Implied Volatility Skew Modelling

Mikhail Borovkov

ID Number: 3146466

Advisor: Claudio Tebaldi

Research thesis presented for the degree of

MSc in Finance

Academic Year 2022/2023

Abstract

This thesis examines the interconnectedness between the equity and stock options markets, specifically studying the impact of the structural risk characteristics on the option-implied volatility skew. A cross-sectional model is constructed, combining several approaches to the skew measure construction, as well as adding industry-specific characteristics related to peers' performance. A cross-sectional regression is used to estimate the market pricing of risk factors, as well as separate the skew into structural and short-term informational components. It is found that the regression performs significantly better during recessions. Furthermore, controlling for the typically used Fama-French factors, returns of factor portfolios are analyzed, with the portfolios based on the informational component of the skew metrics exhibiting the best performance.

Keywords— Option-Implied Volatility – Volatility Surface – Volatility Skew – Factor Portfolio Analysis – Structural Risk – Credit Risk

Table of Contents

1 | Introduction

Although equity markets and options markets are distinctly different systems, they remain vastly interconnected, often influencing one another in multifaceted ways. Given the increasing use of options markets for achieving investment goals by both institutional and retail investors, it becomes ever so critical for investors to understand the underlying intricacies of the markets. Traders and investors in the professional field, as well as researchers in the academic sphere, have extensively studied the phenomenon of price discovery, with the role of the options market being widely explored and debated. Existing research has looked into the degree of that interconnectedness between the two markets, developing diverse methodologies to comprehend the information from the equity markets into signals for the option markets, and vice versa.

Presumably, the information embedded in the price level changes in equity markets can be used to obtain profits by conducting trading activity in the option markets, and vice versa. The novelty in this specific work lies in the introduction of peer-related characteristics, which, to the best of my knowledge, were previously unexplored in a cross-sectional setup, as more robust intra-sector characteristics were utilized.

This thesis aims to measure the degree of predictability of key options market characteristics using features of the equity markets, including fundamental information relating to specific stocks, as well as the overarching metrics related to the performance of a specific industry, in which a specific company operates. The specific options market characteristic that is chosen for the analysis is the option-implied volatility skew, as it is a tangible indicator that has a clear economic intuition, reflecting the general market consensus on the riskiness of the present economic conditions. Using trading jargon, the strategy of 'trading the skew' is one of the most common strategies among the options market participants, as it serves as a reliable hedging instrument, as well as a way to target a specific portfolio of risk exposures.

Hence, the following research questions are set to be answered:

- Can the structural risk factors based on the fundamental company information serve as sufficient predictors of the implied volatility skew?
- Is there potential in capturing industry-related information contained in the sector-wide structural risk characteristics when assessing the skew explainability?
- Does the selected skew measure play a role in the quality of predictability by the structural risk factors? What skew measure yields the best results?
- Is it advantageous to utilize the skew metrics as investment factors in a factor portfolio context? What skew measure proves to be the most advantageous?

In order to answer these questions, a variety of stock-related and options-related data is employed. A cross-sectional model is then created, using a combination of constructed structural characteristics of stocks and various volatility skew metrics. The performance of structural risk factors as skew predictors is then assessed using cross-sectional regressions. Further, the performance of portfolios created using the risk characteristics is studied, using the quantile portfolio approach, as well as the return factor model approach.

The achieved results are in line with the general consensus on the topic. Cross-sectional regression results from the previous research are replicated here, with the introduction of peer-related characteristics proving to be beneficial, as the reported accuracy figures are higher than those in the past papers. Further, decile portfolio analysis reveals various peculiar non-linear behaviors, which are not always consistent with the previous research. Finally, the return factor model showcases the performance of structural risk characteristics, as well as the residual skew, as investment factors.

The remainder of the thesis is structured as follows. Section 2 provides an extensive literature review of the matter, giving a general review of the used concepts related to the equity and option markets, outlining the anomalies that should be taken into account when constructing the empirical setup, as well as going into detail on what predictors of implied volatility skew have been identified before. Section 3 outlines the empirical setup, starting with the overview of the data sources, completed preprocessing, construction of the structural risk characteristics and skew metrics, and the used methodology. Section 4 goes over the results, starting with summary statistics and continuing with market pricing and factor portfolio performance, as well as outlines the next steps, over which the research could be expanded. Section 5 concludes the thesis, reiterating the findings. The Appendix contains the additional analysis outputs that were not included in the main body for brevity purposes, as well as the replication code along with its explanation.

2 | Related Literature

2.1 Implied volatility surface modelling

It is generally accepted that the primary information determining the equity option prices lies in their implied volatility. The first substantial attempt to model the volatility parameter of the equity options came in the form of the (Black & Scholes, 1973) model, which suggested that, when it is derived through backward engineering, the level of implied volatility for all options with the same maturity should converge to a uniform value. In practice, it proved not to be the case, contrary to the created theoretical framework, as there are many phenomena related to market microstructure, such as the supply and demand effects on the derivatives market, that influence the prices, and in turn, the implied volatility levels. Another important underlying assumption of accurate option pricing, which disproves the original ideas of uniform implied volatility level, is liquidity - a continuous trading sequence of a specific derivative at specific prices, facilitated by buyers and sellers present on the market, is necessary for the link between prices and implied volatility to be substantial; lower liquidity can lead to higher prices via liquidity premium (Bongaerts, De Jong, & Driessen, 2011). Wider attention to the implied volatility surface was brought after the market crash of 1987, which saw a sudden drop in equity prices across various markets. The events of the crash suggested that the demand for buying protection against downside ticks in the underlying stock is higher, which was formalized in the derivatives theory by the volatility skew. This modification allowed to account for the observed lognormal, leptokurtic returns, which were highlighted by the 1987 crash. The leptokurtic nature of the returns, otherwise colloquially referred to as 'fat-tailed', is attributed to the occurrent sudden price movements in stock prices, or jumps.

Several extensions of that reasoning followed, such as the local volatility model described in (Derman & Kani, 1994), which produced much more accurate option price valuations across maturities and moneyness levels than the Black-Scholes model, building in a local volatility function at each node of the binomial options pricing model. The asymmetry suggested in their paper helped to bridge the gap between theory and the real market; (Dupire, 1994) and (Rubinstein, 1994) extend the model onto the continuous case. A different approach to extending the (Black & Scholes, 1973) model was taken by (Heston, 1993), which introduced a stochastic component in the widely used stochastic volatility model. That approach allowed to account for the volatility level changes that come from future stock prices and time. In other words, (Heston, 1993) negated the assumption from (Black & Scholes, 1973) that volatility was stable over the duration period of an option. Extensions of stochastic volatility models followed, such as (Bates, 1996), which incorporated the jump-diffusion parameters to asset movements. The overall consensus regarding the volatility skew across all these models, which was also consistent with the empirical observations, was that the skew is persistently negative for the equity indices, but can be either positive or negative for individual stocks given a more discretized nature of the supply and demand trading dynamics, i.e., a relatively smaller number of market participants, as well as lower volume of trading activity can cause certain spikes in the volatility surface for individual stocks.

The measures of implied volatility skew vary across research. (Mixon, 2011) offers various measures used in previous literature, concluding that the most optimal measure is the $(\sigma_{\Delta 25put} - \sigma_{\Delta 25call})/\sigma_{\Delta 50}$ ratio, used in (Carr & Wu, 2007). To dissect the impact of volatility surface features on the fundamental characteristics of a stock, this measure proves to be the most optimal, as according to (Mixon, 2011), it is virtually independent of the implied volatility level itself. This allows standardizing the analysis across individual stock options, which naturally have differing volatility levels at different maturities. Previous research prefers standardized metrics, such as (Foresi & Wu, 2005), which introduces a standardized moneyness measure to study skew behavior in index options. To the best of my knowledge, not much research has been done in terms of comparison of different skew widths, i.e. how does the predictive power of $\Delta 25$ skew compare to that of a tighter $\Delta 40$ skew. This, however, would be a rather trivial distinction in the skew calculation, as the mere difference between out-of-the-money and in-the-money volatility levels is a relatively robust indicator. More advanced skew calculation techniques have been discussed in the existing literature, such as (Schneider & Trojani, 2014), and later in (Schneider, Wagner, & Zechner, 2020). These measures allow to account for the non-linearity of the skew, while still using only the information that is extractable from the option prices, and will be looked at in more detail in Chapter 3.

2.2 Information flow between equity & options markets and fundamental company characteristics

Interdependence between equity and options markets has been widely discussed ever since the 1987 market crash. Information contained in the option implied volatility surface is often used as the source of latent information that could be a predictor of the underlying asset returns. (Bates, 1991) investigates the predictive power of volatility surface parameters (skew premium, jump-diffusion parameters) within the context of the 1987 crash. Stating that the option prices present a clear projection of the market's expectation of an upcoming crash, the researcher reaches conclusive results, noting that it was particularly evident that out-of-the-money puts have been becoming abnormally expensive throughout the year before the crash, which indirectly indicates the predictive power of the volatility skew. (Doran, Peterson, & Tarrant, 2007) examine the predictive power of the option implied volatility levels and skew in predicting both market crashes and upward spikes, concluding that skew has the strongest predictive power in identifying possible short-term market movements. Specifically, researchers state that the skew comprised of implied volatility levels derived from the put option prices strongly predicts short-term market downturns, whereas the call option-derived skew acts as a predictor of short-term upward spikes. Another notable conclusion is that the predictive power of crashes decreases as the time to maturity of the studied implied skew increases. They further underline the hypothesis that the market demand and supply dynamics influence the negativity of the skew. (Bali & Hovakimian, 2009) also investigate the matter, reaching various conclusions: on one hand, the implied volatility levels on their own did not prove to be sufficient predictors of stock returns; on the other hand, the implied volatility skew (called the call-put implied volatility spread in the paper) turns out to be a decent predictor, yielding significant non-zero returns on the quantile portfolios formed using the skew characteristics. Besides that, researchers show a negative persistent link between expected returns and realized volatility skew (call-put realized volatility spread). A positive relationship between the skew and the cross-section of stock returns is also concluded by (Cremers & Weinbaum, 2010). In that paper, the skew phenomenon is closely tied with deviations from the put-call parity, hinting that certain mispricings and arbitrage opportunities could arise in both stock and option markets.

