Forecasting Crashes: Correlated Fund Flows and the Skewness in Stock Returns

Aug 2014

Xun Gong[1]

Chunmei Lin[2]

Remco C.J. Zwinkels[3]

[1] Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands

[2] Erasmus University Rotterdam, P.O.Box 1738, 3000 DR, Rotterdam, The Netherlands, E-mail: lin@ese.eur.nl

[3] Erasmus University Rotterdam and Tinbergen Institute, P.O.Box 1738, 3000 DR, Rotterdam, The Netherlands, E-mail: Zwinkels@ese.eur.nl

Abstract

This paper uses the correlation of money flow among mutual funds to forecast the skewness of stock returns. We show that asset returns are highly negatively skewed when their mutual fund owners experience correlated liquidity shocks. In addition, stocks with high mutual fund ownership are more "crash prone", whereas the returns of stocks with concentrated ownership tend to display more positive skewness.

Keywords: skewness, mutual funds, capital flow

JEL-codes: G12, G17, G23, C58

1. Introduction

Stock returns are asymmetrically distributed and display negative skewness (e.g., Albuquerque, 2010). One interpretation of this statistical artifact is that stocks that display more negative skewness are more prone to crashes (e.g. Chen, Hong, and Stein 2001). Whereas the existence of negative asymmetries in returns is generally not disputed, the economic mechanisms driving this empirical finding are less clear. We posit that correlated demand for liquidity of owners of financial assets is an important driver of the negative skewness of stock returns. We construct a measure of flow-driven–skewness (FDS) that captures the correlation of the liquidity driven trading needs of different mutual fund owners. We find that this measure strongly predicts skewness in the cross-section of stock returns.

A large number of empirical studies show that investor trading unrelated to fundamentals can trigger a significant price impact. For example, Coval and Stafford (2007) show that uninformed capital demand shocks can have a significant price impact on the assets held by distressed funds. When capital demand by investors increases, mutual funds without adequate cash reserves need to sell their holdings to meet the sudden increase of capital demand. Since the sale is immediate, mutual fund managers may sell their holdings at price levels that are significantly lower than their information-efficient values.

Greenwood and Thesmar (2011) develop a measure, which they call “Fragility”, to capture the expected variance of non-fundamental capital demand. The Fragility measure is high if a stock’s current owners face volatile liquidity shocks, if a stock’s ownership is highly concentrated, and if liquidity shocks for the stock’s owners are highly correlated. Greenwood and Thesmar (2011) connect the Fragility measure to the volatility of stock returns and find that Fragility strongly predicts stock volatility.

Brennan, Chordia, Subrahmanyam and Tong (2009) show that the demand for immediacy is stronger for sellers of securities than for buyers since investors are more likely to have a pressing need to raise cash than to exchange cash for securities. Mutual funds experiencing net capital outflows are forced to sell immediately, whereas funds with net inflows are much less pressed to act as they tend to scale up their existing holdings (Coval and Stafford, 2007, Lou, 2010). This asymmetric response to liquidity demand can lead to negative skewness of assets held by distressed funds.

On a similar approach as Greenwood and Thesmar (2011)'s Fragility measure, we propose the Flow-Driven Skewness (FDS) measure to capture the expected skewness caused by non-fundamental asset demand. A higher value of FDS implies that the net flow of a stock's current owners is more negatively skewed. Intuitively, if the aggregate flow of a stock's owners is negatively skewed, the owners will liquidate their position and hence drive down the stock price. As a result, the stock will also displays negative skewness, which indicates to what extend the stock is "crash prone".

The FDS of a stock is the sum of the flow skewness of its current owners and various coskewness terms. The coskewness terms capture the comovement between the owners' conditional mean flows. In other words, the FDS captures not only the movements of a single owner's flow, but also the comovements between the flows of different owners.

To measure the composition of ownership and the ex-ante covariance structure of the liquidity needs faced by its owners, we use the mutual fund ownership of U.S. listed equities. Because mutual funds regularly report their positions, the correlation structure of their liquidity-driven trades can be inferred from their inflows and outflows. Therefore, we implement our measures of FDS on US stocks for the period 1990 to 2010.

Our results indicate that FDS strongly predicts firm-level skewness and remains highly significant after controlling for several factors known to affect skewness. In addition, we find that ownership concentration and the proportion of shares held by mutual funds also play an important role in determining the firm-level skewness.

Accurately forecasting skewness has important economic relevance: not only does positive skewness provide investors higher expected return, but negative skewness measures to what extend a stock is "crash prone" which is essential for risk management (Chen, Hong and Stein 2001). The results of our study provide a novel and valuable methodology to identify "crash prone" assets which can be used as an important tool for both asset management and risk management.

2. Flow-Driven Skewness

Based on the finding that the flow of mutual funds can generate price pressure effects on stock prices, Greenwood and Thesmar (2011) construct the "Fragility" measure to capture the variance of the aggregate flow of stocks' current owners. They show that the "Fragility" measure has both statistically and economically significant power in forecasting the future volatility of stocks. We extend the approach of Greenwood and Thesmar and construct the higher moment variable Flow-Driven Skewness (FDS).

The weights w_{ikt} on stock i held by mutual fund k at date t is given by

$$w_{ikt} = \frac{n_{ikt}P_{it}}{a_{kt}} , \tag{1}$$

where $n_{ikt}$is the number of stocks i held by mutual fund k at time t, a_{kt} is the total net assets (TNA) of mutual fund k at time t, and P_{it} is the price of stock i at time t.

The net purchases of stock i by mutual fund k consists of two parts, active changes in weights and changes driven by flow of new money,

$$P_{it}\Delta n_{ikt} = n_{ikt}P_{it}\left(\frac{\Delta w_{ikt}}{w_{ikt}} - \frac{\Delta P_{it}}{P_{it}}\right) + w_{ikt}\Delta a_{kt}, \tag{2}$$

where Δ denotes the change relative to one period lag.

The net flow of mutual fund k can be denoted as the change in total assets adjusted for the assets appreciation or depreciation

$$f_{kt} = \Delta a_{kt} - \sum_j n_{jkt}\Delta P_{jt}. \tag{3}$$

Inserting Equation (3) into (2) yields

$$P_{it}\Delta n_{ikt} = n_{ikt}P_{it}\left(\frac{\Delta w_{ikt}}{w_{ikt}} - (\frac{\Delta P_{it}}{P_{it}} - \sum_j w_{jkt}\frac{\Delta P_{jt}}{P_{jt}})\right) + w_{ikt}f_{kt}, \tag{4}$$

where subscript j denotes other stocks in the portfolio of mutual fund k.

Equation (4) shows that the net dollar purchase of stock i by mutual fund *k* can be decomposed into two terms. The first term represents trading stemming from active rebalancing by the mutual fund manager. The second term is the contribution of fund flow to the net purchase of stock i, which represents liquidity driven trading. Greenwood and Thesmar (2011) focus on the second term in the decomposition and assume a linear relationship between total liquidity trades and stock returns, based on the fact that liquidity driven trading can trigger price pressure

$$r_{it+1} = \alpha + \gamma \frac{\sum_k w_{ikt} f_{kt}}{\theta_{it}}, \qquad (5)$$

where r_{it+1} is the return of stock i in period *t+1*, θ_{it} is the market capitalization of stock i at time t, and α and γ are the parameters that need to be estimated[4].

Equation (5) shows that the price pressure effect is proportional to the sum of flow driven capital demand of all mutual funds. The coefficient γ capture the magnitude of price impact such that 1/ γ is the price elasticity of demand (Wurgler and Zhuravskaya (2002) and Chacko, Jurek, and Stafford (2008)).

Rewriting Equation (5) into matrix notation yields:

$$r_{it+1} = \alpha + \frac{\gamma}{\theta_{it}} W_{it}' F_t, \qquad (6)$$

where W_{it} is the vector of holding weights of stock i held by all mutual funds in their portfolios, and F_t is the vector of net flows of each mutual funds experienced at time t.

Based on Equation (6), the conditional variance of returns of stock i at time t is given by:

$$\sigma_{it}{}^2 = (\frac{\gamma}{\theta_{it}})^2 W_{it}' \Omega_t W_{it} \qquad (7)$$

where Ω_t is the conditional variance-covariance matrix of net flows of all mutual funds which hold stock i at time t.

