Lottery and bubble stocks and the cross-section of option implied tail risks

Sobhesh Kumar Agarwalla[a,*], Sumit Saurav[a,**], Jayanth R. Varma[a,*]

[a]*Indian Institute of Management Ahmedabad, Vastrapur, Ahmedabad, Gujarat 380015, India,*

Abstract

The option smile provides forward-looking information about the risk at the center of the distribution and at the tails. We investigate the cross-sectional determinants of the option smile using indices that capture firm fundamental risks, heterogeneity of belief, lottery characteristics and bubble characteristics. We find that the ATM volatility is largely explained by historical risks and predicted future risks measured using accounting-based risk measures and firm characteristics. However, the cross-sectional variation in the skew is driven by risk premia and by buying and selling pressure that is influenced by heterogeneity in belief and the underlying's lottery-like and bubble-like characteristics.

Keywords: Volatility smile, Volatility skew, Bubble stocks, Lottery stocks, Emerging Markets

JEL: G11, G12, G13

Declarations of interest: None

---

[*]Professor, Finance and Accounting Area Address: Wing 7, Indian Institute of Management Ahmedabad, Vastrapur, Ahmedabad, Gujarat 380015, India, Phone No. +91.79.71524865.

[**]Doctoral Student, Address: MSH-1402, Indian Institute of Management Ahmedabad, Vastrapur, Ahmedabad, Gujarat 380015, India, Phone No. +91.8809388919.

Email addresses: sobhesh@iima.ac.in (Sobhesh Kumar Agarwalla), phd19sumits@iima.ac.in (Sumit Saurav), jrvarma@iima.ac.in (Jayanth R. Varma)

1. Introduction

The existence of volatility "smile" has been one of the most intriguing phenomena in the derivatives literature. It is known to reveal information about the tail risk of price crashes and booms and about more typical risks (Bakshi et al., 2003; Kim and Zhang, 2014). While at-the-money (ATM) options (with strikes close to the current stock price) reflect information about the center of the distribution, options with significantly lower and higher strikes reveal information about the left and right tail risk, respectively (Agarwalla et al., 2021). However, since option prices are derived from the risk-neutral distribution and not the physical distribution, they also reflect a variety of risk premia. In the presence of trading frictions, they might also be influenced by buying and selling pressure in the spot and derivative markets. Moreover, options markets are forward-looking and do not blindly extrapolate the recently observed historical distribution into the future. Accounting information on firm fundamentals is known to predict future firm risk (Bai and Wu, 2016) and would therefore impact the option smile. Thus the cross-sectional variation of the option smile would depend not only on the variance and other moments of the historical return distribution but also on accounting-based firm fundamentals and variables that capture risk premia and buying/selling pressure. To provide a parsimonious analysis of the smile, we start with a comprehensive set of variables representing all these factors and reduce them to indices that capture fundamental risk, heterogeneity in belief, lottery-like stock characteristics, and bubble-like stock characteristics.

India provides a unique setting for studying the cross-sectional variation of various risk parameters across different firms. The Indian derivatives market is one of the world's largest single stock options (SSO) and single stock futures (SSF) markets by volume and is unique in several ways (FIA, 2019). First, with both SSF and SSO contracts being traded in the same platform and both being very liquid, the pricing efficiency of SSO contracts is found to be good (Jain et al., 2019). Second, the liquid SSF market provides an attractive (through short-selling and leverage advantages) avenue for informed investors to exploit their private information. This leaves the SSO market for investors who desire to exploit the asymmetry in the payoff from option contracts (downside protection) by paying an upfront cost. Third, India has a liquid equity derivative market on about 200 large companies whose stocks are liquid and followed by many analysts. These companies belong to diverse industries, and their heterogeneity can be expected to generate substantial cross-sectional variation in the option smile.

Unlike in developed countries where the smile parameters are readily available in third-party databases, the study required us to estimate the option smiles. We have used a robust and sophisticated methodology to estimate the smiles for each SSO-day. Our sample consists of all the SSOs that traded in India during the period 2011-2019. We realize that all SSOs traded on a stock are not equally liquid and therefore used time-matched prices obtained from the high-frequency trades dataset covering all trades in the SSO/SSF markets.

Our key results are as follows. We find that accounting-based risk measures and firm characteristics related to asset pricing risk factors have high incremental explanatory power in explaining the ATM volatility after controlling for historical volatility. This is consistent with a rational market that impounds a wide variety of fundamental information into a forward-looking estimate of risk in the center of the distribution. By contrast, the left and right tail risks defy the explanation in terms of purely fundamentals-based risk measures. Instead, we find that stocks with characteristics that have been associated in the prior literature with lottery stocks have a pronounced right tail risk (probability of a price boom).

Similarly, stocks with characteristics that have been associated in the prior literature with bubble stocks have a pronounced left tail risk (probability of a price crash). Thus we observe a dichotomy between how the market prices risk in the center of the distribution and how it prices skewness. This dichotomy may appear a little less surprising when we recall the well-known finding in the index options markets that the skewness of the risk-neutral distribution is largely a matter of risk premia rather than risk since the physical distribution exhibits negligible skewness (Boyer and Vorkink, 2014). Our results are consistent with option writers charging significant risk premia when selling call options on lottery stocks and when selling put options on bubble stocks.

Our study contributes to the two strands of literature. First, our finding that the lottery or bubble-like characteristics of stocks explains the skewness of the IV curve is a non-trivial contribution to the larger body of literature (Mayhew, 1995; Pena et al., 1999; Bates, 2000) that tries to explain the existence of "smile-like" or "smirk-like" implied volatility curve. Second, a growing body of literature tries to analyze the influence of gambling like preference over the options pricing. Our results add to the finding of Byun and Kim (2016), and Blau et al. (2016) by providing evidence that gambling-like preference not only affects volume, put-call parity, and pricing of at-the-money option but also influences the pricing of OTM options. To the best of our knowledge, this is perhaps the first study that uses such a wide array of variables to explain the market-implied risks of the stock, and more so in the context of an emerging market. Our study adds to the relatively scant literature on the derivatives market in India and other emerging markets (Webb, 2016). Our findings have implications for developed markets and will be useful for options traders, investors, especially skewness-seeking investors, foreign institutional investors, including hedge funds specializing in option writing.

The remaining part of our paper is structured as follows. Section 2 provides a comprehensive summary of relevant literature along with the research questions. In section 3, we discuss the unique features of the Indian derivatives market, details of the variables used in the study and the index construction. Section 4 provides the descriptive statistics of the sample and the research methodology. In Section 5 and Section 6, we discuss our main results and those from the battery of robustness tests, respectively. We conclude in section 7.

2. Literature Review and Research Questions

The existence of volatility "smile" has been a widely researched topic in the field of derivatives (see Mayhew (1995) for a comprehensive review of literature on volatility smile). As per the widely used Black-Scholes (hereafter, B-S) model (Black and Scholes, 1973), the IV of an option should be independent of its strike price and time to maturity. However, various studies (for example, Rubinstein, 1985) have provided empirical evidence of "smile-like" or "smirk-like" relation between IV and strike price of both index options and SSOs for a given maturity. Past studies have focused on modelling the non-flat IV surface using various deterministic (for instance, Dupire et al., 1994) and stochastic models (for instance, Heston, 1993; Bates, 1996).

Bakshi et al. (1997) provides a comparison of the empirical performance of these models. However, the empirical evidence suggests that these models do not fit the observed volatility data well (Das and Sundaram, 1999) making it difficult to use them for pricing and hedging exotic options. Motivated by the shortcomings of the above models, another set of studies attempted to explain the IV smile using other arguments like leverage effects (Black, 1976b), information aggregation model (Grossman, 1987), risk aversion at the time of market distress (Franke et al., 1998), high buying pressure on put option for hedging purpose (Bollen and Whaley, 2004; Chan et al., 2004), transaction costs (Pena et al., 1999), and investor sentiment for index options (Han, 2008). However, none of these models completely explain the IV smile, in the sense that none can theoretically explain the empirical characteristics of IV-Smile.

