Weather-Induced Mood, Institutional Investors, and Stock Returns*

William N. Goetzmann
Yale University

Dasol Kim
Case Western Reserve University

Alok Kumar
University of Miami

Qin Wang
University of Michigan at Dearborn

First Draft: September 4, 2013
This Draft: September 17, 2013

Abstract

This study shows that weather-based indicators of mood impact perceptions of mispricing and trading decisions of institutional investors. We use survey and disaggregated trade data to show that relatively cloudier days increase perceived overpricing in both individual stocks and the Dow Jones Industrial Index, and increases selling propensities of institutions. We introduce stock-level measures of investor mood, and demonstrate that investor pessimism negatively impacts daily stock returns, mostly among stocks that are costly to arbitrage. Finally, we document comovement in stocks experiencing similar changes in investor mood. These findings complement existing studies on the weather effect on stock index returns, and identify an additional channel through which it can manifest.

---

* Please address all correspondence to Dasol Kim, Department of Banking and Finance, Weatherhead School of Management, Case Western Reserve University, Cleveland, OH 44106-7235; *Email:* dasol.kim@case.edu. William Goetzmann can be reached at william.goetzmann@yale.edu, Alok Kumar can be reached at akumar@miami.edu, and Qin Wang can be reached at qinw@umich.edu. We thank Jawad Addoum for helpful discussions and valuable comments. We are responsible for all remaining errors and omissions.

Weather-Induced Mood, Institutional Investors, and Stock Returns

Abstract

This study shows that weather-based indicators of mood impact perceptions of mispricing and trading decisions of institutional investors. We use survey and disaggregated trade data to show that relatively cloudier days increase perceived overpricing in both individual stocks and the Dow Jones Industrial Index, and increases selling propensities of institutions. We introduce stock-level measures of investor mood, and demonstrate that investor pessimism negatively impacts daily stock returns, mostly among stocks that are costly to arbitrage. Finally, we document comovement in stocks experiencing similar changes in investor mood. These findings complement existing studies on the weather effect on stock index returns, and identify an additional channel through which it can manifest.

1. Introduction

A number of recent studies document a robust relation between weather patterns in major financial centers and stock index returns, providing indirect evidence of the role of investor mood on asset prices.[1] However, less is known in whether and how mood fluctuations impact investors that play a key role in price formation: institutional investors. Mood fluctuations can have a persistent effect on asset prices through their impact on cognitive processes. Directly identifying how these processes are affected by mood is difficult due to data limitations. Examining the impact of weather patterns on actual perceptions of investors is important for establishing the plausibility of the weather effect in the existing evidence. Kramer and Weber (2012) and Bassi, Colacito, and Fulghieri (2013) provide experimental evidence to illustrate that weather impacts cognitive processes related to financial decision-making in non-professional subject pools. Whether such findings can be generalized to professional investors is the focus of this study.

Despite commonly-held assumptions about investor sophistication, recent studies document evidence of cognitive biases among professional investors. Coval and Shumway (2005) document trading patterns that are consistent with loss aversion amongst CBOT traders. Frazinni (2006) finds evidence of the disposition effect amongst U.S. mutual fund managers, while Barber, Lee, Liu, and Odean (2007) show consistent evidence in a broader set of investors using international data. More closely related to this study, Goetzmann and Zhu (2005) provide evidence suggesting that New York City weather patterns affect NYSE market makers, while showing little effect on retail investors. These findings are consistent with predictions made in Bodoh-Creed (2013), who shows that informed market agents are more likely to be susceptible to mood-related biases.

This study matches zip code-level fluctuations in weather to a variety of data sources. We employ survey data from the Yale International Center for Finance that captures institutional investor perceptions of stock market investment. From this data, we construct various measures of perceived mispricing on individual stocks as well as the Dow Jones Industrial Index. Additionally, we also use a disaggregated database that contains daily, institutional investor

[1] These studies include Saunders (1993); Hirshleifer and Shumway (2003); Kamstra, Kramer, and Levi (2003); and Goetzmann and Zhu (2005). Hirschliefer and Shumway (2003) and Kamstra, Kramer and Levi (2003) provide evidence on an international panel data of equity market indices, reflecting the resilience of these findings across country borders.

trades to assess the impact of weather on trading decisions. Finally, we construct a novel, stock-level measure of investor mood that captures weather patterns across locations of investors using institutional holdings data.

Our findings are summarized as follows. First, relatively cloudier days increase perceived overpricing among institutional investors, and increase the likelihood of perceived mispricing in both individual stocks as well as the Dow-Jones Industrial Index (DJIA). A standard deviation change in cloud coverage explains approximately 8% of the total sample variation in perceived DJIA overpricing. Second, using daily trading data on institutional investors, we show that relatively cloudier days increase the propensity of institutional investors to sell. The impact of weather patterns on perceptions of mispricing and trading activities is significant both statistically and economically. Differences in trade imbalance between the top and bottom 10th percentile in cloud coverage represents close to 7% of the total sample variation.

In the last part of the paper, we assess how stock-level measures of investor mood relate to individual stock returns. Because U.S. stock ownership is generally dispersed across investors in different geographical regions, tests using weather as a proxy for mood can suffer from low power due to identification issues, especially when only using weather in firm headquarter locations. We construct relatively cleaner proxies for investor mood using institutional holdings data, aggregating information on weather patterns of investors located in different regions.

Using this new stock-level measure of investor mood, we find that investor pessimism has a negative impact on daily stock returns. The results are statistically insignificant in the pooled sample, but the weather effect is larger and statistically significant among stocks with higher arbitrage costs. Additionally, we document excessive returns comovement related to changes in investor mood. We find that these effects are short-lived, and mostly disappear after three months.

These findings contribute to a growing literature in finance and economics that identifies channels through which mood affects economic and stock market activity.[2] A large number of these studies focus on channels related to sunlight and weather conditions.[3] These studies are

[2] Edmans, Garcia, and Norli (2007) and Agarwal, Duchin, and Sosyura (2012) examines sporting event and singing competition outcomes. Dougal et al. (2012) and Garcia (2013) examines word connotations in financial journalism.
[3] An abbreviated set of studies includes Saunders (1993); Hirshleifer and Shumway (2003); Kamstra, Kramer, and Levi (2003); Goetzmann and Zhu (2005); Lo and Wu (2011); and Chhaochharia, Korniotis, and Kumar (2012).

motivated by findings in the psychology literature of how affective states can modify human emotion and create biases in evaluating risk-based decisions.[4]

Investor behavior may be subconsciously influenced by factors that affect emotion, and may in turn impact financial market performance under certain conditions (Mehra and Sah, 2002 and Bodoh-Creed, 2013). These weather-related triggers may be powerful enough to affect the marginal investor. Mehra and Sah (2002) demonstrate that even small perturbations in preference parameters related to mood can have measurable impact on stock returns and volatility.

Recent experimental studies show that mood-inducing cues related to weather can have a sizable impact on risk preferences and decision making (Kramer and Weber, 2012; Bassi, Colacito, and Fulghieri, 2013). In particular, Bassi, Colacito and Fulghieri (2013) provide experimental evidence directly showing how weather has a sizable impact on risk tolerance. They find that individual risk aversion positively relates to observable and subjective weather conditions in their sample of student participants. Kramer and Weber (2012) present experimental evidence on the impact of Seasonal Affective Disorder (SAD) on risk attitudes. They find that subjects suffering from SAD exhibit significantly larger risk aversion than those who do not. Further, they find that the effect of SAD on risk aversion is transmitted through depression.

However, experimental evidence is based upon participants from the general population, and professional investors may not necessarily be subject to the same dynamics.[5] We offer direct examination of how weather conditions impact professional investor beliefs using survey data, and document consistent evidence using actual trading and returns data.

Other studies examine the effect of weather on investor trading decisions. In particular, Loughran and Schultz (2004) and Goetzmann and Zhu (2005) employ disaggregated data at different geographical locations to examine the effect of weather on trading activities of investors. Variation in localized weather patterns across geographies provides arguably better statistical power and identification of weather patterns to specific investors. Loughran and Schultz (2004) show that relatively cloudier days in a firm's location has little impact on the

[4] Schwartz (1990); Clore and Parrott (1991); Wilson and Schooler (1991); and Clore, Schwarz and Conway (1994) discusses the general role of mood and emotion in decision making. Loewenstein (2000) and Loewenstein et al. (2001) discusses how emotion can be misattributed as information.

[5] For example, see Camerer et al. (1997) for violations of the law of supply on wage data for New York taxi drivers.

firm's trading volume outside of extreme weather conditions, which can be attributable to other factors that may be unrelated to mood.

Goetzmann and Zhu (2005) match retail investor trade data to weather conditions in major cities from 1992 through 1995. They show that weather fluctuations have little impact on investor propensity to buy and sell, but they find a significant relation between bid-ask spreads and index returns on NYSE stocks with weather conditions in New York City. They interpret these findings as evidence that the weather effect has a strong impact on market-makers and other market agents, who are concentrated in New York and contribute more to aggregate trading volume and price formation.

Direct tests on institutional investors have not been formally examined in the literature. Goetzmann and Zhu (2005) suggest that the weather effect may have a stronger impact on institutional rather than retail investors. Using daily trading data on institutional investors, we show that relatively cloudier days negatively impact buy-sell trade imbalances, and the result is robust to a number of alternative explanations. Additionally, we are able to document similar patterns when only allowing for variation across investors within the same stock and date.

Our findings on investor perceptions and trading decisions suggest that mood should also impact price formation, as institutional investors play an important role in price formation. To that end, we construct a new stock-level measure of investor mood using weather fluctuations in locations of investors holding the stock. Because daily, individual stock returns are likely to be affected by liquidity issues, we base our tests upon a dummy variable that corresponds with positive stock return days, similar an approach taken in Hirshleifer and Shumway (2003).

Measurement errors in our stock-level measure are a potential concern given that we can only observe quarterly snapshots of institutional investor holdings. We cannot identify investors who actively trade a stock but do not have any holdings around the SEC filing dates, which may be associated with non-systematic measurement bias. In particular, investors who do not hold long positions in the stock due to bearish opinions cannot be observed in the data, and can bias the test coefficients in the opposite direction of our predictions. Despite these concerns, we find that relatively pessimistic stocks are more likely to experience negative returns days. Together, we regard our results as conservative estimates of the weather effect on individual stocks.

Our findings suggest that sky cloud coverage has a strong impact on trading decisions of institutional investors and also affects the price formation process. This evidence complements the results from existing studies. In particular, we document the weather effect in investor beliefs and trading volume amongst institutional investors. Together with the evidence in Goetzmann and Zhu (2005), this finding suggests that the weather effect is pronounced amongst informed market participants and influences the price formation process.

2. Data and Summary Statistics

2.1 Main Data Sources

Our empirical analysis employs data from several sources. First, respondent-level, institutional investor survey data are collected from the Investor Behavior Project at Yale University. Since 1989, questionnaire survey data have been collected on the perceptions of investors in the U.S. about stock market investment. The respondent-level data are used to construct an aggregate confidence index for both professional money managers and wealthy individual investors.

Respondents are randomly sampled from a directory of institutional investors found in the investment managers section of the "Money Market Directory of Pension Funds and Their Investment Managers."[6] Approximately 100 professional investors are surveyed per month. The respondent data are available from January 2005 to February 2007. To ensure consistency in the data, only survey responses that have responses to most of the survey fields are included in the analysis. Altogether, there are 1,543 "clean" observations in our sample.

Second, the weather data are collected from the Integrated Surface Database (ISD) which is publicly available from the National Oceanic and Atmospheric Administration web site (www.ncdc.noaa.gov). The ISD database records hourly, weather observations from over 20,000 active and inactive weather stations worldwide, and 7,610 weather stations within the US. The dataset includes a number of fields collected from each station including weather fields, such as sky cloud coverage, as well as the coordinate location of the station.

Third, the institutional daily trading data are provided by ANcerno Ltd. (formerly the Abel Noser Corporation). The sample period is from January 1999 to December 2010. ANcerno is a widely recognized consulting firm that monitors equity trading costs of institutional investors such as CalPERS, Putman Investments, and Lazard Asset Management.[7] Accompanied with the client manager code, an institutional client code allows identification of the investor. Additional

---

[6] Additional information about the survey can be found at the International Center for Finance website (icf.yale.edu).
[7] Puckett and Yan (2011) provide a detailed description of the ANcerno data.

fields used in the analysis include the stock historical CUSIP number, trade date, trade direction (buy or sell), quantity of shares traded, and trade execution price.