[4] Scaling total flow driven capital demand by market capitalization is a common procedure. The reason is primarily to make the price impact caused by liquidity driven trading comparable across stocks.

Greenwood and Thesmar (2011) subsequently define the fragility measure G of stock i at time t as:

$$G_{it} = \frac{1}{\theta_{it}{}^2} W_{it}{}' \Omega_t W_{it} \tag{8}$$

Extending the approach of Greenwood and Thesmar (2011), we construct a measure for the third moment of the distribution of flows. Portfolio skewness is given by

$$\sigma_{pt}^3 = E[R_{pt} - \overline{R_{pt}}]^3 = \sum_{l,j,k}^{N} w_{it} w_{jt} w_{kt} E[(R_{it} - \overline{R_l})(R_{jt} - \overline{R_j})(R_{kt} - \overline{R_k})], \tag{9}$$

where σ_{pt}^3 is the multivariate skewness measurement at time t, R_{pt} is the actual portfolio return, and $\overline{R_p}$ is the expected return of the portfolio. w_i, w_j, and w_k are the weights assigned to assets $i, j,$ and k in the portfolio. R_{it}, R_{jt}, and R_{kt} are the returns of assets i, j, and k and $\overline{R_l}$, $\overline{R_j}$, and $\overline{R_k}$ are the average return of assets i, j, and k respectively.

Similarly, we construct the skewness of the flow of mutual funds that hold stock i as follows

$$\sigma_{it}^3 = \sum_{l,j,k}^{N} w_{lt} w_{jt} w_{kt} E[(f_{lt} - \overline{f_{lt}})(f_{jt} - \overline{f_{jt}})(f_{kt} - \overline{f_{kt}})], \tag{10}$$

where σ_{it}^3 is the raw flow-driven-skewness, and $w_{lt}, w_{jt,}$ and w_k are the holding weights of stock i held by mutual funds i, j, and k respectively in their portfolios at time t. f_{lt}, f_{jt}, and f_{kt} are the flows experienced by mutual funds i, j, and k at time t respectively. $\overline{f_{lt}}$,$\overline{f_{jt}}$, and $\overline{f_{kt}}$ are the average fund flows from time *0* to t of mutual funds i, j, and k, respectively.

Recognizing that the coskewness matrix, a $NxNxN$ cube, is difficult to manage we adopt the methodology proposed by Athayde and Flores (1987). Athayde and Flores (1987) suggest transforming the $NxNxN$ cube into a NxN^2 matrix, slicing the cube and laying down the matrices together into a NxN^2 matrix. For example, assume one stock is held by two mutual funds, then the coskewness matrix of flow is given by:

$$\sigma^3 = \begin{bmatrix} \sigma_{111}\sigma_{112}\sigma_{211}\sigma_{212} \\ \sigma_{121}\sigma_{122}\sigma_{221}\sigma_{222} \end{bmatrix},$$

where, for instance, σ_{221} is the coskewness between flows of mutual funds 1 and 2, etc. If one stock is held by n mutual funds, the coskewness matrix is given by (following Satchell and Scowcroft 2003):

$$M_3 = \begin{bmatrix} \sigma_{111} & \sigma_{121} & \dots & \sigma_{1n1} & \sigma_{112} & \sigma_{122} & \dots & \sigma_{1n2} & \sigma_{11n} & \sigma_{12n} & \dots & \sigma_{1nn} \\ \sigma_{211} & \sigma_{221} & \dots & \sigma_{2n1} & \sigma_{212} & \sigma_{222} & \dots & \sigma_{2n2} & \sigma_{21n} & \sigma_{22n} & \dots & \sigma_{2nn} \\ \dots & \dots & & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_{n11} & \sigma_{n21} & \dots & \sigma_{n12} & \sigma_{n12} & \sigma_{n22} & \dots & \sigma_{nn2} & \sigma_{n1n} & \sigma_{n2n} & \dots & \sigma_{nnn} \end{bmatrix}$$

where σ_{ljk} is the coskewness of the flow of mutual funds i, j, and k respectively, and the individual elements in M_3 can be calculated as

$$\sigma_{ljk} = \frac{1}{T-1} \sum_{l,j,k=1}^{N} \sum_{t=1}^{T-1} (f_{l,t} - \overline{f}_l)(f_{j,t} - \overline{f}_j)(f_{k,t} - \overline{f}_k). \tag{11}$$

Following the method suggested by Satchell and Scowcroft (2003), we decompose M_3 into N slices. Let M_{31} denotes the first slice of the cube, then the coskewness matrix can be obtained as:

$$M_3 = [M_{31} M_{32} M_{33} M_{34} \dots M_{3n}] \tag{12}$$

The multivariate skewness of stock i derived from the aggregate flow of its owners is subsequently given by

$$\sigma_{it}^3 = W_{it}{}' M_{3,t} (W_{it} \otimes W_{it}), \tag{13}$$

where W_{it} is the vector of holding weights of stock i of all mutual funds at time t, and $\otimes$ represents the Kronecker product[5].

To make σ_{it}^3 comparable across stocks, , we follow Greene (1993) to scale the raw aggregate skewness by the aggregate standard deviation of flow (the square root of Fragility measure) raised to third power. Applying the normalization as described above yields:

$$\sigma_{it}^3 = W_{it}{}' M_{3,t} (W_{it} \otimes W_{it}) / (W_{it}{}' \Omega_t W_{it})^{\frac{3}{2}} \quad , \tag{14}$$

where Ω_t is the covariance-variance matrix of fund flow, and equivalently $W_{it}{}' \Omega_t W_{it}$ is the Fragility measure as developed in Greenwood and Thesmar (2011).

[5] Kronecker product is operator for two matrices with arbitrary size. For example, if A is a (nxm) matrix and B is a (pxq) matrix, then C=A⊗B is a (npxmq) dimensional matrix.

Given that liquidity driven trading by mutual funds can trigger price pressure and that the relationship between the price impact and the total flow driven capital demand is linear, we posit the following equation

$$Skewness\ r_{i,t+1} = \alpha + \gamma * W_{it}{}' M_{3,t}(W_{it} \otimes W_{it})/(W_{it}{}' \Omega_t W_{it})^{\frac{3}{2}}, \tag{15}$$

where $Skewness\ r_{i,t+1}$ is the skewness of stock i's return in period t+1, and similarly, α and γ are the parameters that need to be estimated. Based on Equation (15), we define our FDS measure as

$$FDS_{i,t} = -W_{it}{}' M_{3,t}(W_{it} \otimes W_{it})/(W_{it}{}' \Omega_t W_{it})^{\frac{3}{2}}. \tag{16}$$

We put a negative sign in front of the aggregate skewness so that stocks with a higher FDS are more "crash prone".

The FDS measure is higher if 1) a stock's ownership is highly concentrated; 2) the liquidity shocks mutual funds face are highly correlated; and 3) the flow of individual mutual funds is negatively skewed. Moreover, Equation (15) shows that γ captures the magnitude of the impact of FDS: a lower (higher) γ implies that the market can absorb the imbalance of liquidity driven trading with little (much) price impact (Greenwood and Thesmar 2010).

3. Data and Methodology

3.1 Mutual fund data

Quarterly domestic equity mutual fund holdings and total assets are extracted from Thomson Financial from December 1989 to December 2009. Index funds are excluded. . We use the Wharton Financial Institutional Center Number (WFICN) as identifier to identify each mutual fund. For mutual funds that offer different share classes, we aggregate the share classes according to the filing date (Thomson Financial FDATE) to derive the total net assets (TNA). We then calculate the mutual fund return based on the TNA-weighted average across different share classes. At the end of each quarter, equity positions in US dollars of mutual funds are obtained, but we only focus on stocks listed on NYSE and stocks larger than or equal to NYSE decile 5. Setting the size threshold of decile 5 is primarily in order to make the skewness-coskewness matrix computable and because liquidity driven trading of mutual funds can generate observable price pressure effect only if a relatively large portion of outstanding shares is captured, which tends to be the case among larger stocks (Greenwood and Thesmar 2010).