(Yan, 2011) examines the relationship between the volatility skew and the jump, concluding that the stock jump size can be estimated by the observed skew, which serves as a method, that is computationally advantageous compared to the estimation of jump distribution. (Han, Liu, & Tang, 2020) construct a latent jump factor from the volatility surface, and conclude that using the jump factor proves to be efficient in predicting downward jumps but not upward jumps. In the same vein as (Doran et al., 2007), (Han et al., 2020) point out that prices of put options alone serve as good predictors of market crash probability, although concluding that call option prices alone are not good predictors of downward market turns. Less conclusive evidence is found in (Muravyev, Pearson, & Broussard, 2013), which offers a conclusion that option prices do not contain additional information that would serve in exploiting the pricing inconsistencies between equity and options markets, based on the analysis of 39 liquid stock and ETF prices and their respective options. More recently, building upon the approach suggested in the (Muravyev et al., 2013) paper, (Liu, Qiu, Hughen, & Lung, 2019) find evidence of price discovery between Chinese equity and options markets, reaching conclusive results after a more thorough separation of options according to the maturity and moneyness dimensions, as well as the use of more granular, high-frequency data. Once again, the strongest impact from the options market into the equity market is noted to occur in the short-term maturities. Overall, extensive research demonstrates that the implied volatility surface provides valuable insights for predicting stock returns.

In turn, researchers in the equity derivatives space have tackled the task of modelling or forecasting the implied volatility surface to varying degrees of success. Research investigating the predictability of implied volatility with strike prices and maturities as parameters (i.e. economic dynamic factor models, or DFMs) proved to be unscalable onto different individual stocks, with the parameter coefficients being unstable (see, for example, (Christoffersen & Jacobs, 2004)). Other research undertook a different approach by modelling implied volatility through various non-options-related metrics, such as fundamental characteristics of the underlying stocks, or macroeconomic variables. One of the earliest attempts to capture that predictive power was done by (Mixon, 2002), which, employing principal component analysis, underlined the statistical interdependence between implied volatility and macro variables, such as short rates and corporate and government bond yield spreads. (Kelly, Pástor, & Veronesi, 2016) study the impact of political uncertainty metrics on implied volatility, concluding that the presence of major political events (and thus, by definition, political uncertainty) throughout the span of option maturity increases the overall volatility, but its impact on the volatility skew is insignificant. (Robe & Wallen, 2016) conduct a similar analysis in the space of oil derivatives, forecasting implied volatility with variables indicating macroeconomic conditions of the world and the US in particular, as well as financial microstructural indicators. More recently, (Tian & Wu, 2020) highlight two structural risk components that influence the variation of the option implied volatility skew: the default risk of a given company, and a measure of the cyclicality of the business.

Researchers have argued for years that default risk and the volatility skew are interconnected, ever since the emergence of (Merton, 1974) model. For instance, (Hull, Nelken, & White, 2004) provides an extension of the Merton model that links the CDS spread to both the ATM volatility level and the volatility skew, concluding that both have a positive impact on credit risk. These conclusions are also present in (Cremers, Driessen, Maenhout, & Weinbaum, 2008), (Cao, Zhong, & Yu, 2009), (Carr & Wu, 2010), and others.

As for the business cyclicality, (Tian & Wu, 2020) propose that the conditional skewness of the stock return distribution is linked to that of the market return distribution through the cubic of the stock-market correlation, which in turn represents the cyclicality of a business compared to the market. The paper suggests that these two factors explain the majority of the long-term variation in skew, while the remaining variation is attributed to the short-term information flow. In this case, both equity market information (market capitalization, realized volatility) and fundamental company data (quarterly outstanding debt figures) are used as predictive features of the implied volatility skew.

Another source for the predictive power of option implied volatility surface could be various contagion effects from volatilities of other indices in the case of index volatility surface modelling, or sub-sector indices in the case of individual stocks volatility surface modelling. For the former, research has found significant evidence of spillover. (Füss, Mager, Wohlenberg, & Zhao, 2011) investigate the link between the German and US macroeconomic events and their impact on the option implied volatilities in their respective markets, VDAX and VIX; their findings suggest that there exists a significant spillover effect, as well as covariance clustering between the two. (Jiang, Konstantinidi, & Skiadopoulos, 2012) examine the spillover between the US and European stock markets using implied volatility indices, with their results supporting the notion of volatility contagion; (Kenourgios, 2014), (C. Y. H. Chen, 2014) and (J. Chen, Han, Ryu, & Tang, 2022) further confirm the results for various US and European volatility indices. Similar conclusions have been drawn for the intra-sector spillover, although liquidity concerns arise for individual stocks, which limits the extent of results scalability. (Bernales & Guidolin, 2014) study the predictability of individual stock options volatility surface using movements in the S&P 500 index options volatility surface. Researchers underline the conclusive evidence of strong cross-sectional interconnection between the index option and stock option volatility surfaces. (Krause & Lien, 2014) analyze the implied volatility contagion between ETF options and their largest component stocks, highlighting the presence of both a market volatility factor and an industry volatility factor; (Bouri, Lucey, & Roubaud, 2020) report similar findings.

(Hann, Kim, & Zheng, 2019) track the information spillover within industries around earnings releases of US firms, using implied volatilities, confirming a significantly positive interconnection between industry peers. Most of the described papers on volatility contagion employ autoregressive time series models (such as GARCH or EGARCH), however, papers such as (Bernales & Guidolin, 2014) or (Hann et al., 2019) use cross-sectional methods as well.

Overall, it is evident that there is potential in dissecting the features of option implied volatility surface in order to infer conclusions on company performance, and vice versa. The practical use of this analysis is identifying the pricing inconsistencies within both markets, which traders could potentially capitalize on by making bets in the derivative markets. Moreover, conclusive evidence on the connection between skew and fundamental company characteristics would mean modified hedging implications for traders.

2.3 Anomalies in the relationship between returns and credit risk

Credit risk is a major structural risk factor affecting financial performance, and perhaps, the most studied one in academic research. Traditionally, models such as the standard Capital Asset Pricing Model (CAPM), poorly capture the premium for the individual default risk, solely focusing on the systematic risk component, measured as beta. (Campbell, Hilscher, & Szilagyi, 2008) examine the so-called distress risk, which is a composite measure of the potential financial failure or bankruptcy of a company, based on the accounting figures and credit ratings. Researchers track the performance of portfolios based on the distress risk factor, reaching conclusions that portfolios with higher produce abnormally low returns, even when controlling for size and momentum, which is contrary to the traditional view that higher risk yields higher returns. (Chang, Christoffersen, & Jacobs, 2013) investigate the link between sensitivity to the option implied volatility skewness and the abnormal returns, showing that higher skewness leads to low returns; (Amaya, Christoffersen, Jacobs, & Vasquez, 2015) reaffirm the findings using intraday data. (Schneider et al., 2020) replicate the results presented in (Campbell et al., 2008), using a framework that implements the option implied volatility skew as a measure of company distress expectations, showing that the low-risk anomalies can be explained by the skewness of the return. Following the methodology of (Schneider & Trojani, 2014), researchers introduce upper and lower skew metrics to compute the ex-ante skewness, which brings the benefit of separating the information from prices of OTM puts and OTM calls. The paper further shows the link between the alpha of a firm and its ex-ante returns skewness, as well as realized ex-post coskewness. They conclude that the abnormally low returns exhibited by the high-risk stocks are only considered abnormal compared to the performance of portfolios based on CAPM beta, which is a fundamentally flawed predictor only accounting for the market-related risk.

3 | Empirical Setup

3.1 Data collection

An empirical analysis was performed on a cross-section of data related to S&P 500 constituents from January 1997 to December 2021, involving daily stock price data, out-of-the-money option prices data, option-implied volatility surface data, and quarterly company fundamental and categorical information. Chronologically, data was sampled to contain observations from 6238 days of trading activity in the span of 25 years.

Several types of data were collected from various sources. First, adjusted closing stock price data on S&P 500 constituents was collected using Yahoo!Finance. Since the S&P 500 composition is dynamic, measures were taken to accommodate the creation of a cross-sectional dataset, meaning that the relevant data was only collected for the dates when stocks were included in the S&P 500 composition. This data, specifically the return series and their 1-year rolling correlation to the S&P 500 index return series, will be relevant in the risk factor creation stage. Next, the 3-month option volatility surfaces for fixed delta levels were extracted for the same stocks, using publicly available data from the data vendor OptionMetrics. This data is sufficient to construct the ATM implied volatility level, as well as the basic skew measures. OTM option price data from OptionMetrics was used for the construction of more complicated skew measures, which take into account information from each available option, rather than just the options with fixed delta levels. This data also allows to decompose the skew into the lower skew and upper skew components. Zero coupon price data was additionally employed in the calculation of advanced skew metrics, which was extracted using Nasdaq Data Link's APIs. Furthermore, the fundamental company data, including the quarterly reported figures of share prices, number of shares outstanding, and total firm debt outstanding, was extracted from Compustat. Additionally, Compustat provides the official classification standards from S&P GICS, which classifies the stocks into 11 economy sectors.

Considering the amount of stock-specific public data required for the analysis, the stock sample was naturally filtered to companies that check all the boxes for data availability. The final sample contained 821 stocks, each of which at some point throughout the studied period was a constituent of the S&P 500 index.

3.2 Preprocessing, construction of structural characteristics and skew measures

Following the methodology of several papers discussed in the literature review, the following features, which represent the structural risk sources, are constructed:

- **Business cyclicality:** Following the methodology of (Tian & Wu, 2020), a measure of business cyclicality is constructed, according to the proposition that the skewnesses of stock and market return distributions are cubically correlated. To represent the scale of the skew measure appropriately, the cubic correlation of stock return and index return is multiplied by the ATM volatility level, which is computed as the average of the $\Delta 50$ call and $\Delta 50$ put volatilities:

$$C_{j,t} = \rho^3_{j,mkt} * \frac{\sigma_{\Delta 50 call,j,t} + \sigma_{\Delta 50 put,j,t}}{2}$$

- **Default risk:** Implementing the (Merton, 1974) model, which uses the market equity value, the equity volatility, and the face value of debt to infer the probability of default of a company. By definition, default occurs when a company is not able to pay off its debts, thus default risk serves as a likelihood measure of that scenario. A standardized measure of default risk is calculated from the following system of non-linear equations for each quarterly company-specific observation:

$$\begin{cases} E_{j,t} = V_{j,t} * N(DD_{j,t} + \sigma_F * \sqrt{\tau}) - D_{j,t} * N(DD_{j,t}) \\ \sigma_E = N(DD_{j,t} + \sigma_F * \sqrt{\tau}) * V_{j,t} * \sigma_F / E_{j,t}, \end{cases}$$

where $E_{j,t}$ is the equity value of the firm, estimated by the market capitalization of equity, $V_{j,t}$ is the firm value, σ_F is the firm return volatility, τ is the debt maturity (taken as 10 by default), σ_E is the one-year stock return volatility, $D_{j,t}$ is the total debt outstanding, and the distance-to-default ($DD_{j,t}$) measure takes the following form:

$$DD_{j,t} = \frac{ln(V_{j,t}/D_{j,t}) - 0.5\sigma_F^2 * \tau}{\sigma_F * \sqrt{\tau}}$$

These two non-linear equations are solved numerically for the two unknowns, the firm value V, and the firm return volatility σ_F. Once the distance-to-default measure is computed, the corresponding probability of default, or default risk, is the following:

$$\pi_{default,j,t} = -\frac{ln(N(DD_{j,t}))}{\tau}$$

It is worth noting that the interest rate is assumed to be zero in this setup, which was deemed suitable for the economic conditions of the studied time period (pre-2022). Interpolation of the analysis over some more recent data should involve some additional interest rate considerations. Furthermore, since this component is constructed based on the quarterly figures, whereas all the other cross-sectional data is daily, this component is linearly interpolated to fit values for dates other than the dates of quarterly reports.

