likely to influence local return comovements. Investors may be more inclined to gamble with stocks that they are more familiar with such as local stocks. Thus, we posit that:

> **H3b:** *Gambling and local comovements:* Local return comovements would be stronger among stocks with lottery features (i.e., low prices, high volatility, and high skewness).

2.3 Uncertainty and Comovements

In the next set of hypotheses, we investigate the relation between uncertainty and return comovements. We conjecture that stock-level and market-level uncertainty would amplify investors' behavioral biases, generate stronger trading correlations, and induce stronger return comovement patterns. This conjecture is motivated by recent theoretical behavioral finance models (Daniel, Hirshleifer, and Subrahmanyam (1998, 2001), Hirshleifer (2001)), which posit that investors' behavioral biases would be stronger when they trade hard-to-value stocks that operate in informationally sparse environments. Kumar (2009a) provides empirical support for this conjecture and also demonstrates that greater market-level uncertainty induces stronger behavioral biases. In particular, investors exhibit greater overconfidence, stronger disposition effect, and gravitate toward familiar domestic and local stocks during periods of greater market-wide uncertainty.

More formally, our first uncertainty-related conjecture is:

> **H4a:** *Uncertainty and trading correlations:* Retail trading correlations are stronger among stocks with higher uncertainty and also increase during periods of greater aggregate uncertainty. In contrast, uncertainty is less likely to influence institutional trading correlations.

Because retail trading correlations induce comovements in stock returns, we conjecture that:

H4b: *Stock-level uncertainty and comovements:* Return comovements would be stronger among stocks with higher valuation uncertainty.

Similarly, there would be an increase in the degree of return comovement during periods of higher aggregate uncertainty.

H4c: *Aggregate uncertainty and comovements:* Return comovements would increase during periods of uncertainty due to an increase in trading correlations.

To test these hypotheses, we use data from several different sources. We describe those data sets in the following sub-section.

2.4 Data and Summary Statistics

We obtain retail trading data from two different sources. First, we obtain the trades and monthly portfolio positions of a sample of 62,387 retail investors at a major U.S. discount brokerage house. The individual investors in the brokerage sample execute 26,000 trades in a typical month or 1,244 trades on a typical day. In a given year, these investors trade about 7,000 stocks, which suggests that their trades span a large set of stocks. The brokerage data also contain investors' demographic characteristics, including age, income, location (zip code), occupation, marital status, gender, etc. The demographic characteristics of investors in the brokerage sample are measured a few months after the end of the sample period (June 1997) and are provided by Infobase, Inc.[3]

The brokerage data covers a relatively short time period from 1991 to 1996. To analyze the impact of retail investor trading over a longer time period, we obtain retail trading data for the 1983 to 2000 time period from the Institute for the Study of Security Markets (ISSM) and the Trade and Quote (TAQ) databases. We use small-sized trades (trade size $\leq$ \$5,000) to proxy for retail trades. Like (Barber, Odean, and Zhu 2009a), we use the ISSM/TAQ data only until 2000 because the assumption that small trades proxy 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.[4]

To study the effects of institutional trading on return comovements, we use data on the quarterly common stock holdings of 13(f) institutions compiled by Thomson Reuters. The sample period is from 1980 to 2005. We infer institutional "trades" from changes in quarterly

[3]See Barber and Odean (2000) for additional details about the brokerage data.

[4]See Barber, Odean, and Zhu (2009a) or Hvidkjaer (2008) for additional details about the ISSM/TAQ data sets, including the procedure for identifying small trades.

holdings. Further, to distinguish between the trades of local and nonlocal institutions, we identify the institutional location (zip code) using the *Nelson's Directory of Investment Managers* and by searching the SEC documents and web sites of institutional managers.

Beyond the retail and institutional data sets, we obtain stock price, return and trading volume data from Center for Research on Security Prices (CRSP). We identify stock splits using CRSP distribution code 5523. We use Compustat to obtain firm headquarter locations, as well as leverage, market-to-book, and other firm characteristics. In some of our tests we use religious affiliation data obtained from the "Churches and Church Membership files from the American Religion Data Archive (ARDA). The county-level demographics data are from the U.S. Census Bureau. Last, we obtain the monthly Fama-French factor returns and monthly risk-free rates from Kenneth French's data library.[5]

Table 1 reports the summary statistics for the main variables used in the empirical analysis. These variables are briefly defined in Appendix Table A.1.

3. Investor Habitats and Correlated Trading

3.1 Retail and Institutional Habitats

Existence of investor habitats is one of the key elements of the BSW model of return comovements. In our first set of empirical tests, we examine whether investors categorize stocks based on their characteristics, which lead to the emergence of characteristic-based investor "habitats". Specifically, we use the Dorn and Huberman (2009) method and assess whether investors are attracted toward stocks with certain characteristics (e.g., volatility). The key intuition is that if investors focus their attention on certain stock attribute, they would concentrate their trading among stocks that are similar according to that attribute. For example, if investors categorize stocks based on their volatility levels and preferred volatility habitats exist, when they sell a stock followed by a purchase of another stock, they are more likely to buy a stock that has a similar volatility level to the stock that was recently sold.

To identify retail habitats, we construct sell-buy transition matrices corresponding to five stock characteristics: price, location, volatility, skewness, and lottery characteristics index. Motivated by the definition of lottery-type stocks in Kumar (2009b), we define a lottery characteristics index as the equal-weighted average of the standardized values of volatility, skewness, and negative of stock price.[6] We consider the sub-sample of investor-months in which an investor

[5]The data library is available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

[6]Kumar (2009b) defines lottery-type stocks as stocks that have low prices, high idiosyncratic volatility, and high idiosyncratic skewness.

both sells and buys stocks. To ensure that buying and selling of the same stock does not artificially influence our results, we use the net trade for the month in a given stock. For a given month, we compute the mean characteristic (e.g., price) of stocks sold and that of stocks bought by the same investor. At the end of each month or quarter, we then assign investors to a sell tercile and to a buy tercile according to the price rank of the stocks they sell and those they buy among all stocks sold and bought in that month.

If investors are insensitive to the price attribute, they would not exhibit a higher propensity to replace stocks in a certain price range with stocks of a similar price level and the assignments to buy and sell categories would be independent. But if investors treat stocks with similar price levels as a category, the diagonal elements in the sell-buy matrix would be larger than the off-diagonal elements. Therefore, the diagonal elements in the sell-buy transition matrices would serve as indicators of retail habitats. We identify the transition matrices for institutional investors using a similar procedure where we aggregate trades each quarter instead of each month.

Table 2 reports the retail and institutional sell-buy transition matrices for various stock characteristics. We find strong evidence of investor habitats among both retail and institutional investors. For instance, when retail investors sell a low priced stock, their subsequent stock purchase is from the low price category in 44.96% of cases, while they buy a high priced stock in only 24.24% of cases. Similarly, when institutions sell a low priced stock, they are three times more likely (61.38% vs. 17.37%) to buy another low priced stocks than a high priced stock. These estimates are consistent with the existence of price-based habitats. We find similar evidence of specialization within stock categories based on volatility, skewness, and location (investor-firm distance).

Collectively, these sell-buy transition matrices indicate that retail and institutional investors categorize stocks based on their salient attributes. These results extend the evidence of volatility habitats among German investors documented in Dorn and Huberman (2009). We provide evidence of similar volatility habitats among U.S. retail and institutional investors. In addition, we show that investors are likely to categorize stocks among other dimensions such as price and location. We also present new evidence of characteristic-based habitats among institutional investors.

3.2 Additional Evidence of Habitats

To better characterize the retail and institutional habitats, we obtain stock-level habitat measures, which allow us to examine the effects of multiple stock characteristics simultaneously. Specifically, using the retail and institutional trading data, we compute each stock's retail trading proportion (RTP) and institutional trading proportion (ITP) to measure the trading intensity of retail and institutional investors, respectively. RTP is defined as the ratio of the total month-t buy- and sell-initiated small trades (trade size below $5,000) dollar volume and the total stock trading dollar volume in the same month. We obtain the RTP measure for each stock at the end of each month and ITP is defined in an analogous manner using quarterly changes in institutional portfolio holdings (i.e., institutional trades).

Table 3, Panels A and B, report the mean RTP and ITP estimates for several stock characteristics. We find that the level of retail trading is high among stocks with low prices, high volatility, and high skewness. In contrast, institutional trading is concentrated more among stocks with high prices, low volatility, and low skewness. The lottery characteristics index sort summarizes these effects. These RTP and ITP estimates indicate that the trading activities of retail investors and institutions are concentrated in non-overlapping segments of the market, i.e., distinct retail and institutional habitats exist. If the trading activities of these investors generate comovements in returns, the comovement patterns are likely to be stronger within their respective habitats.

3.3 Trading Correlations Within Investor Habitats

The second essential element of the BSW model of return comovements is the presence of correlated trading. We measure the level of correlation in the trading activities of retail and institutional investors by estimating time series regressions of the following form:

$$BSI_{it} = \beta_0 + \beta_1 BSI_{pt} + \beta_2 RMRF_t + \varepsilon_{it}, \tag{1}$$

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

$$BSI_{it} = \frac{VB_{it} - VS_{it}}{VB_{it} + VS_{it}}, \tag{2}$$

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 buy-sell imbalance of all stocks that belong to portfolio p. The buy and sell trading volumes of a stock is estimated separately for retail and institutional investors. We estimate the time series regression for retail investors annually using monthly data and for institutions we estimate quarterly regressions using a three-year rolling window.

Portfolio p defines the investor habitat and the coefficient estimate β_1 captures the degree of correlated trading in stock i within that habitat. Specifically, β_1 measures the correlation between the trading activities of investors who hold stock i and the average trading activities of investors who hold other stocks within the habitat characterized by portfolio p. We consider several portfolios and obtain estimates of retail and institutional trading correlations within those investor habitats. Specifically, we consider the following portfolios: low price, high volatility, high skewness, and high lottery characteristics. The low price portfolio contains all stocks below the 30^{th} NYSE stock price percentile (excluding stock i). The volatility, skewness, and lottery characteristics index categories are defined in an analogous manner.

The trading correlation sorting results are reported in Table 3, Panels C and D. The evidence in Panel C indicate that retail trades in low priced stocks are more strongly correlated with trades in other stocks in the low price index. Similarly, retail trading activities in high volatility stocks are more strongly with other stocks in the high volatility category. The trading across skewness quintiles do not exhibit significant variation but the degree of correlated trading is significant across all categories. Institutional trades within a stock category are also positively correlated but there is not much variation across the characteristic-based quintiles.

We find similar results when we measure trading correlations using each stock's own category indices. These results are presented in Table 3, Panels E and F. We find that the retail trades of a low priced stock is positively correlated with retail trading in other low priced stocks (mean RTC = 0.263) and the retail trades of a high priced stock is positively correlated with retail trading in other high priced stocks (mean RTC = 0.159). The RTC differential of −0.104 is statistically significant and indicates that the own-category retail trading correlation is stronger within stock categories in which retail concentration is high. Examining the own-category trading correlation estimates for institutions (see Panel F), we find that the ITC estimates do not vary significantly across the characteristic quintiles.

Overall, our trading correlation estimates indicate that the trades of retail investors are more strongly correlated, especially within low price, high volatility, and high lottery characteristics index categories in which their concentration is high. This evidence supports our first clientele-related hypothesis (H1a).

3.4 Determinants of Investor Habitats and Correlated Trading

Our analysis so far treats each stock characteristic as an independent measure. However, stock characteristics such as stock price, firm size, volatility, and skewness are correlated. Thus, the similarities in the trading patterns across these stock characteristics are not very surprising. In this sub-section, we adopt a multivariate framework and estimate pooled regressions in which we can simultaneously examine the relative influences of various stock characteristics. In these regressions, the dependent variable is either a habitat (i.e., RTP or ITP) or a trading correlation (i.e., RTC or ITC) measure. The set of independent variables includes various stock and regional characteristics, including proxies for informed trading, stock-level uncertainty, local stock preference, and gambling behavior. All the variables are defined in Appendix Table A.1.

The pooled regression estimates are presented in Table 4. In Panel A, we report the results from the baseline regressions and in Panel B we consider an extended specification with proxies for local stock preference and gambling. Consistent with the evidence from the sorting results, we find that the level of retail trading is higher among smaller and lower priced stocks. In contrast, the institutional trading is more concentrated among larger firms. Examining the estimates of informed trading and uncertainty proxies, we find that stocks with high levels of informed trading (as captured by the PIN measure) have lower RTP and higher ITP. Retail concentration is also higher among stocks with higher uncertainty (as captured by the monthly turnover variable).

Further, examining the coefficient estimates of local stock preference and gambling proxies in Panel B, we find that stocks with lottery features have higher levels of retail trading. For example, the RTP level is high for stocks with high volatility or those which are located in regions in which local investors exhibit a greater propensity to invest in stocks with lottery features (i.e., higher CPRATIO). The retail trading intensity does not decrease with distance but the level of institutional trading increases as the average distance to the institutional clientele decreases. This evidence indicates that local institutions trade more actively, perhaps to exploit their local informational advantage.

The trading correlation regressions portray a similar picture. Consistent with the evidence in Kumar (2009a), we find that the retail trading correlation is high when the uncertainty levels are high. Both uncertainty proxies have significant coefficient estimates and the results are similar for both low price and local stock indices. The coefficient estimates of local stock preference and gambling proxies in Panel B indicate that stocks with lower prices and other lottery features have higher RTC. Even local retail trading correlations are stronger among small and low priced stocks. The ITC regression estimates indicate that institutional trading correlations are lower when institutions are located closer to the firm or when the degree of informed trading (i.e., PIN) is high.

Taken together, the habitat regression estimates indicate that RTP is higher for stocks with lottery features and higher uncertainty. In contrast, institutional trading is higher among stocks with more informed trading.[7] Further, the trading correlation regression estimates indicate that retail trades are more strongly correlated among stocks with lottery features and higher uncertainty. In contrast, the institutional trades are higher when the degree of informed trading

[7]When we consider local ITP as a dependent variable, the results are similar to the ITP regression estimates. To save space, we do not report those results.

is lower or firm-institution distance is higher. Thus, retail and institutional habitats appear distinct from each other and investors' trades within their respective habitats are more strongly correlated.

4. Trading-Induced Return Comovements

In this section, we use the retail and institutional habitat and trading correlation measures to investigate their impact on return comovements.

4.1 First Look: Portfolio Based Approach

In our first set of comovement tests, we use a portfolio-based approach to estimate the degree of return comovements in different investor habitats. Following the Kumar and Lee (2006) method, we estimate the following time series regression:

$$\begin{aligned} R_{pt} - R_{ft} = \alpha_p \quad &+ \quad \beta_{1p} RMRF_t + \beta_{2p} SMB_t + \beta_{3p} HML_t + \beta_{4p} UMD_t \\ &+ \quad \beta_{5p} BSI_{pt} + \varepsilon_{pt}; \quad t = 1, 2, \ldots, T. \end{aligned} \tag{3}$$

Here, R_{pt} is the month-t rate of return of portfolio p, R_{ft} is the riskfree rate of return, $RMRF_t$ is the market return in excess of the riskfree rate, SMB_t is the difference between the value-weighted return of a portfolio of small stocks and the value-weighted return of a portfolio of large stocks, HML_t is the difference between the value-weighted return of a portfolio of high B/M stocks and the value-weighted return of a portfolio of low B/M stocks, UMD_t is the difference between the value-weighted return of a portfolio of stocks with high returns during months $t-12$ to $t-2$ and the value-weighted return of a portfolio of stocks with low returns during months $t-12$ to $t-2$, BSI_{pt} is the buy-sell imbalance for portfolio p in month t, and ε_{pt} is the residual return of the portfolio. The coefficient estimate β_{5p} indicates the degree of return comovement induced by correlated retail trading.[8]

The return comovement estimates (i.e., β_{5p}) for different stock characteristic-sorted portfolios are presented in Table 5. The results indicate that return comovements are stronger within the habitat of retail investors. The β_{5p} estimate is statistically significant only for the low price, small-cap, high RTP, and low ITP portfolios. For example, when we consider price-sorted portfolios, the β_{5p} estimate is 0.114 (t-statistic = 4.76) for the lowest price quintile portfolio but only 0.002 (t-statistic = 0.52) for the highest price quintile portfolio, and the difference of -0.112 is statistically significant (t-statistic = -5.33).

[8] We are unable to conduct a similar exercise using institutional trading data due to its coarseness.

When we consider quintile portfolios based on various gambling or local stock preference proxies, we find that the return comovement estimate is positive and significant for portfolios that represent the retail habitat. Specifically, the β_{5p} estimate is significant for high volatility, high lottery characteristics index, and high distance to institutions portfolios. This evidence indicates that the effect of retail trading on return comovements is stronger when there are fewer local institutions. Examining the β_{5p} estimates for portfolios formed on uncertainty proxies, we find weak evidence that the degree of comovement is stronger among younger firms that have higher levels of uncertainty.