Monthly mutual fund return data is from the Center for Research in Security Prices (CRSP). Thomson Financial mutual fund holding data and CRSP mutual fund return data are linked by the MFLinks file provided by WRDS. We only retain fund-quarters for which both TNA and quarterly returns are observable. After obtaining the mutual fund TNA and quarterly returns, the quarterly flow can be calculated. After the data screening procedure mentioned above, we ended up with 187,138 fund-quarter observations, but for 3.6% of observations (6731 fund-quarters) we find that the ratio of total equity holdings to TNA is larger than 1. We set an upper bound of 1.5 in order to mitigate the potential data quality issues. For the fund-quarter observations with total equity holdings to TNA ratio larger than 1.5 (2.09%, 3917 fund-quarter), we sum up the total current holdings to calculate the alternative total net assets to replace original total net assets that we treat as erroneous. Finally, 5,085 unique mutual funds are observed across the sample period. Summary statistics are reported in Table 1.

[Insert Table 1 about here]

3.2 Stock data

Daily stock information is retrieved from CRSP. The NYSE decile information is from the market capitalization portfolio constructed by CRSP. We assume that the decile of each stock remains unchanged within the year. We select the stocks which have at least once been above the NYSE decile 5. Since mutual funds tend to hold larger stocks, we set the decile 5 threshold to capture observable price impact caused by the liquidity driven trading of mutual funds. Given that the FDS is a third moment measure which is vulnerable to outliers in the sample, we exclude the stocks with less than five mutual fund investors throughout the sample period[6].

[6] We also check other thresholds such as 0, 10, 15, and 20; the results are qualitatively similar, results available on request.

3.3 Computing FDS

The computation of the FDS measure is challenging due to the three dimensions of the skewness-coskewness matrix. Implementing the methodology proposed by Athayde and Flores (1987), we can regard the fund flow covariance matrix as a second moment's tensor and the fund flow skewness-coskewness matrix as a third moment's tensor. For a stock-quarter observation with *n*-holders, we can visualize these tensors as a NxN matrix (variance-covariance) and as a $NxNxN$ cube (skewness-coskewness). Computing the skewness-coskewness matrix is difficult due to the "curse of dimensionality" problem. For example, for a stock-quarter with 500 mutual fund holders, the skewness-coskewness matrix would consist of 500x500x500, or 125,000,000 entries. For a stock-quarter with 1,000 mutual fund holders, the skewness-coskewness matrix would have 10^9 entries. The extremely large matrices make the computation extremely time-consuming (Beardsley, Field, and Xiao, 2012). To make the computation feasible and manageable, we set 500 investors as a threshold for each stock-quarter. The screening procedure is implemented as follow: for the each stock with more than 500 mutual fund investors, we rank all mutual fund investors according to the number of their holding shares at each stock–quarter, and select the top 500 investors with the most shares. Although setting the threshold of 500 investors may result in a loss of information, on average the top 500 investors' shareholding still constitutes 97.62% of the total holding of all mutual funds in each stock-quarter. There are 2.44% stock-quarters (134 unique stocks) with more than 500 mutual fund investors in the sample; on average across the whole sample period, 92 of them belong to NYSE breaking point decile 10, and 24, 15, and 3 of them belong to decile 9, 8, and 7, respectively. After the data screening procedure mentioned, we end up with 84,854 stock-quarters.

Ω_t and M_3 are the variance-covariance matrix and skewness-coskewness matrix of *dollar* flows. However, directly computing these two matrices suffers from the heteroskedasticity problem: the direct dollar flow will overestimate the impact of flow into funds that have declined in size, and underestimate the impact of flow into funds that have increased in size (Greenwood and Thesmar 2010). To circumvent the problem of heteroskedasticity, percentage flow is calculated at the end of each quarter *t* as described in Equation (17)

$$f_{k,t}^{\%} = \frac{TNA_{k.t} - TNA_{k.t-1}(1 + R_{k.t})}{TNA_{k.t-1}} \quad (17)$$

Where $TNA_{k.t}$is the total net assets of mutual fund k at time t, and $R_{k.t}$ is the return of mutual fund k at time t. For each quarter, the percentage $\Omega_t^{\%}$ and $M_{3,t}^{\%}$ are calculated on rolling window basis[7], of which the window size T is five years, or equivalently 20 quarterly observations. Thereafter the covariance-variance matrix Ω_t and coskewness-skewness matrix M_3 are constructed using Equations (18) and (19):

$$\Omega_{i,t} = \frac{1}{t-1} diag(TNA_t) * \left(F_t^{\%} - \overline{F_t^{\%}}\right)' * \left(F_t^{\%} - \overline{F_t^{\%}}\right) * diag(TNA_t), \quad (18)$$

where TNA_t is a vector which collects the total net assets of mutual funds which hold stock i at time t, and $diag(TNA_t)$ is the KxK diagonal matrix whose kth term is $TNA_{k,t}$. Then, $F_t^{\%}$ is a TxN matrix which collects the percentage fund flow time series of N mutual funds which hold stock i at time t, whereby the kth column of $F_t^{\%}$ is the percentage flow time series of mutual fund k. Last, $\overline{F_t^{\%}}$ is the mean percentage flow during the rolling window T.

As Equation (12) suggests, the skewness-coskewness matrix can be viewed as N slices of the original skewness-coskewness cube. The kth slice of the cube is calculated as follows:

$$M_{3k} = \frac{1}{t-1} diag(TNA_t) * \left(F_t^{\%} - \overline{F_t^{\%}}\right)' * (TNA_{k,t} * diag\left(\left(f_{k,t}^{\%} - \overline{f_{k,t}^{\%}}\right)\right) * \left(F_t^{\%} - \overline{F_t^{\%}}\right) * diag(TNA_t),$$

where $f_{k,t}^{\%}$ is the vector of percentage flow time series of mutual fund k during the rolling window T. Then based on Equation (12), M_3 is given by:

$$M_3 = [M_{31} M_{32} M_{33} \dots M_{3k} \dots M_{3n}] \quad (19)$$

Once the Ω_t and M_3 matrices have been constructed, the stock-level fragility and FDS measures can be obtained according to Equations (8) and (16).

3.4 Other Empirical Measures

NSKEW

[7] Experiments with an expanding window yield qualitatively similar results; available on request.

The baseline measure of skewness, which we denote as NSKEW in this paper, is the conventional measurement of skewness which is calculated by dividing the third moment of daily stock return by the standard deviation of stock daily return raised to the third power. In order to match the quarterly frequency of the holdings data provided by Thomson Financial, we choose three months as the time interval starting at January, April, July, and October. Mathematically, NSKEW is calculated as

$$NSKEW_{it} = -(n(n-1)^{\frac{3}{2}} \sum R_{it}^3)/((n-1)(n-2)(\sum R_{it}^2)^{\frac{3}{2}}), \tag{20}$$

where R_{it} is the daily de-meaned arithmetic return of stock i during period t, and n represents the number of observation in period t. We add a negative sign to the raw third such that an increase in NSKEW is associated with a stock being more "crash prone". Using arithmetic return instead of log return adopted by the literature is mainly due to the fact that the FDS is calculated by simple percentage flow other than log flow[8]. Furthermore, because the liquidity shocks should be unrelated to fundamentals, it is interesting to investigate the power of FDS in forecasting the excess skewness calculated using risk-adjusted returns. As we will see, FDS has a slightly higher impact in forecasting the excess skewness of stock returns.

DUVOL

Given that higher-moment calculations are sensitive to data outliers, we use an alternative measure of the stock returns asymmetry, DUVOL (down-to-up volatility), proposed by Hong and Stein (2001).

For stock i we separate the above-average returns from the below-average returns over a three month period. We then calculate the standard deviation for the two subsamples and take the log ratio of the standard deviation of down days to the standard deviation of up days

$$DUVOL_{it} = \log\{((n_u - 1)\sum_{DOWN} R_{it}^2)/((n_d - 1)\sum_{UP} R_{it}^2)\}, \tag{21}$$

where n_u is the number of days with returns above the period mean value and n_d is the number of days with returns bellows the period mean value. Compared to NSKEW, the DUVOL

[8] We also ran all analyses using log-returns and log-flows, but results remained qualitatively similar; available on request.

measure is less sensitive to outliers. As we can see from Equation (21), a higher (lower) DUVOL value represents stock returns with a more left (right) skewed distribution.