- **Intra-sector characteristics:** This is a novel introduction to the cross-sectional analysis of skew predictability, as to the best of my knowledge, no industry-specific characteristics aside from the dummy industry variables have been used in cross-sectional analysis. The introduction of these characteristics is an attempt to extract sector-specific information in a way that is less robust than the use of dummy variables. Ideally, volatility surface data for industry-specific ETFs would have been used directly, as they are traded products on the Chicago Board Options Exchange (CBOE). Unfortunately, data on volatility surfaces of industry-specific ETFs is not publicly available through OptionMetrics or other data vendors available on the Wharton Research Data Services (WRDS) portal. Because of that, the intra-sector skew is calculated as the average skew of the industry peers. To avoid direct collinearity issues, the stock j, for which the parameter is calculated, is excluded from the sector sample:

$$C_{peers,j,t} = (\sum_{i=1}^{n} \frac{C_{i,t}}{n} - \frac{C_{j,t}}{n}) * \frac{n}{n-1}$$

$$\pi_{peers,j,t} = (\sum_{i=1}^{n} \frac{\pi_{i,t}}{n} - \frac{\pi_{j,t}}{n}) * \frac{n}{n-1}$$

Further, the analysis will also be controlled for the ATM volatility level (both for single stock and for peers in the economy sector):

$$\sigma_{ATM,j,t} = \frac{\sigma_{\Delta 50call} + \sigma_{\Delta 50put}}{2}$$

$$\sigma_{ATM,peers,j,t} = (\sum_{i=1}^{n} \frac{\sigma_{i,t}}{n} - \frac{\sigma_{j,t}}{n}) * \frac{n}{n-1}$$

Finally, for the factor portfolio performance analysis, several traditional risk factors were introduced — size, book-to-market ratio, and stock return momentum. These factors are commonly included in such studies, and their calculation followed the standard procedures used in previous literature, such as (Fama & French, 1992) or (Bali, Engle, & Murray, 2016).

As for the target variable, i.e. skew, the following several approaches are taken to compute the total of four skew measures:

- Following the methodology of (Tian & Wu, 2020), a *basic skew* measure at maturity T is computed as a difference between the implied volatilities σ of an OTM call and an OTM put.

$$basicSKEW_{j,t,T} = \sigma_{\Delta 25call,j,t,T} - \sigma_{\Delta 25put,j,t,T}$$

  The primary benefits of this skew measure are its trivial computation, as well as its simple interpretability – the resulting figure is exactly how much cheaper the OTM call is compared to the OTM put of the same delta level.

- Following the methodology of (Schneider et al., 2020), separate *lower skew* and *upper skew* measures are computed. The lower skew is calculated using the prices of OTM puts, i.e. put options with delta levels higher than -0.5, and it represents the left half of the skew distribution. Conversely, the upper skew is based on the prices of OTM calls, i.e. call options with delta levels lower than 0.5, and it represents the right half of the skew distribution. The following are the formulas used in lower and upper skew calculation:

$$lowerSKEW_{j,t,T} = -\frac{6}{p_{t,T}} \left(\int_0^{F_{t,T}} \left(ln \frac{F_{t,T}}{K} \right) \frac{\sqrt{\frac{K}{F_{t,T}}} P_{t,T}(K)}{K^2} dK \right),$$

$$upperSKEW_{j,t,T} = \frac{6}{p_{t,T}} \left(\int_{F_{t,T}}^{\infty} \left(ln \frac{K}{F_{t,T}} \right) \frac{\sqrt{\frac{K}{F_{t,T}}} C_{t,T}(K)}{K^2} dK \right),$$

  where $p_{t,T}$ is the price of the zero-coupon bond with maturity T, $F_{t,T}$ is the forward price of the underlying stock, K is the option strike level, and $P_{t,T}(K)$ and $C_{t,T}(K)$ are the put and call option prices at the corresponding strike level K.

  These measures allow separation of the information flow contained in both parts of the skew distribution, as well as incorporate additional dynamics between different OTM options, accounting for differences in strikes of consecutive options. Overall, the integral-based skew measures can be interpreted as measures of the area under the skew density graph, where the impact of option prices closer to the at-the-money level (i.e., the peak of the skew distribution) is the highest, while the impact of option prices that are far out-of-the-money (i.e., the tails of the skew distribution) is lower.

In the computation of these volatility skew metrics, a limitation was introduced for only the options with an absolute delta higher than 0.05 to be considered in the integral, as the impact of such options in the grand scheme of things is negligible. The trading activity of these options is in most cases discrete and irrelevant to the other parts of the volatility surface.

Putting lower and upper skew components together, *composite skew* is calculated as:

$$compSKEW_{j,t,T} = upperSKEW_{j,t,T} + lowerSKEW_{j,t,T} =$$
$$= \frac{6}{p_{t,T}} \left(\int_{F_{t,T}}^{\infty} \left(ln\frac{K}{F_{t,T}} \right) \frac{\sqrt{\frac{K}{F_{t,T}}} C_{t,T}(K)}{K^2} \, dK - \int_{0}^{F_{t,T}} \left(ln\frac{F_{t,T}}{K} \right) \frac{\sqrt{\frac{K}{F_{t,T}}} P_{t,T}(K)}{K^2} \, dK \right).$$

Evidently, the deployment of these integral-based skew measures allows to account for much more information contained in the option prices. As (Schneider et al., 2020) points out, considering the lower skew and upper skew measures separately adds a degree of explainability – the same values of the composite skew can arise from various combinations of lower skew and upper skew. However, the computation of these measures turns out to be significantly more resource-intensive. Moreover, their interpretability is not as intuitive as in the basic skew case.

Another challenge that arises with the use of these skew measures is the unavailability of daily single-stock options that precisely align with the desired days-to-maturity period, such as 3 months. To tackle this limitation, aggregation of data on options with nearby maturities is performed. Specifically, the chosen options are those with maturities with the sample average days-to-maturity period converging to the desired length. It is worth noting that in the case of the basic skew construction using volatility surface data, the problem is not present as OptionMetrics aggregates data to produce volatility skews with fixed days-to-maturity periods, potentially using a similar technique.

In order to improve the interpretability of the used data, extensive preprocessing of the features was conducted – since different features naturally have varying units and scales, such a step is necessary. Specifically, the features data, with the exception of the skew measures, was normalized and winsorized. First, all six structural risk factors, as well as three traditional factors (size, book-to-market ratio, return momentum), were winsorized across the entire data sample at the 1% percentile threshold on both tails of the distribution in order to minimize the impact of extreme outliers. Following that, the data was separated into cross-sectional sub-datasets for each trading day, where the nine variables were standardized by the cross-sectional median values and cross-sectional standard deviations. Finally, a second round of winsorization was completed, again at the 1% percentiles on both distribution ends.

Date	Ticker	C_j	π_j	$\sigma_{ATM,j}$	C_{peers}	π_{peers}	$\sigma_{ATM,peers}$
1997-01-02	AAPL	0.0034	0.0542	0.5140	0.0190	0.0086	0.4790
1997-01-02	EMN	0.0073	0.0001	0.2293	0.0136	0.0010	0.2645
1997-01-02	FMC	0.0215	0.0002	0.2156	0.0126	0.0010	0.2655
...	...	...	...	...	...	...	...

Table 3.1: Structural risk factors sample

Date	Ticker	$basicSKEW_j$	$lowerSKEW_j$	$upperSKEW_j$	$compSKEW_j$
1997-01-02	AAPL	-0.0246	-0.0029	0.0021	-0.0008
1997-01-02	EMN	-0.0037	-0.0004	0.0003	-0.0001
1997-01-02	FMC	-0.0109	-0.0015	0.0012	-0.0003
...	...	...	...	...	...

Table 3.2: Skew measures sample

The resulting dataset, after the applied normalization and winsorization procedures, contains over 2.1 million daily company-specific observations. Samples of both features and skew measures are presented in Tables 3.1 and 3.2.

3.3 Methodology

A cross-sectional regression is going to be used, approximating the following factor model:

$$S_{j,t} = \beta_{cycle} * C_{j,t} + \beta_{default} * \pi_{default,j,t} + \beta_{ATM} * \sigma_{ATM,j,t} +$$
$$+ \beta_{peers,cycle} * C_{peers,t} + \beta_{peers,default} * \pi_{peers,t} + \beta_{peers,ATM} * \sigma_{peers,ATM,t} + \epsilon_{j,t},$$

where $S_{j,t}$ is one of the four identified skew measures (basic skew, lower skew, upper skew, and composite skew), and $\beta_{cycle}, \beta_{default}, \beta_{ATM}, \beta_{peers,cycle}, \beta_{peers,default}, \beta_{peers,ATM}$ are estimated coefficients of each of the six outlined risk factors.

The predictive power of this cross-sectional model is analyzed for options with maturities of roughly three months. As was mentioned before, an aggregation technique, similar to the one employed in (Muravyev et al., 2013), is used to gather options data with expiration date arriving in between 50 and 150 days – the average expiration date of options filtered by these boundaries arrives at around 92 days, or three months.

While the accuracy of the regression is practical when contrasting the influence of specific regressors, the accuracy of a model alone cannot be sufficient to conclude the presence of a viable investment opportunity based on the identified factors. Thus, after constructing the timeline of the cross-sectional regression, market pricing of each risk characteristic is studied via portfolio performance. First, skew-based decile portfolios, formulated using the actual skew metrics, are constructed, as well as the skew measures from the cross-sectional regression analysis, namely the predicted (structural) skew and the residual (informational) skew. Further, stock factor portfolios are formed, according to the (Fama & MacBeth, 1973) approach, widely used and developed in the last half a century of research in financial markets. The performance of structural skew components, as well as the residual skew, is studied and assessed using criteria of statistical significance, controlled for the commonly used traditional risk factors - size, book-to-market ratio, and stock return momentum.