There are two unique features of the ANcerno data that are central to the trade-related tests. First, the ANcerno dataset provides the true trade direction (i.e., buy/sell) for all executed trades, eliminating the need to rely on the Lee and Ready (1991) algorithm to infer a trade's direction. So, our study is free of the worry about the accuracy of inferences about trade direction. Second, the ANcerno dataset provides the names of the institutions. Using the institution names, we hand collect the zip code of each institution's location. We then merge this information with the weather database to obtain the weather patterns at institutional locations. The majority of the funds in the ANcerno database are located in U.S., which are used in our analysis. For these U.S. funds, we are able to match the locations of almost 80 percent of all trades in the ANcerno database.

Fourth, the 13f institutional holdings data are collected from Thompson Reuters for the 1999Q1 to 2010Q4 sample period. The data provide quarterly snapshots of institutional investor positions. The ANcerno database only provides trading information for a subset of investors, while the 13f data provide a more complete picture of institutional investor holdings in a particular stock.

Other data sources used in the analysis are as follows. Stock characteristics are collected from the Center for Research in Security Prices (CRSP). Only common stocks (share code of 10 or 11) from January 1999 to January 2010 are included in the analysis. Sentiment variables constructed in Baker and Wurgler (2004) are collected from Jeffrey Wurgler's web site (http://people.stern.nyu.edu/jwurgler/), which include: IPO volume from Ibbotson, Sindelar, and Ritter (1994) and updates; average first-day returns on IPOs from Ibbotson, Sindelar, and Ritter (1994) and updates; average closed-end fund discount from Herzfeld; the equity share in new issues defined following Baker and Wurgler (2000); and New York Stock Exchange (NYSE) monthly turnover from the NYSE Factbook. The daily values of the VIX index are collected from the Chicago Board Options Exchange (CBOE) web site available at www.cboe.com. The county-level estimates of median household income and population are obtained from the Bureau of Economic Analysis.

Figure 1 presents the geographical distribution of the institutional investors represented in the survey (Panel A), trade (Panel B), and holdings (Panel C) datasets at the county-level. Color-coded counties correspond with regions where at least one investor is represented. All three panels convey considerable geographical heterogeneity in each of the datasets. Panel A shows that the survey respondents are well-represented geographically, and are more heavily represented in regions with higher population. Panel B shows a similar pattern within the trade dataset, though this geographical distribution is relatively sparse relative to the survey data. This evidence is not surprising, given the limited number of investors in the trade database. Panel C shows that the holdings data have greater heterogeneity than the trade data, which is expected since investors in the trade data represent a subset of investors in the holding data.

2.2 Variable Construction

The primary variable of interest from the weather data are the hourly, sky cloud coverage, which takes on integer values from zero (sky clear) to eight (full cloud coverage). Similar data are also used in Hirshleifer and Shumway (2003) and Goetzmann and Zhu (2005).

The investor survey and trade data are merged with the weather data by geography. Because the weather station data present location by location coordinate, the distance between each station and the investor locations is calculated based upon the coordinates of the centroid of the investors' zip codes using the Haversine distance formula. A simple matching criterion could involve choosing matches that minimize pairwise distances. However, the weather stations during the sample period do not necessarily operate over the entire sample period. Further, while the exact location of the weather station is observable, we can only observe zip code level values for the institution's location.

The raw sky cloud coverage measure requires adjustment to be used in the analysis. The average, hourly sky cloud coverage from 6 am to 12 pm is calculated to provide a single value for each day and weather station. For each zip code, the average daily, sky cloud coverage is calculated using all weather stations within a 50 kilometer radius of the zip code centroid.[8] Because the effect of cloudiness on individual mood is likely to take hold after long periods of persistent cloudiness, a rolling average is taken from *x* days before the response or trade date to one day prior for each zip code. Finally, the average amount of sunlight is a decreasing, convex function in sky cloud coverage. As a result, the sky cloud coverage measure used in the analysis (*SKC*) is defined as the natural log of one plus the rolling average of the zip code-level sky cloud

[8] We also examine alternative distance thresholds. The results in the analysis remain similar when using a 30 kilometer threshold, though weaken when increasing the distance to 100 kilometers.

coverage. Partial observations are discarded from the computation. At least one, matched, weather station is required for an institutional investor in either samples to be included in the analysis.

To control for seasonality in the *SKC* measure, the analysis adopts three approaches that help mitigate its effects. First, the primary specifications include, in addition to *SKC*, historical sky cloud coverage, as defined as the natural log of one plus the average, daily sky cloud coverage for the same month in the previous year, or *LastSKC*. Second, the *SKC* measure is substituted with the difference between *SKC* and *LastSKC*, or *DSKC*. The *DSKC* measure is a restrictive measure, as it implicitly, in an OLS regression models, constrains the coefficient estimate on *LastSKC* to be equal to negative of the *SKC* coefficient estimate. Last, fixed effects estimators based on time measures, such as year-quarter or date units, are also estimated for comparison for some of the specifications.

Buy-sell imbalance (*BSI*) using the institutional trade data are constructed similarly to Goetzmann and Zhu (2005), and is defined as the difference between the daily, dollar buy and sell volume across all investors in the dataset in a particular zip code, scaled by the average, daily, total dollar volume over sample period for investors in the same zip code. Where there is no trade on a particular date, neither a buy nor sell, the zip code level *BSI* takes a value of zero. Daily investor-stock level *BSI* is constructed similarly, and is defined as the daily, net buy minus sell dollar volume scaled by the average, daily dollar volume over the sample period for the same zip code.

Aggregating *BSI* on the zip code level across investors helps neutralize idiosyncratic trading behavior by a single investor, which serves to add noise to the *BSI* measure. Because some of the zip codes only include a few investors for certain parts of the sample period, the analysis restricts the sample to only zip codes with at least three investors at each point in time.

2.3 Summary Statistics

Table 1 provides detailed summary statistics of the weather, survey, and trade variables used in the analysis. Panel A describes the zip code level sky cloud coverage measures across different estimation windows. The estimation window of *x* days calculates average sky cloud coverage using data from days *t-x-1* to *t-1*. Because there is little theory to help guide the selection of the estimation window used in the analysis, the comparison helps pinning down an appropriate length that yields well-behaved estimates. As the estimation window increases, the sample average decreases slightly, though not monotonically. The reduction in the sample standard deviation also increases significantly, where the magnitude of the reduction stabilizes from a two week window. In particular, the coefficient of variation using a 1 day window, which is used in Shumway and Hirshleifer (2003), is 0.866 (e.g. 3.252/3.755). The coefficient of variation decreases considerably up to the two week window (0.387), and remains stable up to the four week window (0.353).

Panel B describes the primary variables used in the survey-based tests. SKC is the primary sky cloud coverage measure used in the analysis, and is the natural log of the average sky cloud coverage using a two week estimation window. The summary statistics displayed are on the matched survey sample. DSKC is the difference between SKC and the natural log of the average sky cloud coverage in the same month of the previous year. In addition to portfolio size and county-level characteristics, the panel reports the summary statistics for the primary variables used in the empirical analysis.

Panel C provides descriptions on the trade sample. The SKC and DSKC measures constructed over the trade sample have similar distributional properties as those constructed over the survey sample, though the standard deviations are slightly smaller. The variable of interest is described further in the trade section below.

Finally, Panel D provides descriptions on the holdings sample. The stock-level SKC and changes in stock-level SKC measures are constructed as the average SKC of investors in the same stock at each point in time. Please refer to subsequent sections for further descriptions on the construction. The weather variables have sample means as the trade sample, though the standard deviations are slightly lower, which in part is due to the construction procedure.

3. Weather-Induced Mood and Perceived Mispricing

Experimental evidence shows that weather conditions have a sizable impact on measures of investor risk preferences. For example, Bassi, Colacito, and Fulghieri (2013) infer risk tolerance through experiments designed by Holt and Laury (2002). The survey data do not collect information that can provide relatively clean proxies for risk preferences, though the dataset does include other information that captures investor opinions about investment in the stock market.

Our primary tests focus on perceived investor mispricing. Investors may view stocks as mispriced for a variety of reasons, which include the investor's risk preferences. If variation in weather patterns generates mood-induced biases in investor beliefs about the underlying fundamentals of the economy, then it may affect the investor trading decisions as well. Negative moods may induce investors to examine information with greater scrutiny (Schwarz, 1990; Petty, Gleicher, and Baker, 1991), and may incline investors to view stocks as overvalued. On the other hand, good moods may incline investors to believe that stocks are priced appropriately, or even undervalued, relative to fundamentals (Clore, Schwarz, and Conway, 1994; Forgas, 1995).

3.1 Survey Based Measures of Perceived Mispricing

Towards that end, the analysis focuses on responses from the survey related to perceived stock mispricing. First, to directly assess whether investors believe that stocks are overpriced relative to fundamentals, the analysis begins by examining the association between the weather measures and responses to the question, "Stock prices in the United States, when compared with measures of true fundamental value or sensible investment value, are: (a) Too low, (b) too high, (c) just right, and (d) do not know". An indicator variable, *D(High)* is constructed, which takes the value of one to a response of "too high", and zero otherwise.[9]

Second, a more general question on whether stock prices *can* deviate from their fundamental values comes from the survey question, "In your opinion, how likely is it that the price of an individual stock in this market is higher or lower than its true value? Would you say this is (a) Definitely, (b) Very optimistic, (c) Somewhat optimistic, (d) Not too optimistic, (e) Impossible". An indicator variable, *D(MisPrc)* is constructed to denote likely mispricing, which assumes the value of one for responses of "Definitely", "Very optimistic", and "Somewhat optimistic", and zero otherwise.

Third, the degree of perceived mispricing among survey respondents is constructed from reported values of what they believe to be the intrinsic level of the Dow Jones Industrial Average (DJIA). Analyzing *D(High)* and *D(MisPrc)* requires the respondents to have a consistent approach to mapping the degree of mispricing to the survey responses. However, this is unlikely to be true. Furthermore, the wording of the possible responses to the question related to *D(MisPrc)* may also bias some respondents as they are not necessarily in neutral tone, though the

[9] Altering the definition to include "do not know" slightly strengthens the results in the main analysis.

direction of the bias is unclear. Both of these factors may introduce noise in the measures, reducing the power of the corresponding tests.

To address these issues, a continuous measure of perceived mispricing is constructed based upon their responses to the survey question: "What do you think would be a sensible level for the Dow Jones Industrial Average based on your assessment of U.S. corporate strength (fundamental) [sic]?". The responses provide estimates of the investors' perceptions on the intrinsic value of the DJIA, and can be linked to the actual DJIA level around the survey date. The measure is constructed as the percentage mispricing (*%MisPrc*) as defined as the natural log of the ratio between the investor-supplied response to the average DJIA level over the past seven days. Lower values of *%MisPrc* are associated with relatively greater degrees of overpricing.

%MisPrc is expected to be negatively related to *D(High)*, and indeed a paired *t*-test in *%MisPrc* across values of *D(High)* is statistically significant (*t*-value = 28.81). Aside from greater sample variation in *%MisPrc*, its relation to *SKC* may reveal additional information, as the measure is specifically related to overpricing in larger stocks.

3.2 Perceived Mispricing Regression Specification

The models of investor mispricing is specified as follows:

$$\mathrm{Y_{i,t}} = b_0 + b_1 * \mathrm{SKC_{i,[t-x,t]}} + b_2 * \mathrm{LastSKC_{i,t}} + \mathbf{b} * \mathbf{X_{i,t}} + \varepsilon_{i,t}$$

The dependent variables in the regression models are *D(High)*, *D(MisPrc)* and *%MisPrc*. As mentioned before, the $SKC_{\mathrm{i,[t-x,t]}}$ is constructed as a *x*-day rolling average using information up to date *t* of zip code-level sky cloud coverage. $\mathbf{X}_{\mathrm{i,t}}$ represents a vector of other explanatory variables for respondent *i* at date *t*.

Our choice of the control variables is motivated by Chhaochharia, Korniotis, and Kumar (2012) who show that weather can affect regional economic conditions, which in turn may influence equity investment opinions. Accordingly, proxies for local economic conditions are included, such as county-level population, changes in population, county-level median income, and changes in median income.

Professional investors with greater assets under management may have systematically different opinions about equity market investment, and so the natural log of one plus the respondent's investment portfolio is included in the regression models. Large, recent fluctuations in stock prices may also influence responses, as periods of greater equity market volatility may negatively bias investor responses. Accordingly, the natural log of the volatility on DJIA returns over the past 30 days is included as a conditioning variable. Finally, investor responses to other questions, such as the respondent's estimate of the equity risk premium, long-term corporate earnings growth, and likelihood of a catastrophic stock market crash, are included, as they may systematically relate to the dependent variables of interest.