Ownership Herfindahl Index

The Herfindahl index, or Herfindahl-Hirschman Index, is designed to measure market concentration, calculated by summing the square of market share of each firm competing in the market. If there is only one firm in the market, the Herfindahl index is equal to one, and if there are 1000 firms fairly competing in the market, the Herfindahl index is nearly equal to zero since each firm only obtains a small fraction of market share. The index is given by

$$\text{Herfindahl Index} = \left(\frac{S_1}{S}\right)^2 + \left(\frac{S_2}{S}\right)^2 + \cdots + \left(\frac{S_n}{S}\right)^2,$$

where S is the total number of shares outstanding, S_1 is the number of shares held by fund 1, and S_2 is the number of shares held by fund 2, and so on.

As suggested by Greenwood and Thesmar (2011), ownership concentration plays a key role in determining the stock's non-fundamental risk caused by liquidity driven trading. Assume a stock with only a few investors; if these investors do not sell this stock in a certain period, then this stock is not exposed to the non-fundamental risk caused by liquidity driven trading. On the other hand, however, if one of these investors sells this stock, then his or her transaction is not likely to be balanced by the trading of other investors. Therefore, similar to the construction of Herfindahl Index, an equivalent measure proposed by Greenwood and Thesmar (2011) to characterize the ownership concentration of stock is given as follow:

$$H_{it} = \left(\frac{S_{1t}}{\text{shrout}_{it}}\right)^2 + \left(\frac{S_{2t}}{\text{shrout}_{it}}\right)^2 + \cdots + \left(\frac{S_{nt}}{\text{shrout}_{it}}\right)^2, \tag{22}$$

where H_{it} is the equivalent Herfindahl index to measure the ownership concentration of stock i at the end of quarter t. shrout_{it} is the share outstanding of stock i held by mutual funds at the end of quarter t. S_{1t} and S_{2t} are shares held by mutual funds 1 and 2 at the end of quarter t, and so on.

NetFlow and NCrossSkew

In order to check whether our FDS is robust to more parsimonious measures that have price pressure impact on the stock price, we construct two additional variables net flow (NetFlow) passed to stock i at time t and the negative cross-sectional skewness (NCrossSkew) of fund flows.

Following Frazzini and Lamont (2008), fund flows pass to stocks through the holding portfolio of funds. Thus, the NetFlow variable is constructed as follow

$$NetFlow_{i,t} = \frac{\sum_k StockFlow_{i,t}^k}{Mkcap_{i,t}} \tag{23}$$

where $StockFlow_{i,t}^k = FundFlow_t^k * Proportion\ of\ stock\ in\ portfolio\ k$

We expect to observe that if more fund flow passed to stock i at time t, the current price will be pushed up and will reverse to the mean in the subsequent periods and therefore exhibit negative skewness.

In addition, since the FDS is constructed on a rolling window basis, it might FDS be less responsive to the newly added observations. Therefore, we construct the variable cross-sectional skewness of fund flows, which ignores the correlation, to check whether the cross-sectional skewness of fund flows is superior to FDS. For each stock i at time t, we calculate the NCrossSkew as following:

$$\text{NCrossSkew}_{i,t} = -\text{skewness}(StockFlow_{i,t}^k) \tag{24}$$

We add a negative sign is in order to make the definition consistent with FDS.

3.5 Baseline Specification

We specify two types of regressions using NSKEW and DUVOL as dependent variables. The sample period starts at December 1989 and the rolling window size is five years. Thus, we begin forecasting the negative skewness in December 1995. We pool all the observations together and specify dummy variables for each quarter to run the Least Square Dummy Variable (LSDV) regressions, thus assuming that the coefficients do not vary over time and across stocks.

We adopt a series of variables suggested by the literature[9] as controls to forecast the one period ahead NSKEW and DUVOL measures calculated using arithmetic returns, market-adjusted returns, and beta-adjusted returns. The regression equation is given by

$$NSKEW_{i,t+1}\left(or\ DUVOL_{i,t+1}\right) = \alpha + \gamma * FBS_{i,t} + \boldsymbol{C} * \boldsymbol{X} + \boldsymbol{\delta} * \boldsymbol{Z} + u_{i,t}, \tag{25}$$

where $\boldsymbol{X}$ is the control variable matrix and $\boldsymbol{C}$ is the vector containing the coefficients loaded on control variables. $\boldsymbol{Z}$ is a matrix containing the year dummy variables. γ is interpreted as the magnitude of FDS impact on firm-level skewness in the next period.

3.6 Descriptive Statistics

Figure 1 shows the evolution over time of the average FDS measure across firms. The average FDS is negative over time, indicating that in general the fund flow is positively skewed in the last two decades. The negative average FDS is not surprising given the dramatic expansion experienced by mutual funds (as shown in Table 1, on average the net mutual fund flows are positive, 3.5% quarterly inflow). Even though in general the FDS is negative over time, it is clearly that the FDS increases in the 2000 Dot-Com crisis and 2007 financial crisis, an observation that is consistent with our expectation for FDS.

[Insert Figure 1 about here]

The summary statistics are reported in Table 2. Panel A reports the mean, median and standard deviation of the major variables classified by the stock's capitalization decile on NYSE breaking point. Panel B presents the correlation matrix of major variables in this paper.

[Insert Table 2 about here]

FDS is negative across all quintiles. There is, however, a strong pattern across size deciles; the FDS is monotonically increasing as the stock size increases. This implies that the aggregate flow of small stocks' current owners is more positively skewed than that of larger stocks. Since the

[9] For example, Muralidhar (1993) and Jondeau (2003) provides evidence that the skewness of firm-level stock returns is persistent. Kapadia (2007) and Chen, Hong and Stein (2001) find that higher volatility of stock returns is associated with positive skewness.

FDS measure not only depends on the magnitude of fund flows but also on the correlation of fund flow across all mutual funds, it is not surprising that we obtain lower FDS value for small stocks which tend to have less owners: as the stock size increase, the number of holders also increases and thus ownership concentration decreases. Hence, if one of the owners of a larger stock experiences a liquidity shock, his or her trading is more likely to be offset by other owners. This finding is consistent with the finding of Chen, Hong and Stein (2001) who find that small stocks are more positively skewed than larger stocks. They explain this phenomenon as a managerial issue: small stocks undertake less scrutiny than larger stocks, and thus the manager of small stocks prefers to announce good news immediately and let bad news dribble out slowly. The fact that FDS increases monotonically with stock size is consistent with Chen, Hong and Stein's (2001) finding, but the negative FDS fails to explain why large stocks are negatively skewed as documented by Chen, Hong and Stein. One explanation for this is that mutual funds on average only hold 15% of outstanding shares. It is possible that the aggregate flow of large stocks' mutual fund owners is positively skewed, but the aggregate flow of all of the large stocks' owners is negatively skewed. Unfortunately, obtaining the flow data of all contemporaneous owners is unrealistic. The focus of this paper, however is to forecast the ex-ante firm-level negative skewness.

DUVOL has the same characteristics as NSKEW: the value of DUVOL is increasing as stock size increases. Furthermore, as shown in Panel B, the correlation coefficient between DUVOL and NSKEW is 0.92. Given that the construction of the two measures is quite different, they actually seem to pick up the same information, implying that the DUVOL measure can serve as an alternative measure of asymmetry of stock returns.

Results of STD and past returns shown in Table 2 are as documented by a large body of literature. STD monotonically increases as stock size decreases, therefore higher returns are asked by investors for bearing higher risk (as shown in the RET1 and RET2 measure). This phenomenon refers to "small-firm effect", which is well known in the literature (see Banz, 1981). However, as presented in Panel B, the correlation coefficient between the STD and NSKEW is slightly positive, supporting the volatility-feedback hypothesis of Campbell and Hentschel (1992) that higher levels of volatility are accompanied by more negative skewness.

Therefore, we incorporate the STD variable in all regressions as a control variable to ensure that we forecast the skewness rather than the volatility.

The relatively high negative correlation between the Fragility measure and FDS measure shown in Panel B illustrates that the aggregate volatility of fund flow is mainly caused by fund *inflow*, which is not surprising given the dramatic growth rate in the mutual fund industry since 1990s. Given such relatively high correlation between the Fragility and FDS measures, one may suspect that the FDS virtually picks up the same information as Fragility does. FDS is, however, still statistically and economically significant even after controlling the Fragility variable, showing that FDS, compared with Fragility, indeed picks up other information in forecasting the next period negative skewness. Other variables in the cross-sectional regressions do not suffer from the multicollinearity problem given the low correlations between each pair of assets.