4 | Results

The following is a description of the achieved results. First, summary statistics are presented in order to gain a better understanding of the statistical properties of both predictive features and the skew parameters. Then, the regression in the cross-sectional model is studied, with market pricing of risk characteristics presented. Further, portfolio performance is studied – first, according to the decile portfolio framework, and then in a return factor model setup.

4.1 Summary statistics

	Moments		**Percentiles**						
Variable	Mean	StDev	1	10	25	50	75	90	99
Basic Skew	-4.19	5.28	-20.65	-8.85	-5.90	-3.95	-2.21	0.00	8.76
Comp. Skew	-0.72	3.05	-9.72	-1.38	-0.48	-0.17	-0.07	-0.02	0.22
Lower Skew	-0.99	3.45	-12.07	-1.95	-0.74	-0.29	-0.13	-0.06	-0.02
Upper Skew	0.27	0.71	0.00	0.02	0.04	0.10	0.24	0.58	2.96
C_j	6.43	6.04	0.02	0.75	2.05	4.56	8.83	14.90	29.12
π_j	4.43	12.52	0.00	0.00	0.01	0.34	3.15	12.25	99.22
$\sigma_{ATM,j}$	31.64	13.94	13.87	18.04	21.90	28.06	37.37	49.53	86.43
C_{peers}	6.45	5.23	0.17	1.53	2.96	4.81	8.41	13.72	25.86
π_{peers}	5.08	6.34	0.07	0.47	1.08	2.68	6.85	12.67	36.80
$\sigma_{ATM,peers}$	31.77	11.10	16.81	20.67	24.03	28.89	36.87	45.78	74.96

Table 4.1: Summary statistics of skew metrics and structural risk factors.

The following can be observed from the summary statistics (see Tables 4.1 and 4.2):

- Skew levels on average are negative, as expected. The average basic skew is -4.19, which indicates that on average, the implied volatility of a $\Delta 25$ call option is lower than the implied volatility of a $\Delta 25$ put option by 4.19%. This result is reasonably consistent with (Tian & Wu, 2020) – the average skew value in that paper is slightly higher at -4.81, which suits the intuition that smaller and riskier stocks have steeper corresponding volatility skews, as firms outside of the S&P 500 constituent list are included in the sample

Variable	Correlation			
	Basic Skew	Comp. Skew	Lower Skew	Upper Skew
C_j	-11.85	-7.34	-7.51	5.57
π_j	-7.15	-14.92	-17.40	16.54
$\sigma_{ATM,j}$	-14.48	-47.88	-57.35	60.60
C_{peers}	-10.00	-6.95	-7.62	6.65
π_{peers}	-4.31	-3.48	1.69	-3.48
$\sigma_{ATM,peers}$	-8.85	-16.58	-19.99	20.76

Table 4.2: Correlation statistics between structural risk factors and skew metrics.

there. Moreover, about 90% of skew observations are negative (91.63% are negative in the (Tian & Wu, 2020) paper).

- All six risk factors exhibit a slightly negative correlation with the basic skew parameter, which is also consistent with the described related literature, as well as common sense. The higher the implied volatility level itself is, the lower the skew becomes (or the higher in absolute value for negative values of skew), signifying traders' increased demand for the downtick scenarios.

- As for the advanced skew characteristics, similar dynamics can be observed. By construction, the lower skew figures are all negative, while the upper skew yields positive figures. It is notable that the amplitude of the lower skew is higher than that of the upper skew, which is consistent with the general consensus that the prices of OTM puts are the primary drivers of the implied volatility skew dynamics. Furthermore, all four skew characteristics exhibit variance coefficient ratios over 1 (i.e. standard deviation than the absolute mean), which suggests a high-variance distribution of the skew. Variance coefficients of the advanced skew metrics are significantly higher than that of the basic skew metric.

- From the correlation statistics of the three advanced skew characteristics, the figures regarding the correlation with the implied volatility level jump out, as they are significantly higher than any other correlation figures. Correlation numbers are reasonably similar for the basic and composite skew, outside of correlation with implied volatility level. This difference alone, though, suggests that the regression accuracy could differ vastly for these skew measures.

4.2 Regression results

	(1) Basic Skew	(2) Comp. Skew	(3) Lower Skew	(4) Upper Skew
C_j	-0.079	0.006	0.012	-0.005
	(0.11)	(0.05)	(0.05)	(0.01)
π_j	-0.027	-0.031	-0.039	0.008
	(0.09)	(0.14)	(0.16)	(0.02)
$\sigma_{ATM,j}$	-0.046	-0.066	-0.094	0.028
	(0.07)	(0.06)	(0.07)	(0.02)
C_{peers}	-0.025	0.024	0.028	-0.004
	(0.21)	(0.08)	(0.09)	(0.02)
π_{peers}	-0.097	-0.024	-0.029	0.005
	(0.30)	(0.11)	(0.11)	(0.04)
$\sigma_{ATM,peers}$	-0.002	0.020	0.029	-0.008
	(0.08)	(0.05)	(0.05)	(0.01)
R^2	7.324	30.151	40.167	42.380
	(5.80)	(18.68)	(20.89)	(15.36)

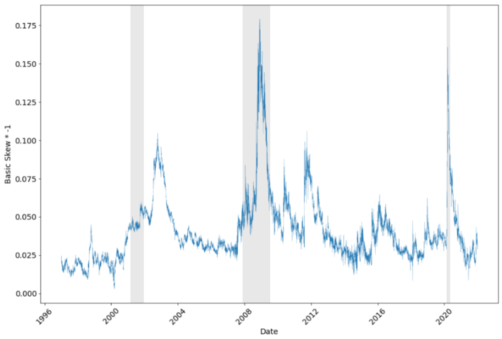
Table 4.3: Market pricing of company risk characteristics on the implied volatility skew

Table 4.3 presents the aggregated results of the regressions in the cross-sectional model. For the basic skew, R^2 summary statistics are consistent with (Tian & Wu, 2020), as single R^2 measures span from as little as 0.14% to 37.19%. The R^2 figures for the three advanced skew characteristics are significantly higher – from 0.35% to 75.58%, from 0.85% to 81.59%, and from 3.29% to 75.87% for composite, lower, and upper skew respectively. It is worth noting that the figures reported in parentheses should not be interpreted as t-statistics of a specific estimated regression coefficient – the reported figures, both the coefficients and the standard deviations are the averages across the cross-sectional regression.

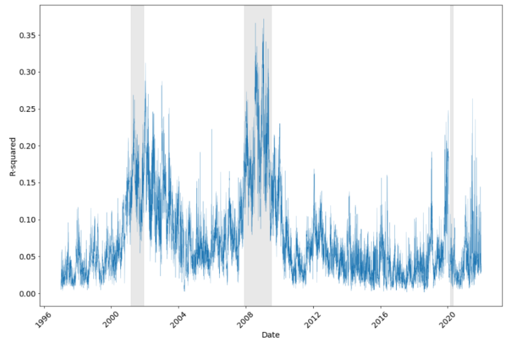
A peculiar difference can be noted in the average market pricing of structural risk factors in modelling of the basic skew metric and the advanced skew metrics - while all six factors exhibit negative mean market pricing estimates in the basic skew case (which is consistent with (Tian & Wu, 2020)), the same does not hold for the advanced skew metrics case. Taking a look at the composite skew, it is noteworthy that for three of the six factors, specifically - stock-specific business cyclicality, industry-specific business cyclicality, and the industry-specific at-the-money implied volatility level, the signs of the market pricing estimates differ from the signs of the correlation coefficients. This circumstance suggests that while the unconditional correspondence of structural risk factors with the skew may be negative, interdependence patterns among the structural risk factors suggest the presence of a confounding variable, the effect of which is not present in the market pricing estimates of the basic skew. More detailed market pricing statistics, including cross-sectional percentiles of regression coefficients and R^2, can be found in Appendix A.

Figure 4.1 contains representations of the time-varying dynamics of the used data. The negative part of the skew is presented in the figures on the left (subfigures *(a), (c), (e), (g)*; the upper skew is the exception – it is represented without changing its sign, as was evidently suggested by the correlation figures discussed earlier), while the predictive power of the cross-sectional model, measured in R^2, is depicted in the right figures (subfigures *(b), (d), (f), (h)*). The grey-shaded areas in the graphs depict the US recessions, as reported by the NBER's Business Cycle Dating Committee.

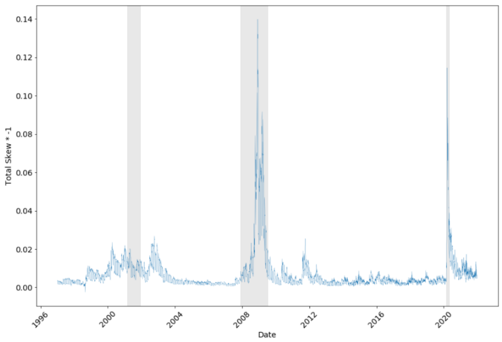
The basic skew data is virtually identical to the one presented in (Tian & Wu, 2020), with the only significant difference coming from the frequency of data (daily here as opposed to weekly in (Tian & Wu, 2020)), which yields outliers in the early years of the sample. The time progression of the R^2 data is also consistent with the findings in the (Tian & Wu, 2020), as skew turns out to be explained the most in times of recessions, spiking massively to all-time high values during the Global Financial Crisis, and reaching similar peaks in 2020 at the