A separate set of models, known as compound option (CO) models (Toft and Prucyk, 1997; Geske et al., 2016), derive the options formula considering the SSOs as a compound option i.e., options written on a firm's stock, which in itself is an option on the firm's asset. This allows the models to consider leverage effects in the option pricing. Since options price is considered a function of the stock price, the stock return variance is not assumed to be constant as in the B-S model. Taking cue from these models, past studies have attempt to explain the shape of the SSO's IV curve using various proxies of firm risks. Previous studies have considered a wide range of variables that include accounting variables that describe firm fundamentals, spot-market based proxies for past risks (Duan and Wei, 2009; Bakshi et al., 2003; Dennis and Mayhew, 2002), and derivatives-market based variables such as net buying pressure on the contracts (Dennis and Mayhew, 2002; Bollen and Whaley, 2004; Garleanu et al., 2008) in their endeavour to explain the cross-sectional variation in the shape of the IV curve. Some of the variables included are historical variance, systematic risk, CAPM Beta, ratio of open interests in put and call contracts, five asset-pricing factors (Fama and French, 1993; Carhart, 1997; Fama and French, 2015), and other firm fundamentals variables such as interest coverage ratio, liquidity, profitability (Chen et al., 2020).

There are theoretical models (for example, Ziegler, 2012; Buraschi and Jiltsov, 2006) that predict the influence of heterogeneity in opinion amongst investors over option prices, skewness of implied volatility smile and risk-neutral density. Past empirical studies have tested the prediction of these models using US options market data. Using CBOE data, studies have found that stocks having greater differences in belief have negatively skewed implied volatility curve (Friesen et al., 2012; Feng et al., 2018) . Past studies have used *share turnover* (Diether et al., 2002), *open interest* in the futures contract (Bessembinder et al., 1996) and *Analysts Dispersion* (Feng et al., 2018) as a proxy of heterogeneity in belief amongst market participants or information asymmetry around a firm.

There is a vast literature that analyzes the pricing implication of preference for lottery-like assets in the stock market (for example, Bali et al., 2011; Boyer et al., 2010; Conrad et al., 2013). Concerning preference for lottery-like SSOs, Byun and Kim (2016); Blau et al. (2016) found that the lottery-like feature is an important determinant of option prices and causes cross-sectional variation in option returns. They found evidence of overvaluation of call options and high trading volume in call options written on stocks with lottery-like characteristics. Boyer and Vorkink (2014) constructed an ex-ante skewness measure of options return, rather than stock returns, and provided evidence of a negative relationship between ex-ante skewness measure (which is a proxy of lottery-like characteristics) and option returns. The preference for lottery stocks from skewness-seeking investors impacts their pricing significantly and away from the B-S prices. While past studies have found that the ATM options for lottery-like stocks are overpriced (Byun and Kim, 2016), these papers fell short of explaining the impact of such demand-supply imbalance on the pricing of OTM options (tail risks), a dominant security for skewness-seeking investors (Filippou et al., 2018).

Another circumstance that can cause decoupling of the option prices from the underlying stock prices is during the phase of the asset price bubble. Such phases are known to persist over a long horizon because positive feedback trading (De Long et al., 1990), short-selling restrictions and costs (Lamont and Thaler, 2003; Ofek and Richardson, 2003), and synchronization risks (Abreu and Brunnermeier, 2002) can deter the informed traders from entering the market. Recent literature has also documented the growing role of retail ("Robinhood") investors in fuelling the trade volumes of equity and options contracts (Jones et al., 2021; van der Beck and Jaunin, 2021). While past studies have also found that the ATM options were efficiently priced even for internet stocks during the dot-com bubble (Battalio and Schultz, 2006), a stock bubble could lead to an increase in the demand for *lower price* OTM put options as a case of insurance-buying. This will cause an increase in the option prices at the left tail. This is more so in the Indian context, where the liquid SSF market easily facilitates leverage short-selling. We conjecture that such an increase in the demand for the OTM put options will increase the left-skew measure of the IV smile more than other parts of the smile.

Similar to emerging market peers, the option pricing literature on the Indian market is very scant (Webb, 2016). Past studies have focused on proving the existence of IV smile in the Nifty-50 index option and empirically checking the efficiency of Black-Scholes model (for instance, Jain et al., 2019; Shaikh and Padhi, 2014). Jain et al. (2019) also examine the micro-efficiency of the Indian options market, but they didn't attempt to explain the factors affecting the shape of the IV curve of SSOs. Our study aims to fill this gap in the existing literature.

Motivated by the literature, we have grouped an array of variables and created four equally-weighted indices and used them to explain the cross-sectional variation in the IV curve of SSOs. The four indices represent fundamental risks, heterogeneity in opinion, and bubbleness and lotteryness of the underlying stocks. Our use of equally-weighted indices mitigates the concern of data mining. The four indices are expected to have varying impacts on the different risk parameters – ATM, Left skew and Right skew. We expect the fundamental risk measure to influence the ATM-IV and the heterogeneity in the opinion index to positively influence the entire smile (ATM-IV than the two tails risks). This can also be related to past evidence of an increase in all the risk measures during extreme macro-level uncertainties like the pandemic (Agarwalla et al., 2021) and micro-level uncertainties like earnings announcements (Dubinsky et al., 2019). We expect the bubble index to positively influence the left-skew measure because of increased demand for OTM put options. Similarly, we expect the lotteryness feature to positively influence the right-skew measure (because of increased demand for OTM call options).

3. Indian Derivatives Market, Data, and Variables Construction

3.1. Indian SSO market

Indian derivatives market ranks amongst the largest options market and is unique in various ways (FIA, 2019). First, in India, all the derivatives products (SSO, SSF, index options, and index futures) and the underlying stocks trade on the same stock exchange, which is different than the US market where spots and derivatives contracts trade at different exchanges. Second, the existence of a liquid SSF market alongside the SSO market at the same stock exchange leads to faster information absorption, thereby improving the pricing efficiency of all the markets. Third, Indian regulation allows about 200 large companies whose stocks are liquid and followed by many analysts to have SSO or SSF. Sometimes large IPOs have SSO/SSF from the beginning. Past research has found huge expiry day effects (Agarwalla and Pandey, 2013; Vipul, 2005), and market manipulations in the (mostly illiquid) SSO/SSF market (Jain et al., 2019).

The SSOs data used in the study is obtained from the National Stock Exchange (NSE), which ranks as one of the largest options trading exchange in the world by volume (FIA, 2019) and accounts for almost 99.9% of the total derivatives contracts traded in India. Our study considers high-frequency trading data of all the near-month SSO contracts traded on the NSE between January 01, 2011 and Dec 31, 2019 (the latest available data at the commencement of the study). NSE shifted from American options to European options in January 2011, post which the SSO market noticed an upward shift in the trade volume (Jain et al., 2019). We, therefore, chose January 2011 as the starting period.

3.2. Fitting Volatility Smile

A robust and sophisticated methodology based on time-matched high-frequency data from both SSO and SSF markets was used to estimate the IV smiles for each SSO-day during our sample period. We broadly followed Jain et al. (2019) for cleaning our dataset and parameterizing the shape of the IV-curve. We applied a battery of filters for estimating the option IVs and then while fitting the smile. We first applied a liquidity filter wherein all SSO contracts that traded for less than (any) five minutes during the day were excluded. We then removed all the option contracts whose prices lied outside the Black models' arbitrage bound. To address the problems arising due to asynchronicity and stale prices, we matched the SSO price with the single stock futures (SSF) prices at the same time (accurate to the minute) and used the SSO price of the last matched SSO contract of the day. We used Black (1976a) formula for estimating IV to avoid the calculation of dividend yield and the cost of carry, instead of the Black and Sholes model (Black and Scholes, 1973; Jain et al., 2019). Following Jain et al. (2019), we fitted volatility smiles for all SSO-days using the IV of the last matched trade of all SSO contracts of a particular underlying that traded during the day provided they satisfy our filter conditions. The risk-free rate required for the IV estimation is computed based on the implied yield of the 91-day T-Bill rate provided by the

Reserve Bank of India.

While fitting IV smile for each SSO-day, we removed SSO-days with less than seven different SSO contracts during a given day and did not have at least one put and one call SSO contract. The IV smile is estimated for every SSO-day pair using a quadratic function, $IV = a\Delta^2 + b\Delta + c$ where Δ is delta of the options contract, following Malz (1997). Using quadratic fitting allows us to control for the two well-known stylized facts about financial data where it diverse from log-linearity assumption –existence of fat tail and skewness (Mixon, 2009). The quadratic smile fits the data well (Jain et al., 2019) and is a parsimonious way of fitting the curve with higher accuracy and low chance of over-fitting (Malz, 1997).

For fitting the smile, the Δ of the put options are converted into Δ of call options so that the Δ ranges between [0,1]. To ensure that IV always lies above the X-axis between $[0, 1]$, a, b and c are estimated with the restriction that a and c are non-negative, and c is shifted by $b^2/4a$ when $-b/2a \in [0, 1]$ else c is shifted by $min(0, a + b)$. Deviating from Jain et al. (2019), we aim to minimize the weighted-mean square errors, using the log of traded value as weights, between the option's market price and the estimated prices recursively. Also, the optimization is done by Nelder-Mead (Nelder and Mead, 1965) simplex optimization function. We removed under-fitted IV curves, where the weighted-mean square errors (WMSE) were greater than twice the median of the WMSE distribution, from our analysis. This has resulted in a sample of 163,589 SSO-day observations with 251 unique firms.