4.2 Investor Trading and Comovements: Sorting Results

To better understand how return comovement patterns vary across investor habitats, we define a stock-level comovement measure. We compute the comovement measure annually for each stock i by estimating the following time series regression:

$$R_{it} - R_{ft} = \beta_0 + \beta_1 CharIdx_{it} + \beta_2 RMRF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_{it} \qquad (4)$$

Here, $CharIdx$ is a return index relative to which the comovement is measured. For example, to measure return comovement relative to an index of low priced stocks, we define a low price index ($LowPrcIdx$), which is the portolio return of stocks priced below the 30th NYSE percentile of price at the end of the prior year (excluding stock i). β_1 from this regression is the comovement measure for stock i relative to the chosen return index. We estimate the beta annually using daily data where the return series used in the regression are standardized so that the comovement estimates relative to different return indices can be meaningfully compared.

In Table 6, we report the average stock-level comovement estimates for habitat- and trading correlation-sorted portfolios. When we measure the comovement relative to the low price index (see Panel A), the return comovement increases monotonically with RTP and RTC. In contrast, the return comovement declines with ITP and exhibits a weakly increasing relation with ITC.[9] These sorting results indicate that low price comovements are stronger within the retail habitat and when retail investors trade more in concert. The negative relation with ITP suggests that high levels of institutional trading might reflect informed trading, which is less likely to induce comovements in returns.

We also use the habitat and trading correlation sorts to investigate whether geography influences the comovement patterns in returns. In Panel B, we report the comovements relative to a local stock index. We find that local comovement is not related to RTP but it increases

[9] We find similar results when we sort stocks using institutional ownership levels instead of institutional trading proportions.

with RTC and exhibits a weak positive relation with ITC. Thus, like the other types of return comovements, higher levels of trading correlations induce greater local comovements.

We also find that local comovement exhibits a weak decreasing relation with ITP and local ITP. This evidence further indicates that the level of institutional trading is likely to be a proxy for informed trading, which is less likely to be correlated. The results from distance sorts are also consistent with the conjecture that local institutions are more likely to be informed. We find that the local return comovement is weaker when the average distance between a firm and its institutional shareholders is small. We do not observe this pattern with retail distance measure. In fact, the local comovement increases as the local retail participation proxy increases, which is consistent with the conjecture that the local stock preference of retail investors is more likely to be induced by familiarity.[10] These results are consistent with our local stock preference hypotheses (H2a and H2b).

To examine whether these comovement patterns are at least partially induced by investors' gambling activities, in Panels C, D, and E, we report the mean comovement estimates relative to other indices, which are constructed using stock characteristics that are more likely to be attractive to gamblers and speculators. Specifically, motivated by the evidence in (Kumar 2009b), we consider high volatility, high skewness, and lottery stocks return indices. We find that these comovement estimates are qualitatively similar to the results obtained using the low price index. This evidence is consistent with our first gambling hypothesis (H3a) and suggests that return comovements would be stronger among stocks in which gambling-motivated trading levels are likely to be higher.

For robustness, in Panel F, we sort stocks according to the retail and institutional habitat and trading correlation measures and report the comovement estimates relative to each stock's own price category. We consider low, medium, and high price categories, which are defined using NYSE price breakpoints. We find that comovement relative to the low price index increases with RTP but the comovement relative to the high price index exhibits an opposite pattern. In contrast, the comovement relative to the low price index decreases with ITP, while the comovement relative to the high price index increases. These results further indicate that return comovements within investor habitats are stronger when investors exhibit more correlated trading.

4.3 Comovement Regression Specification

In the next set of tests, we estimate several pooled regressions with various types of fixed effects to identify the main determinants of return comovements. Specifically, we investigate how the

[10]The local participation proxy is defined in Appendix Table A.1.

retail and institutional habitat and trading correlation measures influence return comovements when we account for other known determinants of return comovements.

Our main explanatory variables in the comovement regressions are the various habitat and trading correlation measures. Specifically, we consider RTP and ITP as indicators of retail and institutional habitats, while RTC and ITC proxies for the level of correlated trading among retail and institutional investors, respectively. Table 7 reports the correlation matrix for the habitat and trading correlation measures. We also define interactions between the habitat and trading correlation variables to test whether the effect of trading correlation on comovement is stronger when habitat effects are strong. Further, to test the uncertainty-based hypotheses, motivated by Jiang, Lee, and Zhang (2005) and Zhang (2006), we use firm age and monthly turnover as information uncertainty proxies to examine whether the comovement patterns are stronger among stocks with higher valuation uncertainty.

In addition to these key independent variables, we consider the lagged values of several firm and regional variables to account for the effects of firm and regional characteristics on return comovements. For example, smaller firms might exhibit stronger return correlations because they are more susceptible to sentiment changes. Similarly, the returns of firms located near an industry cluster could be more strongly correlated due to local information spillovers. Further, the pooled regressions include year and industry effects using the 48 Fama and French (1997) industries and the standard errors are clustered by firm in most specifications.

4.4 Price-Based Comovement Regression Estimates

In the first set of comovement regressions, we examine whether investor habitats and trading correlations influence price-based return comovements. The dependent variable in these regressions is the return comovement measured with respect to the low price index. The regression estimates are reported in Table 8, Panel A.

Consistent with the evidence from the sorting tests, we find that low price comovement is stronger among stocks with high levels of retail trading and low levels of institutional trading. The RTP estimate of 0.023 (t-statistic = 20.55) in specification (1) indicates that a one standard deviation change in RTP would be associated with an $0.023 \times 1.769 = 0.041$ increase in low price return comovement. Relative to the mean low price index comovement estimate of 0.071, this represents an economically significant 57.31% increase. Even when we account for other factors that influence comovements (see specification (6)), the RTP estimate is 0.019 (t-statistic = 16.76) and a one standard deviation change in RTP is associated with a 57.31% increase in low price return comovement.

The low price index comovement is also stronger when retail trades are more strongly correlated, especially when the level of retail trading is also high. The RTC variable as well as the RTP $\times$ RTC interaction term have positive and significant coefficient estimates. The RTC coefficient estimate of 0.006 (t-statistic = 12.44) implies that a one standard deviation change in RTC would correspond to a $0.006 \times 2.258 = 0.014$ shift in the low price return comovement. Relative to the mean low price index comovement, this represents a 19.08% shift. Further, although unconditionally high levels of institutional trading is associated with lower return comovements, high ITP coupled with correlated institutional trading induces stronger comovements in returns. The ITP $\times$ ITC interaction terms has a positive and significant coefficient estimate (estimate = 0.022, t-statistic = 2.06).

Examining the coefficient estimates of the uncertainty proxies, we find that younger firms and firms with high turnover exhibit higher levels of comovement relative to the low price index. In specification (6) where we account for other determinants of comovement, monthly turnover has a coefficient estimate of 0.052 (t-statistic = 16.76) and firm age has a coefficient estimate of -0.003 (t-statistic = -2.30). Relative to the mean comovement estimate of 0.071, the turnover and age coefficient estimates indicate that a one standard deviation change in turnover and age would correspond to a 10.18% and 3.97% change in comovement, respectively. Compared to the effects of habitat and trading correlation measures on comovements, the uncertainty proxies have weaker but still significant influence on comovements.

For robustness, we estimate the comovement regression using each stock's own price category comovement as the dependent variable. The estimates for low, medium, and high price sub-samples are reported in Panels B, C, and D, respectively. The results indicate that RTP has a positive coefficient estimate in the low priced stocks sub-sample, while ITP has a positive coefficient estimate in the high priced stocks sub-sample. Thus, trading activities of retail and institutional investors have a positive effect on return comovements within their own respective habitats.

Overall, the habitat and trading correlation coefficient estimates from comovement regressions indicate that habitat and clientele characteristics are important determinants of price-based return comovements. This evidence supports our main comovement hypothesis (H1b), the second uncertainty hypothesis (H4b), and provides direct support for the BSW model of return comovements.

4.5 Geography-Based Comovement Regression Estimates

In the second set of comovement regressions, we focus on geography-based return comovements. We examine whether location-based comovements are also influenced by investor habitats and trading correlations of investors, particularly those of local investors. The dependent variable in these regressions is the return comovement measured with respect to each stock's local return index. The regression estimates are reported in Table 9.

The regression estimates indicate that unlike price-based comovements, local return comovement is not influenced by the level of retail trading. The RTP variable has an insignificant coefficient estimate (estimate = 0.001, t-statistic = 1.24). This evidence is not very surprising because RTP captures the overall level of retail trading and the trading activities of local and non-local retail investors might have offsetting effects on local return comovements. The retail trading correlation, however, has a positive and significant coefficient estimate. In specification (5) where we account for other determinants of local return comovement, the local index RTC estimate of 0.003 (t-statistic = 6.33) implies that a one standard deviation change in RTC would correspond to a $0.003 \times 1.721 = 0.005$ shift in the local return comovement. Relative to the mean local return index comovement of 0.042, this represents a significant 12.29% increase.

Although our ISSM/TAQ data do not allow us to construct a local RTP measure, we examine the effect of local retail habitat using the brokerage data. For each stock, we obtain its distance to all retail stockholders in the brokerage sample and assume that this would serve as a proxy for the distance to all retail stockholders. When we use the distance to retail investors measure as an independent variable in the comovement regression specification, it has a significantly negative coefficient estimate. This evidence indicates that the local return comovement is stronger when a stock is held by more local retail investors. In contrast, the distance to institutional investors measure has a significantly positive coefficient estimate, which indicates that the local return comovement is stronger when there are fewer institutions in the neighborhood.

To better assess the effects of local retail habitats on local return comovements, we consider an alternative local stock market participation proxy that is constructed using the Census data. We estimate the local return comovement regression for low, medium, and high local participation sub-samples. The results are reported in columns (6) to (8) of Table 9. Consistent with our conjecture that investor habitats influence return comovements, we find that retail distance and trading correlation coefficients have stronger estimates with the appropriate sign for the high local retail participation sub-sample. For example, local index RTC has a coefficient estimate of 0.002 (t-statistic = 2.23) for the low local participation sub-sample and 0.004 (t-statistic = 5.02) for the high local participation sub-sample.

Examining the coefficient estimates of institutional habitat and trading correlation estimates, we find that like the price-based comovement regressions, the ITP coefficient estimate is significantly negatively related to local return comovement and is consistent with the conjecture that high levels of institutional trading reflects informed trading and negatively impacts local return comovements.[11] However, when institutions trade in concert (i.e., ITC is high) and their trades are not information-induced, institutional trading has a positive effect on local comovement.

We also examine the effect of stock-level uncertainty on local return comovements. Like the evidence from price-based comovement regressions, we find that stocks with higher uncertainty comove more strongly with other local stocks. Both uncertainty proxies have significant coefficient estimates. Specifically, in specification (5), monthly turnover has a coefficient estimate of 0.148 (t-statistic = 6.17). This estimate implies that a one standard deviation increase in market turnover would correspond to a $0.148 \times 0.139 = 0.021$ increase in the local return comovement. Relative to the mean local return index comovement estimate, this represents a significant 48.98% increase. Similarly, a one standard deviation reduction in firm age corresponds to a 11.15% increase in local return comovement. The incremental effects of both uncertainty proxies on local return comovement are stronger than their incremental effects on price-based comovement. Interestingly, the effect of stock-level uncertainty on local return comovements is significant only when the local stock market participation is high (see column (8)). This evidence suggests that uncertainty is likely to influence return comovements through its impact on the trading activities of local investors.

Taken together, the results from local return comovement regressions indicate that the local stock preference of retail rather than institutional investors induces stronger correlations in local stock returns. Using trading and location measures of retail and institutional investors, we provide direct support for the broad BSW conjecture. More specifically, the evidence is consistent with our main comovement hypothesis (H1b), the local stock preference hypotheses (H2a and H2b), and the uncertainty hypothesis (H4b).

4.6 Comovement Changes Around Stock Splits

For a sharper test of the impact of retail trading correlation on return comovement, we examine habitat shifts, trading correlation changes, and comovement patterns around stock splits. Stock splits represent a discrete drop in the price of a stock, which is followed by retail clientele shifts and an increase in retail trading (e.g., Dhar, Goetzmann, and Zhu (2004), Brav, Brandt, Graham, and Kumar (2009)). Further, Green and Hwang (2009) demonstrate that return comovement simultaneously increases with lower price stocks and decreases with higher price stocks around stock splits. Motivated by these previous studies, we conduct two tests using stock splits to directly establish the link between clientele changes around stock splits and the change in price-based return comovement.

[11]We find very similar results when we consider local ITP as an independent variable. Like the overall level of institutional trading, local institutional trading is negatively associated with local return comovements.

First, we test whether trading correlations among retail investors within price-based categories change around stock splits are consistent with the observed changes in return comovements. For each split event, we simultaneously estimate the retail trading correlation for each stock with respect to both high- and low-priced stock indices using the following regression:

$$\begin{aligned} BSI_{it} = \beta_0 \quad &+ \quad \beta_{high} PreSplitPrcBSIIdx_{it} + \beta_{low} PostSplitPrcBSIIdx_{it} \\ &+ \quad \beta_{mkt} MKTRF_t + \varepsilon_{it}. \end{aligned} \tag{5}$$

In this equation, BSI_{it} is the buy-sell imbalance of retail trades in stock i during month t, $PreSplitPrcBSIIdx_{it}$ is the equal-weighted average BSI of stocks in the price decile of i prior to the split, and $PostSplitPrcBSIIdx_{it}$ is the equal-weighted portolio BSI of the price decile to which stock i belongs after the split. The pre- and post-split portfolio BSIs exclude stock i itself. The market return is included in the regression to control for changes in BSI related to the overall movements of the market. β_{high} and β_{low} estimates measure the correlations between retail demand for stock i and retail demand for other stocks in its pre-split and post-split price ranges, respectively. We estimate β_{high} and β_{low} separately for k months prior to the split and for k months after the split ($k = 12, 18$, and 24).

The first two columns of Panel A of Table 10 report the mean changes in β_{high} and β_{low} around the split. We find that retail trading correlations shifts around stock splits are consistent with shifts in return comovements documented in Green and Hwang (2009). Retail trading in the stock that experiences a split exhibits a significant decrease in correlations with retail trading in higher-priced stocks and a significant increase in correlations with retail trading in lower-priced stocks.

As an alternative test that is more closely tied to the comovement regressions used in our paper, we estimate retail trading correlation relative to a portfolio of low-priced stocks using the following regression:

$$BSI_{it} = \beta_0 + \beta_{low} LowPrcBSIIdx_{it} + \beta_{mkt} MKTRF_t + \varepsilon_{it}. \tag{6}$$

Here, $LowPrcBSIIdx_{it}$ is the equal-weighted mean BSI of stocks priced below the 30^{th} NYSE percentile (excluding stock i). The last column of Panel A reports the mean change in β_{low} around stock splits. Consistent with the results reported in the first two columns, we find that retail trading correlation with low-priced stocks increases following a stock split.

Our last stock split test examines the influence of retail trading on price-based return comovements more directly. We estimate the sensitivity of stock returns to the BSI of pre-split and post-split price indices simultaneously using the following regression specification:

$$\begin{aligned} r_{it} - r_f = \beta_0 \quad &+ \quad \beta_{high} PreSplitPrcBSIIdx_{it} + \beta_{low} PostSplitPrcBSIIdx_{it} \\ &+ \quad \beta_{mkt} MKTRF_t + \beta_{smb} SMB_t + \beta_{hml} HML_t + \beta_{umd} UMD_t + \varepsilon_{it}. \end{aligned} \tag{7}$$

We estimate the regression separately using a k-month window before the split and after the split. These regressions are analogous to the time-series portfolio regressions reported in Table 5, except that we run stock-by-stock regressions around stock splits.

Panel B of Table 10 shows the mean changes in β_{high} and β_{low} around stock splits. We observe that β_{high} decreases significantly while β_{lw} increase following stock splits. This evidence indicates that the sensitivity of stock returns to the correlated retail demand in higher priced stocks (as captured by portfolio BSI) decreases while the return sensitivity to the retail demand in lower priced stocks simultaneously increases. Thus, consistent with our main comovement hypothesis (H1b), we provide fairly direct evidence the correlated retail demand influences return comovements around stock splits.

5. Comovement Generating Mechanisms

Our evidence so far shows that investor habitats and trading correlations are significant determinants of comovements in stock returns. In this section, we identify the mechanisms through which investors clienteles develop and their trades become correlated. Specifically, motivated by the evidence in recent behavioral finance literature, we examine whether local stock preference of investors, their preference for stocks with lottery features, and changes in aggregate uncertainty amplify the return comovement patterns.