SKC is expected to have a positive impact on *D(High)*, as well as perceived mispricing, or *D(MisPrc)*. *SKC* captures incremental variation in cloudiness relative to historical weather conditions, as *LastSKC* included as a conditioning variable. The models are estimated using OLS estimators. For the binary response dependent variables, the model estimates from probit estimators are also reported, taking the form:

$$\mathrm{P}(\mathrm{Y}_{\mathrm{i,t}} = 1) = \Phi\left(c_0 + c_1 * \mathrm{SKC}_{\mathrm{i,[t-x,t]}} + c_2 * \mathrm{LastSKC}_{\mathrm{i,t}} + \mathbf{c} * \mathbf{X}_{\mathbf{i,t}}\right)$$

Y represents the indicator variables associated with perceived mispricing. In either specification, the standard errors on the parameter estimates are likely to be biased as the residuals are unlikely to be independent given multiple responses for each date across regions and zip codes across time, and so two-way clustered standard errors on the zip code and date levels are reported for both the OLS and probit point estimates. Observations subject to data inputting errors or which represent extreme outliers in the *%MisPrc* measure are removed from the sample.

3.3 Perceived Mispricing Regression Estimates

Our empirical analysis begins with estimating OLS regression models including only the weather measures across different estimation windows for *SKC* and *DSKC*. As mentioned above, we have little guidance on how to choose the duration of the estimation window. The variation in sky cloud coverage decreases in the length of the estimation window, as shown in Table 1. Because the standard deviation of the *SKC* measure begins to stabilize approximately after two weeks, extending the estimation window beyond that point is expected to lead to incremental changes in the model estimates.

Table 2 displays the model estimates. Panels A, B, and C display the results from the overpricing, mispricing, and DJIA percentage mispricing models, respectively. Models 1 through 5 display the results using estimation windows of one day, one week, two weeks, three weeks and four weeks, respectively. As with all subsequent tests using the survey data, standard errors are adjusted using two-way clustering on response date and respondent zip code.

Across Panels A through C, the one day estimation window estimates are statistically indistinguishable from zero. However, as the estimation window lengthens, the *SKC* coefficients increase in absolute magnitude in most of the models, and are all statistically significant at least at the 10% level for estimation windows of at least two weeks, controlling for the previous year's *SKC*. Past two weeks, the adjusted R^2 does not appear to increase reliably across the specifications.

The estimation window for the *SKC* measure for the remaining analysis is defined as the most recent 14 days before the observation date. The 14-day window is chosen to stabilize the *SKC* measure while ensuring that the estimation window is not so long that our findings may be attributed to other factors such as seasonality. When we repeat the analysis using a four-week window, the results are qualitatively similar.

Table 3 repeats the analysis in Table 2 across different specifications to assess the stability of the point estimates when we consider conditioning variables that are expected to explain variation in the dependent variables using the 14-day estimation window for SKC. In Panel A, the dependent variable in the regression model is a measure that captures whether the respondent believes that current stock prices are too high relative to fundamentals overall.

The point estimates remain similar across all the OLS specifications in models 1 through 3. In model 1, the *SKC* coefficient is positive (estimate = 0.076, *t*-value 2.23), and very similar to the 2-week estimate in Panel A of Table 2. Inclusion of the other survey responses in model 3 may introduce proxy control issues, as *SKC* may be related to the dependent variable through its impact on these factors. However, the SKC point estimates remain similar (estimate = 0.077, *t*-value 2.26) with their inclusion. In economic terms, a standard deviation increase in *SKC* is associated with a 0.027 increase in the dependent variable.

When we use an alternative specification using a probit estimator, we find similar results. In model 5, the *SKC* coefficient is positive (estimate = 0.323, *z*-value = 2.27), while the partial effect of *SKC* is 0.086 (*z*-value = 2.34). In this instance, a standard deviation increase in *SKC* leads to a 3.21% increase in the probability that an investor perceives the market to be mispriced.

Panel B of Table 3 examines models in which the dependent variable is an indicator variable for whether the respondent believes that mispricing in the market exists. The results are formatted similarly to those in Panel A. Inclusion of the additional conditioning variables yields similar point estimates on the *SKC* coefficient. In model 3, where the full set of conditioning variables is included, the *SKC* coefficient (estimate = 0.102, *t*-value 3.52) is 50% larger than that of model 3 in Panel A. This evidence suggests that the wording of the question used in Panel A may introduce inconsistent evaluation schemes across the respondents. The *SKC* coefficient estimates from the probit model are qualitatively similar. For example, the partial effect of *SKC* in model (5) is 0.101 (*z*-value = 2.94), and a standard deviation increase in *SKC* leads to a 3.34% increase in the perceived probability of mispricing.

Panel C of Table 3 uses the estimated percentage mispricing given their estimated, instrinsic values for the DJIA on the response date, and models 1 through 3 display the estimates. Model 4 replaces the dependent variable with dummy variable that takes the value of one if the percentage mispricing is above the top 25th percentile of the sample, which would imply that the respondents believe the actual DJIA level is very low. Models 5 through 8 replace the dependent variable with a dummy variable that takes the value one if the percentage mispricing is below the bottom 25th percentile of the sample, which would imply that the respondents believe that the actual DJIA level is very high relative to their estimated, intrinsic values. Models 7 and 8 use probit estimators to obtain the model estimates, while the other models use OLS.

Models 1 through 3 display the *SKC* estimates, and show that the coefficients are relatively stable. For model 3, the *SKC* coefficient is negative (estimate = -0.027, *t*-value = -2.70), as expected. A standard deviation increase in *SKC* yields a decrease of 0.010 in the dependent variable, which represents 8.33% of its sample standard deviation and is economically meaningful.

In model 4, the *SKC* coefficient is negative as well (estimate = -0.054, *t*-value = -1.17), though not statistically significant at the 10% level. In models 5 and 6, the *SKC* coefficient is positive, as expected, and statistically significant at the 1% level. The point estimates in models 5 (estimate = 0.122, *t*-value = 3.56) and 6 (estimate = 0.123, *t*-value = 3.62) are more than 50% larger than model 3 of panel A.

Fitting similar models using probit estimators in specifications 7 and 8, we find that the *SKC* coefficient estimate remains positive. The partial effect of *SKC* in model 8 is 0.133 (*z*-value = 3.17), and a standard deviation increase of *SKC* leads to an increase of 4.95% in the probability that investors perceive greater mispricing.

Interestingly, most of the explanatory power of *SKC* in the percentage mispricing model comes from the respondents who believe the actual DJIA levels are too high relative to fundamentals. The results are consistent with the notion that investors in very bad moods (i.e., high *SKC* values) are more likely to believe the market valuations are too high relative to fundamentals due to heightened scrutiny to details, as conjectured in Hirsheliefer and Shumway (2003). The *SKC* coefficients in models 5 through 8 are also substantially larger than those in Panel A, suggesting that inconsistency in evaluation of extreme mispricing may have attenuated the point estimates in Panel A. When using a consistent standard for mispricing in Panel C, the point estimates become more pronounced. However, this may be due in part to differences in the wording of the questions, as the percentage mispricing measure is specifically related to DJIA levels.

3.4 Perceived Mispricing Regression Estimates: Robustness Checks

The estimators used in Tables 2 and 3 are susceptible to omitted variable issues related to inter-regional as well as seasonal factors. While the models include the previous year's SKC for the same month, so that the explanatory power of SKC is incremental to typical weather conditions to that region, the past SKC may be measured with error. Similar factors may be present related to seasonality in the investor responses, which may be associated with documented seasonal patterns in the stock market.

To guard against these potential biases, we use fixed effects at the state level in models 1 through 4 of Table 4. In models 5 through 8, fixed effects at the year-quarter-level are included. Due to sample size limitations, analysis with regional variables with greater granularity is infeasible, as the respondent locations are diverse. Similarly, the sample period includes nine quarters, with 114 to 190 responses per quarter. All models include market volatility, portfolio size, and county characteristics conditioning variables as well, though the coefficients are not reported to conserve space.

Models 1 through 4 show that the *SKC* coefficients are quite similar to those in Table 3. That is, limiting the variation in the survey responses to intra-regional variation across time yields similar point estimates. In models 5 through 8, the *SKC* estimates are also similar, though somewhat smaller in absolute magnitude for all specifications aside from model 6. In these specifications, the explanatory power in *SKC* is limited to intra-sample period variation across regions. All the *SKC* coefficients across the models are at least significant at the 10% level.

Finally, the analysis further assesses the robustness of the results using alternative cloud coverage measures used in the literature. Specifically, the previous tests are reproduced after replacing *SKC* and *LastSKC* with the difference between the two, or *DSKC*. The specification is more conservative, as the models implicitly constrain the *LastSKC* coefficient to be the negative of the *SKC* coefficient.

Table 5 presents the *DSKC* coefficient estimates across the investor responses. While the *SKC* coefficient in model 1 remains positive, it is no longer statistically significant at the 10% level (estimate = 0.57, *t*-value = 1.58). Outside of model 1, all the coefficients in the other specifications are consistent with the predictions and are statistically significant at least at the 5% level. Additionally, most of the specifications obtain an adjusted R^2 that is lower than those reported in Table 3. The partial effects are similar to those obtained in Table 3, though are somewhat lower in most specifications.

In summary, the evidence from the perceived mispricing regression estimates based on the survey data indicates that *SKC* is positively related to perceived overpricing. The results are robust to alternative specifications that account for potential issues with the survey data mentioned above. The economic magnitude of *SKC* in the percentage DJIA mispricing regression is quite sizable. In regressions with the binary response variables, the economic magnitude of *SKC* is more modest, though still economically meaningful.

4. Weather-Induced Mood and Institutional Investor Trading

The results from the previous section suggest that *SKC* has a strong impact on perceived mispricing amongst professional investors. This evidence suggests that *SKC* may also adversely impact their trading behavior. In particular, investors are likely to exhibit a greater propensity to sell during cloudy days. We construct tests using daily, institutional investor trading data to assess whether this is indeed the case.

This trading analysis follows the method used in Goetzmann and Zhu (2005) and considers the buy-sell imbalance (*BSI*) measure as the dependent variable in the trading regression models. Positive (negative) *BSI* relates to institutional investors that are net buyers (sellers) for the observation date. The daily trade data are aggregated to the zip code-level, which are used to construct the *BSI* measure. *SKC* and *LastSKC* are constructed identically to the procedure described in the survey-based tests.

4.1. Weather and Trading: Univariate Results

A simple way of assessing the impact of *SKC* is to examine differences in the average *BSI* across sunny and cloudy days. *SKC* is expected to have a negative impact on *BSI*, and the magnitudes are expected to increase in more extreme values of *SKC*. To that end, univariate regression models are estimated using zip code-level *BSI* as the dependent variable, with a dummy variable that takes the value one if *SKC* is in the top 50th percentile of the sample, or *D(HighSKC),* and an intercept term as explanatory variables. To explore relative differences in extreme *SKC* values, the univariate models are estimated on subsamples based upon unconditional *SKC* rankings, and *SKC* rankings conditional on date.[10]

Using *SKC* rankings over the entire sample, Panel A of Table 6 provides the results. Estimates on the intercept term are not reported. Models 1 through 4 present the estimates on sample subsets based upon the top and bottom 50th, 33rd, 25th and 10th percentiles of *SKC* over the entire sample, respectively.[11] Model 1 shows a modest difference in average *BSI* between the top (cloudy) and bottom (sunny) 50th percentile (estimate = -0.018, *t*-value = -1.125). In other words, the difference in average *BSI* in the cloudy minus sunny observations is -0.018. However, as the subsample is contracted to include increasingly extreme values of *SKC*, the magnitudes of the differences are monotonically larger. In model 4, the differences in the cloudy-sunny observations is -0.094 (*t*-value = -3.61), which is two-fold larger than that of model 3 (estimate = -0.045, *t*-value = -2.04).

The estimates in Panel A are susceptible to potential biases due to overweighting certain days of the year, as the *SKC* rankings are taken over the entire sample. In an alternative specification, *D(HighSKC)* uses *SKC* rankings within each date. Panel B of Table 6 presents the estimates of the univariate regression models. The estimates are comparable to those in Panel A. As before, the magnitude of the differences in average *BSI* increases in relatively extreme values of *SKC*. Assessing the magnitudes based upon the point estimates, the differences in average BSI across the top and bottom 10th percentile in *SKC* represent 6.35% (model 4 of Panel A) and 6.49% (model 4 of panel B) of the total variation in *BSI*.

[10] Note that testing the statistical significance of the differences is equivalent to calculating the paired *t*-statistics across the *SKC* rankings.

[11] The sample prohibits the analysis from examining more extreme values due to lack of variability in the tails of the empirical *SKC* distribution.

The next part of the trading analysis assesses the robustness of the results from these univariate tests. Our goal is to assess whether potential omitted variables are correlated with seasonal weather patterns, stock characteristics, and other factors affect our findings.