The fitted IV-smile is then used to calculate the following three parameters that characterizes the curve: (a) Level (ATM-IV): IV of options delta equal to 0.5 ($IV_{\Delta=0.5}$), (b) Left-Skew:

computed as the ratio of the IV of options delta equal to 0.75 and 0.50 ($IV_{\Delta=0.75}/IV_{\Delta=0.5}$), (c) Right-Skew: computed as the ratio of the IV of options delta equal to 0.25 and 0.5 ($IV_{\Delta=0.25}/IV_{\Delta=0.5}$). Our parametrization of the IV-curve characteristics is a simple reparameterization of the standard ATM, Risk-Reversal (RR), Butterfly (BF) specification of IV smile that is also used in many studies (see, for example Jain et al., 2019).

To examine the determinants of cross-sectional variation in the IV smile (IV-Smile) of SSOs, we choose to analyze at a monthly frequency as most of our independent variables are at monthly frequency. For this, we averaged the three estimated parameters for each SSO-day over the first 20 calendar days of every month. We removed the observation during the last ten days of a month to avoid the issues around the expiry days like high volume, high volatility and possible manipulations (Agarwalla and Pandey, 2013). Our IV-curve characteristics data set consists of 108 trading months and 11,955 SSO-month observations.

3.3. Index Construction and Control Variables

As mentioned earlier, we create four equally-weighted indices as our explanatory variables. The details of the variables used for constructing the indices and the control variables are given in Table 1 and 2. Table 2 provides the information about the estimation methodology and frequency along with the data source. All the variables are normalized to mean zero and standard deviation one before using them for index computation. All the variables used for the construction of indices and other control variables are taken either from CMIE $Prowess_{dx}$ or from the I/B/E/S database (see Table 2 for individual variables source.).

Our first index —*Fundamental Risk Index* (FRI), is constructed using variables that proxy for the fundamental risks of the underlying stocks. The index value was computed using seven variables, viz. *Leverage*, *Cash flow Volatility*, *Interest Coverage Ratio* (inverted), *Investment* (inverted), *Profitability* (inverted), *Size* (inverted), and *Momentum* (inverted) of the underlying firms. The variables considered are either proxy of firms' credit/default risk or explain the cross-section variation in the stock return. *Investment*, *Size*, *Momentum*, and *Profitability* are four well known factors in the empirical asset pricing literature (Fama and French, 2015; Jegadeesh and Titman, 2001). Variables like *Leverage*, *interest coverage ratio*, and *cash flow volatility* are known proxies of credit/default risk of firms (Altman, 1968; Bai and Wu, 2016). These variables are known to influence the shape of the IV-curve (see, for example Geske et al., 2016; Chen et al., 2020). Thus, for riskier firms the value of FRI would be high.

Our second index —*Heterogeneity in opinion Index* (HOI), is constructed using three variables that are known to capture the divergence in the views of the market players (Diether et al., 2002; Kim et al., 2019; Bessembinder et al., 1996). The variables considered are *Share Turnover*, *Open Interest Future*, and *Analyst Dispersion*. Previous papers in the options pricing literature have documented the influence of heterogeneity in opinion about the future return expectation from the underlying over the shape of the IV-cure (Feng et al., 2018; Friesen et al., 2012). Thus, for firms having higher heterogeneity in the opinion amongst market participants, the value of HOI would be high.

The third index —*Bubble Index* (BI), is constructed using three variables that are used to characterize stocks as bubble stocks in the literature Dass et al. (2008)). Specifically, we use $Price-to-Earnings\ Ratio$, $Price-to-Sales\ Ratio$, and $Put-to-Call\ Ratio$ to construct BI. An inflated stock price (bubble stock) would lead to higher value of BI.

Our fourth and the last index —*Lottery Index* (*LI*), is constructed using five variables used in the literature to identify lottery stocks. For example, Kumar (2009) noted that lottery-like stocks are characterized by high non-systematic variance, idiosyncratic skewness, illiquidity, book-to-market ratio, and low price. Specifically, we use $Non-Systematic\ Variance$, $Idiosyncratic\ Skewness$, *Price* (inverted), *Illiquidity*, and $Book-to-market\ Ratio$. Our selection of the variables for the *LI* is guided by literature.

Two other variables —$Systematic\ Variance$ and $Systematic\ Skewness$, are included in our regression models as control variables. Since previous papers have demonstrated the influence of these variables over IV-curve (for example, Duan and Wei, 2009; Byun and Kim, 2016), we have included them as controls.

All the variables used to construct the indices, and the control variables are winsorized at the 5 and 95 percentile. As is commonly done, the variables constructed from the firm's accounting data are used with the lag of three months to account for the gap with which they become known to the market. Our final data set consists of 11,211 observations from 108 trading months.

[*Insert Table 1 and 2 here*]

4. Descriptive Statistics and Research Method

4.1. Descriptive Statistics

Table 3 provides the overall descriptive statistics of our sample. The ATM ranges from 20.8% to 78.1% with a mean of 39%. The mean values of the left-skew (Right-Skew) variables are 1.041 (1.039) and ranges from 0.951 (0.959) to 1.21 (1.127). Table 4 reports correlation between variables. The low linear correlation between our IV curve parameters (ATM-IV, Left-Skew, & Right-Skew) indicates that they capture different dimension of risk.

As a first cut analysis, we examine the univariate relationship of our independent variables (that include four indices formed) with IV-curve characteristics (ATM-IV, Left-Skew, and Right-Skew). We first formed five sub-samples, separately based on the quintiles of each IV-curve characteristic, and then analyze the trend in the mean value of the independent variables for each sub-sample. Additionally, we also verify if the difference in the mean value for the first and the fifth quintile of each parameter is significantly different than zero. Table 5 reports the result of this analysis. Our univariate analysis shows that almost all the variables show a statistically significant trend across quintiles formed based on IV-curve characteristics and are potentially useful variables to add to the regression. The majority of the variables show a monotonic behaviour as we move from the first to the fifth quintile, suggesting that a linear regression model is appropriate.

[*Insert Table 3 to 5 about here*]

4.2. Cross-sectional Regression Models

In this subsection, we discuss the empirical approach that we employ to analyze the determinants of IV-curve characteristics. Specifically, we seek to explain the cross-sectional variation of three smile characteristics (ATM-IV, Left-Skew, and Right-Skew) using four different sets of variables. As discussed earlier, we constructed four equally-weighted indices, namely FRI, HOI, BI, and LI, one for each set of variables (for the details about the variable construction methodology, see Table 1).

For every dependent variable, we estimate three regression specifications. Equation 1 to 3 show the regression models. In model 1 we only consider the systematic part of the second and third moment of physical return distribution of the underlying. Model 2 includes FRI, and HOI on top of the variables considered in the model 1. Model 3 include BI, and LI along with variables considered in previous models.

$$IV_{it} = \alpha_t + \beta_{0,t}SV_{i,t-1} + \beta_{1,t}SSKEW_{i,t-1} + \epsilon_{it} \tag{1}$$

$$IV_{it} = \alpha_t + \beta_{0,t}SV_{i,t-1} + \beta_{1,t}SSKEW_{i,t-1} + \beta_{2,t}FRI_{i,t-1} + \beta_{3,t}HOI_{i,t-1} + \epsilon_{it} \tag{2}$$

$$IV_{it} = \alpha_t + \beta_{0,t}SV_{i,t-1} + \beta_{1,t}SSKEW_{i,t-1} + \beta_{2,t}FRI_{i,t-1} + \beta_{3,t}HOI_{i,t-1} + \beta_{4,t}BI_{i,t-1} + \beta_{5,t}LI_{i,t-1} + \epsilon_{it} \tag{3}$$

where, IV_{it} = ATM-IV, Left-Skew, or Right-Skew of the IV Smile of stock i in period t, $SV_{i,t-1}$ = Systematic Variance, $SSKEW_{i,t-1}$ = Systematic Skewness, $FRI_{i,t-1}$ = Fundamental Risk Index, $HOI_{i,t-1}$ = Heterogeneity in Opinion Index, $BI_{i,t-1}$ = Bubble Index, and $LI_{i,t-1}$ = Lottery Index. We have standardized all our independent variables to mean zero and standard deviation one before using them in our cross-sectional regression models. Thus regression coefficients in our result tables show both statistical and economic significance.