5.1 Do Stocks with Lottery Features Exhibit Stronger Comovements?

In the first set of tests, we introduce gambling proxies in the comovement regression specification and examine to what extent price- and geography-based comovements are influenced by gambling-motivated trading. The results from extended regression specifications are reported in Table 11. In Panel A, we present the estimates for the low price comovement regressions and in Panel B we report the estimates from local return comovement regressions.

We find that both low price and local comovements are stronger for stocks with lottery features. After accounting for other factors that generate comovements, we find that the degree of comovement is higher for firms with lower prices, higher volatility levels, and higher skewness. This evidence is consistent with our conjecture that gambling-motivated trading induces greater comovement in returns because all three stock characteristics are known to attract investors who exhibit a greater propensity to gamble (e.g., Kumar (2009b)).

When we define a lottery stock dummy or a lottery characteristics index that captures the combined effect of these three stock characteristics, we find that they have positive and strongly significant coefficient estimates in both regression specifications. For instance, the lottery characteristics index has a coefficient estimate of 0.140 (t-statistic = 19.26) in the low price comovement regression, which indicates that a one standard deviation increase in the lottery characteristics index would correspond to a $0.140 \times 0.671 = 0.094$ increase in the low price comovement. In the local comovement regression, a one standard deviation increase in the lottery characteristics index would correspond to a $0.042 \times 0.671 = 0.028$ increase in comovement. Relative to the mean estimates of 0.071 and 0.042 for the low price and local return comovements, both comovement changes are economically significant.

To gather additional support for our gambling hypotheses, motivated by Kumar, Page, and Spalt (2009), we use the ratio of Catholics and Protestants adherents in a county (CPRATIO) as a proxy for people's propensity to gamble.[12] We find that both low price and local comovements are stronger for lottery-type stocks that are located in regions with high CPRATIO. This result provides additional support for the Kumar, Page, and Spalt (2009) conjecture that gambling propensity influences aggregate market outcomes. In a new economic setting, we show that gambling activities of investors influence return comovements.

Overall, our results from extended comovement regression specifications indicate that gambling activities of investors is an important source of comovements in returns. Even local return comovements are significantly influenced by people's propensity to gamble. These findings provide empirical support to both of our gambling hypotheses (H3a and H3b).

5.2 Volatility and Skewness Based Comovements

For robustness, we consider other types of return comovements that are more likely to be influenced by investors' gambling-motivated trading activities. Specifically, like the low price return comovement estimates, we obtain return comovement measures relative to high volatility return index, high skewness return index, and lottery-type stocks return index. Investors who exhibit a higher propensity to gambles find stocks with high volatility and high skewness more attractive (e.g., Kumar (2009b)) and, therefore, return comovement patterns in these stocks are more likely to reflect gambling-motivated trading.

The high volatility, high skewness, and lottery-type stocks comovement regression estimates are reported in Table 12, Panel B. For comparison, in the first column, we report the coefficient estimates from the low price comovement regression estimated in Table 7. In addition, Panel A of Table 12 reports the correlation matrix for the various comovement measures. As expected,

[12]This identification strategy is motivated by the observation that gambling attitudes are strongly determined by one's religious background. In particular, the Roman Catholic church maintains a tolerant attitude towards moderate levels of gambling and is less disapproving of gambling activities, while a strong moral opposition to gambling and lotteries has been an integral part of the Protestant movement since its inception.

the high volatility and high skewness comovement measures are strongly correlated with the low price comovement measure and, therefore, we also define residual comovement measures by including the low price comovement as an independent variable in the comovement regressions.

The comovement regression estimates indicate that returns are more strongly correlated when the level of retail trading is high, retail trades are more correlated, and stock-level uncertainty is high. In contrast, consistent with the informed trading hypothesis, high levels of institutional trading are associated with lower return comovements. These comovement results with high volatility, high skewness, and lottery-type stocks index are similar to our previous evidence from low price comovement regressions. Overall, these results suggest that gambling activities of investors is likely to be a significant source of comovements in the returns of stocks dominated by retail investors.

5.3 Do Stocks Held by "Gamblers" Exhibit Stronger Comovements?

In the next set of tests, we relate clientele characteristics to return comovements directly to further strengthen the link between gambling and return comovements. Again, we follow an indirect approach to examine the gambling-comovement relation because the gambling activities of investors are not directly observable. Our conjecture is that the levels of retail trading correlations and return comovements would be higher among stocks that are held by investors who are more likely to engage in speculative or gambling-motivated trading (e.g., younger, male, less educated, and low-income investors), as documented in Kumar (2009b).[13]

We test this conjecture using the U.S. discount brokerage data. For each stock, first, we measure the average demographic characteristics of brokerage investors who trade the stock during the brokerage sample period. Then, we estimate cross-sectional regressions in which either the sample period average return comovement measure or the retail trading correlation measure is the dependent variable and the average clientele characteristics of the stock are the independent variables. The cross-sectional regression estimates are reported in Table 13. In the first five columns, the dependent variable is a return comovement measure. In the next two columns, the dependent variable is a retail trading correlation measure. In the last column, the dependent variable is the retail trading proportion measure. These estimates have been previously reported in Han and Kumar (2009) and we reproduce them here for completeness.

Our results indicate that the clientele characteristics of stocks with high return comovement levels are similar to the characteristics of investors who are more attracted toward lottery-type stocks (Kumar (2009b)). For example, the return comovement levels are high for stocks that

[13]Investors with these characteristics are less likely to own stocks (e.g., Campbell (2006)). We conjecture that, conditional upon stock market participation, those investors are more likely to hold stocks with lottery features.

are held by male, young, and single investors with lower income, lower education levels, and non-professional occupations. The retail clientele of stocks with high return comovement also hold more concentrated portfolios and live in urban areas and regions with higher per-capita consumption of state lotteries. The high degree of similarity in the retail clienteles of high return comovement and lottery-type stocks suggests that gambling is likely to be an important source of comovements in stock returns.

To examine whether the trades of investors with higher propensity to gamble are more strongly correlated, we estimate two additional cross-sectional regressions in which the dependent variable is either the local stocks RTC or lottery stocks RTC.[14] The estimates indicate that, like the evidence from the comovement cross-sectional regressions reported in columns (1)-(5), the clientele characteristics of stocks with high RTC are similar to the characteristics of investors who are more attracted toward lottery-type stocks. This evidence suggests that the trades of investors with higher propensity to gamble are more strongly correlated, which in turn induce stronger return correlations.

5.4 Uncertainty, Trading Correlations, and Return Comovements

In our last set of tests, we examine the impact of aggregate market-level uncertainty and sentiment on return comovements and trading correlations. Following Kumar (2009a) and Lemmon and Portnaiguina (2006), we use the Chicago Board of Options Exchange volatility index (VIX) as a proxy for aggregate market-level uncertainty and the Michigan consumer sentiment index as a proxy for aggregate market sentiment.[15]

To begin, in Figure 1, we show the mean return comovements time series. It is evident from the plot that the comovement levels are high during periods of greater market-level uncertainty (e.g., in 1987) or when the market sentiment levels are high (e.g., during the late nineties and early 2000s). Next, in Table 14, we report the mean comovement and trading correlation estimates for sentiment sorted periods. In Panels A, B, and C, we consider VIX-sorted time periods, while in Panels D, E, and F, we consider Michigan sentiment index sorted time periods. The sorting results indicate that return comovements as well as retail trading correlations are higher during periods of greater market-level uncertainty and higher market sentiment. In contrast, neither the level of institutional trading nor the degree of institutional trading correlations are significantly affected by time variation in aggregate uncertainty or market sentiment. Although

[14]The results are qualitatively similar when we consider RTC measured relative to the low price, high volatility, or high skewness indices. For brevity, we do not report those results.

[15]For robustness, we also use the Baker and Wurgler (2006) sentiment index and find similar results. Also, see the evidence in Figure 2.

the VIX-sorted results are stronger, both the uncertainty and sentiment proxies yield qualitatively similar results. These results provide support to our uncertainty hypotheses H4a and H4c.

To ensure that the uncertainty-comovement and sentiment-comovement relations hold when we account for other determinants of return comovements, we estimate additional comovement regressions. We add the uncertainty and sentiment measures to the main comovement regression specification estimated earlier (see Section 4). Specifically, we regress the return comovement measures (low price beta, local beta, high volatility beta, high skewness beta, and lottery-type stocks beta) on measures of investor habitats, trading correlations, uncertainty or sentiment, and additional firm and region characteristics used earlier. Because the VIX and sentiment measures are the same for every firm in a given year, we modify our regression specification. Instead of using year and industry effects, which allows for an identification of the determinants of return comovements from the cross-section, we employ firm and industry effects. This change in the specification allows us to identify the determinants of return comovements comes from the time-series variations within each stock. To account for the lack of dependence within a given year, we cluster the standard errors by year.

In addition to allowing us to identify the effect of uncertainty and sentiment on return comovements, the alternative comovement regression specification provides a robustness check. It allows us to examine whether our main habitat and trading correlation measures also predict time-series variation in return comovements.

Table 15, Panel A reports regressions that include average annual level of the VIX index as an additional explanatory variable. We find that the signs and significance of RTP, ITP, RTC and ITC are similar in the time-series as in the cross-section. Further, the coefficient on VIX is strongly positive for all types of return comovement, except for comovement with high volatility stocks, for which the coefficient estimate is only weakly significant. Similarly, Panel B reports the comovement regression estimates in which the average annual level of the Michigan Sentiment index is included as an additional explanatory variable. The sentiment variable also has a relatively strong positive effect on low price and local return comovements, but its effect is weaker on high volatility, high skewness, and lottery-type stocks return comovements. Overall, the results from the alternative comovement regression specifications are consistent with our conjecture that trading correlations and, consequently, return comovements, increase during times of greater aggregate uncertainty and higher market sentiment.

For robustness, we also estimate the low price and local return comovement regressions separately for low, medium, and high sentiment and uncertainty sub-periods. For brevity, we only report the RTC coefficient estimates in Figure 2. We find that for both low price and local return comovement regressions, the RTC coefficient estimate is higher when aggregate market-level uncertainty or market sentiment is high. This evidence is consistent with our third uncertainty hypothesis (H3c), which posits that return comovements are stronger during periods of higher uncertainty or higher sentiment because of an increase in the degree of trading correlations.[16]

Overall, in this section, using investor-level, stock-level, and regional gambling proxies, we show that gambling-motivated trading are important drivers of return comovements. We also show that aggregate market-level uncertainty amplifies the comovement patterns in stock returns. These results provide strong support to our gambling and uncertainty hypotheses.

6. Summary and Conclusion

In this study, we use a large database of portfolio holdings and trades of both retail and institutional investors and provide *direct* empirical support for the habitat-based return comovements model proposed in Barberis, Shleifer, and Wurgler (2005). We focus on geography- and price-based return comovements but also consider other related forms of comovement. We show that correlated trading activities of retail and institutional investors generate return comovements within geography- and price-based stock categories. The comovement patterns are stronger among stocks with lottery features, high retail concentration, and high arbitrage costs. Across time, comovements get amplified during periods of greater aggregate uncertainty and stronger consumer sentiment.

Overall, our evidence contributes to the growing literature in behavioral finance, which demonstrates that retail investors play an important role the return generating process. Our results also indicate that local stock preference and gambling behavior of investors is an important source of comovement patterns in stock returns. In broader terms, our results highlight the usefulness of a habitat-based approach for studying asset prices.

[16]In unreported results, we find that unlike RTC, the ITC coefficient estimate is insignificant in all three sentiment and uncertainty sub-periods.

References

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

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.

———, and Ning Zhu, 2009a, Do retail trades move markets?, *Review of Financial Studies* 22, 151–186.

———, 2009b, Systematic noise, *Journal of Financial Markets* Forthcoming.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283–317.

Battalio, Robert H., and Richard R. Mendenhall, 2005, Earnings expectations, investor trade size, and anomalous returns around earnings announcements, *Journal of Financial Economics* 77, 289–319.

Blackburn, Douglas W., William N. Goetzmann, and Andrey Ukhov, 2007, Risk aversion and clientele effects, Working Paper (February), Yale School of Management; Available at SSRN: http://ssrn.com/abstract=933119.

Brav, Alon, Michael Brandt, John Graham, and Alok Kumar, 2009, The idiosyncratic volatility puzzle: Time trend or speculative episodes?, *Review of Financial Studies* Forthcoming.

Campbell, John Y., 2006, Household finance, *Journal of Finance* 61, 1553–1604.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 1998, A theory of overconfidence, self-attribution, and security market under- and over-reactions, *Journal of Finance* 53, 1835–1885.

———, 2001, Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* 56, 921–965.

Dhar, Ravi, William N. Goetzmann, and Ning Zhu, 2004, The impact of clientele changes: Evidence from stock splits, Yale ICF Working Paper No. 03-14.

Dorn, Daniel, and Gur Huberman, 2009, Preferred risk habitat of individual investors, *Journal of Financial Economics* Forthcoming.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is information risk a determinant of asset returns?, *Journal of Finance* 57, 2185–2221.

Fama, Eugene F., and Kenneth R. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153–193.

Frieder, Laura, and Avanidhar Subrahmanyam, 2006, Executive compensation and investor clientele, Working Paper (October), University of California at Los Angeles.

Graham, John R., and Alok Kumar, 2006, Do dividend clienteles exist? Evidence on dividend preferences of retail investors, *Journal of Finance* 61, 1305–1336.

Green, T. Clifton, and Byoung-Hyoun Hwang, 2009, Price-based return comovement, *Journal of Financial Economics* Forthcoming.

Greenwood, Robin, 2008, Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights, *Review of Financial Studies* 21, 1153–1186.

Han, Bing, and Alok Kumar, 2009, Speculation, realization utility, and volatility-induced retail habitat, Working Paper (May), University of Texas at Austin.

Hartzell, Jay, and Laura Starks, 2003, Institutional investors and executive compensation, *Journal of Finance* 58, 2351–2374.

Hirshleifer, David A., 2001, Investor psychology and asset prices, *Journal of Finance* 56, 1533–1597.

Hvidkjaer, Soeren, 2008, Small trades and the cross-section of stock returns, *Review of Financial Studies* 21, 1123–1151.

Jiang, Guohua, Charles M.C. Lee, and Yi Zhang, 2005, Information uncertainty and expected returns, *Review of Accounting Studies* 10, 185–221.

Kasch, Maria, and Asani Sarkar, 2009, Are comovements excessive?, Working Paper (March), University of Bonn; Available at SSRN: http://ssrn.com/abstract=1364539.

Kumar, Alok, 2009a, Hard-to-value stocks, behavioral biases, and informed trading, *Journal of Financial and Quantitative Analysis* Forthcoming.

———, 2009b, Who gambles in the stock market?, *Journal of Finance* 64, 1889–1933.

———, and Charles M.C. Lee, 2006, Retail investor sentiment and return comovements, *Journal of Finance* 61, 2451–2486.

Kumar, Alok, Jeremy Page, and Oliver Spalt, 2009, Religious beliefs, gambling attitudes, and financial market outcomes, Working Paper (March), University of Texas at Austin.

Lemmon, Michael L., and Evgenia Portnaiguina, 2006, Consumer confidence and asset prices: Some empirical evidence, *Review of Financial Studies* 19, 1499–1529.

Pirinsky, Christo, and Qinghai Wang, 2004, Institutional investors and the comovement of equity prices, Working Paper (October), Georgia Institute of Technology; Available at SSRN: http://ssrn.com/abstract=585884.

———, 2006, Does corporate headquarters location matter for stock returns?, *Journal of Finance* 61, 1991–2015.

Sun, Zheng, 2008, Clustered institutional holdings and stock comovement, Working Paper (April), University of California at Irvine; Available at SSRN: http://ssrn.com/abstract=1332201.

Zhang, X. Frank, 2006, Information uncertainty and stock returns, *Journal of Finance* 61, 105–136.

TABLE 1

Summary Statistics

This table reports summary statistics for all variables used in the empirical analysis. All variables have been defined in Appendix Table A.1. The sample period is from 1980 to 2005, or from 1983 to 2000 for the retail trading proportion (RTP) and the retail trading correlation (RTC) measures.