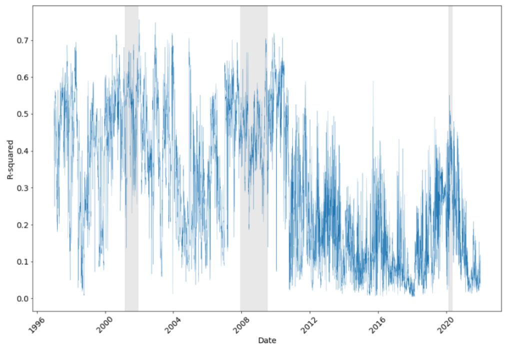
(a) Negative basic skew



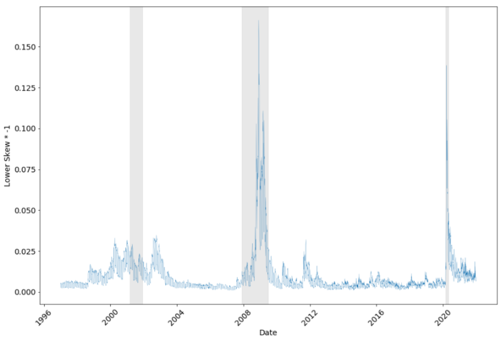
(b) R^2 for basic skew



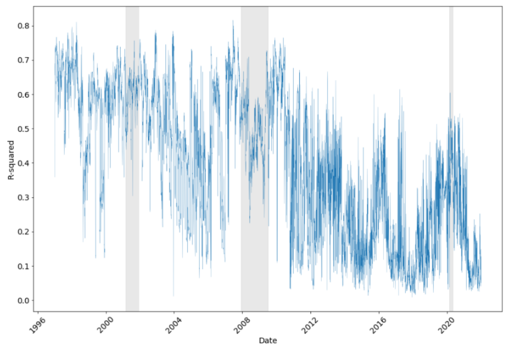
(c) Negative composite skew



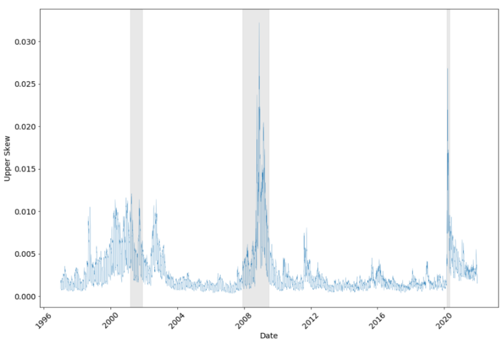
(d) R^2 for composite skew



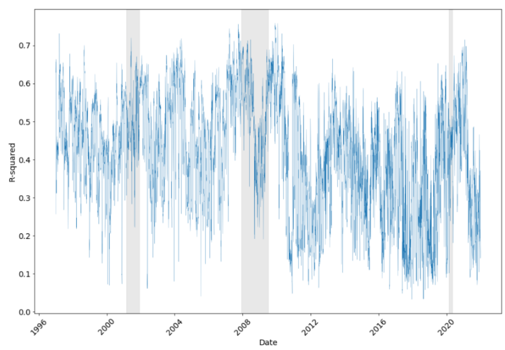
(e) Negative lower skew



(f) R^2 for lower skew



(g) Upper skew



(h) R^2 for upper skew

Figure 4.1: Cross-sectional explanatory power of the pricing model on the implied volatility skew

start of the COVID-19 pandemic. Intuitively, the two figures seem to move simultaneously in the same directions – as the absolute skew figures become higher in times of recessions when downside protection becomes particularly expensive, the explainability of the skew parameter also becomes higher in the cross-sectional model.

The advanced skew characteristics exhibit similar dynamics to that of the basic skew parameter. The upper skew measure is higher in the earlier years when compared to the all-time peaks, which is most likely a result of lesser trading activity in the in-the-money regions (in terms of put options), and hence, lower liquidity and more discrete price gaps between the options restricting the hedging opportunities. As for the time-varying dynamics of the advanced skew measures, slightly less conclusive evidence is reached. While, like the basic skew characteristic, all three advanced skew characteristics spike in value during the recessions, the intuition behind the evolution of the cross-sectional R^2 is less clear. It could be noted from the graph shapes it is evident that the explainability of the lower skew dominates over the upper skew, as the subfigures related to the lower and composite skew are virtually identical, specifically in the earlier years of the studied period. Overall, it is again evident that the skew explainability is at all times higher than that of the basic skew, most probably due to higher correlations between the implied volatility level and skew characteristics than in the case of the basic skew.

Table 4.4 examines the summary statistics of the two skew components obtained from the regressions - structural skew (predicted value of the skew based on the six outlined structural risk factors), and informational skew (prediction residual, which is attributed to the short-term information flow). Intuitively, while the structural skew is supposed to represent the priced-in factors that are deducible from the equity markets and fundamental company information, the informational skew is thought to capture the short-term temporal market effects, which could be more attributed to the market microstructure phenomena.

The results are consistent with (Tian & Wu, 2020). The structural skew figures are consistently lower than those of the informational skew. The informational skew is centered around zero, suggesting that the changes in the skew, which cannot be attributed to the structural factors,

Percentiles	Structural Skew 10	25	50	75	90	Informational Skew 10	25	50	75	90
Mean	-0.264	-0.130	-0.043	0.035	0.112	-0.152	-0.076	-0.002	0.083	0.226
StDev	0.180	0.234	0.326	0.423	0.486	0.183	0.235	0.324	0.414	0.470
Auto	0.606	0.658	0.700	0.750	0.838	0.609	0.663	0.699	0.748	0.829

(a) Basic skew

Percentiles	Structural Skew 10	25	50	75	90	Informational Skew 10	25	50	75	90
Mean	-0.107	-0.069	-0.026	0.009	0.036	-0.075	-0.050	-0.017	0.022	0.059
StDev	0.086	0.110	0.150	0.190	0.305	0.094	0.115	0.152	0.187	0.287
Auto	0.720	0.777	0.838	0.898	0.937	0.679	0.748	0.820	0.873	0.914

(b) Composite skew

Percentiles	Structural Skew 10	25	50	75	90	Informational Skew 10	25	50	75	90
Mean	-0.140	-0.086	-0.033	0.014	0.050	-0.089	-0.055	-0.007	0.038	0.095
StDev	0.098	0.127	0.168	0.219	0.326	0.106	0.131	0.170	0.214	0.314
Auto	0.758	0.819	0.874	0.918	0.944	0.725	0.792	0.853	0.899	0.925

(c) Lower skew

Percentiles	Structural Skew 10	25	50	75	90	Informational Skew 10	25	50	75	90
Mean	-0.139	-0.006	0.006	0.019	0.035	-0.086	-0.065	-0.047	-0.031	-0.020
StDev	0.023	0.028	0.037	0.045	0.056	0.040	0.048	0.060	0.076	0.095
Auto	0.793	0.840	0.876	0.901	0.919	0.540	0.662	0.779	0.864	0.899

(d) Upper skew

Table 4.4: Summary time-series statistics on the structural and informational skew.

can sway the skew value in any direction with more or less equal probability – this finding is intuitive considering the transient nature of the impact of the short-term information flow. Moreover, this is consistent with the widely accepted random walk hypothesis in the underlying stock behavior.

The autocorrelation figures of the structural skew, on the other hand, are comparatively smaller across each specified percentile, meaning that the skew estimates are less persistent than the ones formed in the reference paper. This could be the result of the inclusion of three peer-related characteristics, which diversify the structural component of the skew, leading to less consistent figures over the successive data points. The autocorrelation percentile figures for the informational skew are conversely higher than those in the reference paper on average.

4.3 Decile portfolio performance

The first approach to the assessment of the performance of the risk characteristics is the quantile portfolio approach. Decile portfolios were created in the following way: at each date, the set of present stock entries is sorted based on the chosen skew characteristic. Then the sorted subdataset is sliced into ten sections, forming decile portfolios of stocks. This is done for each of the four skew measures, and then the results are compared.

Tables 4.5 and 4.6 represent the decile portfolio performance expressed in annualized returns across three investment horizons (one week, one month, one quarter) and for each of the three skew estimates (total, structural, informational). t-statistics are presented in the parentheses next to each of the annualized returns estimates.

Overall, the annualized return figures are significantly higher than the ones presented in (Tian & Wu, 2020). This difference can be attributed to a certain survivorship bias of the company sample, as it centers around companies that have been constituents of the S&P 500 index at some point during the studied period, as opposed to company filtering based on market capitalization performed in (Tian & Wu, 2020). For the same reasons, all portfolio performance figures are statistically significant, while that is not the case in (Tian & Wu, 2020).

What is more interesting is the fact that while in (Tian & Wu, 2020), portfolio performance increases with increasing skew, here both low- and high-decile portfolios outperform the middle-decile portfolios based on the total basic skew across each of the three investment horizons (Table 4.5.a). When the skew is decomposed into the structural and informational parts, the portfolio behavior shifts. Decile portfolios based on stocks with the lowest structural basic skew outperform the portfolios with the highest structural basic skew. This result is inconsistent with

(a) Basic skew

Decile	One Week			One Month			One Quarter		
	Total	Structural	Informational	Total	Structural	Informational	Total	Structural	Informational
1	0.23 (7.90)	0.24 (8.13)	0.14 (8.98)	0.22 (17.23)	0.22 (16.25)	0.14 (18.71)	0.22 (31.70)	0.20 (27.50)	0.15 (36.70)
2	0.19 (8.13)	0.28 (12.08)	0.16 (9.73)	0.19 (17.30)	0.26 (23.45)	0.16 (20.16)	0.18 (31.20)	0.25 (35.38)	0.16 (36.33)
3	0.17 (8.69)	0.24 (11.71)	0.17 (9.46)	0.19 (19.21)	0.22 (21.87)	0.16 (18.68)	0.17 (34.02)	0.20 (38.07)	0.17 (35.11)
4	0.18 (9.92)	0.19 (9.71)	0.18 (9.79)	0.17 (19.50)	0.19 (20.87)	0.18 (19.42)	0.17 (35.40)	0.19 (38.99)	0.17 (33.99)
5	0.18 (10.10)	0.18 (9.49)	0.18 (9.57)	0.17 (20.31)	0.18 (19.79)	0.18 (20.21)	0.17 (36.55)	0.18 (36.06)	0.17 (35.24)
6	0.18 (10.77)	0.18 (9.44)	0.19 (9.62)	0.16 (20.63)	0.18 (19.63)	0.18 (19.74)	0.16 (37.83)	0.17 (34.77)	0.18 (37.05)
7	0.19 (11.62)	0.18 (9.18)	0.21 (10.12)	0.17 (22.24)	0.17 (18.77)	0.19 (21.26)	0.16 (39.64)	0.17 (33.64)	0.20 (39.88)
8	0.19 (11.73)	0.18 (9.98)	0.25 (11.54)	0.17 (23.00)	0.16 (18.70)	0.22 (21.67)	0.17 (43.27)	0.17 (34.96)	0.20 (38.02)
9	0.22 (12.07)	0.17 (10.18)	0.27 (12.44)	0.20 (23.92)	0.16 (20.56)	0.25 (23.01)	0.19 (41.52)	0.16 (36.35)	0.24 (36.38)
10	0.23 (11.94)	0.15 (9.31)	0.25 (8.58)	0.23 (24.00)	0.14 (19.22)	0.22 (16.78)	0.21 (40.50)	0.15 (37.48)	0.20 (28.20)