Following Bakshi et al. (2003) and Duan and Wei (2009), while doing our analysis we used Fama and MacBeth (1973) type two-pass regression. In the first step, we estimated separate cross-sectional regressions for each IV-curve parameter and regression model pair on monthly frequency (i.e., 108 such regressions for each IV-curve characteristics and regression model pair). This provides us a time series of coefficients for each independent variable for every regression specification separately. In the second step, these coefficients are then averaged, and the corresponding t-statistics are calculated using the Newey-West standard errors with twelve lags. Each column in the Table 6, shows the mean and Newey-West standard errors (with twelve lags) of the time series of coefficients estimated in the first step.

5. Results and Discussion

Table 6 shows the results from our regression models explained in equation 1-3. The cross-sectional regressions validate most of our findings of the univariate results. As all the independent variables are standardized, one can interpret the economical and statistical significance directly from the regression coefficients. We focus our discussion only on coefficients that are significant at the 1% level.

We find that increase in *systematic variance* increases ATM-IV and Left-Skew but decreases Right-Skew. Specifically, one standard deviation change in the *systematic variance* lead to a change of 3.14%, 0.0078, and -0.0076 in ATM-IV, Left-Skew, and Right-Skew, respectively. This result is along the expected lines and confirms the finding of Duan and Wei (2009) in the Indian options market. *Systematic skewness* is negatively related to ATM-IV and positively related to the Left-Skew. A one standard deviation change in *systematic skewness* leads to -1.52%, and 0.0029 change in ATM-IV and Left-Skew, respectively.

As expected FRI loads positively on ATM-IV and Left-Skew, but not on Right-Skew. The HOI loads positively on all three characteristics of the IV-curve. Our results support the finding in the literature that differences in opinion amongst investors make the IV-Smile more pronounced (Feng et al., 2018; Friesen et al., 2012).

We find that the coefficients of our *Bubble Index* is positive and significant for Left-Skew and negative and significant for both Right-Skew and ATM-IV. Specifically, one standard deviation change in the BI would lead to an increase of 0.01 in Left-Skew, and a reduction of 0.0074 in Right-Skew. This is along the expected lines, as a bubble stock would have a higher risk of a correction. Therefore, bubble stocks would have high Left-Skew and low Right-Skew. One may interpret the result as paying a premium for buying downside protection for these stocks (or insurance buying). Thus, greater hedging demand increases the premium for downside protection.

As expected, our *Lottery Index*, which measures lottery-like characteristics, loads positively on ATM-IV and Right-Skew, whereas it loads negatively in Left-Skew. Specifically, one standard deviation change in our LI results in an increase of 5.5 percentage point in ATM-IV, and 0.013 in Right-Skew, the same change result in a reduction of 0.0046 in Left-Skew. Lottery stocks are widely used for skewness-seeking investors who prefer gambling and therefore resorts to buying lower price OTM call options that are lottery-like.

Table 7 shows the statistical significance of the increase in the explanatory power for Model-3 over Model-1 and 2. For all three IV-curve characteristics, we find that Model-3, wherein we add the BI and LI, has significantly high explanatory power (Adj-R^2) as compared to Model- 1 and 2.

For ATM-IV, adding FRI and HOI improves explanatory power from 0.453 to 0.560, and adding BI and LI improves it further to 0.641. For the Left-Skew and the Right-Skew variables, the explanatory power (Adj-R^2) improves from 0.06 to 0.079 and from 0.036 to 0.096, respectively as we include the FRI and HOI indices to the base model. When one moves from model 2 to 3, by including the LI and BI, the (Adj-R^2) for the two skew variables almost doubles (0.135 for Left-Skew and 0.192 for Right-Skew). This indicates that the *Bubble Index* and *Lottery Index* add significant explanatory power.

[*Insert Table 6 and 7 about here*]

6. Robustness Tests

We conducted a battery of additional tests to examine the robustness of our results. First, instead of constructing our indices of interest (*Bubble Index* and *Lottery Index*) as an equally-weighted average of normalized variables, we constructed them by extracting the first principal component of the variables. Table 8 shows the coefficients and standard errors of equation-3 estimated using the indices created by extracting the first principal component. We find qualitatively similar results.

Second, to alleviate the concern that our results may be sensitive to the choice of method, we replicated all our results using a Fixed-effect panel data model instead of a cross-sectional model. Specifically, we used the following panel data model.

$$IV_{it} = \beta_0 SV_{i,t-1} + \beta_1 SSKEW_{i,t-1} + \beta_2 FRI_{i,t-1} + \beta_3 HOI_{i,t-1} + \beta_{4,t} BI_{i,t-1} + \beta_{5,t} LI_{i,t-1} + \eta_i + \gamma_t + \epsilon_{it} \quad (5)$$

where, IV_{it} = ATM-IV, Left Skew, or Right Skew of the IV Smile of stock i in period t, $SV_{i,t-1}$ = Systematic Variance, $SSKEW_{i,t-1}$ = Systematic Skewness, $FRI_{i,t-1}$ = Fundamental Risk Index, $HOI_{i,t-1}$ = Heterogeneity in Opinion Index, $BI_{i,t-1}$ = Bubble Index, and $LI_{i,t-1}$ = Lottery Index, η_i = Firm level fixed effect, γ_t = Month-Year fixed effect. We have standardized all our independent variables to mean zero and standard deviation one before using them in our panel data regression models.Table 9 shows the results of the full model for all the three dependent variables. Most of our results remained qualitatively similar to those from cross-section regression models.

Third, instead of doing our analysis at a monthly frequency, we ran our regressions at a weekly frequency. We selected one day of the week (in our case, Tuesday) and estimated the coefficients of Equation- 3 at a weekly frequency. Table 10 shows the coefficient and standard error. We find qualitatively similar results. We also did our panel data analysis using weekly frequency data. We also estimated the coefficients of Equation- 5 using this data. Table 11 shows the results. We find a qualitatively similar result.

[*Insert Table 8 to 11 about here*]

7. Conclusion

Theoretical considerations and prior literature indicate that stock options prices depend on (a) the risk characteristics of the historical return distribution, (b) accounting information on firm fundamentals that is predictive of future risk, (c) a variety of risk premia as well as buying and selling pressure in the spot and derivative markets. By studying the cross-sectional determinants of the option smile, we can throw light on the relative importance of these factors in pricing risk in the center and tails of the distribution.

We find that the cross-sectional variation in ATM volatility is largely explained by historical risk and predicted future risk. Accounting-based risk measures and firm characteristics related to asset pricing risk factors that predict future risk have high incremental explanatory power. This is consistent with a rational market that impounds a wide variety of fundamental information into a forward-looking estimate of risk in the center of the distribution. This result strengthens the findings of Jain et al. (2019) by providing more granular evidence of the micro-efficiency of the Indian SSO market.

By contrast, the cross-sectional variation in left and right skew defy the explanation in terms of purely fundamentals-based risk measures. Instead, the evidence suggests that these smile parameters are driven by risk premia and buying and selling pressure. This is consistent with the well-known finding in the index options markets that the skewness of the risk-neutral distribution is largely a matter of risk premia rather than risk since the physical distribution exhibits negligible skewness.

Stocks with characteristics that have been associated in the prior literature with lottery stocks have a pronounced right tail risk (probability of a price boom). Similarly, stocks with characteristics that have been associated in the prior literature with bubble stocks have a pronounced left tail risk (probability of a price crash). Our results are consistent with option writers charging significant risk premia when selling call options on lottery stocks and when selling put options on bubble stocks. These findings contribute to the larger body of literature (Mayhew, 1995; Pena et al., 1999; Bates, 2000) that tries to explain the existence of “smile-like” or “smirk-like” implied volatility curve.

The Indian stock options market has a strong retail presence. Our findings about the impact of gambling-like preferences on the pricing of OTM options assume importance in the context of the growing role of retail (“Robinhood”) investors in fuelling the trade volumes of equity and options contracts (Jones et al., 2021; van der Beck and Jaunin, 2021) in developed markets.

References

Abreu, D., Brunnermeier, M.K., 2002. Synchronization risk and delayed arbitrage. Journal of Financial Economics 66, 341–360.

Agarwalla, S.K., Pandey, A., 2013. Expiration-day effects and the impact of short trading breaks on intraday volatility: Evidence from the indian market. Journal of Futures Markets 33, 1046–1070.

Agarwalla, S.K., Varma, J.R., Virmani, V., 2021. The impact of COVID-19 on tail risk: Evidence from nifty index options. Economics Letters , 109878URL: https://www.sciencedirect.com/science/article/pii/S0165176521001555, doi:https://doi.org/10.1016/j.econlet.2021.109878.

Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of finance 23, 589–609.

Amihud, Y., Hameed, A., Kang, W., Zhang, H., 2015. The illiquidity premium: International evidence. Journal of financial economics 117, 350–368.

Bai, J., Wu, L., 2016. Anchoring credit default swap spreads to firm fundamentals. Journal of Financial and Quantitative Analysis (JFQA) 51.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. The Journal of finance 52, 2003–2049.

Bakshi, G., Kapadia, N., 2003. Delta-hedged gains and the negative market volatility risk premium. The Review of Financial Studies 16, 527–566.

Bakshi, G., Kapadia, N., Madan, D., 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. The Review of Financial Studies 16, 101–143.

Bali, T.G., Cakici, N., Whitelaw, R.F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. Journal of Financial Economics 99, 427–446.

Bates, D.S., 1996. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. The Review of Financial Studies 9, 69–107.

Bates, D.S., 2000. Post-'87 crash fears in the s&p 500 futures option market. Journal of econometrics 94, 181–238.

Battalio, R., Schultz, P., 2006. Options and the bubble. The Journal of Finance 61, 2071–2102.

van der Beck, P., Jaunin, C., 2021. The equity market implications of the retail investment boom. Available at SSRN 3776421 .

Bessembinder, H., Chan, K., Seguin, P.J., 1996. An empirical examination of information, differences of opinion, and trading activity. Journal of Financial Economics 40, 105–134.

Black, F., 1976a. The pricing of commodity contracts. Journal of financial economics 3, 167–179.

Black, F., 1976b. Studies of stock market volatility changes. 1976 Proceedings of the American Statistical Association Bisiness and Economic Statistics Section .

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of political economy 81, 637–654.

Blau, B.M., Bowles, T.B., Whitby, R.J., 2016. Gambling preferences, options markets, and volatility. Journal of Financial and Quantitative Analysis , 515–540.

Bollen, N.P., Whaley, R.E., 2004. Does net buying pressure affect the shape of implied volatility functions? The Journal of Finance 59, 711–753.

Boyer, B., Mitton, T., Vorkink, K., 2010. Expected idiosyncratic skewness. The Review of Financial Studies 23, 169–202.

Boyer, B.H., Vorkink, K., 2014. Stock options as lotteries. The Journal of Finance 69, 1485–1527.

Buraschi, A., Jiltsov, A., 2006. Model uncertainty and option markets with heterogeneous beliefs. The Journal of Finance 61, 2841–2897.

Byun, S.J., Kim, D.H., 2016. Gambling preference and individual equity option returns. Journal of Financial Economics 122, 155–174.

Carhart, M.M., 1997. On persistence in mutual fund performance. The Journal of finance 52, 57–82.

Chan, K.C., Cheng, L.T., Lung, P.P., 2004. Net buying pressure, volatility smile, and abnormal profit of hang seng index options. Journal of Futures Markets: Futures, Options, and Other Derivative Products 24, 1165–1194.

Chen, D., Guo, B., Zhou, G., 2020. Firm fundamentals and the cross section of implied volatility shapes.

Available at SSRN .

Conrad, J., Dittmar, R.F., Ghysels, E., 2013. Ex ante skewness and expected stock returns. The Journal of Finance 68, 85–124.

Das, S.R., Sundaram, R.K., 1999. Of smiles and smirks: A term structure perspective. Journal of financial and quantitative analysis , 211–239.

Dass, N., Massa, M., Patgiri, R., 2008. Mutual funds and bubbles: The surprising role of contractual incentives. The Review of Financial Studies 21, 51–99.

De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Positive feedback investment strategies and destabilizing rational speculation. the Journal of Finance 45, 379–395.

Dennis, P., Mayhew, S., 2002. Risk-neutral skewness: Evidence from stock options. Journal of Financial and Quantitative Analysis 37, 471–493.

Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. The Journal of Finance 57, 2113–2141.

Duan, J.C., Wei, J., 2009. Systematic risk and the price structure of individual equity options. The Review of Financial studies 22, 1981–2006.

Dubinsky, A., Johannes, M., Kaeck, A., Seeger, N.J., 2019. Option pricing of earnings announcement risks. The Review of Financial Studies 32, 646–687.

Dupire, B., et al., 1994. Pricing with a smile. Risk 7, 18–20.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of financial economics 116, 1–22.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of political economy 81, 607–636.

Feng, S., Pu, X., Zhang, Y., 2018. An empirical examination of the relation between the option-implied volatility smile and heterogeneous beliefs. The Journal of Derivatives 25, 36–47.

FIA, 2019. *Top 40 Derivatives Exchanges Worldwide*. https://www.fia.org/media/2407 [Accessed: April

27, 2021].

Filippou, I., Garcia-Ares, P.A., Zapatero, F., 2018. Demand for lotteries: The choice between stocks and options. Available at SSRN 3016462 .

Franke, G., Stapleton, R.C., Subrahmanyam, M.G., 1998. Who buys and who sells options: The role of options in an economy with background risk. journal of economic theory 82, 89–109.

Friesen, G.C., Zhang, Y., Zorn, T.S., 2012. Heterogeneous beliefs and risk-neutral skewness. Journal of Financial and Quantitative Analysis , 851–872.

Garleanu, N., Pedersen, L.H., Poteshman, A.M., 2008. Demand-based option pricing. The Review of Financial Studies 22, 4259–4299.

Geske, R., Subrahmanyam, A., Zhou, Y., 2016. Capital structure effects on the prices of equity call options. Journal of Financial Economics 121, 231–253.

Grossman, S.J., 1987. An analysis of the implications for stock and futures price volatility of program trading and dynamic hedging strategies. Technical Report. National Bureau of Economic Research.

Han, B., 2008. Investor sentiment and option prices. The Review of Financial Studies 21, 387–414.

Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. The Journal of finance 55, 1263–1295.

Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. The review of financial studies 6, 327–343.

Jain, S., Varma, J.R., Agarwalla, S.K., 2019. Indian equity options: Smile, risk premiums, and efficiency. Journal of Futures Markets 39, 150–163.

Jegadeesh, N., Titman, S., 2001. Profitability of momentum strategies: An evaluation of alternative explanations. The Journal of finance 56, 699–720.

Jones, C.M., Reed, A.V., Waller, W., 2021. When brokerages restrict retail investors, does the game stop? Does the Game Stop .

Kim, J.B., Lu, L.Y., Yu, Y., 2019. Analyst coverage and expected crash risk: evidence from exogenous changes in analyst coverage. The Accounting Review 94, 345–364.

Kim, J.B., Zhang, L., 2014. Financial reporting opacity and expected crash risk: Evidence from implied volatility smirks. Contemporary Accounting Research 31, 851–875.

Kumar, A., 2009. Who gambles in the stock market? The Journal of Finance 64, 1889–1933.

Lamont, O.A., Thaler, R.H., 2003. Can the market add and subtract? mispricing in tech stock carve-outs. Journal of Political Economy 111, 227–268.

Malz, A.M., 1997. Estimating the probability distribution of the future exchange rate from option prices. Journal of Derivatives 5, 18–36.

Mayhew, S., 1995. Implied volatility. Financial Analysts Journal 51, 8–20.

Mixon, S., 2009. Option markets and implied volatility: Past versus present. Journal of Financial Economics 94, 171–191.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. The computer journal 7, 308–313.

Ofek, E., Richardson, M., 2003. Dotcom mania: The rise and fall of internet stock prices. The Journal of Finance 58, 1113–1137.

Pena, I., Rubio, G., Serna, G., 1999. Why do we smile? on the determinants of the implied volatility function. Journal of Banking & Finance 23, 1151–1179.

Rubinstein, M., 1985. Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active cboe option classes from august 23, 1976 through august 31, 1978. The Journal of Finance 40, 455–480.

Shaikh, I., Padhi, P., 2014. Stylized patterns of implied volatility in india: a case study of nse nifty options. Journal of Indian Business Research .

Toft, K.B., Prucyk, B., 1997. Options on leveraged equity: Theory and empirical tests. The Journal of Finance 52, 1151–1180.

Vipul, V., 2005. Futures and options expiration-day effects: the indian evidence. Journal of Futures Markets 25, 1045.

Webb, R.I., 2016. Recent advances in the literature: Asia–pacific derivative markets. Asia-Pacific Journal of Financial Studies 45, 34–47.

Ziegler, A.C., 2012. Incomplete information and heterogeneous beliefs in continuous-time finance. Springer Science & Business Media.