4.2 Multivariate BSI Regression Specification

The multivariate *BSI* model estimated over the full sample is specified as follows:

$$\mathrm{BSI}_{\mathrm{z,t}} = d_0 + d_1 * \mathrm{SKC}_{\mathrm{z,[t-14,t]}} + d_2 * \mathrm{LastSKC}_{\mathrm{z,t}} + \mathbf{d} * \mathbf{W}_{\mathbf{z,t}} + \varepsilon_{i,t}$$

For zip code *z* and day *t*, $\mathbf{W}_{\mathbf{z,t}}$ represent a vector of variables that will help isolate the sources of explanatory power on *SKC*. January and Monday dummy variables are included in the models, as in Goetzmann and Zhu (2005). County-level economic variables as described in the survey-based tests are also included. Sentiment-based measures are from Baker and Wurgler (2007), and include: value-weighted dividend premium, rolling count of IPO volume, rolling average of first-day returns on IPOs, rolling average of the closed-end fund discount, equity share in recent, new issues, and recent NYSE monthly turnover from the NYSE Factbook. Kaplanski and Levy (2009) find that sunlight explains the VIX index, as it relates to perceptions on market volatility, and so this variable is included in the set of conditioning variables. In addition to the specification described above, an alternative specification that replaces *SKC* and *LastSKC* with *DSKC*, as described above, is also estimated to facilitate comparisons with the results in the existing literature.

With the larger dataset, the analysis can feasibly exploit intra-zip code and intra-date variation in the *SKC* variable using fixed effects estimators. Fixed effects at the zip code-level help address concerns that time-invariant, regionals factors not properly accounted for in the *BSI* construction. The *SKC* coefficients in the pooled estimator may relate to other factors not sufficiently accounted for in the *LastSKC* conditioning variable, such as seasonal effects, which can be addressed by using fixed effects on the date level.

In the last set of tests, the dataset is disaggregated further into a three-level panel, where the data are aggregated to the zip code (*z*), date (*t*), and stock (*j*) level. *BSI* is reconstructed as the daily, net buy-sell dollar trading volume for each zip code, date and stock, scaled by the average, total daily dollar trading volume for the zip code across the sample period.

The three-level panel is expected to introduce additional noise in the *BSI* variable, as variation in *BSI* may be driven by idiosyncratic trading motives that the aggregation procedure cannot eliminate. Additionally, because the investor's daily positions are not known in the dataset, the tests are conditional on whether the investor trades (e.g. either buys or sells) in the stock for a particular date.

The analysis repeats the date fixed effects model on the three-level panel. This specification allows the estimator to account for individual stock characteristics that are not possible with the two-level panel tests. Specifically, the model also includes the natural log of the stock's market capitalization and the inverse of the stock's share price as conditioning variables.

Finally, a model with fixed effects for each date-stock pair is also estimated. In these tests, the variation in *SKC* is limited to intra-stock-date groupings, allowing the estimator to conditioning for unobservable stock-level factors at each point in time. This specification is expected to significantly reduce power in the tests, and so can be viewed as conservative. The model specification is as follows:

$$\mathrm{BSI}_{\mathrm{z,j,t}} = f_0 + f_1 * \mathrm{SKC}_{\mathrm{z,[t-14,t]}} + f_2 * \mathrm{LastSKC}_{\mathrm{z,t}}$$
$$+ \sum_j \sum_t f_{j,t} * D(Stock = j \bigcap Date = t) + \varepsilon_{i,t}$$

Because the residuals in the model are unlikely to be independent across the panel dimensions, standard errors reported in the results are adjusted for three-way clustering at the zip code-, date- and stock-levels.

4.3 BSI Regression Estimates

The analysis begins by estimating the baseline *BSI* regression model using only the weather variables. Additional conditioning variables are added to assess the stability of the *SKC* coefficients. A summary of the main findings is that *SKC* has a negative impact on *BSI*, as expected, and the point estimates remain stable across most of the specifications.

Table 5 presents the pooled regression model estimates using daily, zip code-level *BSI* as the dependent variable. Models 1 and 2 present the results without any additional conditioning variables. Models 3 and 4 include dummy variables associated with the Monday and January effect. Models 5 and 6 include the sentiment-related measures as well as county-level economic condition variables. Models 1, 3, and 5 present the specification with *SKC* and previous year

SKC as the primary explanatory variables, while models 2, 4, and 6 present the estimates using *DSKC* instead.

In the baseline models of models 1 and 2, *SKC* (estimate = -0.090, *t*-value = -2.04) and *DSKC* (estimate = -0.051, *t*-value = -2.55) are both negative, and remain negative and statistically significant at least at the 5% level across all the remaining specifications. Therefore, the explanatory power of *SKC* is not related to any of our conditioning variables considered. The adjusted R^2 across the specification is low, as expected given the nature of the trade data. In model 1, the adjusted R^2 is 0.019%, while that of model 2 is, as expected, is lower (R^2 = 0.004%).

Using the estimates in model 5, a standard deviation increase in *SKC* decreases *BSI* by approximately 0.025, or 1.732% of the sample *BSI* standard deviation. The economic magnitudes calculated from the *DSKC* estimates are similar, though smaller. The estimates are smaller than those calculated in the univariate tests of Table 6, though this finding is reasonable given the fact that the univariate tests exploit subsamples of relatively extreme values of *SKC*.

To evaluate the robustness of the results in Table 7, additional specifications are considered in Table 8 for the regression models on the two-level panel that include fixed effects based on the zip code groups and, separately, on the date groups. The zip code fixed effects estimator helps evaluate whether the *SKC* explanatory power may be related to other time-invariant, regional effects. The date fixed effects estimator restricts the *SKC* explanatory power to intra-date variation, providing a conservative test that better accounts for potential seasonality in the BSI variable. The centered R^2 is reported instead of the adjusted R^2 for comparison with Table 7.

Models 1 through 4 present the results for the zip code fixed effects models. The *SKC* estimates attenuate slightly, though they are still negative and statistically significant at the 10% level in model 1 (estimate = -0.075, *t*-value = -1.74). Including the full set of conditioning variables, the *SKC* coefficient remains negative (estimate = -0.080, *t*-value = -1.86). Similar results obtain for the *DSKC* specifications. For the date fixed effects models in models 5 through 8, the *SKC* coefficient remains negative (estimate = -0.112, *t*-value = -2.60) in model 1, and is slightly larger than those reported in Table 7. Because the model includes date fixed effects, only county-level economic variables are included in models 7 and 8. The *SKC* coefficient remains stable, along with the *DSKC* specification.

In the final set of specifications, the institutional investor trading data are disaggregated to the zip code-, stock-, and date-level, constituting a three-level panel. In these specifications, the dependent variable is the daily, investor-stock-level *BSI*, constructed as the net buy-sell dollar volume for each investor-stock, scaled by the daily, average total trading volume for each investor. The *BSI* measure is further scaled by a factor of 1,000.

Table 9 presents the models estimates. Models 1 through 4 include only date fixed effects, while models 5 and 6 include date-stock fixed effects. Models 3 and 4 include stock-level characteristics, including the natural log of the stock's market capitalization, as well as the inverse of the stock's share price. The disaggregated data are expected to generate downward bias on the OLS standard error estimates, and so the standard errors are adjusted using a three-way cluster on zip code-, date- and stock-levels.

In the date fixed effects models, the *SKC* coefficient is negative in model 1 (estimate = -1.021, *t*-value = -1.94), and remains similar with stock-level conditioning variables in model 2 (estimate = -1.024, *t*-value = -1.93). The *DSKC* coefficients remain negative as well, though the coefficients are no longer statistically significant at the 10% level in either of the specifications. The results are not surprising, given the additional noise in the investor-stock *BSI* measure.

Finally, the date-stock fixed-effect model results are presented in models 5 and 6 of Table 9. Again, given the large number of fixed effects, the power of these tests is expected to be quite low, as the explanatory power of *SKC* is restricted to within each stock-date group. Even so, the point estimates on *SKC* help address the robustness of the main findings for omitted factors not captured in the previous models.

In model 5, the *SKC* coefficient is negative and statistically significant (estimate = -1.242, *t*-value = -2.00). As in models 2 and 4, the *DSKC* coefficient is negative, though statistically insignificant (estimate = -0.701, *t*-value = -1.24). The *SKC* coefficient is slightly larger in absolute magnitude over those in models 1 and 3, though the differences are statistically insignificant. As expected, the centered R^2 are much smaller than those in Table 8.

In summary, consistent with our key conjecture, we find that weather-induced mood affects institutional trading. Institutions are more likely to engage in selling when the weather makes them relatively more pessimistic and exacerbate their perceptions about market mispricing. Specifically, we find that *SKC* negatively impacts *BSI* across a wide variety of specifications. This evidence is robust to unobservable variation related to regional, seasonal, and stock characteristics.

5. Weather-Induced Investor Mood, Stock Prices, and Returns Comovement

Our results so far demonstrate that weather-based proxies of mood impact institutional investor beliefs and trading behavior. Naturally, a related question is whether mood could have an impact on stock prices. In this section, we examine the potential link between weather and stock returns. The stock-level SKC is statistically insignificant in the daily return regression models, though loads negatively and is statistically significant among stocks with greater arbitrage costs. We also document returns comovement among stocks with larger variation in investor mood.

5.1 Investor Mood and Stock Prices: Regression Specification

Because the institutional investor trading database only represents a fraction of the universe of institutional investors, our tests in this section utilize 13f holdings data to construct stock-level measures of investor mood. Specifically, we construct a stock-level measure of investor mood proxied by the average *SKC* around institutional locations. Institutional investors in stock *i* are identified using 13f holdings data from the previous quarter. The holdings data provide a snapshot of end-of-the-quarter holdings for institutional investors. The *SKC* measurements are calculated as before using the zip code of the institution's location. We require that at least one institutional investor holds stock *i* in the previous quarter in order to be included in the sample. Stocks with institutional ownership of 100% and more are excluded from the analysis. Lastly, we exclude all stocks with share price under $5 in 2010 dollars.

Stock-level *SKC* (or *StockSKC*) is calculated as the log of one plus the average, 14-day SKC of all institutional investors in stock *i* at date *t*. *StockSKC* is updated daily, using the most recent holdings data.[12] Likewise, lagged *StockSKC* is the log of one plus the average SKC in the same month of the previous year, but using same holdings data in the stock at time *t. DStockSKC* is defined as the difference between *StockSKC* and lagged *StockSKC*.

StockSKC is expected to have a negative relationship with daily stock returns, controlling for lagged *StockSKC*. Hirshleifer and Shumway (2003) argue that weather flucutations are more likely to impact the sign of stock returns rather than the magnitude. Because daily, individual

[12] Using weighted averages based on investor position size do not qualitatively alter the results.

stock returns are likely to be affected by various microstructure issues, such as asynchronicity, we employ a test similar to that employed in Hirshleifer and Shumway (2003). In particular, we estimate a linear probability model where the dependent variable of interest is an indicator variable, or *D(R>0)*, that takes the value of one if the stock return is positive, and zero otherwise.[13]

$$D(R_{i,t} > 0) = g_0 + g_1 * StockSKC_{i,t} + g_2 * LaggedStockSKC_{i,t} + \boldsymbol{g} * \boldsymbol{X_{i,t}} + \varepsilon_{i,t}$$

Here, **X** represents a set of conditioning variables that includes the log of the end-of-the-previous-quarter's market capitalization, proportion of shares held by institutional investors, the inverse of the end-of-the-previous-quarter's share price, a Monday dummy variable, and a January dummy variable. We also examine alternative specifications that include interchanging *StockSKC* and lagged *StockSKC* with *DStockSKC*. The coefficient on the *SKC* variable is expected to be negative, or $g_1 < 0$, as before.

Short-sales constraints are also expected to affect the estimates, as they increase the costs for other, mood-neutral institutional investors to arbitrage the stock (Nagel, 2005). Costs to arbitrage, or *ArbCosts*, is measured as the inverse of one plus the proportion of shares held by institutional investors. Thus, pooling the sample of low and high *ArbCosts* stocks will likely limit power in detecting the weather effect on individual stock returns. Accordingly, the sample is divided based upon *ArbCosts* percentile rankings for each date. We expect the weather effect to be more pronounced in stocks with higher levels of *ArbCosts*, and so we allow for greater granularity in the subsamples with higher *ArbCost* values.

We consider the effect of measurement error in the stock-level measure of investor mood, and argue that it is likely to bias the estimates away from our predictions. The frequency of the holdings data does not allow us to observe investor positions within each quarter In particular, we cannot identify investors that may actively trade in the stock, but hold no position at the SEC filing dates. Measurement errors related to the observation frequency are unlikely to be systematic, which could attenuate the bias in our test coefficients toward zero. However, if locations of pessimistic investors with short positions in the stock are less likely to be included in the stock-level measure of mood, this measure will be biased downward. Consequently, our point estimates may be biased in the opposite direction of our predictions. In other words, the stock-

[13] The results are similar when we use a logisitic regression model.

level measure will take on values that are too low in these cases, and will be accompanied by negative stock returns, creating a positive bias in the test coefficient.