4. Forecasting Skewness in the Cross-section of Stock Returns

Table 3 reports the estimation results of the baseline specification.

[Insert Table 3 about here]

The coefficients on FDS are positive and highly significant across all six different forms of asymmetry of stock return. In forecasting one period ahead NSKEW, the high t-statistics indicates that FDS indeed has statistically significant power in forecasting the one period ahead firm-level skewness. All the coefficients on FDS carry positive signs demonstrating that the FDS measure predicts the ex-ante conditional firm-level skewness in the same direction. This implies that if the aggregate flow of a stock's current mutual fund owners is negatively skewed during quarter *t*, then the return of the stock also exhibits negative skewness in the next period. One standard deviation of FDS increase will lead to a 36% increase in negative skewness.

The positive sign for current NSKEW supports the findings of other authors. For example, Muralidhar (1993) uses a bootstrap approach to test the persistence of skewness and finds that for a large portion of stocks the skewness persists over time. Jondeau (2003) also provides evidence that the skewness of firm-level stock returns is persistent.

Current STD is also highly significant, even after controlling for the current skewness. Incorporating the current volatility as control variable is to ensure that we forecast the ex-ante skewness rather than the volatility. The phenomenon that current volatility predicts skewness has been documented by large body of recent literature. Specifically, Kapadia (2007) documents that there are high volatility-high skewness and low volatility-low skewness relationships for individual stocks. In addition, Chen, Hong and Stein (2001) also find similar results that higher volatility of stock returns is associated with positive skewness. One explanation for this phenomenon is that when current volatility is high, investors ask for a higher premium to compensate the risk they are bearing, and hence drive down the current price. Thereafter the price reverses in the next period and exhibits positive skewness. Past returns also have strong predictive power in forecasting skewness. The positive coefficients on current quarter return and two quarter lagged return illustrate that higher past return predict lower skewness in the next period. The result that high past returns are associated with negative skewness is parallel to previous works by Harvey and Siddique (2000) and Chen, Hong and Stein (2001).

It is rather surprising that MF strongly predicts the negative skewness, implying that the more shares held by mutual funds the more negative skewness of the stock returns. Previous studies have argued that institutional investors behave less like “noise traders” than retail investors, and they stabilize the market by trading against irrational investors (e.g. Shiller, 1992, Josef Lakonishok 1994). It is surprising to see that stocks which are held more by mutual funds exhibit more negative skewness, given that mutual funds on average experienced *net inflows* in the last two decades. However, this result reflects a different story consistent with previous work of Sias (1996), who found that an increase in the number of institutional investors is associated with rising volatility. Our results show that the downside volatility even increases more than upside volatility.

There are several reasons why stock holding by mutual funds might destabilize the market. First, in contrast with retail investors, funds tend to engage in large volume transaction, which results in large fluctuation of stock price (Lee, 1992). Second, an increase in the number of owners may lead to information being interpreted differently, which is similar to the theory proposed by

Hong and Stein (1999). Hong and Stein argue that the dispersion of opinions of investors plays an important role in explaining stocks' negative skewness. After controlling for MF, the coefficient on FDS remains positive and highly significant.

Due to the highly negative correlation between MF and the Herfindahl H[10], as shown in Table 2 panel B, we do not put these two variables in the regressions at same time to circumvent multicollinearity issues. However, we also replace the MF by Herfindahl H. The results show that the coefficient of the Herfindahl H is -0.623 with a t-statistic of -8.10, which confirms the previous discussion: stocks with a higher Herfindahl H tend to display higher skewness. One explanation for this phenomenon is that when mutual funds receive capital inflow, they tend to expand their existing position (Coval and Stafford, 2007, and Dong Lou 2012); stocks with concentrated ownership are more sensitive to this capital inflow since the flows of mutual funds are more likely to be correlated. In columns (4), (5), and (6) of Table 3, we use DUVOL calculated using arithmetic return, market-adjusted return, and beta-adjusted return as an alternative measure of asymmetry of stock returns. Compared to NSKEW, all the coefficients have a slightly lower value but this is due to the fact that the order of magnitude of DUVOL is different from NSKEW; the coefficients carry slightly higher t-values. Even though the construction of DUVOL is quite different with that of NSKEW, none of results obtained by using NSKEW as dependent variable changed.

4.2 Robustness Checks

In this section, several estimation methods and different control variables are adopted to check the robustness of the FDS measure. Table 4 reports the results. In column (1), we simply regress FDS on one period ahead NSKEW. Results show that the FDS is statistically significant and strongly predicts firm-level negative skewness, even though the R^2 obtained is slightly lower than other specifications. In column (2), we regress FDS on one period ahead NSKEW but control for the share held by mutual fund (MF) and the ownership concentration (Herfindahl H). Consistent with the results in Table 3, the coefficient on FDS is positive and remains highly

[10] The negative correlation between the MF and Herfindahl H can be interpreted as that as the number of shares held by mutual funds increase, it is more possible that more mutual fund investor get involved into the stock market than because original mutual funds investors expand their existing position.

significant. MF and Herfindahl H also carry the expected sign as discussed before. The results suggest that for stocks with similar ownership, the more negatively skewed the aggregate flow of their current owners, the more negative skewness the stock exhibits in the next period.

[Insert Table 4 about here]

Considering the high correlation between the FDS measure and Fragility measure, one might suspect that these two measures virtually pick up the same information. We include the Fragility measure as control variable in the baseline specification, and with other control variables suggested by conventional literatures such as past negative skewness, past volatility, and past cumulative return. The result is shown in column (3). Both FDS measure and Fragility measure are highly significant and carry the anticipated sign. One standard deviation increase in FDS leads to an increase in negative skewness by 0.02 (0.034*0.6), while one standard deviation increase in Fragility measure leads to an decrease in negative skewness by 0.04 (-0.029*1.4). While FDS remains statistically significant and predicts the negative skewness, Fragility predicts the positive skewness, indicating that FDS does pick up different information.

In column (4) and (5) we include the NetFlow and CrossSkew variable as control variable. None of these two variables wipes out the FDS. Notice that we add a negative sign in front of the cross-sectional skewness of fund flows, the lower NCrossSkew value is, the more *positively* skewed the cross-sectional skewness of fund flows is. Thus, these two variables could share similar interpretation that the more fund flow pass into stock at current period, the stocks' price are pushed up and reverse in the next period, and subsequently exhibit negative skewness.

Chen, Hong and Stein (2001) argue that the dispersion of opinions of investors plays an important role in making stock display negative skewness, and that the turnover of stocks can proxy for the degree of the difference of opinions. Therefore, we add *Turnover* in the baseline regression in column (6). Consistent with Chen, Hong and Stein's (2001) results, the coefficient on *Turnover* is positive and highly significant, demonstrating that *Turnover* strongly predicts negative skewness. FDS is still robust to including *Turnover*.

In addition to the baseline specification in which market-adjusted return and beta-adjusted return are used to calculate the skewness and standard deviation, we add a specification in column (7) using Fama-French 3-factor model adjusted returns to calculate the skewness and standard deviation. As the result show, the coefficients of FDS are persistent across different designs. We expect that FDS has higher predictive power in forecasting the excess skewness, since liquidity driven trading of mutual fund is non-fundamental and short-lived, and therefore FDS should have some stronger power to capture the characteristics of excess skewness based on the nature the construction of FDS. In forecasting the skewness calculated by Fama-French 3-factor model adjusted returns, the coefficient of FDS is slightly higher than skewness calculated by beta-adjusted return (0.059 vs. 0.056), which is consistent with our expectations.

Columns (8), (9), and (10) are regressions using different estimation methods. Specifically, in the Fama-Macbeth regression shown in column (8) we follow the conventional practice and report t-statistics by calculating Newey-West standard errors on the time-series of slope coefficients. After adjustment, t-statistics value drop from 6.42 to 3.51 compared to LSDV estimation, but still remain highly significant and none of the conclusions mentioned previously changed. In column (9), we adjust the clustered standard errors due to the concern that observations across each stock may not be independent. After adjusting for clustered standard errors, all the coefficients remain significant. In column (10), we estimate a fixed effect panel regression using firm fixed effect[11] and the main results remain.