Measure	Mean	Std Dev	10^{th}Pctl	25^{th}Pctl	Median	75^{th}Pctl	90^{th}Pctl	N
Return Comovement Measures								
Low Price Index Beta	0.071	0.235	−0.206	−0.072	0.066	0.208	0.354	78,389
Local Stocks Beta	0.042	0.149	−0.109	−0.043	0.026	0.103	0.201	78,389
High Volatility Index Beta	0.055	0.210	−0.191	−0.070	0.049	0.174	0.309	78,389
High Skewness Index Beta	0.101	0.334	−0.302	−0.107	0.098	0.305	0.512	78,389
Lottery Stocks Index Beta	0.076	0.236	−0.207	−0.070	0.070	0.217	0.366	78,389
Own Price Category Beta	0.162	0.379	−0.238	−0.034	0.142	0.340	0.602	78,389
Own Volatility Category Beta	0.055	0.210	−0.191	−0.070	0.049	0.174	0.309	78,389
Own Skewness Category Beta	0.101	0.334	−0.302	−0.107	0.098	0.305	0.512	78,389
Own Lottery Category Beta	0.132	0.395	−0.325	−0.090	0.124	0.345	0.592	78,389
Retail Trading Correlation (RTC) Measures								
Low Price Index RTC	0.740	2.258	−1.745	−0.408	0.740	1.911	3.236	41,281
Local Stocks RTC	0.427	1.721	−1.337	−0.383	0.339	1.229	2.337	41,221
High Volatility Index RTC	0.160	0.413	−0.410	−0.130	0.188	0.471	0.686	41,082
High Skewness Index RTC	0.163	0.414	−0.399	−0.119	0.189	0.465	0.681	41,069
Lottery Stocks Index RTC	0.158	0.567	−0.395	−0.123	0.182	0.455	0.666	41,057
Own Price Category RTC	0.184	0.400	−0.364	−0.093	0.214	0.483	0.688	41,082
Own Volatility Category RTC	0.181	0.406	−0.373	−0.098	0.208	0.484	0.693	41,093
Own Skewness Category RTC	0.168	0.409	−0.393	−0.114	0.189	0.471	0.687	41,040
Own Lottery Category RTC	0.166	0.395	−0.371	−0.105	0.185	0.458	0.666	41,100
Institutional Trading Correlation (ITC) Measures								
Low Price Index ITC	0.127	1.673	−0.364	−0.108	0.159	0.404	0.606	67,275
Local Stocks ITC	0.097	1.586	−0.386	−0.150	0.112	0.359	0.561	67,230
High Volatility Index ITC	0.132	1.917	−0.367	−0.108	0.164	0.408	0.613	67,277
High Skewness Index ITC	0.132	0.770	−0.377	−0.119	0.151	0.393	0.598	67,275
Lottery Stocks Index ITC	0.105	0.591	−0.391	−0.149	0.114	0.364	0.566	67,290
Own Price Category ITC	0.148	1.636	−0.350	−0.090	0.178	0.424	0.628	67,252
Own Volatility Category ITC	0.142	1.960	−0.354	−0.097	0.177	0.421	0.627	67,278
Own Skewness Category ITC	0.151	1.183	−0.357	−0.099	0.169	0.414	0.619	67,275
Own Lottery Category ITC	0.154	1.721	−0.353	−0.095	0.177	0.422	0.627	67,279

TABLE 1 (Continued)

Summary Statistics

Measure	Mean	Std Dev	10^{th}Pctl	25^{th}Pctl	Median	75^{th}Pctl	90^{th}Pctl	N
Clientele Characteristics								
ln(RTP)	−0.100	1.769	−2.316	−1.374	−0.132	1.169	2.193	41, 515
ln(1+IO)	0.253	0.199	0.006	0.067	0.226	0.416	0.544	78,389
ln(1+ITP)	0.244	0.209	0.002	0.059	0.210	0.383	0.536	78,389
ln(1+Local IO)	0.049	0.080	0	0	0.011	0.060	0.162	78,389
ln(1+Local ITP)	0.048	0.093	0	0	0.008	0.050	0.159	78,389
EW Distance to Retail	1,019	519	357	713	1,001	1,283	1,647	54,903
EW Distance to Institutions	1,113	502	608	753	1,007	1,404	1,864	73,835
PIN	0.217	0.093	0.115	0.157	0.206	0.263	0.325	53,294
Stock Characteristics								
Stock Price	$16.39	$21.75	$1.37	$3.75	$10.25	$23.10	$38.19	78,389
Volatility	0.557	0.281	0.248	0.344	0.500	0.707	0.959	78,389
Skewness	0.371	0.541	−0.125	0.136	0.339	0.578	0.880	78,389
Lottery Stock Dummy	0.240	0.427	0	0	0	0	1	78,389
Lottery Characteristics Index	0.000	0.671	−0.893	−0.409	0.107	0.486	0.741	78,389
ln(Firm Size)	18.396	2.189	15.625	16.798	18.270	19.914	21.360	78,331
Market-To-Book (M/B)	2.824	3.825	0.634	0.986	1.648	3.007	5.691	78,389
Leverage	0.223	0.190	0	0.044	0.200	0.353	0.484	78,148
3-Year R&D Expenditure	0.044	0.081	0	0	0	0.053	0.142	72,025
3-Year Advertising Expenditure	0.014	0.031	0	0	0	0.012	0.044	72,025
3-Year ROA	−0.017	0.172	−0.197	−0.029	0.032	0.069	0.108	72,025
Dividend Yield	0.018	0.213	0	0	0	0.014	0.040	78,389
Past 12-Month Return	0.188	0.944	−0.517	−0.255	0.046	0.384	0.889	78,389
Monthly Turnover	0.099	0.139	0.015	0.029	0.059	0.117	0.223	76,718
ln(Firm Age)	4.76	0.937	3.466	4.043	4.796	5.455	5.973	78,389
Regional Characteristics								
ln(Number of Firms in MSA)	4.285	1.280	2.398	3.367	4.615	5.142	6.014	78,389
Social Capital Index	1.344	0.589	0.622	0.920	1.311	1.752	2.074	67,929
Urban Dummy	0.952	0.072	0.875	0.943	0.978	0.995	1	78,389
Industry Cluster Dummy	0.707	0.455	0	0	1	1	1	78,315
County Catholic-Protestant Ratio	2.152	1.749	0.276	0.676	1.743	3.154	4.845	78,389
Local Market Participation Index	0.502	0.126	0.326	0.411	0.505	0.611	0.653	78,389

TABLE 2

Stock Characteristic Based Specialization Estimates

This table reports the transition matrices among stock categories in trading by retail (Panel A) and institutional (panel B) investors. The sample includes all trades in which a household (institution) sold stock(s) and purchased others within a month (quarter). Each month (quarter), the net trades in each stock are aggregated into household- (institution-) level "buy" and "sell" portfolios. The portfolios are sorted into terciles based on the average price, volatility, etc. of the portfolio. For each sell category, the tables present the fraction of trades into each of the buy categories. We consider stock categories based on price, volatility, skewness, lottery characteristics, and distance between the investor and the firm headquarters. For the lottery-type stocks, "Low" refers to stocks with above-median price and below-median volatility and skewness, "High" refers to stocks with below-median price and above-median volatility and skewness, and "Medium" refers to all other stocks. The sample periods are 1991-1996 for the retail trades and 1980-2008 for the institutional trades.

Panel A: Price Specialization						
	Retail			Institutions		
Tercile	Buy:Low	Buy:Medium	Buy:High	Buy:Low	Buy:Medium	Buy:High
Sell:Low	44.96%	30.80%	24.24%	61.38%	21.25%	17.37%
Sell:Medium	28.64%	38.00%	33.36%	28.25%	38.50%	33.25%
Sell:High	20.46%	33.14%	46.40%	23.20%	38.69%	38.11%
Panel B: Specialization in Local Stocks						
Sell:Low	39.45%	31.31%	29.24%	47.07%	38.12%	14.81%
Sell:Medium	24.50%	47.62%	27.88%	28.10%	49.90%	22.00%
Sell:High	24.15%	29.42%	46.43%	12.73%	23.36%	63.91%
Panel C: Volatility Specialization						
Sell:Low	40.74%	30.06%	24.20%	47.26%	36.99%	15.75%
Sell:Medium	25.05%	37.95%	37.00%	24.25%	44.38%	31.37%
Sell:High	15.37%	31.99%	52.64%	9.95%	21.91%	68.14%
Panel D: Skewness Specialization						
Sell:Low	39.43%	35.20%	25.38%	46.45%	35.79%	17.76%
Sell:Medium	27.09%	37.14%	35.77%	26.41%	43.10%	30.48%
Sell:High	18.39%	33.10%	48.51%	13.70%	24.73%	61.57%
Panel E: Specialization in Lottery-Type Stocks						
Sell:Low	35.94%	35.42%	28.64%	40.14%	38.20%	21.66%
Sell:Medium	31.81%	35.68%	32.51%	29.87%	41.22%	28.91%
Sell:High	27.34%	33.60%	39.06%	17.86%	24.93%	57.21%

TABLE 3

Mean Clientele Characteristics and Trading Correlations Across Stock Categories

This table reports the clientele and trading correlation properties of stock portfolios sorted on various characteristics. Panel A reports the mean retail trading proportion (in logs) in each portfolio formed by sorting stocks into quintiles based on price, volatility, skewness, and lottery characteristics. Panel B reports the mean institutional trading proportion (in logs) for portfolios sorted on the same characteristics. Similarly Panels C, D, E and F report portfolio means of retail trading correlations (with the low price, high volatility, high skewness, and lottery stock portfolios), institutional trading correlations, retail trading correlations with other stocks in each stock's own price, volatility, skewness or lottery category, and institutional own-category trading correlation, respectively. All variables have been defined in Appendix Table A.1. The sample period is from 1983-2000 for the retail measures, and 1980-2005 for the institutional measures.

Panel A: Mean Retail Clientele Measure (ln(1+RTP))

	Stock Characteristic Quintiles					
Stock Characteristic	Low	Q2	Q3	Q4	High	High−Low
Price	1.757	1.032	0.072	−0.828	−1.903	−3.660***
Volatility	−1.344	−0.923	−0.260	0.486	1.446	2.790***
Skewness	−0.873	0.020	0.131	−0.062	−0.207	0.666***
Lottery Characteristics Index	−1.694	−0.823	−0.016	0.749	1.442	3.136***

Panel B: Mean Institutional Clientele Measure (ln(1+ITP))

	Stock Characteristic Quintiles					
Stock Characteristic	Low	Q2	Q3	Q4	High	High−Low
Price	0.080	0.160	0.255	0.358	0.420	0.340***
Volatility	0.415	0.359	0.243	0.162	0.098	−0.317***
Skewness	0.334	0.243	0.215	0.229	0.255	−0.079***
Lottery Characteristics Index	0.422	0.346	0.245	0.160	0.104	−0.318***

Continued . . .

TABLE 3 (Continued)

Mean Clientele Characteristics and Trading Correlations Across Stock Categories

Panel C: Mean High/Low Category Retail Trading Correlations

	Stock Characteristic Quintiles					
Stock Index	Low	Q2	Q3	Q4	High	High−Low
Low Price	1.239	1.068	0.836	0.471	0.332	−0.907***
High Volatility	0.072	0.125	0.177	0.221	0.246	0.174***
High Skewness	0.153	0.186	0.182	0.176	0.153	0.000
High Lottery Characteristics Index	0.096	0.118	0.177	0.214	0.242	0.146***

Panel D: Mean High/Low Category Institutional Trading Correlations

	Stock Characteristic Quintiles					
Stock Index	Low	Q2	Q3	Q4	High	High−Low
Low Price	0.088	0.155	0.143	0.111	0.092	0.004
High Volatility	0.125	0.121	0.090	0.132	0.147	0.022
High Skewness	0.110	0.127	0.133	0.142	0.129	0.020*
High Lottery Characteristics Index	0.100	0.078	0.075	0.088	0.086	−0.014

Panel E: Mean Own Category Retail Trading Correlations

	Stock Characteristic Quintiles					
Stock Characteristic	Low	Q2	Q3	Q4	High	High−Low
Price	0.263	0.241	0.185	0.150	0.159	−0.104***
Volatility	0.142	0.149	0.180	0.222	0.246	0.102***
Skewness	0.175	0.187	0.182	0.177	0.152	−0.023*
Lottery Characteristics Index	0.143	0.144	0.167	0.198	0.223	0.080***

Panel F: Mean Own Category Institutional Trading Correlations

	Stock Characteristic Quintiles					
Stock Characteristic	Low	Q2	Q3	Q4	High	High−Low
Price	0.088	0.155	0.143	0.140	0.156	0.068
Volatility	0.166	0.129	0.087	0.132	0.147	−0.019
Skewness	0.152	0.139	0.145	0.153	0.140	−0.012
Lottery Characteristics Index	0.135	0.132	0.167	0.130	0.136	0.001

TABLE 4

Investor Habitats, Trading Correlations, and Informed Trading: Pooled Regression Estimates

This table reports coefficient estimates from regressions of investor clientele and trading correlation measures on measures of informed trading and uncertainty, firm characteristics, and regional characteristics of the firm's headquarters location. The dependent variables are log retail trading proportion, log institutional trading proportion, low price retail trading correlation, low price institutional trading correlation, local retail trading correlation, and local institutional trading correlation. All variables have been defined in Appendix Table A.1. The sample period is from 1983-2000 for the retail measures, and 1983-2005 for the institutional measures. In addition to the basic regressors considered in Panel A, Panel B includes explanatory variables related to gambling and local bias. The regressions include year and industry effects using the Fama and French (1997) 48 industry definitions, and the standard errors are clustered by firm. The *t*-statistics for the coefficient estimates are reported in the parentheses below the estimates.

Panel A: Basic Regression Specification

Variable	ln RTP	ln (1 + ITP)	Low Pr RTC	Low Pr ITC	Local RTC	Local ITC
Informed Trading and Uncertainty Proxies						
PIN	−1.301	0.306	0.110	−0.122	0.261	−0.072
	(−9.22)	(13.02)	(0.25)	(−3.16)	(0.72)	(−1.41)
Monthly Turnover	0.836	−0.309	0.612	−0.083	0.376	−0.094
	(6.760)	(−11.35)	(4.41)	(−3.64)	(3.55)	(−2.61)
ln(Firm Age)	0.004	0.001	−0.061	0.015	−0.047	0.005
	(0.32)	(0.37)	(−3.26)	(3.97)	(−3.06)	(0.17)

Continued . . .

TABLE 4 (Continued)

Investor Habitats, Trading Correlations, and Informed Trading: Pooled Regression Estimates

Variable	ln RTP	ln (1 + ITP)	Low Pr RTC	Low Pr ITC	Local RTC	Local ITC
Firm Characteristics						
ln(Firm Size)	−0.662	0.056	−0.146	−0.014	−0.050	−0.003
	(−49.47)	(49.38)	(−15.84)	(−5.97)	(−6.25)	(−0.73)
Market-To-Book	0.024	−0.006	−0.007	−0.002	−0.018	0.002
	(5.840	(−14.87)	(−1.44)	(−1.26)	(−4.12)	(0.59)
Leverage	0.209	−0.007	0.138	0.011	0.023	−0.022
	(3.56)	(−0.82)	(1.68)	(0.53)	(0.33)	(−0.33)
3-Year R&D Expenditure	−0.939	−0.092	−0.332	−0.043	0.335	−0.012
	(−4.29)	(−4.21)	(−1.31)	(−0.59)	(1.53)	(−0.22)
3-Year Advertising Expenditure	0.734	−0.164	0.908	−0.424	0.149	−0.082
	(2.20)	(−3.05)	(1.75)	(−1.30)	(0.38)	(−0.55)
3-Year ROA	−0.976	0.041	−0.280	0.019	−0.255	−0.002
	(−9.97)	(4.42)	(−2.17)	(0.49)	(−2.36)	(−0.03)
Dividend Yield	0.232	0.015	0.321	−0.013	0.188	−0.009
	(3.19)	(2.57)	(2.84)	(−1.15)	(1.61)	(−0.59)
Past 12-Month Return	0.016	−0.003	−0.029	0.002	−0.010	−0.001
	(1.30)	(−3.33)	(−1.92)	(0.79)	(−0.79)	(−0.39)
Regional Characteristics						
ln(Number of Firms in MSA)	−0.001	−0.003	0.003	0.000	0.071	−0.001
	(−0.07)	(−1.72)	(0.24)	(0.01)	(7.34)	(−0.17)
Social Capital Index	−0.020	0.002	−0.043	0.003	−0.059	−0.036
	(−1.23)	(0.82)	(−1.62)	(0.51)	(−3.02)	(−2.30)
Urban Dummy	0.003	−0.028	0.079	0.004	0.919	0.190
	(0.02)	(−0.86)	(0.41)	(0.09)	(7.94)	(1.18)
Industry Cluster Dummy	−0.012	−0.010	−0.063	−0.004	−0.006	−0.037
	(−0.52)	(−2.99)	(−2.04)	(−0.65)	(−0.21)	(−1.87)
Number of Obs	24,247	40,634	24,051	39,676	24,026	39,666
Adjusted R^2	0.632	0.446	0.034	0.013	0.021	0.014

TABLE 4 (Continued)

Investor Habitats, Trading Correlations, and Informed Trading: Pooled Regression Estimates