(b) Composite skew

Decile	One Week			One Month			One Quarter		
	Total	Structural	Informational	Total	Structural	Informational	Total	Structural	Informational
1	0.23 (7.90)	0.25 (7.50)	0.17 (10.70)	0.23 (17.23)	0.25 (16.04)	0.16 (20.15)	0.22 (31.70)	0.26 (30.57)	0.15 (36.84)
2	0.19 (8.13)	0.25 (9.60)	0.17 (10.47)	0.19 (17.30)	0.22 (18.72)	0.17 (21.04)	0.18 (31.20)	0.22 (33.08)	0.16 (36.28)
3	0.17 (8.69)	0.21 (9.83)	0.17 (10.20)	0.18 (19.21)	0.21 (21.32)	0.17 (20.47)	0.17 (34.02)	0.20 (37.49)	0.16 (37.64)
4	0.18 (9.92)	0.22 (11.16)	0.18 (10.18)	0.17 (19.50)	0.19 (21.06)	0.17 (20.00)	0.17 (35.40)	0.19 (37.59)	0.16 (36.09)
5	0.18 (10.10)	0.21 (11.16)	0.18 (9.97)	0.17 (20.31)	0.17 (20.09)	0.17 (20.20)	0.17 (36.55)	0.17 (35.75)	0.16 (36.43)
6	0.18 (10.77)	0.16 (9.26)	0.19 (10.00)	0.17 (20.63)	0.16 (19.15)	0.17 (19.86)	0.16 (37.83)	0.16 (35.07)	0.17 (36.64)
7	0.19 (11.62)	0.18 (10.45)	0.22 (11.18)	0.17 (22.24)	0.17 (20.56)	0.20 (21.47)	0.16 (39.64)	0.16 (36.47)	0.19 (38.05)
8	0.19 (11.73)	0.17 (10.42)	0.22 (10.47)	0.18 (23.00)	0.16 (20.18)	0.21 (21.07)	0.17 (43.27)	0.16 (37.09)	0.20 (37.56)
9	0.22 (12.07)	0.19 (11.52)	0.22 (9.04)	0.20 (23.92)	0.17 (21.72)	0.21 (18.35)	0.19 (41.52)	0.16 (37.15)	0.21 (33.04)
10	0.23 (11.94)	0.18 (11.79)	0.26 (8.63)	0.23 (24.00)	0.16 (21.45)	0.24 (17.47)	0.21 (40.50)	0.15 (38.29)	0.25 (32.57)

Table 4.5: Return behaviors of stock decile portfolios based on basic and composite skew.

(c) Lower skew

Decile	One Week						One Month						One Quarter					
	Total		Structural		Informational		Total		Structural		Informational		Total		Structural		Informational	
1	0.22	(7.70)	0.28	(8.24)	0.16	(10.07)	0.22	(16.93)	0.27	(17.13)	0.16	(20.59)	0.22	(31.56)	0.28	(31.51)	0.14	(37.75)
2	0.18	(7.76)	0.24	(9.33)	0.15	(8.96)	0.19	(17.27)	0.23	(19.38)	0.15	(18.87)	0.18	(31.74)	0.22	(33.67)	0.15	(35.53)
3	0.18	(9.07)	0.23	(10.89)	0.16	(9.25)	0.19	(19.33)	0.21	(21.55)	0.16	(19.42)	0.17	(32.97)	0.21	(38.08)	0.16	(36.00)
4	0.18	(9.61)	0.20	(10.29)	0.16	(9.06)	0.17	(19.46)	0.19	(20.83)	0.16	(19.29)	0.16	(34.58)	0.19	(38.16)	0.16	(34.71)
5	0.18	(10.09)	0.19	(10.38)	0.19	(10.67)	0.17	(20.25)	0.17	(20.00)	0.17	(20.29)	0.17	(37.09)	0.17	(36.71)	0.16	(35.64)
6	0.17	(10.00)	0.17	(9.33)	0.19	(9.95)	0.16	(19.70)	0.16	(19.39)	0.17	(19.88)	0.16	(38.47)	0.15	(34.06)	0.17	(37.28)
7	0.18	(10.77)	0.16	(9.02)	0.22	(11.24)	0.17	(21.58)	0.16	(18.94)	0.19	(21.07)	0.16	(39.22)	0.16	(35.18)	0.19	(38.25)
8	0.18	(11.32)	0.17	(9.95)	0.20	(9.52)	0.17	(22.69)	0.17	(20.24)	0.21	(20.70)	0.17	(43.02)	0.16	(36.29)	0.20	(37.86)
9	0.22	(12.07)	0.16	(9.89)	0.23	(9.10)	0.20	(23.94)	0.15	(19.12)	0.22	(18.60)	0.19	(41.65)	0.14	(34.20)	0.21	(33.69)
10	0.26	(12.76)	0.16	(10.31)	0.30	(9.28)	0.23	(23.44)	0.15	(20.45)	0.28	(18.58)	0.22	(40.96)	0.15	(39.39)	0.27	(33.17)

(d) Upper skew

Decile	One Week						One Month						One Quarter					
	Total		Structural		Informational		Total		Structural		Informational		Total		Structural		Informational	
1	0.23	(7.90)	0.14	(10.22)	0.29	(8.81)	0.22	(17.23)	0.15	(21.59)	0.27	(17.77)	0.22	(31.70)	0.15	(39.88)	0.26	(32.24)
2	0.19	(8.13)	0.14	(8.64)	0.23	(9.37)	0.19	(17.30)	0.15	(18.58)	0.22	(18.89)	0.18	(31.20)	0.15	(36.41)	0.21	(34.14)
3	0.17	(8.69)	0.16	(9.39)	0.22	(10.11)	0.18	(19.21)	0.15	(19.31)	0.20	(20.64)	0.17	(34.02)	0.15	(36.60)	0.19	(34.91)
4	0.18	(9.92)	0.14	(8.48)	0.20	(10.42)	0.17	(19.50)	0.15	(19.09)	0.19	(21.33)	0.17	(35.40)	0.15	(35.08)	0.18	(36.53)
5	0.18	(10.10)	0.19	(10.75)	0.21	(11.52)	0.17	(20.31)	0.17	(19.91)	0.18	(21.56)	0.17	(36.55)	0.16	(34.82)	0.17	(37.54)
6	0.18	(10.77)	0.16	(8.95)	0.17	(10.12)	0.17	(20.63)	0.17	(19.45)	0.17	(20.75)	0.16	(37.83)	0.16	(34.59)	0.16	(37.27)
7	0.19	(11.62)	0.22	(11.19)	0.16	(9.57)	0.17	(22.24)	0.19	(21.10)	0.16	(20.06)	0.16	(39.64)	0.18	(37.36)	0.16	(37.29)
8	0.19	(11.73)	0.24	(11.30)	0.16	(9.89)	0.18	(23.00)	0.21	(21.15)	0.15	(19.83)	0.17	(43.27)	0.20	(37.80)	0.15	(37.76)
9	0.22	(12.07)	0.27	(10.30)	0.15	(9.80)	0.20	(23.92)	0.25	(21.00)	0.16	(20.92)	0.19	(41.52)	0.24	(36.73)	0.16	(39.24)
10	0.23	(11.94)	0.33	(9.90)	0.21	(12.00)	0.23	(24.00)	0.31	(19.58)	0.18	(22.58)	0.21	(40.50)	0.29	(33.58)	0.17	(39.40)

Table 4.6: Return behaviors of stock decile portfolios based on lower and upper skew.

the (Tian & Wu, 2020), where the opposite effect is exhibited. The behavior of portfolios based on the informational skew is the only one out of the three that is consistent with the (Tian & Wu, 2020) paper, as return figures tend to rise with the skew component deciles. Results are consistent across all three investment horizons.

Moreover, interesting conclusions can be inferred from the comparison of decile portfolio performance using advanced skew measures. The composite skew measure exhibits the same behavior across all nine combinations of skew components and investment horizons as the basic skew, with the portfolios formed on unprocessed skew metrics ('total' in the table) showing identical results to those of raw basic skew (Table 4.5.b). This result is expected and confirms that these two skew measures (basic and composite) are in many ways equivalent. What is more peculiar is the performance of portfolios based on lower and upper skew (Table 4.6). While separating the stocks based on the total skew leads to the same results as in the case of basic and composite skew – low- and high-decile portfolios outperform the middle-decile ones, the structural and informational skew components show different results. Effectively, the lower and upper skew measures mirror the behavior of one another – larger figures of structural lower and informational upper skew lead to better portfolio performance, while larger figures of informational lower and structural upper skew lead to worse portfolio performance. Once again, these results tend to be consistent across all three chosen investment horizons.

A clear advantage of separating the structural skew and informational skew from each other is evident from the decile portfolio analysis. One of the most common practical suggestions in such analysis is forming a high-low portfolio, which takes a long (short) position in the highest quantile portfolio of stocks and a short (long) position in the lowest quantile portfolio. Using the unprocessed, raw skew as a factor for sorting the portfolios to create a high-low investing strategy proves to be inefficient in this case, as the tail-end portfolios yield similar performances across all investment horizons.

4.4 Return factor portfolio behavior

Finally, Table 4.7 depicts the return factor portfolio behavior, comprised according to the approach introduced as early as (Fama & MacBeth, 1973), where future returns of stocks across various investment horizons (one week, one month, one quarter) are regressed against risk factors. Traditionally used factors include size, constructed as a logarithm of the market capitalization of a company, book-to-market ratio, as well as a measure of returns momentum. These factors, along with the six structural risk characteristics, which comprise the structural skew, as well as with informational skew, are included in the stock factor portfolio model.

Decile	**One Week**		**One Month**		**One Quarter**	
Size	-0.003	(-2.51)	-0.024	(-16.86)	-0.058	(-39.78)
Book-to-Market	-0.003	(-1.71)	-0.002	(-1.28)	-0.005	(-2.77)
Momentum	0.002	(1.08)	-0.000	(-0.21)	-0.002	(-1.20)
C_j	-0.019	(-8.29)	-0.016	(-6.78)	-0.013	(-5.32)
π_j	-0.005	(-2.50)	-0.020	(-10.60)	-0.045	(-24.53)
$\sigma_{ATM,j}$	0.175	(75.43)	0.121	(52.05)	0.095	(42.34)
C_{peers}	0.017	(6.46)	0.021	(7.83)	0.015	(5.54)
π_{peers}	-0.014	(-4.53)	-0.020	(-6.49)	-0.039	(-12.93)
$\sigma_{ATM,peers}$	-0.013	(-6.37)	-0.017	(-8.28)	-0.011	(-5.14)
Informational Basic Skew	0.124	(11.36)	0.114	(10.55)	0.088	(8.14)
Informational Comp. Skew	0.153	(12.37)	0.140	(11.45)	0.123	(10.07)
Informational Lower Skew	0.159	(12.93)	0.143	(11.85)	0.125	(10.39)
Informational Upper Skew	0.128	(10.10)	0.121	(9.72)	0.110	(8.77)

Table 4.7: Return behavior of stock factor portfolios.