6. Conclusion

This paper develops a methodology to forecast the skewness in the daily returns of individual stocks based on its ownership structure. We use US stocks between 1990 and 2009, and quarterly mutual fund ownership data to construct the Flow-Based Skewness (FDS) measure. FDS takes (1) the number of stocks' current owners and ownership concentration, (2) the correlation of flows between stocks' owners, and (3) the asymmetry of flows of stocks' owners into account.

[11] We ran Hausman and Breusch-Pagan Lagrange Multiplier tests to test for fixed versus random effects. Both tests are highly significant (with p-value less than 0.0001), suggesting that a fixed effect regression is preferable. However, we also have tried a random effect regression; all the coefficients are significant and predict the expected direction as discussed previously.

The results show that FDS is highly statistically significant and strongly predicts the conditional firm-level skewness: stocks that display more negative skewness are also the ones with higher FDS value. In addition, the ownership concentration and the number of shares held by mutual funds also play an important role in determining firm-level skewness. The higher ownership concentration of stock is, the higher skewness the stock displays. On the other hand, the number of shares held by mutual funds is negatively related to skewness, which is consistent with the findings of Sias (1996). The coefficient on FDS is highly significant and persistent under different robustness specifications. This finding suggests that liquidity driven trading combined with the ownership structure can be used to forecast the excess movement of stock price.

References

Albuquerque, R. (2010). Skewness in Stock Returns: Reconciling the Evidence on Firm versus Aggregate Returns. *Journal of Economics Literature*.

Almazan, A. B. (2004). Why constrain your mutual fund manager? *Journal of Financial Economics*, 73(2), 289-321.

Bates, D. (1991). The crash of 87: 'was it expected?' The evidence from options markets. *Journal of Finance*, 1009-1044.

Bates, D. (2000). Post-’87 crash fears in S&P 500 futures options. *Journal of Econometrics*, 94(1-2), 181-238.

Beedles, W. L. (1979). On the Asymmetry of Market Returns. *Journal of Financial and Quantitative Analysis*.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1986 Meetings of the American Statistical Association* (pp. 177-181). Business and Economical Statistics Section.

Campbell, J. H. (1992). No news is good news: an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 281-318.

Christie, A. (1982). The stochastic behavior of common stock variances- value, leverage and interest rate effecs. *Jornal of Finanicial Economics*, 407-432.

Coval, J. D. (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 479-512.

Dittmar, R. F. (2002). Nonlinear Pricing Kernels, Kutosis Preference, and Evidence from the Cross Section of Equity Returns. *The Journal of Finance*, 369-403.

Edelen, R. M. (2001). Aggregate price effects of institutional trading: A study of mutual fund flow and market returns. *Journal of Financial Economics*, 195–220.

Frazzini, A. a. (2008). Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics*, 299–322.

French, K. R. (1987). Expected Stock Returns and Volatility. *Journal of Financial Economics*, 3-29.

Glosten, L. R. (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Joumal of Finance*, 1779-1801.

Goetzmann, W. N. (2003). Index funds and stock market growth. *Journal of Business*, 1-28.

Greene. (1993). *Econometric Analysis.* New York: Macmillan.

Gustavo M. de Athayde, R. G. (2004). Finding a maximum skewness portfolio-a general solution to three-moments portfolio choice. *Journal of Economic Dynamics and Control*, 1335-1352.

H. Russel Fogler, R. C. (1974). A Note on Measurement of Skewness. *Journal of Financial and Quantitative Analysis*, 485.

Hong, H. S. (2003). Differences of opinion, short-sales constraints, and market crashes. *Review of financial studies*, 16(2), 487-525.

Jondeau, R. (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics & Control*, 1699-1737.

Josef Lakonishok, A. S. (1994). Contrarian Investment, Extrapolation, and Risk. *The journal of Finance*, 1541-78.

Joseph Chen, H. H. (2001). Forecasting Crashes: trading volume, past returns, and donditional skewness in stock prices. *Journal of Financial Economics*, 345-381.

Kapadia, N. (2007). Skewness, Idiosyncratic Volatility, and Expected Returns. *Chapel Hill*.

Koski, J. P. (1999). How are derivatives used? Evidence from the mutual fund industry. *Journal of Finance*, 791-816.

Lee, C. (1992). Earnings News and Small Traders. *Jounal of Accounting and Economics*, 265-302.

Muralidhar, K. (1993). The Bootstrap Approach for Testing Skewness Persistence. *Management Science*.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 347-370.

Nicholas Barberis, A. S. (2003). Style investing. *Journal of Financial Economics*, 161-199.

Pindyck, R. (1984). Risk, inflation, and the stock market. *American Economic Review*, 334-351.

Scowcroft, S. S. (2000). *Advances in Portfolio Construction and Implementation.* Butterworth-Heinemann, First Edition.

Shiller, R. (1992). Stock Prices and Social Dynamics. *Brookings Papers on Economic Activity*, 235-68.

Sias. (1996). Volatility and the Institutional Investor. *Financial Analysits Journal*.

Siddique, C. R. (2000). Conditional Skewness in Asset Pricing Tests. *The Journal of Finance*, 1263-1295.

Stephen Figlewski, X. W. (2000). Is the 'Leverage Effect' a Leverage Effect? *Journal of Economics Literature*.

Turner, C. M. (1989). A Markov Model of Heteroskedasticity, Risk,and Learning in the Stock Market. *Journal of Financial Economics*, 3-22.

Warther, V. (1995). Aggregate mutual fund flows and security returns. *Journal of Financial Economics*, 209-35.

Figure 1 the evolution of the average FDS over time

The sample period runs from December 1989 to December 2009. FDS stands for the flow-based-skewness, which is calculated from the aggregate skewness of fund flow scaled by the square root of second moment of fund flow raised to third power: $FBS_{i,t} = -W_{it}'M_{3,t}(W_{it} \otimes W_{it})/(W_{it}'\Omega_t W_{it})^{3/2}$. FDS is winsorized at 0.5% level.