Table 1: **Index Construction Details**

Index Name	Variables considered	Details
Fundamental Risk Index (FRI)	1) Leverage 2) Size (Inverted) 3) Cash flow Volatility 4) Investment (Inverted) 5) Profitability (Inverted) 6) Interest Coverage Ratio (Inverted) 7) Momentum (Inverted)	While constructing our Fundamental Risk Index, as a first step we inverted all the variables that are inversely related to firms risk. Therefore, variables like size, interest coverage ratio, profitability, momentum [1], and investment are inverted. In second step, we normalized all these variable to mean zero and standard deviation one, then took average.
Heterogeneity in opinion Index (HOI)	1) Open Interest Future 2) Share Turnover 3) Analyst Dispersion	Stocks having higher heterogeneity of belief among market participants are known to have high stock turnover (Kim et al., 2019), high analyst dispersion (Diether et al., 2002), and open interest in the futures market (Bessembinder et al., 1996). We these three variables to construct our heterogeneity in opinion index. We first normalized these variables to mean zero and standard deviation one, and then took average.
Bubble Index (BI)	1) Price-to-Earnings Ratio 2) Price-to-Sales Ratio 3) Put-to-Call Ratio	Bubble stocks are characterized by high price-to-earnings ratio, and high price-to-sales ratio (Dass et al., 2008). We first, normalized all our variables to mean zero and standard deviation one, and then took average. Thus, our Bubble index is positively related to the Bubble like behavior of stocks.
Lottery Index (LI)	1) Non-Systematic Variance 2) Idiosyncratic Skewness 3) Price (Inverted) 4) Illiquidity 5) Book-to-market Ratio	Lottery stocks are characterised by high non-systematic variance, high idiosyncratic skewness, low price, high illiquidity, and high book-to-market ratio (Kumar, 2009). We first invert variables like price that are inversely related to lottery like behavior of stocks. Then, we normalize all our variables to mean zero and standard deviation one. Finally, we take average of all our variables. Thus, our Lottery index is positively related to the lottery like behavior of stocks.

[1] Bai and Wu (2016) found that the momentum of a stock is inversely related to the credit default swap spread.

Table 2: **Variable Construction Details**

Variable Name	Variable Definition	Frequency	Source	Details
Historical Moments				
Historical Volatility (HV)	$\mathrm{HV}_t^h = \sqrt{(\frac{252}{\tau}\sum_{i=t-\tau}^{t}(r_{i-1,i} - \bar{R})^2)}$	Daily	CMIE	This variable is calculated using Bakshi and Kapadia . (2003). The rolling period τ is taken as 30 days, $r_{i-1,i}$ is the daily return and $\bar{R}$ is the average daily return.
Systematic Variance(SV)	$SV = R^2_{CAPM} \times \mathrm{HV}^2$	Monthly	CMIE	R^2_{CAPM} denotes R^2 of CAPM model. For every stock we ran $$r_{jt} = \alpha_j + \beta_j r_{mt} + \epsilon_{jt} \quad (6)$$ where r_{jt} and r_{mt} are returns for the stock j and the market at time t. We ran 60 months rolling window regression for each stock, therefore, R^2_{CAPM} is available at monthly frequency. To calculate SV, we multiplied HV^2 available at daily frequency with R^2_{CAPM} of that month to get SV at daily frequency which is then averaged for every month to convert it into monthly frequency.
Systematic Skewness ($SSKEW$)	$SSKEW$ of stock i in financial year t is defined as γ_i, estimated by equation-7	Yearly	CMIE	Following Harvey and Siddique (2000), we ran $$R_{i,d} = \alpha_i + \beta_i R_{m,d} + \gamma_i R^2_{m,d} + \epsilon_{i,d} \quad (7)$$ where $R_{i,d}$ is return of stock i on day d, $R_{m,d}$ is the market return on day d, and $\epsilon_{i,d}$ is the residual of the regression model (idiosyncratic return of day d). We estimated this regression for every stock using one financial years data at a time.
Fundamental Risk Index				
Leverage (LV)	$\mathrm{LV} = \frac{Total\ External\ Liability}{Share\ Price \times Number\ of\ Share\ Outstanding}$	Monthly	CMIE	
Size	$\mathrm{Size} = Log(\frac{Share\ Price \times Number\ of\ Share\ Outstanding}{1000})$	Monthly	CMIE	We converted this variable into monthly frequency by taking average of daily market capitalization of every stock.

Table 2: Variable Construction Details

Variable Name	Variable Definition	Frequency	Source	Details
Cash flow Volatility (CFV)		Yearly	CMIE	Following Kim et al. (2019), we estimate cash flow volatility of a particular year as standard deviation of net cash flow from operation scaled by lagged total asset over preceding 5 years.
Investment (INV)	INV= $\frac{Retained\ Earning}{Total\ Asset}$	Yearly	CMIE	
Profitability (PF)	PF = $\frac{Gross\ Income}{Total\ Asset}$	Yearly	CMIE	
Interest Coverage Ratio (ICR)	ICR = $\frac{Gross\ Income}{Interest\ and\ Related\ Expense}$	Yearly	CMIE	
Momentum (MOM)	$MOM_{it} = \frac{P_{i,t-1} - P_{i,t-12}}{P_{i,t-12}}$	Monthly	CMIE	For a particular stock momentum for every month is it's return over the past 11 months.
Heterogeneity in opinion Index				
Open Interest Future (OIF)	OIF = log(monthly average of Open Interest)	Monthly	NSE Bhav Files	
Share Turnover (ST)	ST = $\frac{Daily\ average\ trading\ volume\ in\ a\ month}{Average\ share\ outstanding\ in\ a\ month}$	Monthly	CMIE	
Analyst Dispersion (AD)	$AD = \frac{Standard\ Deviation\ of\ the\ analysts'\ EPS\ forecasts\ of\ a\ firm}{Earning\ Per\ share}$	Monthly	IBES	Std deviation of the analysts' EPS forecasts available at monthly frequency is divided by quarterly EPS of the firm to calculate this variable.
Bubble Index				
Price-to-Earnings Ratio ($PbyE$)	$PbyE = 1/N \sum_{i=1}^{N} \frac{Price\ of\ Stock}{Earnings\ Per\ share}$	Monthly	CMIE	Here N is the number of trading days in a month.
Price-to-Sales Ratio ($PbyS$)	$PbyS = 1/N \sum_{i=1}^{N} \frac{Price\ of\ Stock}{Total\ Sales}$	Monthly	CMIE	Here N is the number of trading days in a month.

Table 2: Variable Construction Details

Variable Name	Variable Definition	Frequency	Source	Details
Put-to-Call Ratio (PbyC)	$PbyC = \frac{\sum_{p=1}^{p=n} OI\ Put\ options_{p,i}}{\sum_{c=1}^{c=n} OI\ Call\ options_{c,i}}$	Monthly	NSE Bhav Files	For each SSO-day pair, we first estimate Put-to-call ratio on daily frequency by dividing the sum of open interest of all Put option contracts by the sum of open interest of all Call options contract. To convert it into monthly frequency we take average of the daily $PbyC$ ratio for each firm-month pair.
Lottery Index				
Non-Systematic Variance (NSV)	$NSV = (1 - R^2_{CAPM}) \times \mathrm{HV}^2$	Monthly	CMIE	R^2_{CAPM} denotes R^2 of Model- 6. To calculate NSV, we multiplied HV^2 available at daily frequency with R^2_{CAPM} of that month to get NSV at daily frequency which is then averaged for every month to convert it into monthly frequency.
Idiosyncratic Skewness ($ISKEW$)	$ISKEW$ of stock i in financial year t is defined as the skewness of daily residual ($\epsilon_{i,d}$) of equation-7 in a financial year	Yearly	CMIE	
Illiquidity ($Illiq$)	$Illiq = (1/N_{it}) \sum(\lvert r_{i,d,t} \rvert / vol_{i,d,t})$	Monthly	CMIE	We follow Amihud et al. (2015) to estimate $Illiq$. Here, $\lvert r_{i,d,t} \rvert$ is absolute return of stock i on day d in period t which is taken as one month, $vol_{i,d,t}$ is the trading volume of stock i on day t, obtained by multiplying traded volume by share price, and N_{it} is number of trading days of stock i in time period t which we take as one month.
Price (SP)	Average stock price over a month	Monthly	CMIE	
Book-to-market Ratio (BbyM)	$\mathrm{BbyM} = \frac{Book\ Value\ of\ Equity}{Market\ Value\ of\ Equity}$	Monthly	CMIE	Previous quarter's Book value is divided by Market value of each day to calculate this ratio at daily frequency which is then converted into monthly frequency by taking monthly average.

All the independent variables used in the study are formed either using firm's account statement or spot market data, and are available either at monthly, or yearly frequency. To make sure all the accounting related information is known to the market we have lagged all the independent variables like ICR, INV, PF, LIQ that available at yearly frequency by three months.