5.2 Investor Mood and Stock Prices: Stock Return Regression Estimates

Table 10 displays the results using the stock-level investor mood measure. The dependent variable in all the regression models is *D(R>0).* The table shows nine columns, where the first is associated with estimates from the pooled sample, and the last eight are associated with estimates from sample splits based upon the *ArbCosts* rankings. The first four of these columns are organized into quintile bins, while the last four columns progress in 5% increments.

The first row displays the estimates using the *StockSKC* variable, while conditioning for lagged *StockSKC* and stock characteristics. The *StockSKC* coefficient in the pooled sample is statistically insignificant. However, the results on the *ArbCosts* sample splits reveal that the weather effect is concentrated among stocks with higher costs to arbitrage. The estimates in the top 20% samples are negative and statistically significant. The coefficients before that point are decreasing, though are statistically insignificant. Using *DStockSKC* in the model instead yield similar results.

Figure 2 graphically represents the coefficients from the sample split regressions from row 2 by *ArbCosts* ranking, along with 90% confidence intervals. Stocks in the highest level of *ArbCosts* increase slightly, creating a smirk-shaped pattern in the coefficients. This is in part due to the fact that the *StockSKC* measure becomes less reliable when the percentage ownership of institutional ownership is very low, or *ArbCosts* is very high, as it becomes very difficult to account for locations of other investors holding those stocks.

5.3 Investor Mood and Stock Prices: Robustness Checks

We examine whether the results are sensitive to our definition of arbitrage costs. In untabulated estimates, we obtain similar results when redefining *ArbCosts* according to other proxies, such as market capitalization or idiosyncratic volatility, though are slightly weaker.

We also consider whether our stock return results can be explained in part by the effect of mood fluctuations of market-makers in NYSE stocks. Goetzmann and Zhu (2005) show that New York City sky cloud coverage negatively impacts daily NYSE index returns, and most of explanatory power is concentrated on days that experience the largest change in spreads in S&P

100 stocks. In untabulated results, we account for these effects by repeating the analysis excluding NYSE stocks, and obtain similar results. We also consider specifications that include New York City sky cloud coverage as an additional conditioning variable, and find that it has little material impact on the findings.

5.4 Investor Mood and Returns Comovement

The previous section confirms that the investor mood measure has a negative impact on daily stock returns. We now examine whether investor mood induces comovement in stock returns. If weather causes correlated trades amongst investors due to mood, for which we provide evidence, weather fluctuations may also generate excessive returns comovement over time.

We test this hypothesis formally by estimating returns comovement relative to a series of portfolios motivated by the findings in the previous section. In particular, four index portfolios are formed based upon stocks experiencing large fluctuations in *StockSKC*, and low and high levels of *ArbCosts*. Specifically, based upon daily ranking of *DStockSKC*, we separate stocks into two groups in the lowest and highest 20th percentile. We further separate these stocks according to whether they are in the lowest and highest 20th percentile based upon unconditional, daily rankings based on *ArbCosts*. Four stock portfolios are constructed from the intersection of the extreme *ArbCosts* and *DStockSKC* rankings: low *ArbCosts* and low *DStockSKC* (LL); low *ArbCosts* and high *DStockSKC* (LH); high *ArbCosts* and low *DStockSKC* (HL); and high *ArbCosts* and high *DStockSKC* (HH). For each portfolio *j*, Idx_j is the portfolio return in excess of the riskless rate. The portfolios are value-weighted and rebalanced daily.[14]

5.5 Investor Mood and Returns Comovement: Regression Specification

For each stock, comovement relative to each of the *Idx* portfolios is estimated using daily stock returns. Specifically, we estimate rolling, time-series regressions that include the *Idx* portfolios, the Fama and French (1993) three factors (MKTRF, HML, and SMB) and the Carhart (1997) momentum factor (UMD), taking the following form:

$$R_{i,t} - R_{f,t} = \beta_0 + \sum\nolimits_{j \in \{LL,LH,HL,HH\}} \beta_j \, Idx_{j,t}$$
$$+\beta_{MKTRF} MKTRF_t + \beta_{HML} HML_t + \beta_{SMB} SMB_t + \beta_{UMD} UMD_t + \varepsilon_{i,t}$$

[14] The results also hold for equal-weighting.

Because mood is expected to have a short-lived effect on stock prices, estimates of the *Idx* comovement parameters, β_j, are estimated using data up 30 calendar days into the future, resulting in a panel of parameter estimates based upon overlapping data. We require that each stock has no missing returns data to be included in the sample.[15]

We compare the parameter estimates across groups based upon *ArbCosts* and *DStockSKC* rankings. Cross-sectional averages of the parameters are calculated for each date within nine groups based upon whether the stock is in the lowest 20th, middle 60th, or highest 20th percentile based upon *ArbCosts* and *DStockSKC* rankings. The comovement estimates are then averaged over time by group.

Our tests on returns comovement focus on differences in the comovement estimates between high and low *DStockSKC* groups, conditional on *ArbCosts*. Because the portfolios are formed using *ArbCosts*, the comovement parameters may be influenced by other factors unrelated to weather that are not observable. We assume that unobservable factors influence returns the same in the estimation within each *ArbCosts* group. Therefore, hypothesis testing for each parameter is based on the difference between the high and low *DStockSKC* groups for each *ArbCosts* group. The overlapping estimation windows used in the regression procedure are likely to generate serial correlation issues, biasing downward standard errors. Accordingly, we use Newey-West standard errors with a lag of 180 to construct the test statistics.

We predict that the comovement estimates should decrease in *DStockSKC* rankings for β_{LL} and β_{HL}, and increase in *DStockSKC* rankings for β_{LH} and β_{HH}. Additionally, these patterns should be pronounced for β_{LL} and β_{LH} in high ArbCost stocks, and for β_{HL} and β_{HH} in low ArbCost stocks.

5.6 Investor Mood and Returns Comovement: Comovement Estimates

Panel A of Table 11 presents the results for the 30-day estimation window. The estimates are reported by *ArbCosts* and *DStockSKC* groups, which are denoted at the top matter of the panel. The differences between high and low *DStockSKC* estimates (*H-L*) are also reported for each *ArbCosts* group. Only the comovement parameters for the *Idx* portfolios are reported for viewing convenience.

---

[15] The results are similar when relaxing the requirement to 15 days.

For the high *ArbCosts* group in the first four columns, β_{HL} decreases monotonically across the *DStockSKC* groups, as predicted, and the difference between the high and low *DStockSKC* groups is negative (estimate = -0.048, t-value = 3.43). Similarly, β_{HH} increases monotonically, and the difference is positive (estimate = 0.040, t-value = 2.86). The differences in β_{LL} or β_{LH} are statistically indistinguishable from zero.

For the low *ArbCosts* group in the last four columns, β_{LL} decreases monotonically across the *DStockSKC* groups, and the difference is negative (estimate = -0.101, t-value = 7.77). β_{LH} increases monotonically, and the difference is positive (estimate = 0.127, t-value = 18.57). While the differences in β_{HL} is statistically indistinguishable from zero, the differences in β_{LH} is slightly positive (estimate = 0.023, t-value = 1.66).

Interestingly, all the parameter differences are statistically significant at the 10% level for the middle *ArbCosts* group with the predicted signs, though the magnitudes are smaller. However, the patterns in the comovement estimates are not monotonic across the *DStockSKC* groups.

These results provide suggestive evidence for weather-induced returns comovement over short estimation windows. As mentioned before, weather is expected to have a short-lived effect on stock returns. To assess comovement over relatively longer windows, we recalculate the estimates extending the estimation window to 90 days, and report the results in Panel B of Table 11. Most of the parameter differences across the nine groups are statistically insignificant, with exception of β_{LH} for the low *ArbCosts* group, and β_{HH} for the high *ArbCosts* group. However, the magnitudes are more than half of those reported in Panel A.

We also consider the differences in the comovement estimates between the high and low *DStockSKC* without conditioning on *ArbCosts*. In addition, all stocks that are included in the *Idx* portfolios are removed from these calculations. The results are displayed in Figure 3. The estimates using a 30-day estimation window are represented in the blue bar series, while those using a 90-day estimation window are represented in the red bar series. Under the 30-day estimation window, in low *DStockSKC* stocks, the average difference is negative for the β_{HL} (estimate = -0.032, t-value = 2.70) and β_{LL} (estimate = -0.017, t-value = 1.71). In high *DStockSKC* stocks, the average difference is positive for β_{HH} (estimate = 0.020, t-value = 2.10) and β_{LH} (estimate = 0.016, t-value = 1.53). Using the 90-day estimation window, all the estimates are much smaller in magnitude, and are mostly statistically indistinguishable from zero, with exception of β_{HL}. Regardless of the estimation window, the comovement parameters associated with low *ArbCosts* are generally smaller in relative magnitude.

6. Summary and Conclusion

Recent studies in finance and economics demonstrate that weather-based mood proxies explain variation in trading volume and stock prices using broad-based indices. Those studies use sky cloud coverage around major stock market locations to identify the marginal investor. In contrast to these earlier studies, we use disaggregated data on the locations and trades of professional investors to examine how sky cloud coverage affect investors' perceptions about the market, their trading behavior, and the impact of weather-induced trading activities on stock returns. To our best knowledge, this study is the first to examine the impact of weather on institutional investor trading decisions and individual stock returns.[16]

We find that the weather-based mood indicators affect institutional investors' perceptions about market mispricing and their trading activities. Specifically, using survey data, we show that relatively cloudier days increase perceived mispricing in both individual stocks and the Dow Jones Industrial Index, while disaggregated trading data show that investors increased their selling propensities. We also find that weather-induced mood impacts individual stock returns using sky coverage data across the locations of institutional investors holding the stock. These results are pronounced among stocks with greater arbitrage costs. Finally, we find evidence of returns comovement relative to stock indices based upon weather fluctuations. Collectively, these findings complement existing studies on the weather effect on stock index returns.

In future work, it would be interesting to examine whether other groups of market participants are also influenced by weather-induced mood shifts. For example, it would be interesting to examine the impact of weather on the behavior of equity analysts and corporate managers. It is possible that corporate policies and analysts' earnings forecasts are both influenced by the local weather variations. It could also be interesting to identify which of these groups are more affected by the local weather and why. We leave these potentially interesting questions for future research.

[16] One notable exception is Loughran and Schultz (2004) who *indirectly* examine the effect of localized trading by firm headquarter locations.

References

Agarwal, S., Duchin, R., and Sosyura, D. (2012). In Mood for a Loan: The Causal Effect of Sentiment on Credit Origination. *Working paper*.

Baker, M., and Wurgler, J. (2000). The Equity Share in New Issues and Aggregate Stock Returns. *Journal of Finance*, 55(5), 2219-2257.

Baker, M., and Wurgler, J. (2004). A Catering Theory of Dividends. *Journal of Finance* 59(3), 1125–1165.

Baker, M., and Wurgler, J. (2007). Investor Sentiment in the Stock Market. *The Journal of Economic Perspectives*, *21*(2), 129-152.

Barber, B., Lee, Y., Liu, J., and Odean, T. (2007). Is the Aggregate Investor Reluctant to Realize Losses? European Financial Management, 13(3), 423–447.

Bassi, A., Colacito, R., and Fulghieri, P. (2013). 'O Sole Mio: An Experimental Analysis of Weather and Risk Attitudes in Financial Decisions, *Review of Financial Studies,* Forthcoming.

Bodoh-Creed (2013). Mood, Memory, and the Evaluation of Asset Prices, *Working paper.*

Camerer, C., Babcock, L., Loewenstein, G., and Thaler, R. (1997). Labor Supply of New York City Cabdrivers: One Day at a Time. *Quarterly Journal of Economics*, 112(2), 407-441.

Chhaochharia, V., Korniotis, G., and Kumar, A. (2012). Prozac for Depressed States? Effect of Mood on Local Economic Recessions, *Working paper*.

Clore, G. L., and Parrott, W. G. (1991). Moods and their Vicissitudes: Thoughts and Feelings as Information. *Emotion and Social Judgments*, 107-123.

Clore, G. L., Schwarz, N., and Conway, M. (1994). Affective Causes and Consequences of Social Information Processing. *Handbook of Social Cognition*,1, 323-417.

Coval, J., and Shumway, T. (2005). Do Behavioral Biases Affect Prices? *Journal of Finance*, 60(1), 1–34.

Edmans, A., Garcia, D., and Norli, Ø. (2007). Sports Sentiment and Stock Returns. *Journal of Finance*, 62(4), 1967-1998.

Dougal, C., Engelberg, J., García, D., and Parsons, C. A. (2012). Journalists and the Stock Market. *Review of Financial Studies*, 25(3), 639-679.