Results regarding the traditional risk factors (size, book-to-market ratio, momentum) are not perfectly recreated. Smaller company size leads to larger returns, in accordance with previous studies; the effect becomes significantly more apparent as the investment horizon increases. A smaller book-to-market ratio, in this company sample, also leads to larger stock returns, while in the previous literature, the consensus is the opposite. While this result objectively contradicts the general consensus, it is worth noting that the results are not strongly conclusive in terms of statistical significance. Moreover, the results are reached using highly noisy daily data as opposed to weekly or even monthly data in most of the previous studies. Results related to the return momentum as a factor are statistically indifferent from zero.

In terms of relevant structural risk factors, the six identified risk components for the implied volatility skew exhibit different behavior in terms of stock return prediction, not all of which is consistent with (Tian & Wu, 2020). Portfolios based on the business cyclicality measure yield slightly negative, albeit statistically significant returns, whereas in the reference paper results regarding this factor are inconclusive. Higher default risk yields negative returns, which is consistent with the previous literature. However, the most remarkable distinction lies in the performance of the implied volatility level factor portfolio, which yields the highest positive returns across six risk factors composing the structural skew. Moreover, these returns noticeably increase as the investment horizon gets shorter. In (Tian & Wu, 2020), the returns of this portfolio are slightly negative, although the results are statistically insignificant in each of the three investment horizons.

Peers-related factors yield significant results, some of which intuitively do not make perfect sense. While the cyclicality of the stock itself yields negative returns, as discussed earlier, the peers' cyclicality yields positive returns of similar magnitude. This result is particularly peculiar, as it is the only one out of the three characteristics that exhibits the directly opposite behavior when comparing the respective stock-specific and industry-specific factors. The peer-average default risk factor is consistent with the stock-related default risk, yielding increasingly negative returns as the investment horizon gets longer – this result is consistent with the expectations that the poorer the economic conditions of an industry are, the smaller the returns will be. Finally, the peer-average implied volatility level risk factor leads to negative returns of much smaller magnitude compared to the stock implied volatility factor, a result that again is counterintuitive.

The results regarding the informational skew are consistent with the ones presented in (Tian & Wu, 2020): portfolios based on the informational skew factor yield the highest returns across all factors, with the results being consistently significant across all three investment horizons. Returns slightly decrease as the investment horizon increases, which is also the case in (Tian & Wu, 2020). That result applies to all four skew factors constructed from regression residuals, with all three advanced skew characteristics performing better than the basic skew measure. Once again, the value in separating the lower and upper skew components is underlined, as the lower skew displays better performance compared to the composite skew.

4.5 Discussion, next steps

Overall, these are the main inference points that could be made based on the results of this research:

- Most of the findings are consistent with the ones described in (Tian & Wu, 2020). Results related to the cross-sectional regression model are virtually identical, from the target-predictor correlations to the time-varying behaviors.
- Overall, the introduction of advanced skew characteristics proved to be beneficial. While the advanced skew metrics provide exhibit higher explainability in the cross-sectional regressions, this does not always directly translate into more apparent portfolio performance distinctions. Decile portfolio performance based on basic and composite skew measures is highly similar. On the other hand, the returns of stock factor portfolios are higher for informational composite skew across all investment horizons. The highest results are exhibited by the lower skew factor portfolio, which suggests that the mispricings in the moneyness region of over 100% are comparatively less relevant.
- Results observed in the decile portfolio composition give several conclusions that con-

tradict the reference paper in some instances; mainly, an apparent monotone increase of returns is not present in the portfolios based on total and structural skew. Instead, a parabolic-shaped performance is exhibited when the total skew is looked at, which would suggest that a simple high-low decile portfolio strategy might not produce the best investment performance. It is particularly interesting that the structural and the informational skew performance dynamics in the decile portfolios tend to mirror each other, underlining the importance of separating the two - using either one as an investment factor rather than the actual skew would yield better results.

- The sample average return figures exhibit higher statistical significance than the ones reported in (Tian & Wu, 2020). This can be mainly attributed to higher data frequency (daily, as opposed to weekly), which leads to lower standard deviation of the cross-sectional results, which, in turn, leads to higher reported t-statistics.

The completed work can be expanded in a variety of ways, building upon the created framework. Firstly, a cross-temporal comparison has not been carried out in this paper, i.e. a comparative review of the results using options data with different maturities was not performed. This was decided to be out of the scope of the thesis, however, it would be a logical continuation of research, given that similar analysis is usually conducted in the literature (an excellent example of such work would be (Liu et al., 2019)), and that the use of created peer-related characteristics would introduce additional new findings in the current state of discourse.

Similarly, the studied stock sample could be broadened and diversified, which would also be an extension that does not disrupt the created framework. This could most likely be adequately achieved by the use of additional data sources, as the threshold for the included companies was that all used company-specific data is available on WRDS for the periods when the companies were S&P 500 constituents. This criterion proved to be somewhat restrictive, as it led to a degree of survivorship bias exhibited in the performance results. In the same vein, the peer-related features could have been deduced from the industry-specific ETF data rather than from the sample averages, but doing that was deemed not feasible in this work as the industry-specific ETF historical data is not available for free among the inspected data vendors.

Beyond that, skew can be further dissected into its predicted and residual components through the introduction of more controlling features. For instance, (Bali & Hovakimian, 2009) introduced market microstructure metrics related to illiquidity, bid-ask spreads, analyst forecast dispersion, trading based on private information and others. Newer studies, such as (Han et al., 2020), used additional firm-related fundamental financial metrics and ratios (among them, R&D over market capitalization ratio, cash holdings ratio, and others), as well as option-related metrics based on implied volatilities and trade volumes. The introduction of extra controlling features would also allow to reduce the informational skew into a less opaque characteristic. Despite these potential areas of enhancement, the comparably narrow introduction of the controlling factors in this thesis could nevertheless be deemed successful, as the practical goals were achieved – created investment factors (primarily, informational skew metrics) proved to be reliable in terms of portfolio performance both in decile portfolio analysis and in factor portfolio analysis.

5 | Conclusion

This research thesis aims to gain insight into the underlying dependencies between the equity markets and options markets. The extensively studied option-implied volatility skew displays changes in the structural risk factors related to the stock as well as the short-term informational flow coming from the equity markets. The choice of the structural factors followed that of (Tian & Wu, 2020), as a measure of business cyclicality, default risk, and at-the-money level of implied volatility were selected. A slight shortcoming of previous literature, however, lies in the overly robust approach to the comprehension of the industry-specific data, which, much like the stock-specific data, carries structural risk information. Hence, an unconventional inclusion of peer-related features was completed in this thesis.

The relative explanatory power of these structural factors and informational flow was explored using a cross-sectional model setup. While the prediction accuracy levels are highly unstable over time, a persistent link between the financial crises and the skew parameters was established. The inclusion of previously underutilized industry-specific information proved to be beneficial, as it contributed to the further dissection of the skew. Improved accuracy figures in comparison to the (Tian & Wu, 2020) paper suggest that the short-term informational skew was singled out to a further extent, leaving the impact of industry-wide structural risks out of the picture. Hence, it can be concluded that the approach taken to extract information from the industry-related features is reliable.

Moreover, two different approaches were taken to the computation of skew measures, for both of which the findings proved to be consistent with each other. The computationally extensive advanced skew metrics performed exhibited significantly higher explainability in the cross-sectional regression setup, with the regression accuracy figures nearing 80% at peak. They were also revealed to be favorable in the context of portfolio performance within both utilized frameworks, as decile portfolio performance of the lower and upper skews separately yielded a monotonous return increase. All three advanced skew metrics also outperformed the basic skew in the stock factor portfolio context, as the lower skew produced the best performance out of the three.

Nevertheless, the work could be further expanded through the use of a differentiated set of stocks, options with varying periods to maturity, as well additional information regarding prices of industry-specific ETFs and their respective options. Another potential way to explore would be further dissection of the structural component of the skew in line with previous literature. The current scope of the work was limited due to data limitations and focus constraints. However, that being said, the utilized framework would not need to be enhanced in a significant way in order to accommodate that research.

Bibliography

Amaya, D., Christoffersen, P., Jacobs, K., & Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, *118*(1), 135-167.

Bali, T. G., Engle, R. F., & Murray, S. (2016). *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons.

Bali, T. G., & Hovakimian, A. (2009). Volatility spreads and expected stock returns. *Management Science*, *55*(11), 1797-1812.

Bates, D. S. (1991). The crash of '87: Was it expected? the evidence from options markets. *The Journal of Finance*, *46*(3), 1009-1044.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options. *The Review of Financial Studies*, *9*(1), 69-107.

Bernales, A., & Guidolin, M. (2014). Can we forecast the implied volatility surface dynamics of equity options? predictability and economic value tests. *Journal of Banking & Finance*, *46*, 326-342.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637-654.

Bongaerts, D., De Jong, F., & Driessen, J. (2011). Derivative pricing with liquidity risk: Theory and evidence from the credit default swap market. *The Journal of Finance*, *66*(1), 203-240.

Bouri, E., Lucey, B., & Roubaud, D. (2020). Dynamics and determinants of spillovers across the option-implied volatilities of US equities. *The Quarterly Review of Economics and Finance*, *75*, 257-264.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, *63*(6), 2899-2939.

Cao, C., Zhong, Z., & Yu, F. (2009). The information content of option-implied volatility for credit default swap valuation. *FDIC Center for Financial Research, Working Paper*(2007-08).

Carr, P., & Wu, L. (2007). Theory and evidence on the dynamic interactions between sovereign credit default swaps and currency options. *Journal of Banking & Finance*, *31*(8), 2383-2403.

Carr, P., & Wu, L. (2010). Stock options and credit default swaps: A joint framework for valuation and estimation. *Journal of Financial Econometrics*, *8*(4), 409-449.

Chang, B. Y., Christoffersen, P., & Jacobs, K. (2013). Market skewness risk and the cross section of stock returns. *Journal of Financial Economics*, *107*(1), 46-68.