Panel B: Extended Regression Specification

Variable	ln RTP	ln (1 + ITP)	Low Pr RTC	Low Pr ITC	Local RTC	Local ITC
Local Stock Preference and Gambling Proxies						
Price	−0.038	0.001	−0.011	−0.000	−0.007	−0.001
	(−14.03)	(2.32)	(−5.42)	(−0.78)	(−5.14)	(−1.54)
Volatility	1.084	−0.301	0.430	−0.026	−0.023	−0.056
	(19.46)	(−31.66)	(4.45)	(−1.32)	(−0.27)	(−1.35)
Skewness	−0.012	−0.002	−0.008	−0.003	−0.036	0.009
	(−0.98)	(−1.02)	(−1.14)	(−0.66)	(−1.51)	(0.73)
EW Distance to Retail	0.024	−0.004	−0.004	−0.005	0.014	−0.017
	(1.40)	(−1.35)	(−0.12)	(−0.71)	(0.52)	(−2.09)
EW Distance to Institutions	−0.015	−0.020	−0.032	0.028	0.084	−0.042
	(−0.57)	(−6.00)	(−0.93)	(3.48)	(3.00)	(−0.70)
Catholic-Protestant Ratio	0.017	−0.001	0.021	−0.000	0.005	0.001
	(2.67)	(−0.54)	(2.23)	(−0.03)	(0.69)	(0.22)
Informed Trading and Uncertainty Proxies						
PIN	−0.522	0.246	0.074	−0.106	0.386	−0.057
	(−5.53)	(10.79)	(0.25)	(−2.58)	(0.74)	(−1.61)
Monthly Turnover	0.237	−0.269	0.605	−0.125	0.318	−0.066
	3.33	−11.45	3.95	−3.86	2.68	−1.78
ln(Firm Age)	0.003	−0.001	−0.042	0.023	−0.036	0.034
	0.20	−0.37	−3.02	4.33	−2.49	1.66
Number of Obs	20,684	31,733	20,488	31,275	20,473	31,270
Adjusted R^2	0.727	0.529	0.042	0.019	0.024	0.020

TABLE 5

Time-Series Factor Model Estimates for Characteristic-Sorted Portfolios

This table reports the factor model estimates for sets of quintile portfolios based on various firm characteristics. The quintile portfolios are formed at the end of each year in December using NYSE size break-points and then held fixed throughout the following year. The following time-series factor model is estimated:

$$R_{pt} - R_{ft} = \alpha_p + \beta_{1p} RMRF_t + \beta_{2p} SMB_t + \beta_{3p} HML_t + \beta_{4p} UMD_t + \beta_{5p} BSI_{pt} + \varepsilon_{pt} \quad t = 1, 2, \ldots, T.$$

Here, R_{pt} is the rate of return on the quintile portfolio, R_{ft} is the riskfree rate of return, $RMRF_t$ is the market return in excess of the riskfree rate, SMB_t is the difference between the value-weighted return of a portfolio of small stocks and the value-weighted return of a portfolio of large stocks, HML_t is the difference between the value-weighted return of a portfolio of high B/M stocks and the value-weighted return of a portfolio of low B/M stocks, UMD_t is the difference between the value-weighted return of a portfolio of stocks with high returns during months $t-12$ to $t-2$ and the value-weighted return of a portfolio of stocks with low returns during months $t-12$ to $t-2$, BSI_{pt} is the buy-sell imbalance for the size-ownership portfolio in month t, and ε_{pt} is the residual return on the portfolio. The portfolio BSI in month t is defined as $BSI_{pt} = \frac{100}{N_p} \sum_{i=1}^{N_p} BSI_{it}$, where BSI_{it} is the buy-sell imbalance of stock i in month t and it is defined in equation (2). The buy-sell imbalance is constructed from the ISSM/TAQ data using small trades, which are available from 1983-2000. The t-statistics for the coefficient estimates are reported in the parentheses below the estimates.

TABLE 5 (Continued)

Time-Series Factor Model Estimates for Characteristic-Sorted Portfolios

	Characteristic-Sorted Portfolio					
Sorting Variable	Low	Q2	Q3	Q4	High	High−Low
Habitat Proxies						
Price	0.114	−0.028	−0.017	−0.010	0.002	−0.112
	(4.76)	(−1.58)	(−0.93)	(−0.88)	(0.52)	(−5.33)
Firm Size	0.128	−0.019	−0.012	0.008	−0.001	−0.129
	(6.24)	(−1.57)	(−0.98)	(0.67)	(−0.29)	(−7.02)
RTP	0.002	0.013	−0.010	0.006	0.088	0.086
	(0.27)	(0.99)	(−0.58)	(0.26)	(3.15)	(3.23)
ITP	0.089	0.001	0.005	−0.003	0.014	−0.075
	(3.46)	(0.09)	(0.32)	(−0.18)	(0.66)	(−2.25)
Gambling Proxies						
Volatility	−0.014	0.006	0.017	0.017	0.057	0.071
	(−0.95)	(0.34)	(0.97)	(0.80)	(1.90)	(2.22)
Skewness	−0.020	0.005	0.025	−0.004	−0.020	0.000
	(−1.52)	(0.44)	(1.54)	(−0.23)	(−1.11)	(0.01)
Lottery Characteristics Index	−0.011	−0.006	0.004	0.014	0.118	0.129
	(−1.28)	(−0.55)	(0.37)	(0.64)	(4.39)	(5.01)
Local Stock Preference Proxies						
Distance to Retail Investors	−0.032	0.006	0.004	0.016	−0.051	−0.019
	(−1.60)	(0.27)	(0.26)	(1.04)	(−2.30)	(−0.64)
Distance to Institutions	−0.012	−0.014	−0.013	0.007	0.049	0.061
	(−0.81)	(−1.10)	(−0.69)	(0.44)	(2.05)	(2.23)
Uncertainty Proxies						
Firm Age	0.032	−0.025	0.009	0.014	−0.008	−0.040
	(1.65)	(−1.16)	(0.38)	(0.86)	(−0.77)	(−1.90)
Monthly Turnover	0.013	0.001	−0.018	−0.005	0.021	0.008
	(0.56)	(0.06)	(−1.03)	(−0.29)	(0.97)	(0.26)

TABLE 6

Mean Return Comovement Estimates For Different Investor Habitats

This table reports the mean return comovements (betas) from the regression of returns on low price, local, high volatility, high skewness, and lottery stock indices, controlling for comovements with the market, SMB, HML,and UMD factors. Mean betas are reported for portfolios of stocks sorted by retail trading proportion, institutional trading proportion, retail trading correlation, and institutional trading correlation. The trading correlation measures are with respect to the same index as the return betas reported in the table. For local comovement, betas are also reported for portfolios sorted by local ITP, a proxy for local participation, and by mean distance to retail and to institutional shareholders. Panel F reports mean betas on the index corresponding to the stocks own price category (low, medium, or high), for portfolios formed by sorting on price category and the retail and institutional trading proportion and trading correlation measures. The difference between high and low portfolios is also reported, with stars to indicate significance. *, **, and *** indicate significance at 1%, 5%, and 10% level, respectively.

Panel A: Low Price Index Comovement						
	Quintiles					
Sorting Variable	Low	Q2	Q3	Q4	High	High−Low
RTP	−0.032	−0.018	0.028	0.091	0.136	0.168***
ITP	0.159	0.132	0.073	0.013	−0.017	−0.176***
RTC	0.005	0.023	0.043	0.062	0.075	0.070***
ITC	0.041	0.042	0.039	0.048	0.055	0.014**
Panel B: Local Stocks Index Comovement						
RTP	0.032	0.026	0.025	0.032	0.034	0.002
ITP	0.047	0.060	0.047	0.033	0.025	−0.022***
Local ITP	0.055	0.050	0.060	0.042	0.022	−0.033***
RTC	0.023	0.020	0.027	0.034	0.046	0.023***
ITC	0.041	0.042	0.040	0.044	0.045	0.004*
Local Retail Part. Proxy	0.016	0.026	0.054	0.057	0.053	0.037***
EW Distance to Retail	0.036	0.037	0.045	0.044	0.035	−0.001
EW Distance to Institutions	0.027	0.023	0.032	0.060	0.070	0.043***
Panel C: High Volatility Index Comovement						
RTP	−0.027	−0.017	0.019	0.071	0.109	0.136***
ITP	0.133	0.115	0.058	0.008	−0.036	−0.169***
RTC	−0.005	0.021	0.040	0.043	0.058	0.063***
ITC	0.033	0.029	0.031	0.040	0.044	0.011*

TABLE 6 (Continued)

Mean Return Comovement Estimates For Different Investor Habitats

	TP or TC Quintiles					
Price Category	Low	Q2	Q3	Q4	High	High−Low
Panel D: High Skewness Index Comovement						
RTP	−0.004	0.001	0.056	0.116	0.160	0.164***
ITP	0.201	0.168	0.094	0.030	0.021	−0.180***
RTC	0.031	0.062	0.069	0.078	0.091	0.060**
ITC	0.075	0.075	0.076	0.075	0.075	0.000
Panel E: Lottery Stocks Index Comovement						
RTP	−0.025	−0.011	0.033	0.100	0.147	0.172***
ITP	0.166	0.146	0.080	0.016	−0.025	−0.191***
RTC	0.005	0.039	0.058	0.065	0.079	0.074***
ITC	0.056	0.053	0.050	0.043	0.048	−0.008
Panel F: Mean Own Price-Category Return Comovements						
RTP Sort						
Low	0.032	0.036	0.066	0.105	0.138	0.106***
Medium	0.137	0.131	0.115	0.108	0.076	−0.061
High	0.213	0.203	0.175	0.091	−0.057	−0.270***
ITP Sort						
Low	0.163	0.159	0.124	0.081	0.064	−0.099***
Medium	0.062	0.111	0.181	0.177	0.165	0.104
High	0.086	0.195	0.223	0.179	0.272	0.186**
RTC Sort						
Low	0.066	0.096	0.108	0.112	0.115	0.049***
Medium	0.111	0.125	0.133	0.146	0.117	0.006
High	0.196	0.217	0.222	0.214	0.225	0.029*
ITC Sort						
Low	0.120	0.116	0.118	0.126	0.121	0.001
Medium	0.165	0.150	0.159	0.170	0.148	−0.017
High	0.261	0.236	0.255	0.240	0.220	−0.041**

TABLE 7

Investor Habitat and Correlated Trading Measures: Correlation Matrix

This table reports the correlation matrix for the set of main explanatory variables, which includes investor habitat and correlated trading measures. All variables have been defined in Appendix Table A.1.

Variable	ln RTP	ln (1 + ITP)	Low Pr RTC	Low Pr ITC	Local RTC	Local ITC	High Vol RTC	High Vol ITC	High Skew RTC	High Skew ITC	Lottery RTC	Lottery ITC
ln(RTP)	1											
ln(1 + ITP)	−0.558	1										
Low Price RTC	0.195	−0.134	1									
Low Price ITC	0.050	−0.032	0.007	1								
Local RTC	0.097	−0.071	0.463	0.014	1							
Local ITC	0.005	0.008	0.007	0.064	0.007	1						
High Vol RTC	0.203	−0.171	0.759	−0.004	0.420	0.007	1					
High Vol ITC	0.032	−0.015	−0.004	0.616	0.005	0.065	−0.009	1				
High Skew RTC	0.185	−0.151	0.705	−0.006	0.419	0.006	0.933	−0.009	1			
High Skew ITC	0.046	−0.029	0.006	0.334	0.001	0.049	−0.001	0.140	−0.002	1		
Lottery Stocks RTC	0.150	−0.120	0.526	−0.005	0.294	0.003	0.725	−0.008	0.694	−0.002	1	
Lottery Stocks ITC	0.022	−0.013	−0.009	0.190	−0.008	0.061	−0.018	0.186	−0.015	0.183	−0.014	1

TABLE 8

Determinants of Price-Based Comovement: Pooled Regression Estimates

This table reports coefficient estimates from pooled OLS regressions of price-based comovement measures on clientele measures, investor trading correlation measures, and other controls. In Panel A, the dependent variable is the annual estimate of the stocks comovement with low-price (below 30^{th} NYSE percentile) stocks, controlling for the standard market, SMB, HML and UMD factors. In panels B, C, and D, the dependent variable is the beta on the index of stocks in the stock's own price category, and results are shown separately for low, medium, and high priced stocks. In Panels B, C, and D, the coefficient estimates of the control variables are suppressed. All variables have been defined in Appendix Table A.1. The data are annual, with the sample period running from 1983-2000 for retail measures and 1980-2005 for institutional measures. The regressions include year and industry effects using the Fama and French (1997) 48 industry definitions, and the standard errors are clustered by firm. The t-statistics for the coefficient estimates are reported in the parentheses below the estimates.

Panel A: Determinants of Low Price Index Comovement

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Habitat and Correlated Trading Measures						
ln(RTP)	0.023				0.020	0.019
	(20.55)				(17.64)	(16.76)
ln(1 + ITP)		−0.117			−0.103	−0.106
		(−16.82)			(−11.56)	(−11.73)
Low Price Index RTC			0.006		0.004	0.004
			(12.44)		(7.97)	(8.34)
Low Price Index ITC				−0.001	−0.000	−0.002
				(−1.10)	(−0.13)	(−1.58)
RTP × RTC						0.001
						(3.56)
ITP × ITC						0.022
						(2.06)
Uncertainty Proxies						
Monthly Turnover	−0.002	0.010	0.018	0.048	0.053	0.052
	(−0.12)	(1.19)	(1.36)	(6.03)	(3.40)	(3.37)
ln(Firm Age)	−0.003	−0.004	−0.003	−0.004	−0.003	−0.003
	(−1.73)	(−2.33)	(−1.49)	(−2.52)	(−2.28)	(−2.30)

Continued . . .

TABLE 8 (Continued)

Determinants of Price-Based Comovement: Pooled Regression Estimates

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Other Firm Characteristics						
ln(Firm Size)	−0.012	−0.020	−0.027	−0.029	−0.008	−0.008
	(−11.07)	(−24.60)	(−29.26)	(−37.80)	(−6.11)	(−6.15)
Market-To-Book	−0.000	−0.002	0.000	−0.001	−0.001	−0.001
	(−0.85)	(−5.04)	(0.51)	(−2.41)	(−1.93)	(−1.91)
Leverage	0.027	0.021	0.031	0.022	0.023	0.023
	(3.25)	(3.18)	(3.76)	(3.14)	(2.81)	(2.80)
3-Year R&D Expenditure	0.038	0.059	0.020	0.055	0.023	0.023
	(1.57)	(3.01)	(0.83)	(2.68)	(0.94)	(0.92)
3-Year Advertising Expenditure	−0.014	−0.008	−0.006	0.007	−0.043	−0.044
	(−0.31)	(−0.22)	(−0.14)	(0.19)	(−1.00)	(−1.01)
3-Year ROA	−0.111	−0.159	−0.130	−0.172	−0.109	−0.108
	(−9.51)	(−18.41)	(−10.97)	(−18.74)	(−9.05)	(−9.02)
Dividend Yield	0.006	0.004	0.012	0.002	0.010	0.011
	(0.57)	(0.81)	(1.10)	(0.40)	(0.94)	(1.04)
Past 12-Month Return	−0.004	0.001	−0.003	0.000	−0.004	−0.004
	(−3.71)	(0.11)	(−3.28)	(0.45)	(−3.83)	(−3.80)
Regional Characteristics						
ln(Number of Firms in MSA)	0.002	0.003	0.001	0.003	0.001	0.001
	(1.19)	(2.81)	(1.08)	(2.58)	(0.72)	(0.73)
Social Capital Index	−0.002	−0.001	−0.002	−0.001	−0.002	−0.003
	(−1.00)	(−0.56)	(−0.91)	(−0.57)	(−1.11)	(−1.09)
Urban Dummy	0.041	0.025	0.041	0.028	0.041	0.041
	(1.92)	(1.31)	(1.86)	(1.39)	(1.97)	(1.96)
Industry Cluster Dummy	0.003	0.000	0.002	0.002	0.002	0.002
	(0.89)	(0.08)	(0.85)	(0.74)	(0.86)	(0.81)
Number of Obs	31,977	54,622	31,764	51,360	30,488	30,488
Adjusted R^2	0.124	0.152	0.115	0.153	0.135	0.135

TABLE 8 (Continued)

Panel B: Determinants of Own Price Category Comovement (Low Priced Stocks)

Variable	(1)	(2)	(3)	(4)	(5)	(6)
ln(RTP)	0.018				0.015	0.016
	(12.38)				(10.08)	(9.99)
ln(1 + ITP)		−0.104			−0.066	−0.076
		(−11.24)			(−5.13)	(−5.71)
Low Price Index RTC			0.005		0.004	0.005
			(8.67)		(6.68)	(6.18)
Low Price Index ITC				−0.001	−0.001	−0.004
				(−3.38)	(−0.48)	(−2.25)
RTP × RTC						−0.001
						(−1.32)
ITP × ITC						0.065
						(2.95)
Number of Obs	16,297	30,095	16,331	26,885	15,102	15,102
Adjusted R^2	0.088	0.109	0.082	0.117	0.096	0.097