Forgas, J. P. (1995). Mood and Judgment: The Affect Infusion Model (AIM). *Psychological Bulletin,* 117(1), 39.

Frazzini, A. (2006). The Disposition Effect and Under-reaction to News. *Journal of Finance*, 61(4), 2017–2046.

Garcia, D. (2013). Sentiment During Recessions. *Journal of Finance,* Forthcoming.

Goetzmann, W. N., and Zhu, N. (2005). Rain or Shine: Where is the Weather Effect?. *European Financial Management,* 11(5), 559-578.

Hirshleifer, D., and Shumway, T. (2003). Good Day Sunshine: Stock Returns and the Weather. *Journal of Finance,* 58(3), 1009-1032.

Holt, C.A., and Laury, S.K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92(5), 1644–1655.

Ibbotson, R., Sindelar, J., and Ritter, J. (1994). The Market's Problems with the Pricing of Initial Public Offerings. *Journal of Applied Corporate Finance*, 7(1), 66-74.

Lo, K. and Wu, S. (2010). The Impact of Seasonal Affective Disorder on Financial Analysts and Equity Market Returns. *Working paper.*

Loewenstein, G. (2000). Emotions in Economic Theory and Economic Behavior. *American Economic Review*, 90(2), 426-432.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., and Welch, N. (2001). Risk as Feelings. *Psychological Bulletin*, 127(2), 267.

Loughran, T., and Schultz, P. (2004). Weather, Stock Returns, and the Impact of Localized Trading Behavior. *Journal of Financial and Quantitative Analysis*, 39(2), 343-364.

Kamstra, M. J., Kramer, L. A., and Levi, M. D. (2003). Winter Blues: A SAD Stock Market Cycle. *American Economic Review,* 93(1), 324-343.

Kaplanski, G., and Levy, H. (2009). Seasonality in Perceived Risk: A Sentiment Effect. *Working paper*.

Kramer, L. A., and Weber, J. M. (2012). This is Your Portfolio on Winter Seasonal Affective Disorder and Risk Aversion in Financial Decision Making. *Social Psychological and Personality Science*, 3(2), 193-199.

Mehra, R. and R. Sah (2002). Mood Fluctuations, Projection Bias, and Volatility of Equity Prices, *Journal of Economic Dynamics and Control*, 26, 869-887.

Petty, R. E., Gleicher, F., and Baker, S. M. (1991). Multiple Roles for Affect in Persuasion. (pp. 181-200). Elmsford, NY, US: Pergamon Press.

Puckett, A., and Yan, X. (2011). The Interim Trading Skills of Institutional Investors. *Journal of Finance*, 66(2), 601-633.

Saunders, E. M. (1993). Stock Prices and Wall Street Weather. *American Economic Review,* 83(5), 1337-1345.

Schwartz, G.E. (1990). Psychobiology of Repression and Health: A Systems Approach. *Repression and Dissociation: Implications for Personality Theory, Psychopathology, and Health*, 405-434, University of Chicago Press.

Wilson, T.D., Schooler, J. (1991). Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decision. Journal of Personality and Social Psychology, 60(2), 181-192.

Table 1: Summary Statistics on Survey, Trade, and Holdings Data

The table reports summary statistics for the average sky-cloud coverage variables that are matched to the survey dataset (Panel A), other characteristics linked to or from the survey dataset (Panel B), variables from and linked to the zip code-level trade dataset (Panel C), and variables from and linked to the zip code-level holding and stock returns dataset (Panel D). Please refer to Section 3 for descriptions of each dataset. Sky cloud coverage is the daily average of sky cloud coverage over 1 day to 4 weeks before the response or trade date. SKC is the natural log of one plus the daily average sky-cloud coverage over 2 week before the response or trade date. DSKC is the difference between SKC and last year's SKC, where last year's SKC is the natural log of the average daily sky cloud coverage for the same month and previous year of the response date. Portfolio size is the sum of the capital across asset classes reported by the investor respondent. Population is the county-level population for the response year. Income is the county-level median income for the response year. %EqPrem is the percentage equity premium estimate of the respondent. %LTGrow is the percentage long-term growth rate estimate of the respondent. %CrashProb is the percentage crash probability estimate of the respondent. %MisPrc is the natural log of the ratio of the respondent estimated, instrinsic level of the DJIA to the average, actual DJIA level over the previous week of the response date. D(High) takes value one if the respondent answers that actual prices in the stock market are too high relative to fundamentals. D(MisPrc) takes value one if the respondent answers that stock prices are unlikely to deviate from its fundamental values, and zero otherwise. D(Over) takes value one if the %MisPrc for a respondent is above the 75th percentile of all respondents. All the trade data variables are aggregated to the zip code-date level. BSI is the daily total, net buy-sell dollar volume, scaled by the average daily dollar trading volume. StockSKC is the average SKC of institutional investors holding within the same stock. DStockSKC is the difference between StockSKC and lagged StockSKC in the same month of the previous year. %IO is the proportion of shares in the stock held by institutional investors. D(R>0) is an indicator variable that takes value one if the stock experiences positive returns for that date, and zero otherwise.

Variable	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
Panel A: Average Sky Cloud Coverage by Estimation Window (Survey Data)					
1 Day	3.755	3.252	0.310	3.071	7.667
3 Days	3.672	2.186	2.000	3.571	5.313
1 Week	3.595	1.710	2.343	3.571	4.771
2 Weeks	3.582	1.388	2.731	3.596	4.527
3 Weeks	3.612	1.300	2.823	3.632	4.415
4 Weeks	3.577	1.264	2.803	3.604	4.343
Panel B: Survey Data Sample					
SKC	1.467	0.355	1.316	1.526	1.711
DSKC	-0.062	0.297	-0.223	-0.044	0.121
ln(Portfolio Size)	14.510	5.163	12.899	14.648	17.910
ln(Population)	8.526	1.311	7.784	8.614	9.323
ln(Income)	5.967	0.274	5.777	5.940	6.133
ln(1+%EqPrem)	0.067	0.085	0.030	0.049	0.068
ln(1+%LTGrow)	0.066	0.064	0.039	0.058	0.077
ln(1+%CrashProb)	0.081	0.094	0.010	0.049	0.095
%MisPrc	-0.003	0.120	-0.031	0.015	0.060
D(High)	0.196	0.397			
D(MisPrc)	0.824	0.381			
D(Over)	0.252	0.435			
Panel C: Trade Data Sample					
SKC	1.582	0.264	1.434	1.615	1.765
DSKC	-0.022	0.268	-0.181	-0.008	0.151
BSI	0.017	1.448	-0.095	0.000	0.107
Panel D: Holdings Data Sample					
StockSKC	1.582	0.139	1.496	1.590	1.676
DStockSKC	-0.017	0.149	-0.112	-0.013	0.079
%IO	0.458	0.239	0.265	0.481	0.648
D(R>0)	0.473	0.499			

Table 2: Investor Responses by SKC Estimation Window

The table displays the OLS regression model estimates for D(High) (Panel A), D(MisPrc) (Panel B) and %MisPrc (Panel C) as the dependent variables across SKC estimation windows. Columns 1 through 5 correspond with SKC calculated as the average daily SKC over an estimation window of 1 day, 1 week, 2 weeks, 3 weeks, and 4 weeks, respectively, up until the response date. The last year SKC is estimated as the average daily SKC over the same month in the previous year. Both explanatory variables are transformed using a natural log transformation on one plus its value. Standard errors are clustered on the zip code and date level, and are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
	SKC Estimation Window				
Variable	1 Day	1 Week	2 Weeks	3 Weeks	4 Weeks
Panel A: Prices Too High					
SKC	0.010	0.043*	0.076**	0.078*	0.083**
	(0.012)	(0.026)	(0.036)	(0.042)	(0.040)
Last year's SKC	-0.021	-0.003	-0.025	-0.027	-0.03
	(0.032)	(0.041)	(0.047)	(0.051)	(0.049)
Adjusted R^2	0.200%	0.300%	0.400%	0.300%	0.400%
N	1,494	1,494	1,494	1,494	1,494
Panel B: Mispricing					
SKC	0.003	0.026	0.092***	0.135***	0.126***
	(0.011)	(0.023)	(0.029)	(0.034)	(0.034)
Last year's SKC	-0.023	-0.039	-0.085**	-0.115***	-0.109***
	(0.034)	(0.033)	(0.036)	(0.039)	(0.039)
Adjusted R^2	0.100%	0.200%	0.600%	0.900%	0.700%
N	1,494	1,494	1,494	1,494	1,494
Panel C: Implied DJIA Percentage Mispricing					
SKC	-0.004	-0.018**	-0.029***	-0.026**	-0.032***
	(0.004)	(0.008)	(0.010)	(0.011)	(0.011)
Last year's SKC	-0.007	0.003	0.01	0.008	0.012
	(0.008)	(0.010)	(0.010)	(0.010)	(0.010)
Adjusted R^2	0.200%	0.500%	0.600%	0.400%	0.500%
N	1,494	1,494	1,494	1,494	1,494

Table 3: Pooled Perceived Mispricing Regression Estimates using the Survey Data

The table displays the regression model estimates of D(High) (Panel A); D(MisPrc) (Panel B); and %MisPrc, D(Over) and D(Under) (Panel C) as the dependent variables. D(Over) and D(Under) take value one if %MisPrc for a respondent is above the 75th percentile and under the 25th percentile, respectively. Binary response models are estimated using OLS or Probit, as indicated. For probit specifications, the partial effects are reported in brackets. Some models include additional explanatory variables as indicated, though the estimates are not displayed. Market volatility is measured as the log of the standard deviation of the DJIA over the previous 90 days. Portfolio size is estimated as the log of one plus the sum of the investor's reported capital across asset classes. County-level economic conditions include the log of population, log of median income, change in log population over the previous year, and change in log median household income over the previous year. The other survey responses include the log of one plus the repondent's equity premium, the market crash probability, and long-term corporate growth rate estimates. Two-way clustered standard errors on the zip code and date level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Prices Too High					
	(1)	(2)	(3)	(4)	(5)
Estimator:	OLS	OLS	OLS	Probit	Probit
Dependent Variable:	D(High)	D(High)	D(High)	D(High)	D(High)
SKC	0.076**	0.073**	0.077**	0.285**	0.323**
	(0.034)	(0.035)	(0.034)	(0.140)	(0.142)
				[-0.079]**	[-0.086]**
Last year's SKC	-0.025	-0.025	-0.029	-0.092	-0.138
	(0.046)	(0.045)	(0.046)	(0.170)	(0.176)
				[0.025]	[0.037]
Adjusted/Pseudo R^2	0.500%	0.600%	3.800%	0.530%	3.910%
N	1,494	1,494	1,494	1,494	1,494
Market Volatility?	X	X	X	X	X
Portfolio Size?	X	X	X	X	X
County Characteristics?		X	X	X	X

Panel B: Mispricing					
	(1)	(2)	(3)	(4)	(5)
Estimator:	OLS	OLS	OLS	Probit	Probit
Dependent Variable:	D(MisPrc)	D(MisPrc)	D(MisPrc)	D(MisPrc)	D(MisPrc)
SKC	0.093***	0.100***	0.102***	0.385***	0.398***
	(0.029)	(0.028)	(0.029)	(0.105)	(0.106)
				[-0.099]***	[-0.101]***
Last year's SKC	-0.085**	-0.080**	-0.081**	-0.313**	-0.316**
	(0.037)	(0.038)	(0.037)	(0.152)	(0.150)
				[0.080]**	[0.080]**
Adjusted/Pseudo R^2	0.600%	0.800%	1.700%	0.790%	1.880%
N	1,494	1,494	1,494	1,494	1,494
Market Volatility?	X	X	X	X	X
Portfolio Size?	X	X	X	X	X
County Characteristics?		X	X	X	X
Other Survey Responses?			X		X

Panel C: Implied DJIA Percentage Mispricing								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator:	OLS	OLS	OLS	OLS	OLS	OLS	Probit	Probit
Dependent Variable:	%MisPrc	%MisPrc	%MisPrc	D(Under)	D(Over)	D(Over)	D(Over)	D(Over)
SKC	-0.027***	-0.026***	-0.027***	-0.054	0.122***	0.123***	0.409***	0.422***
	(0.010)	(0.010)	(0.010)	(0.046)	(0.034)	(0.034)	(0.112)	(0.114)
							[0.130]***	[0.133]***
Last year's SKC	0.01	0.009	0.01	0.063	-0.069	-0.071	-0.224	-0.246
	(0.010)	(0.011)	(0.011)	(0.053)	(0.048)	(0.048)	(0.153)	(0.157)
							[-0.071]	[-0.078]*
Adjusted/Pseudo R^2	1.100%	1.100%	3.800%	2.600%	0.900%	3.500%	0.820%	2.900%
N	1,494	1,494	1,494	1,494	1,494	1,494	1,494	1,494
Market Volatility?	X	X	X	X	X	X	X	X
Portfolio Size?	X	X	X	X	X	X	X	X
County Characteristics?		X	X	X	X	X	X	X
Other Survey Responses?			X			X		X