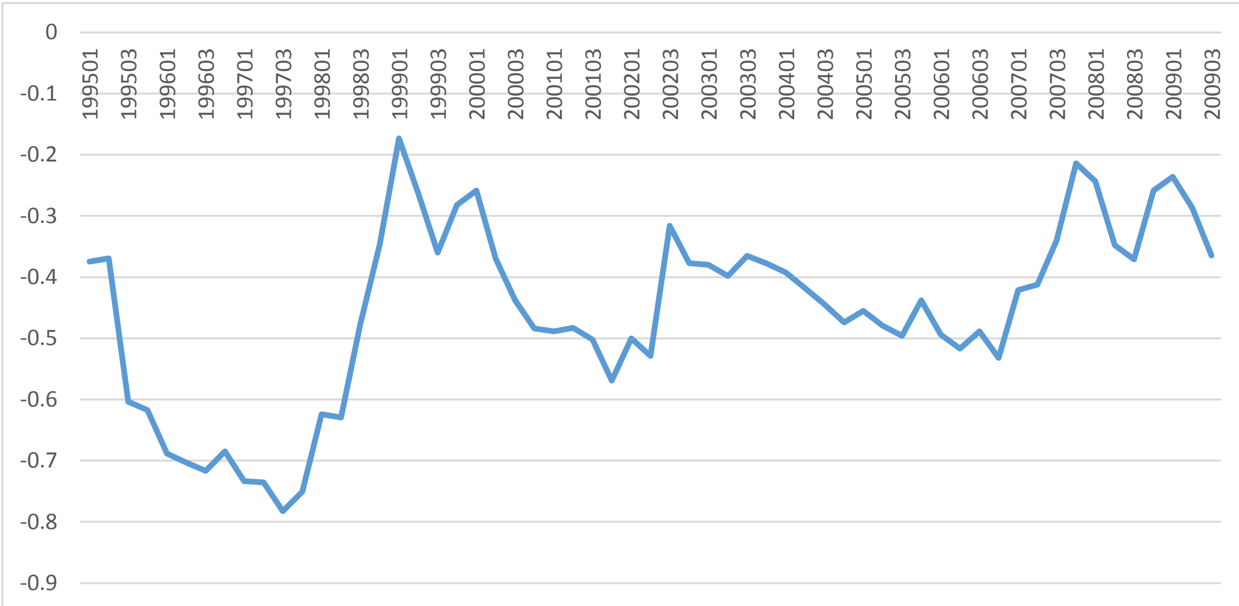

Table 1 summary statistics of Fund Flows

This table reports the summary statistics of mutual fund sample at the end of each year. The sample period covers Dec 1989 to Dec, 2009. Holding information and mutual fund total assets are from Thomson Financial and mutual fund return is from CRSP. These two data sets are merged through the MFLinks file provided by WRDS. No.Funds refers to the number of mutual funds at the end of each year and No.Stocks denotes the number of stocks selected in sample, which are stocks larger than the NYSE breaking point decile 5. Mutual fund quarterly flow is calculated as $f_{k,t}^{\%} = \frac{TNA_{k.t} - TNA_{k.t-1}(1+R_{k.t})}{TNA_{k.t-1}}$, where $TNA_{k.t}$ is the total net asset of mutual fund k at time t, and $R_{k.t}$ is the fund return of mutual fund k at time t. Mutual Fund TNA is the total asset value of mutual fund reported by Thomson Financial. % Stocks Held by Mutual Fund is calculated by dividing the sum of equity dollar position of all mutual fund by the sum of market capitalization value of stocks.

			Mutual Fund Quarterly Flow%		**Mutual Fund TNA ($ Million)**		**% Stocks Held by Mutual Fund**
Year	**No.Funds**	**No.Stocks**	**Mean**	**Median**	**Mean**	**Median**	
1989	647	767	0.83	-1.15	358.95	102.57	3.66
1990	704	804	2.18	-0.87	324.19	89.48	3.91
1991	822	873	6.11	0.97	425.52	111.57	4.45
1992	948	970	7.26	2.26	491.86	124.7	5.16
1993	1269	1091	7.46	2.69	547.86	127.4	6.4
1994	1538	1170	3.43	-0.23	527.63	107.52	7.76
1995	1689	1241	4.7	0.9	707.27	136.1	8.91
1996	1946	1347	6.39	0.72	835.05	147.77	10.29
1997	2180	1456	5.55	1.01	1004.35	162.76	11.35
1998	2482	1521	2.84	-1.07	1090.73	155.7	12.18
1999	2806	1509	5	-1.44	1251.36	161	13.49
2000	2988	1450	3.81	-0.47	1164.48	164.1	14.58
2001	3112	1444	4.67	-0.28	1020.18	149.25	15.2
2002	3097	1481	1.87	-1.3	809.35	118.7	15.64
2003	3090	1490	6.13	1.08	1087.13	168.15	16.25
2004	3102	1536	3.58	-0.89	1270.08	195.4	17
2005	3035	1548	2.46	-1.08	1443.39	219.3	17.7
2006	2960	1550	0.09	-2.12	1745.47	270.1	17.73
2007	2952	1515	-0.95	-2.06	1924.18	300	18.19
2008	2778	1492	-2.18	-3.41	1206.46	194.3	17.83
2009	2486	1400	-2.42	-2.13	1678.44	282.25	17.03

Table 2 Summary Statistics of Major Variables in Regressions

The sample period runs from December 1989 to December 2009. FDS stands for the flow-based-skewness, which is calculated from the aggregate skewness of fund flow scaled by the square root of second moment of fund flow raised to third power: $FDS_{i,t} = -W_{it}{'}M_{3,t}(W_{it} \otimes W_{it})/(W_{it}{'}\Omega_t W_{it})^{3/2}$. Fragility is consistent with measure of Greenwood and Thesmar (2011), which is calculated by $G_{it} = (1/\theta_{it})^2 W_{it}{'}\Omega_t W_{it}$. *Owner* represents the log-transformed number of mutual fund investors and Herfindahl H is equivalent to the Herfindahl Index to measure the ownership concentration. MF is the shares held by mutual fund in percentage form. The NSKEW is the negative standard measure of skewness, which is the raw third moment of stock return scaled by volatility raised to third power. DUVOL is the log ratio of DOWN-DAYS to UP-DAYS standard deviation over one quarter period. STD represent the standard deviation over one quarter period. RET1 and RET2 are the cumulative return over one quarter period and two quarter period respectively. Arithmetic, Market Adjusted, and Beta Adjusted in the parentheses after NSKEW, DUVOL, and STD represent that the according variables are calculated using simple arithmetic return, market adjusted return, and beta adjusted return respectively. Market adjusted return is calculated by simple arithmetic return less the market return which is extracted from Fama-French Website. Beta adjusted return is the residual in the CAPM regression, and the Beta coefficients are allowed varying across individual stocks. In panel B, NSKEW, DUVOL, and STD are calculated by simple arithmetic return. FDS and Fragility are winsorized at 0.5% level.

Panel A: Descriptive statistics of major variables

	Size Quintile	All Firms	Quintile 1 (Smallest)	Quintile 2	Quintile 3	Quintile 4	Quintile 5 (Largest)
FDS	**Mean**	-0.452	-0.692	-0.624	-0.550	-0.451	-0.237
	Median	-0.409	-0.651	-0.569	-0.515	-0.412	-0.210
	Stdev	0.597	0.645	0.618	0.591	0.582	0.528
Fragility	**Mean**	0.913	1.036	1.228	1.243	0.952	0.416
	Median	0.410	0.285	0.568	0.644	0.520	0.221
	Stdev	1.440	1.797	1.825	1.751	1.331	0.651
Investor	**Mean**	128	28	50	72	108	250
	Median	93	22	45	68	103	237
	Stdev	115	21	31	42	60	133
Herfindahl H	**Mean**	0.116	0.249	0.165	0.132	0.107	0.066
	Median	0.080	0.196	0.129	0.097	0.077	0.048
	Stdev	0.115	0.168	0.130	0.118	0.102	0.069
MF	**Mean**	0.184	0.125	0.162	0.186	0.196	0.192
	Median	0.179	0.104	0.147	0.176	0.193	0.191
	Stdev	0.108	0.143	0.119	0.111	0.106	0.090
NSKEW (Arithmetic)	**Mean**	-0.092	-0.535	-0.245	-0.112	-0.061	0.041
	Median	-0.177	-0.423	-0.238	-0.193	-0.158	-0.129
	Stdev	1.381	1.354	1.207	1.353	1.424	1.426
NSKEW (Market Adjusted)	**Mean**	-0.077	-0.515	-0.229	-0.098	-0.044	0.050
	Median	-0.164	-0.392	-0.219	-0.177	-0.138	-0.122
	Stdev	1.464	1.339	1.248	1.440	1.511	1.535
NSKEW (Beta Adjusted)	**Mean**	-0.090	-0.535	-0.247	-0.110	-0.058	0.043
	Median	-0.177	-0.412	-0.239	-0.192	-0.153	-0.127
	Stdev	1.485	1.373	1.276	1.459	1.536	1.549
DUVOL (Arithmetic)	**Mean**	-0.126	-0.397	-0.212	-0.138	-0.105	-0.052
	Median	-0.164	-0.357	-0.217	-0.173	-0.154	-0.124
	Stdev	0.735	0.858	0.685	0.708	0.743	0.740
DUVOL (Market Adjusted)	**Mean**	-0.100	-0.360	-0.182	-0.112	-0.078	-0.029
	Median	-0.139	-0.316	-0.186	-0.146	-0.121	-0.106
	Stdev	0.746	0.765	0.651	0.710	0.759	0.793
DUVOL (Beta Adjusted)	**Mean**	-0.112	-0.383	-0.196	-0.126	-0.091	-0.037
	Median	-0.152	-0.343	-0.207	-0.162	-0.134	-0.113
	Stdev	0.762	0.799	0.674	0.727	0.776	0.798