Panel C: Determinants of Own Price Category Comovement (Medium Priced Stocks)

Variable	(1)	(2)	(3)	(4)	(5)	(6)
ln(RTP)	−0.010				−0.007	−0.008
	(−2.03)				(−1.37)	(−1.54)
ln(1 + ITP)		0.025			0.069	0.080
		(0.83)			(1.99)	(2.20)
Low Price Index RTC			−0.001		−0.000	0.001
			(−0.47)		(−0.03)	(0.62)
Low Price Index ITC				−0.004	−0.044	−0.014
				(−0.61)	(−2.86)	(−0.40)
RTP × RTC						0.002
						(1.25)
ITP × ITC						−0.077
						(−0.98)
Number of Obs	9,884	14,931	9,882	14,886	9,854	9,854
Adjusted R^2	0.049	0.051	0.049	0.051	0.050	0.050

Panel D: Determinants of Own Price Category Comovement (High Priced Stocks)

Variable	(1)	(2)	(3)	(4)	(5)	(6)
ln(RTP)	−0.019				−0.025	−0.025
	(−3.08)				(−3.79)	(−3.71)
ln(1 + ITP)		0.232			0.222	0.210
		(4.99)			(4.36)	(3.93)
Low Price Index RTC			−0.004		−0.002	−0.001
			(−1.43)		(−0.71)	(−0.33)
Low Price Index ITC				−0.043	−0.012	−0.059
				(−2.71)	(−0.63)	(−1.11)
RTP × RTC						0.000
						(0.08)
ITP × ITC						0.104
						(1.00)
Number of Obs	5,796	9,596	5,551	9,589	5,532	5,532
Adjusted R^2	0.126	0.125	0.123	0.122	0.131	0.131

TABLE 9

Determinants of Georgraphy-Based Comovement: Pooled Regression Estimates

This table reports coefficient estimates from pooled OLS regressions of low price return betas on clientele measures, investor trading correlation measures, and other controls. The dependent variable is the annual estimate of return comovement with other stocks headquartered in the same county, controlling for the market, SMB, HML and UMD factors. All variables have been defined in Appendix Table A.1. Data are annual, with the sample period running from 1983-2000 for retail measures and 1980-2005 for institutional measures. The regressions include year and industry effects using the Fama and French (1997) 48 industry definitions, and the standard errors are clustered by firm. The *t*-statistics for the coefficient estimates are reported in the parentheses below the estimates.

						Local Retail Part. Proxy		
						Low	Medium	High
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Habitat and Correlated Trading Measures								
ln(RTP)	0.001				0.000	0.002	−0.002	0.000
	(1.24)				(0.01)	(1.91)	(−0.92)	(0.18)
ln(1 + ITP)		−0.034			−0.030	−0.003	−0.021	−0.089
		(−4.72)			(−4.04)	(−0.42)	(−1.66)	(−5.50)
Local Index RTC			0.003		0.003	0.002	0.002	0.004
			(6.64)		(6.33)	(2.23)	(2.06)	(5.02)
Local Index ITC				0.001	0.001	0.003	0.001	−0.001
				(2.23)	(2.59)	(1.05)	(2.43)	(−0.15)
Distance to Retail	−0.005	−0.005	−0.005	−0.005	−0.005	0.004	0.002	−0.012
	(−2.39)	(−2.14)	(−2.40)	(−2.16)	(−2.46)	(1.71)	(0.67)	(−3.16)
Distance to Institutions	0.024	0.025	0.024	0.026	0.024	0.004	0.014	0.052
	(7.08)	(7.57)	(7.08)	(7.39)	(6.80)	(1.53)	(3.11)	(6.88)
Uncertainty Proxies								
Monthly Turnover	0.159	0.15)	0.159	0.165	0.148	0.014	0.083	0.192
	(6.65)	(7.35)	(6.74)	(7.77)	(6.17)	(0.98)	(3.95)	(4.09)
ln(Firm Age)	−0.005	−0.007	−0.005	−0.007	−0.005	−0.003	−0.005	−0.006
	(−2.75)	(−3.51)	(−2.77)	(−3.67)	(−2.59)	(−1.17)	(−1.34)	(−2.15)

Continued ...

TABLE 9 (Continued)

Determinants of Geography-Based Comovement: Pooled Regression Estimates

						Local Retail Part. Proxy		
						Low	Medium	High
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Other Firm Characteristics								
ln(Firm Size)	0.003	0.006	0.003	0.004	0.004	0.001	0.005	0.008
	(3.04)	(4.81)	(2.92)	(3.81)	(3.69)	(0.52)	(2.73)	(3.62)
Market-To-Book	−0.001	−0.001	−0.001	−0.001	−0.001	−0.00)	−0.002	−0.001
	(−1.85)	(−3.45)	(−1.88)	(−2.90)	(−2.32)	(−0.14)	(−2.30)	(−0.99)
Leverage	0.017	0.021	0.018	0.021	0.017	0.005	0.015	0.008
	(2.64)	(3.27)	(2.74)	(3.27)	(2.53)	(0.65)	(1.27)	(0.61)
3-Year R&D Expenditure	0.086	0.091	0.083	0.092	0.075	−0.01)	−0.031	0.115
	(3.56)	(4.15)	(3.46)	(4.02)	(3.07)	(−0.35)	(−0.98)	(2.58)
3-Year Advertising Expenditure	−0.086	−0.103	−0.087	−0.094	−0.094	−0.021	−0.12)	−0.124
	(−2.68)	(−3.27)	(−2.72)	(−2.88)	(−2.84)	(−0.55)	(−1.61)	(−2.21)
3-Year ROA	−0.009	−0.022	−0.012	−0.024	−0.01)	−0.015	−0.065	0.009
	(−0.81)	(−2.26)	(−1.08)	(−2.34)	(−0.90)	(−1.20)	(−3.58)	(0.49)
Dividend Yield	−0.02)	−0.003	−0.023	−0.003	−0.023	−0.01)	−0.05)	−0.002
	(−2.15)	(−0.75)	(−2.51)	(−0.81)	(−2.37)	(−0.90)	(−2.53)	(−0.17)
Past 12-Month Return	0.001	0.002	0.001	0.002	0.001	0.002	−0.001	0.003
	(0.91)	(1.87)	(1.04)	(2.02)	(0.98)	(1.82)	(−0.24)	(1.46)
Regional Characteristics								
ln(Number of Firms in MSA)	0.004	0.006	0.004	0.006	0.004	0.001	0.004	0.005
	(4.82)	(6.50)	(4.62)	(6.24)	(4.23)	(0.48)	(2.71)	(2.05)
Social Capital Index	0.001	−0.001	0.002	−0.001	0.001	−0.001	0.001	0.003
	(0.82)	(−0.63)	(1.11)	(−0.91)	(0.84)	(−0.38)	(0.38)	(1.03)
Urban Dummy	0.054	0.049	0.051	0.051	0.052	0.028	0.03)	0.088
	(4.54)	(4.13)	(4.23)	(4.09)	(4.27)	(2.40)	(1.76)	(2.58)
Industry Cluster Dummy	0.013	0.015	0.013	0.016	0.013	0.003	0.002	0.021
	(5.63)	(6.36)	(5.53)	(6.43)	(5.51)	(1.30)	(0.40)	(3.26)
Number of Obs	26,856	41,178	26,618	39,964	25,970	8,468	8,723	8,779
Adjusted R^2	0.087	0.114	0.090	0.116	0.093	0.022	0.157	0.158

TABLE 10

Trading Correlation and Return Comovement Changes Around Stock Splits

This table reports changes in retail trading correlations and the sensitivity of returns to retail trading around stock splits. Panel A reports changes in retail trading correlation. For each split event, the stock's own retail BSI is regressed on the portfolio BSI of of the price portfolios consisting of stocks in its own pre-split price decile and its own post-split price decile, controlling for the market return. These BSI betas are estimated for 24 months prior to the split and for 24 months following the split. The first column reports the change in coefficient on the portfolio BSI of the stock's pre-split price decile from before the split to after the split, while the second column shows the change in coefficient on the stock's post-split decile portfolio. The third column reports the change in coefficient on the portfolio BSI of a low price index (below 30^{th} NYSE percentile) from a separate estimation. In Panel B, the dependent variable is monthly returns, and the two columns report the change in coefficient on the portfolio BSI of the stock's pre-split price decile and the change in coefficient on the stock's post-split decile portfolio, respectively, controlling for the market, SMB, HML and UMD factors. The beta estimates for the pre- and post-split periods are winsorized at the 2.5 and 97.5 percentile levels to remove extreme outliers. In each panel, results are shown with increasing restrictions on the number of observations required in the pre- and post-split windows (12, 18 and 20). The t-statistics for the coefficient estimates are reported in the parentheses below the estimates.

Panel A: Change in RTC Around Stock Splits

		BSI Portfolio Index		
Min Nobs	N	Pre-Split Price	Post-Split Price	Low Price
12	3,301	−0.426	0.364	0.060
		(−6.71)	(5.47)	(1.73)
18	2,661	−0.452	0.413	0.105
		(−6.56)	(5.75)	(2.84)
24	1,622	−0.502	0.471	0.194
		(−5.52)	(5.16)	(4.13)

Panel B: Change in Sensitivity to BSI Around Stock Splits

		BSI Portfolio Index	
Min Nobs	N	Pre-Split Price	Post-Split Price
12	4,257	−0.241	0.233
		(−3.51)	(3.28)
18	3,155	−0.247	0.230
		(−3.27)	(2.94)
24	1,877	−0.233	0.202
		(−2.33)	(1.95)

TABLE 11

Firm-Level and Regional Gambling Proxies and Return Comovements

This table reports coefficient estimates from pooled OLS regressions of price-based and local comovement measures on stock and regional characteristics associated with gambling behavior, as well as other (unreported) controls. In Panel A, the dependent variable is the annual estimate of the stocks comovement with low-price (below 30^{th} NYSE percentile) stocks, controlling for the market, SMB, HML and UMD factors. In Panel B, the dependent variable is the annual estimate of the stocks comovement with local (within the same county) stocks. The primary gambling-related variables are price, volatility, skewness, a lottery dummy and more continuous lottery index which each combine price volatility and skewness, and the ratio of Catholics to Protestants in the county where the firm is located. Unreported controls include dividend yield, 12-month past return, firm age, turnover, size, market/book, leverage, 3-year averages of R&D expenditures, advertising expenditures, and ROA, number of firms in the MSA, urban dummy, and an industry cluster dummy. All variables have been defined in Appendix Table A.1. The data are annual, with the sample period running from 1980-2005. The regressions include year and industry effects using the Fama and French (1997) 48 industry definitions, and the standard errors are clustered by firm. The t-statistics for the coefficient estimates are reported in the parentheses below the estimates.

TABLE 11 (Continued)

Firm-Level and Regional Gambling Proxies and Return Comovements

Panel A: Low Price Index Comovement

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Stock Price	−0.001						
	(−2.71)						
Lagged Volatility	0.038						
	(6.88)						
Lagged Skewness	0.010						
	(5.11)						
Lottery Stock Dummy		0.042		0.034		0.028	
		(16.66)		(13.63)		(7.16)	
Lottery Characteristics Index			0.140		0.179		0.156
			(19.26)		(24.07)		(15.79)
Catholic-Protestant Ratio				0.000	0.001	−0.000	−0.005
				(0.52)	(0.60)	(−0.22)	(−2.69)
Lottery Stock Dummy × CPRATIO						0.002	
						(1.99)	
Lottery Characteristics Index × CPRATIO							0.011
							(3.60)
Number of Obs	54,622	52,247	52,247	54,613	54,622	54,613	54,622
Adjusted R^2	0.149	0.151	0.154	0.150	0.158	0.150	0.158

Panel B: Local Stocks Index Comovement

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Stock Price	−0.001						
	(−2.39)						
Lagged Volatility	0.024						
	(5.58)						
Lagged Skewness	0.003						
	(3.00)						
Lottery Stock Dummy		0.008		0.010		−0.004	
		(4.70)		(5.64)		(−1.46)	
Lottery Characteristics Index			0.042		0.042		−0.010
			(7.10)		(6.77)		(−1.11)
Catholic-Protestant Ratio				−0.002	−0.002	−0.004	−0.014
				(−3.40)	(−3.38)	(−4.85)	(−8.27)
Lottery Stock Dummy × CPRATIO						0.006	
						(5.99)	
Lottery Characteristics Index × CPRATIO							0.023
							(8.94)
Number of Obs	54,622	52,247	52,247	54,613	54,622	54,613	54,622
Adjusted R^2	0.109	0.110	0.111	0.109	0.110	0.110	0.114

TABLE 12

Return Comovements Among Stocks with Lottery Features: Pooled Regression Estimates

This table reports cross-sectional estimates for return betas with respect to other gambling related stock characteristics. In addition to betas estimated on a low price index, we also use betas estimated on a high volatility index, betas on a high skewness index, and betas on an index of lottery stocks as alternate dependent variables. The regressions specification is identical to specification (6) in Table 8, but with alternate dependent variables. Panel A reports a correlation matrix of the dependent variables, which are estimates from annual regressions of daily returns on a low price, high volatility, high skewness, or lottery stock index, controlling for the market, SMB, HML, and UMD factors. Panel B reports estimates from the cross-sectional regressions. In the last three column report results from regressions which also included the low price beta as a traditional control. All variables have been defined in Appendix Table A.1. The data are annual, with the sample period running from 1980-2005. The regressions include year and industry effects using the Fama and French (1997) 48 industry definitions, and the standard errors are clustered by firm. The *t*-statistics for the coefficient estimates are reported in the parentheses below the estimates.

Panel A: Correlation Matrix

Comovement Measure	Low Price	High Volatility	High Skewness	Lottery-Type Stocks
Low Price	1			
High Volatility	0.875	1		
High Skewness	0.723	0.565	1	
Lottery-Type Stocks	0.893	0.867	0.714	1

TABLE 12 (Continued)

Return Comovements Among Stocks with Lottery Features: Pooled Regression Estimates

Panel B: Regression Estimates

		Total Comovement			Residual Comovement		
	(From Table 7)	High	High	Lottery	High	High	Lottery
	Low Price	Volatility	Skewness	Type	Volatility	Skewness	Type
Variable	Index	Index	Index	Index	Index	Index	Index
ln(RTP)	0.019	0.015	0.018	0.021	0.012	0.012	0.013
	(16.76)	(15.00)	(9.69)	(15.66)	(5.25)	(3.79)	(5.47)
ln(1 + ITP)	−0.106	−0.093	−0.086	−0.136	−0.048	−0.052	−0.045
	(−11.73)	(−12.52)	(−5.99)	(−14.86)	(−6.90)	(−3.20)	(−12.02)
Own Index RTC	0.004	0.035	0.033	0.020	0.009	0.024	0.022
	(8.34)	(13.72)	(6.98)	(4.04)	(8.71)	(5.28)	(3.35)
Own Index ITC	−0.002	0.000	−0.002	−0.005	−0.001	−0.000	−0.000
	(−1.58)	(0.41)	(−0.89)	(−2.71)	(−1.19)	(−0.21)	(−0.35)
RTP × RTC	0.001	0.015	0.014	0.006	0.003	0.012	0.008
	(3.56)	(10.56)	(5.55)	(2.69)	(5.30)	(4.30)	(2.16)
ITP × ITC	0.022	0.013	−0.025	0.021	0.003	−0.029	0.001
	(2.06)	(1.49)	(−1.50)	(2.54)	(0.91)	(−2.07)	(1.33)
Monthly Turnover	0.052	0.049	0.055	0.043	0.055	0.063	0.053
	(3.57)	(5.70)	(2.55)	(7.21)	(11.94)	(3.83)	(8.78)
ln(Firm Age)	−0.003	−0.002	−0.002	−0.004	−0.002	−0.001	−0.002
	(−2.30)	(−1.48)	(−1.60)	(−2.01)	(−2.12)	(−1.91)	(−2.86)
Number of Obs	30,488	30,325	30,324	30,321	30,325	30,324	30,321
Adjusted R^2	0.135	0.156	0.087	0.154	0.875	0.591	0.816

TABLE 13

Clientele Characteristics, Return Comovements, and Trading Correlations: Cross-Sectional Regression Estimates

This table reports cross-sectional regression estimates, where the dependent variable is either a return comovement measure, retail trading correlation (RTC) measure, or the retail trading proportion (RTP) of a stock measured over the 1991 to 1996 time period. The betas are defined in Appendix Table A.1. The main independent variables are the following characteristics of the retail investor clientele that trades the stock: Age, Annual Income, Education, Professional Occupation, Gender (Proportion Male), Marital Status (Proportion Married), Proportion Catholic, 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 (portfolio variance divided by the average variance of stocks in the portfolio). The clientele characteristic is the equal-weighted average characteristic of retail investors who trade the stock during the brokerage sample period, where stocks with fewer than five trades are excluded from the sample. 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 to 1996 time period. The t-statistics are reported in parentheses below the estimates. To ensure that extreme values are not affecting the results, we winsorize all variables at their 0.5 and 99.5 percentile levels. Both the dependent and the independent variables have been standardized (mean is set to zero and standard deviation is one).