Table 4: Fixed Effects Regression Estimates using the Survey

The table displays the OLS regression model estimates of D(High) (Models 1 and 5), %MisPrc (Models 2 and 6), D(Over) (Models 3 and 7), and D(MisPrc) (Models 4 and 8) as the dependent variables. Though not displayed, some of the models include additional explanatory variables, as indicated. Models 1 through 4 include fixed effects for state groups, and models 5 through 8 include fixed effects for year-quarter grouping pairs. Two-way clustered standard errors on the zip code and date level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Dependent Variable				
Variable	D(High)	D(MisPrc)	%MisPrc	D(Over)	D(High)	D(MisPrc)	%MisPrc	D(Over)
SKC	0.092**	0.091***	-0.029***	0.130***	0.061*	0.116***	-0.018*	0.094***
	(0.037)	(0.033)	(0.011)	(0.039)	(0.035)	(0.029)	(0.009)	(0.033)
Last year's SKC	0.011	-0.081**	0.001	-0.045	-0.042	-0.083**	0.009	-0.069
	(0.046)	(0.040)	(0.011)	(0.051)	(0.045)	(0.038)	(0.011)	(0.046)
Adjusted R^2	0.360%	0.350%	0.400%	-0.230%	1.310%	0.530%	2.540%	2.670%
N	1,494	1,494	1,494	1,494	1,494	1,494	1,494	1,494
Market Volatility?	X	X	X	X	X	X	X	X
Portfolio Size?	X	X	X	X	X	X	X	X
County Characteristics?	X	X	X	X	X	X	X	X
State FE?	X	X	X	X				
Year-Quarter FE?					X	X	X	X

Table 5: Pooled Regression Analysis on Investor Responses using Alternative Weather Measure

The table displays the OLS regression model estimates of D(High) (Model 1), D(MisPrc) (Model 2), %MisPrc (Model 3), and D(Over) (Model 4) as the dependent variable. DSKC is the difference between SKC and past year SKC, which are described in detail in the previous table. Though not displayed, all the models include additional explanatory variables. Two-way clustered standard errors on the zip code and date levels are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
	Dependent Variable:			
Variable	D(High)	D(MisPrc)	%MisPrc	D(Over)
DSKC	0.057	0.093***	-0.021**	0.104***
	(0.036)	(0.027)	(0.009)	(0.034)
Adjusted R^2	0.400%	0.800%	1.000%	0.800%
N	1,494	1,494	1,494	1,494
Market Volatility?	X	X	X	X
Portfolio Size?	X	X	X	X
County Characteristics?	X	X	X	X

Table 6: Regression Analysis of High-Low SKC on Zip code-Date BSI

The table reports panel regression estimates using daily institutional buy-sell imbalance (BSI) as the dependent variable across samples based upon the unconditional (Panel A) and conditional (Panel B) percentile ranking of SKC. Observations are on the zip code-date level. Models 1 through 4 is based upon various subsets of the overall sample, only including observations that fall within the top and bottom 50^{th}, 33^{rd}, 25^{th}, and 10^{th} SKC percentiles, respectively. SKC is defined as the natural log of one plus the average SKC over the past 14 days. D(HighSKC) takes value one if the SKC ranking is greater than the 50^{th} percentile for the respective sample subset, and zero otherwise. Panel A uses SKC rankings based upon the entire sample, while panel B uses SKC rankings conditional on date. Two-way clustered standard errors are on the date- and zip code-level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Unconditional SKC Ranking

	(1)	(2)	(3)	(4)
	Sample Subset:			
Variable	High-Low 50th Percentile	High-Low 33rd Percentile	High-Low 25th Percentile	High-Low 10th Percentile
D(HighSKC)	-0.018	-0.029	-0.045**	-0.094***
	(0.016)	(0.020)	(0.022)	(0.026)
Adjusted R2	0.005%	0.011%	0.026%	0.083%
N	104,749	69,091	54,304	21,248

Panel B: Within-Date SKC Ranking

	(1)	(2)	(3)	(4)
	Sample Subset:			
Variable	High-Low 50th Percentile	High-Low 33rd Percentile	High-Low 25th Percentile	High-Low 10th Percentile
D(HighSKC)	-0.019	-0.041*	-0.054**	-0.092***
	(0.017)	(0.022)	(0.022)	(0.028)
Adjusted R2	0.005%	0.021%	0.036%	0.077%
N	104,749	69,091	54,304	21,248

Table 7: Pooled Regression Analysis on Daily, Zip code-Level BSI

The table reports panel regression estimates using daily, zip code-level institutional buy-sell imbalance (BSI) as the dependent variable. Investor trades are aggregated to the date-zip code level. The explanatory variables of interest are the natural log of one plus the average SKC over the past 14 days (SKC), the natural log of one plus the average SKC over the same month in the previous year (LSKC), and the deseasoned SKC defined as the difference between SKC and LSKC (DSKC). Though not displayed, some of the models include additional explanatory variables, as indicated. D(MONDAY) takes value 1 if the observation date is a Monday, and zero otherwise. D(JANUARY) takes value 1 if the observation month is January, and zero otherwise. Sentiment-based variables are taken from Baker and Wurgler (2004). PDND is the value-weighted dividend premium defined following Baker and Wurgler (2004). NIPO is the natural log of one plus the IPO volume from Ibbotson, Sindelar, and Ritter (1994) and updates. RIPO is the first-day returns on IPOs from Ibbotson, Sindelar, and Ritter (1994) and updates. CEFD is the closed-end fund discount. SE is the natural log of one plus the equity issuance. SD is the natural log of one plus the debt issuance. TURN is the NYSE monthly turnover from NYSE Factbook. Additionally, VIX is the daily VIX index level. County economic condition variables are as follows. POP is the log of the county-level population. DPOP is the change over the previous year. INC is the log of the county-level median household income. DINC is the change in INC over the previous year. Two-way clustered standard errors are on the date and zip code level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
SKC	-0.090**		-0.091**		-0.095**	
	(0.044)		(0.044)		(0.040)	
Last year's SKC	-0.047		-0.049		-0.056	
	(0.061)		(0.061)		(0.062)	
DSKC		-0.051**		-0.050**		-0.052***
		(0.020)		(0.020)		(0.018)
Adjusted R^2	0.019%	0.004%	0.018%	0.003%	0.048%	0.032%
N	104,753	104,753	104,753	104,753	104,753	104,753
Monday / January?			X	X	X	X
Sentiment?					X	X
County Characteristics?					X	X

Table 8: Fixed Effect Regression Analysis on Daily, Zip code-Level BSI

The table reports panel regression estimates using total daily, zip code-level BSI as the dependent variable. Investor trades are aggregated to the date-zip code level. Though not displayed, some of the models include additional conditioning variables, as indicated, and are described in the text. Models (1) through (4) include fixed effects based on zip code groups. Models (5) through (8) include fixed effects based on date groups. Two-way clustered standard errors on the date and zip code levels are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SKC	-0.075*		-0.080*		-0.112***		-0.112***	
	(0.043)		(0.043)		(0.043)		(0.040)	
Last year's SKC	-0.012		-0.021		-0.094		-0.093	
	(0.060)		(0.065)		(0.078)		(0.072)	
DSKC		-0.049**		-0.050***		-0.055***		-0.055***
		(0.020)		(0.019)		(0.020)		(0.020)
Centered R^2	0.010%	0.005%	0.040%	0.003%	0.034%	0.004%	0.062%	0.003%
N	104,753	104,753	104,753	104,753	104,753	104,753	104,753	104,753
Monday / January?			X	X				
Sentiment?			X	X				
County Characteristics?			X	X			X	X
Zip code FE?	X	X	X	X				
Date FE?					X	X	X	X

Table 9: Fixed Effect Regression Analysis on Daily, Zip code-Stock-Level BSI

The table reports regression model estimates using daily BSI as the dependent variable. Observations are on the date-stock-zip code level. Though not displayed, some of the models include additional explanatory variables, as indicated. Models (1) through (4) include fixed effects based on date groups. Models (5) through (6) include fixed effects based on date-stock grouping pairs. Three-way clustered standard errors on the stock, date and zip code levels are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
SKC	-1.021*		-1.024*		-1.242**	
	(0.526)		(0.530)		(0.620)	
Last year's SKC	-0.389		-0.391		-0.715	
	(0.711)		(0.707)		(0.876)	
DSKC		-0.629		-0.631		-0.701
		(0.446)		(0.449)		(0.563)
Centered R^2	0.0004%	0.0001%	0.0004%	0.0001%	0.0005%	0.0001%
N	13,324,421	13,324,421	13,324,405	13,324,405	13,324,421	13,324,421
Stock Characteristics?			X	X		
Date FE?	X	X	X	X		
Date x Stock FE?					X	X

Table 10: Stock-level SKC and Daily Stock Returns

The dependent variable in the models below is indicator variable taking value 1 for positive return days, and 0 otherwise. The top matter of the table presents the sample counts for the pooled and split samples. Each row present regression estimates from nine separate OLS regression models. The first column present the pooled estimates, and the second to ninth columns present the subsample estimates based upon *ArbCosts* rankings over the entire 1999-2010 sample period. We define *ArbCosts* as the inverse of the proportion of shares held by institutional investors. In total, there are 18 (= 2 * 9) regression models estimated. The models include the weather variables and additional conditioning variables mentioned within the text, though not shown. The weather variable is stock-level SKC (StockSKC) for first row, and change in stock-level SKC (DStockSKC) for second row. Lagged StockSKC is also included in also included in the first row, though not shown. Standard errors are estimated using two-way clustering on stock and date, and are shown in parentheses. Statistical significance at the 1%, 5% and 10% levels are denoted next to the coefficients as ***, **, and *.

			ArbCosts Percentile Rankings							
	Sample:	All	[0,20%)	[20,40%)	[40,60%)	[60,80%)	[80,85%)	[85,90%)	[90,95%)	[95,100%]
	N:	8.67E+7	1.73E+7	1.73E+7	1.73E+7	1.73E+7	4.30E+6	4.30E+6	4.30E+6	4.30E+6
(1)	StockSKC	0.012	0.037	0.040	0.027	0.010	-0.020	-0.027*	-0.034**	-0.031***
		(0.023)	(0.036)	(0.031)	(0.027)	(0.020)	(0.018)	(0.016)	(0.014)	(0.009)
	Adjusted R^2	0.25%	0.05%	0.08%	0.11%	0.15%	0.12%	0.13%	0.16%	0.20%
(2)	DStockSKC	-0.003	0.01	0.012	0.004	-0.008	-0.027*	-0.027*	-0.029**	-0.018**
		(0.020)	(0.032)	(0.028)	(0.024)	(0.018)	(0.016)	(0.014)	(0.012)	(0.008)
	Adjusted R^2	0.24%	0.03%	0.05%	0.10%	0.14%	0.12%	0.13%	0.16%	0.18%

Table 11: Returns Comovement in StockSKC Portfolios

The table reports comovement parameters from daily returns regressions that include ArbCosts-DStockSKC portfolio returns. Portfolios based upon ArbCosts and DStockSKC rankings for each date include stocks with in the lowest 20th percentile on ArbCosts and DStockSKC (LL); lowest 20th percentile in ArbCosts and highest 20th percentile in DStockSKC (LH); highest 20th percentile in ArbCosts and lowest 20th percentile in DStockSKC (HL); and highest 20th percentile in ArbCosts and DStockSKC (HH). Each portfolio's returns are value-weighted, measured in excess of the riskless rate, and rebalanced daily. For each sample date, regressions are estimated on the time-series of each stock's daily returns for up to 30 (Panel A) or 90 (Panel B) calendar days into the future. In addition to the portfolios mentioned above, the regression models also include the Fama and French (1993) three factors and the Carhart (1997) momentum factor, though the estimates are not reported here. Each regression coefficient is then averaged across stocks for each date within nine groups formed on ArbCosts and DStockSKC according to lowest 20th percentile (low or L), middle 60th percentile (middle or M) and highest 20th percentile (high or H). For each ArbCosts group, the difference between the high and low DStockSKC (H-L) estimates is also reported. The resulting time-series for each the regression coefficient is then averaged, and reported below. Newey-West standard errors using 180 day lags are reported in parentheses on the time-series averages. Group memberships are denoted at the top of each panel. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: 30-day Rolling Beta Estimates