Chen, C. Y. H. (2014). Does fear spill over? *Asia-Pacific Journal of Financial Studies*, *43*(4), 465-491.

Chen, J., Han, Q., Ryu, D., & Tang, J. (2022). Does the world smile together? a network analysis of global index option implied volatilities. *Journal of International Financial Markets, Institutions and Money*, *77*, 101497.

Christoffersen, P., & Jacobs, K. (2004). The importance of the loss function in option valuation. *Journal of Financial Economics*, *72*(2), 291-318.

Cremers, M., Driessen, J., Maenhout, P., & Weinbaum, D. (2008). Individual stock-option prices and credit spreads. *Journal of Banking & Finance*, *32*(12), 2706-2715.

Cremers, M., & Weinbaum, D. (2010). Deviations from put-call parity and stock return predictability. *Journal of Financial and Quantitative Analysis*, *45*(2), 335-367.

Derman, E., & Kani, I. (1994). Riding on a smile. *Risk*, *7*(2), 32-39.

Doran, J. S., Peterson, D. R., & Tarrant, B. C. (2007). Is there information in the volatility skew? *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *27*(10), 921-959.

Dupire, B. (1994). Pricing with a smile. *Risk*, *7*(1), 18-20.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427-465.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607-636.

Foresi, S., & Wu, L. (2005). Crash–o–phobia: a domestic fear or a worldwide concern? *The Journal of Derivatives*, *13*(2), 8-21.

Füss, R., Mager, F., Wohlenberg, H., & Zhao, L. (2011). The impact of macroeconomic announcements on implied volatility. *Applied Financial Economics*, *21*(21), 1571-1580.

Han, Y., Liu, F., & Tang, X. (2020). The information content of the implied volatility surface: Can option prices predict jumps? *Available at SSRN*(3454330).

Hann, R. N., Kim, H., & Zheng, Y. (2019). Intra-industry information transfers: evidence from changes in implied volatility around earnings announcements. *Review of Accounting Studies*, *24*, 927-971.

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*(2), 327-343.

Hull, J., Nelken, I., & White, A. (2004). Merton's model, credit risk, and volatility skews. *Journal of Credit Risk Volume*, *1*(1), 05.

Jiang, G. J., Konstantinidi, E., & Skiadopoulos, G. (2012). Volatility spillovers and the effect of news announcements. *Journal of Banking & Finance*, *36*(8), 2260-2273.

Kelly, B., Pástor, Ĺ., & Veronesi, P. (2016). The price of political uncertainty: Theory and evidence from the option market. *The Journal of Finance*, *71*(5), 2417-2480.

Kenourgios, D. (2014). On financial contagion and implied market volatility. *International Review of Financial Analysis*, *34*, 21-30.

Krause, T. A., & Lien, D. (2014). Implied volatility dynamics among exchange-traded funds and their largest component stocks. *The Journal of Derivatives*, *22*(1), 7-26.

Liu, D., Qiu, Q., Hughen, J. C., & Lung, P. (2019). Price discovery in the price disagreement between equity and option markets: Evidence from SSE ETF50 options of China. *International Review of Economics & Finance*, *64*, 557-571.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The*

Journal of Finance, *29*(2), 449-470.

Mixon, S. (2002). Factors explaining movements in the implied volatility surface. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *22*(10), 915-937.

Mixon, S. (2011). What does implied volatility skew measure? *Journal of Derivatives*, *18*(4), 9.

Muravyev, D., Pearson, N. D., & Broussard, J. P. (2013). Is there price discovery in equity options? *Journal of Financial Economics*, *107*(2), 259-283.

Robe, M. A., & Wallen, J. (2016). Fundamentals, derivatives market information and oil price volatility. *Journal of Futures Markets*, *36*(4), 317-344.

Rubinstein, M. (1994). Implied binomial trees. *The Journal of Finance*, *49*(3), 771-818.

Schneider, P., & Trojani, F. (2014). *Divergence and the price of risk.* (Working paper, University of Lugano and SFI)

Schneider, P., Wagner, C., & Zechner, J. (2020). Low-risk anomalies? *The Journal of Finance*, *75*(5), 2673-2718.

Tian, M., & Wu, L. (2020). Cross-sectional variation of option implied volatility skew. (Available at SSRN 3707006)

Yan, S. (2011). Jump risk, stock returns, and slope of implied volatility smile. *Journal of Financial Economics*, *99*(1), 216-233.

A | Market pricing of risk characteristics

Variable	Moments			Percentiles						
	Mean	StDev	Auto	1	10	25	50	75	90	99
β_{cycle}	-0.078	0.107	0.603	-0.420	-0.203	-0.135	-0.071	-0.014	0.039	0.147
$\beta_{default}$	-0.027	0.094	0.694	-0.373	-0.132	-0.050	-0.006	0.008	0.046	0.182
β_{ATM}	-0.046	0.072	0.759	-0.246	-0.127	-0.085	-0.046	-0.003	0.031	0.134
$\beta_{peers,cycle}$	-0.025	0.211	0.695	-0.544	-0.267	-0.150	-0.036	0.089	0.248	0.539
$\beta_{peers,default}$	-0.097	0.303	0.727	-1.190	-0.438	-0.196	-0.034	0.042	0.156	0.593
$\beta_{peers,ATM}$	-0.017	0.081	0.602	-0.210	-0.091	-0.046	-0.004	0.039	0.096	0.213
R^2	7.324	5.798	0.829	0.681	1.859	3.228	5.594	9.647	15.719	27.362

(a) Basic skew

Variable	Moments			Percentiles						
	Mean	StDev	Auto	1	10	25	50	75	90	99
β_{cycle}	0.007	0.047	0.742	-0.154	-0.030	-0.009	0.005	0.022	0.046	0.165
$\beta_{default}$	-0.031	0.144	0.961	-0.742	-0.055	-0.014	-0.001	0.003	0.010	0.060
β_{ATM}	-0.066	0.056	0.878	-0.292	-0.124	-0.083	-0.052	-0.032	-0.021	-0.006
$\beta_{peers,cycle}$	0.024	0.083	0.740	-0.133	-0.043	-0.012	0.011	0.041	0.097	0.346
$\beta_{peers,default}$	-0.024	0.105	0.827	-0.343	-0.113	-0.047	-0.008	0.008	0.034	0.287
$\beta_{peers,ATM}$	0.021	0.049	0.783	-0.120	-0.010	0.003	0.014	0.035	0.066	0.170
R^2	30.151	18.676	0.883	1.642	5.967	12.970	28.989	46.271	55.878	68.106

(b) Composite skew

Variable	Moments			Percentiles						
	Mean	StDev	Auto	1	10	25	50	75	90	99
β_{cycle}	0.012	0.054	0.808	-0.197	-0.027	-0.006	0.010	0.031	0.060	0.180
$\beta_{default}$	-0.039	0.159	0.966	-0.837	-0.079	-0.020	-0.002	0.003	0.009	0.050
β_{ATM}	-0.094	0.067	0.894	-0.338	-0.166	-0.117	-0.077	-0.053	-0.038	-0.021
$\beta_{peers,cycle}$	0.028	0.087	0.764	-0.135	-0.044	-0.010	0.017	0.049	0.107	0.368
$\beta_{peers,default}$	-0.029	0.111	0.850	-0.380	-0.124	-0.056	-0.010	0.009	0.036	0.290
$\beta_{peers,ATM}$	0.029	0.054	0.823	-0.127	-0.007	0.006	0.020	0.048	0.083	0.189
R^2	40.167	20.892	0.892	3.432	10.423	21.311	41.640	58.930	66.748	75.128

(c) Lower skew

Variable	Moments			Percentiles						
	Mean	StDev	Auto	1	10	25	50	75	90	99
β_{cycle}	-0.005	0.015	0.799	-0.042	-0.020	-0.011	-0.005	-0.001	0.005	0.056
$\beta_{default}$	0.008	0.024	0.927	-0.021	-0.004	-0.001	0.000	0.009	0.028	0.115
β_{ATM}	0.028	0.016	0.850	0.006	0.011	0.017	0.025	0.035	0.049	0.077
$\beta_{peers,cycle}$	-0.005	0.017	0.830	-0.062	-0.023	-0.013	-0.004	0.004	0.013	0.042
$\beta_{peers,default}$	0.005	0.030	0.875	-0.087	-0.018	-0.005	0.002	0.012	0.034	0.111
$\beta_{peers,ATM}$	-0.008	0.012	0.844	-0.052	-0.021	-0.013	-0.006	-0.001	0.003	0.016
R^2	42.380	15.356	0.849	8.529	19.743	31.596	43.695	54.398	61.767	70.041

(d) Upper skew

Table A.1: Market pricing of company risk characteristics on the implied volatility skew.

B | Replication Code

As mentioned in the introduction, this thesis provides the code written in Python for the reader, available at the `https://github.com/maborovkov/MSc-Thesis-3146466`. This appendix includes two Jupyter Notebook files:

- **Full Pipeline:** a notebook, which contains the entire pipeline of the analysis with the exception of the advanced skew metrics calculation. The following is a description of crucial parts of that document, with *functions* outlined that can be used to yield some of the results that were displayed in this thesis:
  - Options, Stock, and Structural Data Retrieval, Data Merge: retrieval and subsequent recomposition of the features and skew data as described in the main body of the thesis. Does not include the creation of the advanced skew metrics, which is done in the other attached notebook file. This section requires an account on the Wharton Research Data Services (WRDS) website for connection to the WRDS databases.
  - Winsorization, Summary Statistics: yields Tables 4.1 and 4.2 (function $calc\_corrs$ for the latter).
  - Cross-Sectional Regression: yields Table 4.3 and A.1. The function $plot\_skew$ is used to produce Figure 4.1.
  - Summary Time Series Statistics: yields Table 4.4 (function $ts\_summary$).
  - Decile Portfolio Composition: yields Tables 4.5 and 4.6.
    The function $calc\_portfolios$ is used to produce a column in either of these tables.
  - Factor Model: yields Table 4.7 (function $calc\_factor\_portfolios$).
- **Advanced Skew Metrics Calculation:** a notebook, which contains the retrieval of option-related data for the calculation of the advanced skew metrics described in Chapter 3.2.

Like the data retrieval process in the Full Pipeline notebook, this process requires a connection to the WRDS databases.

It should be noted that the replication of the entire analysis would take several hours, with the data retrieval and decile portfolio composition portion taking up the most time. The efficiency of the execution was not made a priority throughout the project, however, the code is meant to be comprehended with the description of the used techniques outlined in the main body of the thesis.