TABLE 13 (Continued)

Clientele Characteristics, Return Comovements, and Trading Correlations: Cross-Sectional Regression Estimates

	Comovement Measure					RTC Measure		From HK09
Variable	Low Price	Local	High Vol	High Skew	Lottery	Local	Lottery	RTP
Intercept	−0.035	0.009	−0.020	−0.005	−0.016	0.007	0.005	−0.047
	(−2.82)	(0.67)	(−1.51)	(−0.41)	(−1.47)	(0.60)	(0.38)	(−4.12)
Age	−0.029	−0.051	−0.022	−0.017	−0.028	−0.023	−0.004	−0.025
	(−1.71)	(−3.68)	(−1.91)	(−2.41)	(−2.47)	(−2.81)	(−0.32)	(−2.52)
Income	0.012	−0.002	0.010	−0.006	−0.019	−0.029	−0.019	−0.036
	(0.89)	(−0.10)	(0.73)	(−1.44)	(−2.50)	(−2.21)	(−2.42)	(−3.50)
Professional Dummy	−0.030	0.011	−0.021	−0.017	−0.018	0.021	−0.012	−0.024
	(−2.17)	(0.90)	(−2.05)	(−1.86)	(−1.92)	(0.55)	(−1.91)	(−1.63)
Proportion Male	0.113	0.017	0.105	0.063	0.094	0.051	0.028	0.078
	(6.65)	(1.47)	(5.53)	(3.35)	(5.37)	(2.94)	(2.61)	(5.93)
Proportion Married	−0.044	−0.023	−0.052	−0.026	−0.034	−0.040	−0.003	−0.059
	(−2.65)	(−2.20)	(−2.85)	(−1.98)	(−2.01)	(−2.38)	(−0.17)	(−4.08)
Portfolio Concentration	0.087	0.031	0.076	0.046	0.082	0.019	0.038	0.112
	(6.35)	(1.97)	(4.90)	(2.69)	(5.76)	(1.34)	(2.64)	(9.39)
Education	−0.065	−0.019	−0.058	−0.049	−0.052	−0.017	−0.044	−0.060
	(−4.74)	(−2.23)	(−3.94)	(−2.99)	(−3.78)	(−1.21)	(−3.18)	(−6.63)
Prop Catholic	0.024	0.031	0.023	0.021	0.029	0.017	0.021	0.091
	(1.87)	(2.06)	(1.68)	(2.16)	(2.12)	(1.55)	(1.97)	(4.93)
Prop African American	0.023	0.008	0.023	0.013	0.028	−0.009	0.003	0.014
	(1.78)	(1.69)	(1.63)	(1.94)	(2.13)	(−0.66)	(0.24)	(1.78)
Prop Hispanic	0.013	0.006	0.011	0.012	0.017	−0.008	0.001	0.017
	(2.11)	(1.40)	(1.77)	(2.17)	(1.48)	(−0.59)	(0.04)	(2.06)
Prop Foreign Born	−0.002	0.059	0.016	−0.001	0.005	−0.013	−0.003	0.003
	(−0.12)	(3.18)	(0.88)	(−0.04)	(0.30)	(−0.79)	(−0.21)	(0.56)
Prop Urban	0.065	0.012	0.076	0.044	0.067	0.028	0.020	0.025
	(3.99)	(0.85)	(4.08)	(2.48)	(4.10)	(1.71)	(2.24)	(2.69)
Avg State Lottery Sales	0.041	0.066	0.035	0.047	0.046	0.029	0.032	0.063
	(3.05)	(4.68)	(3.58)	(3.04)	(3.85)	(3.22)	(2.96)	(5.18)
Number of Stocks	6,053	5,555	5,637	5,637	6,047	5,695	5,872	5,925
Adjusted R^2	0.046	0.034	0.042	0.022	0.048	0.006	0.008	0.059

TABLE 14

Effect of Aggregate Uncertainty and Consumer Sentiment on Trading Correlations and Return Comovements: Sorting Results

This table shows mean values of return comovement estimates, measures of investor clienteles, an trading correlation for stock-year observations sorted according to annual uncertainty and sentiment measures. Panel A shows the mean estimates of return comovement with low-price stocks, high volatility stocks among local stocks with high skewness stocks, and with lottery type stocks for stocks-years in 5 quintiles sorted by the annual mean level of VIX (our proxy for market uncertainty). Panel B shows retail habitat and trading correlation measures for sorts on VIX levels, while Panel C shows average institutional habitat and trading correlation measures for the same VIX sorts. Panels D, E, and F show averages of the same variables for stock-years sorted on the level of the Michigan Consumer Sentiment Index. All comovement, habitat and trading correlation variables have been defined in Appendix Table A.1. The data are annual, with the sample period running from 1980-2005 (retail measures are from 1983-2000).

Panel A: Uncertainty and Return Comovements

	VIX Quintiles					
Comovement	Low	Q2	Q3	Q4	High	High−Low
Low Price	0.047	0.062	0.051	0.073	0.100	0.053***
Local	0.031	0.039	0.028	0.057	0.064	0.033**
High Volatility	0.036	0.048	0.037	0.063	0.092	0.056***
High Skewness	0.071	0.090	0.098	0.115	0.151	0.080***
Lottery Stocks	0.048	0.069	0.064	0.076	0.107	0.059***

Panel B: Uncertainty, Retail Trading Proportion, and Correlations

	VIX Quintiles					
Retail Measure	Low	Q2	Q3	Q4	High	High−Low
ln(RTP)	−0.291	−0.101	−0.261	0.376	0.105	0.396***
Low Price RTC	0.744	0.707	0.710	0.712	0.804	0.060*
Local RTC	0.306	0.305	0.432	0.497	0.584	0.278***
High Volatility RTC	0.121	0.110	0.150	0.171	0.220	0.099***
High Skewness RTC	0.119	0.114	0.157	0.178	0.212	0.093***
Lottery Stocks RTC	0.116	0.106	0.145	0.179	0.208	0.092***

Panel C: Uncertainty, Institutional Trading Proportion, and Correlations

	VIX Quintiles					
Institutional Measure	Low	Q2	Q3	Q4	High	High−Low
ln(1+ITP)	0.266	0.256	0.247	0.251	0.241	−0.025**
ln(1+Local ITP)	0.050	0.050	0.048	0.047	0.046	−0.003*
Low Price ITC	0.142	0.104	0.190	0.162	0.144	0.002
Local ITC	0.095	0.103	0.120	0.134	0.112	0.018
High Volatility ITC	0.142	0.102	0.180	0.165	0.149	0.007
High Skewness ITC	0.135	0.148	0.201	0.139	0.134	−0.002
Lottery Stocks ITC	0.114	0.118	0.131	0.091	0.112	−0.002

TABLE 14 (Continued)

Effect of Aggregate Uncertainty and Consumer Sentiment on Trading Correlations and Return Comovements: Sorting Results

Panel D: Sentiment and Return Comovements

	Michigan Sentiment Quintiles					
Comovement	Low	Q2	Q3	Q4	High	High−Low
Low Price	0.078	0.065	0.077	0.055	0.079	0.001
Local	0.031	0.045	0.047	0.021	0.058	0.027**
High Volatility	0.043	0.045	0.069	0.039	0.069	0.026**
High Skewness	0.083	0.089	0.113	0.078	0.130	0.047**
Lottery Stocks	0.079	0.066	0.083	0.063	0.085	0.006

Panel E: Sentiment, Retail Trading Proportion, and Correlations

	Michigan Sentiment Quintiles					
Retail Measure	Low	Q2	Q3	Q4	High	High−Low
ln(RTP)	−0.107	−0.125	−0.223	−0.381	0.193	0.300*
Low Price RTC	0.484	0.781	0.759	0.830	0.745	0.262***
Local RTC	0.325	0.392	0.414	0.421	0.493	0.169***
High Volatility RTC	0.094	0.158	0.160	0.175	0.175	0.081***
High Skewness RTC	0.094	0.163	0.159	0.184	0.177	0.083***
Lottery Stocks RTC	0.084	0.155	0.152	0.174	0.179	0.095***

Panel F: Sentiment, Institutional Trading Proportion, and Correlations

	Michigan Sentiment Quintiles					
Institutional Measure	Low	Q2	Q3	Q4	High	High−Low
ln(1+ITP)	0.209	0.264	0.250	0.244	0.249	0.040*
ln(1+Local ITP)	0.044	0.051	0.048	0.051	0.046	0.002
Low Price ITC	0.123	0.121	0.146	0.077	0.155	0.031
Local ITC	0.120	0.092	0.103	0.088	0.095	−0.026
High Volatility ITC	0.140	0.125	0.144	0.094	0.151	0.011
High Skewness ITC	0.179	0.120	0.136	0.103	0.142	−0.037
Lottery Stocks ITC	0.127	0.109	0.107	0.056	0.126	−0.001

TABLE 15

Effect of Aggregate Uncertainty and Consumer Sentiment on Return Comovements

This table reports pooled regression estimates for return betas with respect to low price stocks, local stocks high volatility stocks, high skewness stocks, and lottery-type stocks. Each of the return comovement measures is regressed on the primary investor habitat and trading correlation measures, the annual mean of the VIX index (Panel A) or the Michigan Consumer Sentiment Index (Panel B). The regression specification is identical to specification (6) in Table 8, with the addition of the sentiment/uncertainty variables. The coefficients on other control variables have been suppressed for brevity. All variables have been defined in Appendix Table A.1. Data are annual, with the sample period running from 1980-2005. The regressions include firm and industry effects using the Fama and French (1997) 48 industry definitions, and the standard errors are clustered by year. The t-statistics for the coefficient estimates are reported in the parentheses below the estimates.

Panel A: Effect of VIX

	Comovement Measure				
Variable	Low Price	Local	High Vol	High Skew	Lottery
ln(RTP)	0.018	0.005	0.015	0.020	0.018
	(3.49)	(2.73)	(3.80)	(3.83)	(3.93)
ln(1+ITP)	−0.010	−0.026	−0.006	−0.046	−0.022
	(−1.02)	(−2.64)	(−0.63)	(−2.89)	(−2.35)
Retail TC	0.002	0.001	0.024	0.029	0.014
	(2.18)	(0.92)	(2.53)	(3.92)	(4.01)
Institutional TC	−0.004	−0.002	−0.001	−0.002	−0.004
	(−1.54)	(−0.77)	(−0.38)	(−1.21)	(−2.18)
RTP × RTC	0.001	0.001	0.014	0.017	0.005
	(4.22)	(1.87)	(4.59)	(3.96)	(3.85)
ITP × ITC	0.008	0.001	0.001	−0.006	0.014
	(0.53)	(0.28)	(0.09)	(−0.26)	(1.47)
VIX	0.001	0.001	0.001	0.001	0.001
	(2.55)	(3.44)	(1.81)	(2.22)	(4.90)
Number of Obs	27,507	27,487	27,388	27,377	27,388
Adj. R^2	0.367	0.442	0.384	0.315	0.382

Panel B: Effect of Michigan Sentiment Index

	Comovement Measure				
Variable	Low Price	Local	High Vol	High Skew	Lottery
ln(RTP)	0.018	0.005	0.015	0.019	0.019
	(4.19)	(2.76)	(4.32)	(4.18)	(4.59)
ln(1+ITP)	−0.019	−0.027	−0.019	−0.012	−0.029
	(−1.53)	(−2.58)	(−1.33)	(−0.60)	(−2.69)
Retail TC	0.003	0.001	0.026	0.027	0.016
	(2.74)	(0.99)	(3.03)	(3.46)	(4.23)
Institutional TC	−0.004	0.001	0.000	−0.002	−0.005
	(−1.66)	(1.51)	(−0.33)	(−0.93)	(−2.55)
RTP × RTC	0.001	0.001	0.014	0.014	0.006
	(4.31)	(1.99)	(4.87)	(3.54)	(4.12)
ITP × ITC	0.019	−0.002	0.017	−0.037	0.020
	(1.47)	(−5.13)	(1.30)	(−1.55)	(2.01)
Michigan Sentiment Index	0.001	0.001	0.000	0.001	0.001
	(2.17)	(2.72)	(1.33)	(0.66)	(1.67)
Number of Obs	30,488	30,449	30,325	30,324	30,321
Adj. R^2	0.347	0.420	0.368	0.294	0.368

FIGURE 1
Return Comovements Across Time

This figure shows the annual means of the low price and local return comovement measures.

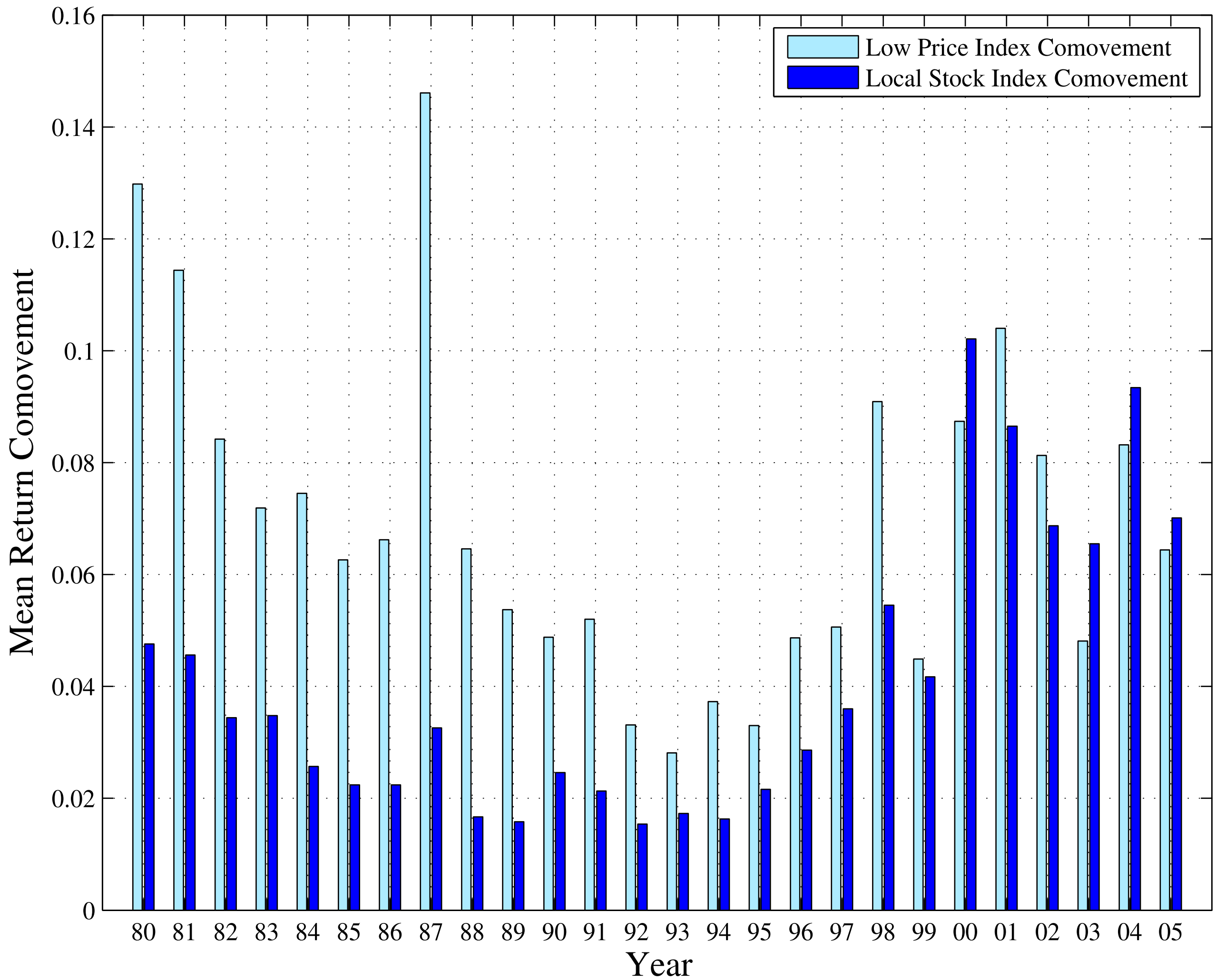
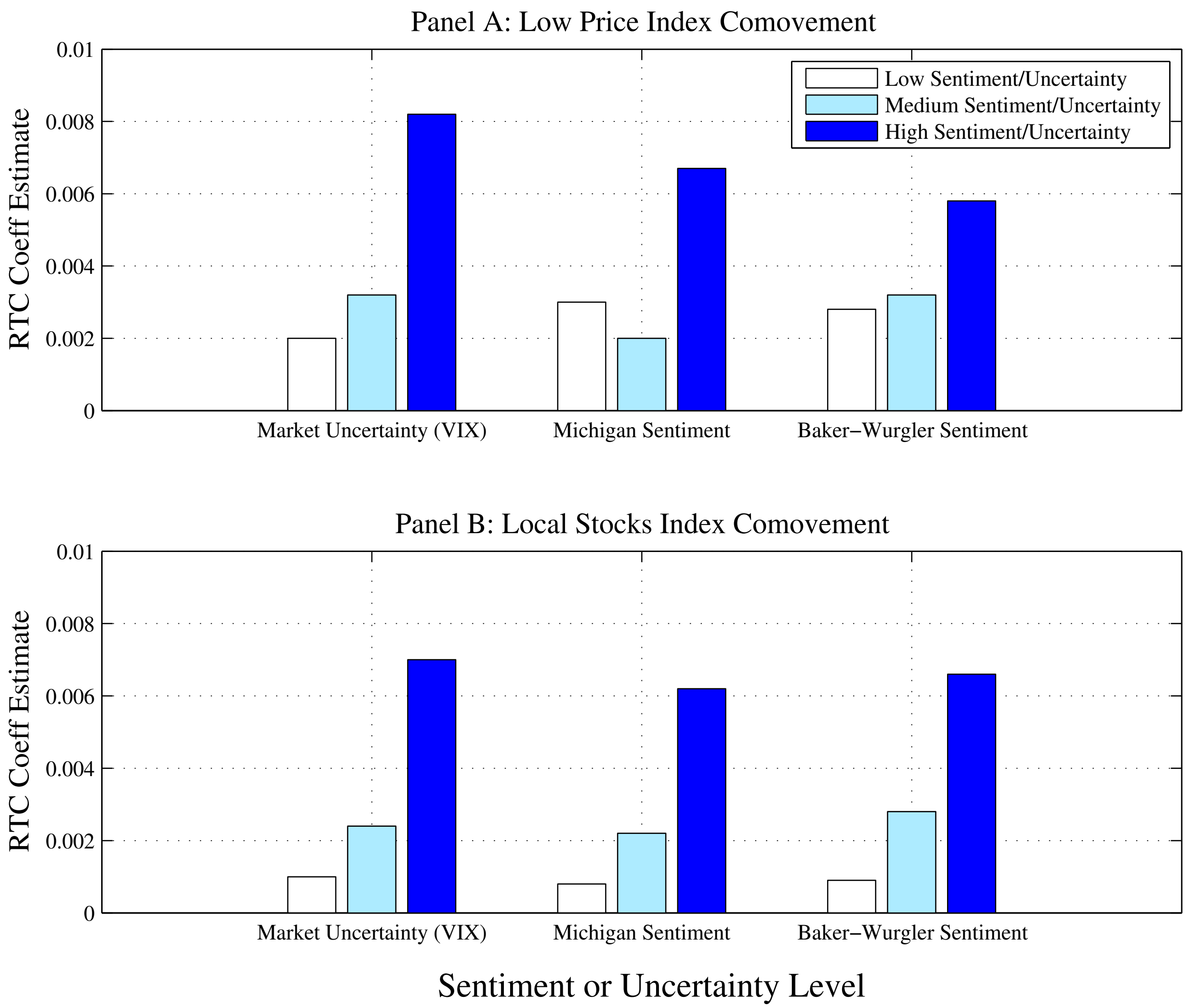