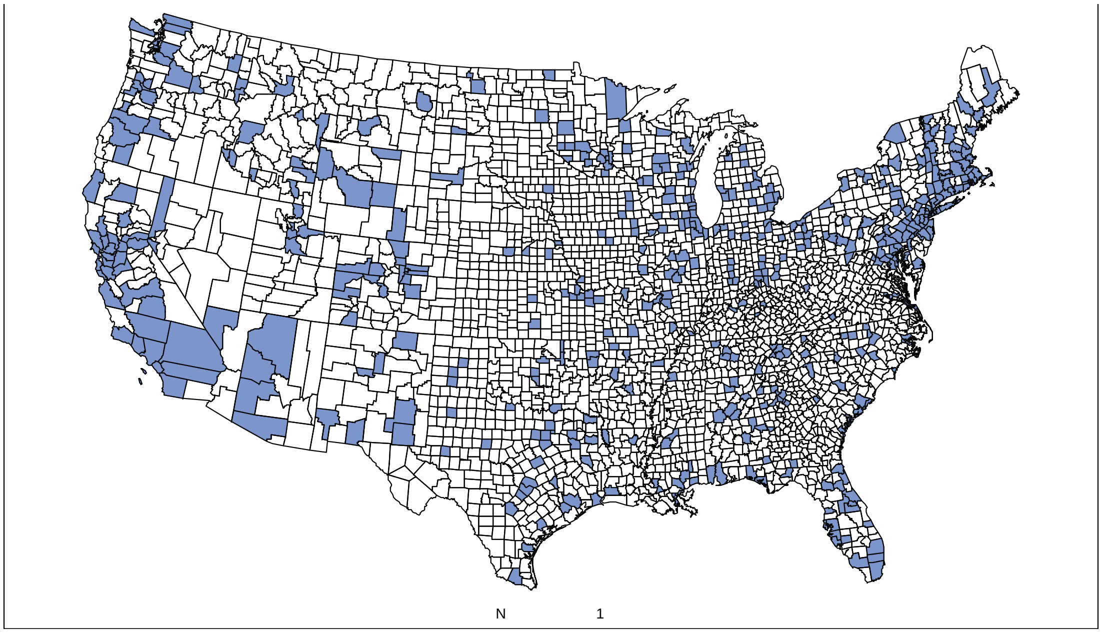
ArbCosts:	High				Middle				Low			
DStockSKC:	L	M	H	H-L	L	M	H	H-L	L	M	H	H-L
β_{HL}	0.181***	0.180***	0.133***	-0.05***	0.085***	0.038***	0.053***	-0.03***	-0.006	-0.024**	-0.008	-0.002
	(0.015)	(0.017)	(0.013)	(0.014)	(0.015)	(0.008)	(0.009)	(0.012)	(0.012)	(0.010)	(0.014)	(0.013)
β_{HH}	0.129***	0.159***	0.173***	0.044***	0.059***	0.030***	0.079***	0.020**	-0.002	-0.013	0.021*	0.023*
	(0.012)	(0.012)	(0.016)	(0.014)	(0.008)	(0.006)	(0.010)	(0.009)	(0.015)	(0.011)	(0.012)	(0.014)
β_{LL}	-0.019**	-0.03***	-0.015	0.004	0.013**	0.019***	-0.004	-0.017**	0.175***	0.099***	0.074***	-0.10***
	(0.008)	(0.011)	(0.012)	(0.010)	(0.007)	(0.006)	(0.008)	(0.008)	(0.014)	(0.011)	(0.009)	(0.013)
β_{LH}	-0.020**	-0.04***	-0.007	0.013	-0.009	0.015***	0.007	0.016*	0.047***	0.098***	0.174***	0.127***
	(0.009)	(0.013)	(0.007)	(0.008)	(0.010)	(0.007)	(0.007)	(0.009)	(0.007)	(0.009)	(0.006)	(0.007)

Panel B: 90-day Rolling Beta Estimates

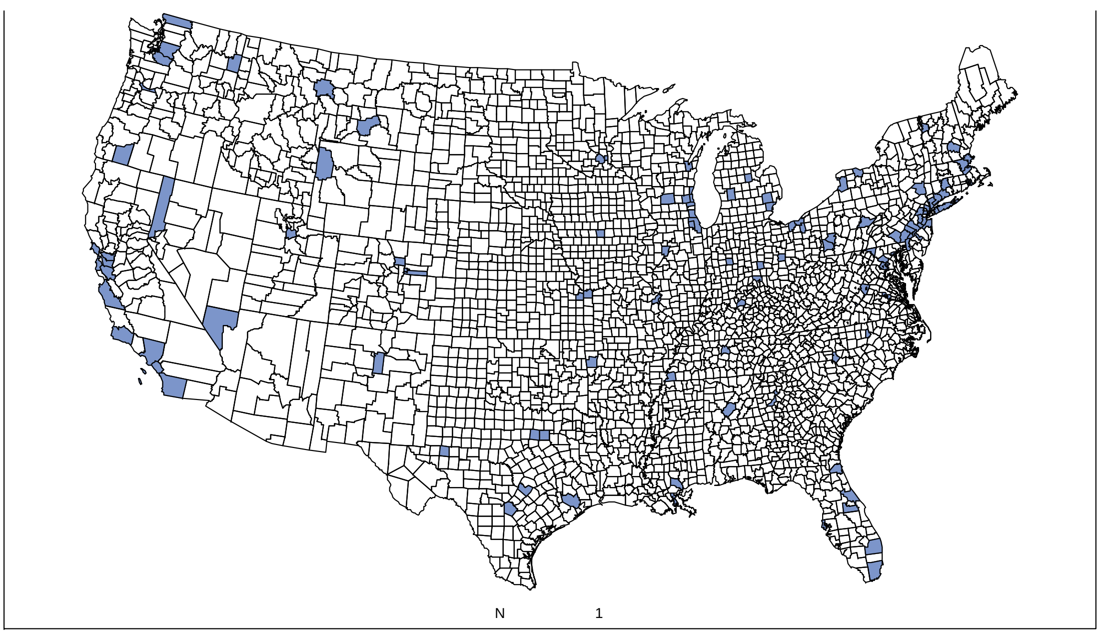
ArbCosts:	High				Middle				Low			
DStockSKC:	L	M	H	H-L	L	M	H	H-L	L	M	H	H-L
β_{HL}	0.148***	0.165***	0.138***	-0.01	0.074***	0.037***	0.058***	-0.016	0.000	-0.02***	-0.007	-0.007
	(0.015)	(0.017)	(0.014)	(0.015)	(0.013)	(0.007)	(0.009)	(0.011)	(0.008)	(0.008)	(0.007)	(0.008)
β_{HH}	0.125***	0.150***	0.145***	0.020**	0.059***	0.030***	0.065***	0.006	0.007	-0.014**	0.015	0.008
	(0.009)	(0.012)	(0.011)	(0.010)	(0.011)	(0.008)	(0.008)	(0.010)	(0.012)	(0.009)	(0.011)	(0.012)
β_{LL}	-0.009	-0.03***	-0.003	0.006	0.008*	0.019***	0.003	-0.005	0.096***	0.089***	0.087***	-0.009
	(0.006)	(0.008)	(0.009)	(0.008)	(0.005)	(0.006)	(0.005)	(0.005)	(0.010)	(0.010)	(0.009)	(0.010)
β_{LH}	-0.009	-0.022**	0.003	0.012	0.008	0.020***	0.017***	0.009	0.066***	0.088***	0.119***	0.053***
	(0.009)	(0.010)	(0.009)	(0.009)	(0.006)	(0.006)	(0.005)	(0.006)	(0.009)	(0.005)	(0.012)	(0.011)

Figure 1: Geographical Dispersion of the Survey and Trade Datasets

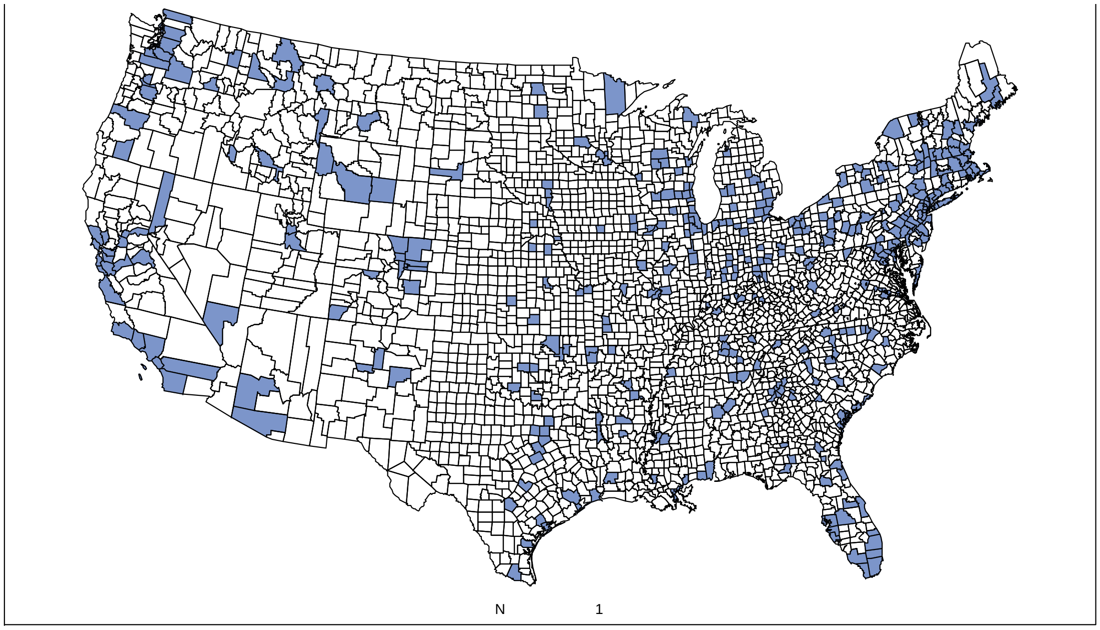
The figure represents the geographical distribution in the United States of the survey (Panel A) institutional investor trade (Panel B), and institutional investor holding (Panel C) data on the county-level. Color-coded counties denote the existence of at least one investor in the corresponding dataset.



Panel A: Survey Data



Panel B: Trade Data



Panel C: Holdings Data

Figure 2: Stock-level SKC Coefficients in D(R>0) Regressions

The figure displays the *DStockSKC* coefficients from row (2) in Table 10 as red dots across different subsamples based upon sample *ArbCosts* percentile rankings, overlaid with 90% confidence intervals.

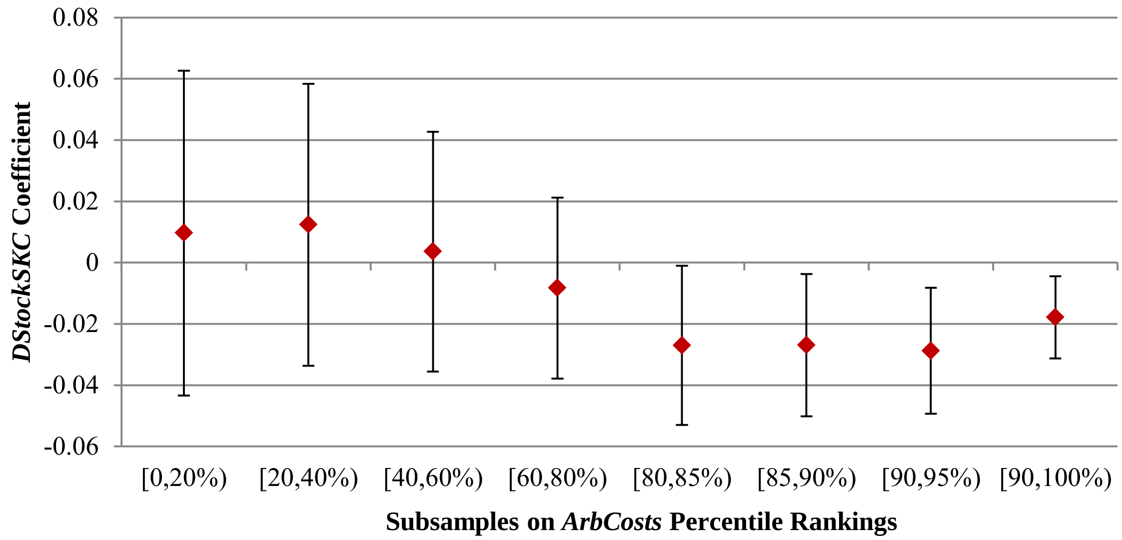


Figure 3: Differenced Comovement Estimates on Mood Portfolios

The figure displays the difference in comovement estimates between high-low *DStockSKC* stocks unconditional on *ArbCosts* ranking and excluding stocks used to form the investor mood (*Idx)* portfolios. The estimates are across the investor mood portfolios, as indicted in the horizontal axis. The blue (red) series denotes estimates using a 30 (90) day window. 90% confidence intervals are overlaid on each series.