Table 3: Summary statistics

The tables shows the summary statistics of all the variables used in the study. P25 and P75 denotes the 25^{th} and 75^{th} percentile of the distribution. The sample period spans from January 2011 to December 2019 and variables are winsorized at 5 and 95 percentile value. For variable definitions see Table 2.

Variables	No. of Observations	Mean	Std. Dev	Min	P25	Median	P75	Max
ATM-IV	11,955	0.391	0.113	0.208	0.308	0.374	0.451	0.781
Left-Skew	11,955	1.041	0.044	0.951	1.010	1.038	1.069	1.214
Right-Skew	11,955	1.039	0.033	0.959	1.016	1.036	1.059	1.127
Systematic Variance (SV)	11,211	0.050	0.046	0.001	0.017	0.037	0.068	0.263
Systematic Skewness	11,809	−6.234	14.259	−54.189	−13.136	−4.609	2.523	28.899
Leverage (LV)	11,282	1.303	1.591	0.031	0.169	0.625	1.748	6.705
Size	11,809	19.210	1.261	16.616	18.256	19.181	20.097	21.771
Cash flow Volatility (CFV)	8,517	0.058	0.051	0.013	0.027	0.042	0.073	0.415
Investment (INV)	11,281	0.285	0.219	0.009	0.074	0.252	0.454	0.770
Profitability (PF)	11,280	0.206	0.157	0.024	0.089	0.144	0.276	0.661
Interest Coverage Ratio (ICR)	10,972	66.176	161.921	−0.236	1.406	3.350	23.384	814.101
Momentum (MOM)	11,801	0.023	0.377	−0.782	−0.239	−0.011	0.244	1.396
Open Interest Future (OIF)	11,949	8.561	0.955	6.679	7.858	8.501	9.284	10.614
Share Turnover (ST)	11,809	0.103	0.105	0.011	0.034	0.065	0.128	0.546
Analyst Dispersion (AD)	11,042	0.289	0.509	0.006	0.050	0.105	0.253	3.588
Price-to-Earnings Ratio ($PByE$)	10,256	30.275	27.074	3.927	13.061	21.166	36.296	157.328
Price-to-Sales Ratio ($PbyS$)	9,463	0.108	0.430	0.0001	0.001	0.003	0.011	3.764
Put-to-call Ratio ($PbyC$)	11,955	0.491	0.186	0.094	0.353	0.461	0.598	1.078
Non-systematic Variance (NSV)	11,211	0.089	0.064	0.017	0.044	0.071	0.114	0.385
Idiosyncratic Skewness ($ISKEW$)	11,808	0.395	0.639	−1.409	0.054	0.365	0.707	2.924
Illiquidity ($Illiq$)	11,809	3.835e-11	3.770e-11	2.194e-12	1.216e-11	2.499e-11	5.109 e-11	2.155e-10
Price (SP)	11,809	390.379	360.583	28.557	110.878	264.931	557.400	1,450.967
Book-to-market Ratio ($BbyM$)	11,779	0.681	0.585	0.066	0.244	0.472	0.952	2.642

Table 4: Correlation Table

	ATM-IV	Left-Skew	Right-Skew
ATM-IV	1		
Right-Skew	-0.114	1	
Left-Skew	0.204	-0.43	1
Systematic Variance (SV)	0.581	-0.097	0.229
Sys Skewness ($SSKEW$)	-0.098	-0.114	0.073
Price-to-Earnings Ratio ($PbyE$)	0.016	0.02	-0.011
Price-to-Sales Ratio ($PbyS$)	-0.047	-0.065	0.072
Put-to-Call Ratio ($PbyC$)	-0.158	-0.293	0.247
Non-Sys Variance (NSV)	0.562	-0.058	0.122
Idio. SKewness ($ISKEW$)	0.008	0.048	-0.053
Illiquidity ($Illiq$)	0.405	0.1	-0.001
Book-to-market ($BbyM$)	0.42	0.054	0.11
Price	-0.45	-0.179	0.029
Momentum (MOM)	-0.161	0.055	-0.24
Profitability (PF)	-0.3	-0.032	-0.094
ICR	-0.174	0.03	-0.067
Investment (INV)	-0.366	-0.006	-0.121
Leverage (LV)	0.342	0.039	0.091
Size	-0.612	-0.055	-0.032
Analyst Dispersion (AD)	0.314	0.064	0.056
Share Turnover (ST)	0.557	-0.035	0.159
Cash flow Volatility (CFV)	0.207	-0.018	0.081
Open Interest Future (OIF)	0.188	0.206	0.036

Table 5: Quintile-wise summary statistics

The table shows the quintile wise mean value of Independent variables used in the study. In Panel-A, Panel-B, & Panel-C quintiles are formed based on ATM-IV, Left-Skew and Right-Skew, respectively. 1 denotes the lowest quintile and 5 denote the highest quintile. The last row of every panel shows t-statistics of difference of mean test between first and fifth quintile. The sample period spans from Jan 2011 to Dec 2019. All Variables for which difference between the mean of first and fifth quintile are significant at 5% is shown in bold.

Panel-A

ATM-IV	Systematic Variance	Non-Systematic Variance	Fundamental Risk Index	Heterogeneity in Opinion Index	Bubble Index	Lottery Index
1	0.023	-0.970	-0.352	-0.411	0.151	-0.540
2	0.036	-5.383	-0.173	-0.278	0.015	-0.317
3	0.048	-7.445	0.010	-0.075	-0.012	-0.025
4	0.062	-8.637	0.206	0.179	-0.058	0.255
5	0.085	-8.957	0.413	0.640	-0.125	0.654
t-stats	**-48.236**	**19.661**	**-52.611**	**-56.826**	**14.977**	**-80.457**

Panel-B

Left-Skew	Systematic Variance	Non-Systematic Variance	Fundamental Risk Index	Heterogeneity in Opinion Index	Bubble Index	Lottery Index
1	0.041	-7.290	-0.002	-0.089	-0.215	0.068
2	0.045	-6.345	-0.036	-0.066	-0.076	-0.052
3	0.051	-5.429	-0.005	-0.019	0.009	-0.048
4	0.055	-5.622	0.024	0.026	0.107	-0.055
5	0.060	-6.443	0.107	0.184	0.161	0.085
t-stats	**-13.493**	**-1.995**	**-7.123**	**-13.107**	**-20.542**	-0.838

Panel-C

Right-Skew	Systematic Variance	Non-Systematic Variance	Fundamental Risk Index	Heterogeneity in Opinion Index	Bubble Index	Lottery Index
1	0.049	-5.582	-0.054	-0.134	0.215	-0.183
2	0.049	-6.118	-0.043	-0.070	0.094	-0.123
3	0.050	-6.062	-0.009	-0.011	-0.020	-0.046
4	0.051	-6.293	0.043	0.047	-0.128	0.063
5	0.052	-7.133	0.153	0.207	-0.195	0.296
t-stats	**-2.211**	**3.641**	**-13.137**	**-16.479**	**22.112**	**-24.829**

Table 6: Cross-sectional Regression

The table shows the cross-sectional regressions results, where dependent variables are ATM-IV, Left-Skew, and Right-Skew. We run cross-sectional regression at monthly frequency, and obtained the coefficients of respective independent variables. Our first step gave us a time series of coefficients for each independent variables. These coefficients are then averaged and the corresponding t-statistics is calculated using the Newey-West standard errors with twelve lags. Each column in the table shows the mean, and Newey-West standard errors (with twelve lags) of time series of coefficients that we got after the first step. All independent variables are standardized to mean zero and standard deviation one. Thus, all regression coefficients show both statistical and economic significance. Avg. Adjusted R^2 reported in the table is the average adjusted R^2 of first pass regressions. ***, **, and * reflect significance at the 1, 5, and 10% levels, respectively.