FIGURE 2
Effect of Correlated Retail Trading on Return Comovements

This figure shows the RTC coefficient estimates from low price index (Panel A) and local (Panel B) comovement regressions, which are estimated separately for low, medium, and high sentiment/uncertainty sub-periods. The captions of Tables 8 and 9 provide details about the comovement regressions.



Appendix Table A.1
Brief Definitions and Sources of Main Variables

This table briefly defines the main variables used in the empirical analysis. The data sources are: (i) ARDA: Association of Religion Data Archives, (ii) Brokerage: Large U.S. discount brokerage, (iii) Census: U.S. Census County Files, (iv) Compustat (iv) CRSP: Center for Research on Security Prices, (v) Estimated: Estimated by the authors, (vi) Putnam: Robert Putnam's website, www.bowlingalone.com (vii) ISSM/TAQ: Institute for the Study of Security Markets (ISSM) and the Trade and Quote (TAQ) databases, (viii) 13(f): 13(f) institutional portfolio holdings data from Thomson Reuters. Table 1 reports the summary statistics for all these variables.

Variable Name	Description	Source
Return Comovement Measures		
Low Price Beta	β_1 from the regression $r_{it} - r_f = \beta_0 + \beta_1 LowPrcIdx_{it} + \beta_2 MKTRF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_{it}$, where $LowPrcIdx$ is the portolio return of stocks priced below the 30^{th} NYSE percentile of price at the end of the prior year (excluding stock i). The beta is estimated annually using daily data.	Estimated
Local Beta	β_1 from the regression $r_{it} - r_f = \beta_0 + \beta_1 LocalIdx_{it} + \beta_2 MKTRF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_{it}$, where $LocalIdx$ is the portolio return of other stocks in the county (excluding stock i). The beta is estimated annually using daily data.	Estimated
High Volatility Beta	β_1 from the regression $r_{it} - r_f = \beta_0 + \beta_1 HighVolIdx_{it} + \beta_2 MKTRF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_{it}$, where $HighVolIdx$ is the portolio return of stocks above the 70^{th} NYSE percentile of total daily return volatility over the prior year (excluding stock i). The beta is estimated annually using daily data.	Estimated
High Skewness Beta	β_1 from the regression $r_{it} - r_f = \beta_0 + \beta_1 HighSkewIdx_{it} + \beta_2 MKTRF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_{it}$, where $HighSkewIdx$ is the portolio return of stocks above the 70^{th} NYSE percentile of total daily return skewness over the prior year (excluding stock i). The beta is estimated annually using daily data.	Estimated
Lottery Stock Beta	β_1 from the regression $r_{it} - r_f = \beta_0 + \beta_1 LottoIdx_{it} + \beta_2 MKTRF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_{it}$, where $LottoIdx$ is the portolio return of stocks with below-medianprice and above median volatility and skewness over the prior year (excluding stock i). The beta is estimated annually using daily data.	Estimated
Own Price Beta	β_1 from the regression $r_{it} - r_f = \beta_0 + \beta_1 PrcIdx_{it} + \beta_2 MKTRF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_{it}$, where $PrcIdx$ is the portolio return of the price category to which stock i belongs (excluding stock i). There are three price categories defined by the 30^{th} and 70^{th} NYSE percentiles of price at the end of the prior year. The beta is estimated annually using daily data.	Estimated

Variable Name	Description	Source
Own Volatility Beta	β_1 from the regression $r_{it} - r_f = \beta_0 + \beta_1 VolIdx_{it} + \beta_2 MKTRF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_{it}$, where $VolIdx$ is the portolio return of the volatility category to which stock i belongs (excluding stock i). There are three volatility categories defined by the 30^{th} and 70^{th} NYSE percentiles of total daily return volatility over the prior year. The beta is estimated annually using daily data.	Estimated
Own Skewness Beta	β_1 from the regression $r_{it} - r_f = \beta_0 + \beta_1 SkewIdx_{it} + \beta_2 MKTRF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_{it}$, where $SkewIdx$ is the portolio return of the skewness category to which stock i belongs (excluding stock i). There are three skewness categories defined by the 30^{th} and 70^{th} NYSE percentiles of total daily return skewness over the prior year. The beta is estimated annually using daily data.	Estimated
Own Lottery Beta	β_1 from the regression $r_{it} - r_f = \beta_0 + \beta_1 LottoIdx_{it} + \beta_2 MKTRF_t + \beta_3 SMB_t + \beta_4 HML_t + \beta_5 UMD_t + \varepsilon_{it}$, where $LottoIdx$ is the portolio return of the lottery stock category (lottery, non-lottery or other) to which stock i belongs (excluding stock i). The beta is estimated annually using daily data.	Estimated
Retail Trading Correlation Measures		
Low Price RTC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 LowPrcBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of small trades in stock i during month t and $LowPrcBSIIdx_{it}$ is the equal-weighted mean BSI of stocks priced below the 30^{th} NYSE percentile (excluding stock i). The BSI beta is estimated annually using monthly data.	Estimated
Local RTC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 LocalBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of small trades in stock i during month t and $LowPrcBSIIdx_{it}$ is the equal-weighted mean BSI of other stocks in located in the same county (excluding stock i). The BSI beta is estimated annually using monthly data.	Estimated
High Volatility RTC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 HighVolBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of small trades in stock i during month t and $HighVolBSIIdx_{it}$ is the equal-weighted mean BSI of stocks above the 70^{th} NYSE percentile of total daily return volatility over the prior year (excluding stock i).	Estimated
High Skewness RTC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 HighSkewBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of small trades in stock i during month t and $HighSkewBSIIdx_{it}$ is the equal-weighted mean BSI of stocks above the 70^{th} NYSE percentile of total daily return skewness over the prior year (excluding stock i).	Estimated
Lottery Stock RTC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 LottoBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of small trades in stock i during month t and $LottoBSIIdx_{it}$ is the equal-weighted mean BSI of lotery-type stocks (excluding stock i).	Estimated

Variable Name	Description	Source
Own Price RTC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 PrcBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of small trades in stock i during month t and $PrcBSIIdx_{it}$ is the portolio return of the price category to which stock i belongs (excluding stock i). There are three price categories defined by the 30^{th} and 70^{th} NYSE percentiles of price at the end of the prior year. The BSI beta is estimated annually using monthly data.	Estimated
Own Volatility RTC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 VolBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the equal-weighted mean BSI of the volatility category to which stock i belongs (excluding stock i). There are three volatility categories defined by the 30^{th} and 70^{th} NYSE percentiles of total daily return volatility over the prior year.	Estimated
Own Skewness RTC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 SkewBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the equal-weighted mean BSI of the skewness category to which stock i belongs (excluding stock i). There are three skewness categories defined by the 30^{th} and 70^{th} NYSE percentiles of total daily return skewness over the prior year.	Estimated
Own Lottery RTC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 LottoBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of small trades in stock i during month t and $LottoBSIIdx_{it}$ is the equal-weighted mean BSI of the lottery stock category (lottery, non-lottery or other) to which stock i belongs (excluding stock i).	Estimated
Institutional Trading Correlation Measures		
Low Price ITC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 LowPrcBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of institutional trades in stock i during month t and $LowPrcBSIIdx_{it}$ is the equal-weighted mean BSI of stocks priced below the 30^{th} NYSE percentile (excluding stock i). The BSI beta is estimated over a three year window using quarterly data.	Estimated
Local ITC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 LocalBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of institutional trades in stock i during month t and $LowPrcBSIIdx_{it}$ is the equal-weighted mean BSI of other stocks in located in the same county (excluding stock i). The BSI beta is estimated over a three-year window using quarterly data.	Estimated
High Volatility ITC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 HighVolBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of institutional trades in stock i during month t and $HighVolBSIIdx_{it}$ is the equal-weighted mean BSI of stocks above the 70^{th} NYSE percentile of total daily return volatility over the prior year (excluding stock i).	Estimated

Variable Name	Description	Source
High Skewness ITC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 HighSkewBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of institutional trades in stock i during month t and $HighSkewBSIIdx_{it}$ is the equal-weighted mean BSI of stocks above the 70^{th} NYSE percentile of total daily return skewness over the prior year (excluding stock i).	Estimated
Lottery Stock ITC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 LottoBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of institutional trades in stock i during month t and $LottoBSIIdx_{it}$ is the equal-weighted mean BSI of lotery-type stocks (excluding stock i).	Estimated
Own Price ITC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 PrcBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of institutional trades in stock i during month t and $PrcBSIIdx_{it}$ is the portolio return of the price category to which stock i belongs (excluding stock i). There are three price categories defined by the 30^{th} and 70^{th} NYSE percentiles of price at the end of the prior year. The BSI beta is estimated annually using monthly data.	Estimated
Own Volatility ITC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 VolBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the equal-weighted mean BSI of the volatility category to which stock i belongs (excluding stock i). There are three volatility categories defined by the 30^{th} and 70^{th} NYSE percentiles of total daily return volatility over the prior year.	Estimated
Own Skewness ITC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 SkewBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the equal-weighted mean BSI of the skewness category to which stock i belongs (excluding stock i). There are three skewness categories defined by the 30^{th} and 70^{th} NYSE percentiles of total daily return skewness over the prior year.	Estimated
Own Lottery ITC	β_1 from the regression $BSI_{it} = \beta_0 + \beta_1 LottoBSIIdx_{it} + \beta_2 MKTRF_t + \varepsilon_{it}$, where BSI_{it} is the buy-sell imbalance of institutional trades in stock i during month t and $LottoBSIIdx_{it}$ is the equal-weighted mean BSI of the lottery stock category (lottery, non-lottery or other) to which stock i belongs (excluding stock i).	Estimated
Clientele Characteristics		
ln(RTP)	Natural log of the ratio of dollar volume of small trades to total dollar volume.	ISSM/TAQ
ln(1+IO)	Natural log of 1 plus the fraction of shares outstanding owned by institutions reported on 13(f) filings.	13(f)
ln(1+Local IO)	Natural log of 1 plus the fraction of shares outstanding owned by institutions located within 250 miles of the firm reported on 13(f) filings.	13(f)
ln(1+ITP)	ITP is defined as the sum of absolute dollar value of quarterly share changes by institutions, scaled by total dollar volume over the quarter.	13(f)

Variable Name	Description	Source
ln(1+Local ITP)	Local ITP is defined as the sum of absolute dollar value of quarterly share changes by institutions located within 250 miles of the firm, scaled by total dollar volume over the quarter.	13(f)
EW Distance to Retail	Equal weighted distance between retail shareholders in the brokerage database and firm headquarters location, in thousands of miles.	Brokerage
EW Distance to Institutions	Equal weighted distance between institutional shareholders in the 13(f) institutional holdings database and firm headquarters location, in thousands of miles.	13(f)
PIN	Probability of informed trading as in Easley, Hvidkjaer, and O'Hara (2002).	ISSM/TAQ
Stock Characteristics		
Stock price	Stock price.	CRSP
Volatility	Total volatility (standard deviation) of daily returns measured over the year.	Estimated, CRSP
Skewness	Total skewness of daily returns measured over the year.	Estimated, CRSP
Lottery Stock Dummy	1 if a stock is above median in both idiosyncratic skewness and idiosyncratic volatility and below median in share price.	Estimated, CRSP
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 $(Score - 3)/(60 - 3)$.	Estimated, CRSP
ln(Firm Size)	Natural log of Price $\times$ Shares outstanding (in millions).	CRSP
Market-To-Book (M/B)	Ratio of market value of equity to book value of equity.	Compustat
Leverage	Total debt in current liabilities plus total long-term debt, divided by total assets.	Compustat
3-Year R&D Expenditure	3-year average of R&D expenses.	Compustat
3-Year Advertising Expenditure	3-year average of advertising expenses	Compustat
3-Year ROA	3-yr average of ROA	Compustat
Dividend Yield	Total dividends paid in the prior year divided by price at the end of the prior year	CRSP
Past 12-Month Return	12-month stock return over the prior year.	CRSP
Monthly Turnover	Average monthly share turnover (share volume/shares outstanding) over the prior year.	CRSP
ln(Firm Age)	Natural log of the number of months since the stock appeared on CRSP.	CRSP

Variable Name	Description	Source
Regional Characteristics		
ln(Num Firms in MSA)	Natural log of the number of firms located within the MSA.	Compustat
Social capital index	Social capital index in the MSA nearest to the firm headquarters location; From Robert Putnam's Bowling Alone book.	Putnam
Urban Dummy	Ratio of population in living in urban areas in the county where firm headquarters is located.	Census
Industry Cluster Dummy	MSA level measure, equals 1 if 10% or more of the market capitalization of firms located in the MSA are from a single industry, and 10% or more of that industrys market capitalization is located in that MSA.	CRSP, Compustat
Catholic-Protestant Ratio	Ratio of Catholic population to Protestant population in the county where the firm headquarters is located.	ARDA
Local Market Participation Index	Stocks are assigned to vigintiles by per capita income, education, urbanicity, median age, and minority population in the county where the firm is headquartered (age and minority enter negatively so that 20 indicates the lowest levels of age and minority population). The income, education, urban, age, and minority vigintile assignments are added for each stock to produce a score ranging from 5 to 100, which is then scaled to range from 0 to 1 using $(Score - 5)/(100 - 5)$.	Estimated, Census