STD (Arithmetic)	**Mean**	0.200	0.339	0.225	0.205	0.187	0.175
	Median	0.162	0.249	0.183	0.171	0.156	0.142
	Stdev	0.218	0.716	0.233	0.164	0.125	0.134
STD (Market Adjusted)	**Mean**	0.186	0.333	0.214	0.191	0.173	0.157
	Median	0.151	0.246	0.176	0.161	0.145	0.128
	Stdev	0.211	0.713	0.225	0.153	0.112	0.123
STD (Beta Adjusted)	**Mean**	0.183	0.329	0.211	0.188	0.170	0.155
	Median	0.148	0.243	0.173	0.158	0.143	0.125
	Stdev	0.210	0.711	0.224	0.153	0.111	0.122
RET1	**Mean**	0.025	0.146	0.044	0.024	0.015	0.008
	Median	0.019	0.044	0.025	0.016	0.018	0.019
	Stdev	0.299	0.763	0.358	0.276	0.220	0.207
RET2	**Mean**	0.045	0.260	0.077	0.042	0.028	0.016
	Median	0.027	0.060	0.035	0.024	0.024	0.026
	Stdev	0.461	1.353	0.492	0.422	0.316	0.285
Number of Observations		84854	3691	12134	21239	23418	24372

Panel B: Correlation Matrix (p-Value reported in bold)

	FDS	Fragility	Owner	Herfindahl H	MF	NSKEW	DUVOL	STD	RET1	RET2
FDS	1									
Fragility	-0.201	1								
	<.0001									
Owner	0.313	-0.046	1							
	<.0001	**<.0001**								
Herfindahl H	-0.179	0.036	-0.729	1						
	<.0001	**<.0001**	**<.0001**							
MF	0.139	0.470	0.513	-0.386	1					
	<.0001	**<.0001**	**<.0001**	**<.0001**						
NSKEW	0.009	-0.010	0.074	-0.058	0.032	1				
	0.0097	**0.0037**	**<.0001**	**<.0001**	**<.0001**					
DUVOL	0.016	-0.020	0.088	-0.062	0.034	0.920	1			
	<.0001	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**				
STD	0.025	0.014	-0.041	0.034	0.017	0.026	-0.030	1		
	<.0001	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**			
RET1	-0.023	-0.004	-0.040	0.020	-0.018	-0.369	-0.412	0.180	1	
	<.0001	**0.2048**	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**	**<.0001**		
RET2	-0.036	-0.004	-0.036	0.011	-0.013	-0.195	-0.216	0.059	0.648	1
	<.0001	**0.2305**	**<.0001**	**0.0017**	**0.0002**	**<.0001**	**<.0001**	**<.0001**	**<.0001**	

Table 3 Forecasting Skewness: Baseline Specification

The sample period runs from December 1995 to December 2009. The dependent variables of column 1 to 3 are one period ahead NSKEW and of column 4 to 6 are one period ahead DUVOL calculated by simple arithmetic return, market adjusted return, and beta adjusted return respectively. FDS stands for the flow-based-skewness, which is calculated from the aggregate skewness of fund flow scaled by the square root of second moment of fund flow raised to third power: $FDS_{i,t} = -W_{it}{}'M_{3,t}(W_{it} \otimes W_{it})/(W_{it}{}'\Omega_t W_{it})^{3/2}$. The NSKEW is the standard measure of skewness, which is the negative raw third moment of stock return scaled by volatility to the power three. DUVOL is the log ratio of UP-DAYS to DOWN-DAYS standard deviation over one quarter period. STD represents the standard deviation over one quarter period. Herfindahl H is the equivalent to Herfindahl Index to measure the ownership concentration. MF is the shares held by mutual fund in percentage form. RET1 and RET2 are the cumulative return over one quarter period and two quarter period respectively. FDS measure is winsorized at 0.5% percent level. All the regressions are Least Square Dummy Variable (LSDV) regressions containing dummy variables for each quarter (not shown); t-statistics is reported in bold.

	$SKEW_{t+1}$ Arithmetic	$SKEW_{t+1}$ Market Adjusted	$SKEW_{t+1}$ Beta Adjusted	$UDVOL_{t+1}$ Arithmetic	$UDVOL_{t+1}$ Market Adjusted	$UDVOL_{t+1}$ Beta Adjusted
FDS_t	0.055	0.053	0.056	0.033	0.031	0.034
	6.42	**5.93**	**6.09**	**7.41**	**6.77**	**7.23**
$NSKEW_t$	0.035	0.030	0.031	0.024	0.023	0.022
	8.94	**7.53**	**7.79**	**11.77**	**11.59**	**11.1**
STD_t	-0.579	-0.662	-0.638	-0.359	-0.428	-0.405
	-21.66	**-23.15**	**-21.9**	**-25.38**	**-29.77**	**-27.4**
MF_t	0.336	0.305	0.341	0.100	0.082	0.111
	7.00	**6.01**	**6.59**	**3.93**	**3.19**	**4.22**
RET1	0.412	0.440	0.439	0.258	0.282	0.279
	15.26	**15.38**	**15.09**	**18.11**	**19.58**	**18.85**
RET2	0.104	0.117	0.113	0.051	0.058	0.056
	6.90	**7.37**	**6.99**	**6.39**	**7.22**	**6.78**
Intercept	-0.259	-0.465	-0.461	-0.134	-0.291	-0.291
	-6.39	**-10.88**	**-10.6**	**-6.25**	**-13.51**	**-13.14**
R^2	0.024	0.024	0.023	0.037	0.038	0.035
No. of Obs.	80751	80751	80751	80749	80749	80749

Table 4 Forecasting Skewness: Robustness Check

The sample period start from December, 1995 to December, 2009. The dependent variable is one period ahead NSKEW calculated by using arithmetic return. In column (9) stock-level skewness is based on the excess return of Fama-French 3-factor model, and all factors are extracted from Fama-French website. In column (10) Fama-Macbeth regression, standard errors are corrected using Newey-West adjustment with 3 lags. FDS and Fragility are winsorized at 0.5% level. Least Square Dummy Variable (LSDV) regressions containing dummy variables for each quarter (not shown); all t-statistics are reported in bold.

	Direct Regression	Control Herfindahl H and MF	Compare with Fragility	Control NetFlow	Control NCrossSkew	Control Turnover	FF 3 factors adjusted	Fama Macbeth Regression	Correct for Stdev Clustering	Panel Fixed effect
	1	2	3	4	5	6	7	8	9	10
FDS_t	0.064	0.041	0.034	0.06	0.058	0.036	0.059	0.052	0.036	0.028
	7.66	**4.8**	**3.78**	**7.01**	**6.79**	**4.12**	**5.83**	**3.51**	**4.21**	**2.86**
$Fragility_t$			-0.029							
			-7.23							
Herfindahl H_t		-0.604								
		-12.99								
MF_t		0.139	0.623	0.286	0.323	0.221	0.372	0.368	0.321	0.744
		2.83	**10.38**	**5.82**	**6.72**	**4.45**	**6.46**	**3.44**	**5.87**	**9.23**
NetFlow				3.907						
				5.08						
NaCrossSkew					-0.01					
					-8.85					
$Turnover_t$						0.03				
						9.51				
$NSKEW_t$			0.034	0.034	0.033	0.034	0.023	0.054	0.036	-0.015
			8.48	**8.74**	**8.51**	**8.73**	**5.9**	**9**	**8.45**	**-3.67**
STD_t			-0.567	-0.577	-0.573	-0.603	-0.588	-0.773	-0.237	-0.463
			-20.69	**-21.58**	**-21.45**	**-22.47**	**-18.11**	**-8.06**	**-3.98**	**-14.98**
RET1			0.403	0.408	0.407	0.419	0.396	0.513	0.253	0.292
			14.58	**15.1**	**15.07**	**15.52**	**12.38**	**8.94**	**8.83**	**10.63**
RET2			0.107	0.1	0.1	0.105	0.115	0.303	0.313	0.139
			7	**6.67**	**6.65**	**6.98**	**6.38**	**5.42**	**15.13**	**9.05**
Intercept	-0.114	-0.097	-0.311	-0.244	-0.254	-0.376	-0.518	0.006	-0.096	1.449
	-3.01	**-2.44**	**-7.52**	**-5.99**	**-6.26**	**-8.89**	**-10.7**	**0.19**	**-5.68**	**1.96**
R^2	0.013	0.016	0.025	0.024	0.025	0.025	0.016	0.029	0.032	0.063
Obs	81663	81663	80001	80751	80751	80751	80751	80751	80751	80751
Estimation	LSDV	LSDV	LSDV	LSDV	LSDV	LSDV	LSDV	FM	Cluster Stdev	Panel FE