Dependent Variable	ATM-IV			Left-Skew			Right-Skew		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Systematic Variance$_{i,t-1}$	0.077***	0.039***	0.031***	0.008***	0.007***	0.007***	0.001***	-0.006***	-0.007***
	(0.002)	(0.001)	(0.001)	(0.0004)	(0.0007)	(0.0008)	(0.0004)	(0.0005)	(0.0005)
Systematic Skewness$_{i,t-1}$	-0.029***	-0.022***	-0.015***	0.004***	0.004***	0.002***	-0.004***	-0.003***	-0.0006
	(0.0017)	(0.0014)	(0.0010)	(0.0008)	(0.0008)	(0.0009)	(0.0005)	(0.0004)	(0.0004)
Fundamental Risk Index$_{i,t-1}$		0.054***	0.021***		-0.0005	0.003**		0.008***	-0.0002
		(0.0024)	(0.0020)		(0.0014)	(0.0012)		(0.0006)	(0.0008)
Heterogeneity in Opinion Index$_{i,t-1}$		0.036***	0.024***		0.002**	0.003***		0.009***	0.007***
		(0.0016)	(0.0014)		(0.0011)	(0.0011)		(0.0004)	(0.0005)
Bubble Index$_{i,t-1}$			-0.011***			0.010***			-0.007***
			(0.0009)			(0.0009)			(0.0006)
Lottery Index$_{i,t-1}$			0.058***			-0.004**			0.013***
			(0.0026)			(0.0018)			(0.0008)
N	11211	11211	11211	11211	11211	11211	11211	11211	11211
Adj. R^2	0.453	0.560	0.641	0.060	0.079	0.135	0.036	0.096	0.192

Table 7: Explanatory Power

The table reports the average of adjusted R^2 of our first pass cross-sectional regressions. Panel- A,B & C reports the adjusted R^2 for models having ATM-IV, Left-Skew, and Right-Skew as dependent variables respectively. First column of each panel reports the adjusted R^2 value of Model - 1 & 2. Second column reports the same value for Model-3, where as third column reports the difference between adjusted R^2 of Model-3 with respect to other two models, and significance of the difference using t-test. ***, **, and * reflect significance at the 1, 5, and 10% levels, respectively.

	Panel A - ATM-IV			Panel B - Left-Skew			Panel C - Right-Skew		
Model	Adj. R^2	Adj. R^2 Model-3	Diff	Adj.R^2	Adj. R^2 Model-3	Diff	Adj. R^2	Adj. R^2 Model-3	Diff
Model-1	0.453	0.641	0.188***	0.060	0.135	0.076***	0.036	0.192	0.156***
Model-2	0.560	0.641	0.081***	0.079	0.135	0.056***	0.096	0.192	0.095***

Table 8: Cross-sectional Regression (PCA)

The table shows the cross-sectional regressions results, where dependent variables are ATM-IV, Left-Skew, and Right-Skew. *Bubble Index* and *Lottery Index* are constructed by extracting first principal component of the constituents. We run cross-sectional regression at monthly frequency, and obtained the coefficients of respective independent variables. Our first step gave us a time series of coefficients for each independent variables. These coefficients are then averaged and the corresponding t-statistics is calculated using the Newey-West standard errors with twelve lags. Each column in the table shows the mean, and Newey-West standard errors (with twelve lags) of time series of coefficients that we got after the first step. All independent variables are standardized to mean zero and standard deviation one. Thus, all regression coefficients show both statistical and economic significance. Avg. Adjusted R^2 reported in the table is the average adjusted R^2 of first pass regressions. ***, **, and * reflect significance at the 1, 5, and 10% levels, respectively.

Dependent Variable	ATM-IV	Left-Skew	Right-Skew
	(1)	(2)	(3)
Systematic Variance$_{i,t-1}$	0.035***	0.008***	-0.008***
	(0.0019)	(0.0010)	(0.0006)
Systematic Skewness$_{i,t-1}$	-0.015***	0.002***	0.0005
	(0.0010)	(0.0010)	(0.0006)
Fundamental Risk Index$_{i,t-1}$	0.046***	0.001	0.001
	(0.0029)	(0.0020)	(0.0010)
Heterogeneity in Opinion Index$_{i,t-1}$	0.013***	0.007***	0.004***
	(0.0022)	(0.0015)	(0.0007)
Bubble Index (PCA)$_{i,t-1}$	-0.007***	0.002***	-0.002***
	(0.0009)	(0.0005)	(0.0004)
Lottery Index (PCA)$_{i,t-1}$	0.021***	-0.003***	0.006***
	(0.0012)	(0.0008)	(0.0003)
N	8072	8072	8072
Adj. R^2	0.628	0.106	0.1587

Table 9: Panel Data Regression

The table shows panel data regression results of Model -5, where dependent variables are ATM-IV, Left-Skew, and Right-Skew. Robust standard errors are reported in the parenthesis. All the models have Firm and Month-Year level fixed effects. All the indices are equally weighted. ***, **, and * reflect significance at the 1, 5, and 10% levels, respectively.

Dependent variable	ATM-IV	Left-Skew	Right-Skew
	(1)	(2)	(3)
Systematic Variance$_{i,t-1}$	0.018^{***}	0.004^{***}	-0.003^{***}
	(0.001)	(0.001)	(0.0005)
Systematic Skewness$_{i,t-1}$	0.00001	-0.001	0.001^{***}
	(0.001)	(0.0004)	(0.0003)
Fundamental Risk Index$_{i,t-1}$	0.020^{***}	-0.002^{*}	0.003^{***}
	(0.002)	(0.001)	(0.001)
Heterogeneity in Opinion Index$_{i,t-1}$	0.029^{***}	-0.001	0.004^{***}
	(0.001)	(0.001)	(0.001)
Bubble Index$_{i,t-1}$	-0.010^{***}	0.013^{***}	-0.010^{***}
	(0.001)	(0.001)	(0.0005)
Lottery Index$_{i,t-1}$	0.057^{***}	-0.003^{***}	0.009^{***}
	(0.002)	(0.001)	(0.001)
Observations	11,211	11,211	11,211
Adj. R^2	0.790	0.441	0.442

Table 10: Cross Sectional Regression using Weekly Frequency

The table shows the cross-sectional regressions results, where dependent variables are ATM-IV, Left-Skew, and Right-Skew. We run cross-sectional regression at weekly (Tuesday of every week) frequency, and obtained the coefficients of respective independent variables. Our first step gave us a time series of coefficients for each independent variables. These coefficients are then averaged and the corresponding t-statistics is calculated using the Newey-West standard errors with twelve lags. Each column in the table shows the mean, and Newey-West standard errors (with twelve lags) of time series of coefficients that we got after the first step. All independent variables are standardized to mean zero and standard deviation one. Thus, all regression coefficients show both statistical and economic significance. Avg. Adjusted R^2 reported in the table is the average adjusted R^2 of first pass regressions. ***, **, and * reflect significance at the 1, 5, and 10% levels, respectively.

Dependent Variable	ATM-IV	Left-Skew	Right-Skew
	(1)	(2)	(3)
Systematic Variance$_{i,t-1}$	0.040***	0.007***	-0.006***
	(0.0020)	(0.0012)	(0.0007)
Systematic Skewness$_{i,t-1}$	-0.010***	0.000	-0.001
	(0.0013)	(0.0009)	(0.0008)
Fundamental Risk Index$_{i,t-1}$	0.017***	0.000	0.0005
	(0.0025)	(0.0014)	(0.0012)
Heterogeneity in Opinion Index$_{i,t-1}$	0.018***	0.004***	0.006***
	(0.0022)	(0.0009)	(0.0009)
Bubble Index$_{i,t-1}$	-0.009***	0.009***	-0.007***
	(0.0013)	(0.0012)	(0.0007)
Lottery Index$_{i,t-1}$	0.049***	-0.0005	0.013***
	(0.0024)	(0.0016)	(0.0015)
N	31183	31183	31183
Adj. R^2	0.579	0.103	0.128

Table 11: Panel Data Regression using weekly frequency

The table shows panel data regression results of Model -5, where dependent variables are ATM-IV, Left-Skew, and Right-Skew. Robust standard errors are reported in the parenthesis. All the models have Firm and Month-Year level fixed effects. All the indices are equally weighted. ***, **, and * reflect significance at the 1, 5, and 10% levels, respectively.

Dependent variable	ATM-IV	Left-Skew	Right-Skew
	(1)	(2)	(3)
Systematic Variance$_{i,t-1}$	0.035***	0.007***	−0.003***
	(0.001)	(0.001)	(0.001)
Systematic Skewness$_{i,t-1}$	−0.001	−0.0004	0.001
	(0.001)	(0.0005)	(0.0004)
Fundamental Risk Index$_{i,t-1}$	0.015***	−0.001	0.003***
	(0.002)	(0.001)	(0.001)
Heterogeneity in Opinion Index$_{i,t-1}$	0.028***	−0.003***	0.004***
	(0.001)	(0.001)	(0.001)
Bubble Index$_{i,t-1}$	−0.010***	0.010***	−0.008***
	(0.001)	(0.001)	(0.001)
Lottery Index$_{i,t-1}$	0.058***	−0.003***	0.010***
	(0.001)	(0.001)	(0.001)
Observations	31,183	31,183	31,183
Adj. R^2	0.664	0.169	0.154